

Modified, threshold-based circulation type classification for Central Europe, on the basis of Lityński's classification

Abstract

There are many classifications of atmospheric circulation types. In Poland, the most important one, used by Polish weather services, is the classification of circulation types developed by Lityński. This paper proposes four modifications to the currently used algorithm for determining atmospheric circulation types. The proposed algorithms have been compared and it has been shown which one produces a catalogue of circulation types in which the division of the distribution of the values of the three indices (the zonal index W_s , meridional index W_p and cyclonicity index C_p) into three classes is the closest to being equally likely. In 1986-2015, the classification scheme that deemed to be the best, differed in above 19% of cases from the currently used classification.

Keywords

Atmospheric circulation • circulation type classification • threshold-based classification • Lityński • Central Europe

© University of Warsaw – Faculty of Geography and Regional Studies

Kinga Kulesza

Faculty of Geography and Regional Studies,
University of Warsaw, Poland
e-mail: kingakulesza@uw.edu.pl

Received: 24 April 2018
Accepted: 11 October 2018

Introduction

The classification of atmospheric states into separate circulation types is a well-known tool for describing and analysing climate conditions. The main idea behind this is to move from continuous information about an atmospheric state (e.g., the pressure field on a given day) towards discrete information. This involves ordering individual atmospheric states and assigning them to groups of types with certain similarities. This is how a circulation type catalogue is created – each type is described with a value on a nominal scale. The main advantage of such an approach is clearly that circulation type catalogues are easy to use and that classification results can be compared at different locations and times, as well as the ability to analyse how different climate elements depend on atmospheric circulation. However, moving from continuous to discrete information entails a certain amount of information loss concerning the atmospheric state, which can make it difficult to interpret results. On the one hand, we thus want to develop the simplest catalogue of types; while on the other, we need to retain as much information as possible on the state of atmosphere. As a result, a number of classifications are being developed worldwide that seek to find an optimal method for classifying atmospheric circulation types – one that would be simple yet, at the same time, provide the most important information about the climate elements' variability. This results in existing classifications differing from one another, mainly in the number of types. European circulation type classifications are characterised by the large variation in the number of types: ranging from 13, (Péczely 1983) up to 40 (Schüepp 1979) or even 80 different predefined types (as in the classification developed by the Central Institute of Meteorology in Vienna). Most of the classifications used

in Europe have between 9 and 40 types, while the most popular ones are those with 9, 18, or 27 types (Philipp et al. 2010).

According to Philipp et al. (2010), circulation type classification schemes can be divided into two main groups: subjective and automated. Subjective classifications include the very well-known *Grosswetterlagen* classification (Hess & Brezowsky 1952; Gerstengarbe & Werner 2005), the Péczely (1983) classification, and the Polish ones by Osuchowska-Klein (1991) and Niedźwiedz (1981). The second type of classification methods, the automated or numerical methods, have the advantage of being independent of the individual experience of the observer; meaning that every researcher obtains the same results with the same data. These methods can be further divided into threshold-based methods (Schüepp 1979; Beck et al. 2007) and those in which the possible types are not defined prior to classification, such as principal component analyses (Buishand & Brandsma 1997; Esteban et al. 2006), leader-based algorithms (Lund 1963), and optimisation algorithms (Enke & Spekat 1997; Petisco et al. 2005). All three of these are based on the idea of identifying 'natural' types, without any presumptions or thresholds. A major drawback with such an approach, and especially of optimisation algorithms, is the fact that it requires the use of a large data set right from the beginning. Moreover, adding new data may entail changes to the previously defined catalogue. Of course, at some point in this process all objects are assigned to their nearest cluster, and reassignment is no longer necessary, so convergence to an optimum is reached. However, this would result in not being able to update the catalogue without changing the previously designated types. A catalogue in which set types are stable over time would seem to be more useful: such

a catalogue can be established using threshold-based methods. In this case, certain threshold values, as well as the number of all possible circulation types, are determined in advance. This group includes, for example, the Polish classifications by Lityński (1969), Ustrnul (1997), and Piotrowski (2009), as well as the Swedish classification by Linderson (2001), and the British one by Jenkinson and Collison (1977), which was based on the Lamb classification (1950).

The original circulation type classification

In the original work by Lityński (1969), the author outlines the main underlying principles of his method, that is, the method for constructing the zonal index, W_s ; the meridional index, W_p ; and the cyclonicity in Poland index, C_p . Lityński proposed dividing the distribution of the values of each index into three equally-likely classes. As a result, the frequency of each class under each index should amount to 33.33%, and all possible class combinations of the three indices produce 27 circulation types: 9 cyclonic types, 9 gradientless types, and 9 anticyclonic types.

Later, the source material used for establishing circulation types, and the algorithm for calculating the thresholds of index classes underwent a change (Pianko-Kluczyńska 2007; Philipp et al. 2010). At present, the Polish weather service uses the following principles for this classification: data from the NCEP/NCAR reanalysis of sea level pressure at 12:00 UTC are used (Kalnay et al. 1996) instead of the synoptic charts used by Lityński. Since reanalyses constitute grid data, the method for calculating the values of the W_s , W_p , and C_p indices has also undergone a change. The grid point, which is taken as the central point, is the one closest to Warsaw, that is, $\varphi=52.5^\circ N$, $\lambda=20^\circ E$. The zonal index, W_s , is determined as a horizontal zonal pressure gradient on the basis of a formula originally proposed by Lityński (1969). It amounts to the difference between the spatially averaged sea level pressure at parallels $\varphi-12.5^\circ$ and $\varphi+12.5^\circ$ within the $\lambda-20^\circ$ to $\lambda+15^\circ$ section, with a 5° step (where φ and λ are the coordinates of the central point). This difference is then divided by 25 (the number of degrees between the parallels) in order to produce the pressure gradient (Equation 1):

$$W_s = \frac{1}{25} \cdot \left(\frac{P(\varphi-12.5^\circ, \lambda-20^\circ) + P(\varphi-12.5^\circ, \lambda-15^\circ) + \dots + P(\varphi-12.5^\circ, \lambda+15^\circ)}{8} - \frac{P(\varphi+12.5^\circ, \lambda-20^\circ) + P(\varphi+12.5^\circ, \lambda-15^\circ) + \dots + P(\varphi+12.5^\circ, \lambda+15^\circ)}{8} \right) \quad (1)$$

The W_s index, as defined in this way, takes positive values for western circulation (western air mass advection) and negative ones for eastern circulation (eastern air mass advection).

