Statistical modelling in movement ecology: the example of the Langevin movement model

Pierre Gloaguen (Equipe MORSE, MIA-Paris, Agroparistech)

AG MIA, 23 Mai 2019

1. Context of movement ecology
2. Home range and utilization distribution
3. The Langevin movement model

Tracking data

Fields

- Ecology:
  - Migration and home range studies;
  - Animal behavior understanding;
  - Species management;

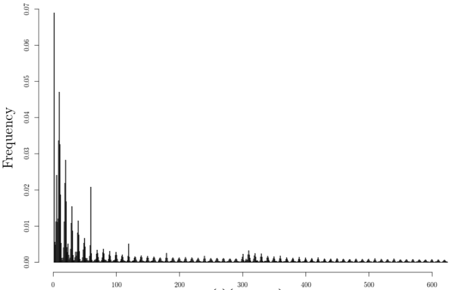
Tracking data (2)

Tracking devices

- Camera traps, Radio collars but mostly GPS;

ID	Longitude	Latitude	Date
Willy	-1.234	49.156	05/19/2010 04:13:12
Willy	-1.456	49.23	05/19/2010 04:14:58
⋮	⋮	⋮	⋮
Papa Youn	-2.314	48.236	05/28/2018 15:40:41

Time lags between 2 observations



Movement ecology

- Active research field;
- Quantitative ecologye;
- Based on increasing data;

J. theor. Biol. (1988) **131,** 419–433

Spatial Analysis of Animals' Movements Using a Correlated Random Walk Model

PIERRE BOVET AND SIMON BENHAMOU

Ecology, 85(9), 2004, pp. 2436–2445
© 2004 by the Ecological Society of America

EXTRACTING MORE OUT OF RELOCATION DATA: BUILDING MOVEMENT MODELS AS MIXTURES OF RANDOM WALKS

JUAN MANUEL MORALES,[1,4] DANIEL T. HAYDON,[2] JACQUI FRAIR,[3] KENT E. HOLSINGER,[1] AND JOHN M. FRYXELL[2]

Integrative modelling of animal movement: incorporating *in situ* habitat and behavioural information for a migratory marine predator

Sophie Bestley, Ian D. Jonsen, Mark A. Hindell, Christophe Guinet and Jean-Benoît Charrassin

Proc. R. Soc. B 2013 **280**, first published online 7 November 2012
doi: 10.1098/rspb.2012.2262

Ecological Modelling 221 (2010) 1757–1769

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

ELSEVIER

Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models

Youen Vermard[a,b,*], Etienne Rivot[b], Stéphanie Mahévas[c], Paul Marchal[a], Didier Gascuel[b]

A lot of data

Increasing amount of data

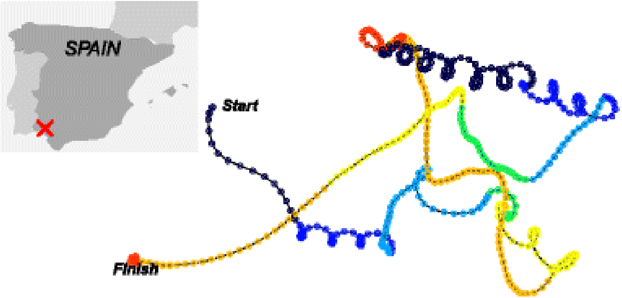
- GPS devices becomes more and more efficient;
- Less and less expensive;


Automated tracking

- Fine scale for tracking:
  - Individual scale;
  - Fine time scale;
  - Fine spatial resolution.

A classical question: Segmenting the movement

- Identifying different behaviors along the trajectory;



Source: Gurarie et al, 2017

Works in MIA-Paris Segmentation through statistical approaches:

- Hidden Markov models (M. Delattre, M.P. Etienne, P. Gloaguen)
- Segmentation in continuous time (J. Chiquet, S. Donnet, M.P. Etienne)

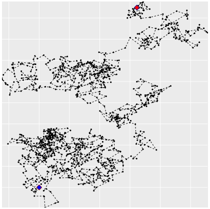
Home range analysis and utilization distribution

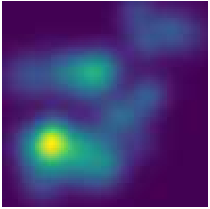
The home range (HR) concept [Burt, 1943]

"That area traversed by an individual in its normal activities of food gathering, matin, and caring for young."

