The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal

A. Gobiet[1,*], M. Suklitsch[2], and G. Heinrich[1,+]

[1]{Wegener Center for Climate and Global Change (WegCenter), University of Graz, Graz, Austria}

[2]{Central Institute for Meteorology and Geodynamics (ZAMG), Department of forecasting models, Vienna, Austria}

[*]{Now at: Central Institute for Meteorology and Geodynamics (ZAMG), Austria}

[+]{Now at: Department of Geography and Regional Science, University of Graz, Graz, Austria}

Correspondence to: A. Gobiet (andreas.gobiet@zamg.ac.at)

Abstract

This study discusses the effect of empirical-statistical bias correction methods like quantile mapping (QM) on the temperature change signals of climate simulations. We show that QM regionally alters the mean temperature climate change signal (CCS) derived from the ENSEMBLES multi-model dataset by up to 15 %. Such modification is currently strongly discussed and is often regarded as deficiency of bias correction methods. However, an analytical analysis reveals that this modification corresponds to the effect of intensity-dependent model errors on the CCS. Such errors cause, if uncorrected, biases in the CCS. QM removes these intensity-dependent errors and can therefore potentially lead to an improved CCS. A similar analysis as for the multi-model mean CCS has been conducted for the variance of CCSs in the multi-model ensemble. It shows that this indicator for model uncertainty is artificially inflated by intensity-dependent model errors. Therefore, QM has also the potential to serve as an empirical constraint on model uncertainty in climate projections. However, any improvement of simulated CCSs by empirical-statistical bias correction methods can only be realized, if the model error characteristics are sufficiently time-invariant.

1 Introduction

Society is increasingly demanding reliable projections of future climate change to analyse adaptation options and costs, to explore climate change mitigation benefits, and to support political decisions. Such climate projections are usually generated with general circulation models (GCMs) of rather coarse spatial resolution, which are refined by dynamical or statistical downscaling methods (e.g., Giorgi and Mearns, 1991; Fowler et al., 2007). Currently, an increasing number of climate change impact investigations rely on dynamical downscaling methods, i.e. the use of regional climate models (RCMs, e.g., Giorgi and Mearns, 1991; 1999; Wang et al., 2004; Rummukainen, 2010). However, even the newest generation of RCMs features considerable systematic errors (e.g., Kotlarski et al, 2014), which complicates the direct application of RCM results in climate change impact research. RCM output is therefore usually post-processed with empirical-statistical "bias correction" methods (e.g., Déqué, 2007; Themeßl et al., 2011) before it is used as input for impact

models, such as hydrological models. Bias correction methods have been demonstrated to successfully reduce systematic model errors (i.e. the difference between historical model output and meteorological observations), but the knowledge about how they influence the climate change signal (CCS; i.e. the long-term average difference between a future and a past climate simulation) is very limited so far.

A relation between model errors and CCS has been discussed by Christensen et al. (2008), who found that monthly temperature errors of RCMs over Europe often depend on the observed monthly mean temperature and that in warmer months errors are often larger than in colder months (or vice versa). Such "intensity-dependent" errors can be shown to alter the temperature CCS (Christensen et al., 2008; Themeßl et al., 2012; Boberg and Christensen, 2012).

Bias correction methods like quantile mapping (QM) modify the CCS. E.g., Themeßl et al. (2012) and Dosio et al. (2012) showed that QM modifies the CCS of RCMs operated over Europe in some regions and seasons and found a lower summer temperature CCS in Eastern Europe as well as a higher winter temperature CCS in Scandinavia after bias correction with QM. Currently, such modifications are often regarded as an undesired deficiency of bias correction methods (e.g., Hempel et al., 2013). However, Maurer and Pierce (2014) recently claimed that QM may have no negative effect on the quality of the CCS and demonstrated that QM doesn't deteriorate the multi-model mean precipitation CCS in a GCM ensemble.

In this paper we go a step further and argue that, under the assumption of time-invariant model error characteristics, the modification of the CCS by QM can be interpreted as improvement, rather than as deterioration, since it is capable of mitigating intensity-dependent model errors. To support this hypothesis, we develop a linearized analytical description of the effect of intensity-dependent model errors on the CCS. This framework allows to investigate the impact of such errors not only on the multi-model mean CCSs in an ensemble of climate simulations, but also on the inter-model variability, which is often used as a measure of uncertainty in climate projections (e.g., Hawkins and Sutton 2009; 2011; Prein et al., 2011). Furthermore, we compare the analytical correction of the CCS to the correction by QM.

In Sect. 2, the QM method is described and its effect on the temperature CCS of the ENSEMBLES multi-model dataset is demonstrated. In Sect. 3, the error characteristics of the ENSEMBLES models are analysed and in Sect. 4 we present an analytical formulation of intensity-dependent model errors and their effects on the CCS. In Sect. 5 these effects are

compared to the effects of QM on CCSs and in Sect. 6 a summary is given and conclusions are drawn.

2 Quantile mapping and its effect on climate change signals over Europe

2.1 Quantile mapping

The basic assumption of QM is that model errors depend on the value of the simulated variable. This concept of intensity-dependent errors is a rough simplification of actual model error characteristics, since model errors are influenced not only by the local value of the simulated variable. However, we will demonstrate that errors and local values correlate well in many cases (Sect. 3). The concept is simple yet powerful, since it separates, e.g., cold from hot regimes, or drizzling from heavy precipitation regimes and therefore accounts for potentially very different model errors under the associated regimes. It should be emphasized that intensity-dependent model errors are equivalent to a miss-representation of variability, i.e. to differences between the observed and modelled width of the density distribution. Fig. 1 demonstrates that intensity-dependent error characteristics with a positive slope correspond to overestimated variability, if the model error is defined as the difference between the inverse modelled and observed empirical cumulative density functions (ECDF). Similarly, a negative error slope corresponds to underestimation of variability.

Fig. 1

QM is a distribution-based bias correction method (e.g., Panofsky and Brier, 1958; Wood et al., 2004) that maps a modelled historical ECDF to an observed ECDF, with the mapping function shown in Fig. 1c for an artificial example. It is a well-established method to prepare climate model output as input for hydrological models (e.g., Déqué 2007; Maraun et al. 2010; Themeßl et al., 2011) and has been successfully applied to daily precipitation sum and air temperature of RCMs and GCMs by Dobler and Ahrens (2008), Piani et al. (2010a; 2010b), Dosio and Parulo (2011), Dosio et al. (2012), Maurer and Pierce (2014) and others. Furthermore, Themeßl et al. (2011) showed for daily precipitation sums that QM outperforms six other prominent bias correction techniques.

In our study, a non-parametric version of QM is used (Themeßl et al., 2011; 2012; Wilcke et al., 2013), as suggested by Gudmundsson et al. (2012). The ECDFs are constructed from 930 values for each day of the year based on modelled and observed data of a 30 year reference period (1961 – 1990) and a 31 day moving window centred on the day under consideration.

Our implementation of QM is not restricted to the range of observed values in the reference period, since the correction is extrapolated beyond the calibration range by using the correction term of the highest and lowest quantile, respectively. Please note, that this implies constant (not intensity-dependent) error characteristics outside the calibration range. As discussed by Bellprat et al. (2013), such constant error at high temperatures outside the calibration range may be more realistic in many cases than a linear extrapolation.

