ACTA UNIVERSITATIS LODZIENSIS
FOLIA GEOGRAPHICA PHYSICA 3, 1998

László Makra, László Stachó***

OPTIMAL SELECTION OF THE STATIONS DESCRIBING THE GLOBAL SEA LEVEL PRESSURE FIELD

OPTYMALIZACJA LOKALIZACJI STACJI POMIAROWYCH ROZKŁADU POLA CIŚNIENIA NA POZIOMIE MORZA

The subject of the study is to examine on a data base, how many measuring stations determine the global mean January sea level pressure field significantly with a given accuracy, how much information they have, furthermore what their spatial distribution is like. The paper contains a new method for optimal selection of measuring stations from an observing network.

INTRODUCTION

Practical methods to determine optimum or satisfactory density of station networks of climate elements have already been applied in a number of studies (e.g. Pokrovskij, Karoly 1988; Dévényi, Radnóti 1989). The latter paper gives a new method for selecting the station network, while the former points out that climate data set, exactly enough in some practical respects, may need relatively few stations. E.g., considering the monthly mean temperature fields in the Northern Hemisphere, in case of reducing the number of stations from 800 to 200–150, accuracy of estimations of mean monthly values decreases only slightly.

Selection of a few stations, with a given information from a dense stations network, can be performed by using statistical methods. If the main duty is to receive information with a given accuracy from the examined field, or in other words, to select stations, to approach the original field with a given accuracy, a linear regression model of the station network is used. In the present study the investigated field is the global mean sea level pressure field in January.

METHOD

The essence of the method is as follows. If X represents the total number of stations to be examined, then it is needed to choose such a minimal number of X* stations, data of which are able to reconstruct the spatial distribution of the field determined from the total X stations with a given level of average square error.

Suppose that $X = \{1, ..., M\}$ is mass of the stations and $T = \{1, ..., N\}$ is mass of observation times. Let $p_{m,t}$ represent the pressure value at the station number m, measured at the time t ($m \in X$, $t \in T$). Standardize these observation values and suppose that $s_{m,t}$ is the standardized value of $p_{m,t}$, namely

$$s_{m,t} = \frac{p_{m,t} - \bar{p}_m}{d_m},$$

where:

$$\bar{p}_m = \frac{1}{N} \sum_{t \in T} p_{m,t} \text{ and } \bar{d}_m^2 = \frac{1}{N-1} \sum_{t \in T} (p_{m,t} - \bar{p}_m)^2 \quad (1)$$

$(m = 1, ..., M;\ t = 1, ..., N)$

Suppose that S is the $(s_{m,t})$ $m \in X$, $t \in T$ matrix and $\sigma_1, ..., \sigma_M$ represent eigenvalues of the $S \cdot S^T$ square (with M columns and M rows) positive definit matrix, moreover, let F represent the matrix with M columns and M rows, columns of which are the eigenvectors of the $S \cdot S^T$ matrix, forming an ortonormal system. Namely,

$$S \cdot S^T = F \cdot \text{diag}(\sigma_1, ..., \sigma_M) \quad (2)$$

(Here $\text{diag}(\sigma_1, ..., \sigma_M)$ represents the $M * M$ matrix, in diagonal of which the numbers of $\sigma_1, ..., \sigma_M$ can be found successively, and all the other elements of the matrix are zero.)

Let $f_{m,n}$ mark the element number n of the row number m in the matrix F $(m, n = 1, ..., M)$.

In order to perform linear regression, it is practical to use those stations, for which the

$$\sum_{n \in X} f_{m,n}^2 \left(\frac{1}{\varepsilon} + \frac{1}{\sigma_n} \right)^{-1} \quad (m = 1, ..., M)\ ,\ (n = 1, ..., M) \quad (3)$$

quantities are the biggest. (That is to say the matrix σ_ε equals the matrix $\varepsilon \cdot I$ and the reconstruction of the original field, by the help of the linear regression method, is expected to be within ε relative error) Namely, this quantity can be considered as information originating from data of the station number m, if there are no *á priori* information about distribution of data of the stations.

