Data exploration, regression, GLM & GAM

Thank you for joining the course on data exploration, regression, GLM and GAM. This website contains the exercises, data sets, solution files and also some useful pdfs.

Copyright statement

Course material

During the course there will be be no hand outs. Please download the following material before the start of the course.

Data exploration (Monday):

- Introduction: P1_STARTRegressionGLMGAM_V3.pdf Password to open this pdf file: Newburgh2015 **Do not copy any blank spaces after the password!!**
- Pdf of powerpoint presentation on data exploration (file name starts with P2). Password to open this pdf file: Newburgh2015
-
- Chapter 4 of Zuur et al. (2007).
- Our 2010 paper in *Methods in Ecology and Evolution*. (See this URL for a nice interview on the 2010 paper).

Multiple linear regression (Monday afternoon and Tuesday):

- Pdf file of the powerpoint presentation on bivariate linear regression (file name starts with P3). Password to open this pdf file: Newburgh2015
- Pdf file of the powerpoint presentation on ANOVA, ANCOVA and interactions (file name starts with P4). Password to open this pdf file: Newburgh2015
- Pdf file of the powerpoint presentation on multiple linear regression. (file name starts with P5). Password to open this pdf file: Newburgh2015
- Chapter 5 of Zuur et al. (2007).
- Slide with four options for model selection
- Pdf file of the Fixed X story. Password to open this pdf file: Newburgh2015
- Appendix A from our 2009 book.

Generalized Linear Modelling (Wednesday & Thursday):

- We will follow Chapter 2 from 'Beginner's Guide to GLM & GLMM with R'. (2013). Zuur, Hilbe, Ieno. This file is the powerpoint presentation (starting with P6) saved as a pdf. The password to open it is: **Newburgh2015**
  - **Do not copy any blank spaces after the password!!**
- More background information is given in Chapters 8, Chapter 9 & Chapter 10 of Zuur et al. (2009a). You can read these chapters after the course.
- Powerpoint file on VIF.

Generalized Additive Modelling (Thursday & Friday):

- We will follow Chapter 2 from 'Beginner's Guide to GAM with R' (2012). Zuur. This file is the powerpoint presentation (starting with P8) saved as a pdf. The password to open it is: Newburgh2015
  - **Do not copy any blank spaces after the password!!**
- More technical background information on GAMs is given in Chapter 3 of Zuur et al. (2009a).

To follow the course there is no need to buy any of our books. However, if you would like to buy one, see: http://www.highstat.com/bookorder.htm.

References:

- Zuur, Ieno, Smith (2007). *Analysing Ecological Data*. Springer.
- Zuur, Ieno, Walker, Saveliev, Smith (2009a). *Mixed effects models and extensions in ecology with R*. Springer.
- Zuur, Ieno, Meesters (2009b). *A Beginner's Guide to R*. Springer.
- Zuur, Ieno, Elphick. (2010). A protocol for data exploration to avoiding common statistical problems. *Methods in Ecology and Evolution*.
- Zuur, Hilbe, Ieno (2013). Beginner's Guide to GLM and GLMM with R.
- Zuur (2012). Beginner's Guide to GAM with R.

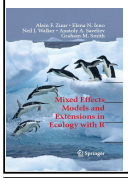
Course preparation

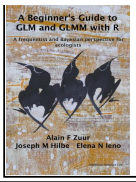
- Please install R and a decent text editor. It is irrelevant which R version you use; anything less than 6 months old is fine. As to the text editor, Tinn-R or RStudio are good choices. But you can also use the internal R text editor. The instructors will be using RStudio (it also runs on a Mac) and the internal R text editor.
- It may be wise to have a quick look at the material for Monday.
- Please download all data sets and R solution code from the table below (just in case there is no WIFI in the classroom).
- Store all files in a directory with a simple and short name. Something line C:/StatsCourse. Do not use accented characters (ń, ű, á, Ø, etc.) in the directory name!

