NONPARAMETRIC ESTIMATION OF SURVIVAL IN THE WILD: APPLICATIONS IN ECOLOGY AND EVOLUTION

Olivier Gimenez

CEFE/CNRS Montpellier, France

olivier.gimenez@cefe.cnrs.fr

Centre d'Ecologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique

Centre d'Ecologie Fonctionnelle & Evolutive

CNRS Centre National de la Recherche Scientifique

Modelling mark-recapture data

- A particular encounter history: 1 0 1

Modelling mark-recapture data

- A particular encounter history: 1 0 1

$$1 \overbrace{\qquad\qquad}^{\phi_1} 0 \overbrace{\qquad\qquad}^{\phi_2} 1$$

- Survival probability ϕ_t

Modelling mark-recapture data

- A particular encounter history: 1 0 1

ϕ_1 ϕ_2

1 0 1

$1-p_2$ p_3

- Survival probability ϕ_t
- Detection probability p_t

Modelling mark-recapture data

- A particular encounter history: 1 0 1

$$\overbrace{1 \qquad 0}^{\phi_1}\!\!\overbrace{\qquad 1}^{\phi_2}$$

$$\begin{array}{ccc} 1 & 0 & 1 \\ & \uparrow & \uparrow \\ & 1-p_2 & p_3 \end{array}$$

$$\Pr(101) = \phi_1(1-p_2)\phi_2 p_3$$

- Survival probability ϕ_t
- Detection probability p_t

Likelihood of a standard mark-recapture model

Example

1111	10;	$\phi^3 p^3$
1110	11;	$\phi^2 p^2(1-\phi p)$
1011	2;	$\phi^3 p^2(1-p)$
1101	3;	……
1100	50;	……
1010	3;	
1001	1;	
1000	120;	
0111	10;	
0110	18;	
0101	4;	
0100	172;	
0011	24;	
0010	76;	

$$L = \prod_{\omega} \Pr(\omega)^{n_\omega} = \left(\phi^3 p^3\right)^{10} \times \left[\phi^2 p^2 (1-\phi p)\right]^{11} \times \ldots$$

$$\log(L) = 10 \cdot \log\left(\phi^3 p^3\right) + 11 \cdot \log\left(\phi^2 p^2 (1-\phi p)\right) + \ldots$$

(Lebreton et al. 1992)

Standard survival/covariate relationships

- Use the logistic link function

$$\text{logit}(\phi) = \beta_0 + \beta_1 x$$

  - +ve: survival estimates within [0,1]
  - -ve 1: variation completely determined by covariate
  - -ve 2: linear and quadratic relationships only

‘Less’ standard relationships…

- Use the logistic link function

$$\text{logit}(\phi) = \beta_0 + \beta_1 x + \varepsilon$$

  - -ve 1: relaxed thanks to random effect ε
    - Cope with unexplained variance, e.g. overdispersion (*Barry et al. 2002 – Biometrics; Royle 2008 - Biometrics*)
    - Allow temporal autocorrelation to be incorporated (*Johnson & Hoeting 2003 - Biometrics*)
  - -ve 2: nonparametric modelling via P-splines

$$\text{logit}(\phi) = m(x) + \varepsilon$$

Flexible covariate modelling: P-splines

- Environmental covariates to assess the *impact of climatic change* on demographic parameter.
- Individual covariates to *investigate natural selection.*
- Bayesian modelling using MCMC simulations: *Consider random effects for automatic smoothing.*

Nonparametric modelling of survival

- Environmental covariates: climatic change

Single covariate

Snow petrel (*Pagodroma nivea*)

Nonlinear interactions via bivariate smoothing

Emperor penguin (*Aptenodytes forsteri*)

Nonparametric modelling of survival

- Individual covariates: natural selection

Single trait

Sociable weavers (*Philetairus socius*)

Fitness surface via bivariate smoothing

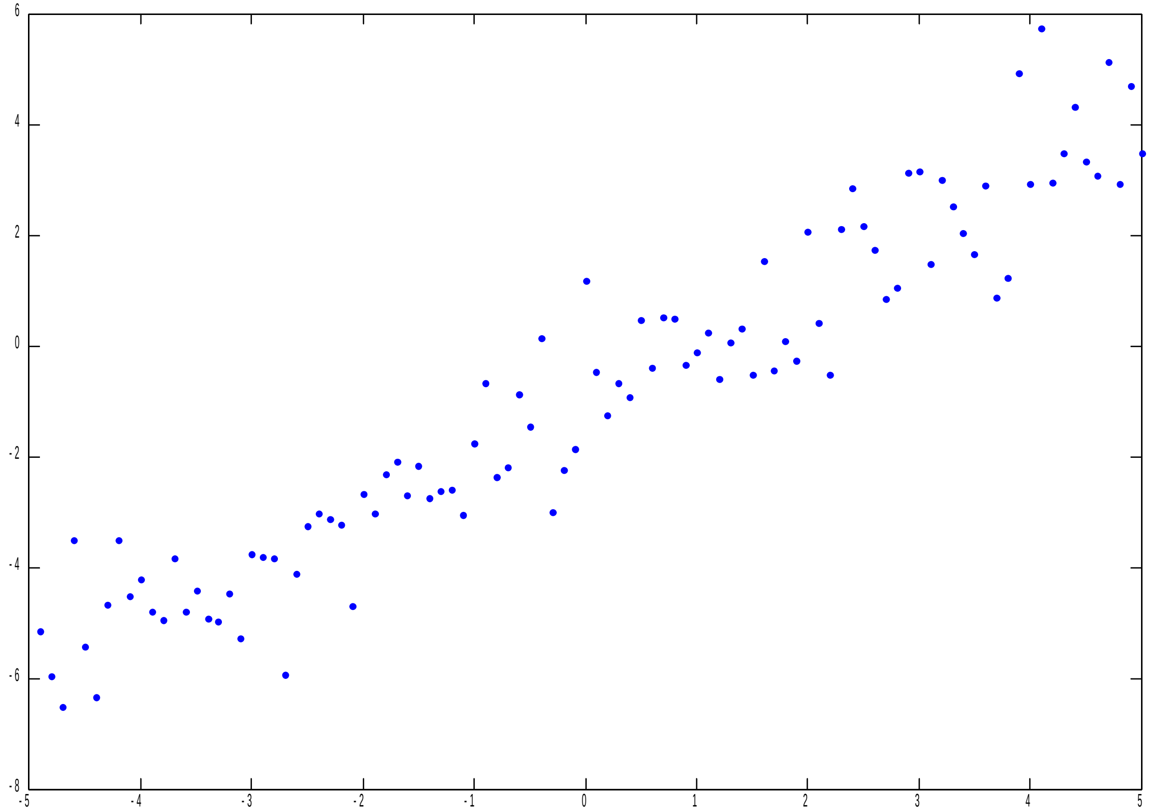
European blackbirds (*Turdus merula*)

Centre d'Ecologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique

1. Introduction to penalized-splines

linearities



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

Centre d'Ecologie Fonctionnelle & Evolutive

CNRS Centre National de la Recherche Scientifique

1. Ordinary least squares

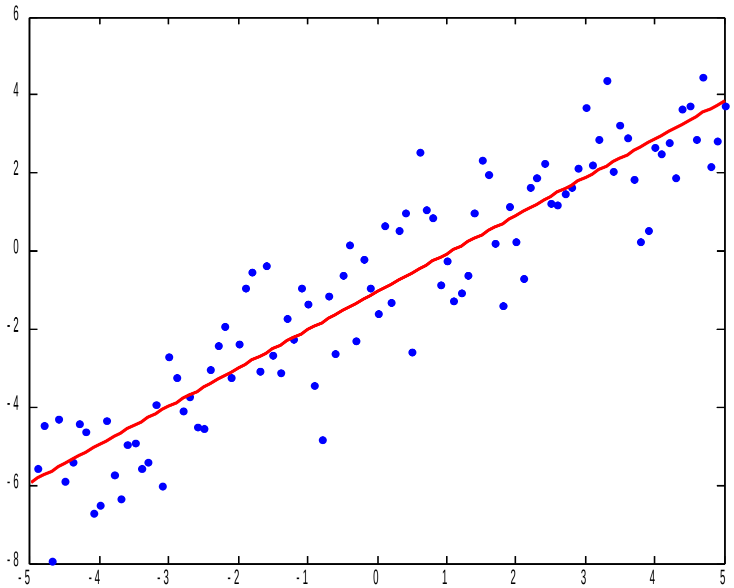
Let us denote $\eta = (\beta_0, \beta_1)^T$;

we search for $\hat{\eta}$ that minimizes $||y - X\eta||^2$

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

1. Introduction to penalized-splines

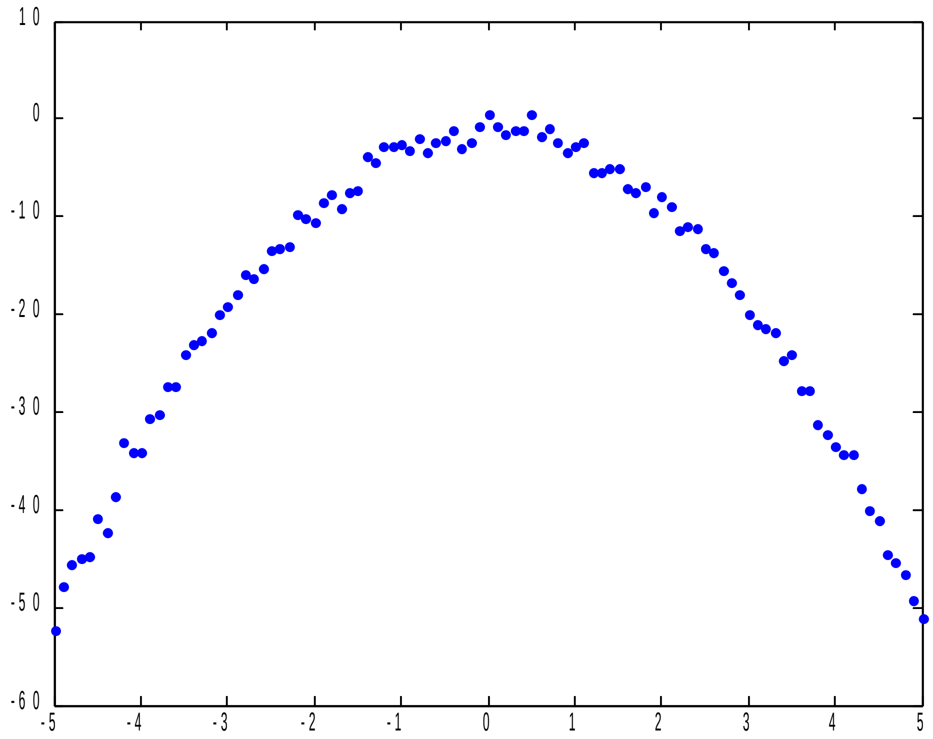
linearities



$$X_i = (1, x_i)$$
$$\eta = (\beta_0, \beta_1)^T$$
$$\widehat{\eta} = (-0.96, 0.98)^T$$

