Compositional data are special in that each set of observations consists of values adding up to 1 (or 100%). This property can lead to paradoxes and contradictions when regular statistical measures such as variances and correlations are computed on compositional data, for example on fatty acid compositions in biochemistry.

Since multivariate methods such as principal component analysis and regression depend on variances and covariances (or correlations for standardized data), these also suffer from the same problems. 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, i.e. so-called "logratios". The object of this 1.5-day workshop is to demonstrate these alternative ways of summarizing and analysing compositional data using logratios, both at the univariate and multivariate levels. The approach taken is essentially pragmatic, aimed principally at solving the research question with the simplest and most appropriate statistical methods and with the maximum participation by the researchers in the field in the statistical learning process.

Since the type of data analysed by the target audience for this course involves a very high number of compositional parts, the topics to be treated will include the following:

- the use of classical SVD-based methods of dimension reduction, called logratio analysis in this context;
- clustering of parts as a way to reduce dimensionality;
- variable selection as a way to reduce dimensionality;
- the problem of zeros in compositional data;
- the correspondence analysis alternative, which can be measurably close to the logratio approach, and which has no problem with data zeros.

The R programming environment will be used in the practical part of the workshop. Participants should preferably have their own laptops, with R and RStudio installed.

Michael Greenacre is Professor of Statistics at the Universitat Pompeu Fabra, affiliated to this university since 1994, where he teaches, among other courses, Data Visualization in the Master of Data Science. He has taught short courses in 15 countries. Author of over 100 peer-reviewed journal articles and book chapters, as well as 10 books, his most recent book is *Compositional Data Analysis in Practice* (Chapman& Hall / CRC Press, 2018), which is accompanied by the R package easy CODA. His recent research activities are centred around ecological research in the Arctic.

Web:  www.econ.upf.edu/~michael
      www.youtube.com/StatisticalSongs
Email:  michael.greenacre@upf.edu