esd-2017-53: A method to preserve trends in quantile mapping bias correction of climate modeled temperature. Answer to anonymous referee #2

We would like to thank Anonymous Referee #2 for the time dedicated to provide us with constructive comments for a second revision. His opinion and his comments helped us improve the manuscript significantly. We would also like to thank the reviewer for acknowledging the improvement of the present manuscript compared to the previously submitted esd-2016-52. The reviewer's comments (shown in italics) have been addressed point-by-point in the following pages.

***General Summary:***

*The authors present a bias correction method that is intended to address the assumption of stationarity in statistical bias correction. The idea behind the presented method is to disaggregate "stationary" and "non-stationary" components of the model-derived time series by the means of a quantile mapping procedure. Then, the "stationary" (or normalised) component is bias corrected, while the un-corrected residuals are added back to the corrected stationary part of the time series I have reviewed a previous version of the manuscript submitted to ESD (esd-2016-52; Reviewer 1). The authors show that this method preserves the trend signal of the original time series but largely reduces biases relative to an observational dataset. I believe that the paper has indeed substantially improved (and appreciate that the authors have taken the revisions seriously), but nonetheless I am still wondering about several potential conceptual and/or technical problems of the presented approach, which are specified below. Given that the idea of separating stationary and non-stationary parts of a time series for bias correction is certainly interesting, I believe it could be beneficial for the paper and the interpretation/understanding of the methodology if the authors would address these in the manuscript.*

We would like to mention here that we have already aligned the scope of the manuscript with the reviewers' recommendations on the first submission (esd-2016-52). Hence, to avoid misconceptions among the readers, the appropriate emphasis was given to the preservation of the temperature trend rather than the treatment of stationary and non-stationary signal. Moreover, in this version, we expanded the analysis of the BC/BC-NM effect on the signal variability in sub-annual and inter-annual scales. Additionally, a simpler trend preservation procedure was adopted and tested along with BC and BC-NM. All the changes are in detail described below.

***Major comments:***

***RC1:*** *Invasiveness of post-processing methods methodological test in light of previous literature*

*While the idea of separating stationary and non-stationary parts of a time series for bias correction is interesting, I am still wondering about potential side effects of the method. The NSM module fits a transfer function to each year of a grid cell based time series, hence the methodology implies a large number of parameters to be estimated (and thus statistical degrees of freedom); I am not sure which potential side-effects this could imply. Hence, I firstly would encourage the authors to discuss potential side effects, and potential disadvantages of a highly invasive method in comparison to simpler methods; for example, what is the advantage of the NSM+quantile mapping in comparison to a case in which one would simply remove the trend prior to quantile mapping (e.g. Cannon et al., Journal of Climate* ***28****, 6938-6959, 2015, among others)? How much are the results different to the case where one would subtract a smoothed time series on a moving window of few years (i.e. subtract inter-annual variability directly, instead of a complex procedure?); and only bias-correct the remainder part? In any case, I believe a comparison to other, simpler trend-preserving bias correction methods would be a crucial aspect of the paper. Potentially, an example based on random/artificial data could help to underscore the differences and advantages of the NSM+bias correction methodology. Moreover, if (specifically) high-frequency variability is corrected towards a reference dataset derived from observations (these obs. datasets could be very noisy at high* *frequencies, because they are derived from individual sites); I wonder whether there is a certain specific sensitivity of the method to the spatial scale of the observational dataset and its high-frequency noise; i.e. whether the method potentially overcorrects sub-annual variability to noise in observations? (i.e. the so-called inflation problem, see e.g. Maraun, Journal of Climate* ***26****:2137-2143, 2013).*

**AC1:** First, a thorough discussion about the potential caveats and disadvantages of the methodology was added to the (newly created) discussion section of the manuscript (provided later on this response). Moreover, to give an insight into the BC-NM results comparing to a simpler trend preservation approach, we added to the central England example one more method for comparison, where the trend is subtracted from a 5-year moving average prior to the application of the BC, while it is returned after the correction in an additive way. We refer to this experiment as BC-TREND. An introduction of the BC-TREND was added to the end of introduction section (line 150):

> "….The two step procedure is examined for its ability to remove the daily biases with simultaneous preservation of the long term statistics. The procedure is compared to the simple quantile mapping and a quantile mapping with combination with a simpler trend preservation procedure".