Data sets and solution files

- Below are the data sets and R solution files. Please download them before the start of the course (just in case when there is no WIFI in the room). To download txt and R files, **YOU MUST USE: RIGHT-MOUSE CLICK and the SAVE AS**.
- Links to scientific publications describing the data are provided. Don't worry if you don't have access to these commercial journals, because each exercise contains a short description of the data.
- R survival guide and corresponding data file as a text file (or as an excel file). The survival guide contains example code for data exploration, regression, GLM and GAM. In order to do the exercises you only have to change the name of the data set and the variable names in the survival guide; the rest is copy-paste.

To download a txt or R solution file, use: RIGHT MOUSE CLICK and SAVE LINK AS

Number	Exercise description	Excel or txt file	R solution file
Support file which you need in each exericse: HighstatLibV9.R			
Data exploration			
1	Loyn exercise 1	loyn.xls 	Ex201A_LoynDataExploration_V3.r
2	Bailey exercise 2	Baileyetal2008.xls Link to paper	Ex202A_Bailey2008DataExploration_3V3.r

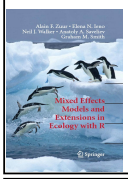
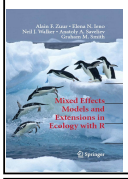
			

Monday evening exercise (or whenever you have internet access). Install packages.
Run the code in this file: Packages2Install.R

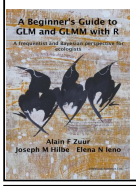
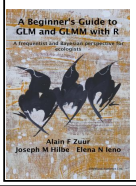
Bivariate linear regression

3	Loyn exercise 3	As above	Ex305A_LoynDataBivariateRegression_V4.r
4	Bailey exercise 4	As above	Ex306A_Bailey2008BivariateRegression_V3.r

Multiple linear regression

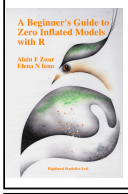
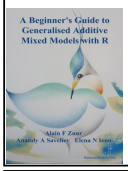
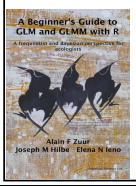
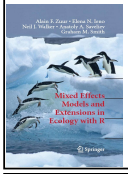
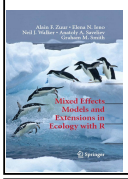
5	Methane	Methane.xls	Ex307A_Methane_V6.R
6	Loyn exercise 6	As above 	Ex308A_LoynMultipleRegression_V4.r
7		As above	Ex309A_Bailey2008MultipleRegression_V5.r
8		IrishPh.xls 	Ex310A_IrishpHMultipleRegression_V3.r
IT Example	Eagles	HumanEagleV3.xls	Eagles5.r Eagles.swf: Video solution file (use your browser to view it)
Bonus	Carbon data	MycorrhizaV1.txt	Ex311A_CarbonData_V2.R

Poisson and NB GLM
Demonstration code Chapter1A_Bailey_GLMGLMM_V3.R

9		Infauna.xls	Ex408A_BenthosAmpeliscidae_V5.R
10	Bailey data (offset)	As above 	Ex409A_BaileyGLM_V3.r
11	Red Squirrels	RedSquirrels.txt Link to paper Link to PhD thesis 	Class exercise:

Binomial GLM

		Crocodiles.txt Crocodiles.xls	

12	Crocodiles	Link to paper 	Ex402A_Crocodiles_V2.r
13	Soay sheep antibodies	SheepAnitbodiesV2.txt Link to paper	Ex410A_SoaySheep_V1.r
14	Drug mites	Drugmites.xls	Ex407A_DrugsMites_V3.R Video solution file (use your browser to view it)
GAM R code and data used during GAM presentation 1: Chapter3A_Bailey.R and BaileyDensity.txt R code and data used during GAM presentation 2: Chapter3B_Heger.R and HegerPierce.txt			
15	Squid isotopes	SquidNorway.txt 	Ex501A_Squid_V7.R
16	Ragweed pollen	Ragweed.xls	Ex502A_Ragweed_V5.R
17	Bailey	As above	Ex503A_Bailey2008AdditiveModelling.r Ex504A_Bailey2008GAM.r
18			
19			
20			
21	Swiss plants	NitrogenDeposition.txt Link to paper	Ex505A_NitrogenDeposition_V3.R
Gamma GLM/GAM			
22	Seedlings	Seedling.txt 	Ex506A_SeedlingsRainforest_V3.R
Extra			
VIF		VegetationX.txt 	VIFVegetationDataV2.r
		Roadkills.txt 	SolutionRoadKillV2.R
	TableGLM.pdf		

Useful pdfs

Some of the following papers are mentioned during the course.

