Title: Novel visualisation tools for exploring compositional data and interpreting statistical models with compositional predictors

Dr. Rishabh Vishwakarma (Trinity College Dublin)

Abstract:

Compositional data are characterised by their variables that are non-negative proportions or percentages with a constant sum constraint (usually 1 or 100%) and reside in the simplex space. Compositional variables arise in multiple scientific domains such as ecology, geology, nutritional studies, demographic studies, etc. and are traditionally modelled as responses. However, they can also be employed as predictors in a model with continuous (or discrete) responses in which case the variability of the response across the simplex space becomes the question of interest. The Diversity-Interactions modelling (DI) framework is designed for such applications and is primarily used in biodiversity and ecosystem function (BEF) research to analyse the effects of proportions of flora, fauna, or microbiota in the ecosystem on its productivity. These models can be fit using the DImodels and DImodelsMulti R packages. However, interpreting them is challenging due to the compositional nature of the predictor space.

We introduce the PieGlyph and DImodelsVis R packages developed to provide novel visualisations for exploring compositional data and interpreting regression models with compositional predictors. PieGlyph can be used to overlay any traditional plot with axis invariant pie-chart glyphs to supplement the plot with additional information about the compositional variables. DImodelsVis introduces novel visualisations such as improved model diagnostic plots for models with compositional predictors, effects-plot for visualising the response-predictor relationship across any dimension in the simplex space and conditional ternary diagrams for visualising the response surface across higher dimensional simplexes. We demonstrate the utility of these visualisations using some landmark BEF datasets. The two visualisation software packages are primarily designed to complement and integrate seamless with the DImodels and DImodelsMulti packages. However, special attention has also been given to ensure that they have broad applicability and are versatile to be used with any standard statistical model object in R (around 300 different model objects are supported) provided the predictor space is compositional in nature.

This is joint work with Caroline Brophy (TCD), Laura Byrne (TCD), and Catherine Hurley (MU).