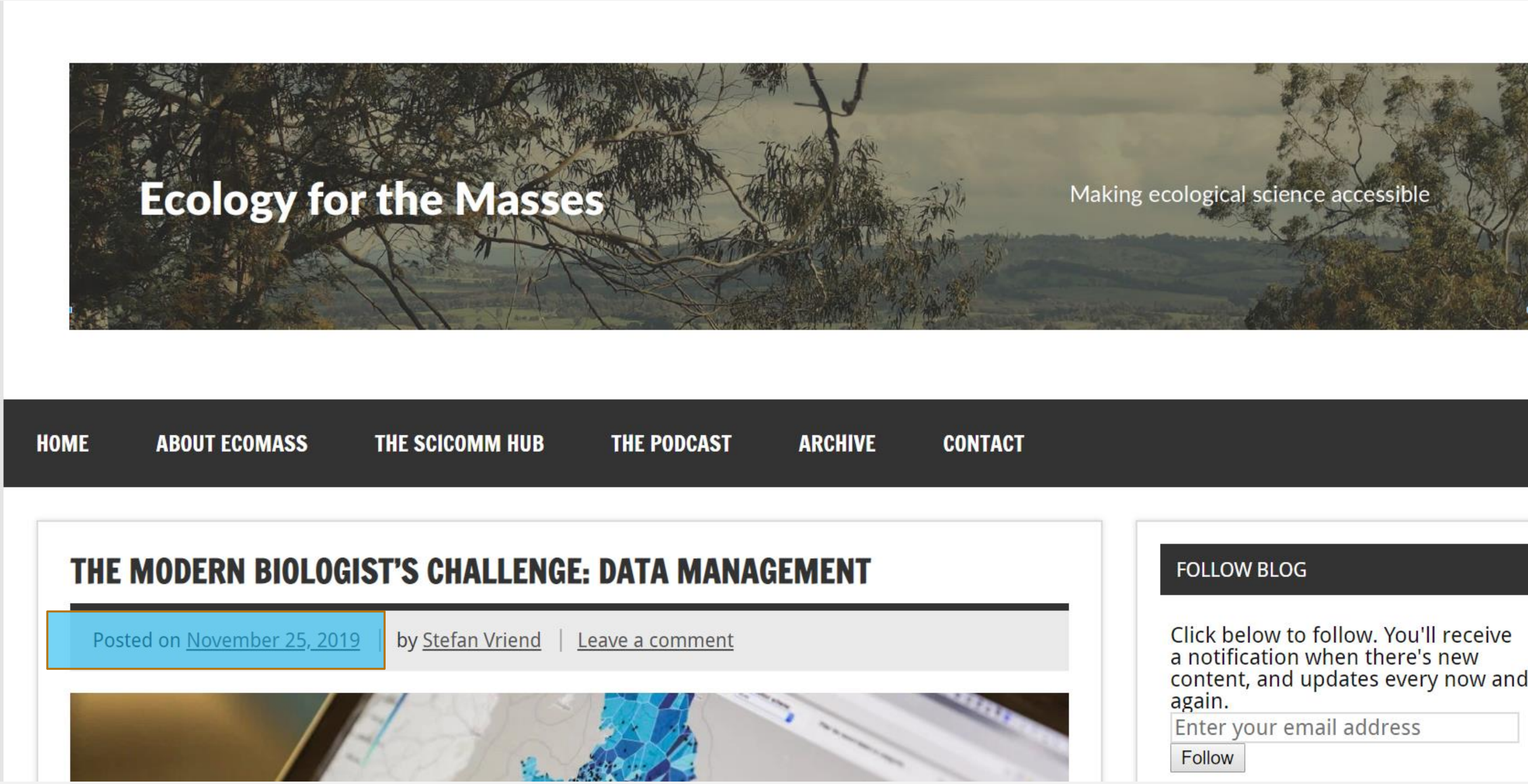
Modelação Ecológica

AULA 20

26th November 2019



*"…these days a large portion of modern biologists embark on very different journeys. Equipped with a computer full of **code** and mathematical **models**, they venture through a jungle of spreadsheets…"*

https://ecologyforthemasses.com/2019/11/25/the-modern-biologists-challenge-data-management/

World Scientists' Warning of a Climate Emergency

WILLIAM J. RIPPLE, CHRISTOPHER WOLF, THOMAS M. NEWSOME, PHOEBE BARNARD, WILLIAM R. MOOMAW, AND 11,258 SCIENTIST SIGNATORIES FROM 153 COUNTRIES (LIST IN SUPPLEMENTAL FILE S1)

Scientists have a moral obligation to clearly warn humanity of any catastrophic threat and to "tell it like it is." On the basis of this obligation and the graphical indicators presented below, we declare, with more than 11,000 scientist signatories from around the world, clearly and unequivocally that planet Earth is facing a climate emergency.