Consequently, the time series of the station number m contains the η_m part of the total information, where

$$\eta_m = \frac{\sum_{n \in x} f_{m,n}^2 \left(\frac{I}{\varepsilon} + \frac{I}{\sigma_n}\right)^{-1}}{\sum_{k,n \in X} f_{km}^2 \left(\frac{I}{\varepsilon} + \frac{1}{\sigma_n}\right)^{-1}} \text{-th share} \tag{4}$$

Accordingly it is practical to arrange the stations in a descending order of the η_m quantities, in case of a given ε:

$$\eta m(1) \geqslant \eta m(2) \geqslant ... \geqslant \eta m(M) \tag{5}$$

and if it is intended to keep a η-th part of the total information, given in advance, then only those stations m(1), m(2), ..., m(K), ($K \leqslant M$) are needed to use, in case of which

$$\sum_{j=1}^{K} \eta_{m(j)} \geqslant \eta, \text{ but } \sum_{j=1}^{K-1} \eta_{m(j)} < \eta \tag{6}$$

Namely η share of information can be gained in this way, using the least number of stations. Of course, the η share of information can be gained from any kind of X^* share mass of X as well, in case of which

$$\sum_{m \in X^*} \eta_m \geqslant \eta \tag{7}$$

It can be seen that the closer η is to 1, the bigger share mass of X is needed to keep the η part of information.

Dependence from ε is a little bit more complicated. If for any m, $\eta_m \rightarrow \frac{1}{M}$, that is to say if it is required a very high relative punctuality, then the stations are practically equally important.

It may happen that by having the ε-parameter changed the order of the η_m shares of information will change, as well. Though the η_m shares of information are practically not shares of information in the sense of the theory of information, but shares of certain variances which measure undoubtedly the actual share of information, too. Bigger value of the

$$\sum_{m \in X} f_{mn}^2 \left(\frac{1}{\varepsilon} + \frac{1}{\sigma_n}\right)^{-1} \tag{8}$$

quantity may represent a bigger possibility to gain information both actually and according to the theory of the information. This also shows that in case of $\varepsilon \rightarrow \sigma$ (namely, if a very high punctuality is required), the information to be gained, approaches towards 0. This can be understood easily, because in order to reach „absolute punctuality" all the data are needed. (On the other hand, it is not possible to reach absolute punctuality from a finite set of data.)

DATA

The data base of the examination are mean January sea level pressure values considering the 30 years period between 1951–1980 from 247 measuring stations, all over the world (Fig. 1).

From 17 stations of the data base – because of their incomplete data – 23 years 1957–1980 are taken into consideration. Among the stations there are weather ships (stations number 39, 65, 66, 85, 91, 108), buoys (166, 184) and interpolated data (84, 150, 198, 203, 208, 215, 236, 237, 245, 247), as well. The pressure data series are taken from the volumes of *World Weather Records* as well as the monthly publications and sea level pressure maps of the *Monthly Climatic Data for the World* and *Die Witterung in Übersee* (Makra 1995).

Spatial distribution of the stations is the most dense in Europe but – for example – there are no data originated from China. Furthermore, the density of stations is little in Siberia, over the Pacific Ocean, as well as in the temperate and polar regions of the Southern Hemisphere.

RESULTS

First the total information of stations on the mean sea level pressure field in January was determined and a descending order was established. The station number 1. has the most information, the station number 2. is the second in this sense, and so on, until the station number 247. which has the least information on the field. The stations with their serial number show a characteristic spatial distribution (Fig. 1). Stations with low serial numbers can be found in the south-eastern part of the Pacific Ocean, the centre of South America, Inner Asia. At the same time the whole Europe, India and Northern Canada, furthermore Greenland show the least information. Stations with high serial numbers show the highest density while those with low serial numbers are furthest from each other.