Some restrictions apply to the application of QM on climate scenarios: As pointed out by Eden et al. (2012), internal variability causes differences between a GCM simulation and observations, which cannot be separated from actual model errors, if QM is applied to GCM-driven RCMs, as in our case (see Sect. 2.2). By using rather long calibration periods (30 years) and by focusing on temperature, which is less affected by natural variability than, e.g., precipitation, we try to minimize this effect. In addition, our multi-model approach further reduces dependence on natural variability. However, in the interpretation of the results, some noise due to natural variability has to be taken into account. Similar to all empirical-statistical downscaling and bias correction methods, the application of QM on future climate simulations is based on the assumption of time-invariant model error characteristics. This stationarity assumption can obviously not be directly assessed for future periods and it can be expected to be violated to some degree. However, several studies demonstrate the skill of empirical-statistical bias correction methods either for past periods independent of the calibration period under on-going climate change (e.g., Piani et al., 2010a; Themeßl et al. 2012; Gudmundsson et al. 2012; Wilcke et al., 2013) or for future periods using a pseudo-reality approach (Maraun, 2012). Furthermore, Teutschbein and Seibert (2013) show that correction methods like QM perform better under non-stationary conditions than widely used linear transformations or the delta-change approaches. This gives confidence that empirical-statistical bias correction with QM is useful not only for historical simulations, but also, though with degraded performance, for future climate simulations. However, in a strict interpretation, the results and conclusions of this study are only valid under the assumption of time-invariant model errors and it is still issue to further investigation to determine the severity of this restriction. Although such investigation is outside the scope of our study, we want to mention that the new centennial re-analyses of ECMWF (ERA-20C) and NOAA-CIRES (V2c) offer a promising new test-bed for the investigation of the long-term stability of model error characteristics.

2.2 Model and observational data

We apply QM to a set of 15 GCM-driven regional climate simulations for Europe from the ENSEMBLES multi-model dataset (van der Linden and Mitchell, 2009). The ENSEMBLES models are operated on a 25 km grid and reach until 2100. In the following, we show the results for daily mean temperature, but the analysis of daily minimum and maximum temperatures gives very similar results. The application of our analysis to other parameters like, e.g., precipitation is basically straight forward, but the linearization applied in section 4 can be expected to be less appropriate for precipitation than for temperature. Further investigation is needed to fully reveal the effect of QM on the precipitation CCS. The major motivation for focusing on temperature here is its relatively simple error characteristic and its significant climate trend, which facilitates the demonstration of the effect of QM on the CCS.

As observational reference, the ENSEMBLES gridded observational dataset (E-OBS, Haylock et al. (2008)) is used. It is a European land-only daily high-resolution (25 km grid spacing) dataset for 5 meteorological parameters, including daily mean temperature.

2.3 The effect of QM on the CCS in ENSEMBLES

Subsequently, we show the effect of QM on the multi-model mean CCS and on the standard deviation of CCSs for the periods 2021 – 2050 and 2070 – 2099, both compared with the reference period 1971 – 2000. In Fig. 2 the spatial patterns of the difference between the uncorrected and the corrected multi-model mean temperature CCS is shown for different seasons in the mid (left) and end (right) of the 21st century. In the end of the century differences exceed +0.5 K in summer (JJA) in larger parts of South-Eastern Europe, France, and the Iberian Peninsula and –0.5 K and in larger regions in Scandinavia, which roughly corresponds to 15% of the uncorrected CCS. These results are consistent with the analyses of Boberg and Christensen (2012) and Dosio et al. (2012) and indicate that summer warming in South-Eastern Europe is projected to be less severe and warming in Scandinavia is projected to be more severe after bias correction with QM. However, the differences remain in the order of 10 % the uncorrected CCS and the basic pattern of temperature change is not strongly altered by QM.

Fig. 2

Fig. 3 shows the spatial pattern of the difference between the uncorrected and the corrected standard deviation of CCSs as a measure of model uncertainty. In most regions, model uncertainty is larger in the uncorrected model ensemble (orange colors), particularly in regions where the CCS is overestimated (see Fig. 2). The overestimation locally peaks at 0.5 K. However, in some regions (e.g., Scandinavia) and periods (e.g., late 21st century winter) model uncertainty is smaller in the uncorrected model ensemble, locally peaking at about -0.4 K.

Fig. 3

After having demonstrated and quantified the effects of QM on the CCS and the model uncertainty in the ENSEMBLES multi-model ensemble, the rest of this paper is devoted to the explanation of these effects.

3 Intensity-dependent model errors in the ENSEMBLES multi-model dataset

Since intensity-dependent model errors are the main suspects to cause the demonstrated effect of QM on the CCS, we investigate whether such errors exist in the ENSEMBLES RCMs. Due to their contrasting error characteristics, two of the RCMs are discussed in more detail: The HadRM3Q3 (driven by the HadCM3Q3 global climate model) operated by the Hadley Centre (HC) and the RCA (driven by run 3 of the ECHAM5 global climate model) operated by the Swedish Meteorological Service (SMHI). Since model error characteristics are known to be regionally very variable, Europe is separated into 8 sub-regions following Rockel and Woth (2007), which are marked in Figs. 2 and 3: The British Islands (BI), France (FR), Central Europe (ME), Scandinavia (SC), Iberian Peninsula (IP), Mediterranean (MD), Alps (AL) and Eastern Europe (EA).

The following characterization of model errors is based on daily mean temperature ECDFs, which are averaged over each month and sub-region. For each model, only the range between the 10th and 90th percentiles is used in order to avoid the noisy tails of empirical distributions. The ECDFs of the grid points in each sub-region are sampled over this range on a daily basis and the daily model error characteristics are derived for each grid point by subtracting the inverse observed from the inverse modelled ECDF (see Fig. 1). Further, the grid point error characteristics are averaged over each sub-region and each month of the year. Fig. 4 exemplarily shows the daily temperature error characteristics of the HC and SMHI models.

Fig. 4

Both models are affected by strongly intensity-dependent errors, but the error characteristics of the two models differ substantially. While the HC model features positive error slopes (up to 0.5) in most seasons and regions, the SMHI model has mainly negative slopes (up to -0.7). Both models are rather extreme examples within the ENSEMBLES multi-model dataset and most other models feature smaller slopes of about +/- 0.1 (Figs. S1 to S8 in the supplementary material).

In order to analyse whether such single-model error slopes cancel out in the multi-model ensemble, the ensemble average error characteristics (bold lines) in SC and EA are shown in Fig. 5 together with those of all 15 individual models (light lines). In SC, a considerable negative multi-model average slope exists in most parts of the year (minimum: in July). Contrary, positive slopes can be found in EA in summer (maximum: in July). Several other regions, like AL, feature only minor multi-model average slopes, but in turn larger slope variability (see supplementary material, Figs. S9 to S12).

Fig. 5

4 Analytical description of the effect of intensity-dependent model errors on the CCS

Having shown and quantified the intensity-dependence of model errors in the ENSEMBLES multi-model dataset, we subsequently give a simplified analytical description to highlight the mechanism how such errors act on the CCS in a multi-model ensemble.

4.1 CCS of a single climate simulation

Let y_j^i be the value of a meteorological variable (e.g., temperature, precipitation sum, or any other simulated variable) on day *j* simulated by model *i*. y^i is the 30 year average for a specific time of the year, e.g., for a month or a season. It can be expressed as a combination of the observed average value x and the deviation of the model from this value due to errors (y_e^i) and due to natural and model internal variability (y_v^i): $y^i = x + y_e^i + y_v^i$. The CCS Δy^i ($\Delta y^i = y_{future}^i - y_{past}^i$) can then be written as:

$$\Delta y^i = \Delta x + \Delta y_e^i + \Delta y_v^i. \quad (1)$$

Δx denotes the deterministic part of the error-free CCS, Δy_e^i the effect of model errors, and Δy_v^i the random effect of internal variability. In many studies, the model error term is neglected ("delta change approach"), since errors are expected to be time-invariant and to cancel out in the CCS. We demonstrate that this is not the case, even for time-invariant error characteristics, if they are intensity-dependent and the CCS is non-zero. The daily intensity-dependent errors can be written as a function of the meteorological variable under consideration: $y_{e,j}^i = f(y_j^i)$. For the sake of simplicity, we assume a linear error function with a constant bias b^i, error slope s^i, and residual ε_j^i:

$$y_{e,j}^i = b^i + s^i y_j^i + \varepsilon_j^i . \quad (2)$$

This linear error function is a good approximation of the error characteristics of the ENSEMBLES multi-model dataset in most cases, since the median coefficient of determination of the linear regression to the error characteristics shown in Sect. 3 is high (R^2 = 0.91). However, it is not always suitable as, e.g., in the case of the HC model in SC in winter and the SMHI model in IP in summer (Fig. 4).

