Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Addressing the assumption of stationarity in statistical bias correction of 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 has become a standard practice in Climate Change Impact (CCI) studies. While various methodologies have been developed, their majority assumes that the statistical characteristics of the biases between the modeled data and the observations remain unchanged in time. However, it is well known that this assumption of stationarity cannot stand in the context of a climate. Here, a method to overcome the assumption of stationarity and its drawbacks is presented. The method is presented as a pre-post processing procedure that can potentially be applied with different bias correction methods. The methodology separates the stationary and the non-stationary components of a time series, in order to adjust the biases only for the former and preserve intact the signal of the later. The results show that the adoption of this method prevents the distortion and allows for the preservation of the originally modeled long-term signal in the mean, the standard deviation, but also the higher and lower percentiles of the climate variable. Daily temperature time series obtained from five Euro CORDEX RCM models are used to illustrate the improvements of this method.

Keywords: temperature non-stationarity, trend preservation, statistical bias correction,

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

1 Introduction

Climate model output consist 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 observational data 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 outcomes have been reported to become unrealistic without a prior adjustment of climate forcing biases (Hansen et al., 2006; Harding et al., 2014; Sharma et al., 2007). These biases may be attributed to a number of sources such as the imperfect representation of the physical processes within the model code and the coarse spatial resolution of output that do not permit the accurate representation of small-scale processes. Furthermore, climate model tuning for global projections focuses on essential elements such as the adequate representation of feedbacks between processes and hence the realistic depiction of a variable, such as temperature, against observations can be 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. This is typically accomplished with the use of a Transfer Function (TF) which minimizes the difference between the cumulative distribution 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. Thus the statistical outcomes of an impact model run using observational data are 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., 2013; Ines and

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Hansen, 2006; Olsson et al., 2015). Under the assumption of climate stationarity, the TF is then applied to data beyond the time-frame of the observations.

A process is characterized as stationary when the probability distribution function (PDF) of its state fluctuates within an unchanging and time invariant interval, or in a looser definition, when the process retains its mean, variance and auto covariance structure constant (Challis and Kitney, 1990). Ergo is the definition of stationary “bias” which refers to the time independent component of the difference between the modeled and the observed values (Haerter et al., 2011). Stationarity has been observed in long observational hydrological time series (Koutsoyiannis, 2002; Koutsoyiannis and Montanari, 2015; Lins and Cohn, 2011; Matalas, 2012). Nevertheless, non-stationarity is “unequivocal and unconditional” to all natural systems (Lins and Cohn, 2011), hence considering long-term climate or other processes involving abrupt system changes as stationary is certainly flowed. Therefore, the stationarity-dependent extrapolation of the TF is often regarded as a leap of faith and may lead to a false certainty about the robustness of the adjusted projection.

As the most obvious effect, the assumption of stationarity in bias correction adds another level of uncertainty in the output (Maraun, 2012). At a more practical level, it may also lead to other unwanted effects, such as changes in the original model derived long-term trend or other higher moments of the climate variable statistics that eventually distort the long-term signal of the climate variable. As an example, Olsson et al. (2015) showed that their distribution based scaling (DBS) bias correction methodology might alter the long-term temperature trends. They attribute the phenomenon in the severity of the biases in the mean or the standard deviation between the uncorrected temperatures and the observations. Similar conclusions were drawn by Hagemann et al. (2011) who showed that a fixed bias correction can alter the climate change signal for specific locations and seasons and concluded that climate parameters require variable adjustment as the distribution between their upper and lower limit changes in time. In their work, Hempel et al. (2013) attempt to provide a solution to the trend changing issue, by preserving the absolute changes in monthly temperature, and relative changes in monthly values of precipitation. The obvious conceptual drawback of this approach is that non-stationarity does not always coincide with a deterministic trend component (Lins, 2012).

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics Discussions

While climate may be considered stationary for the studied time scales (i.e. a few decades), it is uncertain whether this stationarity holds between intervals selected for bias correction. Narrow time windows may obscure information about a recurring periodicity thus attributing non-stationarity to the data. Instead, in a time window sufficiently larger than underlying periodicities, the same signal may reveal to be part of a stationary process. A representative example about the role of timescales in non-stationarity is illustrated by Koutsoyiannis and Montanari (2015). In the context of climate change, the concept of climate non-stationarity is discussed beyond the profound daily and seasonal periodicities. Milly et al., (2008) denote as stationary a climate process response with time-invariant probability distribution in one year periodic. This is a reasonable assumption in the context of CCI studies, as seasonal is the most well defined periodicity of a climate system at least in the time scale of a few decades. Reversely, when the year-to-year distribution of a climate process response changes, the variable can be described as non-stationary, thus leading to the unwanted effects of quantile mapping based bias correction methods described in Olsson et al. (2015). As the TF of the bias correction is estimated between the reference period observations and climate model data, it indicates the different magnitude of correction for the different parts of the probability distribution function (PDF). Considering that the climate data PDF is actually time-dependent (i.e. non stationary), the stationary TF gradually changes its response on the climate data, providing unequal bias correction in different periods as Hagemann et al. (2011) also notice.

