Authors' response to Referee #3

For clarity, authors' responses are presented by blue colour.

We have answered all the comments of the reviewer 3. Answers are attached to this revision note. Along with the answers we are also explaining all the changes we have done.

The goal of this study is to present and evaluate a bias correction of the ECMWF ERA-20c reanalysis for South Korea. The authors apply a combination of transfer functions and wet frequency adjustment methods to correct the bias present in the precipitation time series. Parameters of the obtained transfer functions derived from the relation between the reanalysis grid and observed rain gauge precipitation are interpolated in space to full grid precipitation data. Overall evaluation: This is a potentially interesting paper, but in order to be published a major revision is required. The results presented in this paper are relatively simple and lack of deep analysis. The authors provide a long text on motivation for bias correction but omit the discussion of the bias correction in context of downscaling and do not discuss the constrains and limitations of the parameter interpolation. The authors claim in the title reduction of uncertainty but do not prove that this is the case.

1. Overall I am left unclear on the core contribution of the paper. The evaluation of the ERA precipitation over South Korea is a valuable contribution but it is very short. The applied bias correction is described in detail but a justification is missing. Finally, the spatial interpolation is not correctly validated.

(**Response**) Thank you for the constructive comments. After the preliminary evaluation of the ERA-20c daily precipitation over South Korea, this study mainly focused on the bias correction of ERA-20c daily precipitation, especially for extreme values, because the century-long precipitation dataset could contribute to the reduction of the uncertainty in hydrologic frequency analysis where a limited number of observations were generally given. As indicated, the bias correction is generally involved with downscaling of general circulation models (GCMs). More specifically, the spatial resolutions of GCMs are too coarse to adequately represent regional climate variability so that the direct use of those model outputs is not appropriate, especially for fine-scale hydrological applications. Moreover, most model outputs in climate models are affected by spatio-temporal biases, leading to significant bias in hydrological impact studies. In these contexts, both bias correction and spatial downscaling of model outputs are crucially involved in the use of GCMs for hydrological impact studies. In our case, spatial resolution of ERA-20c (i.e. 0.125°×0.125°) is relatively high enough to be used in practical applications. Therefore, the spatial downscaling has not been considered in the current study and we rather much focused on certain aspects of the bias correction, which might be of importance for a special use of the long-term reanalysis data in hydrologic frequency analysis (Coles et al., 2003; Huard et al., 2010; Overeem et al., 2008; Tung and Wong, 2014; Van de Vyver, 2015). In order to validate the use of long-term reanalysis data for the reduction of uncertainty in estimating design rainfalls, we further explored the uncertainty range of design rainfalls based on GEV distribution for a given return period and a given data length within a Bayesian modelling framework. As illustrated in Figure A1, the uncertainty range of design rainfall is significantly reduced with increasing data. In this perspective, we applied the suggested QM approach in this study.

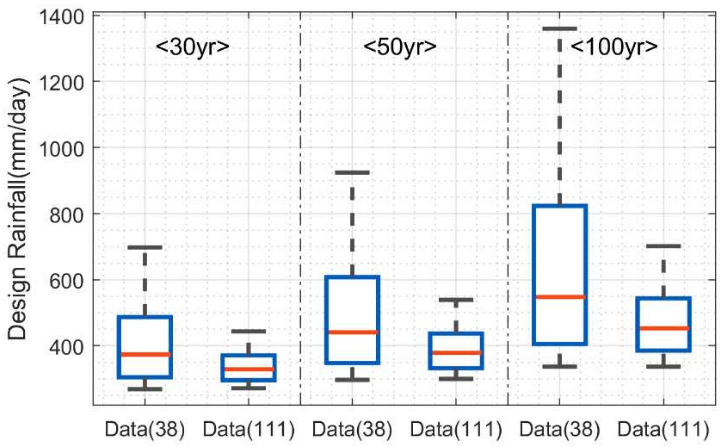
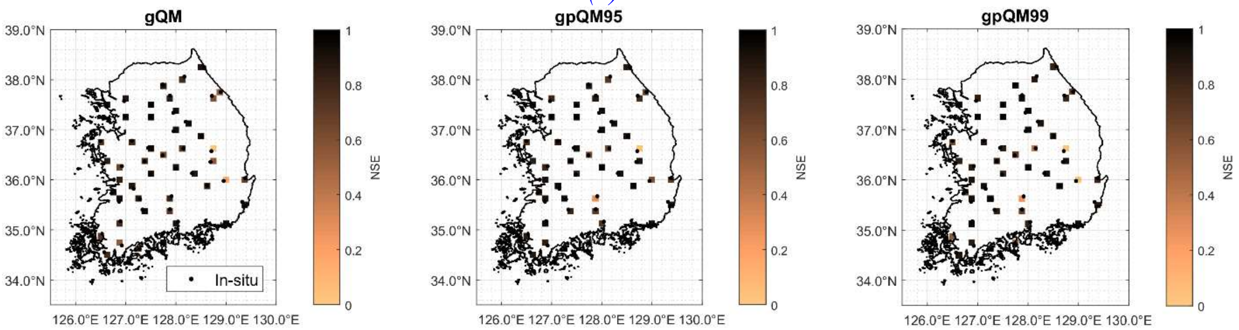


Figure A1. Boxplot for the uncertainty range of design rainfalls with 30-yr, 50-yr and 100-yr return period based on GEV distribution for 38 annual maximum series (AMS) (Data(38)) and 111 AMS data (Data(111)).