In Section 3 - (Case study area and data), a description about the BC-TREND was added after the line 227:

> "…An additional comparison was also performed to a less complicated trend preservation procedure, inspired by (Bürger et al., 2013) and (Cannon et al., 2015). This procedure considers the detrending of the raw data using a 5-year moving average temperature. The detrended data are corrected using the BC methodology, while the trend is additively put back into the time-series after the correction, similarly to the NM. We refer to this as BC-TREND. This comparison is used to benchmark the BC-NM towards a simpler quantile mapping that also approaches the trend preservation."

The respective discussion of the BC-TREND results and comparison to the BC/BC-NM was added between lines 260 and 284, along with the respective changes in Figure 5:

> "The results of the split sample test on the central England example are presented in Figure 5. The NM separates of the raw data into a residuals and a normalized stream (5b). In the annual aggregates the normalized time series do not exhibit any trend or significant fluctuation, since the normalization is performed on annual basis, while the long-term trend and variability are contained in the residual time series. In Figure 5a, annual aggregates obtained via the BC, BC-NM and the BC-TREND procedures are compared to the raw data and the observations. Results show that all three procedures adjust the raw data to better fit the observations in the calibration period 1850-1899. In the validation period, all three procedures produce similar results in terms of mean and standard deviation, but the BC-NM long-term linear trend is slightly lower than that of the BC results and slightly higher than the respective BC-TREND slope. While both BC and BC-TREND slopes are closer to the observations' linear trend, the BC-NM is closer to the raw data trend (Table 2). The BC-TREND validation period trend is found lower relatively to the RAW data, but closer to it, relatively to the BC. This is attributed to the new trend that was introduced to the detrended time series by the differential quantile mapping in each year's CDF, similarly to the example in Figure 1.
>
> Figure 5c shows that in the annual aggregated temperature, the BC-NM resemble the raw data histograms in shape, but shifted in mean towards the observations. A small decrease in the variability is observed in the BC-NM relatively to the raw data but consists a substantially smaller

disturbance relatively to the BC. The annual variability in BC-TREND is closer to the raw data comparing to the BC approach, but BC-NM still outperforms in the annual variability preservation. The transfer of the mean with a simultaneous preservation of the larger part of the variability of the BC consists a nearly idealized behavior for the adjusted data when the long term statistics preservation is a desired characteristic, as the distribution of the annual temperature averages are retained after the correction (trend, standard deviation, and inter-quartile range - Table 2). The respective results generated on daily data (Figure 5d) show that all three procedures adjust the calibration and validation histograms in a similar degree towards the observations. This is also verified by the mean, standard deviation and the 10th and 90th percentile of the daily data of Table 2. An early concluding remark about the NM is that it retained the long-term statistics of the adjusted data towards the climate model signal better than the alternative approaches, without however sacrificing the daily scale quality of the correction."