Figure 1 presents an indicative example where temperature data[1] have a mean bias of 2.02 ^{o}C in the reference period (Figure 1a). The average bias is expressed by the average horizontal distance between the TF and the bisector of the central plot. The histogram on the left 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. In the rightmost histogram, the difference between the reference period and the moving 30-year period is estimated. The red mark

[1] The Figure 1 data were obtained from ICHEC-EC-EARTH r12i1p1 SMHI-RCA4_v1 Euro-CORDEX simulation under the RCP85, for the location Chania International Airport (lon=24.08 lat=35.54).

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 2068-2097 period reaches 3.48 °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 where the TF distance from the bisector is variable, this side 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.

In this study, we present a methodology to account for the non-stationarity of the climate parameters. The methodology takes the form of a pre- and post-processing module that can be applied along with different statistical bias correction methodologies. To account for the non-stationarities, the method separates the stationary from the non-stationary components of a time series before it is bias adjusted. Bias correction is then applied to the stationary-only components of the time series. Finally, the non-stationary components are again merged to the adjusted component to form a single corrected time series. In order to use and test the module, we employ a generalized version of the Multi-segment Statistical Bias Correction (MSBC) methodology (Grillakis et al., 2013) that can be used in a wider set of climate parameters.

2 Methods

2.1 Terminology

As non-stationary components are identified the statistical deviations of each year's data comparing to the average reference period data distribution. Specifically, the differences between the CDF of each year's model climate data comparing to the CDF of the entire reference period of the model data are identified as the non-stationary component. Hence, the first step of the procedure is to normalize each year's data individually, against the average modeled reference period climatology. 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

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

estimated by transferring each year 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 is the non-stationary component S_i^{NS} of the time series (Eq. 2).

$$S_R^{NS} = S_R - S_R^n \qquad \text{Eq. (2)}$$

The original model data S_i can be reconstructed by adding the non-stationary components S_i^{NS} to the normalized data S_i^n as in Eq. 3.

$$S_i = S_i^{NS} + S_i \qquad \text{Eq. (3)}$$

The non-stationary components (S_i^{NS}) contain the random part of the climate signal, as well as the potential long-term changes in the statistics. After the separation, the stationarized climate model data are statistically bias corrected following a suitable methodology. The stationarized components of the modeled data are bias adjusted disregarding the stationarity assumption, as the data to be corrected are stationary. The non-stationary components (S_i^{NS}) are preserved in order later to be added again to the bias corrected time series. We refer to the described method as non-stationarity module (NSM) to hereafter lighten the nomenclature of the paper.

2.2 Bias correction

The NSM is applied along with a modification of the MSBC algorithm proposed by 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 novelty of the method is the partitioning of the data CDF space into discrete segments and the individual quantile mapping correction in each segment, thus achieving better fit of the parametric equations on the data and better correction especially on the CDF edges. The optimal number of the segments is estimated by Schwarz Bayesian Information Criterion (SBIC) to balance between complexity and performance.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics
Open Access
Discussions

Here the methodology is modified to use linear functions instead of the gamma functions used in the original methodology. Moreover, the upper and lower edge segments are explicitly corrected using only the mean difference between the reference period of the model data and the observations. This choice costs to the methodology, the remainder of some bias in the corrected data. However, it 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., 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).

3 Case study

To examine the effect of NSM on the bias correction, 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[2] historical emissions run between 1850 and 2005 for Central England. The Klemes (1986) split sample test methodology was adopted here for verification. 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. To resemble a typical CCI study, the available 50 years of data between 1850 and 1899 served as calibration period, while the rest of the data between 1900 and 2005 was used as pseudo-future period for the validation. The bias correction results of the two procedures, with (BC-NSM) and without (BC) the non-stationarity module, were then compared against the observations. Figure 2a demonstrates the division of the raw data performed by the NSM into non-stationary components and normalized raw data in annual aggregates. The summation of the two time series can reconstruct the initial raw data time series. The normalized time series do not exhibit any trend or significant fluctuation in the annual

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

aggregates, since the normalization is performed at annual basis, while the long-term trend and the variability is contained in the residual time series.

In Figure 2b, annual aggregates obtained via the above two procedures are compared to the raw data and the observations. Results show that both procedures adjust the raw data to better fit the observations in the calibration period 1850-1899. In the validation period, both procedures produce similar results, but the BC-NSM long-term linear trend is slightly lower than that of the BC results. While the latter slope is closer to the observations' linear trend, the former is closer to the raw data trend (Table 1). The persistence of the long-term trend is a desirable characteristic of the NSM procedure as the GCM long-term moments were not distorted by the correction. However, the wider deviation of the BC-NSM trend relatively to the BC depicts the skill of the GCM to simulate the observations' respective trend. Figure 2c shows that the BC-NSM output resemble the raw data histograms in shape, but are shifted in their mean towards the observations. This consists an idealized behavior for the adjusted data, as the distribution of the annual temperature averages are retained after the correction. Similar results generated on daily data (Figure 2d) show that both procedures adjust the calibration and validation histograms in the same degree towards the observations. This can also be verified by the mean, the standard deviation and the 5th and 95th percentile of the daily data (Table 1). An early concluding remark about the NSM is that it improved the long-term statistics of the adjusted data towards the climate model signal, without sacrificing the daily scale quality of the correction.