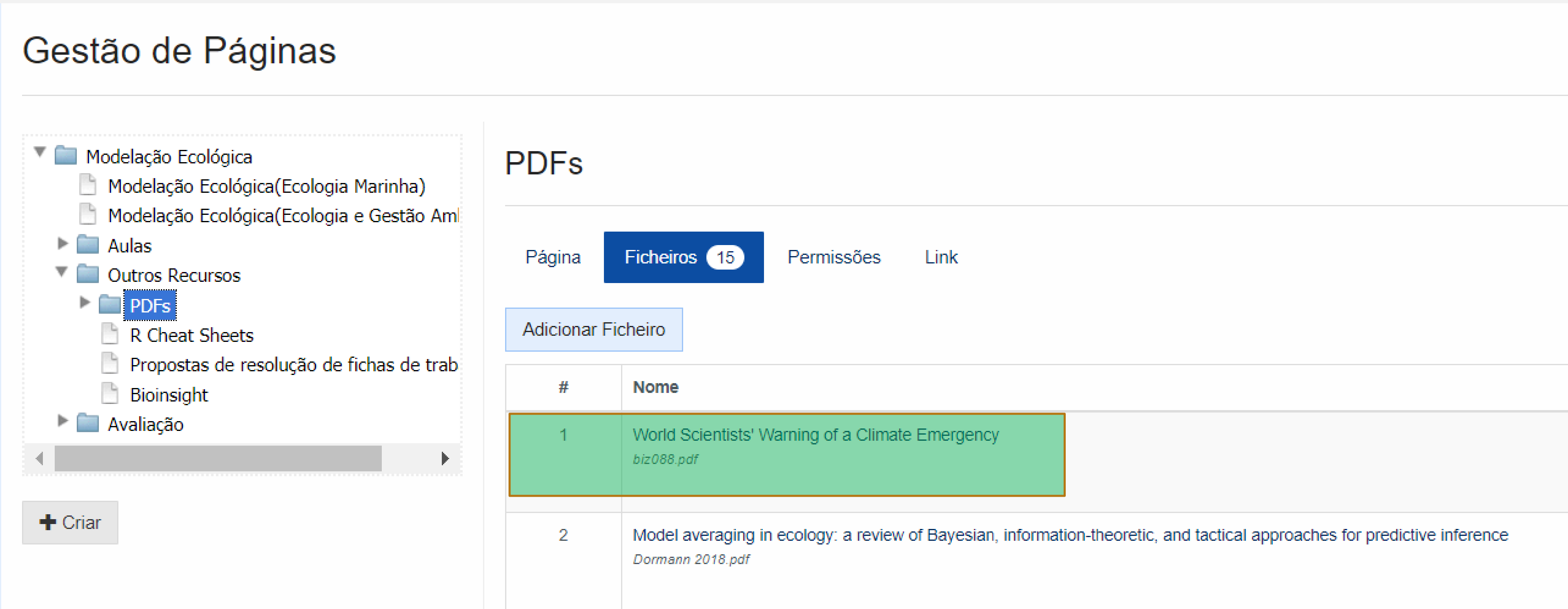
as actual climatic impacts (figure 2). We use only relevant data sets that are clear, understandable, systematically collected for at least the last 5 years, and updated at least annually.

The climate crisis is closely linked to excessive consumption of the wealthy lifestyle. The most affluent countries are mainly responsible for the historical GHG emissions and generally

forest loss in Brazil's Amazon has now started to increase again (figure 1g). Consumption of solar and wind energy has increased 373% per decade, but in 2018, it was still 28 times smaller than fossil fuel consumption (combined gas, coal, oil; figure 1h). As of 2018, approximately 14.0% of global GHG emissions were covered by carbon pricing (figure 1m), but

Scientists have a moral obligation to clearly warn humanity of any catastrophic threat and to "tell it like it is." On the basis of this obligation and the graphical indicators presented below, we declare, with more than 11,000 scientist signatories from around the world, clearly and unequivocally that planet Earth is facing a climate emergency.