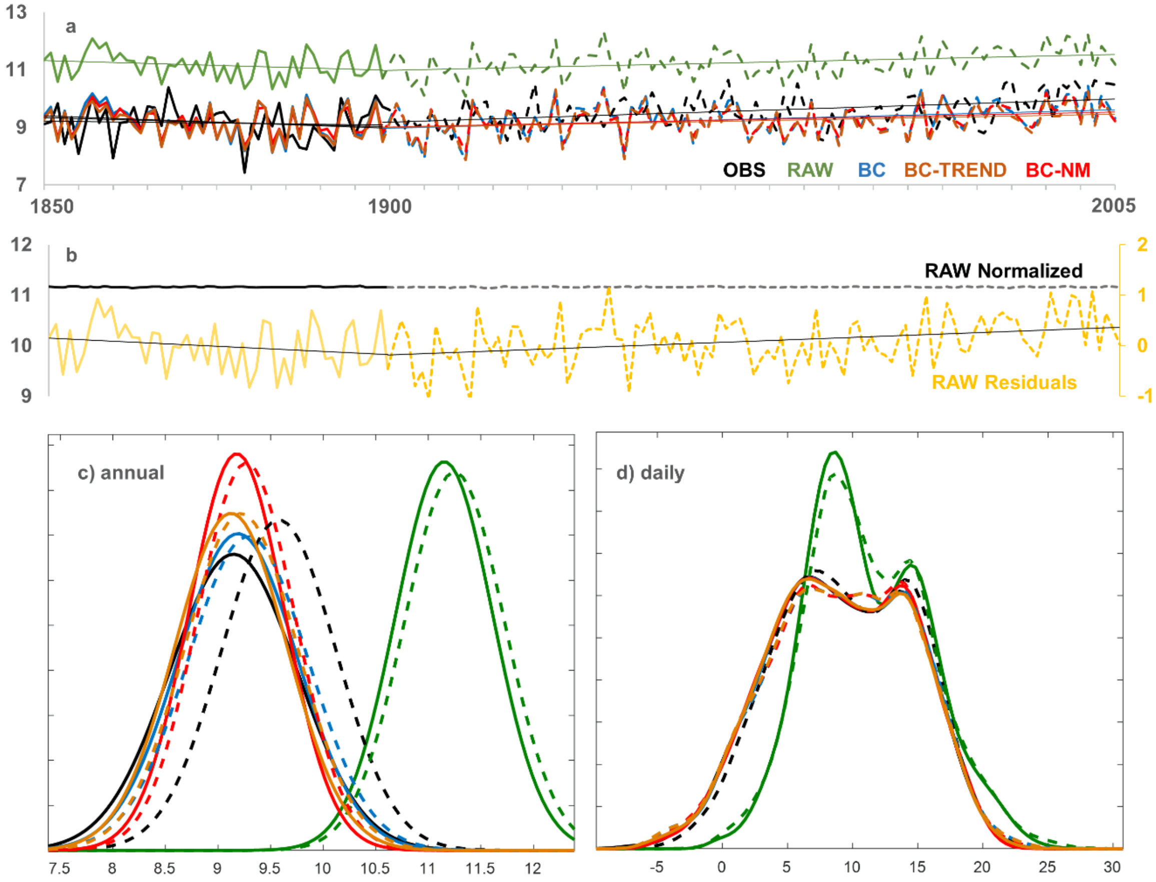


Figure 1: a) annual average temperature of raw model, observations and the bias corrected with, without the NM data and following the BC-TREND approach, for the calibration period 1850 – 1899 (solid lines) and the validation period 1900-2005 (dashed lines). b) Annual averages of the normalized and the residuals of the raw temperature. Probability densities of annual (c) and of daily means (d).

The new section was added right after line 318, and presents the advantages and disadvantages of the BC-NM, and how it compares with other method in the literature:

"The methodology shares similarities to other correction methods found in the literature. Furthermore it exhibits a number of advancements that overcomes drawbacks of other trend preserving methodologies. The fundamental idea of the presented method is also identified in Haerter et al., (2011) method that considers two different timescales and performs a cascade correction of temperature. In the present study annual and daily scales are used for the separation of the temperature signal in two parts. While in the former methodology, the cascade correction benefits the results in both timescales, here the separation offers a correction in the daily scale and an intentional preservation of the raw model statistics in the annual scale. A comparisons can be made to the methodology of (Li et al., 2010) that use the differences in the raw data between the reference period and the projection period. In the present study the differences are defined between the reference period and each year of correction separately. This is an improvement to the technique that overcomes the subjectivity of the future period selection. Additionally, the quantile mapping correction ensures the skillful correction in the higher and lower quantiles, relatively to simpler additive approaches such as (Hempel et al., 2013) that although preserving the trend and year-to-year variability, it marginally improve the tails of the temperature distribution (Sippel et al., 2016). Regarding the simpler BC-TREND version that was used for the central England example, it was found that it tends to preserve the long term statistics as also noted by (Cannon et al., 2015), but still, the 5-year average that was used for the trend preservation cannot encompass the changes in each year's CDF, as the NM can.

