MARKOV SWITCHING MODELS FOR HIGH-FREQUENCY TIME SERIES FROM AUTOMATIC MONITORING OF ANIMALS

Luigi Spezia

Biomathematics & Statistics Scotland

Aberdeen, UK

Cecilia Pinto

School of Biological Sciences

University of Aberdeen

Aberdeen, UK

Overview

- Flapper skate's depth profile
- High-frequency time series

  Long memory process

  Non-linear time series

Dipturus intermedia

- Markov switching autoregressive models
- Results
- Current research

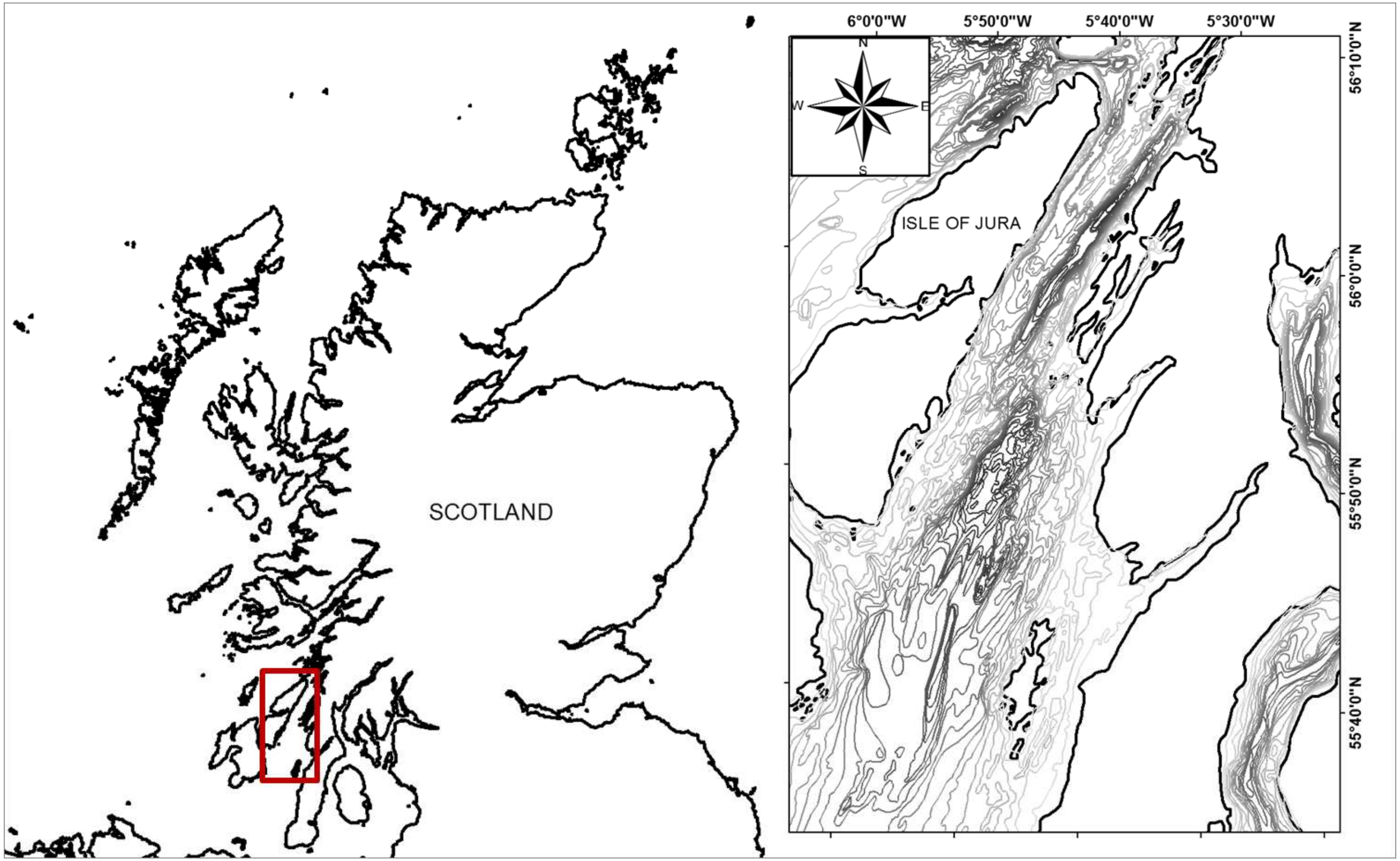
Flapper skate's depth profile

Cecilia Pinto's PhD project:

Estimating the probability of recolonization of endangered marine species integrating demographic and movement parameters.