https://doi.org/10.1093/biosci/biz088


Gestão de Páginas
Modelação Ecológica
Modelação Ecológica(Ecologia Marinha)
Modelação Ecológica(Ecologia e Gestão Aml
Aulas
Outros Recursos
PDFs
R Cheat Sheets
Propostas de resolução de fichas de trab
Bioinsight
Avaliação
Criar
PDFs
Página
Ficheiros 15
Permissões
Link
Adicionar Ficheiro
#
Nome
1
World Scientists' Warning of a Climate Emergency
biz088.pdf
2
Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference
Dormann 2018.pdf

GAM VISUALIZATION

mgcViz 0.1.4 | Reference | Articles

An introduction to mgcViz: visual tools for GAMs

Matteo Fasiolo and Raphael Nedellec

2019-06-20

Source: `vignettes/mgcviz.rmd`

Contents

- mgcViz basics
- Layered smooth effect plots
- Model checking
- Special plots
- Differences-between-smooths plots
- Plotting multiple slices of multi-dimensional smooth effects
- References

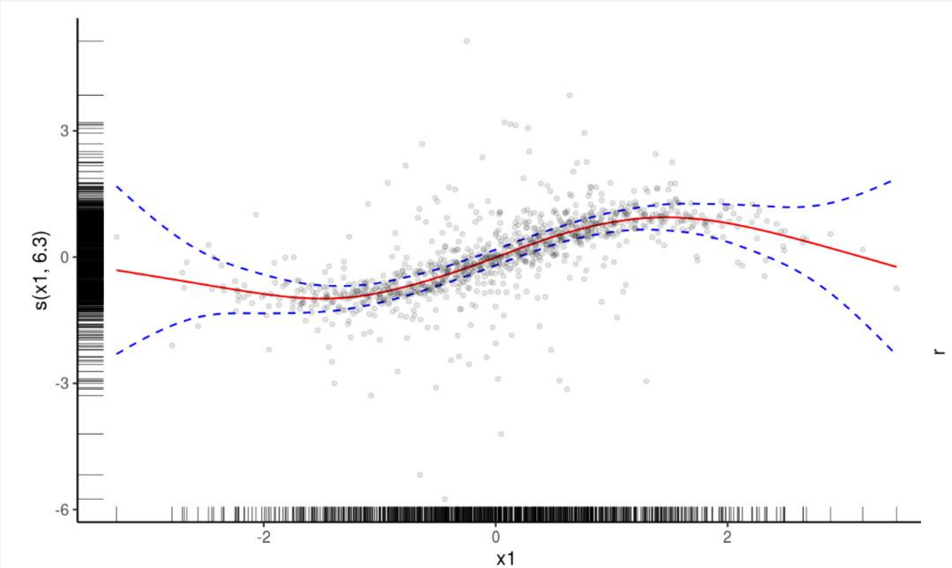
mgcViz basics

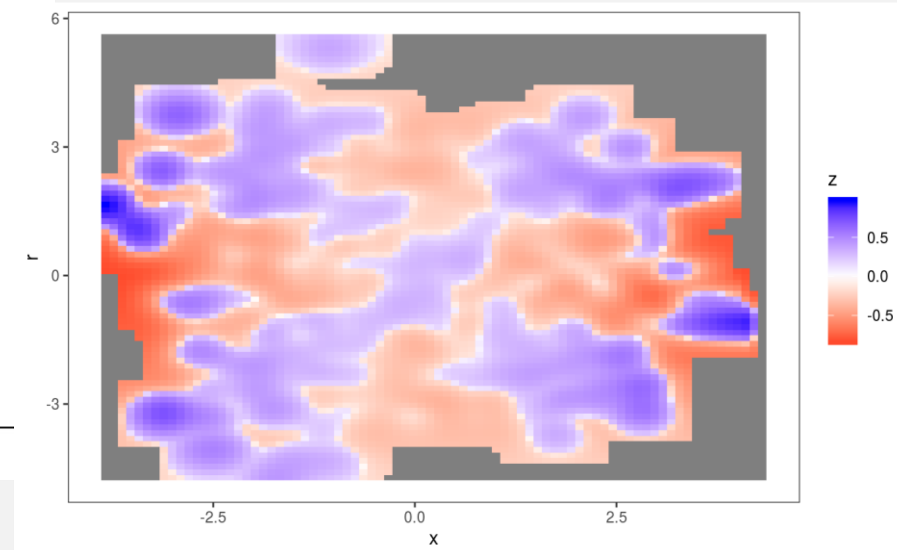
The `mgcViz` R package (Fasiolo et al, 2018) offers visual tools for Generalized Additive Models (GAMs). The visualizations provided by `mgcViz` differs from those implemented in `mgcv`, in that most of the plots are based on `ggplot2`'s powerful layering system. This has been implemented by wrapping several `ggplot2` layers and integrating them with computations specific to GAM models. Further, `mgcViz` uses binning and/or sub-sampling to produce plots that can scale to large datasets ($n \approx 10^7$), and offers a variety of new methods for visual model checking/selection.

