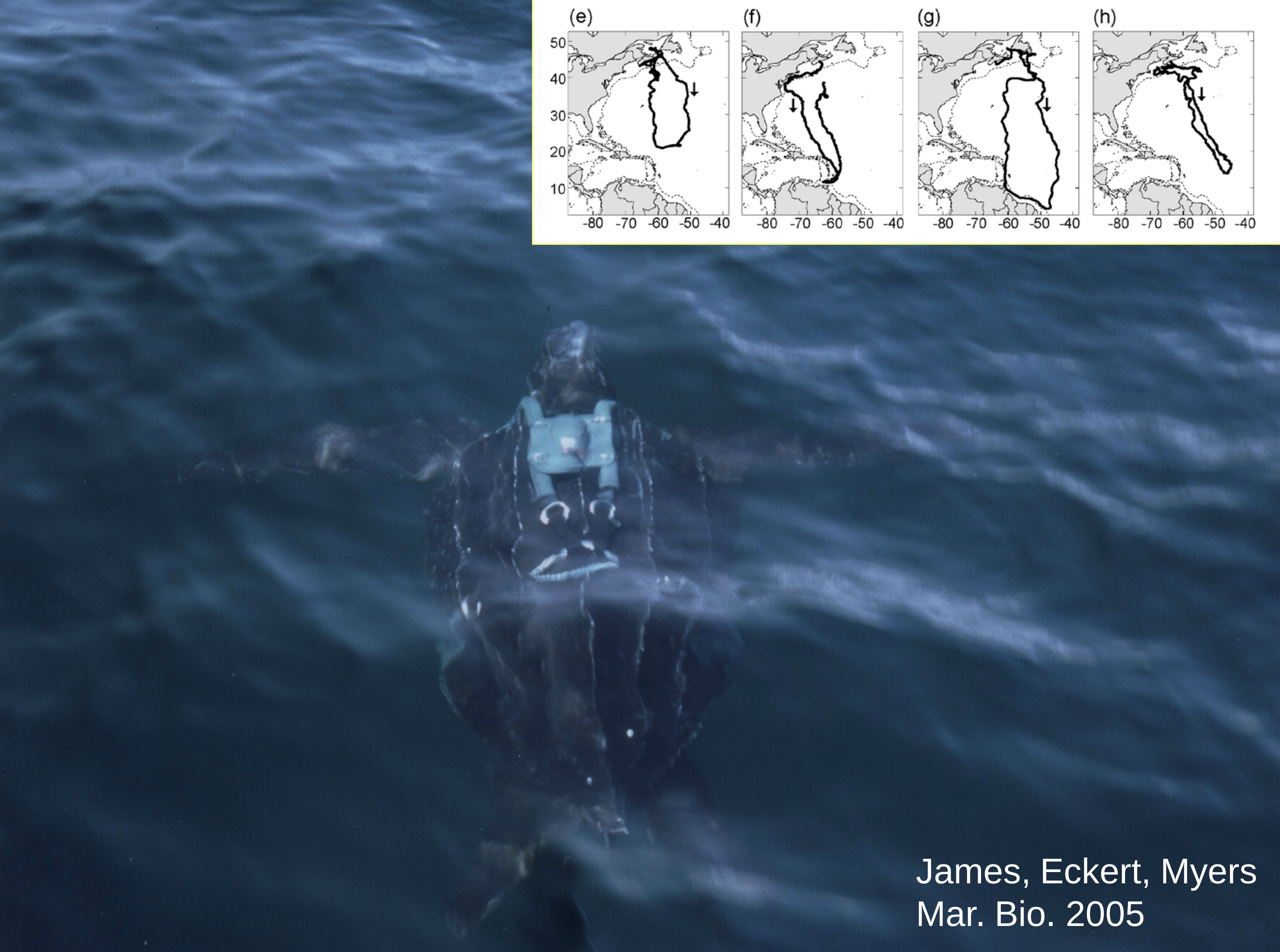
Robust, Meta-analytic, State-Space Switching Models for Movement of Marine Animals

Ransom A. Myers (RAM)
Ian Jonsen, Joanna Flemming
Greg Breed, Chris Field
Mike James

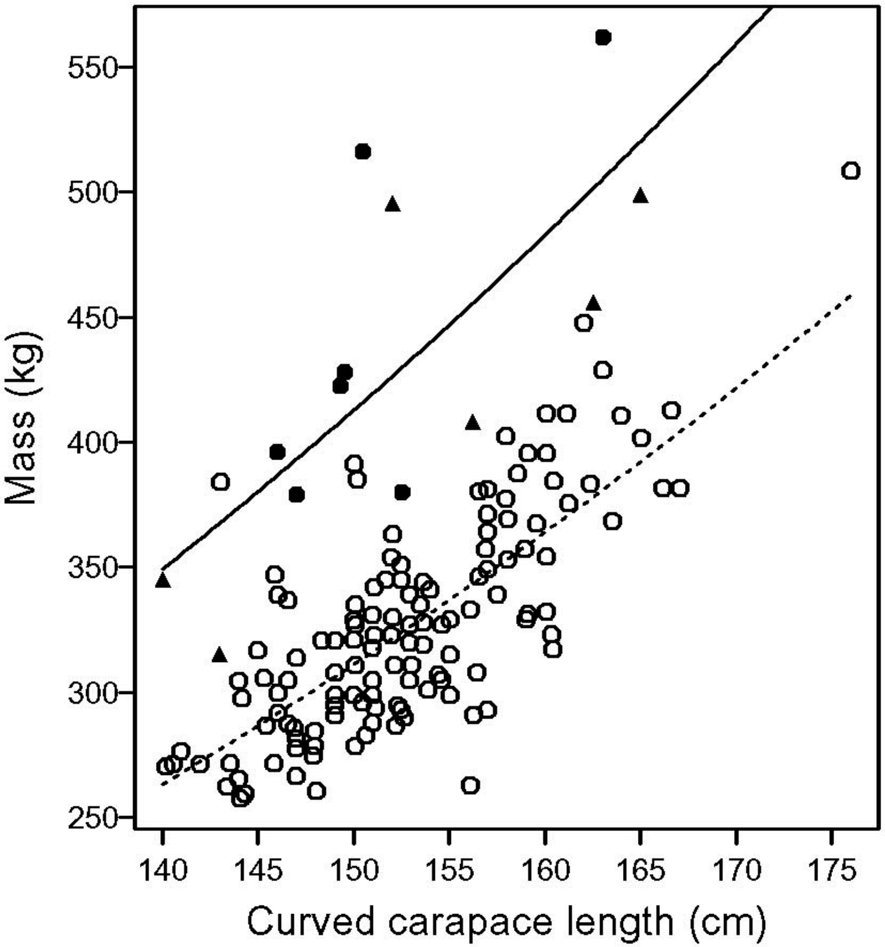
FMAP (Future of Marine Animal Populations))
National Program on Complex Data Structures

Dalhousie University
Halifax, Canada



James, Eckert, Myers
Mar. Bio. 2005

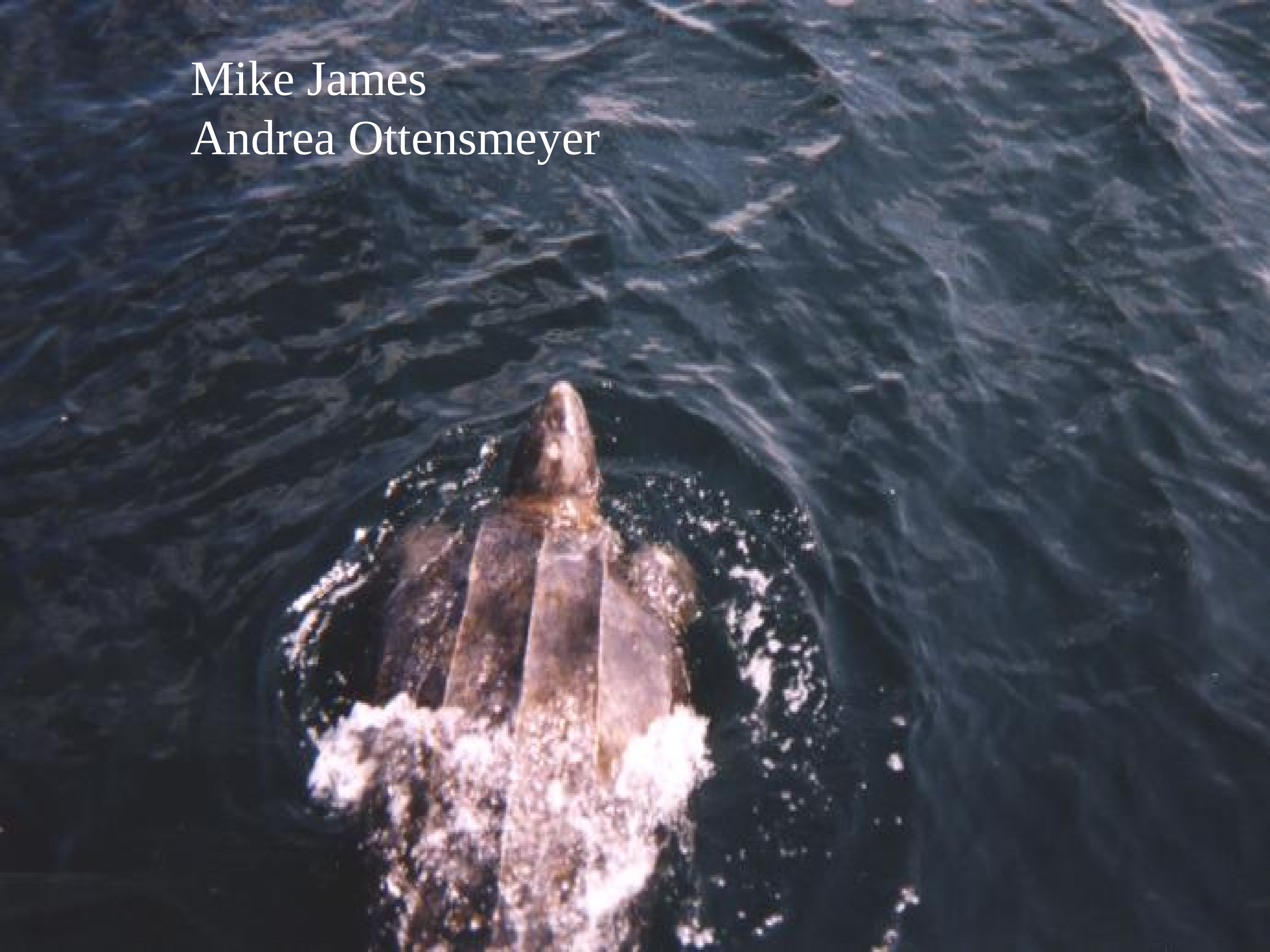
Weights in Canadian waters



Turtles are
33% heavier in Canadian coastal areas versus on the nesting beach

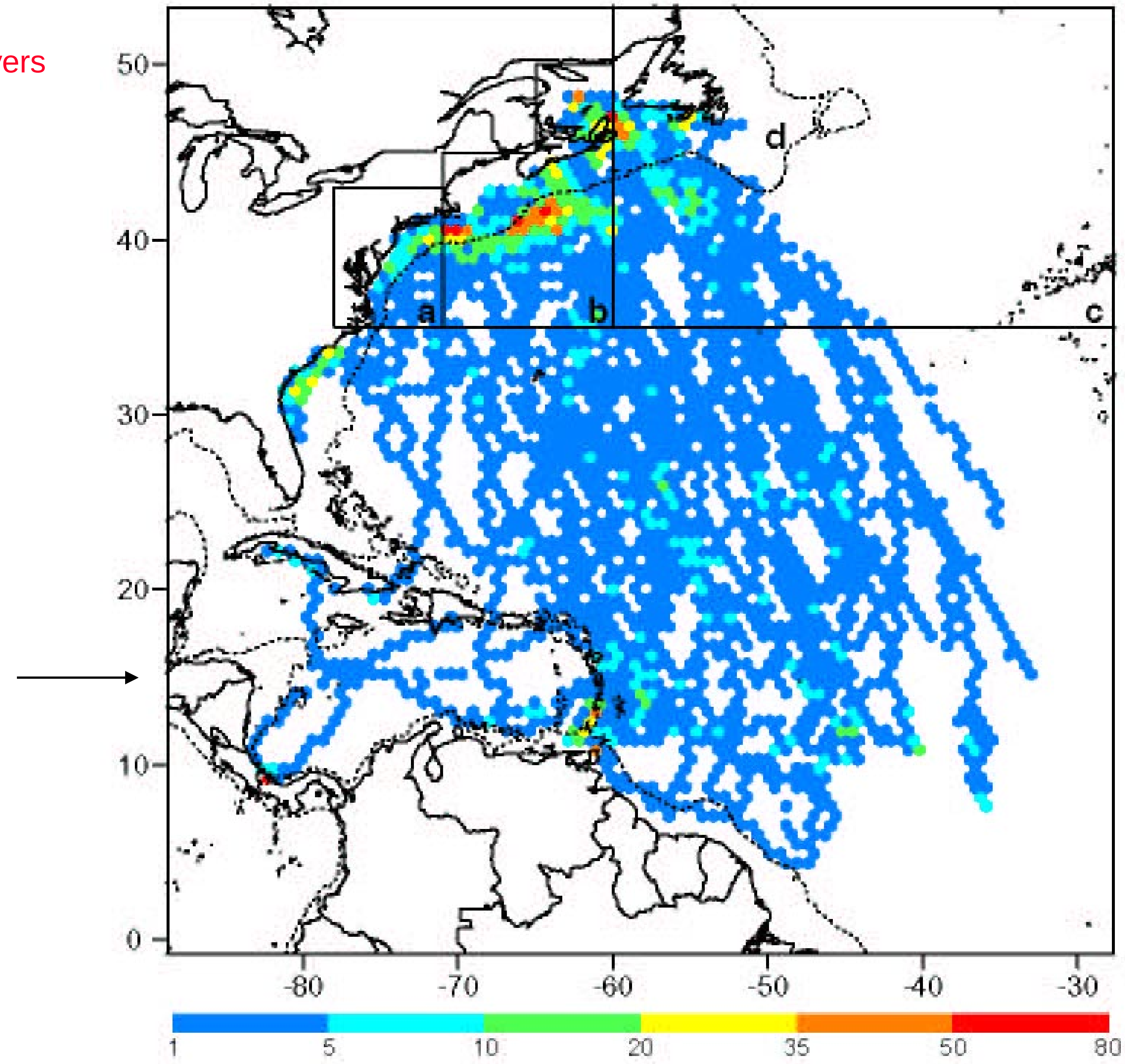
Nesting female morphometrics: St. Croix, U.S.V.I.
Boulon et al. 1996. Chelonian Conserv, Biol. 2:141-147.
Lines fit by constant slope analysis of covariance after log transformation.

• Three commercial fishing vessels are retrofitted seasonally for turtle research.

