An overview of compositional data analysis with applications to the biosciences

Javier Palarea

Biomathematics and Statistics Scotland, Edinburgh, Scotland, UK

Multivariate observations referring to parts of a whole, so-called compositions or compositional data in statistics, are common across varied scientific fields. This is for example the case when dealing with chemical concentrations, food nutritional contents, election vote shares, time-use and behavioural patterns, relative abundances, amongst others. Their distinctive feature is that there is an inherent interplay between the parts of the composition and, hence, the information they convey is fundamentally relative to each other, rather than absolute as ordinary statistical methods assume. This can in first instance cause some technical difficulties with ordinary data analysis and modelling, such as perfect multicollinearity and spurious correlations; but beyond that, it affects the way we think about a scientific problem, the statistical analysis to be conducted, and the interpretation of its results. In this talk I will provide an overview of the topic through some recent developments and applications in different domains of the biosciences.