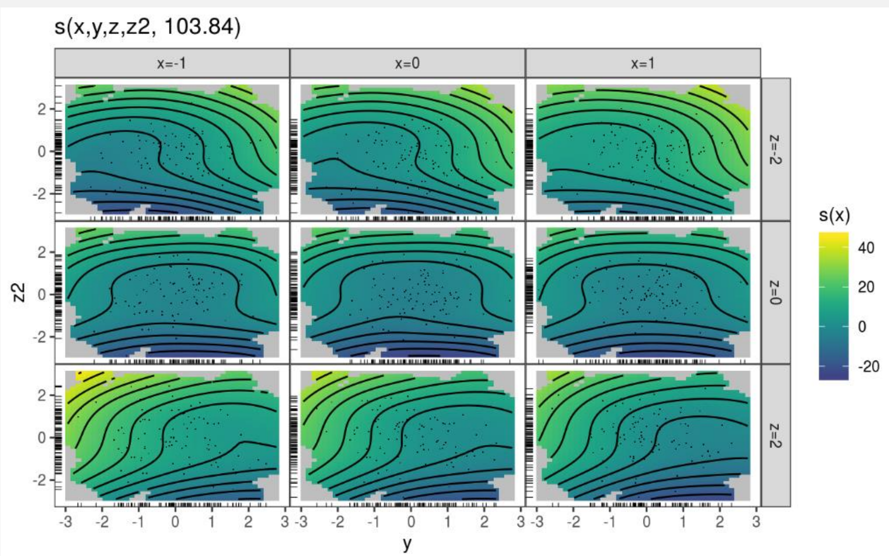
This document introduces the following categories of visualizations:

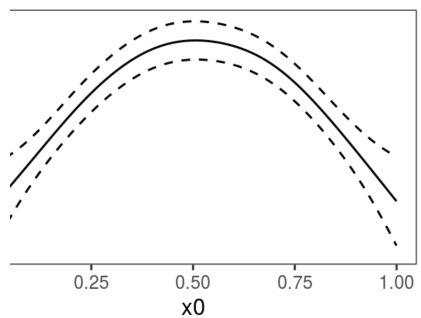
1. **smooth and parametric effect plots**: layered plots based on `ggplot2` and interactive 3d visualizations based on the `rgl` library;

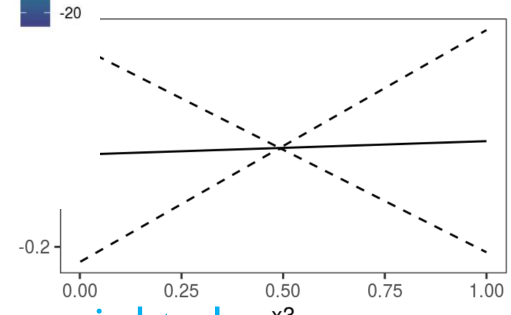
https://mfasiolo.github.io/mgcViz/articles/mgcviz.html

