Compositional Data Analysis In The Geosciences From Theory To Practice Special Publication No 264 Geological Society Special Publication

Modeling and Analysis of Compositional Data presents a practical and comprehensive introduction to the analysis of compositional data along with numerous examples to illustrate both theory and application of each method. Based upon short courses delivered by the authors, it provides a complete and current compendium of fundamental to advanced methodologies along with exercises at the end of each chapter to improve understanding, as well as data and a solutions manual which is available on an accompanying website. Complementing Pawlowsky-Glahn's earlier collective text that provides an overview of the state-of-the-art in this field, Modeling and Analysis of Compositional Data fills a gap in the literature for a much-needed manual for teaching, self learning or consulting.

Compositional Data Analysis in Practice is a user-oriented practical guide to the analysis of data with the property of a constant sum, for example percentages adding up to 100%. Compositional data can give misleading results if regular statistical methods are applied, and are best analysed by first transforming them to logarithms of ratios. This book explains how this transformation affects the analysis, results and interpretation of this very special type of data. All aspects of compositional data analysis are considered: visualization, modelling, dimension-reduction, clustering and variable selection, with many examples in the fields of food science, archaeology, sociology and biochemistry, and a final chapter containing a complete case study using fatty acid compositions in ecology. The applicability of these methods extends to other fields such as linguistics, geochemistry, marketing, economics and finance. R Software The following repository contains data files and R scripts from the book https://github.com/michaelgreenacre/CODAinPractice. The R package easyCODA, which accompanies this book, is available on CRAN -- note that you should have version 0.25 or higher. The latest version of the package will always be available on R-Forge and can be installed from R with this instruction: install.packages("easyCODA", repos="http://R-Forge.R-project.org").

This unique book addresses the statistical modelling and analysis of microbiome data using cutting-edge R software. It includes real-world data from the authors' research and from the public domain, and discusses the implementation of R for data analysis step by step. The data and R computer programs are publicly available, allowing readers to replicate the model development and data analysis presented in each chapter, so that these new methods can be readily applied in their own research. The book also discusses recent developments in statistical modelling and data analysis in microbiome research, as well as the latest advances in next-generation sequencing and big data in methodological development and applications. This timely book will greatly benefit all readers involved in microbiome, ecology and microarray data analyses, as well as other fields of research.

Microcomputer Package for the Statistical Analysis of Compositional Data

This volume contains a selection of papers presented at the Seven~h Confer ence of the International Federation of Classification Societies (IFCS-2000), which was held in Namur, Belgium, July 11-14,2000. From the originally sub mitted papers, a careful review process involving two reviewers per paper, led to the selection of 65 papers that were considered suitable for publication in this book. The present book contains original research contributions, innovative ap plications and overview papers in various fields within data analysis, classifi cation, and related methods. Given the fast publication process, the research results are still up-to-date and coincide with their actual presentation at the IFCS-2000 conference. The topics captured are: • Cluster analysis • Comparison of clusterings • Fuzzy clustering • Discriminant analysis • Mixture models • Analysis of relationships data • Symbolic data analysis • Regression trees • Data mining and neural networks • Pattern recognition • Multivariate data analysis • Robust data analysis • Data science and sampling The IFCS (International Federation of Classification Societies) The IFCS promotes the dissemination of technical and scientific information data analysis, classification, related methods, and their applica concerning tions.

This book presents the statistical analysis of compositional data using the log-ratio approach. It includes a wide range of classical and robust statistical methods adapted for compositional data analysis, such as supervised and unsupervised methods like PCA, correlation analysis, classification and regression. In addition, it considers special data structures like high-dimensional compositions and compositional tables. The methodology introduced is also frequently compared to methods which ignore the specific nature of compositional data. It focuses on practical aspects of compositional data analysis rather than on detailed theoretical derivations, thus issues like graphical visualization and preprocessing (treatment of missing values, zeros, outliers and similar artifacts) form an important part of the book. Since it is primarily intended for researchers and students from applied fields like geochemistry, chemometrics, biology and natural sciences, economics, and social sciences, all the proposed methods are accompanied by worked-out examples in R using the package robCompositions.