For spatial interpolation of a set of parameters associated with transfer functions in quantile mapping approach, this study further evaluated the IM-PCM method by employing a leave-one-out cross validation framework over 48 weather stations for the reference period (1973-2010) and the overall performance has been illustrated in the manuscript for both the extreme and mean. For a more specific analysis in each weather station in the context of cross validation, we generated a map showing the spatial errors in both annual maximum series (AMS) rainfalls and mean. The AMS errors were evaluated by root-mean-square-error (RMSE) and Nash-Sutcliffe efficiency (NSE) in Figure A2. For the mean, we additionally evaluated the IM-PCM method by estimating the relative error between the observed and modelled in Figure A3. As shown in the figures, for the AMS rainfalls, gpQM95 and gpQM99 generally perform well except for a few stations. Most stations showed NSE over 0.8 and RMSE less than 30mm. For the mean daily rainfall, the relative errors are generally below 10%.

(a)



(b)

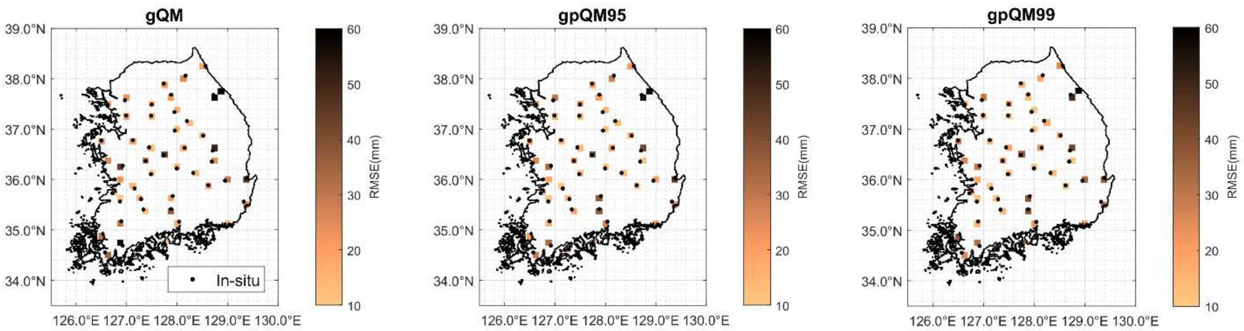


Figure A2. Cross validation results of the IM-PCM for the annual maximum series rainfall of the bias corrected data by QM approaches (gQM, gpQM95 and gpQM99) over 48 grid points. (a) Nash-Sutcliffe efficiency (NSE) and (b) root-mean-square-error (RMSE).

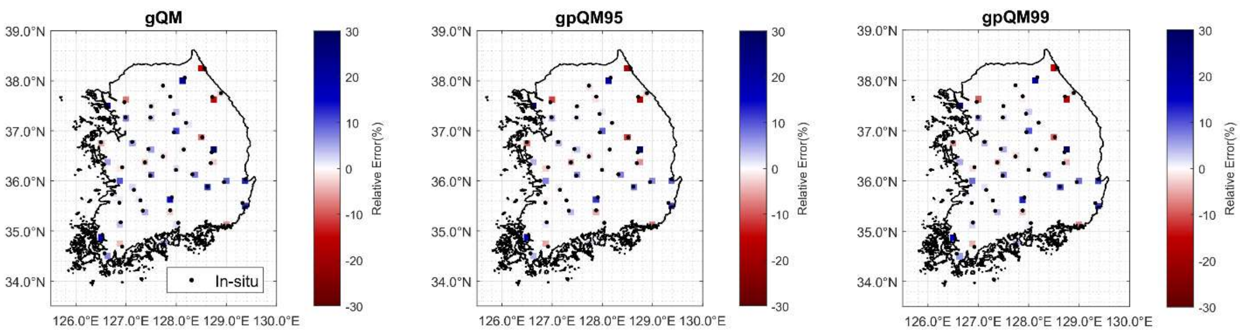


Figure A3.Relative error of the bias-corrected mean rainfall by QM approaches (gQM, gpQM95 and gpQM99) in 48 grid points compared with the corresponding in-situs.

For title, we agree that our study rather focused on the uncertainty range for three different periods. Thus, we have changed the title of this manuscript to "Exploring the Long-Term Reanalysis of Precipitation and <u>its Bias Correction using a Composite Gamma-Pareto Distribution Approach over South Korea</u>" upon your comments. We will further explore the uncertainty reduction in using the long-term reanalysis data in the next study.

