A method to preserve trends in quantile mapping bias correction of climate modeled temperature

Manolis G. Grillakis[1], Aristeidis G. Koutroulis[1], Ioannis N. Daliakopoulos[1], and Ioannis K. Tsanis[1,2]

[1] {Technical University of Crete, School of Environmental Engineering, Chania, Greece}

[2] {McMaster University, Department of Civil Engineering, Hamilton, ON, Canada}

Manolis G. Grillakis Ph.D.

Phone: +30.28210.37728, Fax: +30.28210.37855, e-mail: manolis@hydromech.gr

Aristeidis G. Koutroulis Ph.D.

Phone: +30.28210.37764, Fax: +30.28210.37855, e-mail: aris@hydromech.gr

Ioannis N. Daliakopoulos Ph.D.

Phone: +30.28210.37800, Fax: +30.28210.37855, e-mail: daliakopoulos@hydromech.gr

Ioannis K. Tsanis Ph.D.

Phone: +30.28210.37799, Fax: +30.28210.37849, e-mail: tsanis@hydromech.gr

correspondence email for proofs: manolis@hydromech.gr

Abstract

Bias correction of climate variables is a standard practice in Climate Change Impact (CCI) studies. Various methodologies have been developed within the framework of quantile mapping. However, it is well known that quantile mapping may significantly modify the long term statistics due to the time dependency of the temperature bias. Here, a method to overcome this issue without compromising the day-to-day correction statistics is presented. The methodology separates the modeled temperature signal into a normalized and a residual component relatively to the modeled reference period climatology, in order to adjust the biases only for the former and preserve the signal of the later. The results show that this method allows for the preservation of the originally modeled long-term signal in the mean, the standard deviation and higher and lower percentiles of temperature. To illustrate the improvements, the methodology is tested on daily time series obtained from five Euro CORDEX RCM models.

Keywords: temperature, trend preservation, moment preservation, statistical bias correction,

1 Introduction

Climate model output provides the primary source of information used to quantify the effect of the foreseen anthropogenic climate change on natural systems. One of the most common and technically sound practices in Climate Change Impact (CCI) studies is to calibrate impact models using the most suitable observational data and then to replace them with the climate model data in order to assess the effect of potential changes in the climate regime. Often, raw climate model data cannot be used in CCI models due to the presence of biases in the representation of regional climate (Christensen et al., 2008; Haerter et al., 2011). In fact, hydrological CCI studies outcome have been reported to become unrealistic without a prior adjustment of climate forcing biases (Hagemann et al., 2013; Hansen et al., 2006; Harding et al., 2014; Sharma et al., 2007). Papadimitriou et al. (2017) quantified the effect of the bias in seven forcing parameters on the resulted runoff of a land surface model, emphasizing the necessity of bias adjustments beyond the precipitation and temperature parameters. These biases are attributed to a number of reasons such as the imperfect representation of the physical processes within the model code and the coarse spatial resolution that do not permit the accurate representation of small-scale processes. Furthermore, in some cases, climate model tuning for global projections focuses on the adequate representation of feedbacks between processes and hence the realistic depiction of a variable, such as temperature, against observations is sidelined (Hawkins et al., 2016).

A number of statistical bias correction methods have been developed and successfully applied in CCI studies (e.g. Grillakis et al., 2013; Haerter et al., 2011; Ines and Hansen, 2006; Teutschbein and Seibert, 2012). Their main task is to adjust the statistical properties of climate simulations to resemble those of observations, in a common climatological period. A commonly used type of procedure to accomplish this is using a Transfer Function (TF) which minimizes the difference between the cumulative density function (CDF) of the climate model output and that of the observations, a process also referred to as quantile mapping. As a result of quantile mapping, the reference (calibration) period’s adjusted data are statistically closer, and sometimes near-identical to the observations. Hence, the statistical outcome of an impact model run using observational data is likely to be reproduced by the adjusted data. The good performance of statistical bias correction methods in the reference period is well documented (Grillakis et al., 2011; Grillakis et al., 2013; Ines and Hansen, 2006; Papadimitriou et al., 2015). The procedure however overlooks the time dependency of the distribution and hence the unequal effect of the TF to the varying over time CDF. An indicative example is presented in Figure 1 where modeled temperature data have a mean bias of 2.49 ^{o}C in the reference period (Figure 1a) relatively to the observations. This mean bias is expressed by the average horizontal distance between the TF and the bisector of the central plot. The left histogram illustrates the reference period modeled data for 1981-2010. The histogram at the bottom is derived from observational data. The histogram on the right is derived from a moving 30-year period between 1981 and 2098. The rightmost histogram shows the difference between the reference period and the moving 30-year period. The red mark shows the theoretical change in the average correction applied by the TF, due to the changes in the projected temperature histogram. Hence, the average correction applied for the period 2068-2097 reaches 3.85 ^{o}C, significantly higher than the reference period's bias (Figure 1b). The time-dependency of the correction magnitude introduces a long term signal distortion in the corrected data. In the quantile mapping based correction methodologies in which the TF distance from the bisector is variable, this effect is unavoidable. Nevertheless, in cases where the TF retains a relatively constant distance to the bisector (i.e. parallel to the bisector), the trend of the corrected data remains similar to the raw model data regardless of the temporal change in the model data histogram.

Based on the previous example, the time extrapolation of the TF is regarded as a leap of faith that may lead to a false certainty about the robustness of the adjusted projection. This may significantly change the original modeled long-term trend or other higher moments of the climate variable statistics that eventually change the long-term signal of the climate variable. In their work on distribution based scaling (DBS) bias correction, Olsson et al. (2015) showed that their methodology might alter the long-term temperature trends, attributing the phenomenon in the severity of the biases in the mean or the standard deviation between the uncorrected temperatures and the observations. Maraun (2016), discusses on whether the change in the trend is a desired feature of bias correction, concluding that it is case specific and depends on the skillfulness of the climate model to simulate the correct long term signal. In the case of CCI studies, this implies that climate model data is assessed for its skill to well represent the trend, which is not a common practice. A possible but indirect solution to this is described in Maurer and Pierce (2014) who study the change in precipitation trend over an ensemble of atmospheric general circulation model (AGCM). They conclude that, while individual quantile mapping corrected AGCM data may significantly modify the signal of change, a relatively large ensemble estimation diminishes the problem as individual model trend changes were cancelled out . Li et al. (2010) present a quantile mapping method to adjust temperature biases taking into account the differences of the future and reference period distributions. A drawback of the method is that the difference between the two periods' distributions depends on the future period length. In their work, Hempel et al. (2013) propose a methodology to resolve the trend changing issue by preserving the absolute changes in monthly temperature and relative changes in monthly values of precipitation. A characteristic of their approach is that it maps anomalies instead of absolute values, indicating that specific correction values are attached to each temperature anomaly, while it also has the drawback that the edges of the distribution are not corrected adequately. A similar additive for temperature and multiplicative for precipitation approach was also followed by (Pierce et al., 2015). Bürger et al., (2013) and Cannon et al. (2015) test the de-trending of the data prior their quantile mapping correction, figuring that the removal of the trends prior to the quantile mapping and its reintroduction after the correction tend without absolutely maintain the long term.

In this study, we present a methodology to conserve the long term statistics such as trend and variability of the climate model data in quantile mapping. The methodology considers the separation of the temperature signal relatively to the raw data reference period, producing a normalized and a residuals data stream. The separation is performed on an annual basis. The residuals include the gradual changes in the signal and the year-to-year fluctuations in the distribution of the temperature. The quantile mapping bias correction is then applied to the normalized daily temperature. Finally, the residual components are merged to the bias corrected time series to form the corrected time series. The idea of identifying and using two different timescales in bias correction of temperature was introduced in Haerter et al. (2011), that present a method to separate the different timescales and apply a correction to each one. The methodology presented here is tested along with a generalized version of the Multi-segment Statistical Bias Correction (MSBC) quantile mapping methodology (Grillakis et al., 2013). The methodology takes the form of a pre- and post-processing module that can be applied along with different statistical bias correction methodologies. 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 in combination with a simpler trend preservation procedure.