- Links to papers on model selection:
  - Why do we still use stepwise modelling in ecology and behaviour?
  - Model selection in ecology and evolution
  - Performance of several variable-selection methods applied to real ecological data
  - Alternative regression methods are not considered in Murtaugh (2009) or by ecologists in general
- Link to a paper on post-hoc testing:
  - Time for some a priori thinking about post hoc testing
- Link to a paper on presenting results for factors:
  - factorplot: Improving Presentation of Simple Contrasts in Generalized Linear Models
- Link to a paper on homogeneity:
  - Neglected biological patterns in the residuals
- Link to a famous paper on pseudo-replication
  - PSEUDOREPLICATION AND THE DESIGN OF ECOLOGICAL FIELD EXPERIMENTS
- Link to the paper from Bailey et al. (2008).
  - Long-term changes in deep-water fish populations in the northeast Atlantic: a deeper reaching effect of fisheries?
- R reference card.
- Examples Information Theoretic approach:
  - Long-term population trends of endangered Hawaiian waterbirds
  - Climate as a driver of population variability in breeding Gentoo Penguins Pygoscelis papua at the Falkland Islands

Flowcharts

The following flowcharts will be shown during the course.

- Model building outline.
- What to write in a paper. Second flow chart.
- I have done linear regression, now what?
- Outline linear regression analysis

What if you are not able to participate?

Once participants are given acces to course exercises with R solution codes, pdf files of book chapters, pdf files of powerpoint files and video solution files, all course fees are non-refundable. However, depending for the circumstances for cancellation we may be able to offer you the option to attend a future course or you can authorise a colleague to attend this course.

Follow-up course

In 2012, we ran an online follow-up course for participants who have attended our data exploration, regression, GLM & GAM course in R. During a period of 12 weeks we discussed a series of 12 exercises (2 exercises per 2 weeks). Participants had access to an online meeting website where they could ask questions on the exercises and discuss the problems/solutions with other participants and the instructors. Solution files contain audio and video.

As a bonus we have made the material from this online course available to you. It is slightly outdated as we used an older HighstatLib.R file. There is also a Discussion Board, but it is unlikely that we will monitor it now and answer any questions (after all, it is a bonus).

This material is only open to people who have participated in our data exploration, regression, GLM & GAM course. No statistical theory is 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.

Sharing these video files with non-participants (including uploading these files to the Internet) without our explicit permission is copyright infringement. Copyright infringement will cost you (or your university) a lot of money.

The URL for the online material is: http://www.highstat.com/RGG1/Course.htm

The username to access this website is: Bonus2012RGG

And the password to access this website is: Highstat2012

- Do not copy any blanc space

If you have any questions please email Dr. Alain Zuur at: highstat@highstat.com

Feedback

You can use the form below to provide (anonymous) feedback.

Please do not use accented symbols (e.g. Ø, á, ö, ű, ń).

Email address (optional)

Name (optional)

