COMPOSITIONAL DATA ANALYSIS IN ENVIRONMENTAL DATA INTERPRETATION

Zbigniew Ziembik, Agnieszka Dołhańczuk-Śródka, Andrzej Kłos, Małgorzata Rajfur, Maria Zielińska, Piotr Godyń, Dominik Jerz

University of Opole, Department of Biotechnology and Molecular Biology, Opole, Poland; e-mail: ziembik@uni.opole.pl

ABSTRACT

In environmental studies analysis of biota compositions provides valuable information about the occurring processes and the over all condition of the living organisms. Usually an analysis is based on the temporal and spatial changes in concentrations of chemical compounds or elements in biota and in mineral components of the environment.

Usually in studies only parts of the objects under interest are available. Concentration of some selected components can be determined and then, after the result's analysis, the inferences are drawn. Unfortunately, the conclusions regarding abundances (for example masses) of the system components and relationships between them which are drawn directly from concentration can be delusive or erroneous.

Concentrations, being compositional data, possess some specific features. They are not independent on each other, if content of one of components increases the others have to decrease. It is the reason for which results of standard statistical analysis of the relationships between raw components or parts are spoiled by spurious effects.

Among problems occurring in concentrations analysis the following can be mentioned:

- the loss of information regarding original abundance distribution,
- the scale variance – the analysis result is related to the unit of the concentrations,
- the limited range of compositional variables changes, for ratio it is <0,1>,
- the subcompositional incoherence – a relationship between the selected concentrations is affected by concentrations of other components,
- the relationship between variance of the component concentration and covariances with concentrations of other components.

Some of these problems can be solved moving the data analysis to simplex sample space. In this approach specific properties of simplex space must be considered. The other solution involves the data transformation what enables utilization of the standard statistical methods in the analysis.

Impact of climate change and pollution on vegetation distribution and condition in the temperate, boreal, alpine and arctic zones of northern Europe (http://polar.uw.edu.pl/), Warsaw, 26-27.10.2016