ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN - POLONIA
VOL. LV/LVI, 9 SECTIO B 2000/2001

Zakład Klimatologii
Uniwersytet Jagielloński

ANITA BOKWA

Cluster analysis as a method of typology of thermal conditions in the air layer near the ground

Analiza skupień jako metoda typologii stosunków termicznych przygruntowej warstwy powietrza

INTRODUCTION

The microclimatic measurements of temperature in the air layer near the ground (0-2 m above the ground) at non-standard levels have been usually carried on in short periods or on days with specific weather conditions (Geiger et al. 1995). Therefore it has not been possible to find out more generał regularities of thermal conditions in that air layer for the whole year. Moreover, the methods used so far have concentrated usually on descriptive statistics of a few chosen characteristics, mainly extreme temperatures and daily course of temperature at a certain height above the ground. The paper presents a new approach to the problem: first of all a unique, long-term microclimatological measurements series is analysed and secondly the method of cluster analysis is used for the first time to find the structure and regularities of the thermal conditions in the air layer near the ground.

DATA

The analysed data consist of 15-year long measurement series of temperature and come from the Research Station located in Gaik-Brzezowa (Wieliczka

Foothills, Carpathian Foreland, φ = 49°52'N, λ = 20°04'E), managed by the Institute of Geography and Space Economy, Jagiellonian University, Cracow, Poland. In the years 1976-80 the station was placed in the Raba River valley bottom (h = 259 m a.s.l.) and in the years 1988-97 it was already operating on the flattened ridge top nearby (h = 302 m a.s.l.), due to the construction of the Dobczyce Water Reservoir. Mercury bulb thermometers were placed in two Stevenson's shelters at the heights of 50 and 200 cm above the ground and they were read six times per day: 6 a.m., 8 a.m., 12 a.m. (noon), 2 p.m., 6 p.m. and 8 p.m. GMT. On the basis of air temperature measurements, for every day six values of thermal gradients were calculated (in C/1 m of height) for the air layer 50-200 cm above the ground. The measurements from the two periods, 1976-80 and 1988-97, represent the conditions of a river valley bottom and a top of a flattened ridge, respectively. We cannot compare the temperature directly, but we can compare the thermal gradients. As far as temperature is concemed it is possible to compare the indices of dispersion.

WHY CLUSTER ANALYSIS?

The author's main aim was to find the structure and regularities of thermal conditions in the air layer near the ground in the yearly course. First, descriptive statistics were calculated (mean, minimum, maximum, lower quartile, upper quartile, quartile range, variance, standard deviation, skewness, curtosis), separately for the temperature and gradients data of every month, measurement term, height above the ground, for 5-year periods and the whole 10-year period 1988-97 (Bokwa 2000). The comparison of results for files with temperature and files with gradients allowed to find out that the skewness for gradients is often higher than for the temperature in measurement terms and has both positive and negative values. In the warm half-year the differentiation of the temperature course at both measurement levels decreases, while for the gradients it increases. It is opposite in the cold half-year. For both temperature and gradients it is characteristic that in the valley bottom in the cold half-year the differentiation between measurement terms is larger than at the flattened ridge top. lt means that in the warm half-year the temperature course is more stable than in the cold half-year, and the values significantly different from the mean may be expected rarely, while the gradients in the warm half-year differ significantly from the mean quite often. Temperature describes the thermal energy of the air, while temperature gradient shows the difference in the beat content of two neighbouring air strata, developed due to low beat conduction of the air (Kędziora 1995). That is why temperature and gradient courses differ so much.

In the case discussed, descriptive statistics allow to carry a detailed analysis of the data and find periods of higher or lower dispersion of the values. We can analyse every measurement term or every month separately. However, we should consider a day as a basie time unit, within which the cyclic temperature changes take place. For every day there are six values of vertical thermal gradients available, each of them possessing a sign and a value. The sign informs about the direction of heat flux, and the value about the intensity of the energy exchange process. Therefore the diurnał sequence of thermal gradients defines the conditions of energy exchange for a certain day. There are many methods used to find days with similar weather conditions and to distinguish weather types, but the thermal conditions in the air layer near the ground have not been the subject of typology so far. The author decided to use the cluster analysis to find days with a similar course of thermal gradients.

