Objective weather classification for environmental applications

Janos Mika[1], Andras Razsi [1], Agnieszka Wypych[2], Zbigniew Ustrnul[2]
[1]Eszterhazy Karoly College, Eger, Hungary, [2]Jagiellonian University, Krakow, Poland

Abstract. An objective statistical classification of daily weather is presented for 8 stations of Poland and 5 stations of Hungary for 30 years periods (1966-1995 and 1961-1990, respectively. Initially, eight weather elements were pre-selected and to four variables, according to factor analysis, based on strong correlation between the elements. The further cluster analysis uses four selected weather elements belonging to different rotated factors. They are the diurnal mean values of temperature, of relative humidity, of cloudiness and of wind speed. The omitted (redundant) elements are (logarithm of) precipitation, sunshine duration, diurnal temperature amplitude and water vapour pressure. The omitted elements will be used in independent validation of classification efficiency.

Next, hierarchical cluster analysis (i.e. no a priori number of classes), the method of furthest neighbours is selected, after testing various other approaches. Considering the steep maximum of six for the optimum number of classes this number is fixed, and limits of the types are re-defined by method of K-means for all months and stations of the two investigated countries.

The obtained local classification will be assessed in comparison with efficiency of macro-circulation types. In overwhelming majority of the months, stations and variables the local types reduce the variance more effectively than the compared Péczely (1957) types for Hungary and an amalgamated (Mika et al., 1999) version of the original Hess-Brezowsky (1969) types, based on objective classification of average sea-level pressure maps derived by Bartholy and Kaba (1990).

These local weather types are important tools in understanding the role of weather in various environmental indicators, in climatic generalisation of short samples by stratified sampling and in interpretation of the climate change. Detection of climate change in terms of the frequency of weather types is another possible application of the local weather types.

Key words

weather types, climatology, factor analysis, cluster analysis, Hungary, Poland.

Introduction

Synoptic climatology i.e. classification of the endless variability of the everyday weather states according to the pressure configuration and frontal systems relative to the point, or region of interest has long history in meteorology. Its main advantage is to set a limited number of similar meteorological situations, which is the unavoidable to study any quantity or event of the environment for which its dependence on meteorological conditions should be quantified. Another advantage of this, so called, macro-synoptic classification (Peczely, 1957, Puskas, 2001, Piotrowicz, 2010) is that having the actual class of a given day selected, the same code can be applied for various stations or field-campaign. The price of this convenience is the limited efficiency of such circulation-based classifications for at least two reasons.

The first one is that the circulation objects and their frontal systems, related to them, often change their positions within the 24 hours of the most common classifications. Hence, the same code may hide rather different situations, indeed. The second reason is the lack of mezo-synoptic object due to the large-scale nature of synoptic analysis and the otherwise reasonable trial to keep the number of the individual classes limited. (Otherwise too long samples were needed.)

The logical alternative, i.e. classification of weather according to the observed local weather elements were less popular until the recent times for various reasons. At first, for long time, the numerical weather forecasts were able to outline the synoptic situation, but not the near-surface meteorological variables. At second, there were no computing facilities to operate with multivariate diurnal samples (order of ten variables, at once). Both problems have been resolved in the recent decades as a result of the rapid development in computer technology.

The numerical weather forecasting does not use the synoptic situation any longer, so a local classification may be equally useful, especially if providing better fit of the local weather elements.

Materials and methods

Eight stations from Poland and five stations from Hungary and 30 years periods (1966-1995 for Poland and 1961-1990 for Hungary) were selected with 8 weather elements (see below). The stations are Łeba, Suwałki, Olsztyn, Warszawa-Okęcie, Zielona Góra, Wieluń, Rzeszów-Jasionka, Bielsko-Biała-Aleksandrowice for Poland and Szombathely, Pécs, Budapest, Szeged, Debrecen for Hungary.