The publication of the King James version of the Bible, translated between 1603 and 1611, coincided with an extraordinary flowering of English literature and is universally acknowledged as the greatest influence on English-language literature in history. Now, world-class literary writers introduce the book of the King James Bible in a series of beautifully designed, small-format volumes. The introducers' passionate, provocative, and personal engagements with the spirituality and the language of the text make the Bible come alive as a stunning work of literature and remind us of its overwhelming contemporary relevance.

La diversidad biológica es fruto de la interacción entre numerosas especies, ya sean marinas, vegetales o animales, a la par que de los muchos factores limitantes que caracterizan el medio que habitan. El análisis multivariante utiliza las relaciones entre diferentes variables para ordenar los objetos de estudio según sus propiedades colectivas y luego clasificarlos; es decir, agrupar especies o ecosistemas en distintas clases compuestas cada una por entidades con propiedades parecidas. El fin último es relacionar la variabilidad biológica observada con las correspondientes características medioambientales. Multivariate Analysis of Ecological Data explica de manera completa y estructurada cómo analizar e interpretar los datos ecológicos observados sobre múltiples variables, tanto biológicos como medioambientales. Tras una introducción general a los datos ecológicos multivariantes y la metodología estadística, se abordan en capítulos específicos, métodos como aglomeración (clustering), regresión, biplots, escalado multidimensional, análisis de correspondencias (simple y canónico) y análisis log-ratio, con atención también a sus problemas de modelado y aspectos inferenciales. El libro plantea una serie de aplicaciones a datos reales derivados de investigaciones ecológicas, además de dos casos detallados que llevan al lector a apreciar los retos de análisis, interpretación y comunicación inherentes a los estudios a gran escala y los diseños complejos.

This book presents modern methods and real-world applications of compositional data analysis. It covers a wide variety of topics, ranging from an updated presentation of basic concepts and ideas in compositional data analysis to recent advances in the context of complex data structures. Further, it illustrates real-world applications in numerous scientific disciplines and includes references to the latest software solutions available for compositional data analysis, thus providing a valuable and up-to-date guide for researchers and practitioners working with compositional

data. Featuring selected contributions by leading experts in the field, the book is dedicated to Vera Pawlowsky-Glahn on the occasion of her 70th birthday.

Since the 1960s, testimony by representatives of the Federal Bureau of Investigation in thousands of criminal cases has relied on evidence from Compositional Analysis of Bullet Lead (CABL), a forensic technique that compares the elemental composition of bullets found at a crime scene to the elemental composition of bullets found in a suspectâ€™s possession. Different from ballistics techniques that compare striations on the barrel of a gun to those on a recovered bullet, CABL is used when no gun is recovered or when bullets are too small or mangled to observe striations. Forensic Analysis: Weighing Bullet Lead Evidence assesses the scientific validity of CABL, finding that the FBI should use a different statistical analysis for the technique and that, given variations in bullet manufacturing processes, expert witnesses should make clear the very limited conclusions that CABL results can support. The report also recommends that the FBI take additional measures to ensure the validity of CABL results, which include improving documentation, publishing details, and improving on training and oversight.

As long ago as 1897 Karl Pearson, in a now classic paper on spurious correlation, first pointed out dangers that may befall the analyst who attempts to interpret correlations between ratios whose numerators and denominators contain common parts. He thus implied that the analysis of compositional data, with its concentration on relation ships between proportions of some whole, is likely to be fraught with difficulty. History has proved him correct: over the succeeding years and indeed right up to the present day, there has been no other form of data analysis where more confusion has reigned and where more improper and inadequate statistical methods have been applied. The special and intrinsic feature of compositional data is that the proportions of a composition are naturally subject to a unit-sum constraint. For other forms of constrained data, in particular for directional data where there is a unit-length constraint on each direction vector, scientist and statistician alike have readily appre ciated that new statistical methods, appropriate to the special nature of the data, are required; and there now exists an extensive literature on the successful statistical analysis of directional data. It is paradox ical that for compositional data, subject to an apparently simpler constraint, such an appreciation and development have been slower to emerge. In applications the unit-sum constraint has been widely ignored or wished away and inappropriate 'standard' statistical methods, devised for and successfully applied to unconstrained data, have been used with disastrous consequences.