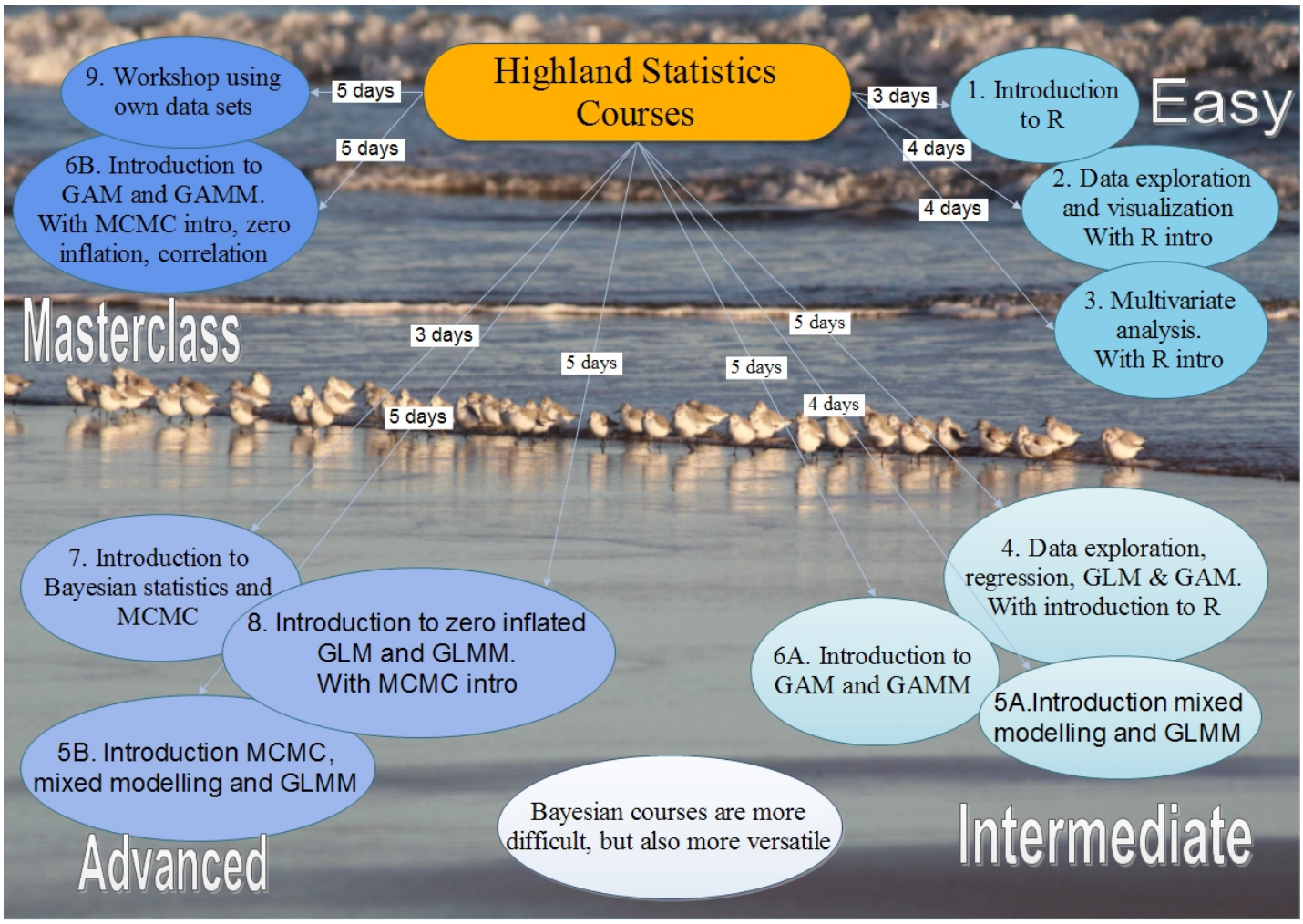
Did you get out of the course what you were looking for?

