



## **PAST: PALEONTOLOGICAL STATISTICS SOFTWARE PACKAGE FOR EDUCATION AND DATA ANALYSIS**

**Øyvind Hammer, David A.T. Harper, and Paul D. Ryan**

Øyvind Hammer. Paleontological Museum, University of Oslo, Sars gate1, 0562 Oslo, Norway

David A. T. Harper. Geological Museum, Øster Voldgade 5-7, University of Copenhagen, DK-1350 Copenhagen K, Denmark

Paul D. Ryan. Department of Geology, National University of Ireland, Galway, Ireland

### **ABSTRACT**

A comprehensive, but simple-to-use software package for executing a range of standard numerical analysis and operations used in quantitative paleontology has been developed. The program, called PAST (PAleontological STatistics), runs on standard Windows computers and is available free of charge. PAST integrates spreadsheet-type data entry with univariate and multivariate statistics, curve fitting, time-series analysis, data plotting, and simple phylogenetic analysis. Many of the functions are specific to paleontology and ecology, and these functions are not found in standard, more extensive, statistical packages. PAST also includes fourteen case studies (data files and exercises) illustrating use of the program for paleontological problems, making it a complete educational package for courses in quantitative methods.

**KEY WORDS:** Software, data analysis, education

Copyright: Palaeontological Association, 22 June 2001

Submission: 28 February 2001 Acceptance: 13 May 2001

---

### **INTRODUCTION**

Even a cursory glance at the recent paleontological literature should convince anyone that quantitative methods in paleontology have arrived at last. Nevertheless, many paleontologists still hesitate in applying such methods to their own data. One of the reasons for this has been the difficulty in acquiring and using appropriate data-analysis software. The 'PALSTAT' program was developed in the 1980s in order to minimize such obstacles and pro-

vide students with a coherent, easy-to-use package that supported a wide range of algorithms while allowing hands-on experience with quantitative methods. The first PALSTAT version was programmed for the BBC microcomputer (Harper and Ryan 1987), while later revisions were made for the PC (Ryan et al. 1995). Incorporating univariate and multivariate statistics and other plotting and analytical functions specific to paleontology and ecology, PAL-

STAT gained a wide user base among both paleontologists and biologists.

After some years of service, however, it was becoming clear that PALSTAT had to undergo major revision. The DOS-based user interface and an architecture designed for computers with miniscule memories (by modern standards) was becoming an obstacle for most users. Also, the field of quantitative paleontology has changed and expanded considerably in the last 15 years, requiring the implementation of many new algorithms. Therefore, in 1999 we decided to redesign the program totally, keeping the general concept but without concern for the original source code. The new program, called PAST (PAleontological STatistics) takes full advantage of the Windows operating system, with a modern, spreadsheet-based, user interface and extensive graphics. Most PAST algorithms produce graphical output automatically, and the high-quality figures can be printed or pasted into other programs. The functionality has been extended substantially with inclusion of important algorithms in the standard PAST toolbox. Functions found in PAST that were not available in PALSTAT include (but are not limited to) parsimony analysis with cladogram plotting, detrended correspondence analysis, principal coordinates analysis, time-series analysis (spectral and autocorrelation), geometrical analysis (point distribution and Fourier shape analysis), rarefaction, modelling by nonlinear functions (e.g., logistic curve, sum-of-sines) and quantitative biostratigraphy using the unitary associations method. We believe that the functions we have implemented reflect the present practice of paleontological data analysis, with the exception of some functionality that we hope to include in future versions (e.g., morphometric analysis with landmark data and more methods for the

validation and correction of diversity curves).

One of the main ideas behind PAST is to include many functions in a single program package while providing for a consistent user interface. This minimizes time spent on searching for, buying, and learning a new program each time a new method is approached. Similar projects are being undertaken in other fields (e.g., systematics and morphometry). One example is Wayne Maddison's 'Mesquite' package (<http://mesquite.biosci.arizona.edu/mesquite/mesquite.html>).

An important aspect of PALSTAT was the inclusion of case studies, including data sets designed to illustrate possible uses of the algorithms. Working through these examples allowed the student to obtain a practical overview of the different methodologies in a very efficient way. Some of these case studies have been adjusted and included in PAST, and new case studies have been added in order to demonstrate the new features. The case studies are primarily designed as student exercises for courses in paleontological data analysis. The PAST program, documentation, and case studies are available free of charge at <http://www.nhm.uio.no/~ohammer/past>.

