Estimation of instantaneous peak flow from maximum mean daily flow by regionalization of catchment model parameters

J.Ding[1,2], U. Haberlandt[2]

[1] Key Laboratory of Tibetan Plateau Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China

[2]Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, University of Hannover, Germany

Abstract

Regionalization methods have been effectively used in many hydrological studies, such as regional flood frequency analysis and low flows. However, there is no study to estimate the instantaneous peak flow (IPF) from maximum mean daily flow (MDF) using hydrological models with regionalized parameters. In this paper, the semi-distributed conceptual hydrological model HBV (Hydrologiska Byråns Vattenbalansavdelning) is operated on a daily time step for 18 catchments in the Aller-Leine basin, Germany. The model is calibrated on four different flow statistics, including winter/summer extremes distribution and flow duration curves. The model parameter values are predefined with the associated catchment descriptors by a transfer function. Two different regionalization schemes are investigated: one is carried out for all the catchments in the study area; the other one is only performed for several catchments within a cluster. The k-means algorithm is used to 12 different catchment characteristics from all 18 catchments as the partitional clustering algorithm. Subsequently, the General Extreme Value (GEV) distributions are fitted to the modeled MDFs, which are then transferred into IPF quantiles using a multiple regression model.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/hyp.11053

The results show that: (1) the uncertainty resulted from model parameter regionalization for the estimation of IPFs is much smaller than the error when using MDFs instead of IPFs (2) the hydrological responses of the clustered catchments located in the flat areas are, in general, not as homogeneous as the ones in high elevated regions; (3) the model with the parameters derived from the same regionalization coefficients within a cluster performs better using the corresponding parameters estimated through all the catchments.

Key words: hydrological modeling; regionalization; instantaneous peak flow (IPF); maximum mean daily flow (MDF)

1 Introduction

Estimates of the design instantaneous peak flow (IPF) are important for solving a number of engineering and environmental problems, such as flood design and water resources management. Our previous study has shown that the often recorded maximum mean daily flow (MDF) data can be used to derive the design peak flow when no IPFs records of sufficient length are available at the target site (see Ding et al., 2015a). Concerning the impact of land use and climate change on extreme runoffs, the flow time series data cannot be used without consistency check. Hydrological modeling with local calibrated parameters has been proven to be a more robust and reliable approach to estimate IPFs from the modeled MDFs when taking into account the dynamic changes in the catchments of interest (see Ding et al., 2015b). However, both of the strategies cannot be carried out to predict the IPFs in ungauged areas due to the lack of hydrological data.

The most commonly adopted methods for estimation of design floods in ungauged areas include Index Flood Method by Hosking and Wallis, (1993), the Quantile Regression

Technique (Haddad and Rahman, 2012; Tasker and Stedinger, 1989) and Probabilistic Rational Method (Rahman et al., 2011; Young et al., 2009). The rainfall-based Design Event Approach has been recommended by many countries around the world, e.g. Australia and England(Hill and Mein, 1996). In this approach, the probabilistic nature of rainfall depth is considered for rainfall-runoff modeling, but the probabilistic behavior of temporal patterns of rainfall and runoff is ignored.

With the enhanced computational power nowadays, hydrological models have been commonly used to estimate the design flood in ungauged catchments given the model structures to be representative of the rainfall-runoff relationship (Athira et al., 2016; Cibin et al., 2014; Razavi and Coulibaly, 2013b). An important aspect of hydrological modeling is the ability to consider the dynamic changes of the properties of the target basins. Besides, the digital spatial datasets obtained by modern techniques could be utilized by the hydrological models to improve their capability of predicting water resource dynamics in ungauged areas. The concept of hydrological similarity assumes that the runoff response to a given rainfall input in two different basins would be similar if similar rainfall-runoff processes occur, therefore, the model parameters for the ungauged area can be estimated using regional information derived from the neighboring gauged catchments (Merz and Blöschl, 2004; Seibert, 1999). This has been proven feasible for regionalization of model parameters on the basis of catchment characteristics (Sellami et al., 2014). Regionalization techniques including the parameter regression approach (Fernandez et al., 2000; Merz and Blöschl, 2004; Seibert, 1999; Servat and Dezetter, 1993) and nearest neighbor approach (Bárdossy and He, 2006; Chiew and Siriwardena, 2005; Merz and Blöschl, 2004) are implemented to transfer model parameters from gauged catchments with calibrated parameters to ungauged catchments of similar hydrological characteristics. McIntyre et al., (2005) doubted the regression-based

approach and suggested investigating the relationship between model parameter and the catchment descriptor to produce a joint distribution of model parameters. However, it requires a large number of gauged catchments. Parajka et al., (2005) reviewed the applications of regionalization methods in a number of studies including their successes and failures**.**

As the main-stream method, there are two steps involved in the regression-based regional approaches: (1) estimation of watershed model parameters at each individual catchment independently, followed by (2) attempts to relate the model parameters to catchment characteristics. Examples of this method can be found in Abdulla and Lettenmaier (1997) and Sefton and Howarth (1998). However, the transfer of parameters is difficult due to the non-uniqueness of the model parameters (see Beven and Freer, 2001). To improve this, an one-step approach is proposed by Hundecha and Bárdossy, (2004). In this approach, calibration is performed without making any direct reference to the model parameters but to the coefficients of the predefined regression function. Model parameters are then regionalized through simultaneous calibration of the same hydrological model on different catchments. This one step regionalization method would be beneficial for estimating IPFs in ungauged catchments.

In this study, a methodology for the regionalization of a lumped conceptual hydrological model is applied. It involves calibration for different catchments simultaneously by using a predefined functional form of the relationship between the model parameters and catchment characteristics. With daily hydrologic data from 18 catchments over a period of more than 30 years, the same hydrological modeling strategy is used as the way we have applied in Ding et al., (2015b). Specifically, we address the following research questions: (a) how the regional

relationships between model parameters and catchment characteristics impact the ability of the hydrological model to estimate IPFs in ungauged area, while the parameter transfer tends to cause deterioration of model performance; (b) how the model performs for the estimation of IPFs in ungauged areas with and without clustering. Note that comparison of different clustering strategies is beyond the scope of this study. (c) is there any spatial pattern of the model performance with the regionalized parameters.

2 Methodology

2.1 Hydrological model and calibration strategy

HBV is a semi-distributed conceptual hydrological model with lumped parameters. It is used in this manuscript to simulate daily streamflow with daily precipitation, air temperature, potential evaporation and monthly estimates of crop coefficients as input. The HBV model has been proven to be an effective and useful tool for flood simulations in many studies (Hlavcova et al., 2005; Velasco et al., 2013; Wallner et al., 2013). A more detailed description of the HBV model can be found in SMHI, (2008).

The optimization technique Dynamically Dimensioned Search (DDS) proposed by Tolson and Shoemaker, (2007) is utilized to calibrate the model parameters. DDS is a stochastic, single objective search algorithm that has been applied in many hydrological studies for model calibration (Arsenault et al., 2014; Muleta, 2012). Compared with the traditional hydrograph calibration, calibrating directly on flood distributions is more favorable for the purpose of design flood simulation (Cameron et al., 1999; Ding et al., 2015b; Haberlandt and

Radtke, 2014). In this study, the objective function is a combination of the Nash-Sutcliffe Coefficient (N_{SC}):

$$OF = 0.275 \cdot NSC_{CDF-SUM} + 0.275 \cdot NSC_{CDF-WIN} + 0.2 \cdot NSC_{FDC} + 0.25 \cdot NSC_{MDF} \quad (1)$$

where $NSC_{CDF-SUM}$ and $NSC_{CDF-WIN}$ is N_{SC} of the cumulative distribution function of daily extremes for summer and winter, respectively; NSC_{FDC} indicates N_{SC} of the flow duration curves (FDC); and NSC_{MDF} denotes N_{SC} of the annual maximum mean daily flow series.

Those four goodness-of-fit measures are used to give more robust parameter estimation for high flows simulation. The weights associated with the above four stream flow statistics are determined by observing the N_{SC} and bias values with a trial-and-error approach. Here, the same values are used as those in our previous work (see Eq. (2) in Ding et al., 2015b). The sum of weights for summer (May-October) and winter (November-April) flood distribution curves is set to 55%. The remaining 45% is portioned between flow duration curve (20%) and annual maximum mean daily flow series (25%). It indicates that the objective function applied here focus more on the extremes. Noted that the model is calibrated for several catchments simultaneously, and the sum of the individual objective functions corresponding to all the catchments is considered in the DDS program.

2.2 Parameter transfer schemes

The transfer function implemented in this study is inspired by the linear transfer function approach applied successfully by Wallner et al., (2013) and Hundecha and Bárdossy, (2004). In this transfer function, the model parameters are linked with the catchment characteristics

and can be estimated uniformly for all catchments based on the selected catchment characteristics. It can be described in a linear form as:

$$\Psi_{p,b} = \sum_{i=1}^{I} \alpha_{p,i} \cdot S_{b,i} + \sum_{j=1}^{J} \beta_{p,j} \cdot L_{b,j} + \gamma_p \cdot slope_b \quad (2)$$

where $\Psi_{p,b}$ is the transferred value of hydrological model parameter *p* for basin *b*; $S_{b,i}$ and $L_{b,j}$ represents the *i*th soil property and *j*th land use characteristics for basin *b*, respectively; $slope_b$ denotes the basin slope; $\alpha_{p,i}$, $\beta_{p,j}$ and γ_p are regression coefficients.

After establishing the relationship between model parameters and catchment characteristics, the model is calibrated by simultaneous calibration of the regression coefficients of the transfer function instead of the model parameters. The regression coefficients of the transfer function keep unique for all the study catchments, while different catchments have different parameter sets. Following our previous research (Ding et al., 2015b), the same six parameters of HBV model (Table 1) are selected for model calibration. The transfer functions with the same catchment characteristics as derived by Wallner et al., (2013) for the same study area are used here. Table 2 shows the six model parameters and their corresponding linked catchment descriptors. The definition of the catchment descriptors can be found in Table 1.

Table 1. Definition for the symbols and abbreviations

Table 2. Relationship between model parameters and their corresponding catchment descriptors for the linear transfer function

2.3 Two regionalization schemes: R_{all} and $R_{clusters}$

There are two different parameter regionalization schemes explored for estimation of IPFs from MDFs using hydrological modeling. In the first regionalization scheme, model parameters of one specific catchment are obtained from simultaneous calibration of all catchments in the whole study area (referred as "R_{all}"). In the second regionalization scheme, model parameters of one catchment are derived from calibration of a cluster of catchments with similar catchment characteristics of their flow regime (referred as "$R_{cluster}$").

A variety of cluster analysis techniques are available to arrange catchments into groups with similar characteristics. However, none of them has been proven universally outperforming the others (Hannah et al., 2005). To identify homogeneous regions with similar flood response characteristics, k-means method and the principal component analysis (PCA) are used. K-means method is one of the simplest unsupervised learning algorithm and its procedure follows a fast and robust way to classify a given data set into a certain number of clusters (Burn and Goel, 2000; Isik and Singh, 2009; Kahya and Demirel, 2007). The main function of PCA is to reduce the dimensionality of a data set that consists of a large number of interrelated variables, while retaining most of the variation in the data set (Jolliffe, 2004). This is realized by transforming variables into a smaller set of principal components which are not correlated. Here, ten watershed attributes, one flow characteristic and one climatic variable are incorporated into the PCA to derive the significant components in pooling homogeneous regions. Since the stream flow data are quite limited in ungauged areas, some easily measurable watershed attributes, such as mean elevation, catchment area, and mean annual rainfall, are used for the cluster analysis of catchments (see Table 1). The leading

principal components with both eigenvalues greater than 1 and cumulative variance greater than 80% are selected for subsequent clustering processes. Then, k-means clustering is carried out with a predefined optimum number of clusters.

