HOW RELIABLE ARE SELECTED METHODS OF PROJECTIONS OF FUTURE THERMAL CONDITIONS? A CASE FROM POLAND

JOANNA WIBIG[1], JOANNA JĘDRUSZKIEWICZ[2]

[1]Department of Meteorology and Climatology, University of Łódź, Poland
[2]Institute of Geography, Pedagogical University, Kraków, Poland

ABSTRACT: The aim of the paper was to assess the robustness of four bias correction techniques: simple bias correction, distribution based bias correction, delta change and distribution based delta change. Data from nine RCM simulations of CORDEX project and 41 Polish weather stations were used. The methods were calibrated in the period 1971–1985 and evaluated in 1991–2005. The improvement in mean, 10th and 90th percentiles was shown, without significant differences among methods. For 1st and 99th percentiles the improvement was generally weaker and simple methods seem to be more robust than the distribution based ones. Strong differences between individual models were found, so the use of model ensemble is recommended.

KEYWORDS: bias correction, delta change, regional climate models, temperature, Poland

Address of the corresponding author: Joanna Wibig, Department of Meteorology and Climatology, University of Łódź, Poland; e-mail: jwibig@geo.uni.lodz.pl

Introduction

It is well known that Regional Climate Models (RCM) temperature output are burdened with systematic biases (Samuelsson et al. 2011). These biases are partly due to coarse resolution and parameterizations. GCMs (Global Climate Models) and RCMs are not perfect. The other reason for biases is that models give us areal averages (sometimes the area is really wide), but in impact studies we need point values (Xu et al. 2005). This areal averaging results in tendency of models to underestimate the highest values and overestimate the lowest ones (Déqué 2007). If realistic output have to be used to force impact models or adaptation processes in local scale it is necessary to cope with these biases. The set of methods called Model Output Statistics (MOS) can be used for it. These methods try to find a statistical link between simulated and observed values and can be divided into two main groups. The first is known as the bias correction (Déqué et al. 2007, Piani et al. 2010). The bias correction factors are calculated by comparison of observations with RCM simulations in the so called reference period, which should be relatively long because two climates are compared, and then these correction factors are applied for simulations in projection period to obtain realistic projections for future climate. Because the bias can be different in different parts of temperature distribution (i.e. the highest values are usually underestimated and the lowest ones are overestimated) in many cases the bias-correction factors are calculated individually for different parts of distribution (so called distribution based bias correction, Déqué

2007, Yang et al. 2010, and Piani et al. 2010, van Roosmalen et al. 2011 and Jaczewski et al. 2014). The second group of MOS methods is referred to as the delta change (DC) method (Hay et al. 2000, Lenderink et al. 2007, van Roosmalen et al. 2011). Here the factors of change in simulated climate from reference to projection periods are calculated, and applied to present climate observations to obtain the projections for the future. As in the case of bias corrections, it can be one factor of change or a series of factors for different parts of distribution (Hay et al. 2000, Lenderink et al. 2007, van Roosmalen et al. 2011). Maraun et al. (2010) give the comprehensive review of MOS methods.

The aim of this paper is to consider a range of bias corection techniques to assess their robustness and determine which one is the most effective when correcting mean daily temperature in Poland simulated by one RCM driven by suite of different GCMs. Four different bias correction techniques were considered: simple bias correction (SB), distribution based bias correction (DB), simple delta change (DC) and distribution based delta change (DDC). These methods are calibrated in the reference period (1971–1985) and then evaluated in the period 1991–2005.

The paper is organized as follows. The first part describes the data used in this study. Then the bias correction methods together with the statistical methods used for the evaluation are outlined. This is followed by presentation of results of the evaluation. Main findings are discussed in the last part of the article.

Data

Data used in this paper consists of series of mean daily temperature from 41 Polish stations from the period 1971–2005 obtained from the Institute of Meteorology and Water Management. Figure 1 shows the location of these stations. The quality of data was tested by standard normal homogeneity Alexandersson test for single shifts (Alexandersson and Moberg 1997). The mean daily near surface temperatures from nine RCM simulations are also used. Their list is presented in Table 1. All simulations were made within CORDEX (Coordinated Regional Climate Downscaling Experiment) project at Swedish Meteorological-Hydrological Institute (SMHI) using the same regional model Rossby Centre regional atmospheric model, version 4 (RCA4) driven by nine different GCMs. The simulations were made within so called historical experiment spanning the time since 1st January 1951 to 31st December 2005 in spatial domain EUR-44 with spatial res-

Fig. 1. Localization of analysed stations.

Table 1. List of simulations used in the analysis.

