ORIGINAL PAPER

A model for estimating rains' area, using the dependence of the time correlation of sites' monthly precipitation totals on the distance between sites

Adam Walanus[1] · Marta Cebulska[2] · Robert Twardosz[3]

Abstract Based on the monthly precipitation series from 16 sites (in the Polish Carpathian Mountains), of 132 years' length, a relatively precise scatterplot of correlation coefficients between sites versus distance between sites is obtained. The "rains" of Gaussian shape, in the spatial sense, are a good model, which produces a scatterplot very closely resembling the observed one. The essential parameter of the model is the area covered by the modeled rains, which results to be of order 30–50 km, though with about a twice lower value for the N-S direction.

1 Introduction

It is hardly unusual that rain records from nearby sites are correlated in time. Similarly obvious is that the amount of such correlation will decrease with the distance between sites. For the distance $d=0$ km, the correlation coefficient would be $r=1$, where there is no measurement error, and for $d>$ 1000 km, r will be much lower, though not 0, because of time fluctuations (seasonal oscillations) in the precipitation.

The shape or the numerical parameters of the functional relation, $r=r(d)$, would give some assistance in estimating of the spatial size of the precipitation.

Poland receives amounts of precipitation that are highly variable in both space and time (Kożuchowski and Żmudzka 2003; Kożuchowski 2004). Nowhere is this feature so evident as in the Polish Carpathian Mountains, a range that stretches over a long distance along the east-west axis, and is thus open to the conflicting influences of maritime and continental air masses. The greatest amounts of precipitation are recorded in the culminating Tatra range, in the south, and in the western-most Beskid Śląski range, which is exposed to rain-heavy winds from the west (Niedźwiedź and Obrębska-Starklowa 1991). Other research (e.g., Cebulak 1997; Cebulak et al. 2000; Niedźwiedź 1999, 2003) suggests that the high monthly values of precipitation observed in southeastern Poland could be the result of (i) local torrential rainfall caused by either free convection in uniform air masses or by dynamic convection in frontal zones, especially on a cold front, or (ii) several-day-long spells of steady rainfall in shallow stationary low-pressure systems. Both the short- and long-lasting rainfall may be the direct cause of the flooding of rivers and streams in the upper Vistula river catchment basin.

This study addresses an apparent gap in the available research on precipitation variability in Poland (Kożuchowski 2004) and Central Europe (Niedźwiedź et al. 2009), which excludes an ample body of data from mountain weather stations, especially from the Polish Carpathian Mountains.

The objective of the study was to develop a model of monthly precipitation on the basis of observed spatial correlations. The precipitation data from 16 weather stations in the Polish Carpathian Mountains and in their foreland were taken into account, spanning the period 1881–2012.

✉ Robert Twardosz
r.twardosz@uj.edu.pl

Adam Walanus
a@adamwalanus.pl

[1] Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, al. A. Mickiewicza 30, 30-059 Kraków, Poland

[2] Institute of Water Engineering and Water Management, Cracow University of Technology, ul. Warszawska 24, 31-155 Kraków, Poland

[3] Department of Climatology, Jagiellonian University, ul. Gronostajowa 7, 30-387 Kraków, Poland
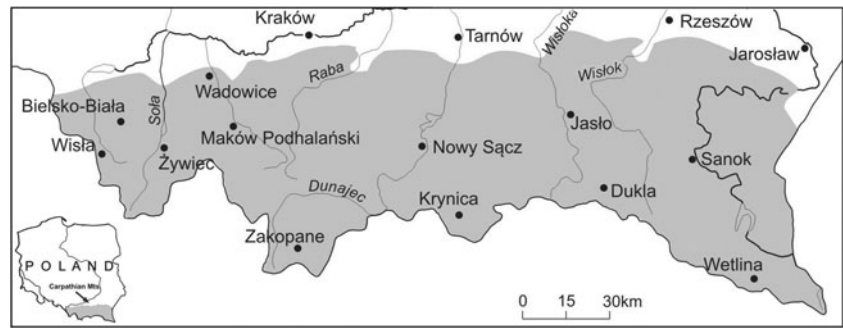
2 Data and methodology

The study employed data recorded at 16 weather stations over a period of 132 years, between 1881 and 2012. Of these stations, 12 were located in the mountains themselves and, according the regional classification of climate proposed by Niedźwiedź and Obrębska-Starklowa (1991), eight of them belonged to the mountain climatic region and four to the Carpathian Foothill climatic region (which, including Nowy Sącz, belongs to a category of inversion-prone mountain basins). The remaining four stations were located in the Carpathian foreland and belonged to the submountain basin climate region (Fig. 1). The database was built using observation data, readily available in hydrographical, meteorological, and precipitation annuals, but not hitherto used in published research.

