BIAS CORRECTION OF CLIMATE DATA BY THE QUANTILE MAPPING FOR TWO RIVER CATCHMENTS IN THE JUŽNA MORAVA RIVER BASIN

Nikola Đokić[1], Borislava Blagojević[2]

Abstract

Planning and design in water resources management and hydraulic engineering is based on climate data in the river catchments of interest. Global climate models (GCM) were originally used to construct climate scenarios for assessing the climate change impact. The improvements in data downscaling are nowadays bringing GCMs close to more precise Regional climate models (RCMs), that could be used in planning and design. Numerous techniques of bias correction for data models have been developed, and four of them are discussed and applied in this paper: quantile mapping with linear transformation function (LTF), delta mapping (DM), mapping using normal (ND) and gamma distribution (GD). Bias correction is performed on the GCM EC-Earth for mean daily temperatures and monthly precipitation sums. Four data gauge periods are used, two per each climate variable, one for data calibration and one for validation. Periods considered for temperature are 1950-2000 and 2001-2010, and for precipitation, 1979-2000 and 2001-2009. The study focuses on two river catchments of the hydrological stations (HS) in the Južna Morava river basin: HS Visoka/ Kosanica and HS Sijarinska Banja/Jablanica. To assess the suitability of the applied techniques, three error measures were used: MAE, MSE and RMSE. The analysis of the bias correction of precipitation data shows that the best results are provided by LTF, slightly worse by GD, and poor results are obtained by applying DM. In the analysis of temperatures, LTF also gave the best results, good results are provided by DM, while ND does not have a good agreement with the gauged data

Key words: *GCM EC-Earth, bias correction, quantile mapping, daily air temperature, monthly precipitation*

[1] Doctoral Academic Studies student, M.Sc. Civ. Eng., Faculty of Civil Engineering and Architecture, University of Niš, nikolavdjokic995@gmail.com

[2] Assist. Prof. Ph.D., Faculty of Civil Engineering and Architecture, University of Niš, borislava.blagojevic@gaf.ni.ac.rs

1. INTRODUCTION

The hydrological cycle, as the basic process on the planet Earth, is sensitive to climate change. Therefore, by disrupting the climate, the hydrological cycle is also disrupted, representing a potential threat to humanity. The governments and authorities worldwide develop strategies for climate change adaptation in their territories that rely upon climate change information. Such information can be obtained by using global climate models (GCM) that simulate climate processes with a low spatial resolution (150 - 300 km). However, this coarse spatial resolution is not suitable to perform analysis at a regional or local level [1]. For this reason, a downscaling procedure is performed to derive high spatial resolution of climate parameters at the regional/local level, which are the basis for various climate change studies. For example, hydrological rainfall-runoff models require climate parameters such as temperature and precipitation with a fine spatial and temporal resolution, which can be obtained from GCM [2]. Therefore, in addition to the spatial downscaling, temporal downscaling is also used in practice, which, for example, represents the downscaling of monthly data into daily values.

GCM and regional climate models (RCM) simulations have systematic errors called bias, which need to be corrected. It is necessary to remove the bias from the models output for projecting the future hydrologic and climatic scenarios correctly [3]. In scientific practice, the term bias correction (BC) is used ambiguously. In some cases, BC is mentioned as a method intended for the correction of raw model data, while in others it is classified as a group of statistical methods of downscaling. It can be said that both understandings are correct, however, the problem is that BC will not allow a 'finer' structure of the change of some quantity in the future, which is potentially conditioned by local characteristics, or even if obtained, it will be insignificant. For this reason, it can be said that BC is a pseudo version of statistical downscaling [4].

In general, there are two techniques of downscaling, dynamic (DD) and statistical (SD), although some sources [1, 5] mention a third group that includes a combination of dynamic and statistical techniques (combined method).

The SD relies on the statistical relationship between the large-scale climate derived from GCM and the local-scale climate obtained from observations, assuming that such a relationship does not change through time. Statistical methods can be divided into three groups: 1) Methods based on time schemes, 2) Time generators, 3) Regression methods.