This Open Access handbook published at the IAMG's 50th anniversary, presents a compilation of invited path-breaking research contributions by award-winning geoscientists who have been instrumental in shaping the IAMG. It contains 45 chapters that are categorized broadly into five parts (i) theory, (ii) general applications, (iii) exploration and resource estimation, (iv) reviews, and (v) reminiscences covering related topics like mathematical geosciences, mathematical morphology, geostatistics, fractals and multifractals, spatial statistics, multipoint geostatistics, compositional data analysis, informatics, geocomputation, numerical methods, and chaos theory in the geosciences.

The authoritative contributions gathered in this volume reflect the state of the art in compositional data analysis (CoDa). The respective chapters cover all aspects of CoDa, ranging from mathematical theory, statistical methods and techniques to its broad range of applications in geochemistry, the life sciences and other disciplines. The selected and peer-reviewed papers were originally presented at the 6th International Workshop on Compositional Data Analysis, CoDaWork 2015, held in L'Escala (Girona), Spain. Compositional data is defined as vectors of positive components and constant sum, and, more generally, all those vectors representing parts of a whole which only carry relative information. Examples of compositional data can be found in many different fields such as geology, chemistry, economics, medicine, ecology and sociology. As most of the classical statistical techniques are incoherent on compositions, in the 1980s John Aitchison proposed the log-ratio approach to CoDa. This became the foundation of modern CoDa, which is now based on a specific geometric structure for the simplex, an appropriate representation of the sample space of compositional data. The International Workshops on Compositional Data Analysis offer a vital discussion forum for researchers and practitioners concerned with the statistical treatment and modelling of compositional data or other constrained data sets and the interpretation of models and their applications. The goal of the workshops is to summarize and share recent developments, and to identify important lines of future research.

Loosely based on the Odyssey, this landmark of modern literature follows ordinary Dubliners through an entire day in 1904. Captivating experimental techniques range from interior monologues to exuberant wordplay and earthy humor.

It is difficult to imagine that the statistical analysis of compositional data has been a major issue of concern for more than 100 years. It is even more difficult to realize that so many statisticians and users of statistics are unaware of the particular problems affecting compositional data, as well as their solutions. The issue of ``spurious correlation'', as the situation was phrased by Karl Pearson back in 1897, affects all data that measures parts of some whole, such as percentages, proportions, ppm and ppb. Such measurements are present in all fields of science, ranging from geology, biology, environmental sciences, forensic sciences, medicine and hydrology. This book presents the history and development of compositional data analysis along with Aitchison's log-ratio approach. Compositional Data Analysis describes the state of the art both in theoretical fields as well as applications in the different fields of science. Key Features: Reflects the state-of-the-art in compositional data analysis. Gives an overview of the historical development of compositional data analysis, as well as basic concepts and procedures. Looks at advances in algebra and calculus on the simplex. Presents applications in different fields of science, including, genomics, ecology, biology, geochemistry, planetology, chemistry and economics. Explores connections to correspondence analysis and the Dirichlet distribution. Presents a summary of three available software packages for compositional data analysis. Supported by an accompanying website featuring R code.

Applied scientists working on compositional data analysis in any field of science, both in academia and professionals will benefit from this book, along with graduate students in any field of science working with compositional data.

