

COMPARING DIFFERENT METHODS FOR COMPLETING THE MISSING DATA ON PRECIPITATION

R. Brázdil and T. Litschmann

Department of Geography, Faculty of Science, J. E. Purkyně University,
Kotlářská 2, Brno, Czechoslovakia

Received for publication: July 1983

SUMMARY

The paper deals with a comparison of several methods of completing the missing monthly sums of precipitation on the example of Moravian stations. The methods verified are the method of simple completion of the missing data, the method of linear regression of all values, the method of linear regression according to individual months, the quotient method and the method of multiple linear regression. From the analysis it follows that the most suitable method from all aspects is the method of multiple linear regression.

1. INTRODUCTION

The beginnings of an extensive network of rainfall measuring stations on the territory of Bohemia and Moravia date back deep into last century, when was organized the basic network of stations. Some of them have kept their observation activity up to the present. But in many cases they have not avoided interferences which for different reasons have resulted in affecting the homogeneity of the observation series (moving the stations, changes in observers, war events, etc.). Therefore stations with complete observation series in this country are relatively very few. This causes trouble in climatic studies in which it is necessary to start from long observation series (such as in studying the variations of the climate). In such cases it is necessary to homogenize the observation series by completing the missing data.

In connection with the calculation of territorial average precipitation the problem cropped up how to homogenize the precipitation series of Moravian stations with longer observation series. The objective of this contribution is to evaluate different methods applicable in completing the missing monthly sums of precipitation.

2. THE MATERIAL USED AND THE METHODS OF PROCESSION

Completing precipitation data, diurnal, monthly or annual sums, is a very complicated matter with respect to a great time and territorial variability of atmospheric precipitation. This fact can only partly be eliminated by a sufficiently dense network of rainfall measuring stations.

In Czechoslovak climatological literature these problems have been paid little attention. In completing the missing values usually the data of the nearest station are used or data following

from the maps of isohyets. In the textbook by M. Nosek (1972) only the method of quotients (see below) is mentioned together with a note saying that for completion also regressive dependences can be used.

In this paper the following methods used for completing the missing monthly sums of precipitation have been used:

- the method of simple completion (SC)
- the quotient method (QM)
- the method of linear regression of all values (LR)
- the method of linear regression in individual months (LM)
- the method of multiple linear regression (MR).

The method of simple completion (SC) is the simplest method consisting in completing the missing data by those from the nearest station, corresponding approximately by the height above sea level and exposition parameters.

The method of quotients (QM) is based on the determination of the quotient of long-term sum of precipitation at the station to be completed and the analogon-station in the respective month. The corresponding monthly sum at the analogon-station is multiplied by the quotient obtained.

The method of linear regression (LR) of all values uses the relation

$$y = a + bx \quad (1)$$

for completing the missing data. In the relation y represents the completed sum of precipitation, x is the precipitation sum at the analogon-station; a , b are parameters determined by the method of least squares from all ordered pairs of precipitation sums by means of relations quoted by J. Anděl (1978).

The method of linear regression according to individual months (LM) starts from the same theoretical basis as the preceding one, the difference consisting in the fact that it states 12 regression equations (one for each month of the year), while in the method of linear regression of all values only one equation was stated for all months.

The method of multiple regression (MR) is employed for completing the missing data on the basis of data from surrounding stations according to the relation

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n, \quad (2)$$

Table 1. A list of stations used (the first of the stations listed in the column stations-analogons was always used for the SC, QM, LR, and LM methods, all stations for the MR method; list of stations — see Tab. 1 in R. Brázdil, 1984)

Completed station	Stations-analogons
Hodonín	Dubňany-Jarohněvice, Prušánky, Břeclav
Horní Bečva	Hutisko-Solanec, Rožnov p. Radhoštěm, Valašská Bystřice
Horní Bečva	Rožnov p. R., Hutisko-Solanec, Valašská Bystřice
Hostýn	Bystřice p. Hostýnem, Holešov, Rajnochovice
Kladeruby n. Osl.	Hrotovice, Náměšť nad Oslavou, Ivančice
Kladeruby n. Osl.	Náměšť nad Oslavou, Hrotovice, Ivančice
Kněževes	Velké Meziříčí, Bohdalov, Skřínářov
Lipník n. Bečvou	Hranice, Přerov
Lysá hora-Mountain	Čeladná-Podolánky, Ostravice, Velké Karlovice-Javorníky
Lysá hora-Mountain	Ostravice, Čeladná-Podolánky, Velké Karlovice-Javorníky
Pavlovice u Přerova	Lipník n. Bečvou, Přerov, Hranice
Pavlovice u Přerova	Přerov, Lipník n. Bečvou, Hranice
Prušánky	Břeclav, Hodonín
Prušánky	Dubňany-Jarohněvice, Břeclav
Rajnochovice	Hošťálková, Kelč, Valašské Meziříčí
Rajnochovice	Kelč, Hošťálková, Valašské Meziříčí
Skřínářov	Bohdalov, Velké Meziříčí, Kněževes
Valtice	Břeclav, Mikulov, Prušánky
Valtice	Mikulov, Břeclav, Prušánky
Zlaté Hory	Zlaté Hory-Rejvíz, Jindřichov, Mikulovice

