## ***Comments to the Author:***

Dear authors,

On my previous decision (9 May), I have pointed the following major issues with your manuscript:

"The reviewers have generally positively evaluated your manuscript but also identified pertinent shortcomings, including (a) insufficient description of the model set-up, particularly regarding initialization (b) insufficient discussion and conclusions (c) reference list is missing many references and (d) language quality need to be strongly enhanced to make the content understandable. Other relevant points can be found in the reviewers comments."

You have re-submitted the manuscript on 10 May (!!!, one day later!!!) and you have not addressed properly at least three of the indicated four points, namely (a), (b), and (d). I am thus giving you a second chance to revise your manuscript in a way that properly addressed the raised concerns.

Please let me know if you would like to re-revise the manuscript. Please mark the additional changes to this version separately.

best regards
Joaquim Pinto
(handling editor)

## ***Response to the editor:***

Dear Dr. Joaquim Pinto,

Thank you for giving us the opportunity to revise the manuscript.

We are sorry that we did not properly address the major issues you mentioned last time. Actually, we began to revise our manuscript point by point according to the reviewers' comments at the end of March (after reviewers uploaded their comments). Therefore, we could quickly submit our revised manuscript after you gave us a positive comment. It is unfortunate that "at least three of the indicated four points" were not addressed properly in the previous manuscript.

This time, we rewrote Section 3.2 about model setup (including initialization) and have made it clearer. In addition, regarding the insufficient discussion and conclusions, we have provided more content in the manuscript, especially in Section 4. In Section 4, we expanded our discussion and rewrote Section 4.3; Table 5 has also been added to clarify our opinions. The manuscript has been edited for language by a professional editing service, and the edit certificate is attached below. We have also checked other points and answered the questions point by point again the reviewers mentioned previously. All amendments are shown with tracked changes in the revised manuscript (below).

Addressing the identified issues truly helped us improve the manuscript, which we hope is now suitable for publication.

We are looking forward to your decision.

Best regards
Dayang Li, Zhongmin Liang, Yan Zhou, Binquan Li, Yupeng Fu

# EDITORIAL CERTIFICATE

This document certifies that the manuscript listed below was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at American Journal Experts.

## Manuscript title:

Multicriteria assessment framework of flood events simulated with the vertically mixed runoff model in semiarid catchments in the middle Yellow River

## Authors:

Dayang Li, Zhongmin Liang, Yan Zhou, Binquan Li, Yupeng Fu

## Date Issued:

June 27, 2019

## Certificate Verification Key:

5D57-6146-3608-39ED-402P

This certificate may be verified at www.aje.com/certificate. This document certifies that the manuscript listed above was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at American Journal Experts. Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes. To verify the final AJE edited version, please visit our verification page. If you have any questions or concerns about this edited document, please contact American Journal Experts at support@aje.com.

## *Response to RC#1:*

**General Comments:**

The paper presents a multicriterion assessment framework for flood events forecasting or warning in semi-arid regions. Four hydrologic models have been used in catchments of the middle Yellow River. The result shows that the VMM model has a better performance of flood modeling than the other three models. The authors believe that flood events in semi-arid and arid regions should have different criteria than that of in humid areas to determine whether a flood forecasting and early warning is acceptable.

The topic of this study is very interesting and the idea is more or less novel. The paper is well-structured. I recommend the publication of this paper with a minor revision. Some specific comments are listed as below.

Answer: Thank you for the positive comments on our manuscript.

**Specific comments:**

(1) The proposed framework has three parts, C1, C2 and C3. C2 is the key part of this framework, and three flow zone (low flow zone, medium flow zone, high flow zone) are divided. I think this simple framework is more important in terms of flood early warning rather than just a performance assessment. This may be real value of the framework but it is not clearly expressed in the manuscript. More explanation and discussion should be added in the paper.

Answer: Thank you for this suggestion. We have almost rewritten Section 4 (Results and discussion) and added more explanations and discussions in the revised manuscript.

(2) The initial condition is very important for a hydrologic model. In this paper, it is reasonable that the daily based model is used to calculate the initial conditions of the event-based model, but the initial condition of daily based model is not mentioned. Please add some explanations.

Answer: Thank you for this good suggestion. The model was run continuously during the period of 1983-2009. The simulation was daily between flood events and hourly within flood events. Hence, there is no need to set initial values for each event. The revised paragraph is as follows:

"The VMM model was run continuously from 1983 to 2009 for each catchment. … Two initial values, the initial tension water storage (W0) and the initial free water storage (S0), were used to describe the initial catchment moisture condition. The initial values are smaller for drier catchments, and the minimum values are zero. In this study, the initial values were assumed to be zero uniformly due to the dry conditions at 00:00:00 on January 1, 1983 for each catchment. It should be noted that continuous simulations for each catchment eliminate the need to set the initial values for each flood event in a catchment."

(3) Conclusion (2): "In the four catchments, by PAWN analysis of VMM, CS, IM, and KE are the most sensitive parameters and are not affected by the choice of objective functions, whereas WM is the most sensitive parameter" make me confusing. Please use clearer and more concise expressions.

Answer: We have rewritten the sentence for clarity.

"In the four catchments, the parameters confluence coefficient of surface flow (CS), impermeable area (IM), and residence time of Muskingum (KE) in VMM model are the most sensitive based on an analysis by the global sensitivity method PAWN; in addition, the sensitivity ranking of the parameter WM related with the soil moisture capacity is the most affected by the objective functions.”

(4) Reference/citation style needs to be revised. For example, a space is missing between Lu and et on p5, line 26; parenthesis is not right on p7, line 4, (Pianosi and Wagener,2015).
Answer: We have corrected this information.

(5) P3, line13:” Streamflow and rainfall data are from 1983 to 2009. Hourly streamflow data came from hydrological stations. Nine. . .”, in this sentence, English tenses should be consistent.
Answer: Thank you for this good suggestion. We have changed this text.

(6) P3, line 26: the runoff is conceptualized as being composed of surface runoff and groundwater flow (notoriously but erroneously called “below-ground off” in the paper).
Answer: Thank you for this good suggestion. We have corrected this text.

(7) Figure 1 is requested to be further processed. The symbols of rain gauge station and hydrological station are not very clear.
Answer: We have improved the figure for clarity.

(8) Figure 3: the y-axis label may be “absolute relative error of peak flow (%)” instead of “peak flow (%)”. The title “Figure 3: Boxplot of peak flows . . .” should be also checked.
Answer: We have corrected the label and checked the figure caption.

(9) Although I am not a native English reviewer, I find some sentences difficult to understand. The authors are encouraged to further polish up the language.
Answer: Thank you for this good suggestion. The manuscript has been polished by a professional service. All amendments are shown with tracked changes in the attached pdf file.

## *Response to RC#2:*

I think you performed a lot work whose results deserve to be published. The topic addressed in your manuscript is interesting and I think that not only statistical hydrologists will benefit from its publication, either from the decision-makers point of view as well as from the perspective of semiarid catchments. However, I noticed several instances in your manuscript that force me not to accept it for publication in its current state. First, I would like to emphasize that all my detailed comments are included in the attached PDF file. Please, kindly see this attached file at the same scale, so you could find the places which my comments point to properly.
The paper is well structured, but a large number of style corrections is required.
Answer: We thank the reviewer for the positive comments on our manuscript. Detailed style corrections are attached below.

**General Comments:**

The discussion of the results is not clear and, in my opinion, could be extended due to the huge effort made in the study. The authors are encouraged to extend the discussion of the results obtained by the explanation of possible causes of differences among hydrological models.
Answer: Thank you for these constructive suggestions. We have added more explanation for clarity.

The initial condition is essential for this type of simulations. Thus, the authors should explain better this section.
Answer: Thank you for this good suggestion. We have rewritten Section3.2 and added more detailed information as follows:
"The VMM model was run continuously from 1983 to 2009 for each catchment. ... Two initial values, the initial tension water storage (W0) and the initial free water storage (S0), were used to describe the initial catchment moisture condition. The initial values are smaller for drier catchments, and the minimum values are zero. In this study, the initial values were assumed to be zero uniformly due to the dry conditions at 00:00:00 on January 1, 1983 for each catchment. It should be noted that continuous simulations for each catchment eliminate the need to set the initial values for each flood event in a catchment."

Regarding the references shown in the manuscript, some are not listed in the reference list, and also there are some errors in the reference list. Please, check the standards of the journal and correct them.
Answer: We have checked all references and corrected the errors.

I strongly recommend to undertake some check of the language (e.g. by some professional service offered directly by Springer on you websites). Without a doubt, there are some places in the text that are hardly understandable due to the language.
Answer: Thank you for this good suggestion. The manuscript has been edited for language by a professional editing service. All amendments are shown with track changes in the attached pdf file

I believe my comments will help you improve your MS, which will lead to is final publication.

Answer: Thank you. Your comments greatly helped us improve the manuscript, which we hope is now suitable for publication.

**Specific comments:**

**Page 1**

Line 10-12: It may be rewritten, although the sentence is understood, I think it is not the best way. Try to avoid double parentheses. Besides, the manuscript is focused on VMM, so this model should be first. E.g.: 'The … (VMM) is compared with three models, one physical based model, the MIKE SHE, and two conceptual models, the XAJ and the Shanbei' (this is just a suggestion, feel free to write your way).

Answer: We accept the reviewer's suggestion.

Change: The vertically mixed runoff model (VMM) is compared with three models, one physical-based model, MIKE SHE model (originated from the Système Hydrologique Européen program), and two conceptual models, the Xinanjiang model (XAJ) and Shanbei model (SBM).

Line 16: "its figure is only 41% in four semiarid catchments", could you explain better this sentence？

Answer: We apologize for our unclear expression, which has been revised.

Change: Our results show that the VMM model has a better flood estimation performance than the other models, and the FCRA framework can provide reasonable flood classification and reliability assessment information, which may help decision makers improve their diagnostic abilities in the early flood warning process.

Line 24: "and", but

Answer: This change has been made.

Change: but

**Page 2**

Line 9-10: Maybe it is better "to be used on a large-scale semiarid area".

Answer: Thank you for this good suggestion. We have changed this phrasing.

Change: The radar costs are too high to be used on a large-scale semiarid area.

Line 14: "severe", what is the meaning of severe here? It should be better explained.

Answer: "Severe rainstorms" means the rainstorms have produced the largest peak flows according to Michaud and Sorooshian (1994). We have added more explanation.

Change: Michaud and Sorooshian (1994) used 24 severe rainstorms that produced the largest peak flows during 1957-1977 to compare three hydrologic models, i.e., the lumped SCS model, simple distributed SCS model and distributed KINEROS model, in the Walnut Gulch catchment.

Line 20-21: "Therefore, it is urgent to search for useful information based on the limited accuracy of modeling results to serve as flood warnings and to improve decision making.". I do not understand this sentence. Please, rewrite the sentence.

Answer: We apologize for the unclear phrasing. We have rewritten the sentence.

Change: Therefore, determining how to use modeling results with limited accuracy to provide guidance for flood early warning is important.

Line 24-25: “Four hydrologic models (the vertically mixed runoff model (VMM), MIKE SHE, Xinanjiang Model (XAJ) and Shanbei model (SBM)) ...”. To much parenthesis; Try to make a continuous sentence: "Four hydrological models: the vertical …, the MIKE SHE model, …, are compared ...

Answer: Thank you for this suggestion. We have rewritten the sentence.

Change: Four hydrological models: the vertically mixed runoff model (VMM), the MIKE SHE model, Xinanjiang model (XAJ) and Shanbei model (SBM), are compared on the....

Line 27: “remainder”. Please, use another word.

Answer: We have rewritten the sentence.

Change: The rest of the paper is organized as follows.

Line 27: “section below”. Specify to which section corresponds exactly, so it is necessary to use capital letters: e.g. "The Methodology Section", or use instead of the name, just the number of the section "Section 3 describes…".

Answer: This is a good suggestion. We have changed this phrasing.

Change: Section 2 describes the study area and the data set used.

Line 28-31: “The VMM model...of the study.”. Please rewrite the whole paragraph, it is difficult to follow it.

Answer: We have rewritten the paragraph to clarify our intent.

Change: Section 3 presents the VMM model methodology, model set, model calibration and validation, multicriteria assessment framework and parameter sensitivity analysis. Section 4 describes the results and discussion of model comparison, sensitivity analysis and analysis of the multicriteria assessment framework for the VMM model. The final section presents the conclusions of the study.

**Page 3**

Line 1: Is there any reference for all these data? I mean the temperature, rainfall and seasonality of the rainfall data.

Answer: The temperature data are obtained from previous papers, and we have added references. In addition, all rainfall data are collected from rain gauging stations, and we have added an explanation.

Change: This change has been made.

Line 3: “is”, between.

Answer: This change has been made.

Change: between.

Line 4:” is”, between.

Answer: This change has been made.

Change: between.

Line4: “65 – 80%”, more less the 65 to 80%.
Answer: This change has been made.
Change: more less the 65 to 80%.

