Modelação Ecológica

AULA 16

12th November 2019

ABOUT AULA 15

- 6 November 2019

- Own work on HMMs project + extra material on gamma and beta GLMs

Your mission for last week, extended to this week (24th November 2019)!

1. Go to Movebank and explore the Movebank Data Repository
2. Find an open access data
3. Use it to fit an HMM to the movement data
4. Write a report about the analysis you made
   * Briefly describe the data
   * Do a simple EDA
   * Fit movement model
   * Present your conclusions
5. Send me a report by the end of the week (week ends Sunday)

Mission: Impossible
Taking down Movebank
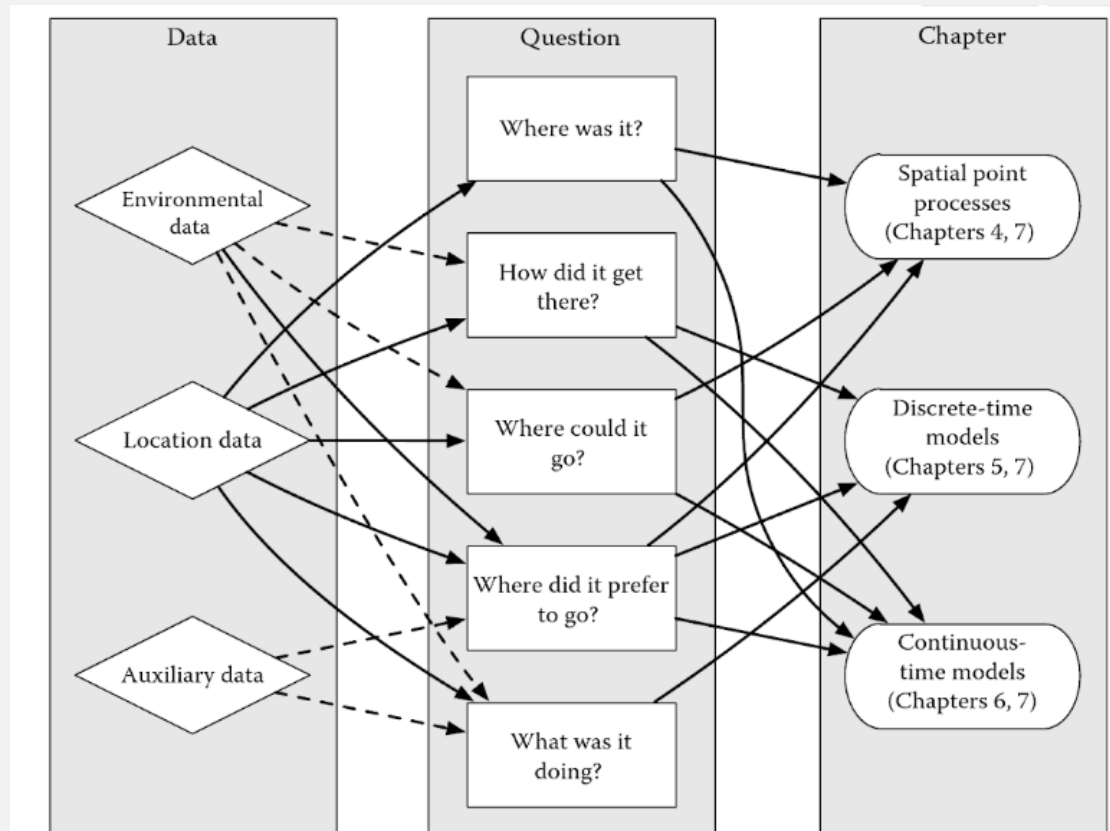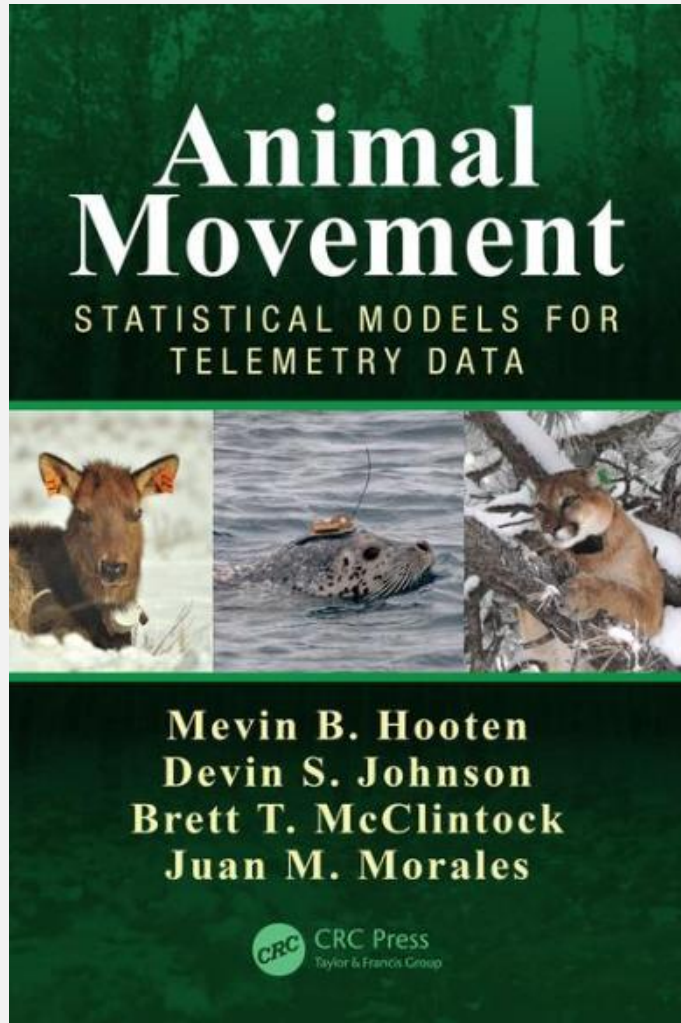(with a random Modelação Ecológica student)

ABOUT MODELLING MOVEMENT

Some final words to wrap up!

Animal Movement Modelling

FIGURE 1.1 Relationships among data types, analytical methods, and some fundamental questions of movement ecology. Location data are the cornerstone of all of the analysis methods described in this book. Environmental data, such as those acquired from remote sensing, are useful in drawing connections between animals and their surroundings. Auxiliary biotelemetry data, such as accelerometer or dive profile data, can help address questions about animal behavior. Dashed lines indicate where data can be helpful for addressing particular questions, but are not essential.

CrossMark

ORIGINAL PAPER

Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges

Toby A. Patterson[1] · Alison Parton[2] · Roland Langrock[3] · Paul G. Blackwell[2] · Len Thomas[4] · Ruth King[5]

Animal Movement Modelling

Guest Editor's Introduction to the Special Issue on "Animal Movement Modeling"

Mevin B. HOOTEN, Ruth KING, and Roland LANGROCK

In this introduction, we provide a brief overview to statistical models for animal trajectories and then summarize the set of invited articles that comprise the issue.

INTRODUCTION

The study of individual animal trajectories is ancient, with a written record dating back more than 2000 years (Nussbaum 1978). The mathematics of trajectories also have a long history (Turchin 1988), but rigorous statistical methods development for animal trajectory data (i.e., telemetry data) only dates back approximately 50 years (e.g., Dunn and Gipson 1977). In more recent times, the technology to collect a massive amount and variety of telemetry data has increased dramatically, leading to unprecedented information on the spatial ecology of animal populations (Cagnacci et al. 2010; Kays et al. 2015). The sheer availability, variety, and volume of telemetry data sets present golden opportunities for the development of new statistical methods that can provide deeper, more meaningful inference