2. Reviewers #1 and #2 provide excellent recommendations and there is no need to repeat them here. Along the lines outlined there the manuscript can be improved.

(**Response**) Thank you for the comments. We have revised the manuscript along with the answers to the comments.

The authors may consider to rename Section 2 to “Material and methods” and to describe in addition to sections 2.1 and 2.2 in two new sections 2.3. and 2.4 the BC and the downscaling issues, possibly with some text from the introduction in which the scientific goals of the study should be clearly identified. Based on the findings and constrains discussed in this section the applied methodology can be justified and presented in detail in section 3. “Applied methodology”. The validation procedure should include an analysis and discussion of the differences between the calculated and observed values at each station when this station is not included into the derivation of the interpolated parameters. This will help to access the possible errors at ungauged grid cells and thus help to judge the entire applied procedure and draw correct conclusions.

(**Response**) Thanks for the constructive comments. The literature review on bias correction methods were described in the introduction Section and the data and study area were illustrated in Section 2. As suggested, we introduced main methodologies used in this study in Section 3, and the proposed bias correction approaches were then applied to daily ERA-20c data for the reference period in Section 3. Therefore, we believe that the current structure seems to be appropriate. However, we agree that the scientific goal of this study should be clearly identified. Thus, we have modified some parts of the introduction in the revised paper as follows:

> “(p. 3|.19) However, although substantial improvements have been made in the modelling process, previous studies have shown that reanalysis datasets still have their own systematic errors which vary in space and time (Bao and Zhang, 2013; Bosilovich et al., 2008; Gao et al., 2016; Kim and Han, 2018; Ma et al., 2009) It is also clear that century-long reanalysis data may misrepresent long-term climatic trends or synoptic scale variability, especially for the first half of twentieth century, and there exists the difference in temporal variability between century-long reanalyses (Befort et al., 2016; Brands et al., 2012; Donat et al., 2016; Krueger et al., 2013; Poli et al., 2013). Nevertheless, if one collects the reliably extended time series for daily precipitation in a certain area, the

uncertainty of the estimated design rainfalls could be affected by uncertainty associated with the sampling error due to the lack of data (time series) (Coles et al., 2003; Huard et al., 2010; Overeem et al., 2008; Tung and Wong, 2014; Van de Vyver, 2015). We further explored the uncertainty range of design rainfalls based on GEV distribution for a given return period and a given data length within a Bayesian modelling framework. As illustrated in Figure A1, the uncertainty range of design rainfall is significantly reduced with increasing data. However, there are limited studies on bias correction for long-term daily reanalysis precipitation data in hydrologic applications. Most of the existing studies have been performed mainly within the context of comparison across different reanalysis data, but not bias correction technique issues (Donat et al., 2016; Poli et al., 2016). Thus, in order to better understand the biases and their roles in hydrologic applications, this study focuses on exploring bias correction methods, especially for extreme value associated with the sampling error in rainfall frequency analysis, in a certain area with spatio-temporally sparse observation network."

"(p. 6|. 1) Thus, a primary question in the statistical bias correction analysis is whether the QM method can reliably improve ERA-20c daily precipitation, especially for extreme value, over 100 years when including the ungauged sites."

For Section 3, in the context of leave-one-out cross validation, we generated a map showing the spatial errors in each station in both annual maximum series (AMS) rainfalls and mean as illustrated in Figures A2 and A3. For the AMS rainfalls, gpQM95 and gpQM99 generally perform well except for a few stations. Most stations showed NSE over 0.8 and RMSE less than 30mm. For the mean daily rainfall, the relative errors are generally below 10%. These results have been included in the revised manuscript.

3. As a minimum requirement before revision, the manuscript has to be professionally revised and edited to correct the language and to remove the unnecessary text repetitions

(**Response**) We have carefully revised the manuscript. Thanks for the constructive comments again.