Nowadays, there is a large number of SD methods, and in [6] alone, 22 different SD techniques were applied. The paper [7] used a combination of all three mentioned groups of statistical methods, using Bayesian Model Averaging. The results of the work [7] are in favor of the fact that more accurate results are provided by the combination of these methods, than the application of each of them separately.

Given that there are many developed BC techniques, it is important to choose the appropriate ones for a given assignment. Using SD is much simpler compared to DD, which is the reason for their more often application in practice. The application of SD is conditioned by the existence of historical (gauge) data, which makes it unsuitable for areas that do not have a dense enough observation network.

DD on the other hand uses the RCM embedded in the GCM to generate fine resolution climate information. Using DD is complex, requires the use of high-performance computers, and is limited to a spatial resolution of 20-50 km [8, 9]. For the hydrological study of an area, it is recommended to use an ensemble of several RCMs if possible, especially in air temperature analyses, because the data of individual RCMs deviate significantly from observed data [10].

In this paper, the BC is used for downscaling data from the GCM EC-Earth for mean daily temperatures and monthly precipitation sums. Four different BC techniques were applied: quantile mapping with linear transformation function (LTF), delta mapping (DM), mapping using normal (ND) and gamma distribution (GD).

SD established a connection between GCM, which has a low resolution with ~ 200 km cells, taken from the Copernicus platform for climate change [11], with observed raster data from the E-OBS database, which has cells of 10 x 10 km, downladed from the Digital Atlas platform of Serbia [12]. Downscaling focuses on two river catchments of the HS in the Južna Morava river basin: HS Visoka/ Kosanica and HS Sijarinska Banja/ Jablanica.

2. METHODOLOGY

2.1. Spatial data processing

The first phase of this research is spatial data processing, performed in the software Quantum Geographic Information System (QGIS), version 3.16. The two catchment borders are taken from previous research [13, 14]. The downloaded raster data of mean daily temperatures and monthly precipitation sums from GCM EC-Earth and observed from E-OBS, were assigned to the two studied basins. In the procedure of overlapping the raster data of these basins, it was concluded that one cell of GCM covers both basins (Figure 1), while on the other hand, 9 cells from the E-OBS database include HS Visoka, and three cells, HS Sijarinska Banja (Figure 1).

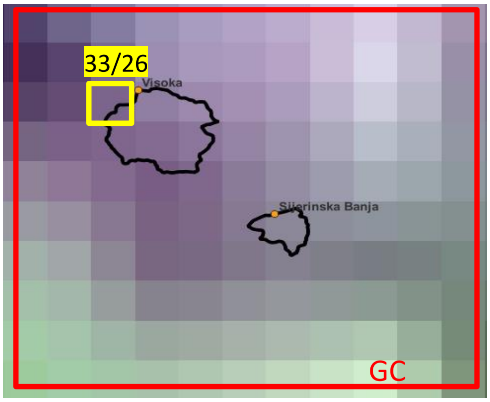

Figure 7. Raster data coverage for the HS Visoka and HS Sijarinska Banja. The red square represents the GCM cell, and the yellow represents the 33/26 cell that covers a part of the HS Visoka. The first number in the label represents the row of the cell, and the second the column of the cell in the E-OBS raster grid.

The Value tool in QGIS is used to ‘read’ the data of each individual cell for the studied climate variables. The data of the cells that cover the studied basins were copied from QGIS and processed in Excel.

2.2. Bias Correction and Spatial Downscaling

Statistical downscaling is based on the establishment of the connection between low resolution data originating from GCM and observed data with fine resolution. In this way, it is possible to establish a connection between GCM - observed data for the reference period, whereby the established connection is maintained for the period of prediction of future climate data contained by GCM.

The reference periods for establishing links between GCM and observed data for temperature and precipitation in this paper, are as follows:

- Calibration period from 1950. to 2000. for temperature;
- Verification period from 2001. to 2010. for temperature;
- Calibration period from 1979. to 2000. for precipitation;
- Verification period from 2001. to 2009. for precipitation.

When linking GCM data to the observed data, chronological data are providing poor results (Figure 2 - left) in contrast to sorting data in the ascending order [13] (Figure 2 - right).

