

A comparative assessment of rainfall-runoff modelling against regional flow duration curves for ungauged catchments

Daeha Kim¹, Ilwon Jung², Jong Ahn Chun¹

¹APEC Climate Center, Busan, 48058, South Korea

5 ²Korea Infrastructure Safety & Technology Corporation, Jinju, Gyeongsangnam-do, 52852, South Korea

Correspondence to: Jong Ahn Chun (jachun @apcc21.org)

Abstract. Rainfall-runoff modelling has long been a special subject in hydrological sciences, but identifying behavioural parameters in ungauged catchments is still challenging. In this study, we comparatively evaluated performance of the local calibration of a rainfall-runoff model against regional flow duration curves (FDC), which is a seemingly alternative method of classical parameter regionalisation for ungauged catchments. We used a parsimonious rainfall-runoff model over 45 Korean catchments under semi-humid climate. The calibration against regional FDCs was compared with the simple proximity-based parameter regionalisation. Results show that transferring behavioural parameters from gauged to ungauged catchments significantly outperformed the local calibration against regional FDCs due to the absence of flow timing information in the regional FDCs. The behavioural parameters gained from observed hydrographs were likely to contain intangible flow timing information affecting predictability in ungauged catchments. Additional constraining with the rising limb density appreciably improved the FDC calibrations, implying that flow signatures in temporal dimensions would supplement the FDCs. As an alternative approach in data-rich regions, we suggest calibrating a rainfall-runoff model against regionalised hydrographs to preserve flow timing information. We also suggest use of flow signatures that can supplement hydrographs for calibrating rainfall-runoff models in gauged and ungauged catchments.

20 **1 Introduction**

A standard method to predict daily streamflow is to employ a rainfall-runoff model that conceptualises catchment functional behaviours, and simulate synthetic hydrographs from atmospheric drivers (Wagener and Wheater, 2006; Blöschl et al., 2013). A prerequisite of this conceptual modelling approach is parameter identification to enable the rainfall-runoff model to imitate actual catchment behaviours. Conventionally, behavioural parameters are estimated via model calibration against observed hydrographs (referred to as the hydrograph calibration hereafter). The hydrograph calibration provides convenience to attain reproducibility of the predictand (i.e., streamflow time series), which is commonly used as a performance measure in rainfall-runoff modelling studies. Because the degree of belief in hydrological models is normally measured by how they can reproduce observations (Westerberg et al., 2011), use of the hydrograph calibration has a long tradition in runoff modelling (Hrachowiz et al., 2013).

30 The hydrograph calibration, however, can be challenged by epistemic errors in input and output data, sensitivity to calibration criteria, and inability under no or poor data availability (Westerberg et al., 2011; Zhang et al., 2008). Importantly, it is difficult to know whether the parameters optimised toward maximising hydrograph reproducibility are unique to represent actual catchment behaviours, since multiple parameter sets possibly show similar predictive performance (Beven, 2006, 1993). This low uniqueness of the optimal parameter set, namely the equifinality problem in conceptual hydrological
35 modelling, can become a significant uncertainty source particularly when extrapolating the optimal parameters to ungauged catchments (Oudin et al., 2008).

To overcome or circumvent those disadvantages, distinctive flow signatures (i.e., metrics or auxiliary data representing catchment behaviours) in lieu of observed hydrographs can be used to identify model parameters. The flow duration curve (FDC) has received particular attention in the signature-based model calibrations as a single criterion (e.g., Westerberg et al.,
40 2014, 2011; Yu and Yang, 2000; Sugawara, 1979) or one of calibration constraints (e.g., Pfannerstill et al., 2014; Kavetski et al., 2011; Hingray et al., 2010; Blazkova and Beven, 2009; Yadav et al., 2007). The FDC, the relationship between flow magnitude and its frequency, provides a summary of temporal streamflow variations in a probabilistic domain (Vogel and Fennessey, 1994). Many FDC-related studies have found that climatological and geophysical characteristics within a catchment determine the shape of the FDC (e.g., Cheng et al., 2012; Ye et al., 2012; Yokoo and Sivaplan, 2011; Bottor et al.,
45 2007). With only few physical parameters, the shape of the period-of-record FDC could be analytically expressed (Botter et al., 2008). Based on this strong relationship between catchment physical properties and the FDC, one may hypothesise that model calibration against the FDC (referred to as the FDC calibration hereafter) can provide parameters that can sufficiently capture actual catchment behaviours. Sugawara (1979) is the first attempt at the FDC calibration, emphasising its advantage to reduce negative effects of epistemic errors in rainfall-runoff data. Westerberg et al. (2011) also highlighted that the FDC
50 calibration may provide robust predictions to moderate disinformation such as the presence of event flows under inconsistency between inputs and outputs.

If it allows rainfall-runoff models to sufficiently capture functional behaviours of catchments, the FDC calibration would have an especial value in comparison to the parameter regionalisation for prediction in ungauged catchment. The parameter regionalisation, which transfers or extrapolates behavioural parameters from gauged to ungauged catchments (e.g., Kim and
55 Kaluarachchi, 2008; Oudin et al., 2008; Parajka et al., 2007; Wagener and Wheeler, 2006; Dunn and Lilly, 2011), conveniently provides a priori estimates of behavioural parameters and thus became a popular approach to parameter identification in ungauged catchments (see a comprehensive review in Parajka et al., 2013). However, it has a critical concern that regionalised parameters are highly dependent on model calibrations at gauged sites that may have substantial equifinality problems. Under no flow information in ungauged catchments, it is impossible to know whether regionalised
60 parameters are behavioural. Thus, regionalised parameters might be insufficiently reliable and highly uncertain (Bárdossy, 2007; Oudin et al., 2008; Zhang et al., 2008).

On the other hand, the calibration against regional FDCs (referred to as RFDC_cal hereafter) may reduce the primary concern in the classical parameter regionalisation scheme. The regional models predicting FDCs at ungauged sites have

showed strong performance, for instance, via regression analyses between quantile flows and catchment properties (e.g., Shu and Ouarda, 2012; Mohammoud, 2008; Smakhtin et al., 1997), geostatistical interpolation of quantile flows (e.g., Pugliese et al., 2014; Westerberg et al., 2014), and regionalisation of theoretical probability distributions (e.g., Atieh et al., 2017; Sadegh et al., 2016) among many variations. The parameters obtained from RFDC_cal are deemed behavioural, because a distinctive flow signature of the target ungauged catchment directly identifies them; however, predicted FDCs should be reliable in this case. A FDC is a compact representation of runoff variability at all time scales from inter-annual to event-scale, embedding various aspects of multiple flow signatures (Blöschl et al., 2013). Based on this strength, several studies already showed promising predictive performance using RFDC_cal for ungauged catchments (e.g., Westerberg et al., 2014; Yu and Yang, 2000).

Nevertheless, practical questions arise when using RFDC_cal for ungauged catchments. First, the FDC is simplified information with flow magnitudes only; hence, the FDC calibration could worsen the equifinality problem relative to the hydrograph calibration. Due to no flow timing information in regional FDCs, one may cast a concern that parameters obtained from RFDC_cal may provide poorer predictive performance than regionalised parameters gained from the hydrograph calibration. Indeed, there is additional uncertainty in predicted FDCs possibly introduced by the regionalisation models (Westerberg et al., 2011; Yu et al., 2002). RFDC_cal may be undesirable when a simple parameter regionalisation can provide better performance, because regionalising observed FDCs may require expensive efforts. Several comparative studies on parameter regionalisation (e.g., Parajka et al., 2013; Oudin et al., 2008) suggested that the simple proximity-based parameter transfer can be competitive in many regions. Second, there may be additional flow signatures to improve predictive performance of the FDC calibration. Additional constraining can lead to better predictive performance of the RFDC (Westerberg et al., 2014); however, it is still an open question which flow signatures can supplement the FDC calibration.

As discussed, RFDC_cal seems promising for prediction in ungauged catchments. However, to our knowledge, RFDC_cal has never been evaluated in a comparative manner with classical parameter regionalisation except Zhang et al. (2015), which assessed its performance in part. Therefore, this study aimed to evaluate predictive performance of RFDC_cal in comparison to a conventional parameter regionalisation. We focused on the absence of flow timing in the FDC and its impacts on rainfall-runoff modelling. In this work, a parsimonious 4-parameter conceptual model was used to simulate daily hydrographs for 45 catchments in South Korea. To predict FDCs in ungauged catchments, a geostatistical regional model was adopted here. The Monte-Carlo sampling was used to identify model parameters and measure equifinality in the hydrograph and the FDC calibrations.

2 Description of the study area and data

45 catchments located across South Korea with no or negligible human-made influences on flow variations were selected for this study (Figure 1). South Korea is characterised as a temperate and semi-humid climate with rainy summer seasons. The

North Pacific high-pressure brings monsoon rainfall with high temperatures during summer seasons, while dry and cold weathers prevail in winter seasons due to the Siberian high-pressure. Typical ranges of annual precipitation are 1200-1500 and 1000-1800 mm in the northern and the southern areas respectively (Rhee and Cho, 2016). Annual mean temperatures in South Korea range between 10 and 15 °C (Korea Meteorological Administration, 2011). Approximately, 60-70 percent of precipitation falls in summer seasons from June to September (Bae et al., 2008). Streamflow usually peaks in the middle of summer seasons because of heavy rainfall or typhoons, and hence information of catchment behaviours is largely concentrated on summer-season hydrographs. Snow accumulation and ablation occurring at high elevations have minor influences on flow variations due to relatively small amount of winter precipitation (Bae et al., 2008).

The study catchments were selected based on availability of streamflow data. High-quality daily streamflow data across South Korea have been produced since establishment of the Hydrological Survey Centre in 2007 (Jung et al., 2010), though river stages have been monitored for an extensive length at a few gauging stations. Thus, we collected streamflow data at 29 river gauging stations from 2007 to 2015 together with inflow data of 16 multi-purpose dams for the same data period from the Water Resources Management Information System operated by the Ministry of Land, Infrastructure, and Transport of the Korean government (available at <http://www.wamis.go.kr/>). The mean annual flow of the study catchments was 739 mm yr⁻¹ with a standard deviation of 185 mm yr⁻¹ during 2007-2015.