The split sample test is also adopted to assess the efficiency of the procedures in a European scale application. Split sample is the most common type of test used for the validation of model efficiency. 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. In order to scale up the split sample test, the k-fold cross validation test (Geisser, 1993)

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

is employed. In 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 transcend climate projection experiment to assess their effect in the long-term attributes of the temperature in a European scale application.

3.1 Data

Temperature data from the European division of Coordinated Regional Downscaling Experiment (CORDEX), openly available through the Earth System Grid Federation (ESGF), are used to evaluate the presented procedures. Data from five RCM models (Table 2) with 0.44° 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). Figure 3 shows the 1951-2005 daily temperature average and standard deviation for the five RCMs of Table 2. The RCMs' mean bias ranges between about -2 °C and 1 °C relatively to the EOBS dataset. 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 is however attributed to the ability of the observational dataset to represent the true temperature. Discussion about the applicability of EOBS to compare temperature of RCMs control climate simulations can be found in Kyselý and Plavcová (2010). For the purposes of this work, the EOBS is assumed to accurately represent the past temperatures over Europe.

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 (Figure 4). Each section is validated once by using the rest five sections for the calibration. A total number of 30 tests are conducted using each procedure.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics
Discussions
Open Access

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 and discussion

Figure 5 shows the mean surface temperature of the cross validation test. 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 correction results show that both bias corrections with and without the NSM, appropriately meet the needs of the correction. 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 NSM increases the residuals due to the exclusion of the non-stationary components 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 ^{o}C. The increased residuals of the NSM are the trade off to the preservation of the model long-term climate change signal, in the transient experiment. Figure 6 presents the long-term change in the signal of the mean temperature, for the 10th and 90th percentiles (in annual aggregates). The trends are estimated by a simple linear least square fit and are expressed in ^{o}C per century. The use of the NSM profoundly better preserved the long-term trend relatively to the raw model data in all three cases. Without using the NSM module, the distortion in the mean annual temperature trend lies between -0.5 and 0.5 degrees per century, while the distortion in the 10th and 90th percentiles are apparently more profound. Additionally, the northeastern Europe's 10th and 90th percentiles reveal a widening of the temperature distribution when NSM is not used. The 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 NSM the change in the trend is considerably reduced in most of the Europe's area.

The impact of NSM on the standard deviation is also significant. Figure 7 shows the evolution of the standard deviation for each model, in the cases of raw data and the bias corrected data using the BC and the BC_{NSM}. The standard deviation is estimated for each

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

grid point and calendar year, and is averaged across the study domain. Results show that standard deviations 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 7 d, e). However, the standard deviation differences between BC_{NSM} and the raw data (Figure 7 f) is significantly more stable than that the respective differences from BC (Figure 7 g), meaning that the signal of standard deviation is better preserved and does not inflate with time in the former case. Additionally, the variation of the standard deviations time series exhibits lower fluctuations.

5 Conclusions

This study elaborates on two correlated issues of statistical bias correction; the assumption of stationarity and the distortion of long-term trends. These challenges are addressed by a pre/post processing module (NSM) that can be applied along with statistical bias correction techniques. The results are validated from several points of view. First, it is shown that the use of the NSM module resulted in the long-term temperature trend preservation in the mean annual aggregates of the temperature, but also in the aggregates of the higher and lower percentiles. Furthermore, the examination of the standard deviation temporal evolution show that is better retained relatively to the raw data. The corrected variable retains some remaining biases in the control period, which however are low to significantly affect CCI study outcomes.

The main advantage of the proposed method compared to other trend preserving methods (e.g. Hempel et al., 2013), is that the preservation of the long-term mean trend is not the objective but rather an ineluctable consequence of excluding the non-stationarity components from the correction process. 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) have not been addressed. Beyond the benefits of statistical bias correction methods, these constrains remain unsurpassable challenges that can only be resolved within a climate model itself. Finally, one should bare in mind that climate data quality prime driver is the climate model skillfulness itself. Statistical post processing methods like bias correction cannot add new information to the data but rather add usefulness to it, depending on the needs of each application.

6 References

Challis, R.E., Kitney, R.I., 1990. Biomedical signal processing (in four parts). Med. Biol. Eng. Comput. 28, 509–524. doi:10.1007/BF02442601

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. CATENA.

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

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, 264–272.