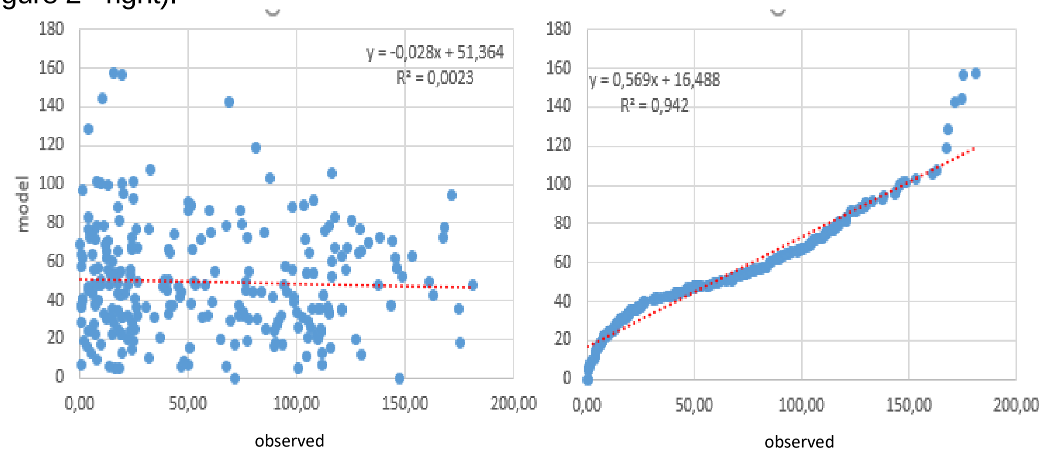

Figure 2. Quantile mapping with raw data (left), and with data sorted in ascending order (right)

From Figure 2 – right it can be seen that the greatest uncertainty in establishing the link between GCM and observed data occurs in the upper and lower zones, respectively, for the occurrence of upper and lower extremes. BC is the process of correcting the output climate model data in order to reduce the effects of systematic errors in climate models and to provide a suitable data source for hydrological models [15]. Four BC techniques with quantile mapping were used in this paper, which are discussed below.

2.2.1. Quantile mapping with linear transformation function

The flexible BC methods adjust the variance of the model distribution to better match the observed variance. Quantile mapping (QM) techniques are among the most popular BC methods. In general, quantile mapping implements statistical transformations for post-processing of climate modeling results. Linear QM establishes a linear relationship between the quantiles of the observed and modeled values of the considered parameters in the reference period:

$$X_{o,h} = a + b * X_{m,h} \quad (1)$$

where $X_{o,h}$ and $X_{m,h}$ represent values of the observed (o) and modeled (m) variables (*x)* for historical period (*h*). The coefficients *a* and *b* are linear regression parameters determined for the historical period. The requested corrected value of the model *Xc,f* at time *t* from the prediction period is obtained from:

$$X_{c,f} = a + b * X_{m,f} \tag{2}$$

where $X_{m,f}$ is the modeled value of the parameter *X* at time *t* from the prediction period.

2.2.2. **Delta mapping**

The simplest BC technique is the delta mapping technique, which establishes the relationship between observed and modeled data as:

$$X_{o,h} = X_{m,h} * \overline{X_{o,h}}/\overline{X_{m,h}} \tag{3}$$

where $X_{o,h}$ and $X_{m,h}$ represent value of the observed and modeled variables for histrorical period, and $\overline{X_{o,h}}/\overline{X_{m,h}}$ is ratio of the mean values of series of observed and modeled data for histrorical period. Sometimes, this technique is not considered as the one from a group of BC, but only uses the model's response to climate change to modify the observations [17].

The corrected value of the considered parameter at a certain moment in the future is obtained as follows:

$$X_{c,f} = X_{m,f} * \overline{X_{o,h}}/\overline{X_{m,h}} \tag{4}$$

where $X_{c,f}$ is the bias-corrected future projection value of model at time *t*, $X_{m,f}$ is the projected value of the model at time *t*, and, and $\overline{X_{o,h}}/\overline{X_{m,h}}$ is ratio of the mean values of series of observed and modeled data for historical period. Therefore, when correcting the model data in the prediction period, the ratio $\overline{X_{o,h}}/\overline{X_{m,h}}$ is used from reference period.

