AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER

10.1002/2016GL067721

Key Points:

- Bias modifications under strong climate forcing may be very large
- Nonmonotonic dependence of biases as a function of the climate forcing is found
- Variance correction schemes do not improve the future climate variability either

Supporting Information:

- Supporting Information S1

Correspondence to:
B. Van Schaeybroeck,
bertvs@meteo.be

Citation:
Van Schaeybroeck, B., and S. Vannitsem (2016), Assessment of calibration assumptions under strong climate changes, *Geophys. Res. Lett.*, *43*, 1314–1322, doi:10.1002/2016GL067721.

Received 8 JAN 2016
Accepted 27 JAN 2016
Accepted article online 1 FEB 2016
Published online 13 FEB 2016

Assessment of calibration assumptions under strong climate changes

Bert Van Schaeybroeck[1] and Stéphane Vannitsem[1]

[1] Royal Meteorological Institute Belgium, Brussels, Belgium

Abstract Climate model calibration relies on different working hypotheses. The simplest bias correction or delta change methods assume the invariance of bias under climate change. Recent works have questioned this hypothesis and proposed linear bias changes with respect to the forcing. However, when the system experiences larger forcings, these schemes could fail. Calibration assumptions are tested within a simplified framework in the context of an intermediate complexity model for which the reference (or "reality") differs from the model by a single parametric model error and climate change is emulated by largely different CO_2 forcings. It appears that calibration does not add value since the variation of bias under climate change is nonmonotonous for almost all variables and large compared to the climate change and the bias, except for the global temperature and sea ice area. For precipitation, calibration provides added value both globally and regionally. The calibration methods used fail to correct climate variability.

1. Introduction

Large discrepancies are known to exist between climatologies of global circulation models (GCMs) and the observed climatology, for the twentieth century, and climate predictions suffer from many shortcomings including model errors (associated for instance to the coarse resolution and subgrid scale modeling), initial condition errors, and scenario or forcing errors. Fortunately, part of the model errors is in some way systematic and this is why calibration has become a common practice [*Meehl et al.*, 2014; *Maraun et al.*, 2010; *Goddard et al.*, 2012].

However, calibration is usually based on assumptions such as the insensitivity to climate change of systematic biases, $\mu_M - \mu_O$, where μ_M and μ_O are the mean of the model variable M and the observation O, respectively. Recent works reassessed this hypothesis, in particular because strong forcing changes are prescribed for long-term (70–100 year) future projections [*Hawkins et al.*, 2014; *Ho et al.*, 2012; *Kerkhoff et al.*, 2014]. In such a situation, nontrivial and nonlinear feedback mechanisms were identified, in particular associated with cryosphere dynamics [*Bracegirdle and Stephenson*, 2012; *Caldeira and Cvijanovic*, 2014] and soil moisture [*Bellprat et al.*, 2013], that are particularly sensitive to the *absolute* value of the variable M under consideration (e.g., temperature) and not to its anomaly, $M - \mu_M$. These works argued that the hypothesis of bias stationarity—which implies that only anomalies must be considered as important—is violated. New correction schemes were then proposed that were based on *monotonous* (linear or step-wise linear) changes [*Buser et al.*, 2009; *Bracegirdle and Stephenson*, 2012; *Bellprat et al.*, 2013] as a function of (lead) time, [*Kharin et al.*, 2012] the state of the climate [*Bellprat et al.*, 2013], and/or the initial observations [*Fučkar et al.*, 2014].

However, it remains unclear to what extent these correction schemes that are tuned on a period of good observational coverage (typically less than 30 years) can be applied on long-term predictions for 70 to 100 years. In this context *Vannitsem* [2011] showed that for different low-order systems, the assumption of *monotonous* bias evolution fails as biases display *nonmonotonous* and irregular variations under climate changes. Therefore, a straightforward application of bias corrections from one system's state (i.e., climate) to another fails. The purpose of the present work is to explore whether the conclusions drawn in *Vannitsem* [2011] for low-order systems extend to Earth-system models of intermediate complexity (EMICs). As compared to GCMs, the physical modules of these models are limited and involve important simplifications. They, however, incorporate the dominant processes with the orographic and the external astronomical forcings in particular, although at low spatial and temporal resolution. Note that it is not the intent of this paper to unravel the physical mechanism underlying the biases but rather to point out and verify the assumptions associated with calibration.