In addition, as atmospheric forcing inputs, we collected daily precipitation and maximum and minimum temperatures for 2005-2015 at 3-km grid resolution produced by spatial interpolations between 60 stations of the automated surface observing system (ASOS) maintained by the Korea Meteorological Administration. The ASOS data were interpolated by the Parameter-elevation Regression on Independent Slope Model (PRISM; Daly et al., 2008), and overestimated pixels of the PRISM grid data were smoothed by the inverse distance method. Jung and Eum (2015) found that this combined method improved the spatial interpolation of precipitation and the temperatures in South Korea. The annual mean precipitation and temperature of the study catchments vary within ranges of 1145–1997 mm yr⁻¹ and 8.0–13.8 °C during 2007-2015. Hydro-climatological features of the 45 catchments are summarised in Table 1.

3 Methodology

3.1 Hydrological model (GR4J)

A parsimonious rainfall-runoff model, GR4J (Perrin et al., 2003), was adopted to simulate daily hydrographs of the 45 catchments for 2007-2015. GR4J conceptualises functional catchment response to rainfall with four free parameters that regulate the water balance and water transfer functions. Figure 2 schematises the structure of GR4J. The four parameters (X1 to X4) conceptualises soil water storage, groundwater exchange, routing storage, and the base time of unit hydrograph respectively. Since its parsimonious and efficient structure allows robust calibration and reliable regionalisation of the parameters, GR4J has been frequently used for modelling daily hydrographs with various purposes under diverse climatic conditions (Zhang et al., 2015). The computation details and discussion are found in Perrin et al. (2003). The potential

evapotranspiration (PE in Figure 2) was estimated by the temperature-based model proposed by Oudin et al. (2005) for lumped rainfall-runoff modelling.

130 3.2 Preliminary data processing

Before rainfall-runoff modelling, we preliminarily processed the grid climatic data to convert precipitation data to liquid water forcing (i.e., rainfall and snowmelt depths) using a physics-based snowmelt model proposed by Walter et al. (2005). The preliminary snowmelt modelling was mainly for reducing systematic errors from no snow component in GR4J, which may affect model performance in catchments at relatively high elevations. We chose this preliminary processing to avoid
135 adding more parameters (e.g., the temperature index) to the existing structure of GR4J. In the case of GR4J, one additional parameter implies 25% complexity increase in terms of the number of parameters, and thus can worsen the equifinality. The snowmelt model uses the same inputs of GR4J to simulate point-scale snow accumulation and ablation processes (i.e., no additional inputs are required). The snowmelt model is a physics-based model but uses empirical methods to estimate its parameters for the energy balance simulation. As outputs, it produces the liquid water depths and the snow water equivalent.
140 For lumped inputs to GR4J, we took spatially averaged pixel values of the liquid water depths and the maximum and minimum temperatures within the boundary of each catchment.

After the snowmelt modelling, consistency between the liquid water depths and the observed flows (i.e., input-output consistency) was checked using the current precipitation index (CPI; Smakhtin and Masse, 2000) defined as:

$$I_t = I_{t-1} \cdot K + R_t \quad (1)$$

145 where I_t is the CPI (mm) at day t , K is a decay coefficient (0.85 d^{-1}), and R_t is the liquid water depth (mm d^{-1}) at day t . CPI mimics temporal variations of typical streamflow data by converting intermittent precipitation data to a continuous time series with an assumption of the linear reservoir. The input-output consistency can be evaluated using correlation between CPI and observed streamflow as in Westerberg et al. (2014) and Kim and Kaluarachchi (2014). The Pearson correlation coefficients between CPI and streamflow data of the 45 catchments had an average of 0.67 with a range of 0.43-0.79, and no
150 outliers were found in the box plot of the correlation coefficients. Hence, we assumed that consistency between climatic forcing and observed hydrographs was acceptable.

3.3 The hydrograph calibration in gauged catchments

To search behavioural parameter sets of GR4J against the streamflow observations (i.e., the hydrograph calibration), we used the Monte-Carlo random sampling with the parameter ranges given by Demirel et al. (2013). The objective function in
155 Zhang et al. (2015) was chosen as the calibration criterion to consider the Nash-Sutcliffe Efficiency (NSE) and the Water Balance Error (WBE) together:

$$\text{OBJ} = (1 - \text{NSE}) + 5|\ln(1 + \text{WBE})|^{2.5} \quad (2a)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^N (Q_{obs,i} - \overline{Q_{obs}})^2} \quad (2b)$$

$$WBE = \frac{\sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})}{\sum_{i=1}^N Q_{obs,i}} \quad (2c)$$

160 where Q_{obs} and Q_{sim} are the observed and simulated flows respectively, $\overline{Q_{obs}}$ is the arithmetic mean of Q_{obs} , and N is the total number of flow observations. The best parameter sets for each study catchment was obtained from minimisation of the OBJ using the Monte-Carlo simulations described below.

To determine sufficient runs for the random simulations, we calibrated GR4J parameters using the shuffled complex evolution (SCE) algorithm (Duan et al., 1992) for one catchment with moderate input-output consistency. Then, the total number of random simulations was iteratively determined by adjusting the number of runs until the minimum OBJ of the random simulations became adequately close to the OBJ value from the SCE algorithm. We found that approximately 20,000 runs could provide the minimum OBJ value equivalent to that from the SCE algorithm. Subsequently, GR4J was calibrated by 20,000 runs of the Monte-Carlo simulations for all 45 catchments, and the parameter sets with the minimum OBJ values were taken for runoff predictions. In addition, we sorted the 20,000 parameter sets in terms of corresponding OBJ values in ascending order, and first 50 sets (0.25% of the total samples) were taken to measure the degree of equifinality. We measured the equifinality simply by the prediction area between 2.5% and 97.5% boundaries of runoff simulations given by the collected 50 parameter sets. This prediction area was later compared to that from the FDC calibration under the same Monte-Carlo framework. Note that we estimated the prediction area to comparatively evaluate the degree of equifinality between the hydrograph and the FDC calibrations under the same sampling size and the same acceptance rate for all the catchments. For more sophisticated and reliable uncertainty estimation, other methods are available such as the Generalised Likelihood Uncertainty Estimation (GLUE; Beven and Bingley, 1992) and the Differential Evolution Adaptive Metropolis (DREAM; Vrugt and Ter Braak, 2011).

175 For the hydrograph calibration, the 9-year streamflow data were divided into two parts for calibration (2011-2015) and for validity check (2007-2010), respectively. A two-year warm-up period was used for initialising all runoff simulations in this study.

3.4 Model calibration against the regional FDC for ungauged catchments

Each catchment was treated ungauged for the comparative evaluation of RFDC_cal in the leave-one-out cross-validation (LOOCV) mode. For regionalising empirical FDCs, the geostatistical method recently proposed by Pugliese et al. (2014) was used. Pugliese et al. (2014) employed the top-kriging method (Skøien et al., 2006) to spatially interpolate the total negative deviation (TND), which is defined as the area between the mean annual flow and below-average flows in a normalised FDC. The top-kriging weights that interpolate TND values were taken as weights to estimate flow quantiles of ungauged catchments from empirical FDCs of surrounding gauged catchments. The FDC of an ungauged catchment in Pugliese et al. (2014) is estimated from normalised FDCs of surrounding gauged catchments as:

$$\hat{\Phi}(w_0, p) = \hat{\phi}(w_0, p) \cdot \bar{Q}(w_0) \quad (3a)$$

$$190 \quad \hat{\Phi}(w_0, p) = \sum_{i=1}^n \lambda_i \cdot \phi_i(w_i, p), \quad p \in (0,1) \quad (3b)$$

where $\hat{\Phi}(w_0, p)$ is the estimated quantile flow ($m^3 s^{-1}$) at an exceedance probability p (unitless) for an ungauged catchment w_0 , $\hat{\phi}(w_0, p)$ is the estimated normalised quantile flow (unitless), $\bar{Q}(w_0)$ is the annual mean streamflow ($m^3 s^{-1}$) of the ungauged catchment, and $\phi_i(w_i, p)$ and λ_i are normalised quantile flows (unitless) and corresponding top-kriging weights (unitless) of gauged catchment w_i , respectively. The unknown mean annual flow of an ungauged catchment, $\bar{Q}(w_0)$, can be
 195 estimated with a rescaled mean annual precipitation defined as:

$$MAP^* = 3.171 \times 10^{-5} \cdot MAP \cdot A \quad (4)$$

where MAP^* is the rescaled mean annual precipitation ($m^3 s^{-1}$), MAP is mean annual precipitation ($mm yr^{-1}$) and A is the area (km^2) of the ungauged catchment, and the constant 3.171×10^{-5} converts the unit of MAP^* from $mm yr^{-1} km^2$ to $m^3 s^{-1}$.

A distinct advantage of the geostatistical method is its ability to estimate the entire flow quantiles in a FDC with a single set
 200 of top-kriging weights. Since a parametric regional FDC (e.g., Yu et al., 2002; Mohamoud, 2008) is obtained from independent models for each flow quantile in many cases, for instance, by multiple regressions between selected quantile flows and catchment properties, fundamental characteristics in a FDC continuum would be entirely or partly lost. The geostatistical method, on the other hand, treats all flow quantiles as a single object; thereby, features in a FDC continuum can be preserved. It showed promising performance to reproduce empirical FDCs only using topological proximity between
 205 catchments. More details on the geostatistical method are found in Pugliese et al. (2014).

For regionalising empirical FDCs of the 45 catchments, we followed the same procedure of Pugliese et al. (2014). We obtained top-kriging weights (λ_i) by the geostatistical interpolation of TND values from observed FDCs for the calibration period (2011-2015). Then, the top-kriging weights were used to interpolate empirical flow quantiles. The number of neighbours for the TND interpolation was iteratively determined as five at which additional neighbouring TNDs are unlikely
 210 to bring better agreement between the estimated and observed TNDs. In other words, normalised flow quantiles of five catchments surrounding the target ungauged catchment were interpolated with the top-kriging weights. Then, MAP^* of the target ungauged catchment was multiplied. We predicted flow quantiles at 103 exceedance probabilities (p of 0.001, 0.005, 99 points between 0.01 and 0.99 at an interval of 0.01, 0.995, and 0.999) for rainfall-runoff modelling against regional FDCs (i.e., RFDC_cal).

