Article

Copula bias correction for extreme precipitation in re-analysis data over a Greek catchment

Georgia Lazoglou [1,*], Christina Anagnostopoulou [1], Charalampos Skoulikaris [2,] and Konstantia Tolika [1]

Received: date; Accepted: date; Published: date
Academic Editor: name

[1] Department of Meteorology Climatology School of Geology Aristotle University of Thessaloniki (Greece)
[2] Department of Civil Engineering Aristotle University of Thessaloniki (Greece)
* Correspondence: glazoglou@geo.auth.gr; Tel.: +30-2310-998414

Abstract: The projection of extreme precipitation events with higher accuracy and reliability, which engender severe socioeconomic impacts more frequently, is considered a priority research topic in the scientific community. Although large scale initiatives for monitoring meteorological and hydrological variables exist, the lack of data is still evident particularly in regions with complex topographic characteristics. The latter results in the use of reanalysis data or data derived from Regional Climate Models, however both datasets are biased to the observations resulting in non-accurate results in hydrological studies. The current research presents a newly developed statistical method for the bias correction of the maximum rainfall amount at watershed scale. In particular, the proposed approach necessitates the coupling of a spatial distribution method, namely Thiessen polygons, with a multivariate probabilistic distribution method, namely copulas, for the bias correction of the maximum precipitation. The case study area is the Nestos river basin where the several extreme episodes that have been recorded have direct impacts to the regional agricultural economy. Thus, using daily data by three monitoring stations and daily reanalysis precipitation values from the grids closest to these stations, the results demonstrated that the bias corrected maximum precipitation totals (greater than 90%) is much closer to the real max precipitation totals, while the respective reanalysis value underestimates the real precipitation totals. The overall improvement of the outputs, shows that the proposed Thiessen-copula method could constitute a significant asset to hydrologic simulations.

Keywords: copula; thiessen polygons; extreme; precipitation; bias correction

PACS: J0101

1. Introduction

Last decades extreme precipitation episodes have been observed more frequently and their intense is greater compared with the past [1]. The impacts of these extremes are obvious in many different fields such as economy, society, agriculture, hydrology, resulting on the need of reliable projections. According to Mao et al. [2] in order to achieve higher accuracy in the hydrological results, is mandatory to correct the bias between model's used real values of several climate parameters and especially precipitations. In cases where lack of observed data is occurred, then proxy but reliable techniques, such as the use of reanalysis data, is proposed as scientifically proven solution. Bastola and Misra [3] in particular, demonstrate that reanalysis data is useful in simulating realistic hydrological response at watershed scales. At the same time, although precipitation estimates from global reanalysis are dynamically consistent with the large-scale circulation, on the other hand, when precipitation is compared with rain gauge estimates the output is poor, since reanalysis products are forecast by the reanalysis system and precipitation is not assimilated [4].

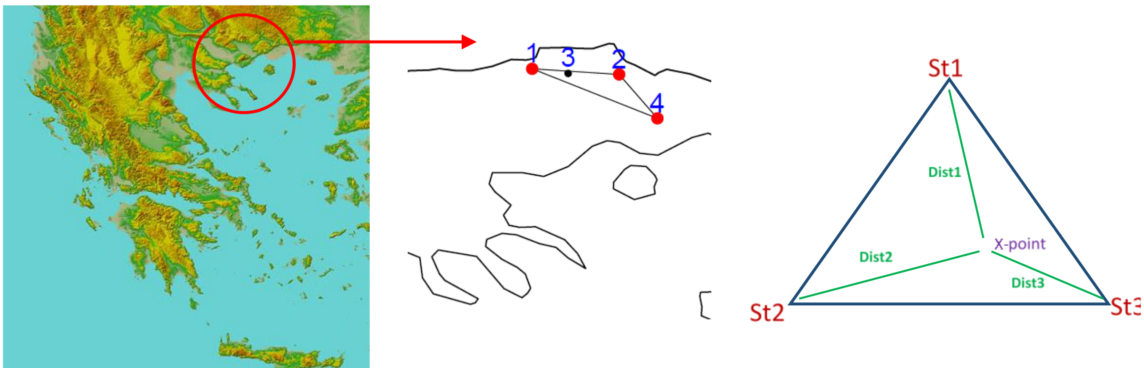
Numerous studies in the fields of insurance and finance have attempted using the Copula method [5,6] technique. The utility of copula method based on its ability to analyze the dependence of two or more random variables that have not necessarily the same distribution [7]. Additionally, copulas have the advantage to capture the features of the dependence [8] and to examine this dependence not only linearly, as the other indices do [2]. Consequently, last decades copulas have been used widely in hydrology. Shiau [9] suggest copulas for drought analysis in order to overcome the problem that drought characteristics have different distributions. Several scientists use copulas for analyzing drought characteristics (e.g. severity and intense) [10, 11] or for correlating drought with other climate parameters such as precipitation [12]. Furthermore Golian [13] used a bivariate copula function for studying the rainfall–runoff simulations of a watershed in the region of Iran, while Perera [14] used this method for studying the interdependence between the Kelani River and Kotte Canal in Sri-Lanka.

