Annales Societatis Geologorum Poloniae (2004), vol. 74: 285–294.

INTEGRATION OF LATE GLACIAL AND HOLOCENE POLLEN DATA FROM POLAND

Adam WALANUS[1] & Dorota NALEPKA[2]

[1] *Institute of Archaeology, University of Rzeszów, Hoffmanowej 8, 35-016 Rzeszów, Poland; walanus@univ.rzeszow.pl*
[2] *W. Szafer Institute of Botany, Polish Academy of Sciences, Lubicz 46, 31-512 Kraków, Poland; nalepka@ib-pan.krakow.pl*

Walanus, A. & Nalepka, D., 2004. Integration of Late Glacial and Holocene pollen data from Poland. *Annales Societatis Geologorum Poloniae*, 74: 285–294.

Abstract: Quaternary palynological data are stored in tables which are typical computer database objects. The individual pollen table, in order to be integrated with other tables using computer-based methods, has to be attributed with (1) geographical coordinates, (2) dates attached to each row (sample) of the table, and (3) taxa names, common to all tables, attached to each column. In this paper, integration of individual lists of palynological taxa is shortly described, and some remarks are given on extracting data for selected time slices. The problem of dating of all pollen spectra, while few radiocarbon dates are available in the profile, is considered. Some mathematical models of age-depth relation are proposed, as well as incorporation of information derived from lithology of the profile. The algorithm of integration (averaging) of pollen percentages on the map of Poland is discussed. The techniques of bootstrap and so-called removed residuals are proposed as tools for assessment of reliability of isopollen lines. A possibility of construction of migration (rate of change) maps is also mentioned. The presented algorithm has been used to obtain hundreds of isopollen maps for the Holocene in the area of Poland.

Key words: palynology, isopollen maps, radiocarbon dating, weighting function, bootstrapping, Vistulian Late Glacial, Holocene, Poland.

Manuscript received 2 May 2004, accepted 6 August 2004

INTRODUCTION

Managing huge amount of data is a typical task for computer. In Quaternary palynology, many authors create pollen tables (Fig. 1) which consist of tens to hundreds of rows (pollen spectra, assigned to coring depth or age) and tens to hundreds of columns (pollen taxa). In typical sample (spectrum) 1,000 or more pollen grains are counted. In the database used for production of isopollen maps for Poland (Polish Committee for Scientific Research grant: KBN no. 6 PO4F 02818; cf. also Ralska-Jasiewiczowa, *in print*), the total amount of 23 million of pollen grains are stored. Each of these grains has been observed, recognized, and counted by palynologist.

Having 100 or more palynologically-investigated sites, more or less uniformly distributed over a region (country), it is worth to exploit the data not only to discuss individual development of vegetation around the site, but also to construct the pattern of vegetation on the map of the region (Szafer, 1935; Huntley & Birks, 1983; Ralska-Jasiewiczowa, 1983; Hoek, 1987a,b). It is a relatively easy task if the data are stored in a computer database; however, some advanced software tools as well as dedicated algorithms are necessary for correct and efficient data handling (Nalepka & Walanus, 2003a).

Even for a comparison of two pollen tables (pollen diagrams), the taxa have to be one-to-one coordinated in both tables. It seems to be trivial, however, bearing in mind that the data were produced by different palynologists, and since the data are to be automatically processed, some problems are easily seen. A minor problem is that of differences in the way of typing taxa names: either *Betula t.* or *Betula type*, or *Betula* (or in invisible space after the last character). More important is the question connected with taxonomical resolution and only local meaning of some taxa.

Compared with the question of taxa names, the problem of spectra "names" is of quite different nature. In a pollen table, depths in the sediment profile are described in rows. However, it is obvious that some dates have to be attributed to each spectrum, since in the Quaternary this is the only way of synchronization of data. The dates have to be "absolute", not relative, i.e. not based on relation to the spectra from another site (Walanus & Nalepka, 1996). While the palynologist has, as a rule, some information or assumption

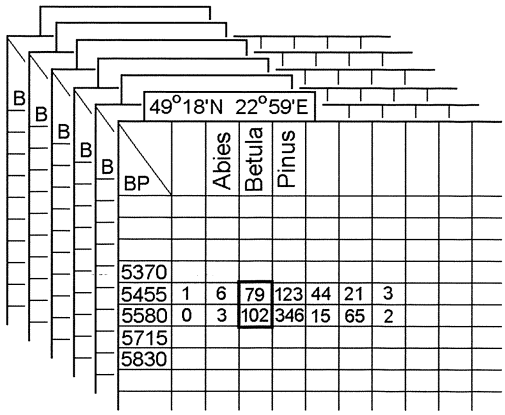

Fig. 1. Pollen data ready for the regional integration: number of pollen tables, with taxa names (columns) belonging to one regional list of taxa, and with spectra (rows) dated (for example, in ^{14}C years BP). A taxon/time slice is marked with bold rectangle. It is for *Betula*, and time: 5,500±100 yrs BP. The geographical coordinates are attached to each table

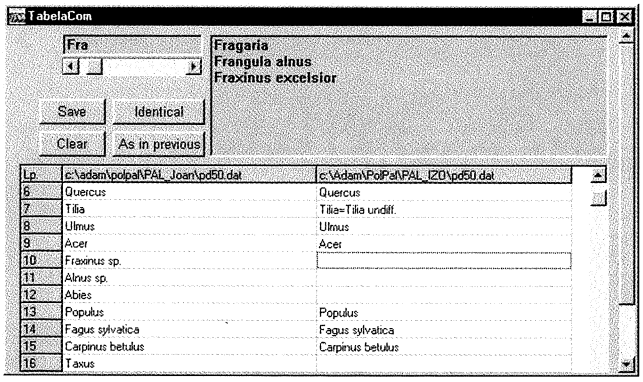

Fig. 2. Computer program, an element of the system POLPAL (Walanus & Nalepka, 1999; Nalepka & Walanus, 2003a) which aids the palynologist to adjust individual list of taxa to the main one. For taxon unrecognized by the program (in this figure: *Fraxinus* sp.), the correct name is to be taken from the list of similar names (here: *Fraxinus excelsior*)

about the age of the spectra, it is much more difficult for him/her to estimate the reliability of age estimation.

The question of location of sites makes no problems at all. Geographical coordinates, given down to arc minute's precision, are detailed enough for drawing maps of pollen percentages in the past. In Central Europe one minute, both in longitude and latitude, corresponds to 1 km distance. It is much less than any regional palaeomap can approach, in the sense of spatial resolution. The average distance between neighbouring sites is generally much larger. Also, the distance of pollen transport for trees is longer; some insect-pollinated herbs and, particularly, the so-called local taxa make some difficulty in interpretation of a palaeomap.

