Analysing Palaeolimnological Data with R

http://www.staff.ncl.ac.uk/staff/stephen.juggins/courses/PalaeoDataWithR.htm

University Marine Biological Station Millport, Isle of Cumbrae, Scotland

16th – 20th August 2012

Course tutors:

Dr Steve Juggins, School of Geography, Politics & Sociology, Newcastle University, Newcastle upon Tyne, NE3 4LP. Stephen.Juggins@ncl.ac.uk

Dr Gavin Simpson, Department of Geography, University College London, Gower Street, London WC1E 6BT. gavin.simpson@ucl.ac.uk

Introduction

Analysing Palaeoenvironmental Data with R is an intensive 4-day course that will provide participants with a training in the theory and practice of analysing palaeolimnological data using the free R software environment. The workshop has been timed to coincide with the 12th International Paleolimnology Symposium (http://paleolim.org/ips2012/), 20-24th August 2012, which will be held in Glasgow, to reduce the expense of attendance for those who will be attending the symposium.

The R statistical language and environment has become increasingly popular in recent years, in part because it is a free, open source application and because it is incredibly powerful and easily extended via add-on packages. This course is aimed at those with little or no experience in R and will address both the essential numerical understanding and the R skills required to handle, process and analyse palaeolimnological data.

The workshop will be led by Dr Steve Juggins and Dr Gavin Simpson. Steve is a co-editor and both are chapter authors in the forthcoming handbook on “Data Handling and Numerical Techniques” to be published by Springer in 2012, and both have authored several R packages for handling palaeolimnological and palaeoenvironmental data.

Prerequisites:

An undergraduate course in statistics, an understanding of basic concepts such as correlation and regression, and familiarity with PC or Mac software for data analysis.

For those new to R we will provide a self-led tutorial on R basics that you must complete before the course.

You must bring your own laptop with R already installed.

Bring your own data to work on!

Logistics and cost

The course will be held at the University Marine Biological Station Millport, on the Isle of Cumbrae in the Clyde Estuary (http://www.gla.ac.uk/centres/marinestation/). The station is an excellent location for the course with easy travel links to Glasgow and convenient for those also attending the 12th IPS.

The course will be limited to c. 30 participants, with accommodation in shared (twin rooms). The total cost of the course is £300 + VAT, which includes food, accommodation and course materials. (Note: there may be limited accommodation in single rooms available at extra cost. Please let us know if you are interested in this option).

The course has generous financial support from PAGES (http://www.pages.unibe.ch/) to cover travel and course costs (up to £1100) for five young researchers from developing countries. If you would like to apply for PAGES financial support please send a CV and short covering letter outlining your research interests and why the course will benefit you to both Gavin and Steve.

The deadline for registration is 31/06/2012 (15/05/2012 for those applying for PAGES financial support). Please see the website listed at the top of this document for registration details.

Resources

A full reading list will be included with the course materials. In the meantime we recommend the following two books:

Borcard, D., Gillet, F., & Legendre, P. (2011) *Numerical Ecology with R* Springer.

Zuur, A., Ieno, E., & Meesters, E. (2009) *A Beginner's Guide to R* Springer.

R may be downloaded from http://cran.r-project.org/

Provisional course schedule:

The provisional schedule is based on 45 minute lectures and 1 hour practicals. The evenings will be used for short presentations on R tips, advanced graphics etc. and for you to work on your own data.

16th - Thursday: Travel to Cumbrae, arrive in time for dinner.

Evening:

- Welcome, logistics and course overview (30 mins)
- Set up laptops, connect to network etc.
- Lecture - basic R revision, data types, conversions, importing data etc. (30 mins)
- Finish evening with informal chat where students introduce themselves and their work

17th – Friday

Morning:

- Basic EDA & graphics, data transformation, outliers etc.
- EDA Practical
- Regression
- Regression Practical

Afternoon:

- Multiple regression (introduce ideas of model building, variance partitioning and variable selection)
- Multiple regression practical
- Modern regression methods
- Modern regression practical

Evening:

- R tips – advanced import, export, common pitfalls
- Finish practicals, work on own data

18th – Saturday

Morning:

- Distance measures and cluster analysis
- Cluster analysis practical
- Direct ordination (PCA, CA, DCA NMDS)
- Direct Ordination practical

Afternoon

- Constrained ordination I (CCA, RDA)
- Constrained Ordination practical
- Constrained ordination II – variable selection, partial ordination, variance partitioning, permutation tests
- Constrained Ordination II practical

Evening:

- Graphics tips

- Finish practicals, work on own data

19th – Sunday

Morning:

- Environmental reconstructions
- Environmental reconstruction practical

Afternoon

- Age-depth modelling
- Age-depth modelling practical

Evening:

- Plotting time series and stratigraphic data in R
- Finish practicals, work on own data

20th – Monday – morning session only, depart for IPS in Glasgow after lunch

Morning:

- Analysis of stratigraphic data (zonation, rate of change, smoothing, interpolation)
- Stratigraphic data practical
- Work on own data