NONPARAMETRIC/SEMIPARAMETRIC SURVIVAL ESTIMATION IN CAPTURE-RECAPTURE MODELS

Olivier GIMENEZ

CENTRE FOR RESEARCH INTO ECOLOGICAL AND ENVIRONMENTAL MODELLING

Snow petrels

Sociable weavers

Joint work with

Christophe BARBRAUD (France)

Ciprian CRAINICEANU (USA)

Stéphanie JENOUVRIER (France)

Byron J. T. MORGAN (England)

Rita COVAS (Scotland)

Mary and Charles BROWN (USA)

Mark ANDERSON (South Africa)

Thomas LENORMAND (France)

Individual encounter/capture histories

individual	**history**	t_1	t_2	t_3	$\ldots$	t_{K-1}	t_K	**sex**
1	h_1	1	1	0	$\ldots$	1	1	*male*
2	h_2	1	1	0	$\ldots$	1	1	*male*
$\vdots$								
j	h_j	0	1	0	$\ldots$	0	1	*female*
$\vdots$								

Parameters :

Probabilities of survival over time interval $[t_i, t_{i+1}]$ for females/males :

$$\phi_i^F, \phi_i^M$$

Probabilities of detection/capture at time t_i for females/males :

$$p_i^F, p_i^M$$

Probability of an individual encounter history

$$[h_1|\boldsymbol{\phi}, \boldsymbol{p}] = \Pr(1 \;\; 1 \;\; 0 \;\; \ldots \;\; 1 \;\; 1 \quad \mathsf{M}) =$$

$$\phi_1^M \times p_2^M \times \phi_2^M \times (1 - p_3^M) \times \ldots \times \phi_{26}^M \times p_{27}^M$$

Probability of an individual encounter history

$$[h_1|\boldsymbol{\phi}, \boldsymbol{p}] = \Pr\left(1 \;\; 1 \;\; 0 \;\; \ldots \;\; 1 \;\; 1 \quad \mathsf{M}\right) =$$

$$\phi_1^M \times p_2^M \times \phi_2^M \times (1 - p_3^M) \times \ldots \times \phi_{26}^M \times p_{27}^M$$

Probability of an individual encounter history

$$[h_1|\boldsymbol{\phi}, \boldsymbol{p}] = \Pr\left(1 \;\; 1 \;\; 0 \;\; \ldots \;\; 1 \;\; 1 \quad \mathsf{M}\right) =$$

$$\phi_1^M \times p_2^M \times \phi_2^M \times (1 - p_3^M) \times \ldots \times \phi_{26}^M \times p_{27}^M$$

Likelihood

$$[\boldsymbol{h}|\boldsymbol{\phi},\boldsymbol{p}] \quad \propto \quad \prod_j [h_j|\boldsymbol{\phi},\boldsymbol{p}]$$

Cormack (1964), Jolly (1965) et Seber (1965)

Classical relationship between survival and a covariate

Lebreton et al. (1992)

$$\log\left(\frac{\phi}{1-\phi}\right) \quad = \quad \beta_0 + \beta_1\, x$$

- Pros : $\phi \in [0;1]$
- Cons 1 : ϕ is completely determined by the covariate
- Cons 2 : the relationship survival/covariate is linear on the logistic scale (or quadratic)
- Objective : β_0 and β_1 are parameters to be estimated

(Less) Classical relationship between survival and a covariate

$$\log\left(\frac{\phi}{1-\phi}\right) \quad = \quad \beta_0 + \beta_1\, x + \varepsilon$$

- Same Pros
- Cons 1 is relaxed thanks to random effects ε
  1. Cope with effects non captured by the covariate, overdispersion (*Barry et al. 2002*)
  2. Allow to cope with temporal autocorrelation (*Johnson et Hoeting 2003*)
- Problem : complex likelihood, yet to be evaluated (approximations, MCMC methods, SIS...)
- Cons 2 : the relationship survival/covariate is linear on the logistic scale (or quadratic)

Nonlinearities?

Snow petrels

Environmental covariates

Effect of climatic conditions

Sociable weavers

Individual covariates

Effect of natural selection

Aims of the talk

To propose a non-parametric relationship survival/covariate

$$\log\left(\frac{\phi}{1-\phi}\right) = m(x) + \varepsilon$$

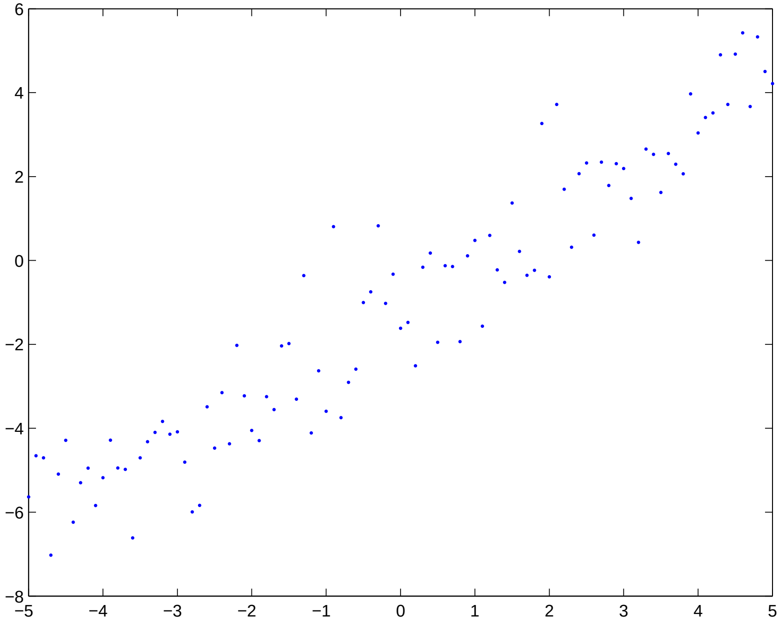
Plan

- Brief introduction to *penalized splines*
- Use of penalized splines in a *mixed model framework*
- *Bayesian approach through MCMC methods (using WinBUGS)*
- Illustrations in ecology and evolution
- Discussion

Plan

- Brief introduction to *penalized splines*
- Use of penalized splines in a *mixed model framework*
- *Bayesian approach through MCMC methods (using WinBUGS)*
- Illustrations in ecology and evolution
- Discussion

linearities...



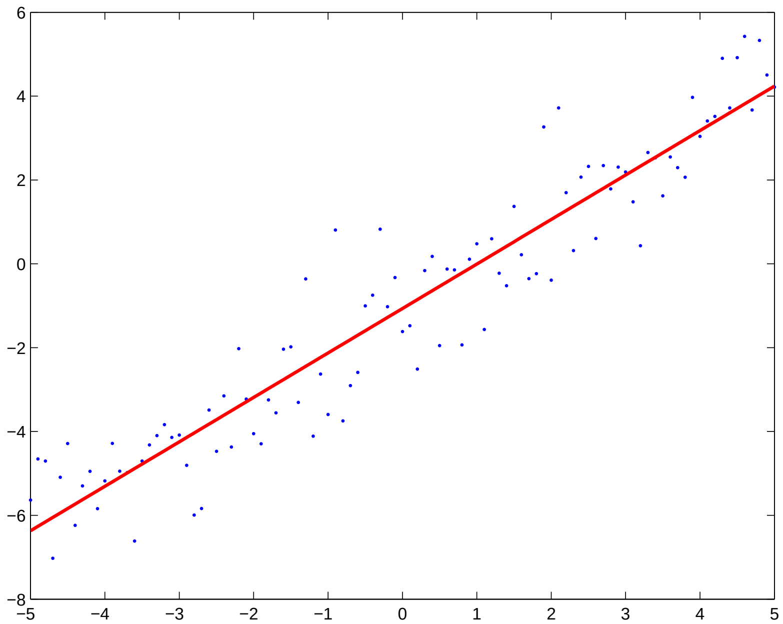
$$y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i?$$

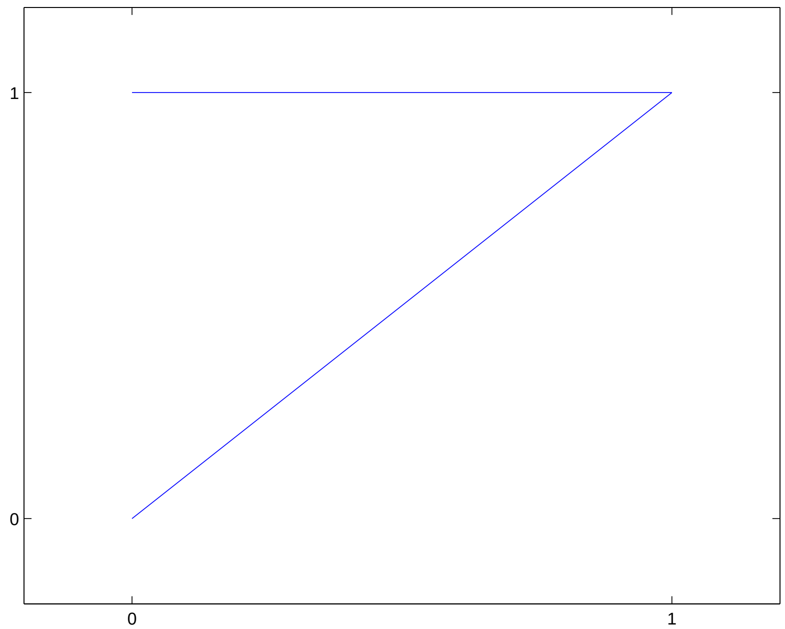
ordinary least squares

search for $\hat{\boldsymbol{\eta}}$ that minimizes $||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\eta}||^2$ with $\boldsymbol{y} = (y_1, \ldots, y_n)^T$

$$\hat{\boldsymbol{\eta}} = (\boldsymbol{X}^T\ \boldsymbol{X})^{-1}\ \boldsymbol{X}^T\ \boldsymbol{y}$$

linearities...