REGIONAL LIST OF TAXA

To integrate data derived from numerous pollen tables, each palynological taxon should be given the same meaning in each table. In principle, this is impossible because of the variety of site types and environments. The palaeobotanical "meaning" of genetically the same taxon can be different in mountains, uplands, and river valleys. From another point of view, however, even the simplest differences in names of the same taxon are to be avoided, if the computer database is to be used. An already mentioned example of three names for one taxon: *Betula t.* - *Betula type* - *Betula* indicates the simplest situation where personal decision is necessary.

Between the two extreme situations: that connected with deep palaeocological meaning, and that of typing conventions and errors, there is a large area of differences in individual opinions on some taxa. Different levels of taxonomical resolutions achieved by different palynologists also create some incompatibility of the taxa.

Anyway, in pollen tables to be integrated into one consistent database, only the taxa names from one generally accepted list of taxa may appear. Such "main" list of taxa of regional meaning is to be prepared by a board of experienced palynologists.

All individual lists of taxa are to be "translated" into the main one. In fact, the pollen tables are translated, because they contain taxa codes which have to be updated. In Fig. 2, the window of special computer program is shown, which is prepared to make translation as simple and straightforward as possible. The pollen table connected to the individual list of taxa is then translated to the main list of taxa. Identical taxa names are processed automatically (also synonyms are accepted). A "dictionary" created for previously done translation may be also applied. For other unidentified taxa, the list of proposals is displayed for personal decision.

TIME SLICE

The question of time scale for pollen tables (profiles) is a difficult one. Typically, profiles have only a few radiocarbon dates; sometimes no absolute dating has been made at all. Even ^{14}C dates are not free of error. One must stress that relative dating of profiles on the basis of comparison of pollen percentages with other dated profiles introduces additional error of another type. These errors are especially dangerous because they are correlated with pollen percentages to be finally plotted on the map. There are virtually no pollen profiles with all pollen spectra ^{14}C-dated. Hence, one must resolve the question of dating tens or hundreds of spectra from the profile with a few available ^{14}C dates only.

DEPTH-AGE RELATION

Dating of spectra is based on the depth-age relation, which is to be established on the basis of ^{14}C dates (Maher, 1992; Lotter & Tzedakis, 1998; Nalepka & Walanus, 2003b). The depth-age relation is plotted on the graph where vertical axis is used for depth and horizontal axis for age (Fig. 3). The line of depth-age relation unequivocally connects depth with age. In mathematical terms, it is a plot of the function *age=f(depth)*. It means that a sample from a given depth has one, strictly defined, however unknown,

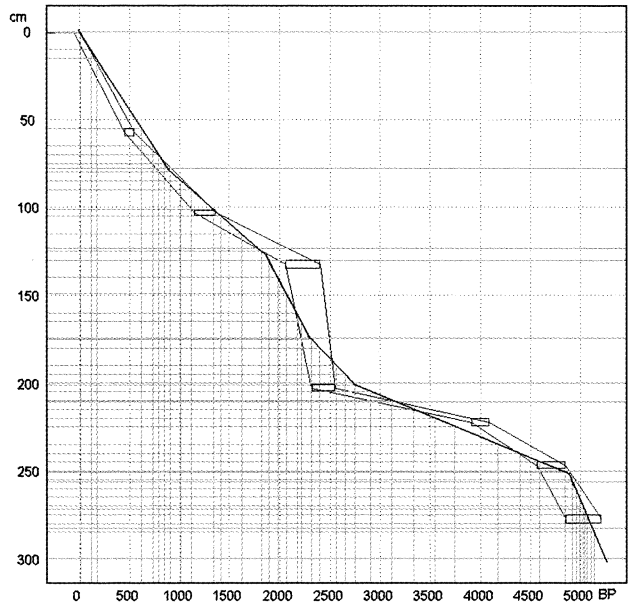

Fig. 3. Subjectively, hand plotted, age-depth relation with lithological boundaries taken into account. About 50 samples are "dated" according to the curve drawn. Radiocarbon dates are represented by rectangles of dimensions defined by dating error and sample thickness

age. The opposite is not true; in principle, it is possible to have two distant samples of equal age due to difficulties in coring or sediment disturbance.

INTERPOLATION

Two points on the depth-age graph, representing two radiocarbon dates, can be connected by a line (see curve 1 on Fig. 4). Using such a line, all spectra comprised between the ^{14}C-dated levels may be dated. A crucial assumption here is that the sedimentation rate is constant. Considering precision of dating of spectra by such interpolation, the error of radiocarbon age ($\pm\sigma$), as well as reliability of radiocarbon date and knowledge about sedimentation rate are to be taken into account.

Information about precision of ^{14}C dates is visualized on the plot by rectangles and lines connecting the right edges of rectangles ($T+\sigma$) and similar lines connecting the left edges ($T-\sigma$). If calibrated radiocarbon dates are used, the boundaries of 68% intervals can be used instead. The resulting band (Fig. 3) gives a good estimate of "real" statistical confidence intervals.

From the statistical point of view, if both neighbouring ^{14}C dates have similar errors, the spectrum in the middle of the levels of two dates is dated most precisely. However, taking into account variability of sedimentation rate, one may expect that the ages of samples placed far from those dated by ^{14}C (in the middle between the dates) are biased most seriously. As a result, a simple band defined by almost parallel lines represents the best first approximation of dating precision.

Any additional (not derived from ^{14}C dates) information about sedimentation rate should be taken into account in the interpolation. Let us assume that the sediment changes from one type to another somewhere between ^{14}C-dated samples, as illustrated by horizontal lines on the age-depth graph in Fig. 3. From the mathematical point of view, if the pollen analyst were able to assess quantitatively the relation of sedimentation rates in one sediment section (A) to that in the other (B), it would be easy for the computer to calculate and plot the line connecting the two dates, although being broken at the point of lithological change. For example, if sedimentation rate in section A is 2 mm/yr and that in section B 1 mm/yr, the slope of the fragment of line in section A should be twice as large as the slope of the fragment of line in section B.

However, basing on typical sediment description, even relative changes in sedimentation rate are difficult to be assessed, so the computer algorithms are not applicable. Therefore, we propose, that having confidence band plotted between the dates, and horizontal lines marking lithological boundaries, the palynologist may draw a broken line on the plot by hand or by mouse clicks (Fig. 3).

