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

L. MAKRA[1], *L. STACHÓ*[2]

[1] Department of Climatology, József Attila University,
Egyetem u. 2, 6722 Szeged, Hungary

[2] Bolyai Institute, József Attila University,
Aradi vértanúk tere 1, 6720 Szeged, Hungary

Received for publication: January 1997

SUMMARY

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.

KEY WORDS

January Sea Level Pressure Field - Significant Information - Reconstruction of the Original Field

INTRODUCTION

Practical methods determining station networks of climate elements have already been analysed 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 determining climate data of the stations, exactly enough in practical respects, needs relatively few stations. In this way, 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 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 of given accuracy from the examined field, with other words to select stations, determining significantly the original field with a given accuracy, a linear regression model of the station network is used. The above mentioned 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 of stations to be examined, then it is needed to choose such a minimal number of X^* stations (X^* is taken from X), data of which reconstruct the spatial distribution of the field determined from the total of X stations with a given level of average square error.

$$s_{m,t} = \frac{p_{m,t} - \bar{p}_m}{\bar{d}_m}, \quad \text{where} \quad \bar{p}_m = \frac{1}{N}\sum_{t \in T} p_{m,t} \quad \text{and}$$

$$\bar{d}_m^2 = \frac{1}{N-1}\sum_{t \in T}\left(p_{m,t} - \bar{p}_m\right)^2 \qquad (m = 1, \ldots, M; \; t = 1, \ldots, N). \qquad (1)$$

Suppose that S is the $(s_{m,t})_{m|X|,t|I|}$ matrix and $s_1, \ldots, s_M$ represent eigenvalues of the $S{\cdot}S^T$ square (with M columnes and M rowes) positive definit matrix, moreover represent F the matrix with M columnes and M rowes, columnes of which are eigenvectors of the $S{\cdot}S^T$ matrix, forming ortonormal system. Namely

$$S{\cdot}S^T = F{\cdot}diag(s_1, \ldots, s_M). \qquad (2)$$

Here $diag(s_1, \ldots, s_M)$ represents the $M * M$ matrix, in diagonal of which the numbers of s_1, ..., s_M can be found successively, and all the other elements of the matrix are zero.
Mark $f_{m,n}$ the element number n of the row number m in the matrix $F(m, n = 1, \ldots, M)$.
In order to perform linear regression it is practical to use those stations, to which the

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

quantities are the biggest. 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 (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.

Consequently the time series of the station number m contains

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

of the total information. Accordingly it is practical to arrange the stations in order of size of the η_m quantities, in case of given ε:

$$\eta_{m(1)} > \eta_{m(2)} > \ldots > \eta_{m(M)} \tag{5}$$

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

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

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

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

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

Dependence from e is a little bit more complicated. If at any case of m, $\eta_m \to \frac{1}{M}$

that is to say if it is required a very high relative punctuality, then the stations are practically important equally.

It may happen that by having the e-parameter changed the order of the h_m shares of information will change, as well. Though the h_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 the bigger possibility of gaining information both actually and in the sense of the theory of the information, as well. This also shows that in case of $\varepsilon \to \sigma$ (namely, if it is required a very high punctuality), the information which can be gained, goes towards 0, too. This can be understood easily, because in order to reach "absolute punctuality" all the data are needed, on the other hand it can not be gained information from the data in order to reach absolute punctuality.

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 - their mean January sea level pressure values considering the 23 years period between 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, 1987).