The present study investigates the combination of the copulas probabilistic distribution method with the Thiessen polygon spatial distribution method to tackle bias correction of extreme precipitations reanalysis data in the important hydrological region of Nestos river basin in Greece. Thiessen polygons are widely used in hydrology for spatial interpolation since 1980 when Lee and Schachter [15] present it. The evaluation of the result was achieved using statistical tools such as Taylor diagrams as well as the relative operating characteristics curves (ROC). The latter are popular in clinical epidemiology as they test the accuracy of a diagnosis [16] and in this study is checked the accuracy of the bias corrected values.

2. Materials and Methods

2.1 Data

The present study uses daily precipitation data from four meteorological stations located in the Nestos catchment (Figure 1a). Apart from the observed precipitation records, Era-Interim reanalysis data with spatial resolution of 12.5 × 12.5 km, were retrieved by the European Centre for Medium-Range Weather Forecasts (ECMWF) for the selected case study area. Thus, for every station the closest continental grid point that presented similar topographic characteristics was selected. Both reanalysis data and observed records cover a time period of 9 sequential years, i.e. from 1987 to 1995.



a) b) c)

Figure 1. a) The map of Greece – the red circle includes the studied region b) zoom of Nestos region and the four studied stations with the studied triangle (1= Achladia, 2= Prasinada, 3= Sidironero, 4= Toxotes), c) St1 to St3 indicate the three stations at the triangle vertices. x_point is the unknown station and dist1 to dist 3 are the distances between x_point and the three stations.

2.2 Methodology

According to Nelsen [17] copulas are multivariate cumulative distribution functions which have the ability to model mathematically the dependence between two or more variables using their marginal distributions. Assuming that X and Y are two random variables with F and G marginal distributions respectively, the joint distribution function of the initial variables is H and is equal to the copula function of their marginal (H(x,y)= C(F(x), G(y)). The centered theorem of copulas is the Sklar's theorem [18]. According to that, if the marginals of the studied variables are continuous, then the Copula function can uniquely be defined. Otherwise, C is unique on RanF × RanG (Ran is the range). The importance of this theorem is that every joint distribution function can be decomposed into the marginal of the variables and into a Copula function which describes their dependence completely.

In the present study, Copula method was combined with Thiessen triangles, which is an alteration of the thiessen polygons method, to achieve a bias correction of total extreme precipitation between real observations and reanalysis data. Firstly, three (stations 1, 2 and 4) of the four available stations, were used for the analysis while the other one (station no 3) for evaluation. Specifically, the three selected stations formed a triangle that includes the tested station (Figure 1b).

For the stations located at the triangle vertices, the absolute maximum and the monthly precipitations were used to model the dependence among them. Twelve copula families (Gaussian, Student t, Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7, BB8, Tawn type 1 and 2) coming from both Archimedean and Ellipitcal categories were tested, in order to select the one which can describe the dependence more satisfactory. The final selection was based on Akaike's information criterion [19] and Bayesian information criterion. Thereafter, using the copula results by the stations located at the triangle vertices and taking also into account the distance between them and the evaluated stations (x-point) a newly copula family was defined (Figure 1c). The new defined copula family can describe mathematically the dependence between the mean and maximum precipitation at the x-point. Consequently, using the new copula family and the reanalysis mean precipitation data, the maximum extreme bias corrected precipitations have been calculated.