School of Biological Sciences, University of Aberdeen

Flapper skate's depth profile



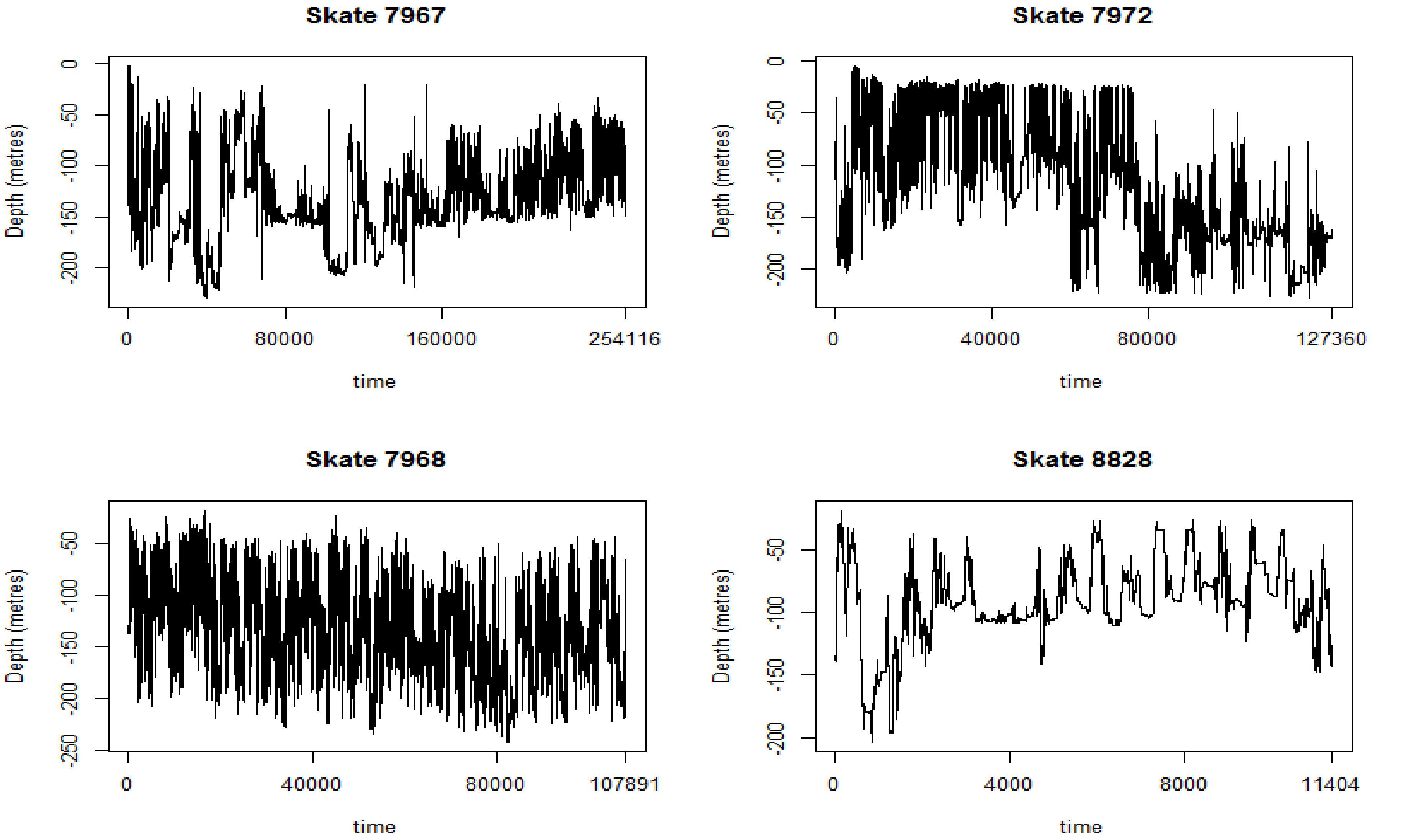
Flapper skates' depth profiles

Flapper skates' depth profiles

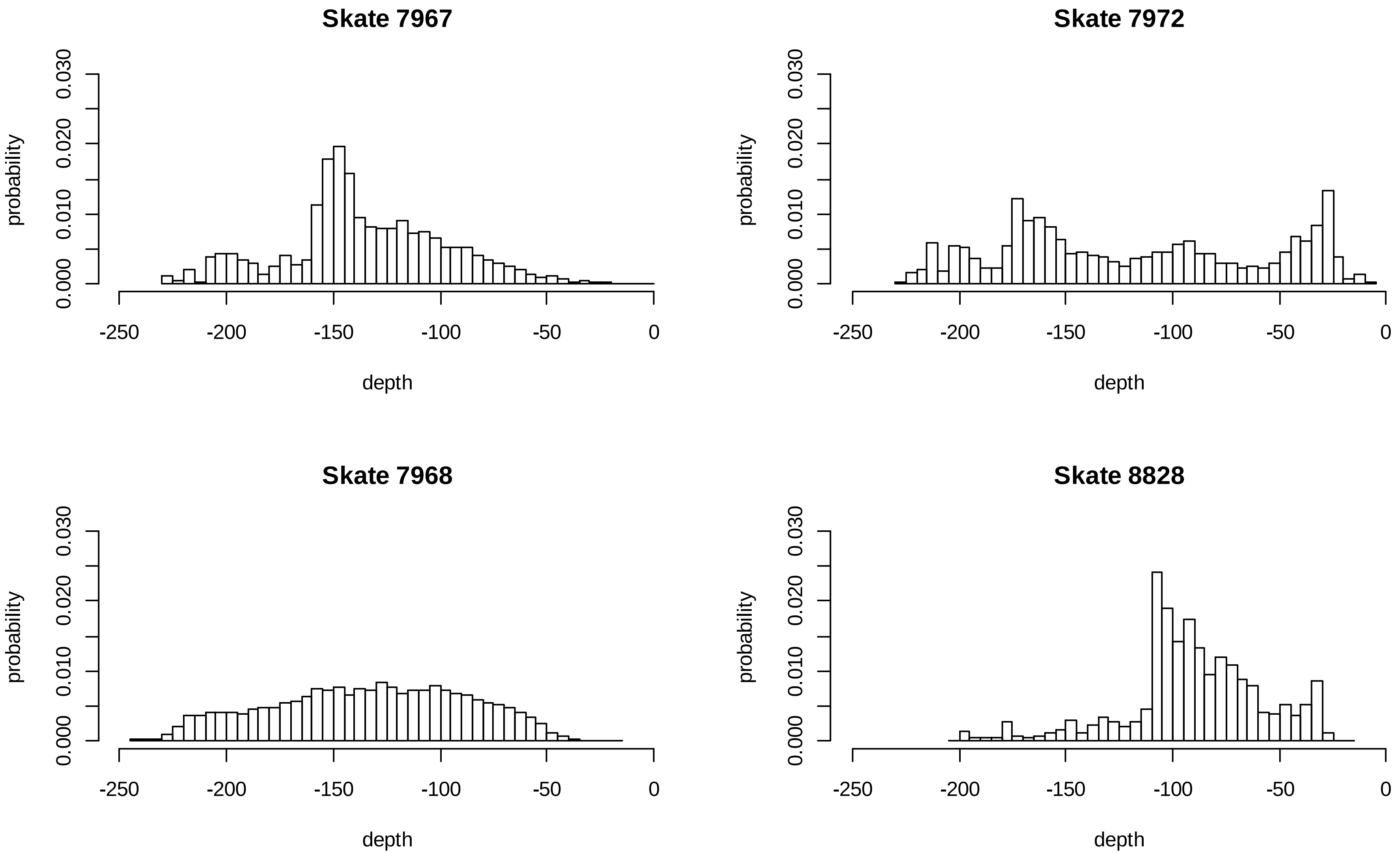
Flapper skates' depth profiles

Skate 7967	male	juvenile	12 months	T=254116
Skate 7972	male	adult	6 months	T=127360
Skate 7968	female	adult	6 months	T=107891
Skate 8828	male	adult	12 days	T=11404

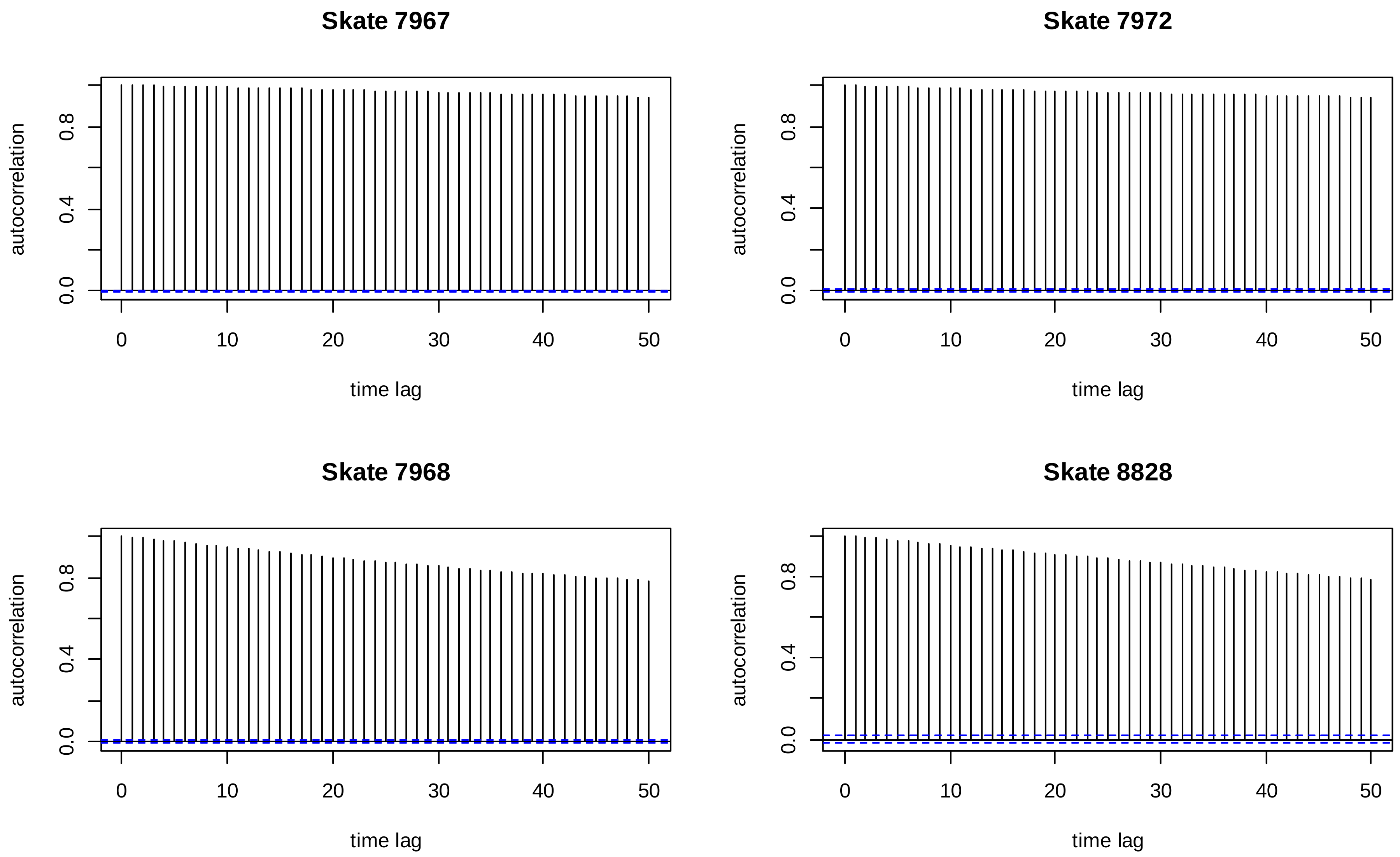
Flapper skates' depth profiles



Flapper skates' depth profiles



Flapper skates' depth profiles



High-frequency time series

High-frequency time series exhibit sample ACF that persists for a long time [i.e., long memory processes]

Modelling strategy 1: fractional differentiation (d)

$$(1-B)^d y_t = e_t \qquad e_t \sim \mathcal{N}(0; \sigma_e^2)$$

$$(1-B)^d y_t = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\,\Gamma(-d)}$$

$$-0.5<d<0.5$$

High-frequency time series

High-frequency time series exhibit sample ACF that persists for a long time [i.e., long memory processes]