## **PLOTTING AND BASIC STATISTICS**

Graphical plotting functions (see <http://www.nhm.uio.no/~ohammer/past/plot.html>) in PAST include different types of graph, histogram, and scatter plots. The program can also produce ternary (triangle) plots and survivorship curves.

Descriptive statistics (see <http://www.nhm.uio.no/~ohammer/past/univar.html>) include minimum, maximum, and mean values, population variance, sample variance, population and sample standard deviations, median, skewness, and kurtosis.

For associations or paleocommunity data, several diversity statistics can be computed: number of taxa, number of individuals, dominance, Simpson index, Shannon index (entropy), Menhinick's and Margalef's richness indices, equitability, and Fisher's  $\alpha$  (Harper 1999).

Rarefaction (Krebs 1989) is a method for estimating the number of taxa in a small sample, when abundance data for a larger sample are given. With this method, the number of taxa in samples of different sizes can be compared. An example application of rarefaction in paleontology is given by Adrain et al. (2000).

The program also includes standard statistical tests (see <http://www.nhm.uio.no/~ohammer/past/twosets.html>) for univariate data, including: tests for normality (chi-squared and Shapiro-Wilk), the  $F$  and  $t$  tests, one-way ANOVA,  $\chi^2$  for comparing binned samples, Mann-Whitney's  $U$  test and Kolmogorov-Smirnov association test (non-parametric), and both Spearman's  $r$  and Kendall's  $\tau$  non-parametric rank-order tests. Dice and Jaccard similarity indices are used for comparing associations limited to absence/presence data. The Raup-Crick randomization method for comparing associations (Raup and Crick 1979) is also implemented. Finally, the program can also compute correlation matrices and perform contingency-table analysis.

### MULTIVARIATE ANALYSIS

Paleontological data sets, whether based on fossil occurrences or morphology, often have high dimensionality. PAST includes several methods for multivariate data analysis (see <http://www.nhm.uio.no/~ohammer/past/multivar.html>), including methods that are specific to paleontology and biology.

Principal components analysis (PCA) is a procedure for finding hypothetical vari-

ables (components) that account for as much of the variance in a multidimensional data set as possible (Davis 1986, Harper 1999). These new variables are linear combinations of the original variables. PCA is a standard method for reducing the dimensionality of morphometric and ecological data. The PCA routine finds the eigenvalues and eigenvectors of the variance-covariance matrix or the correlation matrix. The eigenvalues, giving a measure of the variance accounted for by the corresponding eigenvectors (components), are displayed together with the percentages of variance accounted for by each of these components. A scatter plot of these data projected onto the principal components is provided, along with the option of including the Minimal Spanning Tree, which is the shortest possible set of connected lines joining all points. This may be used as a visual aid in grouping close points (Harper 1999). The component loadings can also be plotted. Bruton and Owen (1988) describe a typical morphometrical application of PCA.

Principal coordinates analysis (PCO) is another ordination method, somewhat similar to PCA. The PCO routine finds the eigenvalues and eigenvectors of a matrix containing the distances between all data points, measured with the Gower distance or the Euclidean distance. The PCO algorithm used in PAST was taken from Davis (1986), which also includes a more detailed description of the method and example analysis.

Correspondence analysis (CA) is a further ordination method, somewhat similar to PCA, but for counted or discrete data. Correspondence analysis can compare associations containing counts of taxa or counted taxa across associations. Also, CA is more suitable if it is expected that species have unimodal responses to the underlying parameters, that is they favor a certain range of the parameter and

become rare under for lower and higher values (this is in contrast to PCA, that assumes a linear response). The CA algorithm employed in PAST is taken from Davis (1986), which also includes a more detailed description of the method and example analysis. Ordination of both samples and taxa can be plotted in the same CA coordinate system, whose axes will normally be interpreted in terms of environmental parameters (e.g., water depth, type of substrate temperature).

The Detrended Correspondence (DCA) module uses the same 'reciprocal averaging' algorithm as the program *Decorana* (Hill and Gauch 1980). It is specialized for use on "ecological" data sets with abundance data (taxa in rows, localities in columns), and it has become a standard method for studying gradients in such data. Detrending is a type of normalization procedure in two steps. The first step involves an attempt to "straighten out" points lying along an arch-like pattern (= Kendall's Horseshoe). The second step involves "spreading out" the points to avoid artificial clustering at the edges of the plot.