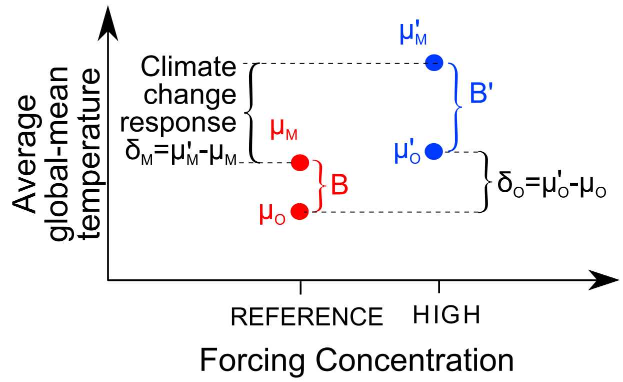

Figure 1. Pictorial representation of global mean biases (*B*) as the difference between average model prediction (μ_M) and observation (μ_O) for two climate systems characterized by two different atmospheric forcing concentrations for CO_2 or other greenhouse gases. The climate change signals δ_M and δ_O that are the difference between the climate prediction of the high and reference concentration runs of model and observations, respectively, are also defined.

To investigate this question, a twin experiment approach [*Maraun*, 2012; *Bellprat et al.*, 2013] is adopted using a climate model of reduced complexity. This choice is made to get a clear answer, overcoming the issue of shortage of observational data and the additional difficulties of state-of-the-art GCMs, namely, the restricted length and number of runs [*Hawkins et al.*, 2014] and the slow drifts that persist in GCM runs due to the limited convergence time of the historical runs toward their natural climatological variability [*Gupta et al.*, 2013].

In section 2 different calibration methods and their underlying assumptions are introduced and error measures are proposed that allow to probe the violation of these assumptions. The climate model setup is explained in section 3, and results are presented in section 4. Finally, a conclusion of the results is given in section 5.

2. Calibration and Underlying Assumptions

The aim of calibration is to produce a *corrected or calibrated climate distribution* based on a distribution of reference observations *O*, reference predictions *M*, and future predictions M'. The variables *O*, *M*, and M' can be either local scalar variables or vectors containing several or all the variables of the model and the reality. Four calibration methods are used here to test the different hypotheses and, along with a discussion on the quantile mapping approach, are detailed in Appendix A. The first calibration method is called *bias correction* (BC) and is a very common method of calibration in climatology. In its most simple form, here called BC1, it affects the *distribution of the future predictions* M' by shifting it over a distance $-B$ [*Wilks*, 2006]. Here *B* is the *bias* of the "reference" model and is equal to the difference between a model mean μ_M and the observational mean μ_O, averaged over the same period of time with constant reference climate conditions, $B = \mu_M - \mu_O$. However, the bias may change depending on the climate conditions, including greenhouse gas forcing. Assuming these conditions are known in the future, one defines the (true) future model bias B' as the difference between the model average μ'_M and the future observations μ'_O. The *first bias correction hypothesis* is therefore (see also Figure 1)

$$\text{BC1 hypothesis:}\quad B' = B. \tag{1}$$

Another calibration method that corrects the mean climatology is the *change factor* (CF) or delta change method. As opposed to bias correction, this method affects the *distribution of the reference observations* and is based on the climate change response δ instead of *B*. The response of the model (δ_M) refers to the difference between the mean future prediction μ'_M and the mean reference model run μ_M [*Ho et al.*, 2012; *Bracegirdle and Stephenson*, 2012] or $\delta_M = \mu'_M - \mu_M$. In order to obtain the corrected future predictions, the simplest CF method, CF1, adds δ_M to the reference observations, and hence, the *first change factor hypothesis* assumes equal climate change response for both model and observations or in other words

$$\text{CF1 hypothesis:}\quad \delta_M = \delta_O, \tag{2}$$

where $\delta_O = \mu'_O - \mu_O$. It is easily verified that the BC1 and CF1 hypotheses are equivalent; that is, equation (1) implies equation (2) and vice versa. However, in case the hypothesis is violated, strongly dissimilar results can be obtained by applying BC and CF as discussed in *Ho et al.* [2012].

A second category of schemes are designed to correct also the climate variance. The corresponding bias correction scheme, BC2, thereby modifies the standard deviation of the *distribution of the future predictions* by a factor σ_O/σ_M [*Vannitsem and Hagedorn*, 2011; *Van Schaeybroeck and Vannitsem*, 2011; *Ho et al.*, 2012]. Similarly, the change factor method CF2 applies a scaling factor σ'_M/σ_M to the *distribution of the reference observations*. *The second bias correction hypothesis* is therefore the equality between the variances of the future observations and the corrected predictions:

$$\text{CF2 hypothesis} = \text{BC2 hypothesis:} \;\; \sigma_O\sigma'_M = \sigma'_O\sigma_M. \tag{3}$$

Note that BC2 and CF2 also correct the climate mean.