1. Introduction to penalized-splines

'quadraticities'



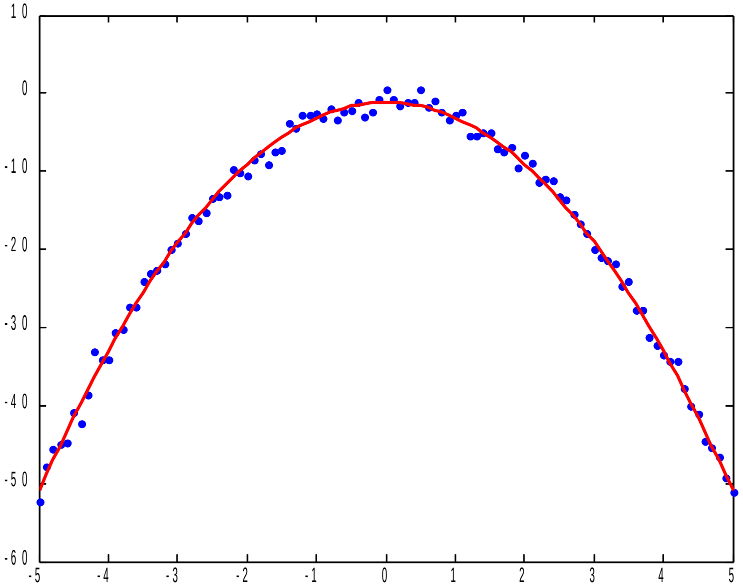
$$y_i = \beta_0 + \beta_1 \times x_i + \beta_2 \times x_i^2 + \varepsilon_i ?$$

Centre d'Ecologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique

1. Introduction to penalized-splines

'quadraticities'



$$X_i = (1, x_i, x_i^2)$$

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

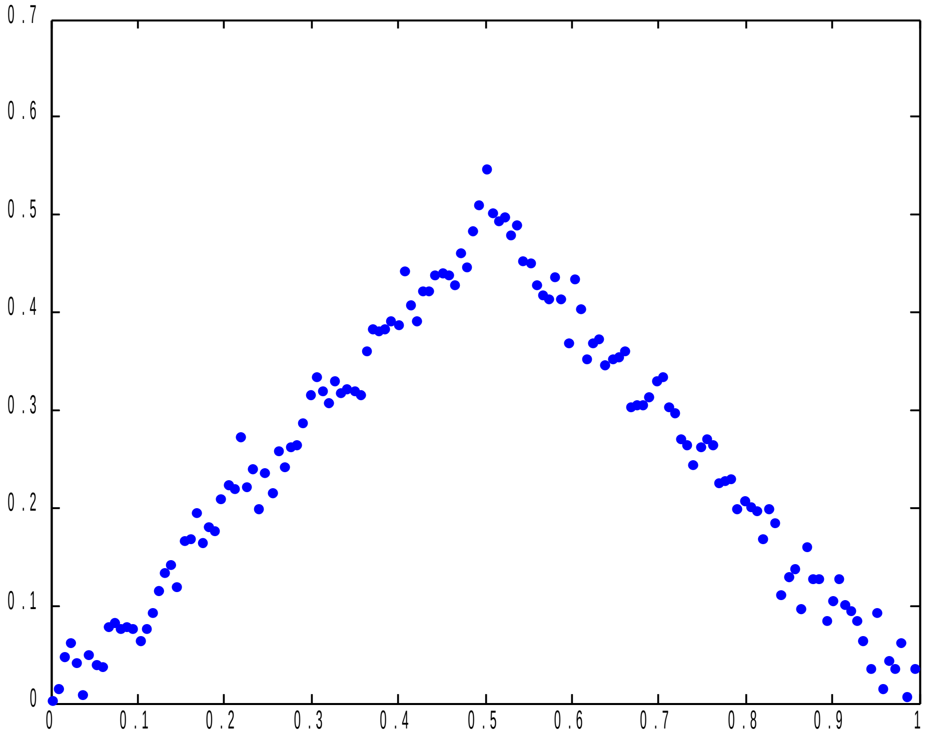
$$\widehat{\eta} = (-1.12, 0.05, -2.00)^T$$

Centre d'Ecologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique

1. Introduction to penalized-splines

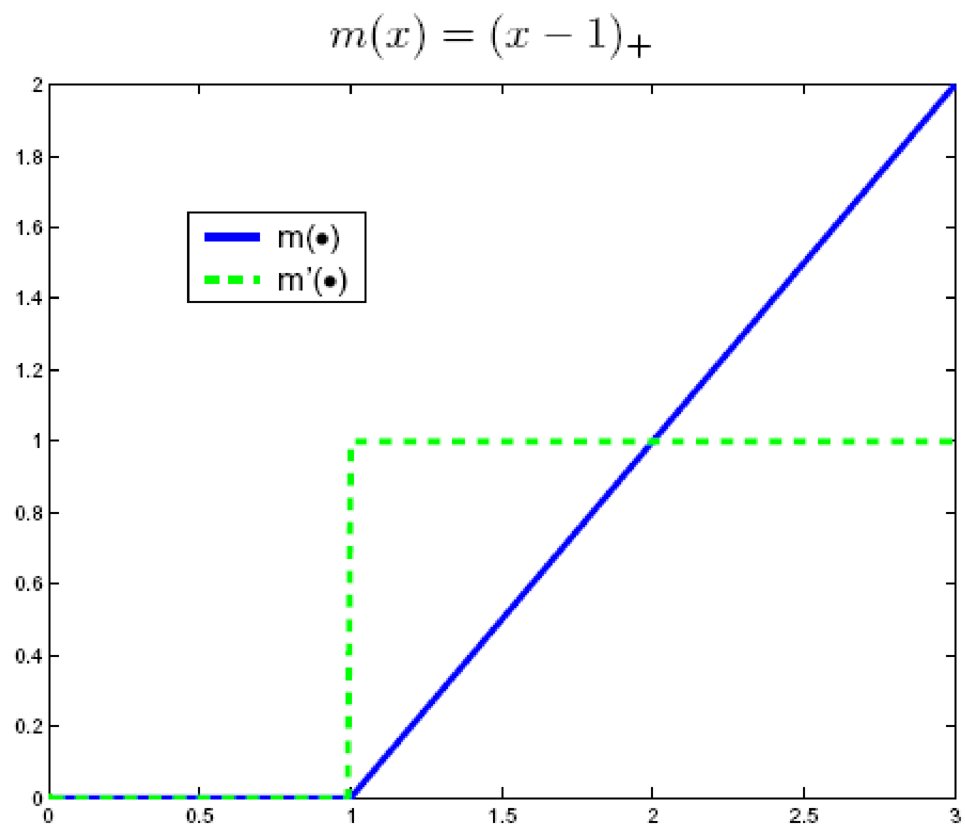
broken line



?????

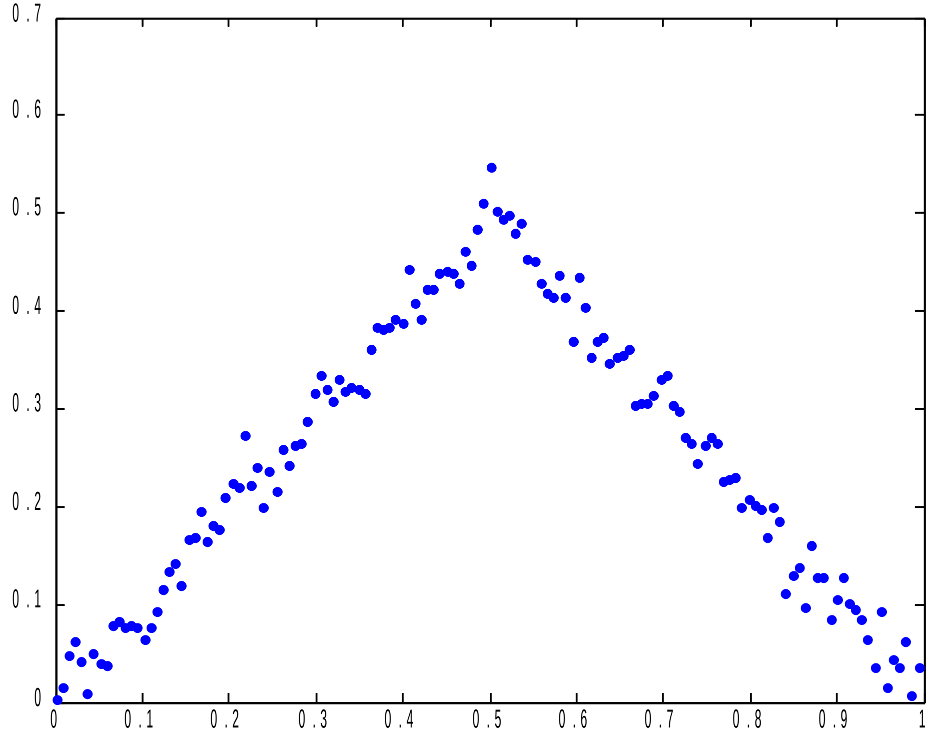
1. Introduction to penalized-splines

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



1. Introduction to penalized-splines

broken line



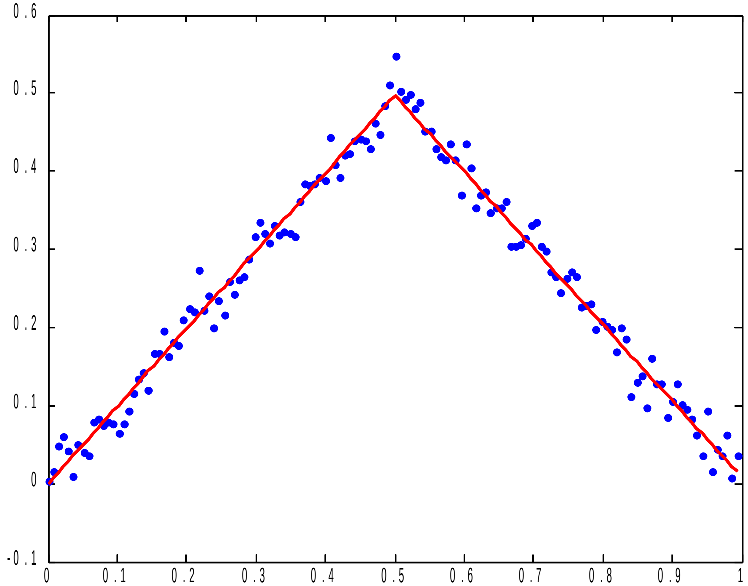
$$y_i = \beta_0 + \beta_1 \times x_i + b_1 \times (x_i - 0.5)_+ + \varepsilon_i?$$