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

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., Haerter, J.O., Heinke, J., Gerten, D., Piani, C., Hagemann, S., Chen, C., Haerter, J.O., Heinke, J., Gerten, D., Piani, C., 2011. Impact of a Statistical

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models. http://dx.doi.org/10.1175/2011JHM1336.1.

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

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

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

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

Koutsoyiannis, D., 2002. The Hurst phenomenon and fractional Gaussian noise made easy. Hydrol. Sci. J. 47, 573–595. doi:10.1080/02626660209492961

Koutsoyiannis, D., Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. Hydrol. Sci. J. 60, 1174–1183. doi:10.1080/02626667.2014.959959

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

Lins, H.F., 2012. A Note on Stationarity and Nonstationarity. WMOCHy-AWG.

Lins, H.F., Cohn, T.A., 2011. Stationarity: Wanted Dead or Alive?1. JAWRA J. Am. Water Resour. Assoc. 47, 475–480. doi:10.1111/j.1752-1688.2011.00542.x

Maraun, D., 2012. Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. Geophys. Res. Lett. 39, n/a-n/a. doi:10.1029/2012GL051210

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

Matalas, N.C., 2012. Comment on the Announced Death of Stationarity. J. Water Resour. Plan. Manag. 138, 311–312. doi:10.1061/(ASCE)WR.1943-5452.0000215

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity Is Dead: Whither Water Management? Science (80-.). 319.

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.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics
Open Access
Discussions

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., 2016. High-end climate change impact on European runoff and low flows – exploring the effects of forcing biases. Hydrol. Earth Syst. Sci. 20, 1785–1808. doi:10.5194/hess-20-1785-2016

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

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

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

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

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics Discussions

List of Figures

Figure 1: The transfer function (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.

Figure 2: a) Annual averages of temperature of the raw model data, the observations and the bias correction with and without the NSM, for the calibration period 1850 – 1899 (solid lines) and the validation period 1900-2005 (dashed). Probability densities of annual (c) and of daily means (d).

Figure 3: The EOBS and the RCM models' mean and standard deviation for the reference period 1951-2005. Differences (DIFF) indicate the [RCM statistic – EOBS statistic].

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: Mean surface temperature of the cross validation test.

Figure 6: Long-term linear trend [^{o}C/100 y] in the raw data (left) for the mean annual temperature (top) and the 10th and 90th (p10 and p90) percentiles the change in the trend relatively to the raw data trend is provided for the BC (middle) and the BC_{NSM} data (right). All values are expressed as degrees per century.

Figure 7: Average of standard deviations for the study domain, for the raw data (a), the BC (b) and the BC_{NSM} (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.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

