

Introduction to Compositional data analysis

Lecturers:

- Prof. Dr. Pawlowsky-Glahn, Vera; Full professor, University of Girona, Girona, Spain. vera.pawlowsky@udg.edu
- Prof. Dr. Egozcue, Juan José; Emeritus professor, Technical University of Catalonia, Barcelona, Spain. juan.jose.egozcue@upc.edu
- Dr. Grunsky, Eric; Professor, China University of Geosciences Beijing; Adjunct Professor, University of Waterloo, Ontario, Canada. egrunsky@gmail.com

Course description

Compositional data are vectors that show the relative importance of the parts of a whole. Typical examples are data in percentages, ppm, ppb, or similar, common in many fields of science, particularly in geosciences. The classical statistical analysis of this type of data suffers from multiple problems, among them the one of spurious correlation. As a solution to these problems, J. Aitchison introduced the logratio approach in the 1980s. Since then, progress has been made in understanding the geometry of the sample space, the simplex of D parts.

The course aims to introduce attendees to the principles and basic methods of compositional data analysis; how to apply them with Codapack; and how to interpret the results obtained. The course combines theoretical classes with practical data analysis.

There will be a focus on the evaluation of multi-element geochemical survey data that will introduce the concepts of process discovery and process prediction based on CODA transforms.

Recommended prerequisites

1. Univariate statistical analysis.
2. Basic knowledge of multivariate statistics.
3. Introductory courses in algebra and calculus.
4. Experience with standard software: MS-Excel, SPSS, Minitab, R or similar.

Free software

1. CoDaPack 2.02.04 (Marc.Comas@udg.edu) is used in the course and is supplemented with MS-Excel. Downloadable from <http://www.compositionaldata.com>.
2. Compositions (in R) (Raimon Tolosana-Delgado r.tolosana@hzdr.de; Gerald van den Boogaart: boogaart@math.tu-freiberg.de).
3. rob-Compositions (in R) (Matthias.Templ@gmail.com).
4. zCompositions (in R) (Josep Antoni Martín-Fernández: josepantoni.martin@udg.edu; Javier Palarea-Albaladejo: javier@bioss.sari.ac.uk)
5. R Project for Statistical Computing: www.r-project.org
6. Quantum GIS: qgis.org/en/site/

Supplementary material (downloadable from www.compositionaldata.com)

- A. Lecture notes.
- B. CoDaPack 2.02.04.

Program

Day 1

- The Aitchison geometry in the simplex (1.5 hours).
- Exploratory analysis with CoDaPack: variation matrix, biplot (1.5 hours).
- The principle of working in coordinates. Regression. (2 hours).
- The CoDa-dendrogram. Orthonormal coordinates with CoDaPack.(2 hours).

Day 2

- The concepts of closure within a mineralogical and rock compositional data. (1 hour)
- Case studies (6 hours)
 - Process discovery and prediction within a diamond bearing kimberlite.
 - Remote predictive mapping of lithologies based on lake sediment geochemistry.
 - Characterization, process discovery and prediction for surface lithologies, thermoclimatic belts, ombrotypes, ecosystems and land cover derived from the continental US soil geochemistry survey.
 - Mapping tectonic assemblages (Age Brackets) from the Tellus Soil Survey data in Northern Ireland.
 - Mapping large tectonic assemblages from a low-density continental scale catchment soil geochemical survey across Australia.

Duration: 14 hours – 2 days, 7 hours daily, schedule 9-13 and 14-17.