Bias correction of EU-ENSEMBLES precipitation data with focus on the effect of sample size

Universität Trier

Rheinland-Pfalz MINISTERIUM FÜR WIRTSCHAFT, KLIMASCHUTZ, ENERGIE UND LANDESPLANUNG

Philipp Reiter[1,2] (philipp.reiter@klimawandel-rlp.de),
Oliver Gutjahr[3], Lukas Schefczyk[3], Günther Heinemann[3], Markus Casper[2]

[1]Rhineland-Palatinate Centre of Excellence for Climate Change Impacts, Trippstadt, Germany [2]Department of Physical Geography, University of Trier, Germany [3]Department of Environmental Meteorology, University of Trier, Germany

Introduction

Bias correction (BC) has become a standard procedure in climate change impact studies due to biases in regional climate model (RCM) output that prevent a direct use. There are numerous assumptions and consequences that are connected with applying a BC. Several of those have been discussed and analysed in studies. The effect of the sample size used for BC calibration on BC performance has however so far not been looked into.

Especially in case of precipitation we expect a strong dependence of BC performance on the sample size. Therefore we apply state-of-the-art BC methods based on different sample sizes for calibration and compare their performances to that of a BC based on 30 years.

Data sets and domain

- precipitation data of 10 RCM reanalysis runs (EU-ENSEMBLES 25 km, 1961–2000, ERA-40)
- E-OBS observational data set (25 km, 1961–2000)
- Germany as area of interest

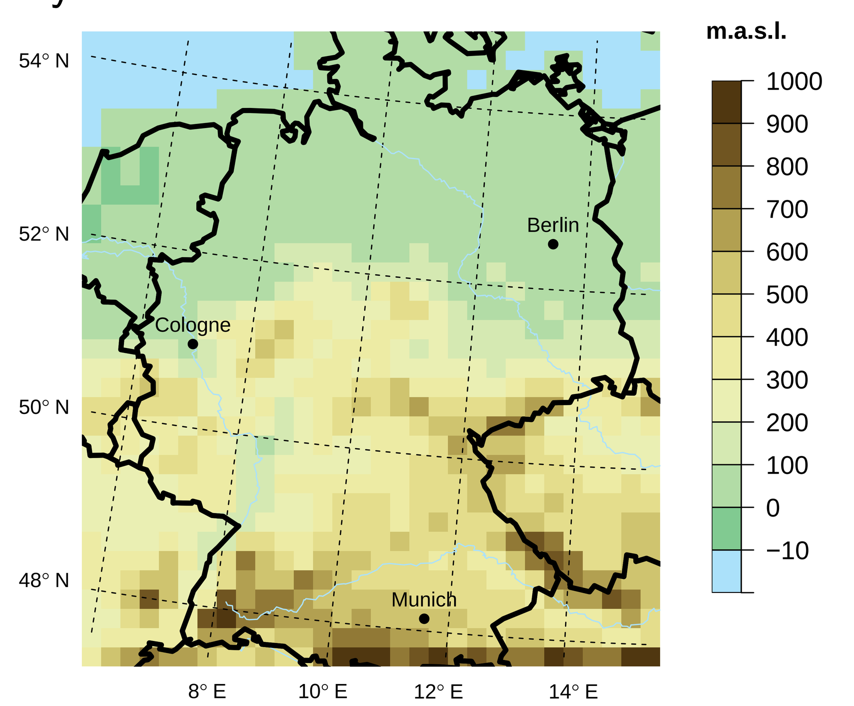


Figure 1: Orography of the domain (Germany and bordering areas) at 25 km.

Methods

BC: 4 quantile matching (QM) approaches

- *eQM*: empirical QM[1]
- *PTF*: 'Piani Transfer Function'[2]
- *gQM*: transfer function based on a gamma distribution[3]
- *GQM*: transfer function is a combination of gamma and generalized Pareto distribution (GPD)[4]

BC performance assessment: 3 skill scores

- *MAE* (mean absolute error)[5]
- *Perkins* skill score[6]
- Ext_{10} (measures deviation in the ten highest values)