Averaging over 30 years, taking the difference between a future and a past period, and neglecting the residual yields the linearized effect of the intensity-dependent model error on the CCS:

$$\Delta y_e^i = s^i \Delta y^i . \quad (3)$$

The bias cancels out, since it is assumed to be time-invariant and not intensity-dependent. From Eqs. (1) and (3) the simulated CCS can be written as:

$$\Delta y^i = \frac{\Delta x + \Delta y_v^i}{1 - s^i} . \quad (4)$$

Eq. (4) shows that intensity-dependent model errors lead to a modeled CCS that is proportional to the error free CCS ($\Delta x + \Delta y_v^i$) and a factor determined by the error slope ($1/(1-s^i)$). Fig. 6a illustrates this effect in relative terms: Positive error slopes lead to an exaggeration of the error-free CCS and negative slopes dampen it, but to a smaller extent. E.g., for slopes of 0.1 and -0.1 the error would amount to about 11 % and -9 %, respectively. The depicted range of error slopes from -0.7 to 0.5 has been selected according to temperature error slopes found in the ENSEMBLES multi-model dataset (see Sect. 5).

Fig. 6

4.2 Multi-model mean CCS

For a multi-model ensemble, the ensemble mean CCS and the multi-model variance of the CCS is relevant. To derive the effect of intensity-dependent errors on these quantities, the error slope can be written as the sum of the ensemble mean error slope ($\bar{s}$) and a model specific residuum error slope (s'^i). Combining this separation with the expanded form of Eq. (4) yields:

$$\Delta y^i = \Delta x + (\bar{s} + s'^i)\Delta y^i + \Delta y_v^i . \quad (5)$$

Accordingly, the multi-model mean CCS is:

$$\overline{\Delta y^i} = \Delta x + \bar{s}\,\overline{\Delta y^i} + \mathrm{cov}(s'^i, \Delta y^i) . \quad (6)$$

In Eq. 6 we could disregard the internal variability Δy_v^i since it has the expectation zero (assuming a large number of models *n*). In addition, the expectation of the product $s'^i \Delta y^i$ equals the covariance of both terms, since the expectation of s'^i is zero under the assumption of normally distributed error slopes. However, it is not independent from Δy^i as the error slope influences the CCS according to Eq. (4). In a similar form as Eq. (4), Eq. (6) reads:

$$\overline{\Delta y^i} = \frac{\Delta x + \mathrm{cov}(s'^i, \Delta y^i)}{1 - \bar{s}} . \quad (7)$$

Eq. (7) shows that intensity-dependent errors influence the multi-model mean CCS via two terms: Firstly, the error slope term, which scales with the error-free CCS (Δx) just like in the single-model case, and secondly the covariance term, which adds an offset. Fig. 6b visualizes the corresponding error in the CCS in relative terms: Positive multi-model mean error slopes lead to an exaggeration of the CCS and negative slopes dampen it, just like in the single model case (black line). The depicted range of multi-model error slopes from –0.16 to +0.13 has been selected according to the multi-model mean temperature error slopes of the ENSEMBLES multi-model dataset (see Sect. 5). Positive and negative covariance terms create positive and negative offsets, respectively. Following Eq. (3), it can be expected that single model error slopes and CCSs are generally positively correlated and that the covariance

term is consequently positive. The depicted range of covariance terms corresponds to values found in the analysis of temperature errors of the ENSEMBLES multi-model dataset, ranging from -0.02 (blue colors) to +0.21 (pink colors) (Sect. 5) and confirms this expectation. The absolute effect of the covariance term (Eq. (7)) is independent from the error-free CCS and thus gets smaller with higher CCS in relative terms, which is indicated by lighter (small CCS) and darker colors (large CCS).

4.3 Variance of CCSs in a multi-model ensemble

The effect of intensity-dependent errors on the second important quantity in a multi-model ensemble, the variance of CCSs (which is often interpreted as a measure of uncertainty), can be described with the linearized model as well. Using Eqs. (5) and (6), the variance can be expressed as:

$$\mathrm{var}(\Delta y^i) = \frac{1}{n}\sum_{i=1}^{n}[\Delta y^i - \overline{\Delta y^i}]^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\bar{s}(\Delta y^i - \overline{\Delta y^i}) + s'^i \Delta y^i + \Delta y_\nu^i - \mathrm{cov}(s'^i, \Delta y^i)\right]^2 . \quad (8)$$

Expanding and simplifying Eq. (8) gives (see supplementary material for a detailed derivation):

$$\mathrm{var}(\Delta y^i) = \mathrm{var}(\Delta y^i{}_\nu) + \bar{s}^2\,\mathrm{var}(\Delta y^i) + \mathrm{var}(s'^i \Delta y^i) + 2\bar{s}\,\mathrm{cov}(\Delta y^i, s'^i \Delta y^i) \quad . \quad (9)$$

Since $\mathrm{var}(\Delta y^i{}_\nu)$ is the effect of natural variability, it can be interpreted as the variance of an error-free model ensemble. Compared to that, the variance of a model ensemble with intensity-dependent errors is always exaggerated by a positive offset $\bar{s}^2\,\mathrm{var}(\Delta y^i)$. E.g., an ensemble mean error slope of ±0.1 results in about 1 % bias in variance. In addition, the positive additive term $\mathrm{var}(s'^i \Delta y^i)$, which represents the variability of the individual model's error slopes times CCSs, further increases the positive bias. The last term $2\bar{s}\,\mathrm{cov}(\Delta y^i, s'^i \Delta y^i)$ is positive for positive slopes and negative for negative slopes, assuming a positive correlation of the simulated CCS and the residual error slope. It is difficult to estimate the relative importance of the different terms and in particular to judge if the possibly negative covariance term can counterbalance the otherwise positive terms, so all terms of Eq. (9) are quantified and analyzed for the ENSEMBLES multi-model ensemble in Sect. 5.

4.4 Linearized correction

The linearized error characterization leads to a simple way to correct the CCS of single models following Eq. (3), the multi-model mean CCS following Eq. (6), and the multi-model variance of CCS following Eq. (9). Error slopes, climate change signals, their variability, and their covariance are calculated based on the comparison of historical simulations with observations and applied to results of future simulations. Such correction assumes not only a linear error-slope, but also time-invariant error characteristics. The linearly corrected multi-model mean temperature CCS is listed in Table 1 (Δx_{LC}) and the variance of the CCSs in Table 2 ($\mathrm{var}(\Delta x)_{LC}$). They are discussed in the following Section.

5 Correction of the CCS and its uncertainty

In Table 1 the terms contributing to errors in the multi model mean CCS (see Eq. (6)) are listed for all sub-regions and seasons. Multi model mean error slopes ($\bar{s}$) are mostly negative in DJF and MAM, mostly positive in JJA and SON, and range from –0.16 in SC in MAM to 0.13 in EA in JJA. Accordingly, they inflate (positive slopes) or dampen (negative slopes) the CCS, depending on season and sub-region. The errors stemming from the slope term ($\bar{s}\overline{\Delta y}$) range from –0.25 K to 0.20 K in the mid-century and from –0.57 K to 0.45 K at the end of the century. Contrary, the covariance-term ($\mathrm{cov}(s'^i, \Delta y^i)$) is, with very few exceptions, positive and increases the CCS. It amounts 0.04 K on average, ranges from –0.02 K to 0.21 K in both periods, but usually does not exceed 0.10 K. Compared to the slope term, the covariance term is smaller in most cases, but cannot be neglected, as it sometimes equals or even exceeds the slope term. Table 1 also lists the uncorrected (Δy) and corrected multi-model mean CCS (linearized correction: Δx_{LC}; quantile mapping: Δx_{QM}) for each season and sub-region. The difference between uncorrected and corrected CCS averaged over all seasons and regions is small (0.01 K), but can reach up to about 0.5 K (about 15% of the uncorrected CCS) in specific regions and seasons. Fig. 7 displays this estimated error in the multi-model mean temperature CCS. With few exceptions, both correction methods feature the similar sign of correction and agree reasonably well in their magnitude. Major differences are found in the later period, when QM often indicates smaller errors than LC. This can be probably explained by the fact that LC extrapolates intensity-dependent errors, while our implementation of QM keeps the error constant outside the calibration range (see Sect. 2.1). This dampens the error slope under severe warming (i.e. at the end of the 21st century) when daily temperatures

outside the calibration range frequently occur. Further discrepancies between QM and LC can be explained by the linear approximation of LC. Both correction methods agree that the uncorrected CCS is regionally biased up to +0.5 K in EA and FR in summer and about –0.5 K in SC. The qualitative agreement of QM with LC can be interpreted as a confirmation that the correction of intensity-dependent errors is the main reason of the modification of the CCS by QM.