2 Methods

2.1 Residual separation

The statistical difference of each individual year's simulated data, compared to the average reference period simulated data is identified as residuals. These are estimated between the CDF of each year's modeled climate data and the CDF of the entire reference period of the model data. Let S_R be the reference period model data and S_i the climate data for year i, then the normalized data S_i^N for year i are estimated by transferring each year's data onto the average reference period CDF through a transfer function TF_{S_i} estimated annually. This can be formulated as Eq.(1).

$$S_i^n = TF_{S_R}^{-1}\left(TF_{S_i}\,(S_i)\right) \qquad \text{Eq. (1)}$$

The difference between the original model data S_i and the normalized data S_i^N are the residual components S_i^D of the time series (Eq. (2)).

$$S_i^D = S_i \; - \; S_i^n \qquad \text{Eq. (2)}$$

The original model data S_i can be reproduced by adding back the residuals S_i^D to the normalized data S_i^n. After the separation, the normalized climate model data are statistically bias corrected following a suitable methodology. The residuals are preserved in order to be later added back to the bias corrected time series. We refer to the described method as Normalization Module (NM) to hereafter lighten the nomenclature of the paper. The normalization procedure is performed on annual basis, as this consists an obvious periodicity to use in the case of temperature, even if it is not so well defined in the tropics. The underlying assumption of the NM procedure is that it considers no major changes in the reference period data, an notion that can hardly fall short due to the usually short length of the reference period.

2.2 Bias correction

Here, the NM is applied along with a modification of the MSBC algorithm that is presented in Grillakis et al. (2013). This methodology follows the principles of quantile mapping correction techniques and was originally designed and tested for GCM precipitation adjustment. The method partitions the CDF data into discrete segments and an individual quantile mapping correction is applied to each segment, achieving a better fitted transfer function. Here the methodology is modified to use linear functions instead of the gamma functions used in the original methodology, in order to facilitate potential negative temperature values but also as a known technique in quantile mapping, as it has also been used elsewhere (Themeßl et al., 2011). An indicative example is shown in Figure 2, where the CDFs are split into discrete segments and linear functions are fit to each of them. In Figure 2, p symbolizes the cumulative probability and s is the slope of the linear function. Then the corrected temperature for each temperature value of the specific segment is estimated as in Eq. (3).

$$T_{corr}^{n} = s_{obs}^{n} * \left(\frac{T_{raw}^{n} - b_{raw}^{n}}{s_{raw}^{n}}\right) + b_{obs}^{n} \quad \text{Eq. (3)}$$

The optimal number of segments is estimated using the Schwarz Bayesian Information Criterion (SBIC) to balance between complexity and performance. Additionally, the upper and lower edge segments are explicitly corrected using the average difference between the reference period of the raw model data and the observations (Figure 2 ΔT). This provides robustness, avoiding unrealistic temperature values at the edges of the model CDF. The bias correction methodology modification has been already used in the Bias Correction Intercomparison Project (BCIP) (Nikulin et al., 2015), while produced adjusted data have been used in a number of CCI studies (Daliakopoulos et al., 2016; Grillakis et al., 2016; Koutroulis et al., 2016; Papadimitriou et al., 2017, 2016). As the MSBC methodology belongs to the parametric quantile mapping techniques, it shares their advantages and drawbacks. A comprehensive shakedown of advantages and disadvantages of quantile mapping in comparison to other methods can be found in Maraun et al. (2010) and Themeßl et al. (2011). A step by step example of the multisegment correction procedure is provided in Appendix A of Grillakis et al. (2013).

2.3 Validation of the results

The Klemes (1986) split sample test methodology was adopted for verification. Split sample is the most common type of test used for the validation of model efficiency. The methodology considers two periods of calibration and validation between the observed and modeled data. The first period is used for the calibration, while the second period is used as a pseudo-future period in which the adjusted data are assessed against the observations. A drawback of the split sample test in bias correction validation operations is that the remaining bias of the validation period is a function of the bias correction methodology deficiency and the model deficiency itself to describe the validation period's climate, in aspects that are not intended to be bias corrected. That said, a skillful bias correction method should deal well in that context, as model "democracy" (Knutti, 2010), i.e. the assumption that all model projections are equally possible, is common in CCI studies with little attention to be given to the model selection.

3 Case study area and data

To examine the effect of NM on the bias correction on a timeseries, the Hadley Center Central England Temperature (HadCET - Parker et al., 1992) observational dataset was considered to adjust the simulated output from the earth system model MIROC-ESM-CHEM (Hasumi and Emori, 2004) historical emissions run between 1850 and 2005 for Central England. This particular case study was chosen due to the large observational record (the longest instrumental record of temperature in the world) that is available for Central England, i.e. the triangular area of the United Kingdom enclosed by Lancashire, London and Bristol. Discussion about dataset related uncertainties can be found in Parker et al. (1992) and Parker and Horton (2005). In the specific application and in order to resemble a typical CCI study, data between 1850 and 1899 serve as calibration period, while the rest of the data between 1900 and 2005 is used as pseudo-future period for the validation. Finally, the bias correction results of the two procedures, with (BC-NM) and without (BC) the normalization module, were compared against the validation period observations. 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 timeseries 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.

Furthermore, to expand the methodology assessment in regional scale, the split sample test is adopted to assess the efficiency of the two procedures in a pan European scale. In order to scale up the split sample test, the k-fold cross validation test (Geisser, 1993) is employed. The procedure has been proposed for evaluating the performance of bias correction procedures in (Maraun, 2016). In the k-fold cross validation test, the data is partitioned into k equal sized folds. Of the k folds, one subsample is retained each time as the validation data for testing the model, and the remaining k-1 subsamples are used as calibration data. In a final test, the procedures are applied on a long-term transient climate projection experiment to assess their effect in the long-term attributes of the temperature in a European scale application.

Temperature data from the European division of Coordinated Regional Downscaling Experiment (CORDEX), openly available through the Earth System Grid Federation (ESGF), are considered. Additional information about the Euro - CORDEX domain can be found on the CORDEX web page (http://wcrp-cordex.ipsl.jussieu.fr/). Data from five RCM models (Table 1) with 0.44^{o} spatial resolution and daily time step between 1951-2100 are used. The projection data are considered under the Representative Concentration Pathway (RCP) 8.5, which projects an 8.5 W m^{-2} average increase in the radiative forcing until 2100. The European domain CORDEX simulations have been evaluated for their performance in previous studies (Kotlarski et al., 2014; Prein et al., 2015). The EOBSv12 temperature data was used (Haylock et al., 2008). Discussion about the applicability of EOBS to compare temperature of RCMs control climate simulations can be found in Kyselý and Plavcová (2010). Figure 3 shows the 1951-2005 daily temperature average and standard deviation for the five RCMs of Table 1. The RCMs' mean bias ranges between about -2 ^{o}C and 1 ^{o}C relatively to the EOBS dataset (individual models data are included to the ESM). The positive mean bias in all RCMs is mainly seen in Eastern Europe, while the same areas exhibit negative bias in standard deviation. Some of the bias may however be attributed to the ability of the observational dataset to represent the true temperature (Hofstra et al., 2010).

For the k-fold cross validation, the RCM data between 1951-2010 are split into 6 ten-year sections, comprising a 6-fold, 5 RCM ensemble experiment of Figure 4. Each section is validated once by using the remaining five sections for the calibration. A total number of 30 tests are conducted using each procedure.

For the transient experiment, the RCM data between 1951 and 2100 are considered, using the 1951-2010 as calibration to correct the 1951-2100 data.

4 Results

The results of the split sample test on the central England example are presented in Figure 5. The NM seperates 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 Figure 1 example.

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 can also be 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 still BC-NM 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, interquartile 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 can also be verified by the mean, the standard deviation, and the 10^{th} and 90^{th} 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.

To further inter-compare the effect of each approach in the data variability beyond the inter-annual and the daily basis, we estimate the power spectral density – PSD (Huybers and Curry, 2006) over their daily temperature signals (Figure 7). The marked spectral peaks associate with the annual and 6-month periodicity is and expected result. Focusing on those regions (Figure 7b), 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 7c shows the standard deviation estimated on temperature aggregates between 1 and 10957 days (i.e. 30 years). Figure 7d 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 comparing to the inflated BC and the deflated BC-TREND results.