Conceptualizing Movement

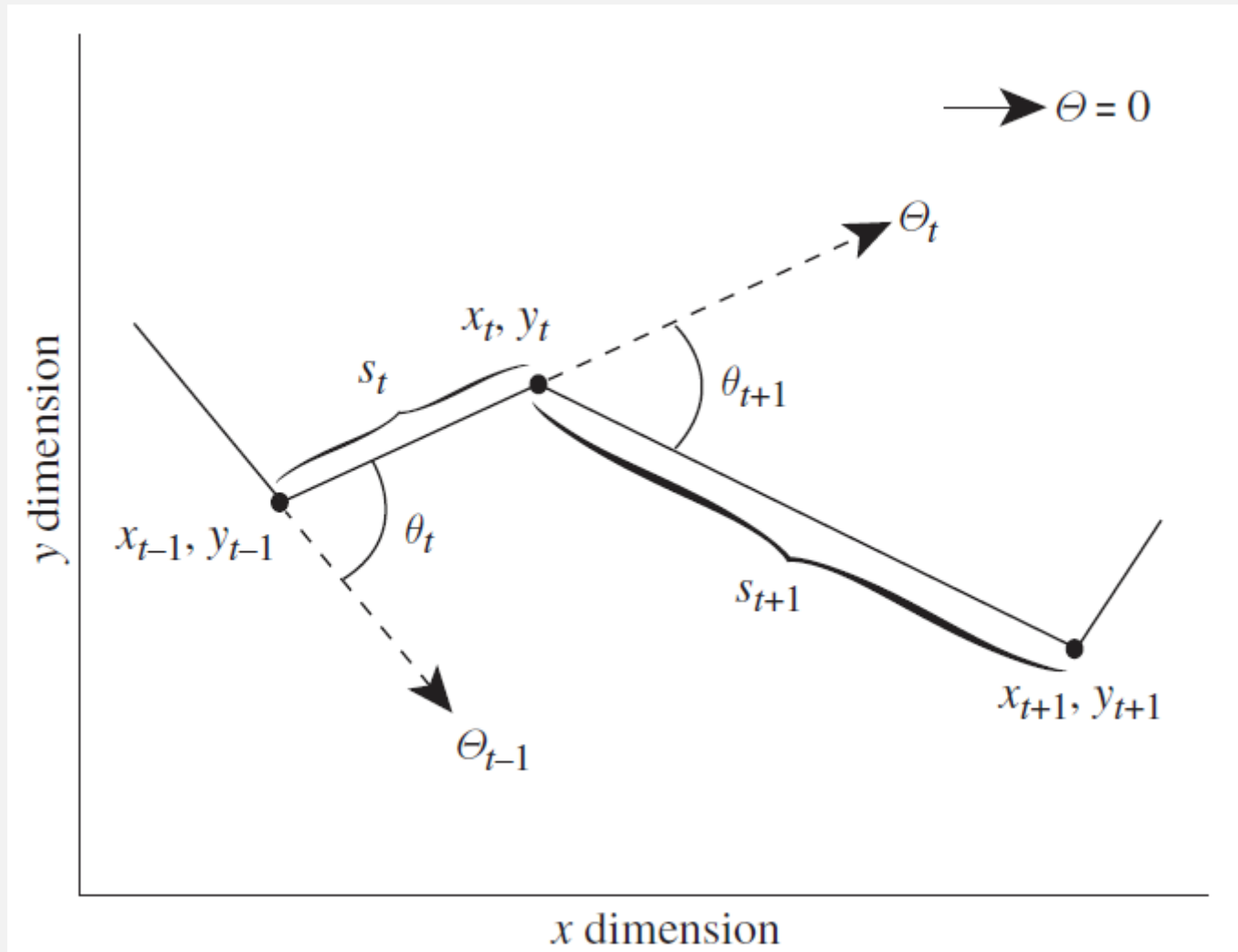Movement is a continuous process in time and space

Data we observe is typically a series of positions collected at regular (or irregular) time points

But we can model it in:

1. Discrete time and discrete space – too rough
2. Discrete time and continuous space – more often
3. Continous time and discrete space – less used
4. Continous time and discrete space – harder to implement

A path is just a starting point plus a succession of turning angles and step lengths!

Polansky, L. & Wittemyer, G. 2010 A framework for understanding the architecture of collective movements using pairwise analyses of animal movement data *Journal of The Royal Society Interface* **8**: 322-333

HIDDEN MARKOV MODELS

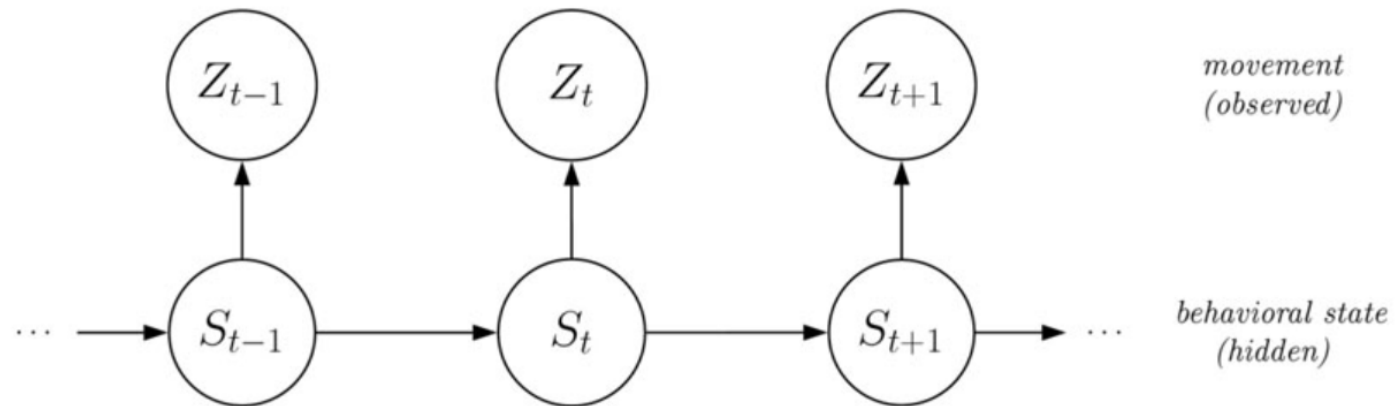APPENDIX A. Graphical model of an HMM for animal movement.

FIG. A1. Dependence structure of an HMM.

