4.9. STATISTICAL METHODS IN THE INTERPRETATION OF THE ANALYTIC RESULTS

*Adam Walanus**

The core G1/87, taken from the deepest part of Lake Gościąż, is 18.6 m long. It comprises almost 13,000 varves. Samples have been taken from the core for many kinds of analyses. Every 50 yr a 10-yr-thick sample of 1 cm^2 surface area was collected. Some parts of the core were sampled more densely. As a result almost 300 samples, more or less equally distributed in time, have been analysed. Analyses of loss on ignition, $CaCO_3$, Fe_2O_3 (Wicik, Chapter 4.3), stable isotopes (Kuc et al., Chapter 4.4), as well as pollen (Ralska-Jasiewiczowa & van Geel, Chapter 4.6.1) have been made. The average values of the varve thickness and thickness ratio of light to (light+dark) laminae were also taken into calculation.

The simplest parameter to be calculated for statistical analysis is the correlation coefficient for pairs of different measurements. However, since there are 16 variables, as many as $15 \cdot 16/2 = 120$ pairs of variables are to be considered. At the first step, to reduce the number of correlation coefficients (r), the average of r's was calculated. However, because r may be positive or negative, and both indicate dependence between variables, the absolute values $|r|$ were averaged (the r's are not too high, and the transformation of r into the normal variable to improve linearity in averaging was not necessary). The average values $|r|_{av}$ were calculated for some groups of measurements.

Some coarse observation of a quantitative nature is that the values of $|r|_{av}$ calculated for measurements performed in one laboratory are in general higher than those obtained for measurements in different institutes. The average of averages $|r|_{av}$ is 0.4, while the value for completely uncorrelated variables is 0.19 (not 0 because of absolute values; the figure 0.19 is from a Monte Carlo experiment). For example, the correlation between $CaCO_3$ measured in two laboratories is not as high as would be expected, and $|r|_{av}$ for some group of measurements was significantly lower than for other groups. The second case resulted in finding an important gross error.

In the Lake Gościąż profile, rapid changes of climate are reflected. The value of $|r|_{av}$, when calculated for the whole profile, shows the response of measured quantities to those changes. So one may expect that $|r|_{av}$ calculated

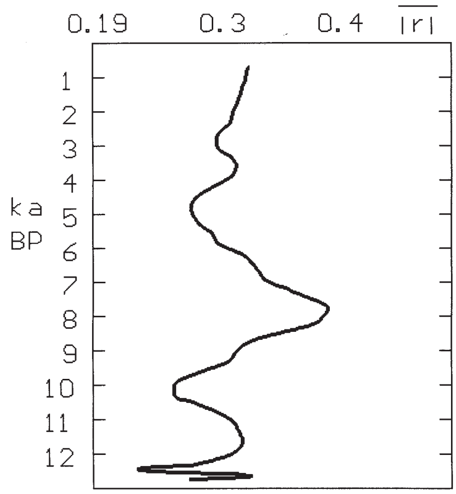

Fig. 4.4. The curve of absolute average correlation coefficient $|r|_{av}$ calculated for 8 variables, within a moving window of 30 samples width. It may be interpreted as an indicator of stable (low $|r|_{av}$) or unstable (high $|r|_{av}$) conditions. The value $|r|_{av} = 0.19$ is minimal, expected for independent variables.

for shorter fragments of the profile would be lower because of avoidance of jumps like that of the YD/PB transition. However, other mechanisms should be taken into consideration. With climate changes, even the sign of correlation of some pairs of variables may reverse. For example, some variables may be positively correlated to dry conditions and negatively under atlantic influences, and the resulting r would be close to zero. It seems reasonable to calculate r's in fragments of the profile.