For both regionalization schemes, the simulated annual maximum mean daily flows (MDFs) are subsequently fitted to the GEV distribution with L-moments (see Hosking and Wallis, (1997)). Quantile values (HQ_{MDF}) for four return periods (T = 10, 20, 50, 100 yrs) can be computed. The simulated MDF quantiles are then transformed into instantaneous peak flow quantiles (HQ_{IPF}) through a multiple regression model:

$$HQ_{IPF} = \alpha_1 \cdot HQ_{MDF} + \alpha_2 \cdot lst\_fp + \alpha_3 \cdot Elv\_ds + \alpha_0 \quad (3)$$

where *lst_fp* denotes the longest flow path; *Elv_ds* indicates the minimum elevation; α_0, α_1, α_2 and α_3 are regression coefficients. We have verified the GEV distribution in fitting the extreme flows and testified the multiple regression model for post correction of MDFs in Ding et al., (2015a).

Four goodness-of-fit measures (as shown in Table 3), namely, the Nash-Sutcliffe Coefficient (N_{SC}), root mean square error (RMSE), Bias of total runoff (Bias) and Bias of simulated quantile of peak flow (Bias-1), are used to evaluate performance of the two regionalization approaches R_{all} and $R_{cluster}$ for estimation of IPFs.

Table 3. Goodness-of-fit measures

3 Study area and data

The investigations are carried out for 18 catchments within the Aller-Leine River basin in northern Germany (Figure 1). The 18 study catchments are located in different geomorphologic areas. 11 catchment descriptors (Table 4 from column 2 to 12) are used for clustering. The first six characteristics (Area – aspR) are derived from a Digital Elevation Model (DEM) with a resolution of 10 meters. The main orientation (aspR) ranges between 0 and 1, the bigger the value, the greater the portion of the basin orient to the north. The soil properties of effective field capacity (FC), saturated hydraulic conductivity (Kf) and total pore volume (TPV) are estimated from the German digital soil data base BÜK1000 (Hartwich et al., 1995). The portion of the forest land use type is derived from the land cover map CORINE2000 (EUR, 1994). Based on the observed runoff with an automatic base flow filter, the recession constants are calculated in different hydrogeological units (HGUs) (Arnold et al., 1995). The mean recession constants for each catchment are weighted according to the contributing area of the HGUs.

Figure 1. The locations of 18 catchments of Aller-Leine in northern Germany

Table 4. Catchment descriptors of the 18 catchments

Observed discharge is available as daily flows and monthly peak flow series within the period from 1965 to 2008 for all the 18 catchments. Other meteorological data applied to force the hydrological model, such as temperature and evaporation, are available for longer time

periods, between 1951 and 2008. The inverse distance weighting (IDW) method is used to estimate the rainfall distribution based on the daily rainfall data from 34 rain gauge stations. An optimal power value of 2.3 for the power parameter is determined by minimizing the lowest RMSE.

4 Results and discussion

4.1 Distinction between groups

Prior to clustering, PCA is first employed to reduce the dimensionality of the data set which gives the Principal Components (PCs). The data set consists of the selected twelve variables shown in Table 4.

Table 5 demonstrates that the first three PCs explain 45.269, 20.89 and 14.48%, respectively, and their cumulative variances account for 80.639% of the total variance of the data set. Besides, their eigenvalues are all greater than 1. Lowest elevation (Elv_ds), Mean elevation (Melv), basin slope (Bslop), ratio of forest (LUF) and annual precipitation (P) have a high correlation with the first principal component. Area (Area), longest flow path (Lst_fp) and soil conductivity (Kf) are correlated with the second and third principal. Therefore, more effect from these attributes is expected in the cluster analysis procedure. The first PC shows a contrast between shape factor of the catchment (with negative coefficients) and soil parameters (positive coefficients). The main contrast for the second component lies between the mean recession constant (BFR) and basin aspect (aspR). The last PC is a trade-off between Area (Area) and soil conductivity (Kf).

Table 5. Summary of PCA for the twelve catchment characteristics

As shown in Figure 2, there is a break (elbow) started at number four which implies 4 clusters for this case. This number will be adjusted further using k-means algorithm by comparing other clusters situations (from 2 to 5 groups). In the end, the entire catchment has been grouped into three homogeneous groups to retain a reasonable number of objects in each group (Figure 3).

As shown in Figure 3, there are four catchments in Group 1, ten catchments in Group 2 and four catchments in Group 3. Some geographic trends could be noticed and the location distribution is the strongest distinguishing feature of the groups. Group 1 is situated in the northern flat area while the other two groups are distributed over the midlands and southern mountainous area, respectively.

Figure 2. Percentage of variance explained as a function of the number of principal components

Figure 3. Spatial distribution of the three groups in Aller-Leine, Germany

More details regarding the catchment characteristic distribution in each cluster are illustrated in Figure 4. Catchments in the northern part of Aller-Leine basin (Group 1) are characterized by lower annual precipitation (733mm/year -872mm/year). Due to the relatively low elevation (less than 100m) and high pore volume (35%), the catchments in Group 1 have slower flow processes than the other two groups. Group 2 is composed of some small catchments covered with high proportion of forest (medians of Area and LUF are 110 km^2 and 40%, respectively). High flows occurring during spring and early summer season from freezing soils and snow cover can be found in these mountainous catchments. Group 3 shows

the smallest range of most catchment characteristics, such as mean conductivity (120mm/h) and total pore volume (27%).

Figure 4. Box plots of the catchment characteristics of the three groups. Thick black line is the median value. The box shows the inter-quantile range between 25^{th} and 75^{th} quantiles of the data. The ends of the whiskers represent 1.5 times the inter-quantile range.

4.2 Comparison of model performances using two different regionalization schemes

Figure 5 shows the comparisons between fitted probability distributions of the observed and simulated annual daily extremes in winter and summer for three selected catchments Br, Pi and De. These catchments are randomly selected from the three classified groups to give a more detailed descriptions. The overall results for all the 18 catchments are shown in Figure 6. In Figure 5, the black solid line denotes the fitted GEV distribution on the observed annual daily extremes (black dots). The black dashed lines enclose the 90% confidence interval for the observations obtained by using a bootstrap method after Efron and Tibshirani, (1986). The purple lines represent distributions of simulated daily extremes, which are generated by the R_{all} parameter regionalization scheme, whereas the green lines correspond to the $R_{clusters}$ parameter regionalization scheme.

It is apparent that the distributions of the simulated extremes generated by $R_{clusters}$ (the green lines) are closer to the distributions of the observations (the black solid lines) than that

generated by R_{all} (the purple lines) in both seasons, especially in winter. This indicates that the simulation of the extreme flows can be significantly improved when calibration is performed separately on catchments of homogeneous clusters rather than on the whole study area. For both regionalization schemes, the simulated distribution curves (green and purple lines) lie below the observed distribution curve (black solid line) at high return periods, which in fact indicates some underestimations of the peak flows. It could be explained that the extremes (caused by intense rainfall) are hard to be well modeled with inaccurate rainfall data. The uncertainty bands of the observed annual extreme values show a considerable scatter with the flow records about 30 years.

Figure 5. CDFs of observed and simulated daily extremes in winter and summer for the three randomly selected sample catchments (Br, Pi & De); the black dots are the observed annual daily extremes; black dashed lines enclose the 90% confidence interval against observed peak flows

The results of goodness-of-fit Chi-square test at 5% significance level for all 18 catchments are presented in the form of violin plots (Figure 6). Performances of the hydrological model using two regionalization schemes are compared between clusters and for the entire catchments. The p value is used for assessing the agreement between the observed and simulated extreme flow frequencies. Larger value of p indicates a better fit.

The benefits of clustering can be seen in the final overall performance (‘all’) with higher median p values and larger quantile ranges than the corresponding results without clustering for both winter and summer season. It means that aggregating the catchments into groups helps improving prediction of flood quantiles. However, the distributions of p values are

quite different for both schemes in the three clusters. The catchments in Group 2 show the lowest level of agreement between the observed and simulated extreme flow frequencies while the catchments in Group 3 have the highest level. This may indicate the transforming equation for estimating the model parameters is more limited in low elevation areas as compared with its performance in high elevation areas. The prediction results of flood frequencies in winter season are better than in summer season for the three clusters, which is consistent with the results presented in Figure 5.

Figure 6. Violin plots of the p value over all catchments for the fitted GEV distribution between observed and simulated daily extremes in winter and summer respectively. The middle black solid line is the median value and the red dashed line is the 5% significance line; the middle black rectangle enclose the 25% and 75% quantile values. The left purple half is R_{all} approach and the right green half is $R_{clusters}$ approach. The width of each violin plot represents the probability distribution

To assist the interpretation of results in Figure 5, statistical analysis (paired t-test at α=0.05) is used to determine if the p values derived from R_{all} and $R_{clusters}$ schemes are significantly different. Table 6 shows the results for all 18 catchments and three clusters in both summer and winter seasons. As can be seen, for summer season, the p values calculated from the two regionalization schemes are proved to be significantly different for “all”, Group 1 and Group 3. Nevertheless, for winter season, this is only true for Group 3.

Table 6. Results of paired t-test on the p values for both regionalization schemes.

The flow duration curves (FDCs) are constructed using average daily flows in terms of 6 quantile values (0.05, 0.25, 0.5, 0.75, 0.95, 0.975 Quantile). Figure 7 shows a summary of the

assessment results of FDCs in all three clusters and 18 catchments by using the two different regionalization schemes. For the three clusters, R_{all} and $R_{clusters}$ perform well with the median values of the Nash-Sutcliffe Coefficient (N_{SC}) around 0.8. However, both schemes show some overestimation with positive median values of bias around 20% and 35% for R_{all} and $R_{clusters}$, respectively.

Overall, the assessment results of all 18 catchments ('all') in Figure 7 show that there is no significant improvement for predicting the FDCs in ungauged areas using the regional calibration with clustering. Moreover, Figure 7 illustrates an opposite pattern across the three groups when compared with the assessment results of flood frequency curves presented in Figure 6. For Group 1, the model performance regarding flow duration curves is best compared with the other two clusters, whereas the model gives the worst results for predicting the flood distributions. This inconsistency is probably because (1) the objective function (Eq. (1)) treats flood frequency curves with higher weight, as compared to the flow duration curves; (2) the interaction between these two different flow statistics (extremes distribution and FDCs) might affect the model performance during optimization procedure.

Figure 7. Regional Calibration results of flow duration curve (FDC) using the Nash-Sutcliffe Coefficient (N_{SC}) and the Bias for two different regionalization schemes. The middle black solid line is the median value; the middle black rectangle enclose the 25% and 75% quantile values. The left purple half is R_{all} approach and the right green half is $R_{clusters}$ approach. The width of each violin plot represents the probability distribution

As above, Table 7 shows the results of paired t-test on the flow duration curve (FDC) using the Nash-Sutcliffe Coefficient (N_{SC}) and bias, before and after clustering for all cases. It shows the simulated FDC results derived from the two different regionalization schemes are not significantly different with p values larger than significance level 0.05. Namely, the regionalization approach with clustering ($R_{clusters}$) has less advantage of simulating FDC than simulating extreme flow distribution.

Table 7. Results of paired t-test on the flow duration curve (FDC) using the Nash-Sutcliffe Coefficient (N_{SC}) and bias for both regionalization schemes.

The pairwise comparison results of daily flows with 97.5% non-exceedance probability (P_{non}) for R_{all} and $R_{clusters}$ are shown in the form of a parallel coordinates plot (Figure 8). The criterion used to formulate the plot is Bias-1 (see Table 3). The estimation is more accurate if the lines ends closer to 0 (Bias-1=0). It is interesting to observe that there is no obvious increase or decrease in the relationship between the estimation error and area for both regionalization schemes. The average absolute values of Bias-1 obtained from all 18 catchments are 13.1% and 10.3% for R_{all} and $R_{clusters}$, respectively. For both regionalization schemes, Bias-1 is within a range from -20% to 20% for most catchments.