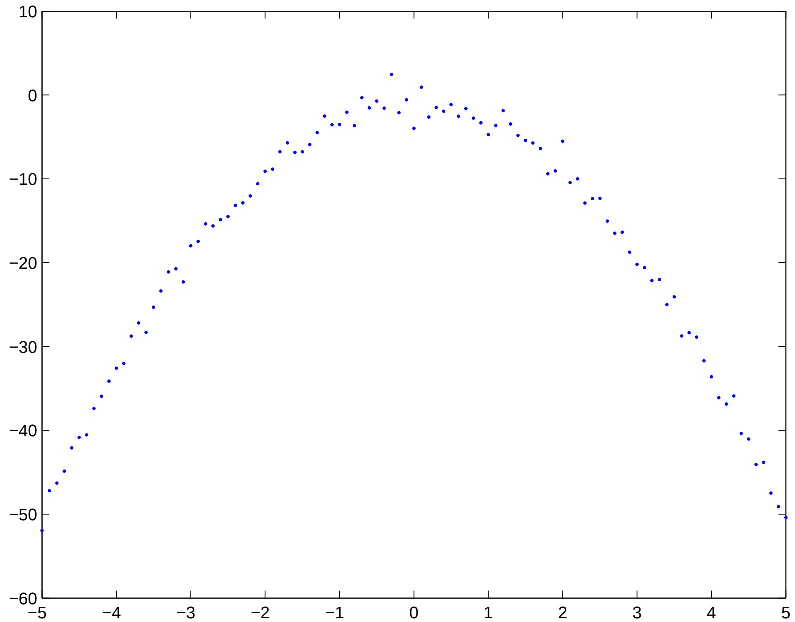


$$\boldsymbol{X_i} = (1, x_i)$$

$$\boldsymbol{\eta} = (\beta_0, \beta_1)^T$$

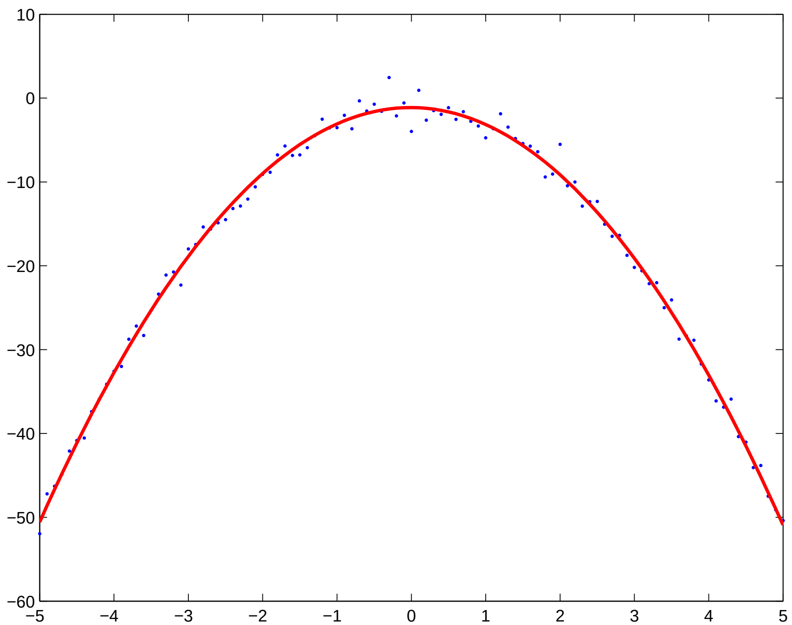
base $\{1, x\}$

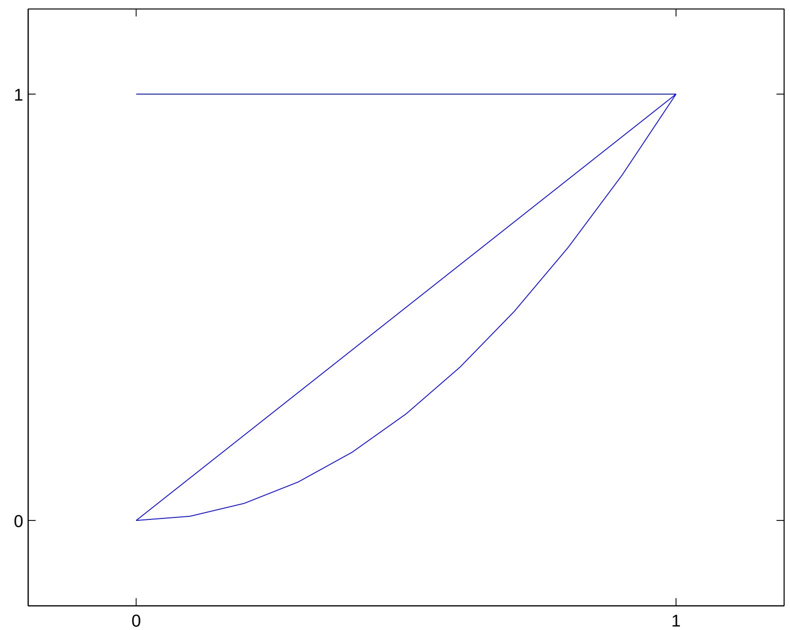
quadraticities...



$$y_i = \beta_0 + \beta_1 \times x_i + \beta_2 \times x_i^2 + \varepsilon_i?$$

quadraticities...