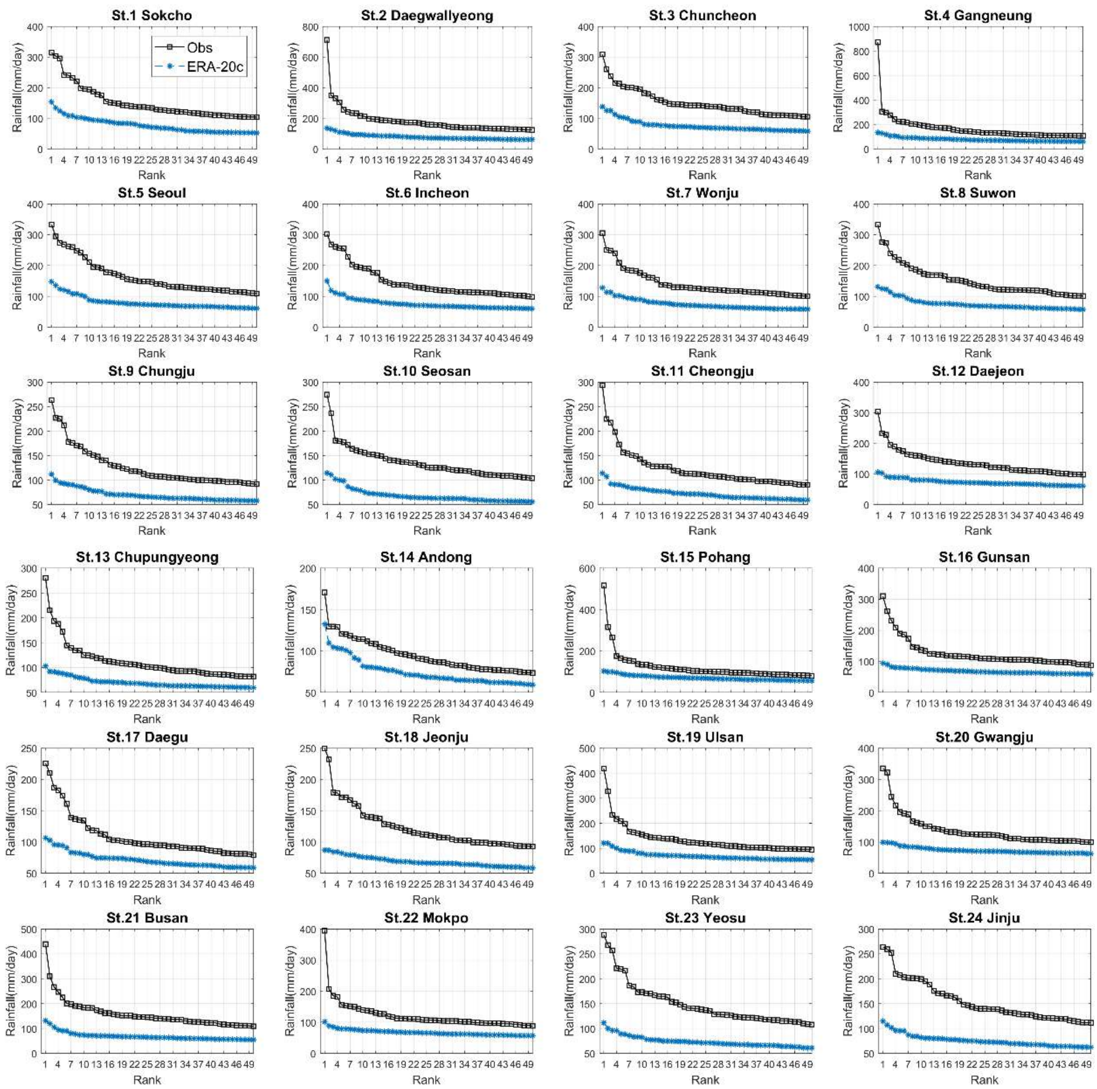
Specific:

p. 9|. 1 - What is the rationale for using stations 4, 16, 28, 40 ?

- Large deviations are also visible in spring

(**Response**) In this study, we analysed the 50 top events in 48 weather stations and the results were generally similar as shown in Figure A4. To efficiently compare the results, we representatively illustrated the comparison for using 4 stations. The extreme rainfalls over South Korea can be generally characterized from two distinct rainfall patterns such as summer monsoon rainfalls ("Changma") and tropical cyclones (Lee et al., 2010; Seo et al., 2015; Son et al., 2017). In this context, we described that the discrepancies in the 50 top extreme rainfalls were largely attributed to differences in rainfall during summer season. We have changed this sentence as follows:

"<u>The relationships between the 50 top extreme rainfalls show that the discrepancies were largely attributed to differences in rainfall during summer season, as noted in Figure 2</u> (Lee et al., 2010; Seo et al., 2015; Son et al., 2017)<u>. The overall relationships are similar to each other, as shown in supplementary material, and the comparisons in the stations 4, 16, 28 and 40 are representatively illustrated in Figure 3(b).</u>"



