Compositional Data Analysis in the Geosciences: From Theory to Practice

Geological Society books refereeing procedures

The Society makes every effort to ensure that the scientific and production quality of its books matches that of its journals. Since 1997, all book proposals have been refereed by specialist reviewers as well as by the Society's Books Editorial Committee. If the referees identify weaknesses in the proposal, these must be addressed before the proposal is accepted.

Once the book is accepted, the Society Book Editors ensure that the volume editors follow strict guidelines on refereeing and quality control. We insist that individual papers can only be accepted after satisfactory review by two independent referees. The questions on the review forms are similar to those for *Journal of the Geological Society*. The referees' forms and comments must be available to the Society's Book Editors on request.

Although many of the books result from meetings, the editors are expected to commission papers that were not presented at the meeting to ensure that the book provides a balanced coverage of the subject. Being accepted for presentation at the meeting does not guarantee inclusion in the book.

More information about submitting a proposal and producing a book for the Society can be found on its web site: www.geolsoc.org.uk.

It is recommended that reference to all or part of this book should be made in one of the following ways:

BUCCIANTI, A., MATEU-FIGUERAS, G. & PAWLOWSKY-GLAHN, V. (eds) 2006. *Compositional Data Analysis in the Geosciences: From Theory to Practice*. Geological Society, London, Special Publications, **264**.

BUCCIANTI, A., TASSI, F. & VASELLI, O. 2006. Compositional changes in a fumarolic field, Vulcano Island: a statistical case study. *In*: BUCCIANTI, A., MATEU-FIGUERAS, G. & PAWLOWSKY-GLAHN, V. (eds) *Compositional Data Analysis in the Geosciences: From Theory to Practice*. Geological Society, London, Special Publications, **264**, 67–77.

GEOLOGICAL SOCIETY SPECIAL PUBLICATION NO. 264

Compositional Data Analysis in the Geosciences: From Theory to Practice

EDITED BY

A. BUCCIANTI

Univerisità degli Studi di Firenze, Italy

G. MATEU-FIGUERAS

Universitat e Girona, Spain

and

V. PAWLOWSKY-GLAHN

Universitat de Girona, Spain

2006

Published by

The Geological Society

London

THE GEOLOGICAL SOCIETY

The Geological Society of London (GSL) was founded in 1807. It is the oldest national geological society in the world and the largest in Europe. It was incorporated under Royal Charter in 1825 and is Registered Charity 210161.

The Society is the UK national learned and professional society for geology with a worldwide Fellowship (FGS) of over 9000. The Society has the power to confer Chartered status on suitably qualified Fellows, and about 2000 of the Fellowship carry the title (CGeol). Chartered Geologists may also obtain the equivalent European title, European Geologist (EurGeol). One fifth of the Society's fellowship resides outside the UK. To find out more about the Society, log on to www.geolsoc.org.uk.

The Geological Society Publishing House (Bath, UK) produces the Society's international journals and books, and acts as European distributor for selected publications of the American Association of Petroleum Geologists (AAPG), the Indonesian Petroleum Association (IPA), the Geological Society of America (GSA), the Society for Sedimentary Geology (SEPM) and the Geologists' Association (GA). Joint marketing agreements ensure that GSL Fellows may purchase these societies' publications at a discount. The Society's online bookshop (accessible from www.geolsoc.org.uk) offers secure book purchasing with your credit or debit card.

To find out about joining the Society and benefiting from substantial discounts on publications of GSL and other societies worldwide, consult www.geolsoc.org.uk, or contact the Fellowship Department at: The Geological Society, Burlington House, Piccadilly, London W1J 0BG: Tel. +44 (0)20 7434 9944; Fax +44 (0)20 7439 8975; E-mail: enquiries@geolsoc.org.uk.

For information about the Society's meetings, consult *Events* on www.geolsoc.org.uk. To find out more about the Society's Corporate Affiliates Scheme, write to enquiries@geolsoc.org.uk

Published by The Geological Society from:
The Geological Society Publishing House, Unit 7, Brassmill Enterprise Centre, Brassmill Lane, Bath BA1 3JN, UK

(*Orders*: Tel. +44 (0)1225 445046, Fax +44 (0)1225 442836)
Online bookshop: www.geolsoc.org.uk/bookshop

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.
ISBN-10: 1-86239-205-6
ISBN-13: 978-1-86239-205-2

Typeset by Techset Composition, Salisbury, UK
Printed by The Cromwell Press, Wiltshire, UK

Distributors

North America
For trade and institutional orders:
The Geological Society, c/o AIDC, 82 Winter Sport Lane, Williston, VT 05495, USA
Orders: Tel +1 800-972-9892
Fax +1 802-864-7626
Email gsl.orders@aidcvt.com

For individual and corporate orders:
AAPG Bookstore, PO Box 979, Tulsa, OK 74101-0979, USA
Orders: Tel +1 918-584-2555
Fax +1 918-560-2652
Email bookstore@aapg.org
Website http://bookstore.aapg.org

India
Affiliated East-West Press Private Ltd, Marketing Division, G-1/16 Ansari Road, Darya Ganj, New Delhi 110 002m India
Orders: Tel. +91 11 2327-9113/2326-4180
Fax +91 11 2326-0538
E-mail affiliat@vsnl.com

Contents

Preface

Compositions are positive vectors whose components represent a relative contribution of different parts to a whole; therefore their sum is a constant, usually 1 or 100. Compositions are a familiar and important kind of data for geologists because they appear in many geological datasets (chemical analyses, geochemical compositions of rocks, sand–silt–clay sediments, etc).

