Batch orographic interpolation of monthly precipitation based on free-of-charge geostatistical tools

Ondrej Ledvinka[1,*]

[1]Czech Hydrometeorological Institute, Hydrology Database and Water Budget Department, Na Sabatce 2050/17, 143 06 Prague 412, Czechia

Abstract. The effects of possible climate change on water resources in prescribed areas (e.g. river basins) are intensively studied in hydrology. These resources are highly dependent on precipitation totals. When focusing on long-term changes in climate variables, one has to rely on station measurements. However, hydrologists need the information on spatial distribution of precipitation over the areas. For this purpose, the spatial interpolation techniques must be employed. In Czechia, where the addition of elevation co-variables proved to be a good choice, several GIS tools exist that are able to produce time series necessary for climate change analyses. Nevertheless, these tools are exclusively based on commercial software and there is a lack of free-of-charge tools that could be used by everyone. Here, selected free-of-charge geostatistical tools were utilized in order to produce monthly precipitation time series representing six river basins in the Ore Mountains located in NW Bohemia, Czechia and SE Saxony, Germany. The produced series span from January 1961 to December 2012. Rain-gauge data from both Czechia and Germany were used. The universal kriging technique was employed where a multiple linear regression (based on elevation and coordinates) residuals were interpolated. The final time series seem to be homogeneous.

1 Introduction

In hydrological studies dealing with climate change and its effect on water resources, the need of having long precipitation time series is of primary focus because precipitation is generally thought of as a basic source of water in many regions. This particularly applies to the territory of Czechia, for instance, which is known to be a source area of many European rivers, while, in fact, no large river crosses this country that would be, on the other hand, beneficial in cases of prolonged drought occurrences. Furthermore, hydrologists prefer to have the information on spatial distribution of precipitation during rainfall events or, more generally, of their means in different places in prescribed time horizons. However, traditionally, the precipitation totals were (and still are) measured at selected points where rain gauges were established. It is clear that this was for economic reasons because we are

---

[*] Corresponding author: ledvinka@chmi.cz

not able to measure everywhere, and it was also found that, in the special case of the mean, it is not necessary to have too many stations spread over an area of interest because there is a limit above which one cannot reach desirable increment of information (e.g. [1]). Radar and satellite precipitation data can be treated only as auxiliary because these measurements started much later (usually in the 1980s) than rain-gauge observations. Therefore, it is necessary to come up with a method capable of estimating the spatial distribution of precipitation using the point records.

For these purposes, spatial interpolation techniques evolved in the past and, especially, the field of geostatistics emerged. Geostatistical applications are not new in hydrology. Rather, it can be said that hydrologists alongside hydrogeologists importantly contributed to the development of geostatistics (e.g. [2–5]). Groundwater hydrologists were probably the first who applied methods of geostatistics but, naturally with the rise of the attention paid to climate change, the shift towards applications to surface water problems and climatology is evidenced by many papers published in the past few decades (e.g. [6,7]). Besides commercial software, the geostatistical techniques were also implemented in freely available tools, which is beneficial to students and other scientists who cannot pay much for them. Thus, the present contribution aims to show how selected tools can be wisely utilized in order to provide valuable information that can be further used in different studies dealing with climate change. The issue of spatial interpolation of precipitation totals in the Ore Mountains located in Central Europe served as an example.

The Ore Mountains stretched along the borders between NW Bohemia, Czechia and SE Saxony, Germany are known to be one of the regions that are sparsely measured regarding precipitation. Also, many gaps occur in time series representing points. This naturally poses many issues to hydrologists studying climate change impacts on small mountainous basins here. Therefore, the task was to estimate reliable time series suitable for climate change studies for selected more or less anthropogenically uninfluenced small river basins. Although so-called technical precipitation series do exist in Czechia that are uninterrupted in the daily time step and should be homogeneous (see [8–10]), similar data could not be gathered from Germany. For this reason, the monthly time step was selected as basic that still allows one to study changes in the annual course of precipitation. The selection of suitable basins was made based on the existence of relatively long discharge series so as to also allow for the comparison of long-term patterns in both precipitation and discharge in some of the future works.