215 For runoff prediction in ungauged catchments, the GR4J parameters were identified by the same Monte-Carlo sampling but toward minimisation of OBJ value between the regional and the modelled flow quantiles at the 103 exceedance probabilities. The best parameter set, which provided the minimum OBJ value, was taken as the best behavioural set of RFDC_cal for each catchment.

3.5 Proximity-based parameter regionalisation for ungauged catchments

220 We selected the proximity-based parameter transfer (referred to as PROX_reg hereafter) to comparatively evaluate predictive performance of RFDC_cal. The parameter regionalisation has three classical categories: (a) proximity-based parameter transfer (i.e., PROX_reg; e.g., Oudin et al., 2008); (b) similarity-based parameter transfer (e.g., McIntyre et al., 2005); and (c) regression between parameters and physical properties of gauged catchments (e.g., Kim and Kaluarachchi, 2008). A comprehensive review on the parameter regionalisation in Parajka et al. (2013) reported that PROX_reg has 225 competitive performance under humid climate with low-complexity models relative to the other categories. Based on modelling conditions in this study (semi-humid climate and 4 parameters), we chose PROX_reg to evaluate RFDC_cal. To predict runoff at the 45 catchments in the LOOCV mode, we transferred the behavioural parameter sets obtained from the hydrograph calibration of the five donor catchments used for the FDC regionalisation. In other words, we used the same donor catchments for FDC regionalisation and PROX_reg. This allows us to have consistency in transferring hydrological information from gauged to ungauged catchments between RFDC_cal and PROX_reg. Using the best behavioural parameter sets of the five donor catchments, we generated five runoff time series and took the arithmetic averages of them to represent runoff predictions by PROX_reg. 230

3.6 Performance evaluation

We used multiple performance metrics to evaluate predictive performance of all modelling approaches applied in this study. 235 Predictive performance of each modelling approach was graphically evaluated using box plots of the performance metrics of the 45 catchments. In addition, we performed several paired t-tests to check the statistical significance of performance differences between the modelling approaches. Following is the description of the performance metrics.

To measure high- and low-flow reproducibility, we chose two traditional performance metrics, (1) the NSE between observed and predicted flows (Eq. 2b) and (2) the NSE of log-transformed flows (LNSE) respectively. LNSE is calculated as:

$$240 \quad \text{LNSE} = 1 - \frac{\sum_{i=1}^N (\ln(Q_{\text{obs},i}) - \ln(Q_{\text{sim},i}))^2}{\sum_{i=1}^N (\ln(Q_{\text{obs},i}) - \ln(Q_{\text{obs}}))^2} \quad (5)$$

Though NSE and LNSE are frequently used for performance evaluation, they may be sensitive to errors in flow observations (Westerberg et al., 2011). Hence, we additionally selected three typical flow metrics that embed dynamic flow variation in a compact manner; the runoff ratio (R_{QP}), the baseflow index (I_{BF}), and the rising limb density (D_{RL}). R_{QP} , I_{BF} , and D_{RL} are proxies of aridity and water holding capacity, contribution of the baseflow to flow variations, and flashiness of catchment 245 behaviours, respectively. They are defined as the ratio of runoff to precipitation, the ratio of baseflow to total runoff, and the inverse of average time to peak (d^{-1}) as:

$$R_{\text{QP}} = \frac{\bar{Q}}{\bar{P}} \quad (6a)$$

$$I_{\text{BF}} = \sum_{t=1}^T \frac{Q_{\text{B},t}}{Q_t} \quad (6b)$$

$$D_{RL} = \frac{N_{RL}}{T_R} \quad (6c)$$

250 where \bar{Q} and \bar{P} are average flow and precipitation for a given period (mm d^{-1}), Q_t and $Q_{B,t}$ (m d^{-1}) is the streamflow and the base flow at time t respectively, N_{RL} is the number of rising limb, and T_R is the total amount of time when the hydrograph is rising (days). $Q_{B,t}$ can be calculated by subtracting direct flow $Q_{D,t}$ from Q_t as:

$$Q_{D,t} = c \cdot Q_{D,t-1} + 0.5 \cdot (1 + c) \cdot (Q_t - Q_{t-1}) \quad (7a)$$

$$Q_{B,t} = Q_t - Q_{D,t} \quad (7b)$$

255 where c is the filter parameter, which was set to 0.925 (Brooks et al., 2011; Eckhardt, 2007).

Flow signature reproducibility of RFDC_cal and PROX_reg were evaluated by the relative absolute bias between modelled and observed signatures as:

$$D_{FS} = \frac{|FS_{sim} - FS_{obs}|}{FS_{obs}} \quad (8)$$

where D_{FS} is the relative absolute bias, FS_{sim} is a flow signature of the modelled flows, and FS_{obs} is that of the observed flows.

260 **4 Results**

4.1 Hydrograph calibration and FDC regionalisation in gauged catchments

Figure 3a displays results of the parameter identification against the observed hydrographs (i.e., the hydrograph calibration). The 45 catchments had the mean NSE and LNSE of 0.66 and 0.65 between the simulated and observed flows for the calibration period, respectively. The average NSE reduction from the calibration to the validation periods was 0.06 with a
 265 standard deviation of 0.10. The temporal transfer of the calibrated parameters did not decrease the mean LNSE value, while a wider LNSE range indicates that uncertainty of low-flow predictions may increase when temporally transferring the calibrated parameters.

The predictive performance was closely related to the input-output consistency (Figure 3b), which was measured by the Pearson correlation coefficient between the CPI and the observed flows. A low input-output consistency implies that the
 270 rainfall-runoff data may include significant epistemic errors such as minimal flow responses to heavy rainfall or excessive response to tiny rainfalls. If the model calibration compensates disinformation from such errors, the parameters would be forced to have biases. Figure 3b shows that consistency in input-output data is a critical factor affecting parameter identification and thus performance. Perhaps, screening catchments with low input-output consistency may provide better predictions in ungauged catchments. However, we did not consider it in the LOOCV for RFDC_cal and PROX_reg, since
 275 variation in input-output consistency would be a common situation. Rather, reducing the number of gauged catchments lowers spatial proximity and thus can cause biases for ungauged catchments too. Overall, 27 catchments and 33 catchments

showed NSE and LNSE values greater than 0.6. We assumed the hydrograph calibration under the Monte-Carlo framework, which was assisted by the SCE optimisation, was able to acceptably identify the behavioural parameters under given data quality.

280 Besides, Figure 4 illustrates the 1:1 scatter plot between the observed and predicted flow quantiles of all the catchments, indicating high applicability of the top-kriging FDC regionalisation. The overall NSE and LNSE values between the observed and regionalised flow quantiles show good applicability of the geostatistical method. The NSE and LNSE values for individual catchments have averages of 0.83 and 0.91 with standard deviations of 0.25 and 0.11, respectively, implying that low-flow predictions were slightly better. The performance of the geostatistical method was relatively poor at locations
285 where gauging density is low. Catchments 4, 10, 35, and 36, which recorded 0.6 or less NSEs are limitedly hatched with or adjacent to the other catchments; nonetheless, LNSEs of those catchments were still greater than 0.7. This result is consistent with a finding of Pugliese et al. (2016) that performance of the geostatistical method was sensitive to river gauging density. Transferring flow quantiles from remote catchments may not sufficiently capture functional similarity between donor and receiver catchments. In spite of the minor shortcomings, the geostatistical FDC regionalisation was deemed acceptable based
290 on the high NSE and LNSE of flow quantiles. Topological proximity was generally a good predictor of flow quantiles for the study catchments.

4.2 Comparing hydrograph predictability between RFDC_cal and PROX_reg

Figure 5 compares the box plots of NSE and LNSE values between RFDC_cal and PROX_reg. PROX_reg generally outperforms RFDC_cal in predicting both high and low flows, suggesting that transferring parameters identified by observed
295 hydrographs would be a better choice than a local calibration against predicted FDCs. The differences between NSE values of PROX_reg and RFDC_cal have an average of 0.22 with a standard deviation of 0.34. Only 8 catchments showed higher NSEs with RFDC_cal. These higher NSE values of PROX_reg imply that PROX_reg is preferable when high-flow predictability is needed such as flood analyses. In the case of LNSE, PROX_reg still had a higher median than RFDC_cal (0.53 and 0.62 for RFDC_cal and PROX_reg respectively). In 25 catchments, PROX_reg provided LNSE values greater than
300 those of RFDC_cal.

The low performance of RFDC_cal was also found in the comparative assessment of Zhang et al. (2015), which evaluated RFDC_cal for 228 Australian catchments using the same GR4J model. Zhang et al. (2015) found that RFDC_cal was inferior to PROX_reg in the Australian catchments, because the FDC calibration poorly reproduced temporal flow variations relative to the hydrograph calibration. This study confirms the difficulty to capture dynamic catchment behaviours with FDCs
305 containing no flow timing information.

A major weakness of RFDC_cal is the absence of flow timing information in the parameter calibration process. Unlike RFDC_cal, PROX_reg did not discard the flow timing information. The regionalised parameters may be able to implicitly transfer the flow timing information from gauged to ungauged catchments (this hypothesis will be discussed later in Section 4.4). Figure 6 illustrates how the absence of flow timing negatively influences on predictive performance. For this

310 comparison, the parameters were recalibrated against the observed FDCs (not regional FDCs) under the same Monte Carlo method to discard errors introduced by the FDC regionalisation (i.e., equivalent to calibrations against perfectly regionalised FDCs). The parameters identified by the observed hydrograph (Figure 6a) brought a good predictability in both high and low flows, resulting in an excellent performance to reproduce the FDC. On the other hand, an excellent FDC reproducibility does not guarantee a good predictability in high flows (Figure 6b). This indicates that reproducing FDCs with rainfall-runoff
315 models would be less sufficient than the hydrograph calibration to capture functional catchment responses.

