Authors would like to thank Dr. Stefan Hagemann for his eponymous review of the manuscript and his insightful comments. His deep knowledge on the field and the positive opinion on the presented methodology encouraged us to provide a comprehensive response to all major and minor comments provided. Here, the reviewer's comments (in italics) are answered point-by-point.

Major remarks

SH1: The authors present a bias correction approach for simulated temperatures where this post processing is separated into three steps. Step 1 comprises the separation of the time series in a normalized and the residuals. In Step 2, a quantile mapping based bias correction is applied to the normalized time series, and then the residuals are added again to this bias corrected time series in the third step. In this way, trends and anomalies of the original time series are better preserved than in the case where the bias correction is applied to the full original time series. The paper clearly presents the new method and its application, but I miss some more information on where does this method stand within general bias correction related research. Given the huge number of papers that have been published on the topic of bias correction within the recent years, the discussion section of the present study needs more comparison to other methods, especially to those that follow similar or related approaches, i.e. those studies that do not only perform quantile mapping based bias correction but also do something more. For example, how does the new method compares to the method of Haerter et al. (2011) who proposed a separation of time scales when applying bias correction, i.e. daily fluctuations are differently corrected than the monthly means. In a way, you are also separating time scales in your approach of separating into a normalized time series and the residuals. How would corrections according to Hempel et al (2013) look different than those obtained in the present study? The authors mentioned the approach of Hempel et al. (2013) in the introduction, but in the moment I don't see how the mentioned conceptual drawback may actually impact the bias corrected time series. With regard to pros and cons, disadvantages and problems with BC, it is referred to previous literature. Even though I think that this generally ok to do so, I also believe that the conclusions section would profit from a paragraph about issues for which also the new method would not lead to improved results or would even lead to misleading results.

AR1: Following the above recommendations, the discussion section (separated from the results section) was expanded. A thorough discussion about the potential caveats and disadvantages of the methodology was added and further comparison about other methods. Moreover, to give an insight into the BC-NM results comparing to a simpler trend preservation approach, we introduced another simpler method for comparison to the central England example. In this method, the trend is subtracted from a 5-year moving average prior the application of the BC, while it is additively returned after the correction. We refer to this experiment as BC-TREND. Starting with the newly introduced methodology, a first notion 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."

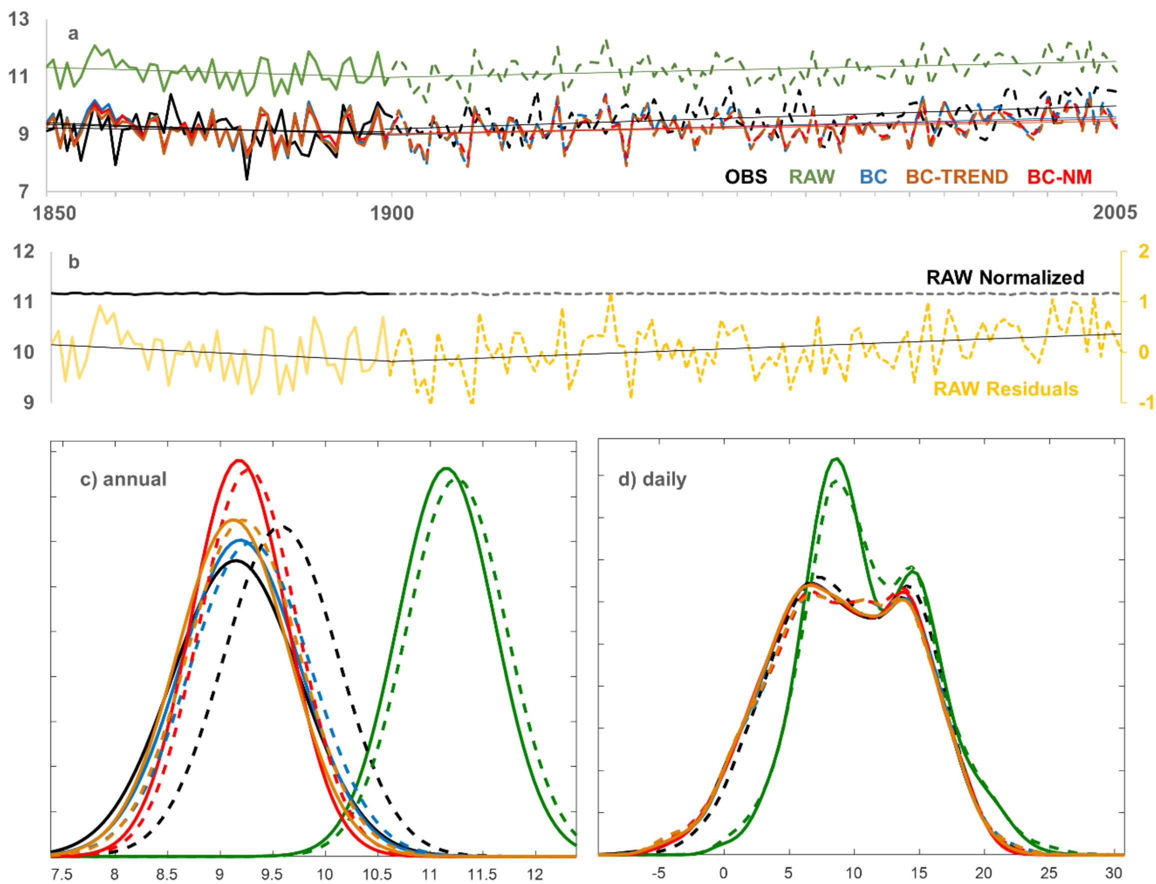

Figure 1: a) annual average temperature of raw model, observations and the bias corrected with, without the NM data and following the BC-TREND approach, for the calibration period 1850 – 1899 (solid lines) and the validation period 1900-2005 (dashed lines). b) Annual averages of the normalized and the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

SH2: In some parts of the manuscript, the English is difficult to read or uses some unusual terms. I recommend proof reading by a native speaker.