Table 1

Fig. 7

In Table 2, the terms contributing to errors in the estimated variance of a multi-model ensemble (Eq. (9)) are listed: Two positive offset terms $\bar{s}^2 \,\mathrm{var}(\Delta y^i)$ and $\mathrm{var}(s'^i \Delta y^i)$, and the term $2\bar{s}\,\mathrm{cov}(\Delta y^i, s'^i \Delta y^i)$, which generally has the same sign as the error slope due to the positive correlation between single model CCS and error slope. While $\bar{s}^2 \,\mathrm{var}(\Delta y^i)$ is very small in both periods (smaller than 0.01 K² in most cases), $\mathrm{var}(s'^i \Delta y^i)$ amounts to 0.041 K² on average (range: 0.006 K² – 0.128 K²) in the earlier period and to 0.223 K² on average (range: 0.026 K² – 0.697 K²) in the later period. Given a modeled average variance of 0.342 K² in the earlier and 1.028 K² in the later period, this means that this term leads to an overestimation of variance by 12 % and 22 % on average, respectively. In specific regions and seasons, the overestimation can amount 50 % and more (e.g., SC in the later period). The average covariance term $2\bar{s}\,\mathrm{cov}(\Delta y^i, s'^i \Delta y^i)$ is very small in both periods (–0.002 K² and +0.002 K², respectively) and ranges from –0.028 K² to 0.136 K². In summary, the positive variability term $\mathrm{var}(s'^i \Delta y^i)$ dominates and is mostly even enhanced by the covariance term. This leads to a general overestimation of ensemble variance.

Table 2

Fig. 8

Table 2 also lists the uncorrected ($\mathrm{var}(\Delta y^i)$) and corrected variance of the CCSs in the multi-model ensemble (LC: $\mathrm{var}(\Delta x_{LC})$; QM: $\mathrm{var}(\Delta x_{QM})$) for each season and sub-region. The average difference between uncorrected and corrected variance over all seasons and regions does not cancel out as in the case of the mean CCS, but amounts on average to 17 % in the case of LC and to 12 % in the case of QM. This demonstrates that time-invariant intensity-dependent errors inflate model uncertainty in multi-model ensembles. In Fig. 8 this error is

expressed as standard deviation, which is overestimated by up to 0.4 K at the end of the century. This is particularly the case in regions where the mean CCS is overestimated like in EA in summer. However, the two correction methods disagree in some cases as, e.g., in SC in winter at the end of the century. These discrepancies are currently not fully understood and require further analysis. They could, e.g., be caused by the linearity assumption of LC, by the constant (not intensity-dependent) correction outside the calibration range of QM, or by time-variant model errors.

6 Summary and Conclusions

The knowledge about the influence of empirical-statistical bias correction methods like QM on the CCS of climate simulations is very limited so far. For the ENSEMBLES multi-model dataset it has been demonstrated that QM dampens projected summer warming in South-Eastern Europe and France by about 0.5 K and enhances projected warming in Scandinavia by about the same amount. This corresponds to about 15 % of the uncorrected CCS. Such modification is currently strongly discussed and is often regarded as deficiency of bias correction methods. However, we argue that under the assumption of time-invariant model errors, QM should generally lead to an improvement of the simulated CCS rather than deterioration.

To support this hypothesis, we analytically formulated the effect of intensity-dependent model errors on the CCS and showed that they erroneously modify the CCS. Positive error slopes lead to an exaggeration of the CCS and negative slopes dampen it. This is the case for a single model's CCS as well as for the multi-model mean CCS in a model ensemble, which is additionally exaggerated by high variability amongst the single model's CCSs. A comparison of this analytically determined error and the effect of QM on the mean CCS in the ENSEMBLES multi-model dataset leads to largely similar results. This confirms that the effect of QM on the CCS is mainly caused by the correction of intensity-dependent errors and that such modification can be regarded as improvement, if roughly time-invariant model error characteristics can be assumed.

With regard to the variance of the CCSs in a multi-model ensemble, the analytical description reveals that intensity-dependent model errors lead to an overestimation of variance. Since variability of CCSs in a multi-model ensemble is often used as indicator for model uncertainty, intensity-dependent model errors can be regarded to be responsible for parts of the model uncertainty in the CCS. This further implies that the correction of intensity-

dependent errors by QM should lead to a smaller variance and therefore constitute an empirical constraint on climate model uncertainty. However, we could only partly demonstrate this very desirable effect by the application of QM on the ENSEMBLES dataset. In most regions and seasons the analytical correction as well as QM reduce the variance as expected, but particularly in the winter season of longer term simulations QM often increases it, which could not be fully explained so far and needs further investigation.

Generally, our results indicate that empirical-statistical bias correction methods that correct for intensity-dependence in model errors can lead to improved estimates of future climate change. The improvements primarily refer to the mean CCS, but also an empirical constraint on uncertainty in multi-model climate projections seems to be feasible. A restriction to these results is the fact that any potential improvement can only be realized if the assumption of time-invariant model error characteristics sufficiently holds. It is still issue to further investigation to determine the severity of this restriction.

Author Contribution

A. Gobiet is responsible for the general concept and conduction of the study, for the analytical description presented in Sect. 4, for the interpretation of the results and for texting. M. Suklitsch contributed the analysis of the error characteristics of the ENSEMBLES models and G. Heinrich the analysis of the effect of QM on the climate change signal in the ENSEMBLES dataset. Both were also involved in the discussion of the results and contributed to parts of the text.

Acknowledgements

This study was partly funded by the EU FP7 projects ACQWA (grant agreement 212250) and IMPACT2C (grant agreement 282746). We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). The ENSEMBLES data used in this work was funded by the EU FP6 Integrated Project ENSEMBLES (Contract number 505539) whose support is gratefully acknowledged.

References

Bellprat, O., Kotlarski, S., Lüthi, D. and Schär, C.: Physical constraints for temperature biases in climate models, Geophys. Res. Lett., 40(15), 4042–4047, doi:10.1002/grl.50737, 2013.

Boberg, F. and Christensen, J. H.: Overestimation of Mediterranean summer temperature projections due to model deficiencies, Nature Clim. Change, 2(6), 433–436, doi:10.1038/nclimate1454, 2012.

Christensen, J. H., Boberg, F., Christensen, O. B. and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, Geophys. Res. Lett., 35(20), L20709, doi:10.1029/2008GL035694, 2008.

Déqué, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, Global and Planetary Change, 57(1–2), 16–26, doi:10.1016/j.gloplacha.2006.11.030, 2007.

Dobler, A. and Ahrens, B.: Precipitation by a regional climate model and bias correction in Europe and South Asia, Meteorologische Zeitschrift, 17(4), 499–509, doi:10.1127/0941-2948/2008/0306, 2008.

Dosio, A. and Paruolo, P.: Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate, J. Geophys. Res., 116(D16), D16106, doi:10.1029/2011JD015934, 2011.

Dosio, A., Paruolo, P. and Rojas, R.: Bias correction of the ENSEMBLES high resolution climate change projections for use by impact models: Analysis of the climate change signal, J. Geophys. Res., 117(D17), D17110, doi:10.1029/2012JD017968, 2012.

Eden, J. M., Widmann, M., Grawe, D. and Rast, S.: Skill, Correction, and Downscaling of GCM-Simulated Precipitation, J. Climate, 25(11), 3970–3984, doi:10.1175/JCLI-D-11-00254.1, 2012.

Fowler, H. J., Blenkinsop, S. and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, Int. J. Climatol., 27(12), 1547–1578, doi:10.1002/joc.1556, 2007.