2.2.3. **Mapping using normal distribution**

Quantile mapping using normal distribution was conducted for mean daily temperature. First, Normal distribution parameters (mean and variance) are estimated separately for the observed *Xo,h* and modeled *Xm,h* data during the historical period. The bias-corrected future projection at time *t* is given by:

$$X_{c,f} = F_{o,h}^{-1}\left[F_{m,f}\left(X_{m,f}(t)\right)\right] \tag{5}$$

Therefore, the corrected value of the model at some time *t* from the prediction period is obtained as the inverse function of the normal cumulative distribution for the specified mean value and standard deviation from the historical period.

To correct the bias of the temperature data, it is suitable to use normal distribution because of the negative values that this climate parameter can have.

2.2.4. Mapping using gamma distribution

The two-parameter gamma distribution is recommended for BC of precipitation data [18, 19]. This distribution has the following distribution parameters:

$$\alpha = \frac{\overline{X_{o,h}}^2}{\sigma^2}; \quad \beta = \frac{\sigma^2}{\overline{X_{o,h}}} \tag{6}$$

where $\overline{X_{o,h}}$ is the mean value of the series of data from the historical period, and σ^2 is the variance of this series.

In the same way as for the normal distribution, via equation (5), the corrected value of monthly precipitation at some point *t* in time in the future is obtained, using the parameters of the gamma distribution in the calculation. The diagram shown in Figure 3 illustrates how simulated value, a quantile of the simulated distribution, is replaced by the quantile of the observed distribution corresponding to the same probability.

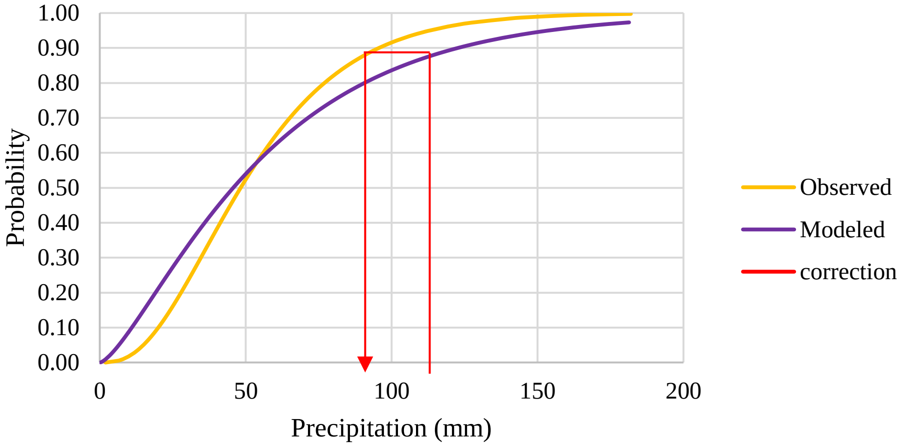

Figure 3. Precipitation bias correction by gamma distribution of modelled and observed values corresponding to the same probability.

2.3. Assessment of the suitability of applied techniques

After calibrating the data using four different BC techniques, the suitability of each technique for fitting the modeled data (from GCM EC-Earth) to the observed values (from E-OBS base) throughout the verification period is assessed by the three error measures: mean absolute error (MAE), mean square error (MSE) and root mean square error (RMSE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_{m,v} - \overline{X_{o,v}}\right| \tag{7}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(X_{m,v} - \overline{X_{o,v}}\right)^2 \tag{8}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{\left(\overline{X_{o,v}}^2 - X_{m,v}\right)}{n}} \tag{9}$$

where $\overline{X_{o,v}}$ the mean value of the data series from the verification period, $X_{m,v}$ represents modeled/corrected data values.

3. RESULTS

The results of the research are presented for each grid cell covering two studied basins through RMSE, MAE and MSE in the GCM validation period. The BC GCM results obtained by the three techniques are shown in parallel to the uncorrected GCM results for temperature (Figure 4) and precipitation (Figure 5). The blue bars in Figure 4 and Figure 5 show error measures of the uncorrected GCM outputs, while green, red and purple bars show corresponding error measures of bias corrected GCM outputs.