AC2: The manuscript has been thoroughly revised to improve readability.

SH3: The paper focuses purely on the new methodology and its implications on the corrected time series. Thus, the paper itself is a solid piece of scientific research and worth publishing. But I am wondering whether the authors chose the right journal for this, as I don't see aspects of Earth System Dynamics (ESD) in the paper. As the content might be interesting for climate impact modelers, especially hydrologists, another Copernicus journal such as HESS seems to be much more appropriate than ESD for the publication of the manuscript. In summary I suggest placing the paper in another Copernicus journal such as HESS, and accepting the paper for publication after some revisions have been conducted. I don't wish do stay anonymous, Stefan Hagemann

AC3: This research work elaborates with the presentation and the analysis of a bias correction methodology which consists the scientific research part, but also analyzes the effect of the methodology on the historical temperature of five RCMs in European scale. According to the authors, the study fits on the ESD focus region of *Earth system change,* in the same context that Hempel et al., (2013) and Sippel et al., (2016) works are published on ESD.

Beyond the above changes, a spectral power density analysis was added to the central England experiment according to the indications of R#2. Furthermore, various small changes across the manuscript were performed in order to include the new results and in order to become clearer.

Minor remarks

p.4 – line 125/126

Do you mean: '… which is not a common practice." ?

AC: Corrected as indicated

p.3 – line 95-96

It is written:

"The procedure however overlooks the time dependency of the biases, i.e the unequal effect of the TF to the varying over time CDF." This sentence is difficult to read and not clear to me. The temperature distribution is varying with time, especially when a climate change signal is present. But that does not necessarily mean that the bias is also varying. Especially in climate change applications of bias correction, it is inherently assumed that the bias does not change with time. If the bias would actually be time dependent, an application of the bias correction to future data may be questionable. Hence, please rephrase and explain more thoroughly.

AC: In accordance with the reviewer's suggestion, we have changed this sentence to ["The procedure however overlooks the time dependency of the distribution and hence the unequal effect of the TF to the varying over time CDF."]

p.4 – line 130

… problem as individual model trend changes were cancelled out.

AC: Changed according to the suggestions

p.11 – line 334-335

Sentence has wrong grammar. Please rewrite!

AC: The sentence was changed according to the suggestions

p.11 – line 336-338

This is a noteworthy feature, but I don't see this as an advantage (or disadvantage).

AC: The sentence was removed as redundant

This sentence is difficult to read, but its content is also rather trivial. It more or less means: 'The main reason for the quality of a climate model is its skilfulness.' Or in other words: 'A model is good because it is well performing.' I suggest removing this sentence.

AC: The sentence was removed accordingly to the suggestions

p.11 – line 345-346

On one hand, this is an important statement. On the other hand it should also be stated that in some cases, where the climate model has some problems, e.g. with circulation feature or subgrid scale processes, the added usefulness may obscure that the bias correction can even deteriorate climate change signals (see also major remarks).

AC: Appropriate statement was added to the indicated point

p.19 – Fig. 3

This figure comprises too many panels. As the panels showing the absolute MEAN and STD values of the five RCMs do not add much valuable information, I suggest removing them.

AC: Figure 3 was simplified by removing the individual models. The original figure was added to the electronic supplementary material.

p.21 – Fig. 5

Panel b) is not cited in the caption. In the curve legend below panel b, there is a typo: BCNSM. This should be corrected to BC-NM.

AC: The typo was corrected. Panels a) and b) were rearranged and the caption was corrected accordingly.

p.22 – Fig. 6

The panels c, d, g, h do not add valuable information. Please remove. Instead I suggest adding a panel showing the difference of RAW-EOBS.

AC: The panels c, d, g and h were omitted from the figure, while a new panel with the difference between RAW and EOBS was added. The old figure was included to the electronic supplementary material.

In the following suggestions for editorial corrections are marked in Italic.

p.2 – line 38

… dependency of the temperature bias.

With regard to this sentence itself, please see remark to p.3 line 95-96 below.

p.2 – line 40

…the modelled reference …

p.2 – line 41-42

… and preserve the signal of the latter.

p.2 – line 45

… improvements due to this method.

p.3 – line 69

… output provides the …

p.4 – line 122

Maraun (2016) discusses on …

p.5 – line 136

… approach is that …

p.6 – line 166

…in order to be added later again …

p.8 – line 235

…long-term transient climate …

p.8 – line 261-263

… in Figure 5, which shows the NM separation of the raw data into residuals and normalized raw data in annual aggregates.

p.9 – line 365

… variability are contained …

p.10 – line 312

… standard deviations of the adjusted …

p.10 – line 321

This study elaborates the issue …

p.10 – line 329-331

… evolution shows that it is better … …residuals from the …

p.11 – line 332

… which comprises an …

p.11 – line 339

… performed on an annual…

p.11 – line 344-345

p.23 – Fig. 7 caption – line 566

…percentiles on an annual …