Beyond the advancements, a drawback of the presented methodology is the use of a large number of parameters to approximate the transfer functions in the two stages of the correction. The methodology can be described as of 'varying complexity' due to the number of the estimated parameters (number of segments) and the added value of the complexity being weighed by an information criterion. In the case of use of high noise observations, it would lead to the transfer of that noise to the corrected data variability. This was marginally detected in the analysis of the standard deviations in Figure 9, even if the effect of BC-NM mitigated the effect comparing to the BC. Another weakness stems from the residuals exclusion from the correction. In the theoretical case where the future projected temperature variability change considerably relative to the reference period, the correction would result to larger remaining biases as it was shown earlier, that could impair the physical continuity of the time series. This should be a consideration in the case that BC-NM was used to correct other types of variables."

***RC2:** Non-correction of inter-annual variability: Discussion about the concept of stationarity and which components actually could/should be corrected*

*Furthermore, I am still wondering about the authors' use of "non-stationary" vs. "stationary", where the former term is used for inter-annual and lower frequencies; whereas the latter term is used for sub-annual variations. To my mind, inter-annual variability (that is not corrected by purpose in the proposed method) could well be stationary, but biased. For example, model deficiencies on the inter-annual time scale are well-known and often related to land-atmosphere interactions (e.g. Fischer et al., Geophysical Research Letters **39**(19), 2012). The illustrative example presented in Fig. 5c shows that the raw model underestimates inter-annual variability compared to the observations that are (by intention) not corrected in the BC-NSM method. I believe it would be worthwhile to discuss in the manuscript, whether this aspect is indeed desired by a bias correction of "non-stationary" components? To this end, I believe that the authors could evaluate and discuss which kind of variability is being corrected by their method, using for example power spectra. Thereby, one could potentially learn or discuss how the spectrum of temperature variability (e.g. see Huybers and Curry, Nature **441**(7091), 2006) produced by a given model is altered by their bias correction methodology (or by other methods).*

Following the concerns of the reviewer and to avoid any potential misunderstanding, we removed the term *stationarity* and *non-stationarity* from lines 143 and 293.

Beyond that, Reviewer #2 points out the importance of inter-annual variability and whether it is a desirable feature to preserve it in the corrected data. Authors believe that this depends on the application on which the bias corrected data are intended to be used. The key idea of the bias correction with a simultaneous intentional preservation of some statistics holds on applications that require those statistics, as for example in biophysical impacts projections (similarly to the purposes that ISIMIP (https://www.isimip.org/ ) initiative use the Hempel et al., (2013) method). Another example is the objectives of HELIX (FP7) project where the preservation of the long term trend was needed for the direct inter comparison of the biophysical impacts between the raw and bias adjusted data simulations. In the case of Figure 5c, there is a clear underestimation of the inter-annual variability of the raw (and BC-NM) data, but this is a case where the raw and observed time-series have the same length. However this is not always the case, hence, the intentional preservation of the inter-annual statistics may be safer than the unintentional correction, especially in cases where the observational data record is not long enough. Motivated by this comment, we added the standard deviation and interquartile range estimates on annual data in Table 2. Discussion on the variability change was added to the discussion section (line 318).

***Table 1: Statistical properties of the calibration and the validation periods for the two bias correction procedures. Variables denoted with * are estimated on annual aggregates. SD stands for standard deviation, pn for the nth quantile and IQR for the interquartile range.***

