A METHOD OF ANALYSING MACROSYNOPTIC TYPES USING ANALOGY INDICES

by

J. Bartholy (Mrs. Szlachányi) and *O. Gulyás*

A makroszinoptikus típusok analízise analógia index segítségével. A tanulmány azt elemzi, hogy a szinoptikus módszerekkel makroszinoptikus típusba sorolt nyomásmezők milyen kapcsolatban állanak a különböző típusokhoz tartozó átlagos nyomásmezőkkel. A vizsgálatban az analóg indexet és az objektumok összetartozásának számszerű jellemzésére szolgáló mérőszámot alkalmazzák.

In the paper ware examined the analogy between individual pressure fields classified into macrosynoptic types by synoptic methods and the mean pressure fields of the different types. By examinations was used the analogy index and the measure for the numerical characterization of the relatedness of the objects.

The relation between the macrosynoptic types used in meteorology and the analogy indices was discussed in our earlier papers. The mathematical definition of the analogy index was formulated, and some problems of principle of the forecasting by using analogies (including the accuracy of forecasts) were analysed in [1] and [2]. The typifyng of meteorological objects with the aid of analogy indexes was the subject of [3]. We considered, among others, the degree of analogy that can be formed between the mean fields of different macrosynoptic types, in other words, how "far" these mean fields are situated from each other. In continuation of these studies, we shall examine in the present paper the analogy between the individual fields classified into macrosynoptic types by synoptic methods and the mean fields of the different types. In the light of the results, conclusions may be drawn regarding the practical uncertainties of the synoptic classifications and about the practicability of typifying by computer methods.

The description of the complicated meteorological (synoptic) situations necessitates the use of a great number of parameters. With a view to simplifying the model, the situations are, therefore, grouped into classes (groups, clusters), corresponding to the above mentioned macrosynoptic types. Each class is denoted by an appropriate codefigure called macrosynoptic code. The characteristics of the various classes are chosen in such a manner that all the situations belonging to one given class (i. e. having the same code) have similar properties, while any two situations belonging to different classes are distinquishable on the basis of their different properties.

Many macrosynoptic code systems have been developed, which differ from each other in the area covered, in the atmospheric elements selected for the meteorological model and in the purpose of application. (It is evident that there can be no "universal" typifying system meeting the most general conditions.) The most widely used classification in Europe is the *Hess—Brezowsky* type-system, comprising 29 macrosynoptic types [4]. Most countries use their own special macrosynoptic code systems, and many countries apply more than one such system. The macrosynoptic classification system developed by *Péczely* for use in Hungary contains 13 types of situation ς]]. The *Péczely*-codes have been determined and are published in [6] for each day from 1877 till practically today. It may be mentioned that other specialpurpose code systems have also been elaborated for Hungary. For example, *B. Bőjti* established a system applicable in meteorological research relating to the Lake Balaton region; *E. Jakus (Mrs. I. Bodolai)* elaborated another code system for the purpose of precipitation forecasting.

For the classification of the meteorological situations by computer, it is necessary to prepare an adequate, unequivocally defined, correct mathematical model. This model, necessary to the formation of types, should include the following:

(a) A clear description of the system of parameters (atmospheric elements, co-ordinates, dimensions, etc.) used for characterizing the object considered;

(b) The analogy index, which is the measure used for the comparison of the objects;

(c) The definition of the measure used for the numerical characterization of the relatedness of the objects.

(a) The method of description of the objects to be classified

In the investigations we used grid-point values of the 500 mb geopotential field (AT_{500}). The area covered and the grid network are shown in *Fig. 1*. The analysis refers to the 4018-day period from 1 January 1956 to 31 December 1966.