The four key weather elements of classification have been selected by factor analysis from the 8 candidates. Considering the skewed distribution of precipitation, its logarithm was further considered (with a 0.1 mm correction to keep the zero precipitation in the sample). All these elements have been standardised against the standard deviation within the monthly samples.

Table 1 indicates approximate results of the factor analysis for Budapest. The main conclusion is that 3 or 4 factors are enough, and, except precipitation, the climate elements belong to the same factors in majority of bimonthly sub-samples. The selected elements are *diurnal mean temperature* (Tm: °C), *cloudiness* (Cl: % of sky), *wind speed* (Ws: m/s) and *relative humidity* (Rh: %).

The omitted (redundant) elements are *precipitation* (Pcp: mm/d), *sunshine duration* (Sd: hour/d), *diurnal temperature amplitude* (ΔT: °C), *water vapour pressure* (Wvp: hPa).

These omitted elements serve for independent quantification of classification efficiency. The monthly sampling has been decided after analysing the standard deviation as a function of the duration of the sample from the annual (1 sample, no separation) to the daily (365 samples, full separation) amalgamation.

The hierarchical cluster analysis has been performed based on the selected four variables with 4-9 clusters in each month and station. The method of furthest neighbours has been selected for rules of joining the groups, based on Euclidean distance, after having tried several other possibilities. No Mahalanobis distance is applied since the retained four elements have no strong cross-correlations.

Since the frequency distribution of optimum numbers has a steep maximum at six classes, further we fixed this number of classes (i.e. weather types) for each station and month. *Fig. 1* indicates it for the Hungarian stations. There has no seasonal cycle been observed in the optimum number of monthly samples. The optimum number of the clusters was established when (i.) after using this number of clusters, the standardized average intra-group variance would be less than 70 % of the original variance of the four variables without clustering, (ii.) the difference between variances of the selected number of classes and of the by one smaller number of classes differs by more than 2 %, but (iii.) the same between the selected number and the one more clusters differ already by less than 1 %.

Table 1. *Results of factor analysis between the eight elements. The numbers indicate the factors.*

Budapest	Tm	Wvp	ΔT	Cl	Sd	Rh	Ws	Pcp
J-F	**1**	**1**	*2*	*2*	*2*	**3**	**3**	*4*
M-A	**1**	**1**	*2*	*2*	*2*	**3**	*4*	*2*
M-J	**1**	**1**	*2*	*2*	*2*	**3**	*4*	**3**
J-A	**1**	**1**	*2*	*2*	*2*	**3**	*4*	**3**
SZ-O	**1**	**1**	*2*	*2*	*2*	**3**	*4*	*2*
N-D	**1**	**1**	*2*	*2*	*2*	*2*	**3**	*4*

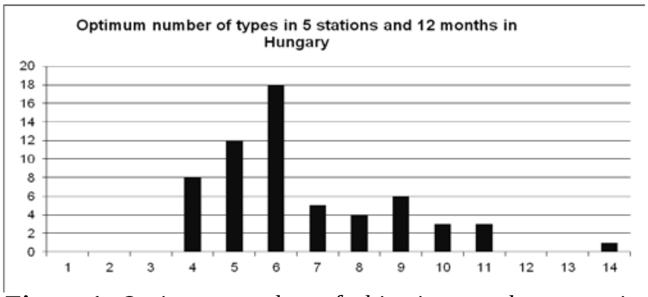


Figure 1. *Optimum number of objective weather types in Hungary according to the applied cluster analysis for 5 stations and 12 months for the applied 30 years period (1961-1990)*

