

SOME PROBLEMS OF REPRESENTATIVITY IN POLLEN ANALYSIS

KNUT FAEGRI

Universitete i Bergen, Botanisk Museum, Norway

ABSTRACT

Three problems of representativity are discussed: A. The classical "statistical" problem to what extent one may conclude from the composition of a pollen spectrum to that of the actual pollen rain. Statistical treatment of this problem chiefly serves as a warning not to draw too definite conclusions from individual spectra. B. To what extent it is possible to conclude from the composition of the pollen rain to that of the vegetation. Doubt is raised about the validity of Davis's negative conclusions which are statistically not well founded. Pollen productivity alone does not give the answer: transport effectivity is equally important. C. To what extent does the actual vegetation represent the potentialities of the area and to what extent are successions dependent on non-climatic factors.

"**P**OLLEN statistics" was the term used (if not exclusively) in the classical literature on what we today call pollen analysis. From a very strict semantical point of view neither term is adequate, but the older one had one advantage lost by the general adoption of the current designation: it is a memento of the fact that statistical evaluations are an integral part of our work. Most of these statistical evaluations are carried out sub-consciously and/or as a routine during the daily work with pollen spectra and diagrams, and thus are subject to all those dangers and fallacies inherent in this kind of procedure.

However, although such problems are usually — as I have repeatedly done myself — referred to as statistical, they should be seen in a broader concept as problems of representativity, which permits of a more comprehensive discussion.

Three different levels of representativity problems can be distinguished — A: How well does the pollen spectrum represent the actual composition of the pollen mixture present in the sample under consideration?, B: How well does the pollen flora of a sample represent the vegetation of the area around the site investigated? and C: How well does the actual flora represent the potential vegetation cover in its climax aspect? Obviously these are three different, although

related, problems, the last of which is not particular to pollen analysis even if its solution is of great importance for the evaluation of pollen-analytic data.

The first problem is a purely statistical one, and can be solved on a general mathematical basis with no reference to pollen analysis. Even so, the statistical problems are far from easy. In a joint paper several years ago Ottestad (FAEGRI & OTTESTAD, 1948), was able to formulate certain rules, but, as he pointed out, their value is rather limited, and one may doubt if the greater work involved actually pays. One of the major snags is that if statistical methods are used to check upon a result (e.g. pollen spectrum) which is, for some reason or other, outstanding, there is already a bias in the material which invalidates the assumption on which the statistical treatment is based. On the other hand, if the whole sequence is subjected to an unbiased statistical check of this kind, the only thing that can come out of it is the establishment that an outstanding sample is outstanding — which anybody could see for themselves. True, one may find that probability values are extremely low, but, again, this may be the one green pea among a thousand yellow ones. Sooner or later it turns up.

It should be kept in mind that statistics can never give absolute answers, only answers in terms of probabilities, and even if probabilities are very low, they are never exclusive. Biology is full of examples of improbabilities come true; after all, it would be very easy to prove that the origin of life itself is a "statistical impossibility". Apparently a good general judgement is as useful as a lot of more or less misapplied statistics when it comes to the interpretation of a pollen diagram.

However, there is one important lesson to be learnt from the treatment of pollen statistics in this sense, viz. the very great possible variation between samples drawn from the same universe and the corresponding care that should be exercised in

drawing conclusions from a scanty material, especially from single spectra. Values may easily vary between 40 and 60 per cent even when as much as 500 pollen grains have been counted, and what is perhaps even more important: the change of small percentages by a factor of three is not unexpected even if 1600 pollen grains had been counted in the sample in question. This matter has been dealt with in Faegri & Iversen (1964: 126-7) and shall not be detailed here, but it may be useful to remind that an increase of the percentage of pollen type A from 90 to 95 per cent in reality reflects that the pollen production of A was doubled in relation to that of non-A (*l.c.* 109). Thus the changes in percentages are far from being proportional to changes in actual pollen output. Again this is an inherent weakness of all percentage calculations of this kind, and not particular to pollen analysis.