Line 5: “is”, between.
Answer: This change has been made.
Change: between

Line 8: “soil erosion”, how much?
Answer: Thank you for this good question. According to Li et al. (2019), the average sediment concentration reaches 126 kg $m^{-3}$ in these regions. We have added this information to the text.
Change: The lack of vegetation in these catchments leads to serious soil erosion, and the average sediment concentration reaches 126 kg $m^{-3}$ according to Li et al. (2019).

Line 9: “few attempts have been applied to model hourly flood flows”. Please rewrite the sentence. It has poor writing.
Answer: We apologize for the poor phrasing. We have rewritten this sentence.
Change: Only a few studies have modeled hourly floods.

Line 11: “Hence, modeling floods and providing a useful method for decision makers in charge of flood defense are essential and urgent.”. Please rewrite the whole sentence, I cannot follow it.
Answer: We apologize for the poor phrasing. We have rewritten this sentence.
Change: Hence, it is important for decision-makers to know how to evaluate the flood risk when a flood is approaching.

Line 12: “Streamflow and rainfall data are from 1983 to 2009.” Please rewrite the sentence. It has poor writing.
Answer: We apologize for the poor phrasing. We have rewritten this sentence.
Change: The period used in the modeling is from 1983 to 2009. Streamflow and rainfall data are collected from streamflow gauging stations and rain gauging stations at an hourly time-step.

Line 12: “stations”, which type of stations?
Answer: They are streamflow gauging stations. We have corrected this phrasing.
Change: Streamflow and rainfall data are collected from streamflow gauging stations and rain gauging stations at an hourly time-step, respectively.

Line 14-15: Thiessen polygon method was...
Answer: This change has been made.
Change: Thiessen polygon method was used to interpolate the rainfall data for each catchment.

Line 15: “...interpolate the rainfall data.”. For each basin? I guess.
Answer: We agree with the reviewer. We have added this information.

Change: Thiessen polygon method was used to interpolate the rainfall data for each catchment.

Line 17: Here you should characterize the model in its whole, I mean, explain here also that it is lumped, continuous or event based, and so on.
Answer: Thank you for this suggestion. We have added this information.
Change: The VMM is a lumped continuous hydrologic model developed by Bao and Wang (1997).

Line 18: ”...conceptual hydrologic model...”, is continuous or event based?
Answer: It a continuous model; we have added this information.
Change: The VMM is a lumped continuous hydrologic model and has been used in many areas in China, especially in semiarid and subhumid catchments.

Line 21: “etc”, better “and others”.
Answer: Thank you for raising this good point. We have corrected this phrasing.
Change: This change has been made.

**Page 4**
Line 24: “in”, at.
Answer: This change has been made.
Change: at.

Line 25-27: “Hence, tension water storage *W*... on daily values of *W* and *S”*. Please, explain it better, because I do not understand what type of variables you used in the case you just simulate a flood event.
Answer: We are apologized for our unclear phrasing. We have rewritten this paragraph.
Change: The VMM model was run continuously from 1983 to 2009 for each catchment. ... Two initial values, the initial tension water storage (W0) and the initial free water storage (S0), were used to describe the initial catchment moisture condition. The initial values are smaller for drier catchments, and the minimum values are zero. In this study, the initial values were assumed to be zero uniformly due to the dry conditions at 00:00:00 on January 1, 1983 for each catchment. It should be noted that continuous simulations for each catchment eliminate the need to set the initial values for each flood event in a catchment.

**Page 5**
Line 1: “Because...”. The sentence has poor writing. Maybe you it is better to write "Due to the fact that only … ". But I recommend rewriting the sentence.
Answer: Thank you for this good suggestion. We have rewritten the sentence.
Change: Due to the fact that only one streamflow gauging station is available for each catchment, the spatial variation in each catchment’s parameters cannot be determined by calibration.

Line 1: “gauge”, gauging.
Answer: This change has been made.
Change: gauging

Line 4: “Li et al., 2018”, 2018a or 2018b?

Answer: We have checked the reference.

Change: Li et al., 2018

Line 8:” ...full fitness...”. I guess you mean "fullfillnes", but I think you should better rewrite the whole sentence in order to get a continuity in the writing.

Answer: Thank you for raising this good point. We have rewritten this sentence.

Change: however, it may not be suitable for semiarid catchments because a good fit is not required between the simulated and observed streamflows.

Line 9: “(McIntyre and Al-Qurashi, 2009: SHARMA and MURTHY, 1998)”. Remove parenthesis, write it properly. "McIntyre and Al-Qurashi (2008) and Sharma and Murphy (1998) used …".

Answer: This change has been made.

Change: McIntyre and Al-Qurashi (2009) and Sharma and Murphy (1998) used...

Line 12: “$Q_{p'}^{i}$”. It is difficult to distinguish the apostrophe. Please, try another mark in order to make out the different variables

Answer: Thank you for this suggestion. We have changed this phrasing.

Change: $Q_{pm}^{i}$.

Line 13: The same as previous.

Answer: Thank you for this suggestion. We have changed this phrasing.

Change: $Q_{vm}^{i}$.

Line 15: Please, try another mark in order to make out the different variables.

Answer: This change has been made.

Change；$Q_{pm}$.

Line 16: The same as previous.

Answer: This change has been made.

Change: $Q_{vm}$

Line 20: “We...”. It is up to you, but you should try to get coherence throughout the text. Consider using the impersonal form, as you have done in the rest of the manuscript.

Answer: Thank you for raising this good point. We have rewritten the sentence.

Change: The number of iterations was set to 2000 in the calibration process.

Line 21: “...step”. Please, use another Word, as step can be misunderstood by iteration.

Answer: Thank you for raising this good point. We have changed this phrasing.

Change: The number of iterations was set to 2000 in the calibration process.

Line 26: ”Wei-jian et al., 2016”. It is not in the reference list.

Answer: We apologize. We have replaced it with another paper due to the reference being in Chinese.

Change: Cheng, C. T., Zhao, M. Y., Chau, K., and Wu, X. Y.: Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure, J. Hydrol., 316, 129-140, 2006.

Line 26-27: “We test the performance... the middle Yellow River.” This sentence is relevant? I mean, you should restructure the whole paragraph in order to get coherence, instead of specifying one by one each model.

Answer: Thank you for this good suggestion. We have made some changes.

Change: Please refer to the revised version of the manuscript.

Line 27: “Zhao (1983)”. It is not in the reference list.

Answer: We have added this information.

Change: Zhao, R.: Watershed Hydrological Model: Xinanjiang Model and Shanbei Model, Water and Power Press, Beijing, China, 1983.

Line 28: “(Bao et al., 2017)”. Could you provide more references?

Answer: We have added some references.

Change:

Li, Z. J., and Zhang, K.: Comparison of three GIS-based hydrological models, J. Hydrol. Eng., 13, 364-370, 2008.

Zhao, L., Xia, J., Xu, C. Y., Wang, Z., Sobkowiak, L., and Long, C.: Evapotranspiration estimation methods in hydrological models, J. Geogr. Sci., 23, 359-369, 10.1007/s11442-013-1015-9, 2013.

Line 28: “MIKE SHE...”. Could you provide more characteristics of the model?

Answer: We have added more characteristics of the model.

Change: MIKE SHE originated from the Système Hydrologique Européen (SHE) program, and it is a deterministic, physically based distributed hydrologic model that can simulate surface water flow, unsaturated flow and saturated flow (Jayatilaka et al., 1998). MIKE SHE has been used to solve water resources and environment problems at different spatiotemporal scales (Li et al., 2018; Rujner et al.,2018; Samaras et al., 2016).

**Page 6**

line 3: There are many grammatical errors. Please, revise the whole point.

Answer: Thank you for pointing out these errors. We have corrected them.

Change: Please refer to the revised version of manuscript.

Line 3: “...of...”. "for" instead of "of”?

Answer: This change has been made.

Change: for.

Line 4: “...characterizes...”, characteristics.

Answer: This change has been made.

Change: characteristics.

Line 4: “and lack of enough rain gauges,”, and also dispersion.

Answer: This change has been made.
Change: and also dispersion.

Line 5: “flood simulation”, flood simulations.
Answer: This change has been made.
Change: flood simulations.

Line 6: “...to...”, for.
Answer: This change has been made.
Change: for.

Line 7: “...flood feature...”. I would remove "flood" as It is obvious.
Answer: This change has been made.
Change: This change has been made.

Line 9: “calculation...”. I would remove this.
Answer: This change has been made.
Change: This word has been removed.

Line 11: “...and Bayesian method...”. Which one? could you give some more information?
Answer: We have added this information.
Change: and Bayesian method with Markov chain Monte Carlo sampling.

Line 12: “ but the way may...”. I do not understand that phrase.
Answer: We have improved the sentence.
Change: although these methods may not lead to clear decisions.

Line 13: “...acquire...”. Consider to change the Word.
Answer: This change has been made.
Change: obtain.

Line 13: “...utility...”, useful.
Answer: This change has been made.
Change: useful.

Line 14: “...Yellow Rivers.”. Please be careful. Remove "s".
Answer: This change has been made.
Change: ...Yellow River.

Line 16: “modeling...”, modeled.
Answer: This change has been made.
Change: modeled

Line 16: “peak flows”. Remove "s".

Answer: This change has been made.
Change: peak flow.

Line 21: “modeling...”, modeled.
Answer: This change has been made.
Change: modeled.

Line 22: “(detailed ...)”. I would not use the parenthesis here. Just use the semicolon.
Answer: Thank you for this good suggestion. We have improved the sentence.
Change: one component of the Bayesian forecasting system is detailed in Krzysztofowicz (1999) and Biondi et al. (2010).

Line 23: “(Krzysztofowicz, 1999; Biondi et al., 2010)”. Remove parenthesis and write them in a correct way.
Answer: Thank you for noting these errors. We have corrected this phrasing.
Change: …Krzysztofowicz (1999) and Biondi et al. (2010).

Line 25: “modelling peak flow...”, the modelling peak.
Answer: This change has been made.
Change: the modelling peak.

**Page 7**
Line 4: “(Pianosi and Wagener, 2015)”. Please write the parenthesis in the correct way. "pianosi and Wagener (2015) proposed …".
Answer: Thank you for pointing out these errors. We have corrected this phrasing.
Change: Pianosi and Wagener (2015)

line 4: “PAWN”, What is the meaning?. Explain the method.
Answer: PAWN is derived from the authors names according to Pianosi and Wagener (2015). We have added more explanation.
Change: Pianosi and Wagener (2015) proposed the novel GSA method PAWN (derived from the authors’ names) based on the cumulative density function.

Line 5-6: “(Khorashadi Zadeh et al., 2017)”. Please, write the reference in the correct way.
Answer: We have corrected this phrasing.
Change: Khorashadi et al. (2017)

Line 21: “5 events...”. "five" instead of 5. Correct the rest of remaining ones throughout the paragraph.
Answer: We have corrected them.
Change: Five

Line 27: Results Section is good, but you have to rewrite it better, and expand it. Please, try to be more clear, and specify every result you have.

Answer: Thank you for these suggestions. We have rewritten Section 4.1.
Change: Please refer to the revised version of the manuscript.

**Page 8**

Line 1: "...VMM performs better...". In both calibration and validation? Please, expand your explanation.

Answer: We apologize for our unclear expression. We have rewritten the sentence.

Change: In terms of the median and average of the absolute relative errors for peak flows, except for the validation period in the Kuye River catchment shown in Figure 3 (h), Figures 3 (a)–(g) reveal that the VMM has lowest values for both calibration and validation.

Line 2: "...for both median and average peak flows." I understand what you are saying, but try to work out your explanation.

Answer: We have rewritten the sentence.

Change: In terms of the median and average of the absolute relative errors for peak flows, except for the validation period in the Kuye River catchment shown in Figure 3 (h), Figures 3 (a)–(g) reveal that the VMM has lowest values for both calibration and validation.

Line 11: "Overall...". Remove that. Rewrite the sentence.

Answer: We have rewritten the sentence.

Change: The analysis of Figure 3, Table 3 and Table 4 shows that the VMM has the best performance for flood modeling in the four studied catchments of the middle Yellow River.⋯

Line 11-12: "in the middle Yellow River". " of the" instead of "in".

Answer: This change has been made.

Change: of the

Line 12: "MIKE SHE...". There is a lack of connectors in some parts of the text. Here, you should use: "Besides", "In addition", etc.

Answer: This change has been made.

Change: In addition

Line 21: "Eq.(9) and Eq.(11)". Substitute the Equation by the variable, or write both of them. "Ep (Eq. 9) and Epv (Eq. 11)".

Answer: Thank you for this good suggestion. We have rewritten them.

Change: $E_p$(Eq. 9) and $E_{pv}$(Eq. 11).

Line 21-22: "The most sensitive ...objective functions.". Try to explain better what you want to mean. Rewrite the sentence.

Answer: We have rewritten the sentence.

Change: The higher the ranking is, the more sensitive the parameters. We can find that the parameters *CS*, *IM* and *KE* have the highest rankings whether the objective function of the VMM model is $E_p$ or $E_{pv}$.

Line 29: "...must meet...". There are two "must" in the sentence. Please, use synonym or rewrite the sentence.

Answer: We have rewritten the sentence.

Change: The framework requires that an accepted flood event should meet one of the requirements of C1 and C2; in addition, C3 needs to be satisfied simultaneously.