1. Series of hidden (unobserved states) states, with Markovian dependence
2. Conditional on which there is a process we observe
3. Conditional on state, process is independent
4. We make inferences on hidden states (e.g. travelling vs. feeding) based on observations (positions)
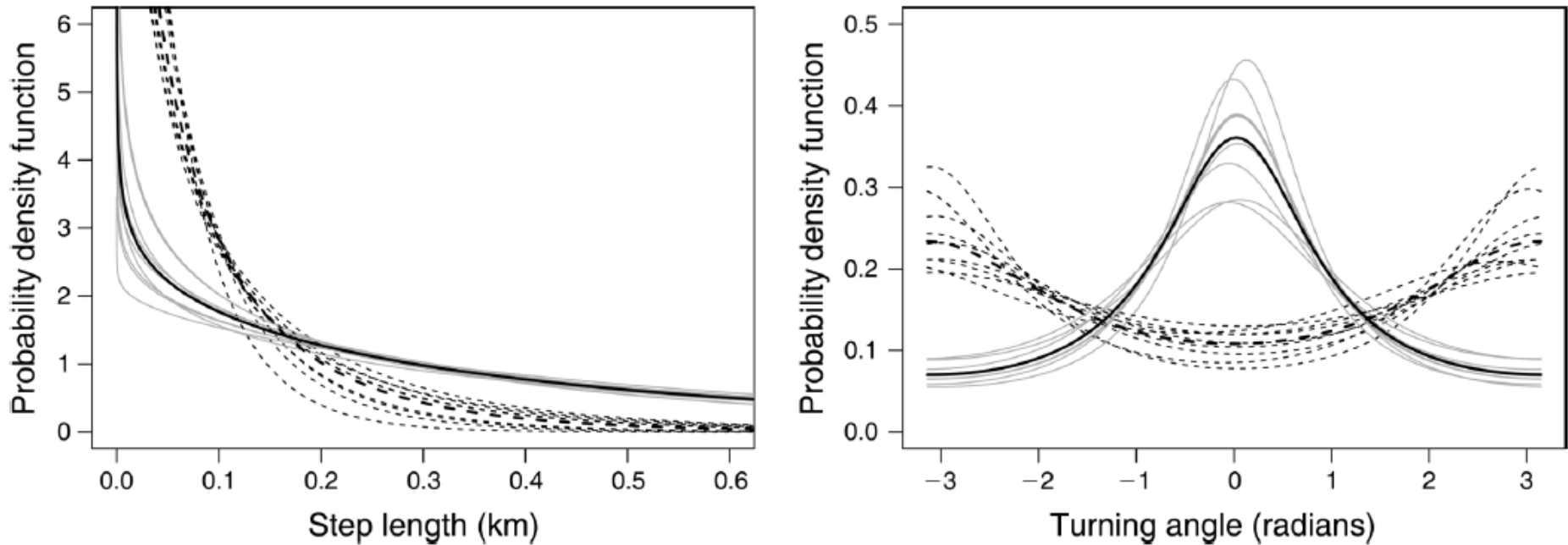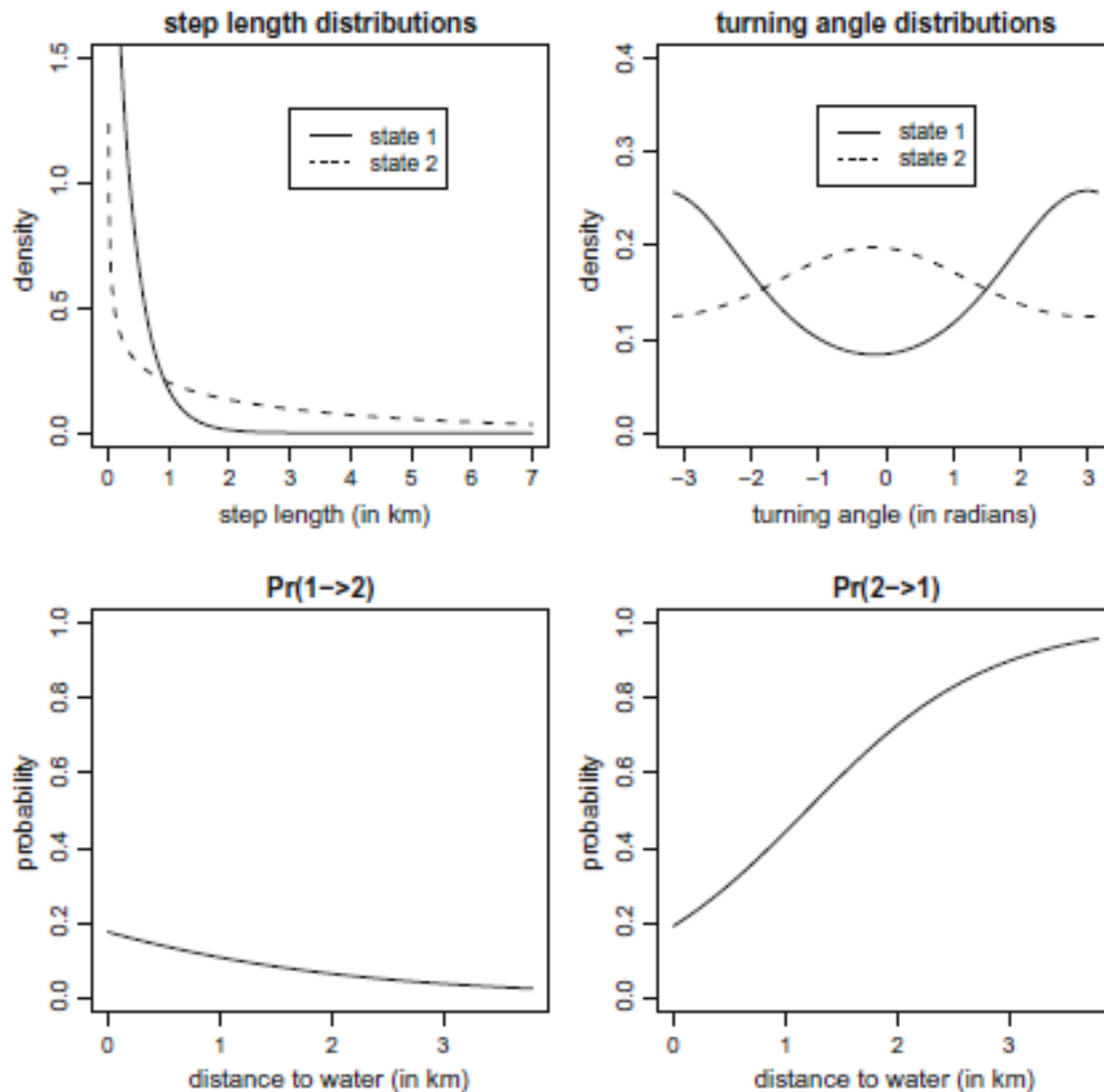
HIDDEN MARKOV MODELS

Fig. 1. Models 2 and 4 for the movement paths of nine American bison: fitted state-dependent Weibull distributions for step lengths (left) and wrapped Cauchy distributions for turning angles (right). Solid lines show "exploratory" movement; dashed lines show "encamped" movement. Gray lines are results for the individual-specific HSMMs, and black lines are results for HSMM with parameters common to all individuals.

Langrock, R.; King, R.; Matthiopoulos, J.; Thomas, L.; Fortin, D. & Morales, J. M. 2012 Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions *Ecology* **93**: 2336-2342

Fig. 2 HMM applied to elk movement data: fitted state-dep. distributions (*top row*) and estimated effect of covariate "distance to water" on state-switching dynamics (*bottom plot*)

Patterson, T. A.; Parton, A.; Langrock, R.; Blackwell, P. G.; Thomas, L. & King, R. 2017 Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges *AStA Adv Stat Anal* **101**:399–438

