

Highland Statistics Ltd

6 Laverock road, AB41 6FN, Newburgh, UK
Tel: +44 (0)1358 788177
Email: highstat@highstat.com
URL: www.highstat.com
VAT No. GB 827 5870 93

Now in its 9th year with over 7000 participants. Highland Statistics Ltd. has provided more than 125 courses at various places around the world. These courses are especially designed for biological and environmental scientists.

Books by the instructors Dr. Alain F. Zuur & Dr. Elena Ieno



Courses provided by Highland Statistics

Course	Content	Days	Pre-knowledge
1	Data exploration and visualization in R	3	-
2	Beginner's Guide to R (classroom or online course)	3	-
3	Data exploration, regression, GLM & GAM in R	5	Basic stats
4	Linear mixed effects modelling with R	5	Courses 2 & 3
5	Zero inflated models and GLMM in R	5	Courses 2 – 4
6	Multivariate analysis in R	5	-
7	Online follow-up course for: Data exploration, regression, GLM & GAM	1-2	Courses 2 & 3



Courses can be run in-house or as open courses, and can be provided at most places in the world. These courses are especially designed for biological and environmental scientists.

Details for each course

Course 1: Data exploration and visualization in R.

This course is based on our upcoming book *Data exploration and visualization with R*, published by Springer, and our 2010 paper: “A protocol for data exploration to avoiding common statistical problems” in *Methods in Ecology and Evolution*.

Course 2: A Beginner’s Guide to R

This can either be a classroom course or an online course. The course material is based on Chapters 1 – 9 of our *Beginner’s Guide to R* book. See the TOC of this book for a detailed outline.

In the online course participants have access to an online meeting website where they can view on-demand video presentations, ask questions on the exercises and discuss the problems/solutions with other participants and the instructors.

Course 3: Data exploration, regression, GLM & GAM in R

An important tool used in statistics is linear regression. Various basic linear regression topics will be explained from a biological point of view. We will discuss potential problems and show how GLM can be used to analyse count data or presence-absence data. Sometimes, parametric models (linear regression, GLM) do not quite fit the data and in such cases GAM (a smoothing technique) can be used.



Course outline:

- Day 1: Introduction to R. Data exploration (outliers, collinearity, transformations, relationships, interactions).
- Days 2 & 3: Linear regression (theory, model selection; stepwise versus information theory approach, interactions, sketching model fits).
- Day 4: GLM for count data, binary data and proportional data (Poisson, negative binomial, binomial distributions) and how to deal with overdispersion.
- Day 5: GAM for count data, binary data and proportional data. What to present in a paper or thesis.

We will follow Chapters 4 & 5 of Zuur et al. (2007) and Chapters 8 – 9 of Zuur et al. (2009). Pdfs will be provided.

Course 4: Linear mixed effects modelling with R

Linear mixed effects models are explained and we apply them on 1-way and 2-way nested data. We also show how to deal with heterogeneity (GLS). Temporal and spatial correlation structures are added to the models. We will follow Chapters 4 – 7 of Zuur et al. (2009).

Course outline:

- Monday: Dealing with heterogeneity in linear regression models (GLS). This is Section 4.1 in Zuur et al. (2009). Exercise 1: Application of GLS, see Section 4.2.
- Tuesday: Explanation of linear mixed effects models for nested data. We will discuss the random intercept model, random intercept and slope model, REML, intraclass correlation, lme versus lmer, p-values. This is based on Sections 5.1 to 5.9 in Zuur et al. (2009). Exercise 2: Application of mixed modelling on bee data. Chapter 19 contains a 12-page solution, including what to present in a paper. Exercise 3: Application of mixed modelling combined with GLS techniques on a rat data set.
- Wednesday: Exercise 4: Application of mixed effects modelling and additive mixed effects modelling on begging behaviour of nestling barn owls. This is Section 5.10. Theory on mixed effect models for 2-way nested data. Exercise 5: Application of mixed effects models on 2-way nested cetacean data. The analyses of this data set is presented as a case study chapter in Zuur et al. (2009), see Chapter 20. We will also hand out a published spin-off paper based on this chapter.