Poster

vCard

Analysis procedure

Bias correction procedure

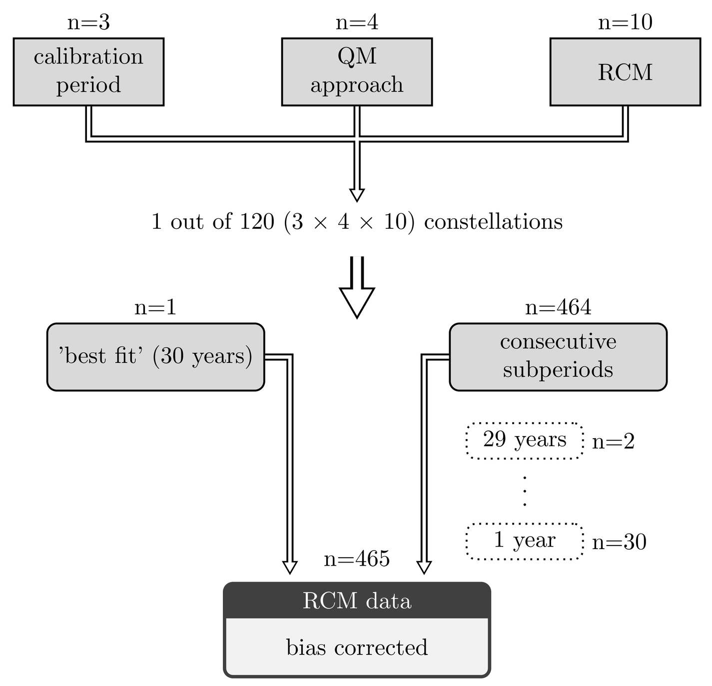


Figure 2: Flowchart showing the BC procedure for the different sample sizes.

- the 40 year period is split into 3 different combinations of a 30 year calibration period and a 10 year validation period
- 120 constellations of calibration period, QM approach and RCM
- for each of these constellations:
  - first of all a BC is done using the complete 30 years for calibration of the BC (the 'best fit' for the constellation)
  - next BCs are done with calibration based on all subperiods out of the calibration period
  - this generates 465 bias corrected data sets for each constellation

Assessment of BC performance

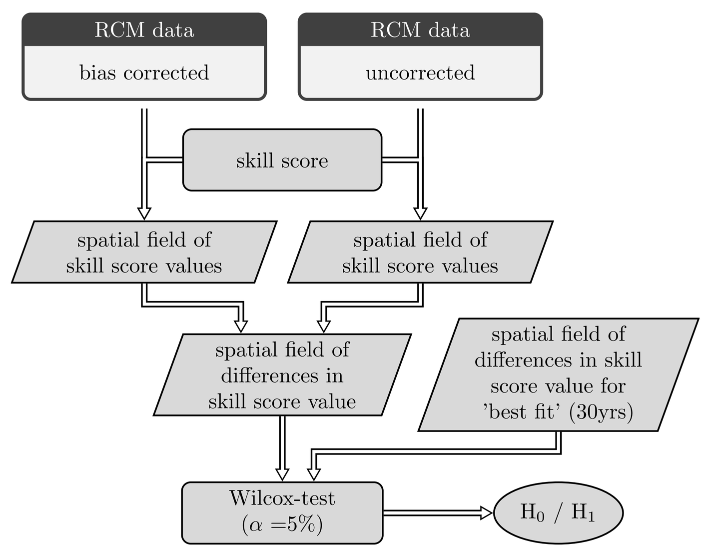


Figure 3: Flowchart showing the assessment of BC performance.

For every constellation (Fig. 2) and every skill score:

- the skill score is applied to the data of the validation period of all bias corrected data sets as well as of the uncorrected RCM data
- for every bias corrected data set the difference in skill score value to those for the uncorrected RCM data is calculated
- these difference fields are tested versus the 'best fit' (one-sided Wilcox-test, $\alpha = 5\,\%$)
- the critical sample size $\boldsymbol{n_{crit}}$ is the largest sample size that shows a significant decrease in skill score values
- n_{crit} quantifies the effect of sample size on BC performance

Bias correction (30 years)