Line 30: "Flood events conforming to conditions...". Is that right?

Answer: We apologize for the unclear expression. We have rewritten the sentence.

Change: The observed peak flows and the modeled peak flows under the conditions C1, C2 or C3 are shown in Figure 5.

**Page 9**

Line 15: "...VMM...". In the conclusions part, you should rewrite the abbreviations. Please, check the rest of references and abbreviations.

Answer: We have checked them and added an explanation for clarification.

Change: Please refer to the revised version of the manuscript.

Line 19-20: "In the four catchments, by...sensitive parameter.". The same as previous.

Answer: We have rewritten the sentence.

Change: In the four catchments, the parameters confluence coefficient of surface flow (*CS*), impermeable area (*IM*), and residence time of Muskingum (*KE*) are the most sensitive based on an analysis by the global sensitivity method PAWN; in addition, the sensitivity ranking of the parameter *WM* related with the soil moisture capacity is the most affected by the objective functions.

Line 23: "The condition C2... ". Explain that. You should bear in mind Conclusions could be read by anyone, so they should not contain references to the rest of the article unless it is completely necessary.

Answer: We have added more explanation.

Change: The condition C2, which divides peak flows into three flow zones, will be affected...

Line 24: "... enrich...". Please, use another word.

Answer: This change has been made.

Change: The framework... provide guidance for decision making.

**Page 10**

Line 5: "Reference:". There are several errors in the reference list. Please, read the journal's rules, and correct every mistake. I just write down some of them. Besides, order the references according to the standards of the journal.

Answer: We have checked the references and corrected them.

Change: Please refer to the revised version of the manuscript.

Line 6: “Andersen....”. This is a PhD Thesis. Please, read journal's rules.
Answer: This change has been made.
Change: Andersen, F. H.: Hydrological modeling in a semi-arid area using remote sensing data, Ph.D. thesis, University of Copenhagen, Copenhagen, Denmark, 2008.

Line 8-9: “Burnash...”. I found the paper, but there is a lack of information here. Please complete the reference.
Answer: This change has been made.
Change: Burnash, R. J., Ferral, R. L., and McGuire, R. A.: A generalized streamflow simulation system, conceptual modeling for digital computers, Report by the Joliet Federal State River Forecasts Center, Sacramento, CA, 204 pp., 1973.

Line 11-12: “Bao, W. and Wang, C....”. Unless, it is completely necessary, please, try to avoid references that are hard to find out. I could not find those references in Chinese.
Answer: Thank you. We believe that this suggestion is very constructive. We have replaced this reference with another paper. However, some references in Chinese are completely necessary, so we have kept them.
Change: Wang, G., and Ren, L.: A Contrastive Study of Simulation Results between GWSC-VMR and Hybrid Runoff Model in Dianzi Basin, in: International Conference on Environmental Science and Information Application Technology, Wuhan, China, 4 – 5 July, 583-588, 2009.

Line 22: “Beven, K.:...”. It is needed more info about these references.
Answer: This change has been made.
Change: Beven, K. J.: Environmental modelling: An uncertain future?, CRC press, London, UK, 328 pp., 2007

Line 23: “Beven, K.:...”. It is needed more info about these references.
Answer: This change has been made.
Change: Beven, K. J.: Rainfall-runoff modelling: the primer, John Wiley & Sons, UK, 488 pp., 2011

**Page 11**
line 4: “3-23, 2007.” DOI?
Answer: This change has been made.
Change: Collier, C. G.: Flash flood forecasting: What are the limits of predictability?, Q. J. Roy. Meteor. Soc., 133(622), 3 – 23, https://doi.org/10.1002/qj.29, 2007

Line 31-32: “Li, D: ...”. Rewrite according to the journal's rules.
Answer: This change has been made.
Change: Li, D.: Hydrologic model: the vertically mixed runoff model (vmm), HydroShare, https://doi.org/10.4211/hs.c523228 7d5c04bfb8cac5ce4e391ea 0f, 2018

**Page 12**

Line 2: "...Qiushui River, Yellow River, 06, 24-28, 2018a.". This is a journal?? Please, again try to avoid these references, but also try to make easy to find them in case you want to add them to the list.

Answer: We have deleted it.

Change: The reference has been deleted.

Line 29: "1-32". Pages range is wrong.

Answer: Thank you for noting these errors. We have corrected this phrasing.

Change: Sharma, K. D., and Murthy, J. S. R.: A practical approach to rainfall-runoff modelling in arid zone drainage basins, Hydrolog. Sci. J., 43(3), 331 – 348, 1998

Line 32: "United Nations Environment...". More information about the Book.

Answer: This change has been made.

Change: United Nations Environment Programme (UNEP): World Atlas of Desertification, Edward Arnold, London, 69 pp., 1992

**Page 13**

Line 3: "...Yellow River.". The same as previous.

Answer: This change has been made.

Change: Please refer to the revised manuscript.

Line 8: "Mathematical modelling and computational experiments.", abbreviation?

Answer: This change has been made.

Change: Sobol, I. M.: Sensitivity estimates for nonlinear mathematical models, Math. Model. Comput. Exp., 1, 407 – 414, 1993.

Line 13: "44(5), 2008.", pages range?

Answer: We have added this information.

Change: Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M. and Stewart, A.: Understanding uncertainty in distributed flash flood forecasting for semiarid regions, Water Resour. Res., 44(5), 61 – 74, 2008.

Line 15: "J.GEOPHYS. RES-ATMOS.". Why in capital letters?

Answer: Thank you for noting these errors. We have corrected this phrasing.

Change: Young, C.B., Nelson, B.R., Bradley, A.A., Smith, J.A., Peters-Lidard, C.D., Kruger, A. and Baeck, M.L.: An evaluation of NEXRAD precipitation estimates in complex terrain, J. Geophys. Res.-Atmos., 104, 19691 – 19703, 1999.

**Page 18**

Figure 1: What is the meaning of the big R? Please, include other references in the map of the Yellow River Basin, maybe the border of the see or the cities included inside the basin. Also, the rain gaunges stations are not clear. Please, change the mark. Finally, the drainage basin is too gross. Please, try to add some more detail.

Answer: Thank you for these suggestions. The big R may be a display error. It should be the North Arrow. In addition, we have checked all figures in case other errors occurred. We have improved Figure 1 for clarity.

Change:

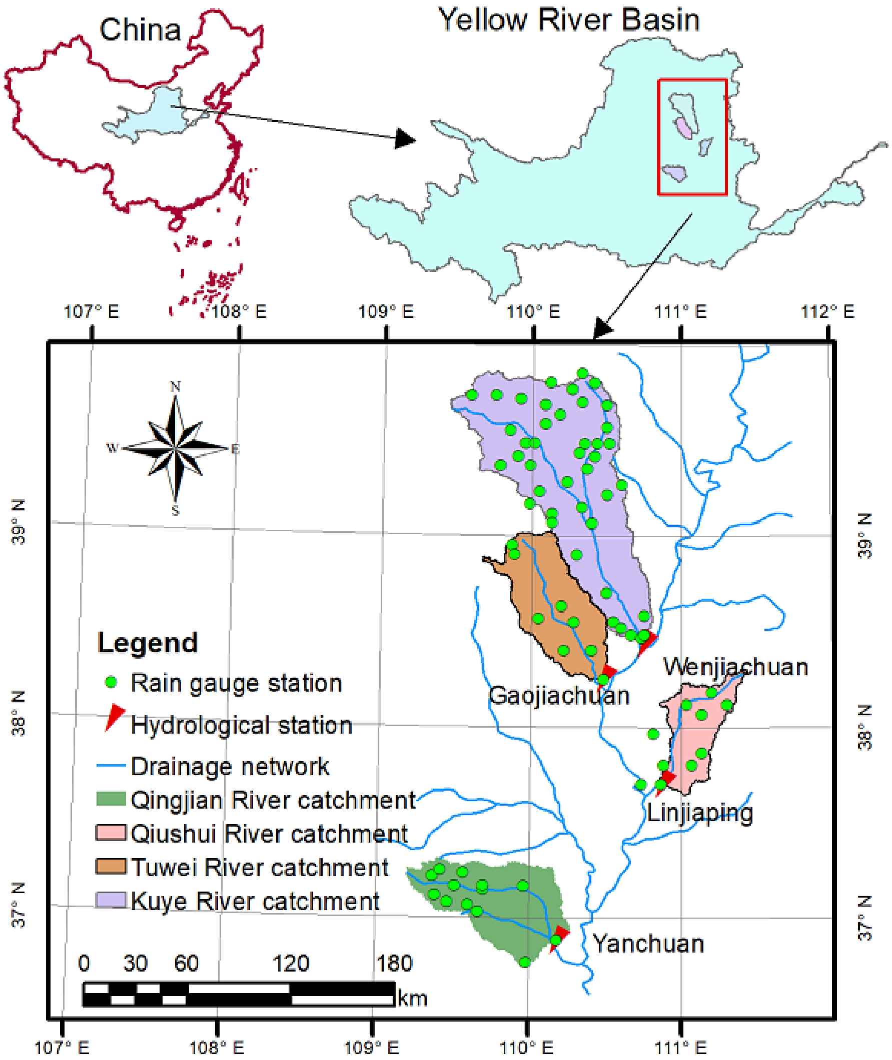


Figure 1: Location of the Qingjian River catchment, Qiushui River catchment, Tuwei River catchment and Kuye River catchment.

**Page 20**

Figure 3: Please, add letters to identify each graph, in order to be able to talk about them in the manuscript. Also, you should add some space among graphs. Regarding the axes, specify clearly what is the variable they are measuring, I guess they talk about errors.

Answer: Thank you for these suggestions. We have improved Figure 3. In addition, the Y axis label should be "Absolute relative errors of peak flows (%)". We have corrected this phrasing.

Change:

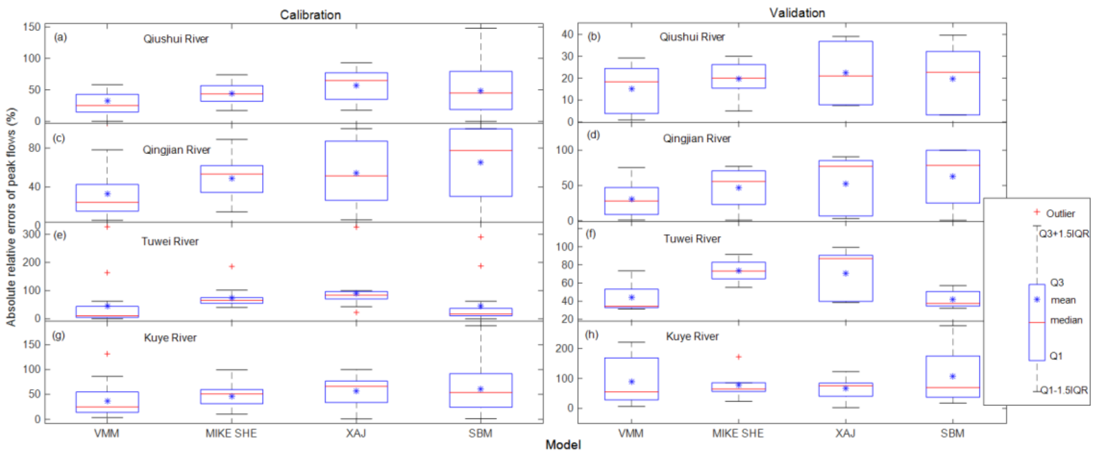


Figure 3: Boxplots of the absolute relative errors of the peak flows in the four catchments; Q1 and Q3 resprent the first quantile and third quantile, respectively; interquartile range (IQR)= Q3 − Q1; and an outlier is defined as an extreme value that exceeds the IQR.

**Page 21**

Figure 4: Please, rewrite the whole figure caption. It is very difficult to understand what you try to say.

Answer: We apologize for the lack of clarity. We have rewritten the figure caption.

Change: Figure 4: Sensitivity rankings of the VMM parameters based on the global sensitivity analysis method PAWN for different objective functions: (a) $E_{pv}$ as the objective function, and (b) $E_p$ as the objective function. The value $P$ is used to assess the sensitivity degree of the parameter with the PAWN method, and a larger value corresponds to greater sensitivity. The numbers on the ordinate represent the sensitivity rankings.

**Page 22**

figure 5: Please, add letters at each graph, in order to be able to talk about them in the text. In addition, clarify the legend, it is difficult to identify each variable. Regarding the caption, please give more info, more detail about the graph. Explain more about we are watching.

Answer: Thank you for this good suggestion. We have added letters at each graph and additional explanations in the figure caption. In addition, additional explanations have been added to Section 4.3.