In case of continuous sediment changes, a smooth curve may be easily drawn. Subjective correction, in case of doubts concerning any particular ^{14}C date, may be also performed. For example, the line may pass not exactly through the radiocarbon dates but through the points moved right or left by 1σ or 2σ from the dates, which is absolutely allowed by the Gaussian distribution of true age (Walanus & Goslar, 2004).

EXTRAPOLATION

Two points representing dates can be connected by a straight line. This line may be plotted also outside the range defined by the dates, i.e. extrapolated to ages younger than the younger date and older than the older one. However, extrapolation far away from the dates may be very unreliable. From the statistical point of view, the error of extrapolation increases much with the distance to the nearest ^{14}C date. Moreover, from the lithological point of view, it may be doubtful to expect the same sedimentation rate outside the dated section. Summing up, extrapolation, i.e. assigning age to pollen spectra outside ^{14}C dates is possible to some extent; however, a reasonable extent cannot be assessed by statistical calculations due to the fuzzy geological factor of sedimentation rate.

MORE THAN TWO ^{14}C DATES IN THE PROFILE

Anyway, it is better to have many "absolute" dates. The third date enables verification of the hypothesis about constant sedimentation rate. On the other hand, verification (falsification) of radiocarbon date which does not agree with well documented assumption of constant sedimentation rate may also be reasonable.

Usually, ^{14}C dates are obtained for samples located at lithological boundaries. In such a case, the broken line passing through the data points would give the best interpolation.

In case when the three oldest dates are positioned along a straight line, extrapolation becomes more reasonable.

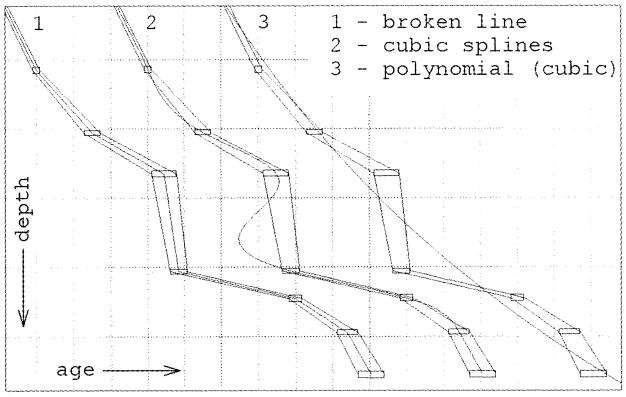

Fig. 4. Comparison of three "mathematical" lines modelling the depth-age relation in case of rather difficult data. Data are the same as in Fig. 3, where "by hand" model is presented

Even in the case of many ^{14}C dates in the profile, it is recommended to take into account the information concerning the lithology, especially in case of some disagreements between the dates. The "subjective" method of drawing a broken line by hand (by mouse clicks) is especially recommended in such a case.

In the case of hiatuses in the profile it would be recommended to treat two (or more) separated fragments independently, as separate profiles.

MATHEMATICAL MODELS

In general, for a given profile the unique depth-age relation does exist. Perhaps this is the reason why so many palaeobotanists expect substantial help from mathematical formulae. In principle it is a right attitude; however, the available models are far too simple to simulate the real world. Anyway, it may be worth to see how such simple models operate with real data.

Therefore, the algorithm to create the depth-age relation for a profile may be as follows:

1. Plot rectangles of ^{14}C dates on the depth-age graph and mark the depths of lithological boundaries.

2. Plot some mathematical functions which are available in the computer software, and supposed to fit the depth-age relation (Fig. 4). With criticism, try to draw some information from the plots obtained.

3. Plot one or more depth-age relations "by hand" (Fig. 3), and choose the best one.

In the software used to obtain the result presented on Fig. 3, three different mathematical models are available (Fig. 4).

a) *Broken line.* The consecutive points of ^{14}C dates are simply connected with straight lines. The drawback of this method is that the depth-age relation has "angles" at ^{14}C dates. However, if the dates are located at levels where changes of sedimentation rate are suggested by lithology, such angles may be justified. Extrapolation is not inherent to this method.

b) *Cubic splines.* Spline functions are smooth, in contrast to the previous option and, similarly as the broken line, the curve strictly passes through ^{14}C data. No extrapolation is possible. A serious drawback is that splines, trying to connect smoothly all dates, may give negative slopes in some parts of the depth-age relation. Such a slope means that deeper sediment is younger. In principle it is not unlikely, however, unreasonable. The spline method gives quite a "nice" smooth curve, but the fact that this curve passes exactly through ^{14}C dates should be treated rather as a drawback than an advantage.

c) *Polynomials (of any order, including straight line).* The maximum order of the polynomial to be fitted is equal to the number of ^{14}C dates minus 1. For three dates, the first order polynomial (straight line) may be fitted as well as the second order one. The latter one will exactly fit the three data points, but not necessarily the other parts of depth-age relation. The straight line would be really good approximation, especially in case of homogeneous sediment and good radiocarbon dates. Such a line, and any other polynomial, naturally provides extrapolation (which may be an advantage as well as a drawback). Polynomials of higher orders fit well to the dates, however, using orders higher than 3 rarely gives good results. Also, the second order polynomial is rarely useful because it is principally non-monotonical. Hence, the third-order polynomial or the straight line seems to be the best from the family of polynomials. The procedure of fitting polynomial to the data would take into account the different "weights" of points (^{14}C dates). It would be useful because ^{14}C dates display errors of different magnitudes. Moreover, it would be a good way for a pollen analyst to quantitatively introduce into calculations subjective doubts considering a radiocarbon date, by attributing low weight to that date.

PRECISION OF TIME SLICES

It is not obvious how many time slices for the Holocene are reasonable. In other words, it is not obvious which time span between time slices is acceptable. For the computer database, with all the pollen spectra dated, it is no problem to produce 100 or more time slices; the question, however, is the reliability of the result. Considering possible time resolution of the reconstruction of past vegetation one should take into account not only the uncertainty of spectra dating, but also the limited representativeness of sites. Since the precise assessment of error is hardly possible, only a general idea of time precision of the database taken as a whole suggests that 500 yrs between the time slices is a good choice. So, the accepted time slices would be: 500, 1000, 1500, ... yrs BP.

The value of the time span of 500 yrs gives the upper limit for the thickness of the time slice – it must be less or equal to 500 yrs. Using time "slices" as thick as the time span between the adjacent time slices is equivalent to the use of all the available spectra. However, such thick "slices" are hardly accepted by pollen data authors. On the other hand, it is mathematically possible to calculate pollen values for the exactly given date. If spectra are assumed to be precisely dated, the simple interpolation between two samples, for example, from 5,455 and 5,580 yrs BP may produce a supposed spectrum at, e.g., 5,500 yrs BP. In this ex-

ample, the 5,500 yrs "spectrum" will be made in 65% from the closer spectrum (5,455) and in 35% from the farther one (5,580).