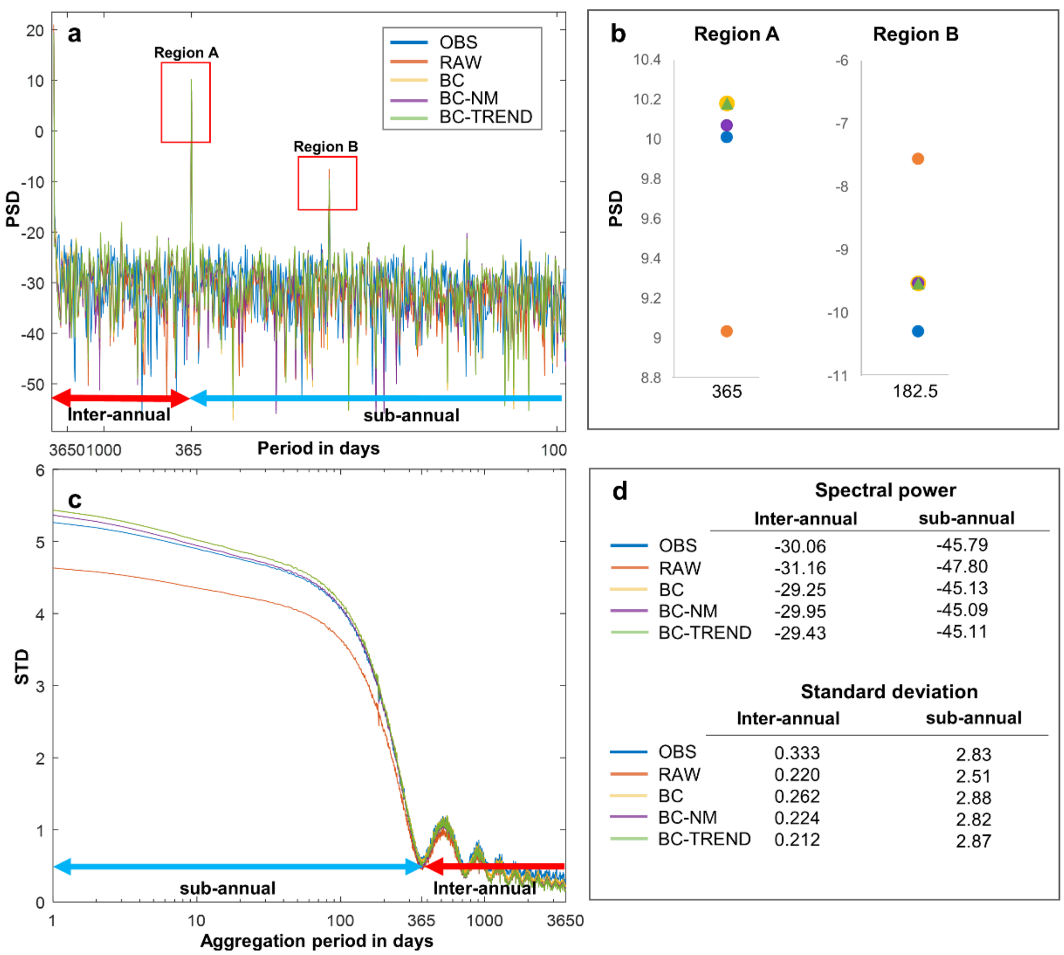
| | Parameter | RAW | Normalized | Residuals | OBS | BC | $BC_{NM}$ | $BC_{TREND}$ |
|---|---|---|---|---|---|---|---|---|
| Calibration | Mean [°C] | 11.2 | 11.2 | 0.0 | 9.1 | 9.2 | 9.2 | 9.1 |
| | SD [°C] | 4.5 | 4.6 | 0.9 | 5.3 | 5.3 | 5.3 | 5.3 |
| | p10 [°C] | 5.7 | 5.7 | -0.9 | 2.1 | 2.2 | 2.2 | 2.1 |
| | p90 [°C°] | 17.4 | 17.2 | 1.0 | 16.3 | 16.3 | 16.2 | 16.2 |
| | Slope [°C/10yr]* | -0.067 | 0.000 | -0.067 | -0.026 | -0.086 | -0.065 | -0.061 |
| | SD [°C]* | 0.46 | 0.46 | 0.01 | 0.61 | 0.57 | 0.45 | 0.53 |
| | IQR* | 0.76 | 0.76 | 0.01 | 0.86 | 0.95 | 0.75 | 0.94 |
| Validation | Mean [°C] | 11.3 | 11.2 | 0.1 | 9.6 | 9.3 | 9.3 | 9.2 |
| | SD [°C] | 4.7 | 4.6 | 0.9 | 5.2 | 5.5 | 5.4 | 5.5 |
| | p10 [°C] | 5.6 | 5.7 | -0.9 | 2.7 | 2.0 | 2.0 | 1.9 |
| | p90 [°C] | 17.4 | 17.2 | 1.0 | 16.3 | 16.3 | 16.2 | 16.2 |
| | Slope [°C/10yr]* | 0.052 | 0.000 | 0.051 | 0.076 | 0.062 | 0.051 | 0.044 |
| | SD [°C]* | 0.48 | 0.47 | 0.01 | 0.54 | 0.57 | 0.46 | 0.53 |
| | IQR* | 0.63 | 0.62 | 0.01 | 0.76 | 0.75 | 0.62 | 0.68 |

