THE COMPUTER USED AS A TOOL IN PALYNOLOGY

Bui-Thi-Mai & Ian Johnson.

Abstract

Contemporary pollen-rains in different parts of a prehistoric rock-shelter were studied; (1) as a control on the extent to which horizontal variations in the pollen spectrum of fossil pollen-samples might be due to local topography rather than anthropological causes, (2) to see whether different horizontal locations of fossil pollen-samples could give rise to differing climatic interpretations for the same external vegetation. In view of the vast number of individual results the computer was for basic data-manipulation and for the application of chi-square tests for the comparison of samples (using a specially written programme giving a graphical representation of the contribution of different species to the statistic). The paper is concluded with a discussion of the application of multi-variate statistics to pollen data.

Aims of the study

The work described in this paper was carried out as part of interdisciplinary research on a Palaeolithic site (the 'Abri Vaufrey' in the Dordogne, excavated by J.-Ph. Rigaud). This particular study was made in the context of a Doctoral Thesis presented by one of the authors (B.T.M.) at the University of Bordeaux[3] (Bui-Thi-Maï 1974). The differences in the composition of the contemporary pollen-rain in different parts of the rock-shelter were investigated with a view to testing the variations in pollen-spectrum which might be caused by site-topography and microclimate. The aim of the study was to provide control for such variations in a subsequent study of fossil pollen-samples taken from different parts of the Palae-

(1) Institut de Biologie à Oran, Algeria.
(2) Institut du Quaternaire, Université de Bordeaux I.
(3) Thesis Director Mlle M.-M. Paquereau.

olithic living floors, in an attempt to isolate horizontal variations due to human activities (Rigaud, in Press).

Several authors (e.g Tauber 1965) have convincingly demonstrated that important variations in the spectrum of the pollen rain for open sites can be due to local topographical effects. However, in the majority of rock-shelter sites, pollen-samples are taken with little regard for the effect that local topography might have, to the extent of sampling different layers in the same sequence at different horizontal locations (not that it is always possible to do otherwise).

A second objective of the study was therefore to see whether horizontal variations in the pollen-rain could give rise to differing climatic interpretations with the same external flora.

Data collection and manipulation

The pollen-rain was sampled by 9 horizontal pollen traps in different parts of the site (see fig. 1), each identified by a letter. The plates were changed 18 times at more or less regular intervals over a period of 11 months. Each interval was identified by a number and is referred to as a 'period'. The pollens counted for a particular plate are termed a **'sample'** e.g. P13, L9, A5. Each sample was classified into 75 species of pollen. The total number of data-values therefore worked out at around 12,000.

It was realised at an early stage of the analysis that there were a lot of repetitive manipulations of the data, such as calculating the percentages of each species. These could be economically handled by computer, even if the programmes were specially written and never re-used (what we will call 'one-off' programmes). Secondly, some sort of statistical assessment of the differences between samples would be neccessary to validate the results.

With these objectives in view, a simple card-deck dataset structure was chosen, which could be punched directly from the pollen count-sheets. By keeping the data on cards it could be manipulated manually when this seemed more economical than programming. 'One-off' programmes were written to calculate percentages for each species within a sample or group of samples, and to group samples for particular periods or particular traps (the latter giving an approximate annual pollen-rain, the omitted month being that of minimum pollination). The experience proved very satisfactory, several weeks of labour being saved by a few hours of programming and manual sorting of cards. Under the right conditions, therefore, 'one-off' programmes can be a worthwhile time-saver.

Statistical tests and graphical representation

Next, a programme for the calculation of the chi-squared statistic in a 2 x n table(4) was used to compare each of the traps in turn within a given period, and also for the whole period of the experiment. This amounted to nearly 700 pairs, so the task was clearly impossible by hand. The first observation was that about 80% of the pairs of samples within the same period were significantly different, i.e. not from the same pollen-rain ($p<.01$), and all the annual samples were significantly different at the same probability. We have made the basic assumption that the 'sample' collected on the traps is an unbiassed sample of the pollen-rain (population) at that point.

The programme used, POLDIF(5), also prints a graphical representation of the contribution of each species to the difference between two samples (fig. 2). This is based on the value of the chi-squared statistic in a 2 x 2 contingency table (for species k the four cells are; sample 1 or sample 2 and species k or non species k), using Yates'correction. Species whose frequencies are insufficient for valid chi-squared tests are distinguished, and the frequencies for these cells are pooled for the calculation of the overall chi-squared statistic. It should be noted that a certain number of 'significant' results will arise by chance alone, owing to the repetition of 2 x 2 calculations.

The graphical representation summarises the data and allows the palynologist to see at a glance where the differences between two samples lie. It proved to be very useful for the formulation of hypotheses about the modes of pollen transportation. It could equally be used to show up the differences between 'samples' of fossil pollen (we cannot consider such 'samples' as unbiassed, but no species which fails to give a 'significant' result can be considered as differently represented in the two

(4) in our case n (≤ 75) is taken as the number of pollen-species present in one and/or the other of the samples, as not all the pollen species are available at any time.
(5) available from Ian Johnson, Relubbus Lane, Penzance, G.B.

samples. A 'significant' result implies either a difference in the populations or sample bias). Similarly it can be applied to other sorts of data where 'samples' are involved, such as attribute data or lithic assemblages, and test applications to the latter have proved a useful aid to interpretation.(Rigaud, Pers.Comm.). We will not dwell on the possibility of using the computer for the graphical representation of pollen diagrams etc., as this is simply a question of suitable programming.