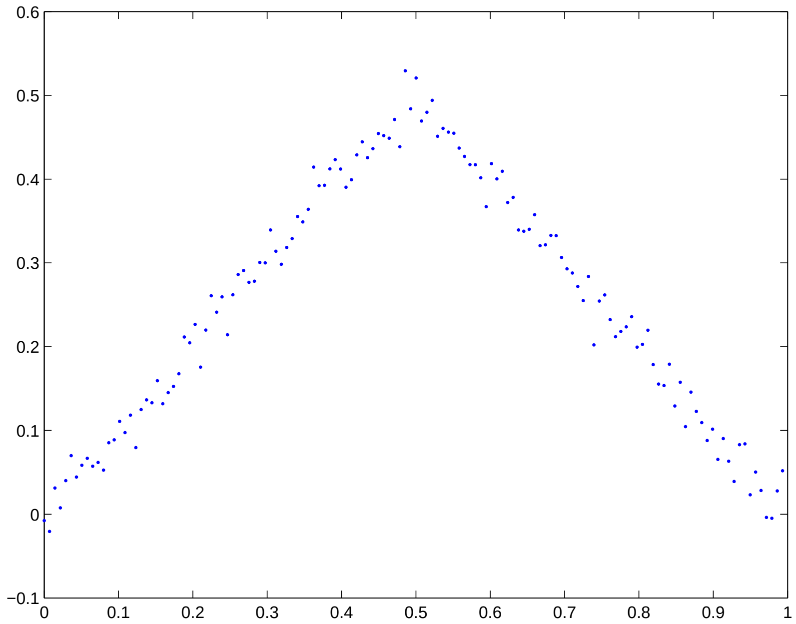


$$\boldsymbol{X_i} = (1, x_i, x_i^2)$$

$$\boldsymbol{\eta} = (\beta_0, \beta_1, \beta_2)^T$$

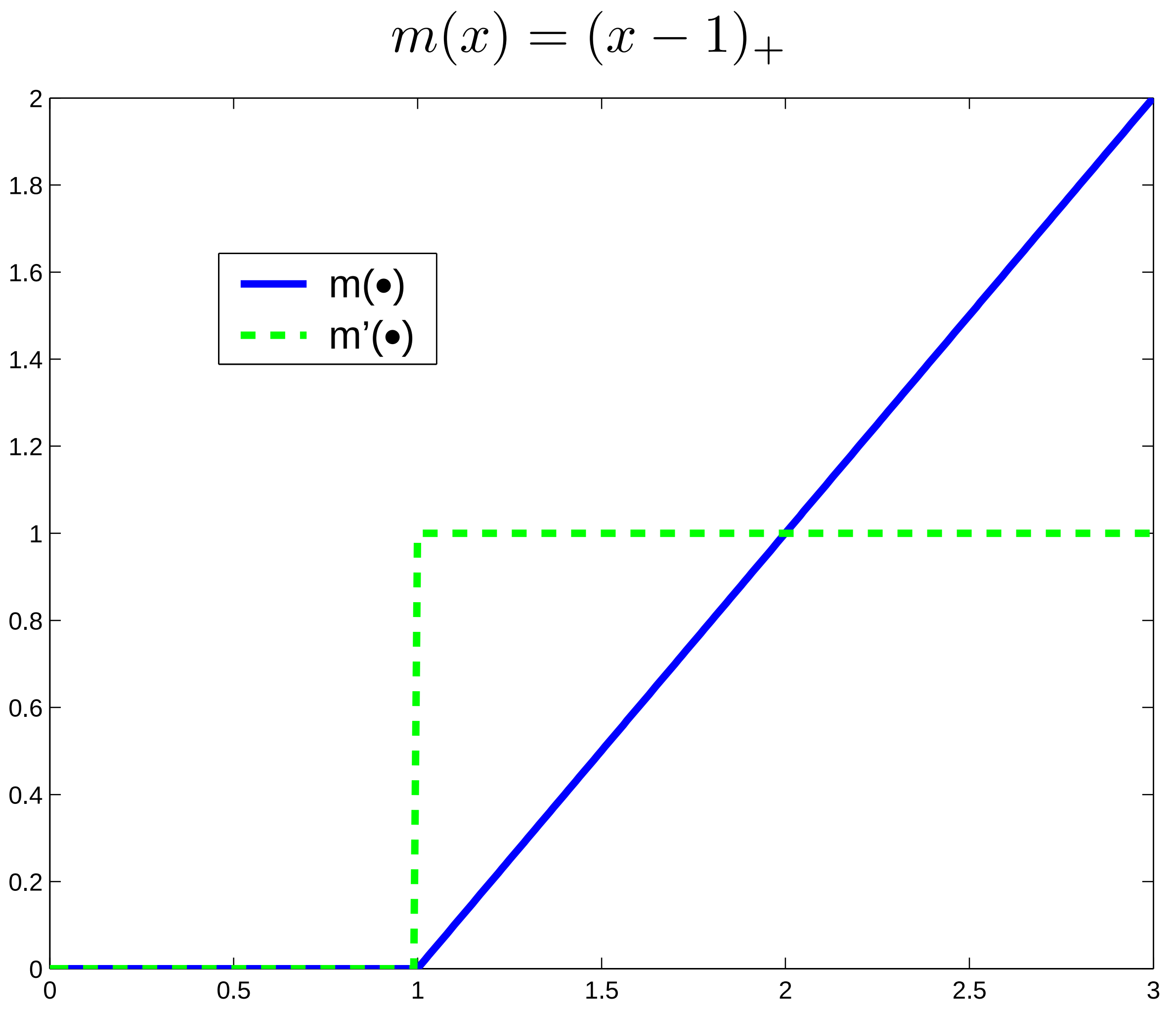
basis $\{1, x, x^2\}$

broken lines...

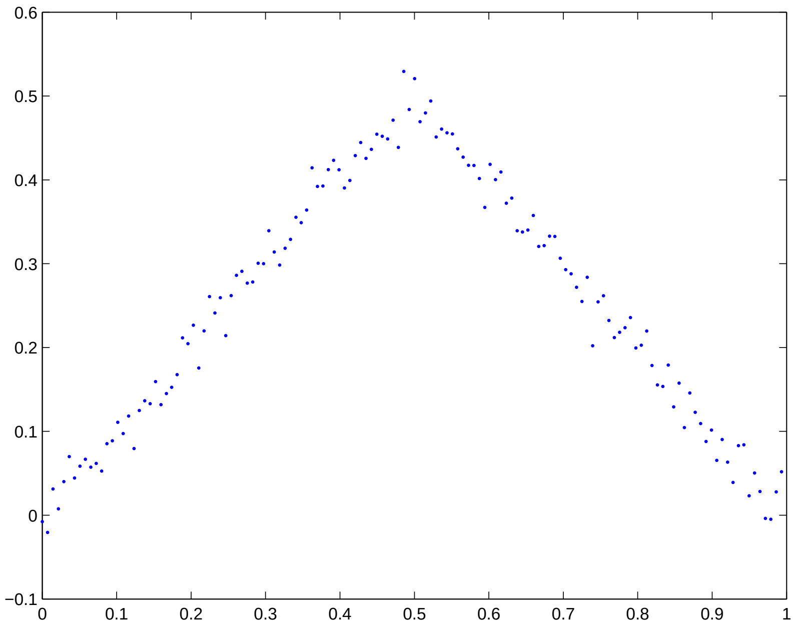


?????

truncated power functions $(u)_+^p = u^p I_{(u \geq 0)}$

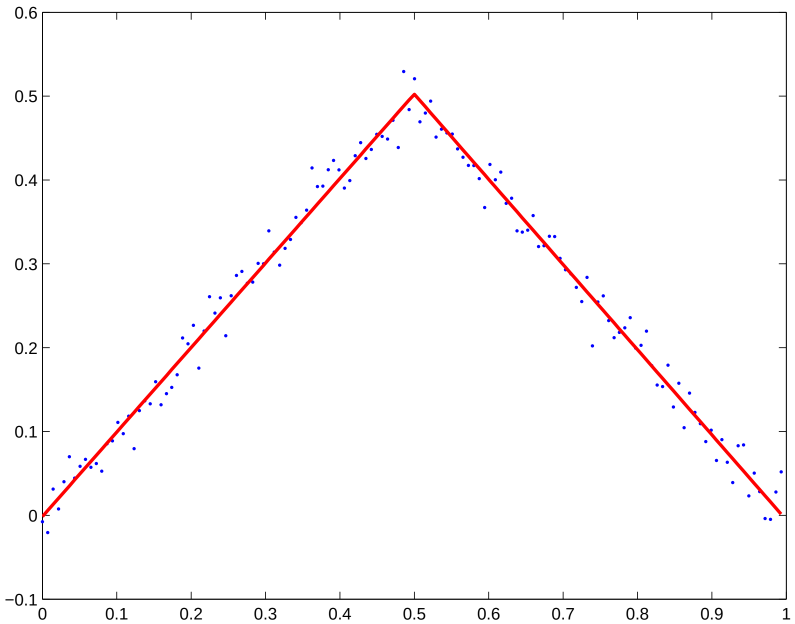


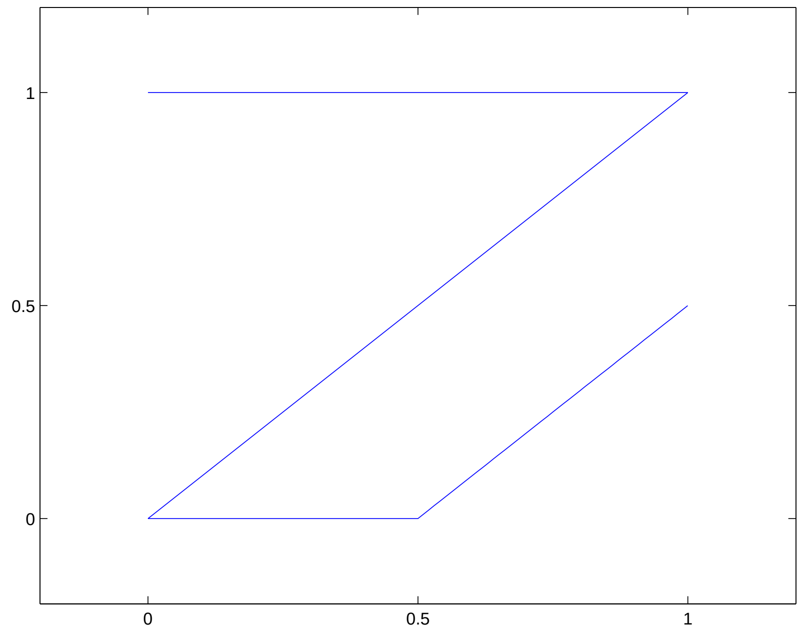
broken lines...



$$y_i = \beta_0 + \beta_1 \times x_i + b_1 \times (x_i - 0.5)_+ + \varepsilon_i?$$

broken lines...





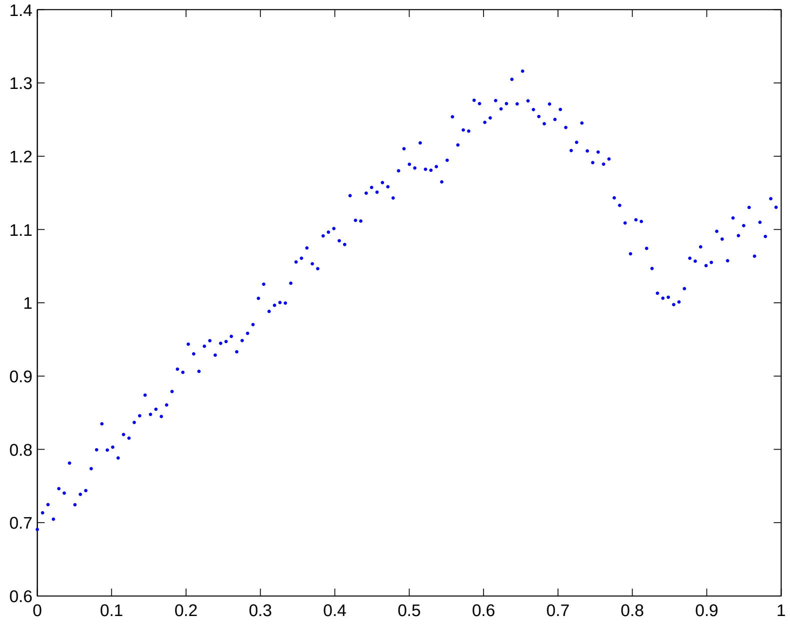
$$\boldsymbol{X_i} = (1, x_i, (x_i - 0.5)_+)$$

$$\boldsymbol{\eta} = (\beta_0, \beta_1, b_1)^T$$

linear splines basis

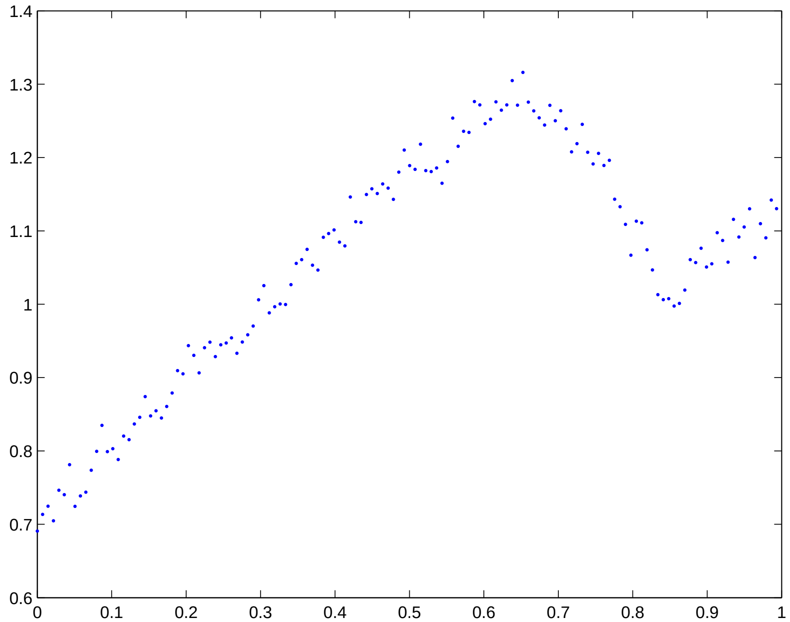
$$\{1, x, (x - 0.5)_+\}$$

non-linearities...