Centre d'Ecologie Fonctionnelle & Evolutive

CNRS Centre National de la Recherche Scientifique

1. Introduction to penalized-splines

broken line



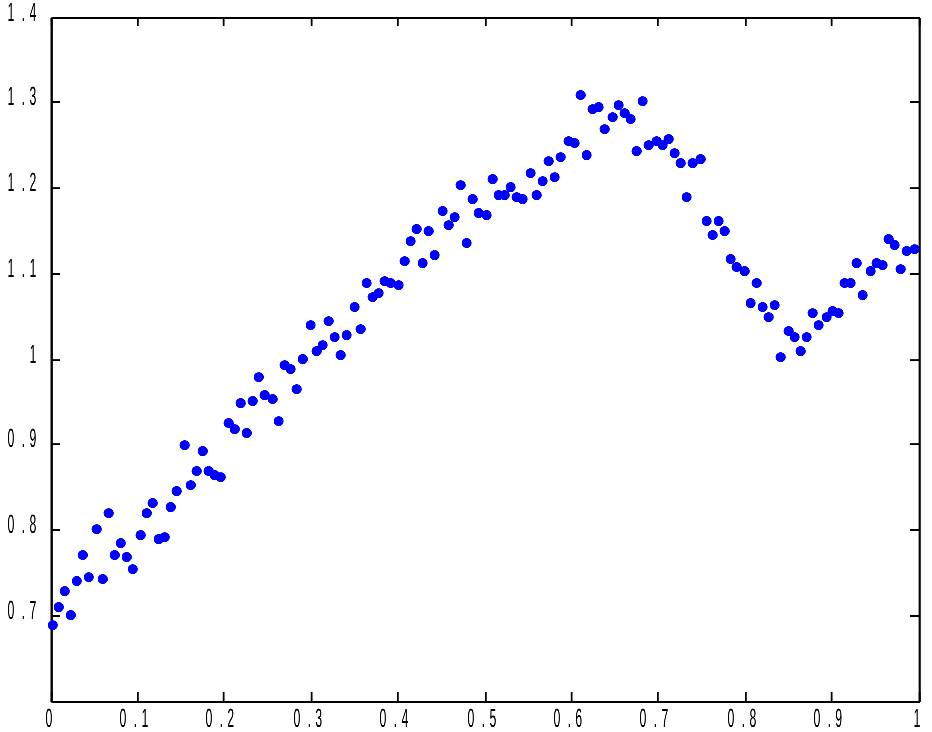
$$X_i = (1, x_i, (x_i - 0.5)_+)$$

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

$$\widehat{\eta} = (0.00, 1.01, -2.04)^T$$

1. Introduction to penalized-splines

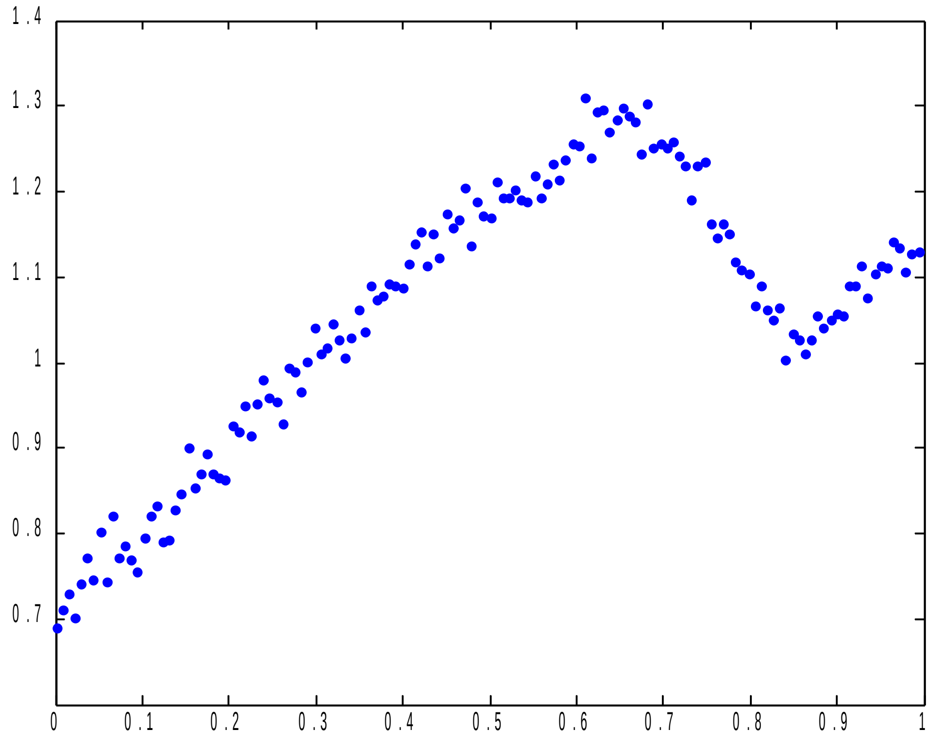
nonlinearities



?????

1. Introduction to penalized-splines

nonlinearities

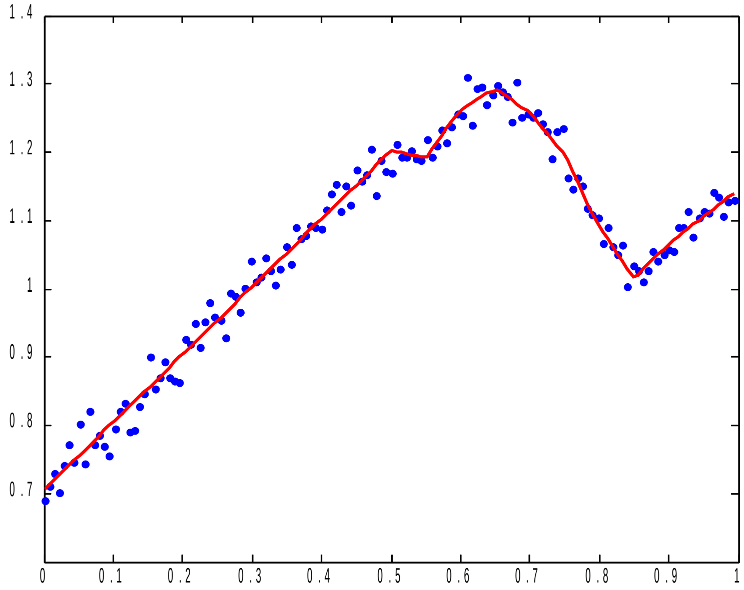


$$y_i = \beta_0 + \beta_1 \times x_i + \sum_{k=1}^{K} b_k \times (x_i - \kappa_k)_+ + \varepsilon_i ?$$

1. Introduction to penalized-splines

nonlinearities

nonparametric fitting using splines



$$X_i = (1, x_i, (x_i - 0.5)_+, \ldots, (x_i - 0.9)_+)$$

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

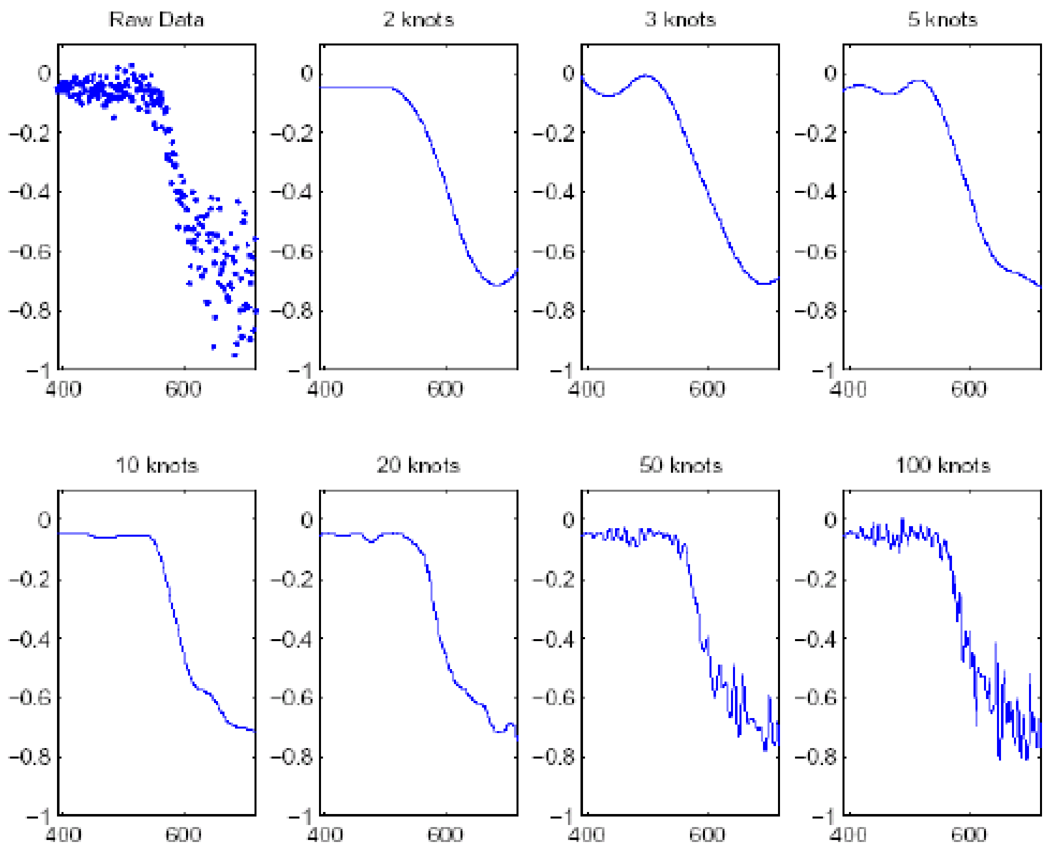
$$\widehat{\eta} = (0.69, 1.04, -1.64, 1.70, \ldots, -0.31)^T$$

