Event and model dependent rainfall adjustments to improve discharge predictions

Diana Fuentes Andino, Keith Beven, Anna Kauffeldt, Chong-Yu Xu, Sven Halldin & Giuliano Di Baldassarre

Event and model dependent rainfall adjustments to improve discharge predictions

Diana Fuentes Andino[1,2]

Keith Beven[1,3]

Anna Kauffeldt[1]

Chong-Yu Xu[1,4]

Sven Halldin[1,2]

Giuliano Di Baldassarre[1]

[1] *Department of Earth Sciences, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden*

[2] *Centre for Natural Disaster Science (CNDS), Sweden.*

[3]*Lancaster Environment Center, Lancaster University, Lancaster LA1 4YQ, UK*

[4]*Department of Geosciences, University of Oslo, P O Box 1047, Blindern, NO-0316, Oslo, Norway*

Corresponding author:

Diana Fuentes Andino
Uppsala University
Dept. of Earth Sciences
Villavägen 16, 752 36 Uppsala, Sweden
diana.fuentes@geo.uu.se

Event and model dependent rainfall adjustments to improve discharge predictions

Abstract Most conceptual rainfall–runoff models use as input spatially averaged rainfall fields which are typically associated with significant errors that affect the model outcome. In this study, it is hypothesized that a simple spatially and temporally averaged event–dependent rainfall multiplier can account for errors in the rainfall input. The potentials and limitations of this lumped multiplier approach are explored by evaluating the effects of multipliers on the accuracy and precision of the predictive distributions. Parameter sets found to be behavioural across a range of different flood events were assumed to be a good representation of the catchment dynamics and were used to identify rainfall multipliers for each of the individual events. An effect of the parameter sets on identified multipliers was found, however it was small compared to the differences between events. Accounting for event–dependent multipliers improved the reliability of the predictions. At the cost of a small decrease in precision, the distribution of identified multipliers for past events can be used to account for possible rainfall errors when predicting future events. By using behavioural parameter sets to identify rainfall multipliers, the method offers a simple and computationally efficient way to address rainfall errors in hydrological modelling.

Key words rainfall multiplier; rainfall input error; reliability of the predictions, precision of predictions, Topmodel, floods.

1 INTRODUCTION

Hydrological models are useful tools to guide decision–making on problems related to e.g. water resource planning, flood forecasting, hydrological design, and impact study of climate change and land–use change, etc. (Kizza *et al.* 2011, Beven 2012, Hailegeorgis and Alfredsen 2015, Emam *et al.* 2016, Jie *et al.* 2016, Zeng *et al.* 2016). Decisions can be facilitated by the use of relatively simple models, such as conceptual rainfall–runoff (CRR) models, that mathematically relate the rainfall upon a catchment to the streamflow at the catchment outlet (Jakeman and Hornberger 1993). However, calibration of CRR models is affected by errors coming from input, evaluation data, model structure and an inability to identify the exact values of model parameters (Beven and Binley 1992, Beven 2009, Montanari and Di Baldassarre 2013). Errors associated with the precipitation inputs can have a major effect on model performance (Kavetski *et al.* 2006a, Moulin *et al.* 2009, Beven *et al.* 2011, Beven and Smith 2015) and are especially common when there are few rain gauges to capture the variability across the catchment (Kutiel and Kay 1996, Michaelides *et al.* 2009, McMillan *et al.* 2011, Xu *et al.* 2013, 2015, Kang 2014, Shafiei *et al.* 2014).

Multipliers to account for errors on input rainfall have been used as an extra calibration parameter in many CRR models ever since the original Stanford Watershed Model (Linsley and Crawford 1966). McMillan *et al.* (2011) show that a multiplicative error is suitable for correcting mean catchment rainfall values for rainfall rates larger than 1 mm h^{-1}. Rainfall multipliers have also been used in Bayesian total error analysis (BATEA) introduced by Kavetski *et al.* (2003) and further developed by Kavetski *et al.* (2006a, 2006b), Kuczera *et al.* (2006), Renard *et al.* (2008, 2010, 2011) and Thyer *et al.* (2009). BATEA offers the possibility to explicitly account for input rainfall and output observational data within the calibration procedure as latent variables additional to the model parameters. In this approach errors are being accounted for as stochastic multiplier variables which follow an assumed distribution. Evaluation of the model results with all available evidence makes it possible to retain likely combinations of model parameters with latent variables representing different sources of errors.

Results from BATEA show that when compared with other sources of uncertainty e.g. output observational and model structure (Kuczera *et al.* 2006, Renard *et al.* 2010) or only output

observational (Thyer *et al.* 2009), uncertainties in rainfall are larger. Yet it should be taken into account that within this approach it is impossible to separate fully the effects of input error from other sources of error such that event by event input multipliers will compensate for model structural error, or output observational error for particular events (see for example, Beven 2005). Results from using storm dependent multipliers (Kuczera *et al.* 2006, Thyer *et al.* 2009), a multiplier per daily step (Thyer *et al.* 2009, Renard *et al.* 2010, 2011) and monthly time step (Huard and Mailhot, 2008) lead to improvements in reliability of predictive distributions. This indicates that CRR models benefit from explicit consideration of rainfall errors (and other sources of uncertainties) but at the cost of more intensive computations (Kavetski *et al.* 2006a).

The aim of this paper is to develop a simple and computationally inexpensive method to account for rainfall errors for sporadic large flood events. It was hypothesised that event–dependent spatially and temporally averaged rainfall depth multipliers could be used for this purpose. And behavioural parameter sets identified within the Generalised Likelihood Uncertainty Estimation (GLUE) methodology can help to identify such multipliers, thus avoiding the computational cost from combining model parameter sets with multipliers and making assumption on prior distribution of the errors. The appropriateness of the method was evaluated based on predictive distributions in terms of accuracy and precision.

2 METHOD

The method was developed within a cross–validation context, i.e. from a total of *N* flood events, *N*–1 were used for calibration leaving one event out at a time for validation. A flowchart of the methodology following the context of *N*-1 calibration events and one validation event is schematized in Fig. 1 and explained in detailed in what follows in this method section. The method is being applied to a data-sparse situation in China, with only a limited number of observed flood events available to allow for model calibration. The method is, however, essentially within the Bayesian paradigm so that as additional events are observed, the posterior estimates of parameter distributions and storm multipliers can be updated.

2.1 Model setup

Multiple parameter sets were generated and each was used with spatially averaged rainfall to simulate *N* flood events using a CRR model. Parameter sets were generated by randomly sampling from assumed prior distributions. Sampling ranges should cover all possible values that the model parameters could take.

2.2 Model evaluation

The widely used Nash-Sutcliffe model efficiency coefficient (NSE) was adopted to evaluate model outputs on all the evaluation points for each event. The more recent GLUE limits of acceptability approach to model evaluation (e.g. Blazkova and Beven 2009; Liu *et al.* 2009) was not used here because no information was available about the discharge uncertainties. Behavioural parameter sets were defined as those for which the weighted score (*WNSE*) of the *N*–1 calibration events was larger than a specified threshold. The *WNSE* was calculated as the average of the NSE values corresponding to the calibration events, weighted by the number of evaluation points of each event (w_i).

$$WNSE = \frac{\sum_{i=1}^{i=N-1} w_i \times NSE_i}{\sum_{i=1}^{i=N-1} w_i} \qquad (1)$$

It is expected that calibrating parameter sets against several different events, each associated with different input errors, will reduce bias in the selection of parameter sets introduced by those errors (Beven 2015, Jie *et al.* 2016). Thus, behavioural parameter sets were identified based on model

3

parameters only, those identified sets were then used to identify multipliers associated with input errors.

The number of simulations was set by using a stopping criteria proposed by Pappenberger et al., (2005). For every 200 behavioural simulations added, an hourly cumulative distribution function (CDF) of the predicted discharge was estimated. This estimated CDF was compared with the previous one using the Kolmogorov-Smirnov two-sample test, the number of runs was considered sufficient when for all hours the addition of behavioural simulations did not change the CDF significantly (i.e. P < 0.05).

2.3 Errors from spatial averaging of rain gauge measurements

Relative errors from spatial averaging of rain gauge measurements were inferred at each rain gauge station by comparing its measured rainfall values for all calibration events with those estimated by averaging the correspondent measured values at the remaining gauge stations. A selected quantile range of the joint relative errors from all rain gauge stations can be used as a sampling range of the relative error from spatial averaging of rainfall at any point in the catchment. This approach to generalize the error across the whole catchment can be reasonable if gauges are representative of different altitudes within the catchment. Furthermore, to consider a variety of events can compensate the lack of more rain gauges, as one could expect the spatial distribution of precipitation to vary widely from event to event in mountainous terrains.