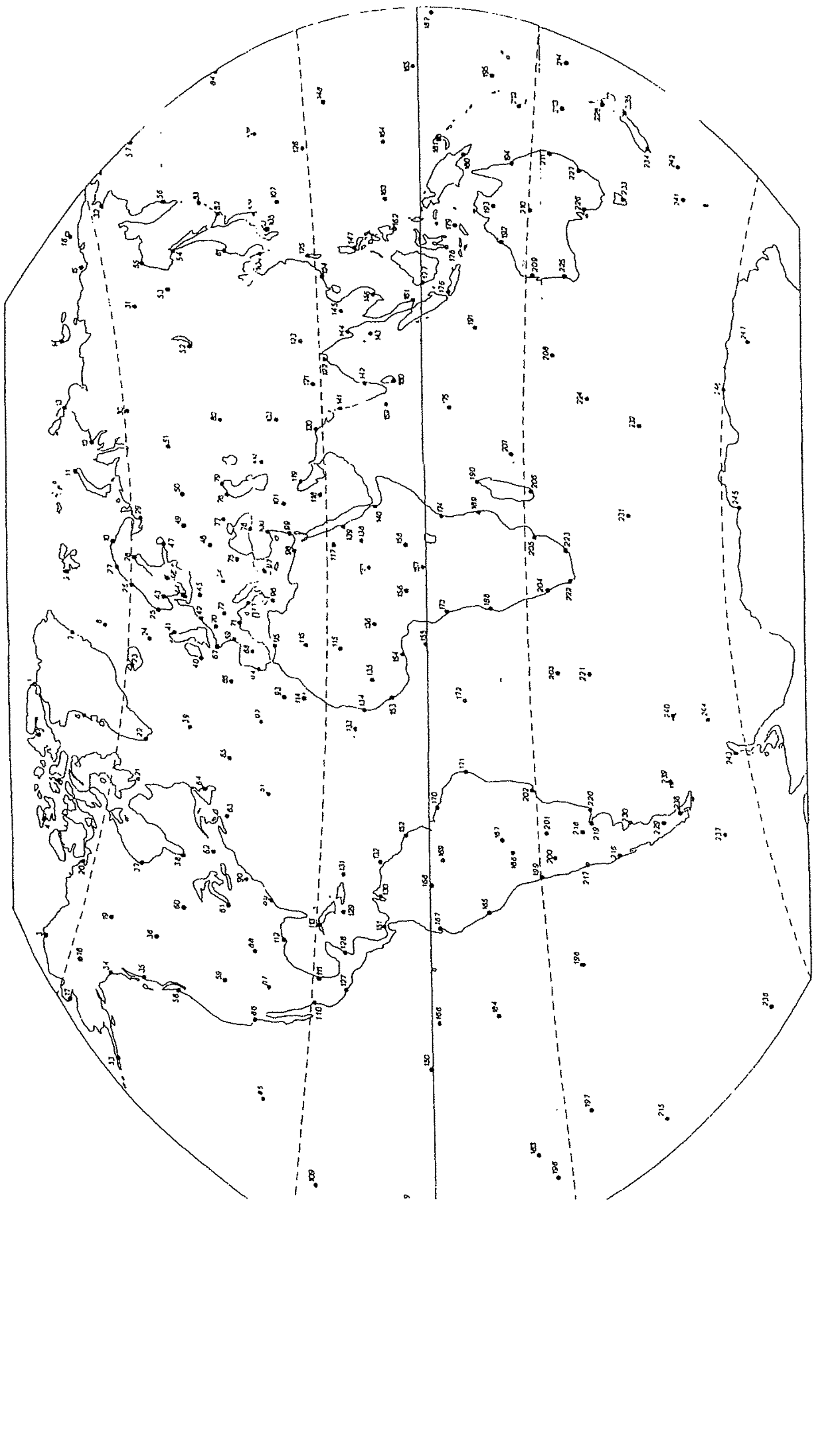
Spatial distribution of the stations is the most dense in Europe but - for example - there are no any data originating from China, furthermore the density of stations is little in Siberia, over the Pacific Ocean as well as in the temperate and polar regions of the Southern Hemisphere.

RESULTS

The base of the examination is the global mean sea level pressure field in January.

It was determined the total information of stations on the mean sea level pressure field in January and an order was established. The station No.1. has the most information, ..., and so on, ..., the station No. 247 has the least information on the field. The stations with their serial number show a characteristic spatial distribution (Fig. 2). 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.

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 all over the world, number of stations were chosen of the all, data of which reconstruct the original mean January global sea level pressure field with a given accuracy (Table 1).



1. Stations

Stations having significant and not significant information (see method, equations 6) show characteristic spatial distribution (Figs. 3-7).

On the basis of the maps it can be distinctly established that there is no role of European and Indian stations as well as stations over 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 of the European stations (Figs. 3-7).