Domain	Global model	Regional model	Institution
EUR-44	CCCma-CanESM2	RCA4	SMHI
EUR-44	CNRM-CERFACS-CNRM-CM5	RCA4	SMHI
EUR-44	ICHEC-EC-EARTH	RCA4	SMHI
EUR-44	IPSL-IPSL-CM5A-MR	RCA4	SMHI
EUR-44	MIROC-MIROC5	RCA4	SMHI
EUR-44	MOHC-HadGEM2-ES	RCA4	SMHI
EUR-44	MPI-M-MPI-ESM-LR	RCA4	SMHI
EUR-44	NCC-NorESM1-M	RCA4	SMHI
EUR-44	NOAA-GFDL-GFDL-ESM2M	RCA4	SMHI

olution of 0.44 degree (about 50 km). For each station the set of data from the closest grid point was extracted from all simulations.

Data were then divided into two subperiods, each lasting 15 years: the reference period 1971–1985 and the evaluation period 1991–2005. Each of analysed projection procedures was calibrated in reference period and then used in the evaluation period. Obtained projections were compared with station data and their quality was assessed. The analysis was performed separately for data from each calendar month.

Methods

Simple bias (SB) correction

When using the SB method, RCM temperatures in the evaluation period T_e are transformed into T_{corr} such that:

$$T_{corr} = T_e + b$$

using a bias b, b = T_{oref} – T_{sref}, where T_{oref} and T_{sref} are the longterm monthly mean temperatures observed and RCM simulated in the reference period, respectively. Here, the bias is applied to each daily simulated temperature of particular month in the evaluation, giving the corrected daily time series. This method has an advantage of simplicity and requires only monthly climatological averages for calculating the bias. However only mean monthly value is corrected and higher moments of distribution stay unchanged.

Distribution based bias (DB) correction

In the DB method bias a correction factors b_i, are calculated independently for each percentile

on the base of quantile-quantile (Q-Q) plots. Correction factors are calculated for 99 percentiles.

$$T_{corr,i} = T_{e,i} + b_i$$

For values between any two percentiles the linear interpolation is applied. This method demands more data than the SB one, and allows for adjustments of higher moments of temperature distribution.

Simple delta change (DC) method

DC method consists of perturbing a reference climate series (T_{oref}) with an absolute factor derived as a difference f_{DC} of RCM simulation data for a projected climate (here temperature in evaluation period, T_{seval}) and RCM simulation data for a temperature in a reference period (T_{sref}). This difference is than added to observed climate series (T_{oref}) in a reference period to obtain the corrected series for future climate:

$$T_{corr} = T_{oref} + f_{DC}$$

This method has an advantage of simplicity, but requires a data set of long daily time series of robust observations with high spatial resolution. The method assumes that the climate variability is unchanged in the scenario projection, because it is inherited from observed variability, i.e. variability in the reference period.

Distribution based delta change (DDC) method

DDC method consists of perturbing a reference climate series (T_{oref}) with an absolute factor derived as a difference $f_{DC,i}$ of percentiles of RCM simulation data for a projected climate (here tem-

perature in evaluation period, $T_{seval,i}$) and appropriate percentiles of RCM simulation data for a temperature in a reference period ($T_{sref,i}$).

$$T_{corr} = T_{oref,i} + f_{DC,i}$$

A projected time series are calculated by addition to temperatures in reference period corrections for appropriate percentile. As in DB method linear interpolation is applied for values between full percentiles.

Evaluation methodology

An assessment of the performance of each method is obtained by comparing observed, RCM simulated and corrected data sets in the evaluation period 1991–2005. For each of nine RCM simulations the absolute differences between simulated and observed data sets were compared with analogous differences between corrected and observed data sets. The results were divided into three categories: improvement, no change and worsening, and the frequencies of all these categories were calculated. The analysis was done separately for each calendar month and for seven distribution parameters: mean, standard deviation, skewness, 1st, 10th, 90th and 99th percentiles.

Results

Frequency of error reduction in the ensemble

A quantitative assessment of the frequency with which each of analysed methods resulted in improvement in selected statistics of temperature predictions was made. Seven statistics were considered: mean, standard deviation, skewness, 1st, 10th, 90th and 99th percentiles. The correction procedures had been calibrated on data from the reference period 1971–1985. Then they were applied in the evaluation period 1991–2005. For each of 41 stations and nine RCM simulations the absolute differences between these statistics in simulated and observed data sets were compared with analogous differences between corrected and observed data sets. An improvement was recognized when the difference was positive and higher than 0.5°C. When the difference was negative and lower than –0.5°C a case was put in category of worsening. In the remaining cases, when difference falled in the interval (–0.5, 0.5), a case fell into category no change. The percentages of each category were calculated for each calendar month separately and results are presented on figures 2–6.