In addition, Figure 6 shows that the prediction area of the 50 behavioural parameters from the Monte-Carlo simulations (indicated by the grey areas and the blue arrows) became much larger when using the FDC calibration instead of the hydrograph calibration. We calculated the ratio of the prediction area of the FDC calibration to that of the hydrograph calibration, and defined it as the equifinality ratio. It quantifies the degree of equifinality augmented by replacing the
320 hydrograph calibration with the FDC calibration. Figure 7 displays the scatter plot between the equifinality ratio and the input-output consistency. The equifinality augmented by the loss of flow timing is likely to increase as the input-output consistency decreases. The average of the equifinality ratios was 1.96, implying that potential equifinality inherent in RFDC_cal could be substantial. This may suggest that the equifinality problem embedded in RFDC_cal could be more significant than that in PROX_reg.

325 **4.3 Comparing flow-signature predictability between RFDC_cal and PROX_reg**

Figure 8 summarises performance of RFDC_cal and PROX_reg to regenerate three flow signatures of R_{QP} , I_{BF} , and D_{RL} . RFDC_cal is competitive in reproducing the averaged-based signatures R_{QP} and I_{BF} , while it showed relatively a weak ability to regenerate the event-based signature D_{RL} . R_{QP} and I_{BF} are flow metrics based on averages of long-term flow and precipitation in which no flow timing information is involved. Especially, RFDC_cal showed strong performance in
330 reproducing I_{BF} relative to PROX_reg. This result can be explained by considering that baseflow has less temporal variations than direct runoff in the Korean catchments under typical monsoonal climate. High seasonality of monsoonal precipitation makes high temporal variations in direct runoff during June to September, while relatively steady baseflow is dominant during dry seasons (October to May). In Catchment 2 whose flow variation is displayed in Figure 6, for example, the coefficient of variance (CV) of direct runoff was 5.86 for 2007-2015, which is approximately 3.5 times as high as that CV of
335 baseflow.

On the other hand, RFDC_cal was poorer to reproduce D_{RL} than PROX_reg. This highlights the weakness of RFDC_cal in which only flow magnitudes were used for identifying model parameters. PROX_reg showed better performance to predict D_{RL} than RFDC_cal. Flow timing information gained from the observed hydrographs might be preserved, even after behavioural parameters were transferred to ungauged catchments. Overall, PROX_reg seems to be better than RFDC_cal to
340 predict the three flow signatures together.

The box plots in Figure 9 provide an indication that D_{RL} is likely to supplement the FDC calibration and thus improve RFDC_cal. From the collection of 50 behavioural parameter sets given by the FDC calibration, we chose the parameter set

providing the lowest bias for each flow signature as the best behavioural sets, and simulated runoff again for all catchments. The high-flow predictability was fairly improved by additional constraining with D_{RL} , suggesting that flow metrics associated with flow timing makes up for the weakness of the FDC calibration. Additional constraining with R_{QP} and I_{BF} did not bring appreciable improvement in the FDC calibration. However, $PROX_{reg}$ was still better than the additional constraining with D_{RL} , indicating that a further study is needed for better constraining rainfall-runoff models using FDCs together with additional flow metrics.

4.4 Paired t-tests between the modelling approaches

For comparative evaluation in this study, we produced several runoff prediction sets using multiple rainfall-modelling approaches. First, we calibrated GR4J against the observed hydrographs (referred to as Q_{cal}), and transferred the behavioural parameters to ungauged catchments in the LOOCV mode ($PROX_{reg}$). We constrained GR4J with the regional FDCs ($RFDC_{cal}$). To evaluate equifinality, we recalibrated the GR4J parameters against the observed FDCs (referred to as FDC_{cal}). Additionally, we constrained the model with observed FDCs plus the flow signatures, and significant performance improvement was found with D_{RL} (referred to as $FDC+D_{RL}_{cal}$). A paired t-test using the performance metrics (NSE , $LNSE$, or D_{FS}) between these modelling approaches can answer various questions beyond the graphical evaluations with box plots. For paired t-tests, we added one more case of transferring parameters gained from FDC_{cal} to ungauged catchments (referred to as $FPROX_{reg}$). $FPROX_{reg}$ transfers behavioural parameters with no flow timing information from gauged to ungauged catchments. The mean NSE of $FPROX_{reg}$ was 0.44 with a standard deviation of 0.49.

A primary hypothesis of this study was that $RFDC_{cal}$ could outperform $PROX_{reg}$. This question can be addressed by NSE differences between $RFDC_{cal}$ and $PROX_{reg}$. The mean NSE difference between them was -0.22 and the standard error was 0.051, providing an evaluation that the NSE differences were less than zero at a 95% confidence level. The paired t-test did not lend support the hypothesis (i.e., $PROX_{reg}$ outperformed $RFDC_{cal}$ significantly). However, we could assume that D_{RL} could improve predictive performance of FDC_{cal} . The mean NSE difference between $FDC+D_{RL}_{cal}$ and FDC_{cal} was 0.12 and the standard error was 0.025, confirming the significance at a 95% confidence level.

Likewise, we tested several questions relevant to rainfall-runoff modelling in ungauged catchments using different combinations. One interesting question would be “Did the behavioural parameters from Q_{cal} contain flow timing information for ungauged catchments?” We addressed this question by comparing between $PROX_{reg}$ and $FPROX_{reg}$ with a hypothesis that predictability in ungauged catchments would decrease if the regionalised parameters were gained only from flow magnitudes. $FPROX_{reg}$ uses FDC_{cal} for searching behavioural parameters at gauged catchments; thereby, it cannot transfer flow timing information to ungauged catchments through the behavioural parameters. The mean NSE difference between $PROX_{reg}$ and $FPROX_{reg}$ was 0.10, and the standard error was 0.031. The NSE differences were greater than zero significantly. The behavioural parameters from Q_{cal} were likely to have flow timing information affecting predictability in ungauged catchments. In Table 3, we summarised the results of paired t-tests for scientific questions that may arise from this study. They could be beneficial information for rainfall-runoff modelling in ungauged catchments.

5 Discussion and conclusions

5.1 RFDC_cal for rainfall-runoff modelling in ungauged catchments

The use of regional FDCs as a single calibration criterion appears to be a good choice for searching behavioural parameters in ungauged sites. As discussed earlier, the FDC is a compact representation of runoff variability at all time scales, and thus
380 able to embed multiple hydrological features in catchment dynamics (Blöschl et al., 2013). A pilot study of Yokoo and Sivapalan (2011) discovered that the upper part of an FDC is controlled by interaction between extreme rainfall and fast runoff, while the lower part is governed by baseflow recession behaviour during dry periods. The middle part connecting the upper and the lower parts is related to the mean within year flow variations, which is controlled by interactions between water availability, energy, and water storage (Yager et al., 2012; Yokoo and Sivapalan, 2011). It is well-documented that
385 hydro-climatological processes within a catchment are reflected in the FDC (e.g., Cheng et al., 2012; Ye et al., 2012; Coopersmith et al., 2012; Yaeger et al., 2012; Botter et al., 2008), and therefore the model parameters identified solely by a regional FDC are expected to provide reliable predictions in ungauged catchments (e.g., Westerberg et al., 2014; Yu and Yang, 2000).

The comparative evaluation in this study, however, provides another expected lesson that the FDC calibration is good to
390 reproduce the FDC itself, but it insufficiently captures functional responses of catchments due to the absence of flow timing information. A hydrograph is the most complete flow signature embedding numerous processes interacting within a catchment (Blöschl et al., 2013), being more informative than an FDC. Since any simplification of a hydrograph, including the FDC, should lose some amount of flow information, it is no surprise that the FDC calibration worsens the equifinality. This study emphasises that the absence of flow timing in RFDC_cal may cause larger prediction errors than regionalised
395 parameters gained against observed hydrographs. The paired t-test between PROX_reg and FPROX_reg highlights that regionalised parameters gained from observed hydrographs were likely to contain intangible flow timing information even for ungauged catchments. The flow timing information implicitly transferred to ungauged catchment is a major gap between PROX_reg and RFDC_cal. The errors introduced by the FDC regionalisation were not significant due to high performance of the geostatistical method in this study.

400 Because the hydrograph calibration can compensate the errors in input-output data, one may convert the hydrograph into the FDC to avoid effects of disinformation on rainfall-runoff modelling. However, in this case, valuable flow timing information should be paid in trade-off. For RFDC_cal in this study, we began with converting the observed hydrographs into the flow quantiles to regionalise them; thus, the flow timing information was initially lost. As shown, the performance of RFDC_cal was generally lower than that of PROX_reg. Therefore, when condensing observed hydrographs into flow signatures,
405 preserving all available flow information in the hydrograph would be a key for a successful rainfall-runoff modelling. This study shows only using regionalised FDCs could lead to less reliable rainfall-runoff modelling in ungauged catchments than regionalised parameters. An FDC is unlikely to preserve all flow information in a hydrograph necessary for rainfall-runoff modelling.

5.2 Suggestions for improving RFDC_cal

410 Westerberg et al. (2014) suggested the necessity of further constraining to reduce predictive uncertainty in RFDC_cal. This study found that RFDC_cal could provide comparable performance to regenerate the flow signatures within which flow magnitudes are only involved (i.e., R_{QP} and I_{BF}). To supplement regional FDCs, flow signatures associated with flow timing seems to be essential. Figure 9 shows potential of additional constraining with D_{RL} , and Q_2 in Table 3 confirms it. Other flow signatures in temporal dimensions such as the high- and the low-flow event durations in Westerberg and McMillan
415 (2015) can be candidates to improve RFDC_cal. However, uncertainty in those flow signatures will be a challenge to build regional models for ungauged catchments (Westerberg et al., 2016).

An alternative method of RFDC_cal is to directly regionalise hydrographs to ungauged catchments (e.g., Viglione et al., 2013). In data-rich regions, topological proximity could better capture spatial variation of daily flows than rainfall-runoff modelling with regionalised parameters (Viglione et al., 2013). Although a dynamic model may be required for regionalising
420 observed daily flows at an expensive computational cost, flow timing information would be contained in regionalised hydrographs. The parameter identification against the regional hydrographs may become a better approach than RFDC_cal and/or other signature-based calibrations.