The main consequences of the correction schemes can be quantified by comparing the averages and the variances of the calibrated distributions with the one of "true" future climate that will be simulated in our twin experiments described in section 3. The error E_μ is defined as the error on the climatological mean. For all calibration methods introduced here, it is easily found that

$$E_\mu = B - B' = \delta_M - \delta_O. \tag{4}$$

Therefore, E_μ is identical for the bias correction and the change factor method. The BC1 and CF1 hypotheses are satisfied when $E_\mu = 0$. In practice, of course, this is never exactly valid and in order to quantify the extent of hypotheses violation, this quantity must be compared to other quantities. First, the bias change E_μ can be considered irrelevant when it is small as compared to the climate change signal δ_M. This implies that the quantity of most interest, i.e., the climate change signal, is not strongly affected by a change of bias. Second, even if the bias change is large as compared to the climate change signal, calibration may still improve the prediction provided that the bias change is small as compared to the bias itself. So the BC1 or CF1 hypotheses are satisfactory if

$$\left|E_\mu/\delta_M\right| \ll 1, \quad \text{and} \quad \left|E_\mu/B\right| \ll 1. \tag{5}$$

The error on the standard deviation of the climate can be checked by a direct comparison of the standard deviation of the calibrated prediction σ'_{cal} with the one of the observed climate σ'_O. Calibration hypotheses BC2 and CF2 are therefore satisfied when

$$\left|\sigma'_{\text{cal}}/\sigma'_O - 1\right| \ll 1. \tag{6}$$

Since BC1 and CF1 simply shift the future model climate distribution and the reference observational distribution, their climate standard deviations are equal to σ'_M and σ_O, respectively. The variance-corrected BC2 and CF2 climate distributions, on the other hand, have equal standard deviation, $\sigma'_M\sigma_O/\sigma_M$. Note that higher-order characteristics of the distribution such as skewness and kurtosis and power law tails remain unaffected by the calibration methods. The validity of conditions (5) and (6) will be evaluated in section 4.

A detailed comparison between bias correction and delta change methods for statistical downscaling in a changing climate has been performed by *Willems and Vrac* [2011] and *Räty et al.* [2014].

3. Model and Setup

In the twin experiment performed in the present work the reality, the reference, and the "model" are generated with the same model LOVECLIM. However, both differ by a single parametric model error. The parameter chosen, expir, is the exponent relating the anomaly of the longwave radiation flux with the anomaly in humidity (further details can be found in *Goosse et al.* [2010]). The reality assumes a fixed parameter of the radiation scheme expir = 0.4, while for the model versions it is fixed within its potential range of uncertainty [0.35, 0.45] as characterized by *Loutre et al.* [2011]. Eleven model versions are then built with different values of expir: 0.35, 0.36, … , 0.44, and 0.45. For each model version and for all three climate forcing regimes, the statistical properties are derived using runs of 8000 years inducing negligibly small confidence bounds. Such long time series allow for getting good statistics. Moreover, by discarding data from an initial relaxation period of 2000 years, long-term trends of the ocean dynamics are discarded [*Gupta et al.*, 2013]. Three climate forcing regimes are considered different by their atmospheric CO_2 concentrations: a *preindustrial climate* (276 ppm), a *reference climate* with CO_2 concentrations of the year 2000 (370 ppm), and a *future climate* with doubled concentration of the reference climate (740 ppm). Note that the same analysis is repeated using a