The acceptable, intermediate solution is to take time slice of, say, 200 yrs thickness. It means that all spectra having dates within 5,500±100 yrs, i.e. older than 5,400 and younger than 5,600 will be taken into account. For example, in the series of 5,370, 5,455, 5,580, 5,715, 5,830 yrs BP there are exactly two spectra within that time slice, i.e. 5,455, and 5,580; and, at that moment they will be taken in equal proportions: 50% - 50%. It would, however, be better to use weighted average, in order to increase significance of the "close" spectrum of 5,501 and decrease significance of the "far" spectrum of 5,599 yrs, in calculation of the 5,500 spectrum. The result of weighing is smooth. However, simplicity and clarity are important when the calculation methods are to be understandable. Some improvement of the final result may be then sacrificed in order to achieve the goal of simplicity.

ISOPOLLEN MAP

The simplest way to present pollen data on the map, for a given time slice and for a given taxon, is to write down on the map the percentages at the places of sites. Such a method has no parameters; there are no doubts how to do it. Some problems arise, however, if there are two profiles of exactly the same geographical coordinates. Two profiles, taken for some reason 2 m one from another, should give only slight differences in pollen percentages (for a given time slice and a given taxon). They might be, for example, 15.8% and 21.7%. The difference between these numbers is only three times larger than pure statistical (Poisson) error. Such a difference is typical, and raises no concern. Probably, even in one profile, two or more spectra within one time slice would be of such or greater difference. If pollen spectra from one profile (table) are simply averaged over a time slice, we can also average percentages from two very close profiles.

Two assumptions are crucial for further considerations: (1) pollen percentages are not exact values – they have inherent errors and, (2) values of those errors cannot be precisely assessed. One important reason for these pessimistic assumptions is that the sites are more or less under local influences. Uncertainty of data is rarely taken into account in typical procedures for displaying data on maps, so standard computer applications may be of little use here.

The creation of a pollen map is equivalent to calculation of pollen percentage at any geographical position, on the basis of percentages at a number of sites, more or less uniformly distributed over an area. The simplest situation of two close sites may be treated using the average value. The average for 15.8% and 21.7% is 18.75%. If the distance between those sites is not 2 m but 2 km, the interpolation – instead of averaging – may be proposed. In the vicinity to the first site, 15.8% is accepted. Close to the second site, the value is 21.7%, and exactly in the middle we have to accept the value of 18.75%, at 0.5 km from the first site (1.5 km from the second one) the value will be 17.3%, and so on. The exact value of 20% will also be somewhere in between the sites, and mathematically **must** be there. Such an approach, however, is in contradiction to the mentioned imprecision of pollen percentage data.

Another approach can be proposed instead. Pollen percentage at any position (x°N y°E), arbitrarily chosen on the map, is to be calculated on the basis of all sites available. Of course, the sites close to (x,y) should have higher influence on the value at that point than the far sites. For the two sites discussed as an example such an approach gives:

1. average value (18.75%) if the sites are very close to each other,
2. average value exactly in the middle, if the sites are not very close to each other,
3. at the position of the first site (x_1,y_1), the value slightly higher than 15.8% because of the influence of the second site (of higher value, namely 21.7%).

In this approach, some weighting function has to be accepted for calculations. The weight (influence) of the site is maximum (equal to 1) if the distance from the point (x,y) to that site is zero, and has to decrease with distance (approaching zero at infinity).

Four different functions may be proposed. In principle, the number of possible mathematical functions is infinite. The proposed four functions are very simple, and they seem to cover well the necessary variety of shapes.

These functions have a scale parameter (s), which decides how fast the given function decreases with a distance. The parameter is a distance (in km) itself. The function and the value of the parameter are to be chosen to define weighting.

Function (1) is simply an inversion of distance, however with 1 added in denominator. The function $w = 1/x$ would give infinitely large weight exactly at the place of a site (where $x = 0$), which, despite of numerical difficulties, does not agree with the assumption that percentage value of the given site has no absolute meaning and has to be averaged with neighbouring sites. Actual formula (see Table 1) of function (1) gives maximum weight equal to 1, and weight decreasing with a distance. That function, decreasing relatively fast in the vicinity of a given place decreases very slowly for longer distances. That is the reason for little use of it. Function (2) is free of such a drawback. Squaring distance gives much faster decrease of weight for long distances. One byproduct of squaring is a flat shape for short distances, i.e. the weight is maximum at a given place and almost the same in the vicinity. The most important feature of the Gaussian function (3) is that if falls to zero very sharply having, at the same moment, some (bell-shaped) flat area at small distances. Function (4) is the simplest exponential function. The characteristic feature of exponential function is that its value decreases e-times (almost three times; $e = 2.718...$) with each step of s kilometers. For example, while for function (1) a move from x = 10s to 11s decreases the weight only negligibly (from 0.09 to 0.08), for exponential function (2), the weight still decreases by a factor of three (from 0.000045 to 0.000016).

Examining the plots in Fig. 5 is not the only way to decide which function and which value of parameter to chose. A more convincing method would be by examining the final

Table 1

Four weighting functions (see Fig. 1). Weight (w) is a function of distance (x), with scaling parameter (s)

No	Name	Symbol in Fig. 4	Formula	Weight at x=0	Weight at x=s/2	Weight at x=s	Weight at x=2s
1	Inverted distance	dist	$w = 1/(1+x/s)$	1	0.67	0.5	0.33
2	Invereted squared distance	dist2	$w = 1/(1+(x/s)^2)$	1	0.80	0.5	0.2
3	Gaussian (bell curve)	Gauss	$w = \exp(-(x/s)^2/2)$	1	0.88	0.61	0.13
4	Expotential	Exp	$w = \exp(-x/s)$	1	0.61	0.37	0.13

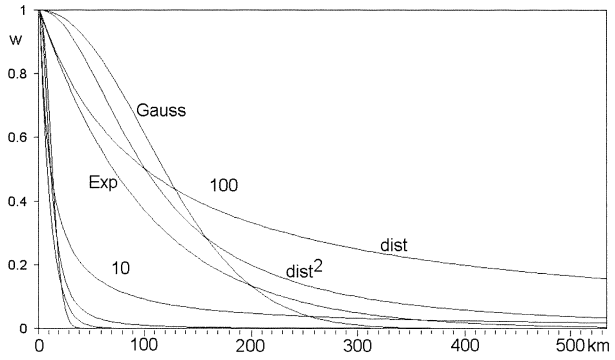

Fig. 5. Weighting functions. Weight (w) of pollen per cent from a given site is dependent on distance (x – km) to a given point on the map. Four functions (Gaussian, exponential, inverse distance, and squared inverse distance) are presented. Two groups of plots are obtained for two values of scaling parameter: s = 10 km and s = 100 km

result, i.e. the map. However, the latter approach has a serious drawback; one would adjust the method in order to obtain the demanded result. On the other hand, it is not easy to find objective, theoretical criteria for choosing the type of function and the value of parameter. Some consideration about nature of pollen dispersion may be relevant here. For example, pollen flux transported by strong wind diminishes approximately with inversed distance, which might be treated as an indication for the first function.