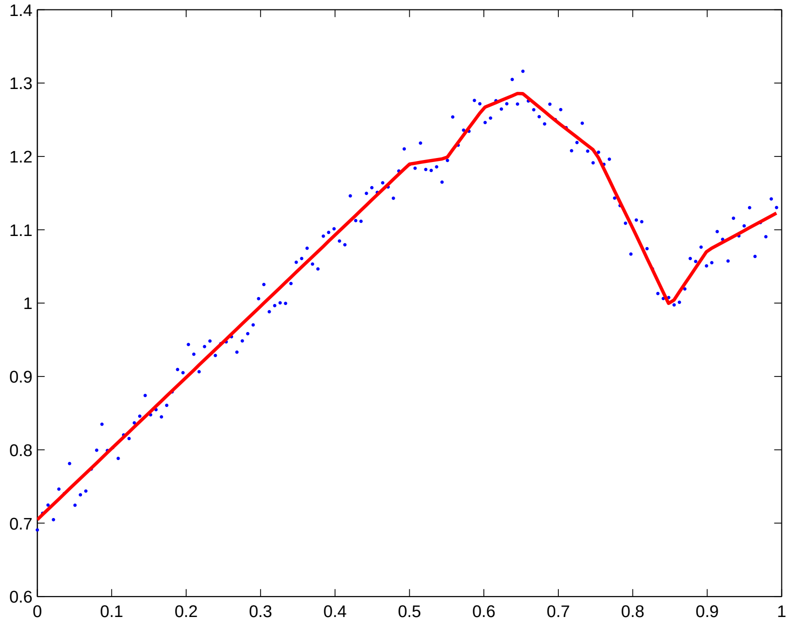
????

non-linearities...



$$y_i = \beta_0 + \beta_1 \times x_i + \sum_{k=1}^{K} b_k \times (x_i - \kappa_k)_+ + \varepsilon_i?$$

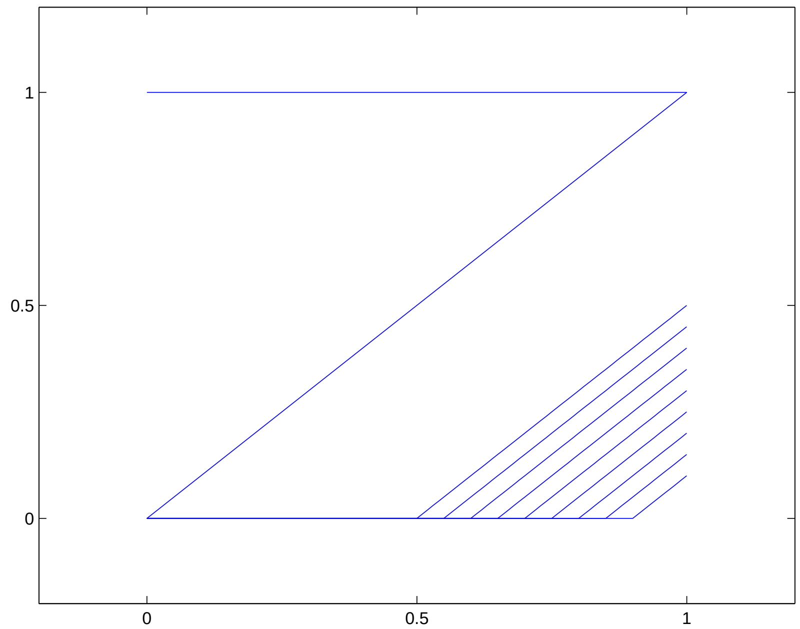
non-linearities...
non-parametric fitting using linear splines



$$\boldsymbol{X_i} = (1, x_i, (x_i - 0.5)_+, (x_i - 0.55)_+, \ldots, (x_i - 0.9)_+)$$

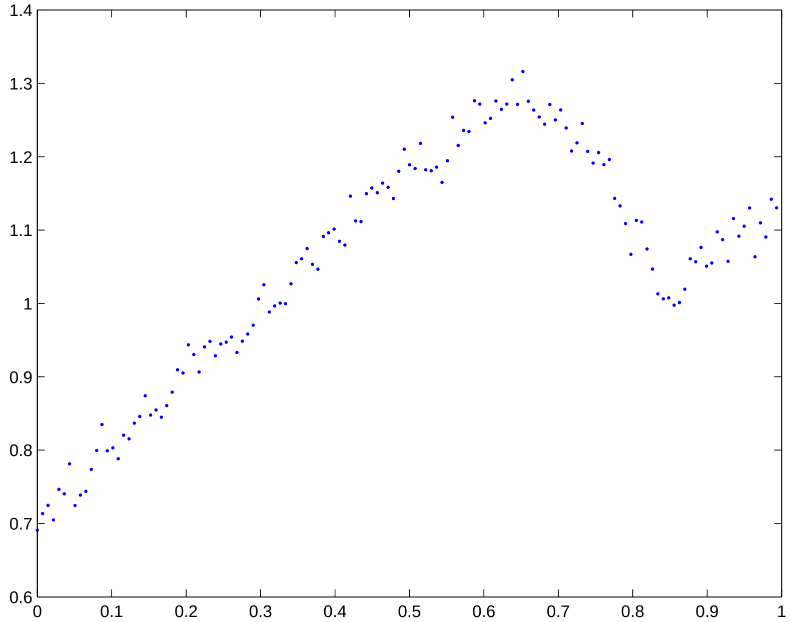
$$\boldsymbol{\eta} = (\beta_0, \beta_1, b_1, b_2, \ldots, b_9)^T$$

non-linearities...
linear splines basis



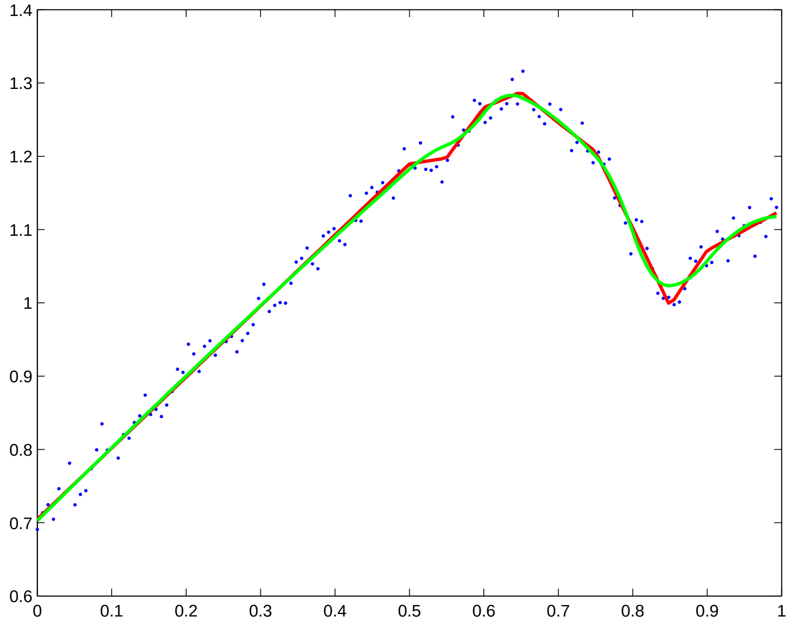
$$\{1, x, (x-0.5)_+, (x-0.55)_+, \ldots, (x-0.9)_+\}$$

non-linearities...



$$y_i = \beta_0 + \beta_1 \times x_i + \beta_2 \times x_i^2 + \sum_{k=1}^{K} b_k \times (x_i - \kappa_k)_+^2 + \varepsilon_i?$$