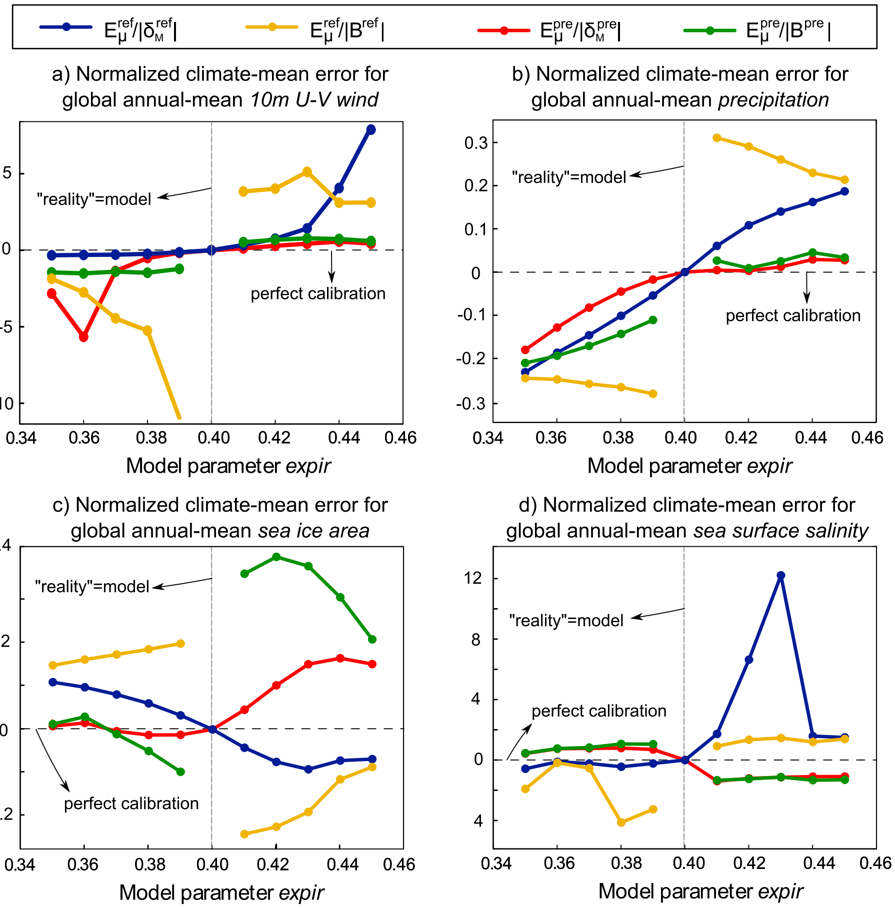

Figure 2. The error of the reference and future climatological means (E_μ^{pre} and E_μ^{ref}) divided by the absolute climate change response $|\delta_M|$ (red and blue lines) or divided by the biases (green and yellow lines) as a function of the model parameter expir. The results are for (a) 10 m U-V wind, (b) precipitation, (c) sea ice area, and (d) sea surface salinity. Note that when expir = 0.40, the model corresponds exactly to the truth such that the bias is zero and $1/B$ is not defined.

parametric error in the wind drag instead of the radiative scheme. Moreover, results with five instead of three climate regimes were considered (see Figures S1 and S2 in the supporting information). All these experiments lead to similar conclusions.

4. Results

Three important questions are addressed here based on a twin experiment with a single model error: (1) Upon strong change of the climate forcing, are bias changes large as compared to the climate change signal and the bias? (2) Do climate variance correction schemes effectively correct the variance? (3) Do biases change monotonically as a function of the forcing parameter?

The first calibration hypothesis (BC1 and CF1) is tested in Figure 2 for four global mean variables and for two different climate change configurations. Different models are considered by a change of the parameter expir. The quantities $E_\mu^{ref}/|\delta_M^{ref}|$ and $E_\mu^{ref}/|B^{ref}|$ refer to climate mean errors of the future climate with a calibration based on the reference climate, while $E_\mu^{pre}/|\delta_M^{pre}|$ and $E_\mu^{pre}/|B^{pre}|$ are the errors of the reference climate with the preindustrial era used as the calibration period. First, for 10 m U-V wind (Figure 2a) and sea surface salinity (Figure 2d), most of the normalized errors are of order 1 for both climate change configurations. This implies a strong violation of the first calibration hypothesis (see equation (5)) such that performing a bias correction is not meaningful and does not generally improve the prediction.

No simple monotonic behavior as a function of the model error amplitude and of the forcing parameter can be found as illustrated in Figure 2. For instance, the curves for $E_\mu^{pre}/|\delta_M^{pre}|$ (red line) and $E_\mu^{ref}/|\delta_M^{ref}|$ (blue line) have opposite signs for sea ice area (Figure 2c) and sea surface salinity (Figure 2d) and complicated