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

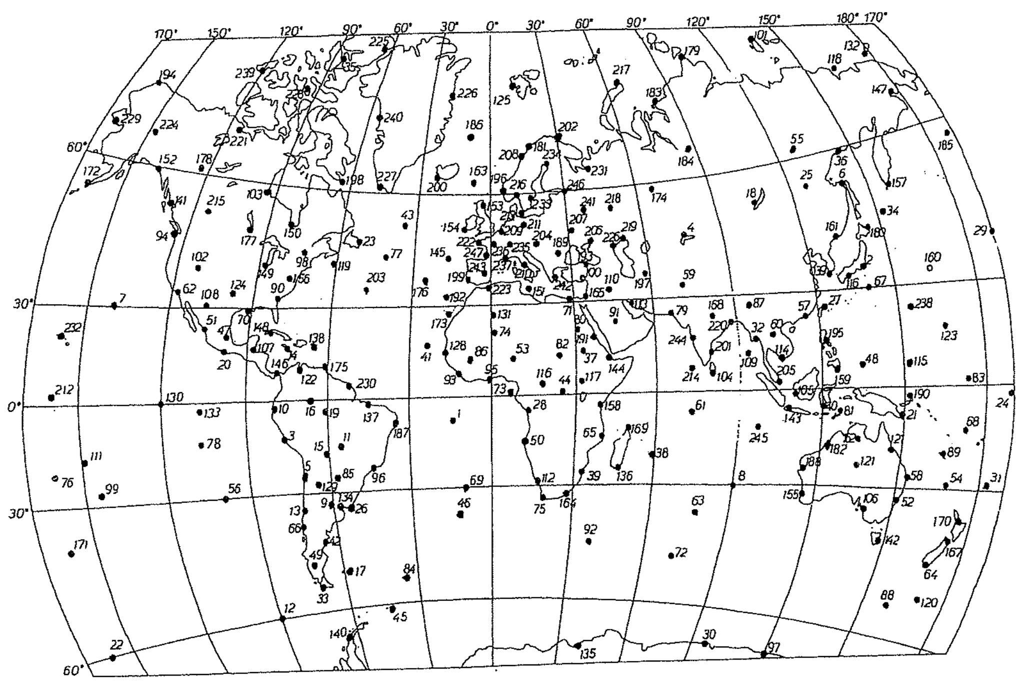


Fig. 2. Stations describing the mean sea level pressure field with their order, January (e.g.

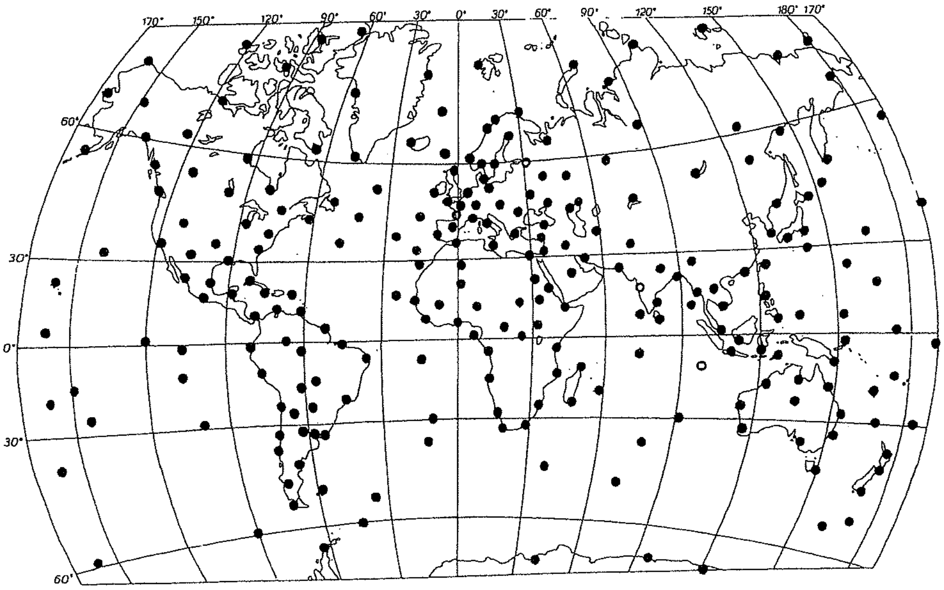
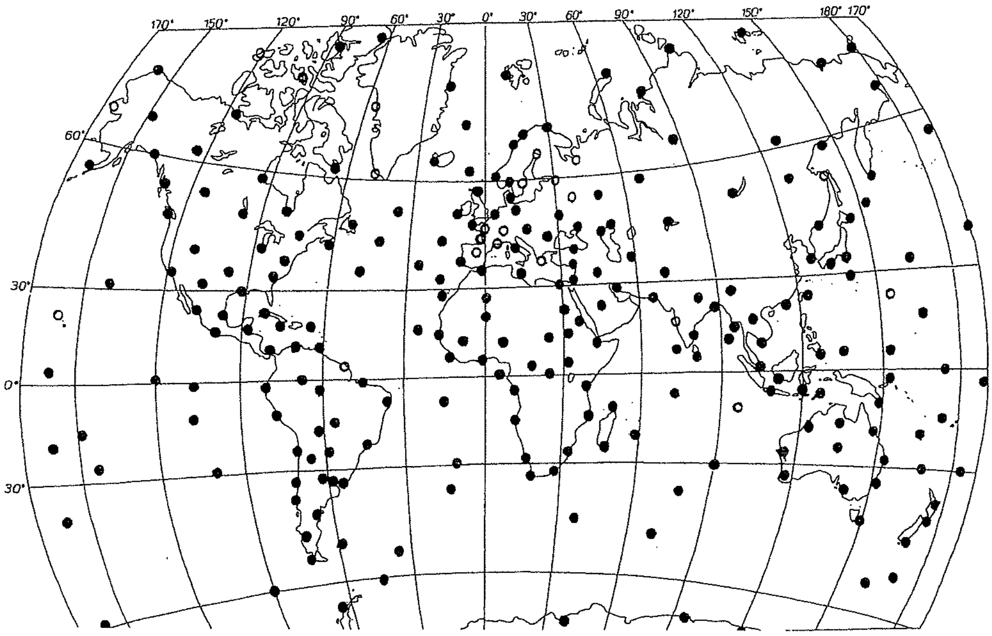