The plot of $|r|_{av}$ calculated for 8 variables within the moving window covering 30 samples is given in Fig. 4.4. The variables are loss on ignition, Fe_2O_3, $CaCO_3$ (two series from different laboratories), $\delta^{18}O$, $\delta^{13}C$, varve thickness, and ratio of laminae thicknesses (light/total). A possible interpretation of the plot is that the high values indicate periods of high variability of sedimentation conditions, whereas the low values indicate low periods. If $|r|_{av}$ is close to 0.19 (the value expected for independent variables), the dominance of measurement errors over real signal would be suspected. The high value of $|r|_{av}$ around the Pleistocene/Holocene transition agrees with the interpretation of high variability of climatic conditions. The most distinctive peak of $|r|_{av}$ is around 8 ka cal BP.

Correlation between measured variables may be subject to changes in time. During climate changes, even the sign of the correlation may change. It is visible on Fig. 4.5, where correlations between 10 variables are given. For every pair, r has been calculated separately within 7 chronozones. In such a way 45 short diagrams were obtained. Division of the profile in chronozones is independent of the data analysed. It is based on the varve chronology. Equally likely the division into millennia could be applied to show how the correlation changes in time.

* The analysis presented was possible thanks to the access to data given by the authors, namely M. Ralska-Jasiewiczowa and B. van Geel – pollen counts, T. Goslar – laminae thickness and chronology, T. Kuc and K. Różański – stable isotopes, B. Wicik – geochemical data. Guidance in palaeoscience and botany of M. Ralska-Jasiewiczowa and discussions with T. Goslar were very important factors in the development of the work presented. The author is much obliged to T. Goslar for the critical revision of the text. Financial support was obtained from the Committee for Scientific Research (Project No. 602529101).

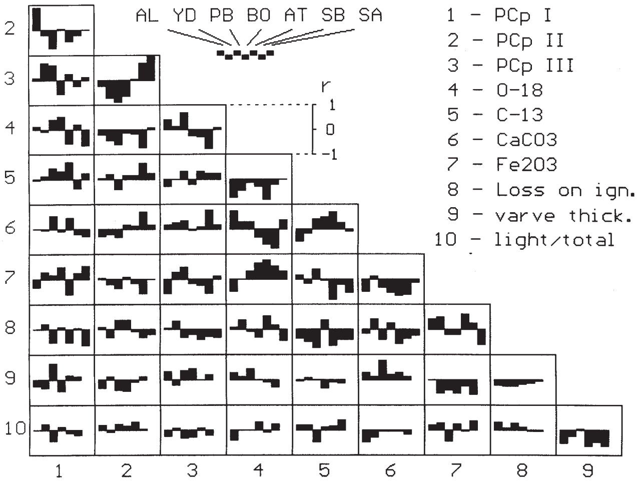

Fig. 4.5. Correlations between 10 variables calculated within the chronozones. In every box, the plot of correlation coefficients for several chronozones is given. "PCp" indicate three principal components of the palynological data. Loadings to the PCp's are given in Fig. 4.6.

The palynological data were also included. However, because of so many pollen taxa (i.e. variables) that should be included, a principal components analysis has been made. The first three principal components (PCp) were used in calculations of r. The meaning of PCp I, II, and III is given by the "loadings" obtained (Fig. 4.6). The really most important (63% of the total information) is the first PCp, with positive participation of *Alnus, Quercus, Corylus, Ulmus, Tilia, Fraxinus,* and *Carpinus* and negative of *Betula, Pinus, Juniperus,* and generally herbs. The second and third PCp's represent, as usual, not such a clear pattern. It must be mentioned that PCp-analysis has been performed for the whole profile, not separately within chronozones. Thus loadings are reflections of general correlations among taxa observed for the Late-Glacial and Holocene. Principal components are completely uncorrelated variables; it is the main feature of the analysis. However, if correlation is calculated only in a fragment of the profile, non-zero values may appear. A good example is the box 3–2 in Fig. 4.5. A surprisingly clear pattern of r's is obtained for the II and the III PCp, both of which carry only 20% of the variance. Positive correlation in SB and SA is probably produced by *Carpinus* and *Corylus*, and earlier negative correlations came from the opposite loadings to II and III PCp of *Betula, Juniperus,* and *Artemisia*. Very high positive correlation between I and II PCp in Allerød results from the lack of other taxa and from loadings of participants other than *Betula*.