The meridional index, W_p , is determined as a horizontal meridional pressure gradient. As with the W_s index, it equals the difference between spatially averaged sea level pressure at meridians $\lambda+15^\circ$ and $\lambda-20^\circ$ within the $\varphi-12.5^\circ$ to $\varphi+12.5^\circ$ section, with a 5° step (Equation 2):

$$W_p = \frac{1}{35} \cdot \left(\frac{P(\varphi-12.5^\circ, \lambda+15^\circ) + P(\varphi-7.5^\circ, \lambda+15^\circ) + \dots + P(\varphi+12.5^\circ, \lambda+15^\circ)}{6} - \frac{P(\varphi-12.5^\circ, \lambda-20^\circ) + P(\varphi-7.5^\circ, \lambda-20^\circ) + \dots + P(\varphi+12.5^\circ, \lambda-20^\circ)}{6} \right) \quad (2)$$

The W_p index, as defined in this way, takes positive values for southern circulation (southern air mass advection) and negative ones for northern circulation (northern air mass advection).

In turn, the C_p index indicates whether the domain is influenced by a cyclonic, anticyclonic, or gradientless circulation. It is determined by the sea level pressure at the central grid point (in hPa); in this case, $\varphi=52.5^\circ N$, $\lambda=20^\circ E$.

The algorithm for determining the thresholds of the circulation index classes has also undergone a change. The mean value \bar{x} and the standard deviation σ are determined for each index, and for each of the months during the whole available time period. Means and standard deviations are calculated on the basis of an increasingly longer period each year; for example, the mean value for January 1948 is calculated on 31 daily values, while the mean value for January 2003 – is based on 1,736 daily values (all 56 Januaries during the period 1948 to 2003). Based on such values for \bar{x} and σ , the thresholds of the three classes (negative, indifferent, and positive) are determined so that the distribution of these classes is close to being equally-likely (the indifferent class is contained between the lower threshold, $t_l = \bar{x} - 0.433\sigma$, and the upper threshold, $t_u = \bar{x} + 0.433\sigma$). Threshold values established in this way are then assigned to the middle dates of each month (the 16th day for 31-day months, the 15th day for 30-day months, and the 14th day for February). For days that are between the middle dates of two adjacent months, the threshold values are linearly interpolated (Pianko-Kluczyńska 2007). At the end, each day is assigned to a circulation type according to its classified index value. Circulation types are defined as combinations of all three index classes. Combining the three classes of W_s and the three classes of W_p results in nine advection types (e.g., $W_s =$ negative and $W_p =$ positive for type SE; $W_s =$ indifferent, and $W_p =$ indifferent for type 0 – no advection). Including the three C_p classes results in a further subdivision of advection types into 27 types according to their cyclonicity characteristics (e.g., $W_s =$ negative, $W_p =$ positive, and $C_p =$ positive for SE anticyclonic type).

Lityński's classification of atmospheric circulation types is one of the most important classifications used in climatological studies in Poland. The researchers who use it usually define it as an equally-likely classification – in reality, however, it differs from the equally-likely division. This paper proposes a number of small modifications to the algorithm currently in use for establishing atmospheric circulation types. The aim, therefore, is to compare the proposed algorithms, and indicate which one produces a catalogue of circulation types in which the division into three classes of the distribution of the values (of the three indices, W_s , W_p , and C_p) is closest to being equally-likely; thus actually fulfilling the principles of Lityński's original classification.

Materials and methods

In order to prepare circulation type catalogues, data originating from the NCEP/NCAR reanalysis on sea level pressure at 12:00 UTC for each day from 1948 to 2015 were used. The variants, which are described below, were compared for the period 1986 to 2015 (30 years). This circulation type classification is calculated for a spatial domain of a specified size. The central grid point, which is specified by the coordinates φ and λ , and the corner points of the domain are shown in Figure 1. In this paper the classification is carried out only for the grid point that is the closest to Warsaw, but the domain can be freely moved to any point in the European middle latitudes. Therefore, the data for the area $40-65^\circ N$ and $0-35^\circ E$ were used, and the grid point, $\varphi=52.5^\circ N$, $\lambda=20^\circ E$ was taken as the central point of the domain.

The compared variants of the classification scheme

Five variants of the classification scheme were prepared. Their main underlying principles are described below.

1. ORG: the classification currently used by the Polish weather service. In this classification, the W_p and W_s indices are determined using a 5° step, and the thresholds of the three index classes are defined based on the above-described algorithm, which was introduced by Pianko-Kluczyńska (on