Some features of different weighting functions are visible on Fig. 6. First of all, the maps in the upper row give isolines of very low values, almost down to 0% (six sites show 0%), whereas the lower maps give no such picture. The upper maps are obtained with smaller scaling factors, which means that there is a weak influence of distant sites. The example of map *d2 10* is especially clear. Each 0% is surrounded by its own isolines. Such a picture, however, does not generalize pollen data from the sites. It is a picture of site distribution rather than that of palaeo-pollen percentage in the area of Poland (see Fig. 7 obtained for real data). From the methodological point of view, the Gaussian weighting function is better recommended because good generalization is obtained; for both values of parameters (*Ga 50* and *Ga 100*) simply the lack of pollen in the country middle is obtained. However, the rectangular shape of the "hole" for *Ga 50* evidently depends on geographical positions of the sites. Interesting is the map *ex 50*; whereupon the influence of "additional" westernmost sites is visible.

VERIFICATION OF DATA AND MAP; REMOVED RESIDUALS AND BOOTSTRAP TECHNIQUE

Isopollen maps may have two goals: to visualize data or to visualize the state of knowledge on the pollen rain in the past (based on the data). The first goal: data visualization, is easily obtained with low values of the scaling parameter (like, e.g., on the map *d2 10* in Fig. 8). In this approach the sites can be easily distinguished, and isolines operate simply as a graphical element. Percentage values at each site (like on Fig. 8) may be added to make the picture unequivocal.

In the second approach, if the image of geographical distribution of past vegetation is the final product, visualization of individual sites seems to be illogical. First of all, the % values should not be present on the map, and the scaling parameter (s) should be large enough to "smooth out" individual sites. In fact, scaling factor has to be adjusted to the average distance between the sites, and the sites have to be more or less uniformly distributed on the area under consideration. The second requirement is unrealistic; the first one is difficult to accept because high value of scaling factor means smoothing over large area, what results in loosing of, may be interesting, fine features of the isolines.

Extremely high value of the scaling factor (for example s=1,000 km) will give uniform, average percentage all over the area. In most cases, it would not be the true picture of palaeo-pollen rain (if reasonable number of sites is available). On the other hand, very low value of scaling factor gives illustration of site distribution rather than that of geographical pattern of vegetation. Since the site percentages are far from absolute precision and are not absolutely representative, such a picture does not represent palaeovegetation. Hence, to produce a good isopollen map the scaling parameter must be chosen in between those two extremes.

It is impossible to obtain isolines of finer structure than that defined by distances between the sites. Starting from that point, resolution of isolines exactly equal to that defined by site distribution should be treated with low confidence. In other words: if one site placed among other sites has pollen percentages much different from those in the surrounding sites, it was probably under local influences and

