Dear Dr. Fabrizio Fenicia

First, let us thank for your efforts in handling our manuscript. We greatly appreciate the constructive comments from you and the anonymous referees. We believe all the comments were helpful to improve the quality of our work.

In this revision, to improve the clarity of this study, we highlighted that the scientific meaning of the model calibration against regional flow duration curves (RFDC_cal), and clearly stated the objective of this study to compare RFDC_cal with a classical parameter regionalization. We introduced strengths of RFDC_cal in line 37-72. Then, we addressed potential questions that can arise when applying RFDC_cal in practice in line 73-84. We emphasized that RFDC_cal has barely compared with conventional parameter regionalization schemes in line 85-87. If RFDC_cal has poorer predictability than the proximity-based parameter transfer (PROX_reg), RFDC_cal would not be pragmatic. The main research question of this study is whether RFDC_cal outperform PROX_reg. We addressed this question by applying two methods to 45 Korean catchments in the jackknife cross validation mode.

As shown in the previous version of our manuscript, RFDC_cal was likely to have weaker predictability than PROX_reg due to the absence of flow timing information in regional FDCs. And, we argued that flow signatures in temporal dimensions should supplement RFDC_cal. In the revision, we attempted one more parameter regionalization that transfers the parameters gained against observed FDCs to ungauged catchments. This approach cannot transfer flow timing information through the model parameters from gauged to ungauged catchments (we referred this approach to as FPROX_reg), because the behavioral parameters were gained against flow magnitudes only. We found that PROX_reg significantly outperformed FPROX_reg via a paired t-test between them. This implies that PROX_reg could transfer flow timing information to ungauged catchments, while it is impossible when using RFDC_cal. In section 4.4 (from line 348), you can find the results of several paired t-tests between modeling approaches applied in this study. We believe they provide clearer indications about performance of RFDC_cal.

In addition, as an alternative method of RFDC_cal in data-rich regions, we suggested use of regional hydrographs (e.g., Viglione et al., 2013) to preserve flow amount and timing information together. And, we emphasized that preserving all flow information inherent in hydrographs would be a key for rainfall-runoff modeling against flow metrics that condense the hydrographs. You may find this context in section 5. To make the manuscript more concise, we combined the discussion and the conclusion sections.

We believe this revision can provide clear lessons and readability. Following are our responses to specific comments from the referees. Again, we thank for all of your editing efforts.

Sincerely,

Jong Ahn Chun
Corresponding author

Response to comments from reviewer 1:

Major comments: The first objective in this study, as stated on page 4 lines 8-10, is to evaluate predictive performance of the hydrograph calibration and the FDC calibration as well as their uncertainty for gauged catchments. I think this idea has been addressed extensively in the literature (some of which are cited in the present manuscript), and therefore, it does not need any further examination. The fact that this study finds FDC-based calibration less promising than hydrograph-based approach (as stated on page 11 lines 13-15) is not of a big surprise, e.g., due to different challenges in FDC estimation and that timing is not handled by FDC, as authors point out in the manuscript as well. Probably, what is more worth studying is how FDC can help to reduce equi-finality. As a result, I suggest that authors remove the first part of the study, or consider FDC as an additional criteria in model calibration and show how its use would improve parameter identifiability (e.g., posterior ranges) and reduce uncertainty (e.g., uncertainty ratio of hydrograph+FDC to only hydrograph).

➔ *We globally revised the manuscript to provide clearer lessons from this study.*
We agreed that it was not a surprise that the FDC calibration has more equifinality than the hydrograph calibration. Therefore, we focused on comparing RFDC_cal and PROX_reg for ungauged catchments (i.e. we removed the comparative assessment for gauged catchments in the previous version).
We did not consider the second option to use FDCs as an additional criterion, because it is already proposed by Pfannerstill et al. (2014).Instead, in this revision, we added one more regionalization approach that transfers parameters gained from observed FDCs to ungauged (FPROX_reg) in order to check whether PROX_reg transfers flow timing information for ungauged catchments. FPROX_reg uses parameters gained from flow magnitudes only, thus it cannot transfer flow timing information to ungauged catchments. A paired t-test showed that the performance difference between PROX_reg and FPROX_reg was significant (i.e., parameters gained from flow magnitudes only may cause predictability losses).
Through several paired t-tests, we found a clearer indication that PROX_reg is better than RFDC_cal for the Korean catchments. We believe that this revision can provide you clear indications.

Authors claim that FDC calibration performs promising for low flow prediction. I would argue that FDC-based approach performs only better than hydrograph-based approach, not good overall. Looking at figure 9, I see that there are several large deviations between simulated and observed BFI (up to 90%) which means that FDC-based method is not that reliable.

➔ *In revision, we withdrew this argument. RFDC_cal could provide better predictability in low flows than high flows due to smaller variability in base flow; however, it was unlikely that RFDC_cal outperformed PROX_reg in low flows. However, it is unclear that PROX_reg outperform RFDC_cal in reproducing BFI as addressed in Q6 in Table 3.*

My other major issue is with how authors set the experiments related to streamflow predictions in ungauged catchments. They first mention three classes of parameter regionalisation in lines 26-30 on page 8, but then mention that they chose the proximity based approach due to its simplicity. I think, given than the first part of the paper can be removed according to my view, authors should focus more on this part and compare different regionalization approaches.

➔ *In section 3.5, we addressed why the proximity-based parameter regionalization was chosen. Modeling conditions in this study were suitable to use PROX_reg. Other regionalization such as similarity-based or regression-based regionalization can be applied too, but our focus was comparing RFDC_cal with the simplest parameter regionalization.*

Also, why not considering the proximity-based transfer of FDCs from donor catchments as an additional approach? Then, a potential topic for the paper can be "comparative evaluation of different regionalization approaches for model calibration in ungauged catchment".

➔ *The geostatistical method applied in this study is a proximity-based transfer (or interpolation) of empirical FDCs. We already transferred observed FDCs to ungauged catchment using the top-kriging weights. And, it showed promising performance for predicting FDCs in ungauged catchments as addressed in section 4.1. The focus of this study is comparison between RFDC_cal and PROX_reg for rainfall-runoff modeling in ungauged catchments.*

Page 7 line 15 says that "Synthetic runoff time series were generated by GR4J for the same 45 catchments by treating each catchment as ungauged.

➔ *Nothing was requested. We globally reviewed the manuscript and used the term "LOOCV mode" to distinguish between approaches for gauged and ungauged catchments.*

Introduction needs to be shorter. Objectives are stated after 6 very long paragraphs in the introduction section. Moreover, discussions sub-sections are too long. I think authors can make them briefer, but still transfer the message to readers.

➔ *In revision, we highlighted the scientific meaning of RFDC_cal in comparison to PROX_reg. The main objective of this study is a comparative assessment of RFDC_cal.*

Minor comments (for improving manuscript quality):

I suggest continuous line numbering in the next version of the manuscript.

➔ *For convenience, we used continuous line numbers in the revised manuscript.*

Page 3, line 34: I suggest that a little explanation is provided here about the proximity-based approach. It is not clear up to this point what that approach actually is. Authors provide a brief description on page 7 line 17. Also, I suggest removing "in truth"

- ➔ *We globally revised the manuscript, and PROX_reg was addressed in section 3.5. We removed the term "in truth".*

Also related to the description of proximity-based approach, section 3.3.2 is not fully understandable. I suggest rewording the paragraph so that the approach is explained in a clearer way. Moreover, please explain at the beginning of this section that when you talk about parameters in the proximity-based approach, you actually mean the parameters of the hydrologic model. Because one can also estimate the parameters of a parametric FDC using this approach.

- ➔ *PROX_reg is now addressed in section 3.5. From line 227, we explained how we transferred behavioral parameters from gauged to ungauged catchments.*

Page 9 line 1: what do you mean by "synchronizing" donor catchments?

- ➔ *It means that we used same donor catchments for the regional FDC and the parameter regionalization. It was for consistency between PROX_reg and RFDC_cal as explained in line 228.*

Page 4 line 3: define "orthogonal"

- ➔ *In revision, we did not use the term "orthogonal".*

Please explain why Monte Carlo is used for parameter estimation, whereas SCE has been used by authors in one of the catchments. I believe that there is the possibility of quantifying uncertainty bounds using the solutions sampled by SCE.

- ➔ *The Monte-Carlo framework was good for us to gauge equifinality across all catchments under the same sampling size and the acceptance rate, though there are other methods for individual catchments. This approach was good to evaluate equifinality under changing input-output consistency across the 45 catchments. It is explained in line 169-175.*

Page 12 line 26-28: the sentence is not understandable. Please reword.

- ➔ *In revision, we did not use this sentence.*

Response to comments from reviewer 2:

The work explores the predictive performance of application of a FDC in comparison with conventional hydrograph calibration and parameter regionalisation for gauged and ungauged catchments. While the manuscript has some interesting results and discussion, it is not clear to me from the text how the work is innovative and unique to the previous studies mentioned in the literature review and discussion. For this reason I suggest major review to lift the manuscript before the work is suitable for publication in HESS. To me the manuscript currently lacks focus in the sense that the key research gaps and innovation should stand out more clearly in the introduction and conclusion. In my opinion the authors should focus on quality and innovation rather than applying existing techniques, and quantity of results and discussion.

- ➔ To improve the clarity of this study, we addressed strengths of RFDC_cal in comparison to the classical parameter regionalization in line 37-72. Then, in line 73-84, we addressed potential questions when applying RFDC_cal in practice. If RFDC_cal has poorer predictability than the proximity-based parameter transfer (PROX_reg), RFDC_cal would not be pragmatic because regional FDC may require expensive efforts. The main research question of this study is whether RFDC_cal outperform PROX_reg for ungauged catchments. We believe the new introduction shows objectives of this study more clearly. In addition, we added the section of paired t-tests for checking our hypotheses. It was emphasized that the flow timing information embedded in parameters gained against observed hydrographs affects predictability for ungauged catchments.

Major comments: The innovation of this work compared to previous studies is not clear to me. Could the authors please state explicitly the innovation of their work compared to previous FDC regionalisation studies and existing methods? The specific research gap/s that the work is addressing should be more prominent in the introduction, and the innovations compared to previous studies need to be more prominent in the summary and conclusions section.

- ➔ *As answered above, the new introduction is now focused on evaluating RFDC_cal in comparison to PROX_reg, which has been barely addressed in previous similar studies. We added paired t-tests between modelling approaches applied in this study. And, we argued that flow timing information can play an important role in prediction even in ungauged catchments. You can find this context throughout the revised manuscript. We also provide a suggestion that regional hydrographs, instead of regional FDCs, would be better to preserve flow timing information for calibration of rainfall-runoff models in ungauged catchments.*

Could the authors also please describe in detail how you improve on your previous 2016 submission to HESS that uses the same 45 South Korean catchments and has a similar goal: “Kim et al. A comparison between parameter regionalization and model calibration with flow duration curves for prediction in ungauged catchments”. Reading the comments from the reviewers on the previous submission there are some points that have not been fully addressed in this submission.

- ➔ *Here, we briefly summarize how we considered the comments given by the previous review process. We believe the comprehensive comments were considered in the revision generally. For example, actual constraining with flow signatures, and replacing the objective function, evaluating low and high flows were considered in the manuscript.*
- ➔ *The referee 1 mainly argued that our study had limited contribution to prediction in ungauged basins because of existing FDC methods for runoff prediction. However, the objective of our study was not to provide a new FDC-based runoff prediction, but a comparative evaluation between existing methods. Hence, we disagreed. The referee 1 also argued that it is no surprise with low performance of the FDC calibration. However, we cannot assure it in the case of ungauged catchments, thus we disagreed. The small number of gauged catchments was pointed out; however, 45 is not a large number, but some parameter regionalization studies used even smaller samples. The reviewer 1 argued that the objective function of NSE is not practical because of its emphasis on high flows. We replaced the objective function with one proposed by Zhang et al. (2015) that considers NES and WBE together. And, we considered all catchments for regionalization instead of only using high performance catchments. Other minor comments were considered as well.*
- ➔ *The referee 2 recommended us to soften conclusions that PROX_reg is better than the other. Nevertheless, in the revision, it was necessary to highlight that RFDC_cal is not as good as PROX_reg, because we received clearer indications that flow timing information in gauged catchments plays an important role in prediction in ungauged catchments too. Use of multiple criteria was recommended as well, thus we used NSE and LNSE together in revision. Some minor suggestions for title, tables, and context were given together. We added new figures and tables. And, the manuscript is retitled.*
- ➔ *The referee 3 provided constructive comments, asking first "why not parameter regionalization gained from observed FDCs?" We did consider this comment to check whether parameters gained against hydrographs can outperform those from FDCs in ungauged catchments. As mentioned, the former significantly outperformed the latter, implying that flow timing information for ungauged catchments might be contained in the parameters from observed hydrographs. The referee3 also suggested including uncertainty evaluation for both approaches for ungauged catchments. The equifinality evaluation using the Monte-Carlo simulations provides a lesson that uncertainty of the FDC calibration would be much larger than in the hydrograph calibration, though this evaluation was not a direct uncertainty comparison between RFDC_cal and PROX_reg. Referee3 also argued that there is no evidence that the rising limb density can supplement the FDC. Hence, we provided actual calibration results conditioned by the rising limb density. This could lend support to the hypothesis. With some minor comments, it was asked to provide more specific examples using flow signatures in runoff modeling. So, we improved the introduction with more literatures about use of FDCs in model calibrations.*

I suggest adding either "ungauged" or "regionalisation " to the title of the manuscript to make the title more descriptive of the work undertaken in the manuscript.

➔ *We agree. We retitled the manuscript as "A comparative assessment of rainfall-runoff modelling against regional flow duration curves for ungauged catchments".*

Minor comments: In the future please line number the manuscript continuously e.g. 1-999 rather than by each page, this will aid the review process.

➔ *Now we used continuous line numbers.*

The first paragraph of Section 3 introduces the GR4J model, and I see no logical progression to Section 3.1. I recommend an opening paragraph describing the structure of the methodology and turning your current paragraph into a new Section e.g. "3.1 Hydrological model (GR4J)". Furthermore I suggest a second section e.g. "3.2. Flow duration curve (FDC)" for consistency and to ensure reproducibility of your work.