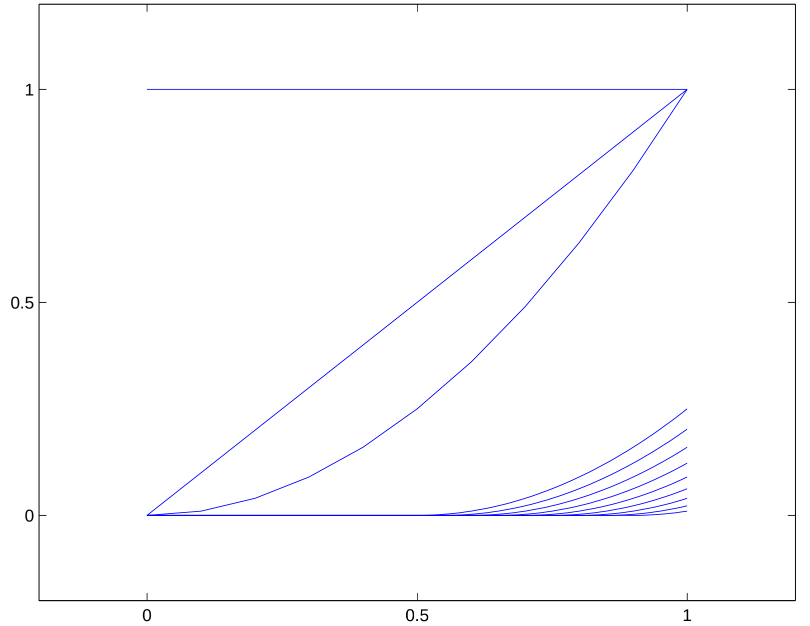
non-linearities...
non-parametric fitting using quadratic splines



$$\boldsymbol{X_i} = (1, x_i, x_i^2, (x_i - 0.5)_+^2, (x_i - 0.55)_+^2, \ldots, (x_i - 0.9)_+^2)$$

$$\boldsymbol{\eta} = (\beta_0, \beta_1, \beta_2, b_1, b_2, \ldots, b_9)^T$$

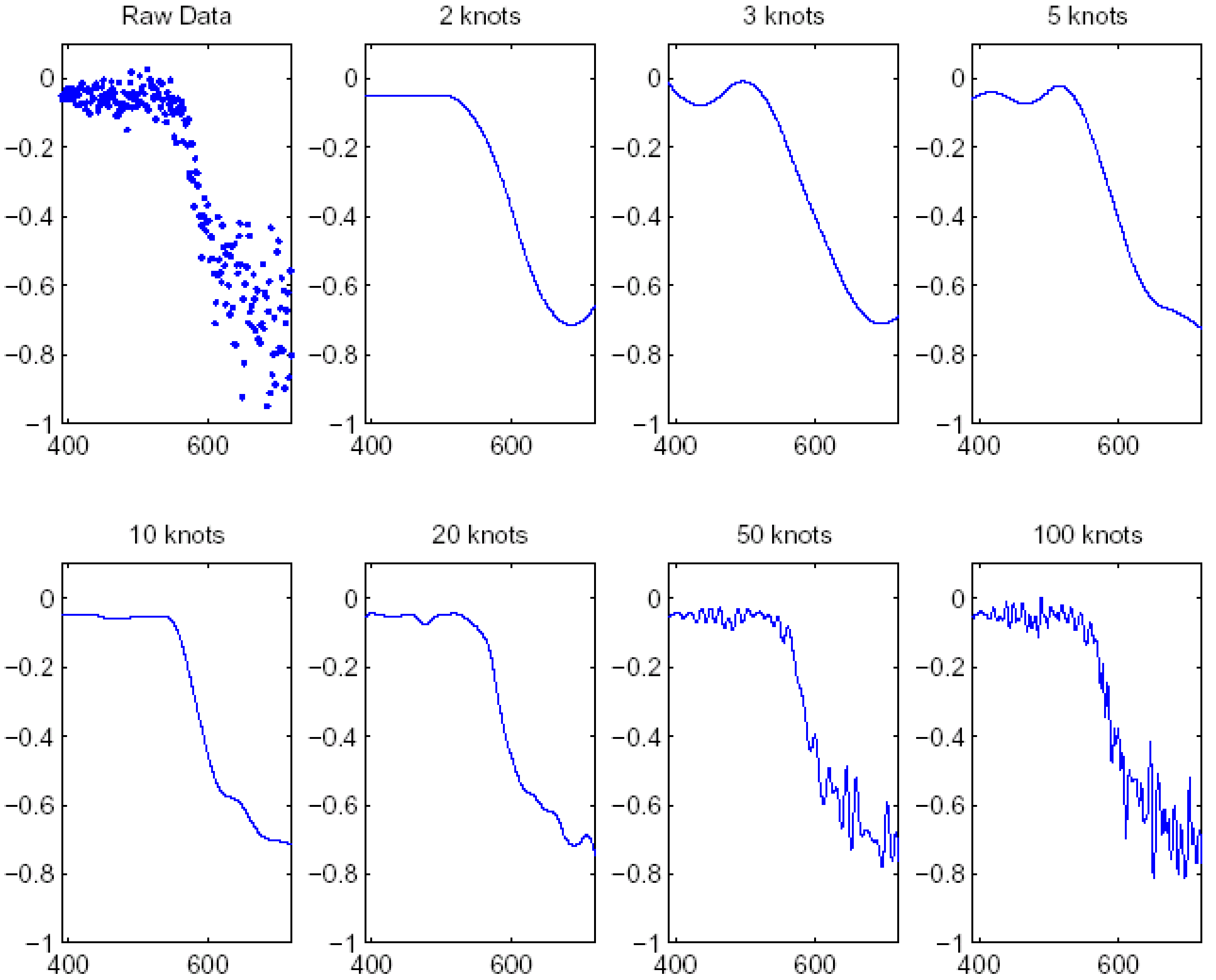
non-linearities...
quadratic splines basis



$$\{1, x, x^2, (x-0.5)_+^2, (x-0.55)_+^2, \ldots, (x-0.9)_+^2\}$$

Influence of the number and location of the knots?

Non-parametric approach via Ordinary Least Squares

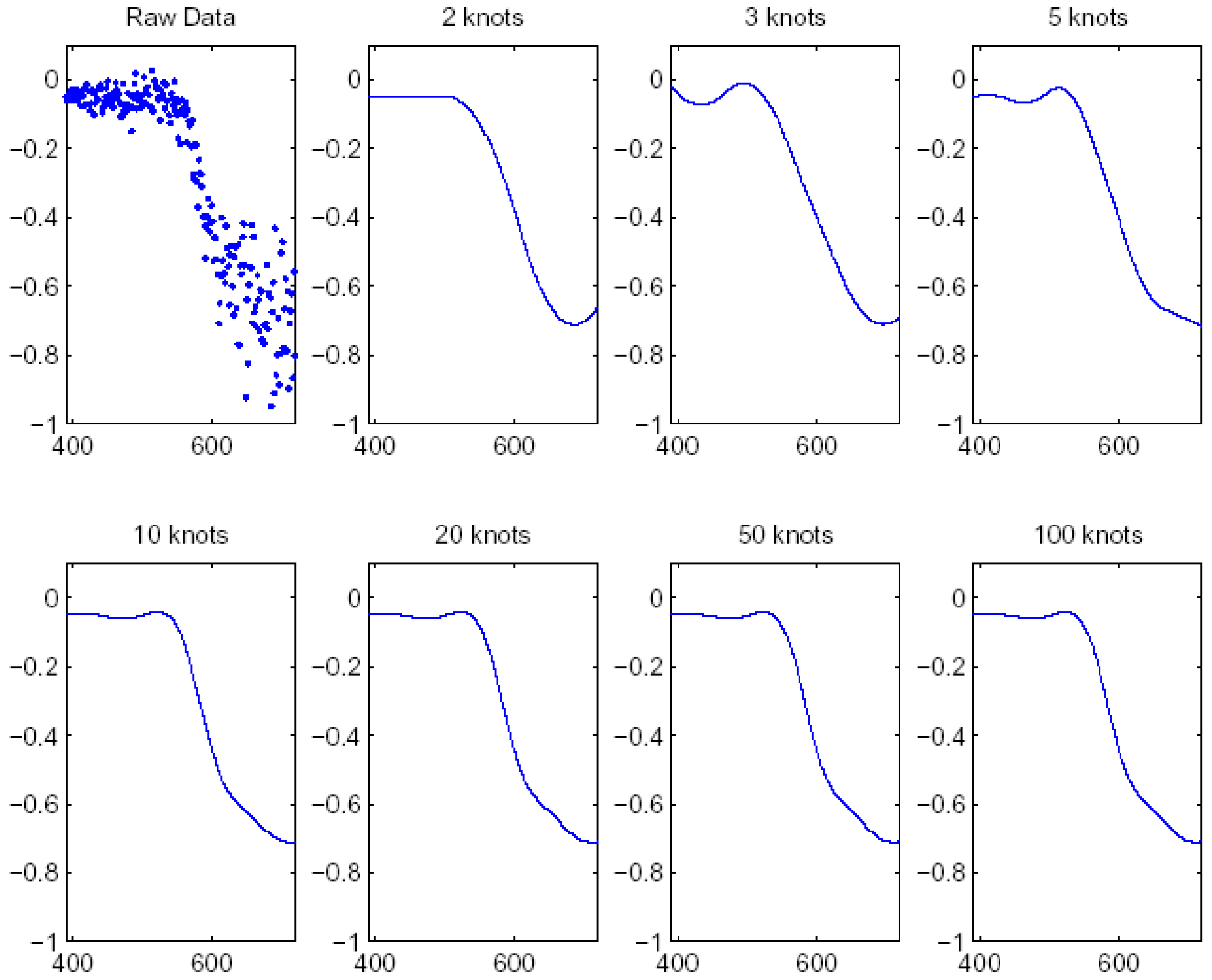


After Ruppert et al. (2003)

Suggestions (Ruppert et al. 2003)