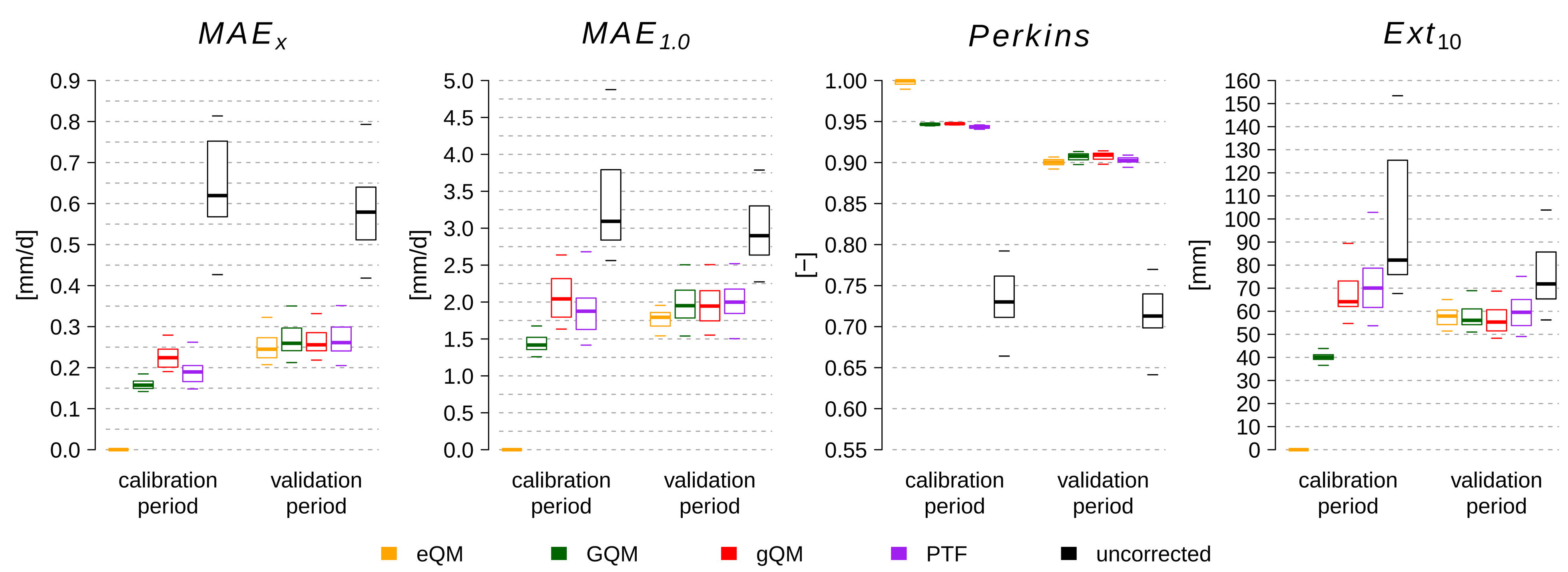
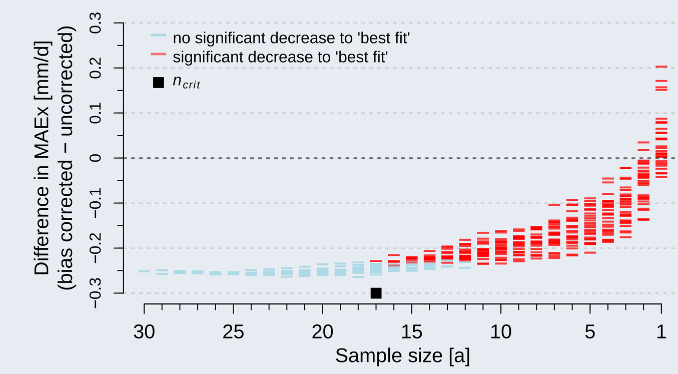


Figure 4: Boxplots of the absolute skill score values for the RCM data, both bias corrected and uncorrected. The results for the bias corrected RCM data are grouped by QM approach and refer to the bias correction run for the 'best fit'. Each boxplot consists of 30 spatial mean values (3 calibration periods, 10 RCMs).

- in the calibration period the more complex QM approaches (especially *eQM*, but also *GQM*) clearly outperform the less complex QM approaches
- in the validation period the performances of all QM approaches are on a comparable level

Effect of sample size on bias correction performance

Results for one exemplary constellation



- the results show a decreasing BC performance with decreasing sample size, especially for sample sizes smaller than 10 years
- for this exemplary constellation at a sample size of $n_{crit} = 17$ years the BC performance is, for the first time, worse than that of the 'best fit'

Figure 5: Medians of the spatial difference fields (Fig. 3; exemplary for calibration period 1961–1990, *eQM*, RCM REMO, skill score MAE_x).