➔ *We considered this comment to improve readability of the methodology section.*

Can you clarify in page 9, lines 4-7 your justification for applying a different objective function for calibration (Eq. 2a, 2b, 2c) OBJ, to the functions used to evaluate predictive performance (Eq. 5) NSE and LNSE?

➔ *The objective function was to consider high-flow reproducibility and long-term water balance in model calibration. NSE and LNSE were to evaluate model predictability in high and low flows. They are addressed in line 156 and 238, respectively.*

Page 10, Line 12 I disagree that the term NSE was used "directly" for calibration, rather I understand that you used a combination of the NSE and the WBE in OBJ. Please clarify.

➔ *We provide new results and discussion sections. This sentence was removed.*

Figure 3: I suggest adding headings "GR4J", and "FDC" to the top panels to ease interpretation.

➔ *Now, Figure 5 compares between RFDC_cal and PROX_reg. GR4J and FDC do not distinguish the two approaches for ungauged catchments.*

Figure 4: If these are 1:1 plots then I suggest adding a 1:1 line to the panels to ease interpretation.

➔ *They were not 1:1 plots. They display the relationship between input-output consistency and model performance. Now it is combined in Figure 3(b) only for the hydrograph calibration.*

Figure 5. Where is the difference between the first and second column of panels described in the caption or figure? I suggest adding headings to describe the difference in a similar manner to my recommendation for Figure 3.

➔ *Instead, we provided Figure 6 to emphasize the equifinality in FDC_cal.*

Could you please provide a more professional title (i.e. remove the phrase "performs good") to Subsection 5.2? e.g. "performs well", or a new title "Suitability of the FDC calibration for prediction of low flows"

- ➔ *Now we mainly focused on comparing RFDC_cal and PROX_reg rather than the performance of the FDC regionalization. Accordingly, we revised all headings.*

In Figure 10a it is very difficult to see the difference between observed and modelled FDCs. If this result is presented then could the authors provide an inset zoom to allow the reader to see the difference between the FDCs for the highest flows?

- ➔ *We did not use this figure in revision.*

Please proof read future submissions in greater detail, see some notes below. Typos and clarifications: Abstract line 11: ". . .Monte-Carlo framework. . ." is a bit vague given the complexity of your calibration (e.g. initial use of the SCE) please be more descriptive.

- ➔ *We rewrote the abstract.*

Page 1, Line 2: Should we not have an "and"?

- ➔ *The given form is unlikely to use "and" between author names.*

Page 2, Line 9: Should "has" be replaced with "is"?

- ➔ *We restructured the introduction.*

Page 2, Line 15: In the papers that you refer to in the previous sentence (i.e. Beven 2006), the term used is "equifinality" rather than "equi-finality". As this is a widely used term in the field of hydrological modelling I think that this consistency is important. Furthermore, the paper referenced (Oudin, 2008) does not refer to the term "equifinality", and so I feel that you may wish to choose a reference that better reflects the implication of the sentence.

- ➔ *We used "equifinality" in the revision. Oudin et al. (2008) did not use the term "equifinality" literally; however, they pointed out that "most models have been shown to have no unique set of parameters to define the best model fit to the flow response of a catchment" (in paragraph 3). In the context, we could find equifinality is an important uncertainty source when extrapolating parameters to ungauged catchments. Thus, we cited it.*

Page 4, Line 3: Please clarify what you mean by "orthogonal" here

- ➔ *In revision, we did not use the term "orthogonal"*

Page 4, Line 13: Why have you used the term "simply"? I suggest removing it.

- ➔ *We removed it.*

Page 4, Line 18: “Characterized”, previously you have used UK English rather than US English, e.g. Page 4, Line 7 “regionalisation“. Another e.g. Figure 1 caption “regionalization”. Another Page 8, Line 25: “regionalization”. Another example when you refer to Figure 2 you use “schematized”, but in the Figure 2 caption you use “schematised”. Please be consistent throughout the paper.

- ➔ *We globally reviewed the expressions.*

Page 4, Line 32: typo “Mistry”, should be “Ministry”

- ➔ *We corrected it.*

Page 7, Line 25: Please choose an alternative wording to: “and thus of consistency”, e.g. “and therefore are consistent”

- ➔ *The context in this sentence is now moved to section 3.5 in line 228.*

Page 8, line 10: “50 parameter sets” I recommend adding “. . .from the Monte-Carlo. . .” to remind the reader what you are referring to here.

- ➔ *We added it in line 316 where it is necessary.*

Page 10, Paragraph starting with line 22. Please clarify what correlation coefficient you are referring to. -->I.e. Pearson correlation.

- ➔ *In revision, we clearly stated “Pearson” correlation coefficient where it is necessary.*

Page 16, line 15. I am not sure if the word “Obviously” is necessary here. How is this future work more “obvious” than the other limitations that you have discussed above? I suggest removing it.

- ➔ *In revision, it was removed.*

Table 1: Typo: “Draiage”

- ➔ *In revision, it was removed.*

References

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: a comparison between of regionalization approaches based on 913 French catchments, Water Resour. Res., 44, W03413, doi:10.1029/2007WR006240, 2008.

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.

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.

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.

~~Comparative evaluation~~A comparative assessment of rainfall-runoff modelling against regional flow duration curves ~~in semi-humid~~for ungauged catchments

Daeha Kim[1], Ilwon Jung[2], Jong Ahn Chun[1]

[1]APEC Climate Center, Busan, 48058, South Korea
[2]Korea Infrastructure Safety & Technology Corporation, Jinju, Gyeongsangnam-do, 52852, South Korea

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

Abstract. ~~Streamflow prediction using rainfall~~Rainfall-runoff ~~models~~modelling has long been a special subject in hydrological sciences, ~~and parameter identification is still challenging~~but identifying behavioural parameters in ungauged catchments~~.~~ is still challenging. In this study, we comparatively evaluated ~~predictive power~~performance of the local calibration of a rainfall-runoff ~~modelling~~model against ~~the~~regional flow duration ~~curve~~curves (FDC), which is ~~gaining attention as signature-based parameter identification, by comparing it with conventional hydrograph-based approaches for gauged and ungauged catchments. Using a parsimonious model GR4J under~~ a ~~Monte Carlo framework, we conducted rainfall-runoff modelling against observed hydrographs and empirical FDCs for 45 gauged catchments in South Korea. By treating each catchment as ungauged, we compared again between parameter calibration against regional FDCs and proximity-based~~seemingly alternative method of classical parameter regionalisation ~~in terms of hydrograph and flow signature reproducibility. Results showed that the FDC calibration could lead to noticeably weaker performance and higher uncertainty in predictions in gauged catchments due to the absence of flow timing.~~for ungauged catchments. We used a parsimonious rainfall-runoff model over 45 Korean catchments under semi-humid climate. The calibration against regional FDCs~~, which were estimated by a geostatistical method, also showed weaker performance than the~~ was compared with the simple proximity-based parameter regionalisation. ~~A relative merit of~~ 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 ~~FDC calibration was high performance in predicting low flows. From~~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 ~~evaluation of signature reproducibility, we found that metrics describing flow dynamics such as~~ rising limb density ~~should be added as complementary constraints for improving~~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 ~~modelling against FDCs~~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.

1 Introduction

~~The runoff hydrograph, a time series of streamflow, is the basis for practical resource management tasks such as water resource allocations, designing infrastructures, flood and drought forecasting, environmental impact assessment (Westerberg et al., 2014; Parajka et al., 2013). It is essential information for investigating physical controls of catchment functional behaviours because a hydrograph aggregates processes interacting within a catchment. Prediction of the runoff hydrograph has long been an important subject in hydrological sciences and is gaining increasing attention with growing concerns about environmental changes (Blöschl et al., 2013). Runoff prediction in ungauged sites has already been a special topic in hydrological sciences, e.g., a decade-long project, Prediction in Ungauged Basins (PUB) by the International Association of Hydrological Sciences (see http://iahs.info/pub/biennia.php). However, predicting hydrograph is a still challenging task due to poor data availability and unknown knowledge of complex catchment responses (Zhang et al., 2015; Blöschl et al., 2013).~~

A standard method ~~for predicting~~to predict daily streamflow is to employ a rainfall-runoff model that conceptualises catchment functional behaviours, and ~~simulates~~simulate synthetic hydrographs from atmospheric ~~forcing inputs~~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 ~~responses, and is commonly achieved via~~behaviours. Conventionally, behavioural parameters are estimated via model calibration against observed hydrographs (referred to as the hydrograph calibration hereafter). ~~On one hand, the~~The hydrograph calibration provides convenience to ~~modellers because~~attain reproducibility of the predictand (i.e., ~~the runoff~~streamflow time series), which is ~~typically taken~~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 ~~be automatically achieved. The~~reproduce observations (Westerberg et al., 2011), use of the hydrograph ~~reproducibility for parameter identification and its validity check~~calibration has a long tradition in ~~rainfall-~~runoff modelling (~~see~~ Hrachowiz et al., 2013).

The hydrograph calibration, ~~on the other hand~~however, can be challenged by epistemic errors in input and output data, sensitivity to calibration criteria, and inability ~~of parameter calibration~~ under no or poor data availability (Westerberg et al, 2011; Zhang et al., 2008). Importantly, it is difficult to know whether ~~or not~~the parameters ~~from the~~optimised toward maximising hydrograph ~~calibration~~reproducibility are unique to represent actual catchment ~~responses~~behaviours, since multiple parameter sets ~~would~~possibly show similar ~~hydrograph reproducibility~~ predictive performance (Beven, 2006, 1993). This low uniqueness of ~~calibrated~~the optimal parameter ~~sets~~set, namely the ~~equi-finality~~equifinality problem in ~~rainfall-runoff~~conceptual hydrological 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~~ of the hydrograph calibration, one can identify the parameters with~~, distinctive flow signatures~~, –~~ (i.e., metrics or auxiliary data representing catchment behaviours~~,~~) in lieu of observed hydrographs ~~(referred to as the signature calibration hereafter). The signature calibration is a good alternative to the hydrograph calibration when suitable~~ can be used to identify model parameters~~ are not easily obtained with observed~~

~~hydrographs alone. Hingray et al. (2010), for instance, calibrated a runoff model with specific flow signatures relevant to its parameters such as snow accumulation and ablation, recession curves, and rising limb, and subsequently found enhanced performance in hourly runoff prediction in Alpine catchments. Yadav et al. (2007) used spatially extrapolated flow metrics for parameter identification, and found major streamflow indices related to catchment functional behaviours. Euser et al. (2013) proposed a framework for structuring a flexible perceptual model with multiple hydrograph signatures, and evaluated model plausibility. Other examples include use of remotely sensed geomorphological metrics (Fang et al., 2010), isotope concentrations (Son and Sivapalan, 2007), the baseflow index (Bulygina et al., 2009), the spectral density of streamflow observations (Montanari and Toth, 2007; Winsemius et al., 2009), and long term hydrograph descriptors (Shamir et al., 2005).~~

~~In particular, the~~. The flow duration curve (FDC) has received ~~great~~particular attention in the signature-based model calibrations as a ~~calibration~~single criterion ~~that can fit model parameters to catchment functional behaviours~~ (e.g., Westerberg et al., 2014, 2011; ~~2014). The FDC, the relationship between the frequency and flow magnitudes, provides a summary of temporal streamflow variations at the outlet of a catchment (Vogel and Fennessey (1994). It has been useful for numerous hydrological applications. Vogel and Fennessey (1995) exemplified potential uses of FDCs in hydrological studies including wetland inundation mapping, lake sedimentation studies, instream flow assessment, hydropower feasibility analysis, contaminant and waste management, water resources allocation, and flood frequency analysis. FDCs has been extensively used for runoff prediction (Zhang et al., 2015; Kim and Kaluarachchi, 2014; Smkhtin and Masse, 2000), land use change assessment (Zhao et al., 2012), design of power plants (Liucci et al, 2014), water quality evaluation (Morrison and Bonta, 2008), and catchment classification (Sawicz et al., 2011) among many variations. Along with those applications, FDCs or metrics from FDCs (e.g., the slope of FDCs) were often used as a single calibration criterion (e.g., Westerberg et al., 2011, 2014;~~ 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; ~~Son and Sivapalan, 2007; Yadav et al., 2007) for identifying behavioural parameters. The rationale behind the model calibration against FDCs is that the catchment functional behaviours can be captured by~~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 ~~FDCs (Vogel and Fennessey, 1995; Yokoo and Sivapalan, 2011). This hypothesis also made it possible to apply runoff models to FDC prediction (Zhang et al., 2014; Yokoo and Sivapalan, 2011) or investigation of physical controls of FDCs (e.g.,~~the FDC (e.g., Cheng et al., 2012; Ye et al., 2012; Yokoo and Sivaplan, 2011; Bottor et al., 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. ~~) in an inverse manner.~~(2011) also highlighted that the FDC calibration may provide robust predictions to moderate disinformation such as the presence of event flows under inconsistency between inputs and outputs.