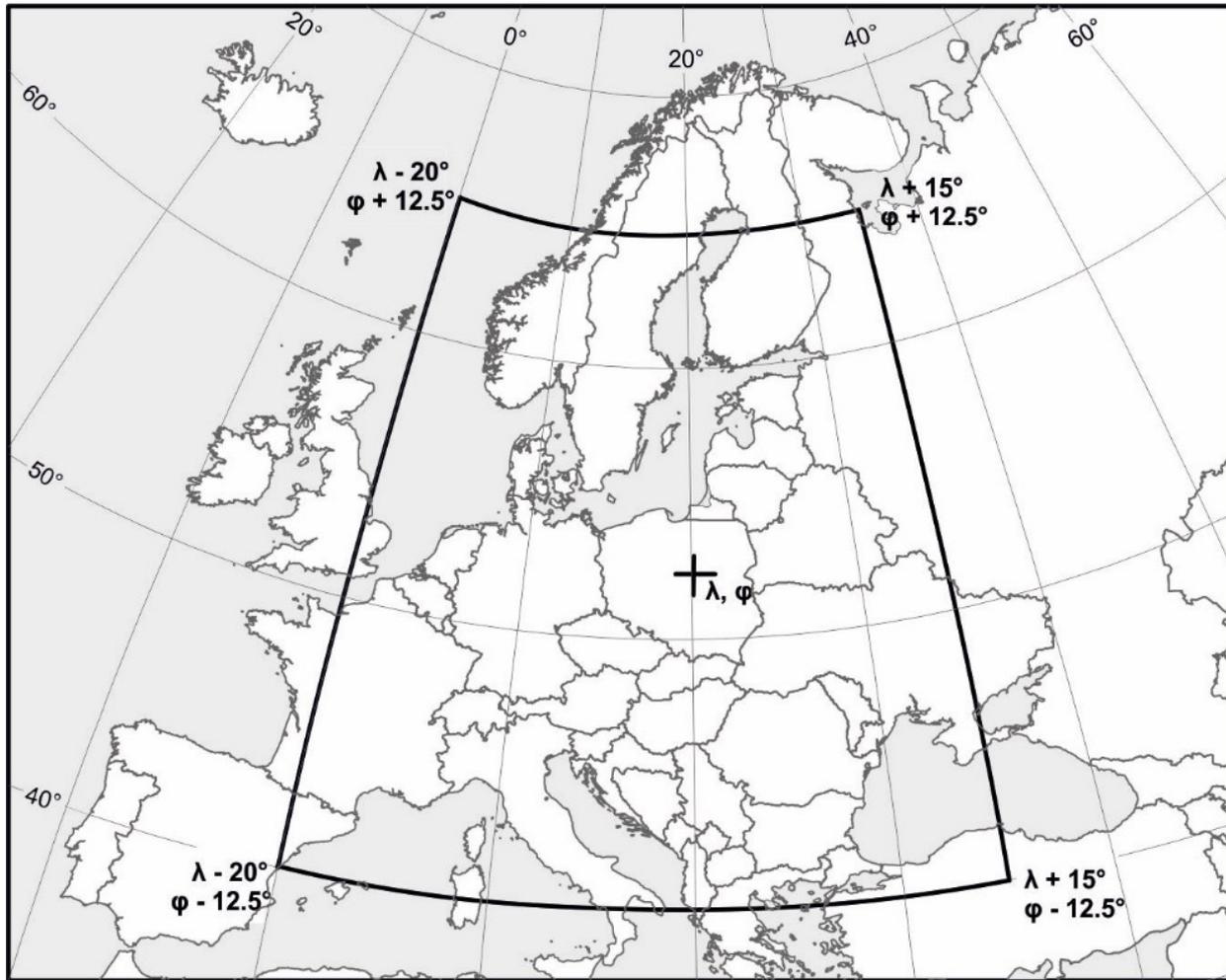


Figure 1. Example of a spatial domain for the Lityński-based circulation type classification. Source: own elaboration.

- the basis of all the material available at a given time; the lower threshold $t_l = \bar{x} - 0.433\sigma$ and the upper threshold $t_u = \bar{x} + 0.433\sigma$).
- NEW: circulation types are determined in a similar way to the ORG classification (the W_p and W_s indices are determined using a 5° step, the lower threshold $t_l = \bar{x} - 0.433\sigma$ and the upper threshold $t_u = \bar{x} + 0.433\sigma$), but with the time period, on the basis of which the thresholds of index classes are calculated, limited to 30 years. This means that instead of using an increasingly longer time period each year to calculate \bar{x} and σ , types are calculated using a moving period of the last 30 years only – the interval constitutes a climatological normals and, as it moves each year, can be considered as ‘running normals’.
 - NEW1: circulation types are determined in the same way as the NEW variant (the class thresholds are determined for a 30 year period, the lower threshold is $t_l = \bar{x} - 0.433\sigma$, and the upper threshold $t_u = \bar{x} + 0.433\sigma$), but the W_p and W_s indices are determined using a 2.5° step (i.e., that corresponding to one grid point of the NCEP/NCAR reanalysis).
 - MOD30: circulation types are determined as for the NEW1 variant (the W_p and W_s indices are determined using a 2.5° step, the thresholds for the circulation index classes are determined for a 30 year period), but using a division based on the percentiles 33 and 67 for calculating the class thresholds.
 - MOD20: circulation types are determined in the same way as for the MOD30 variant (the W_p and W_s indices are determined using a 2.5° step, the thresholds of the circulation index classes are calculated using a division based on the percentiles 33 and 67), but with the period on which the thresholds of the circulation index classes are calculated, limited to 20 years.
- In all variants the calculated threshold values are assigned to the middle date of each month. For days that are between the middle dates of two adjacent months, the threshold values are linearly interpolated.

Method of comparison

We carried out an evaluation of classification schemes in order to decide which variant results in a catalogue of circulation types where the distribution of index values in the three classes is the closest to being equally-likely. The comparison was made by calculating the mean absolute error (MAE). It was adopted as a way of measuring, for each classification scheme, how close

the distribution of values (of an individual index) within the three classes, was to be equally-likely, as shown in Equation 3:

$$MAE = \frac{1}{9} \sum_{i=1}^9 |T_i - E_i| \quad (3)$$

The empirical frequency of each of the nine classes is indicated by E ; while T indicates reference values (i.e., the theoretical equally-likely frequency [33.33%]); i means the subsequent class; and in each classification scheme there are 9 classes (three classes for the W_s index, three for W_p , and three for C_p). The MAE is expressed in percentage points, and the lower its value the closer the distribution of values in the three classes (for each index) is to being equally-likely.

Results

Comparison of ORG and NEW variants. The percentage of agreement between the types, determined using the ORG and NEW classifications, stands at 89.3% (Table 1); which means that, within the 30 year period, this percentage of days was classified identically for both of these variants. Both classifications were prepared from the same source material and using the same algorithm for determining their thresholds. It can therefore be concluded that the almost 11% of classification cases that were different resulted from using a different period (a period extended each year, or ‘running normals’) to calculate class thresholds.

Calculating the mean value, \bar{x} , and the standard deviation, σ , for a given index, for each month, on the basis of an increasingly longer time period, results in the thresholds being quite random at the beginning, and ‘flattened’ at the end of the long period. Another method may be to adopt a fixed reference time period (e.g., 1951–1990), as in the Polish classification by Nowosad (2017). Unfortunately, the choice of such a time period is subjective, and determining the circulation types after the period on which the class thresholds were calculated, results in the classification becoming less and less equally-likely with each year. The third method consists of calculating the threshold values based on a period of fixed length, but which moves each year. Such an approach has one considerable advantage – once the type of circulation is determined it will never change. The basic climatological reference time period that is recommended by the World Meteorological Organization and used in atmospheric circulation classification studies (Beck et al. 2007; Kulesza 2017) is 30 years, so an attempt can be made to use this ‘moving’ period to calculate the mean value and standard deviation. By doing so, each year takes into account only the last 30 years, making the changes in the class threshold values over the medium and long term more visible. In the long term, atmospheric circulation fluctuates – some circulation patterns increase in frequency while others decrease. Fluctuations in atmospheric circulation, which are caused by changes in the frequency of individual atmospheric pressure patterns, make it possible to determine circulation eras, that is, periods when certain macro-forms of circulation prevail. By shortening the period that \bar{x} and σ are calculated from, one can detect changes in these circulation eras, which results in the computed distribution of circulation types being closer to the equally-likely result. The above-mentioned ‘flattening’ of the values during a multi-year period is clearly visible in the case of the class thresholds that are calculated for the middle dates of January for 1948–2015 (i.e., the whole NCEP/NCAR reanalysis data set) (Figure 2). The class thresholds for the W_s index in the ORG variant (Figure 2b), after an initial few years of high volatility (caused by calculating \bar{x} and σ over a very short period), stabilise around 1993 at a level of 0.7–0.75 (the upper threshold) and 0.1–0.15 (the lower threshold). At the same time the thresholds set

Table 1. The matrix of agreement between the results of individual classification schemes (%). Variant names are explained in the text.

ORG				
89.3	NEW			
88.0	95.6	NEW1		
84.9	91.1	93.0	MOD30	
80.9	86.3	87.6	89.2	MOD20

Source: own elaboration.

in the NEW variant, for which only the last 30 years are taken into account when calculating \bar{x} and σ , coincided with those determined in the ORG variant until 1978, but after this clearly began to differ. As from 1993, the NEW variant thresholds, both the upper and the lower, rise to a level of 0.85–0.9 and 0.25–0.28 respectively, but after 2009 begin to decline. In the case of the W_p index (Figure 2a), from 1993 the NEW classification shows significant oscillations in the class thresholds around the threshold values set for the ORG variant; such changes would have been unnoticeable if only the ORG classification had been used for calculating the class thresholds.