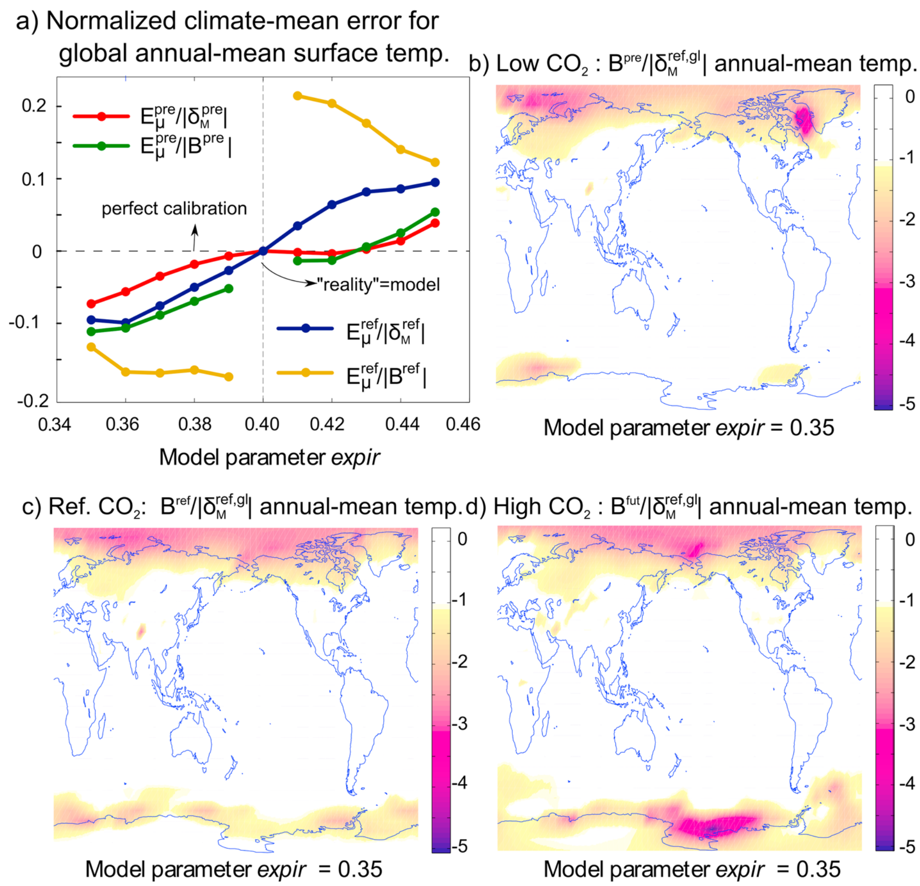

Figure 3. (a) Same as Figure 2 but for surface temperature. (b–d) Spatial patterns of the relative biases of surface temperature corresponding to three different climate regimes: $B^{\mathrm{pre}}/|\delta_M^{\mathrm{ref,gl}}|$ is the preindustrial bias, $B^{\mathrm{ref}}/|\delta_M^{\mathrm{ref,gl}}|$ the bias of the reference regime and $B^{\mathrm{fut}}/|\delta_M^{\mathrm{ref,gl}}|$ the future bias. The climate change signal $\delta_M^{\mathrm{ref,gl}}$ is the global mean of the climate change signal (reference minus future climate) and therefore does not depend on position.

dependences as a function of the model error amplitude. Therefore, the hypothesis that the bias changes linearly with temperature or forcing as found for near-future climate changes fails for long-term climate changes.

For precipitation and sea ice area (Figures 2b and 2c), the normalized errors are of the order of 10%–20% and this is also the case for global mean surface temperature displayed in Figure 3a. Such deviations are substantial but may be considered as acceptable in order to apply bias correction, although here again an opposite trend in biases is found for sea ice extent. However, a closer look indicates that the coincidence of global biases among the different model systems does not derive from the same response mechanisms as illustrated by the spatial bias distributions of the surface temperature for the three climate regimes in Figures 3b–3d. The largest biases of the preindustrial and reference regime resemble one another in the sense that they are mostly situated around the Arctic. Such agreement would encourage the application of calibration to a future era. However, the largest temperature biases of the future regime are located on the Antarctic. The same conclusions are drawn for sea ice area (not shown). Therefore, although the calibration assumptions are reasonably well satisfied for global mean quantities (see Figure 3a), strong violations appear when applied on a more regional scale (see Figure S3 in the supporting information).

Figure 4 shows the ratio of the standard deviation of the corrected climate distribution with the observed one for global mean annual temperature. As concluded in equation (6), this ratio must be approximately equal to one for a well-calibrated prediction. Only the BC2 and CF2 methods calibrate the climate variance, but these methods do not improve systematically the variance. On the contrary, from Figure 4a it appears that the variance of the BC1-corrected prediction of the reference climate (which is equal to the uncorrected one) almost