2.4 Event dependent rainfall multipliers

Mismatches between observed and simulated discharge are expected to result from a combination of errors, mainly in precipitation input, output observational data and model structure. Different sources of errors were accounted for by the following assumptions:

- Input errors coming from spatial averaging of rainfall are assumed to be adequately characterised by an event–dependent, spatially and temporally averaged rainfall multiplier as in Kavetski *et al.* (2006a).
- Errors in discharge observations are assumed to be independent and Gaussian distributed.
- Model structural errors are not expected to be large and assumed to be treated implicitly in the covariation of model parameters.

Possible rainfall multipliers were sampled from a vague prior uniform distribution with limits based on quantiles of the relative errors found from spatial averaging of rain gauge measurements, section 2.3. Sampled rainfall multipliers were combined with each of the behavioural parameter sets to simulate all flood events. For each event and parameter set, the multiplier that resulted in the best fit to the observations (based on the NSE score) was identified, resulting in a number of multipliers equal to the number of behavioural sets for each calibration event.

To test the event dependency of identified multipliers, the Student's t-test was used with the null hypothesis that two groups of multipliers are from the same population with equal means, without assuming that the populations also have equal variances. If the hypothesis is not rejected then two groups of multipliers are statistically similar, thus the multipliers are not event dependent or the two events have similar average rainfall errors. However, it only takes one rejection of the hypothesis to confirm that two groups of multipliers come from different populations and, hence, prove event dependence.

4

2.5   The effect of parameter set on identified multipliers

Parameter sets consistently associated with high (or low) multipliers (relative to the average of the identified best multipliers of an event) across events are a result of an effect of parameter sets on identified multipliers. Then, the multiplier is, at least partly, interacting with the dynamics of the parameter set to improve model performance. Identified multipliers were tested for such effects by searching for a systematic pattern of identified multipliers across calibration events. For each event, multipliers were standardized to have zero mean and unit variance by subtracting their average and dividing them by the standard deviation. An average standardized multiplier for each behavioural parameter set was estimated by averaging N–1 standardized multipliers, leaving one event for validation as before. A systematic pattern was confirmed if the average standardized multipliers were highly correlated with the standardized multipliers for the validation event. If such a pattern is confirmed, the effect of the parameter set on identified multiplier can be reduced at each calibration event by subtracting a value equal to the parameter set average of standardized multipliers times the standard deviation of the multipliers for that calibration event.

2.6   Effect of rainfall multipliers on predictive distributions at calibration and validation

Allowing identified event–dependent multipliers to account for errors on rainfall input, the predictive distribution composed of behavioural parameter sets was evaluated for both calibration and validation events and compared with the case of not using multipliers. The predictive distributions were evaluated considering reliability and precision as in Laio and Tamea (2007), Thyer *et al.* (2009) and Renard *et al.* (2010).

In calibration, multipliers were used to account for event input errors using the average of the multipliers identified for the event, whereas in validation the multiplier is unknown and the identified multipliers for the N–1 calibration events were used to infer likely multipliers. To this end, the distributions of multipliers from the calibration events were joined and fitted to a theoretical distribution used to sample multipliers with equally probable intervals (to have a relatively small number of samples but all along the possible range). All sampled multipliers were combined with each of the behavioural parameter sets for prediction.

2.6.1 Reliability, a predictive distribution (PD) is reliable if it is statistically consistent with the observations, i.e. if the observations are realisations of the PD. In this paper, the reliability of the PD is evaluated using the predictive QQ-plot (Laio and Tamea 2007, Thyer *et al.* 2009, Renard *et al.* 2010), which provides visual clues to the statistical consistency between the observed discharge and the PD. Assuming an observation of discharge at time t, x_t, is a realisation from the PD with cumulative distribution function F_t, then the cumulative probability P-value, $F_t(x_t) = probability(X_t \leq x_t)$, is a realisation from a uniform distribution on [0,1]. Thus, for t varying from 1 to L observations in an event, the series of L P-values, one for each observation, will also be a realisation from a uniform distribution. The predictive QQ-plot is constructed by plotting the quantiles i.e. P-values against the corresponding theoretical quantiles of a uniform distribution on [0, 1]. The closer the points fall to the bisector (1:1 line), the better the agreement of the predictive distribution with the observations, and with all points falling on the line indicating a perfect agreement (Fig. 2). Deviations from the bisector indicate issues with prediction bias and predictive uncertainty. For example, if at the theoretical median, P-values are higher (lower) than the corresponding theoretical quantiles, the model systematically under–predicts (over–predicts) the observed data. A steep (flat) slope of the curve in the midrange (around theoretical quantiles 0.4–0.6) indicates an underestimated (overestimated) predictive uncertainty.

5

2.6.2 Precision, also called resolution in the literature (Renard *et al.* 2010), measures the sharpness of the PD. Two predictive distributions could give reliable estimates but have different precisions. The relative precision (P) of the predictors can be quantified as the average of the relative precision of the predictions X_t over all the observations (L), or could be looked at as the average of the inverse coefficient of variation:

$$P = \frac{1}{L}\sum_{t=1}^{L}\frac{E[X_t]}{Sdev[X_t]} \qquad (2)$$

where $E[\ ]$ and $Sdev[\ ]$ are the expectation and standard deviation operators.

3   APPLICATION

3.1   Study area and data

The study area is the 259 km^2 Jiaokou Reservoir watershed, a sub-basin of the Yongjiang River basin located in Zhejiang province, south–eastern part of China (Fig. 3). The watershed characteristics and data have previously been described in Du *et al.* (2007, 2009). The watershed elevation ranges from 50 to 975 m. The largest part of the catchment is composed of forest (78%) and the remaining area is a mix of agriculture (14.5%), grassland (2.5%), water surface (2.7%) and residential areas (1.9%). Most of the catchment is composed of loam and clay, mainly poorly drained soil with high runoff generation potential. The region has a typical subtropical monsoon climate, with an annual average temperature of 16.3°C and minimum and maximum temperatures of -11.1°C and 39.5°C occurring in January and July, respectively. The mean annual precipitation is about 2000 mm with most of the rainfall occurring between March and September.

In the catchment, there are three stations that measure precipitation and one river flow gauging station (Fig. 3), all of which measure hourly data. Discharge measurements in the catchment were conducted during large flood events only. All the available six isolated large flood events measured from the period 1979 to 2000 (Fig. 4) were used in this work. A 30 m digital elevation model from a digitalized contour map of 1:50 000 scale was generated by Du *et al.* (2007).

3.2   Model description

TOPMODEL is a semi–distributed CRR model originally proposed by Beven and Kirby (1979). TOPMODEL takes into account the dynamics of catchment saturation by means of a time–variable saturated contributing area. Nevertheless, TOPMODEL retains the simplicity of a lumped model for representing the flow through the soil profile. Since its original development, TOPMODEL has been modified and adapted to the specific needs of different modelling applications; thus, instead of a model, it is better known as a collection of concepts (Beven 2012). In the present version of TOPMODEL, the network width function, NWF, concept (Surkan 1969, Kirkby 1976), with further adaptations by Grimaldi *et al.* (2010) was used to route the surface runoff. The main idea behind the NWF is that discharge at the outlet is influenced by the shape of the stream network. Thus, a histogram function is built based on the shape of the flow path distances along the catchment contributing area and the river network. A comprehensive description of the model is given in Appendix A.

3.3   Model set up

By Monte Carlo sampling from the uniform distribution, with sampling ranges as defined in Table 1, 10 000 combinations of model parameters were generated. The initial deficit of the model was estimated by inverting equation (A5) in Appendix A based on measured discharge at the beginning of

6

each event. The root zone storage was considered constant and was estimated based on the characteristics of the catchment. Spatially averaged rainfall in conjunction with the generated parameter sets were used to simulate the six flood events. All six different combinations of five–calibration–events and one validation event per each combination were analysed. Behavioural sets were defined as those sets for which the combined NSE score (*WNSE*, Equation 1) for the five calibration events was larger than *0.90*.

4   RESULTS

4.1   Errors from spatial averaging of rain gauge measurements

No substantial difference was found on the statistics of the relative error for all different calibration events combinations (Fig. 5). Given the small difference between the relative error distributions, the values of the relative errors corresponding to the 20th and 80th quantiles, i.e. containing 60% of the errors, were similar across combinations. Lower estimates of the multiplier (i.e. relative error plus one) ranged from 0.55 to 0.60 and upper estimates from 1.45 to 1.55.