Aeropalynological conclusions

Among other conclusions, it was found that in general there were significant differences between the different traps. These differences could be attributed to a number of factors involving differential transportation of different pollens by the air circulating within the shelter. However, these differences were insufficient to give rise to differing climatic interpretations, within the limits of resolution of interpretations based on fossil pollen. The factors coming into play are highly complex and not neccessarily correlated, and include the size and shape of the pollen grain, the nature of the exine, the relationship between wind direction, force and frequency and the flowering season, humidity and temperature. The least biassed representation of the vegztation surrounding the site appeared to come from the most central traps (H & K), away from the walls (where there are local air currents - traps J, G, L & I), the depths of the shelter (where lightness and transportability are most important - trap A) and areas of turbulence and eddies at the mouth of the shelter (traps P & Q). The shelter seems to act as a sort of integrator of its pollen-environement, and its composition is best represented at its centre. It should be noted, however, that this conclusion only applies to the site analysed, which is unusually deep for a rock-shelter (fig 1).

The study of horizontal variations in the fossil pollen has not yet been carried out.

Multivariate analysis

As a final stage in the computer analysis, a factor analysis was run on the most representative period, in an attempt to group together species with similar transportation patterns.

We were unable to find any coherent explanation for the factors extracted, which confirms, though in no way conclusively, the very complex nature of the factors affecting the differential transportation of the pollen. We feel, however, that it serves to point out that factor analysis, and related techniques, can easily be applied to palynological data. There are two possible approaches, corresponding with R mode and Q mode analysis.

(1) Grouping together of species which covary.

This approach was taken by Anta Montet-White et alia (1973), who equally applied it to sediment samples. For fossil pollen sequences this approach can help to define plant communities or plants with common environemental requirements, detected by their covariation with time and/or from pollen sequence to pollen sequence (i.e. from sample to sample).

(2) Grouping together of samples according to their composition.

For fossil pollen sequences this approach could be used to date pollen sequences relative to one-another by showing up links between samples in different sequences.

Castaing (1973) gives further examples of both these approaches applied to ocean bottom sediment samples.

The advantages of multivariate statistical techniques over an intuitive analysis of pollen data, either fossil or contemporary, rests on the volume and complexity of the data. Such techniques can help one to see the wood for the trees (no pun intended!), by bringing out the underlying pattern of variability. This has been clearly recognised by Montet-White et alia (op. cit. :41), who have been among the few people to use such techniques in palynology.

It should be stressed, however, that though multivariate analysis shows great promise in the field of palynology, where it seems to have been somewhat neglected, it is only a <u>tool</u>. The final interpretation rests firmly with the researcher.

We would like to thank Jean-Philippe Rigaud and Marie-Françoise Diot, who read and commented on the draft of this text. Any errors or omissions are, however, entirely our responsability.
Palynological study; Bui-Thi-Mai 1972 - 74. Computer work and drafting of article; Ian Johnson 1974,1975

Bui-Thi-Mai, M. 1974 — 'Contribution à l'étude du transport et de la sédimentation des pluies polliniques dans un abri sous roche'. Thèse à l'Université de Bordeaux I.

Castaing, P. 1973 — 'Remarques sur l'utilisation de l'analyse factorielle en sédimentologie' BULLETIN DE L'INSTITUT DE GEOLOGIE DU BASSIN D'AQUITAINE N° 13.

Montet-White, A., Paquereau, M.-M., Rogers, R.A., Texier, J.-P. 1973 — 'Le Malpas Rockshelter'. UNIVERSITY OF KANSAS PUBLICATIONS IN ANTHROPOLOGY N° 14.

Rigaud, J.-Ph. In Press. — 'Horizontal control and palaeobotany'. EXCAVATING A PALAEOLITHIC SITE. Tübingen University Publications.

Tauber, H. 1965 — 'Differential Pollen Dispersion and Interpretation of Pollen Diagrams'. DANMARKS GEOL. UNDERSOGELSE 2, 69.