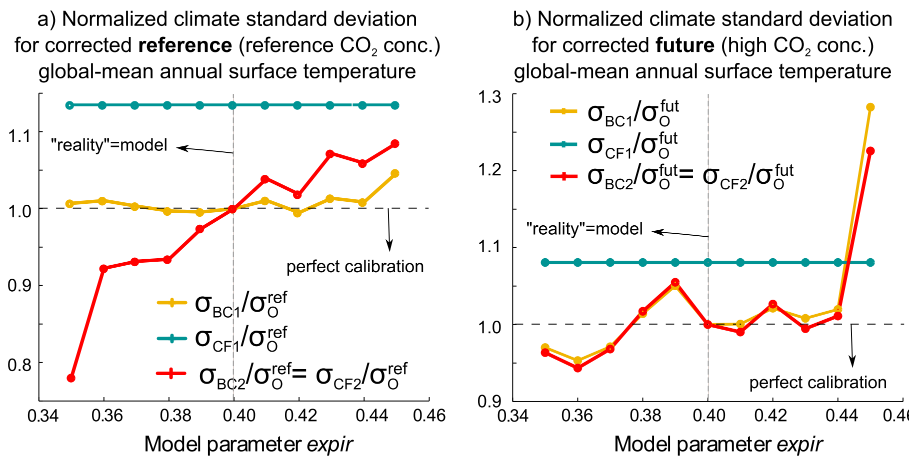

Figure 4. Climate variance of global annual mean surface temperature after calibration normalized by the climate variance of the current climate (a) σ_O^{ref} or (b) σ_O^{fut}. Perfect calibration is found when the ratio is equal to one. Figure 4a shows result for the reference climate where calibration is done using preindustrial climate. Figure 4b shows results for the future climate, calibrated using the reference climate.

coincides with the observed one but this conclusion does not extend to the future climate (see Figure 4b) where BC1 is not better than the other methods. Similar consideration of variance calibration can be made for other climate variables (not shown).

The physical interpretation of the origin of these biases is beyond the scope of this brief summary but will be explored in more detail in the future.

5. Conclusions

As demonstrated in this work, bias modifications under strong climate forcing are very large for most of the variables considered, even for small model errors. Moreover, nonmonotonic variations of these biases as a function of the climate forcings are found that could considerably impact the approaches developed for correcting the model outputs. Recently, monotonic dependences of biases as a function of key variables (e.g., temperature) reflecting the amplitude of the climate forcing (e.g., greenhouse gases) have been proposed as supported by the investigations of current and past behavior of the models [*Christensen et al.*, 2008; *Boberg and Christensen*, 2012; *Kerkhoff et al.*, 2014]. In light of the present results, this approach should be reassessed and care must be taken when extrapolating the use of linear bias corrections (as well as constant bias corrections) to regimes of large climate forcings.

A possible alternative is to adopt a probabilistic approach of the potential biases based, for instance, on an ensemble of possible configurations of a specific GCM and to consider this distribution as an uncertainty measure. The change of this distribution as a function of the climate forcing may then represent the uncertainty of the future climate. If this distribution does not change much as a function of the climate forcing, we have at our disposal a realistic measure of the uncertainty. This analysis can be performed in the context of an EMIC as the one used here.

Note that the spread of a multimodel ensemble projection under climate forcing (such as the one based on Coupled Model Intercomparison Project Phase 5 models) does provide an indication of the impact of model biases present in the reference system, provided that climate forcings and the impact of the initial conditions of the slow components of the system are the same among the different members. But a larger number of members (based on several versions of the same models) would be necessary to explore the ensemble of possible scenarios. The question of the number of members necessary for a reliable ensemble projection would be worth addressing in the context of an EMIC like LOVECLIM in a transient climate change mode and under the presence of a sample of "realistic" model errors.

Note that the present analysis does not take into account spatial and/or intervariable dependences [*Vrac and Friederichs*, 2014], which is worth addressing in the future. In addition, only the impact of small model errors on one parameter were considered. One can expect that in the presence of multiple and large model errors together with initial condition and forcing errors, the behavior will be even more involved. This will be also be explored in the future.

Appendix A: Calibration Methods

Details of the different calibration methods are given in this Appendix. The first bias correction (BC) scheme, named BC1, corrects the climate mean [*Wilks*, 2006], while BC2 corrects the climate variance [*Vannitsem*, 2009; *Vannitsem and Hagedorn*, 2011; *Van Schaeybroeck and Vannitsem*, 2011]). Both methods take each prediction M' from the distribution of the future model runs and change them as follows:

$$O'_{\mathrm{BC1}} = M' - B, \tag{A1a}$$

$$O'_{\mathrm{BC2}} = O'_{\mathrm{BC1}} + \left(\frac{\sigma_O}{\sigma_M} - 1\right)\left(M' - \mu'_M\right). \tag{A1b}$$

For variables featuring bounded values such as precipitation or sea ice area, shifted values that exceed the bounds must be projected onto the bounds [*Barth et al.*, 2015]. In general, the shift will not be equal to the bias but must be such that when applied on the current climate distribution, the mean of the corrected distribution coincides with the mean of the current climate observations. Alternatively, one could apply bias correction after a logarithmic transformation, an approach that is equivalent to a multiplicative correction [*Maraun et al.*, 2010].