4.2   Event dependent multipliers and the effect of parameter set on identified multipliers

The number of behavioural parameter sets for which the CDF of the predictions became stable ranged from 1439 to 5127 for the different combinations of five calibration events. The initial sampling of 10 000 parameter sets was enough to find enough behavioural parameter sets for all these combinations. To reduce computational cost, 480 parameter sets that were behavioural for all the combinations and for which the CDF for all predicted hours was similar (i.e. $P < 0.05$) to those estimated using all behavioural simulations for each combination were identified. For each combination, it was ensured that those 480 sets were representative of the whole range of five–calibration–events weighted NSE score (*WNSE*). For each calibration event, 120 multipliers, sampled from the range found in section 4.1, were combined with each of the 480 behavioural sets. A plot of the multipliers against NSE score resulted in convex sets with one peak for all the cases, confirming that best multipliers were always within the multipliers sampling range.

Some parameter sets were consistently associated with high multipliers (relative to the average of the identified best multipliers of an event) across events whereas others were consistently associated with low multipliers. This was confirmed by high correlation coefficients found between the average of standardized multipliers for the calibration events and the standardized multipliers of each validation event (0.76, 0.89, 0.96, 0.82, 0.70 and 0.74 for validation events 1 to 6 respectively). Accounting for the effect of parameter sets on the multipliers leads to a reduction in the noise of identified best multipliers (Fig. 6). The average values of the multipliers associated with each event after correction for multipliers were 1.24, 0.96, 1.08, 0.96, 0.97 and 0.95 for event 1 to 6 respectively.

Differences of the spread in values of identified multipliers between events were even more noticeable, after correction for multipliers (Fig. 6). Results from the Student's t-test estimated at a 5% significance level showed that events 2 and 4 were samples from the same populations whereas 1, 3, 5 and 6 were from different. Thus four events showed event dependency of identified multipliers.

Correction for the effect of parameter set on multiplier can also be considered at prediction, however, the event standard deviation is unknown and the smallest standard deviation from all the calibration events was used as it improved performance for all calibration events. Improvements will not occur at validation events with smaller deviation than the smallest observed at calibration, thus there will be benefits from having more calibration events with different characteristics.

7

4.3 Effect of rainfall multipliers on predictive distribution at calibration

Predictive ranges for one calibration event obtained from different combination of calibration events were similar. This was because across combinations, the same behavioural parameter sets were used and the different weighted scores to each parameter set for different combinations did not show significant effect on the predictive ranges. Thus in what follows, when referring to a predictive range at a calibration event, it applies to any of the combinations of calibration events.

The use of multipliers increased the proportion of observations inside the predicted range for events 1, 3, 5 and 6. The increase was larger for event 1 which had the largest error (Table 2). For events 1 and 3, the use of multipliers clearly made predictions closer to the bisector 1:1 line in the QQ-plot (Fig. 7). However, for events 2, 4, 5 and 6 there was no general improvement on reliability, and predictions were close to the bisector only at certain sections of the QQ-plot.

4.4 Effect of rainfall multipliers on predictive distribution at validation

The posterior distribution of the event multipliers suggested that they could be usefully treated as fuzzy variables (Fig. 8). Thus the minimum, median and maximum values of the combined distribution were set to correspond to membership values of zero, one and zero respectively of a triangular fuzzy measure. The exclusion of event 1 generated a narrower range than the other cases which had more similar distributions. Thus the distributions of 50 sampled likely multipliers for validation events varied depending on the combination of calibration events. Predictions from a combination of all the sampled multipliers with each of the behavioural parameter sets resulted in a general better fit to the bisector 1:1 line in the QQ-plot for all validation events (Fig. 9). Since, for validation, a distribution of multipliers was considered, the precision of the distribution decreased more when compared to calibration where only one multiplier per set was used (Table 2). However, the proportion of observations inside the predicting range increased considerably by using multipliers.

Including events with large errors at calibration increases considerably the observations within the predictive range for that calibration event (Fig. 10a) as oppose to events with small error (Fig. 10b). When predicting events with errors larger than those observed at calibration, the range of estimated likely multipliers is not wide enough to encompass the error (Fig. 10c). On the contrary, when considering calibration events with large errors, estimated likely multipliers are wide–ranging thus increasing greatly observation inside the predicting range (Fig. 10d).

5 DISCUSSION

The method used here depends on identifying an adequate ensemble of behavioural parameter sets that can then be used to obtain estimates for the event by event rainfall multipliers. Given the data-sparse situation in the study catchment, this clearly raises issues about whether the available data are adequate to provide a robust calibration. The methodology somewhat circumvents this issue since the ensemble of behavioural parameter sets are all consistent with the calibration data and are never interpreted in terms of the marginal parameter distributions. It is the sets of parameters that are used in predicting the next event (a prediction that is still required even given the limited observations). The methodology is also Bayesian in concept, since new data are readily incorporated into the analysis as they become available, ready to predict the next future event. Thus, estimates of distributions of model parameters and multipliers for prediction can be used with few available events but should be part of a process in which events should be added when they become available to update the distributions. The larger the number of events considered at calibration, the more stable and representative the predictive ensemble of model parameter sets and multipliers could be expected to be, assuming the forcing data at prediction to be similar to that used in calibration (Kavetski *et al.* 2006a).

8

In the results presented in this paper, the hypothesis that event identified rainfall multipliers came from the same population was rejected for four events and the results therefore supports the assumption that the multipliers are event–dependent and associated with errors on rainfall input. Since the NSE criterion used for model evaluation is sensitive to extreme values (Hauduc *et al.* 2015), identification of multipliers was expected to be biased towards flood peaks. In other words, multipliers are biased towards compensating for errors during the more intensive time of the storm but this is acceptable as the emphasis is put on the simulation of peaks. Furthermore, multipliers associated with high rainfall intensities are generally better identified (Renard *et al.* 2010) as opposed to those associated with low rainfall depths, which might be not representative of the rainfall error but noise (McMillan *et al.* 2011). The larger average value of identified multipliers was 1.24 for event 1, thus accounting for uncertainty larger than $\pm 24\%$ in rainfall would encompass the error on rainfall input found at any of the evaluated events.

The range of the identified multipliers for each event was substantially narrower than the initially assumed sampling ranges. This is not surprising as the sampling range was estimated based on errors at three particular points in the catchment, while the identified multipliers represented the lumped errors over the catchment. Hence, the method of inferring spatial averaging error from rain gauge data gives a reasonable indication of the range of variation of multipliers when no better source of information is available. Assumed vague priors of the multipliers with wide ranges can be an option to avoid the problem of model given wrong inferences when wrong priors are assumed (Kavetski *et al.* 2006b, Kuczera *et al.* 2006, Renard *et al.* 2010, McMillan *et al.* 2011). It was confirmed that parameter sets influenced the identified multipliers, but to a much lesser degree than the input errors for the different events. The effect of parameter sets on multipliers might be dependent on the CRR model used, but it is worth considering this effect when obtaining the marginal distribution of multipliers (Kavetski *et al.* 2006a), as it can reduce bias produced by the effect of parameter sets in the multipliers. At prediction, if the event–dependent standard deviation of the identified multipliers could be explained by some characteristics of the event, e.g. magnitude or type of storm, then the noise due to the effect of parameter sets could be further reduced.

In calibration, accounting for input errors on event 2, 4, 5 and 6 did not result in a general improvement of reliability of the predictive distribution, at some cases lead to increase in under–prediction (events 5 and 6). Poor fit of the QQ-plot after correction of input errors could be due to the simplistic approach used to consider those errors or due to errors on model structural and output observational data not considered in this study. However, predictive distributions at events 1 and 3 were more statistically consistent at calibration when using multipliers, i.e. issues with under–prediction were decreased. Strong improvements of reliability can be obtained when substantial input errors are considered (event 1).

In validation, the use of multipliers leads to improvement in reliability, however precision decreased for all events, which was expected due to the consideration of a distribution of multipliers. Though, the decrease in precision can be seen as secondary compared to the strong benefit of encompassing more of the observations with the predictive range.

Inferred event–dependent, spatially and temporally averaged rainfall multipliers can be useful to correct errors on rainfall input, improvement of their identification was possible by correcting for the dependence on specific model parameter sets. Once improvement in reliability after correction for multipliers is evidenced, the remnant errors, after correction of the input, should be more representative of model structural errors, and, hence, a better representation of model deficiencies, which can facilitate model improvements. During predictions, identified events–dependent multipliers at calibration could be used to infer distribution of multipliers likely to occur in future events.