In Figure 7, the results of the cross validation test of the BC on the Euro – CORDEX data with and without the use of NM are shown, in terms of mean temperature. The mean of the raw temperature data and the observations are respectively equal for their calibration and the validation periods due to the design of the experiment. The bias correction results show that both the correction with and without the NM, appropriately meet the needs in terms of the mean value. The differences between the calibration and validation averages with the corresponding observations show consistently low residuals. A significant difference between the two tests is that the use of the BC-NM increases the residuals due to the exclusion of some parts of the signal from the correction process. Nonetheless, the scale of the residuals is considered below significance in the context of CCI studies, as it ranges only up to 0.035 ^oC. The increased residuals of the NM are the trade off to the preservation of the model long-term climate change signal, in the transient experiment. Potential drawbacks that arise from the residuals existence are discussed later. Figure 8 shows the long-term change in the signal of the mean temperature, for the 10^{th} and 90^{th} percentiles (estimated on annual basis). The trends are estimated by linear least square regression and are expressed in ^oC per century. The use of the NM was superior in preserving the long-term trend relatively to the raw model data in all three cases. Without using the NM module, the distortion in the mean annual temperature trend lies between -0.5 and 0.5 degrees per century, while the distortion in the 10^{th} and 90^{th} percentiles are apparently more profound. Additionally, the northeastern Europe’s 10^{th} and 90^{th} percentiles reveal a widening of the temperature distribution when NM is not used. This widening is attributed to the considerable negative trend in the p10 and the considerable positive p90 trend in the same areas. The magnitude of the distortion is considerable and can potentially lead to CCI overestimation. In contrast, with the use of NM the change in the trend is reduced in most of the Europe’s area.

The impact of NM on the standard deviation is also significant. Figure 9 shows the evolution of the standard deviations of the adjusted daily data for each model, in the cases of raw data and the bias corrected data using the BC and the BC_{NM}. The standard deviation is estimated for each grid point and calendar year, and is averaged across the study domain. The results show that the standard deviation of the adjusted data differ from the respective standard deviations of the raw data, in both adjustment approaches. This is an expected outcome, as raw model data standard deviations differ from the respective observed data standard deviation (Figure 9 d, e). However, the standard deviation differences between BC_{NM} and the raw data (Figure 9 f) is significantly more stable than that the respective differences from BC (Figure 9 g), meaning that the signal of standard deviation is better preserved and does not inflate significantly with time in the former case. Additionally, the variation of the standard deviations time series exhibits lower fluctuations.

5 Discussion

This study focuses on known issues of bias correction that have been well discussed 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 the temperature trend preservation as the model sensitivity is retained in the corrected timeseries. As pointed 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.

The methodology shares similarities to other correction methods found in the literature. Furthermore it exhibits a number of advancements that overpass drawbacks of other trend preserving methodologies. The fundamental idea of the presented method is also identified in the method of Haerter et al. (2011) method that considers two different timescales and performs a cascade correction of temperature. In the present study, a discrimination of annual and daily scales is 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. Comparisons can also be performed with the methodology of Li et al. (2010) who 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 can be considered as an evolution of 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 that of Hempel et al. (2013) which, while preserving the trend and year-to-year variability, marginally improves 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 these advancements, a critical drawback of the presented methodology is that it uses 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’ as the number of the estimated parameters (number of segments) and the added value of the complexity is weighed by an information criterion. Nonetheless, it is highly invasive, which means that in the case that high noise observations were used, it would lead to 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 radically 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 limitation should be taken into consideration in case that the BC-NM was used to correct other types of variables, without forbidding its use on them.

6 Conclusions

This study elaborates the issue of the distortion of the long term statistics in quantile mapping statistical bias correction relatively to the raw model data. An extra processing step is presented, that can be applied along with quantile mapping statistical bias correction techniques. This step, namely NM, splits the original data into two parts, a normalized one that is bias adjusted using quantile mapping, and the residuals part that is added to the former after the bias correction. The methodology is tested and validated from several points of view, leading to some key remarks about its added value. First, it is shown that the use of the NM module results in the long-term temperature trend preservation of the mean temperature change, but also of the trend in the higher and lower percentiles. Furthermore, the examination of the standard deviation temporal evolution shows that it is better retained relatively to the raw data, as the exclusion of the residuals from the correction minimizes the inflation of the variance. Additionally, the inter-annual variability of the raw data is preserved relatively to the compared simpler quantile mapping methods, which comprises an important feature for climate impact studies that involve carbon cycle simulations (Rubino et al., 2016). Another noteworthy feature of the proposed method is that the normalization is performed on an annual basis, hence the projection period results are not affected by the length of the projection period. Nevertheless, it has to be stressed that a range of issues, such as the disruption of the physical consistency of climate variables, the mass/energy balance and the omission of correction feedback mechanisms to other climate variables (Ehret et al., 2012) were not examined in this work, despite the existence of methods that preserve consistency between specific variables (Sippel et al., 2016). As an epilogue, bias correction cannot add further accuracy to the data but rather add usefulness to it, depending on the needs of each application. Nevertheless, it should not be underestimated that this added usefulness may obscure a deterioration of the climate change signal owed to the bias correction.

Acknowledgements

The authors would like to thank Dr. Stefan Hagemann and the anonymous reviewer for their valuable comments and suggestions to improve the quality of the paper. The research leading to these results has received funding from the HELIX project of the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 603864. We acknowledge the World Climate Research Programme's Working Group on Regional Climate, and the Working Group on Coupled Modelling, former coordinating body of CORDEX and responsible panel for CMIP5. We also thank the climate modelling groups (listed in Table 1 of this paper) for producing and making available their model output. Finally, we acknowledge the E-OBS data set from the ENSEMBLES EU-FP6 project (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu).

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

Christensen, J.H., Boberg, F., Christensen, O.B., Lucas-Picher, P., 2008. On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys. Res. Lett. 35, L20709. doi:10.1029/2008GL035694

Daliakopoulos, I.N., Tsanis, I.K., Koutroulis, A.G., Kourgialas, N.N., Varouchakis, E.A., Karatzas, G.P., Ritsema, C.J., 2016. The Threat of Soil Salinity: a European scale review. Sci. Total Environ. 573, 727–739.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., Liebert, J., 2012. *HESS Opinions* "Should we apply bias correction to global and regional climate model data?" Hydrol. Earth Syst. Sci. 16, 3391–3404. doi:10.5194/hess-16-3391-2012

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

Geisser, S., 1993. Predictive inference. CRC press.

Grillakis, M.G., Koutroulis, A.G., Papadimitriou, L.V., Daliakopoulos, I.N., Tsanis, I.K., 2016. Climate-Induced Shifts in Global Soil Temperature Regimes. Soil Sci. 181. doi:10.1097/SS.0000000000000156

Grillakis, M.G., Koutroulis, A.G., Tsanis, I.K., 2013. Multisegment statistical bias correction of daily GCM precipitation output. J. Geophys. Res. Atmos. 118, 3150–3162. doi:10.1002/jgrd.50323

Grillakis, M.G., Koutroulis, A.G., Tsanis, I.K., 2011. Climate change impact on the hydrology of Spencer Creek watershed in Southern Ontario, Canada. J. Hydrol. 409. doi:10.1016/j.jhydrol.2011.06.018

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

Hagemann, S., Chen, C., Clark, D.B., Folwell, S., Gosling, S.N., Haddeland, I., Hanasaki, N., Heinke, J., Ludwig, F., Voss, F., Wiltshire, a. J., 2013. Climate change impact on available water resources obtained using multiple global climate and hydrology models. Earth Syst. Dyn. 4, 129–144. doi:10.5194/esd-4-129-2013

Hansen, J.W., Challinor, A., Ines, A.V.M., Wheeler, T., Moron, V., 2006. Translating climate forecasts into agricultural terms: advances and challenges. Clim. Res. 33, 27–41.