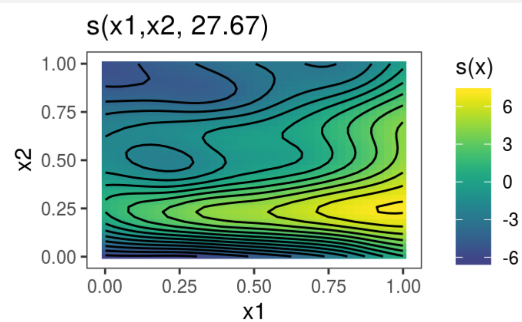


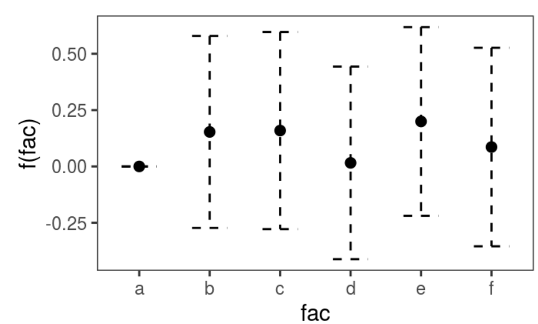








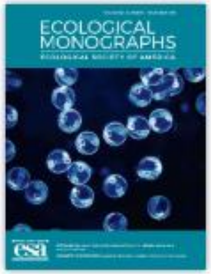




https://mfasiolo.github.io/mgcViz/articles/mgcviz.html

ECOLOGICAL MONOGRAPHS
ECOLOGICAL SOCIETY OF AMERICA

This was last week…so today is "next Tuesday"!



Volume 88, Issue 4
November 2018
Pages 485-504

Review | Full Access

Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference

Carsten F. Dormann, Justin M. Calabrese, Gurutzeta Guillera-Arroita, Eleni Matechou, Volker Bahn, Kamil Bartoń, Colin M. Beale, Simone Ciuti, Jane Elith, Katharina Gerstner ... See all authors

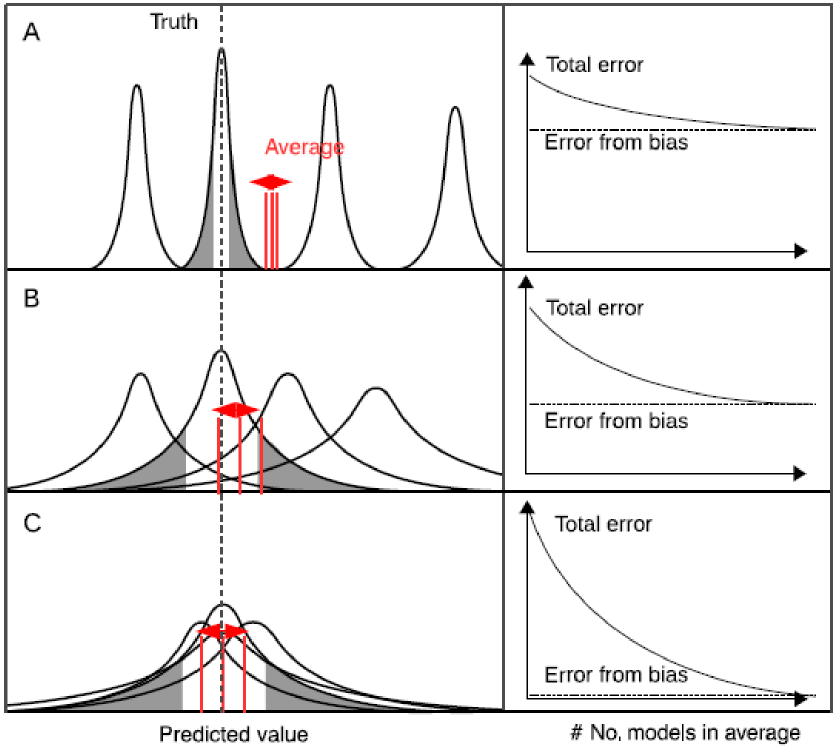
Figures | References | Related | Information

First published: 02 May 2018 | https://doi.org/10.1002/ecm.1309 | Citations: 16

Texto Integral @ b-on

Corresponding Editor: Perry de Valpine.

SECTIONS | PDF



Read this paper for next Tuesday: we will begin the day with a discussion around the topic of model averaging…

Gestão de Páginas

- Modelação Ecológica
  - Modelação Ecológica(Ecologia Marinha)
  - Modelação Ecológica(Ecologia e Gestão Ambiental)
  - Aulas
  - Outros Recursos
    - PDFs
    - R Cheat Sheets
    - Propostas de resolução de fichas de trabalho
    - Bioinsight
  - Avaliação

+ Criar

PDFs

Página | Ficheiros 14 | Permissões | Link

Adicionar Ficheiro

#	Nome
1	Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for *Dormann 2018.pdf*
2	Generalized additive models: Building evidence of air pollution, climate change and human health *Ravindraetal2019.pdf*

BIOMETRICS 53, 603–618
June 1997

Model Selection: An Integral Part of Inference

S. T. Buckland,[1] K. P. Burnham,[2] and N. H. Augustin[1]

[1]School of Mathematical and Computational Sciences, University of St. Andrews,
North Haugh, St. Andrews, Fife KY16 9SS

[2]Colorado Cooperative Fish and Wildlife Research Unit,
Fort Collins, Colorado 80523, U.S.A.