Giorgi, F. and Mearns, L. O.: Approaches to the simulation of regional climate change: A review, Rev. Geophys., 29(2), 191–216, doi:10.1029/90RG02636, 1991.

Giorgi, F. and Mearns, L. O.: Introduction to special section: Regional Climate Modeling Revisited, J. Geophys. Res., 104(D6), 6335–6352, doi:10.1029/98JD02072, 1999.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E. and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a

comparison of methods, Hydrol. Earth Syst. Sci., 16(9), 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Predictions, Bull. Amer. Meteor. Soc., 90(8), 1095–1107, doi:10.1175/2009BAMS2607.1, 2009.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Clim Dyn, 37(1-2), 407–418, doi:10.1007/s00382-010-0810-6, 2011.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D. and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res., 113(D20), D20119, doi:10.1029/2008JD010201, 2008.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J. and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, Earth Syst. Dynam., 4(2), 219–236, doi:10.5194/esd-4-219-2013, 2013.

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. and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7(4), 1297–1333, doi:10.5194/gmd-7-1297-2014, 2014.

Maraun, D.: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, Geophys. Res. Lett., 39(6), L06706, doi:10.1029/2012GL051210, 2012.

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. and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48(3), RG3003, doi:10.1029/2009RG000314, 2010.

Maurer, E. P. and Pierce, D. W.: Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean, Hydrol. Earth Syst. Sci., 18(3), 915–925, doi:10.5194/hess-18-915-2014, 2014.

Panofsky, H. A. and Brier, G. W.: Some Applications of Statistics to Meteorology, Mineral Industries Extension Services, College of Mineral Industries, Pennsylvania State University., 1958.

Piani, C., Haerter, J. O. and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, Theor Appl Climatol, 99(1-2), 187–192, doi:10.1007/s00704-009-0134-9, 2010a.

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S. and Haerter, J. O.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, Journal of Hydrology, 395(3–4), 199–215, doi:10.1016/j.jhydrol.2010.10.024, 2010b.

Prein, A. F., Gobiet, A. and Truhetz, H.: Analysis of uncertainty in large scale climate change projections over Europe, Meteorologische Zeitschrift, 20(4), 383–395, doi:10.1127/0941-2948/2011/0286, 2011.

Rockel, B. and Woth, K.: Extremes of near-surface wind speed over Europe and their future changes as estimated from an ensemble of RCM simulations, Climatic Change, 81(1), 267–280, doi:10.1007/s10584-006-9227-y, 2007.

Rummukainen, M.: State-of-the-art with regional climate models, WIREs Clim Change, 1(1), 82–96, doi:10.1002/wcc.8, 2010.

Teutschbein, C. and Seibert, J.: Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions?, Hydrol. Earth Syst. Sci., 17(12), 5061–5077, doi:10.5194/hess-17-5061-2013, 2013.

Themeßl, Matthias Jakob, M., Gobiet, A. and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, Int. J. Climatol., 31(10), 1530–1544, doi:10.1002/joc.2168, 2011.

Themeßl, M. J., Gobiet, A. and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, Climatic Change, 112(2), 449–468, doi:10.1007/s10584-011-0224-4, 2012.

van der Linden, P. and Mitchell, J.F.B.: ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project — European Environment Agency (EEA), [online] Available from: http://www.eea.europa.eu/data-and-

maps/indicators/global-and-european-temperature/ensembles-climate-change-and-its (Accessed 2 January 2015), n.d.

Wang, Y., Leung, L. R., McGREGOR, J. L., Lee, D.-K., Wang, W.-C., Ding, Y. and Kimura, F.: Regional Climate Modeling: Progress, Challenges, and Prospects, Journal of the Meteorological Society of Japan. Ser. II, 82(6), 1599–1628, doi:10.2151/jmsj.82.1599, 2004.

Wilcke, R. A. I., Mendlik, T. and Gobiet, A.: Multi-variable error correction of regional climate models, Climatic Change, 120(4), 871–887, doi:10.1007/s10584-013-0845-x, 2013.

Wood, A. W., Leung, L. R., Sridhar, V. and Lettenmaier, D. P.: Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs, Climatic Change, 62(1-3), 189–216, doi:10.1023/B:CLIM.0000013685.99609.9e, 2004.

Table 1. Multi model mean temperature error slopes ($\bar{s}$), multi-model mean CCSs ($\overline{\Delta y^i}$), covariance error terms ($\mathrm{cov}(s'^i, \Delta y^i)$), linearly corrected CCSs (Δx_{LC}), and non-linearly corrected CCSs (Δx_{QM}) for the periods 2021 – 2050 (left) and 2070 – 2099 (right) [K].

			mid-century (2021-2050)				end-century (2070-2099)			
region	season	$\bar{s}$	Δy	$\mathrm{cov}(s', \Delta y)$	Δx_{LC}	Δx_{QM}	Δy	$\mathrm{cov}(s', \Delta y)$	Δx_{LC}	Δx_{QM}
BI	DJF	0.07	1.05	0.03	0.96	0.97	2.18	-0.01	2.05	2.05
	MAM	-0.16	0.94	0.03	1.06	0.99	2.08	0.04	2.40	2.23
	JJA	-0.15	0.94	0.06	1.03	1.02	2.31	0.09	2.59	2.48
	SON	0.02	1.10	0.02	1.06	1.11	2.47	0.02	2.41	2.50
FR	DJF	-0.03	1.31	0.03	1.33	1.33	2.54	-0.02	2.65	2.65
	MAM	-0.07	1.03	0.04	1.07	1.09	2.40	0.05	2.52	2.53
	JJA	0.08	1.37	0.13	1.15	1.22	3.68	0.21	3.19	3.37
	SON	0.00	1.22	0.01	1.21	1.15	3.08	0.02	3.07	2.96
ME	DJF	-0.03	1.49	0.03	1.50	1.50	3.01	-0.01	3.11	3.14
	MAM	-0.06	1.01	0.01	1.06	1.04	2.35	0.01	2.47	2.42
	JJA	0.05	1.14	0.07	1.01	1.08	3.00	0.10	2.77	2.85
	SON	0.02	1.23	0.01	1.20	1.16	3.01	0.01	2.95	2.82
SC	DJF	-0.06	1.84	0.06	1.90	1.99	4.37	0.02	4.64	4.80
	MAM	-0.16	1.57	0.07	1.77	1.62	3.59	0.05	4.17	3.85
	JJA	-0.14	1.25	0.01	1.44	1.46	2.68	0.01	3.09	3.11
	SON	-0.10	1.66	0.02	1.81	1.76	3.56	0.02	3.91	3.81
IP	DJF	-0.04	1.24	0.01	1.27	1.21	2.33	0.00	2.41	2.34
	MAM	-0.04	1.22	0.05	1.22	1.25	3.11	0.07	3.16	3.22
	JJA	0.04	1.66	0.05	1.55	1.58	4.44	0.06	4.21	4.32
	SON	0.00	1.45	0.02	1.42	1.41	3.51	0.04	3.45	3.48
MD	DJF	-0.08	1.36	0.02	1.45	1.38	2.78	0.02	3.00	2.96
	MAM	-0.05	1.27	0.05	1.29	1.33	3.04	0.08	3.14	3.20
	JJA	0.00	1.85	0.05	1.80	1.87	4.35	0.09	4.26	4.47
	SON	-0.04	1.44	0.03	1.47	1.38	3.41	0.05	3.50	3.40
AL	DJF	-0.02	1.54	0.03	1.55	1.40	3.15	-0.01	3.22	2.95
	MAM	-0.06	1.29	0.03	1.35	1.37	3.00	0.04	3.17	3.24
	JJA	0.02	1.58	0.06	1.48	1.57	4.10	0.10	3.91	4.08
	SON	0.05	1.35	0.01	1.28	1.24	3.37	0.02	3.20	3.16
EA	DJF	-0.04	1.70	0.03	1.73	1.67	3.39	0.00	3.52	3.44
	MAM	-0.01	1.19	0.02	1.18	1.18	2.79	0.03	2.78	2.80
	JJA	0.13	1.54	0.08	1.29	1.34	3.52	0.12	3.00	3.07
	SON	0.03	1.41	0.01	1.36	1.29	3.23	0.02	3.12	2.97
Mean		-0.03	1.35	0.04	1.35	1.34	3.12	0.04	3.16	3.15