Applied Compositional Data AnalysisWith Worked Examples in RSpringer

This book focuses on recent advances, approaches, theories and applications related to mixture models. In particular, it presents recent unsupervised and semi-supervised frameworks that consider mixture models as their main tool. The chapters considers mixture models involving several interesting and challenging problems such as parameters

estimation, model selection, feature selection, etc. The goal of this book is to summarize the recent advances and modern approaches related to these problems. Each contributor presents novel research, a practical study, or novel applications based on mixture models, or a survey of the literature. Reports advances on classic problems in mixture modeling such as parameter estimation, model selection, and feature selection; Present theoretical and practical developments in mixture-based modeling and their importance in different applications; Discusses perspectives and challenging future works related to mixture modeling.

Compositional Data Analysis (CoDA) refers to the analysis of all those vectors representing parts of a whole which only carry relative information. This type of data is ubiquitous in most applications, especially in geology, chemistry, genetics, and environmental sciences. The last comprehensive treatment was written by John Aitchison in 1986, and this book represents an update of that classic book.

This book highlights major problems in the statistical analysis of compositions that have been known for over a century, as well as the corresponding solutions that have been put forward by specialists over the past 30 years. The basic assumptions of normality or multi-normality are pointed out and methods to test and achieve them are also covered. The conventional major and trace element geochemistry and modeling equations are discussed, and are followed by a more sophisticated multidimensional approach to data handling. The book's main focus is on the use of statistical techniques to facilitate data interpretation. It also highlights the classification (or nomenclature) and tectonic discrimination aspects for both igneous and sedimentary rocks. The book concludes by discussing computer programs that are helping pave the way from geochemistry to geochemometrics. Written by a leading expert in the area of geochemistry, it offers a valuable guide for students and professionals in the area.

This book provides a guided approach to the geostatistical modelling of compositional spatial data. These data are data in proportions, percentages or concentrations distributed in space which exhibit spatial correlation. The book can be divided into four blocks. The first block sets the framework and provides some background on compositional data analysis. Block two introduces compositional exploratory tools for both non-spatial and spatial aspects. Block three covers all necessary facets of multivariate spatial prediction for compositional data: variogram modelling, cokriging and validation. Finally, block four details strategies for simulation of compositional data, including transformations to multivariate normality, Gaussian cosimulation, multipoint simulation of compositional data, and common postprocessing techniques, valid for both Gaussian and multipoint methods. All methods are illustrated via applications to two types of data sets: one a large-scale geochemical survey, comprised of a full suite of geochemical variables, and the other from a mining context, where only the elements of greatest importance are considered. R codes are included for all aspects of the methodology, encapsulated in the R package "gmGeostats", available in CRAN.

Compositional Data Analysis: Theory and Applications Edited by Vera Pawlowsky-Glahn, Department of Computer Science and Applied Mathematics, University of Girona, Spain. Antonella Buccianti, Department of Earth Sciences, University of Florence, Italy It is difficult to imagine that the statistical analysis of compositional data has been a major issue of concern for more than 100 years. It is even more difficult to realize that so many statisticians and users of statistics are unaware of the particular problems affecting compositional data, as well as their solutions. The issue of spurious correlation", as the situation was phrased by Karl Pearson back in 1897, affects all data that measures parts of some whole, such as percentages, proportions, ppm and ppb. Such measurements are present in all fields of science, ranging from geology, biology, environmental sciences, forensic sciences, medicine and hydrology. This book presents the history and development of compositional data analysis along with Aitchison's log-ratio approach.

"Compositional Data Analysis" describes the state of the art both in theoretical fields as well as applications in the different fields of science. Key Features: - Reflects the state-of-the-art in compositional data analysis. - Gives an overview of the historical development of compositional data analysis, as well as basic concepts and procedures. - Looks at advances in algebra and calculus on the simplex. - Presents applications in different fields of science, including, genomics, ecology, biology, geochemistry, planetology, chemistry and economics. - Explores connections to correspondence analysis and the Dirichlet distribution. - Presents a summary of three available software packages for compositional data analysis. - Supported by an accompanying website featuring R code. Applied scientists working on compositional data analysis in any field of science, both in academia and professionals will benefit from this book, along with graduate students in any field of science working with compositional data.