In the case of mean (Fig. 2) percentages of improvements vary from 38 per cent in September to 70 per cent in May, being almost the same for all methods. Percentages of deterioration vary from 1 to 38 per cent. They fall to 1 in June and reach 38 percent in February for all methods. The differences between all methods are extremely low.
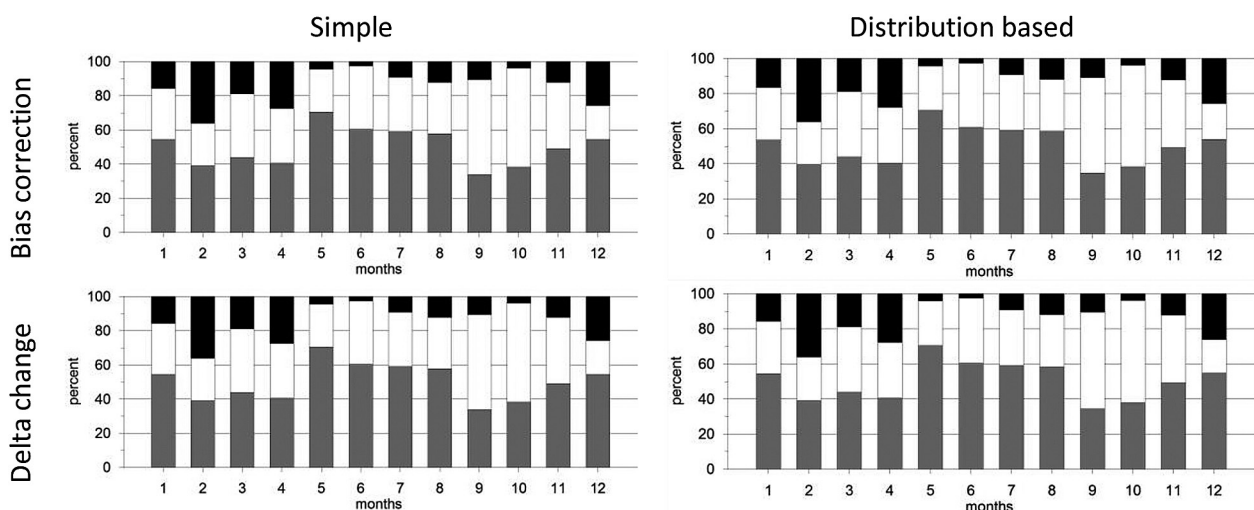

Fig. 2. Frequencies of improvements (grey), no change (white) and worsening (black) after corrections made by different methods for a mean. Bias correction methods in the upper raw and delta change methods in the lower raw, simple variants in the left column and distribution based variants in the right column.
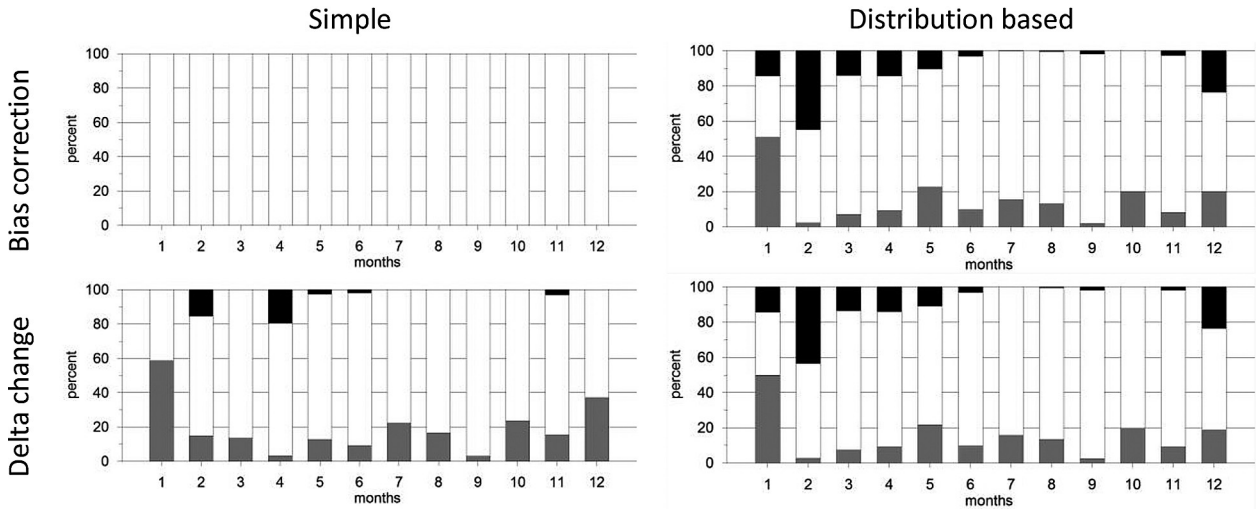
Fig. 3. Frequencies of improvements (grey), no change (white) and worsening (black) after corrections made by different methods for a standard deviation. Bias correction methods in the upper raw and delta change methods in the lower raw, simple variants in the left column and distribution based variants in the right column.