Let

$$\underline{\xi} = (\xi_1, \xi_2, \ldots, \xi_i, \ldots, \xi_{80})$$

denote the AT_{500} values, interpreted as probability variable at the grid points, and Θ the macrosynoptic code for the same time. Accordingly, the value (or realization) of the probability vector variable for one day (the j-th day) of the time period examined will be denoted by

$$\underline{X}_j = (X_{j1}, X_{j2}, \ldots, X_{ji}, \ldots, X_{j80})$$

and the *Péczely* macrosynoptic code for the same day by ϑ_j. The AT_{500} grid point values for the ϑ_j-th day are determined by interpolation from the tables of the Deutscher Wetterdienst and are stored on a special magnetic tape, while the macrosynoptic codes relating to the same time are contained in [6]. The macrosynoptic types have been coded in figures as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13
mCc	AB	CMc	mCw	Ae	CMw	zC	AW	As	An	AF	A	C

The full observational material is thus comprised by (1) the $\underline{\underline{X}}$ matrix formed by the realizations of $\underline{\xi}$, and (2) the $\underline{\vartheta}$ vector representing the realizations of Θ:

$$\underline{\underline{X}} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nN} \end{pmatrix} \qquad \underline{\vartheta} = \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \\ \vdots \\ \vartheta_n \end{pmatrix}$$

$$N = 80, \quad n = 4018$$

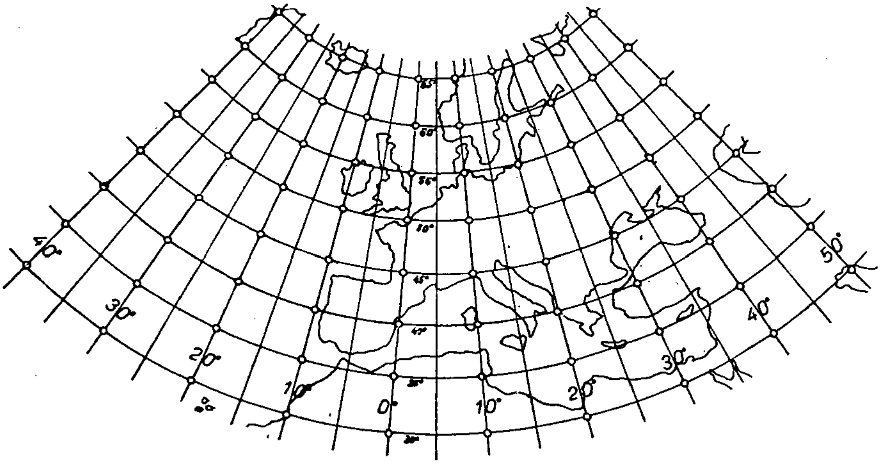

Fig. 1. The area covered and the grid system used

(b) The analogy index

The objects to be classified may be described by points of the N-dimensional space. In order to ensure that, when forming the classes, the points representing similar objects belong to the same class and the points corresponding to objects differing from the earlier ones belong to other classes, it is necessary to define some numerical criterion for the measurement of similarity. Analogy means relationship between two objects considered, it is also called their similarity or their association capability relating to each other. The analogy index is the measure expressing the similarity of two objects in numerical form. This index is an important element of the classification procedure, since it represents the condition of classifying two objects into the same class or into different classes.

Let us consider now, in general, two vectors in the N-dimensional field:

$$\underline{x} = (x_1, x_2, \ldots, x_N),$$

$$\underline{y} = (y_1, y_2, \ldots, y_N).$$

Any $\varrho(\underline{x}, \underline{y})$ function interpretered for the $\underline{x}, \underline{y}$ pair of vectors is called analogy index if it satisfies the following cosditions:

1. $\varrho(\underline{x}, \underline{x}) = 1$ for any $\underline{x}$,
2. $\varrho(\underline{x}, \underline{y}) = \varrho(\underline{y}, \underline{x})$ for any $\underline{x}, \underline{y}$.
3. The $\Delta(\underline{x}, \underline{y}) = 1 - \varrho(\underline{x}, \underline{y})$

function satisfies the so-called triangle inequality, i.e. for any $\underline{x}, \underline{y}, \underline{z}$

$$\varrho(\underline{x}, \underline{y}) \geqq \varrho(\underline{x}, \underline{z}) + \varrho(\underline{z}, \underline{y}) - 1.$$

From these conditions it follows that the analogy index meets the following inequality:

$$0 \leqq \varrho(\underline{x}, \underline{y}) \leqq 1.$$

The analogy of two objects $\underline{x}, \underline{y}$ is defined as the values of their analogy index $\varrho(\underline{x}, \underline{y})$

(In some cases the fulfilment of the triangle inequality is dispensed with.)