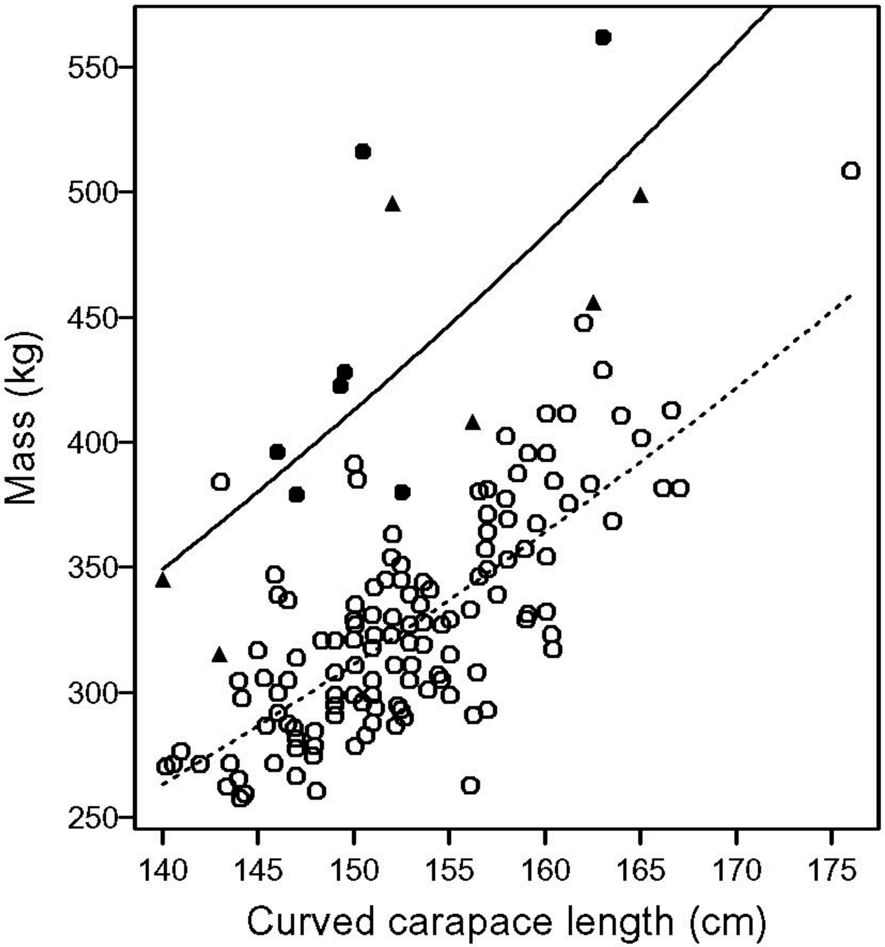

Mike James
Andrea Ottensmeyer

Identification of high-use areas and threats to leatherback sea turtles in northern waters

James, Ottensmeyer and Myers
Ecology Letters (2005)

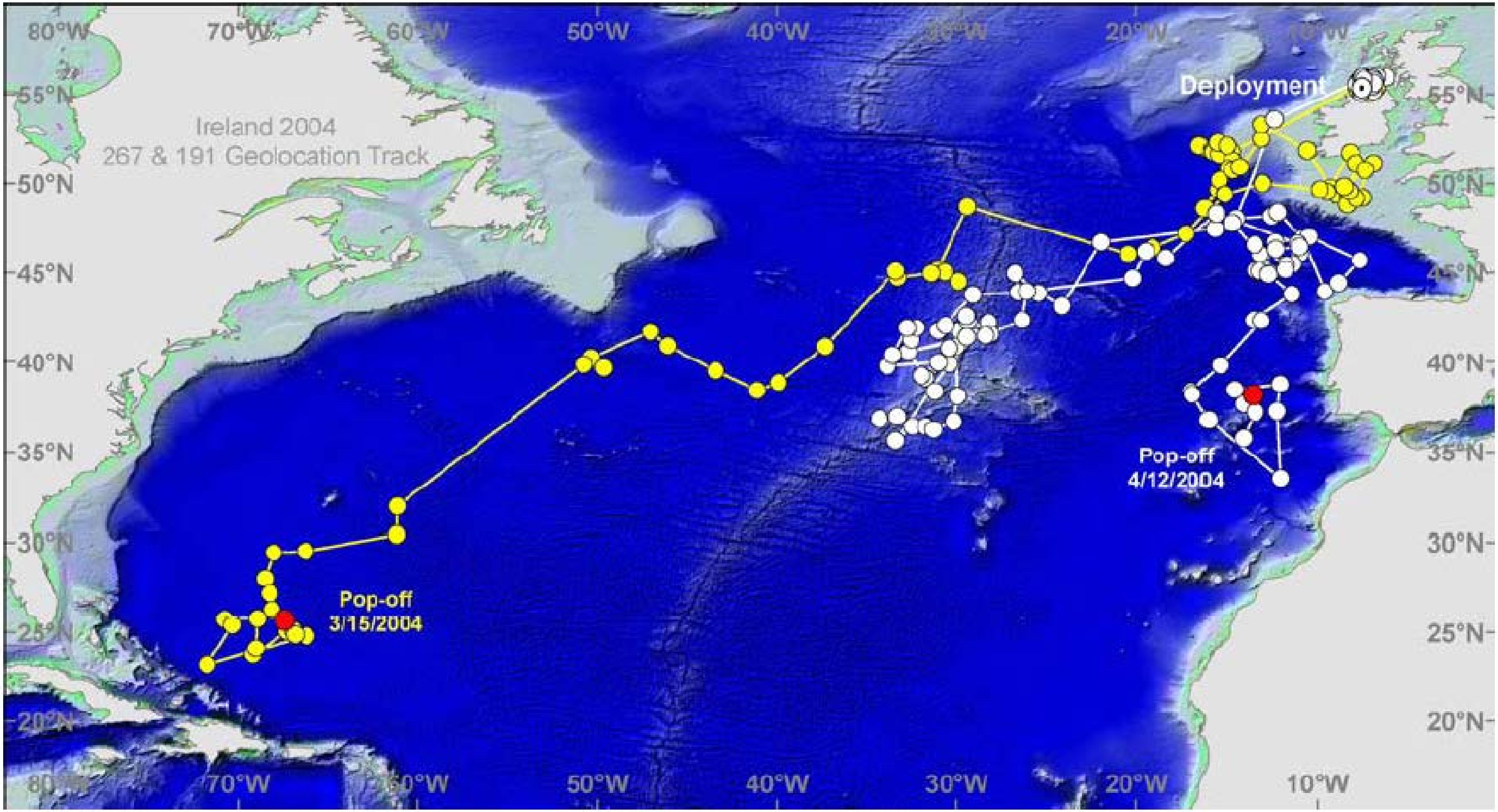


Weights in Canadian waters

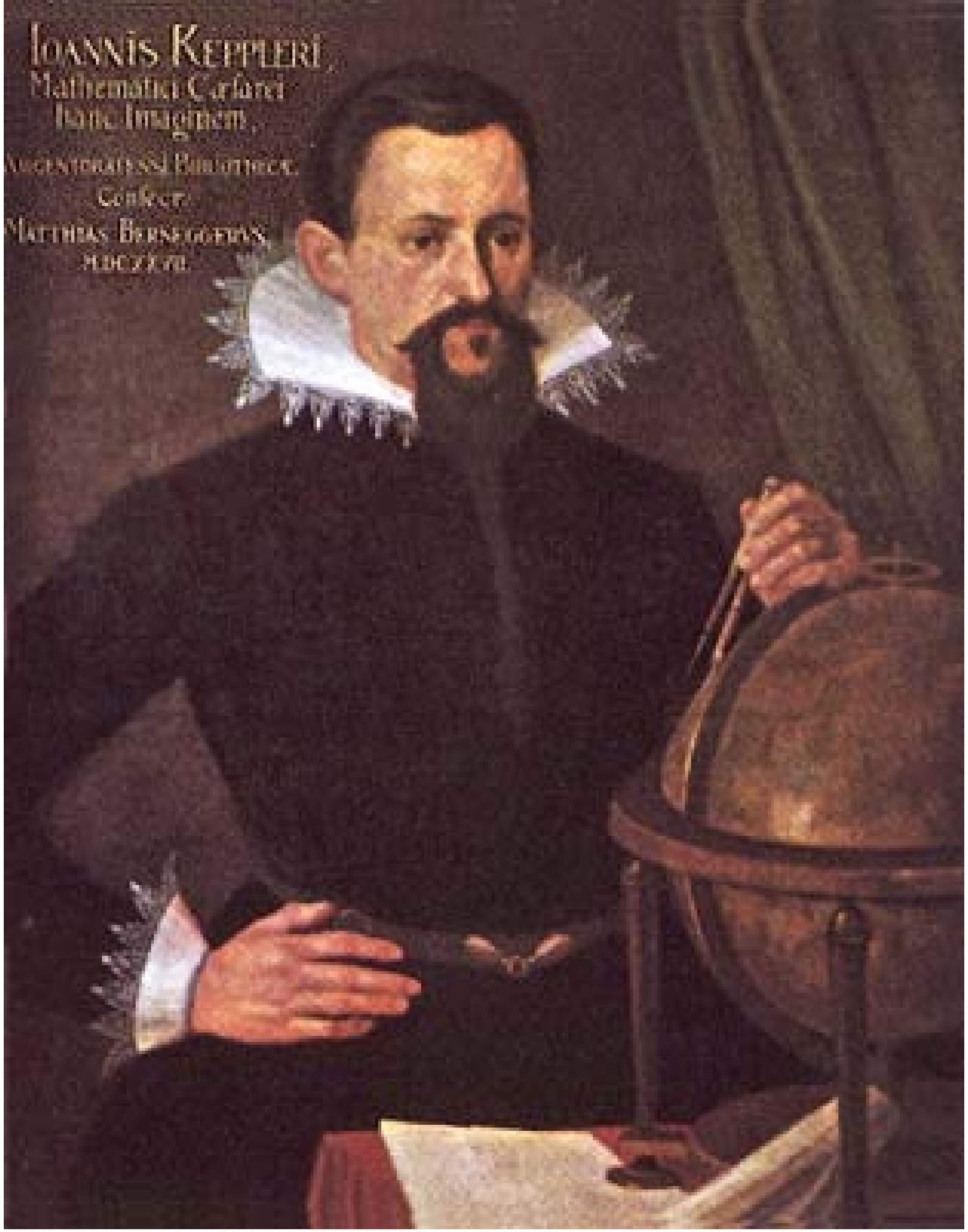


Turtles are
33% heavier in Canadian coastal areas versus on the nesting beach

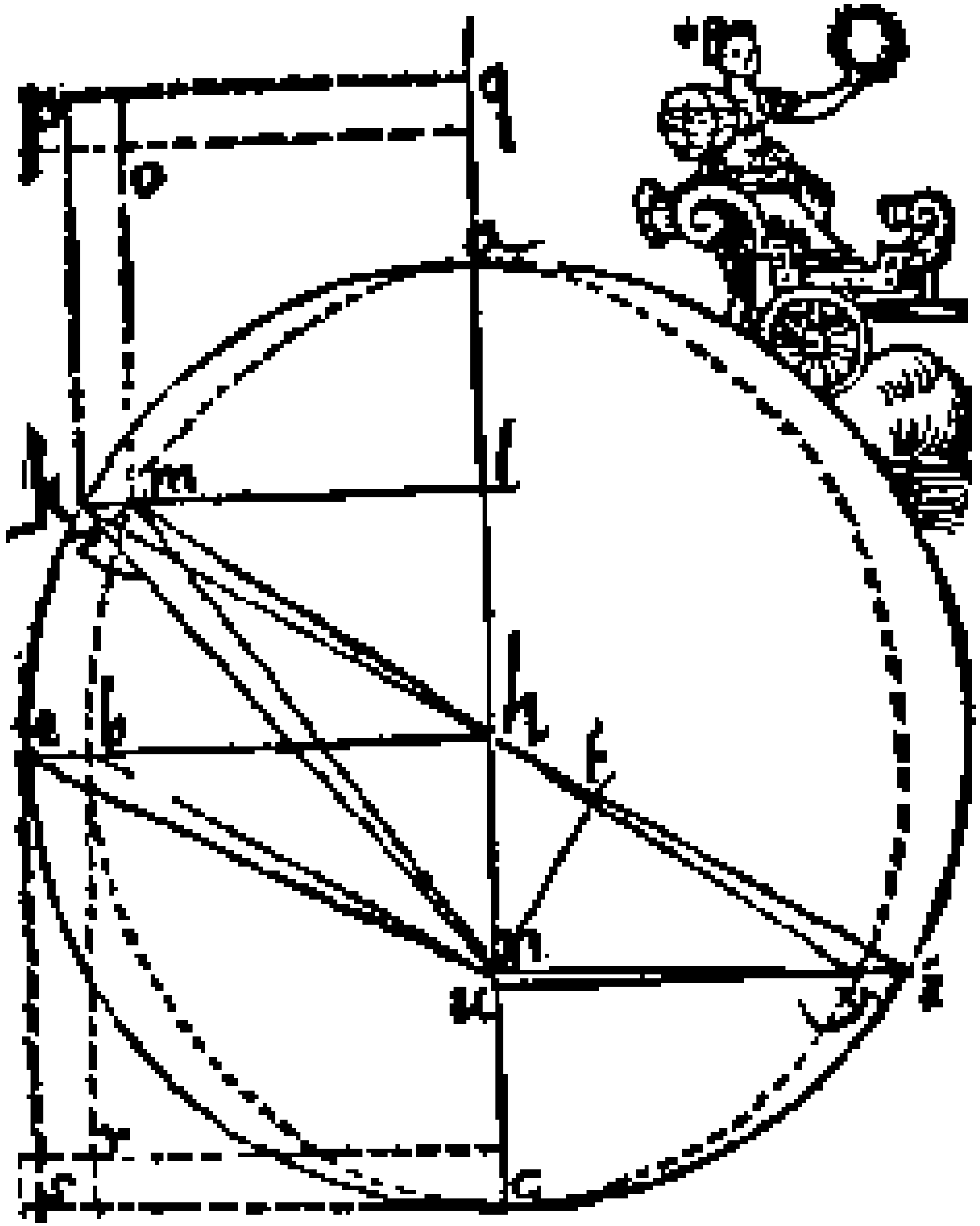
Nesting female morphometrics: St. Croix, U.S.V.I.
Boulon et al. 1996. Chelonian Conserv, Biol. 2:141-147.
Lines fit by constant slope analysis of covariance after log transformation.