Centre d'Écologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique

1. Influence of number/location of knots

non-parametric approach via OLS



Ruppert et al. (2003)

1. Determining number & location of knots

- See book by Ruppert et al. (2003)
- Number of knots K = min(0.25n,35)
- Location of knots according to the data: « equally-spaced sample quantiles » (k-*th* knot at sample quantile of the covariate corresp. to prob k/K+1)
- Idea: penalize the b_k's in order to constraint the influence of the knots -> b_k's are random effects
- The P-spline estimator is actually equal to the BLUP of a mixed model
- Extensive simulation studies show that the procedure works pretty well in a broad context (Ruppert 2002), and in particular…

Centre d'Ecologie Fonctionnelle & Evolutive

CNRS Centre National de la Recherche Scientifique

1. Simulation study

- 2 scenarios
  - Threshold

$$f(x) = \begin{cases} 2.2 & \text{if } x \leq -0.06 \\ 2.08 - 2x & \text{otherwise} \end{cases}$$

  - Nonlinear

$$f(x) = 1.5g\left(\frac{x - 0.35}{0.15}\right) - g\left(\frac{x - 0.6}{0.1}\right)$$

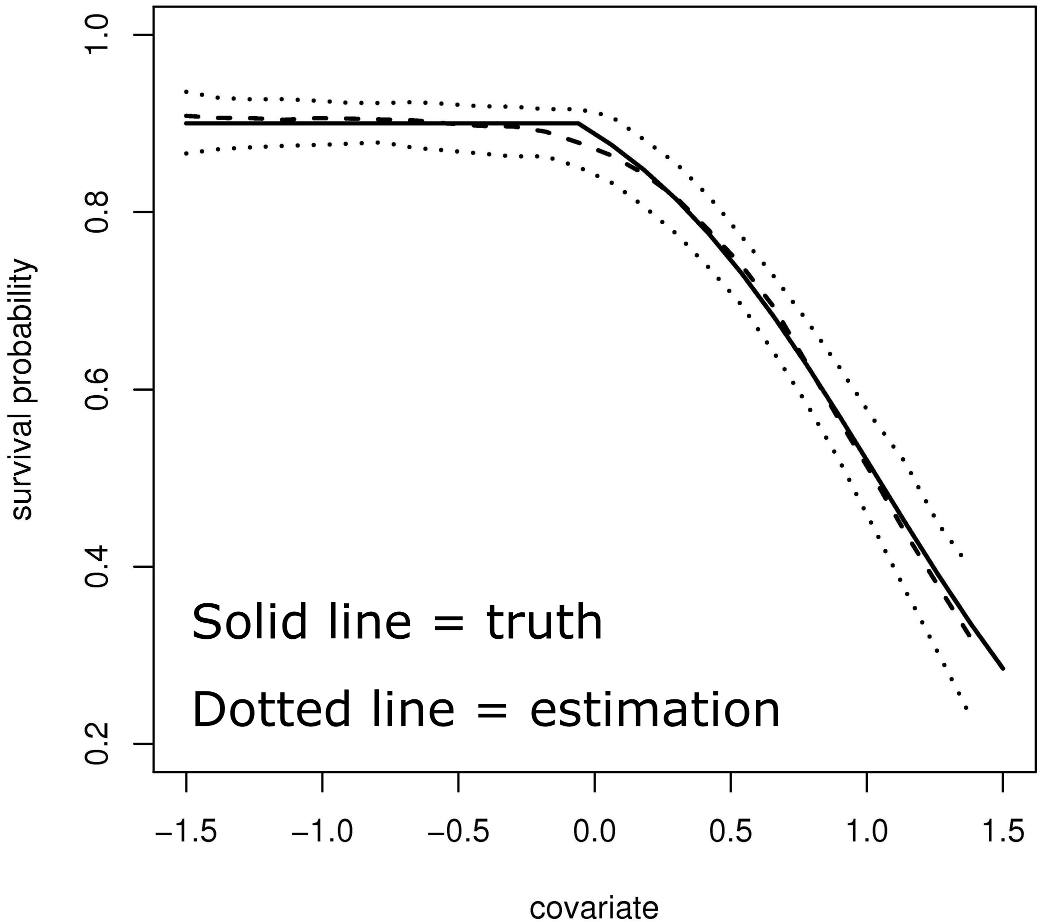
$$g(x) = 1/\sqrt{2\pi} \cdot \exp\left(-x^2/2\right)$$

- $\sigma^2 = 0.02$, $p = 0.7$
- 50 simulations

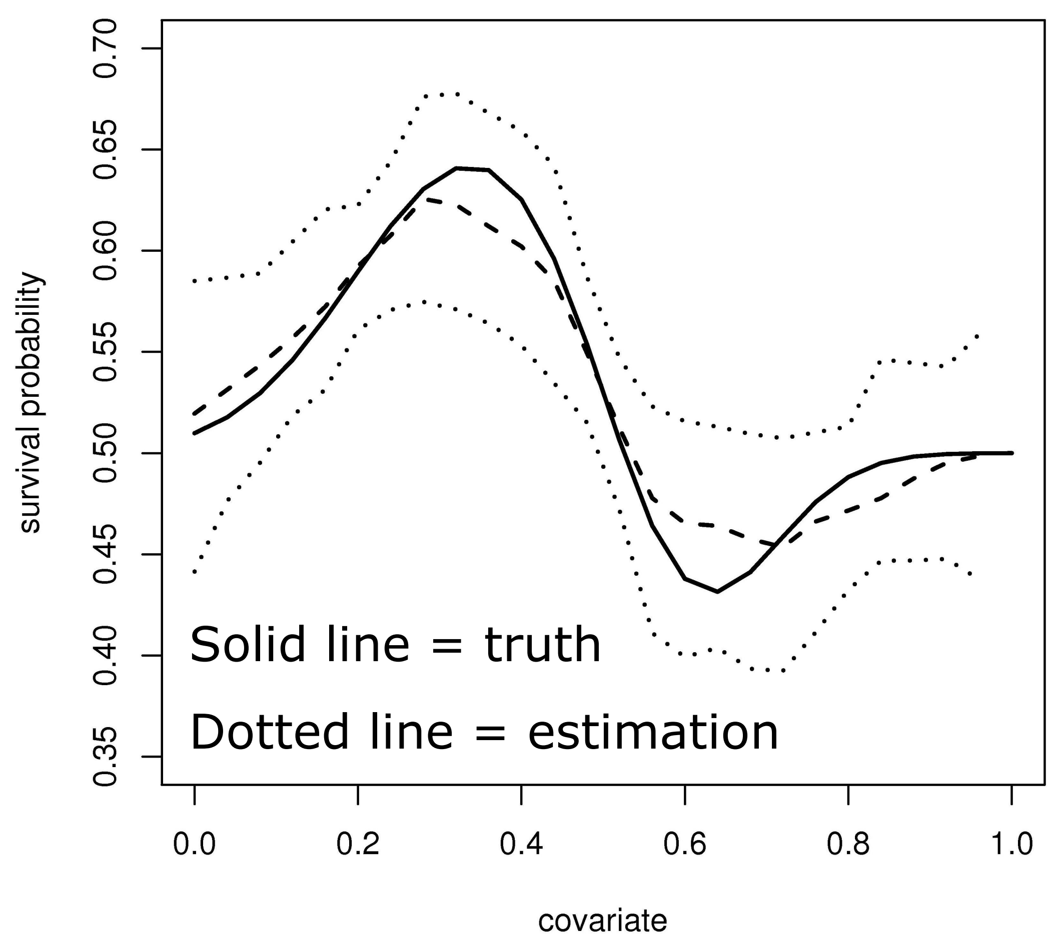
Centre d'Ecologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique

1. Threshold scenario



1. Nonlinear scenario



2. P-splines & environmental covariates

- Environmental covariates: climatic change
  - Single covariate:

Gimenez et al. 2006 – Biometrics

Snow petrel (*Pagodroma nivea*)

Joint work with C. Barbraud, S. Jenouvrier, C. Crainiceanu, B.J.T. Morgan

2. P-splines & environmental covariates

- First, incorporate the effect of Southern Oscillation Index (SOI)

$$
\begin{aligned}
\mathrm{logit}(\phi_i^l) &= \beta_0 + \beta_1 \mathrm{SOI}_i + \\
&\sum_{k=1}^{7} b_k \left(\mathrm{SOI}_i - \kappa_k\right)_+ + \varepsilon_i
\end{aligned}
$$

2. P-splines & environmental covariates

- We assume

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

$$b_k \sim N(0, \sigma_b^2)$$

2. P-splines & environmental covariates

- We assume

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

$$b_k \sim N(0, \sigma_b^2)$$

- Mixed model representation

$$\text{logit}(\phi) = X\boldsymbol{\beta} + Z\boldsymbol{b} + \varepsilon$$

2. P-splines & environmental covariates

- Second, incorporate the effect of sex

$$\begin{aligned}\text{logit}(\phi_i^l) &= \beta_0 + \gamma\text{SEX} + \beta_1\text{SOI}_i + \\ &\sum_{k=1}^{7} b_k\,(\text{SOI}_i - \kappa_k)_+ + \varepsilon_i\end{aligned}$$