In the SB method all values are corrected by addition of the same factor, so it does not change central moments of the order higher than one (Fig. 3). The other methods also exert relatively low impact on the standard deviation. In the case of DB method percentages of improvement vary from almost 0 in February and September to about 50 in January, percentages of deterioration are smaller and vary from 0 in July and October to 42 in February. The strongest improvements are when DC method is applied and vary from 3–4 percent in April and September to almost 60 in January. The increase of bias is the highest in April when it equals 20 percent, but generally the percentage of improvement exceed the frequency of deteriora-tion. Similar changes appear when DDC method is applied. Analysed methods do not affect skew-ness too much. The improvement frequency exceeds 10 percent from January to March and in August and November, and is the highest when DC method is applied (Fig. 4). The worsening is the strongest for the distribution based methods, exceeding 10 percent only in March.

In the case of moderate extremes: 10th and 90th percentiles (Fig. 5–6) in all methods the signifi-cant improvement is observed. In the case of 10th percentile the improvement is the strongest in cold part of the year exceeding 50 percent from October to April for simple methods and in De-cember, January, April and May for distribution
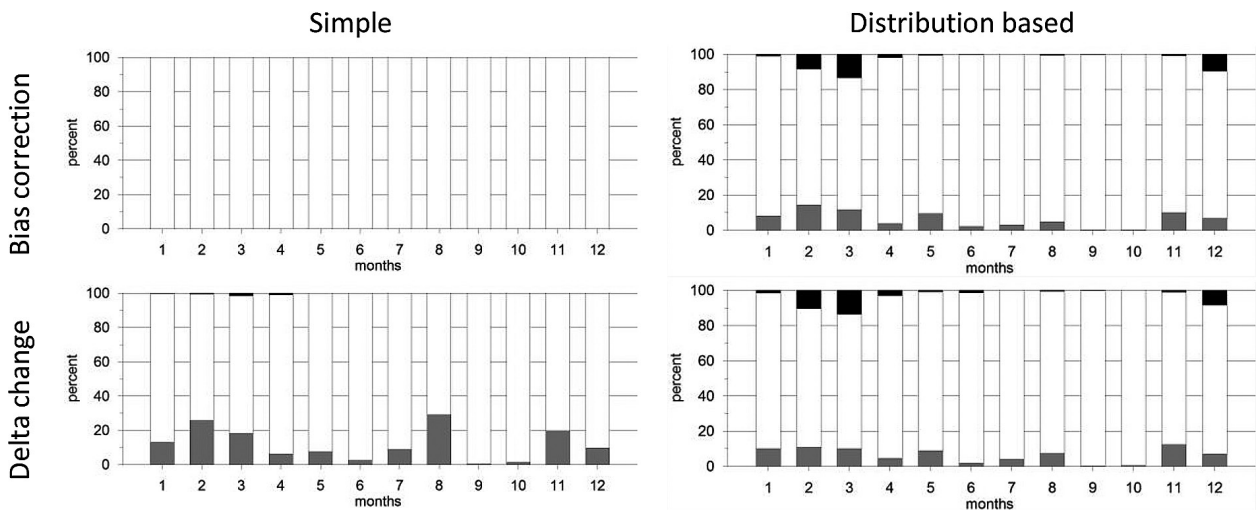