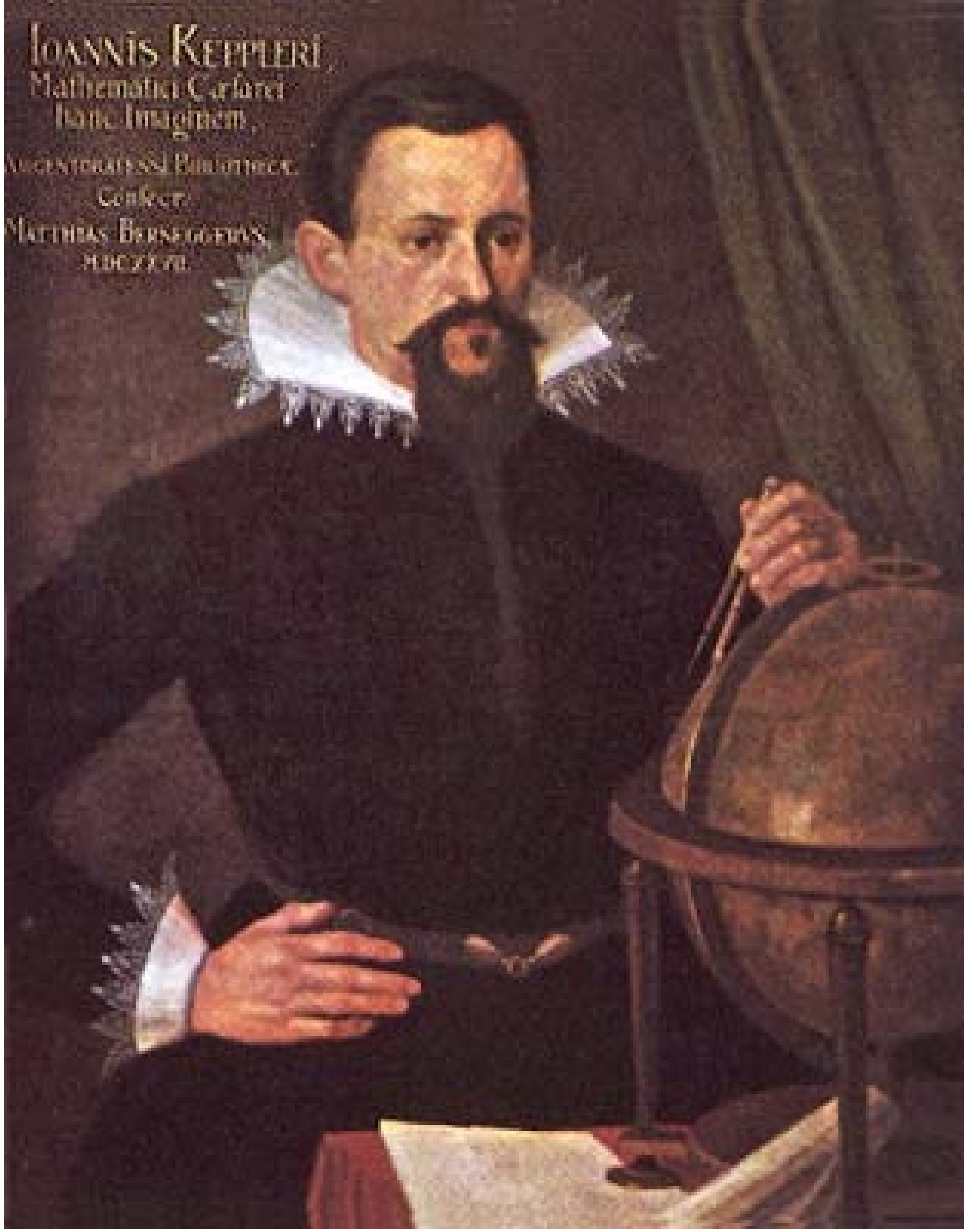
Ireland 2004
267 & 191 Geolocation Track
Deployment
Pop-off
3/15/2004
Pop-off
4/12/2004
80°W
70°W
60°W
50°W
40°W
30°W
20°W
10°W
55°N
50°N
45°N
40°N
35°N
30°N
25°N
20°N



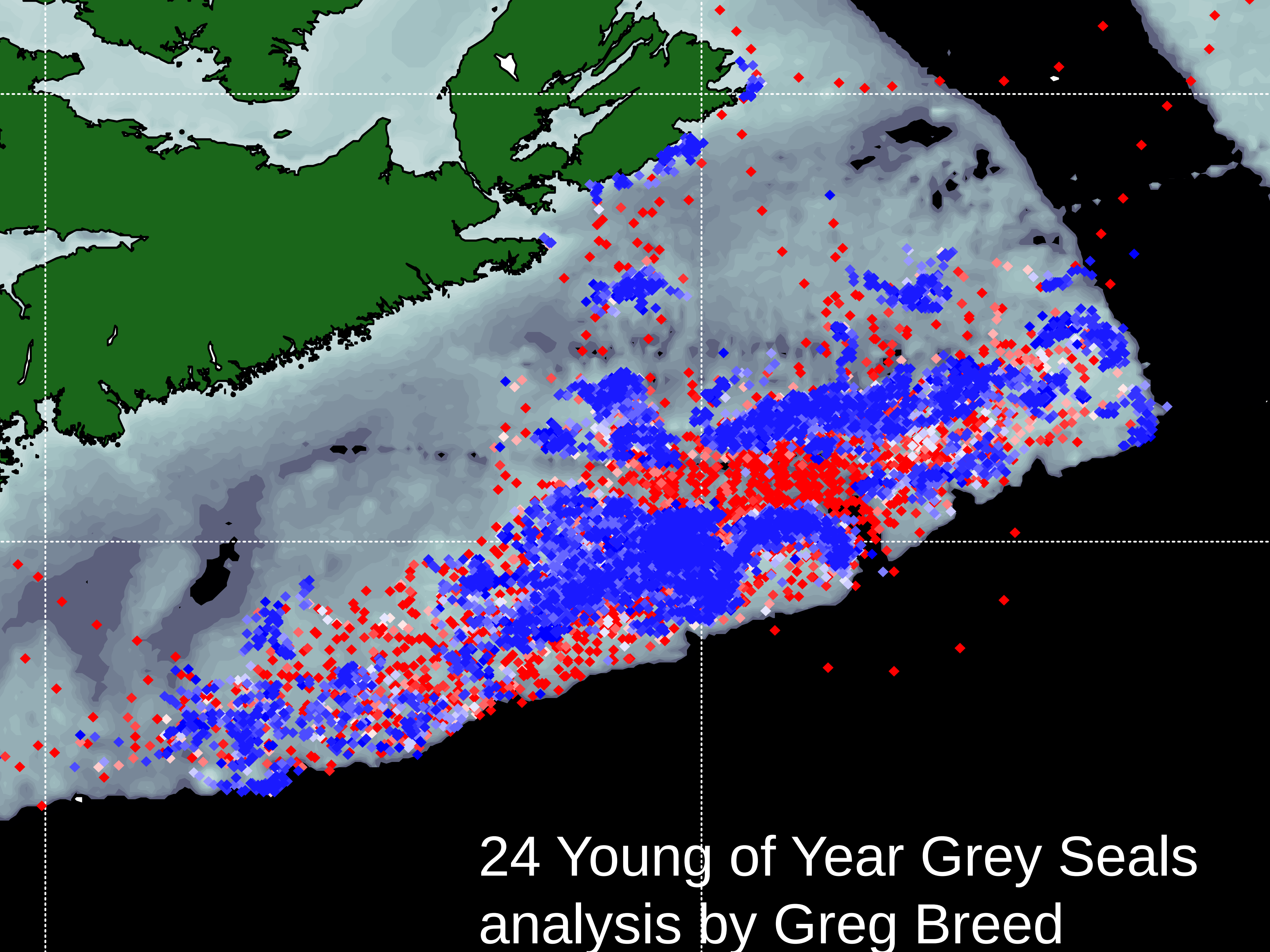
Imperial Mathematician



[Kepler's elliptical orbit for Mars.]().



Imperial Mathematician

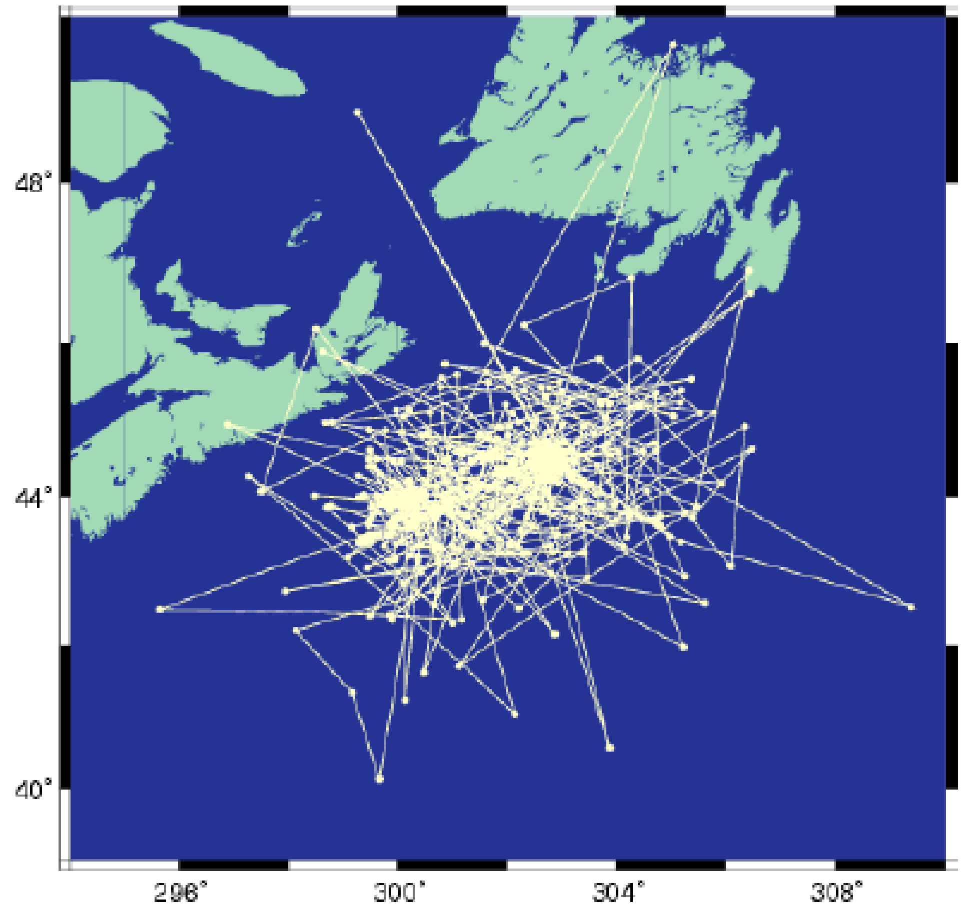

24 Young of Year Grey Seals
analysis by Greg Breed

Argos Satellite Telemetry Data

Getting more out of the data

Goals of State-Space analysis

- Infer true locations from noisy data
- Account for error w/out loss of information
- Infer behaviour, test hypotheses

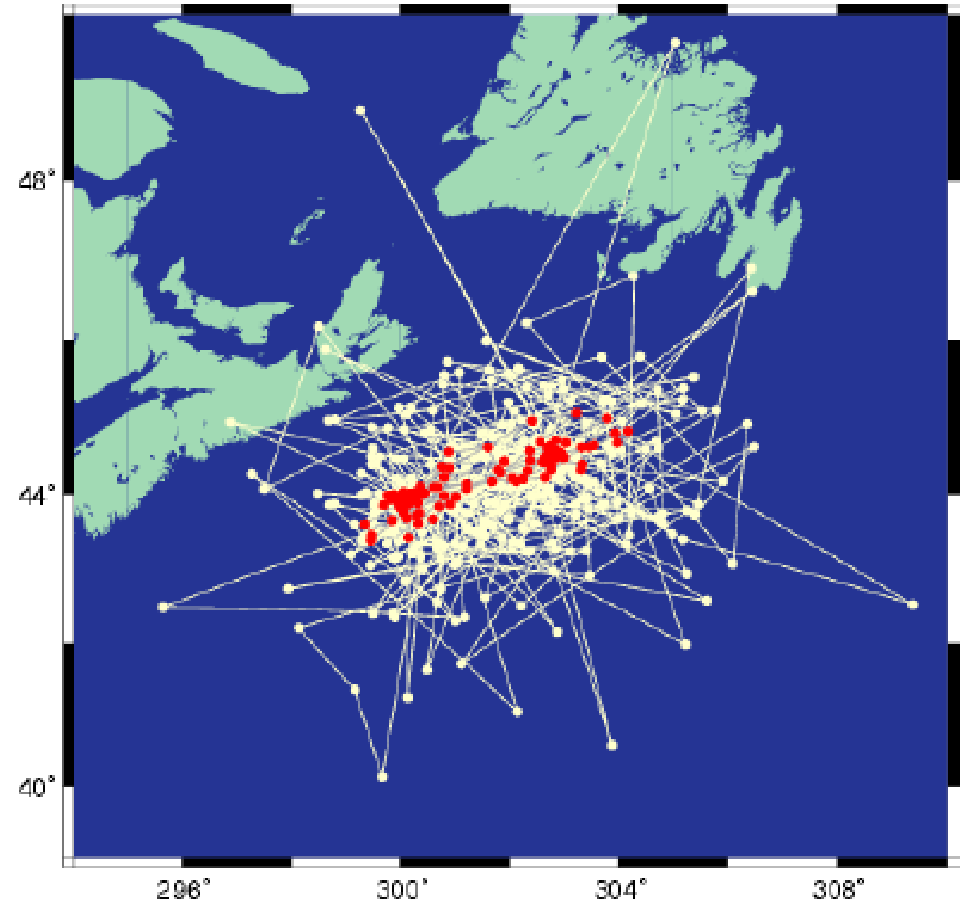


Argos Satellite Telemetry Data

Getting more out of the data

Goals of State-Space analysis

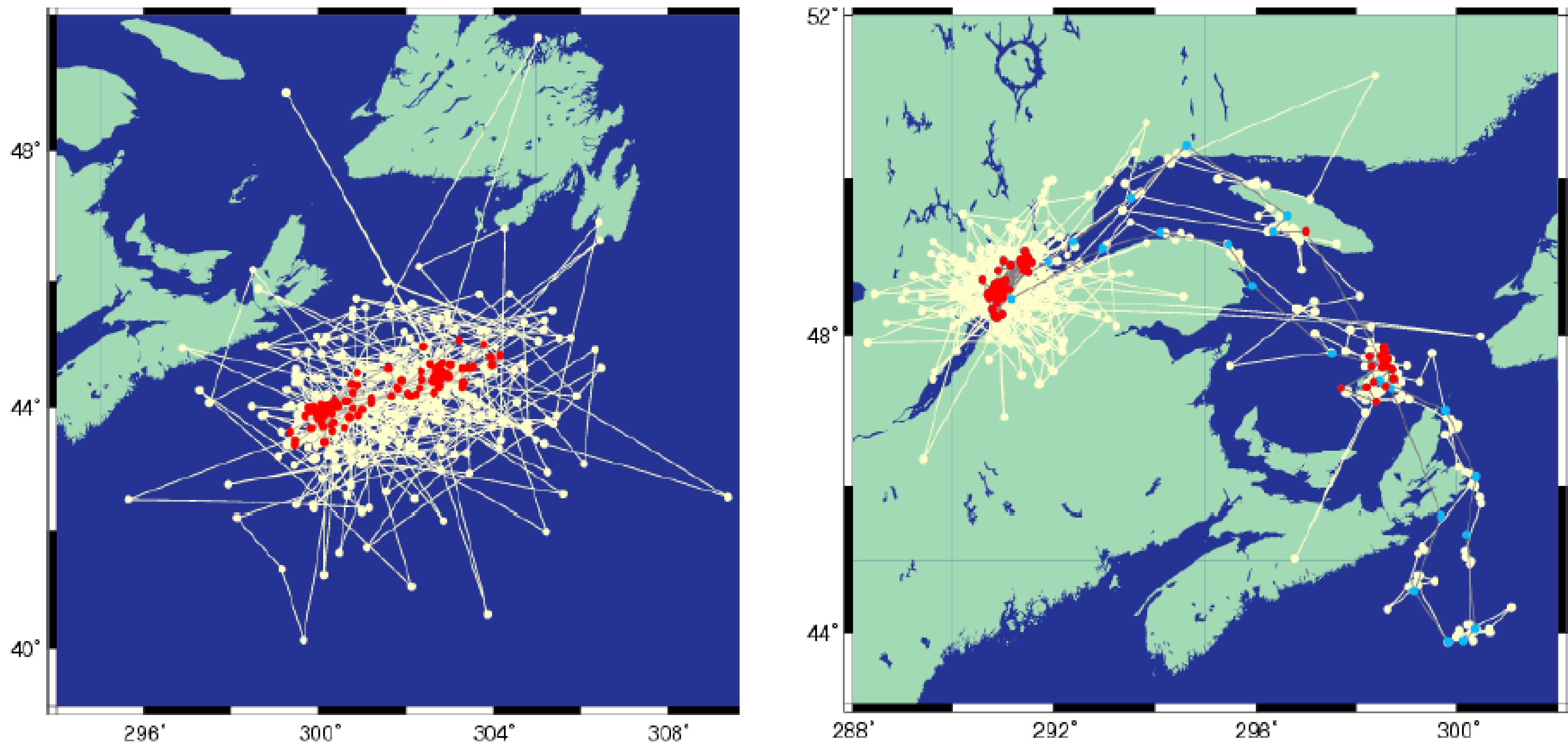
- Infer true locations from noisy data
- Account for error w/out loss of information
- Infer behaviour, test hypotheses