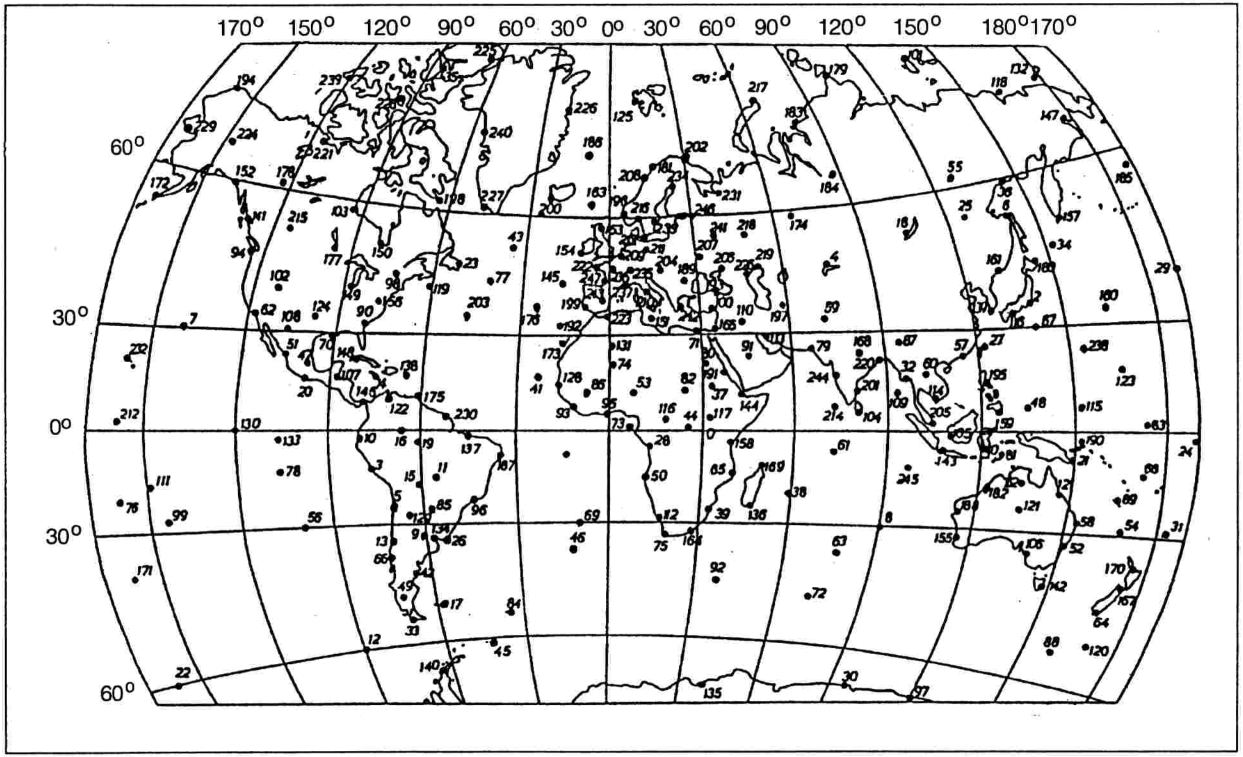

Fig. 1. Network of stations, describing the mean sea level pressure field with their order, January (e.g. station number 1. has the biggest information on the field, etc.). See Eq. 4. and 6. for definition

Rys. 1. Sieć stacji opisujących pole średniego ciśnienia na poziomie morza wg istotności dla stycznia (tj. stacja nr 1 przekazuje najwięcej informacji o polu itd.). Patrz równanie 4 i 6 dla definicji

Applying the method, mentioned above, to the mean January sea level pressure values, considering the 23 years period between 1958–1980 from 247 measuring stations, the number of stations was chosen to reconstruct the original mean January global sea level pressure field with a given accuracy (Tab. 1).

Table 1

Number of stations reconstructing the original pressure field with a given accuracy. See Eq. 4. and 6. for definition

Ilość stacji rekonstruujących oryginalne pole ciśnienia z zadaną dokładnością. Patrz równanie 4 i 6 dla definicji

Accuracy (%)	Number of stations
99	243
95	226
90	208
85	191
80	175
75	160
70	145
65	132
60	119
55	106
50	93

Stations having significant and not significant information (see section 2) show characteristic spatial distribution (Fig. 2–6).

On the basis of the maps it can be distinctly established that there is no role of European and Indian stations or the stations over the Polar Circle in reconstructing the original field with the accuracy of 80%. Reconstruction of the original field with 90 or 95% accuracy also shows little information mentioned the European stations (Fig. 2–6).

CONCLUSION

Summarizing our results, it can be established that stations of the three large regions – mentioned above – having characteristically little information, can be taken out of consideration during further examinations. After leaving stations with little information, the rest show certainly a more uniform distribution. Reconstruction of the original field with 90% accuracy can be