SUMMARY

We argue that model selection uncertainty should be fully incorporated into statistical inference whenever estimation is sensitive to model choice and that choice is made with reference to the data. We consider different philosophies for achieving this goal and suggest strategies for data analysis. We illustrate our methods through three examples. The first is a Poisson regression of bird counts in which a choice is to be made between inclusion of one or both of two covariates. The second is a line transect data set for which different models yield substantially different estimates of abundance. The third is a simulated example in which truth is known.

Dealing with correlation
Random Effects, Mixed Models & Generalized Estimating Equations

Wrapping up Mixed Models

A **random effect** is in a way, a suitable way to reduce the number of parameters to estimate associated with a factor covariate, in particular if you are mostly interested in the variation across the levels of that factor, and not in each factor *per se*.

Therefore, if the estimated variance of the **random effect** is small, that might mean that the random effect is not useful in explaining the variation on the data.

The problem is about defining **what is small**, and the fact that testing it formally involves testing a parameter at its boundary (a variance can't be lower than 0), which raises technical problems.

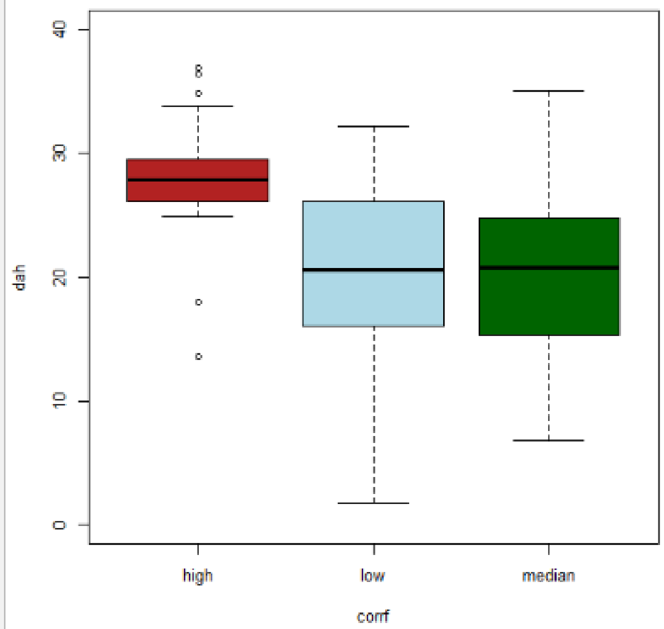
Quite often, we want to include a **random effect** because we know that it captures variability that we do not want to end up in the error term. This comes from the "design", not from the data itself. (e.g. observations collected in clusters, on same subjects over time, etc).