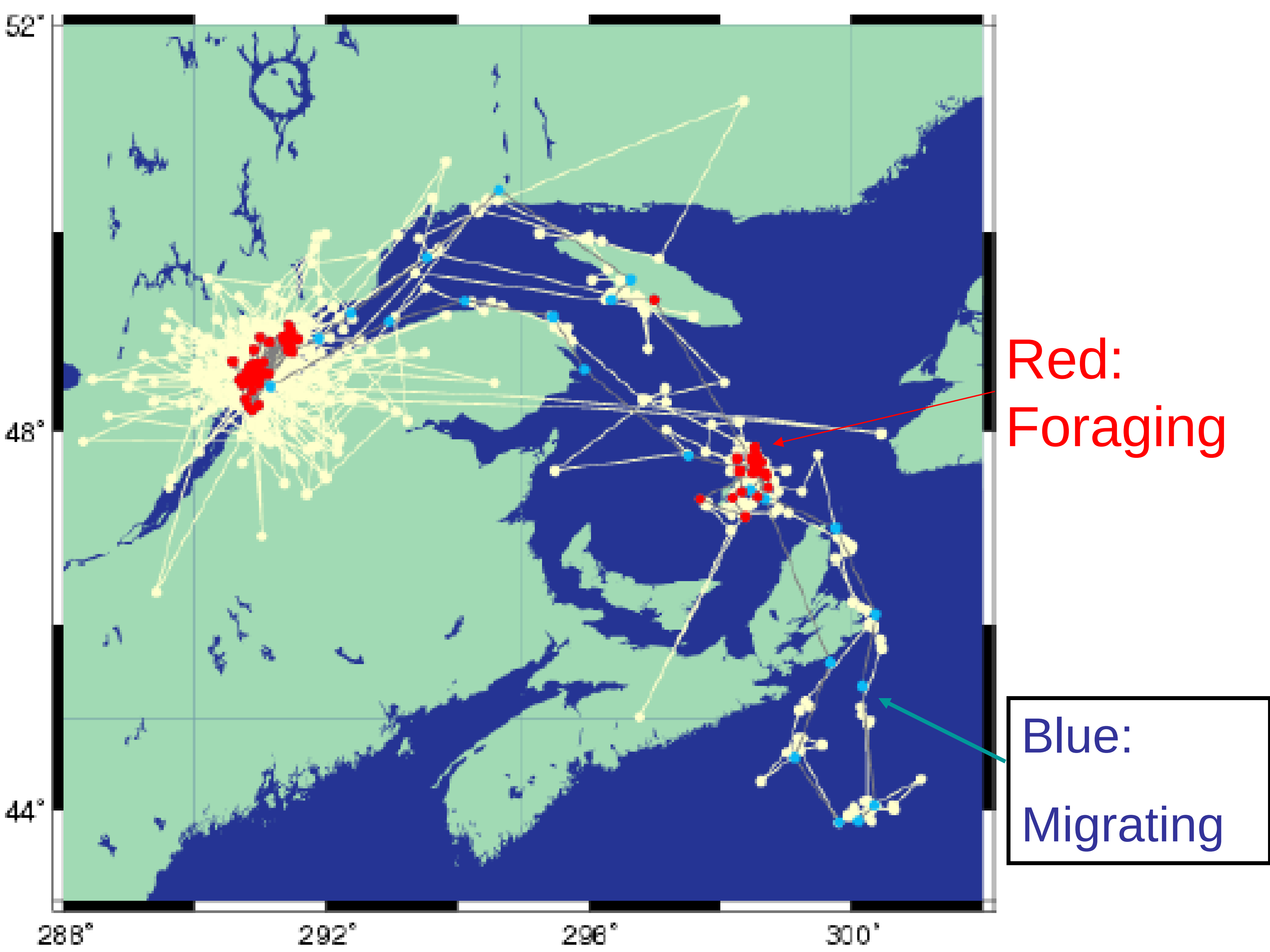
Jonsen, Flemming and Myers (2005) Ecology 86: 2874-2880

Data Filtering & State Estimation

Jonsen et al. 2005. Ecology 86:2874-2880

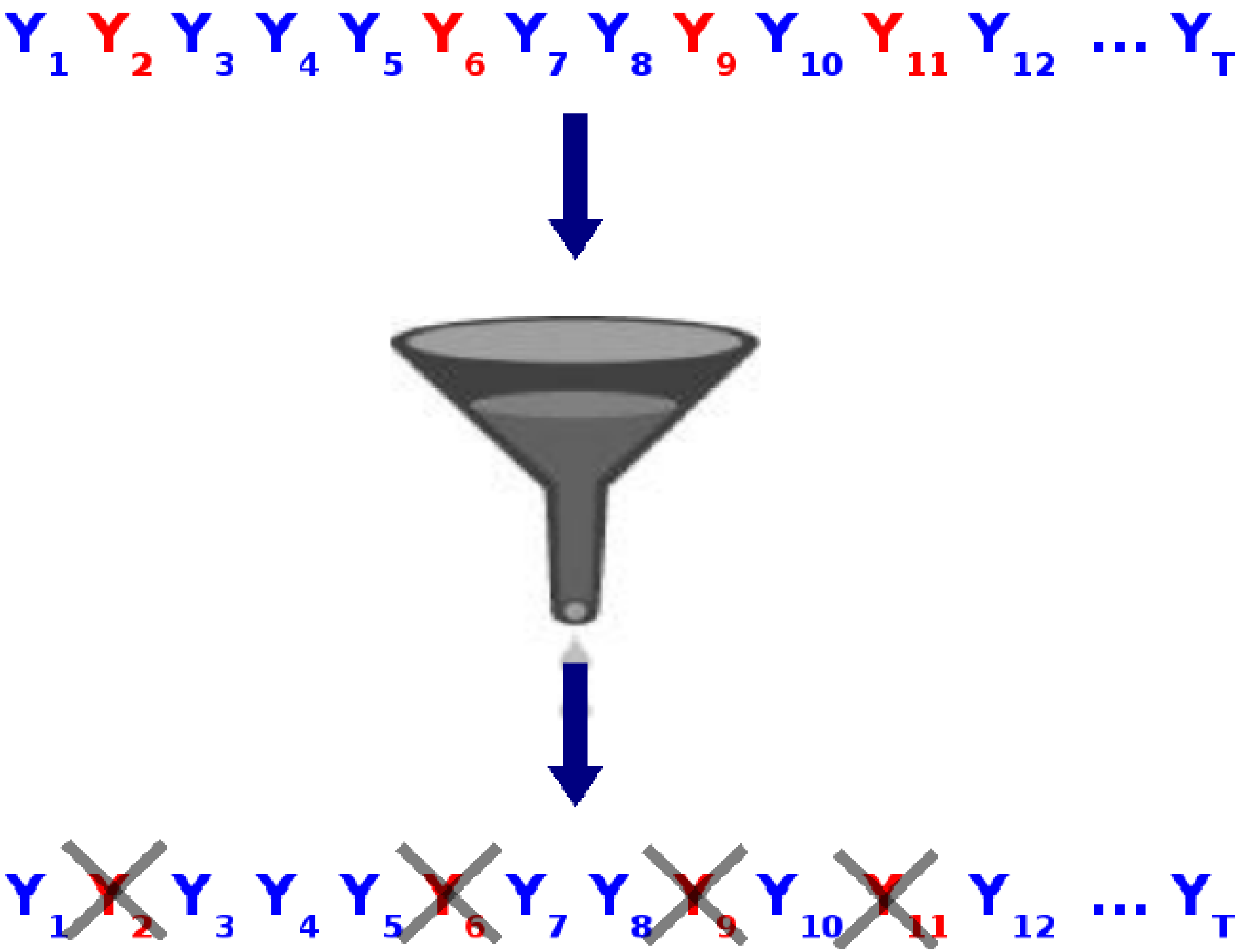


Jonsen, Flemming and Myers (2005) Ecology 86: 2874-2880


52°
48°
44°
288°
292°
296°
300°
Red:
Foraging
Blue:
Migrating

Noisy Data: ad-hoc Filtering

Extreme values removed prior to analysis



State-Space Filtering is Fundamentally Different

Location estimates w CIs & parameter estimation

$$Y_1\ Y_2\ Y_3\ Y_4\ Y_5\ Y_6\ Y_7\ Y_8\ Y_9\ Y_{10}\ Y_{11}\ Y_{12}\ \dots\ Y_T$$

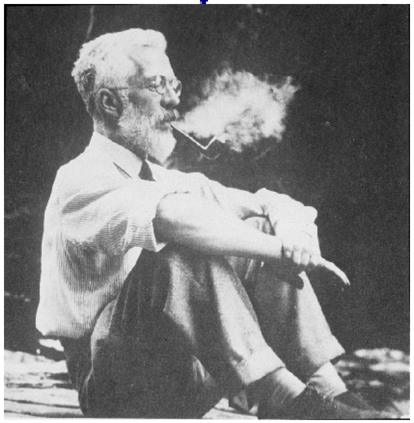
Rev. Thomas Bayes

$$\alpha_1\ \alpha_2\ \alpha_3\ \alpha_4\ \alpha_5\ \dots\ \alpha_T\ ;\ \gamma\ ,\ \sigma\ ,\ \tau$$

State-Space Filtering is Fundamentally Different

Location estimates w CIs & parameter estimation

$Y_1\ Y_2\ Y_3\ Y_4\ Y_5\ Y_6\ Y_7\ Y_8\ Y_9\ Y_{10}\ Y_{11}\ Y_{12}\ \ldots\ Y_T$



We also carried out likelihood analysis

This is Sir Ronald Fisher

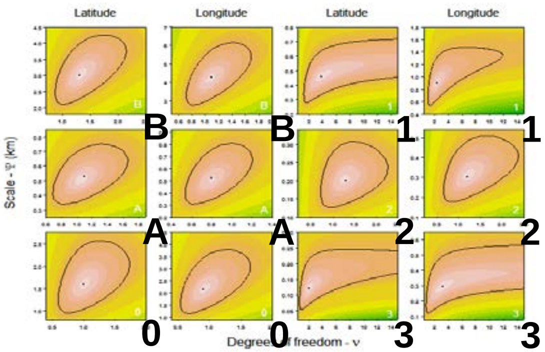
$\alpha_1\ \alpha_2\ \alpha_3\ \alpha_4\ \alpha_5\ \ldots\ \alpha_T\ ;\ \gamma\ ,\ \sigma\ ,\ \tau$

State-Space Filtering is Fundamentally Different

Location estimates w CIs & parameter estimation

Y_1 Y_2 Y_3 Y_4 Y_5 Y_6 Y_7 Y_8 Y_9 Y_{10} Y_{11} Y_{12} ... Y_T

Priors



Rev. Thomas Bayes

Meta-analysis of
data from other animals

Y_1 Y_2 Y_3 Y_4 Y_5 Y_6 Y_7 Y_8 Y_9 Y_{10} Y_{11} Y_{12} ... Y_T

Y_1 Y_2 Y_3 Y_4 Y_5 Y_6 Y_7 Y_8 Y_9 Y_{10} Y_{11} Y_{12} ... Y_T

Y_1 Y_2 Y_3 Y_4 Y_5 Y_6 Y_7 Y_8 Y_9 Y_{10} Y_{11} Y_{12} ... Y_T

α_1 α_2 α_3 α_4 α_5 ... α_T ; γ , σ , τ

State Space Models

Measurement Equation

- Relates imperfect observations to true position
- Estimate ARGOS error with ε

$$y_t = h\,(\alpha_t\,,\,\varepsilon_t)$$

Error function

observed location

true location

Transition Equation

- Predicts next position from behavioral model

$$\alpha_t = f\,(\alpha_{t\text{-}1},\ \eta_t\,;\,\gamma)$$

movement function

parameters

What is important about the state equation?