Modelling strategy 1: fractional differentiation

ARFIMA(p,d,q)

$$\varphi(B)(1-B)^d y_t = \theta(B)e_t \qquad -0.5<d<0.5$$

$$\varphi(B) = 1-\varphi_1 B-...-\varphi_p B^p; \qquad B^j y_t=y_{t-j}; \quad j=1,..,p$$

$$\theta(B) = 1+\theta_1 B+...+\theta_q B^q$$

High-frequency time series

Modelling strategy 2: non-linear time series with structural breaks produce realizations that appear to have long memory

"structural change" and "long memory" are effectively different labels for the same phenomenon

structural changes can be modelled as stochastic regime switching

e.g., Markov switching autoregressive models (non-Normal and non-linear models)

Markov switching autoregressive models

$(\{y_t\}; \{x_t\})$

$\{x_t\}$: m-state *hidden* Markov chain

$S_X = \{1, 2, \ldots, m\}$

$P(x_t = i \mid x_{t-1} = j) = \gamma_{j,i}$ $\quad 0 < \gamma_{j,i} < 1 \quad \forall\, i,j \in S_X$

$\Gamma = [\gamma_{j,i}]_{(m \times m)}$

$\{y_t\}$: conditional autoregressive process of order p

$y_t = \mu_{(i)} + \varphi_{1(i)} y_{t-1} + \varphi_{2(i)} y_{t-2} + \ldots + \varphi_{p(i)} y_{t-p} + e_t \qquad e_t \sim \mathcal{N}(0; \lambda_{(i)}^{-1})$

Non-homogeneous Markov switching autoregressive models

$$\Gamma^t = [\gamma^t_{j,i}]_{(m\times m)}$$

$$\gamma^t_{j,i} = P(x_t=i \mid x_{t-1}=j) \qquad 0<\gamma_{j,i}<1 \qquad \forall\, i,j \in S_X; \quad \forall\, t=2,\ldots,T$$

$$Z=(z_{p+1},\ldots,z_t,\ldots, z_T)' \qquad z_t=(z_{t,1},\ldots, z_{t,n})' \qquad \forall\, t=p+1,\ldots,T$$

$$\alpha=[\alpha_{j,i}] \qquad [\alpha_{j,i}]=(\alpha_{j,i,0},\alpha_{j,i,1},\ldots,\alpha_{j,i,n}) \qquad \forall\, i,j \in S_X$$

$$\text{logit}(\gamma^t_{j,i}) = \ln(\gamma^t_{j,i} / \gamma^t_{j,j}) \qquad \forall\, i,j \in S_X$$

$$\gamma^t_{j,i} = \frac{\exp(z_t\alpha_{j,i})}{1 + \sum_{i\neq j} \exp(z_t\alpha_{j,i})} \qquad \gamma^t_{j,i} = \frac{1}{1 + \sum_{i\neq j} \exp(z_t\alpha_{j,i})}$$

Non-homogeneous Markov switching autoregressive models

$\Gamma^t = [\gamma^t_{j,i}]_{(m\times m)}$ $\gamma^t_{j,i} = P(x_t=i|\, x_{t-1}=j)$

$0< \gamma_{j,i} <1 \quad \forall\, i,j \in S_X; \quad \forall\, t=2,\ldots,T$

At each time t:

Categorical covariate D_t defined on d categories

$G^h = [g^h_{j,i}]_{(m\times m)}$ $g^h_{j,i} = P(x_t=i|\, x_{t-1}=j, D_t=h)$

$h=1,\ldots,d \quad \forall\, i,j \in S_X; \quad \forall\, t=2,\ldots,T$

$$\Gamma^t = \sum_{h=1}^{d} G^h \times \mathrm{I}(D_t=h)$$

Prior specification

$$p(G^h) = \prod_{j=1}^{m} p(G^h_{j\bullet})$$ $$G^h_{j\bullet} = (g^h_{j,1}, g^h_{j,2}, \ldots, g^h_{j,m})$$

$$G^h_{j\bullet} \sim \mathcal{D}(\bullet)$$ for all j=1,…,m and h=1,...,d

$$\mu_{(i)} \sim \mathcal{N}(\bullet;\bullet),$$ for all i=1,…,m

$$\lambda_{(i)} \sim \mathcal{G}(\bullet;\bullet),$$ for all i=1,…,m

$$p(\varphi) = \prod_{j=1}^{p} \prod_{i=1}^{m} p(\varphi_{j(i)}),$$

$$\varphi_{j(i)} \sim \mathcal{N}(\bullet;\bullet),$$ for all i=1,…,m and j=1,…,p

Gibbs sampling

- $\mu_{(i)} \leftarrow$ Normal for all i=1,…,m
- $\lambda_{(i)} \leftarrow$ Gamma for all i=1,…,m
- $\varphi_{j(i)} \leftarrow$ Normal for all i=1,…,m and j=1,…,p
- $G^{h}_{j\bullet} \leftarrow$ Dirichlet for all j=1,…,m and h=1,...,d
- sample permutation
- $(x_1,...x_T) \leftarrow$ forward filtering – backward sampling algorithm

Label switching

Iteration k = 1,...,L: Burn-in

Iteration k = L+1,...,L+M: Posterior mode

$$\{\mu^*, \lambda^*, \varphi^*, G^*\}=\arg\max p(\mu^{(k)}, \lambda^{(k)}, \varphi^{(k)}, G^{(k)})$$

Iteration k=L+M+1,…, L+M+N: Permutations

$$\eta^*=\arg\min_{\eta_j \in \boldsymbol{H}} ||\eta_j(\mu^{(k)}, \lambda^{(k)}, \varphi^{(k)}, G^{(k)}) - (\mu^*, \lambda^*, \varphi^*, G^*)||$$