METHODS

The advantage of the cluster analysis in the case discussed is its usefulness for the multi-dimensional analysis. Every day is a basie unit, defined with six parameters (six thermal gradients). All methods of cluster analysis have only one criterion of selecting objects into clusters and it is the mathematically defined similarity between them (Marek 1989), so they can be considered objective methods. The calculations were carried out for seasons (for 5-year periods and the whole 10-year period 1988-97), unlike for descriptive statistics, where all indices were calculated for months. lt was due to the fact that in cluster analysis the finał result depends strongly on the size of the sample (the larger the better) and additionally the descriptive statistics mentioned earlier indicated larger differentiation of the gradients between the seasons than between the months. Standard seasons were chosen as a most popular time resolution, mainly because the presented analysis cannot be compared with any previous one. However, the application of other time spans would be needed for further investigations, especially in case of transitional seasons, spring and autumn.

The methods of complete cluster analysis can be divided into two groups: non-hierarchical and hierarchical ones. In all methods the first step is creating a matrix of distances between objects, and for the data from Gaik-Brzezowa the Euclidean distance was used. The methods applied in the research were: k-means method (a non-hierarchical method), Ward's method and average linkage method (both hierarchical methods).

As the data come from 15 years of observations, the main purpose of the research was to find days with similar diurnał course of thermal gradients in order to establish a typology, separately for the valley bottom and the top of a flattened ridge. A type of diurnał course of thermal gradients is defined here as the sequence of mean values of gradients in six measurement terms in a cluster.

Therefore it was assumed that the distinguished types (clusters) should differ one from another as much as possible and their number should be the least possible, as the differences between mean values of the gradients in the measurement terms are rather small. When the increase of the clusters' number generated a new cluster with low number of cases and mean values little different from those of another cluster, then it was rejected and the previously obtained number of clusters (types) was accepted.

. In the k-means method the number of clusters to be obtained must be specified first. Therefore for all the files, the triple analysis was carriedout, assuming 4, 5 and 6 finał clusters. The procedure attempts to identify relatively homogenous groups of cases and form clusters of greatest possible distinction (Norusis 1990). For the flattened ridge top, the types of diurnał course of thermal gradients were distinguished on the basis of the data from the period 1988-97. For winter and spring four types were chosen, while for summer and autumn - five. In the case of the river valley bottom for all seasons four types were accepted as sufficient. In every season the types can be put in order from stable to dynamie (Bokwa - in print, a). Analyses carriedout for 5-year periods and the whole 10-year period 1988-97 prove that the differences between the two landforms are larger than the differences between two 5-year periods for the flattened ridge top. Therefore the data for the river valley may be considered sufficient for the comparison with the flattened ridge top, in spite of uneven measurement periods. In all the seasons and both landforms the most common are the stable types of diurnał gradient courses. In all cases they consist of 30% of days in every month of a season, except spring and autumn at the ridge top; in spring such days concentrate in March (about 50%) and in autumn - in November. It means that in other months of those seasons more differentiated gradient courses are more common. It only proves a transitional character of both seasons and the need for the application of other time spans for their investigation (Bokwa 2000).

In hierarchical methods the most usual output of computations is a tree plot, which is then the subject of the researcher's analysis. In the case of the data from Gaik-Brzezowa, due to the large sample sizes, that sort of analysis was impossible. Therefore, according to the results from the k-means method, the composition and size of clusters were analysed at the assumption of their number to be four and five. The average linkage method tumed out to be useless in the case of the discussed data, as in all analysed files, apart from large or medium-size clusters, very small clusters were generated, with objects not representing any extreme conditions. Their elimination and repeated procedure did not give better results. Such results cannot be used for the typology. Ward's method proved to be much more useful. Generally, both the k-means and Ward's methods give similar results; however, there are important differences.

For winter Ward' s method gives clusters in a smaller value range than the k-means method, which might be the result of less differentiation in clusters sizes. In the k-means method the smallest clusters, representing a dynamie course of gradients, are less numerous than in Ward's method and the mean gradient values are more differentiated. We can conclude then that they gather only days with very extreme values, while in the case of Ward's method the tendency to generate clusters of equal sizes results in adding to the dynamie gradient courses those which are only similar to them. Therefore the mean values of gradients in measurement terms are lower in such clusters in Ward's method than the in k-means method. It means that the k-means method generates more distinct clusters than Ward's method and this is a great advantage for the typology in winter. For other seasons the mean values of gradients are very similar in both methods.