~~For prediction in ungauged catchment~~If it allows rainfall-runoff models to sufficiently capture functional behaviours of catchments, the ~~parameter calibration against FDCs (referred to as the~~ FDC calibration ~~hereafter) provides practical advantages~~would have an especial value in comparison to ~~conventional~~the parameter regionalisation~~.~~ for prediction in ungauged catchment. The parameter regionalisation, ~~i.e., transferring calibrated~~which transfers or extrapolates behavioural parameters from gauged to ungauged catchments (e.g., Kim and Kaluarachchi, 2008; Oudin et al., 2008; Parajka et al., 2007; Wagener and Wheater, 2006; Dunn and Lilly, ~~2001), has~~2011), conveniently provides a ~~critical concern~~priori estimates of ~~over-reliance on~~ behavioural parameters ~~of gauged catchments. Although~~and thus became a ~~priori~~popular approach to parameter ~~estimates of~~identification in ungauged catchments (see a comprehensive review in Parajka et al., 2013). ~~are conveniently achieved by the parameter regionalisation, they are indirectly derived from modelling results~~However, it has a critical concern that regionalised parameters are highly dependent on model calibrations at gauged sites ~~with the equi-finality problem.~~that may have substantial equifinality problems. Under no flow information in ungauged catchments, it is impossible to know whether regionalised parameters are behavioural. Thus, regionalised parameters ~~could~~might be insufficiently reliable and highly uncertain (Bárdossy, 2007; Oudin et al., 2008; Zhang et al., 2008).~~ To circumvent those drawbacks of~~

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~~, the FDC-based calibration possibly becomes a good alternative. A number of studies have proposed~~ scheme. The regional models ~~for~~ predicting FDCs at ungauged sites ~~through~~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~~). In general, FDCs~~) 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 ~~by those regional models (referred~~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). ~~as the regional FDCs hereafter) well agreed with empirical FDCs; hence, the model calibration with regional FDCs was~~Based on this strength, several studies already ~~applied and~~ showed promising predictive performance using RFDC_cal for ungauged catchments (e.g., ~~Yu and Yang, 2000;~~ Westerberg et al., 2014~~). The parameter identification against regional FDCs was useful even for gauged catchments in the cases of observed hydrographs with poor quality or no overlap between climatic inputs and hydrographs. Importantly, it may be more reliable than the parameter regionalisation because flow information of the catchment of interest, albeit predicted, is directly used to find behavioural parameter sets.~~ ; Yu and Yang, 2000).

~~However, several~~Nevertheless, practical questions arise when using ~~the FDC calibration~~RFDC_cal for ~~gauged and~~ ungauged catchments. First, the FDC is simplified information with flow magnitudes only; ~~thus~~hence, the FDC calibration could worsen the ~~equi-finality and may be more deficient in~~ equifinality problem relative to the hydrograph calibration. Due to no flow ~~prediction (van Werkhoven et al., 2009). Second,~~ timing information in reginal FDCs, one ~~can~~may cast ~~concerns about~~ 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 ~~regional~~predicted FDCs possibly introduced by ~~errors in streamflow data and the regional~~the regionalisation models (Westerberg et al., 2011; Yu et al., 2002). ~~If the calibration with regional FDCs yields highly uncertain and unreliable quantile flows due to those error sources, it may be less pragmatic than~~ RFDC_cal may be undesirable when a simple parameter regionalisation~~. In truth, several~~ can provide better performance, because regionalising observed FDCs may require expensive efforts. Several comparative studies on parameter regionalisation (e.g., Parajka et al., ~~found that~~ 2013; Oudin et al., 2008) suggested that the simple proximity-based parameter transfer ~~well performed~~can be competitive in many regions~~ (e.g.,~~. Second, there ~~Parajka et al., 2013; Oudin et al., 2008); thus, the calibration against the regional FDCs may be undesirable in the case. Third, there~~ may be additional flow signatures ~~that can~~to improve predictive performance of the FDC calibration. ~~If any flow signatures are found orthogonal to FDCs, additional~~Additional constraining ~~with those signatures will enable~~can lead to ~~alleviate the equi-finality of the FDC calibration and thus enhance~~better predictive performance~~. Nevertheless~~ of the RFDC (Westerberg et al., 2014); however, it is still an open question which flow signatures ~~complement FDCs~~can supplement the FDC calibration.

~~This~~ 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 ~~explored~~aimed to evaluate predictive performance of ~~the FDC calibration in~~ 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 comparison with the conventional approaches, the hydrograph calibration and the parameter regionalisation for gauged and ungauged catchments respectively. To answer the questions given, we (1) evaluated predictive performance of the hydrograph calibration and the FDC calibration with their uncertainty for gauged catchment, (2) assessed the calibration against regional FDCs in comparison with the proximity-based parameter regionalisation for ungauged catchments, and (3) gauged ability of the FDC calibration to reproduce typical flow signatures.~~. In this work, a parsimonious 4-parameter conceptual model was used to simulate daily hydrographs ~~from lumped atmospheric forcing for 45 unregulated~~for 45 catchments in South Korea. To predict FDCs in ungauged ~~catchment~~catchments, a geostatistical regional model was adopted here. The Monte-Carlo sampling was ~~simply~~ used ~~for parameter identification~~to identify model parameters and ~~uncertainty assessment. The following section presents~~measure equifinality in the hydrograph and the FDC calibrations.

~~used in our comparative study.~~

~~2 The study area and data~~ Description of the study area and data

~~The study area is 45 gauged~~45 catchments located across South Korea with no or negligible human-made ~~alterations (e.g., river diversion and dam operations) in~~influences on flow variations were selected for this study (Figure 1). South Korea is ~~characterized~~characterised as a temperate and semi-humid climate with rainy summer seasons. The North Pacific high-pressure brings monsoon rainfall with high temperatures ~~in~~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 ℃ (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 ~~response~~behaviours is largely concentrated on summer-season hydrographs. Snow accumulation and ablation ~~are observed~~occurring at high elevations~~, but their effects~~ have minor influences on ~~temporal~~ flow variations ~~are minor~~ due to ~~the limited~~relatively small amount of winter precipitation (Bae et al., 2008). ~~Annual temperatures range between 10 and 15 ℃ (Korea Meteorological Administration, 2011).~~

The study catchments ~~shown in Figure 1~~ were selected based on availability of streamflow data. ~~Although long streamflow data are available at a few river gauging stations, high~~High-quality daily streamflow data across ~~the~~ South Korea have been produced since establishment of the Hydrological Survey ~~Center~~Centre in 2007 (Jung et al., 2010~~). We~~), 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 ~~Mistry~~Ministry of Land, Infrastructure, and Transport of the Korean government (available at http://www.wamis.go.kr/). The ~~selected~~mean annual flow of the study catchments ~~are listed in Table~~ was 739 mm yr^{-1} with ~~their climatological features~~a standard deviation of 185 mm yr^{-1} during 2007-2015. ~~As the climatic~~In addition, as atmospheric forcing inputs ~~for rainfall-runoff modelling, we used gridded~~, we collected daily precipitation~~,~~ and maximum and minimum temperatures for 2005-2015 at 3-km grid resolution produced by spatial ~~interpolation~~interpolations between 60 stations of the automated surface observing system (ASOS) maintained by the Korea Meteorological Administration. ~~Jung and Eum (2015) combined~~The ASOS data were interpolated by the Parameter-elevation Regression on Independent Slope Model (PRISM; Daly et al., 2008~~) with~~), and overestimated pixels of the PRISM grid data were smoothed by the inverse distance method~~ for~~. Jung and Eum (2015) found that this combined method improved the spatial interpolation~~, and found improved performance for producing grid~~ of precipitation and the temperatures in South Korea. The annual mean precipitation and temperature ~~datasets across South Korea. For simulating streamflow at outlets of the~~ of the study catchments~~, we collected the grid climatic data from 2005 to 2015. Annual precipitation and mean temperature in each catchment range~~ vary within ranges of 1145–1997 mm yr^{-1} and 8.0–13.8 ℃

Formatted: Superscript

~~respectively~~ during 2007-2015. Hydro-climatological features of the ~~climatic data period. Processing the climatic data for rainfall-runoff modelling will appear later~~45 catchments are summarised in ~~the methodology section~~Table 1.

3 Methodology

~~In this work, a conceptual~~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~~, and is schematized in~~. 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. ~~GR4J is classified as a soil moisture accounting model, and computation details are found in Perrin et al. (2003).~~ Since its parsimonious and efficient structure ~~enables~~allows robust calibration and reliable regionalisation of ~~model~~the parameters, GR4J has been frequently used for modelling daily hydrographs with various purposes ~~(e.g., Nepal~~under diverse climatic conditions (Zhang et al., ~~2016; Tian et al., 2013~~2015). The computation details and discussion are found in Perrin et al. (2003). The potential evapotranspiration (PE in Figure ~~1) in this study~~2) was estimated by the temperature-based model ~~of~~proposed by Oudin et al. (2005)~~ that proposed~~ for lumped rainfall-runoff modelling.

3.~~1~~2 Preliminary data processing

Before rainfall-runoff modelling~~ with GR4J~~, we preliminarily processed the ~~gridded~~grid climatic data to convert precipitation data to liquid water ~~depths~~ forcing~~ catchments~~ (i.e., rainfall and snowmelt depths) using a physics-based snowmelt model proposed by Walter et al. (2005). The preliminary ~~processing~~ snowmelt modelling was mainly for reducing systematic errors ~~or bias~~ from no snow component in GR4J, which may affect model ~~efficiencies~~performance in catchments at relatively high elevations. ~~Though combining a~~ We chose this preliminary processing to avoid adding more parameters (e.g., the temperature index~~ snowmelt model with GR4J can be an alternative approach, it increases the number of parameters (i.e., higher equi-finality) and thus model uncertainty. Since contribution of snowmelt to temporal flow variation is insignificant in South Korea as described, maintaining~~) to the ~~parsimonious~~existing structure of GR4J~~ was considered more importantly for~~. In the case of GR4J, one additional parameter ~~calibration and regionalisation in this work.~~implies 25% complexity increase in terms of the number of parameters, and thus can worsen the equifinality. The ~~error sources in the snowmelt model were assumed to yield minor impacts on runoff prediction. The snowmelt model has~~snowmelt model uses the same ~~input requirement as GR4J, thus no additional data are necessary for the processing. It simulates~~inputs of GR4J to simulate point-scale snow accumulation and ablation processes ~~using~~(i.e., no additional inputs are required). The snowmelt model is a physics-based model but uses empirical methods ~~that~~to estimate ~~physical~~its parameters~~ required~~ for the energy balance ~~in snowpack, and~~ simulation. As outputs, it produces the liquid water depths and the snow water equivalent~~ as~~

~~outputs. After the snowmelt modelling, ~~. 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~~ as lumped inputs to GR4J~~.

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

$$I_t = I_{t-1} \cdot K + R_t \tag{1}$$

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~~ that forces the catchment (i.e., rainfall or snowmelt).~~. CPI mimics temporal variations ~~in~~of typical streamflow data by converting intermittent ~~rainfall~~precipitation data to a continuous time series with an assumption of the linear reservoir. ~~The consistency between model~~The input~~ and ~~-output ~~was checked for each catchment~~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 outliers were found in the box plot of the correlation coefficients. Hence, we ~~hypothesised~~assumed that ~~acceptable~~ consistency ~~existed ~~between climatic forcing and observed hydrographs ~~for parameter calibration~~was acceptable.

3.~~2 Rainfall-runoff modelling for~~3 The hydrograph calibration in gauged catchments

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

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

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

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

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~~ a~~ 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 ~~high~~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 ~~one~~that from the SCE algorithm. Subsequently, GR4J was calibrated by 20,000 runs of the Monte-Carlo simulations for ~~remaining 44~~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 were taken for uncertainty assessment (i.e., 0.25% of the rejection threshold). For the parameter identification~~, 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).
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 ~~initializing~~initialising all runoff simulations in this study.

~~The FDC calibration was also conducted by the same Monte-Carlo sampling but for minimisation of OBJ between observed and modelled quantile flows. We used quantile flows 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) to evaluate agreement between observed and simulated FDCs. As did in the hydrograph calibration, the best parameter set was found by 20,000 random simulations and 50 behavioural parameter sets were taken.~~

3.~~3 Rainfall-runoff modelling for ungauged catchments~~4 Model calibration against the regional FDC for ungauged catchments

~~Synthetic runoff time series were generated by GR4J again for the 45 catchments by treating each catchment as ungauged. The parameters of ungauged catchments were identified by (a) local FDCs and by (b) transferring calibrated sets from nearby gauged catchments (i.e., proximity-based parameter regionalisation). Following are descriptions of both approaches.~~

~~3.2.1 Parameter identification against regional flow duration curves~~

~~The~~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~~ for regionalising observed FDCs~~. Pugliese et al. (2014) employed the top-kriging method (Skøien et al., 2006) to spatially interpolate the total negative deviation (TND), which ~~indicates an~~is defined as the area between the mean annual flow and below-~~mean~~average flows in a ~~normalized~~normalised FDC. The top-kriging weights that interpolate TND values were ~~used~~taken as weights to estimate flow quantiles of ungauged catchments from empirical FDCs of ~~neighbouring~~surrounding gauged catchments. ~~Since the top-kriging weights are obtained from topological proximity between catchments, the two methods for ungauged catchments in this study are categorised as proximity-based approaches and thus of consistency.~~ The FDC of an ungauged catchment in Pugliese et al. (2014) is estimated from normalised FDCs of ~~neighbouring~~surrounding gauged catchments as:

$$\widehat{\Phi}(w_0, p) = \widehat{\phi}(w_0, p) \cdot \overline{Q}(w_0) \tag{3a}$$

$$\widehat{\phi}(w_0, p) = \sum_{i=1}^{n} \lambda_i \cdot \phi_i(w_i, p), \quad p\epsilon(0,1) \tag{3b}$$

where $\widehat{\Phi}(w_0, p)$ is the estimated quantile flow (m^3 s^{-1}) at an exceedance probability p (unitless) for an ungauged catchment w_0, $\widehat{\phi}(w_0, p)$ is the estimated ~~normalized~~normalised quantile flow (unitless), $\overline{Q}(w_0)$ is the annual mean streamflow (m^3 s^{-1}) of the ungauged catchment, and $\phi_i(w_i, p)$ and λ_i are ~~normalized~~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, $\overline{Q}(w_0)$, can be estimated with a rescaled mean annual precipitation defined as:

$$\mathrm{MAP}^* = 3.171 \times 10^{-5} \cdot \mathrm{MAP} \cdot \mathrm{A} \tag{4}$$

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

A distinct advantage of the geostatistical method is ~~that it enables~~its ability to estimate the entire flow quantiles in a FDC with a single set 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, ~~e.g.,~~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 catchments~~, and further~~. More details ~~and discussion~~on the geostatistical method are ~~available~~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 ~~empirical~~observed FDCs for the calibration period (2011-2015). Then, the top-kriging weights were used to ~~regionalise~~interpolate empirical flow quantiles. The number of neighbours for the TND interpolation was iteratively determined as five at which additional neighbouring TNDs are unlikely to ~~give~~bring better agreement between the estimated and ~~empirical~~observed TNDs. ~~FDCs for the calibration period~~In other words, normalised flow quantiles of five catchments surrounding the target ungauged catchment were ~~regionalised~~interpolated with the top-kriging weights. Then, MAP* of ~~the TND interpolation at~~ the target ungauged

Formatted: Superscript

catchment was multiplied. We predicted flow quantiles at 103 exceedance probabilities~~. Against the~~ (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~~, parameters of GR4J were directly calibrated for each catchment. The parameters were identified in the same manner of 20,000 runs of the Monte Carlo simulations, but towards minimisation of the OBJ value between regional and modelled FDCs.~~ (i.e., RFDC_cal).