Application | 🔓 Free Access

moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models

⊘ Correction(s) for this article ⌄

Théo Michelot ✉, Roland Langrock, Toby A. Patterson

momentuHMM: R package for generalized hidden Markov models of animal movement

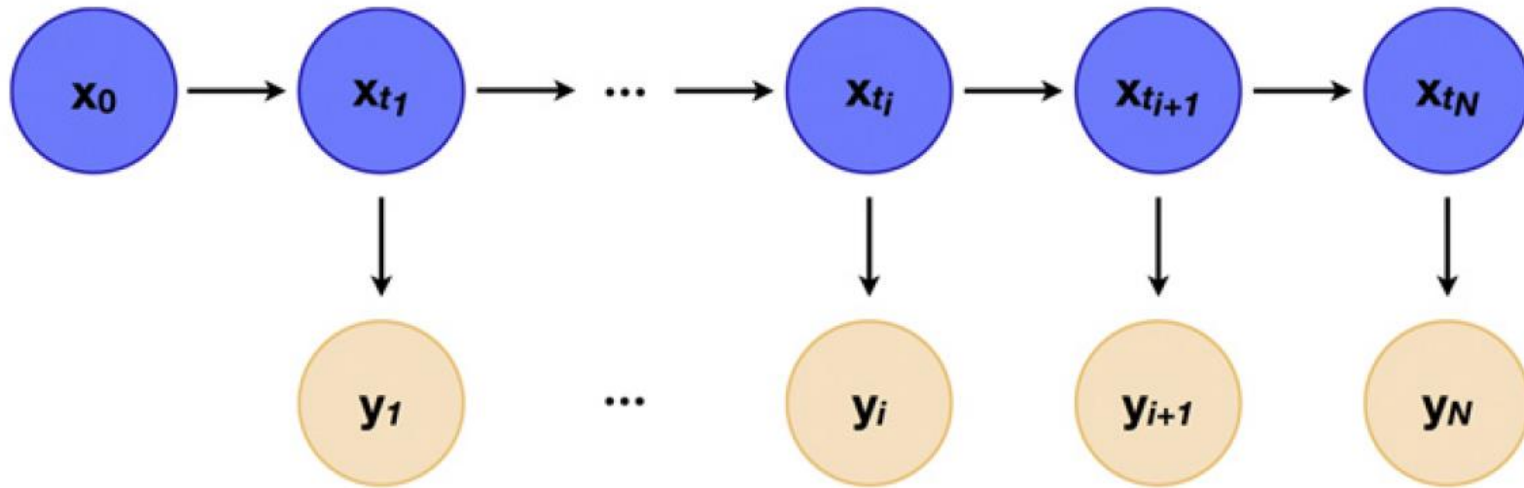Brett T. McClintock ✉, Théo Michelot
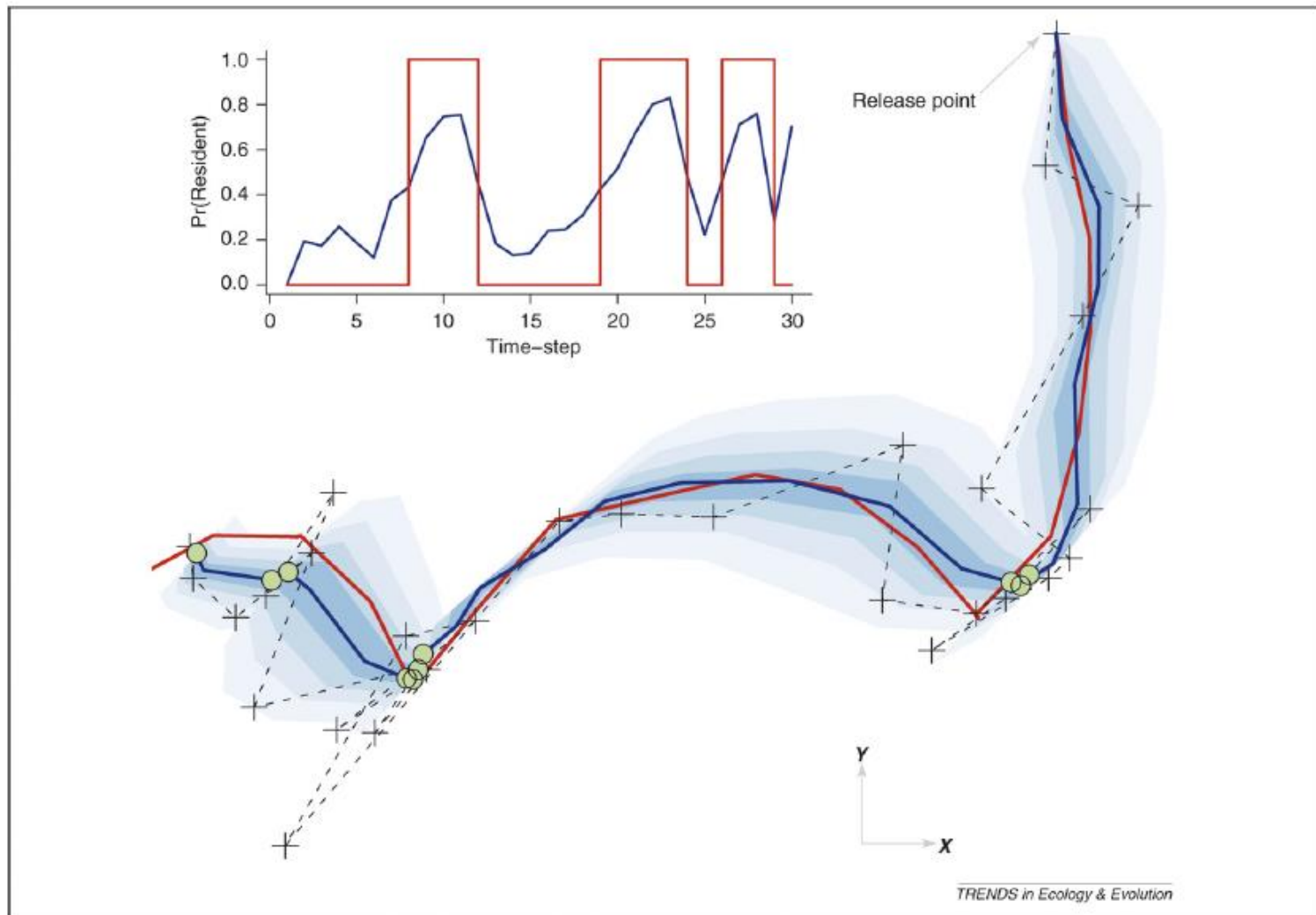
SECTIONS · PDF · TOOLS · SHARE

STATE SPACE MODELS

Fig. 1. Structure of a SSM for estimation. The state variable evolves randomly in time, starting at $\mathbf{x}_0$ and ending at $\mathbf{x}_{t_N}$. At times $t_1, t_2, \ldots, t_N$, measurements $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are taken. In this graphical representation of the model, an arrow from one variable (say, $\mathbf{x}_{t_i}$) to another (say, $\mathbf{y}_i$) indicates that the model is written in terms of the conditional distribution of $\mathbf{y}_i$ given $\mathbf{x}_{t_i}$, and furthermore that conditional on $\mathbf{x}_{t_i}$, $\mathbf{y}_i$ is statistically independent of all other variables since they are not connected with arrows.

Jonsen, I.; Basson, M.; Bestley, S.; Bravington, M.; Patterson, T.; Pedersen, M.; Thomson, R.; Thygesen, U. & Wotherspoon, S. 2012 State-space models for bio-loggers: A methodological road map *Deep Sea Research Part II: Topical Studies in Oceanography*, **88-89**: 34-46

Figure 1. Simultaneous inference of location and behaviour from noisy telemetry data. A hypothetical movement path (red) was simulated from a correlated random walk (Box 1) that switches between two behavioural states – 'transient' and 'resident'. Noisy observations (black crosses) typical of satellite telemetry [22,45] were added. Bayesian particle-filtering techniques (Box 2) were used to estimate movement parameters, the true path and behavioural modes, given moderately informative priors on the parameters. Posterior mean estimates of the true path (dark blue) are shown to vastly improve estimates of the true path. Importantly, uncertainty in the path is quantified (dark to light blue shading indicates 20, 40, 60 and 80% credibility intervals). Locations estimated as being from the resident phase (where the probability of being in resident mode was most likely, Pr[resident] >0.5) are shown as green dots. (Inset) Inferences of behavioural movement mode from the path. The blue line gives the estimated probability of the animal being in resident mode (the probability of transit is simply 1-Pr[resident]); the square red line is the true mode.

Patterson, T. A.; Thomas, L.; Wilcox, C.; Ovaskainen, O. & Matthiopoulos, J. 2008 State-space models of individual animal movement *Trends in Ecology and Evolution* **23**:87-94

There's lots of great movement data out there for grabs!

https://www.movebank.org/

EXTRA STUFF
- on GLMs

- https://rpubs.com/kaz_yos/glm-Gamma (Gamma Regression)

- Gamma regression (improved) : see pdf gammareg.pdf

- Beta regression (see pdf betareg.pdf) and paper

REVIEW

Methods in Ecology and Evolution

Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression

Jacob C. Douma[1,2] | James T. Weedon[3,4]

Gestão de Páginas

- Modelação Ecológica
  - Modelação Ecológica(Ecologia Marinha)
  - Modelação Ecológica(Ecologia e Gestão Ambiental)
  - Aulas
    - Aula1
    - Aula2
    - Aula3
    - Aula4
    - Aula5
    - Aula6
    - Aula7 08 10 2019
    - Aula8 09 10 2019
    - Aula9 15 10 2019
    - Aula10 16 10 2019
    - Aula11 22 10 2019
    - Aula12 23 10 2019
    - Aula13 29 10 2019
    - A14 05 11 2019 HMMs by Théo Michelot
    - A15 06 11 2019

A15 06 11 2019

Página | Ficheiros 3 | Permissões | Link

Adicionar Ficheiro

#	Nome
1	betareg.pdf
2	gammareg.pdf
3	Douma_et_al-2019-Methods_in_Ecology_and_Evolution.pdf

EXTRA STUFF
- on Negative Binomial GLM

http://sherrytowers.com/2018/04/10/negative-binomial-likelihood-fits-for-overdispersed-count-data/