Figure 8. The values of Bias-1 for daily flow duration curve (FDC) at non-exceedance probability P_{non}= 0.975

Figure 9 gives the relationships between multiple regression coefficients (see Eq. (3)) and return periods for all 18 catchments in winter and summer seasons. The first three coefficients (a_1, a_2, a_3) in the upper part of Figure 9 are the correlation coefficients between

the IPFs and MDFs, longest flow path and minimum elevation, respectively. It can be noticed that the sign of all these three partial regression coefficients is the same for the two different regionalization schemes. This conformity with physical principles agrees with the findings of Ding et al., (2015a). It shows a positive relationship between IPFs and MDFs and minimum elevation, whereas IPFs have a negative relationship with longest flow path.

Figure 9. The values of multiple regression coefficients for the two regional calibration schemes

To better understand the differences between R_{all} and $R_{clusters}$, the frequently considered recurrence interval of 100-year design flood is selected here as an exemplary result. Relative errors between the observed and simulated quantile values over all 18 catchments are shown in Figure 10. The observed 100-year flood is estimated from the monthly peak flow discharges between 1965 and 2008. As above, the parallel coordinates plots are used to present the Bias-1 results for winter season (Figure 10a) and summer season (Figure 10b). This plot visualizes relationships between the relative error and area.

In general, the estimations of 100-year flood resulted from the $R_{clusters}$ scheme are slightly better than from the R_{all} scheme for all cases. The randomness of relative error and catchment area patterns is in agreement with the findings in Figure 8 for both schemes. Considerable differences between these two regionalization schemes are found in catchments of Group 2 and Group 3. Accordingly, it can be seen that some catchments in Group 2 contribute considerably larger error in both summer and winter seasons, than the others. This could be due to the uncertainties associated with model inputs and observed flow statistics or bad prediction from the multiple regression model.

Considering the seasonal performance, the model tends to spread errors more widely in winter season than in summer season for both schemes. This can be attributed to less predictors' effects in controlling the prediction of design flood in winter than in summer season. Furthermore, the simulations of Group 2 are generally worse than for the other two groups for both R_{all} and $R_{clusters}$. These findings are in consistent with the results in Figure 6.

Figure 10(a). Comparison between the two regionalization calibration schemes at a return period of 100-year in terms of Bias-1 for winter season for all 18 catchments

Figure 10(b). Comparison between the two regionalization calibration schemes at a return period of 100-year in terms of Bias-1 for summer season for all 18 catchments

4.3 Comparison of two regionalization schemes

Finally, to sum up the IPFs estimation results over the whole study area for four different return periods (T=10, 20, 50, 100yrs) in summer and winter seasons, respectively, the RMSE and bias criteria, are applied to check the overall estimation quality of the design IPFs by calibration on the flow statistics. Figure 11 summarizes the results from both R_{all} and $R_{clusters}$. The first column (MDF-IPF) shows the differences between the observed IPFs and the corresponding observed MDFs without correction. The second column (CDF_d) shows the estimation results by calibrating the HBV model individually for each catchment based on the

same calibration strategy as in this study. We discussed the details of the first two types of results in Ding et al. 2015b. They are used here as the benchmark error to compare with the results from R_{all} (the third column) and $R_{clusters}$ (the fourth column).

The upper section of Figure 11 shows the RMSE values from all 18 catchments in a bar-plot. Compared with directly using observed MDFs, both of regionalization schemes (R_{all} and $R_{clusters}$) show advantages of estimating IPFs regarding smaller RMSE and Bias values. As expected, the local calibration approach (CDF_d) outperforms the two regionalization schemes. However, it can be noticed that the discrepancies of RMSE between the CDF_d and $R_{clusters}$ are smaller than 6% for the four return periods in both seasons. This reveals that the clustering of catchments is useful to improve the model performance using regionalized parameters.

On average, the $R_{clusters}$ gives a RMSE of 23% in both winter and summer season, which is smaller than the corresponding results from R_{all} with a RMSE of 30%. As to the R_{all} method, the prediction of IPFs for high return periods is not significantly improved. The RMSE value showing in winter season at T=100 yrs (30%) is even a little higher than directly using the observed MDFs (28%). It means that the parameter set calculated simultaneously from all the 18 catchments is not as accurate as the one deduced by $R_{clusters}$ method.

As can be seen from the lower part of Figure 11, immediate replacement of IPFs with observed MDFs (MDF-IPF) can lead to significant underestimation of IPFs in both seasons for all return periods. In contrast, the regionalization schemes with additional post correction of simulated MDFs reduce the bias noticeably with slight over estimation for winter and summer seasons.

Figure 11. Comparison of root mean square error (RMSE) and Bias using regionalized parameter sets for clusters ($R_{clusters}$) and the whole study area (R_{all}) in both winter and summer season

4.4 Comparison of leave one out cross validation (LOOCV) performance of R_{all} and $R_{clusters}$

The reliability of the two parameter regionalization schemes is assessed through LOOCV for estimation of IPFs in ungauged area. During the LOOCV procedure, the site of interest is left out to be treated as an ungauged site and the hydrological model is calibrated concurrently using the remaining sites. This step is repeated for all selected catchments in this study. The simulated MDFs will be corrected by the multiple regression model and compared with the observed IPFs. As before, the RMSE and bias are used to assess the model efficiency.

Figure 12 presents the LOOCV comparison results of R_{all} and $R_{cluster}$ for both winter and summer seasons. The black lines representing the overall results for 18 catchments using R_{all} and $R_{clusters,}$ are found to be similar with the corresponding calibration results in Figure 11. However, it shows somewhat higher error due to the cross validation. The overall validation results of both schemes are similar in winter and summer season with an average values of RMSE 28% and bias 5% for the four return periods. Group 2 contributes the most error with RMSE values between 30-40% in winter and 40-50% in summer, which is consistent with the calibration results of flood frequency curves shown in Figure 6. It suggests: (1) $R_{clusters}$ provides better estimation results than R_{all} in both summer and winter season, which implies clustering is necessary for the regionalization in ungauged areas for flood quantile predictions;

(2) the similarity of hydrological responses from the catchments in Group 2 are hardly identified by their similar physical catchment characteristics, such as the elevation, land cover properties and soil properties; (3) the model performance is found to be satisfactory in Group 1 and Group 3 with low values of RMSE and bias values suggesting the regionalization with clustering approach is a useful tool for prediction of IPFs in ungauged basins. In addition, there is a significant improvement noted in the validation results of Group 3 for summer season, which suggests the hydrological responses in these catchments is more similar for summer flooding than for the snowmelt induced winter flooding.

Figure 12. RMSE and bias results by cross-validating both R_{all} and $R_{clusters}$ schemes for the three groups.

5 Conclusions

Estimation of design peak flow in ungauged areas based on daily flow data is essential for hydraulic infrastructures, flood management and planning for future development. In this paper, we have compared two different parameter regionalization schemes to estimate IPFs from MDFs, namely, R_{all} without clustering and $R_{clusters}$ with clustering. The lumped conceptual hydrological model HBV combined with flow statistic calibration strategy is operated on a daily basis for 18 catchments across Aller-Leine, Germany. The model efficiencies for the estimation of flood quantiles are acceptable given the model structure is adequate for the purpose of this paper. The objective function applied in this study is more robust than other ways to measure high flows. The assessment of model performance using R_{all} and $R_{clusters}$ schemes is based on LOOCV comparison results for all 18 catchments. This assessment is a measure of how well the regionalization schemes can be applied for estimation of IPFs in ungauged catchments. The main findings are summarized as follows:

The overall calibration results in Figure 11 shows that the model using the $R_{clusters}$ strategy performs better than using R_{all} strategy. The hydrologic cluster analysis is found to play a central role in estimation of IPFs and different combinations of the catchment predictors may result in quite different results. The $R_{clusters}$ strategy also shows more advantages with respect to extremes distribution compared with the flow duration curve simulation. This may suggest that the impacts of clustering on hydrological modeling are dependent on the chosen flow statistics. Thus, different calibration strategies could influence the predictive performance which is independent of the selection of regionalization scheme.

The average RMSEs for the four different return periods are 30% and 26% by using R_{all} and $R_{clusters,}$ respectively. This suggests inclusion of information on catchment similarity before regionalization of the model parameters helps the prediction performance of IPFs. The scatter of the performances among the three clusters indicated the ability of HBV model for predicting IPFs is higher for clusters with smaller number of catchments (Group 1 and Group 3) than for the ones with larger group members (Group 2). The relatively poor performance in Group2 can be explained as that, the classified catchments in Group 2 have similar watershed attributes but with hydrologically heterogeneous response. Additionally, the predictive performance of $R_{clusters}$ method shows obvious seasonality across the three different clusters, especially for Group 2 and Group 3. It decreases in summer for mountainous regions dominated in Group 2 and 3, while keeps constant in Group 1 characterized with flat areas for both summer and winter season. However, for these two regionalization schemes, the bias is relatively small and the random errors from three different clusters are substantial. It means the spatial transposition of the model parameters for flood quantile estimation is sensitive to the clustering results.

Figure 12 demonstrates that based on the established function between model parameters and readily available information of catchment attributes, regionalization of the HBV model parameters combined with post correction technique is capable of estimating the IPFs from simulated MDFs in ungauged areas. The catchment attributes as used in this paper represent the physiographic and hence hydrologic characteristics of catchments well. In general, regionalization method combined with clustering yields better predictive model performance in both summer and winter season.

For both Figure 11 and Figure 12, it appears that those selected attributes are good predictors of the hydrological dynamics at least for environments such as the study region of this paper. In addition, most of the selected catchment attributes are static attributes while dynamic indicators, such as seasonality measures and storm type indicators, are suggested by some hydrologists for application in ungauged basins (see Merz and Blöschl, 2004). Therefore, more efforts are needed in future to find better predictive variables and similarity measures than those currently used catchment predictors for parameter regionalization.

Acknowledgements

The authors thank their colleagues for their valuable comments and suggestions. We are also grateful for the right to use data from the German National Weather Service (DWD), NLWKN Niedersachsen, Landesamt für Geoinformation und Landentwicklung Niedersachsen, Landesamt für Bergbau and the funding from the China Scholarship Council (CSC).

Reference

Abdulla, F. A., and Lettenmaier, D. P., 1997, Development of regional parameter estimation equations for a macroscale hydrologic model: Journal of Hydrology, v. 197, no. 1–4, p. 230-257.

Arnold, J. G., Allen, P. M., Muttiah, R., and Bernhardt, G., 1995, Automated Base Flow Separation and Recession Analysis Techniques: Ground Water, v. 33, no. 6, p. 1010-1018.

Arsenault, R., Poulin, A., Cote, P., and Brissette, F., 2014, Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration: Journal of Hydrologic Engineering, v. 19, no. 7, p. 1374-1384.

Athira, P., Sudheer, K. P., Cibin, R., and Chaubey, I., 2016, Predictions in ungauged basins: an approach for regionalization of hydrological models considering the probability distribution of model parameters: Stochastic Environmental Research and Risk Assessment, v. 30, no. 4, p. 1131-1149.

Bárdossy, A., and He, Y., 2006, APPLICATION OF A NEAREST NEIGHBOUR METHOD TO A CONCEPTUAL RAINFALL-RUNOFF MODEL: Proceedings of the 7th International Conference on HydroScience and Engineering Philadelphia, USA September 10-13, 2006 (ICHE 2006).

Beven, K., and Freer, J., 2001, Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology: Journal of Hydrology, v. 249, no. 1–4, p. 11-29.

Burn, D. H., and Goel, N. K., 2000, The formation of groups for regional flood frequency analysis: Hydrol. Sci. J., v. 45(1), p. 97–112.

Cameron, D. S., Beven, K. J., Tawn, J., Blazkova, S., and Naden, P., 1999, Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty): Journal of Hydrology, v. 219, no. 3–4, p. 169-187.

Chiew, F. H. S., and Siriwardena, L., 2005, Estimation of SimHyd parameter values for application in ungauged catchments: International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand, p. 2883–2889.