$$\alpha_t = f\ (\alpha_{t-1},\ \eta_t\ ;\ \gamma)$$

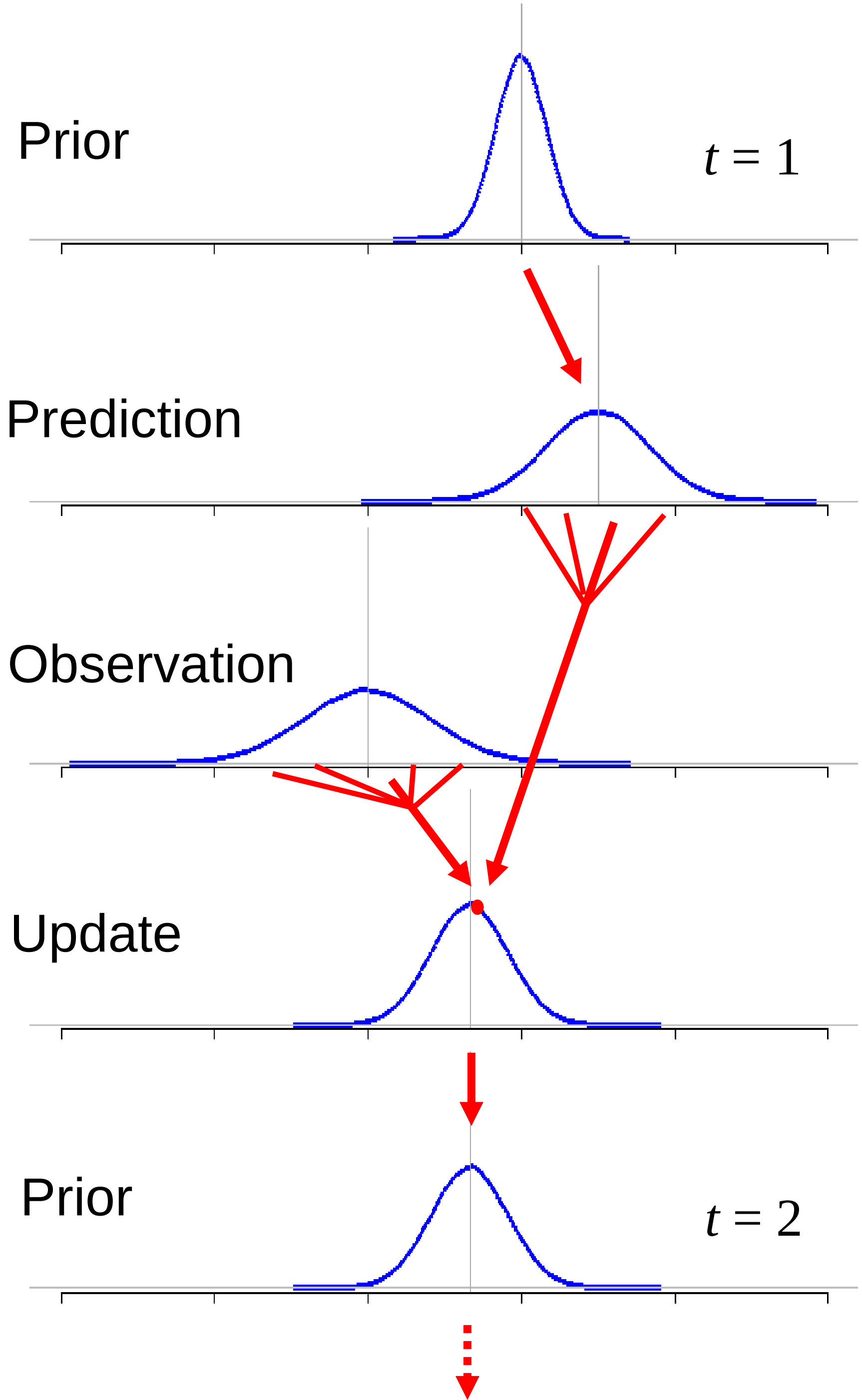
The state variable is random variable, and should NOT be thought of as a simple number. We write the state as a Greek letter, α_t

How is this programmed with BUGS

mean[t] <- *f*(*alpha[t-1]* ; γ)

alpha[t] ~ dlnorm (*mean[t]*, *sigma*)

This symbol means “is distributed as”, and implies that alpha[t] is a random variable.



1[st] location = release point

eg. release location estimated with GPS

Apply dynamics (transition eqn)

Observe a location with error

Integrate over predicted & observed densities (Bayes Rule)

Updated prediction becomes prior for next time step

Software

WinBUGS: Bayesian Analysis Using Gibbs Sampling

Bayes Rule

$$p(\alpha_t \mid \mathrm{Y}_t;\gamma) = \frac{p_y(\mathbf{y}_t \mid \alpha_t)\, p(\alpha_t \mid \mathrm{Y}_{t-1};\gamma)}{\int p_y(\mathbf{y}_t \mid \alpha_t)\, p(\alpha_t \mid \mathrm{Y}_{t-1};\gamma)\, d\mathbf{x}_t}$$

This is the innovation likelihood

Movement (Transition) Equation

- First difference correlated random walk

$$d_t = \gamma T(\theta) d_{t-1} + N_2(0,\Sigma)$$
$$\alpha_t = \alpha_{t-1} + d_t$$

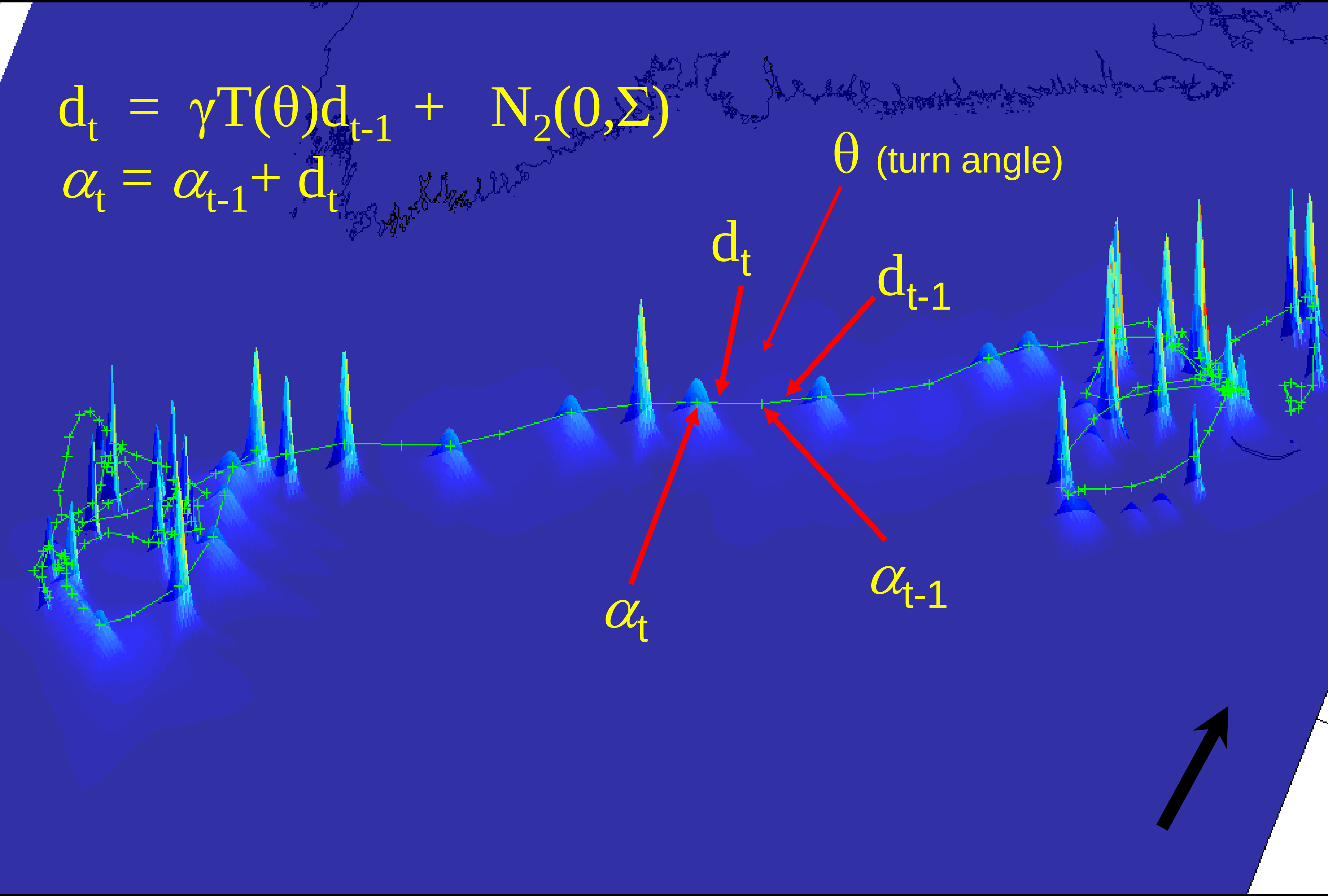
Observation Equation

$$y_t = \textit{t-distribution}\ (\alpha_t\ ,\ \sigma_{t,}\ \upsilon_t\ )$$

Plus an algorithm to regularize estimated locations in time

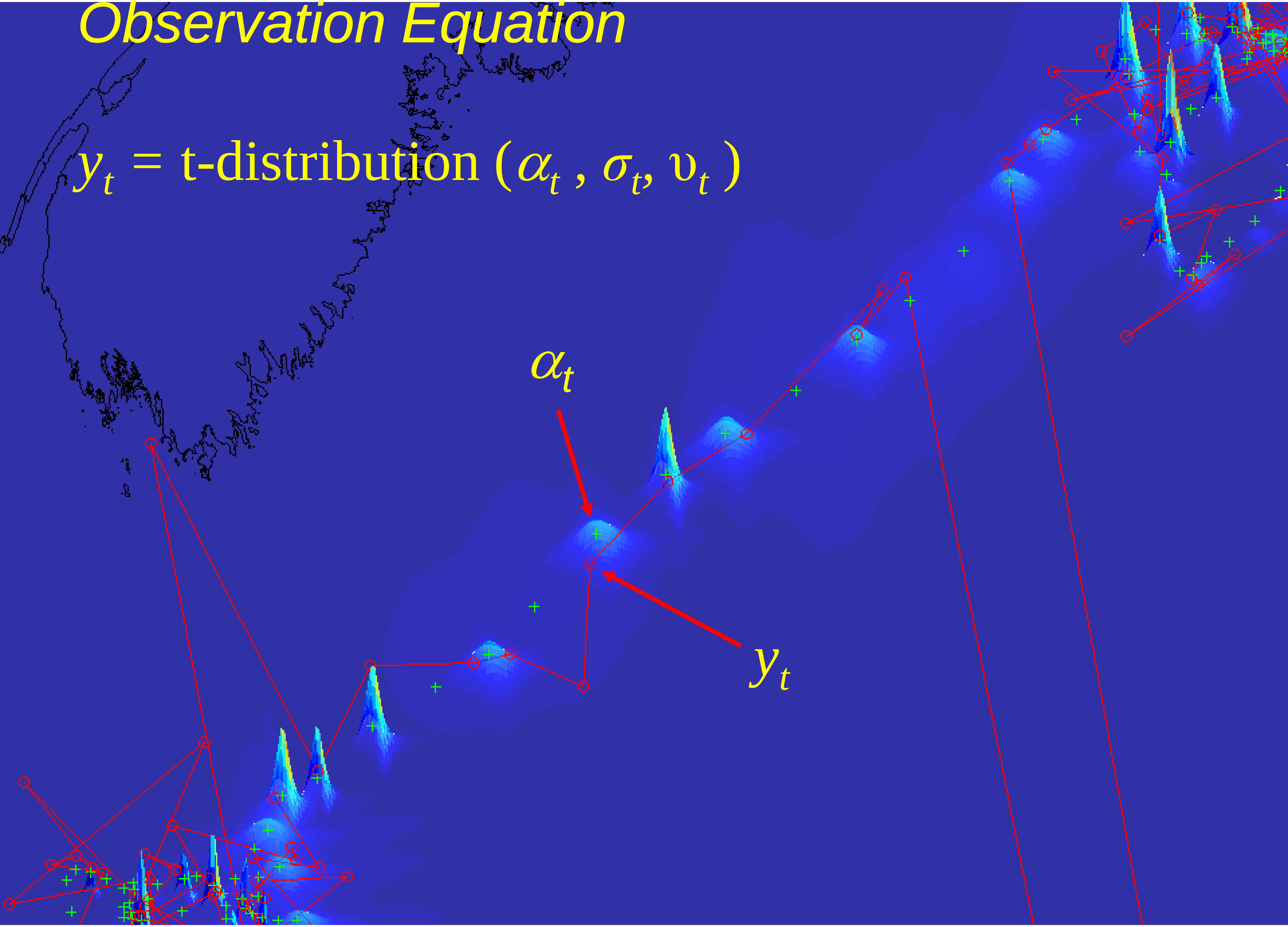
Movement (Transition) Equation

$$d_t = \gamma T(\theta) d_{t-1} + N_2(0,\Sigma)$$
$$\alpha_t = \alpha_{t-1} + d_t$$



Observation Equation

$$y_t = \text{t-distribution}\,(\alpha_t\,,\,\sigma_t,\,\upsilon_t\,)$$



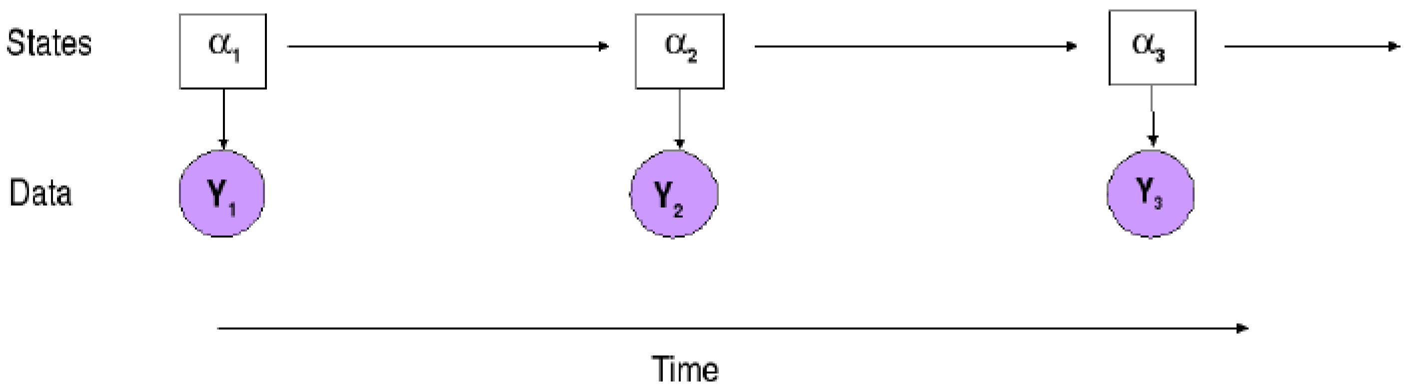
Dealing with Complex Data

Argos telemetry data

- Observations made irregularly through time
- Observation errors non-Gaussian (extreme values)
- Observation errors vary through time