https://stats.idre.ucla.edu/r/dae/negative-binomial-regression/

WORDS OF CAUTION

Model selection techniques and dignostic plots are very important – never understimate the importance of understanding your data
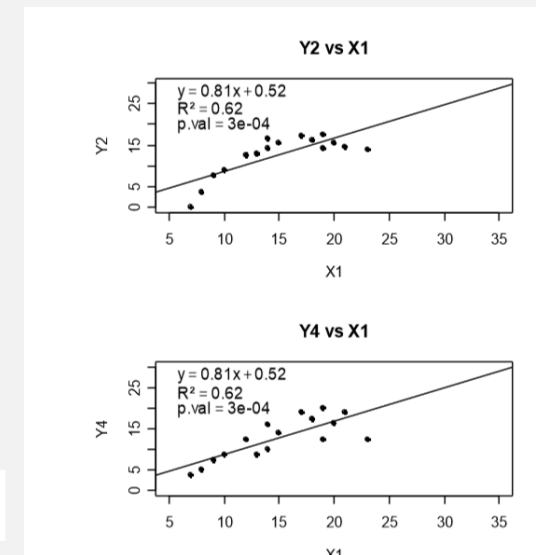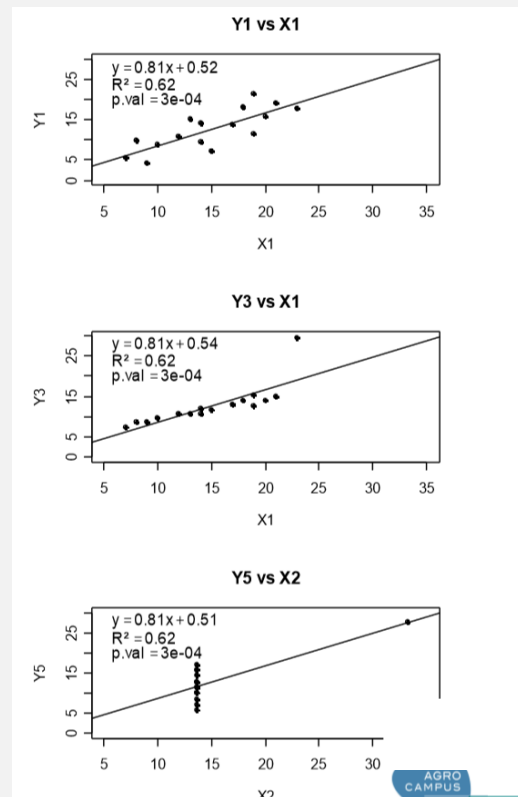
Tomassone et al. (1992)

One should not only evaluate the sensitivity of fitted models to changes in data, one should also try to validate the models considered for further inference.

The validation of a model might be done by a variety of techniques:
- diagnostic plots
- generating data from the fitted models
- re-using data used to fit the model (cross-validation)
- contrasting predictions against observations not used in
    the model fitting (e.g. leave one out cross validation, etc.)

LIMITATIONS OF GLMs

Modelo de crescimento de Gompertz

Modelo stock-recrutamento

$$W_t^{pred} = W_\infty \cdot (1 - e^{-K \cdot (t-t_0)})^3 \qquad \theta = (L_\infty, K, t_0)$$

Beverton & Holt $\qquad R_t^{pred} = \dfrac{S_t}{\alpha + \beta \cdot S_t} \qquad \theta = (\alpha, \beta)$

Ricker $\qquad R_t^{pred} = S_t \cdot e^{\alpha - \beta \cdot S_t} \qquad \theta = (\alpha, \beta)$
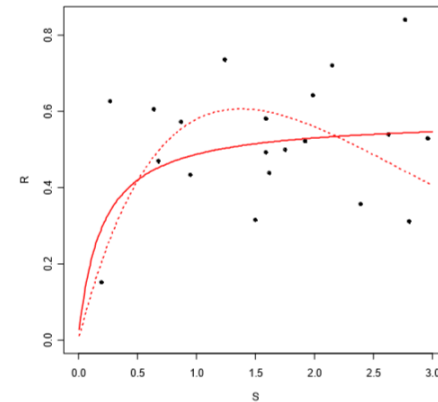
Beverton & Holt

Ricker

Séries temporais climáticas

Beaugrand & Reid (2003)
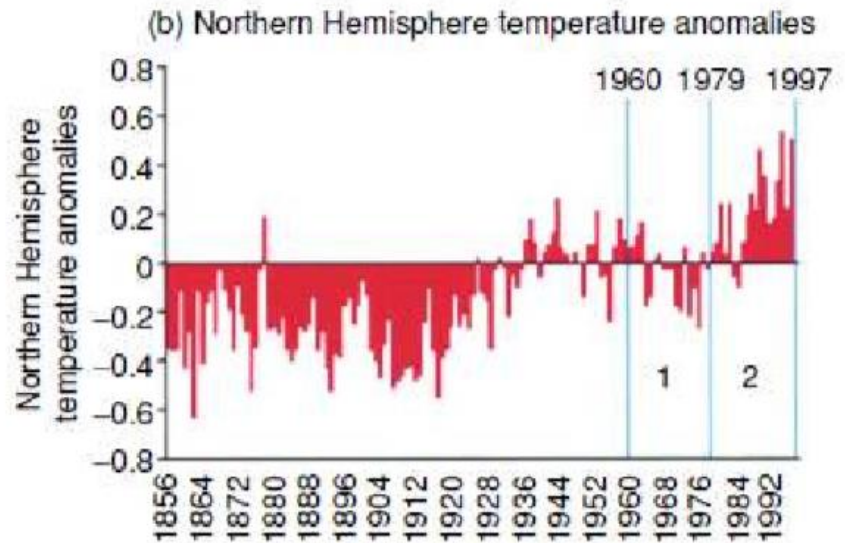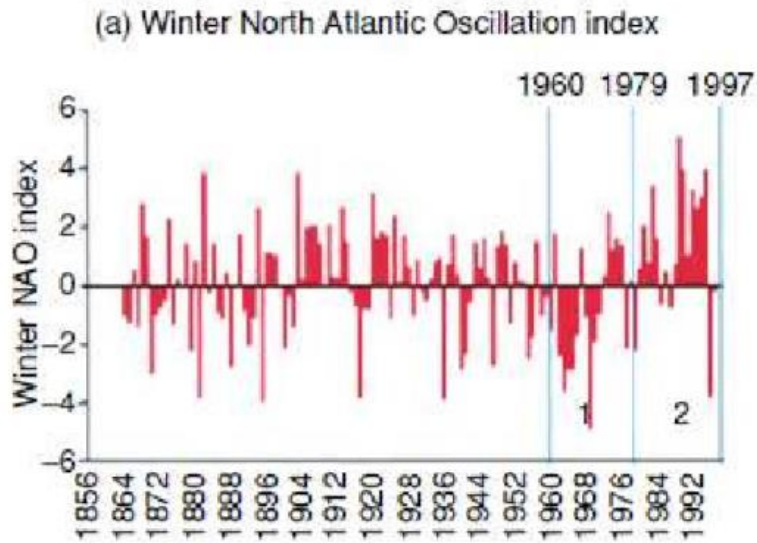
Modelos espaciais

Pavine (2006)

Non-linear relations, not convertible into linear relations → **GAMs**

Non-negligible lack of Independence, due to e.g.:
- clusters of observations (e.g. repeated measures in individuals),
- Measures over space
- Measures over time

GLMMs

GAMMs

Other "time series" models (e.g. autoregressive models, like simple **AR** autoregressive, **ARMA** autoregressive moving average ou **ARIMA** autoregressive integrated moving, **CAR** conditional auto regressive)

GAMs
*
Generalized Additive Models

Generalized Additive Models

When the relation between the response variable and the covariates is non-linear, we might need more than a GLM. This leads us to

As in GLMs

CHAPTER 3

Introducing GAMs

Função de ligação (*link function*)

A non-smooth component

3.1 Introduction

A generalized additive model (Hastie and Tibshirani, 1986, 1990) is a generalized linear model with a linear predictor involving a sum of smooth functions of covariates. In general the model has a structure something like

$$g(\mu_i) = \mathbf{X}_i^* \boldsymbol{\theta} + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \dots \qquad (3.1)$$

where

$\mu_i \equiv \mathbb{E}(Y_i)$ and $Y_i \sim$ some exponential family distribution.

A smooth component

Wood, S. N. 2006 Generalized Additive Models: an introduction with R. CRC/Chapman & Hall

Without entering the computational details, which are extremely complex, what a GAM does is to find a smooth function that fits the data, while penalizing for model complexity.

The most common smooth functions used in practice for GAMSs are splines (used e.g. in the `mgcv` package, function `gam`, which we will use to implement GAMs). These correspond to a set of simple **basis functions** (see below) which are multiplied by constants (parameters) and summed to approximate a **smooth function**.

The objective is to find the constants (i.e., the parameters) which result in the best approximation to the data, while penalizing for the number of parameters (model complexity)!