Table 2. Multi-model variance of the temperature CCSs ($\mathrm{var}(\Delta y^i)$), error terms of the variance ($\bar{s}^2\,\mathrm{var}(\Delta y^i)$, $\mathrm{var}(s'^i \Delta y^i)$, $2\bar{s}\,\mathrm{cov}(\Delta y^i, s'^i \Delta y^i)$), linearly corrected variance ($\mathrm{var}(\Delta x)_{LC}$), and QM corrected variance ($\mathrm{var}(\Delta x)_{QM}$) for the periods 2021 – 2050 (left) and 2070 – 2099 (right) [K²].

		mid-century (2021-2050)						end-century (2070-2099)					
region	season	$\mathrm{var}(\Delta y)$	$s^2\mathrm{var}(\Delta y)$	$\mathrm{var}(s'\Delta y)$	$2s\,\mathrm{cov}(\Delta y, s'\Delta y)$	$\mathrm{var}(\Delta x_{LC})$	$\mathrm{var}(\Delta x_{QM})$	$\mathrm{var}(\Delta y)$	$s^2\mathrm{var}(\Delta y)$	$\mathrm{var}(s'\Delta y)$	$2s\,\mathrm{cov}(\Delta y, s'\Delta y)$	$\mathrm{var}(\Delta x_{LC})$	$\mathrm{var}(\Delta x_{QM})$
BI	DJF	0.283	0.001	0.008	0.004	0.269	0.198	0.719	0.003	0.050	0.002	0.664	0.883
	MAM	0.178	0.004	0.018	-0.008	0.165	0.183	0.409	0.010	0.073	-0.024	0.349	0.430
	JJA	0.308	0.007	0.034	-0.018	0.284	0.309	0.925	0.021	0.193	-0.059	0.770	0.784
	SON	0.266	0.000	0.016	0.001	0.249	0.240	0.814	0.000	0.077	0.002	0.736	0.768
FR	DJF	0.192	0.000	0.019	-0.002	0.174	0.150	0.797	0.001	0.076	0.001	0.719	1.136
	MAM	0.268	0.001	0.027	-0.006	0.246	0.282	0.674	0.003	0.124	-0.020	0.567	0.636
	JJA	0.540	0.004	0.058	0.028	0.451	0.307	1.641	0.011	0.697	0.137	0.796	0.829
	SON	0.445	0.000	0.011	0.000	0.433	0.311	1.571	0.000	0.087	-0.001	1.484	1.063
ME	DJF	0.322	0.000	0.035	-0.003	0.289	0.257	0.860	0.001	0.162	0.000	0.698	1.217
	MAM	0.311	0.001	0.022	-0.002	0.290	0.340	0.651	0.002	0.110	-0.004	0.543	0.706
	JJA	0.379	0.001	0.044	0.009	0.324	0.238	1.210	0.003	0.405	0.036	0.766	0.747
	SON	0.319	0.000	0.012	0.000	0.307	0.264	1.374	0.000	0.083	0.002	1.288	1.092
SC	DJF	0.647	0.002	0.128	-0.016	0.532	0.662	0.567	0.002	0.644	-0.015	-0.064	1.127
	MAM	0.518	0.014	0.098	-0.052	0.457	0.433	0.632	0.017	0.362	-0.073	0.326	0.771
	JJA	0.350	0.006	0.074	-0.002	0.271	0.449	0.775	0.014	0.231	-0.002	0.531	1.034
	SON	0.151	0.001	0.099	-0.008	0.059	0.179	0.534	0.005	0.365	-0.012	0.176	0.805
IP	DJF	0.170	0.000	0.006	-0.001	0.165	0.155	1.025	0.001	0.026	-0.003	1.000	1.142
	MAM	0.371	0.000	0.044	-0.006	0.332	0.308	0.781	0.001	0.229	-0.019	0.569	0.574
	JJA	0.354	0.001	0.028	0.004	0.322	0.267	0.858	0.001	0.274	0.016	0.566	0.739
	SON	0.420	0.000	0.008	0.000	0.411	0.284	1.356	0.000	0.049	0.001	1.305	0.855
MD	DJF	0.178	0.001	0.017	-0.004	0.163	0.177	1.424	0.009	0.084	-0.011	1.342	1.717
	MAM	0.381	0.001	0.060	-0.009	0.329	0.323	0.885	0.003	0.264	-0.028	0.646	0.709
	JJA	0.410	0.000	0.102	0.000	0.307	0.417	1.308	0.000	0.437	0.001	0.870	1.598
	SON	0.272	0.000	0.014	-0.002	0.260	0.195	1.076	0.002	0.082	-0.012	1.004	0.726
AL	DJF	0.304	0.000	0.023	-0.002	0.282	0.216	1.175	0.001	0.113	0.001	1.060	1.377
	MAM	0.410	0.002	0.049	-0.006	0.364	0.427	0.919	0.004	0.202	-0.019	0.732	0.917
	JJA	0.485	0.000	0.057	0.005	0.423	0.371	1.376	0.001	0.476	0.019	0.880	1.196
	SON	0.403	0.001	0.007	0.001	0.393	0.287	1.608	0.004	0.058	0.008	1.539	1.061
EA	DJF	0.332	0.000	0.080	-0.005	0.256	0.265	1.228	0.002	0.346	0.000	0.881	1.453
	MAM	0.350	0.000	0.040	0.000	0.311	0.335	0.711	0.000	0.206	-0.001	0.506	0.679
	JJA	0.461	0.008	0.060	0.041	0.353	0.207	1.597	0.027	0.432	0.136	1.002	0.816
	SON	0.182	0.000	0.022	0.001	0.158	0.134	1.427	0.001	0.129	0.006	1.291	0.973
Mean		0.342	0.002	0.041	-0.002	0.301	0.287	1.028	0.005	0.223	0.002	0.798	0.955

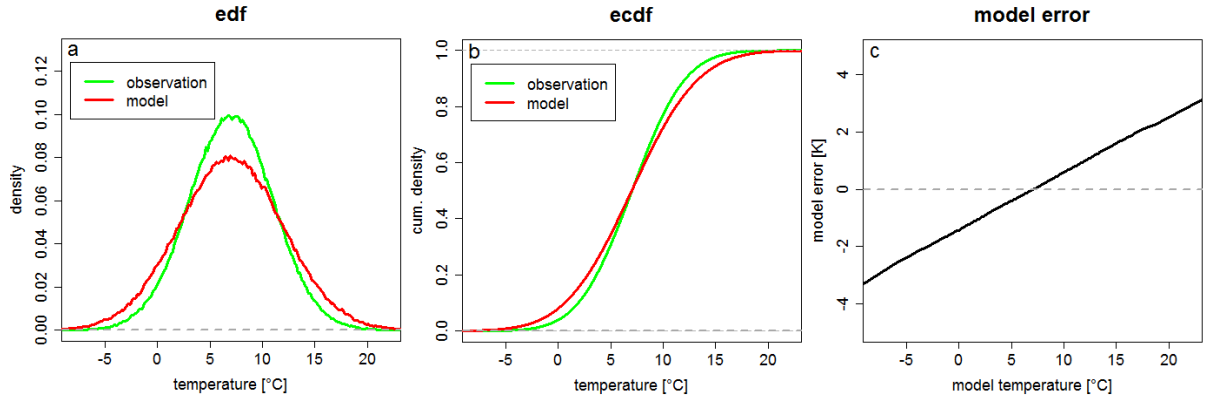

Figure 1. Intensity-dependent model errors of a model that overestimates daily temperature variability (artificial data). a: modeled (red, standard deviation = 5 °C) and observed (green, standard deviation = 4 °C) empirical density functions; b: modeled (red) and observed (green) ECDFs; c: model error at different modeled values.