Harding, R.J., Weedon, G.P., van Lanen, H.A.J., Clark, D.B., 2014. The future for global water assessment. J. Hydrol. 518, 186–193. doi:10.1016/j.jhydrol.2014.05.014

Hasumi, H., Emori, S., 2004. K-1 Coupled GCM (MIROC) Description K-1 model developers.

Hawkins, E., Sutton, R., Hawkins, E., Sutton, R., 2016. Connecting Climate Model Projections of Global Temperature Change with the Real World. Bull. Am. Meteorol. Soc. 97, 963–980. doi:10.1175/BAMS-D-14-00154.1

Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J. Geophys. Res. 113, D20119. doi:10.1029/2008JD010201

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

Ines, A.V.M., Hansen, J.W., 2006. Bias correction of daily GCM rainfall for crop simulation studies. Agric. For. Meteorol. 138, 44–53. doi:10.1016/j.agrformet.2006.03.009

Klemes, V., 1986. Operational testing of hydrological simulation models. Hydrol. Sci. J. 31, 13–24. doi:10.1080/02626668609491024

Knutti, R., 2010. The end of model democracy? Clim. Change 102, 395–404. doi:10.1007/s10584-010-9800-2

Kotlarski, S., Keuler, K., Christensen, O.B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., Wulfmeyer, V., 2014. Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci. Model Dev. 7, 1297–1333. doi:10.5194/gmd-7-1297-2014

Koutroulis, A.G., Grillakis, M.G., Daliakopoulos, I.N., Tsanis, I.K., Jacob, D., 2016. Cross sectoral impacts on water availability at +2°C and +3°C for east Mediterranean island states: The case of Crete. J. Hydrol. 532, 16–28. doi:10.1016/j.jhydrol.2015.11.015

Kyselý, J., Plavcová, E., 2010. A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations. J. Geophys. Res. 115, D23118. doi:10.1029/2010JD014123

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

Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., Thiele-Eich, I., 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. 48, RG3003. doi:10.1029/2009RG000314

Maurer, E.P., Pierce, D.W., 2014. Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. Hydrol. Earth Syst. Sci. 18, 915–925. doi:10.5194/hess-18-915-2014

Nikulin, G., Bosshard, T., Yang, W., Bärring, L., Wilcke, R., Vrac, M., Vautard, R., Noel, T., Gutiérrez, J.M., Herrera, S., Others, 2015. Bias Correction Intercomparison Project (BCIP): an introduction and the first results, in: EGU General Assembly Conference Abstracts. p. 2250.

Olsson, T., Jakkila, J., Veijalainen, N., Backman, L., Kaurola, J., Vehviläinen, B., 2015. Impacts of climate change on temperature, precipitation and hydrology in Finland – studies using bias corrected Regional Climate Model data. Hydrol. Earth Syst. Sci. 19, 3217–3238. doi:10.5194/hess-19-3217-2015

Papadimitriou, L. V., Koutroulis, A.G., Grillakis, M.G., Tsanis, I.K., 2017. The effect of GCM biases on global runoff simulations of a land surface model. Hydrol. Earth Syst. Sci. Discuss. 1–43. doi:10.5194/hess-2017-208

Papadimitriou, L. V., Koutroulis, A.G., Grillakis, M.G., Tsanis, I.K., 2015. High-end climate change impact on European water availability and stress: exploring the presence of biases. Hydrol. Earth Syst. Sci. Discuss. 12, 7267–7325. doi:10.5194/hessd-12-7267-2015

Parker, D., Horton, B., 2005. UNCERTAINTIES IN CENTRAL ENGLAND TEMPERATURE 1878–2003 AND SOME IMPROVEMENTS TO THE MAXIMUM AND MINIMUM SERIES. Int. J. Climatol. Int. J. Clim. 25, 1173–1188. doi:10.1002/joc.1190

Parker, D.E., Legg, T.P., Folland, C.K., 1992. A new daily central England temperature series, 1772–1991. Int. J. Climatol. 12, 317–342. doi:10.1002/joc.3370120402

Pierce, D.W., Cayan, D.R., Maurer, E.P., Abatzoglou, J.T., Hegewisch, K.C., Pierce, D.W., Cayan, D.R., Maurer, E.P., Abatzoglou, J.T., Hegewisch, K.C., 2015. Improved Bias

Correction Techniques for Hydrological Simulations of Climate Change*. J. Hydrometeorol. 16, 2421–2442. doi:10.1175/JHM-D-14-0236.1

Prein, A.F., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., Fox Maule, C., van Meijgaard, E., Déqué, M., Nikulin, G., Vautard, R., Colette, A., Kjellström, E., Jacob, D., 2015. Precipitation in the EURO-CORDEX $$0.11^{\circ }$$ 0 . 11 ∘ and $$0.44^{\circ }$$ 0 . 44 ∘ simulations: high resolution, high benefits? Clim. Dyn. 46, 383–412. doi:10.1007/s00382-015-2589-y

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

Sharma, D., Das Gupta, A., Babel, M.S., 2007. Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand. Hydrol. Earth Syst. Sci. 11, 1373–1390. doi:10.5194/hess-11-1373-2007

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

Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. J. Hydrol. 456, 12–29. doi:10.1016/j.jhydrol.2012.05.052

Themeßl, M.J., Gobiet, A., Heinrich, G., 2011. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. Clim. Change 112, 449–468. doi:10.1007/s10584-011-0224-4

List of Figures

Figure 1: The transfer function (TF - heavy black line) between observed (bottom histograms) and modelled (histograms on the left) for the reference period (1981-2010) is used to adjust bias of a 30-year moving window starting from 1981-2010 to 2068-2097. The rightmost plot shows the residual histogram after bias correction. The change in the average correction (red mark) on the TF in comparison to the reference period mean correction (square) is shown. The animated version provided in the supplemental information shows the temporal evolution of the bias as the 30-year time window moves on the projection data. Data were obtained from ICHEC-EC-EARTH r12i1p1 SMHI-RCA4_v1 RCM model of Euro-CORDEX experiment (0.11 degrees resolution) simulation under the representative concentration pathway of RCP85, for the location Chania International Airport (lon=24.08, lat=35.54). Observational data were obtained from the E-OBS v14 dataset (Haylock et al., 2008) of 0.25 degrees spatial resolution.

Figure 2: MSBC methodology on temperature correction using linear functions (borrowed from Grillakis et al., (2013); modified) in one of the data segments.

Figure 3: Mean temperature (upper) and standard deviation (lower) for EOBS, RCM model ensemble (ENS) and for their difference (model - obs) (DIFF) for the reference period 1951-2005.

Figure 4: The 6-fold cross validation scheme with the calibration (C) and the validation (V) periods of each fold. Each experiment (Exp) was replicated for all five RCM models.

Figure 5: 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).

Figure 6: 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).

Figure 7: Mean surface temperature of the cross validation test. Panels (a) and (b) show the ensemble mean of the 5 raw models data and the EOBS respectively, while panel (c) their difference. Panels (d) and (e) show the ensemble mean remaining bias of the 5 RCM models after the correction with and without the NM module respectively, for the calibration periods' data. Panels (f) and (g) are the same as (d) and € but for the validation period data.

Figure 8: Ensemble long-term linear trend of the 5 RCM models' data. The trend is estimated on the mean temperature (top) and the 10th (mid) and 90th (bottom) percentiles on an annual basis. The change in the corrected data trend relatively to the raw data trend is provided for the BC (middle panels) and the BCNM data (right panels). All values are expressed as degrees per century [^{oC}/100 y].

Figure 9: Average of standard deviations for the study domain, for the raw data (a), the BC (b) and the BC-NM (c) for the different models and the observations, in annual basis. Differences between the raw and the bias corrected standard deviations are shown in (d) and (e). Plots (f) and (g) correspond to the same data as (d) and (e), but normalized for their 1951-2005 mean.

