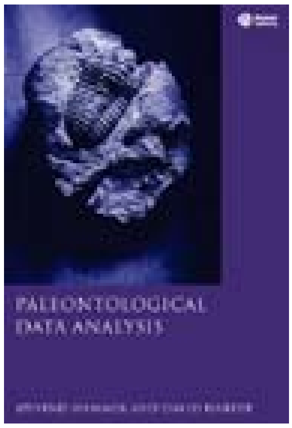
Journal of Sedimentary Research

An International Journal of SEPM

Colin P. North and Kitty L. Milliken, Editors

A.J. (Tom) van Loon, Associate Editor for Book Reviews

Review accepted 28 November 2005



Paleontological Data Analysis, by Oyvind Hammer & David Harper, 2005. Blackwell Publishing, 350 Main Street, Malden, MA 02148-5020, USA. Paperback, 351 pp. Price USD 89.95, GBP 39.99. ISBN 1405115440.

The days are long gone when geology in general, and paleontology in particular, were attractive to students considering a career in science because they were “non-mathematical” disciplines. Today’s palaeontologist is as likely to be found discussing maximum-likelihood models of phylogenetic networks or geometric morphometrics with colleagues as they are the details of invertebrate stratigraphic zonations or vertebrate taxonomy; and even these topics have their own quantitative literature. Given the broad interdisciplinary scope the topical diversity of 21st century paleontology then, the problem becomes how to teach the analytic skills all paleontologists need to come to conceptual grips with to new (and old) practitioners. This book tries to bridge the gap between mathematical treatises on specific methods and the technical applications literature in which results of a specific analysis are presented. The authors also clearly intend their book to be used as a kind of “extended user’s guide” to their public-domain PAlentological Statistics (PAST) program (http://folk.uio.no/ohammer/past/).

Paleontological Data Analysis is subdivided into sections on basic statistical methods (mostly highlights from the of univariate and bivariate parametric and non-parametric techniques covered in a standard statistics textbook), multivariate data analysis, morphometric methods, phylogenetic analysis, diversity and spatial methods termed “Paleobiogeography and Palaeoecology” here, time-series methods, and quantitative biostratigraphy. Most of the commonly used techniques in palaeontological data analysis receive (at least) a mention and a reference though discussions of some older (e.g., true factor analysis) and newer (e.g., partial least squares analysis) approaches are conspicuously absent. Odd too is the book’s assignment of presentations to these chapters. For example, the general-purpose multivariate methods of principal components analysis (PCA) and correspondence analysis are discussed not in the very brief chapter on multivariate methods—which only contains discussions of tests for multivariate normality and, inexplicably, cluster analysis—but in specialist applications chapters (e.g., morphometrics and biogeography–paleoecology for these two, respectively) where they are placed side-by-side with genuinely specialized techniques (e.g., geometric morphometrics, rarefaction).

To cover such a broad topical range in a scant 300 pages (excluding introduction, appendices, and index) requires authors to make hard decisions. Hammer & Harper have consistently favored breadth over depth. This is consonant with their stated purpose of providing a manual which "can be quickly consulted for initial information as to how to attack a specific problem" and "a dictionary briefly explaining methods à commonly mentioned in scientific papers" (p. ix). Nevertheless, a book that stuck to this remit rigorously could have been considerably shorter. Much of the book is taken up with "Example" boxes that refer to a dataset (available online at the book's web site, http://folk.uio.no/ohammer/past/book.html) and briefly review some of the results. These are not done as well as the now classic Sokal & Rohlf "Biometry" textbook and one can't avoid the impression these examples could have been handled better if they were presented entirely online, thus saving many book pages—and some proportion of the cost—with little or no loss of genuine information. As it stands their presentation contains much solid information and many oddities, along with the occasional howler. For example, in the discussion of population variance (p. 13–14) the reader is are told it is "necessary" to calculate the standard deviation from the variance, but not told why; no mention of units, no mention of Chebyshev's inequality. Similarly, in the section on statistical assumptions contains a cryptic, almost Rumsfeldian, direction, to wit "we will sometimes say that a statistical method makes no assumptions, but by this we mean that it makes the normal 'obvious' assumptions" (p. 6, emphasis theirs). Of course a list of what these 'obvious' assumptions might be is nowhere to be found in the text and no reference is made to textbooks or review papers that do contain such a discussion.

The authors clearly have broad practical experience of the literature, if not with the actual use of many of the techniques they describe (though even here there are many lapses in appropriate attribution). The book is, in the end, just what I suspect its authors intended it to be, somewhere between a well-referenced user's guide and an extended review paper. It can be read profitably by a broad audience and will properly serve as a point of entry into the far richer technical literature. But readers should be aware they will need to consult these additional sources to gain more than a cursory understanding of most of the topics mentioned.

Norman MacLeod
Palaeontology Department
Natural History Museum (London)
E-mail: N.MacLeod@nhm.ac.uk

SEPM - The Society for Sedimentary Geology