Interpretation of a compositional time series

R. Tolosana-Delgado (1) and K.G. van den Boogaart (2)

(1) Universitat Politècnica de Catalunya, LIM - CIIRC, Spain (raimon.tolosana@upc.edu), (2) Technische Universität Bergakademie Freiberg, Institute for Stochastics, Germany (boogaart@math.tu-freiberg.de)

Common methods for multivariate time series analysis use linear operations, from the definition of a time-lagged covariance/correlation to the prediction of new outcomes. However, when the time series response is a composition (a vector of positive components showing the *relative* importance of a set of parts in a total, like percentages and proportions), then linear operations are afflicted of several problems. For instance, it has been long recognised that (auto/cross-)correlations between raw percentages are spurious, more dependent on which other components are being considered than on any natural link between the components of interest. Also, a long-term forecast of a composition in models with a linear trend will ultimately predict negative components.

In general terms, compositional data should not be treated in a raw scale, but after a log-ratio transformation (Aitchison, 1986: *The statistical analysis of compositional data*. Chapman and Hill). This is so because the information conveyed by a compositional data is *relative*, as stated in their definition. The principle of working in coordinates allows to apply any sort of multivariate analysis to a log-ratio transformed composition, as long as this transformation is invertible. This principle is of full application to time series analysis.

We will discuss how results (both auto/cross-correlation functions and predictions) can be back-transformed, viewed and interpreted in a meaningful way. One view is to use the exhaustive set of all possible pairwise log-ratios, which allows to express the results into $D(D-1)/2$ separate, interpretable sets of one-dimensional models showing the behaviour of each possible pairwise log-ratios. Another view is the interpretation of estimated coefficients or correlations back-transformed in terms of compositions. These two views are compatible and complementary.

These issues are illustrated with time series of seasonal precipitation patterns at different rain gauges of the USA. In this data set, the proportion of annual precipitation falling in winter, spring, summer and autumn is considered a 4-component time series. Three invertible log-ratios are defined for calculations, balancing rainfall in autumn vs. winter, in summer vs. spring, and in autumn×winter vs. spring×summer. Results suggest a 2-year correlation range, and certain oscillatory behaviour in the last balance, which does not occur in the other two.