In this book we are concerned with Bayesian learning and forecast ing in dynamic environments. We describe the structure and theory of classes of dynamic models, and their uses in Bayesian forecasting. The principles, models and methods of Bayesian forecasting have been developed extensively during the last twenty years. This devel opment has involved thorough investigation of mathematical and sta tistical aspects of forecasting models and related techniques. With this has come experience with application in a variety of areas in commercial and industrial, scientific and socio-economic fields. In deed much of the technical development has been driven by the needs of forecasting practitioners. As a result, there now exists a relatively complete statistical and mathematical framework, although much of this is either not properly documented or not easily accessible. Our primary goals in writing this book have been to present our view of this approach to modelling and forecasting, and to provide a rea sonably complete text for advanced university students and research workers. The text is primarily intended for advanced undergraduate and postgraduate students in statistics and mathematics. In line with this objective we present thorough discussion of mathematical and statistical features of Bayesian analyses of dynamic models, with illustrations, examples and exercises in each Chapter.

Comprehensive overview of the spectroscopic, mineralogical, and geochemical techniques used in planetary remote sensing.

This book presents an overview of computational and statistical design and analysis of mass spectrometry-based proteomics, metabolomics, and lipidomics data. This contributed volume provides an introduction to the special aspects of statistical design and analysis with mass spectrometry data for the new omic sciences. The text discusses common aspects of design and

analysis between and across all (or most) forms of mass spectrometry, while also providing special examples of application with the most common forms of mass spectrometry. Also covered are applications of computational mass spectrometry not only in clinical study but also in the interpretation of omics data in plant biology studies. Omics research fields are expected to revolutionize biomolecular research by the ability to simultaneously profile many compounds within either patient blood, urine, tissue, or other biological samples. Mass spectrometry is one of the key analytical techniques used in these new omic sciences. Liquid chromatography mass spectrometry, time-of-flight data, and Fourier transform mass spectrometry are but a selection of the measurement platforms available to the modern analyst. Thus in practical proteomics or metabolomics, researchers will not only be confronted with new high dimensional data types—as opposed to the familiar data structures in more classical genomics—but also with great variation between distinct types of mass spectral measurements derived from different platforms, which may complicate analyses, comparison, and interpretation of results.

This book presents the statistical analysis of compositional data sets, i.e., data in percentages, proportions, concentrations, etc. The subject is covered from its grounding principles to the practical use in descriptive exploratory analysis, robust linear models and advanced multivariate statistical methods, including zeros and missing values, and paying special attention to data visualization and model display issues. Many illustrated examples and code chunks guide the reader into their modeling and interpretation. And, though the book primarily serves as a reference guide for the R package "compositions," it is also a general introductory text on Compositional Data Analysis. Awareness of their special characteristics spread in the Geosciences in the early sixties, but a strategy for properly dealing with them was not available until the works of Aitchison in the eighties. Since then, research has expanded our understanding of their theoretical principles and the potentials and limitations of their interpretation. This is the first comprehensive textbook addressing these issues, as well as their practical implications with regard to software. The book is intended for scientists interested in statistically analyzing their compositional data. The subject enjoys relatively broad awareness in the geosciences and environmental sciences, but the spectrum of recent applications also covers areas like medicine, official statistics, and economics. Readers should be familiar with basic univariate and multivariate statistics. Knowledge of R is recommended but not required, as the book is self-contained.

Since Karl Pearson wrote his paper on spurious correlation in 1897, a lot has been said about the statistical analysis of compositional data, mainly by geologists such as Felix Chayes. The solution appeared in the 1980s, when John Aitchison proposed to use logratios. Since then, the approach has seen a great expansion, mainly building on the idea of the `natural geometry' of the sample space. Statistics is expected to give sense to our perception of the natural scale of the data, and this is made possible for compositional data using logratios. This publication will be a milestone in this process.