Table 1. Basic morphometric characteristics of six selected small river basins/catchments located in the Ore Mountains and delineated by the water-gauging stations represented by long discharge series. Crosses indicate the east. For the definition of Graveli coefficient see [11].

Station (ID)	River	Area [km²]	Max. elevation [m a.s.l.]	Outlet elevation [m a.s.l.]	Mean elevation [m a.s.l.]	Mean slope [decimal degrees]	Graveli coefficient
Klingenthal (530020)	Zwota (Svatava)	58.62	930	541	707	9.17	1.52
Svatava (208200)	Svatava	291.64	988	406	673	8.81	1.81
Chaloupky (209100)	Rolava	20.06	950	815	904	2.28	1.50
Stará Role (210100)	Rolava	126.35	1008	392	735	4.43	1.83
Rothenthal[+] (568350)	Natzschung (Načetín B.)	75.94	918	540	758	5.56	1.48
Třetí Mlýn[+] (218000)	Chomutovka	44.91	914	485	770	7.76	1.77
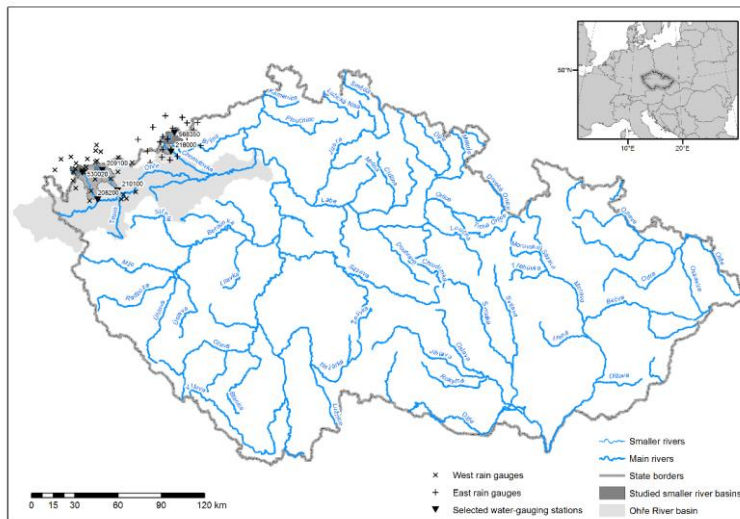
Fig. 1. Location of six studied small mountainous river basins in relation to the borders of Czechia and the larger Ohře (Eger) River basin. Except for the Natzschung, all investigated rivers are left-hand tributaries of the Ohře. All German and Czech rain gauges are depicted from which the data were used, but it does not mean that each of them was available for every month during interpolations.

2 Material and methods

2.1 Area of interest and gathered material

As described above, smaller river basins located in the Ore Mountains were selected. The final selection of six basins was determined by relatively long instrumental observations of discharge that was, at the same time, believed to be not influenced, or only slightly influenced, by human activities in such places. During processing the accessible material, the area of interest was divided into two parts, simply four basins in the western part delineated by water-gauging stations 530020, 208200, 209100 and 210100, and two basins in the eastern part closed by stations 568350 and 218000. The basic information on the morphometric characteristics of these basins is given in Table 1. Note that some of the values provided in the table may slightly differ from those officially published (e.g. at http://hydro.chmi.cz/ismnozstvi/ or in [12]). This is because the values in Table 1 come from the author's own derivation from the GIS material described further below. The location of the basins relative to the borders of Czechia and the larger, and thus better known, Ohře (Eger) River basin is depicted in Fig. 1. Two of the western basins, particularly 530020 and 209100, are in fact the subbasins of the Svatava River basin closed at the profile 208200 and the Rolava River basin closed at the profile 210100, respectively.