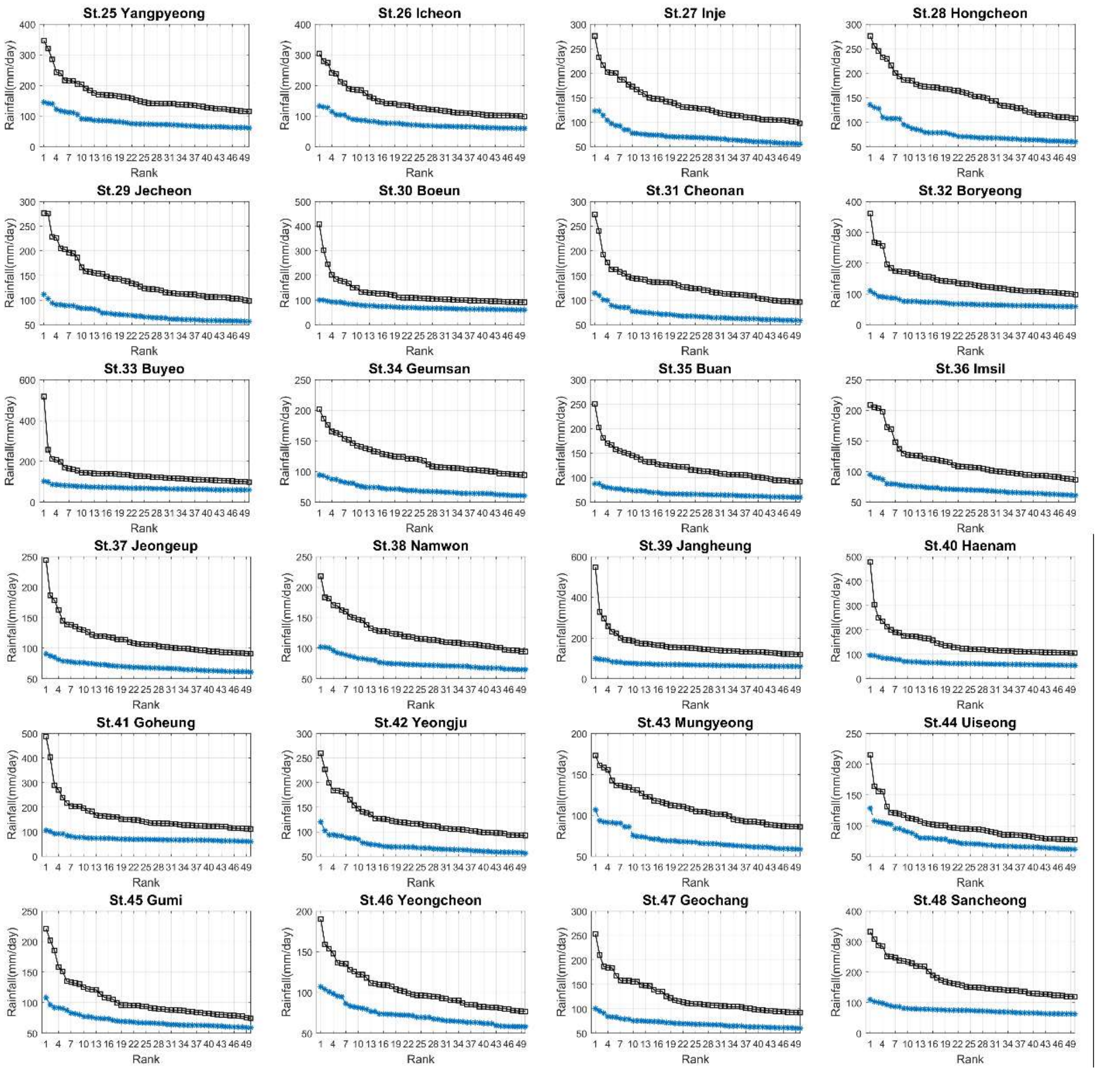


Figure A4. Comparison of the deviation corresponding to the rank in top 50 events for the baseline period (1973-2010) in 48 stations

p.9|. 3 - The bias in extreme is proportional … : I cannot see this, and even would argue that the maximum rain at station 4 is a mistake in station reading

(**Response**) As seen in Figures 3(b) and A4, the higher the rank, the more there generally exists the absolute difference between the observed and ERA-20c. In this context, we have changed this sentence as follow:

> "The biases in extreme values are <u>generally</u> proportional to the amount of rainfall, and the biases are likely to be higher in the upper tails of the distribution than that of the middle layer, as shown in Figure 3(b)."

The maximum rainfall (870.5mm/day) in Station 4. Gangneung was caused by a tropical cyclone (“Typhoon Rusa”) on 31 August 2002 and it was record rainfall in South Korea. The detail information is found in Seo et al. (2017).

p.9|. 6 -This paragraph is supposed to summarize the section 2.2, but after the summary it introduces a new investigated item: wet-day. This should be presented after line 5 on page 3. Also, I suggest to add a short description of the applied evaluation statistics after the introduction of ERA-20c.

(**Response**) In this Section, we preliminary explored two deficiencies in the ERA-20c daily precipitation over South Korea: the overestimation of the wet-day frequency and underestimation of the extreme values. As indicated in the section 2.2, the overestimation of the wet-day frequency has been a well-known issue in climate models, so that we applied a relatively simple way to adjust the wet-day frequency as a pre-processing step for the bias correction. Rather, we provided a set of relevant references for the the over-pronounced frequency of light precipitation in climate models. However, we agree that it is valuable to add quantitative results so that some evaluations have been added to this Section in the revised manuscript as follows:

“(p.9|. 7) In summary, the ERA-20c precipitation data are capable of reliably reproducing the mean values with 0.968 for NSE and 15.59mm for RMSE, while the extreme values in the 50 top records are consistently underestimated with -1.088 for NSE and 76.69mm for RMSE.”