From positions to maps

- Quantifying HR from GPS tracking;





Utilization distribution

Probability of finding an animal:

- X_t : animal's location at time t;
- For an area $\mathcal{A}$,

$$\mathbb{P}(X_t \in \mathcal{A}) = \int_{\mathcal{A}} \pi(z)\mathrm{d}z$$

- π is the **utilization distribution** (UD)

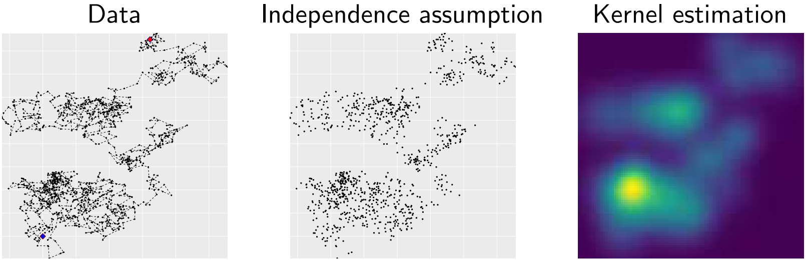
Utilization distribution estimation

Utilization distribution

$$\mathbb{P}(X_t \in \mathcal{A}) = \int_{\mathcal{A}} \pi(z)\mathrm{d}z$$

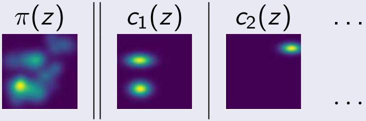
Kernel density estimation [Worton, 1989]

- Assuming all observed locations $X_{t_1}, \dots X_{t_n}$ i.i.d;
- Estimating π via kernel density methods;



Linking spatial distribution to covariates

Linking the UD to the environment



For environmental spatial covariates $c_1(z), c_2(z), \dots, c_J(z)$, assuming that:

$$\pi(z) = f(c_1(z), c_2(z), \dots, c_J(z), \theta)$$

f is a **resource selection function** (RSF).

Classical example of RSF

$$\pi(z) \propto \exp\left(\sum_{j=1}^{J} \beta_j c_j(z)\right)$$

Each β_j determines the influence of covariate j on the UD.

Link with movement ecology

Problems

- The UD estimation forgets the movement (independence assumption);
- The UD should "rise" from individual's movements;

Idea: Introducing a movement model

- GPS observations are not independent:
  - They are issued from a process;
- Movements leads to a state occupancy, the UD:
  - The UD is the density of the process observations (after a while);
- The movement is driven by spatial covariates;
  - Thus, the covariates rule the UD through the movement process;

Langevin movement model (Michelot et al. submitted)

Langevin diffusion

Let $\pi(x, \theta)$ be a smooth probability density function.
Let $(X_t)_{t \geq 0}$ be the position process of an individual, starting at X_0.
Suppose that $(X_t)_{t \geq 0}$ is solution to

$$\mathrm{d}X_t = \frac{1}{2} \nabla \log \pi(X_t, \theta) \mathrm{d}t + dW_t, \; X_0 = x_0 \tag{1}$$

Then, $(X_t)_{t \geq 0}$ is a (asymptotic) stationary process, with $\pi(x, \theta)$ as stationary distribution, i.e.,

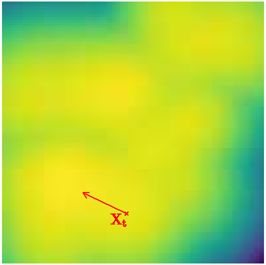
$$\mathbb{P}(X_t \in \mathcal{A}) \underset{t \to \infty}{\longrightarrow} \int_{\mathcal{A}} \pi(z) \mathrm{d}z$$

[Roberts and Tweedie, 1996]

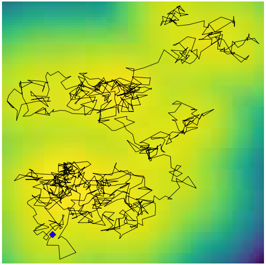
The model defined by (1) links the movement to the utilization distribution.

Langevin movement model (Michelot et al)

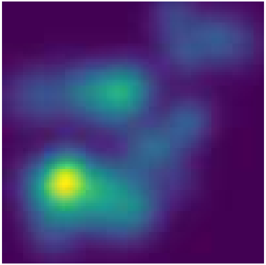
$\log \pi(z)$



$dX_t = \frac{1}{2}\nabla \log \pi(X_t, \theta)dt + dW_t$



$\rightarrow X_t \overset{t\rightarrow\infty}{\sim} \pi(z)$

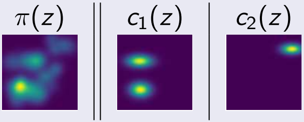


Parametric example of the Langevin movement model

$$\mathrm{d}X_t = \frac{1}{2}\nabla \log \pi(X_t, \theta)\mathrm{d}t + dW_t,\ X_0 = x_0$$

Back to a popular resource selection function

$\pi(z)$ || $c_1(z)$ | $c_2(z)$ ...