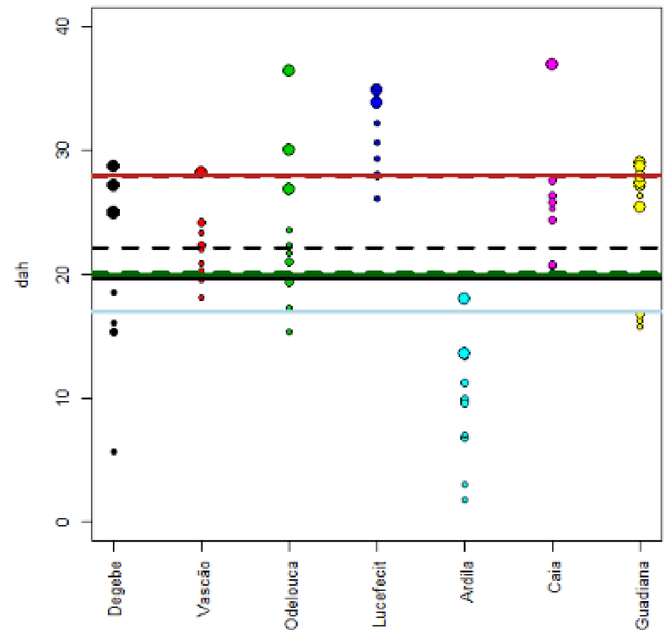
AN EXAMPLE WITH TWO FACTORS

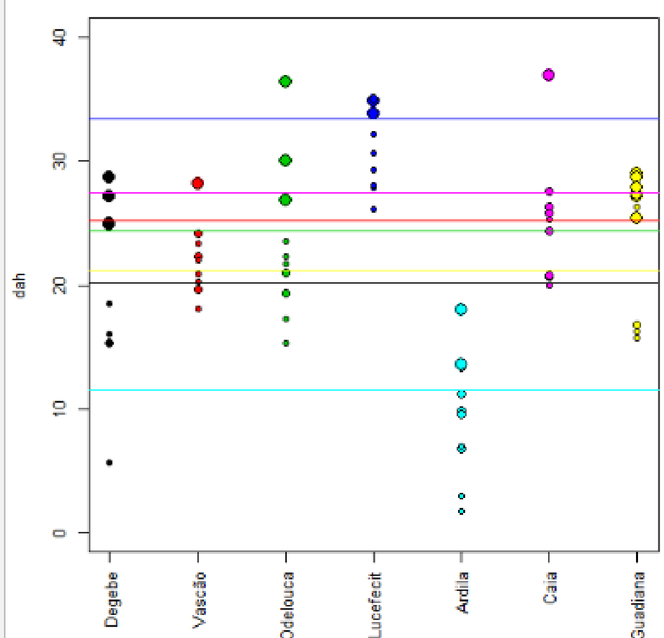
- Density of *Anaecypris hispanica* as a function of current velocity (`corrf`) and river (`riverI`) - dataAHD.txt
- One might be random… you have to explore
- Create a small report describing the data
- Model *A. hispanica* density as a function of the covariates
- Take you conclusions

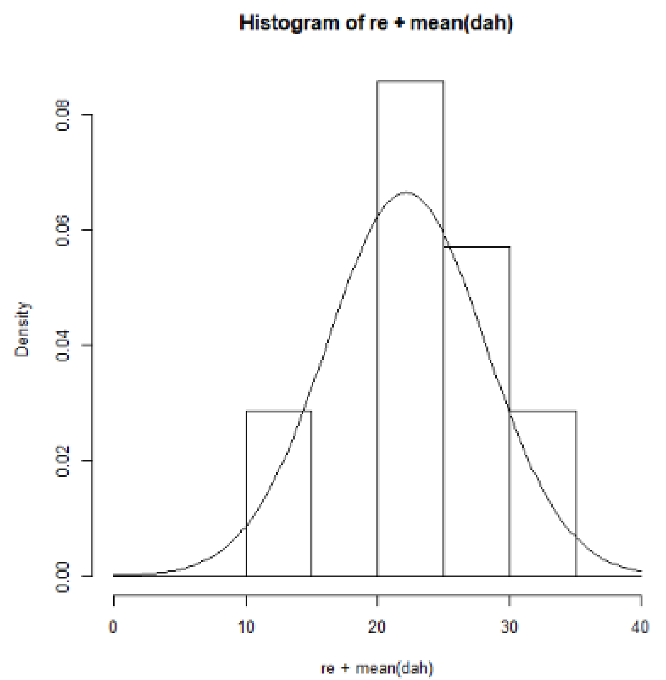
```
> head(dataAHD)
       dah riverI riverI2  corrf     corr
1 28.68104 Degebe       1   high 9.138035
2 24.91680 Degebe       1   high 7.011055
3 15.30073 Degebe       1 median 5.146127
4 16.01281 Degebe       1    low 1.667494
5 27.19677 Degebe       1   high 7.318869
6 25.01242 Degebe       1   high 9.406168
```

At the end of the class I'll give you (I've given you ;) my code that allows you to see how I generated the data and how the different models retrieve different components of the "truth".


dah
0
10
20
30
40
high
low
median
corrf


dah
0
10
20
30
40
Degebe
Vascão
Odelouca
Lucefecit
Ardila
Caia
Guadiana


dah
0
10
20
30
40
Degebe
Vascão
Odelouca
Lucefecit
Ardila
Caia
Guadiana


Histogram of re + mean(dah)
Density
0.00
0.02
0.04
0.06
0.08
0
10
20
30
40
re + mean(dah)

Using a fixed effects model

```
> summary(lm(dah~riverI+corrf))

Call:
lm(formula = dah ~ riverI + corrf)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.9709 -2.3200  0.1354  1.9468  8.5143 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)       16.009      1.309  12.226  < 2e-16 ***
riverICaia        16.629      1.484  11.205  < 2e-16 ***
riverIDegebe       9.563      1.502   6.366 2.82e-08 ***
riverIGuadiana    11.757      1.543   7.619 1.99e-10 ***
riverILucefecit   22.429      1.528  14.681  < 2e-16 ***
riverIOdelouca    13.810      1.499   9.211 3.75e-13 ***
riverIVascão      13.981      1.494   9.358 2.12e-13 ***
corrflow          -9.956      1.021  -9.756 4.54e-14 ***
corrfmedian       -7.289      1.163  -6.266 4.15e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.29 on 61 degrees of freedom
Multiple R-squared:  0.8457,    Adjusted R-squared:  0.8254 
F-statistic: 41.78 on 8 and 61 DF,  p-value: < 2.2e-16
```

