LegumeLegacy Doctoral Network
SS4 Foundations of Statistics Training Event
17th September 2024

LegumeLegacy

SS4 – Introduction to Diversity-Interactions modelling
Lab Session

A dataset has been provided in the file: `Dataset_4species_simulated.csv`. The data in this file is simulating a biodiversity and ecosystem function (BEF) experiment. The diversity of four grassland species was manipulated across 50 plots and treated with low nitrogen fertiliser (lowN) and repeated at regular nitrogen fertiliser (regN), giving a total of 100 plots. The first two species are grasses, and the second two legumes. The yield of each plot over a growing season was recorded.

1) Explore the dataset, e.g., check the design and what each variable in the dataset represents, and generate some graphs of the response and its trends with the manipulated variables.

2) Carry out an initial model selection process to identify the best Diversity-Interactions model for this data. For this part, use the `autoDI` function in the `DImodels` R package.
   a) Write out (on paper) the equation of each model that is fitted within the `autoDI` model selection process (not the R code version, the mathematical algebraic version!).
   b) For the best model selected using `autoDI`, use the model to predict the yield response for a 50:50 mixture of the two grasses managed at low N (do the calculation by hand on paper).
   c) Generate a graphic to illustrate the outcome of the model selected by `autoDI`.

3) Continue the model selection process, this time using the `DI` function in the `DImodels` R package, to explore models not fitted via `autoDI`.
   a) Write out (on paper) the equation of each model that you fit as you work through the selection process (the mathematical algebraic version).
   b) For your final selected model, fit the model again, this time using the `lm` function in R.

4) For the best model from the previous step, write out (on paper) the equation of the model (the mathematical algebraic version) and state the model assumptions. Test the model diagnostics to verify if the assumptions are reasonable. Continue with model fitting if required.

5) For the final model, examine the model coefficient estimates and explain the meaning of each. Can you explain the parameterisation of the model and consider alternative parameterisations for fitting same model? (For example, with DI models, we don't normally fit an intercept, but we can fit an intercept and it will be the same model fitted, just a different parameterisation.)

6) For the final model, what is the predicted yield for the following communities? (Generate by hand / on paper): a grass (species 1) monoculture treated with low and (separately) regular N; an equi-proportional mixture of the two legumes treated with low and (separately) regular N. Generate a few other predictions of your choice.

7) Generate graphs that will show the interpretation of the fitted final model. What is the effect of the species diversity gradient? Of the N fertiliser treatment?

8) Write up (e.g., in Word) a 1-page report on your analysis. Include at least one figure that illustrates the key result(s). (It can go to two pages if needed, but aim for one page.)

The LegumeLegacy project has received funding from the European Union's Horizon 2021 Doctoral Network programme under the Marie Skłodowska-Curie grant agreement No. 101072579