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

- Semiparametric model

2. P-splines & environmental covariates

- Mixed model representation: fixed effects

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

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

2. P-splines & environmental covariates

- Mixed model representation: random effects

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

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

2. P-splines & environmental covariates

- Prior distributions for all parameters:

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

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

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

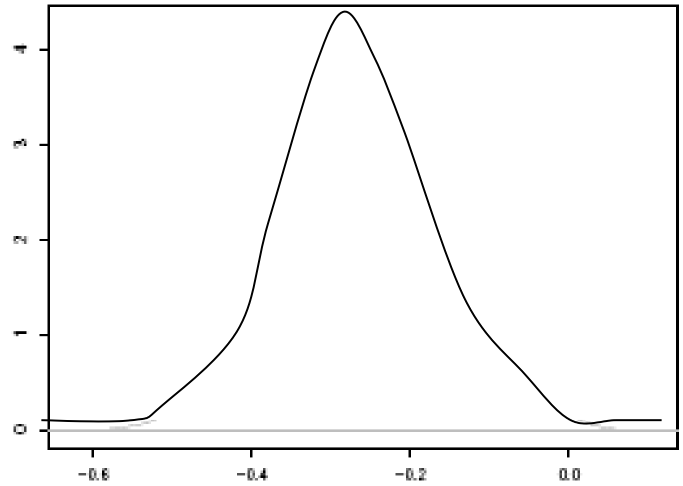
$$[\varepsilon_i] = N(0,\sigma_\varepsilon^2)$$

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

2. Results: sex effect?

- Males survive better than females

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



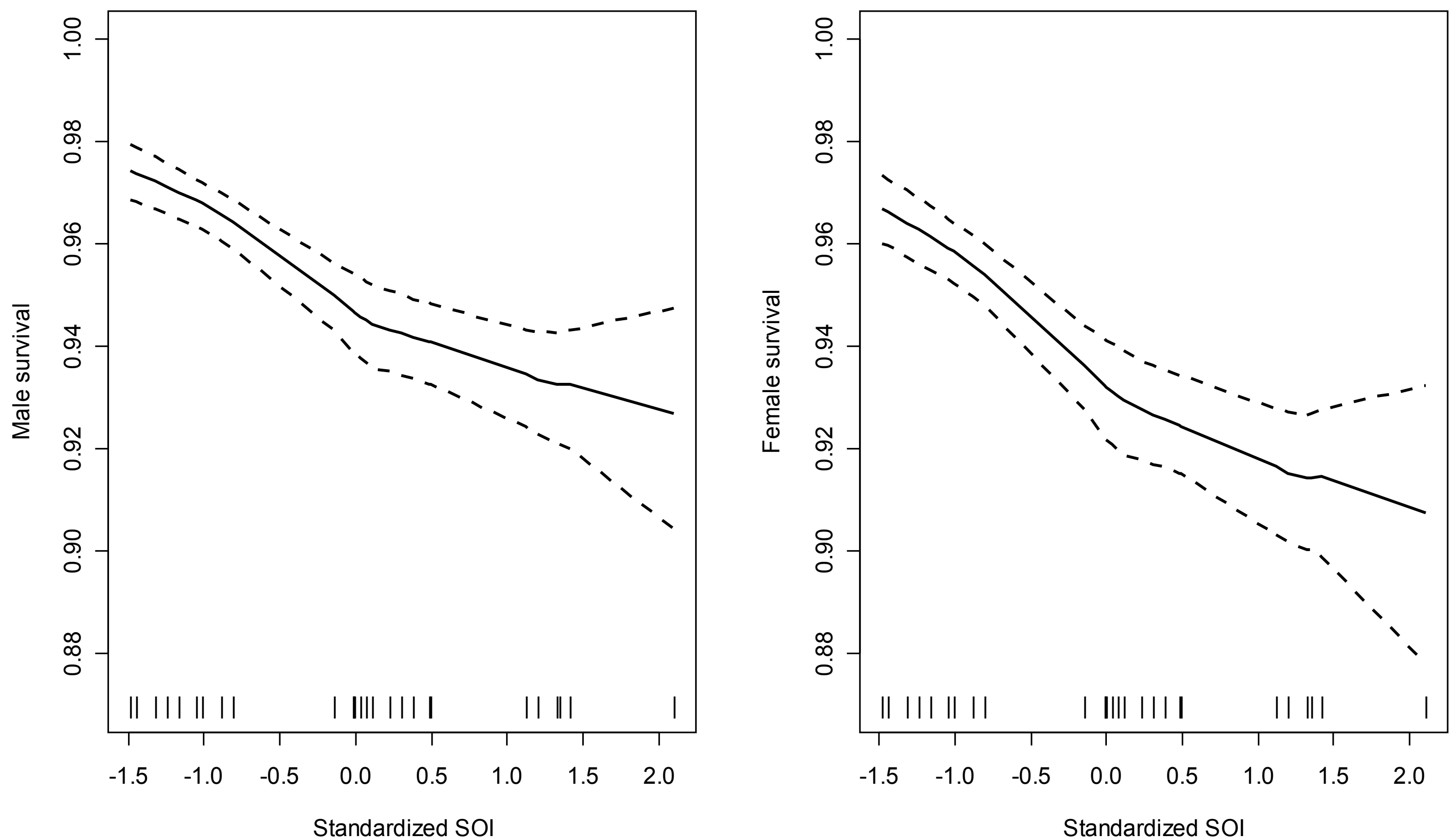
Posterior distribution of γ

Centre d'Ecologie Fonctionnelle & Evolutive

CNRS Centre National de la Recherche Scientifique

2. Results

- There is a negative effect of SOI on survival

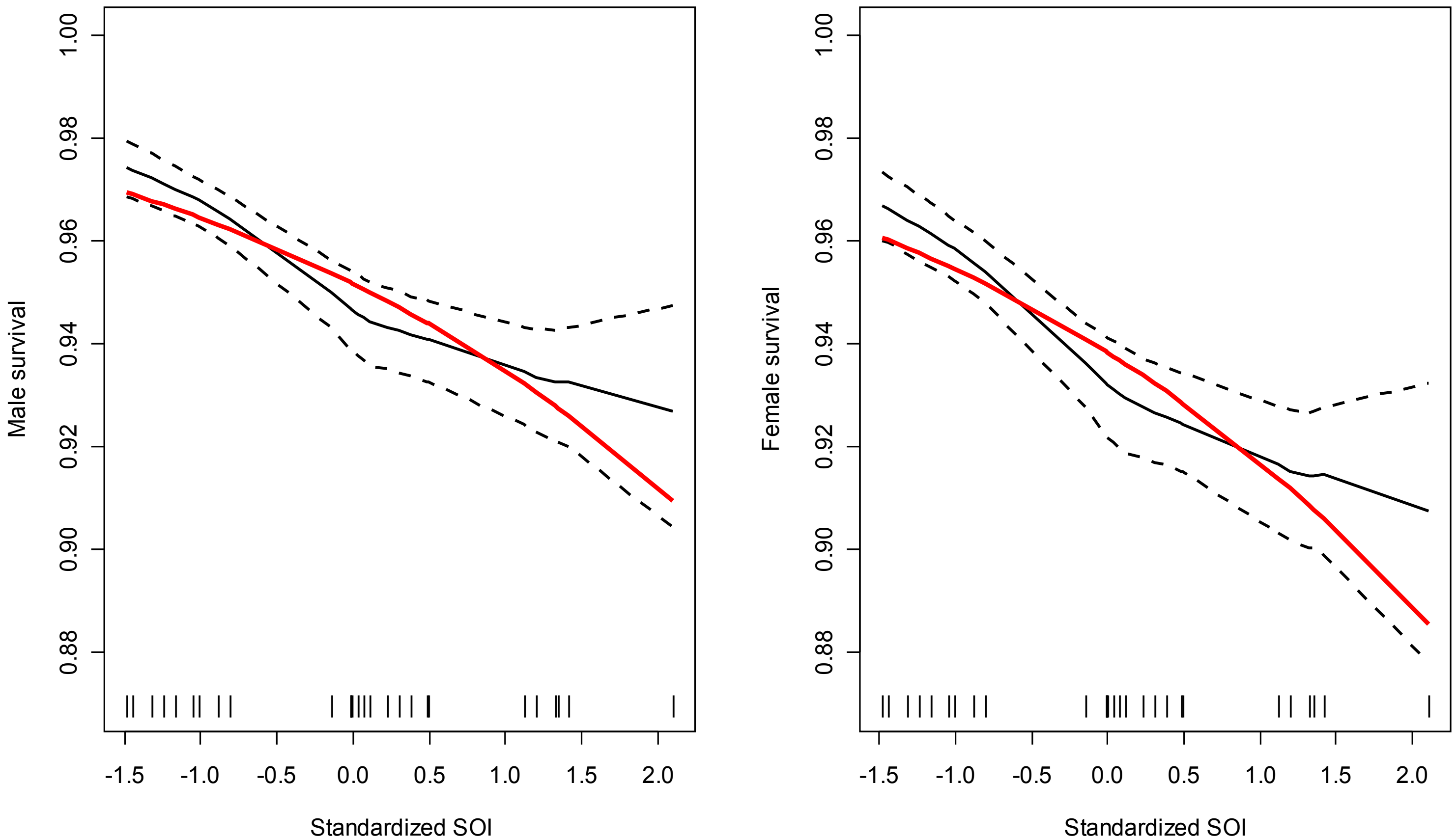


Centre d'Ecologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique

2. Results

- Is this effect nonlinear?



2. P-splines & environmental covariates

- Environmental covariates: climatic change
  - Bivariate smoothing

Gimenez & Barbraud 2008 – Env. and Ecol. Stat.