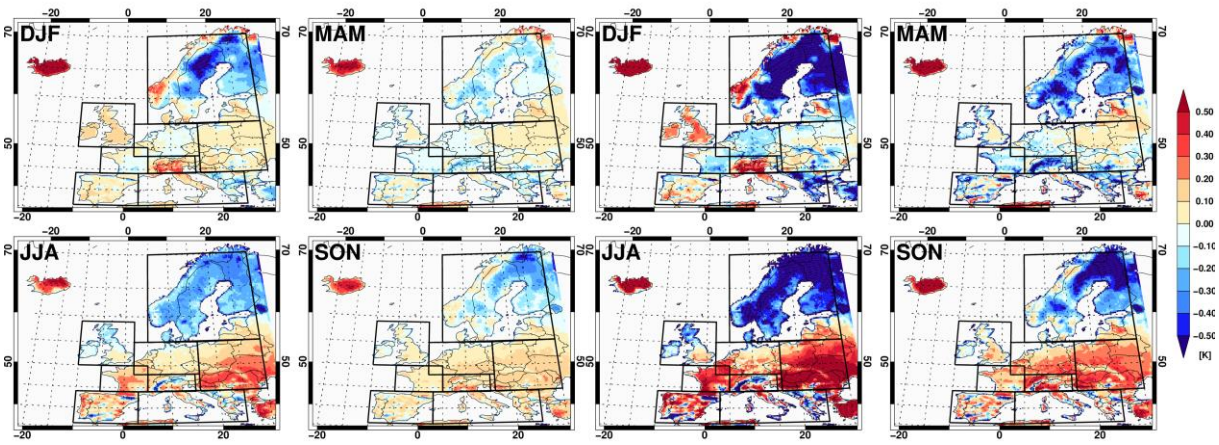

Figure 2. Differences between uncorrected and corrected (QM) multi-model mean temperature CCS. The reference period is 1971 – 2000. The left panels refer to CCSs in the mid-21st century (2021 – 2050), the right panels to the end-21st century (2070 – 2099). Blue colors indicate areas where the uncorrected model is colder than the corrected, red colors vice-versa.

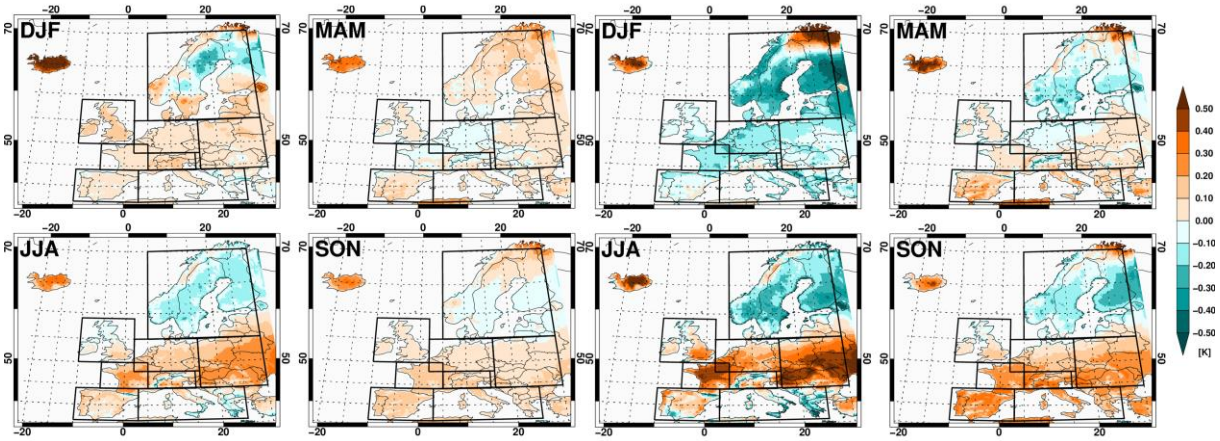

Figure 3. Differences between uncorrected and corrected (QM) multi-model standard deviation. The reference period is 1971 – 2000. The left panels refer to CCSs in the mid-21st century (2021 – 2050), the right panels to the end-21st century (2070 – 2099). Blue colors indicate areas where the uncorrected ensemble features a smaller standard deviation, orange colors vice-versa.

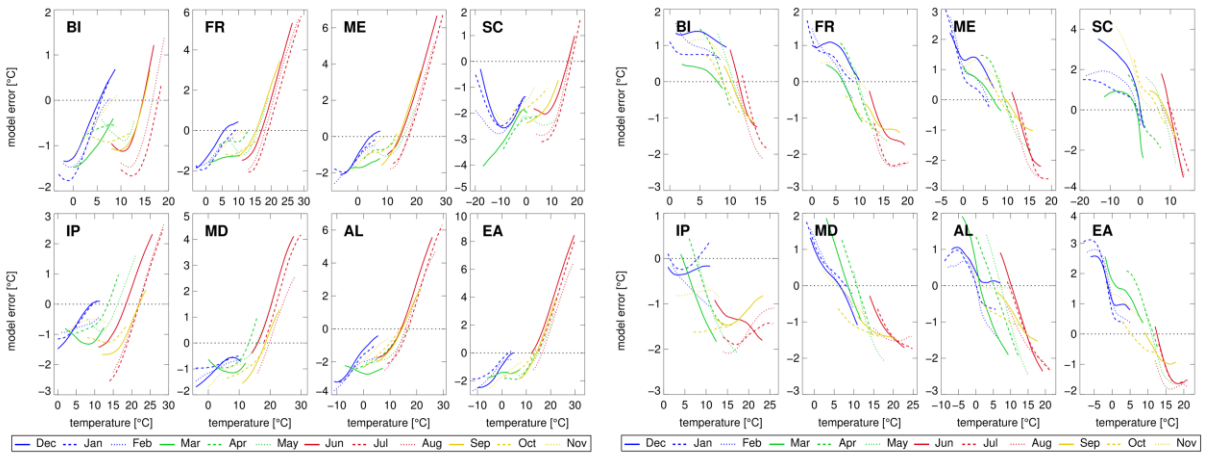

Figure 4. Temperature error characteristics (model minus observation) of the HC (left panels) and SMHI (right panels) RCMs in eight sub-regions of Europe (sub-panels) and each month of the year.

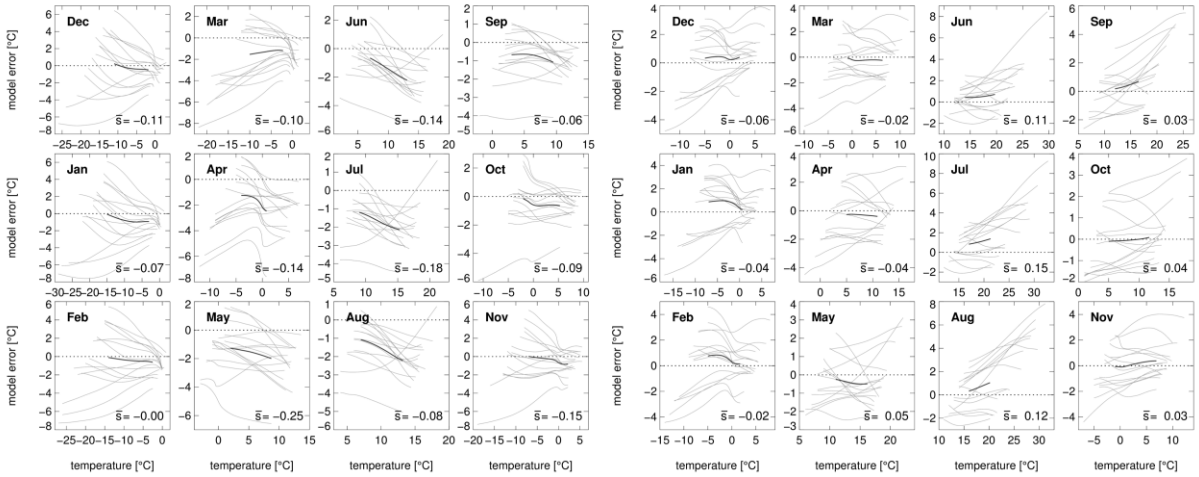

Figure 5. Temperature error characteristics (modeled minus observed) of the ENSEMBLES models in SC (left panels) and EA (right panels). The light lines show the error characteristics of the individual models, the bold line shows the ensemble average. The number in the lower right corner of each panel denotes multi-model average error slope.