Change:

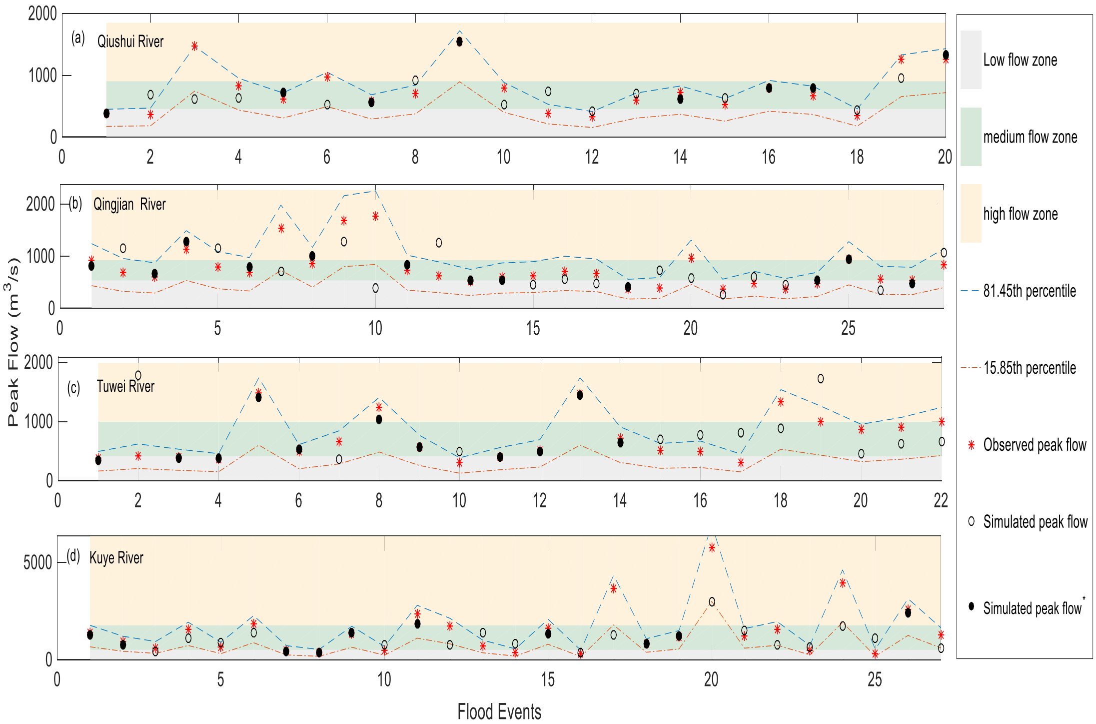


Figure 5: Observed peak flows (red asterisk) and simulated peak flows (solid ball and circle) with the VMM for each catchment under the conditions C1, C2 and C3. Flood peaks conforming to the condition C1 is represented by a solid ball; the three flow zones (low, medium and high flow zones) divided by the condition C2 are shown in gray, green and off-white, respectively; 68.3% confidence interval of peak flows estimated by the condition C3 is between the blue dashed line (81.45th percentile) and the red dash-dotted line (15.85th percentile).

# Multicriteria assessment framework of flood events simulated with vertically mixed runoff model in semiarid catchments in the middle Yellow River

Dayang Li, Zhongmin Liang, Yan Zhou, Binquan Li, Yupeng Fu

College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China

*Correspondence to*: Binquan Li (libinquan@hhu.edu.cn)

Deleted: the

Deleted: a

Deleted: reaches of the

**Abstract.** Flood forecasting in semiarid regions is always poor, and a single criterion assessment provides limited information for decision making. Here, we propose a multicriteria assessment framework called flood "classification–reliability" assessment (FCRA) that combines the absolute relative error, flow classification and uncertainty interval estimated by the Hydrologic Uncertainty Processor (HUP) to assess the most striking feature of an event-based flood: the peak flow. One hundred flood events in four catchments of the middle reaches of the Yellow River are modeled with hydrological models over the period of 1983–2009. The vertically mixed runoff model (VMM) is compared with one physical-based model, the MIKE SHE model (originated from the Système Hydrologique Européen program), and two conceptual models, the Xinanjiang model (XAJ) and the Shanbei model (SBM). Our results show that the VMM model has a better flood estimation performance than the other models, and the FCRA framework can provide reasonable flood classification and reliability assessment information, which may help decision makers improve their diagnostic abilities in the early flood warning process.

Deleted: and simulation

Deleted:

Deleted: are

Deleted: partitioning

Deleted: confidence

Deleted: The vertically mixed runoff model (VMM) is compared with three models, one physical-based model, the MIKE SHE and two conceptual models, the Xinanjiang model (XAJ) and the Shanbei model (SBM). The 100

Deleted: The 100

Deleted: the

Deleted: three models,

Deleted: under the

Deleted:

Deleted: multicriteria assessment

Deleted: FCRA

Deleted: -

Deleted: flood

Deleted: , the average acceptance of flood events is improved. In addition, the framework may provide reasonable flood early warning information for decision-makers.

## 1 Introduction

Arid and semiarid regions account for approximately one-third of the global land surface and half of China's land surface. A trend towards a warmer climate has increased the global incidence of intense precipitation events. Arid and semiarid regions, i.e., areas where the annual rainfall is less than 250 and 250–500 mm/a, respectively, are particularly vulnerable to this change in climate (Khomsi et al., 2016; Yatheendradas et al., 2008). More than 50% of flood-related casualties occur in these regions worldwide (Brito and Evers, 2016).

Hydrological models play an important role in flood simulation and forecasting (Devia et al., 2015). Many studies have focused on the improvement and estimation of hydrologic models in humid catchments, although there are fewer similar studies for semiarid catchments (Jiang et al., 2015). The runoff generation mechanisms for semiarid catchments are complex and may be simultaneously dominated by infiltration excess and saturation excess mechanisms (Beven, 1983; Beven and Freer, 2001).

Deleted: work

Deleted: is relatively few

Deleted: of

Deleted: is

Modeling semiarid catchments is a difficult task due to the strong spatial variability in rainfall and complexity of landscape characteristics (vegetation, soil, etc.) (Pilgrim et al., 1988). Compared with humid catchments, the rainfall of semiarid catchments is characterized by a high intensity and short duration (Andersen, 2008). In certain areas with developing economies and small populations, the raingauge networks are generally sparse. Rainfall data are important inputs for hydrologic models, and the high temporal-spatial rainfall variability combined with sparse raingauges makes modeling runoff more difficult (Hao et al., 2018; Li and Huang, 2017; Mwakalila et al., 2001).

Satellite technology has the possibility to solve the issue of low raingauge densities, although the low spatial and temporal resolutions of the products limit their applicability to subdaily rainstorms (Dinku et al., 2007). Weather radar has high spatial resolution (1 km) and temporal resolution (15 min). However, the radar costs are too high to be used for large-scale semiarid areas (Young et al., 1999).

Literature on the subdaily modeling of rainfall runoff is limited in semiarid catchments. Due to rapid times-to-peak and scarce rainfall data, capturing rainstorm flood responses is more difficult than estimating daily, monthly or annual runoff (Andersen, 2008; McMichael et al., 2006). Flood simulation results in semiarid catchments are often poor. Michaud and Sorooshian (1994) used 24 severe rainstorms that produced the largest peak flows from 1957 to 1977 to compare three hydrologic models, i.e., the lumped Soil Conservation Service (SCS) model, the simple distributed SCS model, and the distributed kinematic runoff and erosion (KINEROS) model, in the Walnut Gulch catchment, and none of them were able to accurately simulate flood events. McIntyre and Al-Qurashi (2009) analyzed 27 flood events with three hydrologic models (the lumped Identification of Hydrographs and Components from Rainfall, Evaporation and Streamflow (IHACRES) model, the distributed IHACRES model, and a 2-parameter regression model) in a catchment in Oman. The average absolute relative errors in the flow peak and flow volume were 53% and 36%, respectively, for the best performing models. Under current technical conditions, it seems difficult to achieve an acceptable simulation/forecasting result for flood events in semiarid catchments. Therefore, determining how to use modeling results with limited accuracy to provide guidance for early flood warnings is important.

In this study, a multicriteria assessment framework that combines the absolute relative error, flow classification and the uncertainty interval estimated by Hydrologic Uncertainty Processor (HUP) is proposed to provide more information for engineers' decision making. Four hydrological models, the vertically mixed runoff model (VMM), the MIKE SHE model (originated from the Système Hydrologique Européen program), the Xinanjiang model (XAJ) and the Shanbei model (SBM), are compared based on the performance of the modeling results in four catchments in the middle reaches of the Yellow River. The global sensitivity analysis (GSA) method PAWN is used to analyze the parametric sensitivity of the VMM model. The rest of the paper is organized as follows. Section 2 describes the study area and the data set used. Section 3 presents the VMM model methodology, model set, model calibration and validation, multicriteria assessment framework and parameter sensitivity analysis. Section 4 describes the results and discussion of model comparison, sensitivity analysis and analysis of the

**Deleted:** network of

**Deleted:**

**Deleted:** s

**Deleted:** is

**Deleted:**

**Deleted:**

**Deleted:** quick

**Deleted:** –

**Deleted:** flood

**Deleted:** partitioning

**Deleted:** confidence

**Deleted:** :

**Deleted:** the

**Deleted:** e

**Deleted:** initial condition

**Deleted:**

multicriteria assessment framework for the VMM model. The final section presents the conclusions of the study.

## 2 Study area and data

The four selected study catchments are all key tributaries located in the middle reaches of the Yellow River, China (Fig. 1). The maximum and minimum areas of the catchments are 1989 km$^2$ and 8706 km$^2$, respectively. The average annual temperature ranges from 6 to 14 ℃. The average annual precipitation ranges from 1010 to 1150 mm, and 65 to 80% is concentrated in summer (Li et al., 2019; Li and Huang, 2017; Xiao et al., 2019). The rainfall is generally characterized by high intensity and short duration. The average annual evaporation ranges from 1010 to 1150 mm. All selected catchments are semiarid based on an aridity index between 2.31 and 2.78 (UNEP, 1992). This catchment information is listed in Table 1.

The lack of vegetation in these catchments leads to severe soil erosion, and the average sediment concentration reaches 126 kg m$^{-3}$ according to Li et al. (2019). Some hydrologists have studied daily and monthly rainfall runoff, although few studies have modeled hourly floods. With the rapid increase in population and economic development, flood disasters have received increasing attention. Hence, it is important for decision makers to know how to evaluate the flood risk when a flood is approaching.

The period used in the model was from 1983 to 2009. Continuous streamflow and rainfall data were collected from streamflow gauging stations and rain gauging stations at a daily time step, respectively; streamflow and rainfall data for each of the flood events were collected at an hourly time step. Nine rainfall gauging stations in the Qiushui River catchment, 15 rainfall gauging stations in the Qingjian River catchment, 12 rainfall gauging stations in the Tuwei River and 41 rainfall gauging stations in the Kuye River were selected. The Thiessen polygon method was used to interpolate the rainfall data for each catchment.

## 3 Methodology

### 3.1 Vertically mixed runoff model

The VMM is a lumped, continuous hydrologic model and has been used in many areas in China, especially in semiarid and subhumid catchments (Bao and Zhao, 2014; Li, 2018; Wang and Ren, 2009). Compared with other conceptual models, such as the XAJ model (Zhao, 1992) and the Sacramento Soil Moisture Accounting (SSMA) model (Burnash et al. 1973), among others, the VMM model is capable of simulating the saturation excess and infiltration excess runoff generation mechanisms simultaneously. As shown in Fig. 2, the VMM model combines the infiltration capacity curve and tension water content storage capacity curve in the vertical direction. Net rainfall (observed rainfall after removal of evaporation, *PE*) is partitioned into surface runoff (*RS*) and infiltration flow (*FA*) by the infiltration capacity curve in the VMM model. *FA* is regulated by the tension water storage capacity curve, part of which supplements the tension water storage (*W*), with the remainder of the

**Deleted:** 4
**Deleted:** ure
**Deleted:** .
**Deleted:** –
**Deleted:** –
**Deleted:** –
**Deleted:** due to
**Deleted:** serious
**Deleted:** -
**Deleted:** ing
**Deleted:** is
**Deleted:** are
**Deleted:** are
**Deleted:** ,
**Deleted:** Model
**Deleted:** able
**Deleted:** to
**Deleted:** e
**Deleted:** ure
**Deleted:** .
**Deleted:** rest

rainfall forming groundwater flow (*RB*) (including unsaturated flow and saturated flow). Here, the calculation of runoff generation is described briefly. More detailed information about the VMM model is contained in Bao and Zhao (2014).

The improved Green-Ampt infiltration curve (Bao, 1993) is applied in the VMM model as the infiltration capacity curve, and the equation is as follows:

$$FM = FC\left(1 + K\frac{WM - W}{WM}\right) \tag{1}$$

where *FM* is the average point infiltration capacity of the catchment, and the descriptions of *WM*, *K*, and *FC* are shown in Table 2.

*FA* is calculated by Equation (2):

$$FA \quad = \begin{cases} FM - FM\left(1 - \frac{PE}{(FMM)^{1+BF}}\right) & PE < FMM \\ FM & PE \geq FMM \end{cases} \tag{2}$$

where

$$FMM = FM(1 + BF) \tag{3}$$

in which *FMM* is the maximum point infiltration capacity of the catchment and *BF* is defined in Table 2.