"This study focuses on known issues of bias correction that appear in the literature. Whether the long term signal of temperature should be preserved or not, has been discussed in a more theoretical level in (Maraun, 2016), while (Haerter et al., 2011) mention that a credible bias correction methodology should involve the consequences of greenhouse gas concentration changes. This is somehow consistent with temperature trend preservation as the model sensitivity is retained in the corrected timeseries. As stated in (Fischer et al., 2012), models tend to underestimate the inter-annual variability due to deficiencies between land-atmosphere interactions, which urge for its correction. Nevertheless, the long-term statistics preservation may be necessitated in cases that

> temperature is used in biophysical impact modeling (Rubino et al., 2016), or may be preferred as a safer option than the unintentional alteration, especially in cases where the observational data record is not long enough."

Next, following the concerns about the variability correction in the different sub-annual and inter-annual time scales, we adopted the indicated example of (Huybers and Curry, 2006), introducing a power spectral density analysis of the BC, BC-NM and BC-TREND results. A new figure (Figure 6) was added, while discussion about its results were added after line 287:

> "To further inter-compare the effect of each approach in the data variability beyond the inter-annual and the daily basis, the power spectral density – PSD was estimated (Huybers and Curry, 2006) over their daily temperature signals (Figure 6). The marked spectral peaks associate with the annual and 6-month periodicity is and expected result. Focusing on those regions (Figure 6b), it is shown that the BC-NM is closer to the observational variability relatively to the other two correction techniques, while in the 6-months all techniques provide similar results. The average power density of the domain beyond the annual periodic shows that BC-NM is closer to the raw data, while the respective sub-annual average is almost equal to the BC and the BC-TREND averages. Figure 6c shows the standard deviation estimated on temperature aggregates between 1 and 10957 days (i.e. 30 years). Figure 6d shows the average variability and average spectral power of the two scaling regimes, above and below annum. The sub-annual scales average variability of BC-NM resembles the observational variability, outperforming the BC and BC-TREND approaches that show higher values. More importantly, the NM works well in the inter-annual scale where the average variability is found to be closer to the raw data variability compared to the inflated BC and the deflated BC-TREND results."



Spectral power

| | Inter-annual | sub-annual |
|---|---|---|
| OBS | -30.06 | -45.79 |
| RAW | -31.16 | -47.80 |
| BC | -29.25 | -45.13 |
| BC-NM | -29.95 | -45.09 |
| BC-TREND | -29.43 | -45.11 |

Standard deviation

| | Inter-annual | sub-annual |
|---|---|---|
| OBS | 0.333 | 2.83 |
| RAW | 0.220 | 2.51 |
| BC | 0.262 | 2.88 |
| BC-NM | 0.224 | 2.82 |
| BC-TREND | 0.212 | 2.87 |

Figure 2: Power spectral density of temperature (a) and high power regions of annual and half year periods (b). Standard deviation of temperature aggregates between 1 and 10957 days (horizontal axis visible between 1 day and 10 years) in (c). In (d), the Inter-annual and sub-annual periods average (denoted with red and cyan arrows respectively) spectral power (a) and standard deviation (c).

### *RC3: Generalisation to other variables*

*Finally, it would be interesting if the authors could discuss (at some point in the manuscript) whether their method is intended for bias correction of temperature only, or whether the method could be extended towards other climatic variables as well?*

**AC3:** The presented methodology was designed and tested on RCM temperature results, without bounding its use to temperature only. However, attention should be paid in the limitations of the methodology. Appropriate mention about the use of the methodology in different variables was added at the end of the discussion (last lines of **AC1**).

