AJ SMIT

QUANTITATIVE ECOLOGY

Contents

List of Tables

List of Figures

Preface

Placeholder

1
Introduction

> *"We have become, by the power of a glorious evolutionary accident called intelligence, the stewards of life's continuity on earth. We did not ask for this role, but we cannot abjure it. We may not be suited to it, but here we are."*
>
> — Stephen J. Gould

1.1 Modern ecological problems

This is a course about community ecology and not so much about population ecology. Community ecology underpins the vast fields of biodiversity and biogeography, and concerns spatial scales from squares of meters to all of Earth. We can look at historical, contemporary, and future processes that have been implicated in shaping the distribution of life on our planet.

Community ecologists tend to analyse how multiple environmental factors act as drivers that influence the distribution of tens or hundreds of species. These data tend to often be messy (not in the sense of untidy data as per the 'tidyverse' definition of tidy data, but it can be that too!) and statistical considerations need to be understood within the context of the data available to us. This translates to errors of measurement and errors due to extreme values, the presence of a few very rare or very abundant species, autocorrelated residuals (due to repeated sampling, for example), colinearity, etc. These challenges make to application of 'basic' statistical approaches problematic, and a new branch of inferential and exploratory statistical needs to be followed. These approaches involve techniques that allow us to work with all the data at once, and because it can simultaneously analyse all the variables (multiple environmental drivers acting on multiple species at multiple places and across multiple times), this group of statistics is called 'multivariate statistics.' There are two main groups of multivariate statistics: 'classifications' and 'ordinations.' Classification generally concerns placing samples (species or environments) into units or groups, while ordination is best suited for analyses that involve arranging samples along gradients. Irrespective of the analysis, the data share a few characteristics.

1.2 Where do ecological data come from?

Information is all around us. It has existed before sentient humans began questioning "Life, the Universe and Everything." During these early times, humans often invented silly answers, especially during the time before the scientific age when the tools and ways of thinking about problems became available. Today we have access to the ways and means to question the world around us, and we may arrive at objective answers (which are hopefully no longer silly). This module concerns large amounts of quantitative data (information turned into numbers) about our world. These quantitative data have been collected over many hundreds of years (give examples of long data sets), and they continue to be collected at increasing rates, over increasing spatial scales, and at finer and finer resolution. And because of advanced deterministic general circulation models that offer a predictive capability, which may be coupled via an ecophysiological understanding of how plants and animals and things that are neither plants nor animals react to environmental stimuli, we may project how the biota may respond in the future.

Let us consider some of the sources of information (data) that we will be able to analyse and turn into knowledge.

1.2.1 Field sampling

1.2.2 Historical data

1.2.3 Remotely sensed data

1.2.4 Modelled data (projections)

1.3 What do we do with these data?

We follow the principles of reproducible research, and throughout we will implement and practice modern data analytical methods. These approaches concern i) data entry, ii) data management, iii) data-wrangling, iv) analysis, and v) reporting, and we shall discuss each under the next headings.

1.3.1 Initial data entry

1.3.2 Meta-data and data management

1.3.3 Data-wrangling: pre-processing and quality assurance

1.3.4 Analysis

1.3.5 Reporting

2
Ecological data

Placeholder

2.1 Properties of the data sets

2.2 Field sampling

2.3 Historical data

2.4 Remotely sensed data

2.5 Modelled data (projections)

3
Exploring the data

3.1 In the beginning

At the start of the analysis, before we get to the hypothesis testing and multivariate approaches, we need to explore the data and compute the various synthetic descriptors (refer to the Basic Statistics Workshop). Recall that this exploratory data analysis includes the various measures of central tendency and dispersion/variation around the mean or median. Using the techniques that we already know, we can produce data summaries and statistical visualisations for each variable independently. So, we say that we produce univariate summaries, and if there is a need, we may also wish to include some of the univariate inferential statistics. Be guided by the research questions as to what is required. Typically, I don't like to produce too many detailed inferential statistics of the univariate data, choosing in stead to see which relationships and patterns emerge from the exploratory summary plots before testing their statistical significance using multivariate approaches. But that is me. Sometimes, some hypotheses call for a few univariate inferential analyses.

3.2 Describe the environment

3.3 Describe the community data

4
Pairwise association matrices

5

Cluster analysis

6

Unconstrained ordination

7

Canonical ordination

8
Bibliography