5.3 Limitations and future research directions

There are caveats in our comparative evaluation. First, uncertainty in input-output data was not considered in our assessment.
425 McMillan et al. (2012) reported typical ranges of relative errors in discharge data as 10-20% for medium to high flow and 50-100% for low flows. We assumed that quality of the discharge data was adequate. However, other methods objectively considering uncertainty could better estimate model performance and the equifinality (e.g., Westerberg et al., 2011, 2014). Second, we used a conceptual runoff model with a fixed structure for all the catchments. Uncertainty from the model structure would vary across the study catchments; nevertheless, the structural uncertainty was not measured here. Our
430 comparative assessment was based on the basic premise that modelling conditions should be fixed for all study catchments. Finally, though the proximity-based parameter regionalisation was good for the Korean catchments, comparison between RFDC_cal and other regionalisation methods, such as the regional calibration and the similarity-based parameter transfer, may provide beneficial information for rainfall-runoff modelling in ungauged catchments. Comparative assessment between RFDC_cal and other parameter regionalisation using more sample catchments under diverse climates will provide more
435 meaningful lessons.

We can no longer hypothesise that the parameters gained against regionalised FDCs would perform sufficiently, because an FDC contains less information than a hydrograph (i.e., the absence of flow timing). For improving RFDC_cal, we suggested to supplement RFDC_cal with flow signatures in temporal dimensions. Then, a question should be addressed on how to make flow signatures more informative than (or equally informative to) hydrographs. It may be impossible only using flow
440 signatures originated from hydrographs (e.g., mean annual flow, baseflow index, recession rates, FDCs, etc.). Combinations

of those signatures are unlikely more informative than their origins (i.e., hydrographs), though it depends on how much disinformation is present in the observed flows. Future research topics may include finding new signatures that supplement hydrographs, and how to combine them with existing flow signatures for rainfall-runoff modelling in ungauged catchments.

5.4 Conclusions

445 While the rainfall-runoff modelling against regional FDCs appeared a good approach for prediction in ungauged catchments, this study highlights its weakness in the absence of flow timing information, which may cause poorer predictive performance than the simple proximity-based parameter regionalisation. The following conclusions are worth emphasising:

- 450 (1) For ungauged catchments in South Korea where spatial proximity well captured functional similarity between gauged catchments, the model calibration against regional FDCs is unlikely to outperform the conventional proximity-based parameter transfer for daily runoff prediction. The absence of flow timing information in regional FDCs seems to cause a substantial equifinality problem in the parameter identification process and thus lower predictability.
- 455 (2) The model parameters gained from observed hydrographs would contain flow timing information even for ungauged catchments. This intangible flow timing information should be discarded if one calibrates a rainfall-runoff model against regional FDCs. This information loss may reduce predictability in ungauged catchments significantly.
- (3) To improve the calibration against regional FDCs, flow metrics in temporal dimensions, such as the rising limb density, need to be included as additional constraints. As an alternative approach, if river gauging density is high, regionalised hydrographs preserving flow timing information can be used for local calibrations at ungauged catchments.
- 460 (4) For better predictions in ungauged catchments, it is necessary to find new flow signatures that can supplement the observed hydrographs. How to combining them with existing information will be a future research topic for rainfall-runoff modelling in ungauged catchments.

Acknowledgements

This study was supported by APEC Climate Center. We send special thanks to Ms. Yoe-min Jeong and Dr. Hyungil Eum for 465 their PRISM climate data sets. Data needed to reproduce the modelling results in this study are available upon request from the authors (d.kim@apcc21.org, jachum@apcc21.org).

References

Atieh, M., Taylor, G., Sttar, A. M. A., Gharadaghi, B.: Prediction of flow duration curves for ungauged basins, *J. Hydrol.*, 545, 383-394, 2017.

- 470 Bae, D.-H., Jung, I.-W., and Chang, H: Long-term trend of precipitation and runoff in Korean river basins. *Hydrol. Process.*, 22, 2644-2656, 2008.
- Bárdossy, A.: Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, 11, 703-710, doi:10.5194/hess-11-703-2007, 2007.
- Beven, K. J., and Bingley, A.: The future of distributed models. Model calibration and uncertainty prediction, *Hydrol.*
475 *Process.*, 6, 279-298, 1992.
- Beven, K. J.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18-36, 2006.
- Beven, K.J.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16, 41-51, 1993.
- Blazkova, S., and Beven, K.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood
frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resour. Res.*, 45, W00b16,
480 doi:10.1029/2007wr006726, 2009.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H.: *Runoff Prediction in Ungauged Basins, Synthesis across Processes, Places, and Scales.* Cambridge University Press, New York, USA, 2013.
- Botter, G., Porporato, A., Rodriguez-Iturbe, I. and Rinaldo, A.: Basin-scale soil moisture dynamics and the probabilistic
characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response, *Water Resour.*
485 *Res.*, 43, W02417, doi:10.1029/2006WR005043, 2007.
- Botter, G., Zanardo, S., Porporato, A., Rodriguez-Iturbe, I., and Rinaldo, A.: Ecohydrological model of flow duration curves
and annual minima, *Water Resour. Res.*, 44, W08418, doi:10.1029/2008WR006814., 2008.
- Brooks, P. D., Troch P. A., Durcik, M., Gallo, E., and Schlegel, M.: Quantifying regional-scale ecosystem response to
changes in precipitation: Not all rain is created equal, *Water Resour. Res.*, 47, W00J08, doi:10.1029/2010WR009762, 2011.
- 490 Cheng, L., Yaeger, M., Viglione, A., Coopersmith, E., Ye, S., and Sivapalan, M.: Exploring the physical controls of regional
patterns of flow duration curves – Part 1: Insights from statistical analyses, *Hydrol. Earth Syst. Sci.*, 16, 4435-4446,
doi:10.5194/hess-16-4435-2012, 2012.
- Coopersmith, E., Yaeger, M. A., Ye, S., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patterns of
flow duration curves – Part 3: A catchment classification system based on regime curve indicators, *Hydrol. Earth Syst. Sci.*,
495 16, 4467-4482, doi: 10.5194/hess-16-4467-2012, 2012.
- Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., and Pasteris, P. P.:
Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States,
Int. J. Climatol., 28, 2031-2064, doi:10.1002/joc.1688, 2008.
- Demirel, M. C., Booiji, M. J., and Hoekstra, A. Y.: Effect of different uncertainty sources on the skill of 10 day ensemble
500 low flow forecasts for two hydrological models, *Water Resour. Res.*, 49, 4035-4053, doi:10.1002/wrcr.20294, 2013.
- Duan, Q., Sorooshian, S., and Gupta, V. K.: Effective and efficient global optimisation for conceptual rainfall–runoff models,
Water Resour. Res., 28, 1015–1031, 1992.

- Dunn, S. M. and Lilly, A.: Investigating the relationship between a soils classification and the spatial parameters of a conceptual catchment scale hydrological model, *J. Hydrol.*, 252, 157-173, doi:10.1016/S0022-1694(01)00462-0, 2001.
- 505 Eckhardt, K.: A comparison of baseflow indices, which were calculated with seven different baseflow separation methods, *J. Hydrol.*, 352, 168-173, doi: 10.1016/j.jhydrol.2008.01.005, 2007.
- Hingray, B., Schaefli, B., Mezghani, A., and Hamdi, Y.: Signature-based model calibration for hydrological prediction in mesoscale Alpine catchments, *Hydrolog. Sci. J.*, 55, 1002-1016, doi:10.1080/02626667.2010.505572, 2010.
- Hrachowitz, M. et al.: A decade of Predictions in Ungauged Basins (PUB) - A review. *Hydrolog. Sci. J.*, 58, 1198–1255.
510 Doi:10.1080/02626667.2013.803183, 2013.
- Jung, S. et al.: Standardisation of methods and criteria for hydrological survey in South Korea. Report Number: 11-1611492-00058-01. Ministry of Land, Transport, and Maritime Affairs. Gyeonggido, South Korea, Written in Korean, 2010.
- Jung, Y. and Eum, H.-I.: Application of a statistical interpolation method to correct extreme values in high-resolution gridded climate variables, *J. Clim. Chang. Re.*, 6, 331-334, 2016.
- 515 Kavetski, D., Fnlcia, F., and Clark, M.: Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment, *Water Resour. Res.*, 47, W05501, doi:10.1029/2010WR009525, 2011.
- Kim, D. and Kaluarachchi, J.: Predicting streamflows in snowmelt-driven watersheds using the flow duration curve method, *Hydrol. Earth Syst. Sci.*, 18, 1679-1693, doi:10.5194/hess-18-1679-2014, 2014.
- 520 Kim, U. and Kaluarachchi, J. J.: Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia, *J. Hydrol.*, 362, 39-56, doi:10.1029/j.jhydrol.2008.08.016, 2008.
- Korea Meteorological Administration: Climatological normals of Korea (1981–2010). Publ. 11-1360000-000077-14, 678 pp. Available at http://www.kma.go.kr/down/ Climatological_2010.pdf, 2011.
- McIntyre, N., Lee, H., Wheeler, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments,
525 *Water Resour. Res.*, 41., W12434, doi:10.1029/2005WR004289, 2005.
- McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge, and water quality, *Hydrol. Process.*, 26, 4078-4111, 2012.
- Mohamoud, Y. M.: Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves, *Hydrolog. Sci. J.*, 53, 706–724, 2008.
- 530 Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungauged catchments: a comparison between of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, W03413, doi:10.1029/2007WR006240, 2008.
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C. : Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 – Towards a simple and efficient potential
535 evapotranspiration model for rainfall-runoff modelling. *J. Hydrol.*, 303, 290-306, 2005.