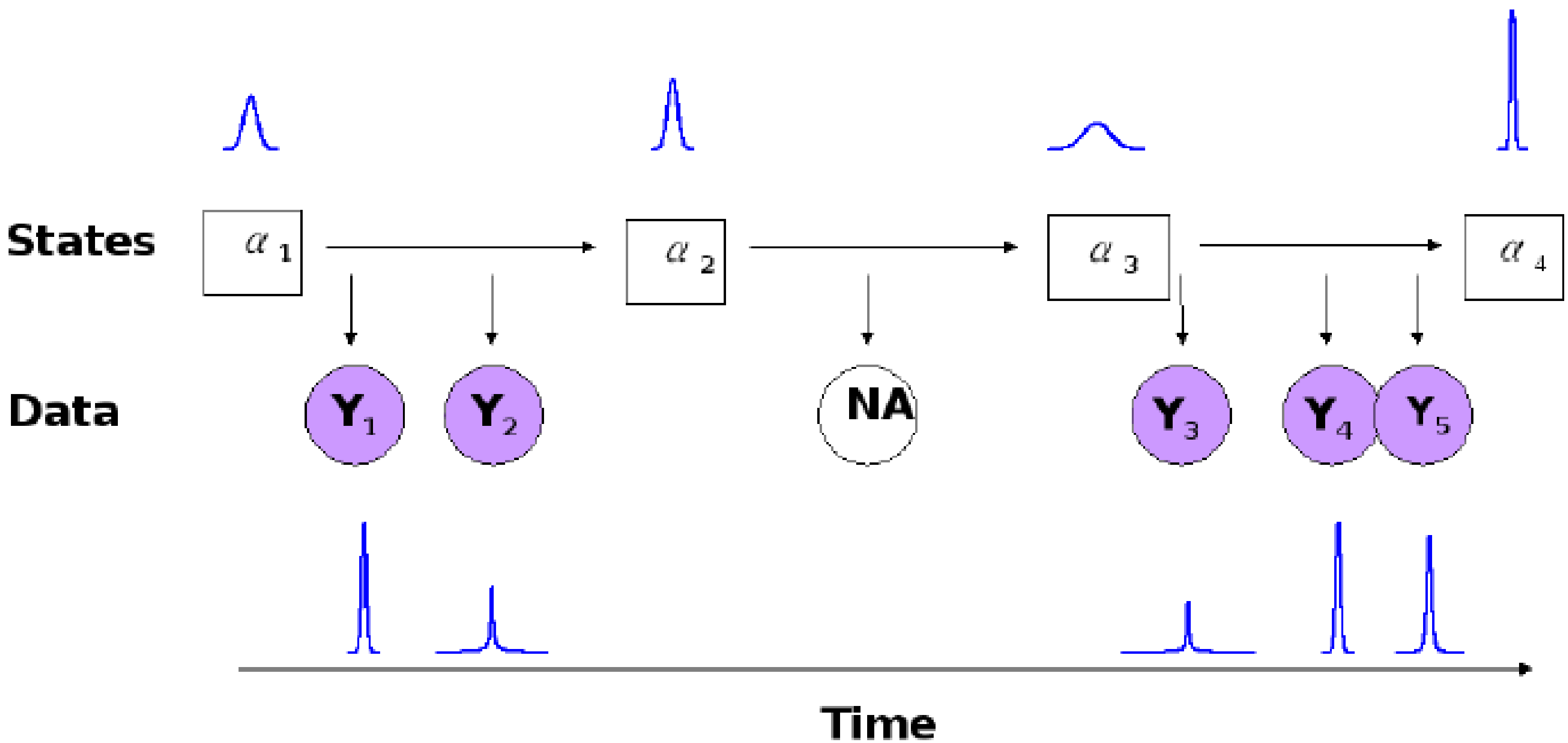
“Ideal” Data

Regular time intervals with constant Gaussian errors, & no missing data



“Real” Data

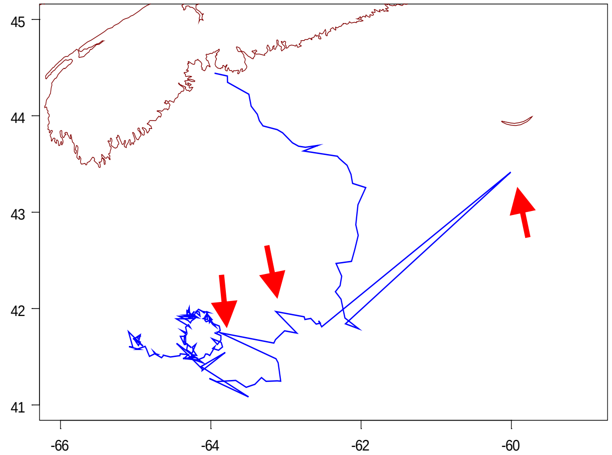
Irregular time intervals with varying non-Gaussian errors, & missing data

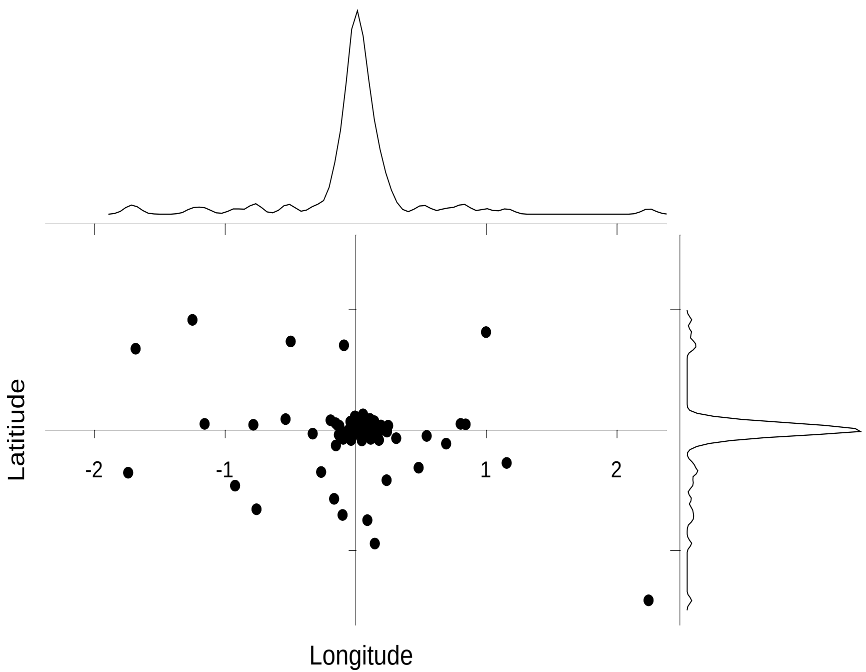


Sources of Uncertainty

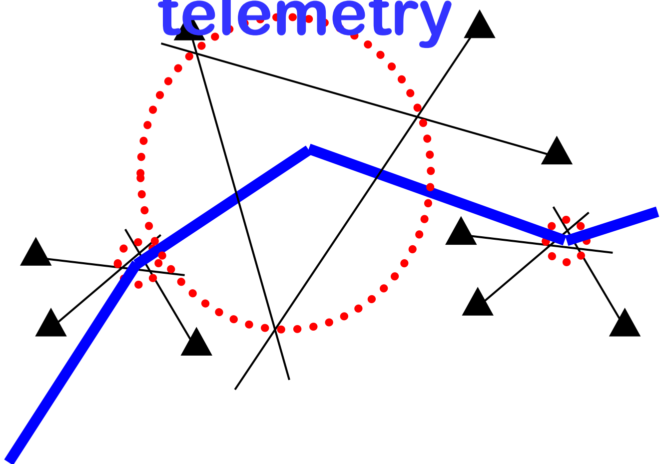
- Estimation error
  - Data observed with error
  - Errors can be non-Gaussian