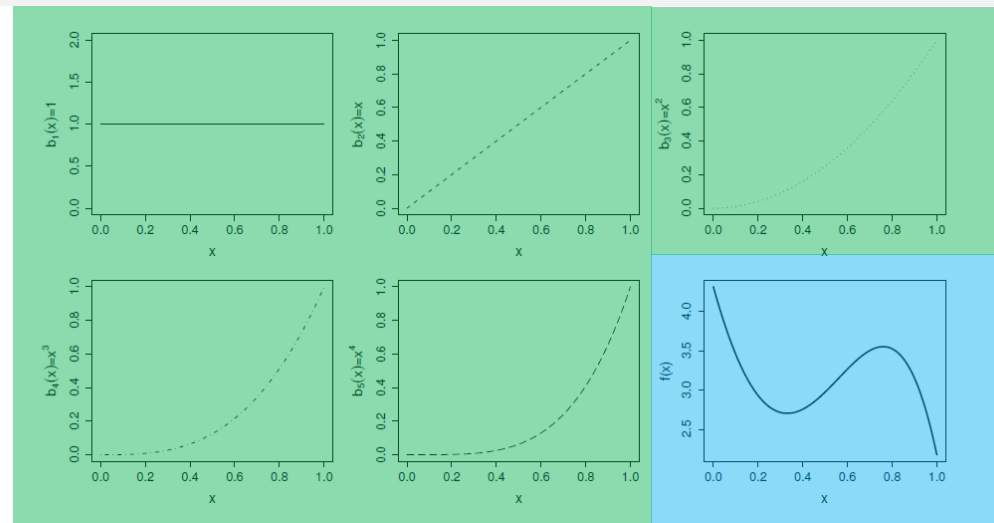This is done within the software, so you don't have to worry about it!

Figure 3.1 *Illustration of the idea of representing a function in terms of basis functions, using a polynomial basis. The first 5 panels (starting from top left), illustrate the 5 basis functions, $b_j(x)$, for a 4th order polynomial basis. The basis functions are each multiplied by a real valued parameter, β_j, and are then summed to give the final curve $f(x)$, an example of which is shown in the bottom right panel. By varying the β_j, we can vary the form of $f(x)$, to produce any polynomial function of order 4 or lower. See also figure 3.2*

Wood, S. N. 2006 Generalized Additive Models: an introduction with R. CRC/Chapman & Hall

LMs, GLMs and GAMs

Linear Model

$$E[y|\mathbf{x}] = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

Additive Model (1985)

$$E[y|\mathbf{x}] = \beta_0 + f_1(x_1) + \ldots + f_p(x_p)$$

Generalized Linear Model (1972)

$$g\left(E[y|\mathbf{x}]\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

Generalized Additive Model (1990)

$$g\left(E[y|\mathbf{x}]\right) = \beta_0 + f_1(x_1) + \ldots + f_p(x_p)$$

parametric   nonparametric
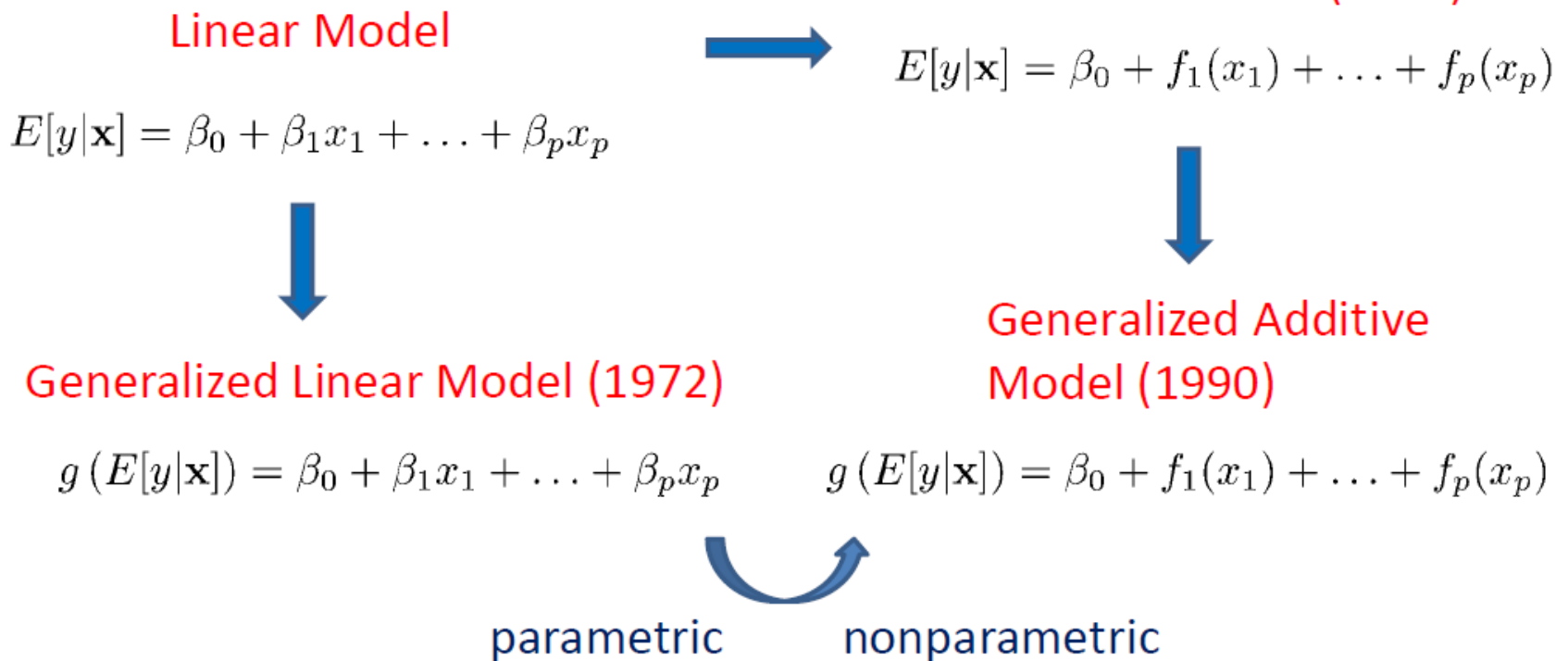
Reminder: on the first tutorial about R there was a (very simple) comparison between a LM, a GLM and a GAM!

LMs, GLMs and GAMs

The key differences between the simple regression models (LM) and the generalized linear models (GLM) or generalized additive models (GAM) are that the last 2:

1. allow to consider non-Gaussian responses;

2. The relation between the response variable and the independent variables does not need to be linear (in a GLM it is still linear, but only on the scale of the link function)