The part that exceeds the average point infiltration capacity of the catchment *FM* forms *RS*. *RS* can be calculated by Equation (4).

$$RS = PE - FA \tag{4}$$

*RB* can be calculated by Equation (5):

$$RB \quad = \begin{cases} FA - WM + W + WM\left(1 - \frac{W^* + FA}{WMM}^{B+1}\right) & FA + W^* < WMM \\ FM - WM + W & FA + W^* \geq WMM \end{cases} \tag{5}$$

where

$$W^* = WMM\left[1 - \left(1 - \frac{W}{WM}\right)\right]^{\left(\frac{1}{B}+1\right)} \tag{6}$$

$$WMM = WM(1 + B) \tag{7}$$

in which *WMM* is the maximum point tension water storage capacity of the catchment; $W^*$ is the ordinate of Fig. 2 (b); which represents the point tension water content capacity in the catchment; and *B* is defined in Table 2.

The outlet runoff *R* can be calculated as follows:

$$R = RS + RB \tag{8}$$

### 3.2 Model set of the VMM model

The VMM model was run continuously from 1983 to 2009 for each catchment. Rainfall data were available only at an hourly time step over the periods of flood events, and for other periods, they were available at a daily time step. Hence, the time step

**Deleted:** the

**Deleted:** .

**Deleted:** shown

**Deleted:** .

**Deleted:** .

**Deleted:** ,

**Deleted:** ure

**Deleted:** .

**Deleted:** ,

**Deleted:** ,

**Deleted:** shown

**Deleted:** Initial condition

**Deleted:** The initial condition has important effects in modeling flood events.

of simulations was set as daily between flood events and hourly within flood events for each catchment. To consider the spatial variation in rainfall, the subcatchments were divided according to the stream networks, and each subcatchment contained at least one rainfall gauging station. The areal mean rainfall of each subcatchment was calculated using the Thiessen polygon method. Because streamflow data were only available in the outlet streamflow gauging station for each catchment, the spatial variation in each catchment's parameters could not be determined by calibration. Thus, the parameters (Table 2) were set uniformly in all subcatchments. Two initial values, the initial tension water storage (*W0*) and the initial free water storage (*S0*), were used to describe the initial catchment moisture condition. The initial values are smaller for drier catchments, and the minimum values are zero. In this study, the initial values were assumed to be zero uniformly due to the dry conditions at 00:00:00 on January 1, 1983 for each catchment. It should be noted that continuous simulations for each catchment eliminate the need to set the initial values for each flood event in a catchment.

### 3.3 Model calibration

The fourteen parameters (Table 2) of the VMM model were calibrated using the Shuffled Complex Evolution (SCE-UA) global optimization algorithm (Duan et al., 1993). The ranges of parameters were determined based on previous literature and prior knowledge (Bao and Zhao, 2014; Li et al., 2018). Due to the rapid rise and fall of floods (usually less than 24 h) in semiarid catchments, accurate simulations of the full hydrograph are not needed and cannot be realized. The Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) is widely used as an objective function of calibration in humid catchments; however, it may not be suitable for semiarid catchments because a good fit is not required between the simulated and observed streamflows. McIntyre and Al-Qurashi (2009) and Sharma and Murphy (1998) used the absolute relative error to evaluate model outputs (flow peak and flow volume) for semiarid areas, and the calibrated results indicated that the peak flow results are more accurate than the suggested results based on the NSE. Thus, the simulated hydrograph is reasonable for the majority of flood events. The equations are as follows:

$$E_p = \frac{1}{n}\sum_{i=1}^{n}\frac{|Q_p^i - Q_{pm}^i|}{Q_{pm}^i} \tag{9}$$

$$E_v = \frac{1}{n}\sum_{i=1}^{n}\frac{|Q_v^i - Q_{vm}^i|}{Q_{vm}^i} \tag{10}$$

where $E_p$ and $E_v$ are the average performances (in terms of absolute relative error) for peak flows and flow volumes in each catchment, respectively; $n$ is the number of events; the index $i$ denotes each event; $Q_p$ and $Q_{pm}$ are the simulated and measured values of peak flow per event, respectively; and $Q_v$ and $Q_{vm}$ are the simulated and measured values of flow volume per event,

**Deleted:** was guaranteed to have

**Deleted:** with

**Deleted:** Due to the fact that

**Deleted:** cannot

**Deleted:** are

**Deleted:** initial values

**Deleted:** are

**Deleted:**

**Deleted:** The drier the catchment, the smaller t

**Deleted:**

**Deleted:** should

**Deleted:** be determined.

**Deleted:** Both of them represent the moisture content of the soil and were assumed to

**Deleted:**

**Deleted:** is worth mentioning

**Deleted:** of

**Deleted:** Rainfall data were available only at an hourly time step over the periods of flood events, and for other periods, they were available at a daily time step. Hence, the time step of simulation was daily between flood events and hourly within flood events.

**Deleted:** To consider the spatial variation in rainfall, the subcatchments are divided and the VMM model is applied to each subcatchment. Due to the fact that only one streamflow gauging station is available for each catchment, the spatial variation in each catchment's parameters cannot be determined by calibration. Thus, the parameters are set uniformly in all subcatchments.

**Formatted:** Indent: First line: 0 ch

**Deleted:** are

**Deleted:** by

**Deleted:** global optimization algorithm

**Deleted:** are

**Deleted:** ¶

**Deleted:** In semiarid catchments, due

**Deleted:** ;

**Deleted:** flow peak

respectively.

Constraining the model output with peak flows and flow volumes can be expressed as follows:

$$E_{pv} = \frac{E_p + E_v}{2} \tag{11}$$

where $E_{pv}$ is the objective value. The model outputs become better as the value of $E_{pv}$ approaches 0. The number of iterations was set to 2000 in the calibration process.

### 3.4 Model comparison

To achieve a better performance in rainstorm flood simulations, three hydrologic models, including two conceptual models, XAJ and SBM, and one distributed model, MIKE SHE, were used for comparison with the VMM model. The XAJ model was developed by (Zhao, 1992) and has a single saturation excess runoff generation mechanism. The XAJ model has been successfully applied in humid and subhumid catchments (Cheng et al., 2006; Lü et al., 2013;). The SBM model was developed by Zhao (1983) and has a single infiltration excess runoff generation mechanism. The SBM model is generally used in semiarid or arid catchments in China (Bao et al., 2017; Li and Zhang, 2008; Zhao et al., 2013). In addition, the MIKE SHE model is a deterministic, physically-based distributed hydrologic model that can simulate surface water flow, unsaturated flow and saturated flow (Jayatilaka et al., 1998). The MIKE SHE model has been used to solve water resources and environment problems at different spatiotemporal scales (Li et al., 2018; Rujner et al., 2018; Samaras et al., 2016).

### 3.5 Multicriteria assessment framework: flood "classification–reliability" assessment for flood events

Due to strong spatially the variability of rainfall, complex landscape characteristics, insufficient raingauges, and dispersion, flood simulations and forecasting in semiarid catchments are very difficult. Although some hydrologists improve flood simulations and forecasting by improving hydrologic models, the improvements are always limited or are suitable for only specific regions (Collier, 2007). The flood peak is the most significant feature in semiarid regions. Determining the extent to which the calculation of flood peaks can be accepted is crucial. Generally, the absolute relative error is used to measure the calculation of flood peak accuracy; for example, 20%, 30% or similar values are acceptable (Li et al., 2014; McIntyre and Al-Qurashi, 2009). To provide more information for flood defense management, the generalized likelihood uncertainty estimation (GLUE) and the Bayesian framework with Markov Chain Monte Carlo sampling are used to provide probabilistic forecasting, such as the 95% uncertainty interval (Christiaens and Feyen, 2002; Li et al., 2017), although these methods may not lead to clear decisions (Beven, 2007).

In this study, to obtain a better diagnostic and discriminatory method for the decision maker, we propose a multicriteria assessment framework called the flood "classification—reliability" assessment (FCRA) in the catchments of the middle reaches of the Yellow River. The FCRA framework consists two parts: flood classification and flood reliability assessment; floods are classified with percentiles and the absolute relative error; the reliability of flood modeling is evaluated with the

**Deleted:** The closer $E_{pv}$ is to 0, the better

**Deleted:** t

**Deleted:** are

**Deleted:** originated from the Système Hydrologique Européen (SHE) program, and it

**Deleted:** lly based d

**Deleted:**

**Deleted:** event

**Deleted:**

**Deleted:**

**Deleted:** certain

**Deleted:** ,

**Deleted:** the

**Deleted:** confidence

**Deleted:** useful

**Deleted:**

**Deleted:**

**Deleted:** for flood forecasting

**Deleted:** framework

Bayesian method. Peak flows, as the most prominent features of flood events, are assessed with the FCRA framework. Detailed descriptions can be found as follows:

(C1) The absolute relative error of peak flow should be less than 20%.

(C2) The modeled and observed peak flows should be in the same flow zone: the observed peak flow $Q_p$ for all flood events in a catchment are divided into three zones (low flow zone, medium flow zone, high flow zone), with 25th percentiles $Q_{p25}$ and 75th percentiles $Q_{p75}$ as the boundary points; if $Q_p \leq Q_{p25}$, then the peak flow $Q_p$ belongs to the low flow zone; if $Q_p \geq Q_{p75}$, then the peak flow $Q_p$ belongs to the high flow zone; the remaining flow peaks belong to the medium flow zone. Both the 25th percentile and 75th percentile are commonly used to distinguish zones.

(C3) The observed peak flows should fall within one standard deviation (σ) of the mean (approximately 68.3% uncertainty interval) peak flow estimated by the Hydrologic Uncertainty Processor (HUP), one component of the Bayesian forecasting system detailed in Krzysztofowicz (1999) and Biondi et al. (2010).

Conditions C1 and C2 are flood classification criteria. If the observed and modeled peak flows meet one of the two conditions, we believe that they are the same types of floods. The key of the FCRA framework is condition C2, and condition C1 is used to avoid errors caused by flow zone boundaries. For example, when $Q_{p75} = 200$ m$^3$/s, the modeled peak flow equals 198 m$^3$/s and the observed peak flow equals 201 m$^3$/s. However, using only condition C2 may lead to inappropriate model results; adding condition C1 can help address the problem. Condition C3 is used to assess the reliability of peak flows modeling; it uses a small uncertainty interval (68.3%) that has narrow upper and lower limits. This interval may reduce the numbers of observed peak flows that fall within the confidence level. A modeled peak flow that can be accepted should satisfy condition C1 or condition C2 and then condition C3.

### 3.6 Parameter sensitivity analysis

A sensitivity analysis (SA) (Saltelli et al., 1989) was proposed to assess the effects of inputs on the model output. The SA can be classified into a GSA and local sensitivity analysis (LSA). Compared with the LSA, the GSA is capable of analyzing the effects of inputs within the entire input domain. The Fourier amplitude sensitivity test (Cukier et al., 1973), Sobol method (Sobol, 1993) and Morris screening method (Morris, 1991) are the most widely used GSA methods in the assessment of parameter sensitivity in hydrologic models. Pianosi and Wagener (2015) proposed the novel GSA method PAWN (derived from the authors' names), which is based on the cumulative density function. PAWN has advantages over the parameter ranking and time-consuming nature of other GSA methods (Khorashadi et al., 2017). In this study, we used the PAWN method to perform a GSA on the VMM model.

Considering $x_{i,j}$ $(i, j = 1, 2, \cdots$, where $i$ and $j$ represent the $i$-th input parameters and the $j$-th sampling, respectively) as sensitivity inputs, then the sensitivity of $x_{i,j}$ can be measured by the distance between $F_{(y_i|x_{i,j})}(y_i)$ (the cumulative probability distribution function of $y_i$ when $x_{i,j}$ changes between the upper bound and lower bound) and $F_{y_i}(y_i)$ (the

Deleted: FCRA
Deleted: This framework
Deleted: described
Deleted: the
Deleted: of
Deleted:
Deleted: :
Deleted: and
Deleted: rest
Deleted: s
Deleted: modeled
Deleted: confidence
Deleted: The c
Deleted: think
Deleted: framework
Deleted:
Deleted: ,
Deleted: which
Deleted: ;
Deleted: t
Deleted: which
Deleted: C3 is used to test the confidence level of modeled peak flows. …
Deleted: To assess the effects of inputs on the model output, a
Deleted: (Saltelli et al., 1989)
Deleted: able to
Deleted: e

cumulative probability distribution function of $y_i$ when $x_i = \frac{1}{n}\sum_{j=1}^{n} x_{i,j}$, where $n$ is the number of samplings per input parameter). The Kolmogorov–Smirnov statistic (Simard and Ecuyer, 2011) is used to measure the distance between $F_{(y_i|x_i)}(y_i)$ and $F_{y_i}(y_i)$:

$$KS(x_{i,j}) = \max_{1 \le j \le n} \left| F_{y_i}(y_i) - F_{(y_i|x_{i,j})}(y_i) \right| \quad (12)$$

As $KS$ varies with $x_{i,j}$, the maximum of all possible $KS$ values is included in the PAWN index $P_i$:

$$P_i = \max_{1 \le j \le n} KS(x_{i,j}) \quad (13)$$

$P_i$ ranges from 0 to 1 and $x_i$ becomes more sensitive as $P_i$ approaches 1. A $P_i$ equal to 1 indicates that $x_i$ has no effect on the model. For more information about PAWN, please refer to Pianosi and Wagener (2015). In this study, the number of evaluations was set to 500, as suggested by Pianosi and Wagener (2018).

