Homework 15

The file 'macroinverts.csv' contains data from a survey of stream invertebrate communites in South-East Australia, as reported in the attached paper by Szöcs et al. The authors were interested primarily in effects of agriculture on the invertebrate community, particularly whether salinization of freshwater and pesticide use alter community composition. There are 60 samples in the dataset, which were taken at three different times, at a number of locations that differ in environmental characteristics. The attached file includes data on many enviromental parameters. The water characteristics are temperature (T), pH, conductivity (Cond = salinity), oxygen, phosphate, and turbidity. And the substrate characteristics are percent bedrock, boulder, cobble, pebble, gravel, sand, and clay/silt. There is also a column 'tmax', which records pesticide toxicity as measured with a bioassay. Finally, there are counts identified by taxonomic family (all the remaining columns in the dataset).

To start off, see how the environmental parameters covary with each other. Do a PCA on the environmental predictors. To make the results more meaningful, you'll probably want to log- or square-root-transform some of the predictors, because some of them are highly skewed. How much variation is explained by the PC axes? Are there dominant axes of variation, or is the environmental variation more complex than that? Make a biplot to look at how the variables load onto the first two axes. Also make a biplot for the third and fourth axes.

Eventually you're going to do a constrained ordination, using environmental variables to predict invertebrate community structure. But right now there are just too many predictors (and I didn't even include all those the authors looked at). Reduce the number of variables in the environmental data to a maximum of 6. Picking variables to exclude is a judgment call, but if two variables seems pretty strongly correlated with each other, then dropping one of them makes sense. Or you can use envfit() or capscale() to drop variables that don't seem important for community composition. It's up to you.

Now use NMDS to ordinate the invertebrate community data. Some of the families are quite rare, so drop those that occur less than 10 times. You should also transform the count data before using it to calculate dissimilarities. I've been using the wisconsin square root transformation in class, but it's not the only option. Try NMDS with 2-4 dimensions. How much does the stress change with the # of dimensions?

Perform a principal coordinates analysis (with capscale), and compare the ordination to the NMDS ordination using procrustes(). Do the two methods give similar results? How much variation in community composition is explained by the first few PCoA axes?

Using the subset of environmental variables you chose earlier, use envfit() to see how the environmental variables relate to the NMDS ordination axes. Plot the results. Which variables seem most strongly related to the major axes of community

variation?

Now use constrained ordination (canonical analysis of principal coordinates) to more directly look for axes of community variation that are explained by the environment. Plot the site scores and the biplot arrows for the environmental axes. How are the environmental variables aligned along the most important axes? How much variation in community composition is explained by those axes? Test whether the axes are 'significant', using the permutation tests provided by the anova() function. Use a plot of the species scores, in combination with the environmental vectors, to see which invertebrate families respond most strongly to the different environmental axes.

You've been looking at overall variation in community composition, but the authors of the study were most interested in the effects of salinity and pesticides. Use permanova (i.e., adonis()) to test the effect of these predictors on community composition. Are there significant effects? How much variation is explained? Is there an interaction between the two predictors? Would you consider these to be strong environmental impacts, at least in terms of changing the structure of the stream invertebrate communities?