Huard and Mailhot (2008) and Renard *et al.* (2010, 2011) have discussed the dependency of posterior distributions of multipliers on model structural errors. Such a dependency is expected as all

9

sources of errors interact among each other during calibration influencing model outcome (Beven 2005, 2006, 2009). Thus, the multipliers identified here will, to a certain degree, be dependent of the model used. Therefore, comparison of this method using different model structures should be done to become confident in generalization of the obtained results. Research to differentiate input errors from model structural errors and/or output observational errors is still on-going and seems to be possible only when precise priors of the error model are given (Renard *et al.* 2008). However, precise priors are virtually impossible to define when data are scarce and information on the uncertainties associated with available sources is limited (Li *et al.* 2011).

In this paper, errors in discharge measurements and model structure are not explicitly considered. If those errors are expected to be large in comparison to precipitation errors, proper account for them should be done.

6   CONCLUSIONS

Effects of precipitation input errors on CRR models are known to be significant, especially when models are driven by spatially averaged rainfall. Consideration of those errors is important to improve predictions and to decrease bias on the remnant errors which helps to identify better flaws on the model structure. Here a simple approach to account for those errors through event–dependent spatial and temporal averaged rainfall multipliers identified from an ensemble of behavioural parameter sets was developed and tested. The main findings were:

- Event–dependent multipliers could be identified and were found to be associated with the error on the event rainfall input.
- An effect of parameter sets on the identified multipliers was found but was inferior to the effect of rainfall errors.
- At calibration, for some of the flood events analysed, accounting for rainfall errors through multipliers improved predictive performance by an increase in accuracy and in the proportion of observation inside the predicting range.
- At validation, for all flood events, accounting for rainfall errors through multipliers improved predictive performance by an increase in accuracy and in the proportion of observation inside the predicting range at only a small cost in terms of precision.

The method considers parameter uncertainty and rainfall input uncertainty in a simpler way than e.g. BATEA, which makes it less computational demanding, but still provides a way to account for rainfall input errors, which can be used to infer on possible multipliers for prediction. This method can be used with few events and can be updated as more events become available. It is not known what effects the chosen model structure might have on the results in this paper and the methodology should be further tested with different model structures.

References

Ambroise, B., Freer, J., and Beven, K., 1996. Application of a Generalized TOPMODEL to the Small Ringelbach Catchment, Vosges, France. *Water Resources Research*, 32 (7), 2147–2159.

Beven, K., 1997. TOPMODEL: A critique. *Hydrological Processes*, 11 (9), 1069–1085.

Beven, K., 2005. On the concept of model structural error. *Water Science and Technology: A Journal of the International Association on Water Pollution Research*, 52 (6), 167–175.

Beven, K., 2006. A manifesto for the equifinality thesis. *Journal of Hydrology*, 320 (1-2), 18–36.

Beven, K., 2012. *Rainfall-Runoff Modelling: The Primer*. 2nd ed. Chichester, West Sussex ; Hoboken, NJ: John Wiley & Sons.

Beven, K., 2015. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal*, 0 (J.).

Beven, K., Gilman, K., and Newson, M., 1979. Flow and flow routing in uplands channel networks. *Hydrological Sciences Bulletin*, 24, 303–325.

Beven, K.J., 2009. *Environmental Modelling: An Uncertain Future? An introduction to techniques for uncertainty estimation in environmental prediction*. Routledge: London.

Beven, K.J. and Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological processes*, 6, 279–298.

Beven, K.J. and Kirby, M.J., 1979. A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, 24 (1), 43–69.

Beven, K., Leedal, D., McCarthy, S., Hunter, N.M., Keef, C., Bates, P.D., Neal, J., and Wicks, J., 2011. *Framework for Assessing Uncertainty in Fluvial Flood Risk Mapping*. No. FRMRC Research Report SWP1.7.

Beven, K. and Smith, P., 2015. Concepts of Information Content and Likelihood in Parameter Calibration for Hydrological Simulation Models. *Journal of Hydrologic Engineering*, 20 (1), A4014010.

Blazkova, S. and Beven, K., 2009. A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research*, 45 (12), W00B16.

Di Lazzaro, M., 2009. Regional analysis of storm hydrographs in the Rescaled Width Function framework. *Journal of Hydrology*, 373 (3–4), 352–365.

Dingman, S.L., 2002. *Physical Hydrology*. 2nd ed. Upper Saddle River, New Jersey (USA): Prentice Hall.

Duan, J. and Miller, N.L., 1997. A generalized power function for the subsurface transmissivity profile in TOPMODEL. *Water Resources Research*, 33 (11), 2559–2562.

Du, J., Xie, H., Hu, Y., Xu, Y., and Xu, C.-Y., 2009. Development and testing of a new storm runoff routing approach based on time variant spatially distributed travel time method. *Journal of Hydrology*, 369 (1–2), 44–54.

Du, J., Xie, S., Xu, Y., Xu, C., and Singh, V.P., 2007. Development and testing of a simple physically-based distributed rainfall-runoff model for storm runoff simulation in humid forested basins. *Journal of Hydrology*, 336 (3–4), 334–346.

Emam, A.R., Kappas, M., and Hosseini, S.Z., 2016. Assessing the impact of climate change on water resources, crop production and land degradation in a semi-arid river basin. *Hydrology Research*, In press, DOI: 10.2166/nh.2015.143.

Grimaldi, S., Petroselli, A., Alonso, G., and Nardi, F., 2010. Flow time estimation with spatially variable hillslope velocity in ungauged basins. *Advances in Water Resources*, 33 (10), 1216–1223.

Haan, C.T., Barfield, B.J., and Hayes, J.C., 1994. *Design Hydrology and Sedimentology for Small Catchments*. San Diego, California, USA: Academic Press.

Hailegeorgis, T. and Alfredsen, K., 2015. Comparative evaluation of performances of different conceptualizations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments. *Hydrology Research*, 46 (4), 607–628.

Hauduc, H., Neumann, M.B., Muschalla, D., Gamerith, V., Gillot, S., and Vanrolleghem, P.A., 2015. Efficiency criteria for environmental model quality assessment: A review and its application to wastewater treatment. *Environmental Modelling & Software*, 68, 196–204.

11

Huard, D. and Mailhot, A., 2008. Calibration of hydrological model GR2M using Bayesian uncertainty analysis. *Water Resources Research*, 44 (2), W02424.

Iorgulescu, I. and Musy, A., 1997. Generalization of TOPMODEL for a power law transmissivity profile. *Hydrological Processes*, 11 (9), 1353–1355.

Jakeman, A.J. and Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29 (8), 2637–2649.

Jie, M.X., Chen, H., Xu, C.-Y., Zeng, Q., and Tao, X.E., 2016. A comparative study of different objective functions to improve the flood forecasting accuracy. *Hydrology Research, in press. doi:10.2166/nh.2015.078.*

Kang, K., 2014. The effect of spatially uniform and non-uniform precipitation bias correction methods on improving NEXRAD rainfall accuracy for distributed hydrologic modeling. *Hydrology Research*, 45 (1), 23–42.

Kavetski, D., Franks, S.W., and Kuczera, G., 2003. Confronting Input Uncertainty in Environmental Modelling. *In*: Q. Duan, H.V. Gupta, S. Sorooshian, A.N. Rousseau, and R. Turcotte, eds. *Calibration of Watershed Models*. American Geophysical Union, 49–68.

Kavetski, D., Kuczera, G., and Franks, S.W., 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, 42 (3), W03407.

Kavetski, D., Kuczera, G., and Franks, S.W., 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resources Research*, 42 (3), W03408.

Kirkby, M.J., 1976. Tests of the random network model, and its application to basin hydrology. *Earth Surface Processes*, 1 (3), 197–212.

Kirkby, M.J., 1997. TOPMODEL: A personal view. *Hydrological Processes*, 11 (9), 1087–1097.

Kizza, M., Rodhe, A., Xu, C.-Y., and Ntale, H.K., 2011. Modelling catchment inflows into Lake Victoria: uncertainties in rainfall–runoff modelling for the Nzoia River. *Hydrological Sciences Journal*, 56 (7), 1210–1226.

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M., 2006. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology*, 331 (1–2), 161–177.

Kutiel, H. and Kay, P.A., 1996. Effects of network design on climatic maps of precipitation. *Climate Research*, 7 (1), 1–10.

Laio, F. and Tamea, S., 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrol. Earth Syst. Sci.*, 11 (4), 1267–1277.

Li, L., Xu, C.-Y., Xia, J., Engeland, K., and Reggiani, P., 2011. Uncertainty estimates by Bayesian method with likelihood of AR (1) plus Normal model and AR (1) plus Multi-Normal model in different time-scales hydrological models. *Journal of Hydrology*, 406 (1–2), 54–65.

Linsley, R.K. and Crawford, N., 1966. *Digital Simulation in Hydrology: Stanford Watershed Model IV*. Palo Alto, California: Department of Civil Engineering, Stanford University, No. 39.