To address the relocation of some of the stations, the monthly precipitation records were checked for homogeneity using the Standard Normalised Homogeneity Test (SNHT),

by Alexanderson (1986). The records were tested using the homogeneous precipitation record from Cracow. Anklim software package, available online at http://www.sci.muni.cz/~pest (Štěpánek 2006), was used. Taken at a significance level of 0.05, the results of the exercise did not falsify the hypothesis that the records were homogenous.

3 General characteristic of precipitation

During the study period, the area received, on average, between 650 mm (Rzeszów) and 1127 mm (Zakopane) of precipitation (Table 1). The maximum annual total of 1694 mm was recorded at the westernmost station of Wisła, while the minimum annual total of below 400 mm was recorded at the foreland stations in Rzeszów and Tarnów. This indicates that precipitation increases not just with growing altitude but also with decreasing continentality, i.e., from east to west. Other important features of the precipitation include the influence of

Table 1 Statistical characteristics of annual precipitation totals (1881–2012)

| No. of station	Name of station	Altitude (m a.s.l.)	Annual totals $P_{av.}$ (mm)	Coef var (%) ±6.2	Asymmetry ±0.21
1	Wisła	433	1185±16	15.2	0.32
2	Bielsko-Biała	322	994±15	17.0	0.30
3	Żywiec	354	879±14	18.1	0.28
4	Wadowice	268	748±11	17.5	0.19
5	Maków Podhalański	359	908±14	17.9	0.13
6	Kraków	206	681±18	16.5	0.68
7	Zakopane	844	1128±16	16.2	0.21
8	Nowy Sącz	292	728±11	16.6	0.28
9	Krynica	613	861±12	16.5	0.41
10	Tarnów	225	706±12	20.1	0.34
11	Jasło	240	720±12	19.0	0.76
12	Dukla	351	833±13	17.9	0.44
13	Rzeszów	214	652±10	18.4	0.10
14	Sanok	314	786±13	19.2	0.39
15	Wetlina	700	1075±18	19.0	0.54
16	Jarosław	204	697±12	19.1	0.26
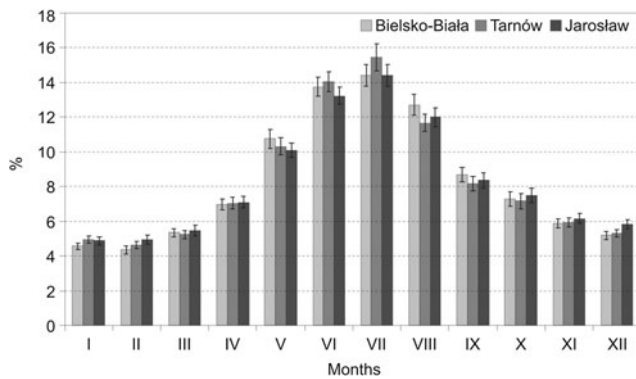
Fig. 2 Annual variation of precipitation (as a percentage of annual totals)

the landform, i.e., clearly lower precipitation values in mid-mountain basins, e.g., at Nowy Sącz, and an increased precipitation dispersion from west to east, as determined with the variability coefficient.

The annual precipitation cycle involves a clear maximum in July, recorded at all of the stations, and a minimum in February or January (Fig. 2). The annual cycle is similar across the whole area.

The general statistical characteristics are presented in Table 1.