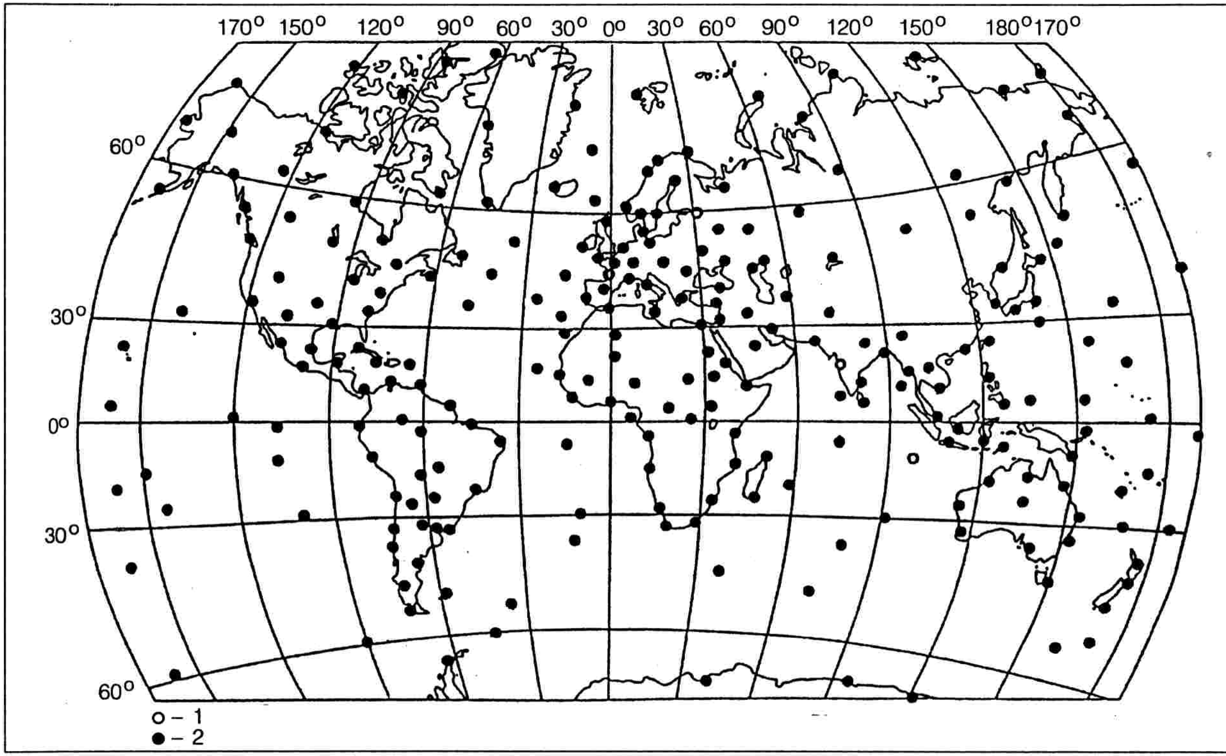

Fig. 2. Stations reconstructing the original pressure field with 99% accuracy, January
1 – values statistically significant, 2 – values statistically insignificant

Rys. 2. Stacje odtwarzające oryginalne pole ciśnienia z dokładnością 99%, styczeń
1 – wartości istotne statystycznie, 2 – wartości nieistotne statystycznie