### 3.7 Model validation

The modeling period was between 1983 and 2009. In the Qiushui River, 20 flood events were selected, with the first 15 events used for calibration and the remaining five events used for validation. Similarly, in the Qingjian River, 29 flood events were selected, with 24 events used for calibration and the remaining five events used for validation. In the Tuwei River, 23 flood events were selected, with 18 events used for calibration and the remaining five events used for validation. Finally, in the Kuye River, 28 flood events were selected, with 23 events used for calibration and the remaining five events used for validation.

## 4 Results and discussion

### 4.1 Comparison of model results

Boxplots of the absolute relative errors of the peak flows for each model in the four catchments are shown in Fig. 3. In terms of the median and average of the absolute relative errors for peak flows, the VMM model has the lowest values for both calibration and validation (as shown in Figs. 3 (a)–(g)) except for the validation period in the Kuye River catchment (shown in Fig. 3 (h)); in most cases, the MIKE SHE model has lower median and average values than the XAJ and SBM models, i.e., Figs. 3 (a), (b), (c), (d), (g), (h). Low median and average values indicate that more modeled flood events have good performance in a catchment. Except for the good performance in the Tuwei River catchment, the results using the SBM model are as poor as those using the XAJ model in other catchments. In terms of ranges of the absolute relative errors for peak flows, the VMM and MIKE SHE models have relatively small ranges (Figs. 3(a), (c), (d), (g)), and the SBM and XAJ models have large ranges in most cases (Figs. 3(a), (b), (c), (d), (g)). This indicates that the VMM and MIKE SHE models are more robust to reproduce the peak flows in the middle reaches of the Yellow River.

Tables 3 and 4 show the average performance in terms of the absolute relative error for flow volume $E_v$ and the lag time for the four models in each catchment, respectively. Low $E_v$ and lag time values indicate that the model is highly capable of

**Deleted:** i

**Deleted:**

**Deleted:** .

**Deleted:** The closer $P_i$ is to 1, the more sensitive

**Deleted:** is

**Deleted:** as Pianosi and Wagener (2018) suggested,

**Deleted:** is

**Deleted:** The modeling time step was hourly, and the

**Deleted:** ure

**Deleted:** .

**Deleted:** ,

**Deleted:** ure

**Deleted:** ure

**Deleted:** .

**Deleted:** , Figures 3 (a)–(g) reveal that the VMM has lowest values for both calibration and validation

**Deleted:** performs better

**Deleted:** ure

**Deleted:** represent

**Deleted:** t, results

**Deleted:** is

**Deleted:** ure

**Deleted:** s.

**Deleted:** ;

**Deleted:** ure

**Deleted:** s.

**Deleted:** means

**Deleted:** have high ability to

reproducing the flow volumes and time-to-peak values. The VMM model has the minimum average $E_v$ and lag time, with values of 39.01% and 3.05 h, respectively (Tables 3 and 4). In contrast, the XAJ model has the maximum average $E_v$ and lag time, with values of 58.93% and 4.51 h, respectively. The MIKE SHE and SBM models have similar performances in terms of average $E_v$ and lag time.

The analysis of Fig. 3, Table 3 and Table 4 above indicates that the VMM model has the best performance to reproduce the peak flows, flow volumes and lag times in the four studied catchments of the middle reaches of the Yellow River and the XAJ model has the worst performance. In addition, the MIKE SHE model is superior for reproducing the peak flows but exhibits similar performance compared with the SBM model for reproducing the flow volume and lag time. Although the MIKE SHE model is a distributed hydrologic model with more complex structures and more explicit physical meaning than the conceptual VMM model, it does not achieve better results than the conceptual VMM model due to a lack of sufficiently high-resolution data, and this is consistent with other studies (Beven, 2002, 2011; Michaud and Sorooshian, 1994; Seyfried and Wilcox, 1995). Both infiltration excess and saturation excess can be simulated via the VMM model; this simulation may be why it performs better than the other two conceptual models (XAJ and SBM), which have single runoff generation mechanisms (saturation excess and infiltration excess, respectively).

### 4.2 Sensitivity analysis of the VMM model

The GSA method PAWN is applied to estimate the influence of parameter uncertainty on the model output results. Figure 4 (a) and Figure 4 (b) show the average SA results of all study catchments for the objective function $E_p$ (Equation 9) and $E_{pv}$ (Equation 11), respectively. The parameters become more sensitive as the ranking becomes higher. Parameters *CS*, *IM* and *KE* have the highest rankings whether the objective function of the VMM model is $E_p$ or $E_{pv}$. The rankings of other parameters are influenced slightly by different objective functions, such as *CG*, except for *WM*. *WM* ranks sixth when $E_{pv}$ is the objective function and 12th when $E_p$ is the objective function. This ranking is because *WM* controls the tension water content in the soil, which determines the amount of rainfall stored in the soil and the generation of runoff. There may be a strong relationship between flow volume and *WM*. Therefore, *WM* has a higher ranking when the objective function considers the effect of flow volume.

### 4.3 Flood “classification—reliability” assessment of the VMM model

The FCRA framework we propose is applied to assess the ability of the VMM model to model flood events in the four catchments. FCRA requires that the accepted modeled peak flows have the same flood types (high flow, medium flow or low flow) as the observed peak flows; in addition, the modeled peak flows should be reliable. Similar types of peak flows that represent the modeled peak flows should meet one of the requirements of conditions C1 and C2; the modeled peak flows that are reliably represented need to meet condition C3. The observed peak flows and the modeled peak flows under condition C1,

**Deleted:** e
**Deleted:** ure
**Deleted:** ….3, Table 3 and Table 4 above indicates that the VMM model has the best performance to reproduce the peak flows, flow volumes and lag times ...
**Deleted:** for flood modeling…in the four studied catchments of the ...
**Deleted:** …uperior ...
**Deleted:** to the SBM
**Deleted:** compared with the SBM… …lthough the MIKE SHE ...
**Deleted:** necessarily
**Deleted:** model
**Deleted:** s…model due to a because distributed models …ack of ...
**Deleted:** , which
**Deleted:** .
**Deleted:** or
**Deleted:** ….11), respectively. The higher the ranking is, the more ...
**Deleted:** S…rong relationship may have ...
**Formatted:** Font: Italic
**Deleted:** it…has a higher ranking place ...
**Deleted:** is reasonable to conclude that when the weight of the flow ...
**Deleted:** **Multicriteria assessment framework**
**Deleted:** multicriteria assessment…FCRA framework ...
**Deleted:** FCRA
**Deleted:** peaks …vents in the four catchments. . The framework ...
**Deleted:**
**Deleted:** a…the accepted modeled peak flows should …ave the ...
**Deleted:** an accepted
**Deleted:**
**Deleted:** flood event… …hould meet one of the requirements of ...
**Deleted:** is…re reliablye…represent they ...
**Deleted:** C3
**Deleted:**
**Deleted:** s…to be satisfied ...
**Deleted:** .
**Deleted:** simultaneously
**Deleted:** the conditions

C2 or C3 are shown in Figure 5. The percentages of modeled peak flows that meet the conditions are presented in Table 5. Although the percentages of the modeled peak flows that meet condition C1 are less than 50% (Table 5), they reduce the boundary effects of flood classification. Taking the 13th flood event of the Kuye River catchment as an example, the observed and modeled peak flows are 1230 $m^3$/s and 1510 $m^3$/s, respectively. As shown in Fig. 5 (d), the absolute relative error for peak flow is greater than 20%. In addition, for the Kuye River catchment, it is reasonable to believe that the peak flows 1230 $m^3$/s and 1510 $m^3$/s may have the same risk according to the known flood peak data, which can be classified as the same flood type (medium flow) according the condition C2.

From the Table 5, we find that 95% or more of modeled peak flows meet condition C3; this indicates that almost all modeled peak flows have less uncertainty and more reliability in the selected catchments. Figure 5 shows more directly that the majority of peak flows for the observations and modeling fall between the 15.85th percentile and the 81.45th percentile (68.3% uncertainty interval) estimated by HUP, which is consistent with Table 5. In addition, the percentages of modeled flood events and observed peak flows that are the same flood types (shown in Table 5 with $C^*$) equal the acceptance rate (shown in Table 5 with $C^{**}$) for each catchment due to the high reliability of modeled peak flows. Under the FCRA framework, the acceptance rates ($C^{**}$) for the catchments are more than 65% except for the Qingjian River catchment. This indicates that the FCRA framework may have the diagnostic capability to assess the modeled flood events in the four semiarid catchments.

Under the FCRA framework, a modeled flood event could be assessed to determine what flood type (high flow, medium flow, low flow) it is and how reliable it is. This information is meaningful in the early flood warning process in the semiarid catchments. Although FCRA is simple and even coarse, it is convenient and beneficial in helping engineers make decisions when a flood is approaching.

## 5 Conclusions

In this study, a multicriteria assessment framework of flood events called the flood "classification—reliability" assessment (FCRA) is proposed with the VMM model in four semiarid catchments of the middle reaches of the Yellow River. The main conclusions are as follows.

(1) Compared with the distributed model MIKE SHE and the two conceptual models, XAJ and SBM, the VMM model has better performance for modeling flood events in the middle reaches of the Yellow River.

(2) In the four catchments, the parameters confluence coefficient of surface flow (*CS*), impermeable area (*IM*), and residence time of Muskingum (*KE*) in VMM model are the most sensitive parameters based on an analysis by the global sensitivity method PAWN; in addition, the sensitivity ranking of the parameter *WM* related with the soil moisture capacity is the most affected by the objective functions.

(3) The FCRA framework combining flood classification and reliability assessment may have the reliable diagnostic

**Deleted:** ….5. The percentages of modeled peak flows that meet the conditions are show

**Deleted:** ure

**Deleted:** .

**Formatted:** Indent: First line: 2 ch

**Deleted:** the…ondition C3,… this indicates that almost all modeled peak flows have less uncertainty and more reliability in the selected catchments;… Figure 5 shows more directly

**Deleted:** We find…that the majority of peak flows for the observations and modeling fall between the 15.85th percentile and the 81.45th percentile (68.3% uncertaintyconfidence…interval) estimated by HUP, which is consistent with Table 5. means that the VMM modeling results satisfy C3 fairly well.

**Deleted:** that modeled …nd observed peak flows that are the same flood types (shown in Table 5 with $C^*$) equal the acceptance rate (shown in Table 5 with $C^{**}$) for each catchment due to the high reliability of modeled peak flows. Under the FCRA framework FCRA… except for Qingjian River catchment, …he acceptance rates ($C^{**}$) $C^{**}$ of…

**Formatted:** Superscript

**Deleted:** remaining three …atchments are more than 65% except for the Qingjian River catchment. This indicates that the FCRA framework FCRA …

**Formatted:** Indent: First line: 0.79 cm

**Deleted:** FCRA… a modeled flood event could be assessed that…o determine what flood type (high flow, medium flow, low flow) …it is,…and how reliable it is. This information is meaningful in the process of flood …arly flood warning process in the semiarid catchments. Although FCRA is simple and even coarse, it is convenient and beneficial to

**Deleted:** Under the premise of satisfying C3, the number of modeling events satisfying C2 is slightly greater than that satisfying C1. Under the multicriteria assessment framework, 15 of the 20 (75.0%) modeled flood events in the Qiushui River catchment, 12 of the 29 (41.4%) of that in the Qingjian River catchment, 15 of the 2…

**Deleted:** peaks

**Deleted:** vertically mixed runoff model (…MM)

**Deleted:** in

**Deleted:** Xinanjiang (…AJ)…and Shanbei (…BM)

**Deleted:** semiarid catchments of

**Deleted:** multicriteria assessment

**Deleted:** FCRA

capability to assess flood events in the early flood warning process. It should be noted that condition C2, which divides peak flows into three flow zones, will be affected by the number of observed peak flows when data availability is limited. The framework is suitable for semiarid regions with poor modeling results and provides guidance for decision making.

**Code availability**

We have shared the MATLAB code for the VMM model at https://doi.org/10.4211/hs.c5232287d5c04bfb8cac5ce4e391ea0f.

**Acknowledgments**

This study was supported by the National Key Research and Development Program of China (grant no. 2016YFC0402706), the National Natural Science Foundation of China (grant nos. 41730750, 41877147), and Special Scientific Research Fund of Public Welfare Industry of Ministry of Water Resources, China (201501004), sponsored by Qing Lan Project. We would like to thank Francesca Pianosi (University of Bristol) for providing the program code of PAWN at https://www.safetoolbox.info/pawn-method/. We also thank the editor and the anonymous reviewers, whose comments have largely improved this work.

**References:**

Andersen, F. H.: Hydrological modeling in a semi-arid area using remote sensing data, Ph.D. thesis, University of Copenhagen, Copenhagen, Denmark, 2008.

Bao, H., Wang L., Zhang, K. and Li, Z.: Application of a developed distributed hydrological model based on the mixed runoff generation model and 2D kinematic wave flow routing model for better flood forecasting, Atmos. Sci. Lett., 18(7), 284–293, 2017.