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.~~2.2~~5 Proximity-based parameter ~~regionalization~~regionalisation for ungauged catchments

Formatted: Heading 2

~~As a counterpart of the calibration against regional FDCs, we used~~We selected the proximity-based parameter transfer ~~for prediction in ungauged catchments.~~(referred to as PROX_reg hereafter) to comparatively evaluate predictive performance of RFDC_cal. The parameter regionalisation ~~can be classified into~~has three ~~typical~~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). ~~Based on its~~A comprehensive review on the parameter regionalisation in Parajka et al. (2013) reported that PROX_reg has 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 ~~simplicity (Oudin et al., 2008; Parajka et al., 2013~~4 parameters), we chose ~~the proximity-based parameter regionalisation~~PROX_reg to evaluate RFDC_cal.

~~For prediction in ungauged catchment, five donor catchments chosen for the FDC regionalisation were again used for transferring their parameter sets to each catchment of interest. To be consistent between two proximity-based approaches, we synchronised donor catchments. The five runoff simulations were averaged for representing modelled hydrographs for each catchment.~~

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.

3.~~3 Evaluation of~~ 6 Performance evaluation

We used multiple performance metrics to evaluate predictive performance ~~and uncertainty~~ ~~Two performance measures were used for evaluating model predictive~~of all modelling approaches applied in this study. Predictive performance~~. One is NSE~~

~~in Eq. 2b~~ 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 ~~observed and modelled flows and the other is the logarithmic Nash Sutcliffe Efficiency (LNSE)~~ 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 ~~simulated flows. They are traditional measures of model performance in hydrological modelling studies, and evaluate reproducibility of high~~predicted flows (Eq. 2b) and ~~medium flows (~~(2) the NSE~~) and low~~ of log-transformed flows (LNSE) respectively. LNSE is ~~defined~~calculated as:

$$LNSE = 1 - \frac{\cancel{\sum_{i=1}^{N}[\ln Q_{obs,i} - \ln Q_{sim,i}]^2}\ \sum_{i=1}^{N}(\ln(Q_{obs,i}) - \ln(Q_{sim,i}))^2}{\cancel{\sum_{i=1}^{N}[\ln Q_{obs,i} - \ln(\overline{Q_{obs}})]^2}\ \sum_{i=1}^{N}(\ln(Q_{obs,i}) - \overline{\ln(Q_{obs})})^2} \qquad (5)$$

~~For uncertainty assessment, the lower and upper bounds were drawn at the values of 2.5 and 97.5 percentiles of predicted hydrographs with the collection of 50 parameter sets. Uncertainty~~Though NSE and LNSE are frequently used for performance evaluation, they may be sensitive to errors in ~~predicted flows was quantified by the area between the lower and upper bounds of simulated hydrographs. We took a ratio of uncertainty of the FDC calibration to that of the hydrograph calibration for each catchment and defined it as the uncertainty ratio. Note that this assessment was not to estimate absolute uncertainty but to measure relative uncertainty gained by replacing a hydrograph with a FDC for model calibration.~~

~~We~~flow observations (Westerberg et al., 2011). Hence, we additionally selected three typical flow metrics ~~to evaluate flow signature predictability~~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}). ~~The three typical signatures describe~~ R_{QP}, I_{BF}, and D_{RL} are proxies of aridity ~~in a catchment, long-term baseflow~~and water holding capacity, contribution of the baseflow to flow variations, and ~~the~~ flashness of catchment ~~response~~behaviours, respectively. They are defined as the ratio of runoff to precipitation, the ratio of ~~long-term~~ baseflow to total runoff, and the inverse of average time to peak (d^{-1}) as:

$$R_{QP} = \frac{\overline{Q}}{\overline{P}} \qquad (6a)$$

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

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

where $\overline{Q}$ and $\overline{P}$ are ~~the~~ average flow and precipitation ~~during~~for a given period~~,~~ (mm d^{-1}), Q_t and $Q_{B,t}$ (m d^{-1}) is the~~ total~~ 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). $Q_{B,t}$ ~~is~~can be calculated by subtracting direct flow $Q_{D,t}$ from Q_t as:

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

$$Q_{B,t} = Q_t - Q_{D,t} \tag{7b}$$

where c is the filter parameter ~~c is a value of~~, which was set to 0.925 ~~from a comprehensive case study by~~(Brooks et al., 2011; Eckhardt ~~(~~, 2007). ~~Reproducibility of R_{QP}, I_{BF},~~

Flow signature reproducibility of RFDC_cal and ~~D_{RL} are~~PROX_reg were evaluated by the relative absolute bias between modelled and observed signatures as:

$$D_{FS} = \frac{|FS_{sim} - FS_{obs}|}{FS_{obs}} \tag{8}$$

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.

4 Results

4.1 ~~Streamflow prediction in gauged catchments~~ Hydrograph calibration and FDC regionalisation in gauged catchments

~~The box plots in Figure 3 comparatively show distributions of NSE and LNSE values between observed and modelled flows. It was clearly indicated that the hydrograph outperformed the FDC calibration in prediction of high flows. The NSEs of the hydrograph calibration were generally greater than those of the FDC calibration for the both calibration and validation periods. The FDC calibration was of much wider NSE ranges than the hydrograph calibration and thus greater uncertainty in high flow prediction. The prediction results tend to have greater medians of NSEs for the calibration periods than the validation period. Because the term NSE is directly used for calibration, the parameter identification could be slightly inclined towards reproduction of high flows for the calibration period. The significantly increasing NSE ranges from the calibration to validation periods in Figure 3b may imply that the FDC calibration has weaker temporal parameter transferability from one period to another. In low-flow prediction, the FDC calibration showed slightly weaker performance than the hydrograph calibration. Although the LNSE medians of the FDC calibration were comparable to those of the hydrograph calibration, LNSEs of the FDC calibration also showed wider ranges than the hydrograph calibration. The FDC calibration was still likely to yield significant uncertainty in low-flow predictions when parameters were temporally transferred. Unlike the NSE comparison, the median LNSE values did not decrease from the calibration to the validation periods for the both hydrograph and the FDC calibrations. This would imply that the behavioural parameter sets have more temporal consistency in low flows than high flows.~~

~~Figure 4 illustrates 1:1 scatter plots between the performance measures and correlation between CPI and observed hydrographs, indicating that consistency between model input and output meaningfully affects predictive performance of rainfall-runoff models. The performance measures were generally in positive relationships with correlation between CPI and observed hydrographs. Adequate input-output consistency seems to be a prerequisite of parameter identification to attain good high-flow predictability especially for the hydrograph calibration. For having 0.6 or more NSE, the correlation~~

~~coefficient between CPI and observed flows should be greater than 0.6 approximately. On the other hand, predictability of low flows was achieved with relatively low input-out consistency. LNSEs less than 0.4 were rarely observed than NSEs for the both hydrograph and FDC calibrations. Interestingly, the FDC calibration appears to have better predictability in low flows despite the use of NSE for parameter calibration, which is a sensitive measure to high-flow reproducibility. It implies that the FDC calibration has some deficiency to capture catchment response to storm events even with adequate model input-output consistency whereas it performs well for long-term low flow or baseflow predictions.~~

~~Shortly, the FDC calibration could lead to relatively low predictive power with increased uncertainty when adopted as an alternative of the hydrograph calibration. Low predictability in high flows can be a particular concern of the FDC calibration. The simplification of flow information appears to exacerbate the equi-finality in parameter identification. This weakness of the FDC calibration was confirmed by the uncertainty bounds of modelled hydrographs in Figure 5. The collection of 50 parameter sets from the FDC calibration showed less robust simulations than the hydrograph calibration for the three catchments even though their FDCs were fairly well reproduced by the FDC calibration. For the 45 catchments, the mean NSE between observed and modelled FDCs was 0.95 when using the FDC calibration. In other words, parameters reproducing observed FDCs generally were less unique to represent catchment functional behaviours than ones reproducing observed hydrographs. The equi-finality in the FDC calibration is likely to get worse with decreasing performance of the hydrograph calibration (Figure 6). On average, uncertainty of predicted hydrographs was doubled for the 45 catchments when the FDC calibration substitutes for the hydrograph calibration. The prediction results from the 45 gauged catchments, hence, suggest that parameter identification with compact information of FDCs could yield weaker performance and less parameter identifiability than the hydrograph calibration.~~

~~4.2 Geostatistical FDC regionalisation~~

~~Figure 7a illustrates the 1:1 scatter plot between observed and estimated TNDs of the 45 catchments. The correlation coefficient between empirical and estimated TNDs was~~ 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 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 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 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.

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 where gauging density is low.~~0.56 (equivalent to 0.30 NSE). The relatively poor prediction of TNDs was likely from the use of annual precipitation for normalising quantile flows. In the original study of the geostatistical method (Pugliese et al., 2014), the TND prediction became poorer (NSE was decreased from 0.81 to 0.60) when using the rescaled annual precipitation instead of observed mean annual flow. Uncertainty introduced by estimation of mean annual flows might influence predictive power of the geostatistical TND interpolation. Another likely reason is that TND is a complex signature of streamflow regime; yet, it could be descriptive in terms of functional similarity between catchments (Pugliese et al., 2016). It may be difficult to completely capture spatial variation of TNDs with topological proximity only. However, Pugliese et al. (2016) also argued that poor prediction of TND did not automatically result in poor quantile flow predictions. Their comparative study achieved successful FDC predictions for 182 catchments in the United States (0.95 of median NSE) using the top kriging weights of TNDs in spite of low TND predictability. Though it is an outside scope of this study, a further study needs to be directed towards effects of TND prediction on the FDC regionalisation. Because it is still unclear whether or not descriptors from FDCs well predict flow quantiles, top-kriging weights of various flow signatures need to be tested for improving the geostatistical FDC prediction as well.~~

~~The high performance in FDC prediction with poor TND prediction was replicated in this study. Overall NSE and LNSE values between observed and predicted quantile flows of the 45 catchments suggest good applicability of the geostatistical method to the study catchments (Fig 7b). The averages of individual NSEs and LNSEs for each catchment were 0.83 and 0.91 with standard deviations of 0.25 and 0.11 respectively. The higher LNSEs imply that performance of the geostatistical method is better for low flows. This might be because the top-kriging weights interpolating TNDs were obtained from below-average flows only. No information of above-average flows reflected in TNDs might incline the FDC regionalisation towards low flow predictions. Low predictive power of the regional FDC model was found at locations with low gauging density.~~ Catchments 4, 10, 35, and 36, which recorded 0.6 or less NSEs~~, were~~ are limitedly hatched with ~~no hatching catchments and/~~or ~~limited~~ adjacent to the other catchments; nonetheless, LNSEs of those catchments were still greater than 0.7. This result ~~was~~is consistent with a finding of Pugliese et al. (2016) that performance of the geostatistical method was

~~highly~~ sensitive to river gauging density. Transferring ~~quantile flows of~~ flow quantiles from remote catchments ~~can yield significant errors because~~may not sufficiently capture functional similarity ~~would not be captured~~ between donor and receiver catchments. ~~Overall, in~~In spite of ~~abovementioned~~the minor shortcomings, the geostatistical FDC regionalisation was deemed acceptable based on the high NSE and ~~topological~~LNSE of flow quantiles. Topological proximity ~~would to be~~was generally a good predictor of ~~FDCs across~~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 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 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 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 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 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 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.

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 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 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 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+DRL_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 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 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 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 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.

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, 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

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 Q2 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 (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 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. 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~~4.3 Streamflow prediction for ungauged catchments~~

~~The box plots in Figure 8 present predictive performance of the calibration against regional FDCs (referred to as RFDC_cal hereafter) in comparison with the proximity-based parameter regionalisation (referred to as PROX_reg hereafter). The performance measures between observed and modelled hydrographs were computed for the entire period of streamflow data (2007-2011). Distributions of NSEs clearly showed that PROX_reg outperforms the FDC calibration in prediction of high flows (Figure 8a), indicating that a priori parameter sets from neighbouring catchments should perform even better than local calibrations against observed FDCs. The average difference between NSEs of PROX_reg and RFDC_cal was 0.18 with a standard deviation of 0.25. RFDC_cal outperformed PROX_reg only for 8 out of the 45 catchments. LNSEs with PROX_reg were still of a slightly higher median than RFDC_cal. Although RFDC_cal appears to have comparable predictability in low flows, 31 out of 45 catchments were having greater LNSEs with PROX_reg. The results for ungauged catchments bring the same intuition as the case for gauged catchment that a priori parameter sets obtained from nearby gauged catchments seem to be more desirable than local parameter identification against regional FDCs.~~

~~The weaker performance of RFDC_cal in this work is consistent with the comparative study of~~). 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 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 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 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

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:

(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.

(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.

(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.