3. Results

3.1. Total and Extreme precipitation analysis

The dependence between absolute maximum extreme and monthly precipitation was assessed using copulas to three of the four studied stations. After testing several copula families, it was found that the Survival Clayton family reflects the dependence for both 4 and 1 stations, while Frank copula for station 2. The dependence is more power for Prasinada station, as the Kendall's tau correlation index is almost 0.87, while for the other two stations is almost 0.78. Figure 2 visualize that relationship, confirming the fact that Survival Clayton copula presents upper tail dependence (stations 1 and 4) while Frank presents no upper or lower tail dependence (station 2).

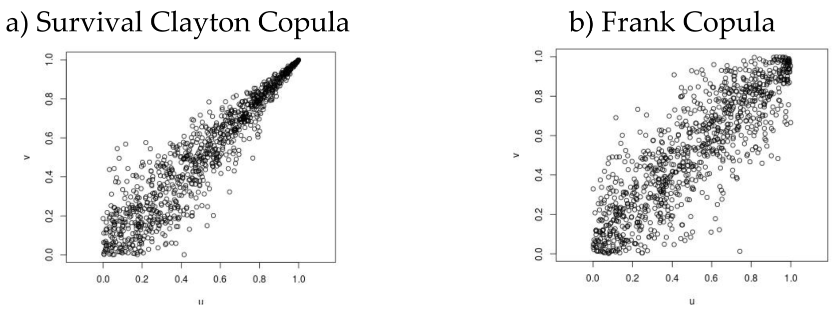


Figure 2. Structure of values which dependence is modelled by Survival Clayton or Frank Copulas.

3.2 Bias correction results

Reanalysis data present important biases from real observations especially for precipitation parameter and especially for extreme events. As a consequence the main purpose of this study is to reduce extreme rainfall biases between real and reanalysis data at the scale of a hydrological basin. Table 1 presents the observed, the reanalysis and the bias corrected extreme precipitation indices, for the station of Sidironero. Extreme precipitation indices defined as the 90th, 95th and 99th percentile of the monthly precipitation [20].

Table 1. Extreme precipitation indices (mm) from observed, reanalysis and bias corrected data

	90%	95%	99%
Observations	32.5	48	67
Reanalysis	26.3	31.5	37.4
Bias Corrected	36.3	42.5	49

As it can be seen from Table 1 for all indices the bias corrected values are closer to the observed ones compared with reanalysis data. Additionally as the percentile became higher, reanalysis present higher differences compared to observe while the bias corrected were closer to the observed ones. This is also proved by the Taylor diagram Figure 3a of the studied data sets. Particularly, Taylor diagram shows that the correlation between observed and reanalysis data was almost zero while after the bias correction the correlation increased to 0.5. Additionally, the Root Mean Square Error has been reduced while there is an increase of the variation.

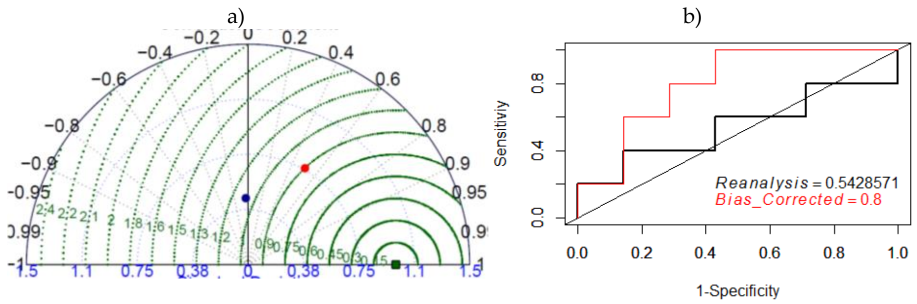


Figure 3. a) Taylor diagram of extreme precipitation at Sidironero. Blue circle presents the reanalysis data and red circle the bias corrected data b) ROC curves of extreme precipitation at Sidironero. Black line concerns the observed extreme values and red line the bias corrected values.

An additional evaluation of the results was conducted with the use of the ROC curves. According to Figure 3b the area under curve –which is an effective measure of accuracy [21] - is bigger for the bias corrected values proving that the Thiessen- Copula method can be used for the bias correction of extreme precipitation.

4. Discussion- Conclusions

The bias correction of extreme rainfall with the coupling of copula method and an alteration of the Thiessen polygon method is presented. The proposed method adjusts the extreme reanalysis precipitation data to observed data in a selected river basin in Greece. The method's evaluation was achieved after the comparison of the bias corrected values with observed datasets using different statistical and optional methods.