Precipitation totals measured at several rain gauges close to or within the basins were gathered from the Czech Hydrometeorological Institute (CHMI) and the German weather service called Deutscher Wetterdienst (DWD). Regarding the Czech data, the so-called technical precipitation series were acquired with the assistance of their main author Petr Štěpánek. These data were in the daily time step, were homogenous and without any missing values. They were and still are derived from the originally recorded precipitation series (see [8,9]). Since the technical series start on 1 January 1961, all the remaining analyses focused

on the period 1961–2012. The year 2012 was chosen arbitrarily, but it was also due to the accessibility of the German discharge series from stations 530020 and 568350 (see [12]). However, there is no problem with exteding the length of precipitation series towards recent times in some of the future works. The German precipitation data were somewhat problematic since no technical series (or their equivalent) were available for the purposes of this study. Instead, the data published on the FTP server linked to on the website of the DWD's Climate Data Centre (see http://www.dwd.de/) were used. The daily data were too gappy and therefore monthly data were rather processed in the subsequent analyses. This means that also the Czech data had to be aggregated so that they finally formed monthly data. The total number of rain gauges included in the analyses was 32 (west) + 24 (east) = 56, of which 18 (west) + 11 (east) = 29 rain gauges represented the German territory. However, it is worth noting that the German rain gauges without any data for a given month did not enter the interpolation process.

Because the chosen interpolation technique required elevation data, a digital elevation model (DEM) represented as a grid layer was used to provide the necessary information. Most of DEMs available for the territory of Czechia are unfortunately clipped by its borders, which is undesirable when investigating river basins that cross the borders. Therefore, an alternative DEM had to be used. Here, a somewhat older DEM, downloaded at the end of the 2000s by the author from the website of the Czech company called ARCDATA PRAHA (https://www.arcdata.cz/), served as an elevation input to the geostatistical models. This DEM is based on the well-known SRTM product, but it was further processed by ARCDATA experts so that a number of possible errors was reduced. The approximate spatial resolution of the DEM was 75 m.

The last input to the geostatistical models were the geographical coordinates pertaining to the rain gauges. This information was provided as the metadata, for the German stations available online together with the precipitation data themselves. For the Czech territory, this information was acquired from the CHMI's CLIDATA database or directly from Petr Štěpánek.

2.2 Methods and tools utilized

All computations were carried out within the environment of the R statistical software [13] which is freely available online for all. Besides a number of packages that are intended to handle spatial data such as grids and vector features and are helpful when tranforming various coordinate systems to a common working system (see e.g. [14] for an overview), mainly the functions of packages 'automap' [15], which needs the package 'gstat' [16–18] to automatically perform kriging in general, and 'MASS' [19] were applied to the precipitation and elevation data and coordinates. It is known that the precipitation totals measured at points spread over a mountainous area correlate with elevation, but the correlation is higher if the elevation representative of a wider vicinity of a rain gauge is studied instead of the elevation corresponding to the place where the rain gauge is situated (see e.g. [20]). Therefore, also the 'raster' package (and its function 'focal()'; see [21]) had to be applied in order to compute mean elevation values representative of places up to 6 and 12 km away from the rain gauges. These distances were derived experimentally, while looking at the values of the Akaike Information Criterion (AIC; [22]) after bulding multiple linear regression (MLR) models. In particular, the package 'MASS' was used to determine the values of AIC, which should be the lowest for the right model. The package was also helpful when deciding on the explanatory variables that should enter the MLR models fed to the universal kriging models to which the 'krige()' function of the 'gstat' package switches according to the formula specified.