“(p.9|. 12) On the other hand, as shown in Figure 4, ERA-20c has a much higher frequency of wet-days (>0mm/day), varying from 11.75 to 26.64 days per month, than that of observation (6.07 to 14.5 days) for all months in South Korea.”

p.9|. 9 What is the role of climate models here?

(**Response**) In this sentence, the climate models mean the ERA-20c modelling process. We have changed “the climate models” into “the ERA-20c modelling process” in the revised manuscript.

p.9. |. 12 Explain wet-day (RR > 1 mm/d ?)

(**Response**) In this rainfall frequency comparison, all wet days with the rainfall lager than 0mm/day were considered. We have changed “wet-days” to “wet-days (>0mm/day)” in this sentence in the revised manuscript.

p. 3|. 21 This paragraph is an unnecessary repetition of the summary in the last paragraph.

(**Response**) Thank you for the comments. We have removed this sentence in the revised manuscript.

p.12|. 12-20 Some explanation of AIC and BIC and discussion why DGP was chosen is needed here. I cannot understand the title and the content of Table 2.

(**Response**) We summarized the selected distributions among six distributions based on AIC and BIC values for the extremes from observed and ERA-20c daily precipitation over the 95th and 99th percentiles for all 48 stations. More specifically, the numbers in Table 2 indicate the number of stations which belong to a certain distribution. We have changed the sentence to better explain the results as follows:

> “(p.12|. 16) … . To ensure the suitability of the GPD, we first evaluated six different distributions, GPD, GEV, GUM, WEI, LOGN and gamma, for the extremes in both the observed and ERA-20c over the 95th and 99th percentiles using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). <u>The model with the lowest AIC and BIC is preferred, as the best-fit distribution.</u> For a given threshold, the GPD was <u>mostly selected</u> as the best-fit distribution for the extremes as shown in Table 2. <u>The numbers in Table 2 indicate the number of stations which belong to a certain distribution.</u>

p. 13|. 5 “Again, : : :” repeats line 3

(**Response**) We have deleted this sentence in the revised manuscript.

p. 15|. 5 ” … the suitability …” for what? This goal of the study has not been mentioned in the introduction.

(**Response**) We have removed the sentence in the revised manuscript.

p. 15|. 19 "… leave-one-out procedure… " The procedure definitely needs a longer explanation and discussion. Usually one period is used for training and an another for validation.

(**Response**) In our study, the cross-validation scheme has been applied in the context of spatial interpolation. As illustrated in the manuscript, the cross-validation scheme first estimates a set of parameters for the observation of daily precipitation for 47 stations out of 48 stations, and the estimated parameters were further used to build contour maps as shown in Figure 6. The set of parameters of the grid point corresponding to the excluded station were taken from the maps, and the proposed bias correction approaches were then applied.

p. 16|. 6 Where is section 3.4.1 ?. In my opinion the section "Evaluation criteria "should be in section 2. Material and methods

(**Response**) Thanks for the suggestion. Section 3.4.1 was the typo. We have changed "as described in Section 3.4.1" to "as described above". Regarding the evaluation criteria, however, we thought that the current structure is relevant.

p. 17|. 1 As illustrated in the previous section… The range 0.-4.66 is not mentioned in the previous section.

(**Response**) We have added the range in the revised manuscript as follows:
"(p.10|. 19) (TH1) 0>mm/day, (TH2) 0.1>mm/day, (TH3) 1>mm/day, and (TH4), the frequency of wet days was set to the observed value, which varied from 0 to 4.66."

p. 17|. 5 What is "the degree of bias" ?

(**Response**) It means the amount of bias in this sentence. It seems that the term "degree" is redundant so that we have removed.

p. 17|. 6 "… significantly varied…" add some numbers here to quantify this variation

(**Response**) We have added the range of variation in the revised manuscript.

"We also found that the bias associated with the cut-off thresholds significantly varied within a specific season, especially in the summer. The biases for both TH1 and TH2 range from 2.21 to 10.49 and from 1.92 to 10.09 during the summer, respectively, while TH3 and TH4 varied from 0.16 to 6.27 and from -1.06 to 2.97, respectively."

p.17|. 21. "This study introduces ..." rewrite to This study applies

(**Response**) We have changed as suggested.

p. 19|. 10-13 " In other words.. " This is trivial. If there is no difference between model and observation then there is no need for a bias correction

(**Response**) This part is a specific explanation why the bias still remains after bias-correction.

p. 21|. 16 "The bias correction : : : improved the quality : : :" Perhaps the mean over the region. What can be said for ungauged regions?