Emperor penguin (*Aptenodytes forsteri*)

Joint work with C. Barbraud

CENTRE D'ECOLOGIE FONCTIONNELLE & EVOLUTIVE

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

2. P-splines & environmental covariates

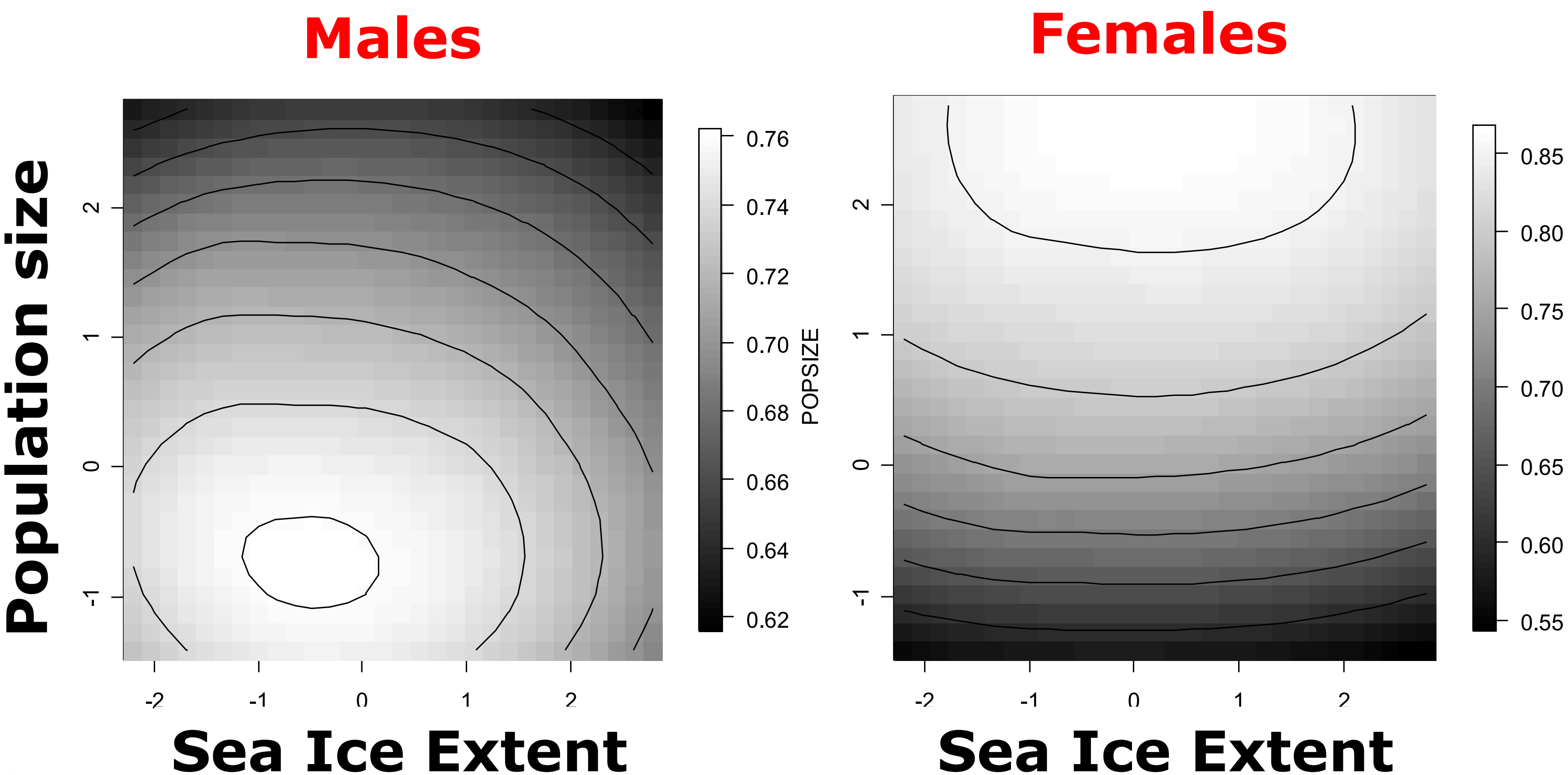
- Nonparametric modelling of interactions between 2 continuous covariates.
- Use of thin-plate splines – details omitted.
- Example of the emperor penguin survival as a function of sea-ice extent (SIE) and number of breeding pairs (POPSIZE).

Centre d'Ecologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique

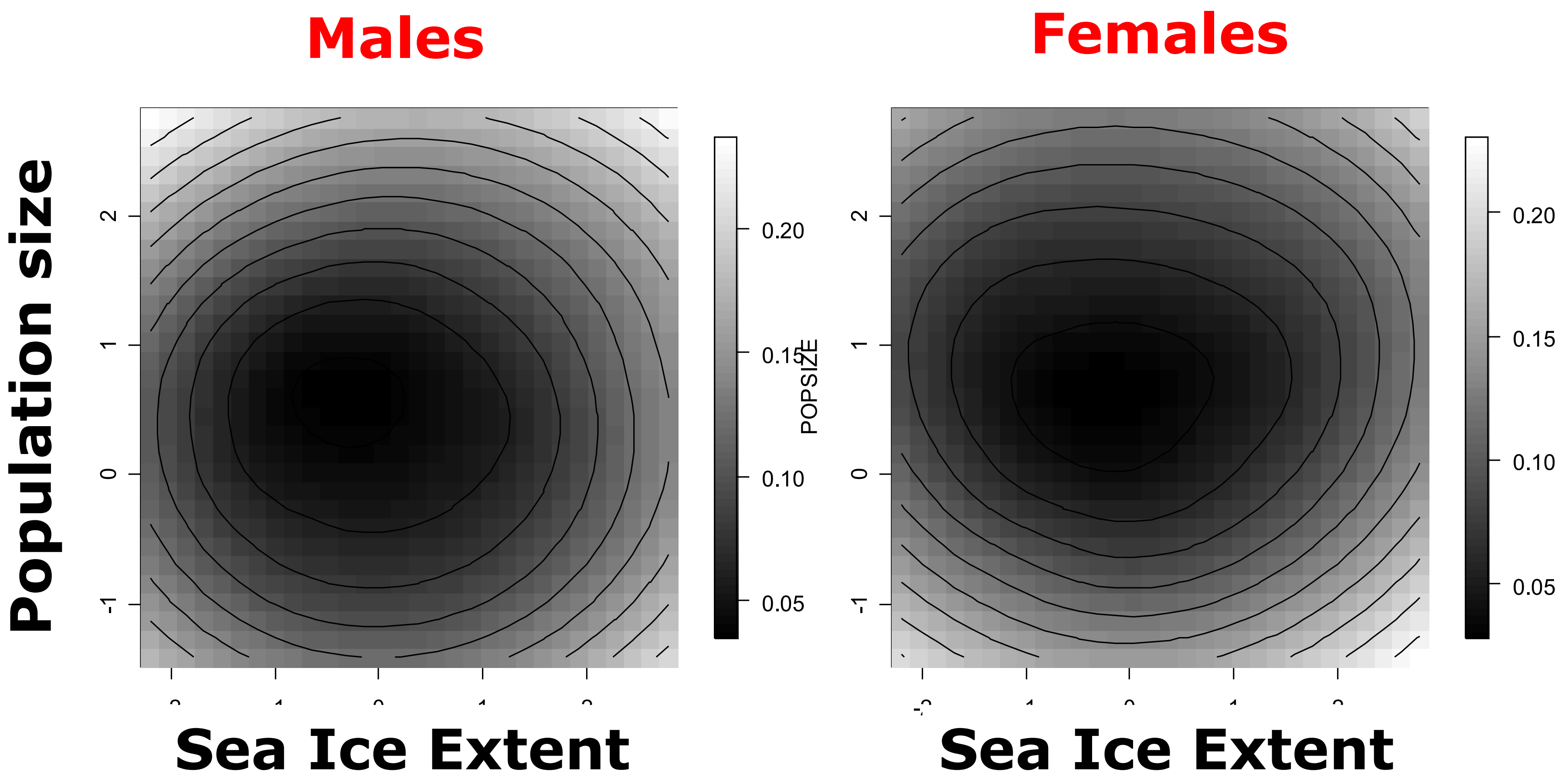
2. P-splines & environmental covariates

- (Posterior mean) survival vs. SIE & POPSIZE



2. P-splines & environmental covariates

- Precision associated with survival surfaces



3. Nonparametric modelling of survival

- Individual covariates: natural selection
  - Single trait

Gimenez et al. 2006 – Evolution

Sociable weavers (*Philetairus socius*)