...

- For covariates $c_1, \dots, c_J$, assume that $\pi(z) \propto \exp\left(\sum_{j=1}^{J} \beta_j c_j(z)\right)$
- Then, assuming that covariate fields are smooth:

$$\mathrm{d}X_t = \frac{1}{2}\left(\beta_1 \nabla c_1(X_i) + \beta_2 \nabla c_2(X_i) + \dots + \beta_J \nabla c_J(X_i)\right)\mathrm{d}t + dW_t,\ X_0 = x_0$$

- β_j determines the influence of covariate j on movement (then, on the UD!).

Estimation of θ

Observations

- The continuous process $(X_t)_{t\geq 0}$ is observed at discrete times $t_0, \dots, t_n,\ X_{obs} = X_0, \dots, X_n$;

By Markov property of the solution to the SDE, the loglikelihood is:

$$\ell(\theta|X_{obs}) = \sum_{i=0}^{n-1} \log p_\theta(X_{i+1}|X_i, \Delta_i)$$

where

- $\Delta_i := t_{i+1} - t_i$
- $p_\theta(x|X_i, \Delta_i)$ is the transition density, i.e., the p.d.f. of $X_{i+1}|X_i$;

Problem

- In general, p_θ has no analytic expression, (even when θ is known);
- $\Rightarrow$ The likelihood can't be computed;

An approximated inference scheme

The continous time process:

$$\mathrm{d}X_t = \frac{1}{2}\nabla \log \pi(X_t, \theta)\mathrm{d}t + dW_t,\ X_0 = x_0$$

is approximated by the discrete time (with irregular time steps) process:

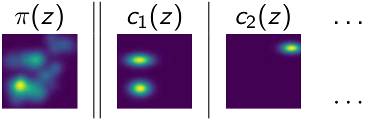
$$X_{i+1} - X_i = \frac{1}{2}\nabla \log \pi(X_i|\theta) \times \Delta_i + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N\left(0,\ \Delta_i I_2\right),$$

which becomes

$$\underbrace{\frac{X_{i+1} - X_i}{\sqrt{\Delta_i}}}_{:=Y_i} = \frac{\sqrt{\Delta_i}}{2}\nabla \log \pi(X_i|\theta) + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N\left(0,\ I_2\right).$$

Only involves Gaussian distribution, thus, maximum likelihood can pe performed.

Estimation in resource selection function example



The Euler approximation then gives

$$Y_i = \frac{\sqrt{\Delta_i}}{2}\left(\beta_1 \nabla c_1(X_i) + \beta_2 \nabla c_2(X_i) + \cdots + \beta_J \nabla c_J(X_i)\right) + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N\left(0,\ I_2\right).$$

Estimators

Classical ones of the linear model!

- Explicit estimators;
- Explicit confidence intervals;
- Model checking;

The quality depends on Euler approximation quality!

- Must be interpreted with caution;

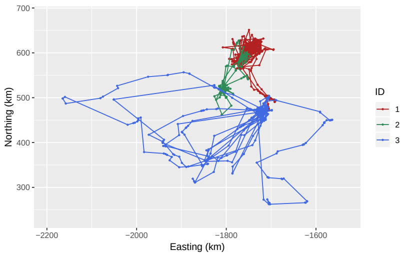
Ecological application advisory

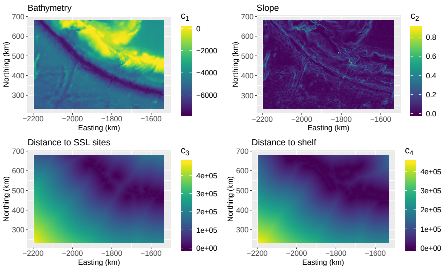
The next slide is extremely cute!

Meet the Stellar sea lion (*Eumetopias jubatus*)

Data set application

- Data set described in Wilson et al. [2018], and provided in their R package.
- Steller sea lions (*Eumetopias jubatus*) in Alaska.
- 3 trajectories, (3 different individuals), total of 2672 Argos locations.
- Time intervals were highly irregular, with percentiles $P_{0.025} = 6$min, $P_{0.5} = 1.28$h, $P_{0.975} = 17.4$h.
- 4 spatial covariates over the study region;
- Fitting of Langevin movement model, with the resource selection function;