In our investigations we used the analogy index derived from the *Euclidean* metric:

$$\varrho(\underline{x}, \underline{y}) = \varrho_E(\underline{x}, \underline{y}) = \frac{1}{1+\sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}}.$$

We performed the calculations also for several other analogy indices, manuely for the *Bagrovean* index and a number of its modified versions, but the results obtained for $\varrho_E(\underline{x}, \underline{y})$ only are discussed here.

(c) The criterion of class formation

For checking the optimization of the grouping, we applied the procedure known as centroid method, which is used in from recognition and cluster analysis.

To this end, the mean AT_{500} fields of the days having the same macrosynoptic code have been calculated on the basis of the data the 4018 days considered. The mean field of the days with the *i*-th macrosynoptic type is denoted by

$$\underline{\eta}^i = (\eta_1^i, \eta_2^i, \ldots, \eta_N^i) \qquad i = 1, 2, \ldots, L$$

The $\underline{\eta}^1, \underline{\eta}^2, \ldots, \underline{\eta}^L$ type mean fields may also be represented in the 80-dimensiona field corresponding to the number of the grid points. Since the *Péczely* macrosynopitc codes were used in our study, L=13.

We examined, in particular, to what extent the L=13 mean fields resembled the corresponding pictures which served as a basis for the definition of the types and which are published in [5]. Our analysis showed that there were no significant differences, and the averaging procedure even led to the statistical levelling of certain errors. One of the mean fields is shown in *Fig. 2,* for the others reference is made to [8].

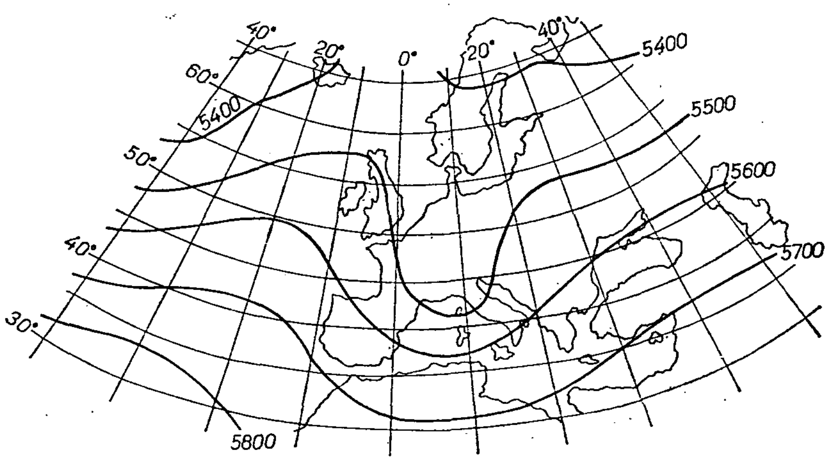

Fig. 2. The mean field of the CMw types (No. 6 in the Péczely classification)

These type means may be interpreted as the fields which are most characteristic of their respective types, etalons of the classes represented by them. It may, therefore be expected that if the AT_{500} fields of the individual days are represented by points of, the 80-dimensional field, then these points will be situated, in the form of clouds of points, around the points corresponding to the type mean fields, and that these clouds will be far from each other. The picture will thus be similar to that of the galaxies in the space. Such a distribution of prints, considered as optimal, is represented in *Fig. 3.*

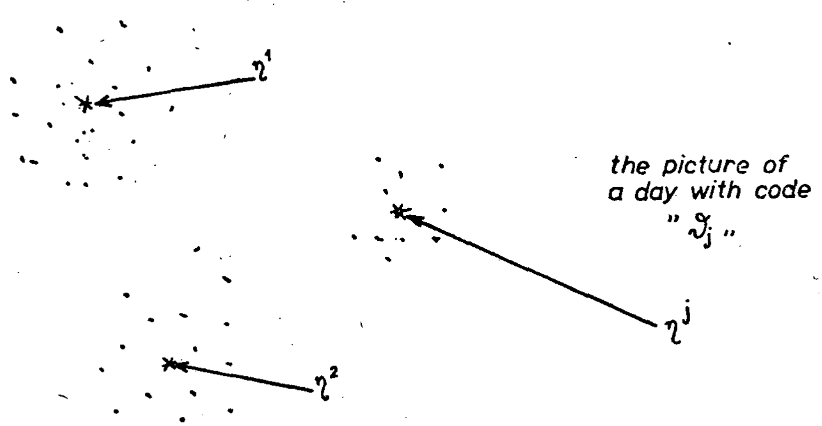

Fig. 3. Two-dimensional representation of the expected distribution of the mean fields and of the individual days

In reality the situation will not, of course, be so nice. The expected ideal distribution may be distored by a number of error sources, e.g. errors made in the course of the classifications, the selection of the analogy index, the fact that inour study we considered only the AT_{500} field, etc.