The procedure itself can be summarized as follows:

- Having four explanatory variables (E_6 – elevation typical of places up to 6 km far from a rain gauge; E_{12} – elevation typical of places up to 12 km far from a rain gauge; X – longitude of a rain gauge; Y – latitude of a rain gauge) try to build a MLR model:

$$P = \alpha + \beta_1 E_6 + \beta_2 E_6^2 + \beta_3 E_6^3 + \beta_4 E_{12} + \beta_5 E_{12}^2 + \beta_6 E_{12}^3 + \beta_7 X + \beta_8 Y + \varepsilon \qquad (1)$$

where α and β indicate nine parameters to be estimated (i.e. intercept and different slopes specified by the subscripts), ε is the residual component and the superscripts indicate powers whose inclusion has been shown beneficial for several mountainous areas (see e.g. [20]). The estimated parameters can in turn be used for the derivation of precipitation estimates $\hat{P}$ for every cell of the DEM where also the coordinates are known. Here, only two rectangular areas for the western and the eastern part of the Ore Mountains were selected (Fig. 2).

- It is not necessary to have all the explanatory variables in the model. Therefore, using the package 'MASS' and its function called 'stepAIC()', decide which of the variables should be included in the MLR model.
- The model with the lowest value of AIC should enter the 'autoKrige()' function, where the (semi)variogram is estimated by the generalized least squares (GLS) method based on the work [23].
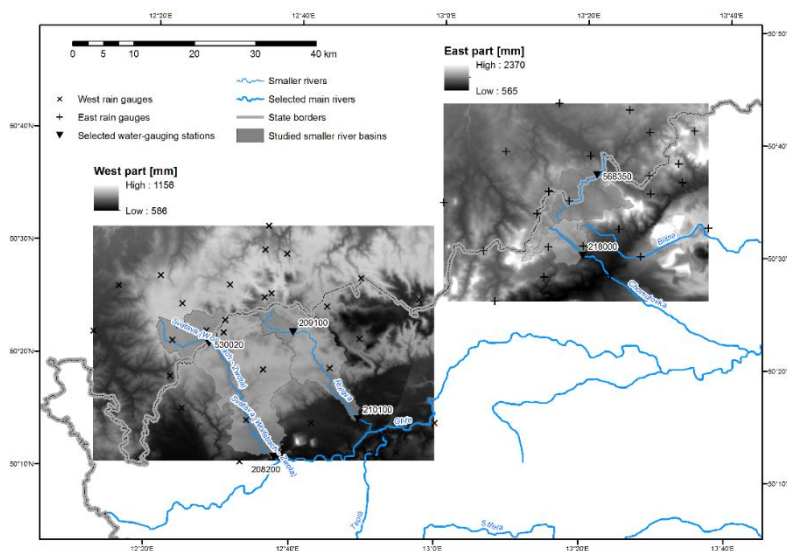
Fig. 2. Closer look at the area of interest with resulting long-term (1961–2012) grids of annual precipitation totals in greyscale separately for the western part and the eastern part of the Ore Mountains. Note that some of the extreme precipitation values may be far from true due to the absence of rain gauges in some places. However, it is believed that within the catchments the totals are much more reliable.

- Because the maximum number of the explanatory variables here was 8, the minimum number of rain gauges, when estimating the value of precipitation for a given DEM cell, was 9. The maximum radius, where the suitable rain gauges had to be searched, was set to 50 km.
- Finally, it is necessary to determine the precipitation totals that should represent the selected river basins. This may be accomplished using the 'zonal.stats()' function of the 'spatialEco' package [24].
- Repeat the procedure for all months. Since there were 624 months in the period 1961–2012 and, in fact, two regions were investigated, the procedure was run 1248 times. Because R is known to be relatively slow, the author recommends to parallelize the computation for larger data sets.

2.3 Homogeneity of produced series

Climatological time series should be homogeneous, which means that the different measuring equipment, the vicinity of a station or changing observers should not have any influences on the population from which the time series is drawn. Therefore, it was desirable to test the homogeneity of the new precipitation series. For this purpose, another R tool, developed in Canada [25], was used. Because this tool works with daily precipitation data, the monthly time series were disaggregated to daily, dividing the monthly totals by the corresponding number of days, which gave the average totals for every day. Then, the tool could be applied to these artificial daily precipitation series.