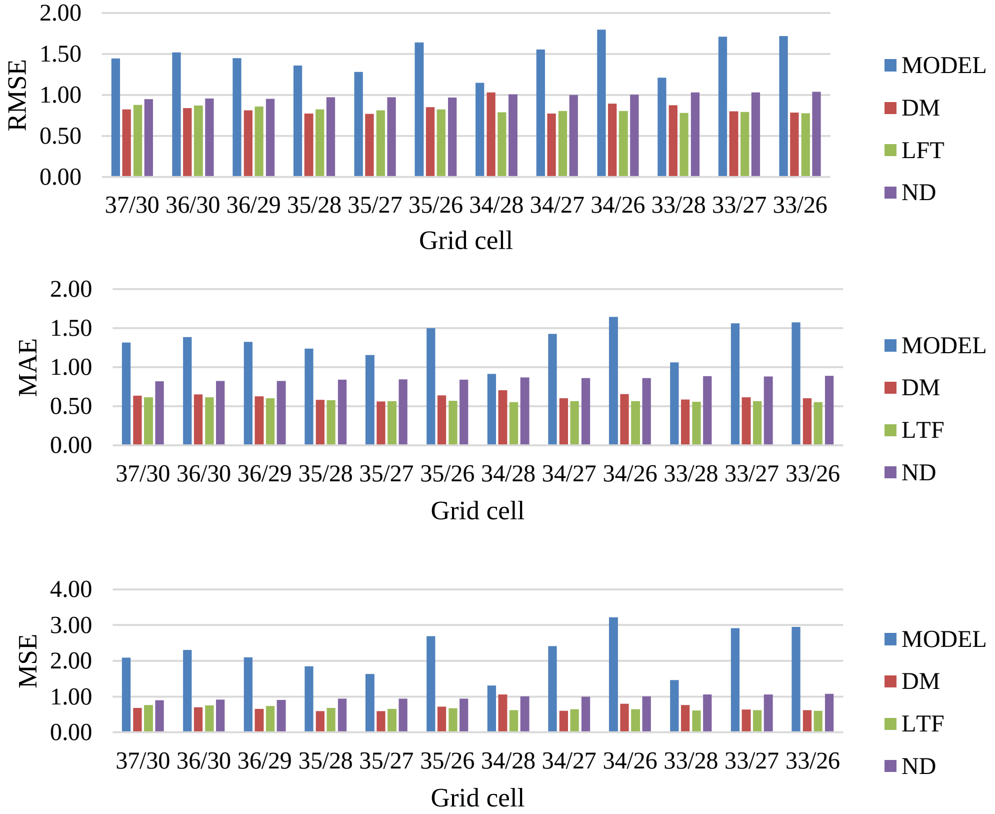

Figure 4. RMSE, MAE and MSE for uncorrected and bias corrected model outputs for temperature data