Hierarchical clustering routines produce a dendrogram showing how and where data points can be clustered (Davis 1986, Harper 1999). Clustering is one of the most commonly used methods of multivariate data analysis in paleontology. Both R-mode clustering (groupings of taxa), and Q-mode clustering (grouping variables or associations) can be carried out within PAST by transposing the data matrix. Three different clustering algorithms are available: the unweighted pair-group average (UPGMA) algorithm, the single linkage (nearest neighbor) algorithm, and Ward's method. The similarity-association matrix upon which the clusters are based can be computed using nine different indices: Euclidean distance, correlation (using Pearson's  $r$  or Spearman's  $\rho$ ,

Bray-Curtis, chord and Morisita indices for abundance data, and Dice, Jaccard, and Raup-Crick indices for presence-absence data.

Seriation of an absence-presence matrix can be performed using the algorithm described by Brower and Kyle (1988). For constrained seriation, columns should be ordered according to some external criterion (normally stratigraphic level) or positioned along a presumed faunal gradient. Seriation routines attempt to reorganize the data matrix such that the presences are concentrated along the diagonal. Also, in the constrained mode, the program runs a 'Monte Carlo' simulation to determine whether the original matrix is more informative than a random matrix. In the unconstrained mode both rows and columns are free to move: the method then amounts to a simple form of ordination.

The degree of separation between to hypothesized groups (e.g., species or morphs) can be investigated using discriminant analysis (Davis 1986). Given two sets of multivariate data, an axis is constructed that maximizes the differences between the sets. The two sets are then plotted along this axis using a histogram. The null hypothesis of group means equality is tested using Hotelling's  $T^2$  test.

#### **CURVE FITTING AND TIME-SERIES ANALYSIS**

Curve fitting (see <http://www.nhm.uio.no/~ohammer/past/fitting.html>) in PAST includes a range of linear and non-linear functions.

Linear regression can be performed with two different algorithms: standard (least-squares) regression and the "Reduced Major Axis" method. Least-squares regression keeps the  $x$  values fixed, and it finds the line that minimizes the squared errors in the  $y$  values. Reduced Major Axis minimizes both the  $x$

and the  $y$  errors simultaneously. Both  $x$  and  $y$  values can also be log-transformed, in effect fitting the data to the “allometric” function  $y=10^b x^a$ . An allometric slope value around 1.0 indicates that an “isometric” fit may be more applicable to the data than an allometric fit. Values for the regression slope and intercepts, their errors, a  $\chi^2$  correlation value, Pearson’s  $r$  coefficient, and the probability that the columns are not correlated are given.

In addition, the sum of up to six sinusoids (not necessarily harmonically related) with frequencies specified by the user, but with unknown amplitudes and phases, can be fitted to bivariate data. This method can be useful for modeling periodicities in time series, such as annual growth cycles or climatic cycles, usually in combination with spectral analysis (see below). The algorithm is based on a least-squares criterion and singular value decomposition (Press et al. 1992). Frequencies can also be estimated by trial and error, by adjusting the frequency so that amplitude is maximized.

Further, PAST allows fitting of data to the logistic equation  $y=a/(1+be^{-cx})$ , using Levenberg-Marquardt nonlinear optimization (Press et al. 1992). The logistic equation can model growth with saturation, and it was used by Sepkoski (1984) to describe the proposed stabilization of marine diversity in the late Palaeozoic. Another option is fitting to the von Bertalanffy growth equation  $y=a(1-be^{-cx})$ . This equation is used for modeling growth of multi-celled animals (Brown and Rothery 1993).

Searching for periodicities in time series (data sampled as a function of time) has been an important and controversial subject in paleontology in the last few decades, and we have therefore implemented two methods for such analysis in the program: spectral analysis and autocorrelation. Spectral (harmonic) analysis

of time series can be performed using the Lomb periodogram algorithm, which is more appropriate than the standard Fast Fourier Transform for paleontological data (which are often unevenly sampled; Press et al. 1992). Evenly-spaced data are of course also accepted. In addition to the plotting of the periodogram, the highest peak in the spectrum is presented with its frequency and power value, together with a probability that the peak could occur from random data. The data set can be optionally detrended (linear component removed) prior to analysis. Applications include detection of Milankovitch cycles in isotopic data (Muller and MacDonald 2000) and searching for periodicities in diversity curves (Raup and Sepkoski 1984). Autocorrelation (Davis 1986) can be carried out on evenly sampled temporal-stratigraphical data. A predominantly zero autocorrelation signifies random data—periodicities turn up as peaks.