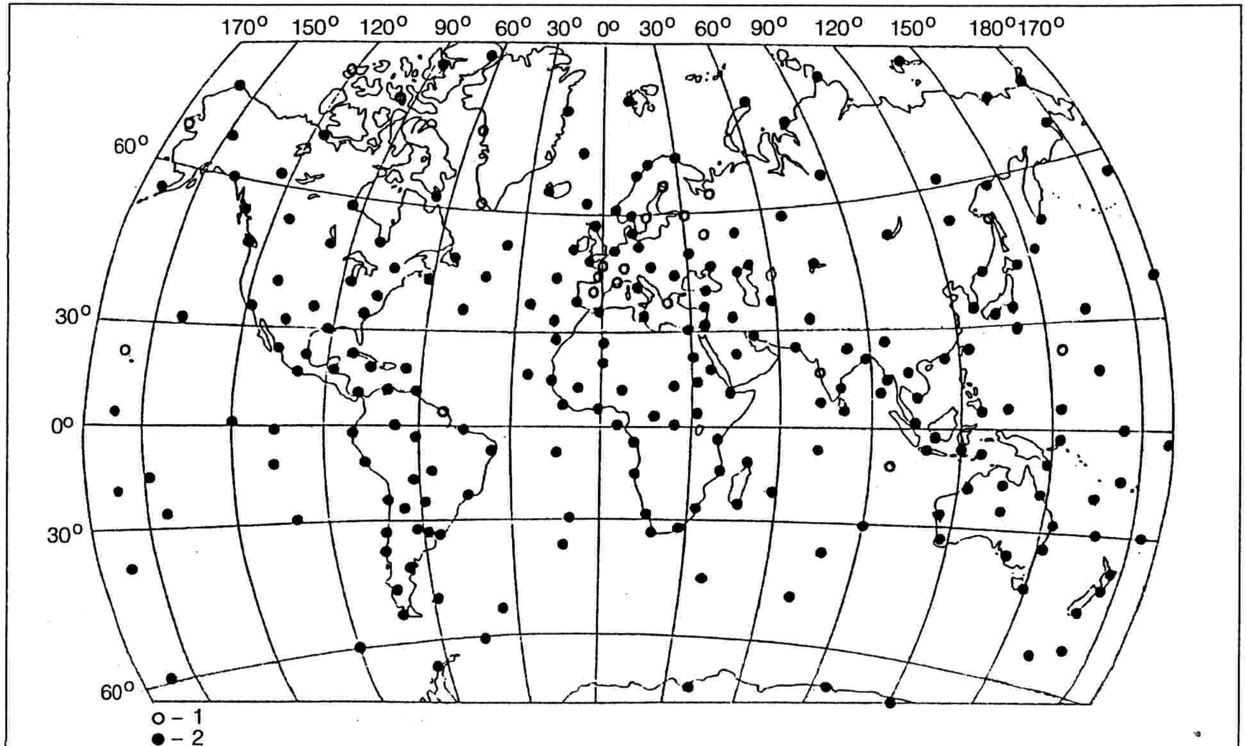

Fig. 3. Stations reconstructing the original pressure field with 95% accuracy, January. Explanations as in Fig. 2

Rys. 3. Stacje odtwarzające oryginalne pole ciśnienia z dokładnością 95%, styczeń. Objaśnienia jak na rys. 2

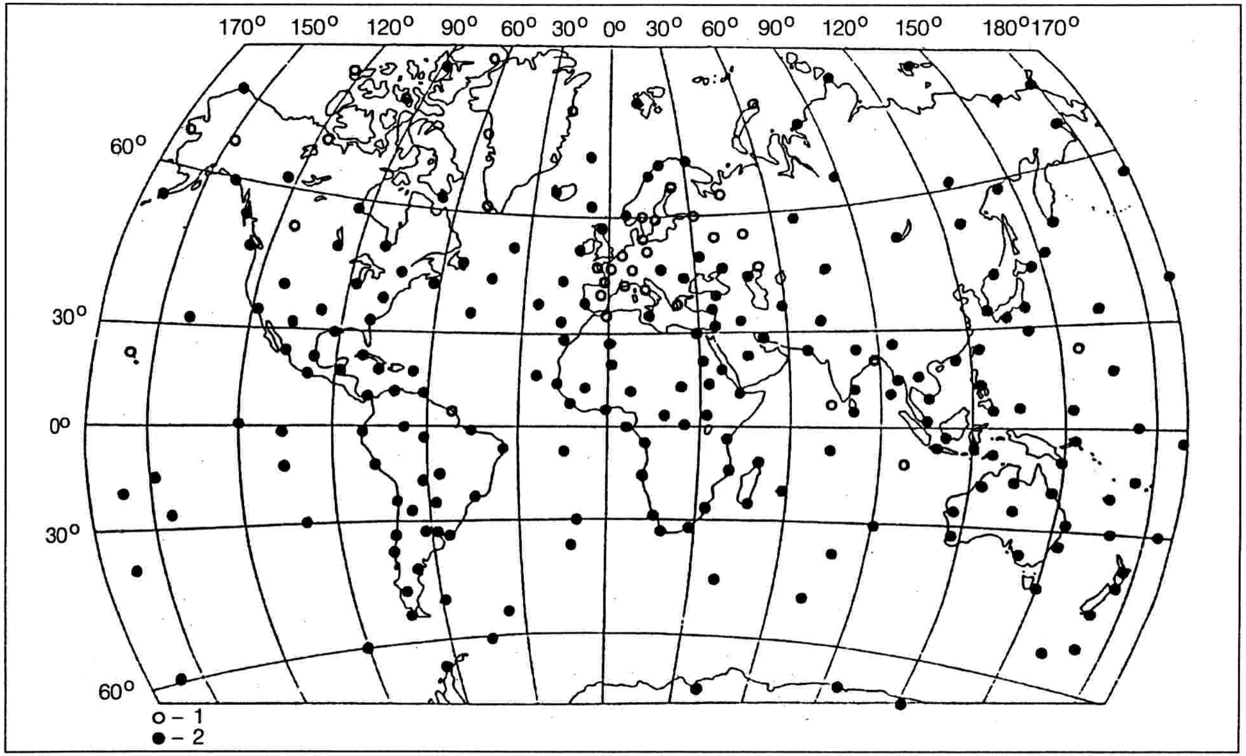

Fig. 4. Stations reconstructing the original pressure field with 90% accuracy, January. Explanations as in Fig. 2