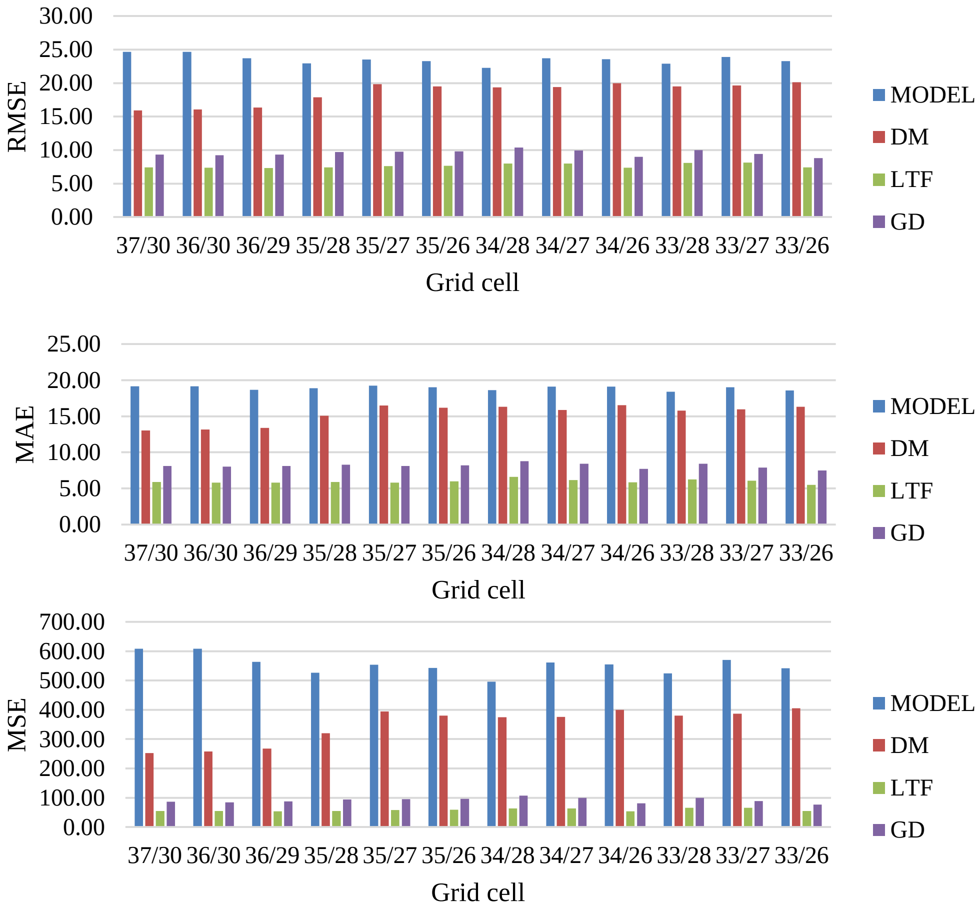

Figure 5. RMSE, MAE and MSE for uncorrected/corrected model outputs for monthly precipitation data

4. DISCUSION

Based on the obtained error measures shown in the Figure 5. and Figure 6., it can be seen that by correcting the bias of the raw output data of the model, better agreement with the observed values is achieved in all cases. In the temperature analysis (Figure 4), the difference between the three used BC techniques was not large, i.e. all three techniques were equally successful in the process of adapting the data model outputs to the observed values. In the case of precipitation, the DM (red bar) slightly corrected the model data, but not enough to be reliably used for future precipitation projections. On the other hand, GD and LTF gave very good results, where GD was slightly better compared to LTF.

Luo et al. [20], used 7 BC techniques for the correction of RCM precipitation: 5 techniques for the correction of average, and one per minimum and maximum daily temperatures for the period 1965-2004. The results of the research [20] indicated a poor fit of precipitation in wet seasons, and also that original RCM outputs are very biased. All used methods, such as Linear Scaling, Distribution mapping, Empirical Quantile Mapping and others, give correct results in the correction of RCM bias.

Bigger differences are obtained in the analyzes of precipitation, which is generally also shown in this paper [20]. The methods based on probability distribution performed best [20], which is the case in this research as well. Beyer et al. [21] found an excellent performance of the DM method, which proved to be better than quantile mapping techniques. The paper also mentions the accuracy of techniques based on probability distributions. This technique performed relatively well in higher latitudes and elevations where there is less precipitation, which brings about less uncertainty compared to humid and subtropical areas [22] .

A Q-Q (P-P) plot in Figure 6 shows precipitation values for the historical period for cell 33/27, where observed monthly precipitation values are plotted on the x-axis, and the uncorrected values from the model on the y-axis, together with the corrected model values of the three different BC techniques. The diagonal represents ideally fitted data with observed values. The values above the diagonal represent higher values compared to the observed ones, and values below the diagonal represent lower precipitation values compared to the observed values. By correcting the data from the model, they approach the diagonal. DM provides better results only for extremely high precipitation compared to LTF and GD, and for precipitation lower than 55 mm/month it performs worse even when compared to uncorrected model values. The precipitation values obtained by the LTF method are located closest to the diagonal (purple circles). However, as already discussed (Figure 2), these techniques give poor results when considering extremes. For the lower extremes, only the GD technique proved to be solid, while with LTF, departure from the diagonal can be seen in the lower part of the diagonal.