Bao, W.: Improvement and application of the Green-Ampt infiltration curve, Yellow River, 9, 1–3, 1993. (In Chinese)

Bao, W. and Zhao, L.: Application of Linearized Calibration Method for Vertically Mixed Runoff Model Parameters, J. Hydrol. Eng., 33(4), 85–91, 2014.

Beven, K. J.: Surface water hydrology—runoff generation and basin structure, Rev. Geophys., 21(3), 721–730, 1983.

Beven, K. J.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, Hydrol. Process., 16(2), 189–206, 2002.

Beven, K. J.: Environmental modelling: An uncertain future?, CRC press, London, UK, 328 pp., 2007.

Beven, K. J.: Rainfall-runoff modelling: the primer, John Wiley & Sons, UK, 488 pp., 2011.Beven, K. J., and Freer, J.: A dynamic TOPMODEL, Hydrol. Process., 15(10), 1993–2011, 2001.

Biondi, D., Versace, P., and Sirangelo, B.: Uncertainty assessment through a precipitation dependent hydrologic uncertainty

**Deleted:** ability

**Deleted:** process of flood

**Deleted:** provide more reliable flood early warning information than single criterion (such as absolute relative error of peak flows) when engineers need to make decisions in semiarid catchments.

**Formatted:** Font: Times New Roman, 10 pt

**Deleted:** mentioned

**Deleted:** The

**Deleted:** the

**Deleted:** the

**Formatted:** Font: Times New Roman, 10 pt

**Deleted:** s

**Deleted:** known

**Deleted:** can

**Deleted:**

**Deleted:** of

**Deleted:** and

**Deleted:** .

**Formatted:** Font: (Asian) 宋体, Not Bold, Font color: Auto

processor: An application to a small catchment in southern Italy, J. Hydrol., 386(1–4), 38–54, 2010.

Brito, M. and Evers, M.: Multi-criteria decision-making for flood risk management: A survey of the current state of the art, Nat. Hazards Earth Syst. Sci., 16(4), 1019–1033, 2016.

Burnash, R. J., Ferral, R. L., and McGuire, R. A.: A generalized streamflow simulation system, conceptual modeling for digital computers, Report by the Joliet Federal State River Forecasts Center, Sacramento, CA, 204 pp., 1973.

Cheng, C. T., Zhao, M. Y., Chau, K., and Wu, X. Y.: Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure, J. Hydrol., 316, 129-140, 2006.

Christiaens, K., and Feyen, J.: Constraining soil hydraulic parameter and output uncertainty of the distributed hydrological MIKE SHE model using the GLUE framework, Hydrol. Process., 16(2), 373–391, 2002.

Collier, C. G.: Flash flood forecasting: What are the limits of predictability?, Q. J. Roy. Meteor. Soc., 133(622), 3–23, https://doi.org/10.1002/qj.29, 2007.

Cukier. R., Fortuin, C., Shuler, K., and Petschek, A., and Schaibly. J.: Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory, J. Chem. Phys., 59(8), 3873–3878, 1973.

Devia, G. K., Ganasri, B. P., and Dwarakish, G. S.: A review on hydrological models. Aquatic Procedia, 4, 1001–1007, 2015.

Dinku, T., Ceccato, P., Kopec, E. G., Lemma, M., Connor, S. J., and Ropelewski, C. F.: Validation of satellite rainfall products over East Africa's complex topography, Int. J. Remote Sens., 28(7), 1503–1526, 2007.

Duan, Q. Y., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, J. Optimiz. Theory App., 76(3), 501–521, 1993.

Hao, G., Li, J., Song, L., Li, H., and Li, Z.: Comparison between the TOPMODEL and the Xin'anjiang model and their application to rainfall runoff simulation in semi-humid regions, Environ. Earth Sci., 77(7), 279, 2018.

Jayatilaka, C., Storm, B., and Mudgway, L.: Simulation of water flow on irrigation bay scale with MIKE-SHE, J. Hydrol., 208, 108-130, 1998.

Jiang, Y., Liu, C., Li, X., Liu, L., and Wang, H.: Rainfall-runoff modeling, parameter estimation and sensitivity analysis in a semiarid catchment, Environ. Modell. Softw., 67, 72–88, 2015.

Khomsi, K., Mahe, G., Tramblay, Y., Sinan, M., and Snoussi, M.: Regional impacts of global change: seasonal trends in extreme rainfall, run-off and temperature in two contrasting regions of Morocco, Nat. Hazards Earth Syst. Sci., 16(5), 1079–1090, 2016.

Khorashadi, Z. F., Nossent, J., Sarrazin F., Pianosi, F., Griensven, V. A., Wagener, T. and Bauwens, W.: Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model, Environ. Modell. Softw., 91, 210–222, 2017.

Krzysztofowicz, R.: Bayesian theory of probabilistic forecasting via deterministic hydrologic model, Water Resour. Res., 35(9), 2739–2750, 1999.

Li, B., Liang, Z., He, Y., Hu, L., Zhao, W. and Acharya K.: Comparison of parameter uncertainty analysis techniques for a TOPMODEL application, Stoch. Env. Res. Risk A., 31(5), 1045–1059, 2017.

Li, B., Liang, Z., Bao, Z., Wang, J., and Hu, Y.: Changes in streamflow and sediment for a planned large reservoir in the middle Yellow River, Land Degrad. Dev., 30(7), 878–893, https://doi.org/10.1002/ldr.3274, 2019.

Li, B., Yu., Z., Liang., Z., Song. K., Li, H., Wang, Y., Zhang, W. and Acharya. K.: Effects of Climate Variations and Human Activities on Runoff in the Zoige Alpine Wetland in the Eastern Edge of the Tibetan Plateau, J. Hydrol. Eng., 19(5), 1026–1035, 2014.

Li, D.: Hydrologic model: the vertically mixed runoff model (vmm), HydroShare, https://doi.org/10.4211/hs.c523228 7d5c04bfb8cac5ce4e391ea 0f, 2018.

Li, D., Liang, Z., Li, B., Lei X., and Zhou Y.: Multi-objective calibration of MIKE SHE with SMAP soil moisture datasets, Hydrol. Res., 50(2), 644–654, https://doi.org/10.2166/nh.2018.110, 2018.

Li, X. and Huang, C. C.: Holocene palaeoflood events recorded by slackwater deposits along the Jin-shan Gorges of the middle Yellow River, China, Quatern. Int., 453, 85–95, 2017.

Li, Z. J., and Zhang, K.: Comparison of three GIS-based hydrological models, J. Hydrol. Eng., 13, 364–370, 2008.

Lü, H., Hou, T., Horton, R., Zhu, Y., Chen, X., Jia, Y., Wang, W. and Fu, X.: The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method, J. Hydrol., 480, 102–114, 2013.

McIntyre, N., and Al-Qurashi, A.: Performance of ten rainfall–runoff models applied to an arid catchment in Oman, Environ. Modell. Softw., 24, 726–738, 2009.

McMichael, C. E., Hope, A. S., and Loaiciga, H. A.: Distributed hydrological modelling in California semi-arid shrublands: MIKE SHE model calibration and uncertainty estimation, J. Hydrol., 317, 307–324, 2006.

Michaud, J., and Sorooshian, S.: Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed, Water Resour. Res., 30, 593–605, 1994.

Morris, M. D.: Factorial sampling plans for preliminary computational experiments, Technometrics, 33, 161–174, 1991.

Mwakalila, S., Campling, P., Feyen, J., Wyseure, G. and Beven, K.: Application of a data-based mechanistic modelling (DBM) approach for predicting runoff generation in semi-arid regions, Hydrol. Process., 15, 2281–2295, 2001.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, J. Hydrol., 10, 282–290, 1970.

Pianosi, F., and Wagener, T.: A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, Environ. Modell. Softw., 67, 1–11, 2015.

Pianosi, F., and Wagener, T.: Distribution-based sensitivity analysis from a generic input-output sample, Environ. Modell. Softw., 108, 197–207, 2018.

Pilgrim, D. H., Chapman, T. G. and Doran, D. G.: Problems of rainfall-runoff modelling in arid and semiarid regions, Hydrolog.

Deleted:

Sci. J., 33, 379–400, 1988.

Rujner, H., Uuml, G., Leonhardt, N., Marsalek, J., and Viklander, M.: High-resolution modelling of the grass swale response to runoff inflows with Mike SHE, J. Hydrol., 562, 411–422, 2018.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto M.: Sensitivity Analysis in Practice. J. Am. Stat. Assoc., 101, 398–399, 1989.

Samaras, A. G., Gaeta, M. G., Moreno, M. A., and Archetti R.: High-resolution wave and hydrodynamics modelling in coastal areas: operational applications for coastal planning, decision support and assessment, Nat. Hazards Earth Syst. Sci., 16(6), 1499–1518, 2016.

Seyfried, M. S., and Wilcox, B. P.: Scale and the nature of spatial variability: Field examples having implications for hydrologic modeling, Water Resour. Res., 31, 173–184, 1995.

Sharma, K. D., and Murthy, J. S. R.: A practical approach to rainfall-runoff modelling in arid zone drainage basins, Hydrolog. Sci. J., 43(3), 331–348, 1998.

Simard, R., and Ecuyer, P. L.: Computing the two-sided Kolmogorov-Smirnov distribution, J. Stat. Softw., 39(11), 1–18, https://doi.org/10.18637/jss.v039.i11, 2011.

Sobol, I. M.: Sensitivity estimates for nonlinear mathematical models, Math. Model. Comput. Exp., 1, 407–414, 1993.

United Nations Environment Programme (UNEP): World Atlas of Desertification, Edward Arnold, London, 69 pp., 1992.

Wang, G., and Ren, L.: A Contrastive Study of Simulation Results between GWSC-VMR and Hybrid Runoff Model in Dianzi Basin, in: International Conference on Environmental Science and Information Application Technology, Wuhan, China, 4–5 July, 583–588, 2009.

Xiao, Z., Liang, Z., Li, B., Hou, B., Hu, Y., Wang J.: New flood early warning and forecasting method based on similarity theory, J. Hydrol. Eng., https://10.1061/(ASCE)HE.1943-5584.0001811.

Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M. and Stewart, A.: Understanding uncertainty in distributed flash flood forecasting for semiarid regions, Water Resour. Res., 44(5), 61–74, 2008.

Young, C. B., Nelson, B. R., Bradley, A. A., Smith, J. A., Peters-Lidard, C. D., Kruger, A. and Baeck, M.L.: An evaluation of NEXRAD precipitation estimates in complex terrain, J. Geophys. Res.-Atmos., 104, 19691–19703, 1999.

Zhao, L., Xia, J., Xu, C. Y., Wang, Z., Sobkowiak, L., and Long, C.: Evapotranspiration estimation methods in hydrological models, J. Geogr. Sci., 23, 359-369, https://doi.org/10.1007/s11442-013-1015-9, 2013.

Zhao, R. J.: Watershed Hydrological Model: Xin’anjiang Model and Shanbei Model, Water and Power Press, Beijing, China, 1983. (In Chinese)

Zhao, R. J.: The Xinanjiang model applied in China, J. Hydrol., 135, 371–381, 1992.

**Table 1. Characteristics of the four catchments**

| Catchment | Area ($km^2$) | Outlet station | Area* ($km^2$) | Mean annual precipitation (mm) | Mean evaporation (mm) | Aridity index |
|---|---|---|---|---|---|---|
| Qiushui River | 1989 | Linjiaping | 1873 | 499 | 1150 | 2.31 |
| Qingjian River | 4080 | Yanchuan | 3468 | 451 | 1080 | 2.4 |
| Tuwei River | 3294 | Gaojiachuan | 2095 | 377 | 1050 | 2.78 |
| Kuye River | 8706 | Wenjiachuan | 8645 | 410 | 1010 | 2.46 |

* The area of the catchment controlled by the outlet station indicated in the table.

**Table 2. Calibrated parameters of the VMM**

| Symbol | Definition | Range* |
|---|---|---|
| *KC* | Ratio of potential evapotranspiration to pan evaporation | [0.5, 1.5] |
| *WM* | Mean areal maximum possible soil moisture, mm | [50, 200] |
| *FC* | Stable infiltration capacity, mm/h | [5, 100] |
| *K* | Infiltration index related to soil permeability, /h | [0.05, 1] |
| *BF* | Index of the watershed infiltration capacity curve | [0, 0.5] |
| *B* | Index of the watershed water storage capacity curve | [1, 2] |
| *KI* | Outflow coefficient of interflow, d | [0.1, 0.5] |
| *KG* | Outflow coefficient of groundwater, d | [0.5, 2] |
| *CS* | Confluence coefficient of surface flow | [0.05, 0.9] |
| *CI* | Recession coefficient of interflow, d | [0.5, 0.95] |
| *CG* | Recession coefficient of groundflow, d | [0.90, 0.99] |
| *KE* | Residence time of Muskingum, h | [0.5, 5] |
| *XE* | Muskingum coefficient | [0.01, 0.49] |
| *IM* | Impermeable area | [0, 1] |

*In [a, b], a and b represent the lower and upper bounds of the parameters, respectively.