(**Response**) This study evaluated the IM-PCM method by employing a leave-one-out cross validation framework over 48 weather stations for the reference period, because this approach is commonly used for validation of spatial interpolation for an ungauged catchment in hydrological studies (Gutjahr and Heinemann, 2013; Rabiei and Haberlandt, 2015). For a more specific analysis in each weather station in the context of cross validation, we generated a map showing the spatial errors in both annual maximum series (AMS) rainfalls and mean, as described in Figures A2 and A3. The AMS errors were illustrated by root-mean-square-error (RMSE) and Nash-Sutcliffe efficiency (NSE) in Figure A2. For the mean, we further evaluated the IM-PCM method by estimating the relative error between the observed and modelled in Figure A3. As shown in the figures, for the AMS rainfalls, gpQM95 and gpQM99 generally perform well except for a few stations. Most stations showed NSE over 0.8 and RMSE less than 30mm. For the mean daily rainfall, the relative errors are generally below 10%.

Figure 1. Indicate the location of the gauges 4, 16, 28, and 40 used in evaluation.

(**Response**) We have additionally indicated the location of Stations 4, 16, 28, and 40 in Figure 1.

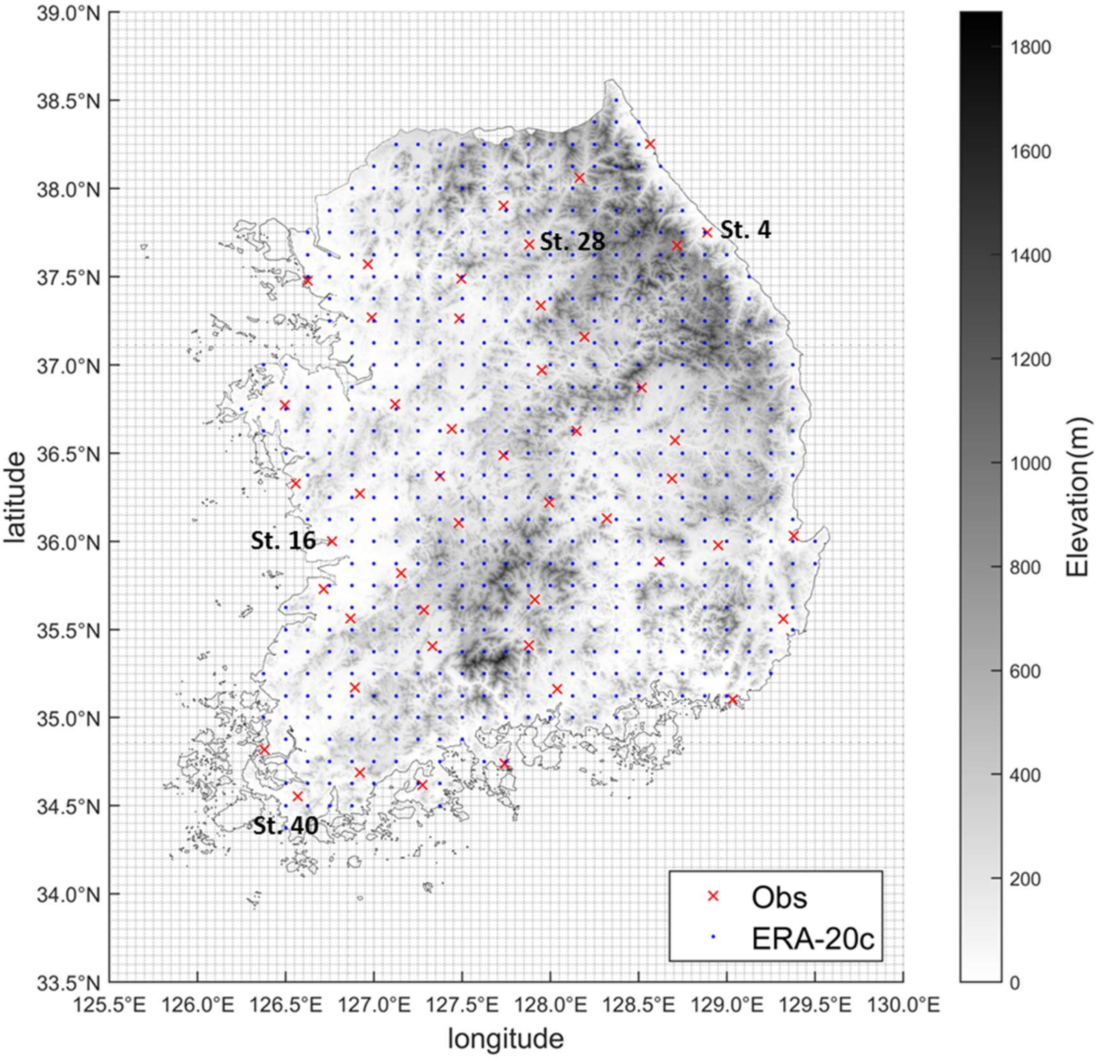


Figure 1. A map showing the study area, local gauging stations and grid points of ERA-20c. The grey shading on the map indicates elevations

[Reference]

Bao, X. and Zhang, F.: Evaluation of NCEP–CFSR, NCEP–NCAR, ERA-Interim, and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan Plateau, J. Clim., 26(1), 206–214, 2013.