Fig. 4. Frequencies of improvements (grey), no change (white) and worsening (black) after corrections made by different methods for a skewness. Bias correction methods in the upper raw and delta change methods in the lower raw, simple vari-ants in the left column and distribution based variants in the right column.
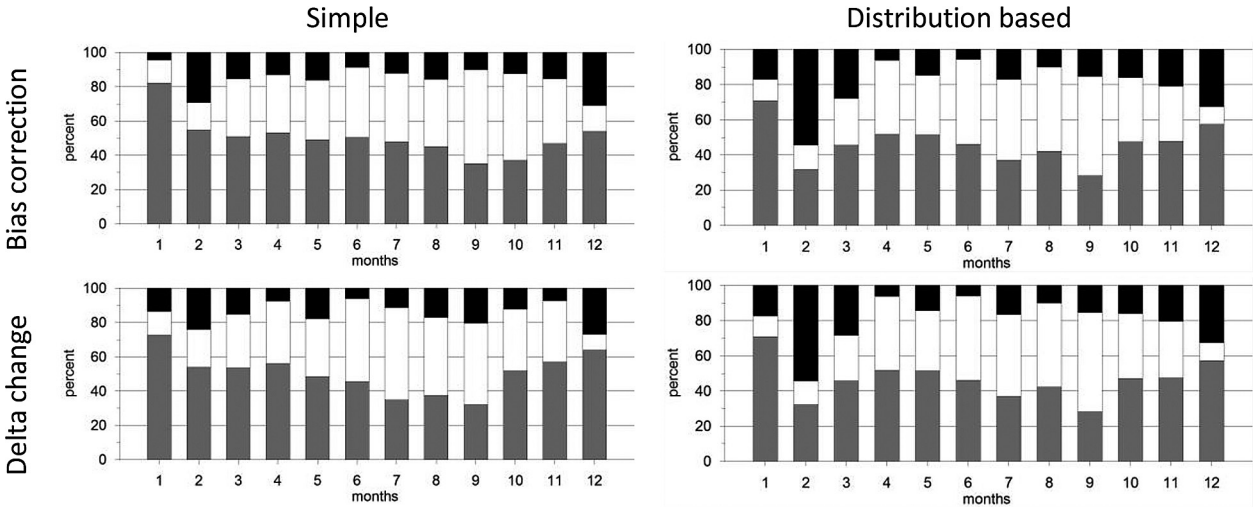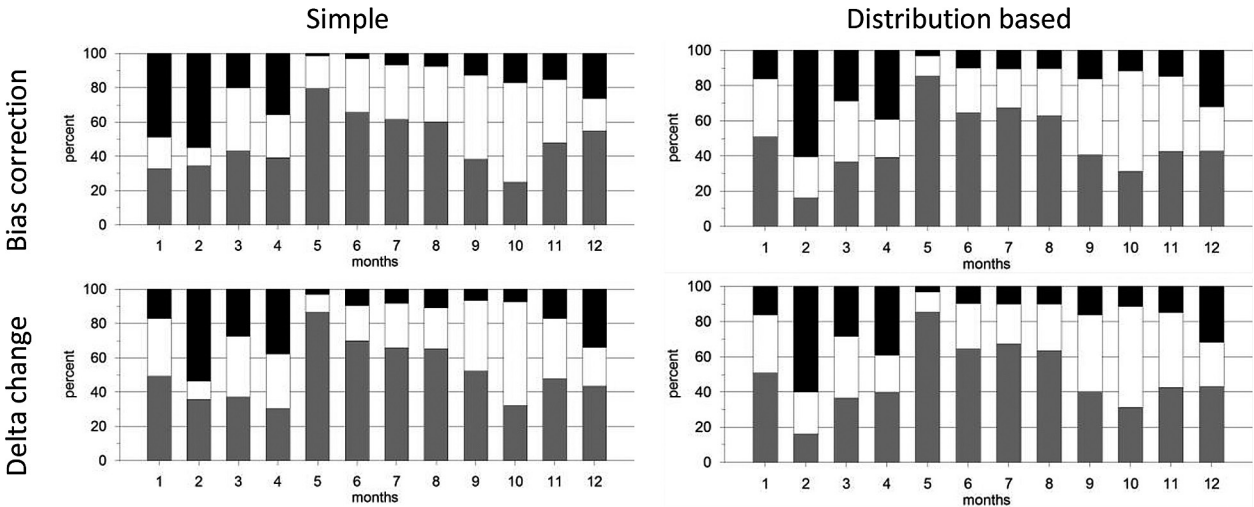
Fig. 5. Frequencies of improvements (grey), no change (white) and worsening (black) after corrections made by different methods for 10th percentile. Bias correction methods in the upper raw and delta change methods in the lower raw, simple variants in the left column and distribution based variants in the right column.

Fig. 6. Frequencies of improvements (grey), no change (white) and worsening (black) after corrections made by different methods for 90th percentile. Bias correction methods in the upper raw and delta change methods in the lower raw, simple variants in the left column and distribution based variants in the right column.

based methods. Surprisingly, the application of distribution based methods do not improve the situation. The number of deterioration vary from a few to more than 50 percent in February for distribution based methods, generally being higher in cold season. In the case of 90th percentile the improvement frequency has evident annual cycle being the highest in warm part of the year, and exceeding 80 percent in May and 50 percent from June to August in all methods. The frequency of deterioration is generally much lower, but in February and April exceeds the frequency of improvements. The highest frequency of worsening (about 60 percent) appears in February when distribution based methods are applied. The dif-

ferences between efficiency of analysed methods are relatively low.