Figure 7 shows a Q-Q (T-T) plot for the daily temperatures in the verification period for the cell 33/27. Given that the historical period has a large amount of data (18263), and a much smaller range of variation of values compared to precipitation, the plots are different. The differences between the corrected values and the observed values for extreme cases are not as pronounced as for precipitation, moreover, ND gives excellent results. Here, ND proved to be the best of the BC techniques used. The model itself, as expected, gives the largest deviation from the diagonal, and between LTF and DM, the DM technique better adapts the model values to the observed values at temperatures below 10℃, while for higher values, LTF is better. Here, the DM method proved to be somewhat more acceptable for model correction unlike in the case of precipitation.

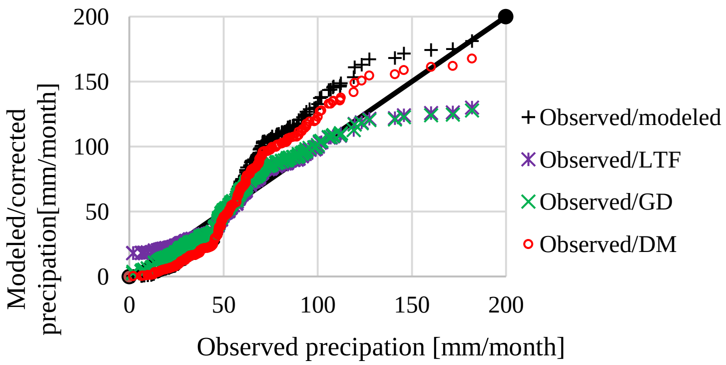

Figure 6. Q-Q plot for observed and modeled/corrected data of precipitation for the historic period

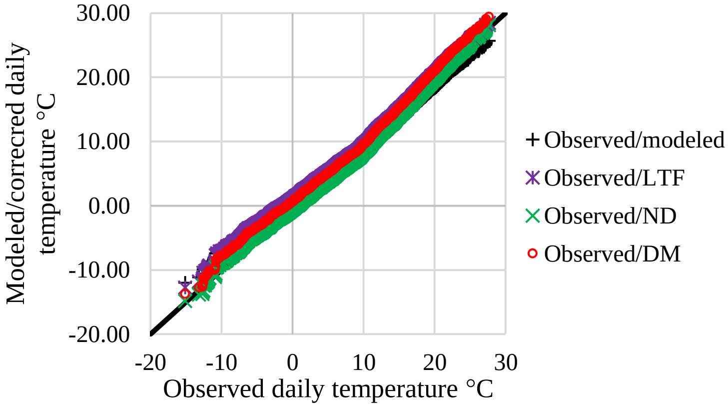

Figure7. Q-Q plot for observed and modeled/corrected data of temperature for the historic period

5. CONCLUSION

The efficiency of four BC techniques on the temperature and precipitation output data of GCM EC-Earth was investigated for two catchments in the Južna Morava river basin. The results show that all investigated techniques provide much better match to the observed values compared to the raw GCM output data.

The results of the obtained error measures indicated that BC methods based on theoretical probability distributions provide good, and in the case of temperatures, the best agreement of the model data with the observed values. DM was poorest for precipitation, while it gave acceptable results in temperature analyses. LTF performed best in precipitation analyses.

In general, it was shown that the extreme values significantly differ from the observed values, especially in the case of precipitation, which indicates that the zones of extremes should be treated separately, i.e. additional action is needed in these zones.

REFERENCES

[1] Marija Sudarić: **Uticaj klimatskih promena na pojavu plamenjače vinove loze u Srbiji**. Master rad. *Prirodno-matematički fakultet Univerziteta u Novom Sadu*, 2015.

[2] Teutschbein, Claudia. Hydrological modeling for climate change impact assessment: transferring large-scale information from global climate models to the catchment scale. *Diss. Department of Physical Geography and Quaternary Geology, Stockholm University*, Stockholm, 2013.

[3] Kaustubh Salvi, S Kannan and Subimal Ghosh: **Statistical Downscaling and Bias Correction for Projections of Indian Rainfall and Temperature in Climate Change Studies.** *International Conference on Environmental and Computer Science.* Singapore, 2011.