- Parajka, J., Blöschl, G., and Merz, R.: Regional calibration of catchment models: potential for ungauged catchments, *Water Resour. Res.*, 43, W06406, DOI: 10.1029/2006WR005271, 2007.
- Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, *Hydrol. Earth Syst. Sci.*, 17, 1783-1795, doi:10.5194/hess-17-1783-2013, 2013.
- 540 Perrin, C., Michel, C., Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, 279, 275-289, 2003.
- Pfannerstill, M., Guse, B., and Fohrer N.: Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, *J. Hydrol.*, 510, 447-458, 2014.
- 545 Pugliese, A., Castellarin, A., and Brath, A.: Geostatistical prediction of flow-duration curves in an index-flow framework, *Hydrol. Earth Syst. Sci.*, 18, 3801-3816, doi:10.5194/hess-18-3801-2014, 2014.
- Pugliese, A., Farmer, W. H., Castellarin, A., Archfield, S. A., Vogel, R. M.: Regional flow duration curves: Geostatistical techniques versus multivariate regression, *Adv. Water Resour.*, 96, 11-22, 2016.
- Rhee, J., and Cho, J.: Future changes in drought characteristics: regional analysis for South Korea under CMIP5 projections. *J. Hydrometeorol.*, 17, 437-450, 2016.
- 550 Sadegh, M., Vrugt, J. A., Gupta, H. V., and Xu, C.: The soil water characteristics as new class of closed-form parametric expressions for the flow duration curve, *J. Hydrol.*, 535, 438-456, 2016.
- Shu, C. and Ouarda, T. B. M. J.: Improved methods for daily streamflow estimates at ungauged sites, *Water Resour. Res.*, 48, W02523, doi:10.1029/2011WR011501, 2012.
- 555 Smakhtin, V. P. and Masse, B.: Continuous daily hydrograph simulation using duration curves of a precipitation index, *Hydrol. Process.*, 14, 1083-1100, 2000.
- Smakhtin., V. Y., Hughes, D. A., and Creuse-Naudine, E.: Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa, *Hydrolog. Sci. J.*, 42, 919–936, 1997.
- Sugawara, M.: Automatic calibration of the tank model, *Hydrological Sciences Bulletin*, 24, 375-388, doi:10.1080/02626667909491876, 1979
- 560 Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria, *Hydrol. Earth Syst. Sci.*, 17, 2263-2279, doi: 10.5194/hess-17-2263-2013, 2013.
- Vrugt, J. A., and Ter Braak, C. J. F.: DREAM(D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems, *Hydrol. Earth Syst. Sci.*, 15, 3701–3713, doi:10.5194/hess-15-3701-2011, 2011.
- 565 Wagener, T. and Wheater, H. S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, 320, 132-154, 2006.

- Walter, M. T., Brooks, E. S., McCool, D. K., King, L. G., Molnau, M., and Boll, J.: Process-based snowmelt modeling: does it require more input data than temperature-index modeling?, *J. Hydrol.*, 300, 65-75, doi: 10.1016/j.jhydrol.2004.05.002, 2005.
- Westerberg, I. K. and McMillan, H. K.: Uncertainty in hydrological signatures, *Hydrol. Earth Syst. Sci.*, 19, 3951-3968, doi: 10.5194/hess-19-3951-2015, 2015.
- Westerberg, I. K., Gong, L., Beven, K. J., Seibert, J., Semedo, A., Xu, C.-Y., and Halldin, S.: Regional water balance modelling using flow-duration curves with observational uncertainties, *Hydrol. Earth Syst. Sci.*, 18, 2993-3013, doi:10.5194/hess-18-2993-2014, 2014.
- Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halladin, S., Freer, J. E., and Xu, C.-Y.: Calibration of hydrological models using flow-duration curves, *Hydrol. Earth Syst. Sci.*, 15, 2205-2227, doi:10.5194/hess-15-2205-2011, 2011.
- Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., and Freer, J.: Uncertainty in hydrological signatures for gauged and ungauged catchments, *Water Resour. Res.*, 52, 1847-1865, doi:10.1002/2015WR017635., 2016
- Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, 30, 1756-1774, 2007.
- Yaeger, M., Coopersmith, E., Ye, S., Cheng, L., Viglione, A., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 4: A synthesis of empirical analysis, process modeling and catchment classification, *Hydrol. Earth Syst. Sci.*, 16, 4483-4498, doi: 10.5194/hess-16-4483-2012, 2012.
- Ye, S., Yaeger, M., Coopersmith, E., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 2. Role of seasonality, the regime curve and associated process controls, *Hydrol. Earth Syst. Sci.*, 16, 4447-4465, doi:10.5194/hess-16-4447-2012, 2012.
- Yokoo, Y. and Sivapalan, M.: Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis, *Hydrol. Earth Syst. Sci.*, 15, 2805–2819, doi:10.5194/hess-15-2805-2011, 2011.
- Yu, P. S., Yang, T. C., and Wang, Y. C.: Uncertainty analysis of regional flow duration curves, *J. Water Resour. Plann. Manage.*, 128, 424–430, 2002.
- Yu, P.-S., and Yang, T.-C.: Using synthetic flow duration curves for rainfall–runoff model calibration at ungauged sites. *Hydrol. Process.*, 14: 117–133. doi:10.1002/(SICI)1099-1085(200001)14:1<117::AID-HYP914>3.0.CO;2-Q, 2000.
- Zhang, Y., Vaze, J., Chiew, F. H. S., and Li, M.: Comparing flow duration curve and rainfall-runoff modelling for predicting daily runoff in ungauged catchments, *J. Hydrol.*, 525, 72-86, 2015.
- Zhang, Z., Wagener, T., Reed, P., and Bhushan, R.: Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, *Water Resour. Res.*, 44, W00B04, doi:10.1029/2008WR006833, 2008.

Table 1: Summary of hydrological features of the study catchments

	Average	CV	minimum	25%	median	75%	Maximum
Area (km ²)	890	1.39	57	208	495	1013	6705
Elevation (m a.s.l.)	339	0.63	39	193	255	495	996
Mean annual prcp. (mm yr ⁻¹)	1359	0.14	1145	1247	1286	1388	1997
Mean annual temp. (°C)	11.9	0.13	7.9	11.3	12.3	13.0	13.8
Aridity index ¹ (-)	0.66	0.11	0.44	0.61	0.68	0.71	0.76
P _{snow} ²	35	0.66	6	23	28	50	141
Mean annual flow (mm yr ⁻¹)	739	0.25	232	624	740	838	1159
R _{PQ} (-)	0.55	0.27	0.18	0.45	0.54	0.63	0.91
I _{BF} (-)	0.49	0.16	0.27	0.44	0.49	0.56	0.62
D _{RL} (day ⁻¹)	0.63	0.10	0.50	0.60	0.63	0.66	0.77

605 ¹Ratio of potential ET to total precipitation, ²Percentage of snowfall to total precipitation. Climatological features were calculated using spatial averages of the grid data, while the flow metrics were from the daily hydrographs for 2007-2015 as explained in Section 3.6.

Table 2: Ranges of GR4J parameters used for parameter calibration (Demirel et al., 2013)

Parameter	Range
X1 (mm)	10 to 2000
X2 (mm)	-8 to +6
X3 (mm)	10 to 500
X4 (days)	0.5 to 4.0

610

Table 3: Results of the paired t-tests for potential questions on rainfall-runoff modelling in ungauged catchments

Questions	Corresponding pair	¹ PM	² Δ PM	³ std. err.	Answer
Q1. Did RFDC_cal outperform PROX_reg?	RFDC_cal – PROX_reg	NSE	-0.22	0.051	No*
Q2. Did D _{RL} improve FDC_cal?	FDC+DRL_cal – FDC_cal	NSE	0.12	0.025	Yes*
Q3. Did parameters from Q_cal contain flow timing information for ungauged catchments?	PROX_reg – FPROX_reg	NSE	0.10	0.031	Yes*
Q4. Did absence of flow timing affect model efficiency?	Q_cal – FDC_cal	NSE	0.23	0.026	Yes*
Q5. Did PROX_reg outperform RFDC_cal in predicting low flows?	PROX_reg – RFDC_cal	LNSE	0.09	0.031	Yes*
Q6. Did PROX_reg outperform RFDC_cal in reproducing I _{BF} ?	PROX_reg – RFDC_cal	D _{FS} (I _{BF})	0.06	0.028	No
Q7. Did errors in regional FDCs affect RFDC_cal significantly?	RFDC_cal – FDC_cal	NSE	-0.09	0.069	No

¹Performance metric used for t-test, ²Mean PM difference between the corresponding pair, ³Standard error of Δ PM. * Δ PM is significantly different from zero. The significance was evaluated at 95% confidence levels.