A special statistical problem in pollen analysis has been the calculation "outside" the pollen sum. The historic reason for this principle of calculation — and the only reason there is, is the historic one — goes back to the earliest days of pollen analysis, 50 years ago. At that time pollen analysis was considered exclusively as a method of studying the development of forests of a landscape that in principle was assumed to be completely forested. The few NAP grains observed were frequently not identified, and certainly were not brought into any calculation. In a completely forest-covered landscape where the existence of forest margins, etc. are discounted *Corylus* in Scandinavia only occurs as an under-shrub. Thus, von Post pointed out, it does not occupy an area of its own, and consequently does not belong to the pollen types that should be taken into account. On the other hand, *Corylus* pollen sometimes was rather frequent and had to be distinguished from that of *Betula* — and it could therefore not be completely ignored, so it was accounted for in this rather arbitrary manner which in practice did not matter as long as the *Corylus* values were as low as in the first Swedish diagrams (cf. VON POST, 1918: 442).

Only later when material came in from other regions, especially western Europe, the excessive *Corylus* values appeared, approaching or exceeding 100 per cent. By the time pollen analysts took interest in

NAP, they were already accustomed to this abnormal way of calculating percentages, and went forth in the same manner, as I have done myself in my first papers.

However, there are two good geobotanical reasons against this habitual procedure. One is that von Post's original reasoning as to *Corylus* is fallacious. Practically all *Corylus* pollen present in the general pollen rain derives from shrubs flowering in the open, thus actually representing separate areas. Under-story *Corylus* flowers very sparsely, and its pollen, like that of other forest-bottom plants, practically never comes outside the forest. The second reason is that the idea of excluding pollen not representing a separate area is certainly not applicable to the NAP, which is only produced in appreciable quantities in deforested or forest-less regions. Thus there are no geobotanic reasons advocating the calculation of percentages of *Corylus* or NAP "outside" the pollen sum, and mathematically the habitual procedure is even less acceptable, as shown by Mosimann (1963) who concludes (p. 53) that "coefficients among inverse counts. . . are estimates of correlations among indices which are not useful quantities". In advanced pollen-analytic work this method of calculation has been abandoned altogether, and it is about time that it disappears.

This line of argument inevitably leads on to one of the fundamental problems in pollen analysis: the pollen sum, the basis for percentage calculation. Let us at once state that there is nothing like *the* pollen sum (cf. WRIGHT & PATTON, 1963), established once and for all. Pollen sums must be adapted to the problem they are supposed to elucidate, and then the basic rule is extremely simple: they should contain the pollen grains from those plants that are of interest in elucidating the actual question. Any grains may be brought in and any grains excluded, only the problem at hand can give an indication as to what should be done. It is perfectly acceptable, for instance, to omit from the consideration all conifer pollen as has been done by Wenner (1948), only the problem under investigation is then the variation in composition of deciduous forest. Whether these variations have any bearing on more fundamental problems or not, is a phytogeographical question, not a pollen analytic one. Similarly, one may even within a partially

forest-less area make an exclusive AP sum — as is frequently done. Variations in the AP composition then give information about variation in the composition of the (residual) forests, and again, general phyto-geographic reasoning must decide whether this information is of any value or not. In the extreme cases none of the forest trees involved occur near the site of investigation, and the result is an analysis of the changes of composition of forests perhaps 50 or 100 km away. The AP/NAP ratio may give indications, but again, a high AP ratio may indicate either the presence of forest (AP) or the absence of pollen-producing non-forested areas (NAP), e.g. a high arctic lichen tundra.

There is one pollen type which should be treated very cautiously, if at all involved in percentage calculations except for extremely restricted local use. The basic assumption of pollen analysis is the even sifting of individual pollen grains over the surroundings of the pollen producer. This holds true for most anemophilous plants. Non-coherence between individual grains is a major point in the syndrome of anemophily (FAEGRI & v. D. PIJL, 1966). On the other hand, many zoophilous plants have sticky pollen that does not separate in individual grains, but (if not used for pollination) falls to the ground in a few, large lumps, culminating with orchid and asclepiad pollinia. In such plants, the basic assumption of an even sifting is completely fallacious, and the number of grains met with in analysis is not a function of the representation of the species, but of the size of the pollen lump accidentally incorporated in the sample. The incidence and quantity of pollen of extreme zoophilous plants therefore varies according to rules different from those valid for anemophilous ones, and a percentage calculation is, strictly speaking, not permissible. It would be more correct simply to state absolute occurrences more or less like the statements of occurrence of mega-fossils.