List of Tables

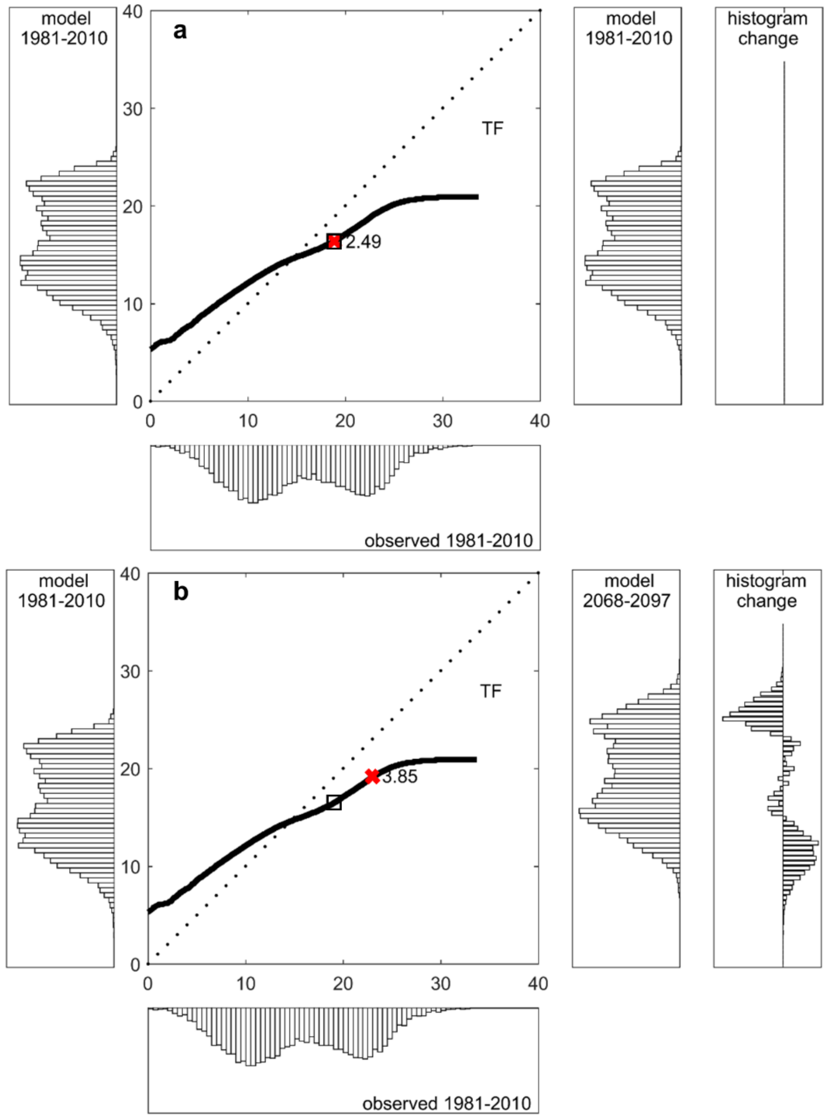

Figure 1: The transfer function (TF - heavy black line) between observed (bottom histograms) and modelled (histograms on the left) for the reference period (1981-2010) is used to adjust bias of a 30-year moving window starting from 1981-2010 to 2068-2097. The rightmost plot shows the residual histogram after bias correction. The change in the average correction (red mark) on the TF in comparison to the reference period mean correction (square) is shown. The animated version provided in the supplemental information shows the temporal evolution of the bias as the 30-year time window moves on the projection data. Data were obtained from ICHEC-EC-EARTH r12i1p1 SMHI-RCA4_v1 RCM model of Euro-CORDEX experiment (0.11 degrees resolution) simulation under the representative concentration pathway of RCP85, for the location Chania International Airport (lon=24.08, lat=35.54). Observational data were obtained from the E-OBS v14 dataset (Haylock et al., 2008) of 0.25 degrees spatial resolution.

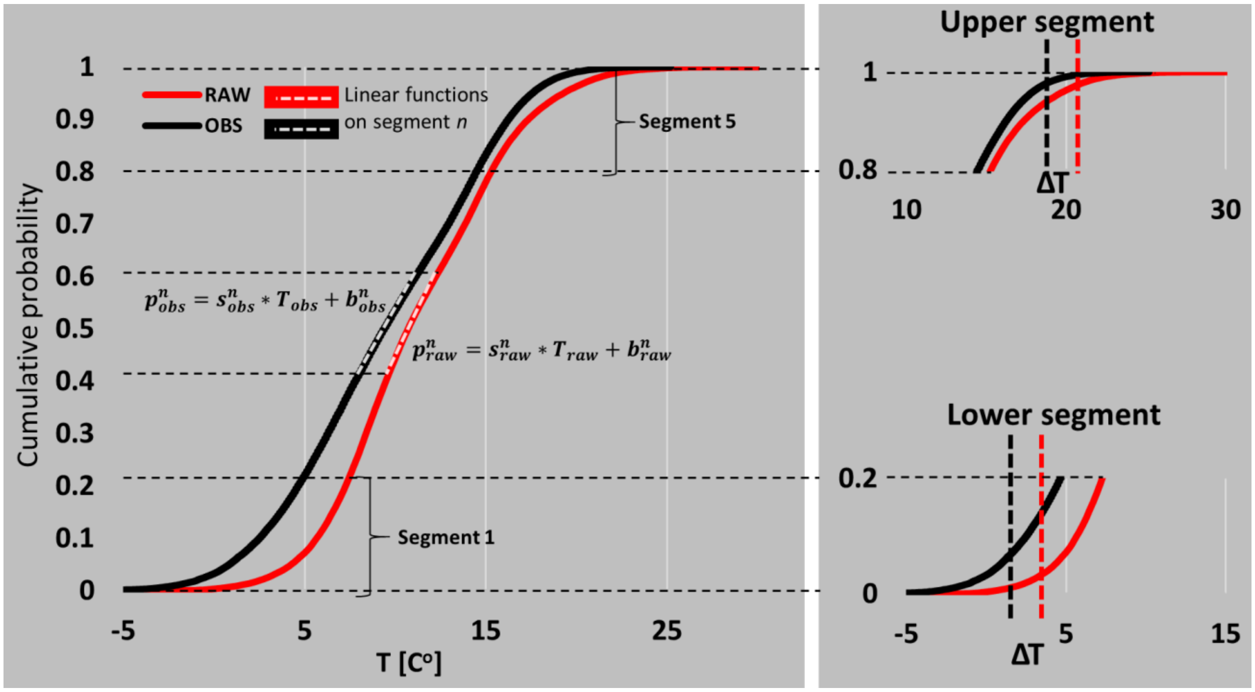

Figure 2: MSBC methodology on temperature correction using linear functions (borrowed from Grillakis et al., (2013); modified) in one of the data segments.

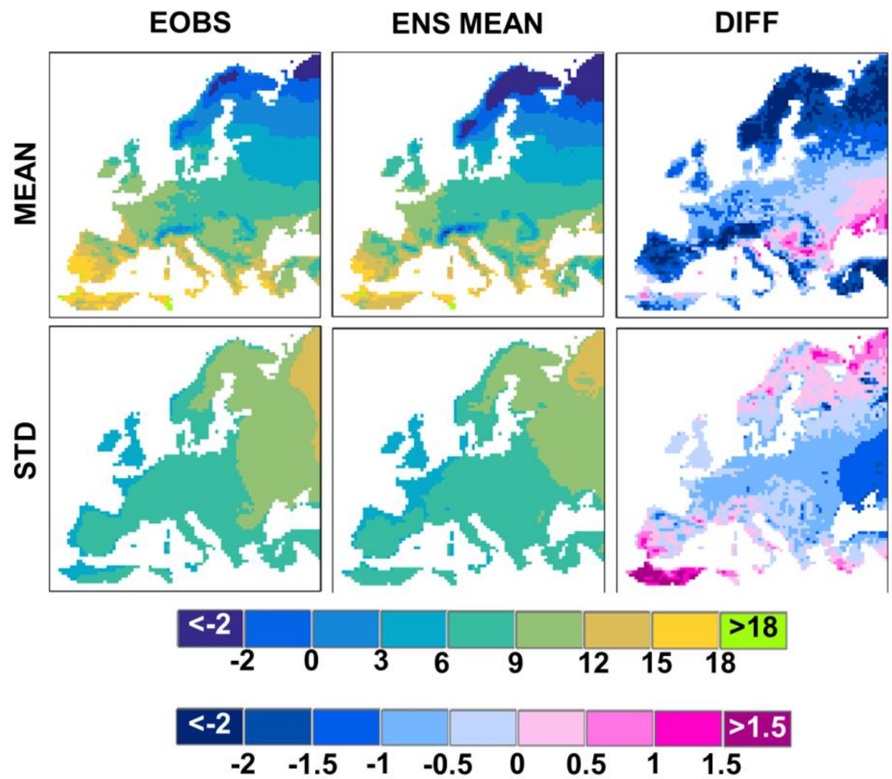

Figure 3: Mean temperature (upper) and standard deviation (lower) for EOBS, RCM model ensemble (ENS) and for their difference (model - obs) (DIFF) for the reference period 1951-2005.