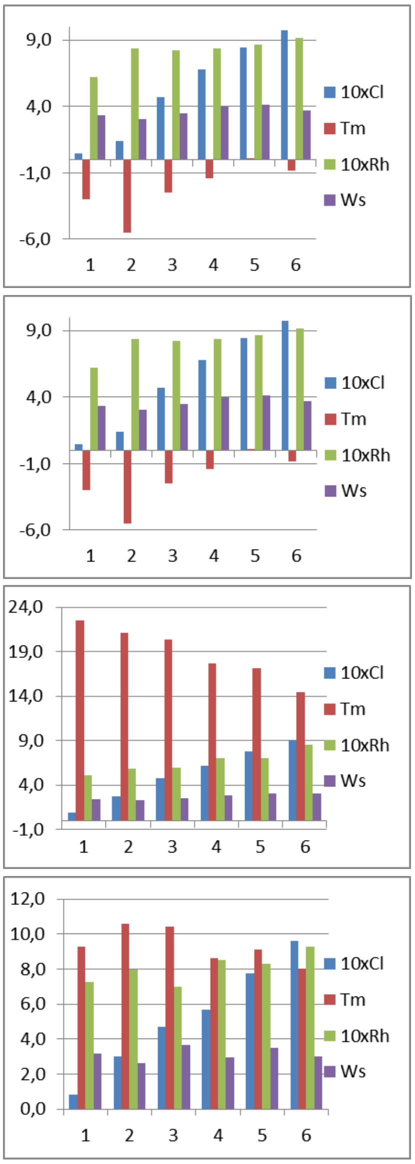
Results

K-Means cluster analysis with six classes

Considering the steep maximum of six for the optimum number of classes this number is fixed, and limits of the types are re-defined by method of K-means for all months and stations of the two investigated countries. This method provides the best separation of the four-component vectors in the space, i.e. yields the smallest average intra-class variance, consequently the largest inter-class variance.

Fig. 2 presents the six weather types for Warszawa in the central months of seasons. The types are characterised by their cluster means, sorted in increasing sequence of cloudiness. Some elements in a part of the seasons exhibit synoptically reasonable coincidence, but other cases they do not.

The days are distributed among the types fairly equally. The minimum and the maximum sample sizes were 12 and 376 in December. This proportion (1.2 %) is the least one from all cases. The frequency maximum of all cases occurred in Suwalki in December with 471 days (50.6 %).



January

April

July

October

Figure 2. *Examples of the objectively defined classes in Warsawa in the central month of the seasons. The types are sorted according to increasing sequence of cloudiness. The abbreviations are resolved on page 2.*

Validation of the classes

The obtained local classification is assessed in comparison with efficiency of macro-circulation types. In overwhelming majority of the months, stations and variables the local types reduce the variance more effectively than the compared Peczely (1957) types for Hungary and an amalgamated (Mika et al., 1999) version of the original Hess-Brezowsky (1969) types, based on objective classification of average sea-level pressure maps derived by Bartholy and Kaba (1990).

The two tables below demonstrate that the 6 objective types delimit the diurnal variance of the local weather elements than the 13 Peczely-types for Hungary and the 9 amalgamated Hess-Brezowsky types for Poland.

Table 2. *The variance explaining capacity of the objective local weather types at the 5 Hungarian stations as compared with the same capacity by the Peczely-types. The quotients of the two variances remained after using the centre of the groups as approaches are always smaller than 1. (The abbreviations stand for Budapest, Debrecen, Pecs, Szeged and Szombathely, respectively.) The two extra variables are the numbers of hours with over 80 % of relative humidity (r80) and over 15 m/sec wind speed (w15), respectively.*

Classes	Mean of the 4 basic elements (Tm, Rh, Ws, Cl)					Mean of the 6 independent elements (ΔT, Pcp, Sd, Wvp, r80, w15)					All the 10 elements in average				
	Bp	De	Pe	Se	So	Bp	De	Pe	Se	So	Bp	De	Pe	Se	So
Péczely	82%	86%	83%	84%	84%	86%	87%	86%	88%	87%	84%	88%	84%	85%	85%
Local	47%	56%	48%	50%	48%	66%	84%	66%	67%	61%	58%	74%	58%	59%	55%
quotient	**0,57**	**0,65**	**0,58**	**0,60**	**0,57**	**0,77**	**0,96**	**0,77**	**0,77**	**0,71**	**0,69**	**0,84**	**0,69**	**0,69**	**0,65**