Whereas the fundamental statistical principles of pollen analysis are simple enough, putting them into practice may meet with complications, above all because of unequal pollen productivity, also among anemophilous species. This inequality of pollen production is, partly, an inherent quality, to which will be referred below, partly, a local phenomenon which has caused

difficulties in interpretation of the diagrams of the majority of active investigators. Locally occurring great pollen (and spore) producers belonging to the mother formation (lake side vegetation), to seral stages (mangroves), or whatever it may be, may completely distort a diagram. It is one of the inherent weaknesses of percentage calculations that such cases cannot be accounted for mathematically. The only remedy is to exclude the pollen types involved with consequent loss of information, and calculate only for the other groups represented, at any rate within those parts of the diagram involved. Unnecessary to say, the calculation basis (the pollen sum) should be uniform within each diagram. If it is considered necessary to exclude a certain type within part of the sequence, a diagram calculated on that basis should not be intercalated in a sequence calculated on a wider basis. If it is desirable to use the same vertical column in order to save printing space, a very clear distinction should be made, e.g. by a heavy horizontal line through the diagram(s) in question.

In cases like these, discriminating identification is the best guarantee against loss of information. If, for example, the excess production of *Phragmites* pollen in a reed-swamp peat forces us to discard the grass pollen curve altogether, a great deal of information about dry-ground vegetation may also be lost. If *Phragmites* pollen can be identified apart from other grass pollen types, and excluded, practically no information is lost. Therefore, whenever locally produced pollen affects diagrams, the primary remedy is to restrict the damage as much as possible by careful identification. But naturally, if one of the regular constituents of the diagram is the cause of excessive pollen accumulation, exclusion can only lead to a "truncated" diagram.

Always, when it has been found necessary to omit from a diagram a pollen type that should ordinarily "belong" there, an auxiliary diagram should be added, containing the complete information. All exclusions etc. represent subjective opinions on the part of the author, and, however well founded they may be, a reader has the right to demand the necessary data for forming his own opinion.

A special problem is the treatment of secondary or supposed secondary pollen (IVERSEN, 1936). However, usually these

pollen grains form a rather insignificant part of the total quantity, and thus it is not so much a problem of representativity as of phytogeography.

The problems represented by differences between various species with regard to pollen productivity have already been mentioned. They lead us to the second main problem: To what extent does the pollen rain represent the vegetation? Davis has recently (1963) dealt with this problem at some length and has arrived at rather negative conclusions. However, I can in no way agree with these conclusions, which form a very distinct warning of what happens when a pollen analyst turns mathematician instead of phytogeographer.

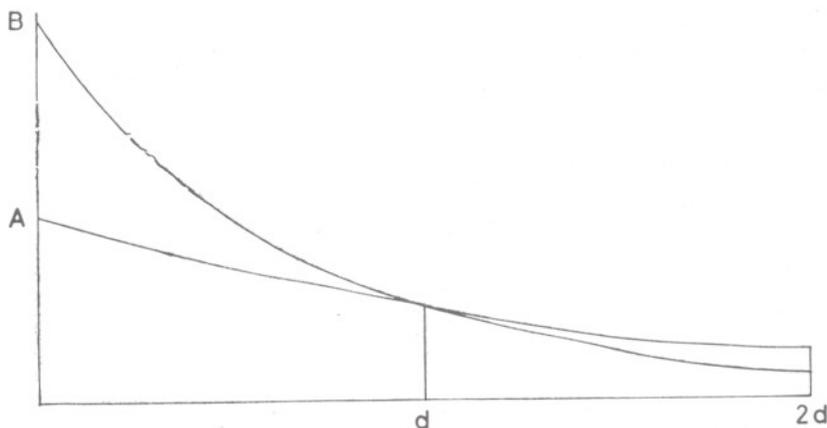
A comparatively simple problem is that of plants that do not produce recognizable pollen, e.g. *Larix* according to Davis (*l.c.*¹). It simply means that what we are studying, is the area covered by other trees (and herbaceous plants, as the case may be), leaving the unknown area of larch aside. If the species in question (larch in this case) covers large areas or is of great phytogeographical interest, this again represents a serious loss of information which simply cannot be helped. After all, there are many other species of great interest that never occur in pollen diagrams. On the other hand, this loss of information may attain formidable proportions in vegetation types the major constituents of which do not contribute to the regular pollen rain, as is evidently the case with many tropical forest types, consisting chiefly of zoophilous plants. Their pollen is not freely liberated into the air and, in addition, forms heavy lumps that fall straight down and is lost among the foliage. Pollen analysis is simply not a method for investigating those vegetational types, unless indirect conclusion can be arrived at from the presence or absence of important indicators that do contribute to the pollen rain. Unnecessary to say this presumes a very intimate knowledge of the ecology and sociology of the vegetation types concerned. To which may be added that pollen analysis of any vegetation type without such knowledge is bound to become at its best a lifeless stratigraphical tool, at its worst useless altogether.