Comparison of NEW and NEW1 variants. A minor change was introduced in the next step, namely, an increase in the density of the grid points on the basis of which the circulation indices are calculated. The NEW1 variant is, therefore, very similar to the NEW variant, and the only difference is the application of a 2.5° step (rather than a 5° step) when calculating the W_s and W_p indices. Consequently, the percentage of agreement between the types determined using these two variants is the highest of all variant pairs and amounts to 95.6% (Table 1).

Comparison of NEW1 and MOD30 variants. One more change has been introduced to the classification algorithm: percentiles have been used instead of the mean value and standard deviation to calculate class thresholds (the MOD30 variant). When using the formula $\bar{x} \pm 0.433\sigma$, we assume that the distribution of index values for a month is symmetrical around the mean value. In reality, this does not always have to be the case. Using the percentiles 33 and 67, it is still possible to divide a distribution into three classes of equal frequency, even if the distribution is asymmetric (which is often observed, for example, in relation to the zonal index, W_s , in January). Figure 3 shows the distribution of index values for selected months. The index values were calculated based on the NCEP/NCAR reanalysis’ data, using a 2.5° step. The presented class thresholds were calculated on the basis of a 30 year period, using the formula $\bar{x} \pm 0.433\sigma$ (NEW1 variant) or the percentiles 33 and 67 (MOD30 variant). The differences between the classes determined in this way are small, but the use of percentiles helps prevent the set catalogue from possibly becoming less equally-likely in the future.

Assuming that the distribution of values is divided into equally-likely classes means that the frequency of each class should be the same, that is, 33.33%. However, in reality the frequency of individual classes is not identical. In the ORG classification, the ‘western’ class was much more frequent (36.3%) than the ‘eastern’ (32.3%) and ‘zero’ classes (31.4%) (Figure 4b). Similarly, the frequency of the ‘northern’ class stood at 32.2%, while that of the ‘southern’ class was 34.5%, and of the ‘zero’ class, 33.2% (Figure 4a). In the NEW classification, the differences in the frequency of individual classes have been reduced. The spread of the W and E classes’ frequency amounted to only just under 1.0 percentage point, while the spread of N and S classes was

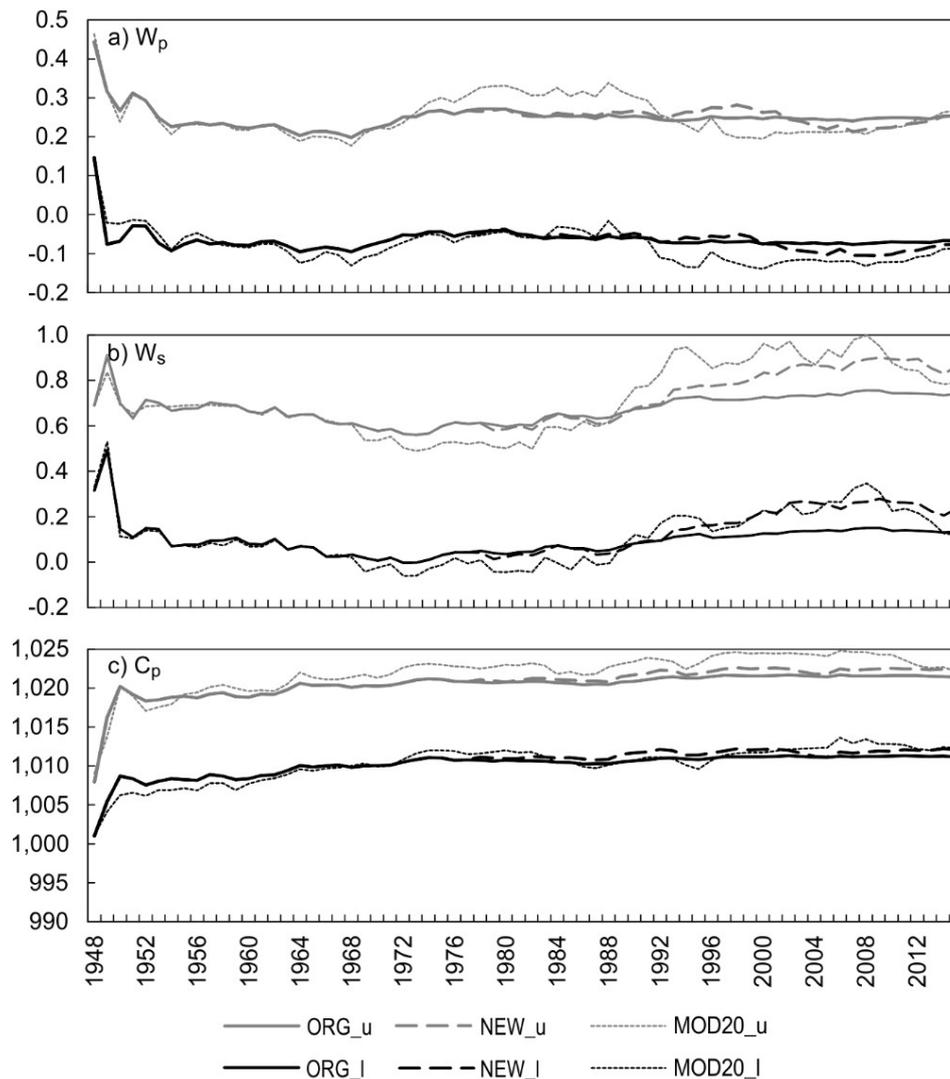


Figure 2. The thresholds for the classes of the indices W_p (a), W_s (b) and C_p (c) for 16 January 1948-2015. ORG_u – the upper threshold for the ORG classification; NEW_u – the upper threshold for the NEW classification; $MOD20_u$ – the upper threshold for the MOD20 classification; ORG_l – the lower threshold for the ORG classification; NEW_l – the lower threshold for the NEW classification; and $MOD20_l$ – the lower threshold for the MOD20 classification.