Rys. 4. Stacje odtwarzające oryginalne pole ciśnienia z dokładnością 90%, styczeń. Objaśnienia jak na rys. 2

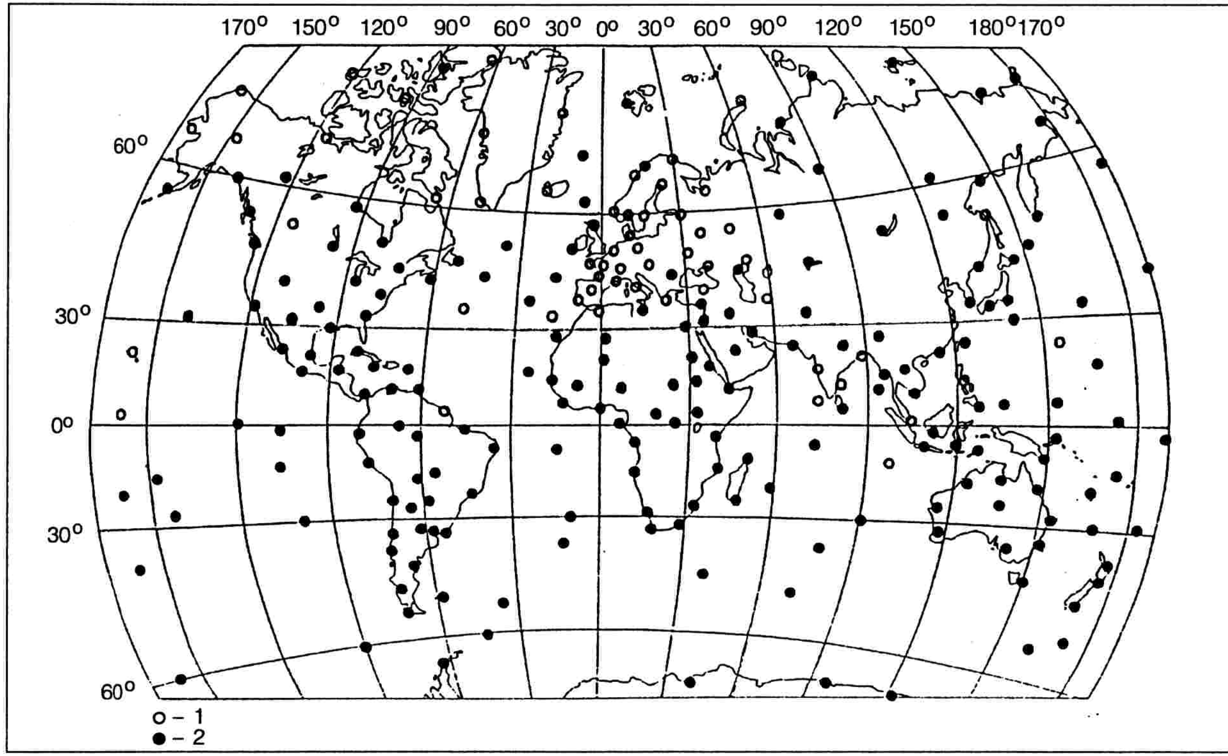

Fig. 5. Stations reconstructing the original pressure field with 85% accuracy, January. Explanations as in Fig. 2

Rys. 5. Stacje odtwarzające oryginalne pole ciśnienia z dokładnością 85%, styczeń. Objaśnienia jak na rys. 2