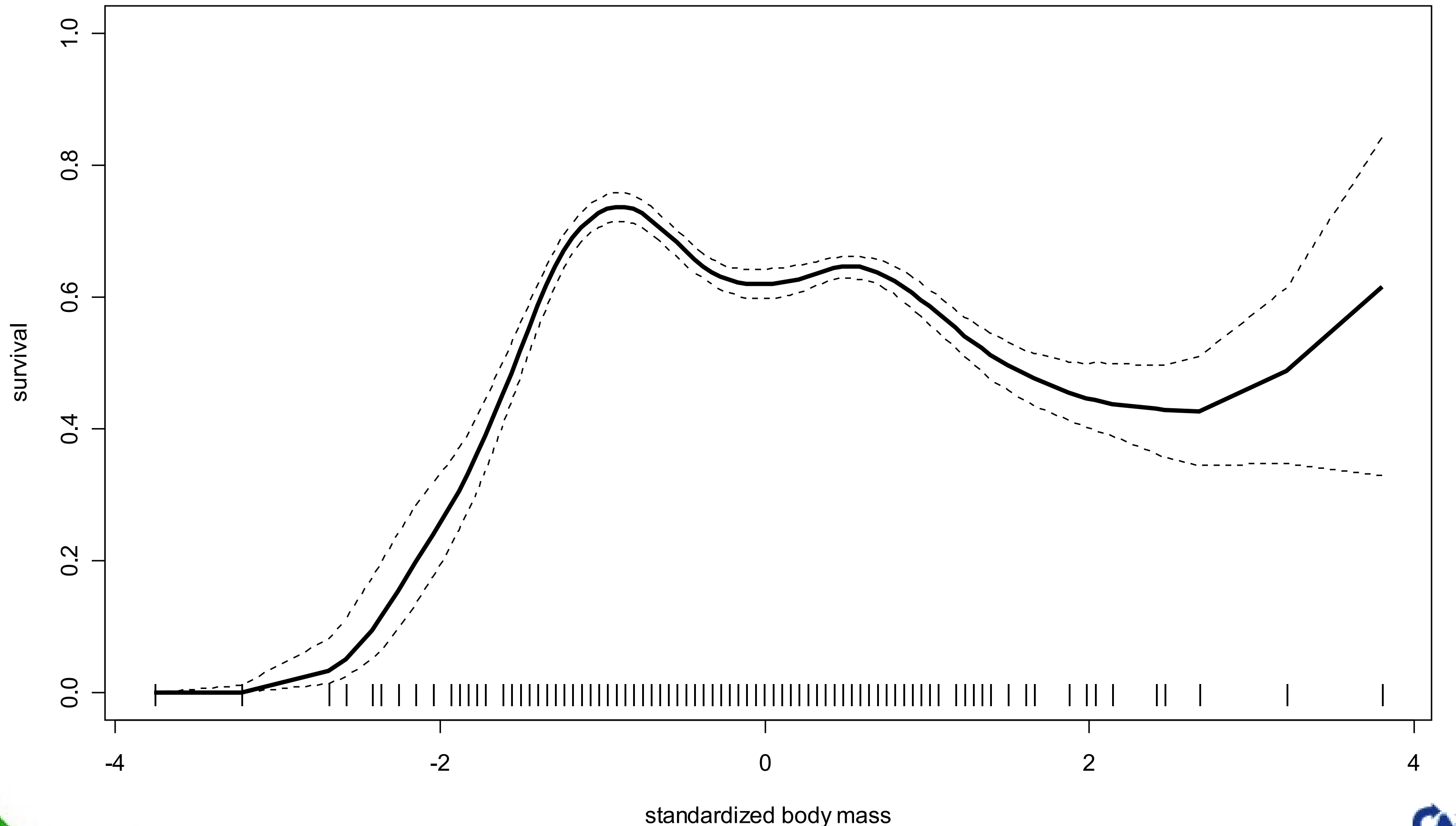
Joint work with T. Lenormand, C. Brown et al.

Centre d'Ecologie Fonctionnelle & Evolutive

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

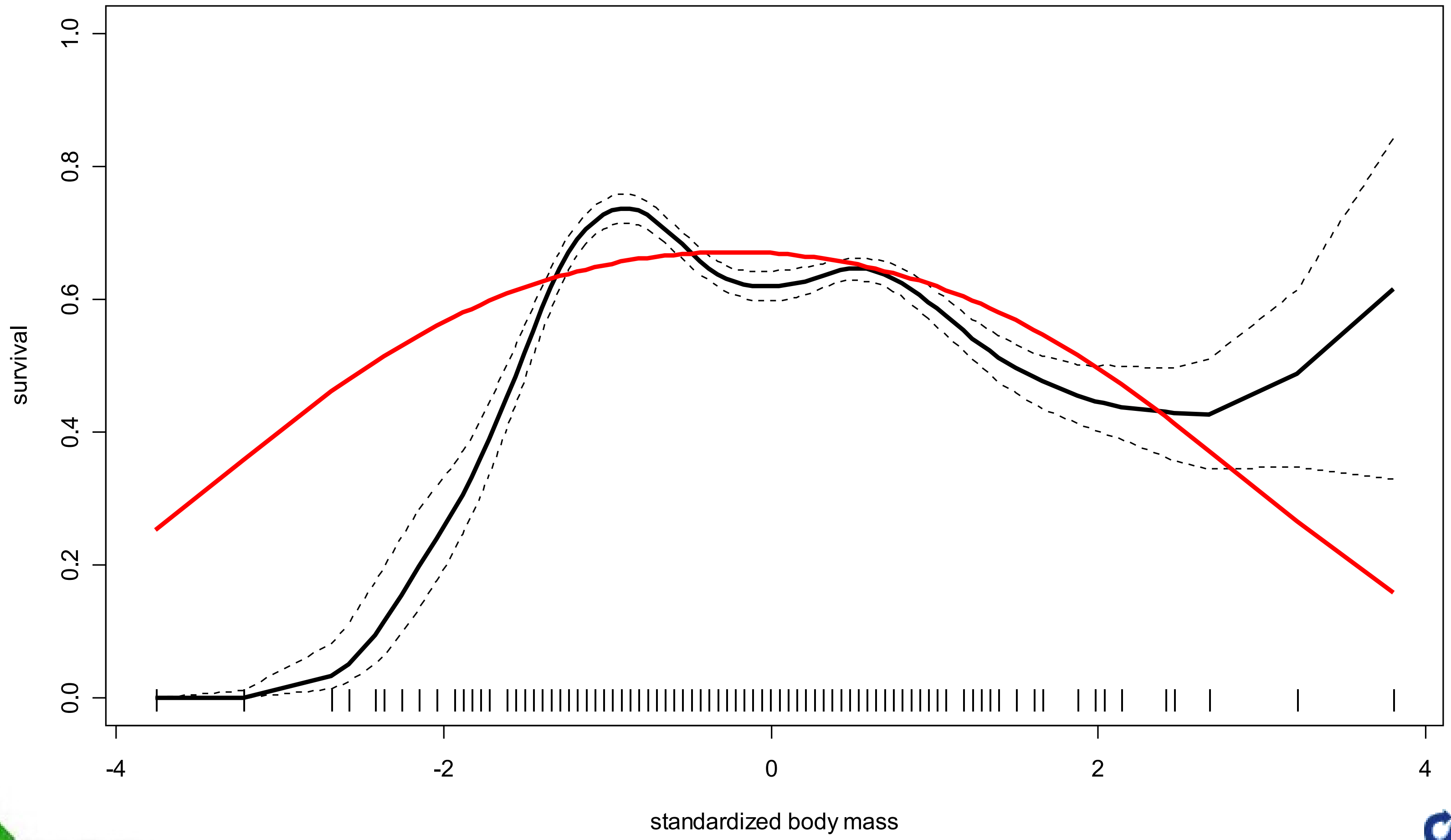
3. P-splines & individual covariates

- Survival vs. body mass via P-splines



3. P-splines & individual covariates

- Survival vs. body mass via quadratic relationship

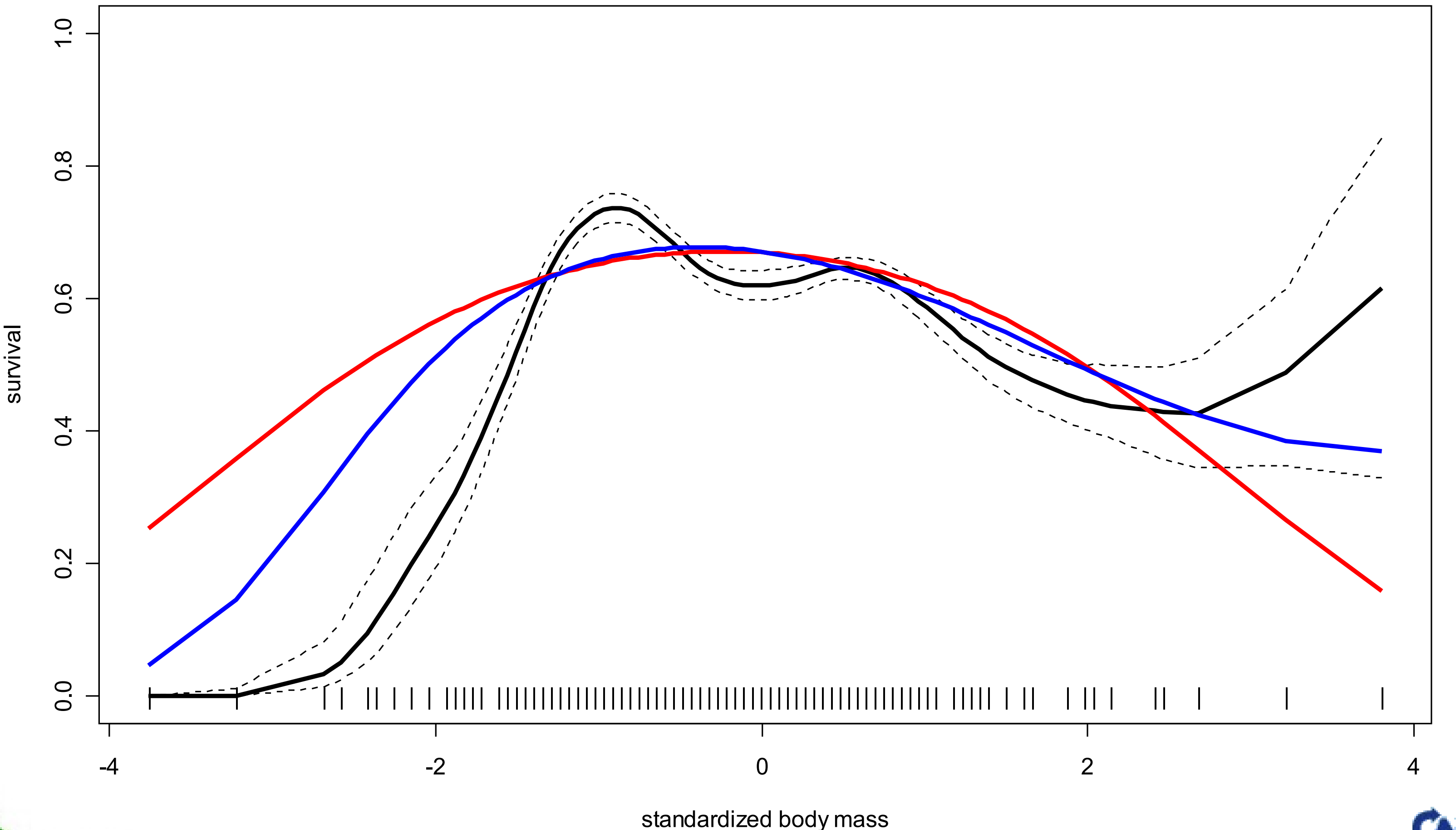


CENTRE D'ECOLOGIE FONCTIONNELLE & EVOLUTIVE

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

3. P-splines & individual covariates

- Survival vs. body mass via cubic relationship



3. Nonparametric modelling of survival

- Individual covariates: natural selection
  - Fitness surface

    Gimenez et al. (subm.)