List of Tables

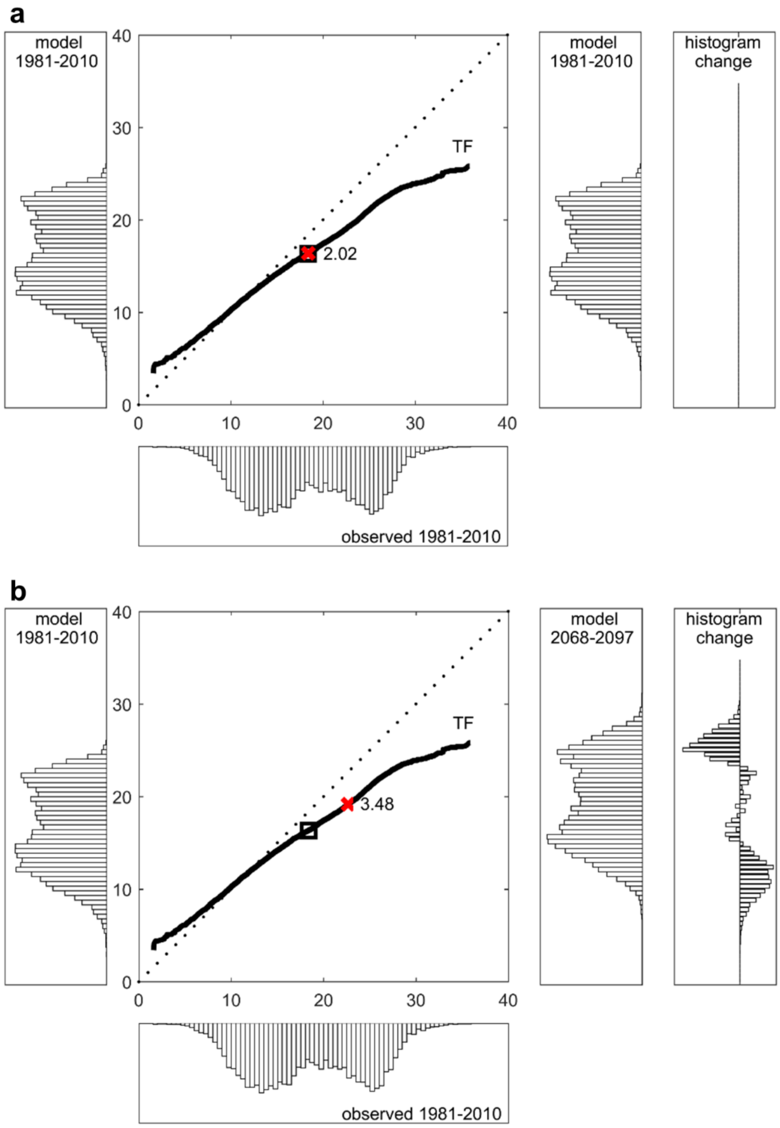


Figure 1: The transfer function (heavy black line) between observed (bottom histograms) and modelled (histograms on the left) for the reference period (1981-2010) is used to adjust bias of a 30-year moving window starting from 1981-2010 to 2068-2097. The rightmost plot shows the