~~Zhang et al. (2015), which evaluated performance of RFDC_cal using GR4J in 228 Australian catchments. Zhang et al. (2015) argued that RFDC_cal is not good enough for predicting daily hydrographs in the Australian catchments due to its much worse performance than the hydrograph calibration in gauged catchments. The information loss from simplifying hydrographs can be attributed to weaker performance and higher uncertainty of rainfall-runoff modelling against in FDCs. In recognition of good agreement between empirical and regional FDCs for the study catchments, prediction errors in regional FDCs would influence minor impacts on performance of RFDC_cal.~~

~~4.4 Evaluation of flow signature reproducibility~~

~~Figure 9 summarises performance of the four methods applied in this study to regenerate three flow signatures of R_{QP}, I_{BF}, and D_{RL}. The box plots of absolute biases between observed and modelled signatures indicate that parameter identification against FDCs showed competitive reproducibility in the long term signatures R_{QP} and I_{BF}, while its ability was relatively weak to regenerate the event based signature D_{RL}. R_{QP} biases seem to be sensitively affected by additional uncertainty sources in the FDC regionalisation and in spatial and temporal parameter transfer, but their medians and box heights were similar between FDC-based and hydrograph-based approaches. Given their relatively competitive performance in low flows, FDC-based approaches would show strong performance to reproduce I_{BF}.~~

~~In contrast, the FDC-based approaches were poorer to reproduce the event-based flow signature, D_{RL}. It is not surprising because a FDC aggregates information of flow magnitude only. No information of flow timing in FDCs is likely a main factor that resulted in poor predictions of peak flow timing for both gauged and ungauged catchments. The FDC based approaches could be insufficient for hydrological applications that require specific flow timings (e.g., flood forecasting). The conventional parameter regionalisation would be a more pragmatic option for the Korean catchments. From Figure 9c, we also had an indication that predictability in peak flow timing of the hydrograph calibration was well preserved even when parameter sets were transferred to neighbouring catchments.~~

~~5 Discussion~~

~~5.1 Evaluation of rainfall-runoff modelling against regional FDCs~~

~~Regionalised flow signatures have frequently used for constraining rainfall runoff models (e.g., Bárdossy, 2007; Boughton and Chiew, 2007; Bulygina et al., 2009). Advantages of the approaches are that they are complementary to a priori estimation of model parameters and are similar to usual methods to directly find the model parameters from dynamic catchment response data (Blöschl et al., 2013). An important lesson learned from previous studies was that the models would dominantly work for reproducing the flow signature of interest (Blöschl et al., 2013), albeit it appears self-evident. Thus, if one forces the model to reproduce low flow signatures, use of the model would be appropriate for a drought forecasting rather than a flood analysis. Likewise, multiple signatures are obviously necessary for constraining runoff models to consider various aspects of flow variation.~~

~~In this context, use of a FDC as a single calibration criterion appears to be a great choice for searching model parameters suitable for dynamic catchment behaviours. A FDC is a compact representation of runoff variability in frequency domain at all time scales from inter-annual to event scale, and thus it embeds various aspects of multiple flow signatures (Blöschl et al., 2013). A pilot study of Yokoo and Sivapalan (2011) discovered that the upper part of a FDC with high flows is controlled by interaction between extreme rainfall and fast runoff, while the middle and lower parts are governed by interactions between water availability, energy and water storage and by baseflow recession behaviour during dry periods respectively. The major~~

~~hydrological processes within a catchment are reflected in a FDC, and therefore a runoff model constrained by a FDC can be expected to provide reliable flow predictions. Westerberg et al. (2014, 2011) and Yu and Yang (2000) are successful examples that applied FDCs to rainfall-runoff modelling as a single calibration criterion.~~

~~The comparative evaluation in this study, however, provides the same lesson that rainfall-runoff modelling against FDCs is good to reproduce the FDC itself, but it was insufficient to be comparable to the hydrograph calibration in gauged catchments. For 41 out of the 45 catchments, NSEs between observed and modelled FDCs were greater than 0.9; nonetheless, hydrograph reproducibility of the FDC calibration was generally weaker. The hydrograph is an output of numerous hydrological processes interacting within a catchment, and is regarded as the most complete flow signature (Blöschl et al., 2013). Since any simplification of the hydrograph including FDCs would lose some amount of flow information, it is no surprise that the FDCs calibration worsens the equi-finality problem in conceptual rainfall-runoff modelling. If one has a runoff time series with acceptable data quality and length, there should be no reason to adopt the FDC calibration in replacement of the hydrograph calibration. The weaker D_{RL} reproducibility confirms that the absence of flow timing in FDCs would lead to poorer runoff predictions of the FDC calibration. Instead, the FDC calibration may be good for prediction of compact flow signatures which are less informative than FDCs (e.g., mean annual runoff and seasonal flow regime).~~

~~For ungauged or poorly gauged catchments, on the other hand, rainfall-runoff modelling against regionalised FDCs (RFDC_cal) can bring advantages. As aforementioned, a priori parameter sets derived from the outside of a catchment of interest may be more uncertain and thus less reliable than ones achieved from independently predicted flow signatures. Nevertheless, RFDC_cal was less powerful than use of parameter sets transferred from neighbouring catchments despite well-regionalised FDCs. The deficiency in RFDC_cal was likely to come not only from the absence of flow information in FDCs, but from powerful performance of PROX_reg. Modelling conditions of this study were very suitable for proximity-based parameter transfer based on a comparative study of Parajka et al. (2013), which extensively reviewed literature on parameter identification in ungauged catchments. Parajka et al. (2013) reported that parameter regionalisation generally showed higher NSE performance under humid conditions than in arid and tropical regions. They argued that PROX_reg can be competitive with or better than similarity-based and regression-based regionalisation (e.g., Oudin et al., 2008; Parajka et al., 2005). Parajka et al. (2013) also provided a relationship between model complexity and performance, indicating that the complexity of GR4J (4 parameters) used in this study was desirable for parameter regionalisation. Given the knowledge in Parajka et al. (2013), aridity and temperature conditions of the 45 study catchments were suited to provide good predictive performance with PROX_reg. The strong performance of PROX_reg in this study suggests that functional similarity between Korean catchments may be changing gradually in space and thus found with spatial proximity. This could be confirmed by good performance of the geostatistical FDCs regionalisation in this study. Under these conditions, it may be difficult to produce better predictions using RFDC_cal with much higher equi-finality.~~

~~5.2 Why the FDC calibration performs good for low flow prediction~~

~~Although we showed its weaknesses, this paper is not intended to leave negative messages on hydrological modelling against FDCs. It should be emphasised that the FDC calibration may provide advantages for applications aiming at assessing long-term flow regime under projected environmental conditions (e.g., climate change impact assessment). In particular, its powerful predictability in low flows needs to be underlined. The objective function used in the parameter calibration includes the NSE, which can lead to overemphasis on high or peak flows due to squared residuals (Hrachowitz et al., 2013), albeit it is combined with the WBE. The calibration against FDCs, however, well reproduced low flows and I_{BF} with no logarithmic transformation of observed flows, and hence could be a good choice for a low flow analysis or a long-term water resources management in both gauged and ungauged catchments.~~

~~In regard of flow variation condensed into quantile flows of a FDC, predictability of the FDC calibration may be explained. In Korean catchments under a typical monsoonal climate, low flows governed by baseflow during dry seasons have less temporal variation than high flows generated by intermittent storm events. Thus, information loss of low flows is much smaller than high flows when a hydrograph is summarised in frequency domain. Figure 10a and b illustrate that high flows modelled by the collection of 50 parameter sets have flow timing errors and low robustness in medium to high flows in spite of fairly good agreement between observed and modelled FDCs across all flow magnitudes. The ranges of baseflow and direct runoff (i.e., main controls of low and flows) for the calibration period are shown together in Figure 10c. It indicates that direct runoff is more significantly condensed into a FDC. Because of the flow regime with small low-flow variability of the Korean catchments, the FDC calibration could automatically incline the model parameter towards reproduction of low flows. Should considerable variability exist in baseflow (e.g., snow fed catchments), performance of the FDC calibration may differ.~~

~~5.3 Flow signatures for improving calibration against FDCs~~

~~As evaluated, rainfall-runoff modelling against FDCs has strength in baselow or low flow prediction in South Korea while high flows were not well captured due to the absence of flow timing. It was confirmed by the flow signature reproducibility in Figure 9 and the low robustness of direct runoff simulations in Figure 10b. Hence, additional constraining may fill the gap in FDC calibration as discussed in Westerberg et al. (2014). Westerberg et al (2014) emphasised the necessity of further constraining to reduce predictive uncertainty despite their sophisticated modelling against FDCs. The comparative evaluation of this study simply suggests that orthogonal (or complementary) flow signatures to a FDC should explain temporal flow variation (e.g., D_{RL}, falling limb density, and recession rate).~~

~~The box plots in Figure 11 show how the FDC calibration can be improved by additional constraints of the three flow signatures (R_{QP}, I_{BF}, and D_{RL}). For runoff predictions, we simply chose one parameter set with the best reproducibility of each signature from the collection of 50 parameter sets of the FDC calibration. As expected from the competitive reproducibility of the FDC calibration in R_{QP} and I_{BF}, no meaningful improvement was found with both signatures. On the~~

~~contrary, the parameter sets constrained by D_{RL} resulted in fairly improved performance, suggesting the need of metrics associated with temporal flow variation in the FDC calibration. A further study needs to be directed for regionalising flow metrics representing flow dynamics together with a framework to combine multiple signatures as it could fill the gap in model calibration against FDCs.~~

~~5.4 Limitations and future research directions~~

~~This study provides a meaningful lesson that modelling against regional FDCs may not be an attractive option where proximity-based parameter regionalisation performs greatly. In our knowledge, the topic of runoff prediction in ungauged catchments has been rarely dealt in South Korea due to limited availability of quality streamflow data, thus this study may become a good reference for scientific community. Nonetheless, there are several limitations in our comparative evaluation. First, we never considered uncertainty in discharge data for constructing FDCs and model calibration. McMillan et al. (2012) reported typical ranges of relative errors in discharge data as around 10-20% for medium to high flow and 50-100% for low flows. The measurement errors and epistemic uncertainty in input and output data may cause a disinformation effect on model calibration. Especially for the hydrograph calibration, if the model is significantly forced to compensate disinformation in high flows, calibrated parameters can be biased (Westerberg et al., 2011). We assumed that quality of the discharge data was adequate based on rigorous controls of the data distribution centre, but consideration of such errors will clarify their relative effects on the hydrograph- and FDC-based runoff modelling. Second, we used a conceptual runoff model with a fixed structure for all catchments, but it could be a structural error source for some catchments. Blöschl et al. (2013) recommended that structuring a conceptual model needs to be considered in a realistic manner for reliable predictions. If this step was included in this study, predictive power might be better in catchments with relatively low NSE performance. Finally, though the proximity based parameter regionalisation was powerful, other regionalisation methods such as regional calibration and spatial similarity parameter transfer would provide comprehensive information for selection between the parameter regionalisation and the signature calibration for ungauged catchments.~~

~~Obviously, one research direction stemming from this study is how to regionalise metrics related to flow timing and dynamics. The signature calibration inherently removes the concern in conventional parameter regionalisation approaches, but should be based on well-regionalised signatures. Candidate flow signatures that can enhance the FDC calibration would be the overall flow variability, the flow autocorrelation, the rising and falling limb densities, and the slope of fast recession curve among other metrics. Unfortunately, the task of regionalising these signatures will be challenging. Westerberg et al. (2016) found that metrics gauging flow dynamics could be more uncertain than one measuring flow distribution (e.g., quantile flows). A new framework beyond conventional regionalisation methods may be needed to reduce uncertainty in regional flow signatures.~~

~~6 Summary and conclusions~~

~~In this study, we investigated performance of the FDC calibration by comparing it with hydrograph-based methods for gauged and ungauged catchments. We began with parameter calibration of GR4J model against observed hydrographs and empirical FDCs at 45 catchments in South Korea using random simulations. Predictive performance and uncertainty of each catchment were evaluated using parameter sets obtained. For evaluation for ungauged catchments, hydrographs of the 45 catchments were again predicted by treating each catchment as ungauged. In doing so, we estimated regional FDCs of the catchments using a promising geostatistical method, and calibrated model parameters against the regional FDCs. Predictive performance of the model based on regional FDCs was evaluated in comparison to hydrographs simulated with parameters transferred from neighbouring catchments. The key findings from our comparative evaluation are summarized as follows:~~

~~(1) For gauged catchments, predictive performance and uncertainty of the FDC calibration can be significantly degraded by lose of flow timing information from a hydrograph to a FDC. Parameter identifiability would be reduced since the equi-finality increased. Uncertainty in hydrographs predicted by the FDC calibration was doubled on average.~~

~~(2) The geostatistical FDC regionalization showed good performance in prediction of quantile flows despite its low TND reproducibility. The top-kriging weights interpolating TNDs had high potential for predicting quantile flows. Topological proximity is likely to well explain functional similarity between catchments in South Korea. However, it is notable that considering topological proximity only can bring bias where gauging density is low.~~

~~(3) The typical proximity-based parameter transfer was of strong performance to regenerate hydrographs, and outperformed model calibration with regional FDCs. Although regional FDCs would have potential for capturing functional behaviour of ungauged catchments, the absence of flow timing would lead to less robust and less predictive performance than transferring parameters.~~

~~(4) Relative merits of model calibration with regional FDCs were strong performance in baseflow prediction. Without logarithmic transformation of observed flows, parameters with regional FDCs seem to be forced to reproduce low flows because of relatively low temporal variation in baseflow of Korean catchments.~~

~~(5) Complementary flow signatures for the FDC calibration could be metrics describing flow timing and dynamics. Additional constraining with~~ D_{RL} ~~showed fairly improved performance with the FDC calibration. A further study for regionalising those metrics will improve the model calibration against regional FDCs.~~

~~In short, we suggest that classical parameter regionalisation is pragmatic for predicting hydrographs in ungauged catchments in South Korea where spatial proximity well captures functional similarity between catchments. Nonetheless, we believe that further studies on regionalisation of relevant flow signatures will inherently improve runoff modelling in ungauged catchments using the FDC-based calibration. The FDC calibration still has a major advantage that it can directly identify parameters against plausible flow information of the catchment of interest unlike the parameter regionalisation.~~

Acknowledgements

This study was supported by APEC Climate Center. ~~The authors~~We send special thanks to Ms. Yoe-min Jeong and Dr. Hyungil Eum for 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.

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. Process., 6, 279-298, 1992.

Beven, K. J.: A manifesto for the equifanality 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, doi:10.1029/2007wr006726, 2009.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije,. H.: Runoff Prediction in Ungauged Basins, Simthesis 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. Res., 43, ~~Boughton, W., and Chiew, F.: Estimating runoff in ungauged catchments from rainfall, PET and the AWBM model, Environ. Model. Softw., 22, 476-487, 2007.~~

~~Bulygina, N., McIntyre, N~~W02417, doi:10.1029/2006WR005043, 2007.