- penalize the b_ks in order to constraint the influence of knots : penalized splines or P-splines
- search for $\hat{\boldsymbol{\eta}}$ that minimizes $||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\eta}||^2$ given the constraint $\boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{b} \leq C$
- it can be shown that it is equivalent to search for $\hat{\boldsymbol{\eta}}$ that minimizes $||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\eta}||^2 + \lambda^2\boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{b}$
- λ is the smoothing parameter
- $\hat{\boldsymbol{\eta}}_\lambda = (\boldsymbol{X}^T \boldsymbol{X} + \lambda^2 \boldsymbol{D})^{-1} \boldsymbol{X}^T \boldsymbol{y}$ with $\boldsymbol{D} = \begin{pmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times K} \\ \mathbf{0}_{K\times 2} & \mathbf{I}_{K\times K} \end{pmatrix}$
- the number of knots is $K = \min\left(\frac{1}{4}n, 35\right)$
- the knot κ_k are the "equally-spaced sample quantiles" i.e. the sample quantiles of the x_i's corresponding to probabilities $k/(K+1)$

Penalized non-parametric approach



After Ruppert et al. (2003)

Plan

- Brief introduction to *penalized splines*
- Use of penalized splines in a *mixed model framework*
- *Bayesian approach through MCMC methods (using WinBUGS)*
- Illustrations in ecology and evolution
- Discussion

Penalized splines within a Mixed model framework

$$y_i = m(x_i) + \varepsilon_i, \qquad \varepsilon_i \text{ i.i.d.} \sim N(0, \sigma_\varepsilon^2)$$

with

$$m(x_i) = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} b_k (x_i - \kappa_k)_+$$

Penalized splines within a Mixed model framework

$$\boldsymbol{\beta} = (\beta_0, \beta_1)^T \text{ et } \boldsymbol{b} = (b_1, \ldots, b_K)^T \; ; \quad b_k \text{ i.i.d.} \sim N(0, \sigma_b^2)$$

$$\boldsymbol{X}_i = (1, x_i) \text{ et } \boldsymbol{Z}_i = ((x_i - \kappa_1)_+, \ldots, (x_i - \kappa_K)_+)$$

$$\lambda = \frac{\sigma_\varepsilon}{\sigma_b}$$

$$\mathbf{y} \quad = \quad \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \varepsilon, \quad \text{cov}\begin{pmatrix} \mathbf{b} \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \sigma_b^2 \mathbf{I}_{K\times K} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I}_{n\times n} \end{pmatrix}$$

Comments

- With Ordinary Least Squares, $\sigma_b^2 \longrightarrow +\infty$
- Having b_k random allows to attenuate the effect of $(x_i - \kappa_k)_+$
- The implementation is much easier in standard software
- Unifying framework useful for extensions

Plan

- Brief introduction to *penalized splines*
- Use of penalized splines in a *mixed model framework*
- *Bayesian approach through MCMC methods (using WinBUGS)*
- Illustrations in ecology and evolution
- Discussion

Bayesian approach :

We assume the components of $\boldsymbol{\theta} = \left(\beta, \mathbf{b}, \varepsilon, \sigma_b^2, \sigma_\varepsilon^2, \mathbf{p}\right)$ are random variables.

We'd like to determine the aposteriori distribution given the data, using Bayes's theorem:

$$[\boldsymbol{\theta}|\boldsymbol{h}] \propto [\boldsymbol{h}|\boldsymbol{\phi}, \boldsymbol{p}][\boldsymbol{\phi}|\boldsymbol{\beta}, \boldsymbol{b}, \varepsilon][\boldsymbol{\beta}][\boldsymbol{b}|\sigma_b^2][\varepsilon|\sigma_\varepsilon^2][\sigma_b^2][\sigma_\varepsilon^2][\boldsymbol{p}]$$

Use of MCMC methods : generate observations from a Markov chain having the target distribution as stationnary distribution $[\boldsymbol{\theta}|\boldsymbol{h}]$.

WinBUGS implementation

Example 1: Snow petrels

Environmental covariates

Environmental covariates: Snow petrels




Nonparametric model

Southern Oscillation Index (SOI)

$$\text{logit}(\phi_i^l) \quad = \quad \beta_0 + \beta_1 \text{SOI}_i + \sum_{k=1}^{7} b_k \left(\text{SOI}_i - \kappa_k\right)_+ + \varepsilon_i$$

Semiparametric model

$$\text{logit}(\phi_i^l) \quad = \quad \beta_0 + \gamma \mathsf{SEX} + \beta_1 \mathsf{SOI}_i + \sum_{k=1}^{7} b_k \left(\mathsf{SOI}_i - \kappa_k\right)_+ + \varepsilon_i$$

$$\mathsf{SEX} = \begin{cases} 1 & \text{si} \quad l = F \quad \text{i.e. female} \\ 0 & \text{otherwise} \end{cases}$$

Semiparametric model
mixed model representation: fixed effects

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \gamma & \beta_1 \end{pmatrix}^T$$

$$X = \begin{pmatrix} 1 & 1 & \mathsf{SOI}_1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & \mathsf{SOI}_{28} \\ 1 & 0 & \mathsf{SOI}_1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & \mathsf{SOI}_{28} \end{pmatrix}$$

Semiparametric model
mixed model representation: random effects

$$\boldsymbol{b} = \begin{pmatrix} b_1 & \ldots & b_7 \end{pmatrix}^T$$

$$Z = \begin{pmatrix} (\mathsf{SOI}_1 - \kappa_1)_+ & \ldots & (\mathsf{SOI}_1 - \kappa_7)_+ \\ \vdots & \vdots & \vdots \\ (\mathsf{SOI}_{28} - \kappa_1)_+ & \ldots & (\mathsf{SOI}_{28} - \kappa_7)_+ \end{pmatrix}$$

A priori

$$[p] = U[0,1]$$

$$[\gamma]\,,[\beta_0]\,,[\beta_1] = N(0,10^6)$$

$$[b_k] = N(0,\sigma_b^2)$$

$$[\varepsilon_i] = N(0,\sigma_\varepsilon^2)$$

$$\left[\sigma_b^2\right],\quad \left[\sigma_\varepsilon^2\right] = \Gamma^{-1}(10^{-6},10^{-6})$$

Technically

- 2 chains run in parallel with overdispersed initial values
- 1100000 iterations with 100000 burn-in iterations 100000
- compute the Gelman-Rubin - $\hat{R}$ - to check for convergence ($\hat{R} = 1$)

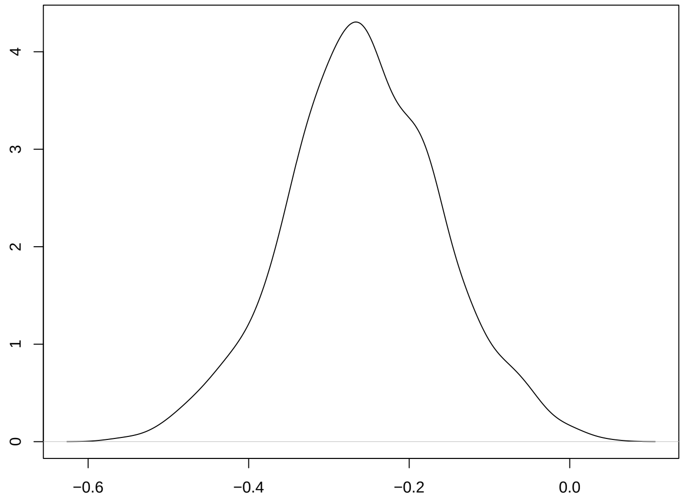
RESULTS

Snow petrel

Difference in survival according to sex

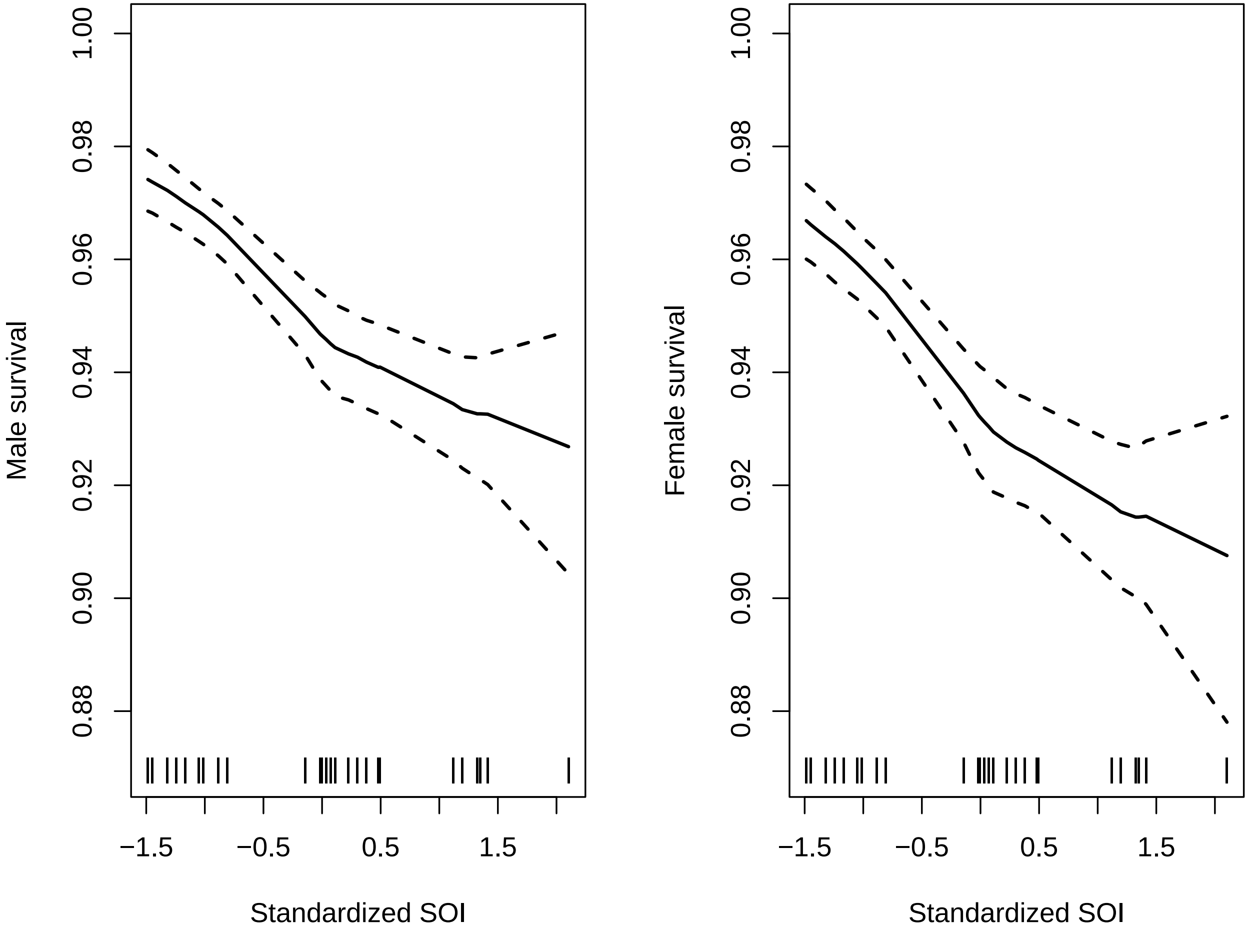
Males survive better than females:

$$\gamma = -0.26 \ \ (-0.45; -0.06)$$

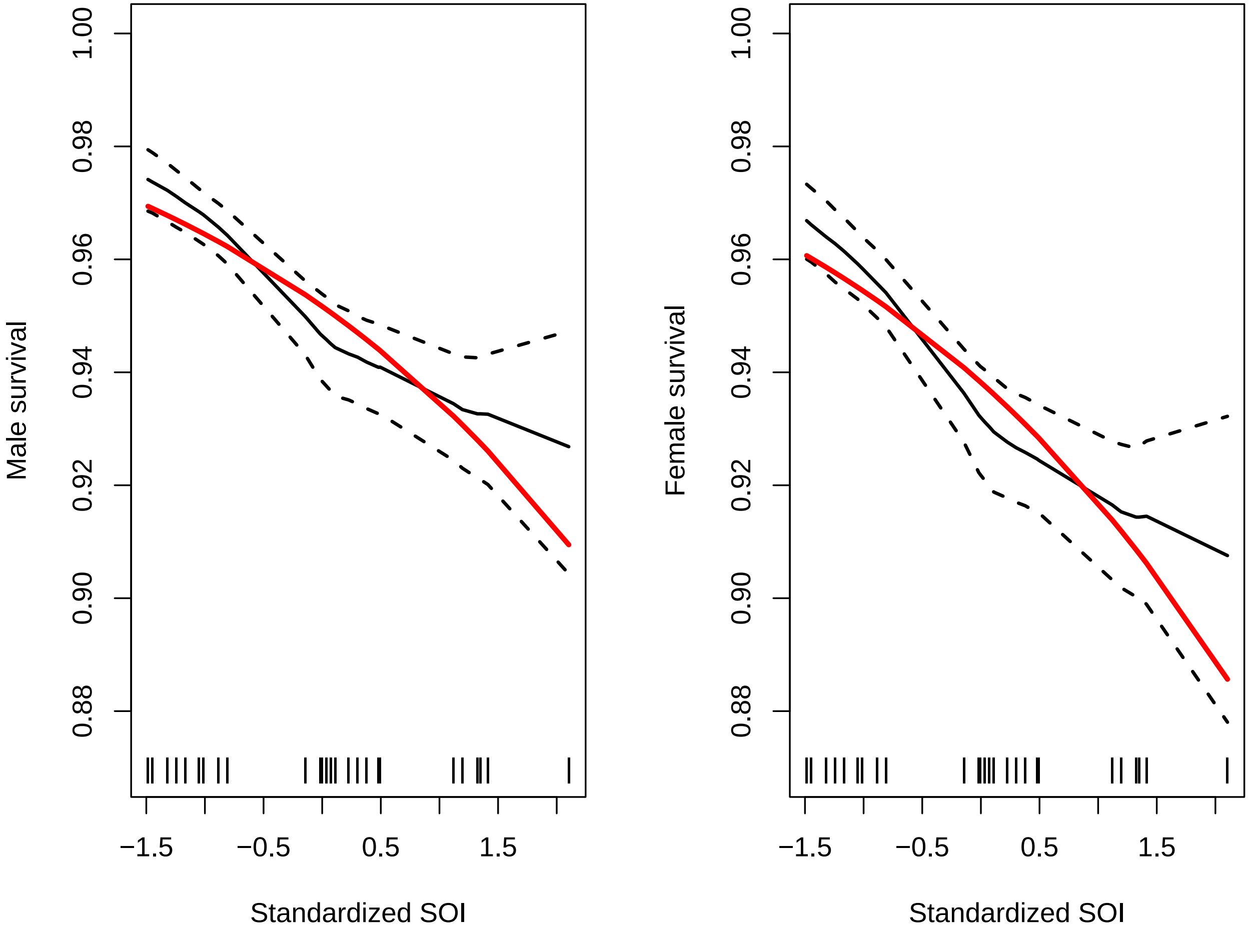


Posterior distribution of γ

Survival/SOI using a P-spline approach



Survival/SOI using a linear-logistic approach

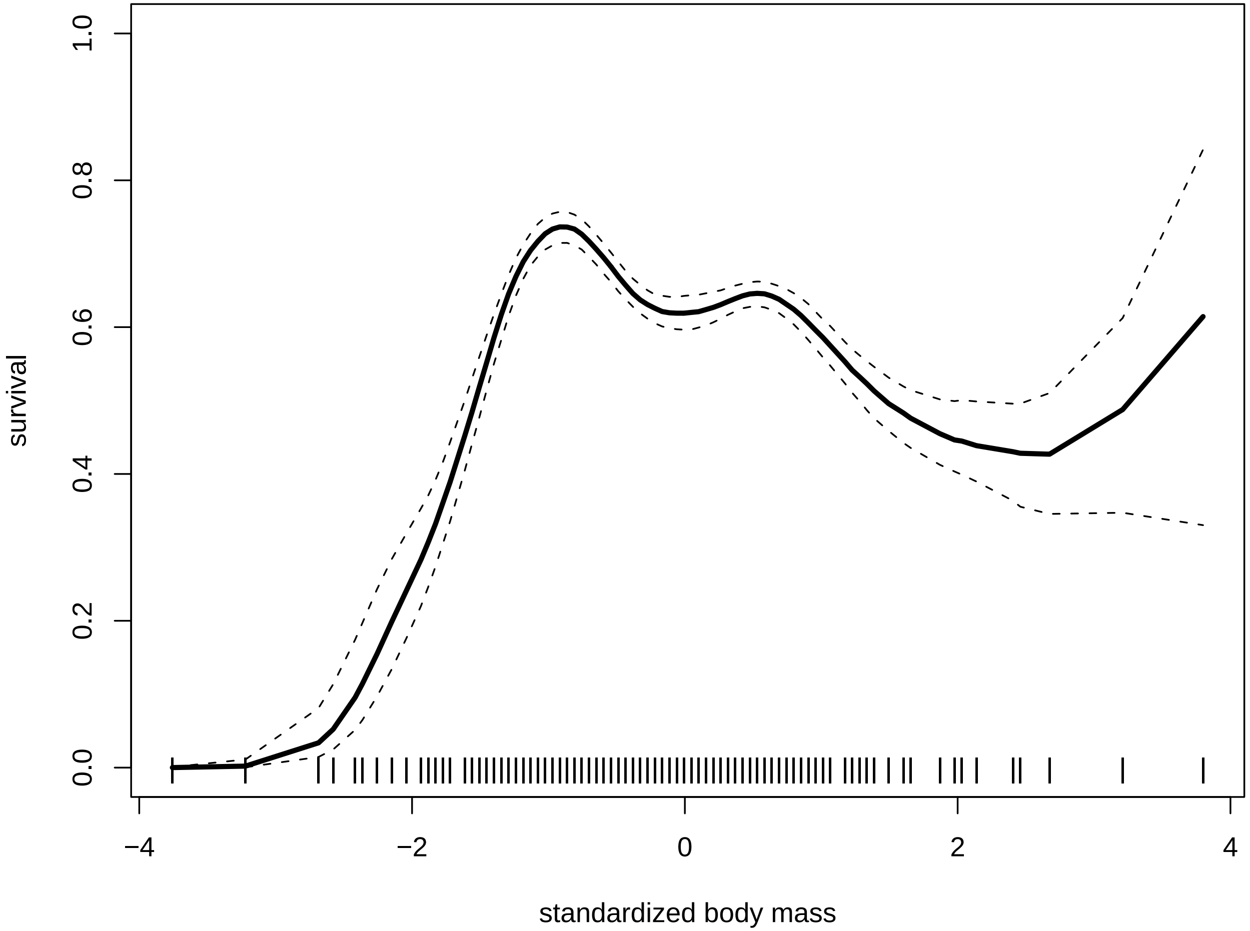


Example 2: Sociable weavers

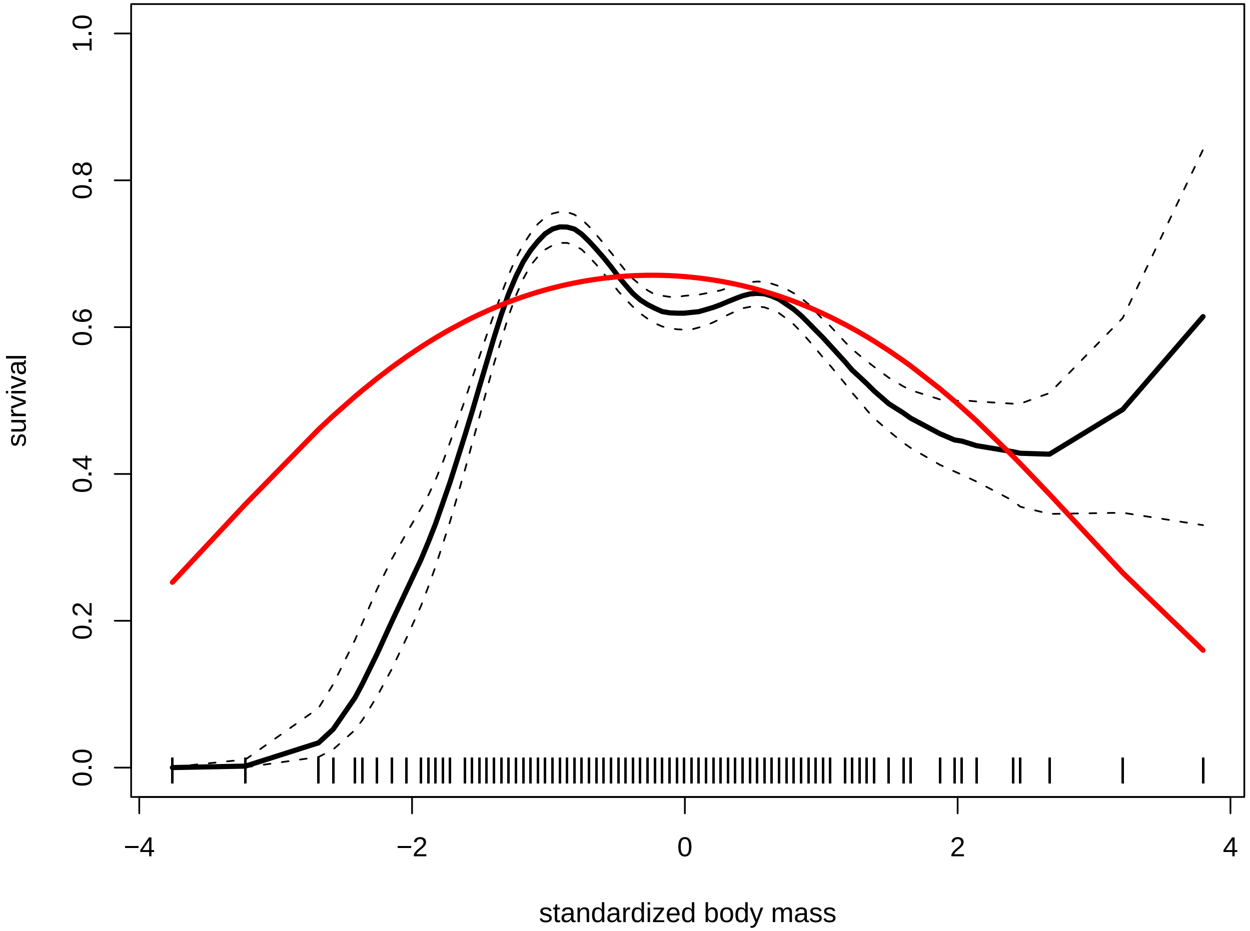
Individual covariates



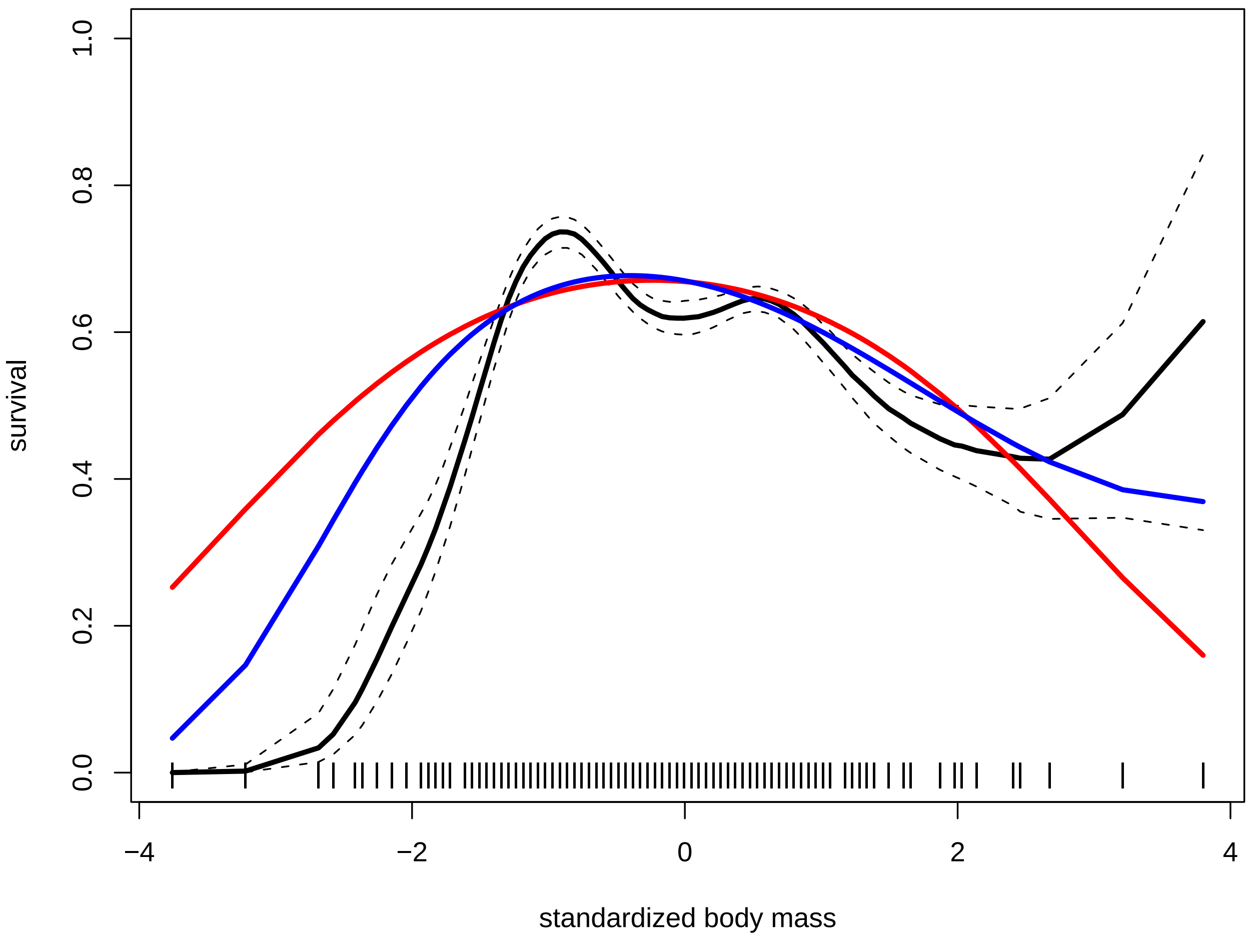

Survival/body mass using a P-spline approach



Survival/body mass using a quadratic-logistic approach



Survival/body mass using a cubic-logistic approach



Summary

- Nonparametric/semiparametric approach for estimating survival in capture-recapture models
- Bayesian inference through MCMC methods (WinBUGS)
- Splines functions useful by themselves, or can suggest a parametric alternative if desired

Perspectives

- Coping with interactions
- Dealing with multiple covariates
- Modelling senescence
- Performing model selection (DIC, RJMCMC, MSPE, Bayes factor...)

By order of apparition:

Snow petrels

Sociable weavers

and others...

- Gimenez O., C. Crainiceanu, C. Barbraud, S. Jenouvrier et B.J.T. Morgan. Semiparametric regression in capture-recapture modelling. (under revision for Biometrics).
- Gimenez O., C.R. Brown and T. Lenormand. Nonparametric estimation of natural selection on a quantitative trait using capture-mark-recapture data. (subm. to Evolution).