#### GEOMETRICAL ANALYSIS

PAST includes some functionality for geometrical analysis (see <http://www.nhm.uio.no/~ohammer/past/morpho.html>), even if an extensive morphometrics module has not yet been implemented. We hope to implement more extensive functionality, such as landmark-based methods, in future versions of the program.

The program can plot rose diagrams (polar histograms) of directions. These can be used for plotting current-oriented specimens, orientations of trackways, orientations of morphological features (e.g., trilobite terrace lines), etc. The mean angle together with Rayleigh’s spread are given. Rayleigh’s spread is further tested against a random distribution using Rayleigh’s test for directional data (Davis 1986). A  $\chi^2$  test is also available, giving

the probability that the directions are randomly and evenly distributed.

Point distribution statistics using nearest neighbor analysis (modified from Davis 1986) are also provided. The area is estimated using the convex hull, which is the smallest convex polygon enclosing the points. The probability that the distribution is random (Poisson process, giving an exponential nearest neighbor distribution) is presented, together with the 'R' value. Clustered points give  $R < 1$ , Poisson patterns give  $R \sim 1$ , while over-dispersed points give  $R > 1$ . Applications of this module include spatial ecology (are in-situ brachiopods clustered) and morphology (are trilobite tubercles over-dispersed; see Hammer 2000).

The Fourier shape analysis module (Davis 1986) accepts x-y coordinates digitized around an outline. More than one shape can be analyzed simultaneously. Points do not need to be evenly spaced. The sine and cosine components are given for the first ten harmonics, and the coefficients can then be copied to the main spreadsheet for further analysis (e.g., by PCA). Elliptic Fourier shape analysis is also provided (Kuhl and Giardina 1982). For an application of elliptic Fourier shape analysis in paleontology, see Renaud et al. (1996).

#### **PHYLOGENETIC ANALYSIS (PARSIMONY)**

The cladistics package (see <http://www.nhm.uio.no/~ohammer/past/cladist.html>) in PAST is fully operational, but is lacking comprehensive functionality. For example, there is no character reconstruction (plotting of steps on the cladogram). The use of PAST in parsimony analysis should probably be limited to entry-level education and preliminary investigations. The parsimony algorithms used in PAST are from Kitching et al. (1998).

Character states are coded using integers in the range 0 to 255. The first taxon is treated as the outgroup and will be placed at the root of the tree. Missing values are coded with a question mark. There are four algorithms available for finding short trees: branch-and-bound (finds all shortest trees), exhaustive (finds all shortest trees, and allows the plotting of tree-length distribution), heuristic nearest neighbor interchange (NNI) and heuristic subtree pruning and regrafting (SPR). Three different optimality criteria are available: Wagner (reversible and ordered characters), Fitch (reversible and unordered characters), and Dollo (irreversible and ordered). Bootstrapping can be performed with a given number of replicates.

All shortest (most parsimonious) trees can be viewed. If bootstrapping has been performed, a bootstrap value is given at the root of the subtree specifying each group.

The consensus tree of all shortest (most parsimonious) trees can also be viewed. Two consensus rules are implemented: strict (groups must be supported by all trees) and majority (groups must be supported by more than 50% of the trees). PAST can read and export files in the NEXUS format, making it compatible with packages such as PAUP and MacClade.

#### **BIOSTRATIGRAPHICAL CORRELATION WITH UNITARY ASSOCIATIONS**

Quantitative or semi-quantitative methods for biostratigraphy are not yet in common use, except for the relatively subjective approach of graphical correlation. Such methods are, however, well developed, and we hope that the inclusion of one method in PAST will help introduce more paleontologists to this field. We have chosen to implement Unitary Associations analysis (see <http://www.nhm.uio.no/~ohammer/past/unitary.html>) (Guex 1991)

because of its solid theoretical basis and minimum of statistical assumptions.

The data input consists of a presence-absence matrix with samples in rows and taxa in columns. Samples belong to a set of sections (localities), where the stratigraphical relationships within each section are known. The basic idea is to generate a set of assemblage zones (similar to 'Ospel zones') that are optimal in the sense that they give maximal stratigraphic resolution with a minimum of superpositional contradictions. An example of such a contradiction would be a section containing species A above species B, while assemblage 1 (containing species A) is placed below assemblage 2 (containing species B). The method of Unitary Associations is a logical but somewhat complicated procedure, consisting of several steps. Its implementation in PAST does not include all the features found in the standard program, called BioGraph (Savary and Guex 1999), and advanced users are referred to that package.