1. The calculation of a conversion factor based upon one *Larix* pollen in a total of 6925 is, of course, statistically inadmissible. This to a certain extent invalidates the results summarized in Davis' Fig. 1B.

Whereas the problems posed by non-contributors are, in principle, comparatively simple, those inherent in disparate pollen productivity and transport distances of regular contributors are more complicated. There is no doubt that a vegetation consisting of one great producer together with a number of less productive species will give a pollen diagram that would appear quite distorted if one works from the supposition that pollen percentages should show any direct correspondence to the number of specimens of the various species concerned present in the area. It is perfectly possible that a pollen-analytic pine period, to use the case discussed by Davis (*l.c.*) is in reality the expression of a *Populus-Acer* dominated community with rather sparse pines. However, if the composition of the community is relatively constant, the pine zone will be stratigraphically consistent, and thus will present a good stratigraphical horizon.

A much more interesting and important problem is the actual vegetation represented by a pollen spectrum of such a character. The easiest way out would be to find conversion factors that would reduce the pollen production per unit area to the same magnitude for all species concerned. Thus, each grain of pine pollen would be assumed to represent an area *a*, a grain of maple or poplar an area, *b*, many times as great. With rather arbitrary correction factors this was proposed many years ago by Iversen (FAEGRI & IVERSEN, 1950: 87), and has later been discussed by various authors including Davis (*l.c.*).

However, in practice also this meets with considerable difficulties. The pollen production of a species is not invariable: climate, exposure, competition, cultivation measures, and other factors influence it deeply. Further, the differential transport effectivity of various pollen types come into the picture. A species, *A*, producing only 50 per cent per unit area in comparison to species *B* may be better distributed with the result that at some distance the pollen rain due to equal areas of *A* and *B* may be the same, and at even greater distances *A* may surpass *B* (TEXT-FIG. 1). This especially pertains to NAP species in relation to AP. Considering that the pollen rain usually consists of local and long-distance pollen in various proportions, it is easily seen that both a calculation of the primary conversion factors and their later applica-



TEXT-FIG. 1 — Pollen incidence at various distances from the site of production. Species A produces half as much as B, but is transported twice as effectively. At distance d the two pollen types are represented by 50 per cent each, at $2d$ the proportions are reversed: $2/3$ A and $1/3$ B. The site of production is here considered as point; the decrease in sedimentation with distance from the source is assumed to follow a logarithmic curve.

tion in a strict mathematical sense meets with insuperable difficulties. Studies on the relation between recent pollen rain and actual vegetation composition are highly commendable, and many more are needed, but I feel that they should be used qualitatively only, and only for local comparison. It will hardly ever be possible to deduce general conversion factors from such — or any other — studies.

Conversion factors of not too great arbitrariness² are very useful, especially to the investigator himself in his attempts to visualize what kind of vegetation could be represented by a sample, but they cannot give any information that is not already there in the primary count. The "Pine zone" is a fact, the "Pine period" likewise, the fallacy comes in when it is unreservedly identified with a *Pine forest* period. To deduce from a pine-dominated diagram to a pine-dominated vegetation is simply an error of pollen analytic judgement. To deduce from a pine-dominated spectrum to the vegetation one has to see what else there is in the spectrum. If the rest of the pollen derives from equally productive species, a

diagram with 70 per cent *Pinus* represents a pine forest. If the rest of the pollen derives from inefficient producers even a few per cent of them may be a warning that the "pine zone" may be the pollen representation of a completely different vegetation type. To base further conclusions on the simple assumption that *Pinus* zone = *Pinus* forest is a fault of the investigator's, not of the method's. And above all, it is a weakness common to all indirect methods of vegetation study.