CONCLUSIONS

Hitherto, taxonomie methods were applied in Polish climatology rnainly by Woś (1996), who used the Wrocław dendrite method to distinguish climatie seasons and to define their spatial variation in Poland. Also Olszewski and Żarnowiecki (1999) used that method to find out that the weather type controls the temperature and humidity conditions in forest communities to a far greater extent than the differences between the communities themselves. The present study reveals new possibilities of the application of the taxonomic methods, especially of cluster analysis. It is a useful tool in climatologieal research and it allows to create a typology of the diurnał course of thermal gradients in microscale in Wieliczka Foothills. Its application should be preceded by the calculation of descriptive statistics, whieh form its background. From the three methods of cluster analysis used here, the k-means one tumed out to be the most useful for the typology, Ward's method is estimated to be less useful, although it gives similar results, while the average linkage method should not be used for that purpose. Cluster analysis gives the opportunity to treat the whole day as a basie unit and analyse different weather elements together, in order to find similar pattems. The obtained typology has been further analysed to find out what weather conditions accompany different types of diurnał course of therrnal gradients and it can also be used as a part of the therrnal structure of the air near the ground (Bokwa 2000; Bokwa - in print, b). Cluster analysis demands relatively long measurement series if we want to obtain more generał results, but small samples can also be treated with it. For every data set and every purpose a few methods should be tested and the best one should be chosen.

REFERENCES

B o k w a A. 2000; Struktura termiczna przygruntowej warstwy powietrza na Pogórzu Wielickim (na przykładzie Gaika-Brzewwej). Maszynopis pracy doktorskiej, Bibl. Jagieł., Kraków.

B o k w a A. (in print, a); Typologia gradientów termicznych w przygruntowej warstwie powietrza. Rocz. AR w Poznaniu.

B o k w a A. (in print, b); Wpływ cyrkulacji annosferycznej na dzienny przebieg pionowych gradientów termicznych w przygruntowej warstwie powietrza. Dok. Geogr., PAN, Warszawa.

G e i g e r R., A r o n R. H., T o d h u n t e r P. 1995; The Climate Near the Ground. Vieweg, Braunschweig/Wiesbaden.

K ę d z i o r a A. 1995; Podstawy agrometeorologii. PWRiL, Poznań.

M a r e k T. 1989; Analiza skupień w badaniach empirycznych. Metody SAHN. PWN, Warszawa.

N o r u ś i s M. J. 1990; SPSS/PC+ Statistics 4.0 for the IBM PC/XT/AT and PS/2. SPSS Inc., Chicago.

O l s z e w s k i J. L., Ż a r n o w i e c k i G. 1999; Metody badań zróżnicowania fitoklimatu zbiorowisk leśnych w Białowieskim Parku Narodowym. Konf. nauk. "Klimatyczne uwarunkowania życia lasu", Za.kopane, 21-22 maja 1999: 183-190.

W o ś A. 1996; Struktura sezonowa klimatu Polski. Bogucki Wydawnictwo Naukowe, Poznań.

STRESZCZENIE

W pracy wykorzystano dane ze Stacji Naukowej Instytutu Geografii i Gospodarki Przestrzennej Uniwersytetu Jagiellońskiego, mieszczącej się w Gaiku-Brzezowej (Pogórze Wielickie, φ = 49°52'N, A = 20°04'E), z lat 1976-1980 (stacja w dnie doliny Raby, h = 259 m n.p.m.) i 1988-1997 (stacja na wierzchowinie, h = 302 m n.p.m.). Na podstawie pomiarów temperatury w sześciu terminach pomiarowych (6, 8, 12, 14, 18 i 20 GMT) w klatkach meteorologicznych na wysokości *50* i 200 cm n.p.g. obliczono pionowe gradienty termiczne (w °C/l m wys.). Dla wartości temperatury i gradientów, dla każdego miesiąca, terminu pomiarowego i wysokości n.p.g. obliczono podstawowe charakterystyki opisowe, osobno dla okresów pięcioletnich i całego dziesięciolecia 1988-1997. Wartości gradientów poddano analizie skupień w celu znalezienia typów ich dziennego przebiegu. Obliczenia te wykonano dla pór roku. Do obliczenia macierzy współczynników odległości użyto odległości euklidesowej. Wykorzystano metody k-średnich, Warda i średnich ważonych połączeń. Przy pomocy metodyk-średnich uzyskano po cztery typy przebiegu gradientów w każdej porze roku w dnie doliny i po cztery typy zimą i wiosną, a po pięć latem i jesienią na wierzchowinie. Metoda Warda okazała się nieco mniej przydatna do tworzenia typologii, a metoda średnich ważonych połączeń - zupełnie nieprzydatna. Analiza skupień stanowi cenne uzupełnienie statystyk opisowych, stosowanych dotychczas w badaniach klimatologicznych. Dla każdej serii obserwacyjnej i zależnie od celu badania należy zawsze wypróbować kilka metod analizy skupień i wybrać najwłaściwszą. Otrzymana typologia dziennego przebiegu gradientów może być wykorzystana do opracowania struktury termicznej przygruntowej warstwy powietrza, a ponadto można analizować, jakie elementy pogodowe wpływają na występowanie poszczególnych typów przebiegu gradientów.