Compositional Data Analysis in Practice is a user-oriented practical guide to the analysis of data with the property of a constant sum, for example percentages adding up to 100%. Compositional data can give misleading results if regular statistical methods are applied, and are best analysed by first transforming them to logarithms of ratios. This book explains how this transformation affects the analysis, results and interpretation of this very special type of data. All aspects of compositional data analysis are considered: visualization, modelling, dimension-reduction, clustering and variable selection, with many examples in the fields of food science, archaeology, sociology and biochemistry, and a final chapter containing a complete case study using fatty acid compositions in ecology. The applicability of these methods extends to other fields such as linguistics, geochemistry, marketing, economics and finance. R Software The R package easyCODA, which accompanies this book, can be downloaded from R-Forge as follows: install.packages("easyCODA", repos="") and will be available on CRAN soon. Notice that the R packages ca and vegan also have to be installed (from CRAN in the usual way).

Based on a course taught by the author, this book combines the theoretical underpinnings of statistics with the practical analysis of Earth sciences data using MATLAB. The book is organized to introduce the underlying concepts, and then extends these to the data, covering methods that are most applicable to Earth sciences. Topics include classical parametric estimation and hypothesis testing, and more advanced least squares-based, nonparametric, and resampling estimators. Multivariate data analysis, not often encountered in introductory texts, is presented later in the book, and compositional data is treated at the end. Datasets and bespoke MATLAB scripts used in the book are available online, as well as additional datasets and suggested questions for use by instructors. Aimed at entering graduate students and practicing researchers in the Earth and ocean sciences, this book is ideal for those who want to learn how to analyse data using MATLAB in a statistically-rigorous manner.

About twenty years ago, there was a recognition in Europe that real benefits would flow from coordinating the manner in which food composition tables were produced in the various countries of Europe. Subsequent development of computerised nutritional data bases has further highlighted the potential advantages of working together. Such cooperation could lead to improved quality and compatibility of the various European nutrient data bases and the values within them. This realisation was one of the driving forces behind the development of the Eurofoods initiative in the 1980's when those people in Europe interested in data on food composition began working together. This initiative received further impetus with the establishment of the Eurofoods-Enfant Concerted Action Project within the framework of the FLAIR (Food-Linked Agro-Industrial Research) Programme of the Commission of the European Communities. It was quickly recognised that the draft guidelines for the production, management and use of food composition data which had been prepared under the aegis of INFOODS (International Network of Food Data Systems, a project of the United Nations University), would be especially applicable to the objectives of the Concerted Action. The guidelines have been written by two recognised experts. Many people associated with FLAIR Eurofoods-Enfant have added constructive criticism and advice to that offered previously by those associated with INFOODS. Thus the guidelines are backed by a consensus in the community of those responsible for the production and use of food composition tables and nutrient data bases.

This book presents a unique collection of contributions on modern topics in statistics and econometrics, written by leading experts in the respective disciplines and their intersections. It addresses nonparametric statistics and econometrics, quantiles and expectiles, and advanced methods for complex data, including spatial and compositional data, as well as tools for empirical studies in economics and the social sciences. The book was written in honor of Christine Thomas-Agnan on the occasion of her 65th birthday. Given its scope, it will appeal to researchers and PhD students in statistics and econometrics alike who are interested in the latest developments in their field.

BIG DATA, ARTIFICIAL INTELLIGENCE AND DATA ANALYSIS SET Coordinated by Jacques Janssen Data analysis is a scientific field that continues to grow enormously, most notably over the last few decades, following rapid growth within the tech industry, as well as the wide applicability of computational techniques alongside new advances in analytic tools. Modeling enables data analysts to identify relationships, make predictions, and to understand, interpret and visualize the extracted information more strategically. This book includes the most recent advances on this topic, meeting increasing demand from wide circles of the scientific community. Applied Modeling Techniques and Data Analysis 2 is a collective work by a number of leading scientists, analysts, engineers, mathematicians and statisticians, working on the front end of data analysis and modeling applications. The chapters cover a cross section of current concerns and research interests in the above scientific areas. The collected material is divided into appropriate sections to provide the reader with both theoretical and applied information on data analysis methods, models and techniques, along with appropriate applications.