4 Normalization of precipitation

What is clear, as is visible in Fig. 2, is the seasonality of the data. Such a feature creates correlation between sites, independently of how far they are from each other. Also, the random variability of precipitation from month to month supports a positive correlation, if the fluctuations cover a large enough

area. For example, the existence of dry and wet years produces correlation. To avoid the influence on the correlation coefficient of the time-dependent factor, which is not our interest here, we performed a kind of normalization of the data (Dutka and Hanson 1989). The operation consists simply of the division of the precipitation for the given "time point" (the "point" is a month here) for each site, by the sum for all sites (for that time point). The resulting transformed data are a series of fractions of precipitation for the given site, related to all sites. Such a quantity has little general meaning but carries exact information about space relations.

The transformation described moves the correlations between sites toward the negative values of the correlation coefficient. Clearly, this is the result of normalization. In the case of two sites only, the resulting correlation coefficient would be precisely $r=-1$, since the time series would be of the type 0.7 and 0.3, 0.1 and 0.9, 0.5 and 0.5. The normalization of 16 (our case) random number series (normally distributed white noise, i.e., not correlated at all) produces the average correlation among series of $r=-0.067$. In the case of our real, correlated data, the correlations have shifted from the ranges $r=0.90$ (close sites) to $r=0.60$ (far sites), to the ranges of $r=0.45$ to $r=-0.38$ (Fig. 3). The precipitation is seemingly not a Gaussian white noise, which is not strange.

5 Dependence of the correlation between the sites' precipitation on the distance between them

Using the multivariate time series of 1584 monthly precipitations for the 16 sites, it is possible to search for the time as well as space dependences. In this article, we pay attention to the
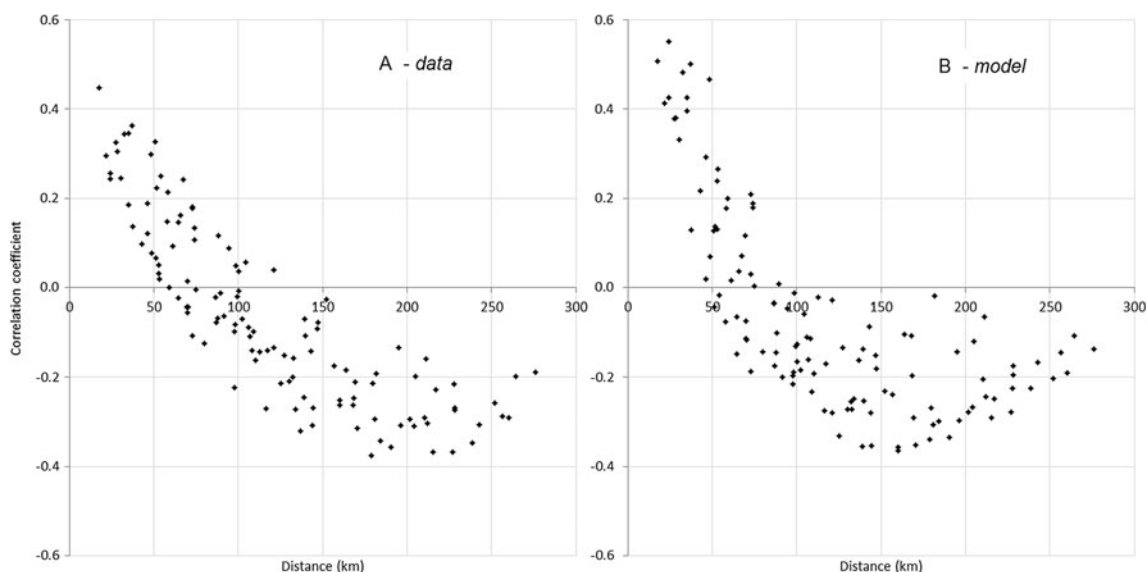

Fig. 3 Scatterplots of the correlation coefficient between sites and distance between sites. For 16 sites, there are 120 (=15*16/2) pairs of sites. Presented are both reality and its best model
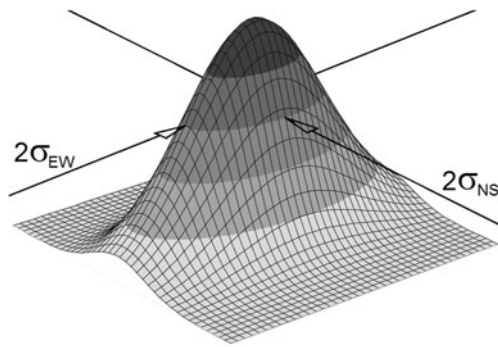
Fig. 4 The single "rain" intensity as modeled by a two-dimensional Gaussian function. The presented shape is an asymmetrical one ($\sigma_{EW}= 2\sigma_{NS}$)