Since Karl Pearson wrote his famous paper on spurious correlation back in 1897, much has been said and written about the statistical analysis of compositional data, mainly by geologists such as Felix Chayes. His famous work concerned the G-2 granite sample and is used for comparison and standardization of geochemical analytical techniques between different laboratories. As with most igneous (and metamorphic) rocks that have achieved a stable mineralogy and minimized chemical-free energy, the number of minerals is limited by the phase rule to ≤ 6. In G-2 the minerals are, in order of decreasing abundance, plagioclase feldspar (43%), microcline feldspar (27%), quartz (21%), biotite (6%) and some minor accessory minerals. Following Chayes, the sympathetic variation between the percentage volumes of these minerals is predictable for quartz and plagioclase feldspar, showing a strong apparent inverse correlation since an increase in the proportion of one necessarily displaces the content of the other to some degree. Minor components may show an apparently positive correlation with a major one, e.g. biotite with plagioclase. Moreover, even modestly abundant components may show an apparent positive correlation, such as microcline with quartz. As Chayes pointed out, such variables pose intractable problems for conventional statistical methodologies generating, for example, bias in the sign of the correlation values. Consequently, the spurious positive correlations are just as predictable and misleading (and meaningless from a statistical point of view) as the negative ones, expected for a small number of major components.

The most important step towards a solution of the issue was made in the early 1980s, when John Aitchison proposed the use of log-ratios. His seminal work has brought a completely new perspective to the statistical analysis of data in general, not only compositional in nature, but also strictly positive data, or a combination thereof. This new approach is based on the idea that the sample space has a **natural geometry**, a geometry that is coherent with the intuitive concept of difference associated to the particular type of data. If we think that 5% is half of 10%, while 45% is 0.9 of 50%, why should we use methods which are based on the idea that the difference between 5% and 10% is the same as between 45% and 50%? Statistics is expected to make sense in our perception of the natural scale of data, and this is possible for compositional data using the log-ratio approach.

There is a long history to the search for a proper approach to the statistical analysis of compositional data, with many publications warning of the potential misuse of standard statistical methods. Many publications have also illustrated the potential of the log-ratio approach, but, despite this, there are still many research groups who are not aware of the existence of a solution to the problem identified by Karl Pearson. We intend this Special Publication to be of interest to geologists using statistical methods with compositional data, mainly geochemists and petrologists. It includes the intuitive justification of the proposed methodology and presents case studies in different fields and includes free software. There is also a section for the mathematically skilled, for those who need to see the proof of the mathematical consistency of the methods used. This last aspect is necessary, since many advances have been made in the last 20 years, and there is no book available up to now which provides a synthesis of the progress made.

Summarizing, it could be said that the main aim of this book is the diffusion of the state of the art in this field, emphasizing practical applications to the geological sciences. To introduce the reader to the subject coherently, the book comprises three general parts, following an introductory chapter. This illustrates with simple examples the potential usefulness of the method and, at the same time, brings the basic concepts common to the subsequent case studies.

Part I 'Applications to the solution of real geological problems' presents the study of some real geological problems. It forms the core of the book, as it is devoted to illustrating the application of the new methodology in the investigation of real geological problems in some different fields of research. In particular, it includes case studies concerning the chemistry of different geological matrices, such as minerals, rocks and sediments, as well as fluids, collected, respectively, in different geodynamical circumstances and environments. An example concerning research in palaeontology, in which reciprocal abundances of different species are usually managed, is also included.

Part II 'Software and related issues' brings the necessary tools for the better understanding and application of the methods proposed, i.e. computer programs with illustrative examples on how to use them. It includes contributions about available software to deal with compositional data, one of them developed in Visual Basic associated with Excel® and the other developed with the R package.

Part III 'General theory and methods' presents some mathematical contributions aimed at giving a useful summary and justifying the appropriateness of the methodology used. Using simple geological examples it shows how reasonable results and interpretations can be obtained with this methodology. Other issues addressed in the theoretical section are the definition of parametric models, and the treatment of missing values. It also includes a contribution on the special topology of the simplex, to help in the many discussions about whether the simplex is opened or closed.

This book was initially based on contributions presented to sessions G13.01 and G03.08 at the International Geological Congress, which took place in Florence in August 2004. Session G13.01 dealt with compositional analysis, while G03.08 was devoted to the importance of statistical analysis in the solution of environmental problems. Later, it was opened up to other contributions with the aim of bringing more case studies, more software and a better insight into the mathematics behind the whole approach. The editors would like to express here their thanks for the effort made by all the authors.

A. Buccianti
G. Mateu-Figueras
V. Pawlowsky-Glahn