Liu, Y., Freer, J., Beven, K., and Matgen, P., 2009. Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error. *Journal of Hydrology*, 367 (1–2), 93–103.

McDonnell, J.J. and Beven, K., 2014. Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. *Water Resources Research*, 50 (6), 5342–5350.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R., 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *Journal of Hydrology*, 400 (1–2), 83–94.

Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., and Lane, J.E., 2009. Precipitation: Measurement, remote sensing, climatology and modeling. *Atmospheric Research*, 94 (4), 512–533.

Montanari, A. and Di Baldassarre, G., 2013. Data errors and hydrological modelling: The role of model structure to propagate observation uncertainty. *Advances in Water Resources*, 51, 498–504.

Moulin, L., Gaume, E., and Obled, C., 2009. Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. *Hydrol. Earth Syst. Sci.*, 13 (2), 99–114.

Pappenberger, F., Beven, K., Horritt, M., and Blazkova, S., 2005. Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. *Journal of Hydrology*, 302 (1-4), 46–69.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46 (5), W05521.

Renard, B., Kavetski, D., Leblois, E., Thyer, M., George Kuczera, and Franks, S.W., 2011. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resources Research*, 47 (11), W11516.

Renard, B., Kuczera, G., Kavetski, D., Thyer, M., and Franks, S., 2008. Bayesian total error analysis for hydrologic models: quantifying uncertainities arising from input, output and structural error. *In International Conference on Water Resources and Environmental Research, 4th edition*, 608–619.

Shafiei, M., Ghahraman, B., and Davary, K., 2014. Assessment of rain-gauge networks using a probabilistic GIS based approach. *Hydrology Research*, 45 (4-5), 551–562.

Surkan, A.J., 1969. Synthetic Hydrographs: Effects of Network Geometry. *Water Resources Research*, 5 (1), 112–128.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., and Srikanthan, S., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resources Research*, 45 (12), W00B14.

Xu, H., Xu, C.-Y., Chen, H., Zhang, Z., and Li, L., 2013. Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. *Journal of Hydrology*, 505, 1–12.

Xu, H., Xu, C.-Y., Sælthun, N.R., Xu, Y., Zhou, B., and Chen, H., 2015. Entropy theory based multi-criteria resampling of rain gauge networks for hydrological modelling – A case study of humid area in southern China. *Journal of Hydrology*, 525, 138–151.

Zeng, Q., Xu, C.-Y., Jie, M.X., and Hou, Y.K., 2016. Feasibility and uncertainty of using conceptual rainfall-runoff models in design flood estimation. *Hydrology Research, in press. doi: 10.2166/nh.2015.069.*

APPENDIX A:

Model description

In the present TOPMODEL scheme, a catchment is grid–cell distributed. For the nth cell, precipitation firstly infiltrates through the root zone storage, S_{1_n}. Where S_{1_n} is the minimum value between a constant value and the local initial deficit (D_n), both with dimension of length (L). The rate at which water infiltrates is i $[LT^{-1}]$, equal to the precipitation depth (L) over the duration of the time step (T), unless i is larger than a specified maximum infiltration rate i_{max}. In that case i_{max} gets infiltrated and the excess rainfall $i - i_{max}$ is routed as surface runoff. Once the maximum capacity of the root zone storage, S_{1max}, is filled, water infiltrates towards the unsaturated zone storage, S_{2_n}. The maximum capacity for S_{2_n} at each time step is equal to the local saturated zone deficit minus the root zone storage, $D_n - S_{1_n}$. Once the local maximum capacity is exceeded, again excess water is considered to reach the outlet by surface runoff. There is a leakage q_{v_n} $[LT^{-1}]$ from the unsaturated zone storage, S_{2_n}, towards the catchment lumped subsurface storage, S_3, the leakage is a function of the water at the unsaturated zone storage and a local residence time i.e. a factor, t_d $[TL^{-1}]$, multiplied by the local unsaturated zone deficit, D_n. Thus the leakage or vertical flux is $q_{v_n} = S_{2_n}/(D_n \times t_d)$. The leakage is summed up over all the cells and instantaneously redistributed over the subsurface storage, S_3.

In TOPMODEL a relationship between the shape of the saturated area and the subsurface storage is established based on the following assumptions (Kirkby 1997): (a) the saturated zone takes a configuration as if it is in equilibrium with a steady recharge rate, r, over an upslope contributing area, a_n; (b) the water table is almost parallel to the subsurface such that the effective hydraulic gradient is

equal to the local surface slope, $\tan \beta_n$; (c) a transmissivity profile is described by a function, with a value of T_o [L^2/T] when the soil is just saturated to the surface (zero deficit). The exponential function is the most commonly used, although other options have been described in Ambroise *et al.* (1996), Duan and Miller (1997) and Iorgulescu and Musy (1997). From assumptions (b) and (c), at any cell n on the catchment, the downslope saturated subsurface flow rate per unit contour length, q_n [L^2/T], is described using an exponential transmissivity soil profile with a rate of decline controlled by a parameter m.

$$q_n = T_o \tan \beta_n \, e^{-D_n/m} \qquad (A1)$$

From assumption (a)

$$q_n = a_n r \qquad (A2)$$

Combining of equation (A1) and equation (A2) leads to:

$$D_n = -m \times \ln(\frac{r a_n}{T_o \tan \beta_n}) \qquad (A3)$$

For $\bar{D}$ equal to the average of D_n over the catchment, a relation between the local deficit, D_n, and the catchment average storage deficit ($\bar{D}$) is expressed as

$$D_n = \bar{D} + m \left[\gamma - \ln(\frac{a_n}{T_o \tan \beta_n}) \right] \qquad (A4)$$

Where $\ln(a_n/(T_o \tan \beta_n))$ is known as the soil–topographic index and can be calculated from the catchment topographic information and a value of T_o; and γ is the average of the wetness index over all the cells in the catchment.

The baseflow compounded volume, Q_b in units L^3/T, over all the catchment at any time step is estimated as the sum of the downslope subsurface flow rate over all the channel cells and simplified as in equation (A5), for A equal to the catchment area

$$Q_b = \sum q_n = A e^{-\gamma} e^{-\bar{D}/m} \qquad (A5)$$

The compounded unsaturated zone recharge volume, Q_v [L^3/T], is estimated as the sum of vertical flows at all the cells.

$$Q_v = A \times \sum q_{v_n} \qquad (A6)$$

The catchment average storage deficit at each time step is updated by subtracting the unsaturated zone recharge ($Q_{v_{t-1}}$) and adding the baseflow ($Q_{b_{t-1}}$) both calculated from the previous time step:

$$\bar{D}_t = \bar{D}_{t-1} + \frac{[Q_{b_{t-1}} - Q_{v_{t-1}}]}{A} \qquad (A7)$$

A simulation can be initialized by assuming an initial value of $\bar{D}$ estimated by inverting equation (A5) assuming an initial baseflow. Then, the shape of the saturated area, D_n can be estimated by equation (A4).

14

There are two sources of water surface runoff: excess rainfall when the rate of water exceeds a specified maximum infiltration rate and water excess after the unsaturated zone storage has reach its maximum capacity (saturation overland flow). The network width function, NWF, concept (Surkan 1969, Kirkby 1976) with further adaptations by Grimaldi *et al.* (2010) was used to route the surface runoff. The main idea behind the NWF is that discharge at the outlet is influenced by the shape of the river network. Thus, a histogram function is built based on the shape of the flow path distances along the catchment contributing area and river network. Assuming constant celerity through the river network (Beven *et al.* 1979, McDonnell and Beven 2014) and a slope–varying celerity on the hillslopes (Haan *et al.* 1994, Grimaldi *et al.* 2010).

If we assume that a quantity of surface runoff $Q(n, t - \tau_n)$ is exceeded at cell n during a time period $t - \tau_n$, where τ_n is the time that takes to travel from the nth cell to the outlet. The hydrograph, $U(t)$ [L^3/T], at the outlet cell is the sum of $Q(n, t - \tau_n)$ at all the cells which contributions arrived at the same time period (t) at the outlet :

$$U(t) = \sum_1^N Q(n, t - \tau_n) \qquad (A8)$$

Where τ_n [T] is estimated by equation (A9) as the sum of the time spent on the hillslope and time spent in the channel network. Flow velocity is considered slower on the hillslope than in the channel network.

$$\tau_n = \frac{L_{c_n}}{v_c} + \frac{L_{h_n}}{v_{h_n}} \qquad (A9)$$

where,

-Following the direction of flow, L_{h_n} [L] is the distance from the nth cell at a hillslope towards the junction at the channel; L_{c_n} [L] is the distance from the junction at the channel towards the catchment outlet.

-v_c, Celerity in channel (L/T), spatially uniform.