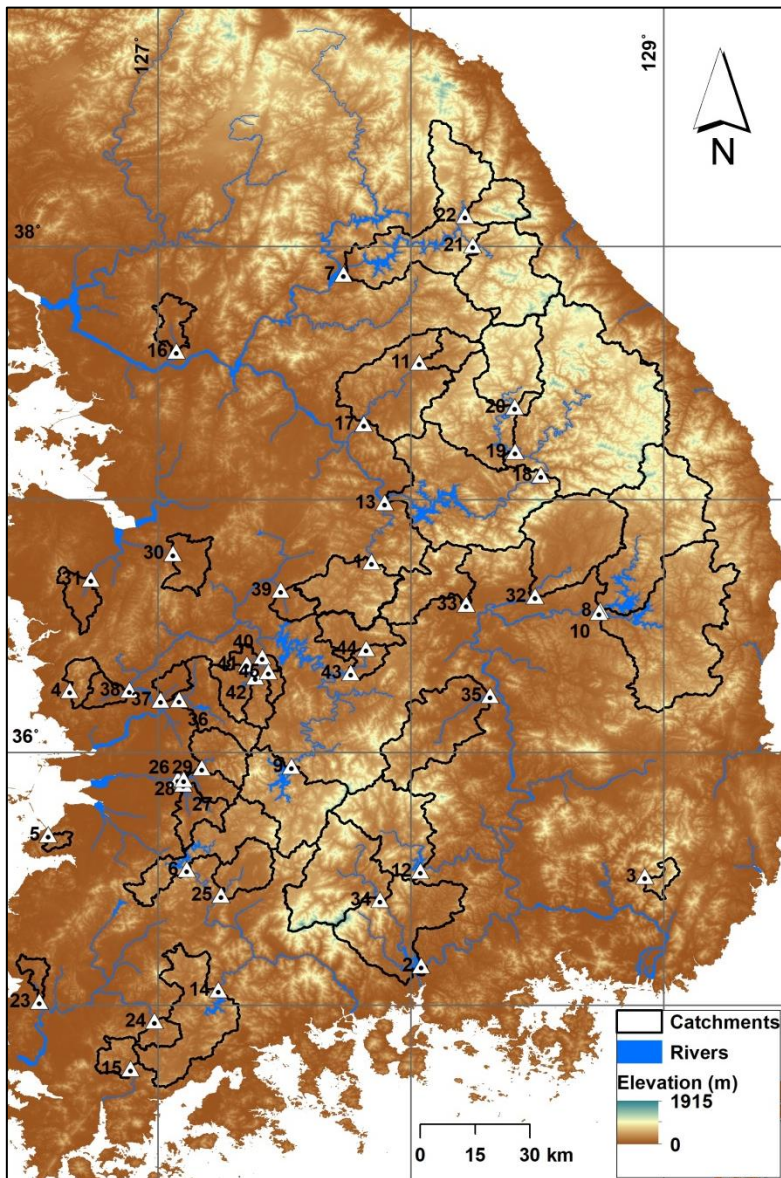
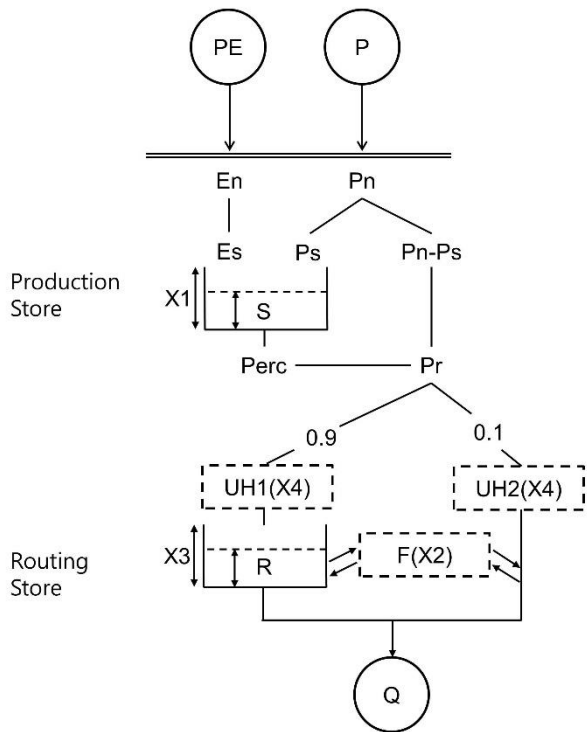
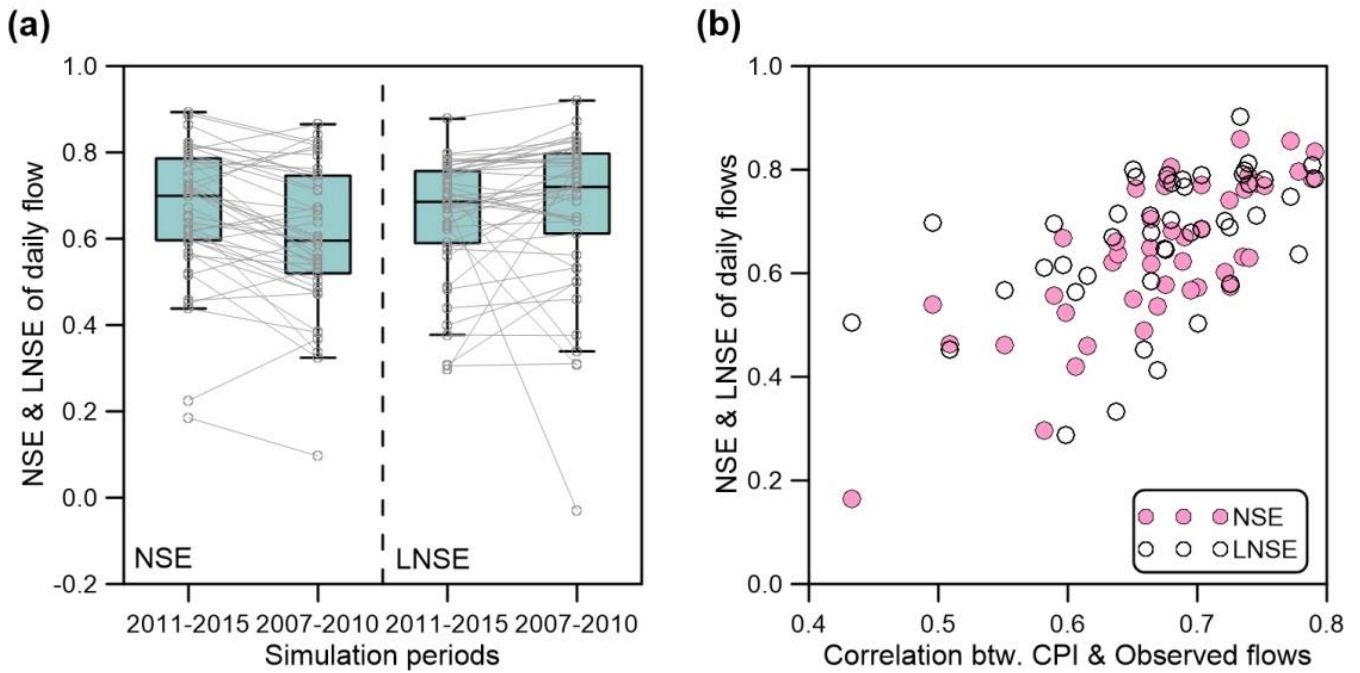


Figure 1: Locations of the study catchments in South Korea. The numbers are labelled at the outlet of each catchment.



620 **Figure 2: The schematised structure of GR4J (X1-X4: model parameters, PE: potential evapotranspiration, P: precipitation, Q: runoff, other letters indicate variables conceptualising internal catchment processes).**



625 **Figure 3: (a) box plots of high flow (NSE) and low flow (LNSE) reproducibility of the behavioural parameters obtained from the hydrograph calibration at the 45 catchments, (b) the relationship between the input-output consistency and the model performance. The straight lines in the box plots connect the performance metrics for the calibration (2011-2015) and the validation periods (2007-2010) in each catchment.**

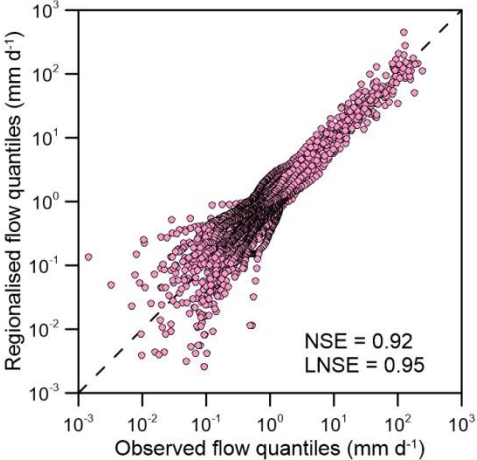


Figure 4: 1:1 scatter plot between the empirical flow quantiles and the flow quantiles predicted by the top-kriging FDC regionalisation method.

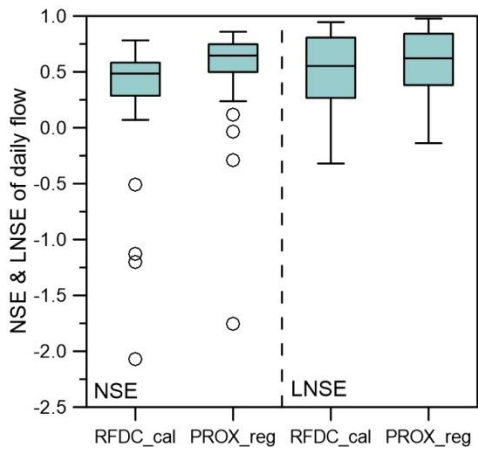
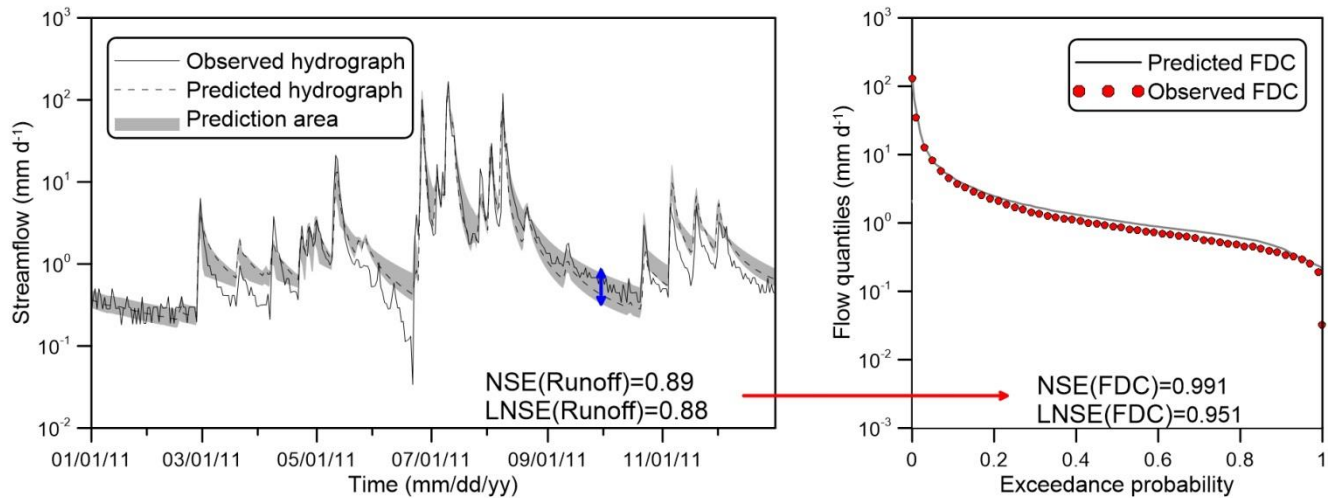
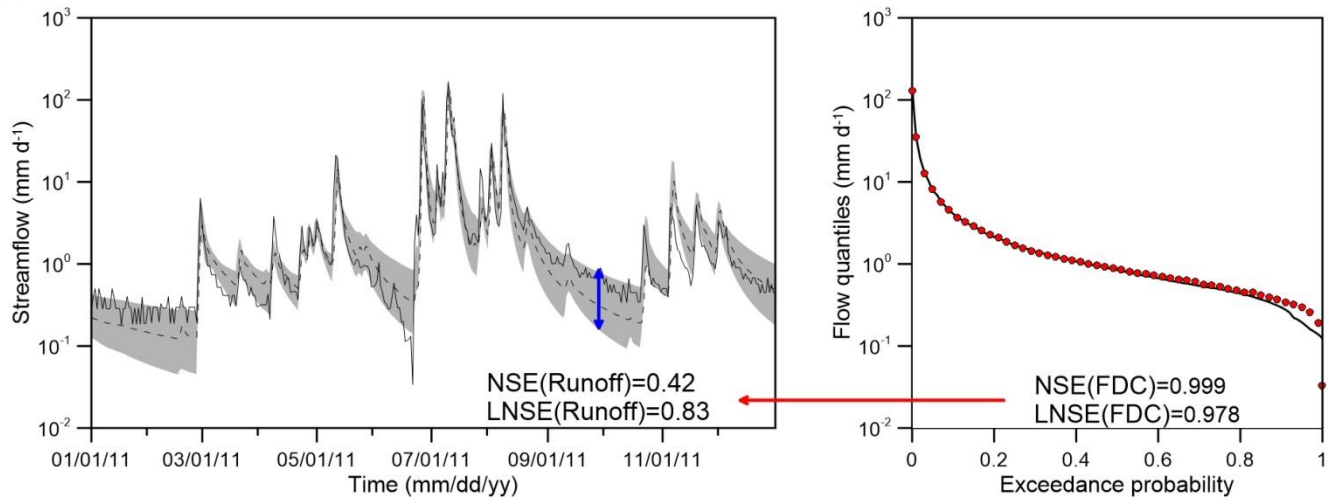