**Table 3. Performance (in terms of absolute relative error) for peak flow $E_v$ in each catchment in the four models**

Unit: %

| | Qiushui River | Qingjian River | Tuwei River | Kuye River | Average* |
|---|---|---|---|---|---|
| VMM | 26.52 | 58.50 | 40.20 | 30.80 | 39.01 |
| MIKE SHE | 40.50 | 60.70 | 45.30 | 38.20 | 46.18 |
| XAJ | 56.60 | 66.61 | 60.20 | 52.30 | 58.93 |
| SBM | 38.14 | 55.82 | 35.50 | 45.2 | 43.15 |

*The average $E_v$ of the four catchments for each model

**Table 4. Lag time of the peak flow in the four catchments in the four models** Unit: h

| | Qiushui River | Qingjian River | Tuwei River | Kuye River | Average* |
|---|---|---|---|---|---|
| VMM | 2.20 | 3.02 | 3.46 | 3.50 | 3.05 |
| MIKE SHE | 2.50 | 3.50 | 4.20 | 3.90 | 3.53 |
| XAJ | 4.10 | 3.81 | 5.62 | 4.50 | 4.51 |
| SBM | 4.00 | 2.95 | 3.46 | 4.20 | 3.65 |

*The average lag time in the four catchments for each model

Formatted: Line spacing: single

**Table 5. The percentage of modeled peak flows that meets various conditions for the VMM model**

Unit: %

| | C1 | C2 | C3 | C* | C** |
|---|---|---|---|---|---|
| Qiushui River | 40.00 | 75.00 | 95.00 | 75.00 | 75.00 |
| Qingjian River | 41.38 | 44.83 | 100.00 | 58.62 | 58.62 |
| Tuwei River | 47.83 | 69.57 | 100.00 | 69.57 | 69.57 |
| Kuye River | 35.71 | 64.29 | 100.00 | 67.86 | 67.86 |

C* represent the modeled peak flows that meet one of the conditions C1 and C2; this means modeled and observed peak flows are the same type

C** represent the modeled peak flows that meet the conditions C* and C3

Deleted: Table 5. The percentage of modeled peak flows that meets various conditions for the VMM model Unit: %¶

Formatted: Font: (Default) Times New Roman, 10 pt

Formatted: Font: (Default) Times New Roman, 10 pt

Formatted: Font: (Default) Times New Roman, 10 pt

Formatted: Font: (Default) Times New Roman, 10 pt

Formatted: Font: (Default) Times New Roman, 10 pt

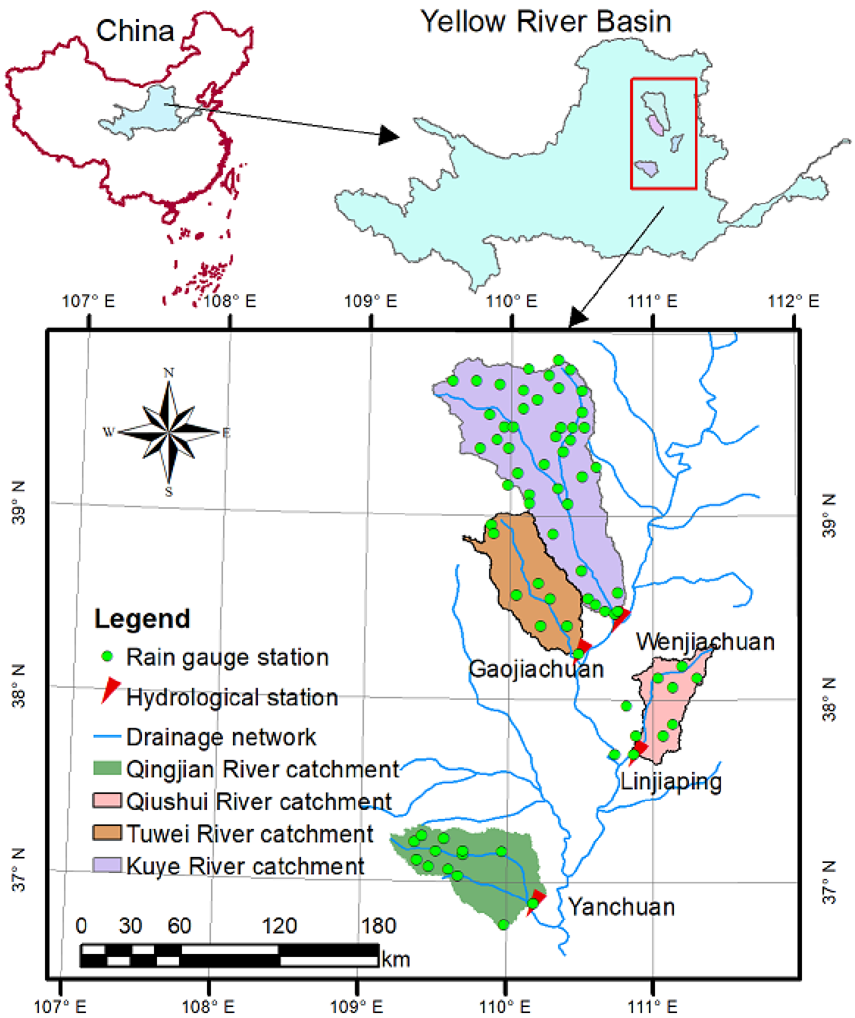


**Figure 1: Location of the Qingjian River catchment, Qiushui River catchment, Tuwei River catchment and Kuye River catchment.**

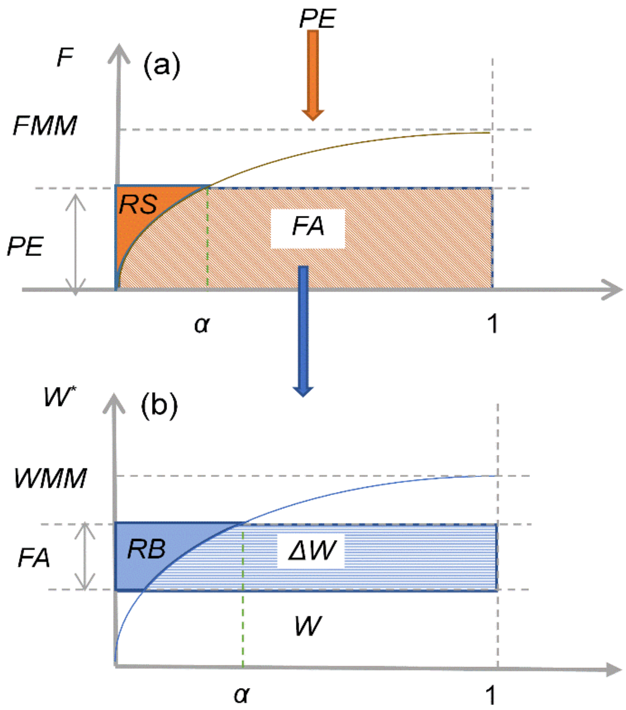


**Figure 2: Runoff generation module in the VMM model. (a) Infiltration capacity curve; and (b) tension water content storage capacity curve. $\alpha$ is the fracture area that is saturated and $F$ represents the point infiltration capacity.**

Deleted: ,

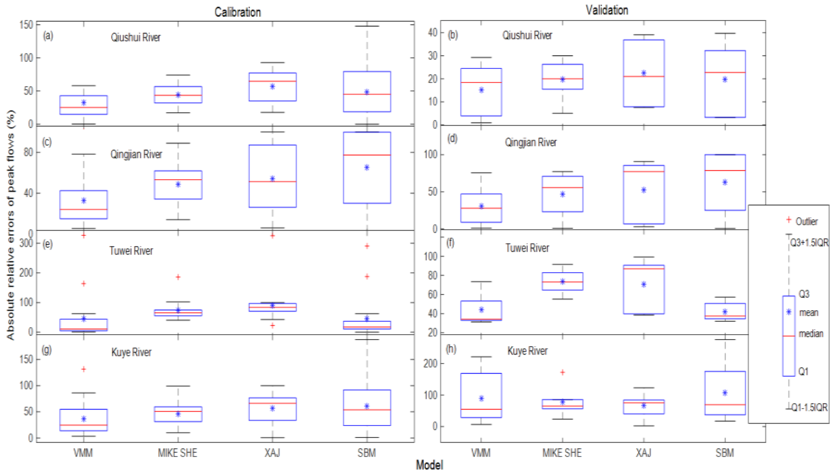


**Figure 3: Boxplot of absolute relative errors of peak flows in the four catchments, Q1 and Q3 represent the first quantile and third quantile, respectively; interquartile range (IQR) = Q3 − Q1, and an outlier is defined as an extreme value that exceeds the IQR.**

**Deleted:** ;

**Deleted:** mean

**Deleted:** ;

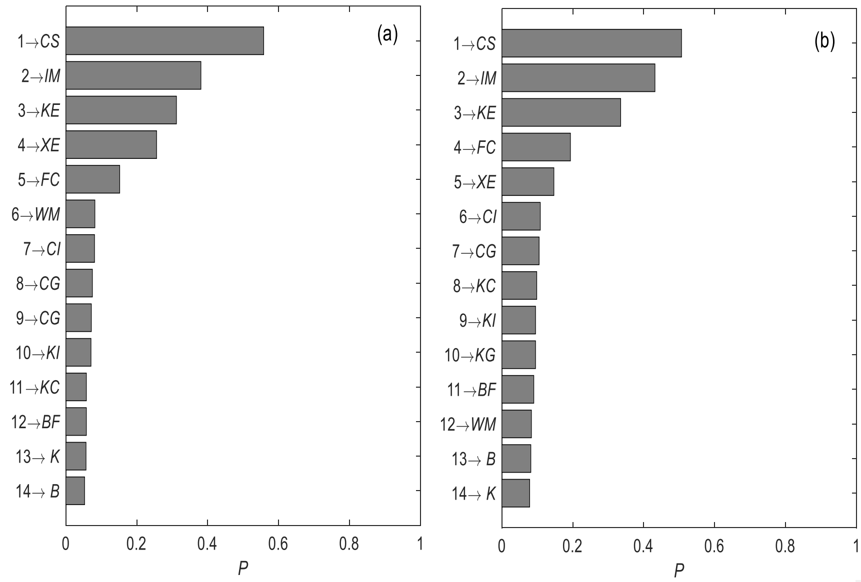


**Figure 4: Sensitivity rankings of the VMM parameters based on the global sensitivity analysis method PAWN for different objective functions. (a) $E_{pv}$ as the objective function and (b) $E_p$ as the objective function. The value $P$ is used to assess the sensitivity degree of the parameter with the PAWN method, and a larger value corresponds to greater sensitivity. The numbers on the ordinate represent the sensitivity rankings.**

Deleted: :

Deleted: ,

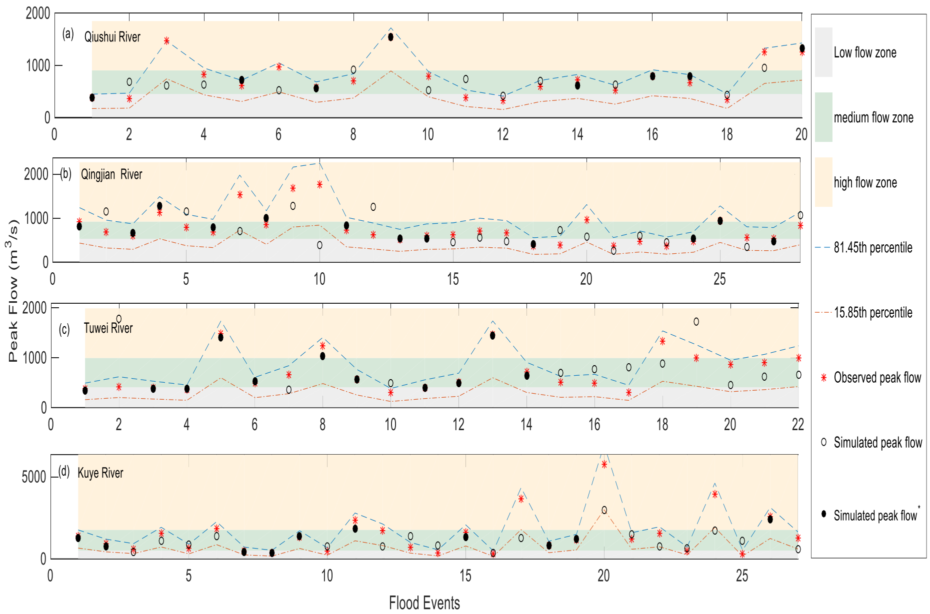


**Figure 5: Observed peak flows (red asterisk) and simulated peak flows (solid ball and circle) with the VMM model for each catchment under conditions C1, C2 and C3. Flood peaks conforming to condition C1 are represented by solid balls; the three flow zones (low, medium and high flow zones) classified by condition C2 are shown in gray, green and off-white, respectively; the 68.3% uncertainty interval of peak flows estimated by condition C3 is shown between the blue dashed line (81.45th percentile) and the red dashed-dotted line (15.85th percentile).**

**Formatted:** Line spacing: single

**Deleted:** the

**Deleted:** the

**Deleted:** is

**Deleted:** a

**Deleted:** divided

**Deleted:** the

**Deleted:** confidence

**Deleted:** the