

## Compositional data Analysis

*Prof. Dr. Michael Greenacre, Universitat Pompeu Fabra, Barcelona*

<b>Date &amp; Time:</b>	Oct 4 – 5, 2016	9 a.m. - 5 p.m.
<b>Location:</b>	AWI Building E	
	Oct 4: room E - 4025	
	Oct 5: room E - 2470	
<b>Language:</b>	English	
<b>POLMAR credit points:</b>	2	
<b>Registration:</b>	<a href="mailto:info.polmar@awi.de">info.polmar@awi.de</a>	

### Course content:

Compositional data are special in that each set of observations is set to add up to 1 (or 100%). This property leads to paradoxes and contradictions when regular statistics such as variances and correlations are computed on compositional data such as fatty acid compositions, or chemical or sedimentary compositions in geology.

Since multivariate methods such as principal component analysis depend on variances and covariances (or correlations for standardized data), these also suffer from the same problems. This issue has been known for a long time in the statistical literature but has so far not yet permeated into the biological literature, with some few exceptions. Ratios of compositions are resistant to the above-mentioned problems, so the modern approach to compositional data analysis relies on analysing ratios, usually on a logarithmic scale. The subject is treated at a more advanced level in the geological and geochemical literature, for example in the journal *Mathematical Geosciences*. The object of this two-day workshop is to explain these issues to an interdisciplinary audience and to demonstrate alternative ways of summarizing and analysing compositional data, both at the univariate and multivariate levels.

The R programming environment will generally be used in the practical part of the workshop, although other software packages can be used to do most (but not all) of the data analyses.

**Course program:**

**Day 1:** • Why are compositional data special? • The consequences of data "normalization" • Subcompositions and subcompositional coherence • Univariate analysis • Ratios and log-ratios • The problem of data zeros • Log-ratio distance • Component weighting • Applications and practical implementation in R and other software

**Day 2:** • Multivariate analysis of compositional data. • Background to principal component analysis and correspondence analysis • Log-ratio analysis • The effect of zero replacement • Correspondence analysis approximation to log-ratio analysis • Applications and practical implementation in R and other software

**Target group:** The course is primarily, but not exclusively, directed at advanced undergraduate and graduate students in ecology, marine biology, the geosciences and oceanography. For the most part and where possible, the level of instruction will take into consideration the spread of disciplinary backgrounds of course participants. For additional guidance please see the recommended reading or contact the instructor.

The language of instruction and discussion in class will be English.

**Pre-requisites:** The participants are expected to have a basic background in descriptive and inferential statistics, and some experience with statistical software. Some experience with the R package is recommended, but we will nevertheless start the computing sessions "from scratch".

**More information:** The textbook '*Multivariate Analysis of Ecological Data*' (Greenacre & Primicerio, 2014) provides some of the main background reading for the course (especially Chapter 14: Compositional Data and Log-Ratio Analysis). Participants are invited to inspect the opening chapters of the book before the beginning of the course (all the PDFs of this book can be freely downloaded from [www.multivariatestatistics.org](http://www.multivariatestatistics.org), as well as Michael Greenacre's book '*Biplots in Practice*' (2010), which is also relevant for this course (especially Chapter 7: Log-ratio biplots).

Course participants are kindly asked to bring a laptop (with wireless network card) to access data and to be able to complete the R sessions during the afternoon PC labs.

**Contact details:**      michael.greenacre@gmail.com

---

*Our courses are generally free of charge for all participants. However, they do have a price and can cost POLMAR / the AWI as much as 150 € per day per student. Please take this into account when cancelling your place at the last minute.*

---