Source: own elaboration.

only 1.7 percentage points. Cyclonic circulation was also the least frequent in the NEW classification, but the differences in the frequency of individual circulation types have been reduced in this classification: cyclonic circulation was observed almost 32.1% of the time, while anticyclonic circulation was for 35.3% of the time. These differences in the frequency of the individual classes of circulation indices within these two classifications result from the fact that in the ORG variant the threshold values are calculated on the basis of an increasingly longer period each year. However, even though in the NEW variant running normals is used instead, the frequency of each class is also not identical in this case. This is due, among other things, to the method adopted for defining class thresholds (using the mean value and standard deviation), as well as to the use of linear interpolation. Using linear interpolation to determine class thresholds for individual days between the middle dates of two adjacent months, results in the distribution of the indices being divided into three classes

that are never equally-likely. Yet this method of calculating the thresholds has been adopted so as to avoid a situation where two consecutive days that belong to adjacent months and have identical circulation parameters (identical values of all three indices) are classified as being completely different types. In subsequent variants, the differences in the frequencies of individual index classes have been reduced even more. In the MOD30 variant, the frequency spread of the W and E classes amounts to only 1.1 percentage points, while the spread of the C and A classes stands at only 0.6 percentage point (Figure 4). The differences in the frequencies of classes are even smaller in the MOD20 variant (described below).

Taking into account all the variants compared so far, a synthetic comparison of all classification schemes has been made using the mean absolute error (MAE). In this case, the theoretical equally-likely frequency (where the frequency of each class amounts to 33.33%) has been used as a model value.

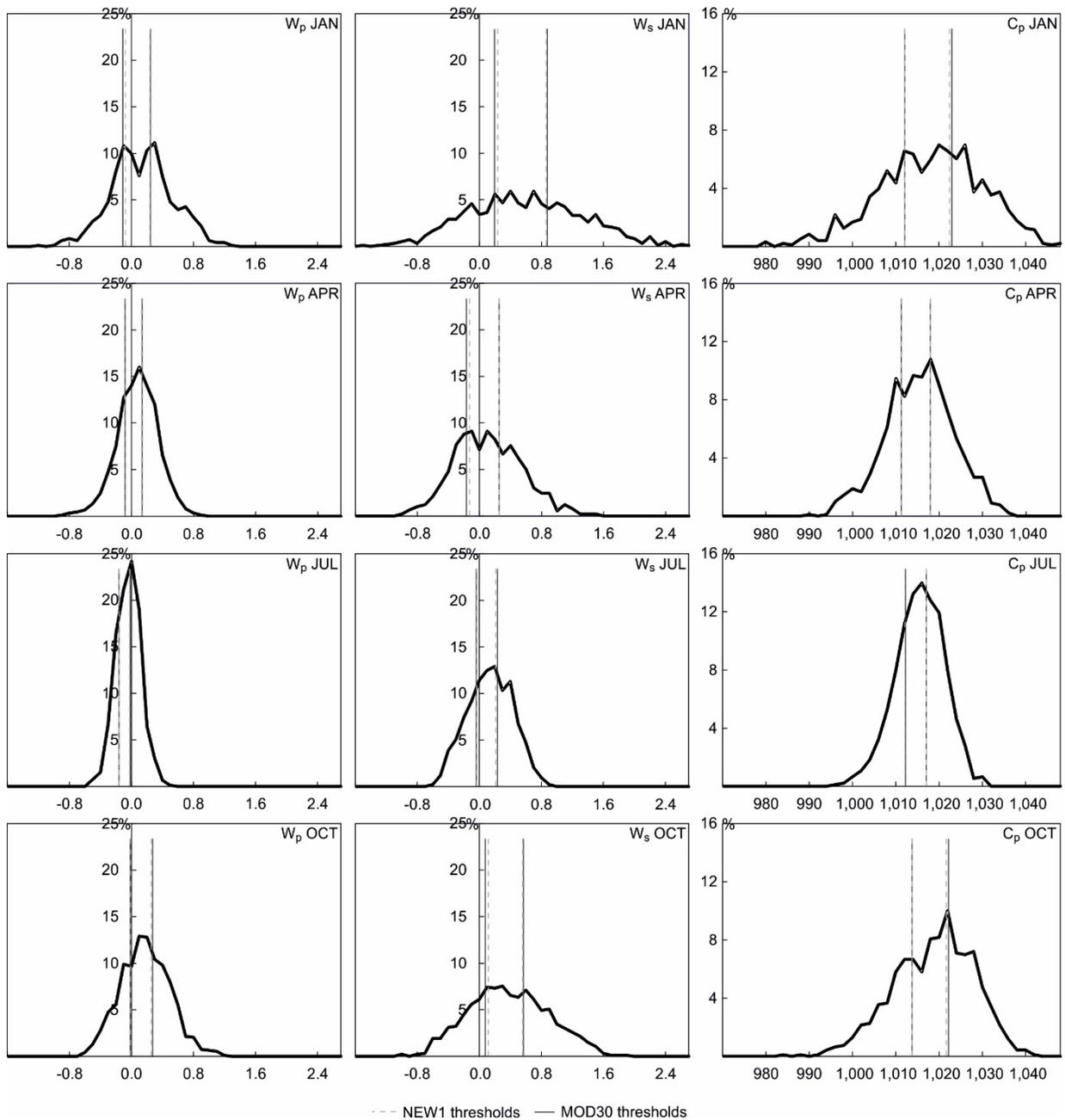


Figure 3. The value distribution of the W_p , W_s , and C_p indices for January (JAN), April (APR), July (JUL), and October (OCT) 1986-2015. Vertical dashed lines denote the class thresholds according to the formula $\bar{x} \pm 0.433\sigma$ (the NEW1 variant); vertical continuous lines denote the class thresholds according to the percentiles 33 and 67 (the MOD30 variant). Variant names are explained in the text. Source: own elaboration.

Table 2 shows that the smallest mean absolute error is found in the MOD30 classification.

In Table 2, three variants (NEW, NEW1, MOD30) adopted a 30 year period for determining class thresholds; however, using a 30 year period does not automatically mean that the classification generated on this basis will be closest to being equally-likely. This is why the MOD30 variant has been tested at various time period lengths on the basis of which the class thresholds are calculated. Periods of 5, 10, 15, 20, 25, 30 and 35 years were chosen for comparison; they are referred to, respectively, as MOD5, MOD10,

MOD15, MOD20, MOD25, MOD30, and MOD35. The smallest MAE calculated was for the MOD20 variant, 0.55 percentage points (Figure 5). This means that the MOD20 classification is closest to being equally-likely.