residual histogram after 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 material shows the temporal evolution of the bias as the 30-year time window moves on the projection data.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics
Discussions

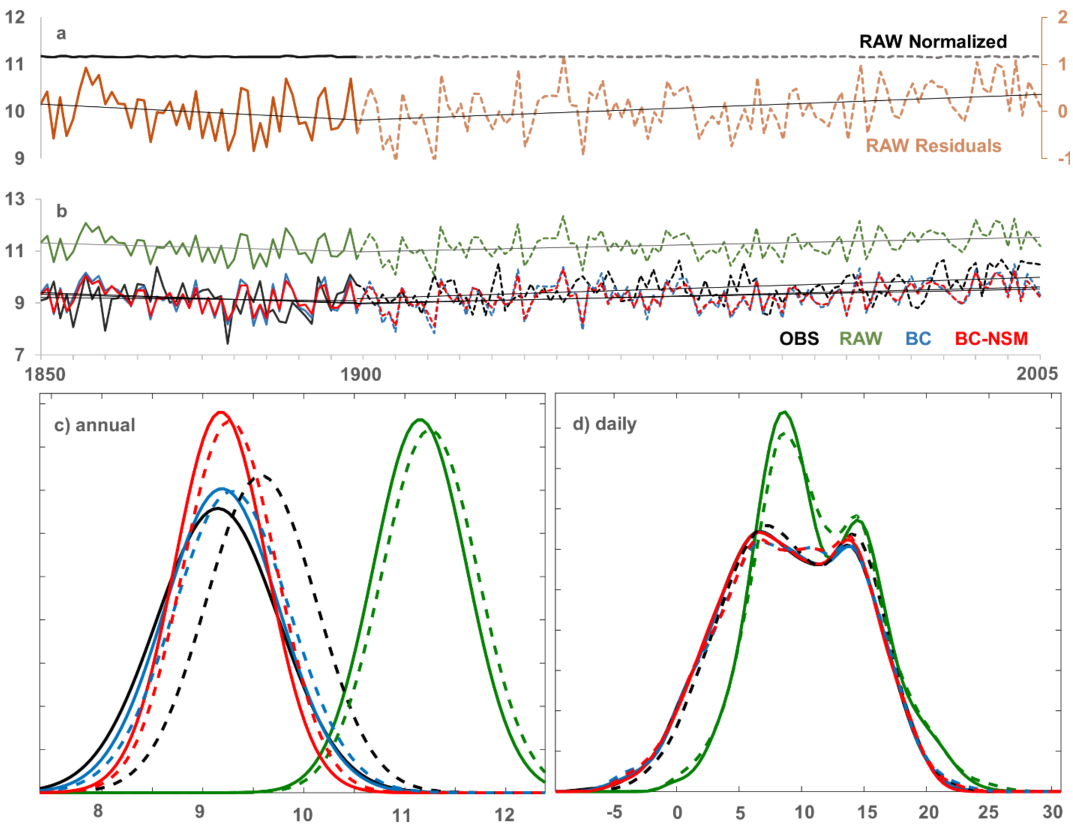
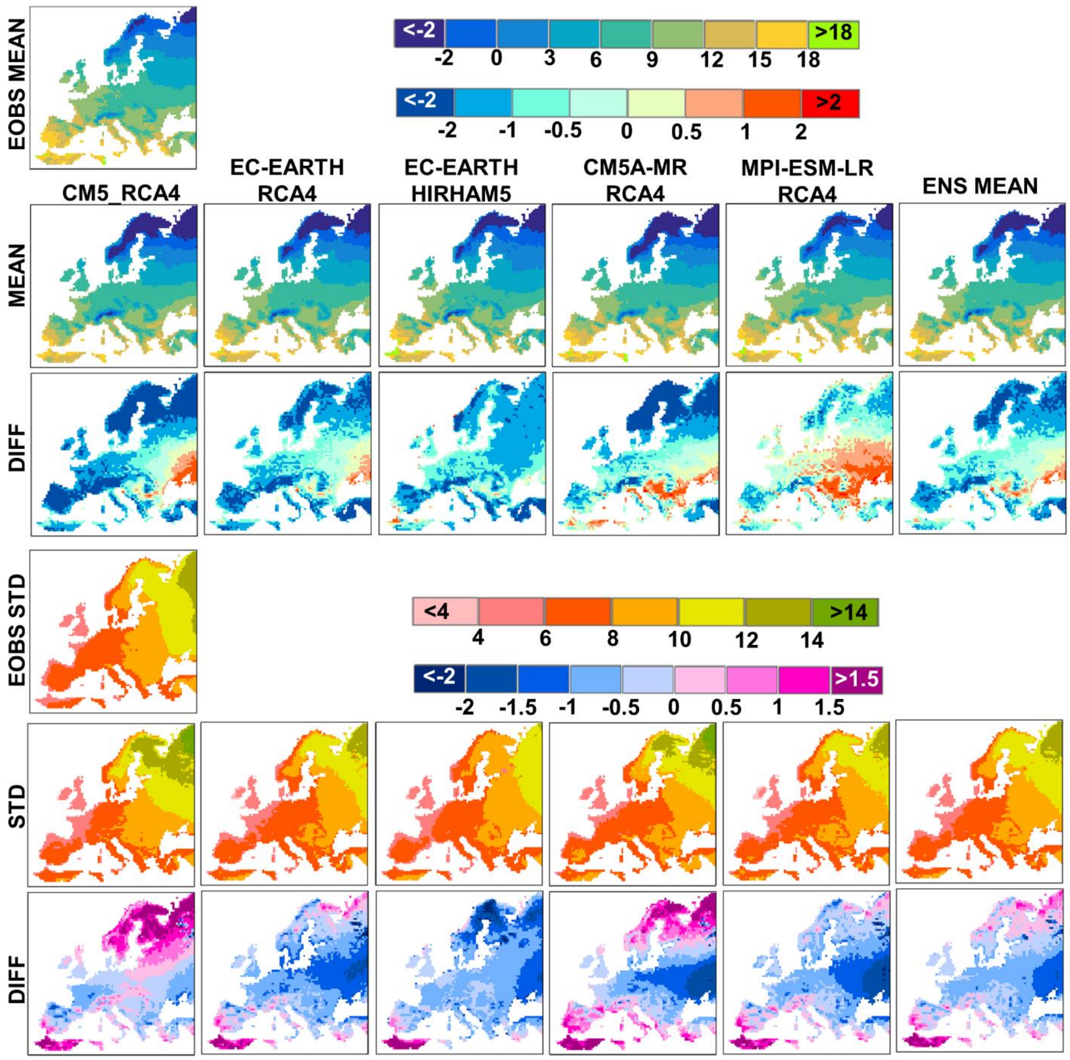