In the case of extreme extremes, 1st and 99th percentiles (not shown) the annual cycle of frequencies of improvements is similar to those of moderate extremes. In the case of 1st percentile higher number of improvements is observed in cold part of the year but for 99th percentile in warm part of the year. The opposite is true for frequencies of deterioration. However the proportions between improvement and deterioration frequencies are more uniform. Surpisingly, the distribution based methods seem to work worse then simple ones both in the case of bias correction and delta change methods.
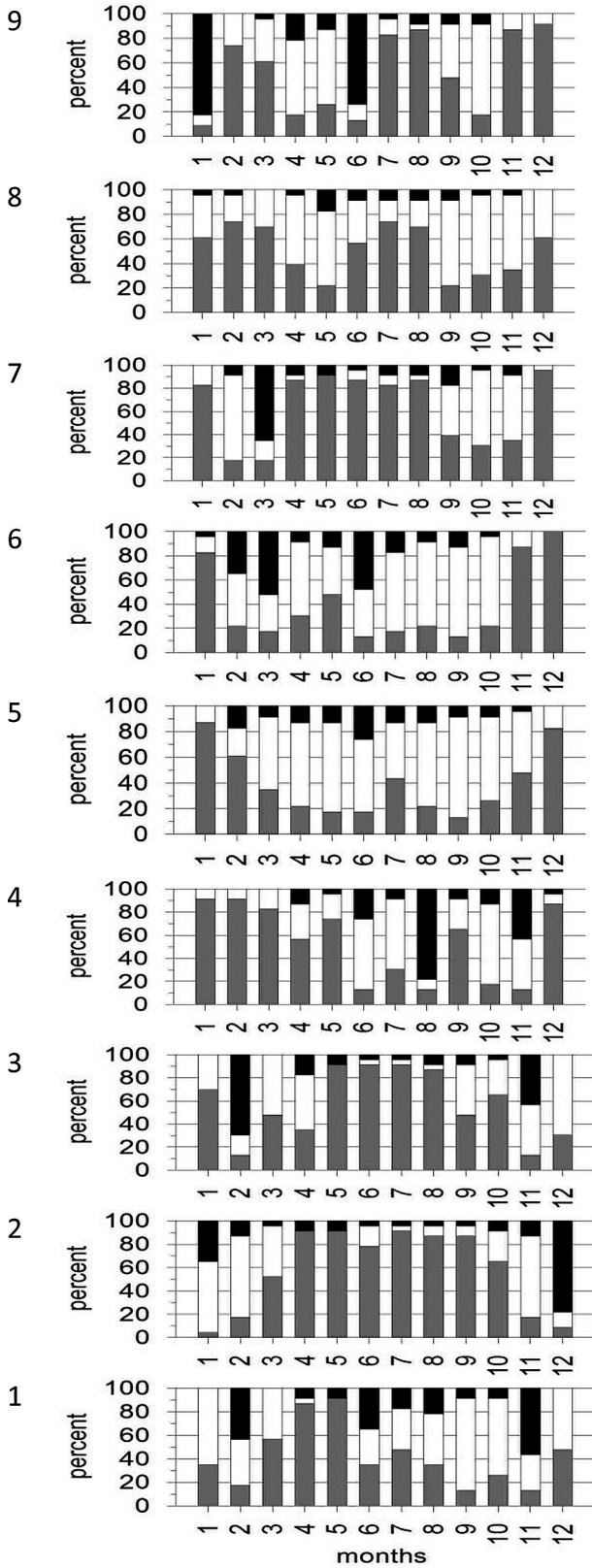
Fig. 7. Frequencies of improvements (grey), no change (white) and worsening (black) after corrections of mean made by the DDC method for individual models (numbers the same as in Table 1).
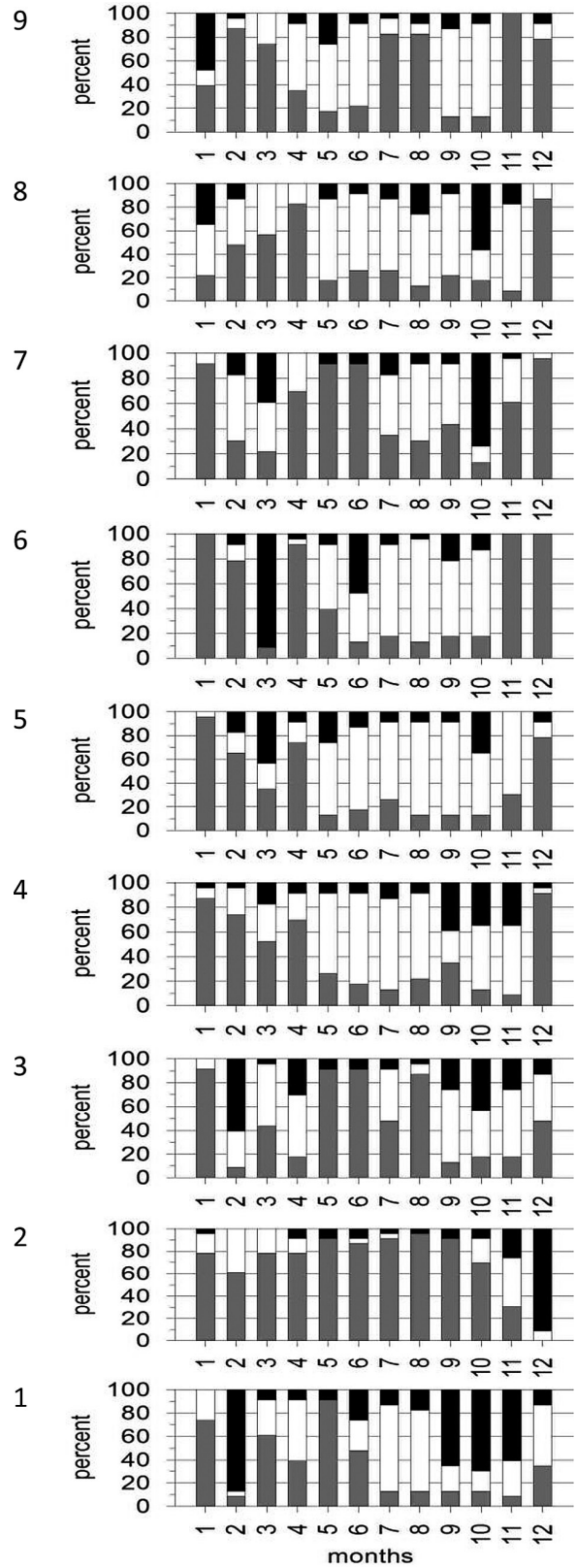
Fig. 8. Frequencies of improvements (grey), no change (white) and worsening (black) after corrections of the 10th percentile made by the DDC method for individual models (numbers the same as in Table 1).