PAST produces a detailed report of the analysis, including maximal cliques, unitary associations, correlation table, reproducibility matrix, contradictions between cliques, biostratigraphic graph, graph of superpositional relationships between maximal cliques, and strong components (cycles) in the graphs (Guex 1991). It is important to inspect these results thoroughly in order to assess the quality of the correlation and to improve the quality of the data, if necessary. Angiolini and Bucher (1999) give an example of such careful use of the method of Unitary Associations.

### CASE STUDIES

The fourteen case studies have been designed to demonstrate both the use of different data analysis methods in paleontology and the specific use of the functions

in the program. The cases are taken from such diverse fields as morphology, taxonomy, paleoecology, paleoclimatology, sedimentology, extinction studies, and biostratigraphy. The examples are taken from both vertebrate and invertebrate paleontology, and they cover the whole of the Phanerozoic. These case studies are well suited for an introductory course in paleontological data analysis and have been tested in classroom situations. The cases are organized into four main subject areas: morphology and taxonomy, biogeography and paleoecology, time-series analysis, and biostratigraphy.

Case studies 1-5<sup>1</sup> involve the description and analysis of morphological variation of different sorts, while case study 6 targets some phylogenetic problems in a group of Cambrian trilobites and the mammals.

Case Study 1 investigates the external morphology of the Permian brachiopod *Dielasma*, developing ontogenic models for the genus and comparing the growth rates and outlines of different samples from in and around a Permian reef complex. In a more focused exercise, Case Study 2 uses spatial statistics to assess the mode of distribution of tubercles on the cranium of the trilobite *Paradoxides* from the middle Cambrian.

Case Study 3 tackles the multivariate morphometrics of the Ordovician illaenid trilobite *Stenopareia* using Principal Components Analysis (PCA), Principal Coordinate Analysis (PCO), cluster and discriminant analyses to determine the validity of two species from Scandinavia.

---

1. PE Note: The Case Study files are available from the PE site, and also directly from the author. The links below point to the author's site, which will, as time and the author proceed, contain updates and newer versions. The author's site is: <http://www.nhm.uio.no/~ohammer/past/>.

Case Study 4 demonstrates the use of Elliptic Fourier shape analysis and principal components for detecting changes in trilobite cephalon shape through ontogeny.

In Case Study 5, aspects of the allometric growth of the Triassic rhynchosaur *Scaphonyx* are investigated using regression analysis.

Case Study 6 investigates the phylogenetic structure of the middle Cambrian Paradoxididae through cladistic analysis, using parsimony analysis and bootstrapping. Similar techniques can be applied to a matrix of 20 taxa of mammal; cladograms generated by the program can be compared with a cluster analysis of the data matrix.

Case studies 7-11 cover aspects of paleobiogeography and paleoecology. Case Study 7 analyzes a global dataset of late Ordovician brachiopod distributions. A series of provincial faunas were developed against a background of regression and cooler surface waters during the first strike of the late Ordovician (Hirnantian) glaciation. Through the calculation of similarity and distance coefficients together with cluster analysis, these data can be organized into a set of latitudinally controlled provinces. Seriation helps to develop any faunal, possibly climatically generated, gradients within the data structure.

In Case Study 8 faunal changes through a well-documented section in the upper Llanvirn rocks of central Wales are investigated graphically and by the calculation of diversity, dominance, and related parameters for each of ten horizons in the sections. The changes in faunas fingerprint environmental shifts through the section, shadowed by marked changes in lithofacies. This dataset is ripe for considerable experimentation.

Case Study 9 involves a re-evaluation of Ziegler's classic Lower Paleozoic depth-related communities from the Anglo-Welsh area. Using a range of multi-

variate techniques (similarity and distance coefficients, cluster analysis, detrended correspondence analysis, and seriation) the reality and mutual relationships of these benthic associations can be tested using a modified dataset.

Case Study 10 discusses some well-known Jurassic shelly faunas from England and France. The integrity and onshore – offshore distribution of six Corallian bivalve-dominated communities is investigated with diversity measures, cluster analysis and detrended correspondence analysis.

Case Study 11 completes the analysis of biotic assemblages with an investigation of the direction and orientation of a bedding-plane sample of brachiopod shells from the upper Ordovician rocks of Scotland.

Two cases involve the study of time series data. Case Study 12 investigates the periodicity of mass extinctions during the Permian to Recent time interval using spectral analysis. A number of diversity curves can be modeled for the Paleozoic and post-Paleozoic datasets available in *Fossil Record 2*, and turnover rates can be viewed for Phanerozoic biotas.