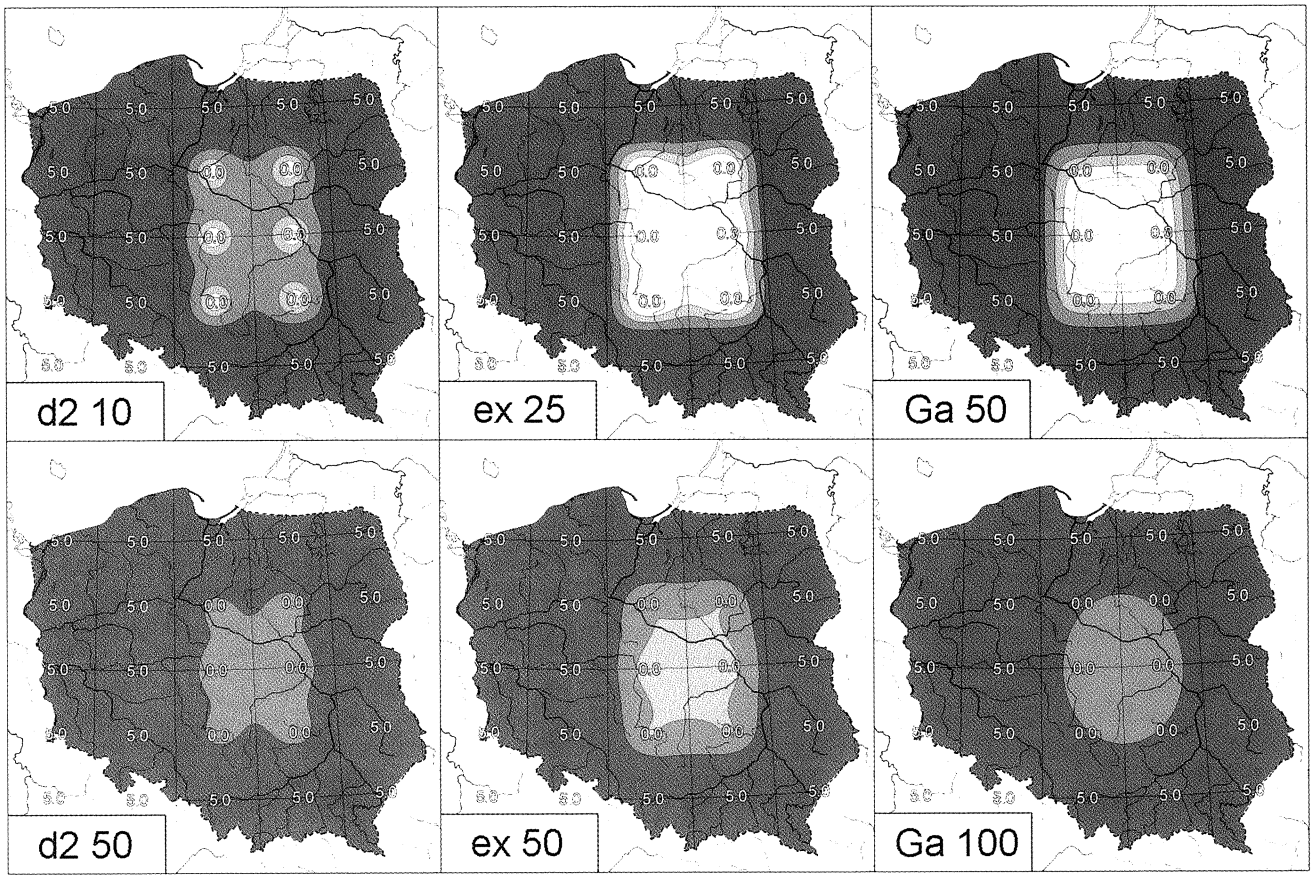

Fig. 6. Maps obtained for artificial data. The analyzed sites are equidistant in the sense of geographic coordinates. Six neighbouring sites in the middle of Poland are of 0%, while the rest of sites are of 5%. Three weighting functions are applied: d^2 –inversed squared distance, exponential, and Gaussian (inversed distance gives picture with almost no structure, too smooth). Two scaling factors are applied to each function (expressed in kilometres, below the maps)

then it should be removed or, perhaps, its dating should be revised.

This idea is applied to a simple algorithm of "removed residuals". For each site the map is produced with the given site removed from the data. If the percentage obtained at the location of the site differs much from the value actually carried by that site (for a given taxon or for all taxa), this site is to be revised. However, the meaning of the term "differ much" is out of any statistical reasoning. Too many factors derived from different levels of abstraction are involved. In practice, the computer program gives two numbers at the location of each site; that corresponding to the given site, and the weighted average calculated basing on percentages of all other sites.

Differences between the two mentioned types of percentages may be averaged (in their absolute values, or their squares) over all the sites, and the resulting value (divided by the number of sites) is an estimate of map reliability.

Another approach to the question of map reliability is the use of a bootstrap technique. The idea of bootstrap is simple, however, people who treat data too literally would have difficulties to accept it. A very good method of appreciation of reliability of any result is to get new data and cross-check the old result. Since it is frequently impossible to obtain new data, the same way with other direction is applicable. In any case it is possible to remove, randomly cho-

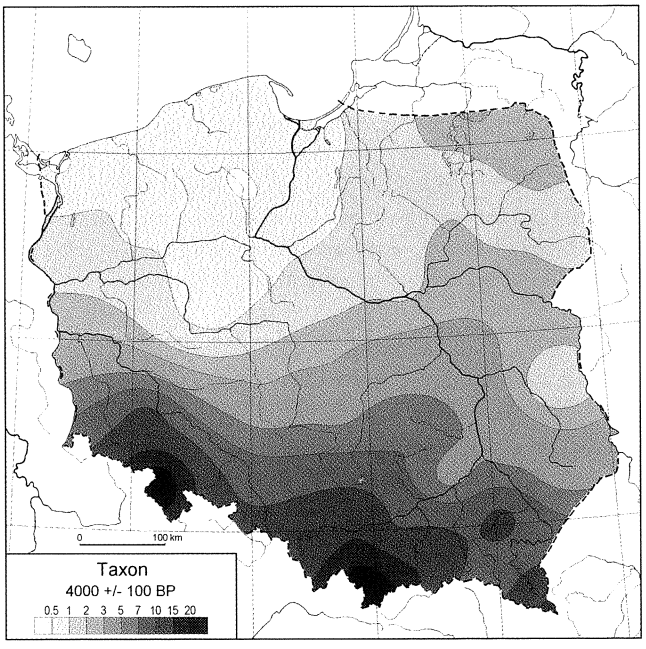

Fig. 7. An example of isopollen map. Weighting function d^2 (inversed squared distance) has been applied, with scaling parameter equal to 50 km. Influence of single sites is visible in the south-eastern region. For more maps see, for example, Ralska- Jasiewiczowa *et al.* (2003)

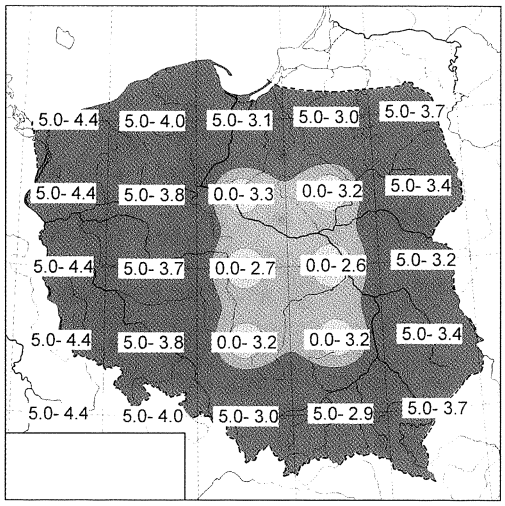

Fig. 8. Isopollen map with "removed residuals", obtained for artificial data comprised of sites with 0% and 5%. For each site the value at a given site is shown, as well as the value calculated on the basis of all other sites (with the given site temporarily removed from the data). The weighting function *d2-10* has been used here. The largest removed residual (0.0–3.3 = –3.3) is obtained for the 0% site, which is almost completely surrounded by the 5% sites

sen (!) one third, or one fourth of the data (or sites). If such a bootstrap-type map (with less sites) differs much from the full-data map or the other bootstrap maps (from other tossing), the map (full-data) has to be treated as poorly representing the reality. The estimate of palaeo-pollen percentage should not depend deeply on site availability.

MAP OF MIGRATION OF VEGETATION

The earlier accepted definition of the time slice (the slices are taken with 500 yrs interval and the slice width is 200 yrs) means that less than half of the pollen data are used. It is reasonable to plot isopollen maps at every 100 yrs, with the slice width ±100 yrs. Then, the neighbouring maps would be partly based on the same pollen spectra; as a result, the neighbouring maps would be to some extent similar. It would be easy, using weighting averages (in time domain), to obtain maps changing absolutely smoothly if plotted, for example, in 10 yrs time resolution. For the entire Holocene it would give 1,000 maps for one taxon; what is enough to obtain one minute of a standard video film.

Anyway, it may be instructive to see all the Holocene vegetation history compressed into a few minutes. However, the idea of assessment the reliability of the maps seems more important. In each map, many subjective decisions are involved (from profile dating to weighting function). Hence, subjective impression after examining a series of, say, 5 maps per 500 yrs may result in changing mind toward more safe (higher) values of the scaling factor of weighting function.