Changing the period on which the class thresholds are determined, as well as changing the method of their determination (e.g., percentiles, grid density) results in a situation where, on a given day, individual classification variants may designate completely different circulation types. Figure 6 includes several examples that illustrate this issue; it shows the circulation types

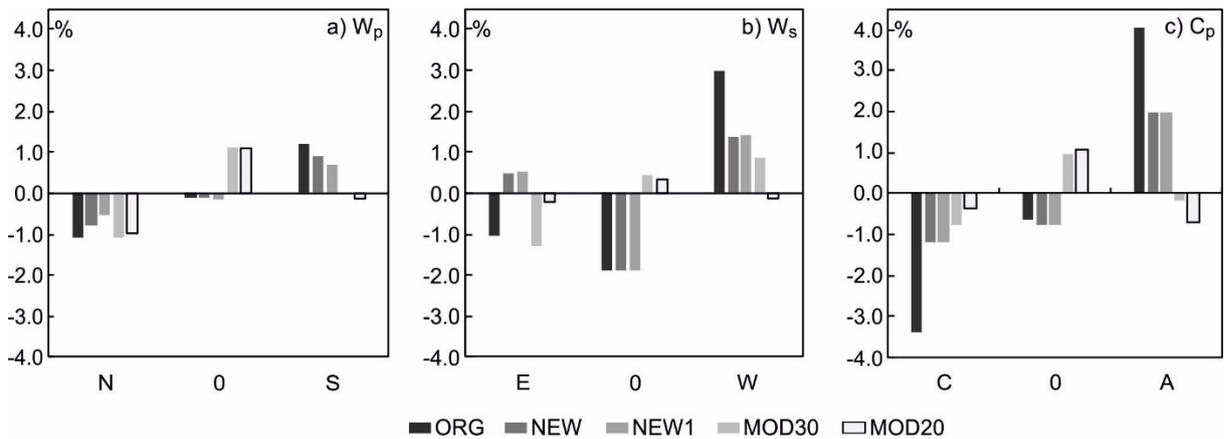


Figure 4. Deviations (%) of the frequencies of individual classes for the indices W_p (a), W_s (b), and C_p (c) from the theoretical equally-likely frequency (where the frequency of each class equals 33.33%), in individual classification variants. Variant names are explained in the text.

Source: own elaboration.

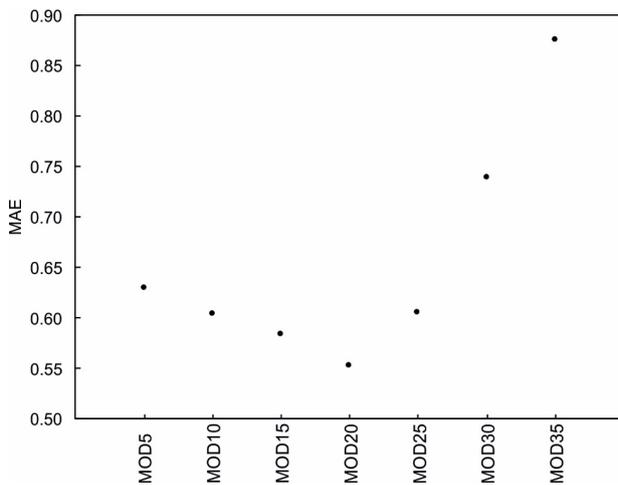


Figure 5. Mean absolute error (MAE) for the sub-variants MOD5, MOD10, MOD15, MOD20, MOD25, MOD30, and MOD35 at different periods of time on which the class thresholds are calculated. Variant names are explained in the text.

Source: own elaboration.

Table 2. Mean absolute error (MAE) for the different variants of the classification algorithm. Variant names are explained in the text.

Variant	ORG	NEW	NEW1	MOD30
MAB (pp)	1.82	1.05	1.01	0.74

Source: own elaboration.

determined on eight different days in the ORG and MOD20 classifications. For example, on 28 April 2011 (Figure 6d), the circulation type according to the ORG classification (EA) was totally different to the one determined in the MOD20 classification (NE0). Air pressure in Warsaw stood at below 1017 hPa and both the isobars' structure and the location of the cold front indicated air mass inflow from the north-east. On 3 August 2013 (Figure 6f), the col was located above central Europe and the

air pressure in Warsaw amounted to 1018 hPa. On that day, the ORG classification indicated a type WA, while the MOD20 classification identified gradientless circulation without air mass inflow from any direction (00). Another example is 8 October 2008 (Figure 6g), when Poland was under the influence of a high-pressure system centred over the Baltic Sea. Both classifications showed no air mass inflow on that day; however, the MOD20 variant, in contrast to the ORG classification, indicated a 'zero' (gradientless) circulation type (air pressure in Warsaw stood at 1021 hPa). These are only some of the examples of days that were classified differently between the ORG and MOD20 classifications, but it should be noted that in total, 19.1% of days were classified in a different way (Table 1). However, the differences between the types that were calculated in the ORG and the MOD20 classifications for any given day are not greater than one class (e.g., S may change to 0 but not to N, or C may change to 0 but not to A).

Finally, let us add a few words on the difference between the determined and actual direction of air mass inflow. In the equally-likely distribution, both thresholds that indicate the division into the three classes may happen to be on the same side of zero. In such situations, the 'zero' class will not contain 'neutral' values that are on both sides of zero, and an additionally designated negative class may also contain positive values or vice versa (e.g., low positive values of the W_s index may also be found in January in class E). During the period under consideration, a positive value of the W_s index's lower threshold in the MOD20 variant was observed in 45.6% of cases, mainly during the cold half of the year from the end of August or the beginning of September to January or February. This means that during this time period, the 'eastern' circulation types are not necessarily linked with an inflow of air from the east. It should be noted, however, that in the whole 30 year period there were only 413 days (3.8%) when the value of the W_s index was positive, yet despite this, the 'eastern' circulation type was determined. On the other hand, the upper threshold of the W_s index was always positive. As for the W_p index, its lower threshold took a positive value in only 16 cases (0.2%), and in the case of all circulation types classified as 'northern' on these days, the W_p index was far below zero. It should therefore be concluded that the 'northern' types are always linked with an inflow of air from the north. On the other hand, the upper threshold of the W_p index took values below zero in 17.1% of cases, mainly in the summer months