second category. For 16 sites, there are 120 pairs of them: Dukla-Jasło, Dukla-Sanok, Sanok-Krynica, Zakopane-Wadowice, and so on. There are as many correlation coefficients for precipitation series and as many distances between sites. The dependence of the correlation on distance is presented in Fig. 3a. The correlation coefficient is simply that of Pearson. The closest sites (17.5 km) are the best correlated, i.e., $r=0.45$ (the data being normalized, see the previous chapter). Starting from the distance of about 80 km, the correlation vanished and, from about 150 km, became constantly negative at an average level of $r=-0.27$.

6 Modeling the dependence of correlation between sites on the distance between sites

The aim of the model is to model data well. We need to artificially generate time series that will produce a relation of the correlation coefficient to the distance close to that observed when the real data are used in calculations. Jumping

ahead, the scatterplots in Fig. 3 present the results for both data and model. There seems to be a reasonable convergence of the two plots.

The model consists of Gaussian-shaped (Fig. 4) "precipitation" (see, for example, Apputhurai and Stephenson 2013). Since what is to be modeled are monthly totals, many individual precipitation, occurring within given months, are necessary. We assumed that the Gaussian "hump" may model the precipitation, including its movement across the area (where W-E winds prevail). The sample "month" presented in Fig. 5 consists of 30 precipitation which are Gaussian, with standard deviations (sigmas) different in the W-E and N-S directions. That may be accepted as taking into account, heuristically, the predominant wind direction, as well as some observed features of the scatterplot (Fig. 3). In the case of the symmetrical bell curves, i.e., with the same sigma along the W-E and N-S axes, the model produces, for close sites, too precise a relation of correlation to distance, with almost no spread. As is rather obvious, if the distance between sites is measured isotropically and the "rains" are not isotropic, the relation of the correlation to the distance is poorer because the W-E neighboring sites are better correlated than their N-S neighbors. For a similar reason, in order to lower the correlation for the close sites, we applied a random factor for the amplitude of each Gaussian rain. The factor is not very sophisticated, being simply a random variable uniformly distributed between 0 and 1. That is connected to another interesting, though understandable, feature of the model, namely that the final result (Fig. 3b) does not depend on the number of individual Gaussian rains applied per month.

The proportion of σ_{WE} to σ_{NS} is not really essential to the model. We accepted $\sigma_{WE}/\sigma_{NS}=2$. As a result, the "clouds" visible in Fig. 5 (and more precisely in Fig. 4) are ellipsoidal rather than circular. Essential to the final result (i.e., the
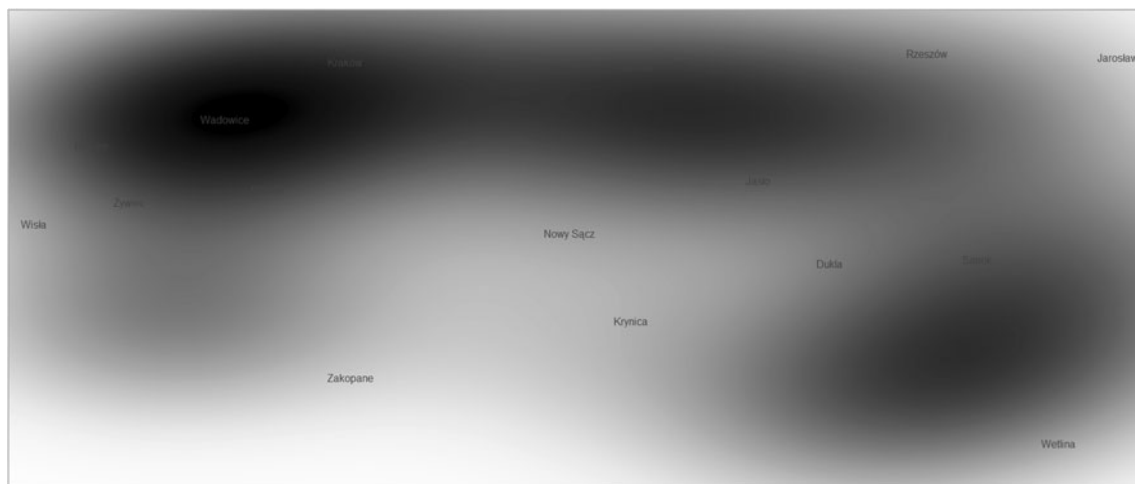