Most pollen diagrams comprising some span of time will show changes in pollen representation, presumably reflecting changes in vegetation type³. Only rarely is vegetation the ultimate object of a pollen-analytic study, so even if the difficulties of translating pollen data into vegetation have been mastered, there remains the obstacles inherent in deducing from composition of vegetation to the various factors responsible for this composition. These factors are partly historical, partly ecological, both concepts taken in a wide sense. The most important of the former is the composition of the flora as due to migrations, differentiation *in loco*, and geologic events. Among the latter are changes of climate or of soil

2. Whereas Iversen used the factor of 0.25 to reduce the pollen count of large produces, Davis' correction factors vary between 240 and 0.036 (not counting the *Larix* figure). However, there are reasons to believe that this variation is too great, and that effects of transport may have influenced the results.

3. Changes due to differential destruction are not discussed here, and in the following I also neglect the effects of local over-representation discussed above.

status and influence of man. There is no doubt that there has been a tendency to over-rate the effects of climate and to try to find climatic explanations for changes which are, in reality, more dependent on various other causes. In Europe, influence of man was the next factor to come to prominence, and it was gradually realized that after the advent of agriculture its effects had an over-riding influence. The retarding effect of long immigration routes was stressed by me as long ago as 1940 (pp. 48, 49), but it has frequently been over-looked. The classical examples are the late arrivals of *Picea* and *Fagus* in northwestern Europe, long after the climate within part of that region at least apparently became suitable for these trees. The floristic poverty of the early Post-Glacial again can hardly be explained by anything but delay because of long immigration routes. There is reason to believe that as for the climate demanding

species might have grown in Scandinavia much earlier than they did arrive.

Recently, Danish investigators have been stressing the influence of soil development as a factor in large-scale vegetational succession (ANDERSEN, 1964) and have shown that many changes interpreted as climatic may be primarily edaphic. The influence of climate on soil development remains to be elucidated.

But these are not problems of pollen analysis. Pollen analysis is a technique used for elucidating certain problems, and the application of the technique must be geared to the problem under investigation. There is not one all-embracing method: methods and problems are interrelated. But whatever the problem it must not be forgotten that the medium of production of the pollen analysed, is vegetation. Knowledge of vegetation is a paramount demand for utilization of pollen-analytic data.

REFERENCES

- ANDERSEN, S. T. (1964). Interglacial vegetational succession and lake development in Denmark. *Tenth int. bot. Congr. Abstracts* 25.
- DAVIS, M. B. (1963). On the theory of pollen analysis. *Amer. J. Sci.*, **261**: 897.
- FAEGRI, K. (1940). Quartärgeologische Untersuchungen im westlichen Norwegen. II. Zur spätquartären Geschichte Jaerens. *Bergens Mus. årb.* 1939-40. *Natv. rk.* 7.
- FAEGRI, K. & IVERSEN, J. (1950). Text-book of modern pollen analysis. 2nd ed: Textbook of pollen analysis. *Copenhagen (Munksgaard)*. (1964)
- FAEGRI, K. & OTTESTAD, P. (1948). Statistical problems in pollen analysis. *Univ. Bergen. Årb. naturv. R.*, **3**.
- FAEGRI, K. & VAN DER PIJL, L. (1966). Principles of pollination ecology. *Pergamon Press, London*.
- IVERSEN, J. (1936). Sekundäres pollen als Fehlerquelle.— *Danm. geol. Unders.* 4. rk. 2, 15.
- MOSIMANN, J. E. (1963). On the compound negative multinomial distribution and correlation among inversely sampled pollen counts. *Biometrika*, **50**: 47.
- VON POST, L. (1918). Skogsträdpollen i sydsvenska torvmosselagerföljder. *Forh.* 16. *skand. naturf. Møte* 1916: 433.
- WENNER, C. G. (1948). Pollen diagrams from Labrador. *Geogr. Ann. Stockh.*, **28-29** (1946-47): 137.
- WRIGHT, H. E. JR. & PATTERN, H. L. (1963). The pollen sum. *Pollen et Spores*, **5**: 445.