- Cormack, Richard M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*, 51: 429–438
- Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, 52: 225–247
- Seber, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika*, 52: 249–259
- Lebreton, Jean-Dominique and Burnham, Kenneth P. and Clobert, Jean and Anderson, David R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs*, 62:67–118
- Barry, S. C. and Brooks, S. P. and Catchpole, E. A. and Morgan, B. J. T. (2003). The Analysis of Ring-Recovery Data Using Random Effects. *Biometrics*, 59: 54–65
- Johnson, D. S. and Hoeting, J. A. (2003). Autoregressive Models for Capture-Recapture Data: A Bayesian Approach. *Biometrics*, 59: 341–350
- Ruppert, D. and Wand, M. P. and Carroll, R. (2003). *Semiparametric Regression.*. University Press. Cambridge.
- Spiegelhalter, D. and Thomas, A. and Best, N. and Lunn, D. (2003). WinBUGS User Manual. Version 1.4 (http://www.mrc-bsu.cam.ac.uk/bugs.). Medical Research Council Biostatistics Unit. Cambridge.
- Brooks, S. P. and Catchpole, E. A. and Morgan, B. J. T. (2000). Bayesian animal survival estimation. *Statistical Science*, 15: 357–376.
- Crainiceanu, M., Ruppert, D. and Wand, M.P. (2004). Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software*.