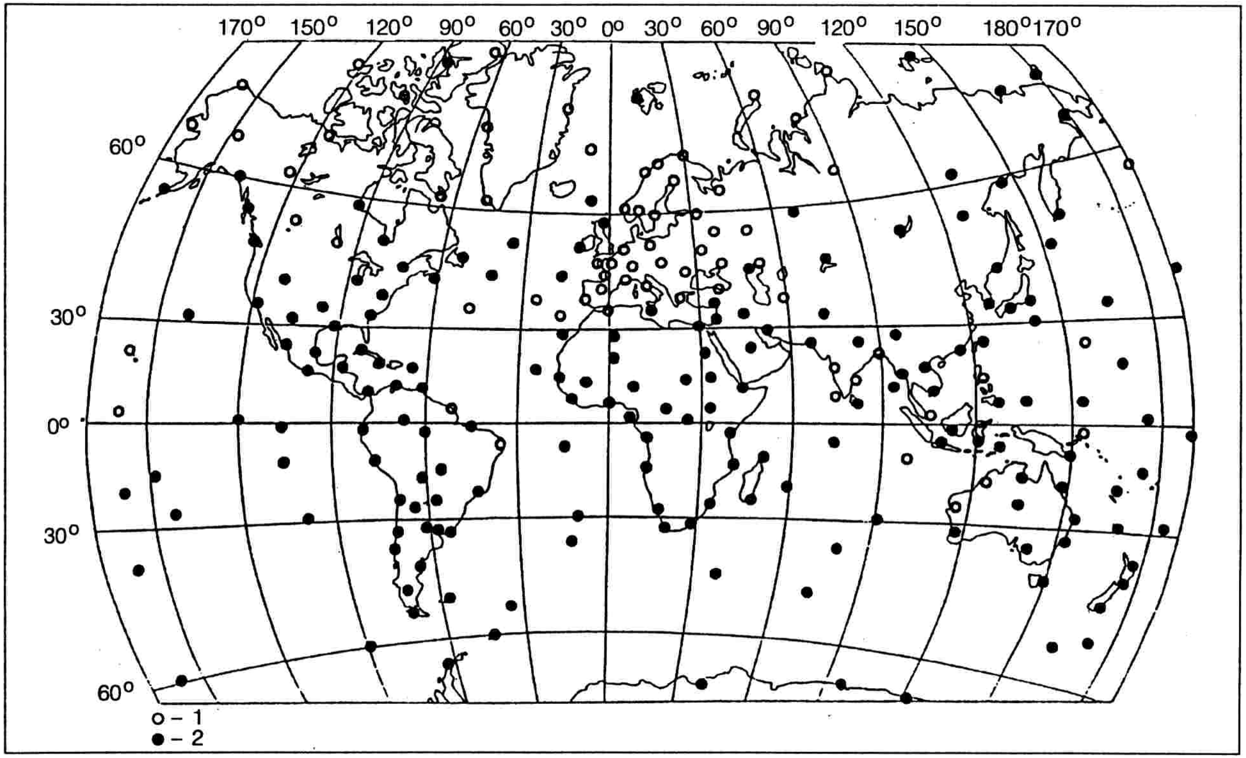

Fig. 6. Stations reconstructing the original pressure field with 80% accuracy, January. Explanations as in Fig. 2

Rys. 6. Stacje odtwarzające oryginalne pole ciśnienia z dokładnością 80%, styczeń. Objaśnienia jak na rys. 2

produced with leaving 20 stations in Europe and 17 more stations in the rest of the regions. This fact, considering spatial distribution of the stations, gives reason for leaving those European stations which have not significant information.

Practical importance of this kind of procedure (reducing data base for further analysis) is to use as little input data as possible without losing a significant share of information.

REFERENCES

Dévényi D., Radnóti G., 1989, *A New Measure for Characterizing Observing Networks*, Időjárás, p. 333–339

Makra L., 1995, *Examination of Statistical Characteristics of the Global Sea Level Pressure Field*, 124 pp. (manuscript)

Pokrovskij O. M., Karoly S. I., 1988, *Optimal Choice of Stations for Climatic Monitoring of Surface Temperature Over the Northern Hemisphere*, Meteorologiya i Gidrologiya, No 9, p. 60–71

* Department of Climatology, József Attila University
Szeged, Hungary
** Bolyai Institute, József Attila University
Szeged, Hungary

STRESZCZENIE

Przedmiotem badań jest ustalenie na podstawie danych, ile stacji pomiarowych określa średni rozkład pola ciśnienia na poziomie morza dla stycznia z żądaną dokładnością, jakimi danymi dysponują oraz jak wygląda ich przestrzenny rozkład. Artykuł zawiera opis nowej metody optymalnego wyboru stacji pomiarowych dla celów sieci obserwacyjnej.