In the investigations described below, the field data of 465 days were used. The *Euclidean* analogy indexes were determined for each individual day ($\underline{x}_j=(x_{j1}, x_{j2}, \ldots \ldots x_{jN})$, $j=1, 2,\ldots,465$) and for each type mean (etalon) field η^i, $i=i=1,2,\ldots 13$. As shown above, the analogy between the mean of the i-th type and the day with index j is

$$\varrho_E(\eta^i, \underline{x}_j)=\frac{L}{1+\sqrt{\sum_{k=i}^{80} \eta_k^i-x_{jk})^2}}$$

For each $\varrho_E(\eta^i \underline{x}_j)$ of the days considered, the values ($i=1,2,\ldots,13$) were then arranged in order of magnitude, and the corresponding code numbers were recorded. In other words, the first place in the resulting series was occupied by the code number of the type showing the greatest analogy with the field of the given day, while the last figure in the series was the code number of the type with the least analogy. This classification, obtained by the application of the *Euclidean* metric, was then compared with the ϑ_j synoptic type of the day in question. For each of the 465 days, we have calculated and recorded the serial number (place) of the *Péczely* code of the day as determined by the synoptician, in the above series of code numbers.

The results are summarized in *Table 1*. By way of explanation the meaning of the line marked by an asteriste in *Table 1* includes the following:

According to the classification made by the synoptician, the *CMw* type (No. 6) occurred 16 times out of the 465 examined cases; in our classification (using *Euclidean* analogy indexes) this same type ranked 4 times first and 5 times second (out of the 16 cases).

Table 1

Frequency of occurence of the serial numbers of classification of the Péczely codes (based on a sample of 465 days)

Péczely code	Serial number of classification: 1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	3	—	3	—	—	6	1	3	—	—	1	—
2	6	6	2	6	1	4	3	1	1	2	1	3	1
3	8	1	—	—	—	—	—	—	—	4	1	—	—
4	7	13	4	3	2	—	—	—	1	3	—	—	—
5	4	12	4	8	3	10	1	3	1	—	1	1	—
6	4	5	4	1	—	—	1	—	—	1	—	—	—
7	2	3	6	4	3	2	—	—	2	—	—	—	—
8	25	10	8	4	2	5	2	3	2	2	10	4	2
9	27	1	—	—	—	—	1	—	—	1	1	1	—
10	—	8	2	9	8	4	4	5	3	3	1	—	—
11	18	12	6	2	1	4	—	1	1	—	—	1	—
12	11	7	11	10	9	11	3	3	1	2	—	—	—
13	—	—	1	2	1	—	1	—	—	—	—	—	—

The analysis of the table shows that:

(a) The code determined by the synoptician (here in after called "proper code") is in surprisingly good agreement with the code determined on the basis of the AT_{500} field with the aid of the analogy derived from the *Euclidean* metric. (This latter code might be referred to as "machine code"). The index of the most analogous mean field is identical with the proper code in 25 per cent of the cases.

(b) In certain cases, for example for the types Nos. 4, 6, 9, 11, the ϑ_j proper code is in particularly conspiciuous agreement with the most analogous mean field.

(c) There is, however, a strikingly great disagreement in the case of codes Nos. 8 and 10. There may be a number of reasons for this, for example, these types may not be well distinguishable in the AT_{500} field, the analogy index we have chosen may not be suitable for distinguishing them, and the ϑ_j "proper code" may also be erroneous.

A detailed discussion of the data in *Table 1* for each individual type is not given here. The information contained in the Table is, however, represented in condensed form in *Fig. 4*. The graph shows the frequency of occurrence of the *Péczely* ("proper") code of the fields in the lst, 2nd, 3rd,...13th places.

The figure indicates, for example, that during the 465 day period there were 81 days for which, among the codes ranked on the basis of the *Euclidean* analogy indexes, the "proper code" of the day as determined by the synopticians occupied the 2nd place.