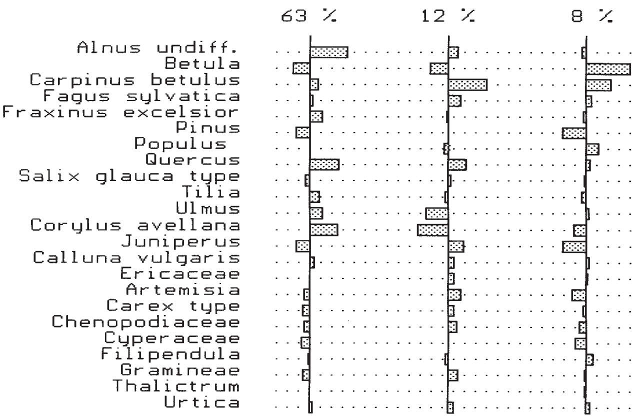

Fig. 4.6. Loadings to the first three principal components of the palynological data, according to the standard PC analysis, performed for all the samples at once (not within the chronozones). Principal components were used in the calculation of correlation coefficients; see Fig. 4.5. Percentages of the total variance carried by PC's are given.

A few clear patterns may be found it Fig. 4.5. The best one is connected with the relation between Fe_2O_3 (column 7) and varve thickness (row 9). The correlation is negative, with quite a high absolute value, but Allerød makes an exception, showing no dependence between variables. It should be emphasized that the interpretation of the value of r is more valid if some comparison to the other r's is possible. For example, correlation of varve thickness with loss on ignition is very small, and only the fact that it is permanently negative seems to indicate real dependence. Another interesting correlation is that of Fe_2O_3 and $\delta^{18}O$. The plot is smooth, with a significant negative correlation in Allerød and the maximum positive correlation in the Atlantic period. The correlation between the stable isotopes (row 5 – column 4) changes smoothly in time, probably also indicating real dependencies. Interesting is the diagram 6–1, i.e. the main variability of pollen data correlated with $CaCO_3$.

REFERENCES

Aitchinson T. C., Leese M., Michczyńska D. J., Mook W. G., Otlet R. L., Ottaway B., Pazdur M. F., van der Plicht J., Reimer P., Robinson S. W., Scott E. M., Stuiver M. & Weninger B. 1989. A comparison of methods used for the calibration of radiocarbon dates. *Radiocarbon* 31: 846–864.

Arnold M., Bard E., Maurice P., Valladas H. & Duplessy J.-C. 1989. C-14 dating with the Gif-sur-Yvette Tanderon accelerator: status report and study of isotopic fractionations in the sputter ion source. *Radiocarbon* 31: 191–199.

Berglund B. E. & Ralska-Jasiewiczowa M. 1986. Pollen analysis and pollen diagrams. In: B. E. Berglund (ed.), *Handbook of Holocene Palaeoecology and Palaeohydrology*, pp. 455–484. J. Wiley & Sons. Chichester.

Calibration 1993. In: A. Long, M. Stuiver & R. E. Kra (eds), *Calibration issue, Radiocarbon* 35: 1–244.

Craig H. 1957. Isotopic standards for carbon and oxygen and correction factor for mass-spectrometric analysis of carbon dioxide. *Geochimica Cosmochimica Acta* 12: 133–149.

Demske D. 1995. Development of the local environment at Lake Gościąż (central Poland) during Late Vistulian and Holocene: vegetation, hydrological changes and influence of man. Ph. D. Thesis, W. Szafer Institute of Botany, Polish Academy of Sciences, Cracow.

Faegri K. & Iversen J. 1975. Textbook of pollen analysis. Munksgaard, Copenhagen.