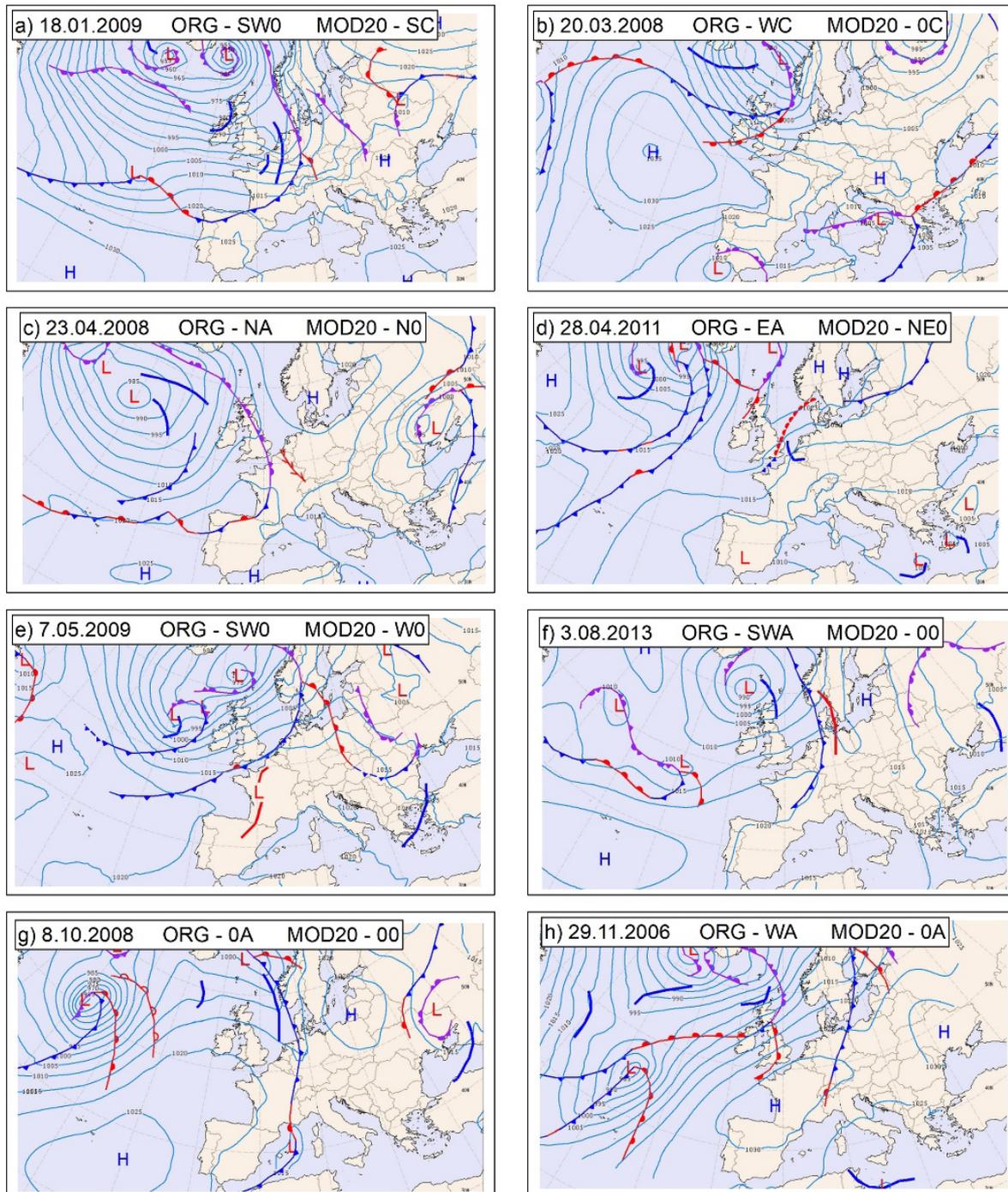


Figure 6. Atmospheric circulation types determined using the ORG and MOD20 classifications in comparison with synoptic charts for selected days at 12:00 UTC. Variant names are explained in the text.

Source: own work on the basis of synoptic charts obtained from: Koninklijk Nederlands Meteorologisch Instituut 2018.

(June, July, August); while the minimum value of this threshold was not lower than -0.054. This means that in summer, 'southern' types are not necessarily connected with the inflow of air from the south; but in the whole 30 year period only 87 cases (0.8%) were classified as 'southern' types, while the actual value of the W_p index on these days was negative.

Discussion and conclusion

This paper compares the different variants of circulation type classification by Lityński. Five different variants of the classification algorithm that were based on different underlying principles, were compared. This made it possible to indicate which algorithm generates a catalogue of circulation types in which the division into three classes of the distribution of the

values of the W_s , W_p and C_p indices is the closest to being equally-likely. During 1986–2015, the MOD20 classification differed from the ORG classification in more than 19% of cases. In subsequent variants of the classification algorithm, the differences in the frequencies of each class (in relation to individual circulation indices) became smaller and smaller. Ultimately, the use of the MOD20 classification resulted in generating a catalogue of circulation types in which the division of the distribution of the values in the W_s , W_p and C_p indices into three classes was the closest to being equally-likely.

MOD20 is a threshold-based classification. In general, threshold-based methods perform surprisingly well in comparison to more complicated algorithms (Huth 2010). The relative simplicity of the MOD20 classification algorithm is its big advantage, in contradiction to methods using principal component analysis or optimisation methods. Even though, in recent times, computing time and algorithm complexity no longer constitute a problem, by having a simpler method, it is also easier to understand what exactly is being produced by the classification algorithm. According to Beck and Philipp (2010) the original Lityński classification (ORG) is one of the threshold-based methods that produces the best results in Central Europe.

The advantage of the MOD20 classification is its use of the empirical distribution of index values – the percentiles 33 and 67 – instead of the theoretical distribution $\bar{x} \pm 0.433\sigma$. In this classification scheme, the thresholds are calculated based on of the last 20 years, which is the optimum balance between the stability of the percentiles and circulation types. The presented MOD20 circulation type classification has 27 predefined types, which is the number of types frequently used in European classifications (Philipp et al. 2010). Owing to this, each synoptic situation can be classified clearly and there are no unclassified or transitional situations. With such an approach, the catalogue of types is easy to use. On the other hand, it should be remembered that indeterminate situations also occur in nature, and these have to be assigned to a certain type anyway. The MOD20 classification scheme is automated and makes it possible to do recalculations, so it is applicable to different regions and time periods. However, it has certain limitations. It may not give proper results in regions

of certain topography, for example, mountain basins, where atmospheric circulation and the advection of air masses are determined mostly by local factors. On the other hand, it is a macroscale classification, so it provides information concerning a very big area around a central grid point. Even though it is calculated for a given point, it defines the atmospheric circulation pattern for a domain of a hundred square kilometres. The MOD20 classification should be used for data from 1967 onwards (i.e., for the years that have the whole preceding fixed-length period on which thresholds are calculated). A major limitation of the MOD20 classification scheme is its use of the moving period on which the thresholds are calculated. This can lead to a situation where two days, for example, at the beginning and at the end of a long time period, that have the same values of the three indices, may be classified differently. A solution to this problem would be to establish a fixed period on which the thresholds are calculated, for example, 1951–2000. Yet it should be remembered that the output catalogue of circulation types would be decreasingly equally-likely after the year 2000. Additionally, other changes to the ways for calculating thresholds (e.g., calculating them for each day separately, instead of calculating for the middle dates of months and interpolating the results) or to the classification scheme itself, could be introduced. This should be the subject of future research.

The algorithm for generating atmospheric circulation types according to the MOD20 method (with the thresholds based on either moving 20 year period or the fixed period, 1951–2000) is available at MATLAB File Exchange's web page.

Funding

This research was carried out with the support of the Interdisciplinary Centre for Mathematical and Computational Modelling (ICM) of the University of Warsaw under grant no. G67-1.

Acknowledgements

The author expresses her gratitude to Justyna Lisok for her help in writing the function in MATLAB scripting language.