Watching a film gives immediate impression of how things are changing (DOS trials; Walanus, 1995). However, if the maps constituting the film are reliable, there exists simple mathematical operation of transformation of movement into static picture. Having two maps of, say, 100 yrs time distance, the rate of change of pollen percentage may be calculated at each point of the area. The result also has a form of a map, however, isolines are not labelled with [%], but with [%/yr] or [%/100yr]. Since negative values are possible here, the zero isoline may exist on the map, and it may be very informative.

The map of the rate of change of pollen percentages would give immediate picture of vegetation migration. The area showing positive values (increasing percentage) is that where the taxon immigrates. However, it is a question of interpretation where the taxon comes from; probably from the neighbouring area of zero rate of change, i.e. wherein the taxon is stabilized.

CONCLUSIONS

The principal goal of the analytical steps described here is to obtain good, informative maps, like isopollen maps shown on Fig. 7. It is, of course, difficult to fulfill this task, because the definition of which map is good cannot be precise. The main source of vagueness is not the obvious existence of measurement errors in the data, but difficulties in their appreciation.

Subjective decisions are necessary at every level of scientific research: from the level of pollen grain recognition, which can be relatively objective, through the use of radiocarbon dates, to the level of isopollen map parameters. Using the example of pollen grain, while the personal decision is inevitable in considering the individual grain, it is to be avoided in case of maps to consider them individually. The use of computer not only shortens the time necessary for map creation; but, what is more important, moves subjective decisions from the level of individual maps to the level of all the maps taken as a whole. It is, in principle, impossible to change or correct isolines on a given map, what used to be normal in time of "by hand" drawn isolines.

It is the choice of weighting function and scale parameter that mainly defines the final appearance of isopollen maps, including also boundaries of percentages and palette of colours. These parameters are selected to maximize the amount of information on the map, in opposition to the random noise which is to be averaged off. The information concerns the percentage of pollen in the area of Poland, and – from such a point of view – even the site distribution is to be treated as random.

Having a well-organized data base (filled with the data) and the algorithm of map construction it is, in principle, one mouse click necessary to obtain hundreds of maps for many taxa and many time slices. The palaeobotanical interpretation of the resulting maps is a different matter (Ralska-Jasiewiczowa, in print). However, both parts cannot and should never be separated.

REFERENCES

Hoek, W., 1997a. Palaeogeography of Lateglacial Vegetations. Aspects of Lateglacial and Early Holocene vegetation, abiotic landscape, and climate in The Netherlands. Ph.D. Thesis, Free

University Amsterdam. *Netherlands Geographical Studies*, 230: 1– 147.

Hoek, W., 1997b. Atlas to Palaeogeography of Lateglacial Vegetations. Maps of Lateglacial and Early Holocene landscape and vegetation in the Netherlands, with an extensive review of available palynological data. Ph.D. Thesis, Free University Amsterdam. *Netherlands Geographical Studies*, 231: 1–165.

Huntley, B. & Birks, H. J. B., 1983. *An atlas of past and present pollen maps for Europe: 0-13,000 years ago*. Cambridge Univ. Press, Cambridge, 667 pp.

Lotter, A. & Tzedakis, C., 1998. Chronologies. European Pollen Database – Newsletter, 8. (*www.ngdc.noaa.gov/paleo/epd/letter8/*).

Maher, L. J., Jr., 1992. Depth-Age Conversion of Pollen Data. *INQUA Commission for the Study of the Holocene, Working Group on Data - Handling Methods, Newsletter*, 7: 13–17.

Nalepka, D. & Walanus, A., 2003a. Data processing in pollen analysis. *Acta Palaeobotanica*, 43: 125–134.

Nalepka, D. & Walanus, A., 2003b. Age determination of individual spectra in pollen diagram based on small number of radiocabon dates. (In Polish, English summary). In: Zastawniak, E. (Ed.), *Paleobotanika na przełomie wieków. Botanical Guidebook*, 26: 295–307.

Ralska-Jasiewiczowa, M., 1983. Isopollen maps for Poland: 0-11,000 years BP. *New Phytologist*, 94: 133–175.

Ralska-Jasiewiczowa, M. (Ed.), in print. *Late Glacial and Holocene history of vegetation in Poland based on isopollen maps*. W. Szafer Institute of Botany, Polish Academy of Sciences, Kraków.

Ralska-Jasiewiczowa, M., Nalepka, D. & Goslar, T., 2003. Some problems connected with the vegetation development during the oligocratic/Homo sapiens phase of Holocene interglacial in central Europe. *Vegetation History and Archaeobotany*, 12: 233–247.

Szafer, W., 1935. The significance of isopollen lines for the investigation of the geographical distribiution of trees in the Post-Glacial period. *Bulletin de l'Academie des Sciences et des Lettres*, Série B, 1: 235–239.

Walanus, A., 1995. Pollen data in space and time – local approach. *INQUA Comission for the Study of the Holocene, Working Group on Data-Handling Methods, Newsletter*, 13: 13.

Walanus, A. & Goslar, T., 2004. *Wyznaczanie wieku metodą ^{14}C dla archeologów*. (In Polish). Wydawnictwa Uniwersytetu Rzeszowskiego, Rzeszów, 120 pp.

Walanus, A. & Nalepka, D., 1996. Palynological profiles synchronization tools in the computer program POLPAL. (In Polish, English summary). *Zeszyty Naukowe Politechniki Śląskiej, seria Matematyka-Fizyka 79, Geochronometria*, 13: 215–226.

Walanus, A. & Nalepka, D., 1999. POLPAL. Program for counting pollen grains, diagrams plotting and numerical analysis. *Acta Palaeobotanica, Suppl.*, 2: 659–661.

Streszczenie

SCALANIE PÓŹNOGLACJALNYCH I HOLOCEŃSKICH DANYCH PALINOLOGICZNYCH Z OBSZARU POLSKI

Adam Walanus & Dorota Nalepka

Typową, podstawową jednostką danych palinologicznych, o których mowa w tytule, jest tabela, czyli arkusz liczb. Wiersze tabeli odpowiadają próbkom pobranym z profilu z głębokości zapisanej w pierwszej kolumnie, natomiast kolumny to poszczególne taksony palinologiczne oznaczone w profilu. Właściwą, liczbową zawartością tabeli są zliczenia ziaren pyłku danego taksonu w próbce z danej głębokości (Fig. 1).

W celu wykorzystania dostępnych dla obszaru Polski tabel do otrzymania pewnego, regionalnego obrazu roślinności w przeszłości konieczne jest ujednolicenie danych. Pierwszym, stosunkowo łatwym krokiem jest ujednolicenie nazewnictwa taksonów palinologicznych dla wszystkich tabel, pochodzących oczywiście od różnych autorów. Warto zwrócić uwagę, że nie jest to tylko kwestia słownictwa, gdyż ziarna pyłku były (i są) oznaczane z różną rozdzielczością taksonomiczną (Fig. 2).