Therefore, in general GLMs are useful to a much wider variety of data sets, in particular in Ecology (e.g. counts, proportions, presence/absence, positive continuous values, etc.) and in the case of GAMs, the relations can even be highly non-linear.

```
set.seed(123)
xs=runif(1000,0,10)
ys=3.5+cos(xs)+rnorm(1000,sd=0.8)
par(mfrow=c(1,1))
plot(xs,ys,cex=0.5)
```
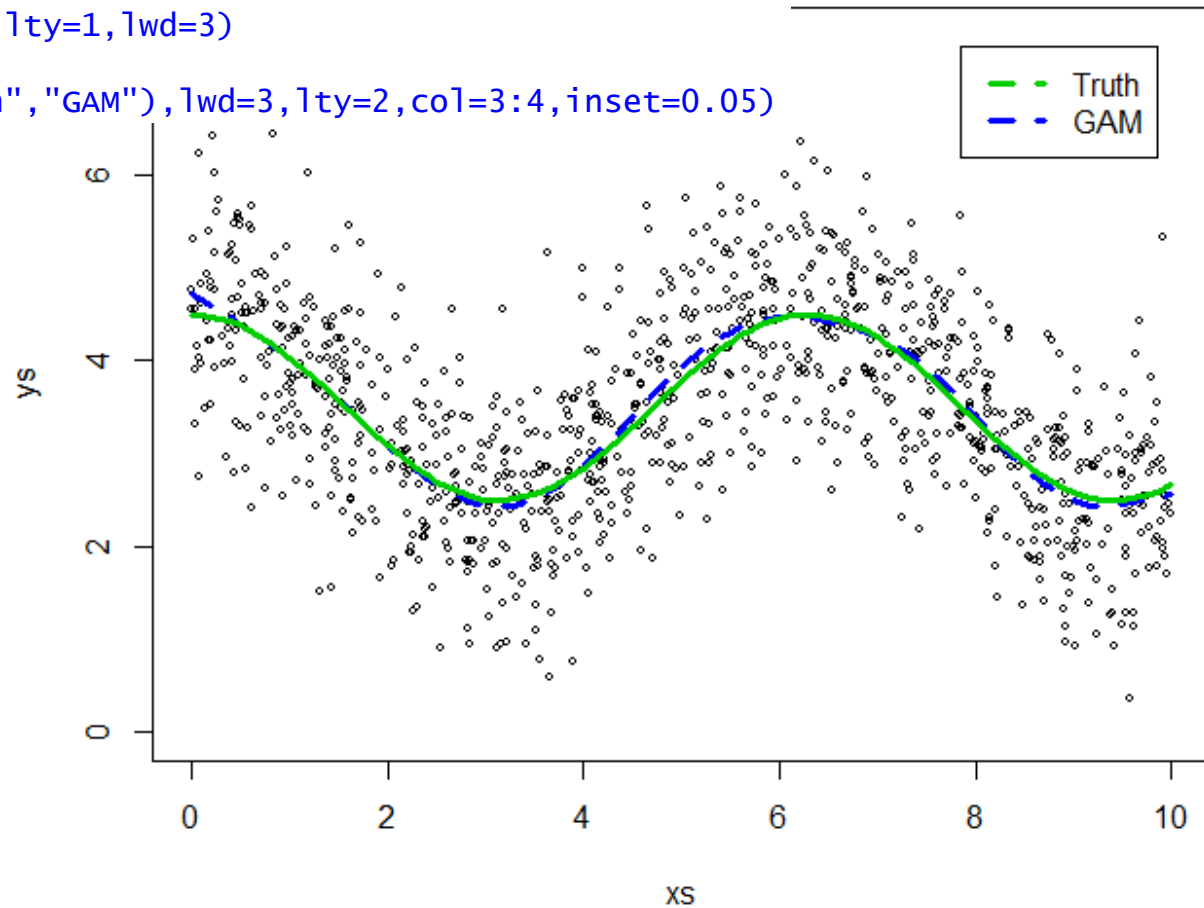
Clearly non-linear relation, no LM or GLM will fit the data
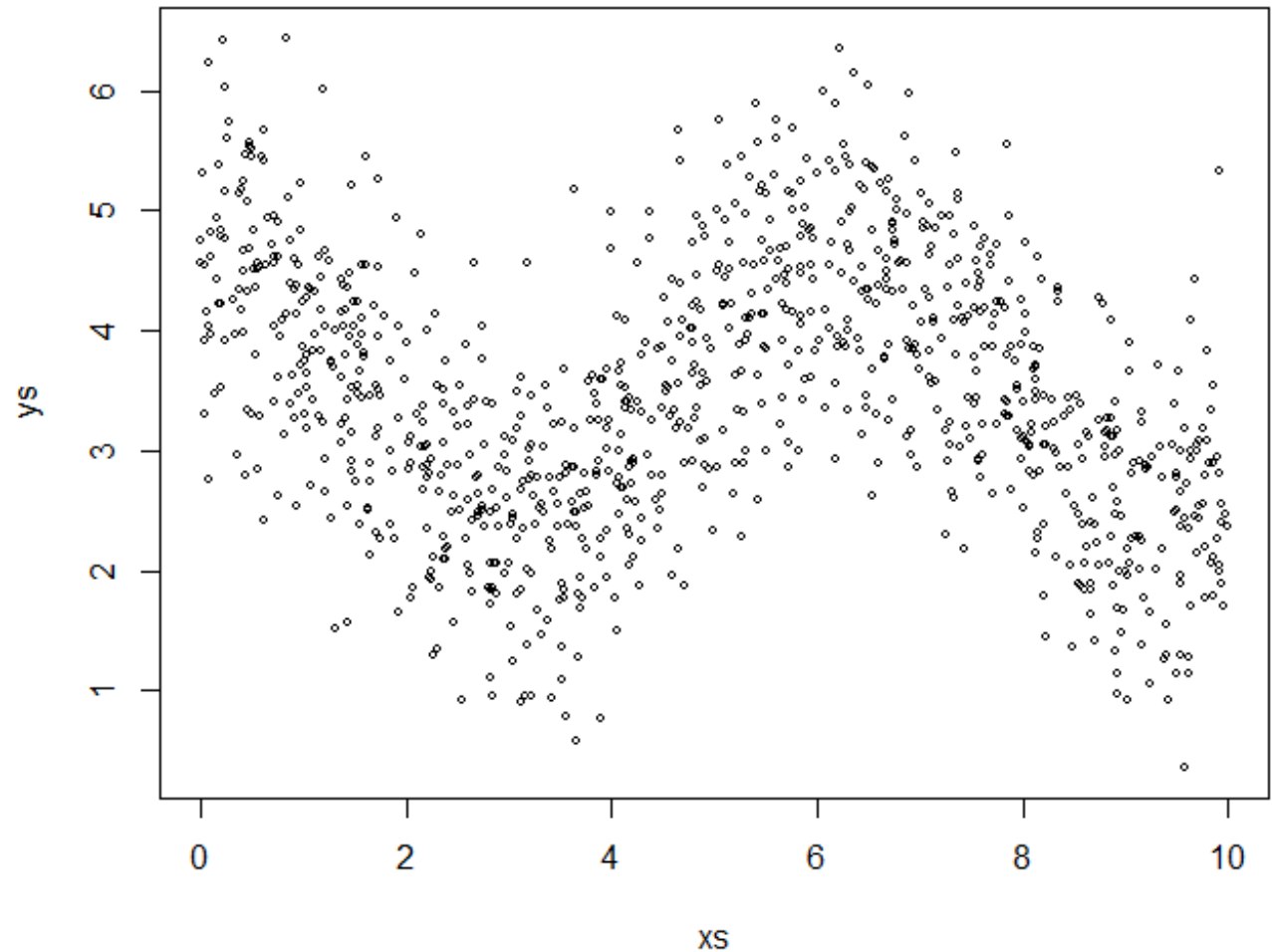
```
library(mgcv)
#fit the model
gam1=gam(ys~s(xs),family=gaussian)
#get a reasonable range of xs for predictions
xpred=seq(0,10,by=0.1)
#predict, on the response scale
preds=predict(gam1,newdata=data.frame(xs=xpred),type="response")
par(mfrow=c(1,1))
#plot the data
plot(xs,ys,cex=0.5,ylim=c(0,7.5))
#add the predictions
lines(xpred,preds,col=4,lty=2,lwd=3)
#add the tue model
lines(xpred,3.5+cos(xpred),col=3,lty=1,lwd=3)
#add a legend
legend("topright",legend=c("Truth","GAM"),lwd=3,lty=2,col=3:4,inset=0.05)
```

Familia de distribuições usada e respectiva função de ligação

Modelo usado

Estimadores da parte não suave (non smooth) (intercept=3.5)

Estimadores da parte suave

Tamanho da amostra

GOF: R^2, Percentagem da desviância explicada e Generalized Cross Validation

Estimativa do desvio padrão dos resíduos ($\sigma=0.7$)

```
> summary(gam1)

Family: gaussian
Link function: identity

Formula:
ys ~ s(xs)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.45364    0.02533   136.3   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
        edf Ref.df      F p-value
s(xs) 8.066  8.766 103.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.475   Deviance explained =   48%
GCV = 0.64755  Scale est. = 0.64168   n = 1000
```

Review article

Generalized additive models: Building evidence of air pollution, climate change and human health

Check for updates

Khaiwal Ravindra[a,*], Preety Rattan[a], Suman Mor[b], Ashutosh Nath Aggarwal[c]