Lafon et al.[22], Teutschbein and Seibert [23] have studied and compared different bias correction methods in hydrology. Methods such as the linear scaling approach, delta change method [24] or the local intensity scaling [25] use simple mathematical equation for bias correction. However, they mainly focus on estimating mean values without expanding to the whole distribution [2].

Additionally as Yang et al. [26] mention, the accuracy for extremes is much lower even with more dynamical methods. At the same concept, Berg, et al. [27], indicate the importance of bias correction to reanalysis products, regarding soil moisture, runoff, and snow water equivalence, at simulations covering the geographic area of N. America.

The results of the present study shows that the Copula method combining with Thiessen triangles can be an accurate tool for rainfall extremes bias correction. Copula method present also satisfying results at Mao's [2] study, who used it for the bias correction of model's precipitation values in Germany. Additionally, Piani [28] propose the use of Copula for a two dimensional bias correction for the parameters of temperature and precipitation in climate models. In accordance with this study, the success of the method derived from the ability of copula to represent the dependence structure of the studied variables satisfactorily. Consequently, the dependence structure is not the same in every station in a specific region, as it is also observed in the three studied stations of the presented case study. Furthermore the bias corrected extreme precipitations not only were much closer to the real ones but they also have higher correlation with the observed extremes, as well as the Root Mean Square Error is lower compared with the reanalysis data. The need of bias correction is more obvious in the case of climate change, where the models outputs need to be bias corrected before it can be used for climate change impact studies. In that case various methods, such as the quantile mapping, the cumulative distribution function transform (CDF-t), equidistant quantile matching are presented in the literature [29]. Concluding, the resent investigation propose the copula method in combination with Thiessen triangles technique as a useful tool for the bias correction of extreme precipitations. It is also believed that this method would be a fruitful area for further work and it is proposed its application in large river basins.

Acknowledgments: This work was supported by a grant by the State Scholarship Foundation (IKY) in Greece

Author Contributions: All authors conceived and initiated the study. Georgia Lazoglou analyzed the data. All authors contributed to the discussion and interpretation of the results and the writing of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest

References

1. Pachauri, R.K.; Allen, M.R.; Barros, V.R.; Broome, J.; Cramer, W.; Christ, R.; Dubash, N.K. Climate Change2014: Synthesis Report; Contribution of Working Groups I, II and III to the Fifth Assessment Report of theIntergovernmental Panel on Climate Change (IPCC): Geneva, Switzerland: 2014
2. Mao, G.; Vogl, S.; Laux, P.; Wagner, S.; Kunstmann, H. Stochastic bias correction of dynamically downscaled precipitation fields for Germany through Copula-based integration of gridded observation data. *Hydrology and Earth System Sciences* **2015**, *19(4)*, 1787-1806, DOI: 10.5194/hess-19-1787-2015.
3. Bastola, S. and Misra, V. Evaluation of dynamically downscaled reanalysis precipitation data for hydrological application. *Hydrol. Process*. **2014**, 28, 1989-2002. DOI:10.1002/hyp.9734.
4. Gehne, M., T.M. Hamill, G.N. Kiladis, and K.E. Trenberth, Comparison of Global Precipitation Estimates across a Range of Temporal and Spatial Scales. *J. Climate* **2016**, 29, 7773–7795. DOI: 10.1175/JCLI-D-15-0618.1
5. Giacomini, E. Risk Management with Copula, Master's thesis - Hulmboldt University, Berlin, 2005.
6. Malevergne, Y.; Sornette, D. Testing the Gaussian copula hypothesis for financial assets dependences. *Quantitative Finance* **2003**, 3(4), 231-250, DOI: 10.1088/1469-7688/3/4/301.
7. Zhang, L.; Singh, V.P. Bivariate rainfall frequency distributions using Archimedean copulas. *J Hydrol* **2007**, *332(1)*,93–109, DOI: 10.1016/j.jhydrol.2006.06.033.
8. Canela, M.A.; Collazo, E.P. Modelling dependence in Latin American markets using copula functions. *J Emerg Market Finance* **2012,** *11(3)*, 231–270, DOI: 10.1177/0972652712466493.
9. Shiau, J.T. Fitting drought duration and severity with two-dimensional copulas. *Water Resour. Manage* **2006**, *20(5)*, 795–815, DOI: 10.1007/s11269-005-9008-9
10. Yusof, F., Hui-Mean, F., Suhaila, J., Yusof, Z. Characterisation of drought properties with bivariate copula analysis. *Water resources management* **2013,** *27(12)*, 4183-4207. DOI: 10.1007/s11269-013-0402-4
11. Mirabbasi, R., Fakheri-Fard, A., Dinpashoh, Y. Bivariate drought frequency analysis using the copula method. *Theor Appl Climatol* **2012**, *108(1)*, 191-206, DOI: 10.1007/s00704-011-0524-7