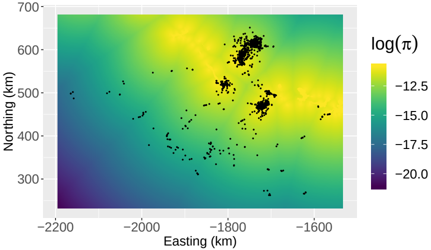




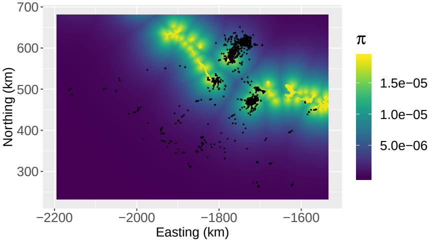
Results

	Estimate	95% CI
β_1 (Bathy)	$1.39 \cdot 10^{-4}$	$(-3.87 \cdot 10^{-7}, 2.79 \cdot 10^{-4})$
β_2 (Slope)	0.12	$(-0.14, 0.37)$
β_3 (DistSite)	$-2.50 \cdot 10^{-5}$	$(-3.58 \cdot 10^{-5}, -1.41 \cdot 10^{-5})$
β_4 (DistShelf)	$3.47 \cdot 10^{-6}$	$(2.05 \cdot 10^{-7}, 6.73 \cdot 10^{-6})$

$\log \pi(z, \theta)$



$\pi(z, \theta)$



What can we say about the inference quality? The goodness of fit?

Is the Euler approximation reliable?

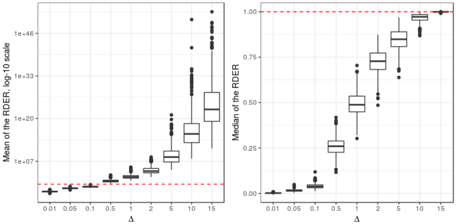
The Euler approximation giving (that gives $\hat{q}_\theta$) is only valid when $\Delta \to 0$.
How estimates can be trusted depending on Δ?

An interesting statistic? The MALA ratio

For two observations x_t and x_{t+1}, let's consider:

$$z_t = \frac{\pi(x_t)\hat{q}_\theta(x_t|x_{t+1})}{\pi(x_{t+1})\hat{q}_\theta(x_{t+1}|x_t)}$$

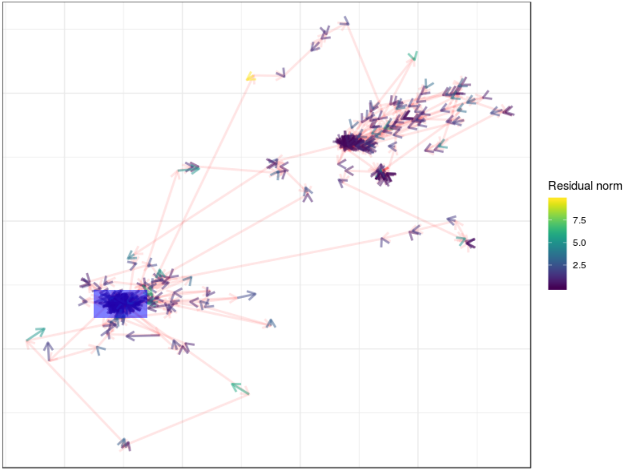
If $\hat{q}_\theta = q_\theta \Rightarrow z_t = 1 \Leftrightarrow (1 - z_t)^2 = 1$.
On the data set, for the chosen $\hat{q}_\theta$ every $(1 - z_t)^2, t = 0, \ldots, n$ can be computed.



The median tends to one when the Euler approximation gets ugly.

Checking model's residuals

Does any structure remain in the residuals?



What can we do to improve this?

Conclusions

The Langevin movement model

- Links movement and utilization distribution;
- Formulated in continuous time (sampling independent formulation);
- Naturally allows inclusion of classical resource selection function;
- (Approximated) Estimation framework based on classical linear regression;
- "User-friendly" inference framework for collaboration with ecologists

Statistical challenges

- Exact inference?
  - General tools based on importance sampling techniques recently developped, the continuous time importance sampling (CIS) [Fearnhead et al., 2017]
- What if the position is observed with error?
  - Inference could be performed using CIS and algorithms for HMMs based on SDEs (joint work with S. Le Corff and Jimmy Olsson).
- What if the position is observed with error and with regime switching?
  - ???

Merci!

References

W. H. Burt. Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy*, 24(3):pp. 346–352, 1943.

P. Fearnhead, K. Latuszynski, G. O. Roberts, and G. Sermaidis. Continuous-time importance sampling: Monte carlo methods which avoid time-discretisation error. *arXiv preprint arXiv:1712.06201*, 2017.

G. O. Roberts and R. L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.

K. Wilson, E. Hanks, and D. Johnson. Estimating animal utilization densities using continuous-time markov chain models. *Methods in Ecology and Evolution*, 9(5):1232–1240, 2018.

B. J. Worton. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, 70(1):164–168, 1989.