***Minor comments:***

*p. 6, l. 183-185: If the CDF is split in segments and then the correction of each segment is performed according to Eq. 3, isn't there a possibility that there could be gap changes in the correction at the edges of the segments?*

> **AC:**Yes it is possible, but its size and effect is negligible to arbitrarily alter the statistics of the corrected data. This effect might be pronounced in the edge segments, but as it is described in the methods, edges are explicitly corrected using the average difference between the reference period of the raw model data and the observations.

*p. 7, l. 212-236: I believe the explanation of the split-sample test could be separated from the introduction of the case study area and data in the paper?*

> **AC:** The split sample methodology was moved in Section 2 Methods (2.3 Validation of the results).

*p. 8, l. 250-252: Is this statement "... some of the bias is attributed to the ability of the observation dataset to represent temperature..." based on evidence or speculation?*

> **AC:** This is the case of E-OBS dataset. As it is based on station data interpolation, it is highly affected by the station density. When fewer than a sufficient number of stations are used to estimate the variance, it is likely to be larger than the true variance (Hofstra et al., 2010). In the Eastern Europe, the station density used for the EOBS dataset is lower, hence it is likely to have increase variance in comparison to that it would have if the number of stations was grater. A reference was added to the manuscript.

**References:**

Bürger, G., Sobie, S.R., Cannon, A.J., Werner, A.T., Murdock, T.Q., Bürger, G., Sobie, S.R., Cannon, A.J., Werner, A.T., Murdock, T.Q., 2013. Downscaling Extremes: An Intercomparison of Multiple Methods for Future Climate. J. Clim. 26, 3429–3449. doi:10.1175/JCLI-D-12-00249.1

Cannon, A.J., Sobie, S.R., Murdock, T.Q., Cannon, A.J., Sobie, S.R., Murdock, T.Q., 2015. Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? J. Clim. 28, 6938–6959. doi:10.1175/JCLI-D-14-00754.1

Fischer, E.M., Rajczak, J., Schär, C., 2012. Changes in European summer temperature variability revisited. Geophys. Res. Lett. 39, n/a-n/a. doi:10.1029/2012GL052730

Haerter, J.O., Hagemann, S., Moseley, C., Piani, C., 2011. Climate model bias correction and the role of timescales. Hydrol. Earth Syst. Sci. 15, 1065–1079. doi:10.5194/hess-15-1065-2011

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., Piontek, F., 2013. A trend-preserving bias correction – the ISI-MIP approach. Earth Syst. Dyn. 4, 219–236. doi:10.5194/esd-4-219-2013

Hofstra, N., New, M., McSweeney, C., 2010. The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. Clim. Dyn. 35, 841–858. doi:10.1007/s00382-009-0698-1

Huybers, P., Curry, W., 2006. Links between annual, Milankovitch and continuum temperature variability. Nature 441, 329–332. doi:10.1038/nature04745

Li, H., Sheffield, J., Wood, E.F., 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching.

J. Geophys. Res. 115, D10101. doi:10.1029/2009JD012882

Maraun, D., 2016. Bias Correcting Climate Change Simulations - a Critical Review. Curr. Clim. Chang. Reports 2, 211–220. doi:10.1007/s40641-016-0050-x

Rubino, M., Etheridge, D.M., Trudinger, C.M., Allison, C.E., Rayner, P.J., Enting, I., Mulvaney, R., Steele, L.P., Langenfelds, R.L., Sturges, W.T., Curran, M.A.J., Smith, A.M., 2016. Low atmospheric CO2 levels during the Little Ice Age due to cooling-induced terrestrial uptake. Nat. Geosci. 9, 691–694. doi:10.1038/ngeo2769

Sippel, S., Otto, F.E.L., Forkel, M., Allen, M.R., Guillod, B.P., Heimann, M., Reichstein, M., Seneviratne, S.I., Thonicke, K., Mahecha, M.D., 2016. A novel bias correction methodology for climate impact simulations. Earth Syst. Dyn. 7, 71–88. doi:10.5194/esd-7-71-2016