Satellite telemetry

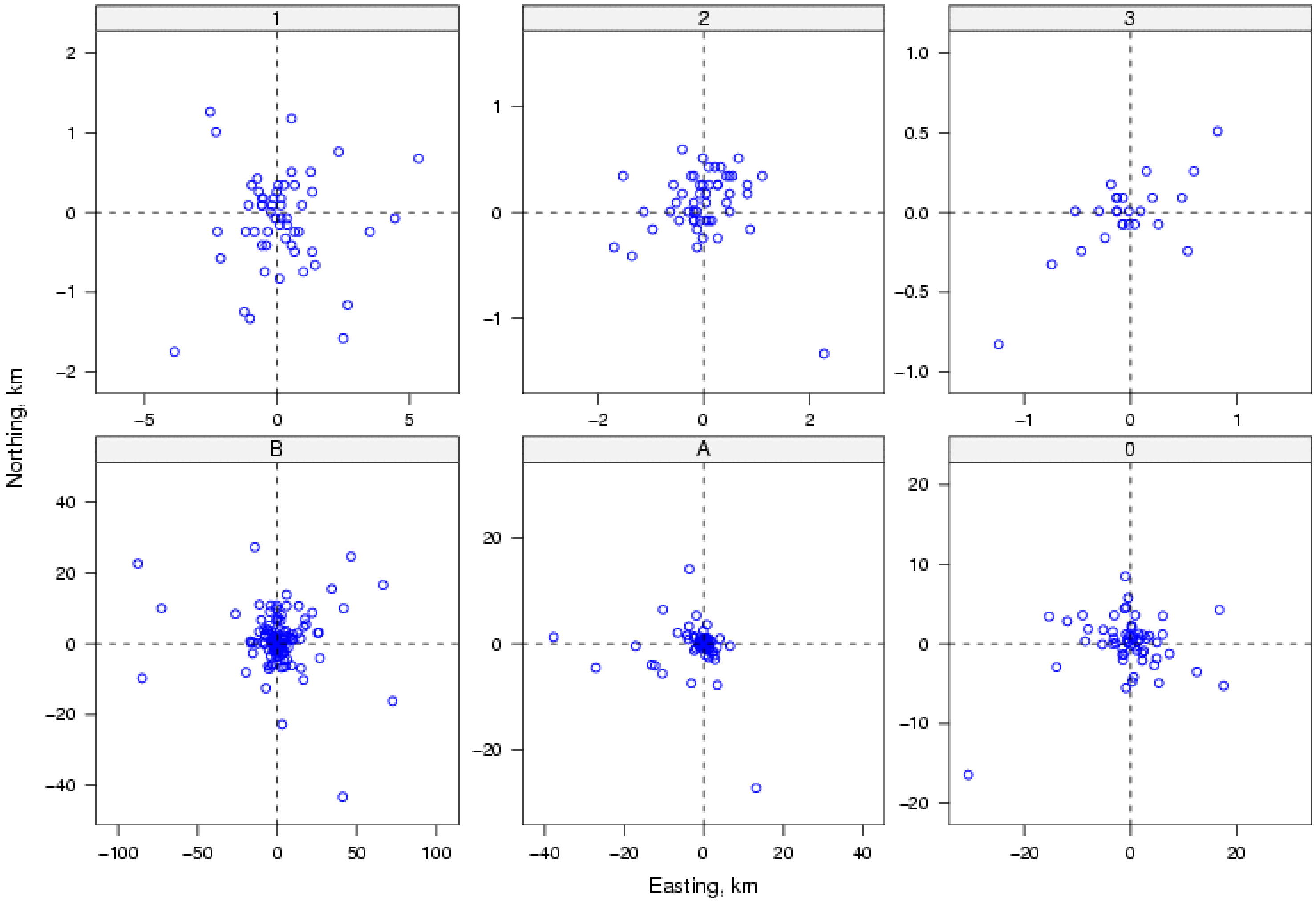




Radio or acoustic telemetry

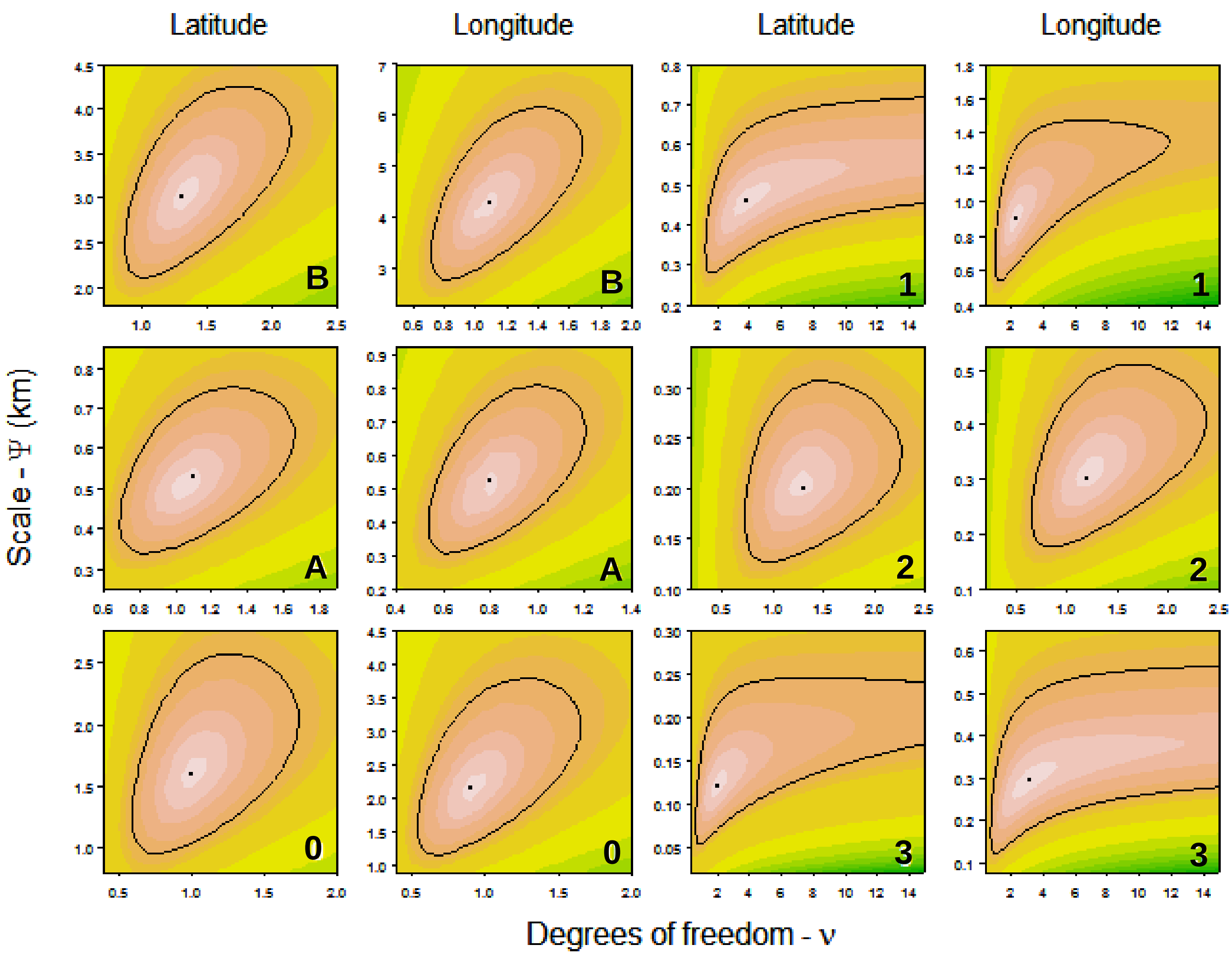


Argos location errors

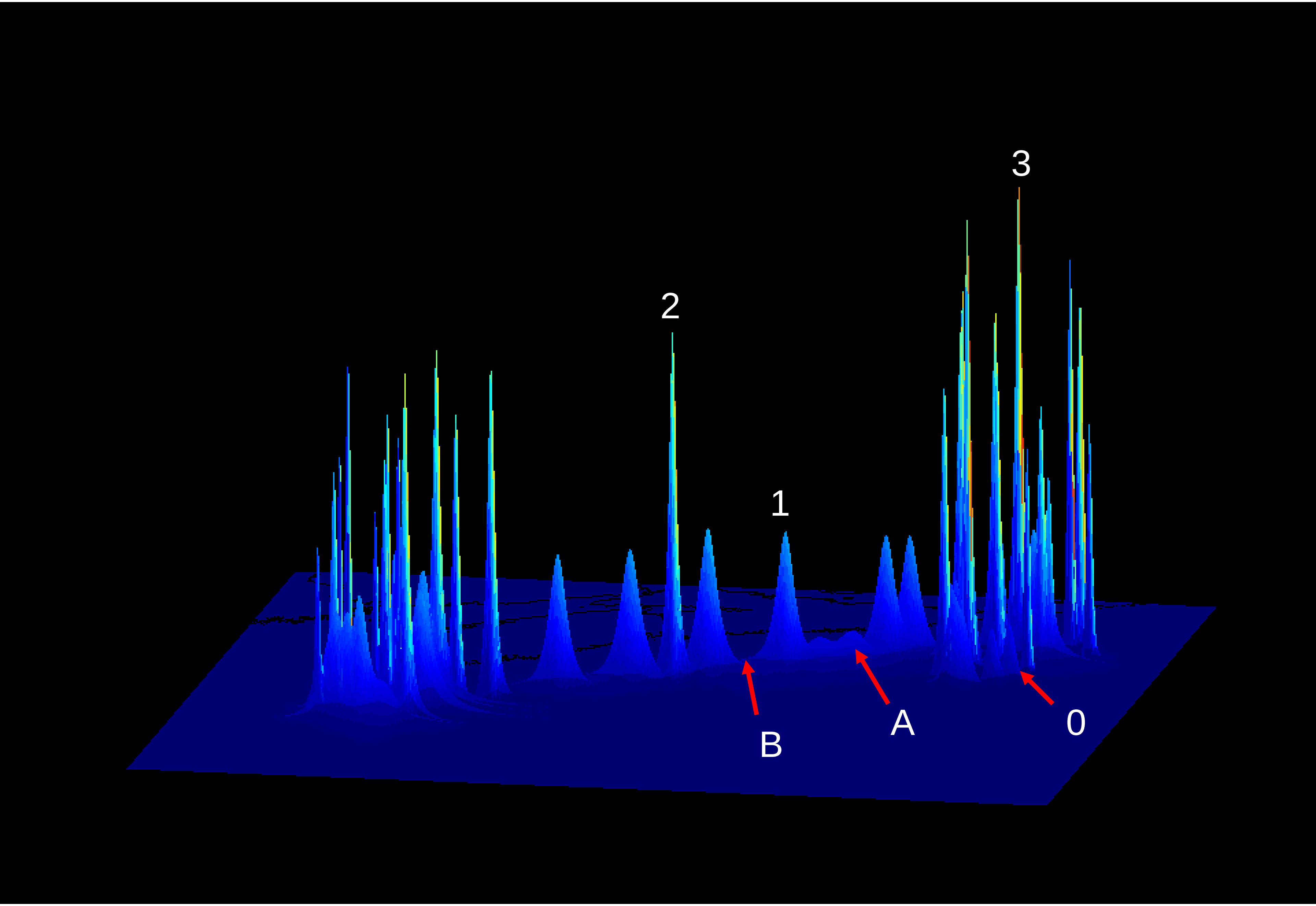


data from Vincent et al. 2002

Argos errors follow *t*-distributions:

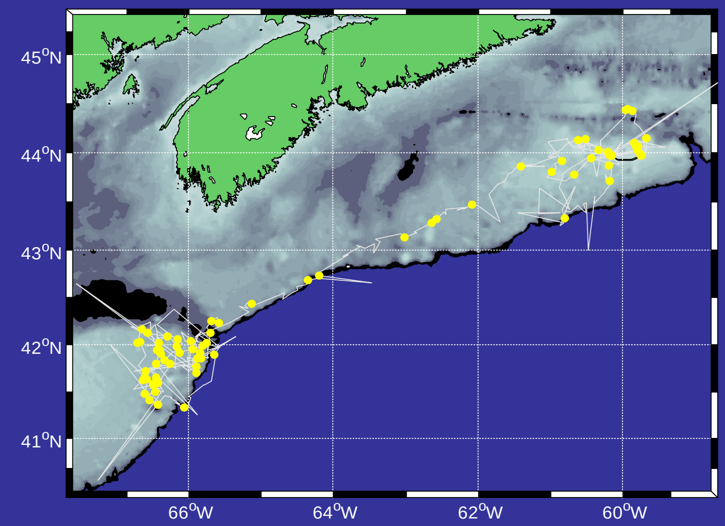


Jonsen, Flemming, Myers, Ecology, 2005


3
2
1
B
A
0

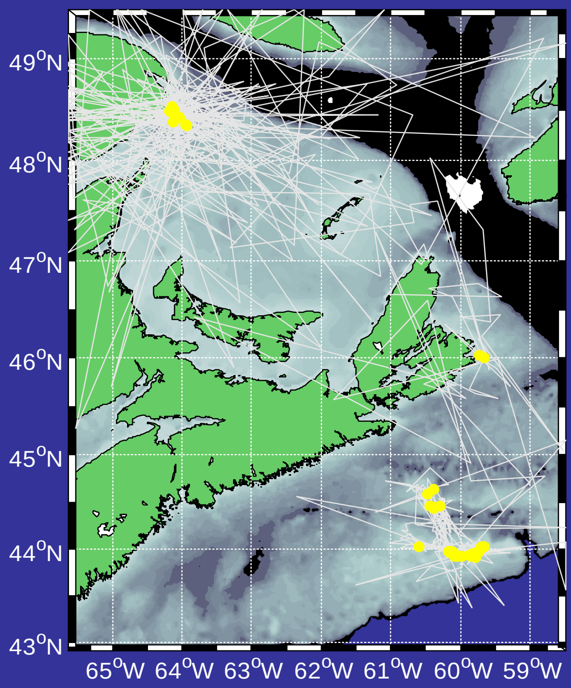
Tag Precision

$$y_t = \text{t-distribution}\,(\alpha_t\ ,c\sigma_t,\ \upsilon_t\ )$$



σ for each location class is assumed known (Vincent et al. 2002, Jonsen et al 2005)

c is an estimated parameter that scales the variance to each tag



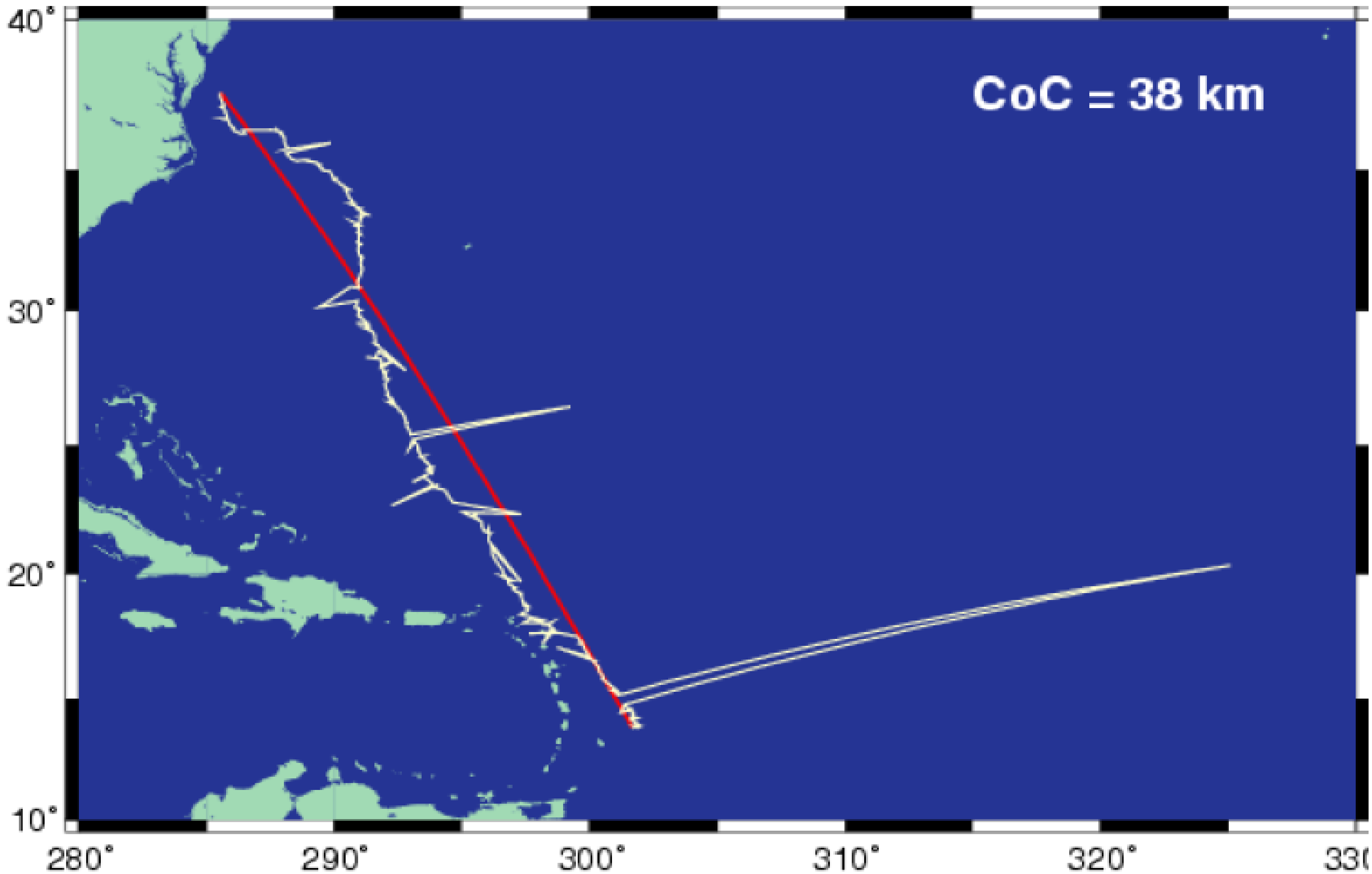
Best tags are **1000's** of times more precise than the worst tags

Best tags in a single lot can be **100's** of times more precise than worst tags from a lot

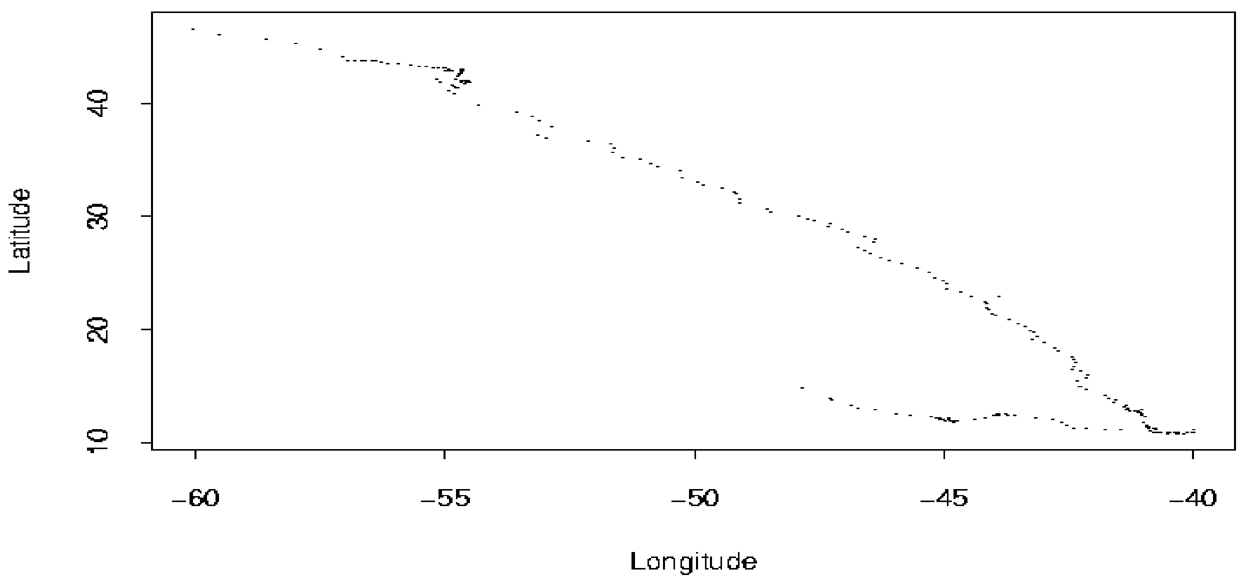
State-space models allow you to think about things, that it is very difficult to think about otherwise

Navigation: Estimating the "Circle of Confusion"

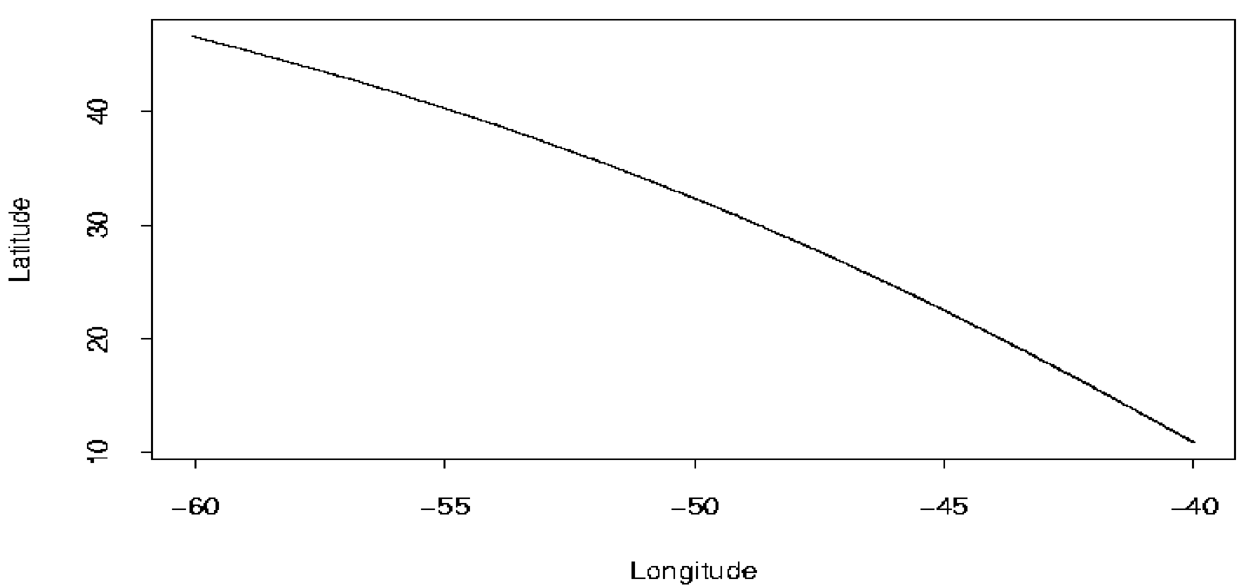
Flemming et al. in press. Environmetrics