- Thursday & Friday: Adding more complicated correlation structures (spatial/temporal) to linear regression and mixed effects models, and smoothing models. This is based on Chapters 6 and 7 in Zuur et al. (2009). Exercise 6: Analysis of a time series data. Exercise 7: Analysis of a spatial forestry data set.

Course 5: Zero inflated models and GLMM in R

What is zero inflation? Suppose you want to study hippos and the effect of habitat variables on their distribution. When sampling, you may count zero hippos at many sites and as a result standard statistical techniques like regression and GLM are not applicable, therefore zero inflated models should be used.

Keywords: Poisson and negative binomial GLM. Overdispersion. Zero inflation. True and false zeros. Hurdle models. ZIP, ZAP, ZINB, ZANB GLM and GAM models. R code. Bayesian statistics, MCMC, 1-way and 2-way nested data. Temporal and spatial correlation.

Course outline:

- Monday: General introduction. Presentation of the Poisson, negative binomial, Gamma and binomial distributions. This is Chapter 8 in Zuur et al. (2009). Presentation of GLM. This is Chapter 9 in Zuur et al. (2009). Two exercises on Poisson GLM and negative binomial GLM. Presentation of binomial GLM for binary data. This is part of Chapter 10 in Zuur et al. (2009). One exercise on binomial, Poisson or negative binomial GLM.
- Tuesday: Presentation of models for zero truncated data. Exercise on zero truncated data. Presentation of models for zero inflated data. This is Chapter 11 in Zuur et al. (2009). We will discuss hurdle models (ZAP and ZANB) and zero inflated models (ZIP and ZINB).
- Thursday: Introduction to Bayesian statistics and MCMC. Fitting GLM and zero inflated models in WinBUGS.
- Friday: Zero inflated models (and GLMMs) for 1-way, 2-way nested data and adding temporal or spatial correlation structures.

We will explain zero inflated models for count data (and density data). We will follow Chapter 11 in Zuur et al. (2009), and Chapters 1, 2 and 6 in Zuur et al. (2012).



Course 6: Multivariate analysis in R

Multivariate techniques are extremely useful and popular tools in biology. There are many multivariate techniques available. For example principal component analysis, correspondence analysis, redundancy analysis, canonical correspondence analysis, canonical correlation analysis, discriminant analysis and multidimensional scaling. Most of these techniques have a biological rationale, but the justifications differ between techniques. Therefore it is extremely crucial to know exactly why, and when to apply a particular technique, and when not to. Even more crucial is, how to correctly interpret the results in biological terms!

Course outline:

- Day 1: Introduction to R. Data exploration following Chapter 4 in Zuur et al. (2007). Outliers, collinearity, transformations, relationships, interactions. Exercises in R.
- Day 2: Measures of similarity (following Chapter 11 in Zuur et al. (2007). ANOSIM and the Mantel test to detect effects of covariates. MDS and NMDS. Exercises in R.
- Days 3 & 4: Principal component analysis and redundancy analysis (Chapter 12 in Zuur et al. 2007). Exercises in R.
- Day 4 & 5: Correspondence analysis and canonical correspondence analysis (Chapter 14 in Zuur et al (2007). Exercises in R.
- Day 5: Discriminant analysis (Chapter 15 in Zuur et al. 2007). Exercises in R.

Course 7: Online follow-up course for: Data exploration, regression, GLM & GAM

This is an online follow-up course for people who have participated in our data exploration, regression, GLM and GAM course in R. During a period of 12 weeks we discuss a series of 12 exercises (2 exercises per 2 weeks). Participants have access to an online meeting website where they can ask questions on the exercises and discuss the problems/solutions with other participants and the instructors. This course is **only open** to people who have participated in our data exploration, regression, GLM & GAM course. No statistical theory will be discussed.

Video files with audio comments of the instructor are available on-demand. You can see the instructor's screen, R solution code and audio comments on the R coding and statistical approaches.