Lastly, we introduce BC using quantile mapping (BC-QM), again correcting the future model climate distribution based on a relation between the climate model observational distributions for the current-day climate. Instead of applying a shift and a linear inflating/deflating as with BC2, the BC-QM transformations are based on relations between the distributional quantile functions $O(q)$ and $F(q)$ with $q \in [0, 1]$. The most commonly used approach is to map a set of quantiles of the model distribution to the corresponding ones of the observation, supplemented by additional assumptions outside the range of these quantiles, *Maraun et al.* [2010]. The underlying assumptions are therefore difficult to assess. Here we propose the use of a ordinary least squares (OLS) regression using $O(q)$ and $F(q)$. Since, in the absence of a model error, $O(q) = F(q)$ for all q, the simplest parametric transformation fits the difference $O(q) - F(q)$ against $F(q)$. The quantile mapping with one regression parameter, i.e., approximating $F(q) - O(q)$ by a constant value using OLS regression, is exactly equivalent with BC1. Therefore, the BC-QM approach under consideration fits $O(q) - F(q)$ using two regression parameters, one of which represents a constant shift and one that multiplies $F(q)$.

In Figure S4 in the supporting information the equivalent of Figure 3a is shown with corrections applied by quantile mapping. Whereas the conditions of equation (5) are reasonably satisfied for BC1, BC2, CF1, and CF2, this is no longer the case when using BC-QM. Therefore, this addition strengthens our conclusion that calibration introduces additional uncertainty in case of large climate changes.

Likewise, the change factor method CF1 corrects only the climate mean, whereas CF2 also adapts the variance. Both correct each reference observations O:

$$O'_{\mathrm{CF1}} = O + \delta_M, \tag{A2}$$

$$O'_{\mathrm{CF2}} = O'_{\mathrm{CF1}} + \left(\frac{\sigma'_M}{\sigma_M} - 1\right)\left(O - \mu_O\right). \tag{A3}$$

The main consequences of the correction schemes can be quantified by the errors on the averages and the variances of the calibrated distributions of the future predictions. The error on the mean is denoted here as E_μ and for the BC and CF methods is given in equation (4).

The standard deviations of the BC and CF calibrated climate distributions read

$$\sigma_{\mathrm{BC1}} = \sigma'_M \tag{A4a}$$

$$\sigma_{\mathrm{CF1}} = \sigma_O, \tag{A4b}$$

$$\sigma_{\mathrm{BC2}} = \sigma_{\mathrm{CF2}} = \frac{\sigma'_M \sigma_O}{\sigma_M}. \tag{A4c}$$

These must be compared with the true standard deviation of the future climate σ'_O in order to see to what extent calibration improves the prediction.

Note that the variance correction methods used here differ from the ones used in *Ho et al.* [2012] for which

$$O'_{\mathrm{BC3}} = O'_{\mathrm{BC1}} + \left(\frac{\sigma_O}{\sigma_M} - 1\right)\left(M' - \mu_M\right), \tag{A5a}$$

$$O'_{\mathrm{CF3}} = O'_{\mathrm{CF1}} + \left(\frac{\sigma'_M}{\sigma_M} - 1\right)\left(O - \mu_M\right). \tag{A5b}$$

Both methods correct the variance in the same way as CF2 and BC2 (see equation (A4c)); however, the error of the climate mean is different:

$$E_{\mu,\mathrm{BC3}} = \delta_M\left(\frac{\sigma_O}{\sigma_M}\right) - \delta_O, \tag{A6a}$$

$$E_{\mu,\mathrm{CF3}} = B' - B\left(\frac{\sigma'_M}{\sigma_M}\right). \tag{A6b}$$

We have tested these calibration schemes, and their implementation also supports our findings.

Acknowledgments
The authors wish to thank Violette Zunz and Hugues Goosse for their suggestions and their kind assistance with the use of LOVECLIM. Discussions with Olivier Giot are acknowledged. The comments of two anonymous referees are highly appreciated. This work is partially supported by the Belgian Federal Science policy office under contract SD/CA/04A. Information on the model LOVECLIM and the model itself are available at http://www.elic.ucl.ac.be/modx/elic/index.php?id=81. The simulation data necessary to reproduce the results are available from the authors upon request (bertvs@meteo.be) and are archived at the Royal Meteorological Institute of Belgium (RMIB). The analysis has been carried out using MATLAB.

References

Barth, A., M. Canter, B. Van Schaeybroeck, S. Vannitsem, F. Massonnet, V. Zunz, P. Mathiot, and J.-M. Beckers (2015), Assimilation of sea surface temperature, sea ice concentration and sea ice drift in a model of the Southern Ocean, *Ocean Model.*, *93*, 22–39.

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

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

Bracegirdle, T. J., and D. B. Stephenson (2012), Higher precision estimates of regional polar warming by ensemble regression of climate model projections, *Clim. Dyn.*, *39*(12), 2805–2821, doi:10.1007/s00382-012-1330-3.