Botter, G., Zanardo, S., Porporato, A., Rodriguez-Iturbe, I., and ~~Wheater, H.: Conditioning rainfall-runoff model parameters for ungauged catchmenets~~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.

Cheng, L., Yaeger, M., Viglione, A., Coopersmith, E., Ye, S., and Sivapalan, M.: ~~land management impacts~~

~~analysis~~Exploring the physical controls of regional patterns of flow duration curves – Part 1: Insights from statistical analyses, Hydrol. Earth Syst. Sci., ~~13, 893-904~~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. ~~13-893-2009, 2009~~Sci., 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 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.

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.

~~Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci., 17, 1893-1912, doi:10.5194/hess-17-1893-2013, 2013.~~

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. 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.

Kavetski, D., Fnicia, 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 Reour. 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.

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.

~~Liucci, L., Valigi, D. & Casadei, S.: A new Application of flow duration curve (FDC) in designing run-of-river power plants, Water Resour Manage, 28, 881-895, doi:10.1007/s11269-014-0523-4, 2014.~~

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, 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.

~~Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, Water Resour. Res., W05434, doi:10.1029/2006WR005184, 2007.~~

~~Morrison, M., and Bonta, J.: Development of Duration-Curve Based Methods for Quantifying Variability and Change in Watershed Hydrology and Water Quality. Report No: EPA/600/R-08/065, Land Remediation and Pollution Control Division, National Risk Management Research Laboratory, Environmental Protection Agency, Cincinanati, Ohio, USA, 2008.~~

~~Nepal, S., Chen, J., Penton, D. J., Neumann, L. E., Zheng, H., and Wahid, S: Spatial GR4J conceptualization of the Tamor glaciated alpine catchment in Eastern Nepal: evaluation of GR4JSG against streamflow and MODIS snow extent, Hydrol. Process., 31, 51–68. doi: 10.1002/hyp.10962, 2017.~~

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged 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 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.

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.

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.

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.

~~Sawicz, K., Wagener, T., Sivapalan, M., Troc, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15, 2895-2911, doi:/10.5194/hess-15-2895-2011, 2011.~~

~~Shamir, E., Imam, B., Gupta, H. V., and Sorooshian, S.: Application of temporal streamflow descriptors in hydrologic model parameter estimation, Water Resour. Res., 41, W06021, doi:10.1029/2004WR003409, 2005.~~

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.

~~Skøien, J. O., Blöschl, G., Laaha, G., Pebesma, E., Parajka, J., and Viglione, A.: rtop: An R package for interpolation of data with a variable spatial support, with an example from river networks, Comput. Geosci. 67, 180-190, 2014.~~

~~Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, Hydro. Earth Syst. Sci., 10, 277-287, doi:10.5194/hess-10-277-2006, 2006.~~

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

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. ~~Son, K.~~Sci., 15, 3701–3713, doi:10.5194/hess-15-3701-2011, 2011.

~~and Sivapalan, M.: Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, Water Resour. Res., 43, W01415, doi:10.1029/2006WR005032, 2007.~~

~~Tian, Y., Xu, Y.-P., and Zhang, X. J.: Assessing of climate change impacts on river high flow through comparative use of GR4J, HBV, and Xinanjiang models. Water Reour. Manage., 27, 2871-2888, doi:01.1007/s11269-013-0321-4, 2013.~~

~~van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Sensitivity guided reduction of parametric dimensionality for multi-objective calibration of watershed models, Adv. Water Resour., 32, 1154-1169, doi:10.1016/j.advwatres.2009.03.002, 2009.~~

~~Vogel, R. M., and Fennessey, N. M.: Flow duration curves I: A new interpretation and confidence intervals, J. Water Resour. Plann. Manage., 120, 485-504. 1994.~~

~~Vogel, R. M., and Fennessey, N. M.: Flow duration curves II: A review of applications in water resources planning, Water Resour. Bull., 31, 1029-1039, 1995.~~

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

~~Winsemius, H. C., Schaefli, B., Montanari, A., and Savenije, H. H. G.: On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, Water Resour. Res., 45, W12422, doi:10.1029/2009wr007706, 2009.~~

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, ~~Y., Vaze, J., Chiew, F. H., Teng, J., and Li, M: Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall-runoff modelling, J. Hydrol., 517, 936-948, 2014.~~

~~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.

~~Zhao, F., Xu, Z., and Zhang, L.: Changes in streamflow regime following vegetation changes from paired catchments, Hydrol.~~

~~Process., 26, 1561-1573, 2012.~~

Table 1: List of the gauged catchments and hydrological features (2007-2015)

ID	Name	Ar[1]	Elv[2]	P_a[3]	T_a[4]	Ard[5]	P_s[6]	ID	Name	Ar	Elv	P_a	T_a	Ard	P_s
1	Goesan Dam	677	363	1223	11.0	.69	29.5	24	Chunyang	145	201	1611	13.2	58	12.8
2	Namgang Dam	2293	431	1558	13.8	.61	5.7	25	Osu	360	255	1434	11.7	61	49.6
3	Miryang Dam	104	512	1824	13.3	.50	20.1	26	Daecheon	816	198	1336	13.2	70	23.4
4	Boryeong Dam	162	244	1997	11.4	.44	140.8	27	Jeonju	276	176	1312	12.9	71	29.5
5	Buan Dam	57	177	1253	13.7	.76	39.3	28	Hari	528	197	1332	13.4	71	20.8
6	Seomjingang Dam	763	357	1487	11.4	.58	54.7	29	Bongdong	345	245	1354	13.2	69	19..4
7	Soyanggang Dam	2783	634	1231	9.5	.64	50.6	30	Hannaedari	284	126	1218	12.6	75	31.2
8	Andong Dam	1629	543	1330	10.0	.61	51.5	31	Suchon	224	94	1254	12.4	72	42.4
9	Yongdam Dam	930	510	1508	12.6	.60	22.6	32	Wolpo	1158	315	1303	11.3	66	30.1
10	Imha Dam	1976	388	1319	10.1	.63	50.6	33	Jeomchon	615	371	1230	11.5	71	29.9
11	Hoengseong Dam	208	436	1247	11.1	.68	28.5	34	Sancheong	1131	554	1608	13.8	59	14.1
12	Habcheon Dam	929	495	1470	12.9	.62	17.1	35	Seonsan	988	298	1202	12.0	73	27.7
13	Chungju Dam	6705	608	1289	9.9	.62	51.5	36	Nonsan	477	151	1309	13.0	71	19.4
14	Juam Dam	1029	269	1765	12.7	.52	19.5	37	Ugon	134	39	1272	13.2	73	19.3
15	Jangheung Dam	192	198	1733	13.4	.54	17.6	38	Seokdong	156	71	1268	12.8	72	29.5
16	Jungranggyo	209	131	1388	12.7	.66	22.9	39	Cheongju	165	149	1235	12.3	73	24.8
17	Munmak	1138	303	1286	11.9	.69	25.1	40	Heodeok	609	193	1266	12.4	71	23.0
18	Yeongchun	4775	996	1145	7.9	.62	83.3	41	Yuseong	246	193	1253	12.6	73	23.0
19	Yeongwol 1	1614	625	1263	9.7	.62	51.3	42	Boksu	162	216	1267	12.2	71	23.6
20	Pyeongchang	696	720	1235	9.3	.62	62.3	43	Sangyeogyo	495	255	1267	12.2	71	23.6
21	Naerincheon	1013	752	1231	9.5	.64	50.6	44	Gidaegyo	361	250	1218	11.3	70	30.6
22	Wontong	300	707	1283	8.6	.59	71.0	45	Indong	68	203	1229	12.0	72	24.8
23	Hampyeong	105	87	1327	13.7	.72	23.7								

[1]Draiage Area (km^2), [2]Mean elevation (m), [3]Mean annual precipitation (mm), [4]Mean annual temperature (℃), [5]Aridity (unitless) defined by the sum of potential evapotranspiration divided by the sum of precipitation, and [6]Mean annual snowfall (mm) defined by mean annual precipitation when mean temperatures were below 0℃. All climatological features were calculated by spatial averages of the grid data.

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

[1]Ratio of potential ET to total precipitation, [2]Percentage 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

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

Questions	Corresponding pair	[1]PM	[2]$\overline{\Delta PM}$	[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

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

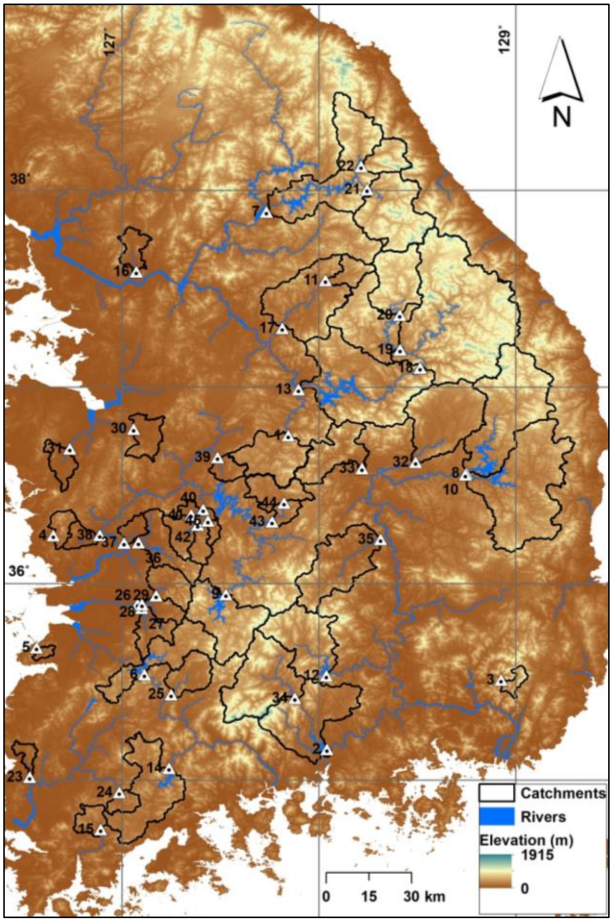


Figure 1: Locations of the ~~gauged~~study catchments ~~for GR4J model and FDC regionalization. Catchment~~in South Korea. The numbers are labelled at the ~~centroid~~outlet of each catchment.

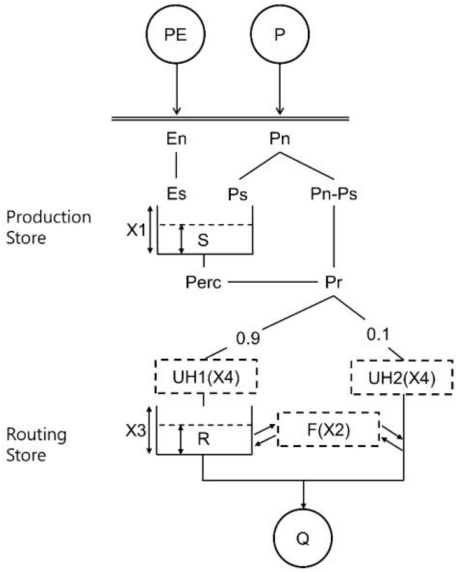
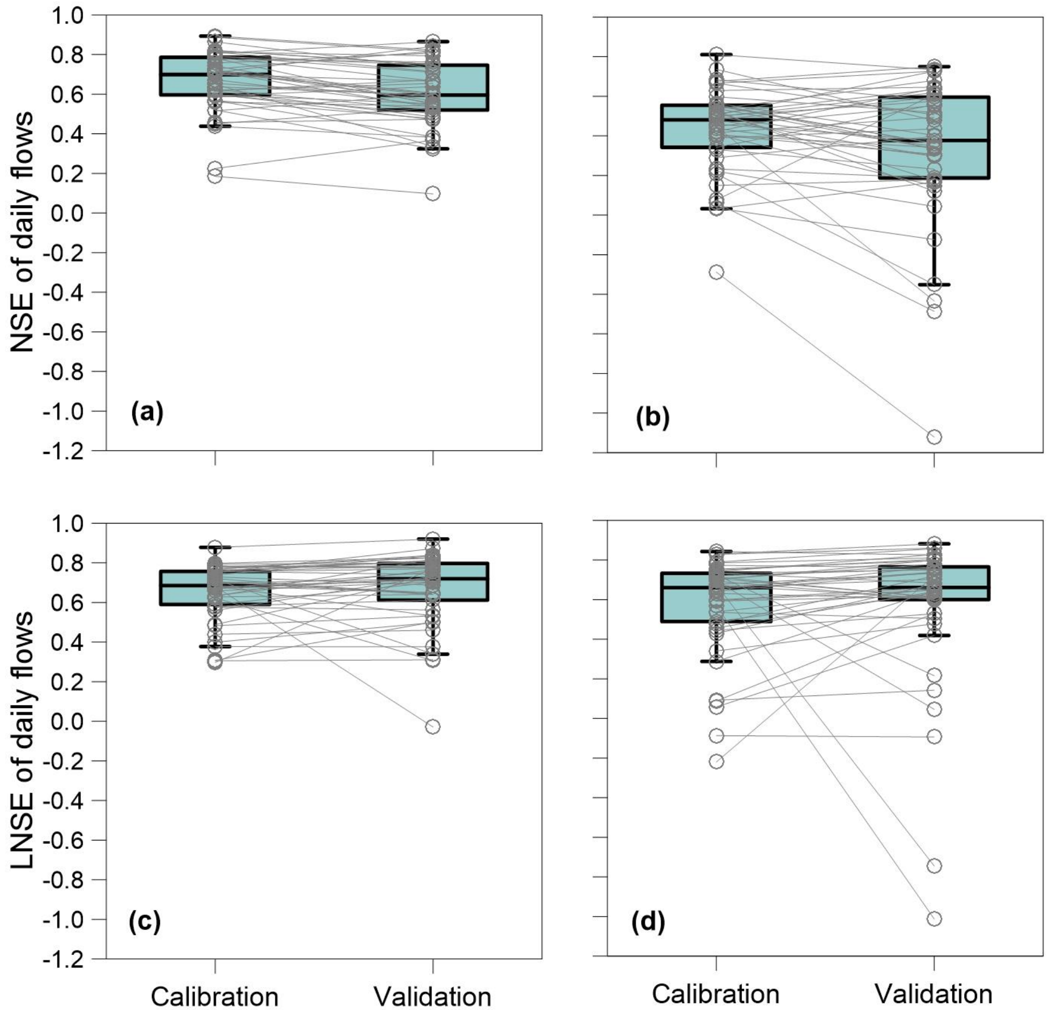


Figure 2: The schematised structure of GR4J (X1-X4: model parameters, PE: potential evapotranspiration, P: precipitation, Q: runoff, other letters indicate variables ~~conceptualizing~~conceptualising internal catchment processes).