Gestão de Páginas

Modelação Ecológica
- Modelação Ecológica(Ecologia Marinha)
- Modelação Ecológica(Ecologia e Gestão Ambiental)
- **Aulas**
  - Aula1 - Intorduction to Ecological Modelling, and Models
  - Aula2 - Relevant statistical concepts and Hands on Tutorial
  - Aula3 - Hands on tutorial, continued
  - Aula4 - Hands on tutorial, continued
  - Aula5 - Useful statitstical concepts
  - Aula6 - Statistical Modelling
  - Aula7 08 10 2019 - Introduction to Regression
  - Aula8 09 10 2019 - Regression, ANOVA and ANCOVA
  - Aula9 15 10 2019 - assessment, maximum likelihood and intro to GLMs
  - Aula10 16 10 2019 - more on GLMs
  - Aula11 22 10 2019 - Hands on GLMs and interactions
  - Aula12 23 10 2019 - Logistic regression, correlated variables and hand
  - Aula13 29 10 2019 - GLM extensions for modelling counts
  - A14 05 11 2019 HMMs by Théo Michelot
  - A15 06 11 2019 - Own work on HMMs + gamma + beta regression
  - A16 12 11 2019
  - A17 13 11 2019
  - A18 19 11 2019
  - A19 20 11 2019
  - A20 26 11 2019
  - A21 27 11 2019
  - A22 03 12 2019 - hand in and present work about R package
  - A23 04 12 2019
  - A24 10 12 2019
  - A25 11 12 2019
  - A26 17 12 2019
  - A27 18 12 2019
- **Outros Recursos**
  - **PDFs**

PDFs

Página | Ficheiros 13 | Permissões | Link

Adicionar Ficheiro

#	Nome
1	Generalized additive models: Building evidence of air pollution, climate change and human health *Ravindraetal2019.pdf*
2	Significance: What does it all mean? *Sheldon-2019-Significance.pdf*
3	Publication bias: What are the challenges and can they be overcome? *jpn-37-149.pdf*
4	The R Book.pdf
5	Mixed Effects Models And Extensions In Ecology With R *Zuur_Mixed-effects-models-and-extensions-in-ecology-with-R.pdf*
6	Analyzing Ecological Data *zuur_2007.pdf*
7	A Beginner's Guide to R *Zuuretal2009useR.pdf*
8	Numerical Ecology with R *Borcardetal2001EcologyUseR.pdf*
9	Introduction to Probability and Statistics Using R *IPSUR.pdf*
10	Ecological Models and Data in R *Bolker2007.pdf*
11	Norberg_et_al-2019 A comprehensive evaluation of predictive performance of 33 species distribution mode *Norberg_et_al-2019-Ecological_Monographs.pdf*

Generalised additive models (GAMs): an introduction

Many data in the environmental sciences do not fit simple linear models and are best described by "wiggly models", also known as Generalised Additive Models (GAMs).
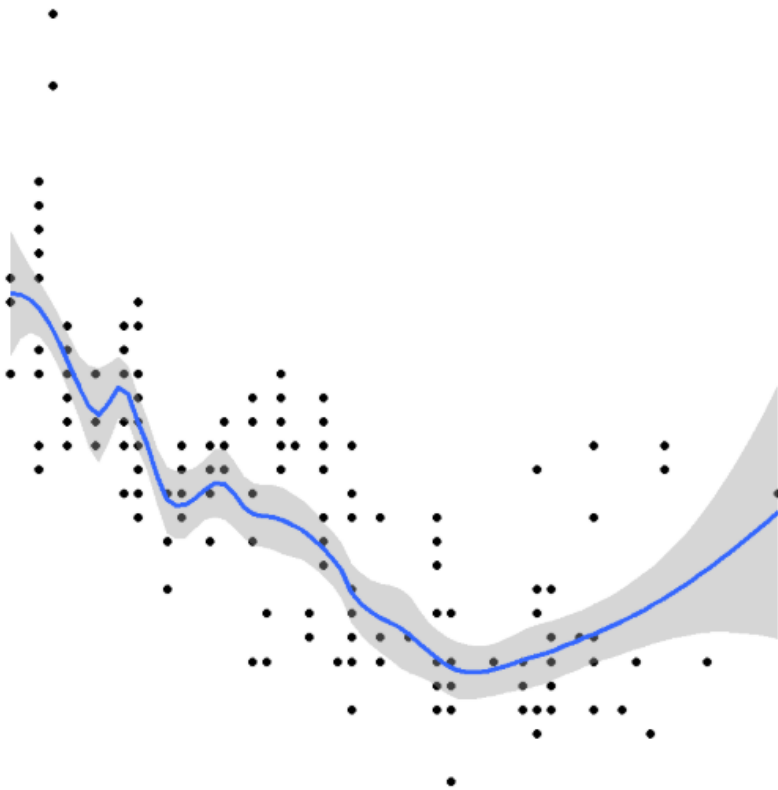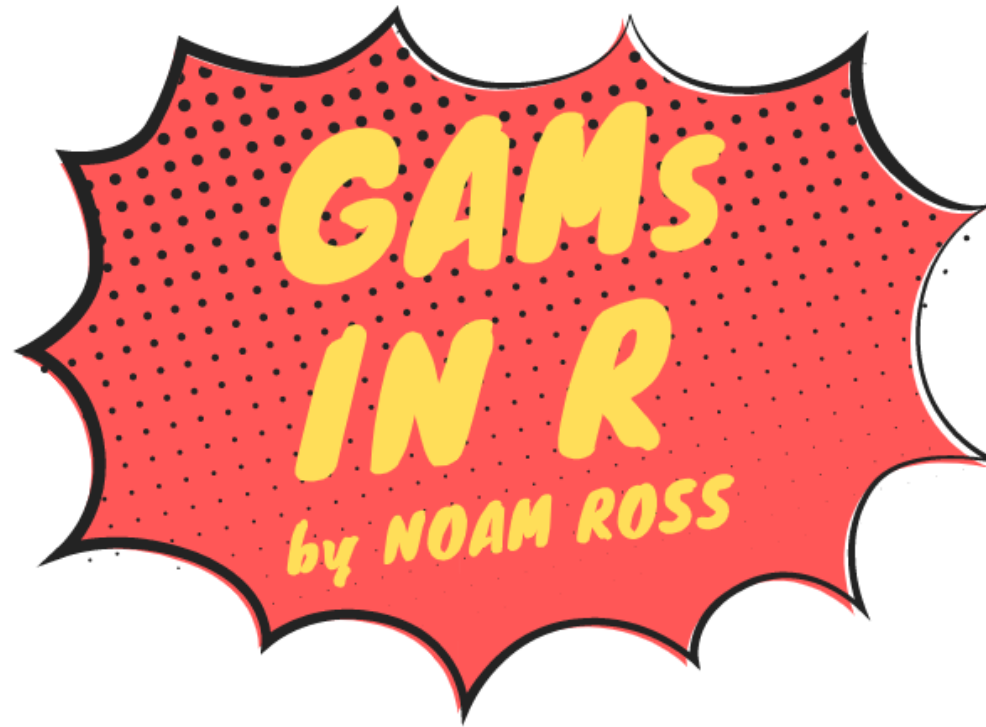
https://noamross.github.io/gams-in-r-course/

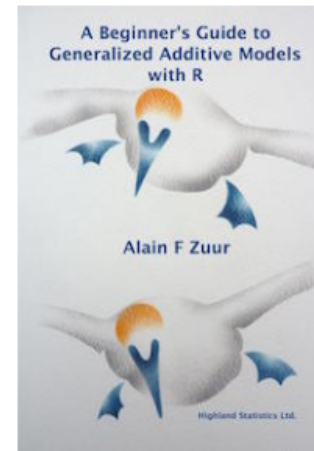# Beginner's Guide to Generalized Additive Models with R (2012)
Zuur, AF

A Beginner's Guide to Generalized Additive Models with R is, as the title implies, a practical handbook for the non-statistician. The author's philosophy is that the shortest path to comprehension of a statistical technique without delving into extensive mathematical detail is through programming its basic principles in, for example, R.

Not a series of cookbook exercises, the author uses data from biological studies to go beyond theory and immerse the reader in real-world analysis with its inherent untidiness and challenges.

The book begins with a review of multiple linear regression using research on human crania size and ambient light levels and continues with an introduction to additive models based on deep sea fishery data. Research on pelagic bioluminescent organisms demonstrates simple linear regression techniques to program a smoother.

In Chapter 4 the deep sea fishery study is revisited for a discussion of generalized additive models.

The remaining chapters present detailed case studies illustrating the application of Gaussian, Poisson, negative

https://highstat.com/index.php/beginner-s-guide-to-generalized-additive-models-with-r

Let's play with real data!

https://highstat.com/Books/BGS/GAM/Data/Parasites3.txt

https://highstat.com/Books/BGS/GAM/Data/BaileyDensity.txt

https://highstat.com/Books/BGS/GAM/Data/Gannets2.txt