Figure 2: a) Annual averages of temperature of the raw model data, the observations and the bias correction with and without the NSM, for the calibration period 1850 – 1899 (solid lines) and the validation period 1900-2005 (dashed). Probability densities of annual (c) and of daily means (d).

Earth System Dynamics Discussions



497

498 **Figure 3: The EOBS and the RCM models' mean and standard deviation for the reference period**
499 **1951-2005. Differences (DIFF) indicate the [RCM statistic – EOBS statistic].**

500

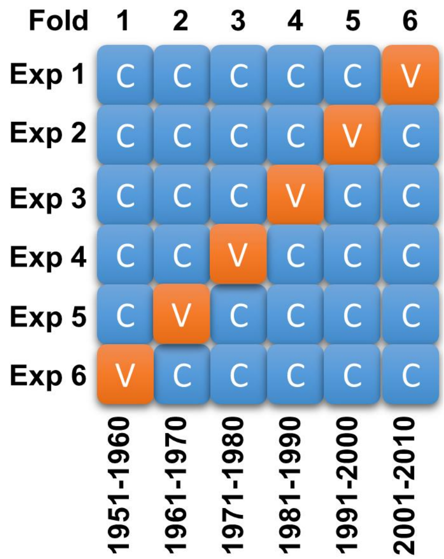
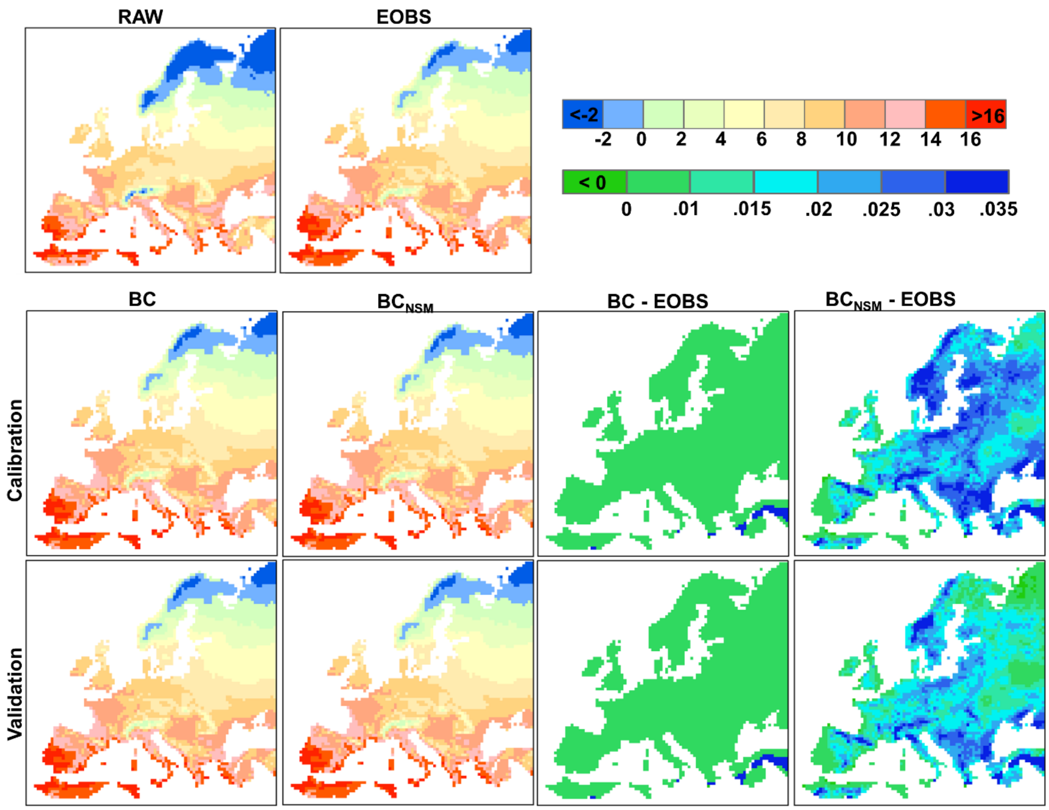