&

$$(\mu^{(k)}, \lambda^{(k)}, \varphi^{(k)}, G^{(k)}) = \eta^*(\mu^{(k)}, \lambda^{(k)}, \varphi^{(k)}, G^{(k)})$$

3 steps: model choice
variable selection
parameter estimation and hidden chain reconstruction

$m = 1,\ldots,4$; $p = 0,\ldots,6$; 3 covariates $\Rightarrow$ 224 competing models

Bayesian inference

3 steps: model choice
variable selection
parameter estimation and hidden chain reconstruction

$m = 1,\ldots,4$; $p = 0,\ldots,6$; 3 covariates $\Rightarrow$ 224 competing models

$\Gamma^t = [\gamma^t_{j,i}]_{(m \times m)}$ $\qquad \gamma^t_{j,i} = P(x_t = i \mid x_{t-1} = j)$

3 covariates: Solar cycle S $\qquad s \in \{0; 1\}$
Lunar cycle L_2 $\qquad l_2 \in \{0; 1\}$
Lunar phase L_4 $\qquad l_4 \in \{1; 2; 3; 4\}$

Indicator D_t $\quad d \in \{1,2,4,8,16\}$

$G^h = [g^h_{j,i}]_{(m \times m)}$ $\qquad g^h_{j,i} = P(x_t = i \mid x_{t-1} = j, D_t = h)$ $\qquad \Gamma^t = \quad G^h \times \mathrm{I}(D_t = h)$

Bayesian inference

3 steps: model choice
variable selection
parameter estimation and hidden chain reconstruction

m = 1,…,4; p = 0,…,6; 3 covariates $\Rightarrow$ 224 competing models

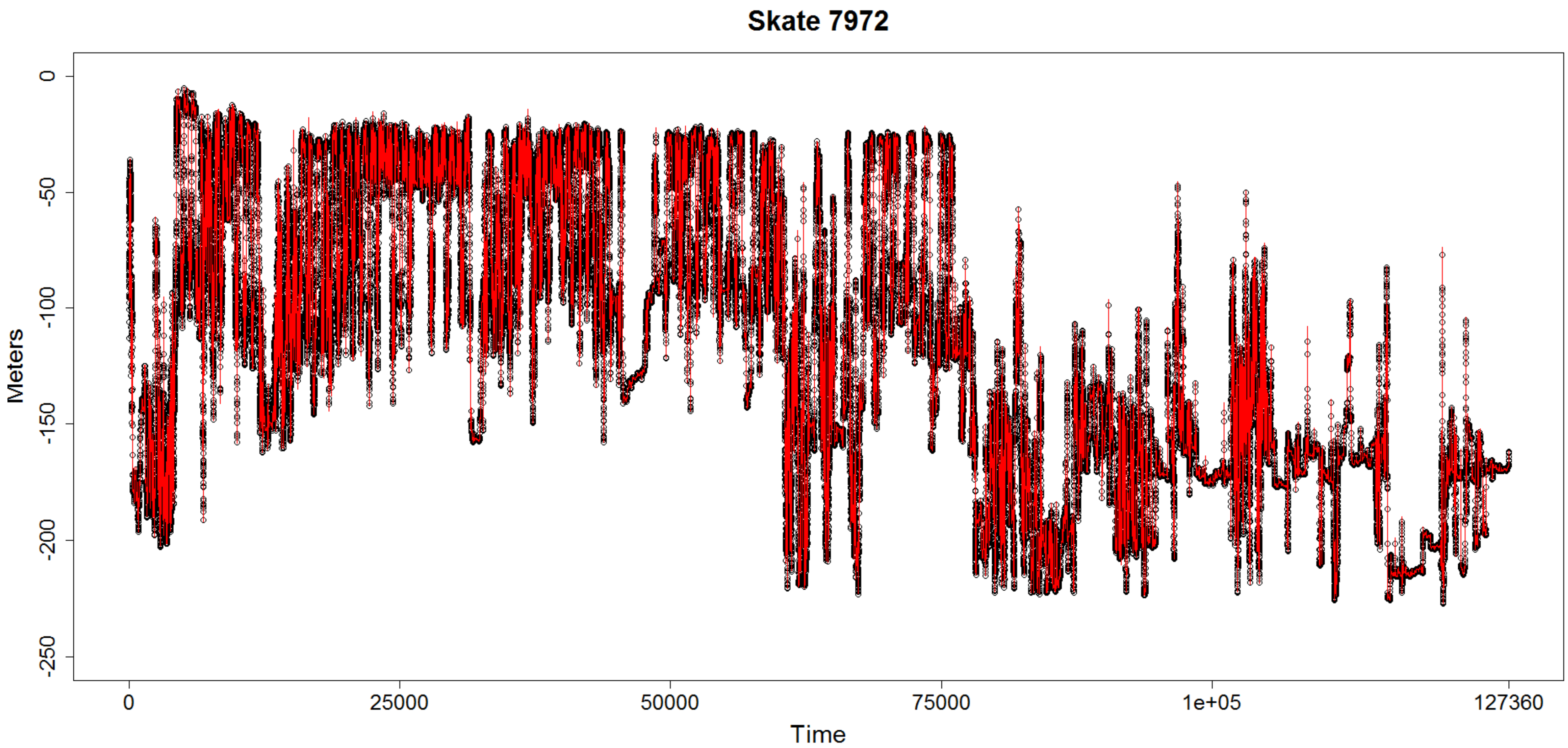
Pragmatic alternative:
[1] best(m;p), with d=1
[2] best combination of covariates $\Rightarrow$ 35 competing models