Table 3. *The same as Tab. 2 for 8 Polish stations, as compared with the capacity of the amalgamated Hess-Brezowsky types. (The abbreviations stand for Leba, Suwalki, Olsztyn, Warszawa, Zielona Gora, Wielun, Rzeszow and Bielsko Biala, respectively.)*

Classes	Mean of the 4 basic elements (Tm, Rh, Ws, Cl)				Mean of the 4 independent elements (ΔT, Pcp, Sd, Wvp)				All the 8 elements in average			
	Le	Su	Ol	Wa	Le	Su	Ol	Wa	Le	Su	Ol	Wa
am. HB	90%	94%	93%	90%	100%	106%	99%	106%	95%	100%	96%	98%
Local	61%	58%	58%	56%	80%	62%	63%	85%	69%	60%	60%	70%
quotient	**0,68**	**0,63**	**0,63**	**0,63**	**0,83**	**0,60**	**0,64**	**0,84**	**0,74**	**0,61**	**0,64**	**0,73**

Classes	Mean of the 4 basic elements (Tm, Rh, Ws, Cl)				Mean of the 4 independent elements (ΔT, Pcp, Sd, Wvp)				All the 8 elements in average			
	Zg	Wi	Rz	Bb	Zg	Wi	Rz	Bb	Zg	Wi	Rz	Bb
am. HB	91%	91%	93%	87%	102%	105%	106%	102%	97%	98%	100%	95%
Local	61%	91%	60%	95%	77%	92%	61%	96%	69%	91%	61%	95%
quotient	**0,67**	**1,01**	**0,65**	**1,10**	**0,78**	**0,91**	**0,60**	**95**	**0,72**	**0,95**	**0,62**	**1,01**

Discussion

Derivation of the local weather types opens new perspectives in various applications. Moreover, these new scientific tools can be used to detect climate change in terms of diurnal weather, i.e. the process will be better transformed into everyday life of the society.

A sub-topic of these applications, the identification of extremes in terms of these newly defined weather types, also contributes to express our continuous process of man-made climate change in terms of probability of the harmful weather events that may occur at the mid-temperate latitudes.

Finally, education of meteorology and climatology for various professions, including geography and environmental sciences may gain a brain new scientific approach and methodology to express the local weather in less abstract and probably more efficient terms than circulation types.

Acknowledgements. The authors are grateful to the Polish-Hungarian bi-lateral project, TÉT_10-1-2011-0037 for the partial support.

References

Bartholy J. and M. Kaba, (1987): Meteorological and statistical analysis and correction of the Hess-Brezowsky macrosynoptic types. *Meteorological Studies* No.57 Hungar. Meteorol. Service 1-76 (In Hung.)

Hess, P., and H. Brezowsky, (1969): *Katalog der Grosswetterlagen Europas.* Berichte des Deutschen Wetterdienst, No. 113. Band. 15., Offenbach am Main

Mika J., Dunkel Z., Borbas É., G. Szenyan I. and Ferenczi Z., 1999: Scarab and ERBE radiation climatology resampled. *Advances in Space Research,* Vol. 24, No. 7, 959-962

Peczely G. (1957): *Grosswetterlagen in Ungarn.* Kleinere Veröffentl. Zentralanst. Meteor. N. 30. Budapest

Piotrowicz, K. (2010), *Sezonowa i wieloletnia zmienność typów pogody w Krakowie*, IGiGP UJ, Kraków.

Puskás J. (2001): New weather front types and catalogue for the Carpathian Basin. In. Nowinszky L. (ed.): *Light trapping of insects influenced by abiotic factors. Part III.* Savaria University Press Szombathely 87-118.