Cibin, R., Athira, P., Sudheer, K. P., and Chaubey, I., 2014, Application of distributed hydrological models for predictions in ungauged basins: a method to quantify predictive uncertainty: Hydrological Processes, v. 28, no. 4, p. 2033-2045.

Ding, J., Haberlandt, U., and Dietrich, J., 2015a, Estimation of the instantaneous peak flow from maximum daily flow: a comparison of three methods: Hydrology Research v. 46, no. 5, p. 671-688.

Ding, J., Wallner, M., Müller, H., and Haberlandt, U., 2015b, Estimation of instantaneous peak flows from maximum mean daily flows using the HBV hydrological model: Hydrological Processes

Efron, B., and Tibshirani, R., 1986, Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy: Statistical Science, v. 1, no. 1, p. 54-75.

EUR, 1994, EUR 12585 - CORINE land cover project - Technical guide. European Commission. Office for Official Publications of the European Communities. Luxembourg.

Fernandez, W., Vogel, R. M., and Sankarasubramanian, A., 2000, Regional calibration of a watershed model: Hydrological Sciences Journal, v. 45, no. 5, p. 689-707.

Haberlandt, U., and Radtke, I., 2014, Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows: Hydrol. Earth Syst. Sci., v. 18, no. 1, p. 353-365.

Haddad, K., and Rahman, A., 2012, Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework – Quantile Regression vs. Parameter Regression Technique: Journal of Hydrology, v. 430–431, no. 0, p. 142-161.

Hannah, D. M., Kansakar, S. R., Gerrard, A. J., and Rees, G., 2005, Flow regimes of Himalayan rivers of Nepal: nature and spatial patterns: Journal of Hydrology, v. 308, no. 1–4, p. 18-32.

Hartwich, R., Behrens, J., Eckelmann, W., Haase, G., Richter, A., Roeschmann, G., and Schmidt, R., 1995, Bodenübersichtskarte der Bundesrepublik Deutschland 1:1000000 (BÜK 1000). Karte mit Erläuterungen, Textlegende und Leitprofilen Bimdesamstaöt für Geowissenschaften und Rohstoffe, Hannover.

Hill, P., and Mein, R., 1996, Incompatibilities between storm temporal patters and losses for design flood estimation: Hydrology and Water Resources Symposium, Hobart, Tasmania, Australia, p. 445–451.

Hlavcova, K., Kohnova, S., Kubes, R., Szolgay, J., and Zvolensky, M., 2005, An empirical method for estimating future flood risks for flood warnings: Hydrology and Earth System Sciences, v. 9, no. 4, p. 431-448.

Hosking, J. R. M., and Wallis, J. R., 1993, Some statistics useful in regional frequency analysis: Water Resources Research, v. 29, no. 2, p. 271-281.

Hosking, J. R. M., and Wallis, J. R., 1997, Regional frequency analysis:an approach based on L-moments: Cambridge University Press ,New York, USA.

Hundecha, Y., and Bárdossy, A., 2004, Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model: Journal of Hydrology, v. 292, no. 1–4, p. 281-295.

Isik, S., and Singh, V. P., 2009, Hydrologic Regionalization of watersheds in Turkey: J. Hydrol. Eng, v. 13(9), p. 824–834.

Jolliffe, I., 2004, Principal component analysis, 2nd edn., USA, Springer Series in Statistics.

Kahya, E., and Demirel, M. C., 2007, A Comparison of low-flow clustering methods: Streamflow grouping: J. Eng. Appl. Sci, v. 2(3), p. 524–530.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T., 2005, Ensemble predictions of runoff in ungauged catchments: Water Resources Research, v. 41, no. 12, p. W12434.

Merz, R., and Blöschl, G., 2004, Regionalisation of catchment model parameters: Journal of Hydrology, v. 287, no. 1–4, p. 95-123.

Muleta, M. K., 2012, Improving Model Performance Using Season-Based Evaluation: Journal of Hydrologic Engineering, v. 17, no. 1, p. 191-200.

Parajka, J., Merz, R., and Blöschl, G., 2005, A comparison of regionalisation methods for catchment model parameters: Hydrol. Earth Syst. Sci., v. 9, no. 3, p. 157-171.

Rahman, A., Haddad, K., Zaman, M., Kuczera, G., and Weinmann, P., 2011, Design flood estimation in ungauged catchments: A comparison between the probabilistic rational method and quantile regression technique for NSW, Volume 14 (2), p. 127 - 139.

Razavi, T., and Coulibaly, P., 2013a, Classification of Ontario watersheds based on physical attributes and streamflow series: Journal of Hydrology, v. 493, p. 81-94.

Razavi, T., and Coulibaly, P., 2013b, Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods: Journal of Hydrologic Engineering, v. 18, no. 8, p. 958-975.

Sefton, C. E. M., and Howarth, S. M., 1998, Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales: Journal of Hydrology, v. 211, no. 1–4, p. 1-16.

Seibert, J., 1999, Regionalisation of parameters for a conceptual rainfall-runoff model: Agricultural and Forest Meteorology, v. 98–99, no. 0, p. 279-293.

Sellami, H., La Jeunesse, I., Benabdallah, S., Baghdadi, N., and Vanclooster, M., 2014, Uncertainty analysis in model parameters regionalization: a case study involving the SWAT model in Mediterranean catchments (Southern France): Hydrology and Earth System Sciences Discussions, p. doi: 10.5194/hess-5118-2393-2014.

Servat, E., and Dezetter, A., 1993, Rainfall-runoff modelling and water resources assessment in northwestern Ivory Coast. Tentative extension to ungauged catchments: Journal of Hydrology, v. 148, no. 1–4, p. 231-248.

SMHI, 2008, Integrated Hydrological Modelling System - Manual Version 6.0,

Swedish Meteorological and Hydrological Institute.

Tasker, G. D., and Stedinger, J. R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology, v. 111, no. 1–4, p. 361-375.

Tolson, B. A., and Shoemaker, C. A., 2007, Dynamically dimensioned search algorithm for computationally efficient watershed model calibration: Water Resources Research, v. 43, no. 1, p. W01413.

Velasco, M., Versini, P. A., Cabello, A., and Barrera-Escoda, A., 2013, Assessment of flash floods taking into account climate change scenarios in the Llobregat River basin: Natural Hazards and Earth System Sciences, v. 13, no. 12, p. 3145-3156.

Wallner, M., Haberlandt, U., and Dietrich, J., 2013, A one-step similarity approach for the regionalization of hydrological model parameters based on Self-Organizing Maps: Journal of Hydrology, v. 494, no. 0, p. 59-71.

Wigington, P. J., Leibowitz, S. G., Comeleo, R. L., and Ebersole, J. L., 2013, Oregon Hydrologic Landscapes: A Classification Framework: Journal of the American Water Resources Association, v. 49, no. 1, p. 163-182.

Young, C. B., McEnroe, B. M., and Rome, A. C., 2009, Empirical Determination of Rational Method Runoff Coefficients: Journal of Hydrologic Engineering, v. 14, no. 12, p. 1283-1289.

Table 1. Definition for the symbols and abbreviations

	Abbreviation	Descriptions	Units
Optimized HBV model parameters	tt	threshold temperature for snowmelt	[°C]
	fc	the maximum soil moisture storage	[mm]
	β	a shape coefficient	[-]
	hl	a threshold value of water content in the upper reservoir	[mm]
	K0	the storage coefficient of the surface runoff	[d]
	Perc	storage coefficient of the percolation	[d]
Catchment descriptors related with model parameter	Melv	Mean elevation	[m]
	aspR	Main aspect	[-]
	FC	Field capacity	[vol.%]
	TPV	Total Pore Volume	[vol.%]
	Bslop	Basin slope	[‰]
	LUF	Ratio of Forest	[%]
	Lst_fp	Longest flow path	[km]
	Kf	Conductivity	[mm h^{-1}]
	BFR	Mean recession constant	[d]
Other catchment descriptors	Area	Catchment areas	[km^2]
	Elv_ds	Minimum elevation	[m]
	P	annual precipitation	[mm/year]

Table 2. Relationship between model parameters and their corresponding catchment descriptors for the linear transfer function

Model parameter	Combination of catchment descriptors		
tt	Melv	aspR	
fc	FC	TPV	
β	Bslop	FC	LUF
hl	Bslop	FC	LUF
K0	Bslop	Lst_fp	Kf
Perc	Bslop	Kf	BFR

Table 3. Goodness-of-fit measures

Evaluation criterion	Calculation	Value of a 'perfect' fit
NSC	$1-\left[\frac{\sum\left(HQ_{obs}-HQ_{sim}\right)^2}{\sum\left(HQ_{obs}-\overline{HQ}_{obs}\right)^2}\right]$	1
RMSE	$\sqrt{\sum\left(\frac{HQ_{sim}-HQ_{obs}}{HQ_{obs}}\right)^2}$	0
Bias	$\overline{\sum\left(\frac{HQ_{sim}-HQ_{obs}}{HQ_{obs}}\right)}$	0
Bias-1	$\frac{HQ_{sim}-HQ_{obs}}{HQ_{obs}}$	0

(HQ_{sim} and HQ_{obs} are the observed and simulated flows)

Table 4. Catchment descriptors of the 18 catchments

Clusters	Name (-)	Area (Km^2)	Elv_ds (m)	Melv (m)	lst_fp (km)	Bslop (‰)	aspR (-)	LUF (%)	FC (vol.%)	Kf (mm $h^{-1)}$	TPV vol.%	BFR (d)	P (mm/year)
Group 1	Br	285	41	58.9	41.8	76.5	0.43	51.9	12.3	282.1	35.7	102	872.3
	Lh	100	24.8	54.9	26	53.6	0.52	41.8	13.9	239	32.6	102	854.3
	BP	116	67.5	90.6	25.7	95.4	0.85	6.3	20.8	62.8	33.9	37	733.9
	NP	334	55.5	65	40.6	77.1	0.67	34.4	14	248.7	43.4	56	737.2
Group 2	BS	127	96.8	156.9	19.7	199.1	0.65	27.4	18.3	82.5	24.3	63	840.2
	Gr	125	129.4	208.4	23.9	269.3	0.56	43.2	15.9	92.6	29.5	56.4	1004.7
	Ha	104	74.8	93.4	25.2	152.6	0.42	38	19.2	66.5	37.7	42.1	838.6
	Ku	61.8	130.2	219.9	13.5	238	0.38	27.7	14.3	110	37.8	37.5	910.9
	Ma	45	196.2	275.6	12	312.6	0.56	31.5	12.3	126	19.9	34	835.9
	Mt	242	36.6	61.8	36.7	53.8	0.63	16.7	14.4	206.3	36.7	64.6	749
	Ol	149	128.6	284.6	25.3	328	0.63	66.4	17.4	130.9	30.5	56	1004.2
	Pi	44.5	339.6	586.1	17.1	634.7	0.32	99.1	7.8	166.9	20.4	35	1537.9
	VR	57.5	133.1	467	22.2	435.9	0.79	65.2	14	145.3	20.1	35	1136.4
	Me	136	81.9	395.5	27.9	199.9	0.75	35.3	17.4	84.3	36.3	38	889.2
Group 3	De	309	90.9	258	49.1	208.9	0.56	38.1	17.3	84.4	29.5	56.4	899.6
	RH	184	154.7	206.1	24.9	254.4	0.56	20.8	17	150.8	26.5	100	781.7
	Re	321	182.9	273.7	43.4	341.5	0.49	28.6	16	127.1	27.3	100	793.7
	Go	633	141.5	315.8	62.9	299.3	0.49	26	16.6	117.4	27.3	100	791.6

Table 5. Weightings of the variables and summary of characteristics on the three principal components

Original variables	Principal components		
	PC1	PC2	PC3
Area	0.195	0.399	-0.429
Elv_ds	-0.377	0.072	-0.203
Melv	-0.368	-0.004	-0.283
lst_fp	0.205	0.389	-0.387
Bslop	-0.399	0.042	-0.249
aspR	0.136	-0.364	-0.122
LUF	-0.342	0.176	0.179
FC	0.258	-0.344	-0.347
Kf	0.056	0.441	0.451
TPV	0.298	0.004	0.295
BFR	0.181	0.449	-0.125
P	-0.396	0.060	0.100
Eigenvalues	5.432	2.506	1.738
Variance (%)	45.269	20.890	14.480
Cumulative Variance (%)	45.269	66.159	80.639

Table 6. Results of paired t-test on the p values for both regionalization schemes.