Figure 4: The 6-fold cross validation scheme with the calibration (C) and the validation (V) periods of each fold. Each experiment (Exp) was replicated for all five RCM models.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

505 

506 **Figure 5: Mean surface temperature of the cross validation test.**

507

Earth System Dynamics
Discussions
Open Access

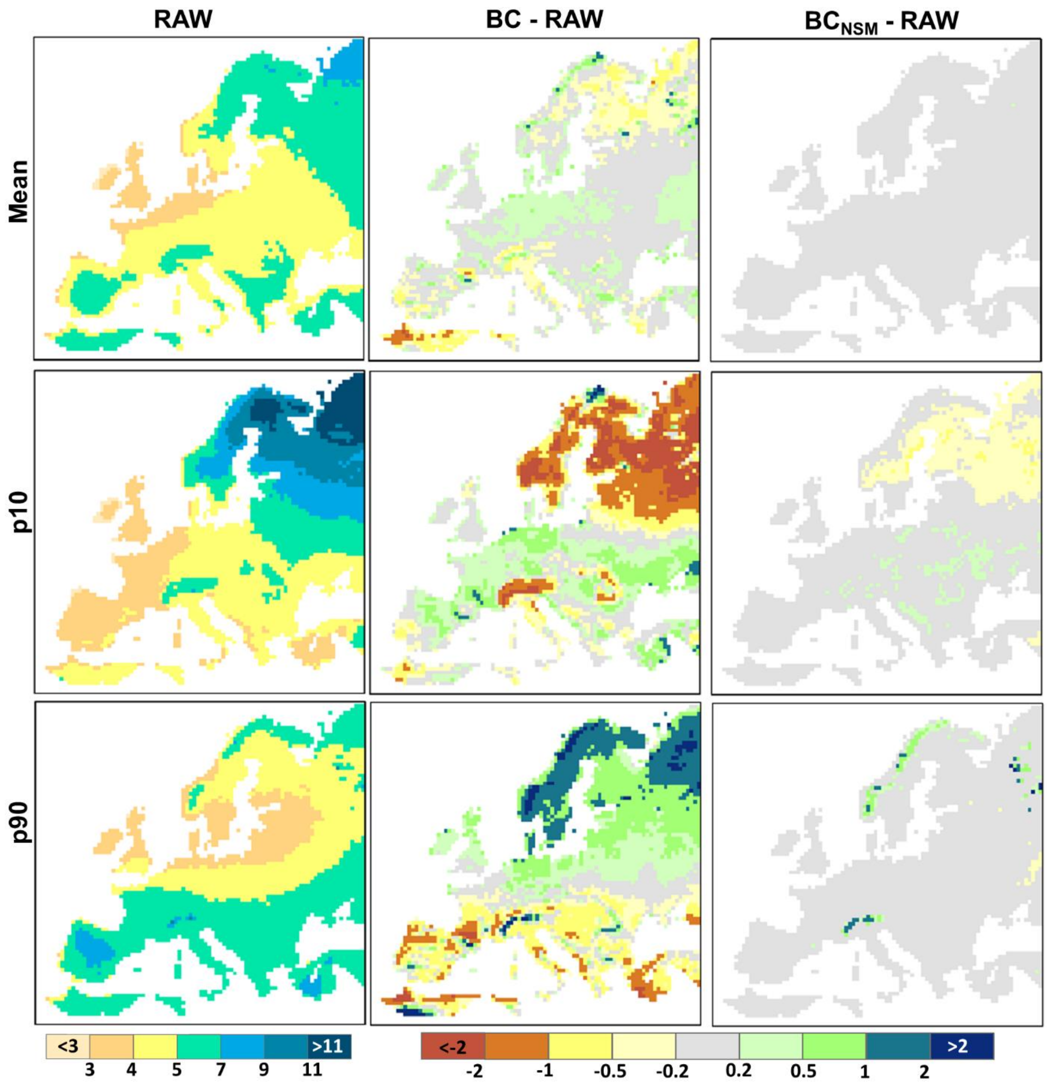


Figure 6: Long-term linear trend [ºC/100 y] in the raw data (left) for the mean annual temperature (top) and the 10th and 90th (p10 and p90) percentiles the change in the trend relatively to the raw data trend is provided for the BC (middle) and the BC_{NSM} data (right). All values are expressed as degrees per century.

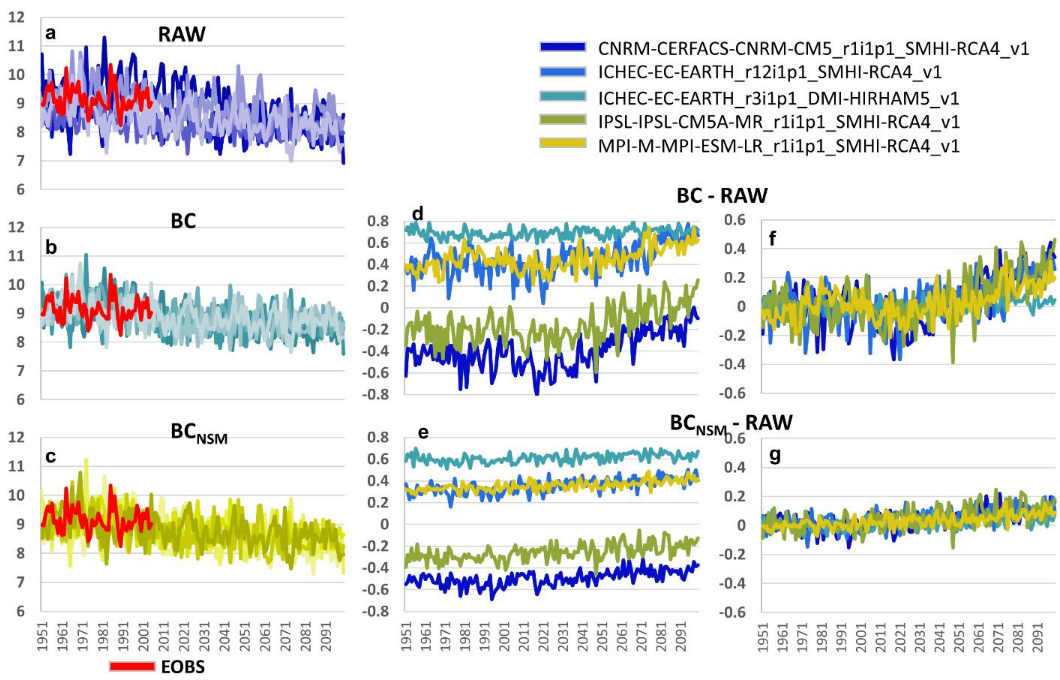


Figure 7: Average of standard deviations for the study domain, for the raw data (a), the BC (b) and the BC$_{NSM}$ (c) for the different models and the observations, in annual basis. Differences between the raw and the bias corrected standard deviations are shown in (d) and (e). Plots (f) and (g) correspond to the same data as (d) and (e), but normalized for their 1951-2005 mean.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-52, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

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 and pn for the nth quantile.

	Parameter	RAW	Normalized	Residuals	OBS	BC	BC$_{NSM}$
Calibration	Slope [ºC/10yr]*	-0.067	0.000	-0.067	-0.026	-0.086	-0.065
	Mean [ºC]	11.2	11.2	0.0	9.1	9.2	9.2
	SD [ºC]	4.5	4.6	0.9	5.3	5.3	5.3
	p5 [ºC]	4.4	4.4	-1.4	0.4	0.5	0.5
	p95 [ºCº]	19.4	19.1	1.5	17.6	17.7	17.6
Validation	Slope [ºC/10yr]*	0.052	0.000	0.051	0.076	0.062	0.051
	Mean [ºC]	11.3	11.2	0.1	9.6	9.3	9.3
	SD [ºC]	4.7	4.6	0.9	5.2	5.5	5.4
	p5 [ºC]	4.2	4.4	-1.3	1.1	0.2	0.3
	p95 [ºC]	19.4	19.1	1.5	17.6	17.7	17.6

Table 2: 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-
4	IPSL-CM5A-MR_r1i1p1_SMHI-
5	MPI-ESM-LR_r1i1p1_SMHI-