-v_{h_n}, Celerity in hillslope (L/T), based on results from (Grimaldi et al. 2010) where various options to estimate celerity at hillslope scale were tested, and an equation for locally variable velocity dependent on hillslope slope from Haan *et al.* (1994) was shown to give better results.

$$v_{h_n} = \text{luc} * \sqrt{s_n} \qquad (A10)$$

For s_n corrected for steep slopes, larger than 0.04.

-s_n, slope from one cell to the next in the direction of flow (L/L).

-luc, land use coefficient that controls the velocity on hillslope (L/T)

15

Tables

Table 1 Model parameters and sample ranges.

Parameter	Variable name	Sampled range	Source
Rate of decline of transmissivity (m)	m	0.005-0.035*	(Beven 1997)
Horizontal transmissivity (m^2 h^{-1})	T_o	0.005-50*	(Beven 1997)
Maximum infiltration capacity rate (m h^{-1})	i_{max}	0.02-0.13	(Dingman 2002)
Time constant factor (h m^{-1})	t_d	1-80*	(Beven 1997)
Channel velocity (m s^{-1})	v_c	1.5-3.5	(Di Lazzaro 2009)
Land use coefficient (m s^{-1})	luc	0.02-0.3	(Di Lazzaro 2009)

[*]Obtained by considering catchments with similar characteristics as the one studied here

16

Table 2 For a 95% probability limits, relative precision (*rp*) and proportion of observations inside the predicting range (*pr*) for calibration and validation events using and not using multipliers.

Event		1		2		3		4		5		6	
		rp	*pr*	*rp*	*pr*	*rp*	*pr*	*rp*	*pr*	*rp*	*po*	*rp*	*po*
Calibration event	With multipliers	14.3	0.79	6.6	0.79	12.3	0.87	15.4	0.71	7.4	0.91	13.2	0.67
	Without multipliers	14.3	0.60	6.6	0.84	12.2	0.85	15.7	0.75	7.2	0.89	13.4	0.62
Validation event	With multipliers	8.4	0.72	5.4	0.91	6.9	1.00	7.3	0.96	5.5	0.93	6.8	0.80
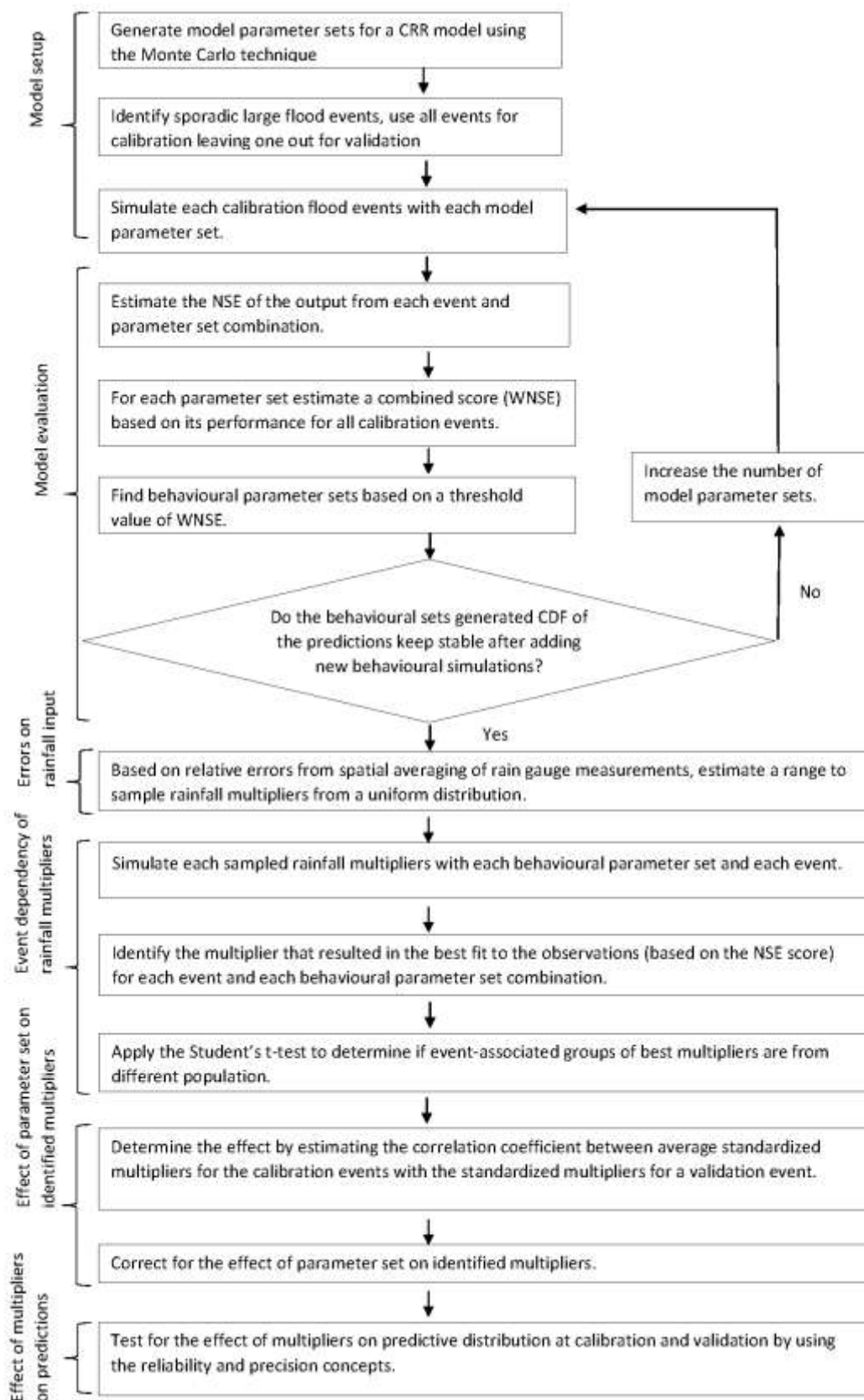
Fig. 1 Flowchart of the procedure for using rainfall multipliers to improve discharge predictions for large flood events
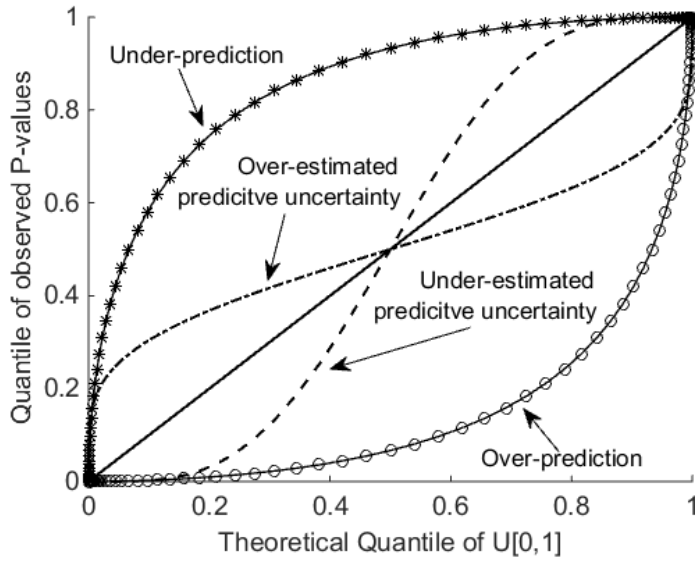
Fig. 2 Interpretation of possible outcomes of a predictive QQ-plot (modified from (Laio and Tamea 2007))
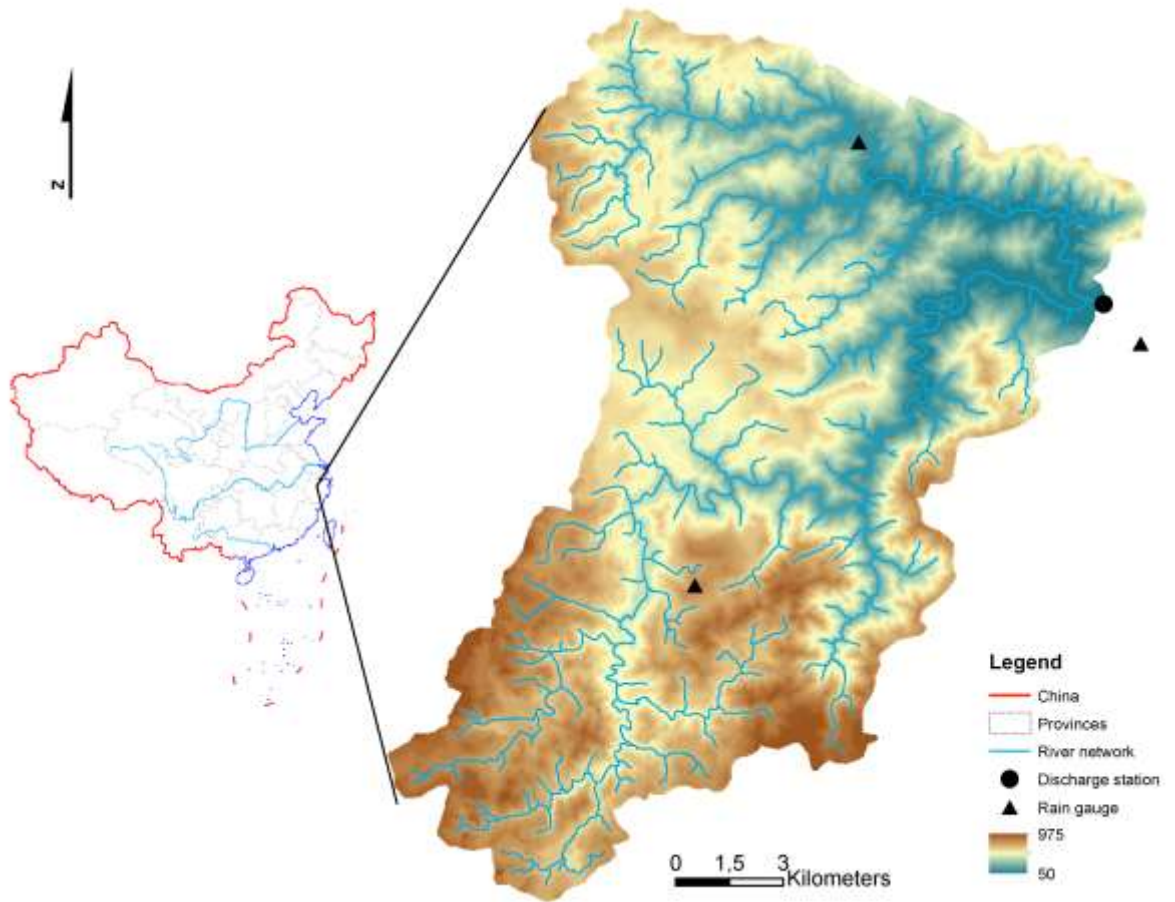
Fig. 3 Map of China, Jiaokou catchment and stations location
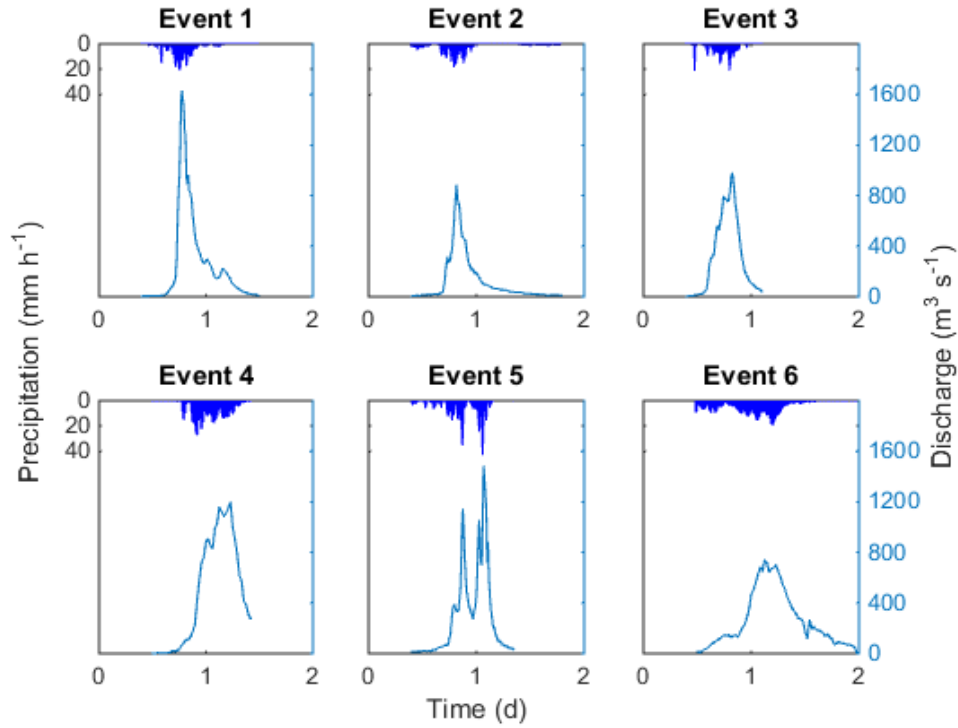
Fig. 4 Measured rainfall intensity (bars) and stream flow (continues line) for six flood events at the Jiaokou catchment
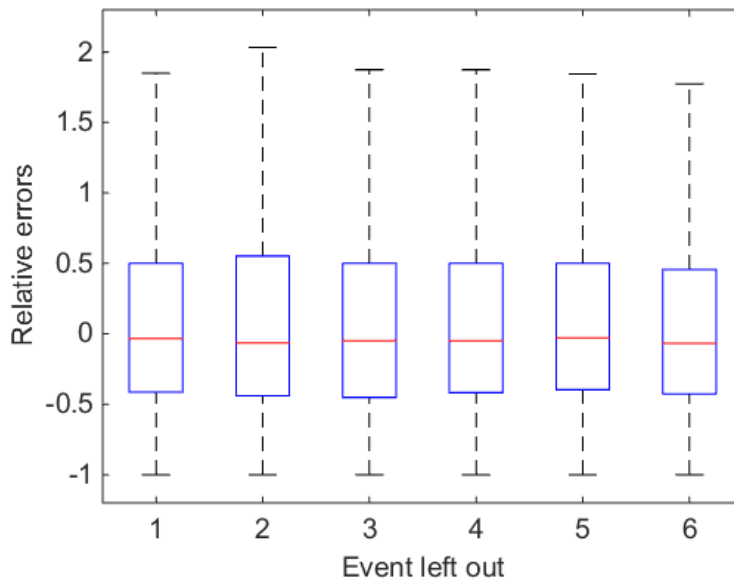