Fig. 3. Stations reconstructing the original pressure field with 99% accuracy, January
Legend to Figs 3-7: • significant stations, ◦ not significant stations



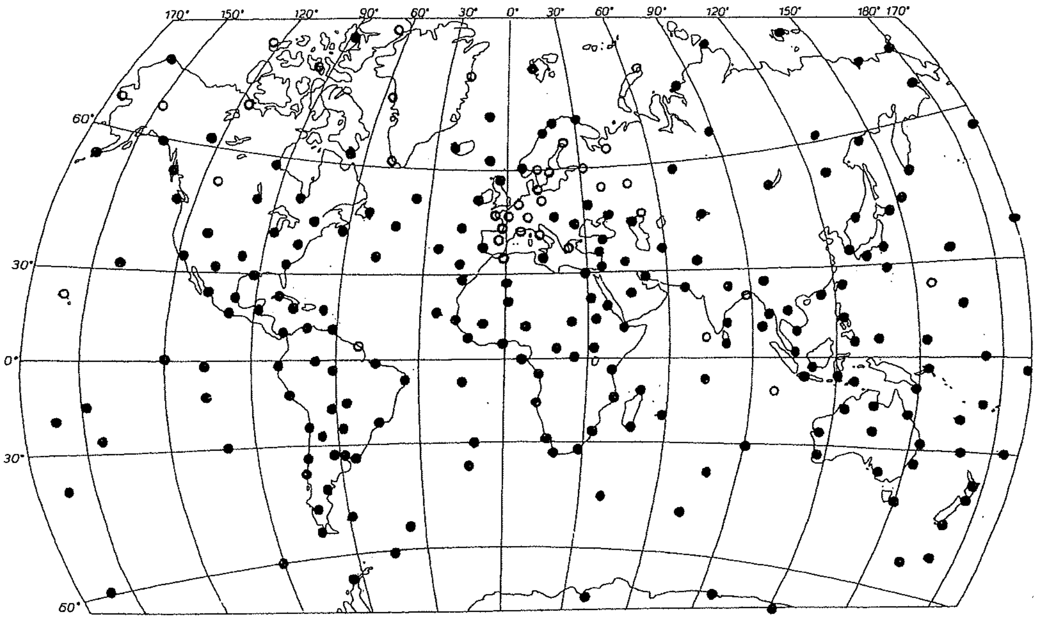


Fig. 5. Stations reconstructing the original pressure field with 90% accuracy, January