		Std.error mean	95% confidence interval		t	df	p-value
			Lower	upper			
All	Winter	0.034	-0.054	0.12	0.823	17	0.422
	Summer	0.067	0.021	0.115	3.065	17	**0.007***
Group 1	Winter	0.172	-0.182	0.525	1.542	3	0.221
	Summer	0.086	0.01	0.161	3.608	3	**0.036***
Group 2	Winter	-0.113	-0.113	0.091	-0.24	9	0.816
	Summer	0.079	-0.008	0.166	2.062	9	0.069
Group 3	Winter	0.037	0.007	0.067	3.921	3	**0.029***
	Summer	0.022	0.005	0.038	4.303	3	**0.023***

Table 7. Results of paired t-test on the flow duration curve (FDC) using the Nash-Sutcliffe Coefficient (N_{SC}) and bias for both regionalization schemes.

		Std.error mean	95% confidence interval		t	df	p-value
			Lower	upper			
All	N_{SC}	0.007	-0.041	0.056	0.312	17	0.758
	bias	0.04	-0.043	0.123	1.008	17	0.327
Group 1	N_{SC}	-0.04	-0.119	0.039	-1.61	3	0.205
	bias	0.012	-0.145	0.17	0.252	3	0.817
Group 2	N_{SC}	0.01	-0.071	0.091	0.278	9	0.787
	bias	0.05	-0.098	0.198	0.762	9	0.465
Group 3	N_{SC}	0.047	-0.096	0.191	1.051	3	0.37
	bias	0.042	-0.182	0.267	0.6	3	0.59

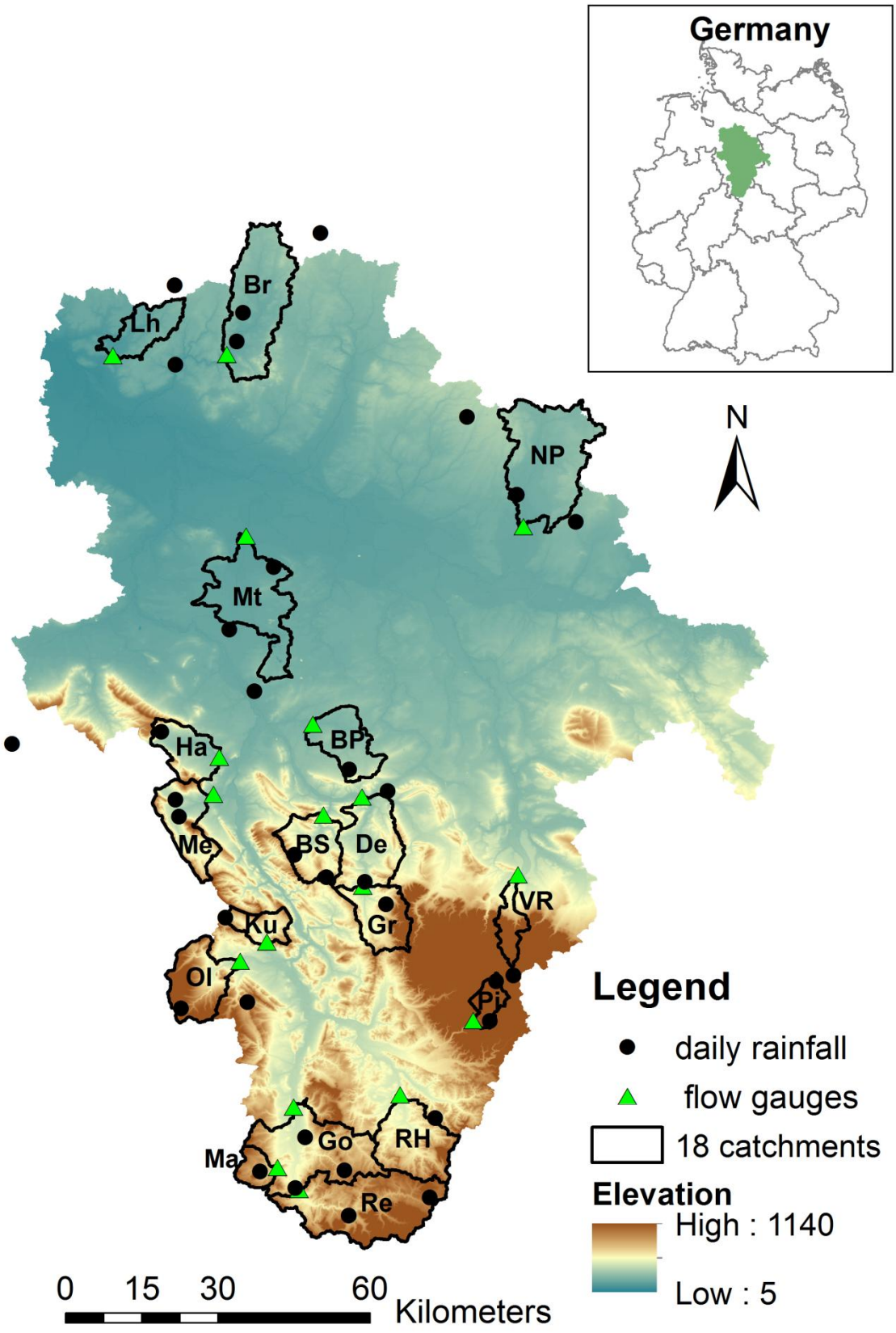

Figure 1. The locations of 18 catchments of Aller-Leine in northern Germany

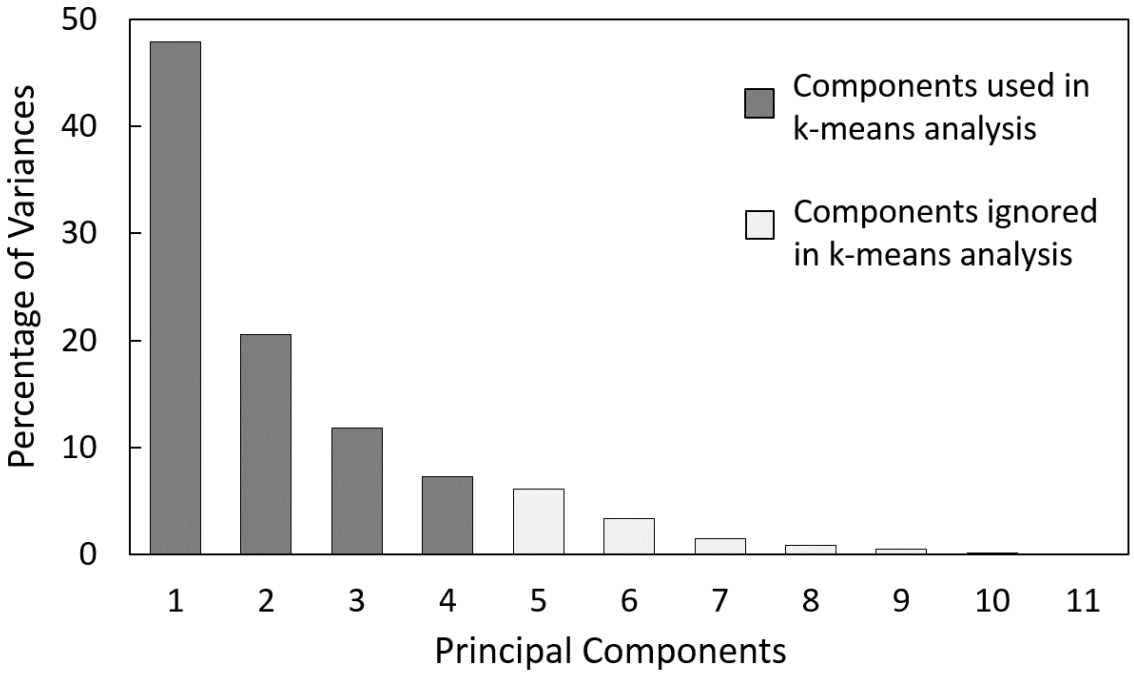

Figure 2. Percentage of variance explained as a function of the number of principal components