Differences between frequency of error reduction in particular models

Bias calculated in this paper was only estimated on the basis of finite time series, which represent examples of possible climate realisations.

On the average in all models in the case of mean the frequencies of improvement are higher then the frequencies of deterioration in each month, however there are exceptions from this rule significantly varying from one model to another (Fig. 7). There is no singular model with the efective estimation of bias giving high frequency of improvement in all months of the year. It is also impossible to show the worst model giving high frequency of deterioration throughout the whole year. That is why, the multi-model ensemble usage is recommended. As can be seen from Figure 7, the frequency of improvements in many cases exceeds 80 percent, but for almost each model it also happens at least in one month that the frequency of deterioration is also close to 80 percent. The bias of the mean is calculated from the sample of about 450 elements (15 years × 30 days/month). Slightly different situation is in the case of extremes. For the moderate extremes (10th and 90th percentiles), their values are calculated from the much shorter sample of about 45 elements (450 × 0.1), so the error of bias estimation is much larger. It is one reason, why the frequency of improvements is generally lower for 10th and 90th percentiles than for the mean. Fig. 8 shows that these values for particular models can differ significantly from the averages. In the case of extreme extremes (1st and 99th percentiles) the share of improvements is lower, and deteriorations larger than in the case of moderate extremes, but in this case the biases are assessed on the basis of 7–8 the most extreme elements, so errors of their assessment are relatively high.

Discussion and conclusions

In the paper four MOS techniques have been compared to determine which is the most effective and robust to correct temperature simulations made by RCMs to obtain point values able for use in impact models. Two of these techniques were bias corrections with constant and distribution based correction factor, two others were delta change methods also with constant and distribution based delta change factors. Both groups of methods have some advantages and disadvantages. The main advantage of the bias corection is that this method is less demanding in the respect of data, because it uses climatological statistics instead of daily values. Its main disadvantage is that it is based on the assumption of bias stationarity i.e. that it is assumed that statistical relationship between RCM simulation and observations identified in the reference period will be the same in the future, which may not be true as was shown by Christensen et al. (2007) and Boberg and Christensen (2012). The delta change methods are more data demanding, but assume that the climate variability does not change considerably from reference to scenario period (Déqué 2007).

Comparison of the results of all methods indicate that the differences between their effectivity and robustness are relatively low. In case of the mean the differences are negligible small. The SB method shifts the whole temperature distribution and does not affect higher moments of distribution, so both standard deviation and skewness remain exactly the same. All other methods also do not change the quality of standard deviation and skewness predictions. Similar result was obtained also by Lafon et al. (2013) for correction of precipitation using a suite of bias correction methods.

Comparison of effects of correction of moderate extremes (10th and 90th percentiles) indicates that the effectiveness of presented methods is slightly lower than for the mean, but much higher than for standard deviation and skewness. However the differences between analysed method are insignificant and it is not possible to designate the best one. In the case of extreme extremes (1st and 99th percentiles) all methods work worse than in the case of the moderate ones. At the same time it seems that distribution based methods (DB and DDC) fail more often than simple ones (SB and DC), but there is no significant difference between the last two. Such situation is probably because biases are estimated from finite time series and the less numerous samples cause the higher uncertainty. Maraun (2012) called this uncertainty of bias estimation as *variability related apparent bias changes*. It means that results of application of these methods for correction

of extreme extremes may not be robust. As was shown by Deser et al. (2012) that natural climate variability constitute the additional source of uncertainty of climate projections for future alongside uncertainty related to emission scenario and model-response. They show that in regions of high natural variability this factor can introduce strong uncertainty.

Figures 7 and 8 present some examples of the span of internal climate variability range and its effect on bias reduction. The same method can decrease bias for some models and increase for the others. It is caused be small length of time series used in comparison with high natural climate variability. It is also the reason why it is generally recommended to prepare climate projections on the base of ensemble of RCM simulations, than on the singular ones.