Fig. 4 shows also that in 25% of the cases the proper code of the day appears on the lst place, which means that in these cases the *Euclidean* analogy classification fully coincided with the "synoptic" classification.

It is in 42 per cent of the total number of days that the synoptic "proper code" of the day stands on the lst or the 2nd place.

It is only 20 per cent of the cases of the 465 day sample that the "proper code" was ranked by the program to a place with serial number 6 or greater. This result may be regarded as very favourable.

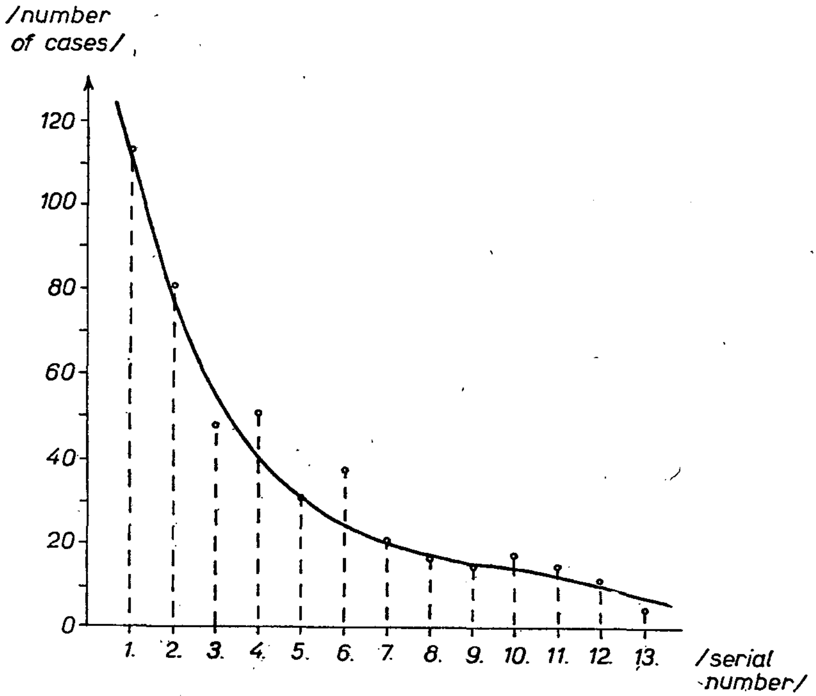

Fig. 4. Frequency of occurence according to the serial number of the proper code (in the classification using Euclidean analogy indexes)

References

[1] *Kaba, M.—Faragó, T.—Gulyás, O.:* Az analógia elvén alapuló prognosztikai módszerek matematikai modellje. (A mathematical model of prognostic methods based on the principle of analogy.) Időjárás *79* (1975), 3. 166—176.

[2] *Gulyás, O.:* Az analógia fogalma és felhasználása típusok képzésére I. (The notion of analogy and its use for type formation I.) Időjárás *81* (1977), 1. 11—18.

[3] *Gulyás, O.—Bartholy, J.—Légrády, G.:* Az analógia fogalma és felhasználása típusok képzésére II. (The notion of analogy and its use for type formation II.) Időjárás *81* (1977), 6.

[4] *Hess, P.—Brezowsky, H.:* Katalog der Grosswetterlagen Europas. — Berichte des Deutschen Wetterdienstes in der US Zone, Nr. 33. Bad Kissingen 1952.

[5] *Péczely, G.:* Grosswetterlagen in Ungarn. — Kleinere Veröffentlichungen der Zentralanstalt für Meteorologie, Nr. 30. Budapest 1957.

[6] *Péczely, G.:* Magyarország makroszinoptikus helyzeteinek éghajlati jellemzése. (Climatological description of macrosynoptic situations for Hungary.) OMSZ Kisebb Kiadványai 32. Budapest 1961.

[7] *Rákóczi, F.:* Meteorológiai folyamatok vizsgálata analógia index segítségével. (Study of meteorological processes with the aid of analogy index.) Beszámoló az 1974. évben végzett kutatásokról, OMSZ, Budapest 1974.

[8] *Bartholy, J.:* Egyetemi doktori értekezes. (Thesis for university doctorate.) OMSZ, 1977. (Manuscript)