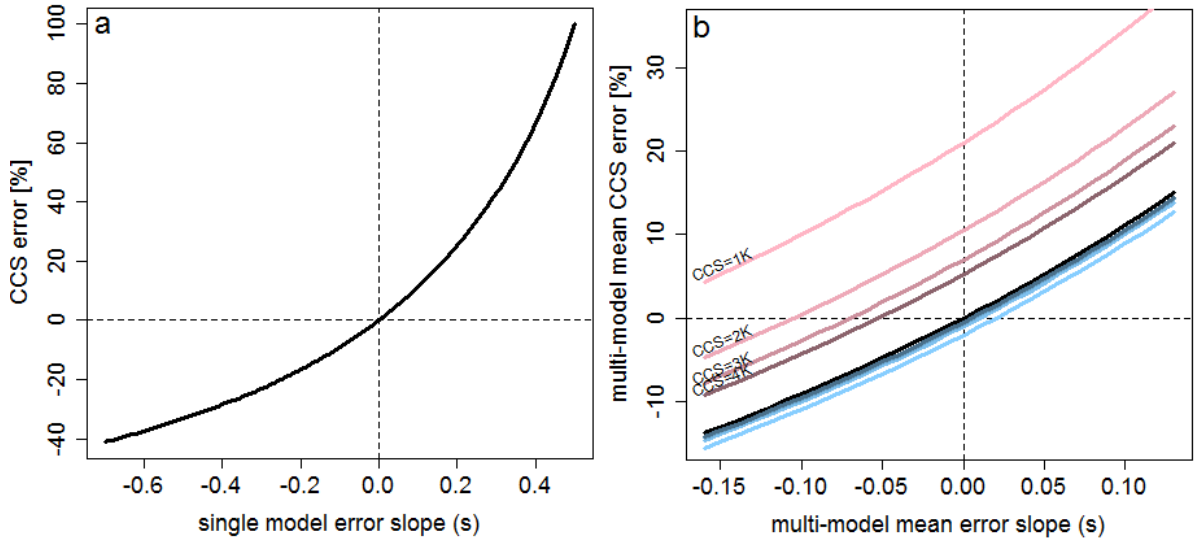

Figure 6. a: Effect of the error slope on the single model CCS. b: Effect of the error slope on the multi-model mean CCS. Black line: covariance term = 0 K; blue lines: covariance term = –0.02 K; pink lines: covariance term = +0.21 K. The lightest colors correspond to an error free CCS of 1 K, the darkest colors to a CCS of 4 K.

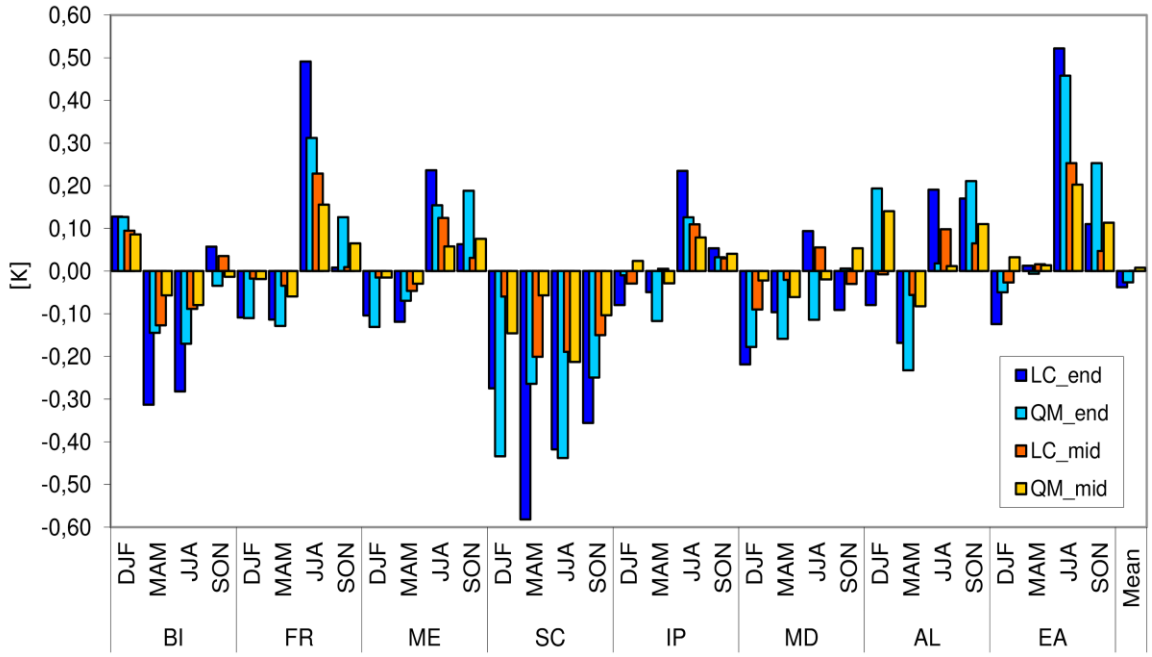

Figure 7. Estimated errors in the multi-model mean CCS due to intensity-dependent model errors. The reference period is 1971 – 2000. The orange colors refer to CCSs in the mid-21st century (2021 – 2050), the blue colors to the end-21st century (2070 – 2099). Light colors correspond to estimation of the error by QM, dark colors to LC.

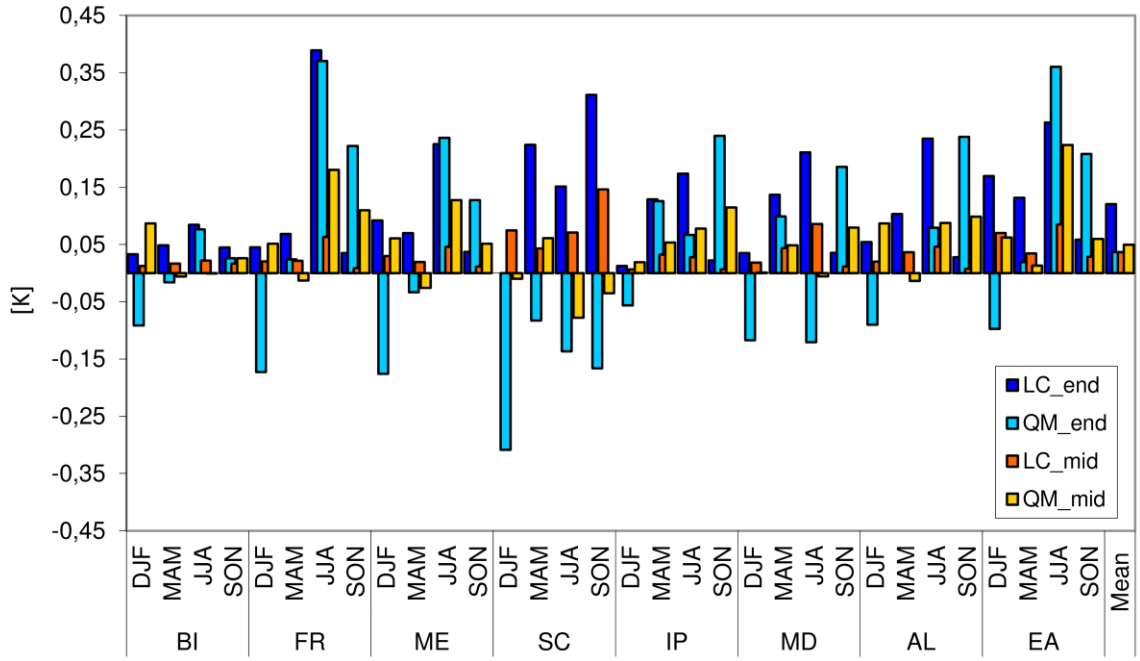

Figure 8. Estimated errors in the multi-model standard deviation of the temperature CCS due to intensity-dependent model errors. The reference period is 1971 – 2000. The orange colors refer to CCSs until the mid-21st century (2021 – 2050), the blue colors until the end-21st century (2070 – 2099). Light colors correspond to estimation of the error by QM, dark colors to LC.