    European blackbirds (*Turdus merula*)
  - 5 morphological traits were considered (Tarsus length, phalanx length, beak height , wing length and rectrice length) – PCA was used to cope with multicollinearity

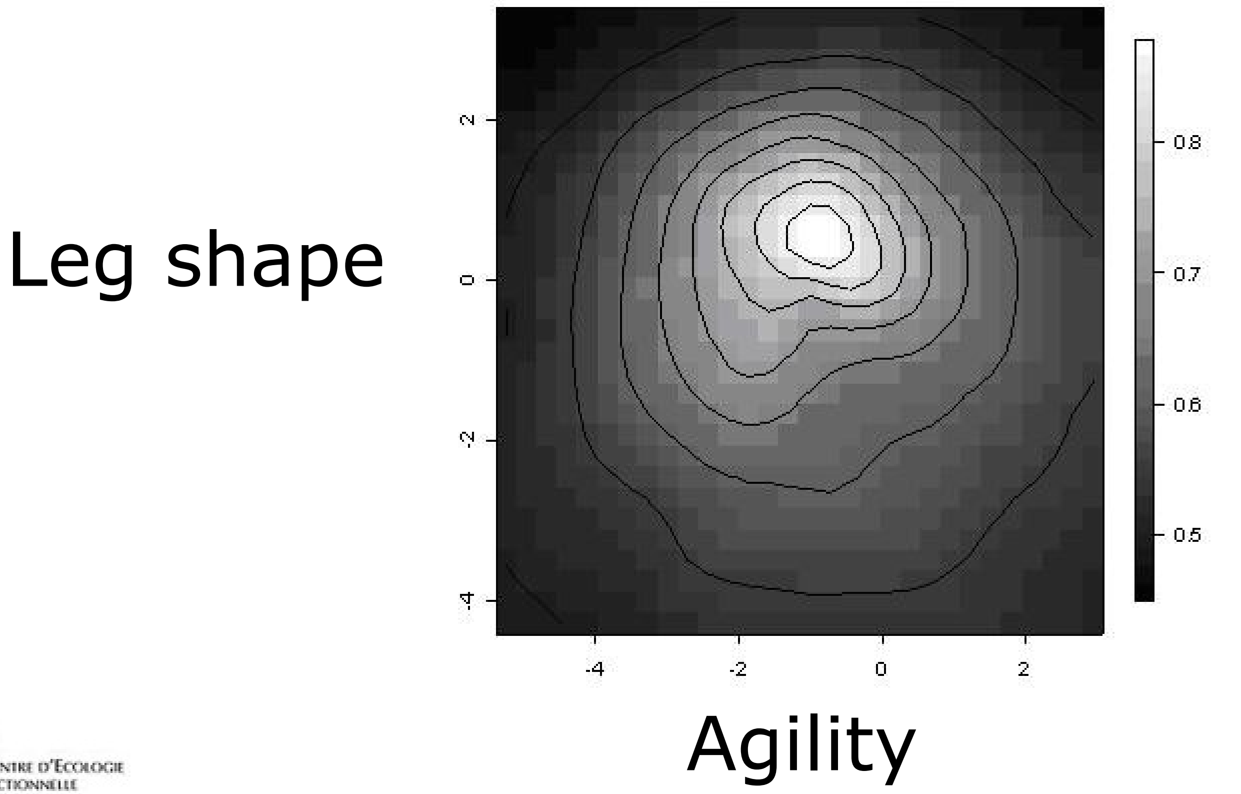
Joint work with A. Grégoire and T. Lenormand

Centre d'Ecologie Fonctionnelle & Evolutive

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

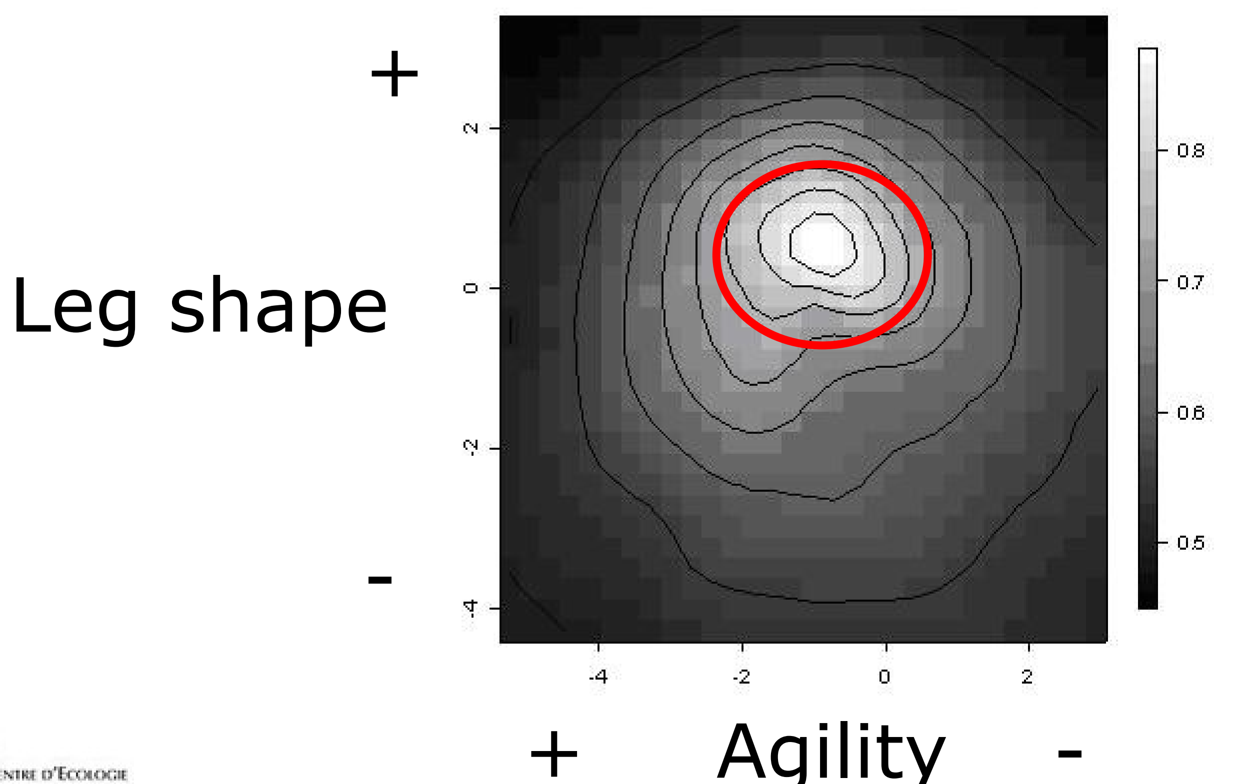
3. P-splines & individual covariates

- Visualization of the survival surface for the European blackbird as a function of two important principal components:



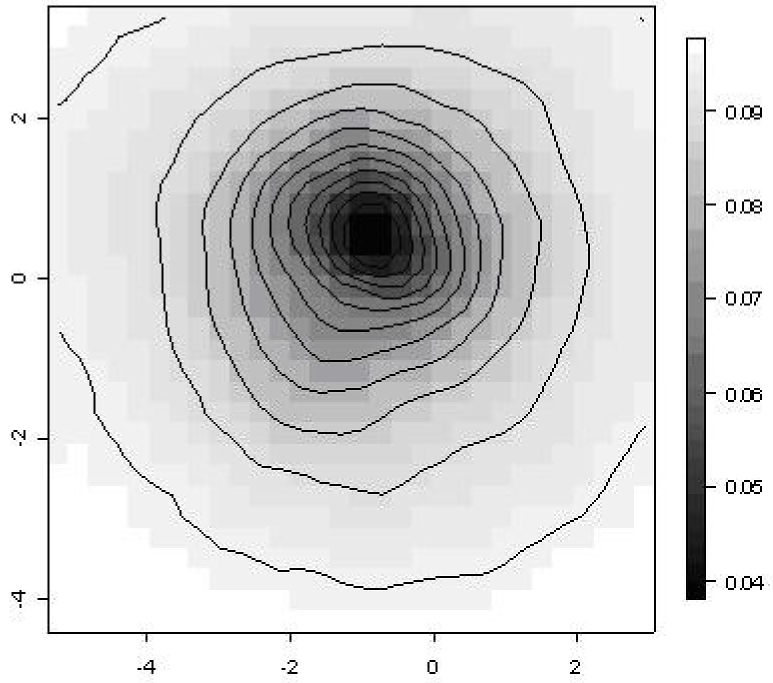
3. P-splines & individual covariates

- Visualization of the survival surface for the European blackbird:



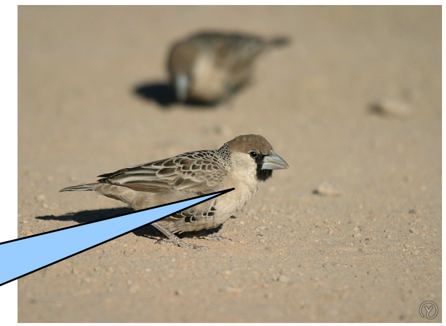
3. P-splines & individual covariates

- Precision associated to the survival surface



Conclusions

- Non- and semi-parametric modelling
  - Very flexible: does not assume a prior relationship
  - GLMM formulation allows automatic calculation of the optimal amount of smoothing
  - Allows tackling biological questions of fundamental importance, while accounting for detectability < 1
- Limits
  - Computational burden: numerical integration vs. MCMC
  - Formally test nonlinearities?



Thank you for your attention

Centre d'Ecologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique

Workshop at Montpellier (France), 17-21 November 2008 : MODELLING INDIVIDUAL HISTORIES WITH STATE UNCERTAINTY

- Multievent (hidden-Markov) mark-recapture models
- E-SURGE software
- Instructors: Pradel, Lebreton, Gimenez, Rouan, Choquet
- http://www.cefe.cnrs.fr/biom/Workshops/

Centre d'Ecologie Fonctionnelle & Evolutive

Centre National de la Recherche Scientifique