Using velocity as a factor variable in the mixed effects model

```
> summary(lme(dah~corrf,random=~1|riverI,data=dataAHD))
Linear mixed-effects model fit by REML
 Data: dataAHD
       AIC      BIC    logLik
  391.4507 402.4742 -190.7254

Random effects:
 Formula: ~1 | riverI
        (Intercept) Residual
StdDev:    6.811164  3.28974

Fixed effects: dah ~ corrf
                Value Std.Error DF   t-value p-value
(Intercept) 28.588381  2.695333 61 10.606622       0
corrflow    -9.906773  1.019084 61 -9.721255       0
corrfmedian -7.305759  1.159858 61 -6.298837       0
 Correlation:
            (Intr) crrflw
corrflow    -0.235
corrfmedian -0.220  0.563

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.06202233 -0.68268447  0.04239849  0.58462158  2.57127416

Number of Observations: 70
Number of Groups: 7
```

Using velocity as a continuous variable in the mixed effects model

```
> summary(lme(dah~corr,random=~1|riverI,data=dataAHD))
Linear mixed-effects model fit by REML
 Data: dataAHD
       AIC      BIC    logLik
  410.7242 419.6022 -201.3621

Random effects:
 Formula: ~1 | riverI
        (Intercept) Residual
StdDev:    6.727382 3.706336

Fixed effects: dah ~ corr
                Value Std.Error DF  t-value p-value
(Intercept) 16.184240  2.691200 62 6.013763       0
corr         1.316578  0.168182 62 7.828293       0
 Correlation:
     (Intr)
corr -0.283

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-2.37019597 -0.65192157  0.02859504  0.55753777  2.35722099

Number of Observations: 70
Number of Groups: 7
```

Testing interactions in the fixed effects model

```
> summary(lm(dah~riverI*corrf))

Call:
lm(formula = dah ~ riverI * corrf)

Residuals:
   Min     1Q Median     3Q    Max 
-8.208 -1.511  0.000  1.684  6.891 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  15.7787     2.2506   7.011 6.38e-09 ***
riverICaia                   21.1612     3.8981   5.429 1.75e-06 ***
riverIDegebe                 10.6730     2.7564   3.872 0.000320 ***
riverIGuadiana               11.7973     2.5987   4.540 3.67e-05 ***
riverILucefecit              18.5553     3.1828   5.830 4.27e-07 ***
riverIOdelouca               15.3126     2.9055   5.270 3.04e-06 ***
riverIVascão                 12.3796     3.8981   3.176 0.002585 ** 
corrflow                    -11.9049     2.9055  -4.097 0.000156 ***
corrfmedian                  -5.6593     2.6629  -2.125 0.038633 *  
riverICaia:corrflow          -2.4137     4.8618  -0.496 0.621793    
riverIDegebe:corrflow        -0.6941     3.6752  -0.189 0.850978    
riverIGuadiana:corrflow       3.7617     3.6752   1.024 0.311071    
riverILucefecit:corrflow      6.6563     3.8670   1.721 0.091506 .  
riverIOdelouca:corrflow       0.8461     3.7208   0.227 0.821069    
riverIVascão:corrflow         4.4068     4.5011   0.979 0.332368    
riverICaia:corrfmedian       -6.5282     4.3207  -1.511 0.137233    
riverIDegebe:corrfmedian     -0.7446     3.8326  -0.194 0.846758    
riverIGuadiana:corrfmedian   -5.1701     4.3485  -1.189 0.240193    
riverILucefecit:corrfmedian   6.4526     4.7208   1.367 0.177916    
riverIOdelouca:corrfmedian   -5.2346     3.9412  -1.328 0.190270    
riverIVascão:corrfmedian     -0.4398     4.5385  -0.097 0.923192    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.183 on 49 degrees of freedom
Multiple R-squared:  0.884,     Adjusted R-squared:  0.8366 
F-statistic: 18.66 on 20 and 49 DF,  p-value: < 2.2e-16
```