Model choice and variable selection: Bayes factors

ln(marginal likelihood) computed by the method of Chib (2001)

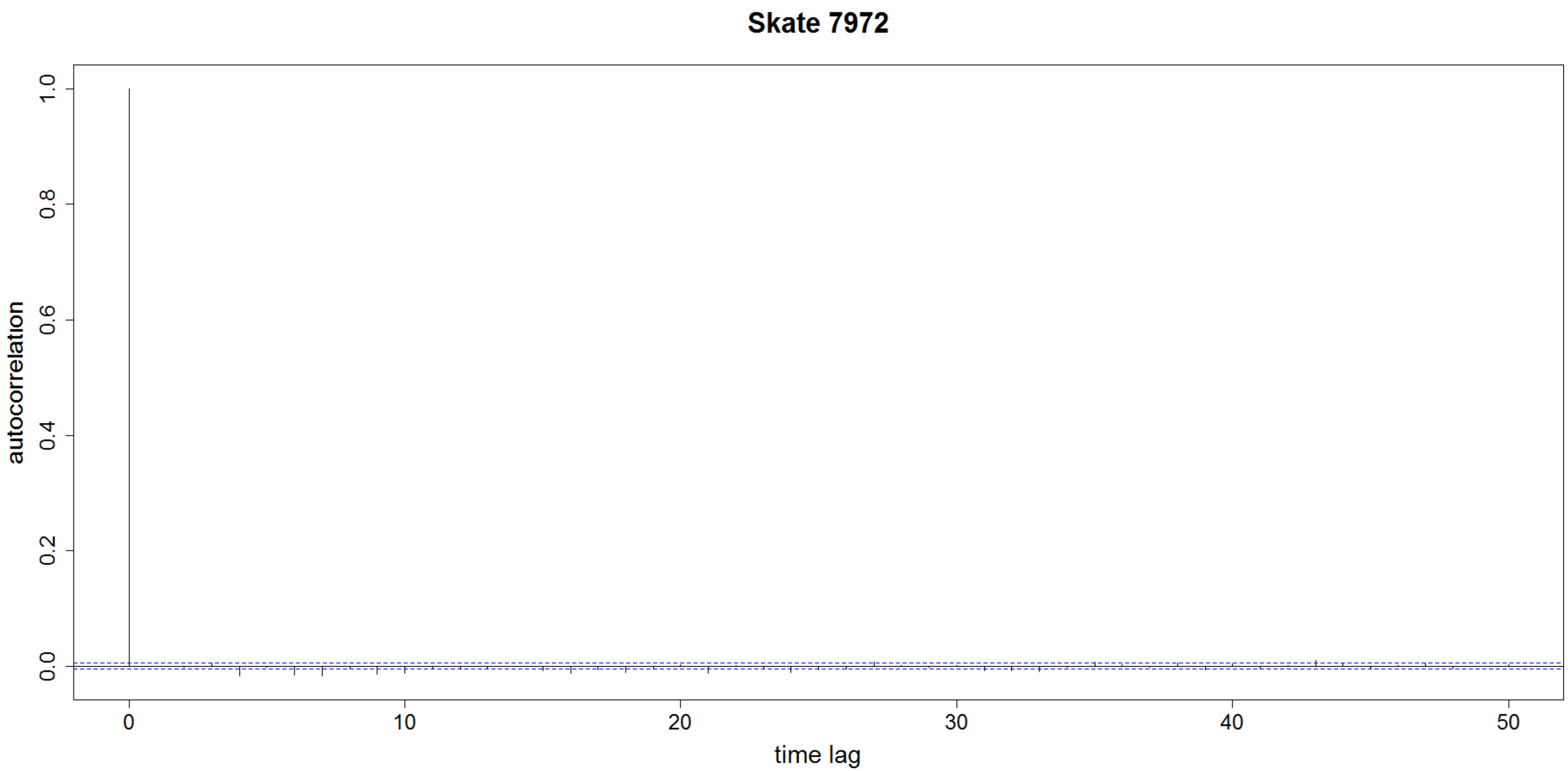
Skate 7967 (ma; ju;12 months)	m=2;p=4; L_4
Skate 7972 (ma; ad; 6 months)	m=2;p=3; L_4
Skate 7968 (fe; ad; 6 months)	m=2;p=3; L_4;L_2
Skate 8828 (ma; ad; 12 days)	m=2;p=4; L_4;L_2;S

Results



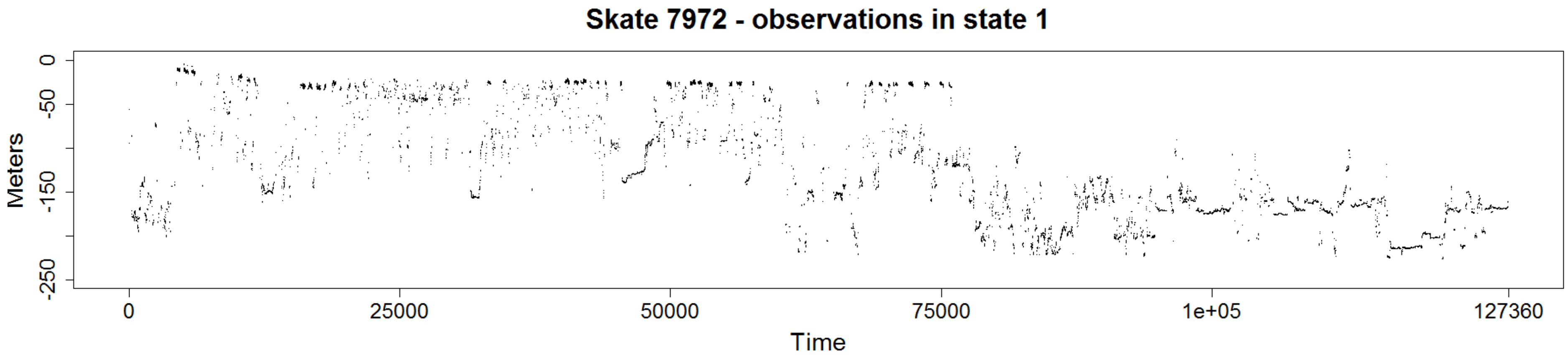
actual (black dots)   and   fitted (red line)