[4] Vladimir Đurđević - private correspondence *(29.05.2023.)*

[5] Xu, Z., Han, Y. & Yang, Z. Dynamical downscaling of regional climate: A review of methods and limitations. *Sci. China Earth Sci*., China *62*, 365–375, 2019.

[6] Goodess, C. M. et al.: STARDEX - Recommendations on the more robust statistical and dynamical downscaling methods for the construction of scenarios of extremes. *Climatic Research Unit, University of East Anglia, Norwich, NR4 7TJ,*UK, 2005.

[7] Jiaming Liu, Di Yuan, Liping Zhang, Xia Zou and Xingyuan Song: **Comparison of three statistical downscaling methods and ensemble downscaling method based on bayesian model averaging in upper Hanjiang river basin, China**. *Advances in Meteorolog*, Volume 2016.

[8] https://agrimetsoft.com/sd-gcm *(09.01.2023.)*

[9] Lianlian Xu, Aihui Wang: Application of the Bias Correction and Spatial Downscaling Algorithm on the Temperature Extremes From CMIP5 Multimodel Ensembles in China. *Earth and Space Science*, December 2019.

[10] Teutschbein, C., and J. Seibert: Regional Climate Models for Hydrological Impact Studies at the Catchment Scale: A Review of Recent Modeling Strategies, *Geography Compass*, 4(7), 834–860, doi:10.1111/j.1749- 8198.2010.00357.x., 2010.

[11] https://cds.climate.copernicus.eu/cdsapp#!/home *(26.01.2023.)*

[12] https://atlas-klime.eko.gov.rs *(19.01.2023.)*

[13] Nikola Đokić, Borislava Blagojević: PRIPREMA KLIMATOLOŠKIH PODLOGA ZA OCENU PROSEČNIH PROTOKA U USLOVIMA BUDUĆE KLIME U SLIVU JUŽNE MORAVE. *Nauka+Praksa - broj 25/2022*. Niš, 2022.

[14] Nikola Đokić, Borislava Blagojević, Slaviša Trajković: **OCENA PROSEČNIH GODIŠNJIH PROTOKA U BUDUĆIM KLIMATSKIM USLOVIMA – STUDIJA SLUČAJA U SLIVU JUŽNE MORAVE.** *Zbornik radova Građevinsko-arhitektonskog fakulteta - broj 37/2022*. Niš, 2022.

[15] https://rcmes.jpl.nasa.gov/content/statistical-downscaling *(25.02.2023.)*

[16] Hakala Assendelft, Kirsti & Addor, Nans & Teutschbein, Claudia & Vis, Marc & Dakhlaoui, Hammouda & Seibert, Jan: **Hydrological Modeling of Climate Change Impacts**. *Encyclopedia of Water: Science, Technology, and Society,* 2020.

[17] Maraun, Douglas: Bias Correcting Climate Change Simulations - a Critical Review. *Curr Clim Change Rep 2, 211–220*, 2016.

[18] Heo, Jun-Haeng, Hyunjun Ahn, Ju-Young Shin, Thomas Rodding Kjeldsen, and Changsam Jeong: Probability Distributions for a Quantile Mapping Technique for a Bias Correction of Precipitation Data: A Case Study to Precipitation Data Under Climate Change. *Water* 11, no. 7, 2019.

[19] Antonios Mamalakis, Andreas Langousis, Roberto Deidda, Marino Marrocu: A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall. *Water Resources Research,* 2017.

[20] Luo, M.; Liu, T.; Meng, F.; Duan, Y.; Frankl, A.; Bao, A.; De Maeyer, P. Comparing Bias Correction Methods Used in Downscaling Precipitation and Temperature from Regional Climate Models: A Case Study from the Kaidu River Basin in Western China. *Water* 2018, *10*, 1046. https://doi.org/10.3390/w10081046.

[21] Robert Beyer, Mario Krapp, and Andrea Manica: A systematic comparison of bias correction methods for paleoclimate simulations. *Climate of the Past*. 2019.

[22] Tan Y, Guzman SM, Dong Z, Tan L.: Selection of Effective GCM Bias Correction Methods and Evaluation of Hydrological Response under Future Climate Scenarios. *Climate*. 2020.