12. Fan, L.; Wang, H.; Wang, C.; Lai, W.; Zhao, Y. Exploration of Use of Copulas in Analysing the Relationship between Precipitation and Meteorological Drought in Beijing, China. *Advances in Meteorology 2017*, **2017**,DOI: 10.1155/2017/4650284
13. Golian, S.; Saghafian, B.; Farokhnia, A. Copula-based interpretation of continuous rainfall–runoff simulations of a watershed in northern Iran. *Canadian Journal of Earth Sciences* **2012**, *49(5)*, 681-691, DOI:10.1139/e2012-011
14. Perera, K.; Sumathipala, P.L.N.S.; Wikramanayake, P.N. Interdependence between the Water Levels of Kotte Canal and Kelani River using Copulas. *Sri Lankan Journal of Applied Statistics* **2015**, 16(2), 135:146.
15. Lee, D.T.; Schachter, B.J. Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences* **1980**, 9(3), 219-242. DOI: 10.1007/BF00977785.
16. Metz, C.E. ROC methodology in radiologic imaging. *Investigative radiology* **1986**, 21(9), 720-733.
17. Nelsen, R.B. An introduction to copulas, Springer, New York, 1999.
18. Sklar, A. Fonctions de R´epartition `a n Dimensions et Leurs Marges. Publications de l'Institut de Statistique de L'Universit´e de Paris **1959**, 8, 229–231.
19. Akaike, H. Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory* **1973**, 267-281.
20. WMO:https://www.wmo.int/pages/prog/wcp/ccl/opace/opace2/documents/DraftversionoftheGuidelinesontheDefinitionandMonitoringofExtremeWeatherandClimateEvents.pdf
21. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, 143(1), 29-36. DOI: 10.1148/radiology.143.1.7063747.
22. Lafon, T.; Dadson, S.; Buys, G.; and Prudhomme, C. Bias correction of daily precipi- tation simulated by a regional climate model: a comparison of methods, *International Journal of Climatology* **2013**, 33, 1367–1381, doi:10.1002/joc.3518
23. Teutschbein, C. Seibert, J. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different meth- ods, *Journal of Hydrology* **2012**, 456-457, 12–29, doi:10.1016/j.jhydrol.2012.05.052, 2012.
24. Hay, L.E.; Wilby, R.L.;Leavesley, G.H. A Comparison of Delta Change and Downscaled GCM Scenarios for Three Mountainous Basins in the United States, *Journal of the American Water Resources Association* **2000**, 36, 387–397, doi:10.1111/j. 1752-1688.2000.tb04276.x.
25. Schmidli, J.; Frei, C.; Vidale, P.L. Downscaling from GCM precipitation: a bench- mark for dynamical and statistical downscaling methods, *International Journal of Climatology* **2006**, 26, 679–689, doi:10.1002/joc.1287.
26. Yang, W.; Andr´easson, J.; Phil Graham, L.; Olsson, J.; Rosberg, J.; Wetterhall, F. Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, *Hydrology Research* **2010**, 41, 211, doi: 10.2166/nh.2010.004
27. Berg, A.A.; Famiglietti, J.S.; Walker, J.P.; Houser, P.R. Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes. *Journal of Geophysical Research: Atmospheres* **2003**, 108,4490, DOI: 10.1029/2002JD003334, D16.
28. Piani, C.; Haerter, J.O. Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophysical Research Letters* **2012**, 39(20), DOI: 10.1029/2012GL053839
29. Pierce, D.W.; Cayan, D.R.; Maurer, E.P.; Abatzoglou, J.T.; Hegewisch, K.C. Improved Bias Correction Techniques for Hydrological Simulations of Climate Change. *J. Hydrometeor* **2015**, 16, 2421–2442, DOI:10.1175/JHM-D-14-0236.1.

Abbreviations

The following abbreviations are used in this manuscript:

ROC: Relative Operating Characteristics curves