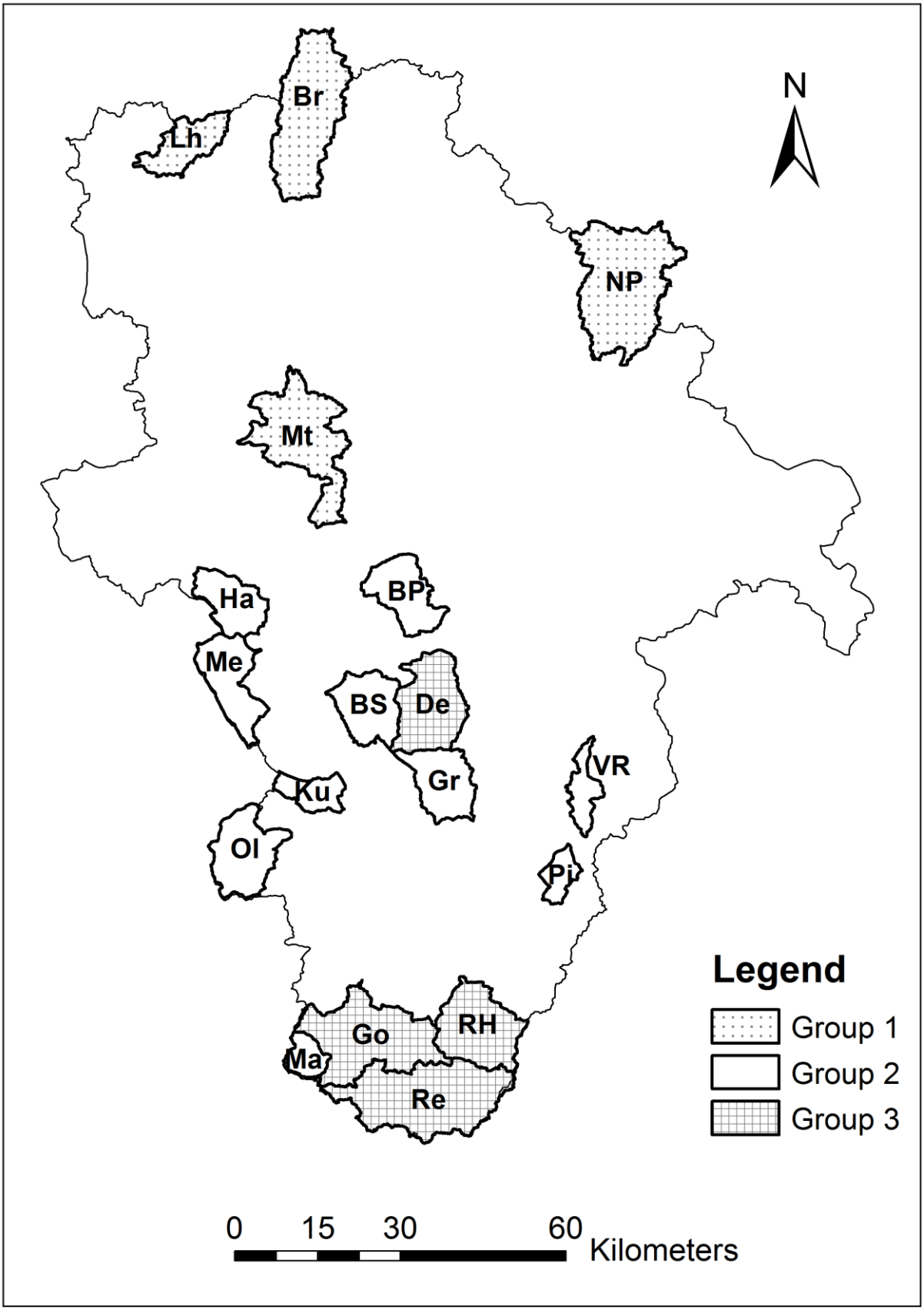

Figure 3. Spatial distribution of the three groups in Aller-Leine, Germany

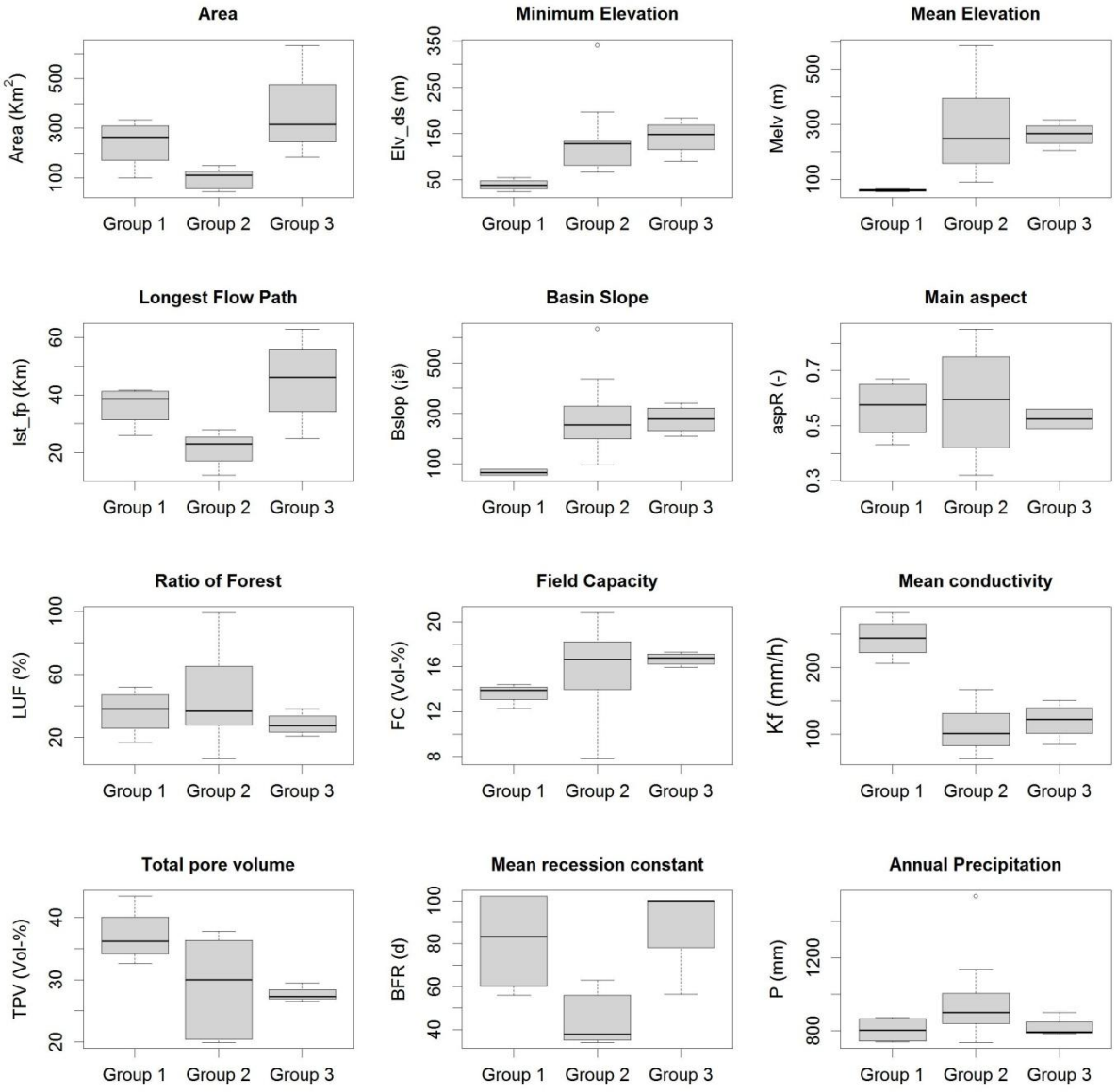

Figure 4. Box plots of the catchment characteristics of the three groups. Thick black line is the median value. The box shows the inter-quantile range between 25th and 75th quantiles of the data. The ends of the whiskers represent 1.5 times the inter-quantile range.

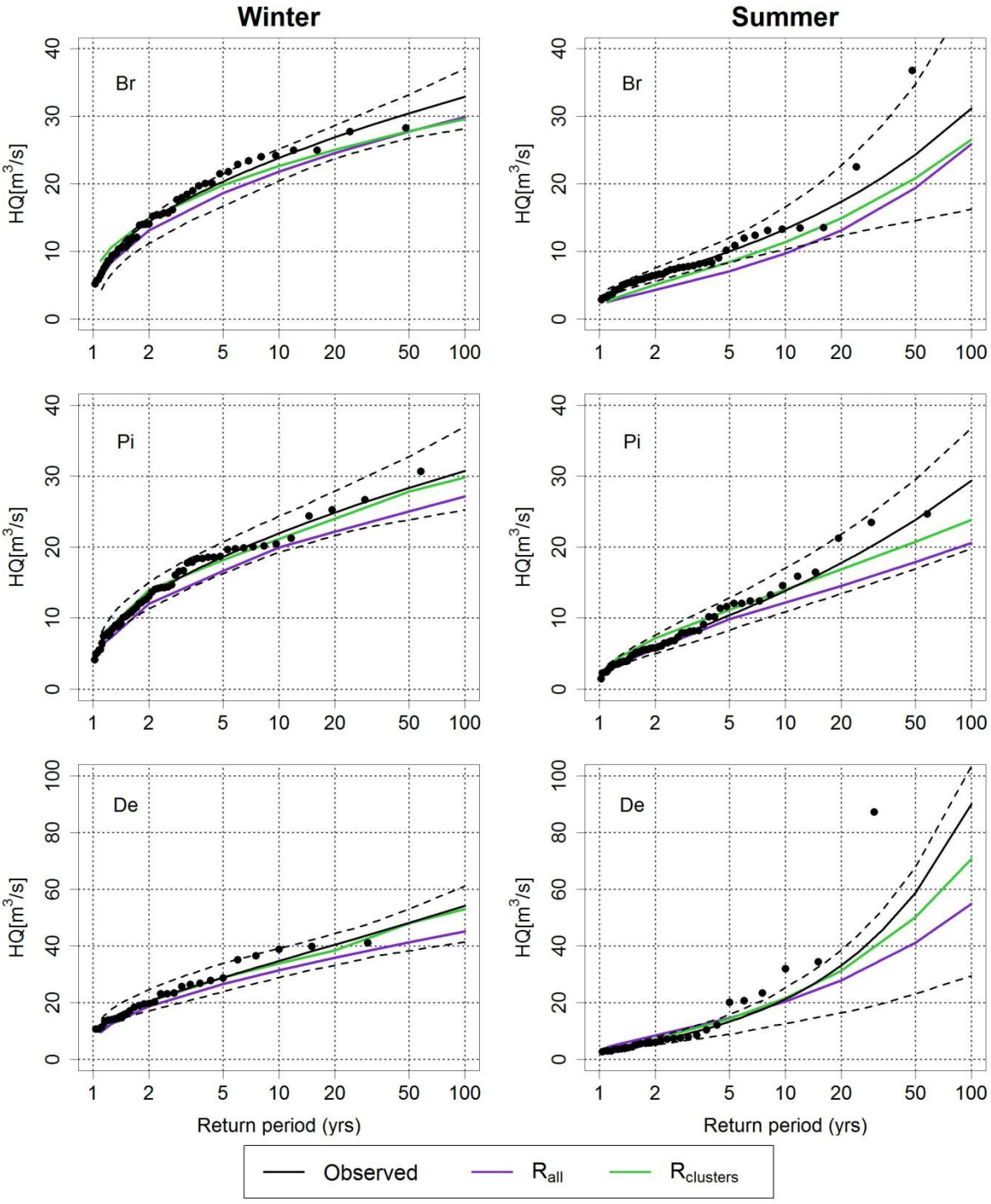

Figure 5. CDFs of observed and simulated daily extremes in winter and summer for the three randomly selected sample catchments (Br, Pi & De); the black dots are the observed annual daily extremes; black dashed lines enclose the 90% confidence interval against observed peak flows

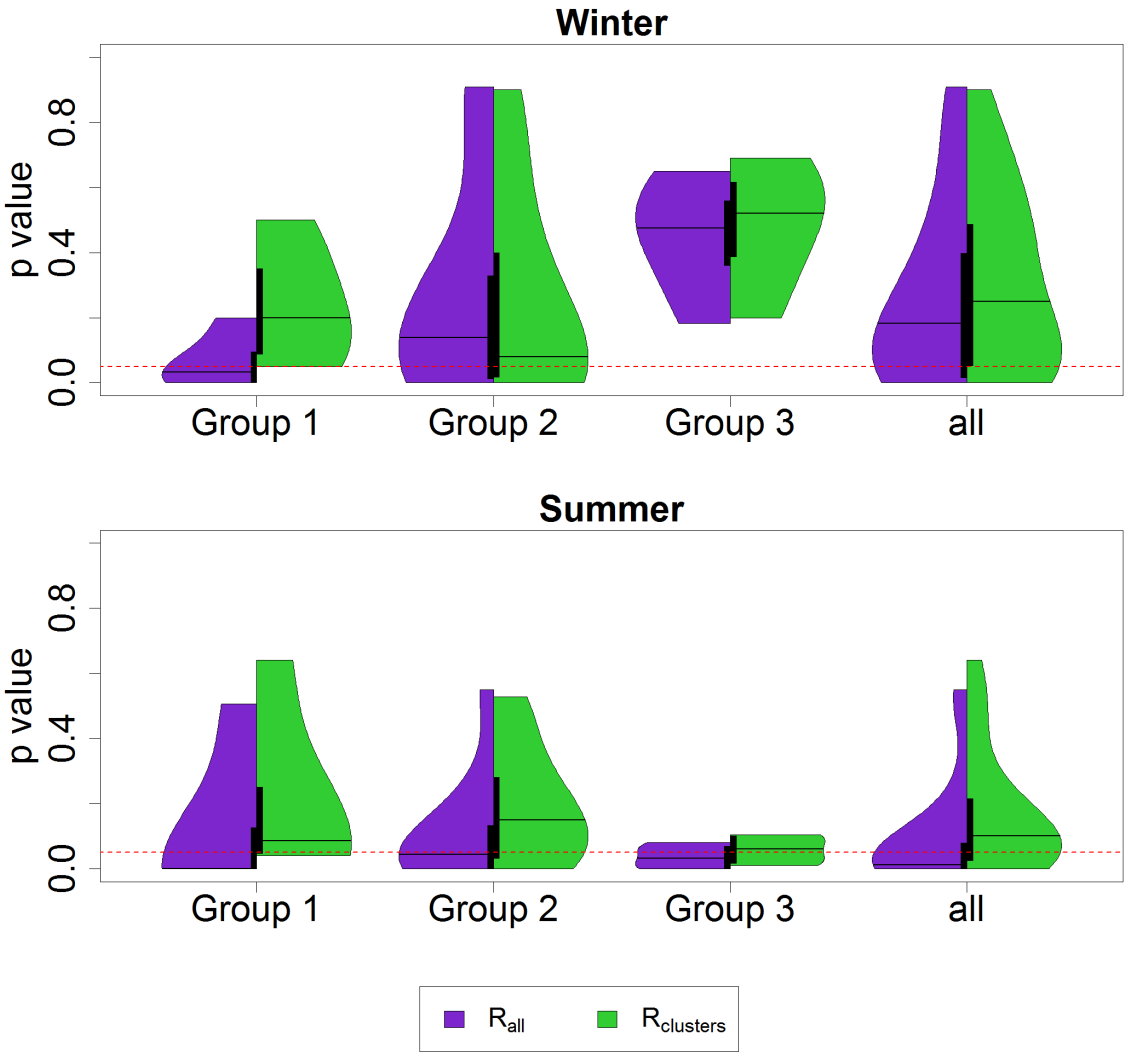

Figure 6. Violin plots of the p value over all catchments for the fitted GEV distribution between observed and simulated daily extremes in winter and summer respectively. The middle black solid line is the median value and the red dashed line is the 5% significance line; the middle black rectangle enclose the 25% and 75% quantile values. The left purple half is R_{all} approach and the right green half is $R_{clusters}$ approach. The width of each violin plot represents the probability distribution