NSE of daily flows
LNSE of daily flows
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
(a)
(b)
(c)
(d)
Calibration
Validation
Calibration
Validation

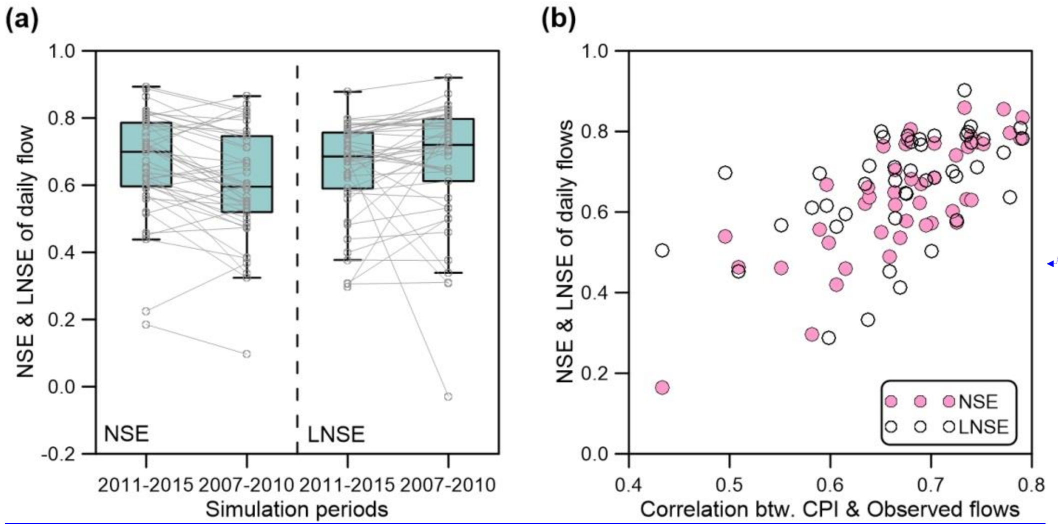
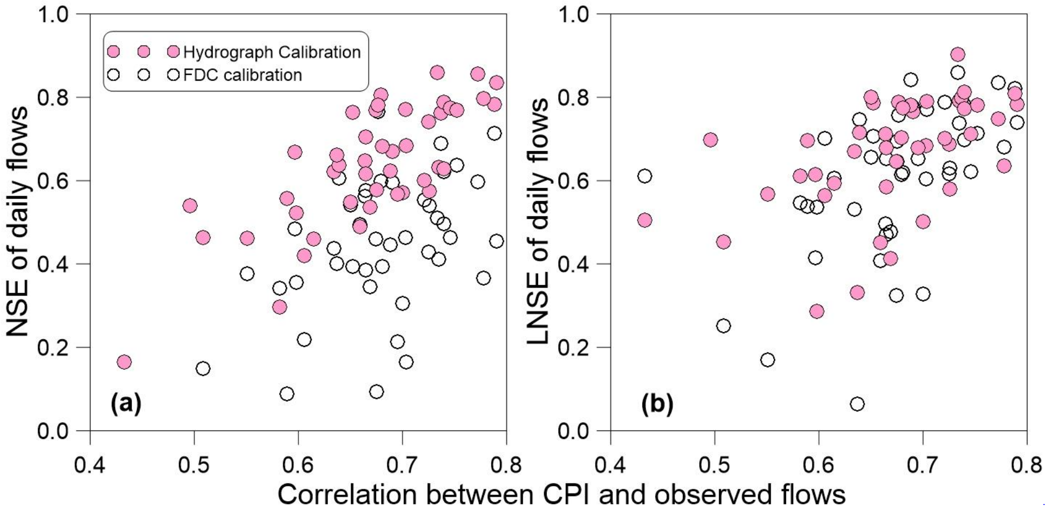


Figure 3: (a) box plots of high flow (NSE) and low flow (LNSE) reproducibility of the behavioural parameters obtained from the hydrograph calibration at the 45 catchments, (b) the relationship between the input-output consistency and the model performance. The straight lines in the box plots connect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

Formatted: Left, Line spacing: single



~~**Figure 4: The relationships between model the behavioural parameters obtained from the hydrograph calibration at the 45 catchments, (b) the relationship between the input-output consistency and (a) high flow reproducibility (NSEs) and (b) low flow reproducibility (LNSEs)**~~

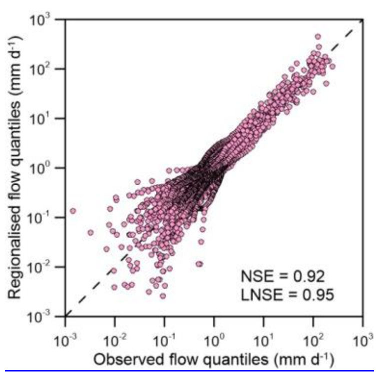
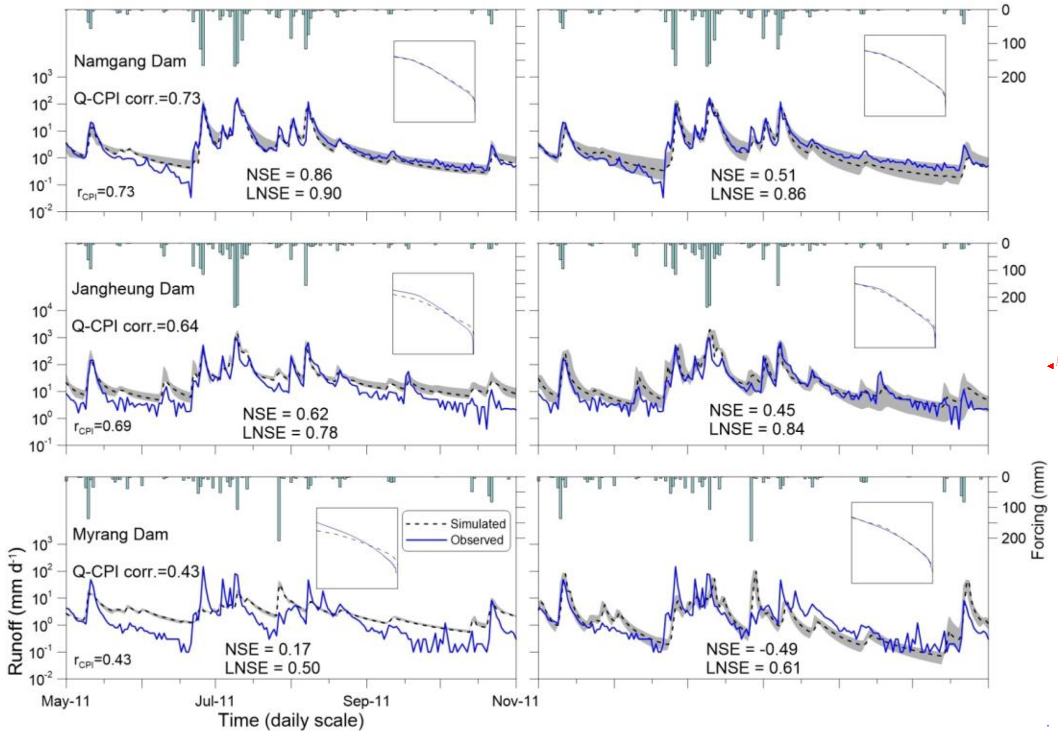


Figure 4: 1:1 scatter plot between the empirical flow quantiles and the flow quantiles predicted by the top-kriging FDC regionalisation method.



~~the model~~

~~**Figure 5: Observed and predicted hydrographs (continuous and dashed lines) with estimated uncertainties (shaded area) at three stations with best (top), intermediate (middle), and worst (bottom) predictive performance respectively. The plot inside of each hydrograph present agreement between observed and modelled FDCs in log-log space in which its horizontal and vertical axes are for exceedance probability (range of 0-1) and runoff (same range of each hydrograph) respectively. . 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.**~~

Formatted: Normal

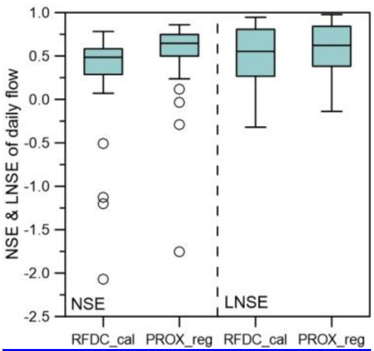
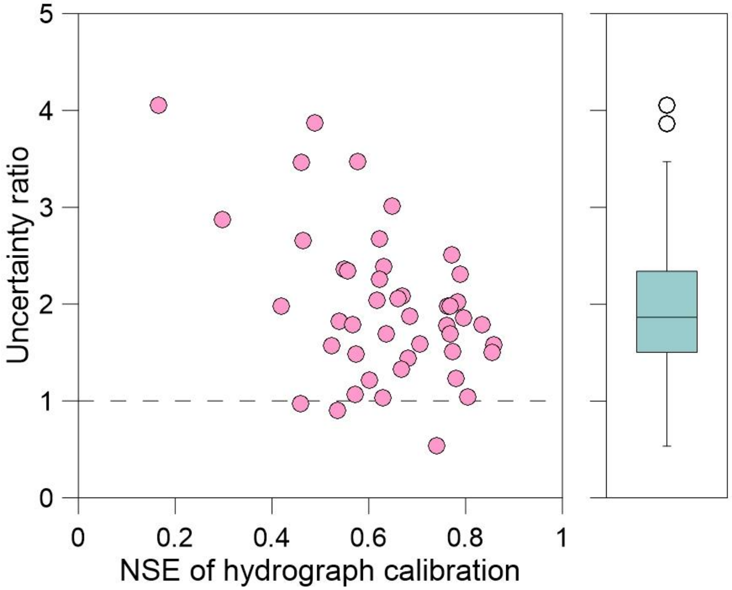


Figure 5: Box plots of NSE and LNSE values between the observed and the predicted hydrographs by RFDC_cal and PROX_reg for the 45 catchments under the cross validation mode.



Uncertainty ratio
NSE of hydrograph calibration
0
1
2
3
4
5
0
0.2
0.4
0.6
0.8
1

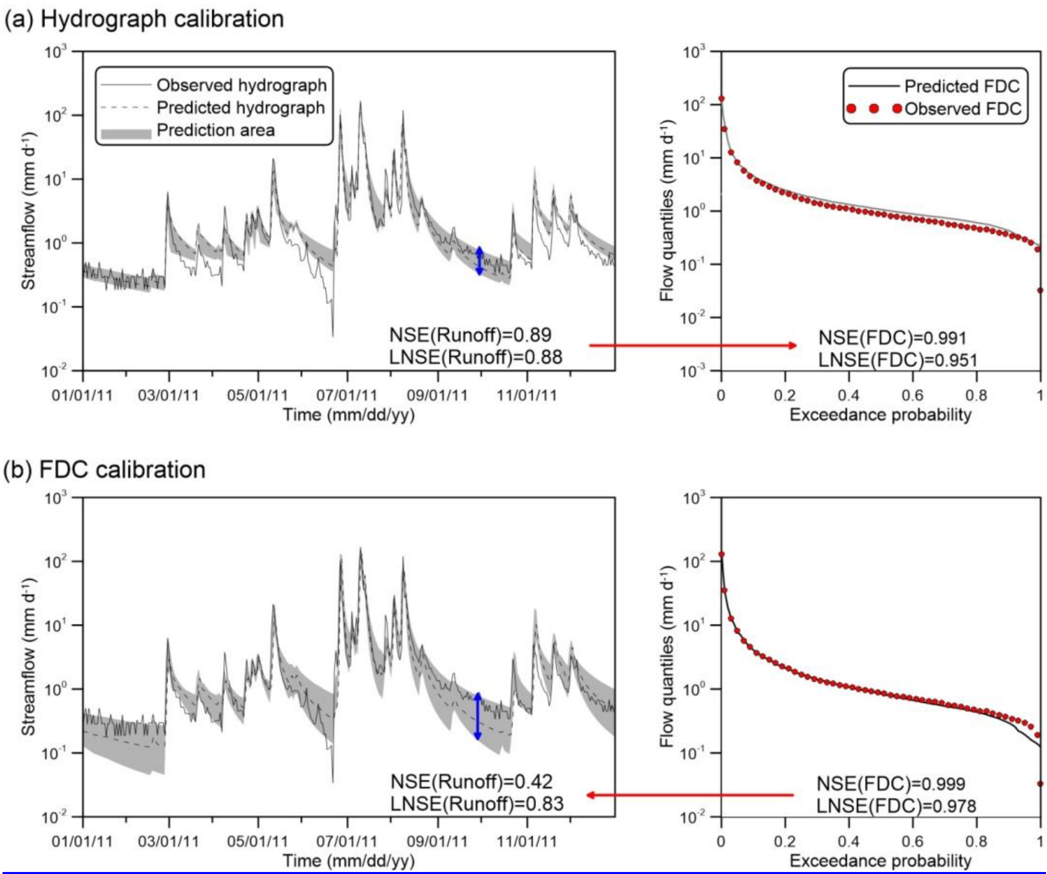
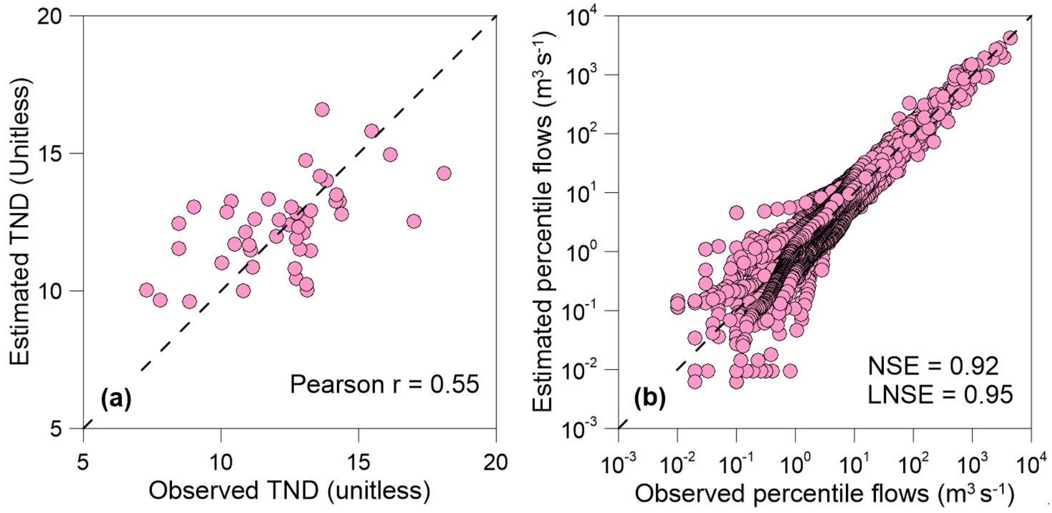


Figure 6: The observed and predicted hydrographs, the prediction areas, and the observed and predicted FDCs given by (a) the hydrograph calibration and (b) the FDC calibration for the Catchment 2.