Table 2. Long-term monthly and annual precipitation totals (in mm) gained after computing spatial means for each month of the period 1961–2012 following universal kriging applied to precipitation totals known for rain gauges depicted in Figs. 1 and 2. The computations run separately for the western basins and for the eastern basins (these indicated by cross).

Station/ Season	530020	208200	209100	210100	568350[+]	218000[+]
01	80.3	73.8	85.5	75.6	69.8	68.5
02	67.8	60.7	67.7	60.6	63.7	60.4
03	73.4	64.1	64.8	62.7	71.8	66.9
04	61.9	53.4	51.0	49.8	66.4	61.1
05	74.9	68.9	61.5	63.6	87.4	81.1
06	93.6	82.5	73.4	75.0	98.5	94.8
07	98.9	88.6	82.6	83.3	111.5	103.4
08	95.5	83.6	80.9	77.2	106.0	99.7
09	76.6	69.3	66.3	66.2	80.6	75.5
10	67.9	61.5	63.2	58.4	64.2	60.5
11	77.4	69.6	72.0	67.7	75.6	71.7
12	93.5	84.3	86.5	80.6	90.8	90.4
Year	961.7	860.4	855.5	820.8	986.3	934.1

3 Results and discussion

According to the main task, complete monthly precipitation time series representing six smaller river basins located in the Ore Mountains were produced, considering both Czech and German rain gauges. The series cover the period 1961–2012, but it does not mean that the analyses could not be conducted also for more recent years. The original rain-gauge data do exist for the years 2013–2016 (see e.g. http://www.dwd.de/). Not all the aspects of the time series can be presented here. Rather, the resulting data were summarized in the form of long-term annual precipitation as the grid layers added to the other geographical features in Fig. 2. Also, long-term monthly precipitation totals for all the six basins are presented in Table 2. The final data are prepared for further analyses.

From Fig. 2, it can be seen that the pattern of the long-term annual precipitation somewhat mimic the distribution of elevation. It is caused by the method of universal kriging itself where the dependence of precipitation on elevation, besides coordinates, was the main source of information when building the models. Moreover, some extremes can be observed in Fig. 2. Whereas the western part looks more realistic, the eastern part includes questionable values (cf. [26]). It may be caused by the fact that there are some places without any rain

gauges at all. The resulting grid for the eastern part of the Ore Mountains can then reflect some overestimation of precipitation totals by the geostatistical model utilized. Nevertheless, it is believed that regarding the investigated river basins, the estimated totals are reliable as shown in Table 2.

Table 2 reveals interesting features of the precipitation distribution in the Ore Mountains. It seems that the eastern part is more humid that the western part, which may be supported by the less annual precipitation over the catchment 209100 (also located higher) compared to the situation in the east. Among other aspects, one may notice that the summer months are dominating. However, one exception occurs. It is the catchment 209100 where the winter precipitation is higher from a long-term perspective.

Regarding their quality, all the new series revealed homogeneity according to the Canadian R tool. This feature of the series signify that the next analyses will be valuable also concerning the assessment of the possible anthropogenic influences on discharge.

Last but not least, it is important to note that the selection of the monthly time step actually allowed this type of analysis. Choosing a finer step such as daily, would cause several problems. For instance, the spatial dependence of precipitation totals at different rain gauges would vanish, especially for summer months when predominantly convective rainstorms occur. During these situations, only small areas are usually hit and the kriging methods combined with the regression based on the elevation are no longer recommended (see e.g. [7, 27]).

4 Conclusion and recommendations

For the purposes of future analyses dealing with possible climate change effects in the Ore Mountains on their water resources, six smaller river basins were selected that are believed to be anthropogenically uninfluenced (or at least very slightly influenced). Because the future analyses are planned so that the precipitation totals should be compared with discharge, it was necessary first to pay attention to the issue of the spatial interpolation of data recorded at various rain gauges that existed or still exist in the area of interest (both in Czech and German territories). The method of universal kriging was chosen that incorporated the dependence of monthly precipitation on elevation, longitude and latitude. The monthly time step was selected because of the good accessibility of such data. Free-of-charge geostatistical tools available online in the form of different packages for the R statistical software were utilized. Finally, the monthly precipitation series representing all the six basins were produced that span from January 1961 to December 2012. Furthermore, all the new series were tested for their homogeneity, and it may be concluded that the series are homogeneous, which undoubtedly signify a good start of the next analyses.