Fig. 5 Single case (single month) of the model of precipitation in the region discussed. The scatterplots presented (Figs. 3b and 6) are based on 10,000 such types of "month"
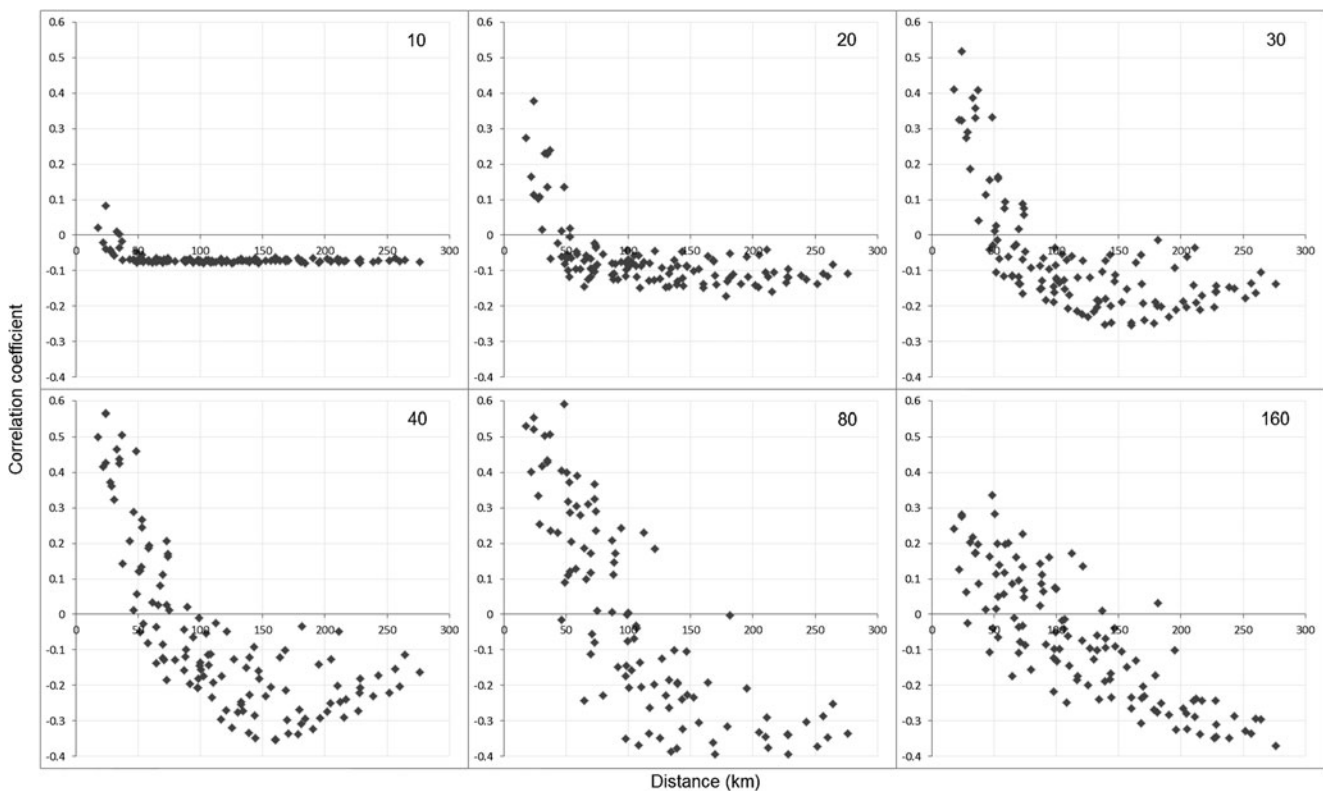
Fig. 6 Modeled scatterplots of the correlation coefficient between sites and distance between sites, as obtained for σ_{NS}=20 km and σ_{EW}=10, 20, 30, 40, 80, and 160 km

convergence of Fig. 3a, b) is the value of the sigma, i.e., the Gaussian spread (measured in km). The best result is obtained for σ_{WE}=40 km (σ_{NS}=20 km).