Case Study 13 addresses the periodicity of oxygen isotope data from ice cores representing the last million years of Earth history.

The final case study demonstrates the use of quantitative biostratigraphical correlation with the method of Unitary Associations. Eleven sections from the Eocene of Slovenia are correlated using alveolinid foraminiferans studied by Drobne.

## CONCLUSION

Statistical and other quantitative methods are now very much part of the paleontologists' tool kit. PAST is a free, user-friendly and comprehensive package of statistical and graphical algorithms, tailor



made for the scientific investigation of paleontological material. PAST provides a window on current and future developments in this rapidly evolving research area. Together with a simple manual and linked case histories and datasets, the package is an ideal educational aid and first-approximation research tool. Planned future developments include extended functionality for morphometrics and the extension of available algorithms within the cladistics and unitary associations modules.

## REFERENCES

- Adrain, J.M., Westrop, S.R. and Chatterton, D.E. 2000. Silurian trilobite alpha diversity and the end-Ordovician mass extinction. **Paleobiology**, 26:625-646.
- Angiolini, L. and Bucher, H. 1999. Taxonomy and quantitative biochronology of Guadalupian brachiopods from the Khuff Formation, Southeastern Oman. **Geobios**, 32:665-699.
- Brower, J.C. and Kyle, K.M. 1988. Seriation of an original data matrix as applied to palaeoecology. **Lethaia**, 21:79-93.
- Brown, D. and Rothery, P. 1993. *Models in biology: mathematics, statistics and computing*. John Wiley & Sons, New York.
- Bruton, D.L. and Owen, A.W. 1988. The Norwegian Upper Ordovician illaenid trilobites. **Norsk Geologisk Tidsskrift**, 68:241-258.
- Davis, J.C. 1986. *Statistics and Data Analysis in Geology*. John Wiley & Sons, New York.
- Guex, J. 1991. *Biochronological Correlations*. Springer Verlag, Berlin.
- Hammer, Ø. 2000. Spatial organisation of tubercles and terrace lines in *Paradoxides forchhammeri* - evidence of lateral inhibition. **Acta Palaeontologica Polonica**, 45:251-270.
- Harper, D.A.T. (ed.). 1999. *Numerical Palaeobiology*. John Wiley & Sons, New York.
- Harper, D.A.T. and Ryan, P.D. 1987. *PALSTAT. A statistical package for palaeontologists*. Lochee Publications and the Palaeontological Association.
- Hill, M.O. and Gauch Jr, H.G. 1980. Detrended Correspondence analysis: an improved ordination technique. **Vegetation**, 42:47-58.
- Kitching, I.J., Forey, P.L., Humphries, C.J. and Williams, D.M. 1998. *Cladistics*. Oxford University Press, Oxford.
- Krebs, C.J. 1989. *Ecological Methodology*. Harper & Row, New York.
- Kuhl, F.P. and Giardina, C.R. 1982. Elliptic Fourier analysis of a closed contour. **Computer Graphics and Image Processing**, 18:259-278.
- Muller, R.A. and MacDonald, G.J. 2000. *Ice ages and astronomical causes: Data, Spectral Analysis, and Mechanisms*. Springer Praxis, Berlin.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. 1992. *Numerical Recipes in C*. Cambridge University Press, Cambridge.
- Raup, D. and Crick, R.E. 1979. Measurement of faunal similarity in paleontology. **Journal of Paleontology**, 53:1213-1227.
- Raup, D. and Sepkoski, J.J. 1984. Periodicities of extinctions in the geologic past. **Proceedings of the National Academy of Science**, 81:801-805.
- Renaud, S., Michaux, J., Jaeger, J.-J. and Auffray, J.-C. 1996. Fourier analysis applied to *Stephanomys* (Rodentia, Muridae) molars: nonprogressive evolutionary pattern in a gradual lineage. **Paleobiology**, 22:255-265.
- Ryan, P.D., Harper, D.A.T. and Whalley, J.S. 1995. *PALSTAT, Statistics for palaeontologists*. Chapman & Hall (now Kluwer Academic Publishers).
- Sepkoski, J.J. 1984. A kinetic model of Phanerozoic taxonomic diversity. **Paleobiology**, 10:246-267.
- Savary, J. and Guex, J. 1999. Discrete Biochronological Scales and Unitary Associations: Description of the BioGraph Computer Program. **Mémoires de Géologie (Lausanne)**, 34.