Fig. 5 Boxplots of the relative errors at three locations from spatial averaging of rainfall. Boxes correspond to 20th, 50th and 80th quantiles, and whiskers extend to 1.5 times the interquartile range.

20
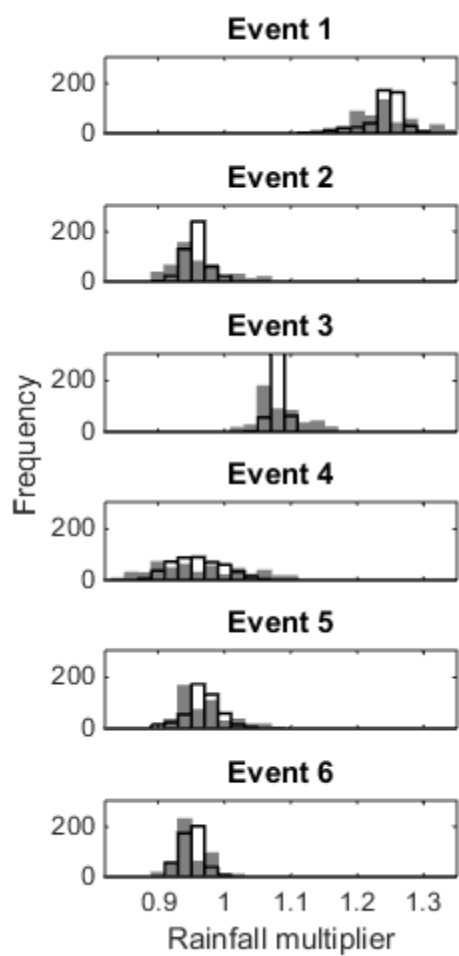
Fig. 6 Frequency distribution of identified best multipliers for different events (top to bottom), without and with adjustment for the effect of parameter sets on the multipliers (grey and black outlined bars respectively)