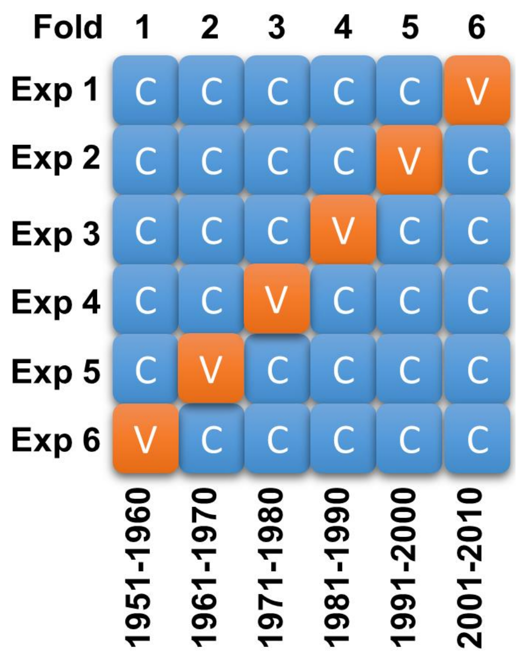

Figure 4: The 6-fold cross validation scheme with the calibration (C) and the validation (V) periods of each fold. Each experiment (Exp) was replicated for all five RCM models.

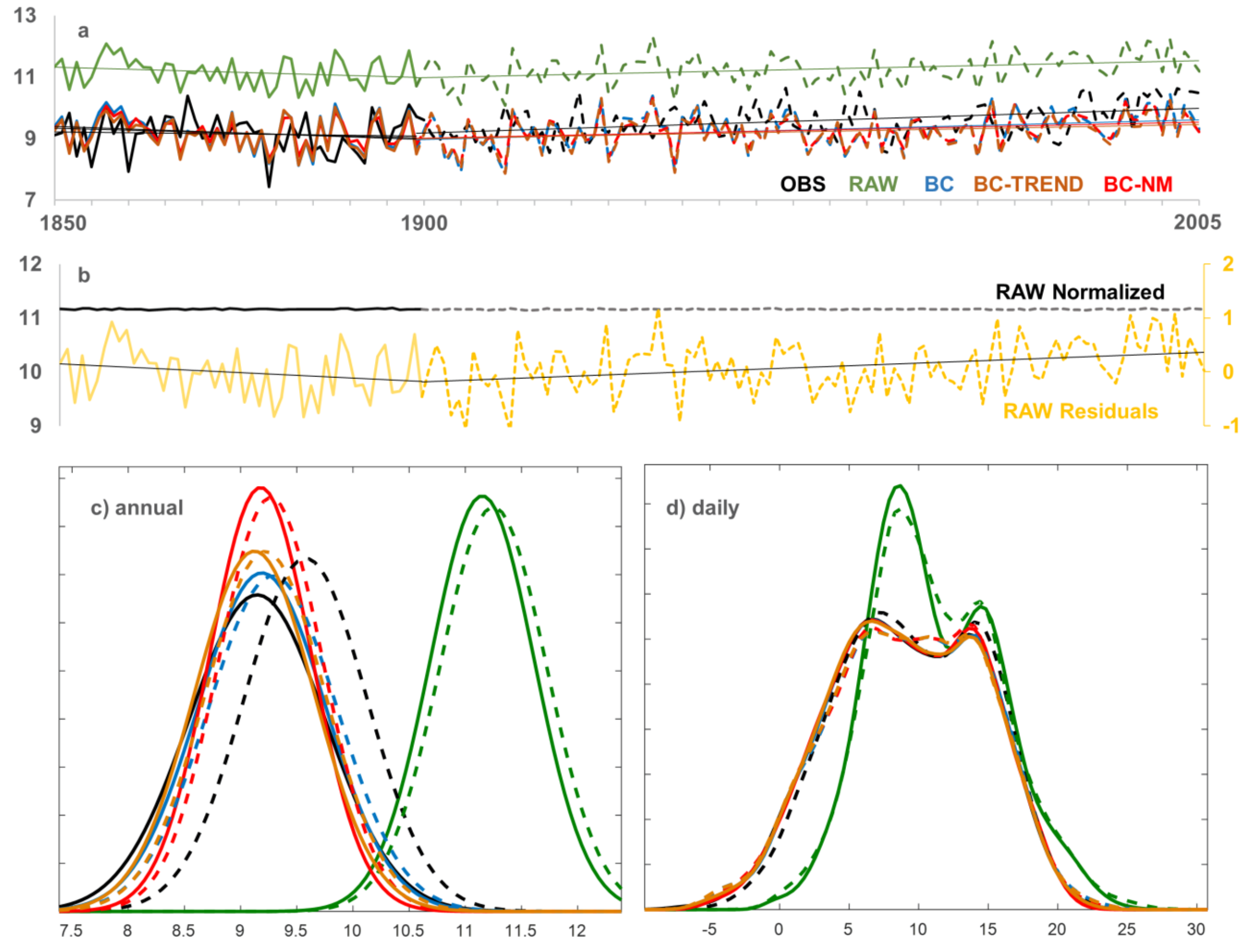

Figure 5: a) annual average temperature of raw model, observations and the bias corrected with, without the NM data and following the BC-TREND approach, for the calibration period 1850 – 1899 (solid lines) and the validation period 1900-2005 (dashed lines). b) Annual averages of the normalized and the residuals of the raw temperature. Probability densities of annual (c) and of daily means (d).