On the other hand, some important features of the new series were not studied, but it is highly recommended to do so before their definite use. For instance, the utilized geostatistical technique was not compared to any of its deterministic counterparts. The validation of the results was not performed either. This may be done via a cross-validation approach or through a hydrological model as described in [7].

References

1. D.P. Lettenmaier, Water Resour. Res. **12**, 1037 (1976)

2. J.P. Delhomme, Adv. Water Resour. **1**, 251 (1978)

3. P.K. Kitanidis, in *Handb. Hydrol.*, edited by D. R. Maidment (McGraw-Hill, New York, 1993), p. 20.1-20.39

4. P. K. Kitanidis, *Introduction to Geostatistics: Applications to Hydrogeology* (Cambridge University Press, Cambridge; New York, 1997)

5. Y. Rubin, *Applied Stochastic Hydrogeology* (Oxford University Press, Oxford, New York, 2003)

6. G. Hudson, H. Wackernagel, Int. J. Climatol. **14**, 77 (1994)

7. S. Ly, C. Charles, A. Degré, Biotechnol. Agron. Société Environ. **17**, 392 (2013)

8. P. Štěpánek, P. Zahradníček, R. Huth, Idojaras **115**, 87 (2011)

9. P. Štěpánek, P. Zahradníček, A. Farda, Idojaras **117**, 123 (2013)

10. O. Ledvinka, Acta Hydrol. Slovaca **16**, 199 (2015)

11. J. Krešl, in *For. Amelior.*, edited by O. Riedl and D. Zachar (Elsevier, Amsterdam, 1984), pp. 7–90

12. LfULG, *Hydrologisches Handbuch* (Landesamt für Umwelt, Landwirtschaft und Geologie, 2012)

13. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2017)

14. R. Bivand, E.J. Pebesma, V. Gómez-Rubio, *Applied Spatial Data Analysis with R*, 2nd ed. (Springer, New York, 2013)

15. P.H. Hiemstra, E.J. Pebesma, C.J.W. Twenhöfel, G.B.M. Heuvelink, Comput. Geosci. **35**, 1711 (2009)

16. E.J. Pebesma, C. G. Wesseling, Comput. Geosci. **24**, 17 (1998)

17. E.J. Pebesma, Comput. Geosci. **30**, 683 (2004)

18. B. Gräler, E. Pebesma, G. Heuvelink, R J. **8**, 204 (2016)

19. W.N. Venables, B.D. Ripley, *Modern Applied Statistics with S*, 4th ed. (Springer, New York, 2002)

20. C. Portalés, N. Boronat, J.E. Pardo-Pascual, A. Balaguer-Beser, Int. J. Climatol. **30**, 1547 (2010)

21. R.J. Hijmans, *Raster: Geographic Data Analysis and Modeling* (2016)

22. S. Konishi, G. Kitagawa, *Information Criteria and Statistical Modeling* (Springer, New York, 2008)

23. N. Cressie, J. Int. Assoc. Math. Geol. **17**, 563 (1985)

24. J.S. Evans, *spatialEco* (2016).

25. X.L. Wang, H. Chen, Y. Wu, Y. Feng, Q. Pu, J. Appl. Meteorol. Climatol. **49**, 2416 (2010)

26. R. Tolasz, editor, *Climate Atlas of Czechia* (Czech Hydrometeorological Institute; Palacký University Olomouc, Praha; Olomouc, 2007)

27. S. Ly, C. Charles, A. Degré, Hydrol. Earth Syst. Sci. **15**, 2259 (2011)