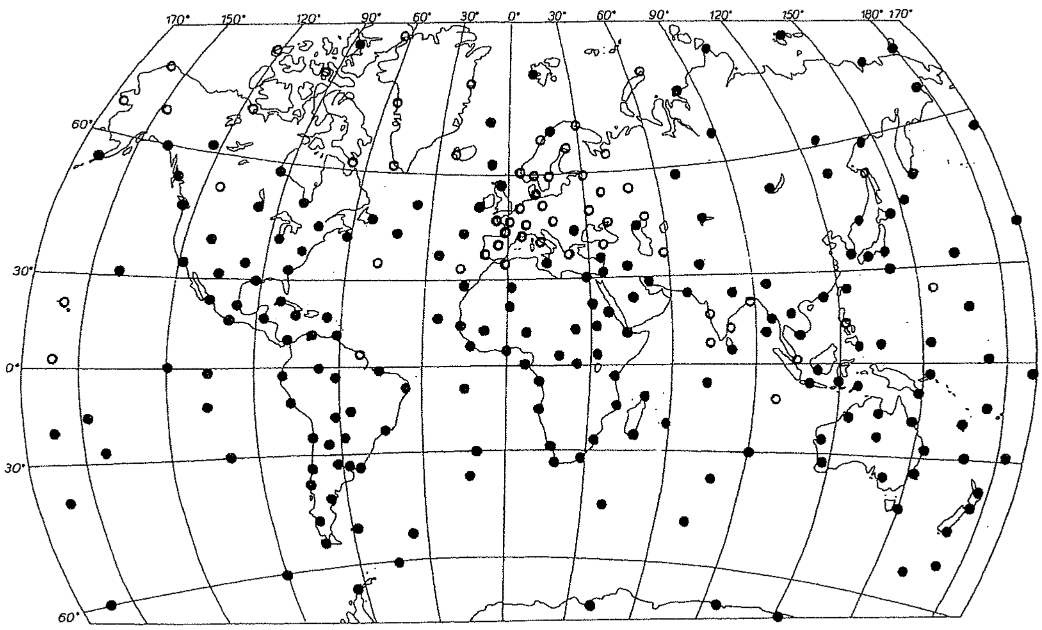


Fig. 6. Stations reconstructing the original pressure field with 85% accuracy, January

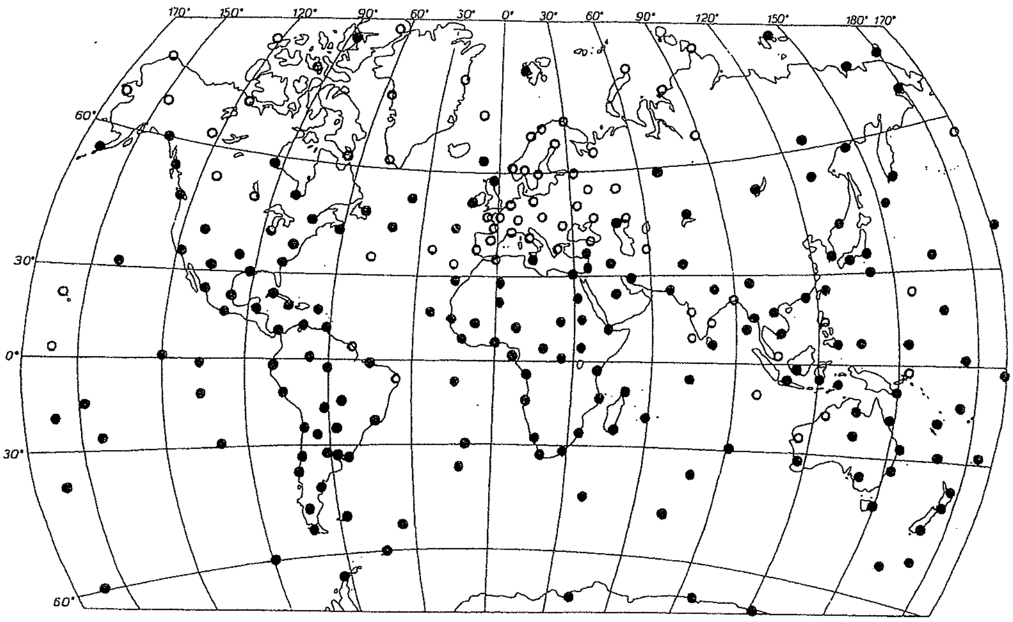


Fig. 7. Stations reconstructing the original pressure field with 80% accuracy, January

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

Table 1. Number of stations reconstructing the original pressure field with a given accuracy

REFERENCES

DÉVÉNYI, D., RADNÓTI, G. (1989): A new measure for characterizing observing networks. Időjárás, 93, 6, 333-339.

MAKRA, L. (1995): Examination of statistical characteristics of the global sea level pressure field. PhD Dissertation, Szeged, 176 pp.

POKROVSKIJ, O. M., KAROLY, S. I. (1988): Ob optimalnom otb[illegible] tantsiiy dlya klimaticheskogo monitoringa prisemnoy temperatury v Sever[illegible]n Polusharii. Meteorologiya i Gidrologiya, 9, 60-71.