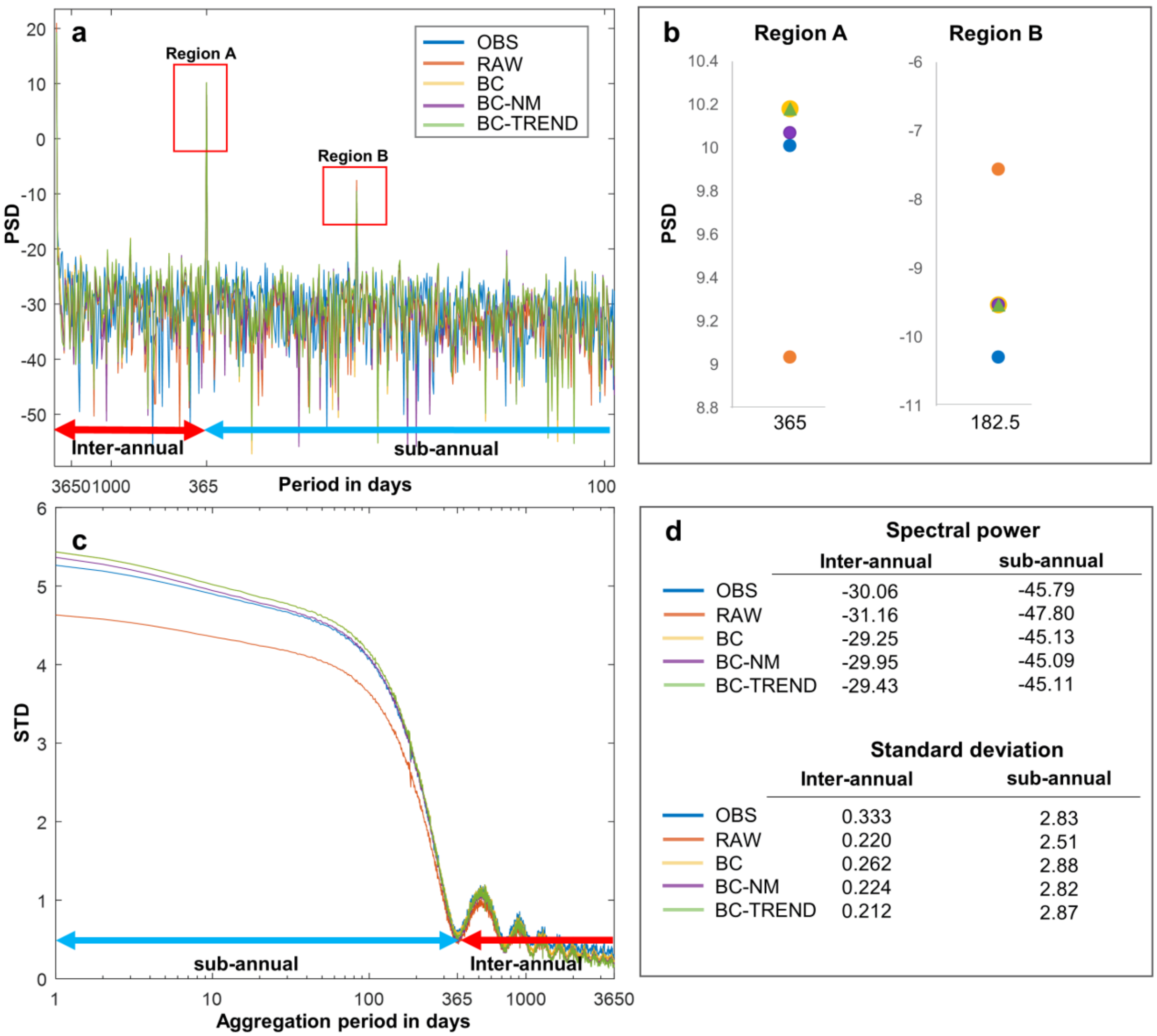

Figure 6: Power spectral density of temperature (a) and high power regions of annual and half year periods (b). Standard deviation of temperature aggregates between 1 and 10957 days (horizontal axis visible between 1 day and 10 years) in (c). In (d), the inter-annual and sub-annual periods average (denoted with red and cyan arrows respectively) spectral power (a) and standard deviation (c).

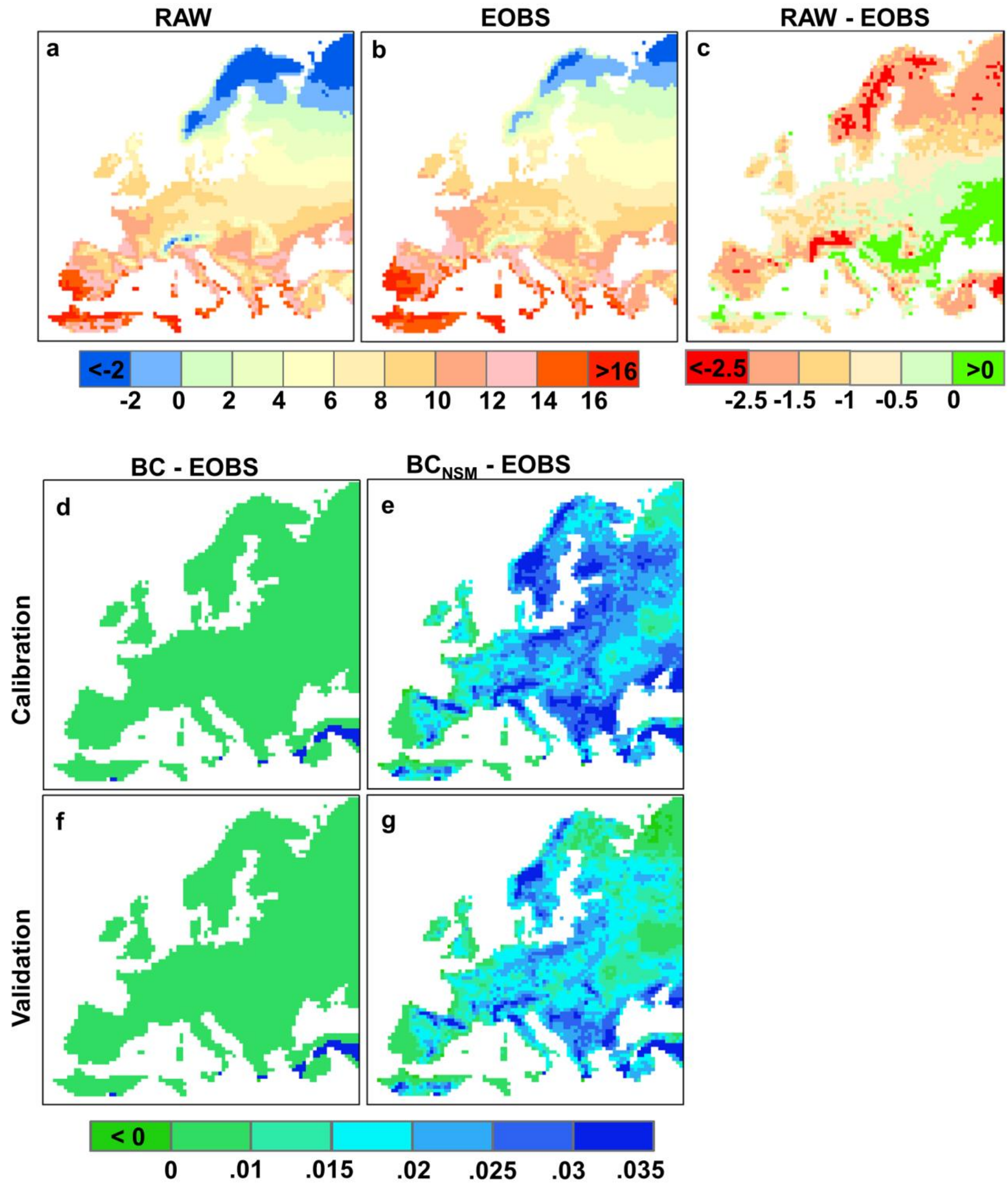

Figure 7: Mean surface temperature of the cross validation test. Panels (a) and (b) show the ensemble mean of the 5 raw models data and the EOBS respectively, while panel (c) their difference. Panels (d) and (e) show the ensemble mean remaining bias of the 5 RCM models after the correction with and without the NM module respectively, for the calibration periods' data. Panels (f) and (g) are the same as (d) and € but for the validation period data.