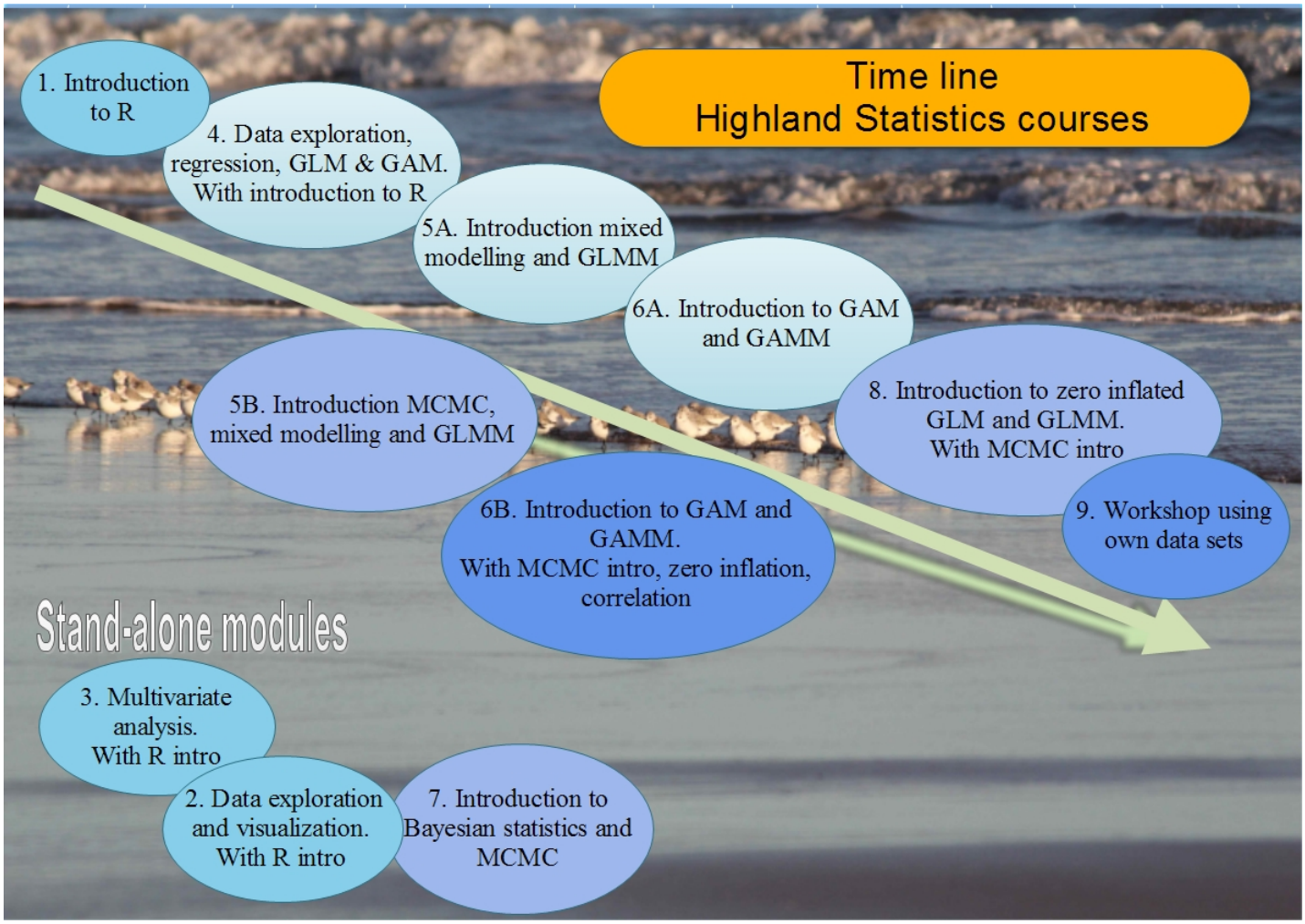
What did you like?

What would you change?

Email Feedback

Other courses and order of courses


Highland Statistics Courses
Easy
1. Introduction to R
2. Data exploration and visualization With R intro
3. Multivariate analysis. With R intro
4. Data exploration, regression, GLM & GAM. With introduction to R
5A.Introduction mixed modelling and GLMM
6A. Introduction to GAM and GAMM
Intermediate
Masterclass
6B. Introduction to GAM and GAMM. With MCMC intro, zero inflation, correlation
9. Workshop using own data sets
Advanced
5B. Introduction MCMC, mixed modelling and GLMM
8. Introduction to zero inflated GLM and GLMM. With MCMC intro
7. Introduction to Bayesian statistics and MCMC
Bayesian courses are more difficult, but also more versatile
3 days
4 days
5 days


Time line
Highland Statistics courses
1. Introduction to R
4. Data exploration, regression, GLM & GAM. With introduction to R
5A. Introduction mixed modelling and GLMM
6A. Introduction to GAM and GAMM
8. Introduction to zero inflated GLM and GLMM. With MCMC intro
9. Workshop using own data sets
6B. Introduction to GAM and GAMM. With MCMC intro, zero inflation, correlation
5B. Introduction MCMC, mixed modelling and GLMM
Stand-alone modules
3. Multivariate analysis. With R intro
2. Data exploration and visualization. With R intro
7. Introduction to Bayesian statistics and MCMC