Acknowledgements

References

Alexandersson H., Moberg A., 1997. Homogenization of Swedish temperature data. Part 1: Homogeneity test for linear trends. *International Journal of Climatology* 17: 25–34.

Boberg F., Christensen J.H., 2012. Overestimation of Mediterranean summer temperature projections due to model deficiencies. *Nature Climate Change* 2: 433–436.

Christensen J.H., Boberg F., Christensen O.B., Lucas-Picher P., 2008. On the need for bias correction of regional climate change projections of temperature and precipitation, *Geophysical Research Letters* 35: L.20709. DOI: 10.1029/2008GL035694.

Christensen J.H., Christensen O.B., 2007. A summary of the PRUDENCE model projections of changes in European climate by the end of the century. *Climatic Change* 81 (Suppl. 1): 7–30.

Christensen J.H., Hewitson B., Busuioc A., Chen A., Gao X., Held I., Jones R., Kolli R.K., Kwon W.-T., Laprise R., Magaňa Rueda V., Mearns L., Menéndez C.G., Räisänen J., Rinke A., Sarr A., Whetton P., 2007. Regional climate projections. In:. Solomon S. et al. (ed.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Déqué M., 2007. Frequency of precipitation and temperature extremes over France in an antrophogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change* 57: 16–26.

Déqué M., Rowell D.P., Lüthi D., Giorgi F., Christensen J.H., Rockel B., Jacob D., Kjelström E., De Castro M., van den Hurk B., 2007. An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Climatic Change* 81 (Suppl. 1): 53–70.

Deser C., Knutti R., Solomon S., Phillips A.S., 2012. Communication of the role of natural variability in future North American climate. *Nature Climate Change* 2: 775–779.

Hay L.E., Wilby R.L., Leavesley G.H., 2000. A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *Journal of American Water Resources Association* 36: 387–398.

Jacob D., Bärring L., Christensen O.B., Christensen J.H., de Castro M., Déqué M., Giorgi F., Hagemann S., Hirschi M., Jones R., Kjellström E., Lenderink G., Rockel B., Sánchez E., Schär C., Seneviratne S.I., Somot S., van Ulden A., van den Hurk B., 2007. An intercomparison of regional climate models for Europe: model performance in present-day climate. *Climatic Change* 81: 31–52.

Jaczewski A., Brzóska B., Wibig J., 2014. Comparison of temperature indices for three IPCC SRES scenarios based on RegCM simulations for Poland in 2011–2030 period. *Meteorologische Zeitschrift* 21: 99–106.

Lafon T., Dadson S., Buys G., Prudhomme C., 2013. Bias correction of daily precipitation simulated by regional climate model: a comparison of methods. *International Journal of Climatology* 33: 1367–1381. DOI: 10.1002/joc.3518.

Lenderink G., Buishand A., van Deursen W., 2007. Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Science* 11: 1145–1159.

Maraun D., 2012. Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophysical Research Letters* 39: L06706. DOI: 10.1029/2012GL051210.

Maraun D., Wetterhall F., Chandler R.E., Kendon E.J., Widmann M., Brienen S., Rust H.W., Sauter T., Themeβl M., Venema V.K.C., Chun K.P., Goodess C.M., Jones R.G., Onof C., Vrac M., Thiele-Eich I., 2010. Precipitation downscaling under climate change: Recent developements to bridge the gap between dynamical models and the end user. *Reviews of Geophysics* 48: 2009RG000314.

Piani C., Haerter J.O., Coppola E., 2010. Statistical bias correction for daily precipitation in regional climate mod-

els over Europe. *Theoretical and Applied Climatology* 99: 187–192.

Rummukainen M., 2010. State-of-the-art with regional climate models. *Wiley Interdisciplinary Reviews: Climate Change* 1: 82–96.

Samuelsson P., Jones C.G., Willén U., Ullerstig A., Gollvik S., Hansson U., Jansson C., Kjellström E., Nikulin G., Weser K., 2011. The Rossby Centre Regional Climate Model RCA3: Model description and performance. *Tellus* A 63: 4–23. DOI: 10.1111/j.1600-0870.2010.00478.x.

van Roosmalen L., Sonnenborg T.O., Jensen K.H., Christensen J.H., 2011. Comparison of hydrological simulations of climate change using perturbation of observation and distribution-based scaling. *Vadose Zone Journal* 10: 136–150.

Xu C., Widen E., Halldin S., 2005. Modelling Hydrological Consequences of Climate Change – Progress and Challenges. *Advances of Atmospheric Sciences* 22: 789–797.

Yang W., Andreasson J., Graham L.P., Olsson J., Rosberg J., Wetterhall F., 2010. Distribution based scaling to improve usability of RCM regional climate projections for hydrological climate change impact studies. *Hydrological Research* 42: 211–220.