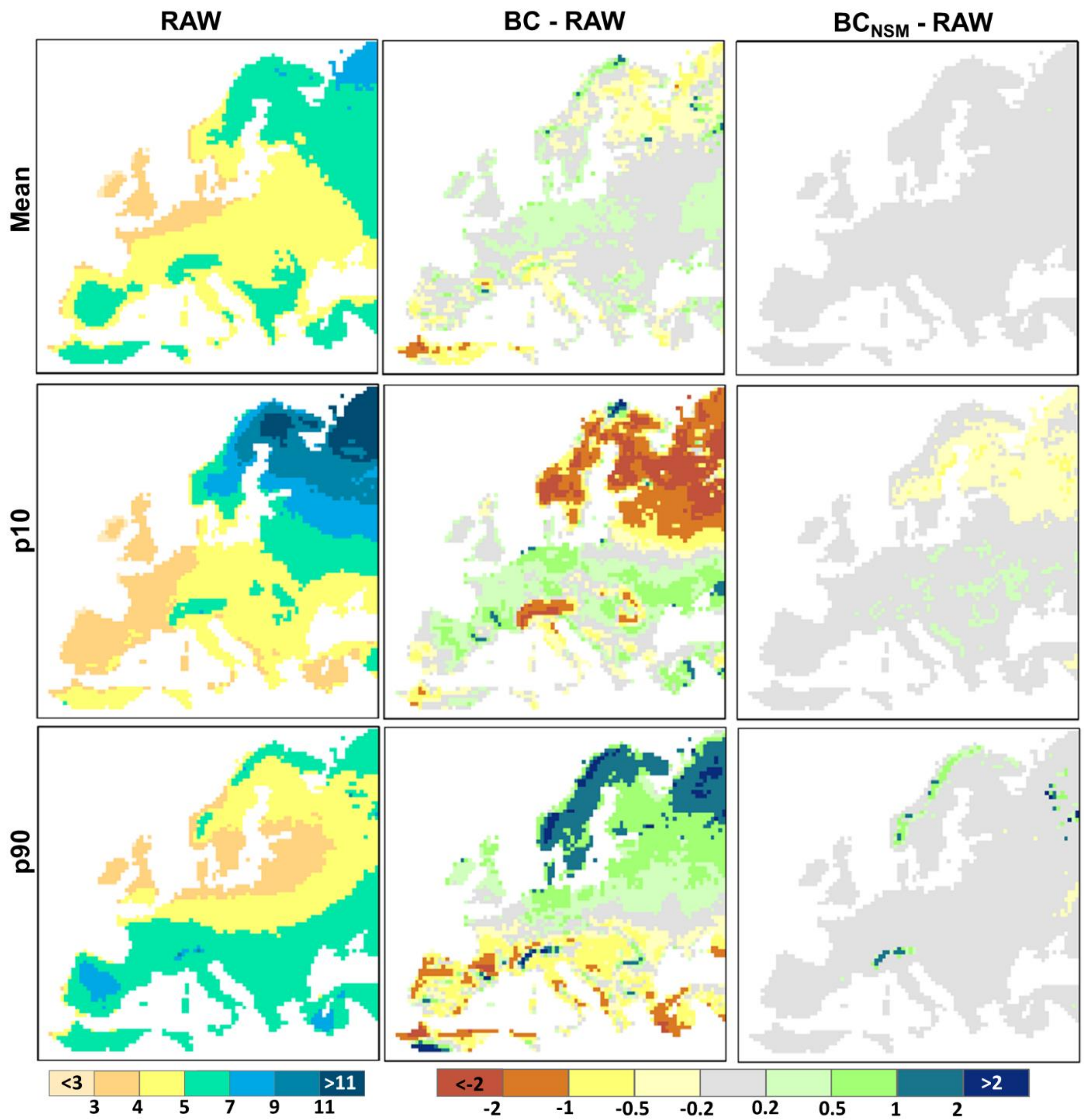

Figure 8: Ensemble long-term linear trend of the 5 RCM models' data. The trend is estimated on the mean temperature (top) and the 10th (mid) and 90th (bottom) percentiles on an annual basis. The change in the corrected data trend relatively to the raw data trend is provided for the BC (middle panels) and the BCNM data (right panels). All values are expressed as degrees per century [ºC/100 y].

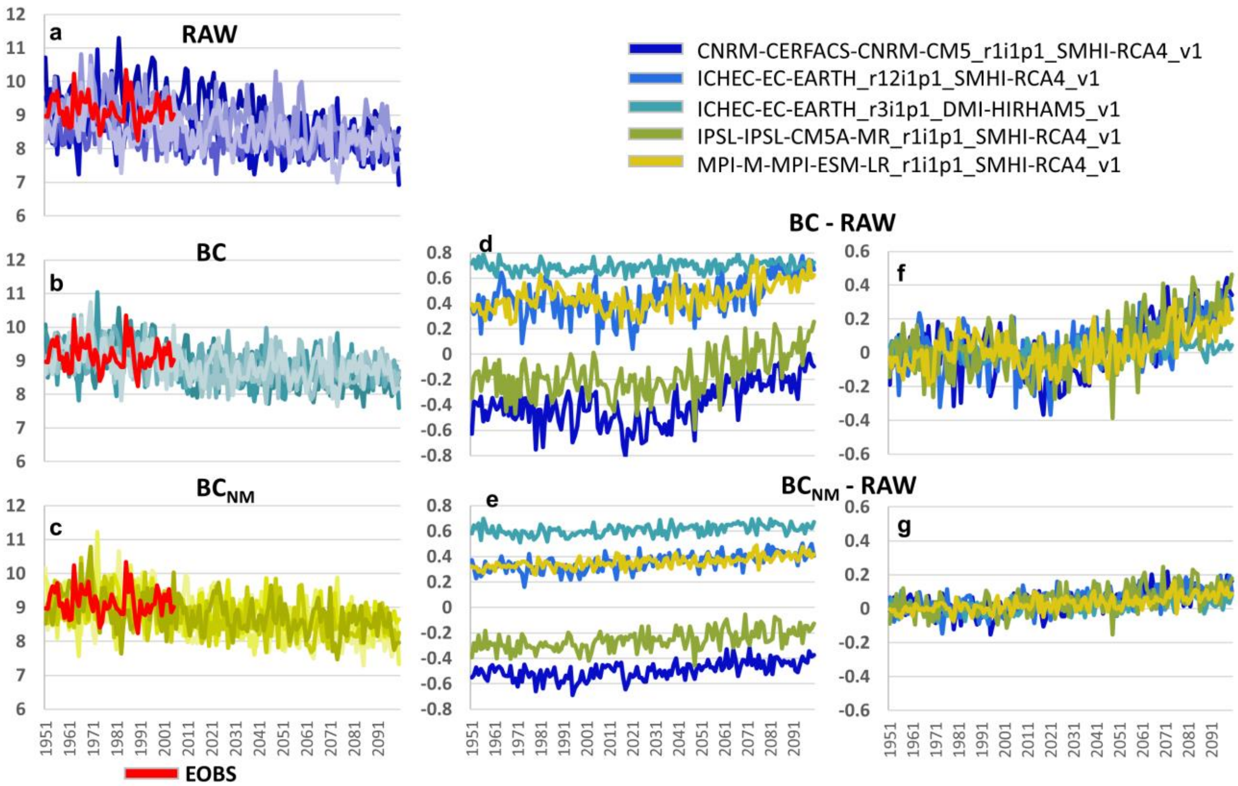

Figure 9: Average of standard deviations for the study domain, for the raw data (a), the BC (b) and the BC-NM (c) for the different models and the observations, in annual basis. Differences between the raw and the bias corrected standard deviations are shown in (d) and (e). Plots (f) and (g) correspond to the same data as (d) and (e), but normalized for their 1951-2005 mean.

Table 1: RCM models used in this experiment.

#	{GCM}_{realization}_{RCM}
1	CNRM-CM5_r1i1p1_SMHI-RCA4_v1
2	EC-EARTH_r12i1p1_SMHI-RCA4_v1
3	EC-EARTH_r3i1p1_DMI-HIRHAM5_v1
4	IPSL-CM5A-MR_r1i1p1_SMHI-RCA4_v1
5	MPI-ESM-LR_r1i1p1_SMHI-RCA4_v1

Table 2: 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, p*n *for the* nth *quantile and IQR for the interquartile range.

Parameter	RAW	Normalized	Residuals	OBS	BC	BC$_{NM}$	BC$_{TREND}$
Mean [^{o}C]	11.2	11.2	0.0	9.1	9.2	9.2	9.1
SD [^{o}C]	4.5	4.6	0.9	5.3	5.3	5.3	5.3
p10 [^{o}C]	5.7	5.7	-0.9	2.1	2.2	2.2	2.1
p90 [^{o}C^{o}]	17.4	17.2	1.0	16.3	16.3	16.2	16.2
Slope [^{o}C/10yr]*	-0.067	0.000	-0.067	-0.026	-0.086	-0.065	-0.061
SD [^{o}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
Mean [^{o}C]	11.3	11.2	0.1	9.6	9.3	9.3	9.2
SD [^{o}C]	4.7	4.6	0.9	5.2	5.5	5.4	5.5
p10 [^{o}C]	5.6	5.7	-0.9	2.7	2.0	2.0	1.9
p90 [^{o}C]	17.4	17.2	1.0	16.3	16.3	16.2	16.2
Slope [^{o}C/10yr]*	0.052	0.000	0.051	0.076	0.062	0.051	0.044
SD [^{o}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