Figure 5: Box plots of NSE and LNSE values between the observed and the predicted hydrographs by RFDC_cal and PROX_reg for the 45 catchments under the cross validation mode.

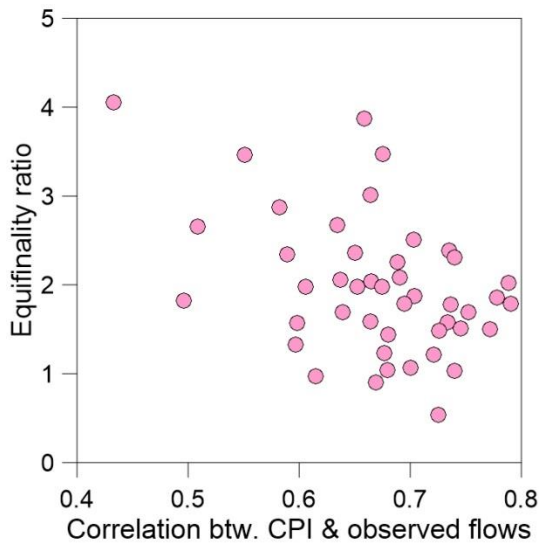
(a) Hydrograph calibration



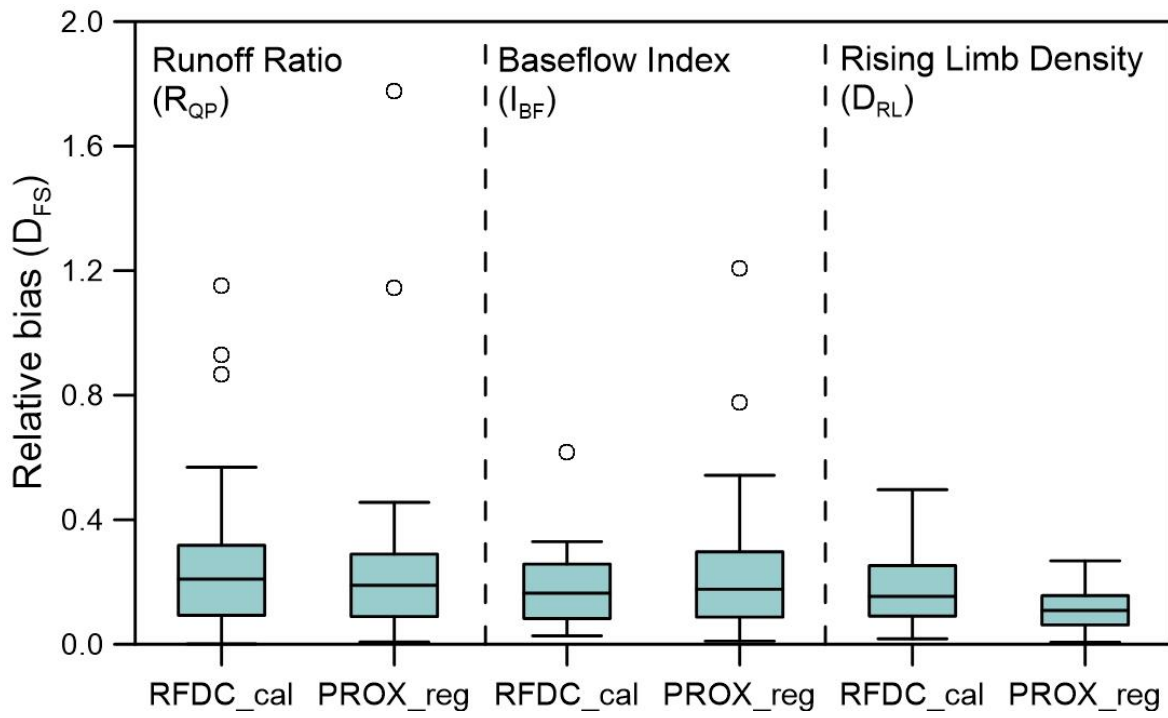
(b) FDC calibration



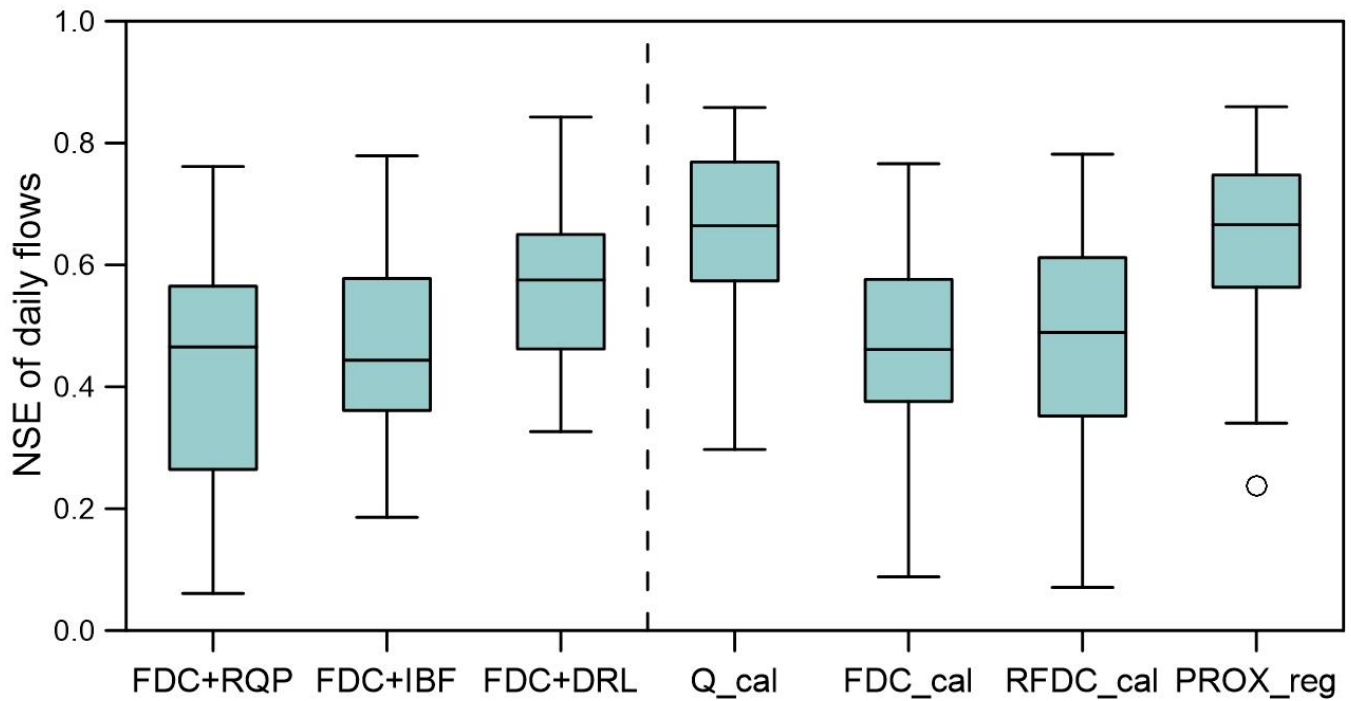
645 **Figure 6: The observed and predicted hydrographs, the prediction areas, and the observed and predicted FDCs given by (a) the hydrograph calibration and (b) the FDC calibration for the Catchment 2.**



650 **Figure 7: The input-output consistency vs. equifinality increased by replacing the hydrograph calibration with the FDC calibration. The equifinality ratio is defined as the ratio between the prediction areas of the 50 behavioural parameters gained from the FDC calibration and the hydrograph calibration.**



655 **Figure 8:** Flow signature reproducibility comparison between RFDC_cal and PROX_reg in terms of R_{QP} (a), I_{BF} (b), and D_{RL} (c).



660 **Figure 9: Predictive performance of the FDC calibrations additionally conditioned by R_{QP} (FDC+RQP), I_{BF} (FDC+IBF), and D_{RL} (FDC+DRL) in comparison to the other modelling approaches. Q_{cal} and FDC_{cal} refer to the hydrograph and the FDC calibration in gauged catchments respectively. 38 catchments with positive NSEs for all the modelling approaches were used in the box-plots.**