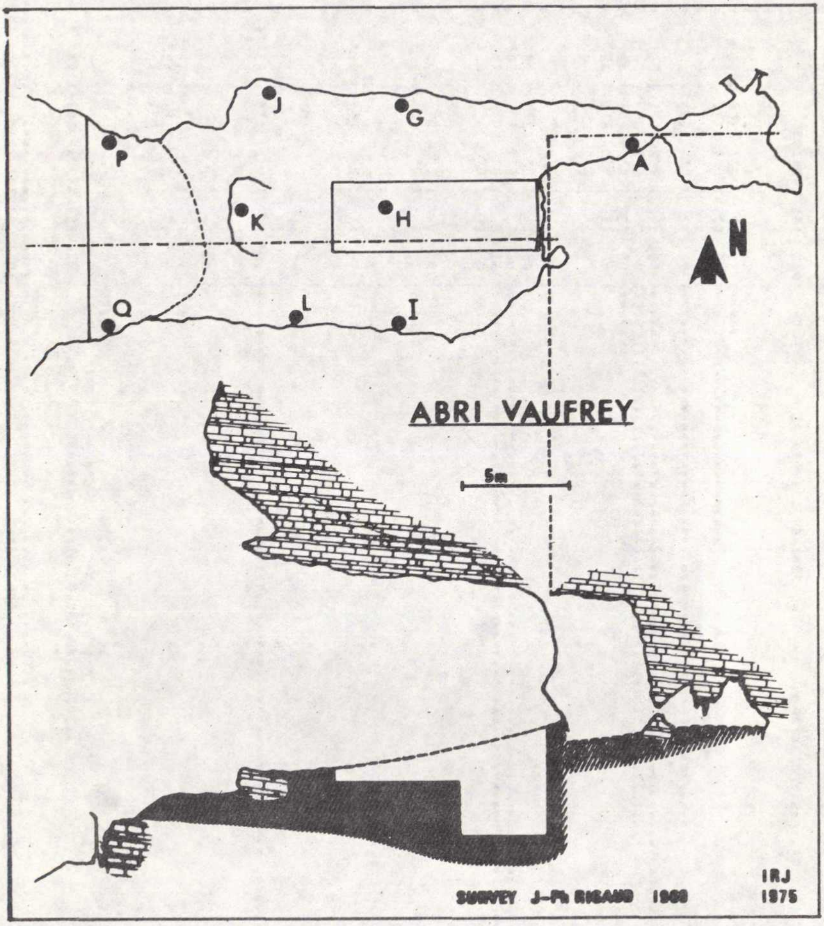

Fig 1. Section and plan of the Abri Vaufrey, showing locations of the pollen traps.

```
VAUFREY POLLEN RAINS 1970.   TRAP A, PERIOD 7.     FIRST 40 SPECIES.
****************************************************************************

VAUFREY POLLEN RAINS 1970.   TRAP P, PERIOD 7.     FIRST 40 SPECIES.
****************************************************************************


OBS * X SQD !PERCENTAGES  SAMPLE !            X SQUARED STATISTIC AND PROBABILITY
            !                    !        1.0       2.0       3.0       4.0       5.0       6.0       7.0       8.0       9.0  >9
            !                    ! 32%                      10%        5%                          1%
            !                    !  *                        *         *                           *
            !                    !
 1 *   .11 ! 1.6  1.2 !   9   7 !*                                    .
 2 *   .00 !  .2   .0 !   1   0 !                                     .
 3 *   .46 !  .2   .2 !   1   1 !????                                 .
 4 *       !          !         !                                     .
 5 *       !          !         !                                     .
 6 *       !          !         !                                     .
 7 *       !          !         !                                     .
 8 * 13.82 ! 9.7 17.5 !  54 101 !****************************************************************************************   ***
 9 *       !          !         !                                     .
10 *       !          !         !                                     .
11 *       !          !         !                                     .
12 *       !          !         !                                     .
13 *       !          !         !                                     .
14 *   .46 !  .2   .2 !   1   1 !????                                 .
15 *       !          !         !                                     .
16 *       !          !         !                                     .
17 *       !          !         !                                     .
18 *   .66 !  .9   .3 !   5   2 !??????                               .
19 *   .00 !  .0   .2 !   0   1 !                                     .
20 *       !          !         !                                     .
21 *       !          !         !                                     .
22 *   .54 !  .4   .0 !   2   0 !?????                                .
23 * 18.63 !  .0  3.6 !   0  21 !****************************************************************************************   ***
24 *  9.85 !  .5  3.3 !   3  19 !****************************************************************************************   ***
25 *   .80 !  .2   .0 !   1   0 !                                     .
26 *       !          !         !                                     .
27 *       !          !         !                                     .
28 *       !          !         !                                     .
29 *       !          !         !                                     .
30 *   .00 ! 9.2  9.4 !  51  54 !                                     .
31 *   .79 ! 4.7  3.5 !  26  20 !*******                              .
32 *  5.90 !  .0  1.4 !   0   8 !?????????????????????????????????????????????????????????
33 * 15.10 !43.5 32.1 ! 241 185 !****************************************************************************************   ***
34 *       !          !         !                                     .
35 *  5.87 ! 4.3  8.0 !  24  46 !*********************************************************
36 *       !          !         !                                     .
37 *  2.74 ! 5.1  3.0 !  28  17 !***************************          .
38 *   .55 ! 8.7  7.3 !  48  42 !*****                                .
39 *   .13 ! 1.3   .9 !   7   5 !*                                    .
40 *       !          !         !                                     .

                  TOTALS  554  576          18 DEGREES OF FREEDOM          OVERALL X SQUARED STATISTIC =  79.32

LUMPING CELLS WITH INSUFFICIENT FREQUENCIES    11 DEGREES OF FREEDOM          OVERALL X SQUARED STATISTIC =  65.43
```

Fig 2. Typical output from POLDIF.