Befort, D. J., Wild, S., Kruschke, T., Ulbrich, U. and Leckebusch, G. C.: Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses, Atmos. Sci. Lett., 17(11), 586–595, doi:10.1002/asl.694, 2016.

Bosilovich, M. G., Chen, J., Robertson, F. R. and Adler, R. F.: Evaluation of global precipitation in reanalyses, J. Appl. Meteorol. Climatol., 47(9), 2279–2299, 2008.

Brands, S., Gutiérrez, J. M., Herrera, S. and Cofiño, A. S.: On the use of reanalysis data for downscaling, J. Clim., 25(7), 2517–2526, doi:10.1175/JCLI-D-11-00251.1, 2012.

Coles, S., Pericchi, L. R. and Sisson, S.: A fully probabilistic approach to extreme rainfall modeling, J. Hydrol., 273(1–4), 35–50, doi:10.1016/S0022-1694(02)00353-0, 2003.
Donat, M. G., Alexander, L. V, Herold, N. and Dittus, A. J.: Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations, J. Geophys. Res. Atmos., 121(19), 2016.
Gao, L., Bernhardt, M., Schulz, K., Chen, X. W., Chen, Y. and Liu, M. B.: A First Evaluation of ERA-20CM over China, Mon. Weather Rev., 144(1), 45–57, doi:10.1175/Mwr-D-15-0195.1, 2016.
Gutjahr, O. and Heinemann, G.: Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM, Theor. Appl. Climatol., 114(3), 511–529, doi:10.1007/s00704-013-0834-z, 2013.
Huard, D., Mailhot, A. and Duchesne, S.: Bayesian estimation of intensity-duration-frequency curves and of the return period associated to a given rainfall event, Stoch. Environ. Res. Risk Assess., 24(3), 337–347, doi:10.1007/s00477-009-0323-1, 2010.
Kim, D.-I. and Han, D.: Comparative study on long term climate data sources over South Korea, J. Water Clim. Chang. [online] Available from: http://jwcc.iwaponline.com/content/early/2018/03/14/wcc.2018.032.abstract, 2018.
Krueger, O., Schenk, F., Feser, F. and Weisse, R.: Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations, J. Clim., 26(3), 868–874, doi:10.1175/JCLI-D-12-00309.1, 2013.
Lee, S., Vinayachandran, P. N., Ha, K. and Jhun, J.: Shift of peak in summer monsoon rainfall over Korea and its association with El Niño–Southern Oscillation, J. Geophys. Res. Atmos., 115(D2), doi:10.1029/2009JD011717, 2010.
Ma, L., Zhang, T., Frauenfeld, O. W., Ye, B., Yang, D. and Qin, D.: Evaluation of precipitation from the ERA-40, NCEP-1, and NCEP-2 Reanalyses and CMAP-1, CMAP-2, and GPCP-2 with ground-based measurements in China, J. Geophys. Res. Atmos., 114(D9), 2009.
Overeem, A., Buishand, A. and Holleman, I.: Rainfall depth-duration-frequency curves and their uncertainties, J. Hydrol., 348(1–2), 124–134, doi:10.1016/j.jhydrol.2007.09.044, 2008.
Poli, P., Hersbach, H., Tan, D., Dee, D., Thépaut, J.-N., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P., Dragani, R., Trémolet, Y., Holm, E., Bonavita, M., Isaksen, L. and Fisher, M.: The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C)., 2013.
Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C. and Thépaut, J.-N.: ERA-20C: An Atmospheric Reanalysis of the Twentieth Century, J. Clim., 29(11), 4083–4097, 2016.
Rabiei, E. and Haberlandt, U.: Applying bias correction for merging rain gauge and radar data, J. Hydrol., 522, 544–557, 2015.
Seo, Y., Hwang, J. and Kim, B.: Extreme precipitation frequency analysis using a minimum density power divergence estimator, Water (Switzerland), 9(2), 1–17, doi:10.3390/w9020081, 2017.
Seo, Y. A., Lee, Y., Park, J. S., Kim, M. K., Cho, C. and Baek, H. J.: Assessing changes in observed and future projected precipitation extremes in South Korea, Int. J. Climatol., 35(6), 1069–1078, doi:10.1002/joc.4039, 2015.
Son, C., Lee, T. and Kwon, H. H.: Integrating nonstationary behaviors of typhoon and non-typhoon extreme rainfall events in East Asia, Sci. Rep., 7(1), 1–9, doi:10.1038/s41598-017-04629-1, 2017.
Tung, Y. koung and Wong, C. leung: Assessment of design rainfall uncertainty for hydrologic engineering applications in Hong Kong, Stoch. Environ. Res. Risk Assess., 28(3), 583–592, doi:10.1007/s00477-013-0774-2, 2014.
Van de Vyver, H.: Bayesian estimation of rainfall intensity-duration-frequency relationships, J. Hydrol., 529, 1451–1463, doi:10.1016/j.jhydrol.2015.08.036, 2015.