PhD course: Advanced statistics in practice

(February - May 2021)

Aim
This course aims to fill this gap by preparing students to analyze, interpret and report their data using the most up-to-date methods in R. Special attention will be given to discussion of questions springing from the students' own work and the biological interpretation of data. The course is intended to deepen the students' understanding of all aspects of ecological inferences, not as a cookbook of "how to".

Credits: 5 ECTS/HEC (each unit equals one credit; students can choose to participate in one or more units)

Language: English

Prerequisites: Admitted to PhD-studies. Basic knowledge of statistics and R

Course structure
We will use a set of activities including initial readings, classroom discussions, hands-on activities, followed by home exercises. Each unit corresponds to one theme/workshop and is planned for one full day (8:45 - 17:00; including 1 hour lunch and 2 fika breaks, *note exception Meta-analysis=2 days*), equaling one credit (ca 27 hours of work for the PhD student). Example datasets will be provided, but students could use their own datasets as well.

Schedule

February 1	Time series analysis
February 15	GIS and spatial analysis
April 12-13	Meta-analysis
May 3	Getting more out of community data
May 31	Dealing with some complexities of GLMs

Location: The course is planned to be delivered either in real life (given an end to social distancing) in Ultuna campus or online (Slack/Zoom) depending on the pandemic situation.

Learning outcomes
By the end of the course, students are expected to be able to

1- Demonstrate the ability to identify relevant functions and packages in R for analyzing their own datasets.
2- Analyze data using R, including reporting advanced statistics, univariate and multivariate modelling and generation of graphs.
3- Interpret, think critically and draw conclusions on data analysis results.

Assessments
No formal examination will be used, and students will be evaluated based on their overall engagement in classroom activities and fulfilling the assignments.

Course content
The course is divided in five units corresponding to five topics as follow:

Unit 1: **Time series analysis**. *Teachers: Jonas Knape and Örjan Östman.*
1 February 2021

1. Introduction to time series
   - What a time series is, how it differs from other data types, autocorrelation.
   - Basic time series models such as white noise, random walks and autoregressive models.
   - Connections to simple models of population dynamics, such as stochastic exponential growth.
2. AR, ARMA, structural changes, MAR, and multivariate (communities) statistics in time-series analysis.
3. State space models
4. Tools for time series analysis
5. Time series analysis of Swedish bird survey data.

Unit 2: **GIS and spatial analysis.** *Teachers: Alistair Auffret and Mohammad Bahram.*
15 February 2021.

1. Intro: Why GIS in R? Which packages?
2. Vector analysis: Buffer zones, Intersections, Summary statistics etc.
3. Raster analysis: Overlays, Reclassification, Raster calculator
4. Spatial autocorrelation and its effect on your results, strategies and methods for accounting or using spatial autocorrelation
5. Analysing spatially-structured ecological data

Unit 3: **Meta-analysis** *Teacher: Julia Koricheva*
12-13 April 2021

1. Formulating the research question for a meta-analysis
2. Data extraction from primary studies
3. Calculation of effect sizes
4. Combining effect sizes across studies using fixed and random effect models
5. Exploring causes of heterogeneity across studies (meta-regression)
6. Sensitivity analysis and testing for publication bias

Unit 4: **Getting more out of community data** *Teachers: Tomas Roslin and Giovanni Strona.*
3 May 2021

Matching the focus of the course on interpretation rather than methods, this unit will depart not from a specific method but from a series of ecological questions. Thus, we will start by defining fundamental questions on communities, then point to methods for approaching them:

1. *Fundamental question:* What can co-occurrence data reveal about ecological interactions?
Methodological application: How do we quantify co-occurrence? How do we separate a "real" pattern from one caused by chance alone (null model analysis)?

2. *Fundamental question:* More generally: if there is structure, then what fundamental forces are behind it? What signals of such forces are hidden in the data?
Methodological application: How can the signals of community assembly be translated into a parameterized statistical model? (HMSC)
3. *Fundamental question:* Moving from patterns in the abundance and distribution of species to direct observations of who interacts with whom: can we identify the forces structuring interaction networks?
Methodological application: Analyses of trait-matching using recent techniques

Unit 5: **Dealing with some complexities of GLMs** *Teacher: Matt Low.*
31 May 2021

1. Understanding and comparing the Normal, Poisson and Binomial distributions.
2. What is overdispersion and what effect does it have on your results? How can it be quantified, corrected for, and what does it all mean biologically?
3. What is zero inflation? How zero-inflation can be thought of as a form of overdispersion (and hence dealt with using the previous solutions). How zero-inflation can be considered as a combination of two separate processes and modelled using either hurdle models or mixture models.

Teachers: Alistair Auffret (SLU), Mohammad Bahram (SLU), Jonas Knape (SLU), Julia Koricheva (Royal Holloway University of London, UK), Matt Low (SLU), Tomas Roslin(SLU), Giovanni Strona (University of Helsinki, Finland), and Örjan Östman (SLU)

Number of participants per unit: maximum 20

Register for the course (all or separate units) by sending an email to the course organizer Mohammad Bahram (mohammad.bahram@slu.se) **no later than November 29 and indicate what units you are interested in.**