References

- Beck, C, Jacobbeit, J & Jones, PD 2007, 'Frequency and within-type variations of large-scale circulation types and their effects on low-frequency climate variability in central Europe since 1780', *International Journal of Climatology*, vol. 27, no.4, pp. 473-491.
- Beck, C, Philipp, A 2010, 'Evaluation and comparison of circulation type classifications for the European domain', *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 35, no 9-12, pp. 374-387.
- Buishand, TA & Brandsma, T 1997, 'Comparison of circulation classification schemes for predicting temperature and precipitation in the Netherlands', *International Journal of Climatology*, vol. 17, no. 8, pp. 875-889.
- Enke, W & Spekat, A 1997, 'Downscaling climate model outputs into local and regional weather elements by classification and regression', *Climate Research*, vol. 8, pp. 195-207.
- Esteban, P, Martin-Vide, J & Mases, M 2006, 'Daily atmospheric circulation catalogue for Western Europe using multivariate techniques', *International Journal of Climatology*, vol. 26, no. 11, pp. 1501-1515.
- Gerstengarbe, FW & Werner, PC 2005 'Katalog der Großwetterlagen Europas (1881-2004) nach Paul Hess und Helmuth Brezowsky', Selbstverlag des Deutschen Wetterdienstes, Offenbach, Potsdam.
- Hess, P & Brezowsky, H 1952, 'Katalog der Großwetterlagen Europas', Berichte des Deutschen Wetterdienstes in der US Zone 33, Bad Kissingen.
- Huth, R 2010, 'Synoptic-climatological applicability of circulation classifications from the COST733 collection: First results', *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 35, no. 9-12, pp. 388-394.
- Jenkinson, AF & Collison, BP 1977, 'An initial climatology of gales over the North Sea', Synoptic Climatology Branch, Meteorological Office, London.
- Kalnay, E, Kanamitsu, M, Kistler, R, Collins, W, Deaven, D, Gandin, L, Iredell, M, Saha, S, White, G, Woollen, J, Zhu, Y, Leetmaa, A, Reynolds, R, Chelliah, M, Ebisuzaki, W, Higgins, W, Janowiak, J, Mo, KC, Ropelewski, C, Wang, J, Jenne, R & Joseph, D 1996, 'The NCEP/NCAR 40-Year Reanalysis Project', *Bulletin of the American Meteorological Society*, vol. 77, no. 3, pp. 437-471.
- Koninklijk Nederlands Meteorologisch Instituut, Available from: <<http://www.knmi.nl>>. [24 April 2018]

- Kulesza, K 2017, 'Nowe spojrzenie na klasyfikację typów cyrkulacji atmosfery J. Lityńskiego' [New look on the classification of atmospheric circulation types by J. Lityński], *Prace Geograficzne [Geographical Studies]*, vol. 150, pp. 79-94.
- Lamb, HH 1950, 'Types and spells of weather around the year in the British Isles: Annual trends, seasonal structure of the year, singularities', *Quarterly Journal of the Royal Meteorological Society*, vol. 76, no. 330, pp. 393-429.
- Linderson, ML 2001, 'Objective classification of atmospheric circulation over southern Scandinavia', *International Journal of Climatology*, vol. 21, no. 2, pp. 155-169.
- Lityński, J 1969, 'Liczbowa klasyfikacja typów cyrkulacji i typów pogody dla Polski' [A numeral classification of circulation and weather types for Poland], *Prace Państwowego Instytutu Hydrologiczno-Meteorologicznego [Works of the National Hydro-Meteorological Institute]*, vol. 97, pp. 3-15.
- Lund, IA 1963, 'Map-Pattern Classification by Statistical Methods', *Journal of Applied Meteorology*, vol. 2, no. 1, pp. 56-65.
- NCEP/NCAR Reanalysis, Available from: <<https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html>>. [24 April 2018]
- Niedźwiedz, T 1981, 'Sytuacje synoptyczne i ich wpływ na zróżnicowanie przestrzenne wybranych elementów klimatu w dorzeczu górnej Wisły' [Synoptic situations and their influence on the spatial diversity of selected climate elements in the upper Vistula basin], *Rozprawy Habilitacyjne UJ [Jagiellonian University habilitation dissertations]*, Cracow.
- Nowosad, M 2017, 'Kalendarz wskaźników cyrkulacji i typów cyrkulacji nad Polską według formuły Lityńskiego. Zbiór komputerowy' [Calendar of circulation indicators and circulation types over Poland according to the Lityński formula. Digital collection], Uniwersytet Marii Curie-Skłodowskiej, Zakład Meteorologii i Klimatologii [Maria Curie Skłodowska University, Department of Meteorology and Climatology].
- Osuchowska-Klein, B 1991, 'Katalog typów cyrkulacji atmosferycznej (1976-1990)' [Catalogue of atmospheric circulation types (1976-1990)], IMGW, Warsaw.
- Péczely, G 1983, 'Magyarország makroszinoptikus helyzeteinek katalógusa (1881-1983)', *Országos Meteorológiai Szolgálat Kisebbségi Kiadványai*, vol. 53.
- Petisco, SE, Martín, JM & Gil, D 2005, 'Método de estima de precipitación mediante "downscaling"'. Nota técnica no. 11 del Servicio de Variabilidad y Predicción del Clima, INM, Madrid.
- Philipp, A, Bartholy, J, Beck, C, Erpicum, M, Esteban, P, Fettweis, X, Huth, R, James, P, Jourdain, S, Kreienkamp, F, Krennert, T, Lykoudis, S, Michalides, SC, Pianko-Kluczyńska, K, Post, P, Álvarez, DR, Schiemann, R, Spekat, A & Tymvios, FS 2010, 'Cost733cat – A database of weather and circulation type classifications', *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 35, no. 9-12, pp. 360-373.
- Pianko-Kluczyńska, K 2007, 'Nowy kalendarz typów cyrkulacji atmosfery według J. Lityńskiego' [New calendar of atmosphere circulation types according to J. Lityński], *Wiadomości meteorologii, hydrologii, gospodarki wodnej [News of the meteorology, hydrology and water management]*, vol. I(LI), no. 4, pp. 65-85.
- Piotrowski, P 2009, 'Obiektywna metoda klasyfikacji cyrkulacji atmosferycznej dla Polski' [Objective method of classification of atmospheric circulation for Poland], *Acta Universitatis Lodzensis, Folia Geographica Physica*, vol. 10.
- Schüepp, M 1979, 'Witterungsklimatologie - Klimatologie der Schweiz III', *Beihefte zu den Annalen der Schweizerischen Meteorologischen Anstalt 1978*.
- Ustrnul, Z 1997, 'Zmienność cyrkulacji atmosfery na półkuli północnej w XX wieku' [The variability of atmospheric circulation in the northern hemisphere in the 20th century], *Materiały Badawcze IMGW, Meteorologia [Research materials of IMWM, Meteorology]*, vol. 27.