~~**Figure 7: (a) 1:1 scatter plots between the observed and estimated TNDs, and (b) the observed and estimated quantile flows of 45 catchments.**~~

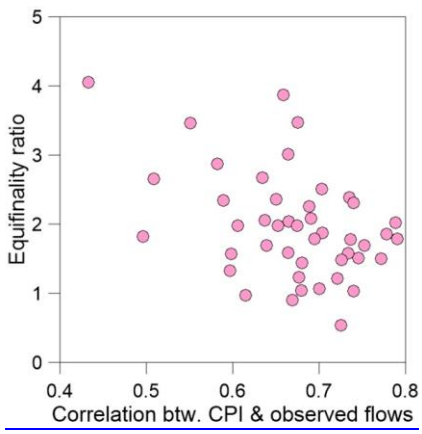
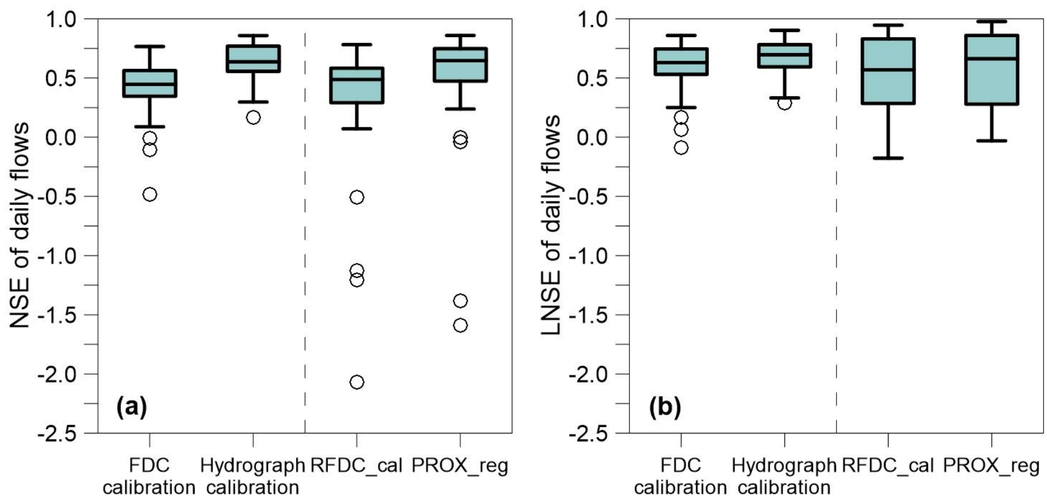


Figure 7: The input-output consistency vs. equifinality increased by replacing the hydrograph calibration with the FDC calibration. The equifinality ratio is defined as the ratio between the prediction areas of the 50 behavioural parameters gained from the FDC calibration and the hydrograph calibration.



~~Figure 8: (a) boxplots of NSEs (high flow reproducibility) of methods for gauged catchments (~~top-kriging **FDC** ~~and Hydrograph calibrations) and for ungauged catchments (RFDC_cal and PROX_reg), (b) boxplots of LSNEs (low flow reproducibility) gained from the same methods. The dashed lines distinguish between~~regionalisation **method**~~for gauged and ungauged catchments~~.

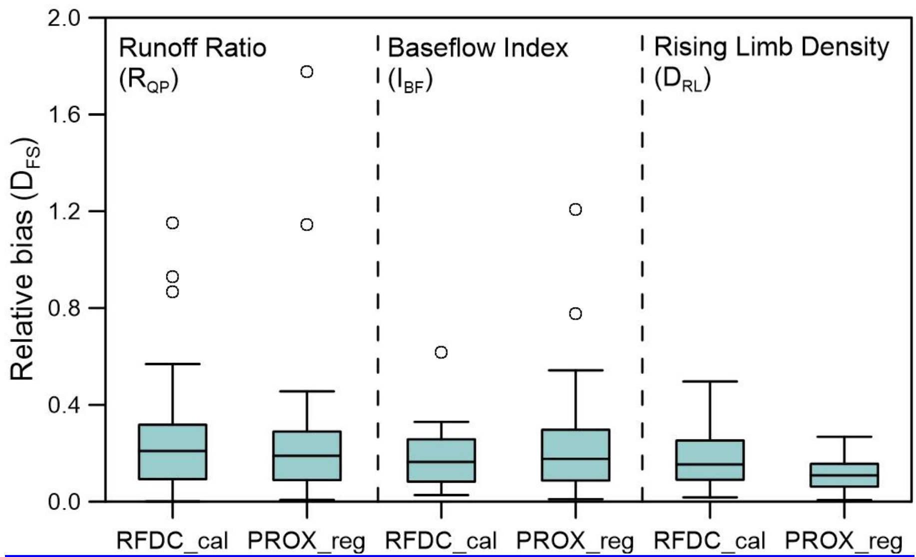
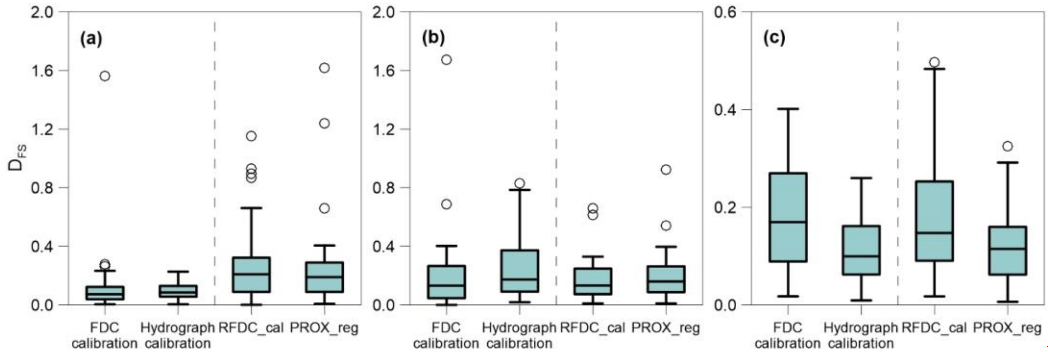


Figure 8: Flow signature reproducibility comparison between RFDC_cal and PROX_reg in terms of R_{QP} (a), I_{BF} (b), and D_{RL} (c).

Formatted: Font: 10 pt, Not Bold

Formatted: Normal



~~Figure 9: Flow signature reproducibility of methods for gauged catchments (FDC and Hydrograph calibrations) and for ungauged catchments (RFDC_cal and PROX_reg) in terms of (a) R_{QP}, (b) I_{BF}, and (c) D_{RL}. The dashed lines distinguish between method for gauged and ungauged catchments.~~

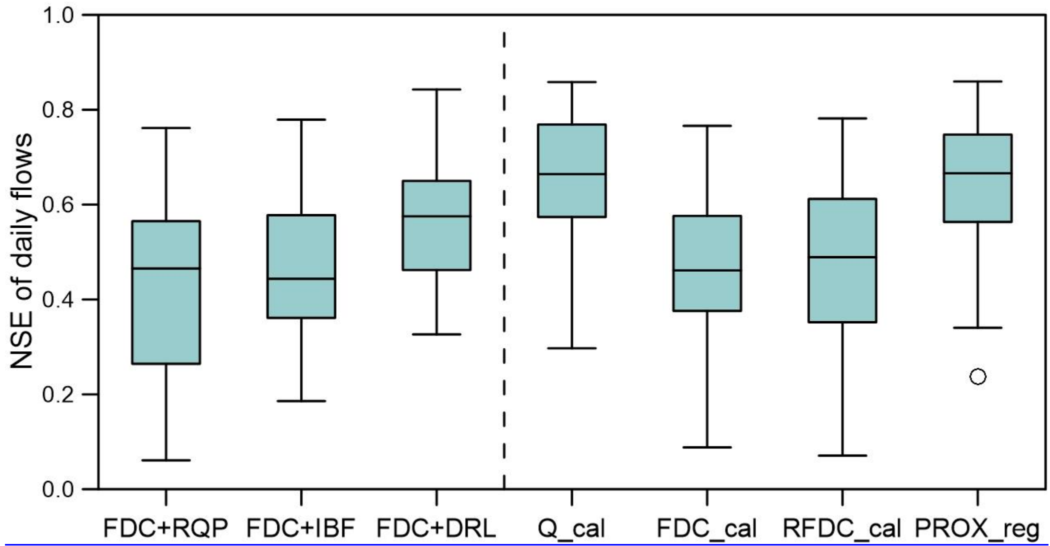
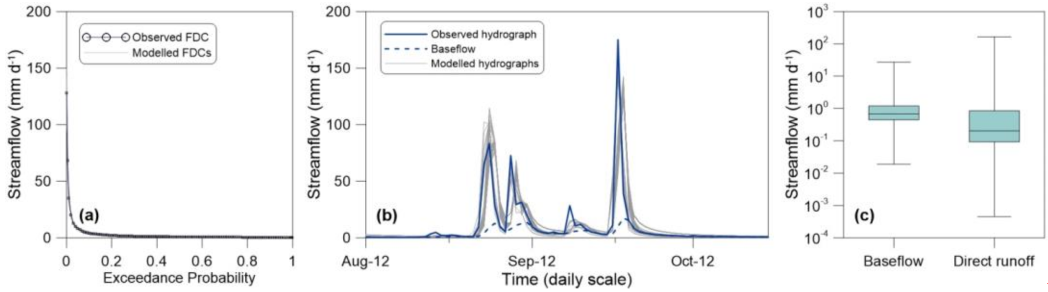
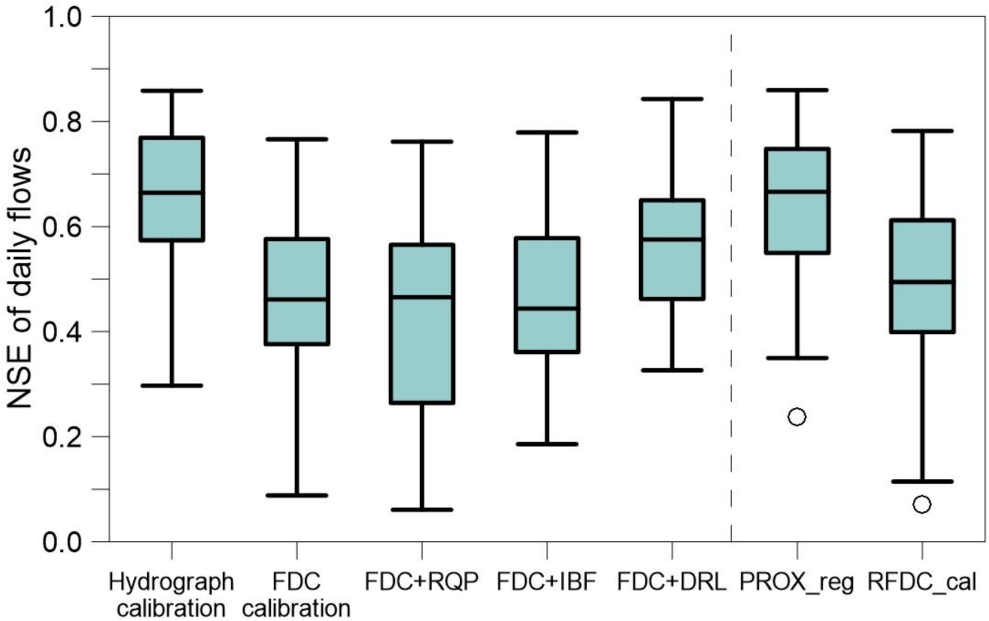


Figure 9: Predictive performance of the FDC calibrations additionally conditioned by R_{QP} (FDC+RQP), I_{BF} (FDC+IBF), and D_{RL} (FDC+DRL) in comparison to the other modelling approaches. Q_cal and FDC_cal refer to the hydrograph and the FDC calibration in gauged catchments respectively. 38 catchments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



~~**Figure 10: (a) observed FDC and FDCs modelled by the 50 parameter sets from the FDC calibration, (b) sample observed hydrograph, and hydrograph modelled by the same 50 parameter sets, and (c) Box plots of observed baseflow and direct runoff. The whiskers indicate maximum and minimum values. All panels are for Namgang dam (catchment 2) with 0.86 and 0.51 NSEs of daily flows using the hydrograph calibration and the FDC calibration respectively.**~~



~~**Figure 11: 8: Flow signature reproducibility comparison between RFDC_cal and PROX_reg in terms of R_{QP} (a), I_{BF} (b), and D_{RL} (c).**~~