

Professor Alan Welsh

The Australian National University (ANU)

Colours and Cocktails: Compositional Data Analysis

J.L. Scealy & A.H. Welsh

The Australian National University

The different constituents of physical mixtures such as coloured paint, cocktails, geological and other samples can be represented by d -dimensional vectors called compositions with non-negative components that sum to one. Data in which the observations are compositions are called compositional data. There are a number of different ways of thinking about and consequently analysing compositional data. The log-ratio methods proposed by Aitchison in the 1980s have become the dominant methods in the field. One reason for this is the development of normative arguments converting the properties of log-ratio methods to 'essential requirements' or Principles for any method of analysis to satisfy. We discuss different ways of thinking about compositional data and interpret the development of the Principles in terms of these different viewpoints. We illustrate the properties on which the Principles are based, focussing particularly on the key subcompositional coherence property. We show that this Principle is based on implicit assumptions and beliefs that do not always hold. Moreover, it is applied selectively because it is not actually satisfied by the log-ratio methods it is intended to justify. This implies that a more open statistical approach to compositional data analysis should be adopted.