Buser, C. M., H. R. Künsch, D. Lüthi, M. Wild, and C. Schär (2009), Bayesian multi-model projection of climate: Bias assumptions and interannual variability, *Clim. Dyn.*, *33*(6), 849–868, doi:10.1007/s00382-009-0588-6.

Caldeira, K., and I. Cvijanovic (2014), Estimating the contribution of sea ice response to climate sensitivity in a climate model, *J. Clim.*, *27*(22), 8597–8607, doi:10.1175/JCLI-D-14-00042.1.

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher (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.

Fučkar, N. S., D. Volpi, V. Guemas, and F. J. Doblas-Reyes (2014), A posteriori adjustment of near-term climate predictions: Accounting for the drift dependence on the initial conditions, *Geophys. Res. Lett.*, *41*, 5200–5207, doi:10.1002/2014GL060815.

Goddard, L., et al. (2012), A verification framework for interannual-to-decadal predictions experiments, *Clim. Dyn.*, *40*, 245–272, doi:10.1007/s00382-012-1481-2.

Loutre, M. F., A. Mouchet, T. Fichefet, H. Goosse, H. Goelzer, and P. Huybrechts (2011), Evaluating climate model performance with various parameter sets using observations over the recent past, *Clim. Past*, *2*(7), 511–526.

Goosse, H., et al. (2010), Description of the Earth system model of intermediate complexity LOVECLIM version 1.2, *Geosci. Model Dev.*, *3*(2), 603–633, doi:10.5194/gmd-3-603-2010.

Gupta, A. S., N. C. Jourdain, J. N. Brown, and D. Monselesan (2013), Climate drift in the CMIP5 models, *J. Clim.*, *26*(21), 8597–8615.

Hawkins, E., B. Dong, J. Robson, R. Sutton, and D. Smith (2014), The interpretation and use of biases in decadal climate predictions, *J. Clim.*, *27*(8), 2931–2947, doi:10.1175/JCLI-D-13-00473.1.

Ho, C. K., D. B. Stephenson, M. Collins, C. a. T. Ferro, and S. J. Brown (2012), Calibration strategies: A source of additional uncertainty in climate change projections, *Bull. Am. Meteorol. Soc.*, *93*(1), 21–26, doi:10.1175/2011BAMS3110.1.

Kerkhoff, C., H. R. Künsch, and C. Schär (2014), Assessment of bias assumptions for climate models, *J. Clim.*, *27*, 6799–6818, doi:10.1175/JCLI-D-13-00716.1.

Kharin, V. V., G. J. Boer, W. J. Merryfield, J. F. Scinocca, and W.-S. Lee (2012), Statistical adjustment of decadal predictions in a changing climate, *Geophys. Res. Lett.*, *39*, L19705, doi:10.1029/2012GL052647.

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

Maraun, D., et al. (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.

Meehl, G., et al. (2014), Decadal climate prediction: An update from the trenches, *Bull. Am. Meteorol. Soc.*, *95*(2), 243–267, doi:10.1175/BAMS-D-12-00241.1.

Räty, O., J. Räisänen, and J. S. Ylhäisi (2014), Evaluation of delta change and bias correction methods for future daily precipitation: Intermodel cross-validation using ENSEMBLES simulations, *Clim. Dyn.*, *42*(9–10), 2287–2303, doi:10.1007/s00382-014-2130-8.

Van Schaeybroeck, B., and S. Vannitsem (2011), Post-processing through linear regression, *Nonlinear Process. Geophys.*, *18*(2), 147–160, doi:10.5194/npg-18-147-2011.

Vannitsem, S. (2009), A unified Linear Model Output Statistics scheme for both deterministic and ensemble forecasts, *Q. J. R. Meteorol. Soc.*, *135*, 1801–1815.

Vannitsem, S. (2011), Bias correction and post-processing under climate change, *Nonlinear Process. Geophys.*, *18*(6), 911–924.

Vannitsem, S., and R. Hagedorn (2011), Ensemble forecast post-processing over Belgium: Comparison of deterministic-like and ensemble regression methods, *Meteorol. Appl.*, *18*(1), 94–104, doi:10.1002/met.217.

Vrac, M., and P. Friederichs (2014), Multivariate-intervariable, spatial, and temporal-bias correction, *J. Clim.*, *28*(1), 218–237.

Wilks, D. S. (2006), *Statistical Methods in the Atmosphere Sciences*, Academic Press, 464 pp., New-York.

Willems, P., and M. Vrac (2011), Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change, *J. Hydrol.*, *402*(3–4), 193–205, doi:10.1016/j.jhydrol.2011.02.030.