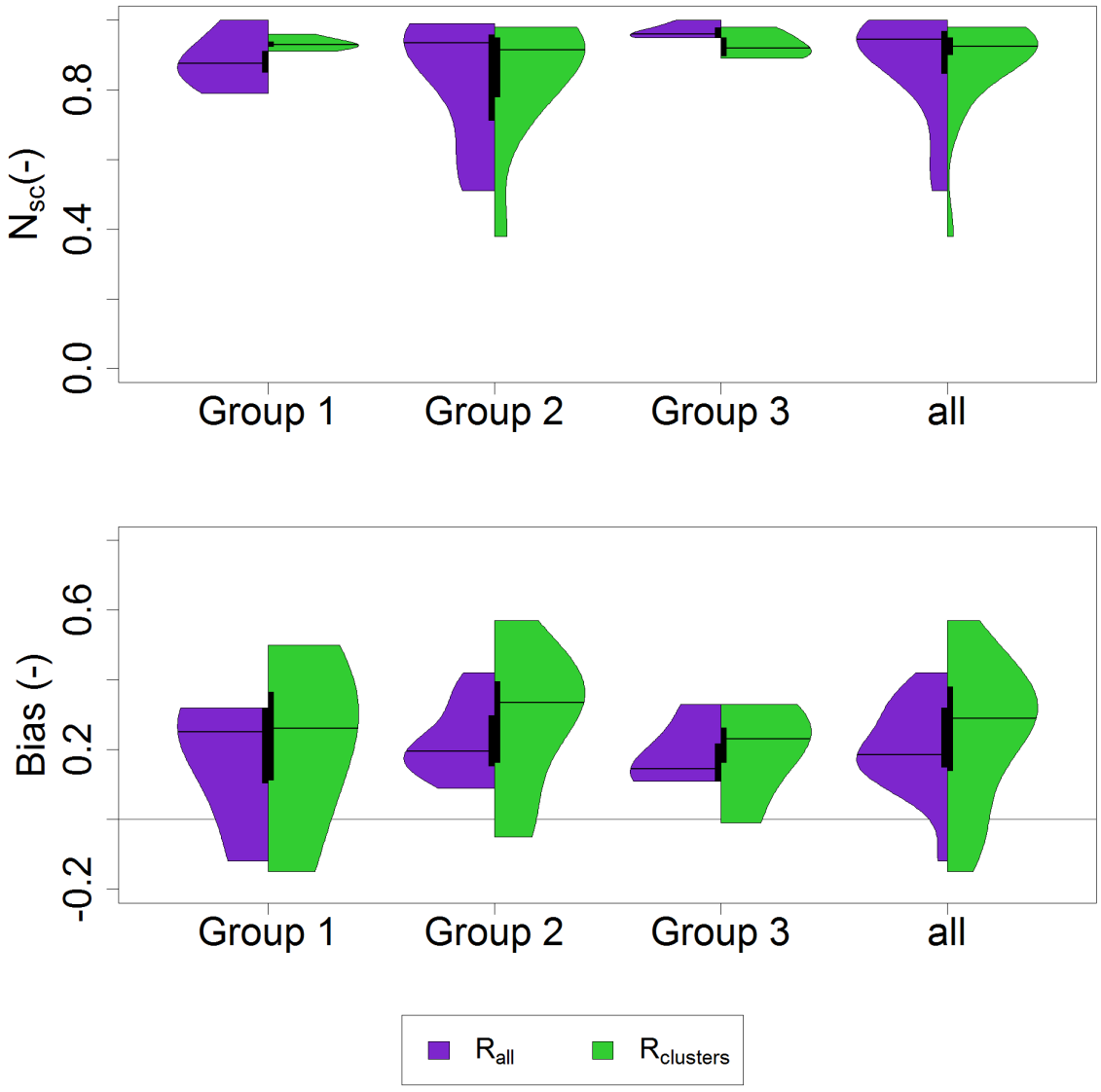

Figure 7. Regional Calibration results of flow duration curve (FDC) using the Nash-Sutcliffe Coefficient (N_{SC}) and the Bias for two different regionalization schemes. The middle black solid line is the median value; the middle black rectangle enclose the 25% and 75% quantile values. The left purple half is R_{all} approach and the right green half is $R_{clusters}$ approach. The width of each violin plot represents the probability distribution

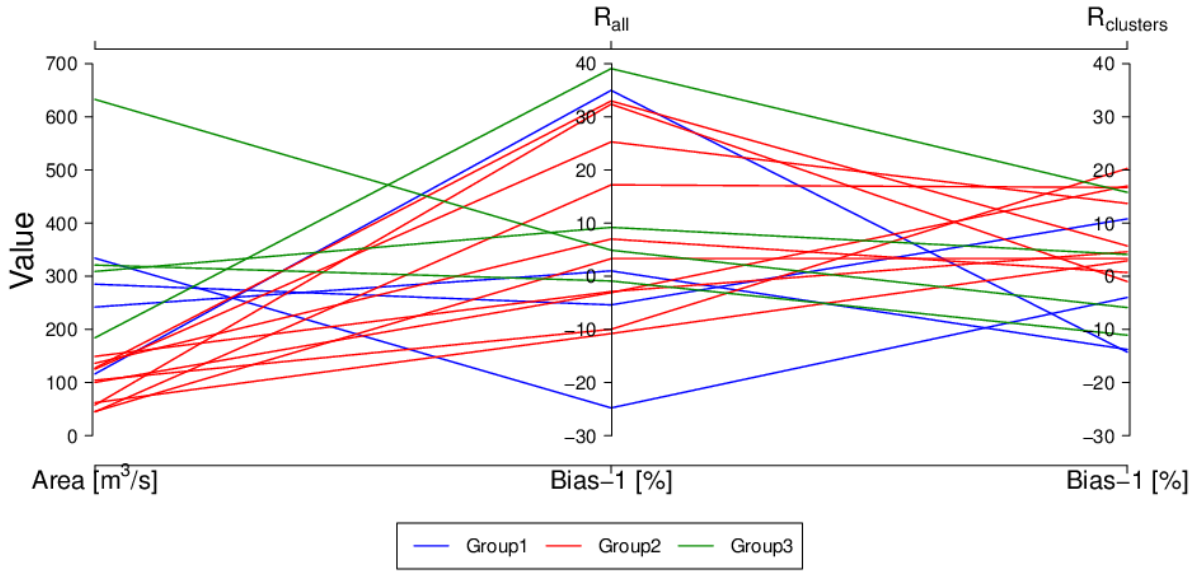

Figure 8. The values of Bias-1 for daily flow duration curve (FDC) at non-exceedance probability P_{non}= 0.975

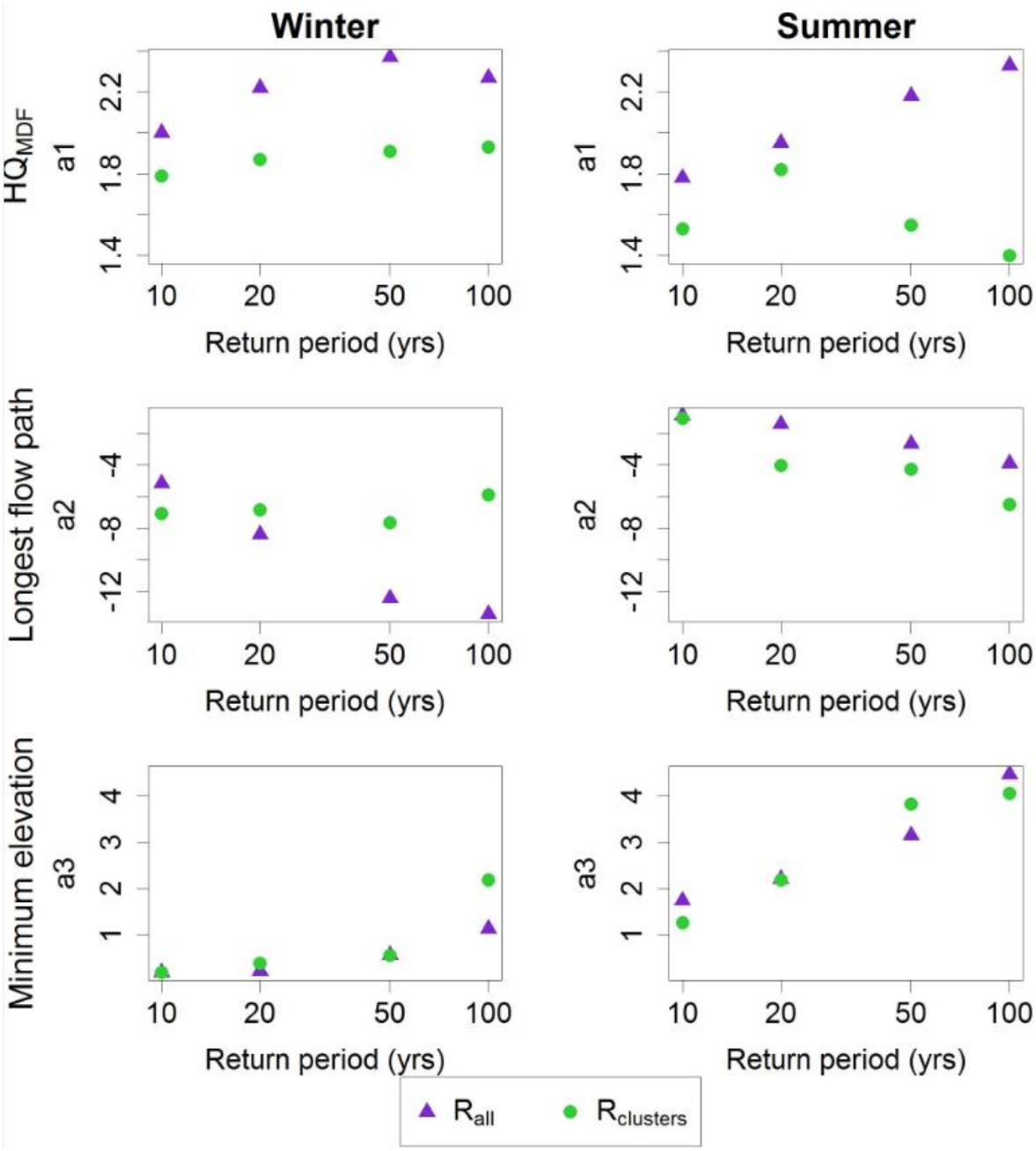

Figure 9. The values of multiple regression coefficients for the two regional calibration schemes