We have not applied any numerical criterion for quantifying the similarity between the empirical and modeled scatterplots such as that in Fig. 3. It seems to be difficult, though it is probably possible, to encapsulate all the features of the scatterplots. Certainly, the simple correlation coefficient is not an option here. One approach to the conclusion about the precision of the final result (σ_{WE}=40 km) is presented in Fig. 6. Presented is the series of modeled scatterplots obtained with different values of σ_{WE}. Bearing in mind Fig. 3a, the versions of plots in Fig. 6, with σ_{WE}=20 km, as well as σ_{WE}=80 km, are unacceptable. This leads to the final result in the form σ_{WE}=40±10 km or, alternatively, σ_{WE}=30–50 km. The latter is preferable, since the quantity ±10 km by no means has the sense or precision of the standard deviation (another sigma).

result (Fig. 3) by assuming σ_{WE}=40 km. Our rains are longitudinally elongated, σ_{NS}=σ_{WE}/2; however, the value of proportion equal to 2 is not of a high precision. Other parameters of the model, such as the random amplitude of rain, as well as the number of rains per month, have almost no influence on the final result, i.e., the scatterplot (Fig. 3b).

The final numerical result is a rough estimation of the space-rains dimension to be 30–50 km WE/15–25 km NS, while the numbers have the sense of the Gaussian sigma. Converting to 95 % of precipitation, the figures four times larger are to be presented (with some additional rounding off being necessary, in order not to suggest spuriously high precision):

100–200 km WE, 50–100 km NS

It is worth reiterating that the result is obtained only on the basis of the data of monthly precipitation for a certain number of neighboring sites (16 of them, measured for 132 years).

7 Conclusion

Modeling the rains using the two-dimensional Gaussian function, it is possible to obtain a relation of correlation between sites to distance between sites very close to the observed one (as calculated from the data). The essential parameter of the Gaussian "rain" is its size in space (in km). We obtain the best

References

Alexandersson H (1986) A homogeneity test applied to precipitation data. J Climatol 6:661–675

Apputhurai P, Stephenson AG (2013) Spatiotemporal hierarchical modelling of extreme precipitation in Western Australia using anisotropic Gaussian random fields. Environ Ecol Stat 20(4):667–677

Cebulak E (1997) Variability of precipitation in selected regions of the Carpathian in the years 1951–1995. Geogr Pol 70:65–76

Cebulak E, Fasko P, Lapin M, Stastný P (2000) Extreme precipitation events in the Western Carpathians. Prace Geograficzne 108:117–124

Dutka AF, Hanson HH (1989) Fundamentals of data normalization, Addison-Wesley, 196

Kożuchowski K (2004) Skala, uwarunkowania i perspektywy współczesnych zmian klimatycznych w Polsce. Zakład Dynamiki Środowiska i Bioklimatologii UŁ, Łódź

Kożuchowski K, Żmudzka E (2003) 100-year series of areally averaged temperatures and precipitation totals in Poland. Studia Geograficzne 75, Acta Univ Wratislaviensis 2542:116–122

Niedźwiedź T (1999) Rainfall characteristics in southern Poland during the severe flooding event of July 1997. Stud Geomorphol Carpatho-Balc 33:5–25

Niedźwiedź T (2003) Extreme precipitation events on ten northern side of the Tatra Mountains. Geogr Pol 76(2):13–21

Niedźwiedź T, Obrębska-Starklowa B (1991) Klimat. In: Dynowska I, Maciejewski M (eds.) Dorzecze górnej Wisły. Wydawnictwo PWN, Warszawa-Kraków, pp 68–84

Niedźwiedź T, Twardosz R, Walanus A (2009) Long-term variability of precipitation series in east central Europe in relation to circulation patterns. Theor Appl Climatol 98:337–350

Štěpánek P (2006) AnClim—software for time series analysis (for Windows). Department of Geography, Faculty of Natural Sciences, Masaryk University, Brno. 1.47 MB. Available at http://www.climahom.eu/AnClim.html