Results for all constellations

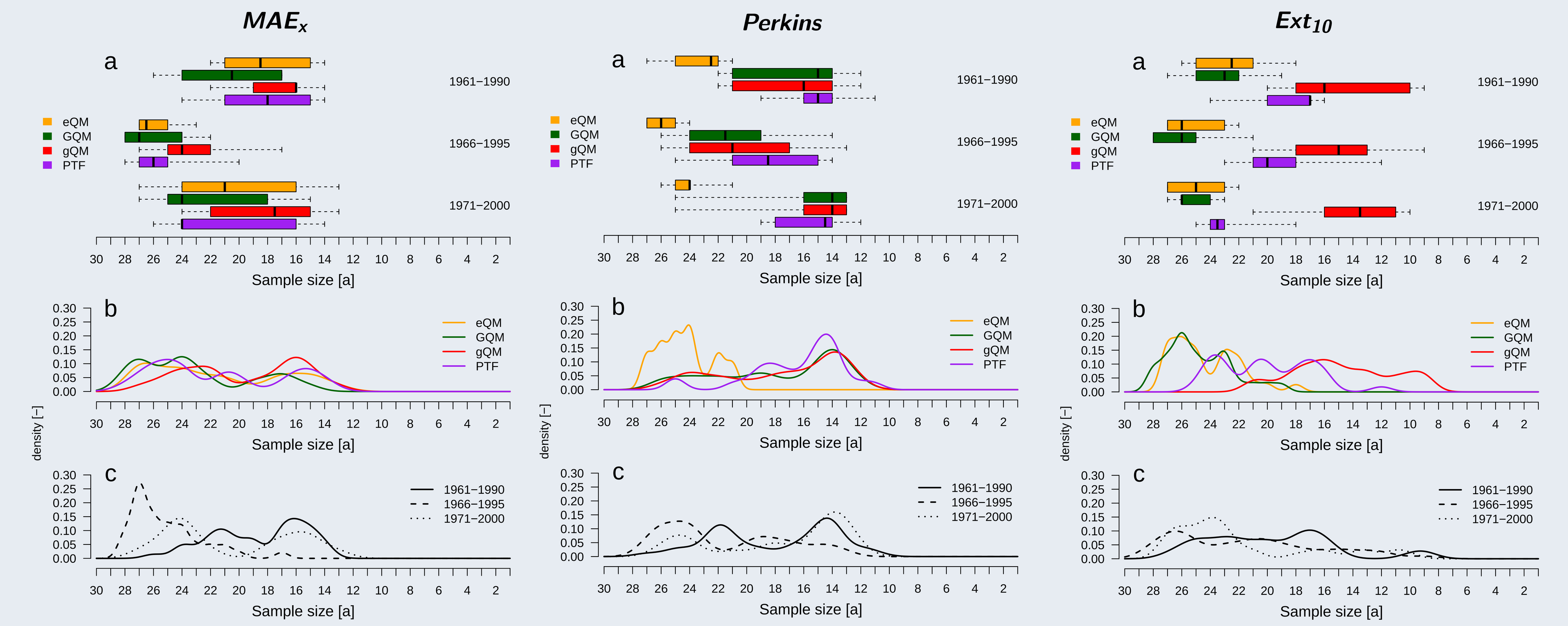


Figure 6: Summary of the n_{crit} values for the skill scores.
a: n_{crit} values for the 10 RCMs, grouped by QM approach and calibration period. **b**: probability density function of n_{crit} values in dependence of QM approach. **c**: probability density function of n_{crit} values in dependence of calibration period.

- we found a large spread of n_{crit} values that is comparable for all skill scores ($\approx$ 28-10 years)
- the n_{crit} value, that shows the effect of sample size, is found to be strongly influenced by the choice of the QM approach but also by the choice of the calibration period
- overall, QM approach *eQM* shows the largest n_{crit} values, followed by *GQM*, *PTF* and finally *gQM* (this ranking matches the ranking by complexity with *eQM* being the most complex approach and most vulnerable to over-fitting)

Conclusions

- a small decrease in sample size can result in a significant worse BC performance
- a general critical sample size can not be assessed, since the n_{crit} values vary strongly and are especially influenced by the choice of QM approach and also by the calibration period, but are always $\geq$ 10 years
- complex QM approaches (*eQM*, *GQM*) show larger n_{crit} values
- if unknown data are bias corrected, less complex QM approaches (*gQM*, *PTF*) are found to be more robust, show a comparable performance to the more complex ones and are hence favourable

References

[1]**Boé J.**, L. Terray, F. Habets, E. Martin (2007): Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology*, 27, 1643–1655, doi:10.1002/joc.1602. [2]**Piani C.**, G. P. Weedon, M. Best , S. M. Gomes, P. Viterbo, S. Hagemann, J. O. Haerter (2010): Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology*, 395, 199–215, doi:10.1016/j.jhydrol.2010.10.024. [3]**Piani C.**, J. O. Haerter, E. Coppola (2010): Statistical bias correction for daily precipitation in regional climate models over europe. *Theoretical and Applied Climatology*, 99, 187–192, doi:10.1007/s00704-009-0134-9. [4]**Gutjahr O.**, G. Heinemann (2013): Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM. *Theoretical and Applied Climatology*, 114, 511–529, doi:10.1007/s00704-013-0834-z. [5]**Gudmundsson L.**, J. B. Bremnes, J. E. Haugen, T. Engen-Skaugen (2012): Technical note: Downscaling RCM precipitation to the station scale using statistical transformations: a comparison of methods. *Hydrology and Earth System Sciences*, 16, 3383–3390, doi:10.5194/hess-16-3383-2012. [6]**Perkins S. E.**, A. J. Pitman, N. J. Holbrook, J. McAneney (2007): Evaluation of the AR4 climate models simulated daily maximum temperature, minimum temperature, and precipitation over australia using probability density functions. *Journal of Climate*, 20, 4356–4376, doi:10.1175/jcli4253.1.