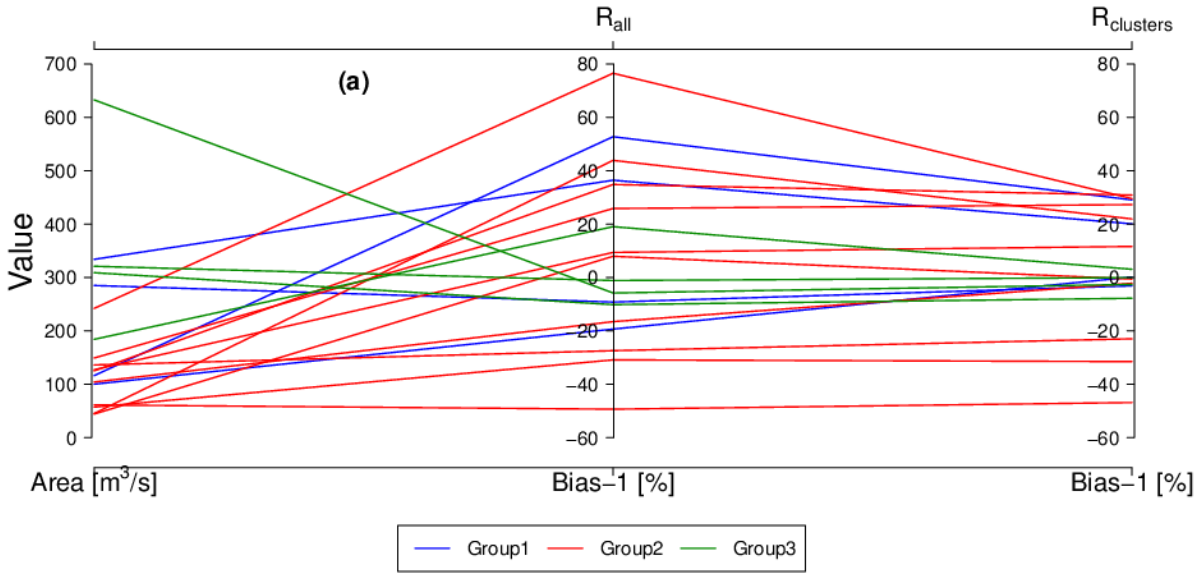

Figure 10(a). Comparison between the two regionalization calibration schemes at a return period of 100-year in terms of Bias-1 for winter season for all 18 catchments

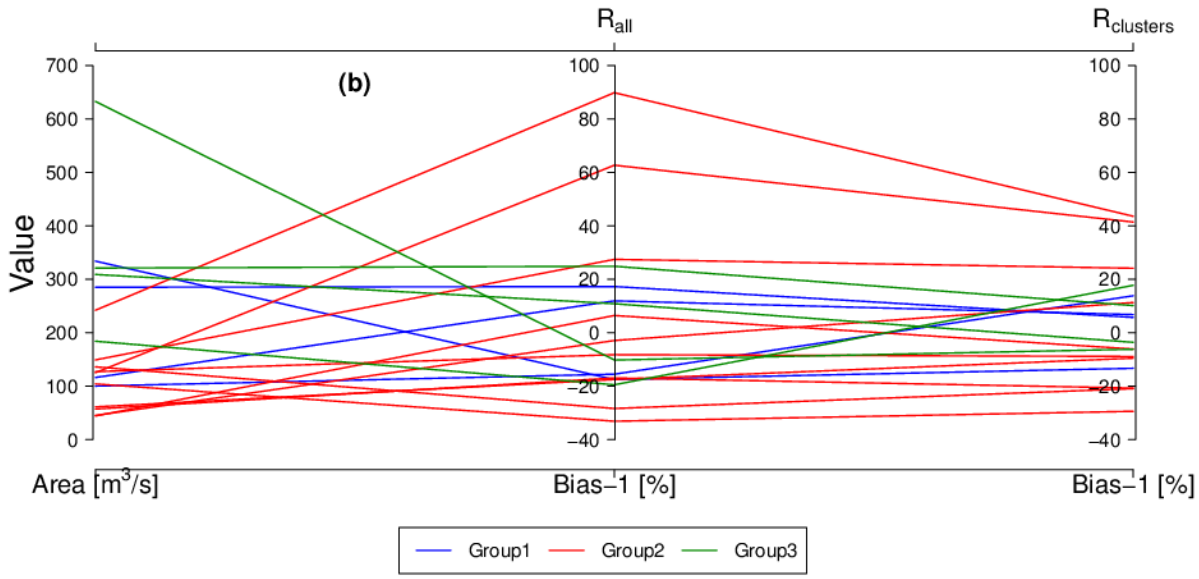

Figure 10(b). Comparison between the two regionalization calibration schemes at a return period of 100-year in terms of Bias-1 for summer season for all 18 catchments

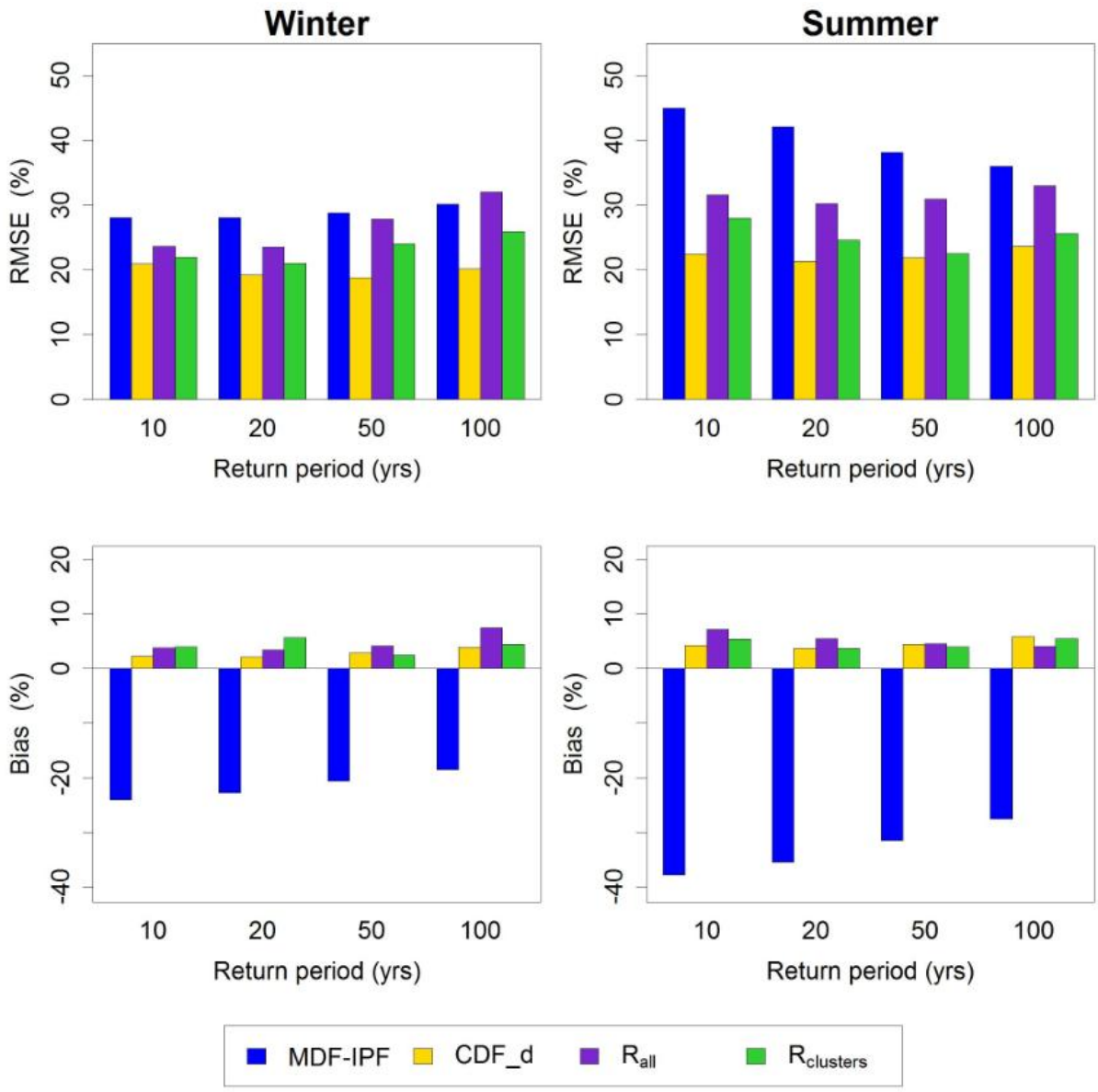

Figure 11. Comparison of root mean square error (RMSE) and Bias using regionalized parameter sets for clusters ($R_{clusters}$) and the whole study area (R_{all}) in both winter and summer season

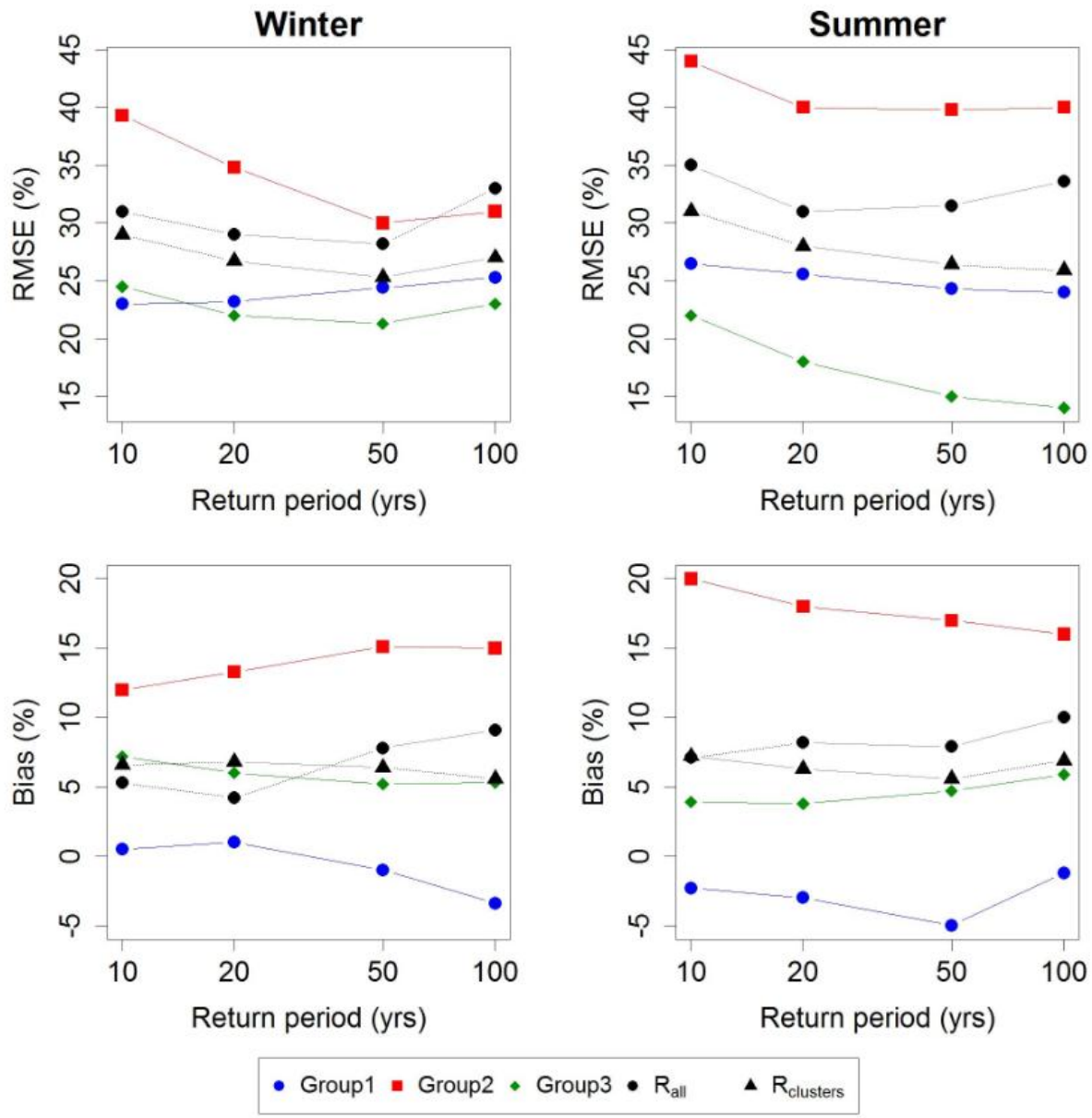

Figure 12. RMSE and bias results by cross-validating both R_{all} and $R_{clusters}$ schemes for the three groups.