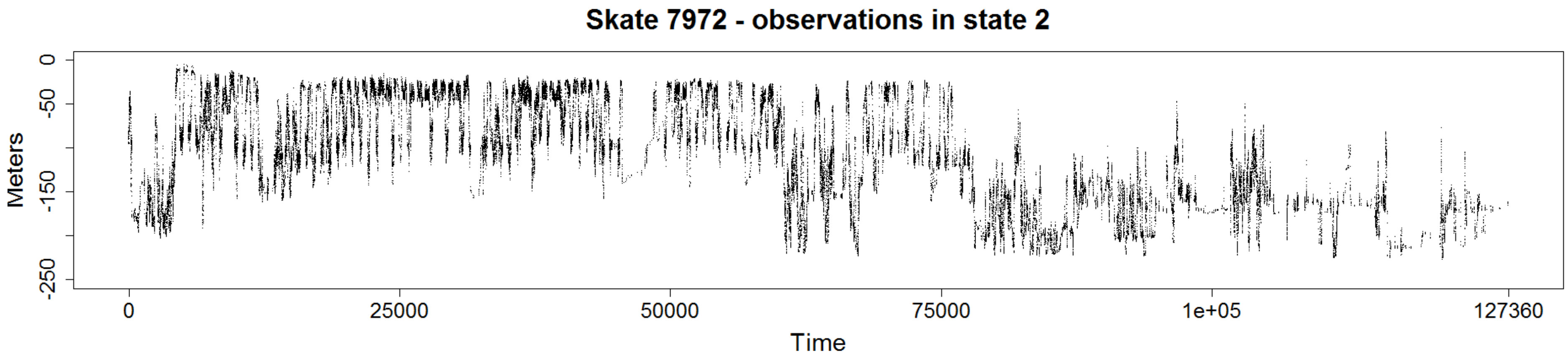
Results



ACF of the residuals

Results

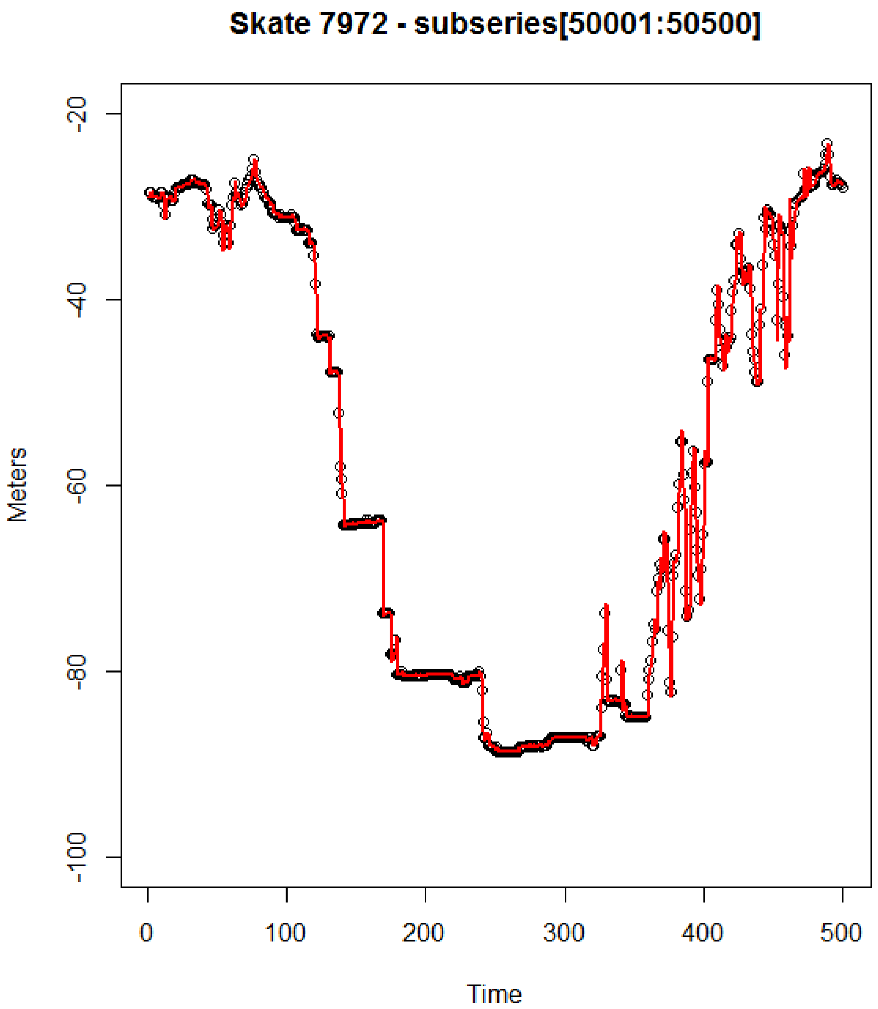




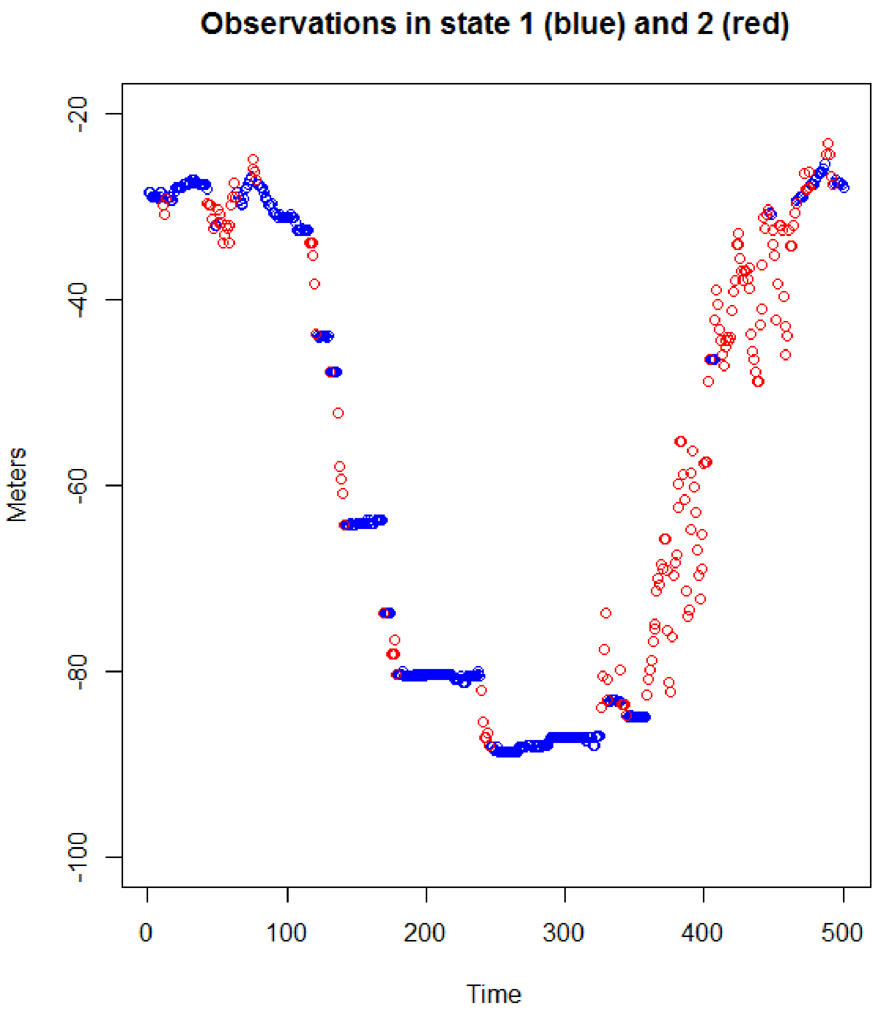
State 1 (68,337 visits, 54%) = low variability (resting, horizontal movement, slow ascending and descending)

State 2 (59,019 visits, 46%) = high variability (fast ascending and descending)

Results



Skate 7972 - subseries[50001:50500]
Meters
-20
-40
-60
-80
-100
0
100
200
300
400
500
Time



Observations in state 1 (blue) and 2 (red)
Meters
-20
-40
-60
-80
-100
0
100
200
300
400
500
Time

Extending the model

$(\{y_t\}; \{x_t\})$

State-dependent autoregressive orders:

$p = (\mathbf{p_1, p_2, \ldots, p_m})$

$\{x_t\}$: m-state *hidden* Markov chain

$\{y_t\}$: conditional autoregressive process of order p_{x_t}

$$y_t = \mu_{(i)} + \varphi_{1(i)} y_{t-1} + \varphi_{2(i)} y_{t-2} + \ldots + \varphi_{\mathbf{p_i}(i)} y_{t-\mathbf{p_i}} + e_t \qquad e_t \sim \mathcal{N}(0; \lambda_{(i)}^{-1})$$

Extending the model

Markov switching ARCH noises:

$$e_t = \sqrt{h_t}\, u_t \qquad u_t \sim \mathcal{N}(0; 1)$$

$$h_t = \eta_{(i)} + \alpha_{1(i)} e^2_{t-1} + \alpha_{2(i)} e^2_{t-2} + \ldots + \alpha_{q(i)} e^2_{t-q}$$

Local stationarity of each state-dependent ARCH process:

$$\eta_{(i)} > 0;\ \alpha_{1(i)}, \ldots, \alpha_{q(i)} \geq 0;$$

$$\sum_{j=1}^{q} \alpha_{j(i)} \leq 1$$

Extending the model

Markov switching ARCH noises:

$$e_t = \sqrt{h_t}\, u_t \qquad u_t \sim \mathcal{N}(0; 1)$$

$$h_t = \eta_{(i)} + \alpha_{1(i)} e^2_{t-1} + \alpha_{2(i)} e^2_{t-2} + \ldots + \alpha_{q_i(i)} e^2_{t-q_i}$$

with state-dependent autoregressive orders:

$$q = (\mathbf{q_1, q_2, \ldots, q_m})$$

[auxiliary variable MCMC method]

Extending the model

Model and variable selection:

Bayesian Information Criterion (BIC)
Deviance Information Criterion (DIC)

on a single subspace.

Extending the model

Model and variable selection:

Bayesian Information Criterion (BIC)
Deviance Information Criterion (DIC)

on a single subspace.

Algorithm:

From Gibbs sampling to …

Summary

Biologgers applied to animals produce long memory processes due to a non-linear dynamics.

Flapper skate's depth profile can be modelled effciently by Markov switching autoregressive models with a non-Homogeneous Markov chain, where the time-varying transition probabilities depend on the dynamics of categorical covariates.

Depth values can be classiefied into two regimes representing two different classes of skates' behaviours.