21
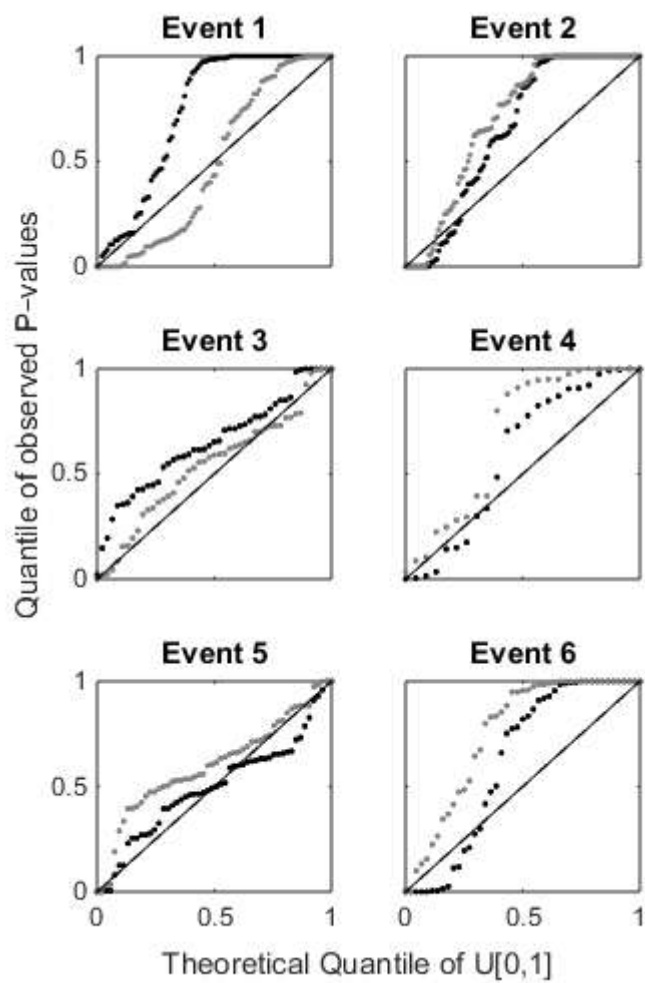
Fig. 7 Predictive QQ-plot for events 1 to 6 when they are part of the calibration events, for the cases of using multiplier (grey dots) and without use of multipliers (black dots)
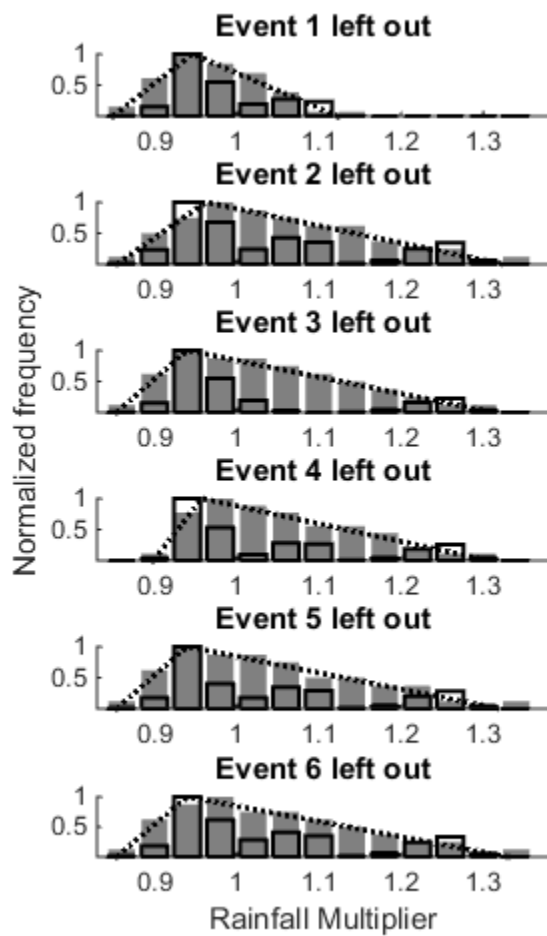
22

Fig. 8 Fuzzy measure (dotted line) based on the probability distribution of identified best multipliers for all calibration events (black outlined bars) and the probability distribution of sampled multipliers according to the fuzzy measure (grey bars).
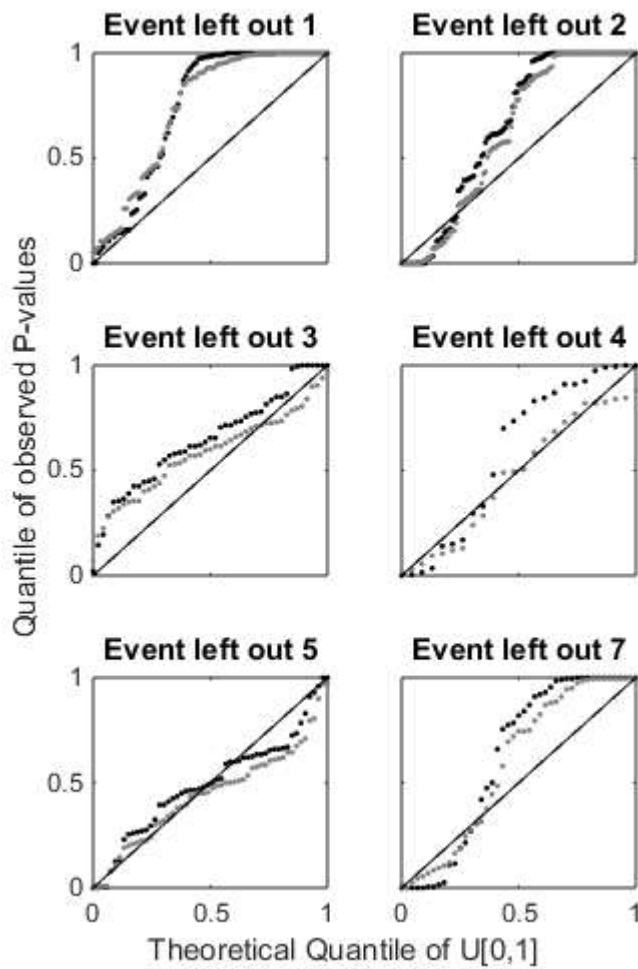
Fig. 9 Predictive QQ-plot for the six different validation events (event left out) for the cases of using multiplier (grey dots) and without use of multipliers (black dots)
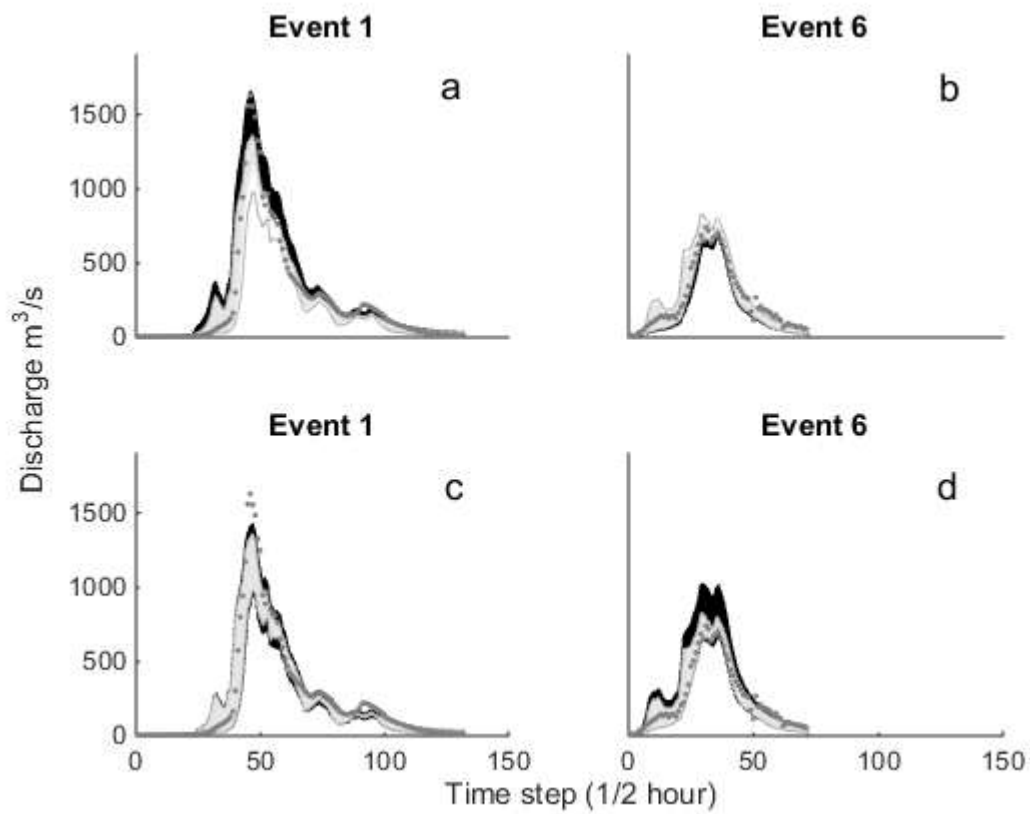
24

Fig. 10 Predictive ranges of the 95% probability limits, using multipliers (black), and not using multipliers (grey) and observed flows (grey dots) for events1 and 6 at calibration (a and b respectively) and same events at validation (c and d respectively).

25