Drugim etapem jest datowanie próbek, czyli przypisanie wierszom tabeli, opisanym głębokością w konkretnym profilu, wieku. Pozostając przy wieku radiowęglowym (konwencjonalnym ^{14}C BP) wyznacza się krzywą głębokość-wiek na podstawie kilku dat ^{14}C. Niestety dla niektórych, cennych palinologicznie profili, z powodu braku dat radiometrycznych zachodzi konieczność posłużenia się porównawczymi "datowaniami" palinologicznymi, wykorzystującymi sąsiednie profile datowane ^{14}C. Krzywą głębokość-wiek wyznaczano korzystając z kilku dopasowań matematycznych (łamana, funkcje sklejane trzeciego stopnia, wielomiany różnych stopni – najlepiej trzeciego; por. Fig. 4). Ponadto, w celu uwzględnienia dodatkowej wiedzy o litologii profilu, uwzględnia się możliwość "ręcznego" poprowadzenia krzywej głębokość-wiek (Fig. 3).

Mając wydatowane spektra palinologiczne można przystąpić do scalania regionalnego dla danego wieku. Trzeba jednak zdać sobie sprawę, iż ze względu na ograniczoną, a ponadto trudną do oszacowania dokładność tego datowania, nie ma sensu pobieranie z tabel danych dla ściśle określonego wieku (np. 2500 BP), choć byłoby to możliwe za pomocą interpolacji. "Cięcia czasowe" powinny mieć pewną szerokość, np. 200 lat, tak, że dla wieku np. 2500 BP pobiera się z tabel spektra, których daty mieszczą się w przedziale 2500±100 lat BP. Oczywiście celowe byłoby zastosowanie ważenia płynnie zmniejszającego znaczenie spektrów odległych od środka przedziału. Szerokość "cięć czasowych" implikuje rozdzielczość czasową metody. Dla danego cięcia czasowego i danego taksonu palinologicznego otrzymuje się z bazy danych sieć punktów na mapie Polski, charakteryzujących się wynikowym procentem pyłku danego taksonu. Oczywiście, nie każde stanowisko ma spektra w każdym cięciu czasowym. Pojedyncze punkty z reguły pochodzą z więcej niż jednego spektrum, z danego profilu.

Dokładność otrzymanych wartości procentu pyłku jest ograniczona. Jest to skutkiem następujących czynników: (1) naturalnego, poissonowskiego rozrzutu liczby zaobserwowanych ziaren, (2) niepewności datowania spektrów, co powoduje, że do "cięcia czasowego" wejść mogły spektra o rzeczywistym wieku nie mieszczącym się w nim, a także (3) lokalnego charakteru danych palinologicznych, np. pyłkowych wskaźników antropogenicznych, podczas gdy punkt reprezentować musi duży obszar (determinowany odległością do sąsiednich stanowisk). Trudna do oszacowania, lecz na pewno stosunkowo niska dokładność liczb, które mają być bezpośrednio podstawą rysowania map izopolowych powoduje, że zastosować trzeba algorytm uśredniający w pewnym stopniu wartości z sąsiednich stanowisk. Nie ma sensu upieranie się, że w miejscu geograficznym danego stanowiska (i jego najbliższym sąsiedztwie) wartość jest dokładnie taka, jaka wynika z danych pochodzących z tego jedynego profilu.

Tak więc, wartość procentu pyłku w dowolnym punkcie obszaru Polski otrzymuje się jako średnią ważoną z wszystkich stanowisk, przy czym waga jest odwrotnie proporcjonalna do odległości wybranego punktu od danego stanowiska. Odwrotną proporcjonalność należy tu traktować symbolicznie: chodzi o to, że waga

maleje z odległością. Konkretna zależność funkcyjna wagi od odległości zależy od postaci funkcji i parametru skalującego odległość. Jako funkcję wagową wybrać można różne funkcje matematyczne (Fig. 5–7); w przypadku prezentowanych tu map zastosowano odwrotność kwadratu odległości (z polem martwym, dla uniknięcia nieskończoności przy zerowej odległości). Parametr skali zależy przede wszystkim od gęstości stanowisk na obszarze Polski. O doborze funkcji wagowej decyduje wygląd map; ich zgodność z intuicją. Podkreślić trzeba, że warto zdecydować się na jedną funkcję wagową dla wszystkich taksonów i wszystkich cięć czasowych, co eliminuje zarzut subiektywizmu. O ile dałoby się manipulując funkcją wagową modyfikować mapę według subiektywnych oczekiwań, to już druga mapa będzie niezależna, a wszystkich map są setki.

Podobnie subiektywny – w pewnym stopniu – jak wybór funkcji wagowej jest wybór granic procentów pyłku dla izolinii, pomiędzy którymi obszar wypełniany jest jednolitym kolorem. Czynnik ten ma dość duży wpływ na końcowy obraz. Zbyt gęste izolinie, podobnie jak zbyt mała skala przestrzennego uśredniania danych powodują, że na mapie zaznaczają się poszczególne stanowiska, co nie jest właściwe, gdyż zamiast obrazu roślinności otrzymuje się raczej obraz rozmieszczenia stanowisk.

Zaproponować można dwie metody weryfikacji danych. Jedna z tych metod dotyczy konfrontowania danych z danego stanowiska ze stanowiskami sąsiednimi. Otóż dla punktu na mapie, w którym znajduje się kolejne stanowisko oblicza się w normalny sposób procent pyłku, lecz nie uwzględniając procentu pyłku z tego stanowiska (Fig. 8). Porównanie tak otrzymanej (średniej) wartości z konkretną wartością z wyłączonego stanowiska może, w przypadku dużej różnicy, sugerować zrezygnowanie ze stanowiska, ze względu na prawdopodobne błędy.

Inna metoda weryfikacji istotności otrzymanego obrazu izopolowego na mapach bazuje na metodzie bootstrap. Polega ona na losowym usuwaniu części danych (stanowisk) i otrzymywaniu w ten sposób serii map. Jeżeli obraz na mapach nie różni się zbytnio i zachowana jest podstawowa struktura, to strukturę tę można uznać za istotnie odzwierciedlającą przyrodniczą rzeczywistość. W przeciwnym razie stwierdzić można, że w danych w przeważającej mierze zawarty jest raczej losowy szum (stanowiska dla danego taksonu mają na przykład charakter zbyt lokalny, by rysować mapę Polski).