Regularized Track of Turtle 18284

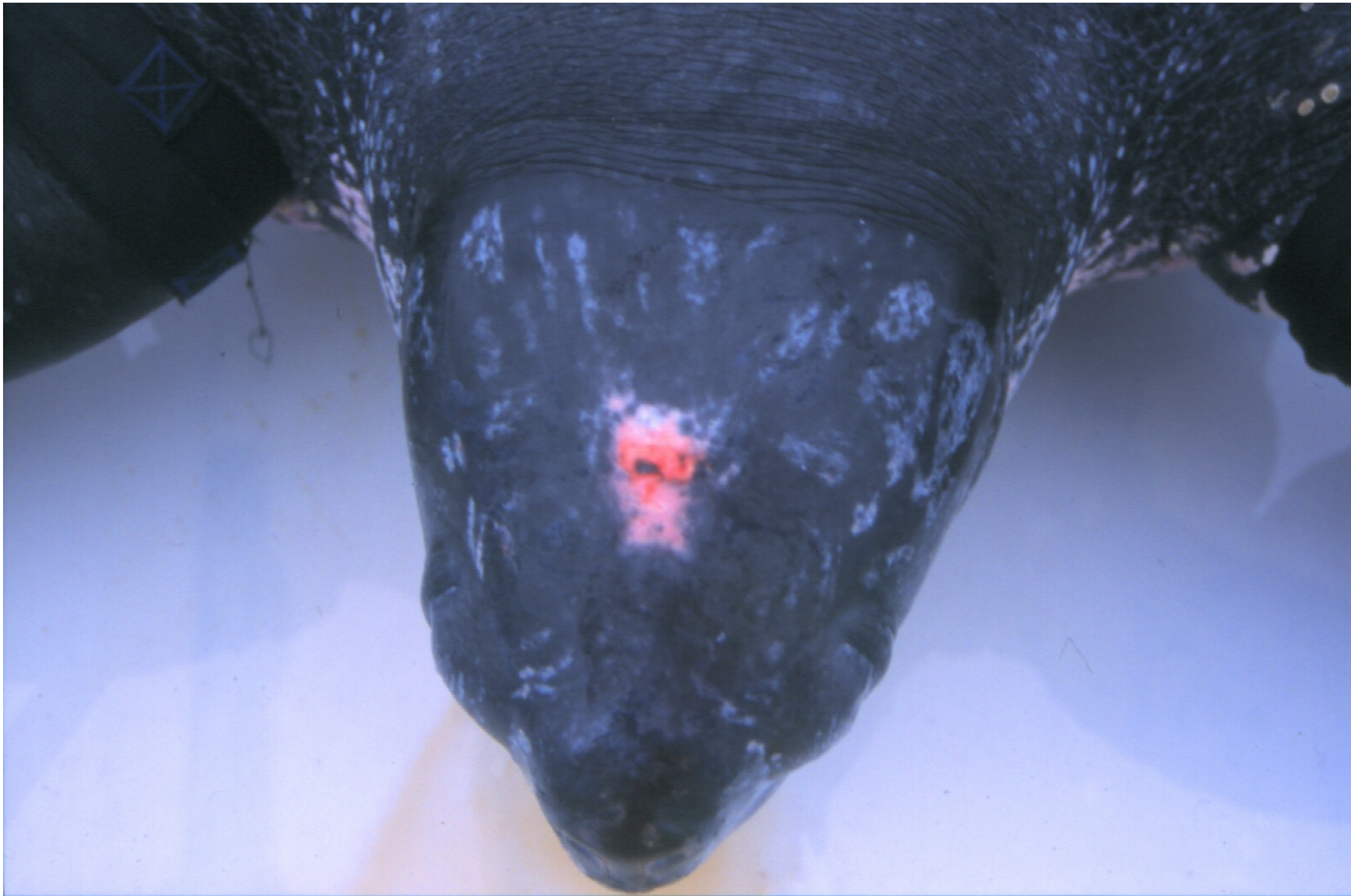


Corresponding GC Route

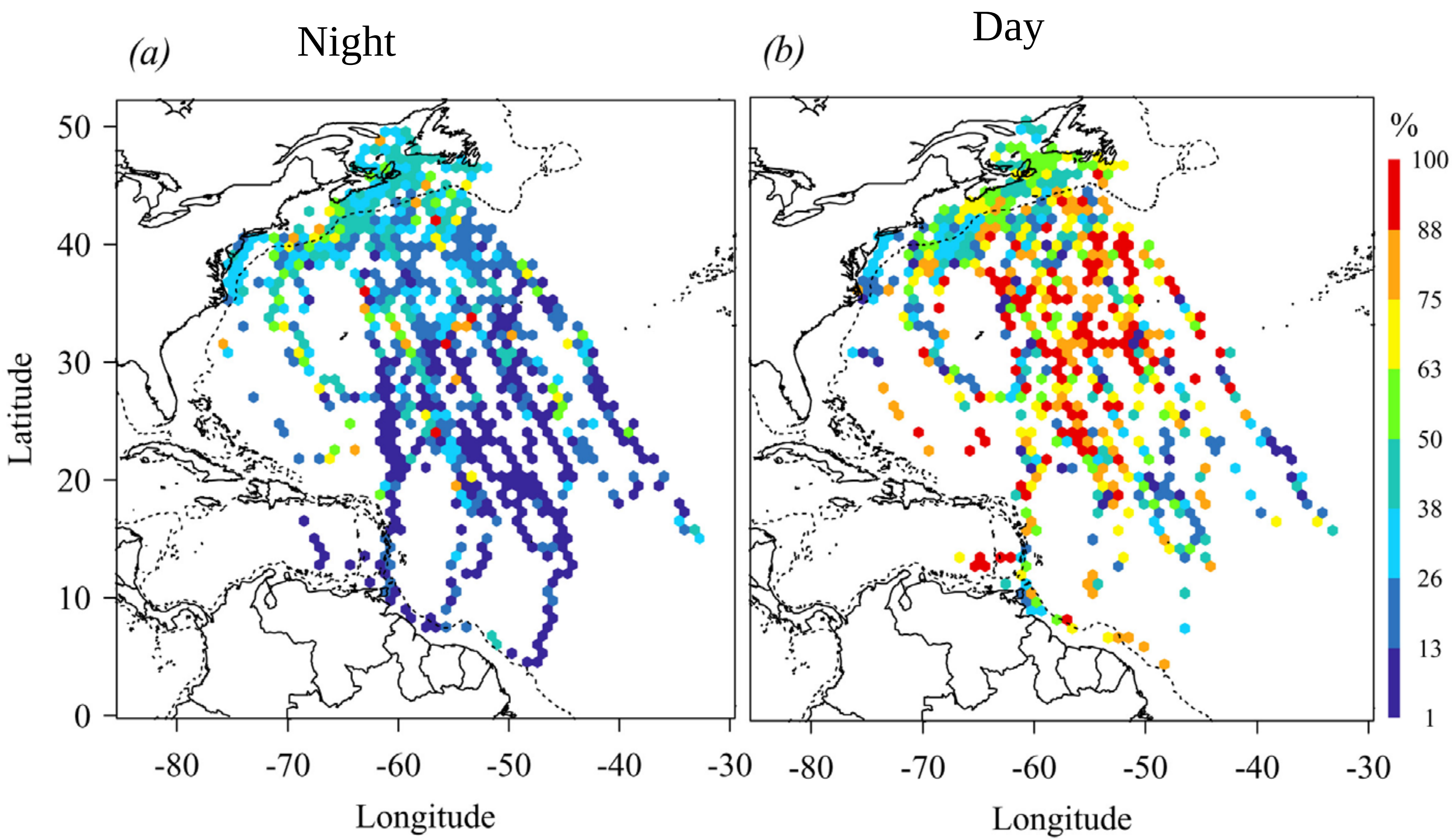


It is essential to treat groups of animals simultaneously for maximum utility of the data.

Leatherback turtles are unique in that they expose their pineal spot to sunlight.

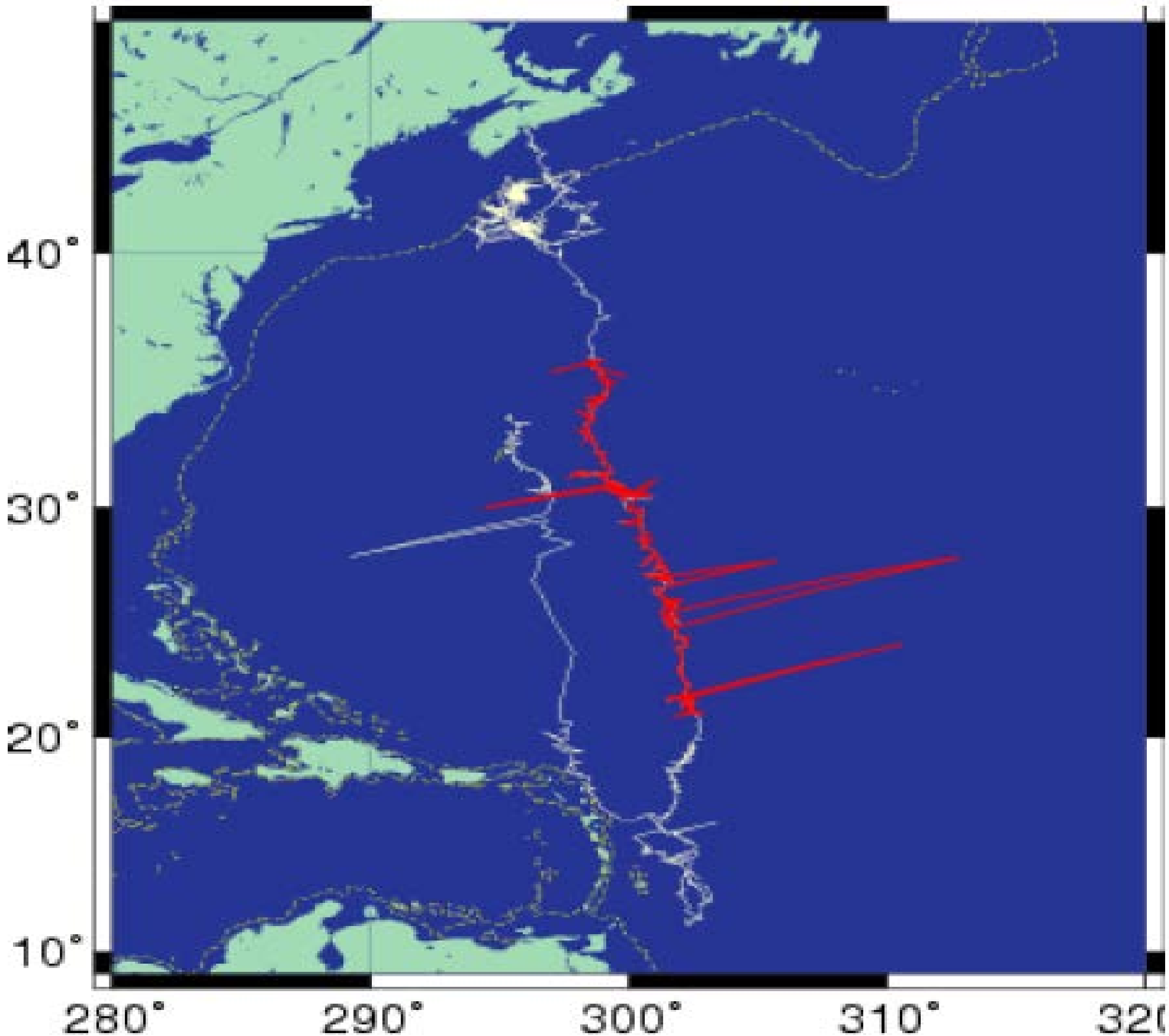


Turtles are close to the surface during the day during migration

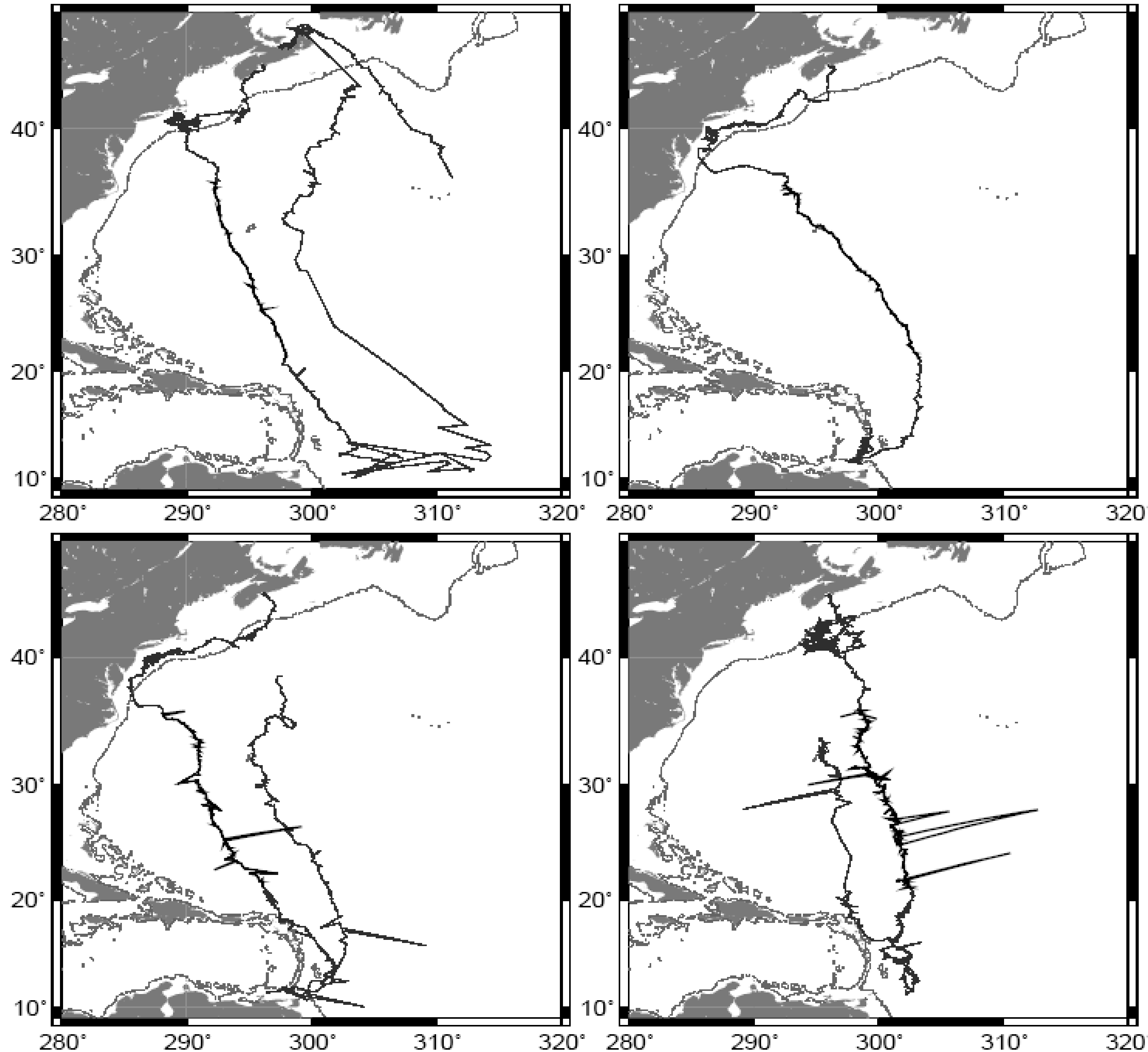


James Ottensmeyer Eckert Myers in press Can. J. of Zoo.

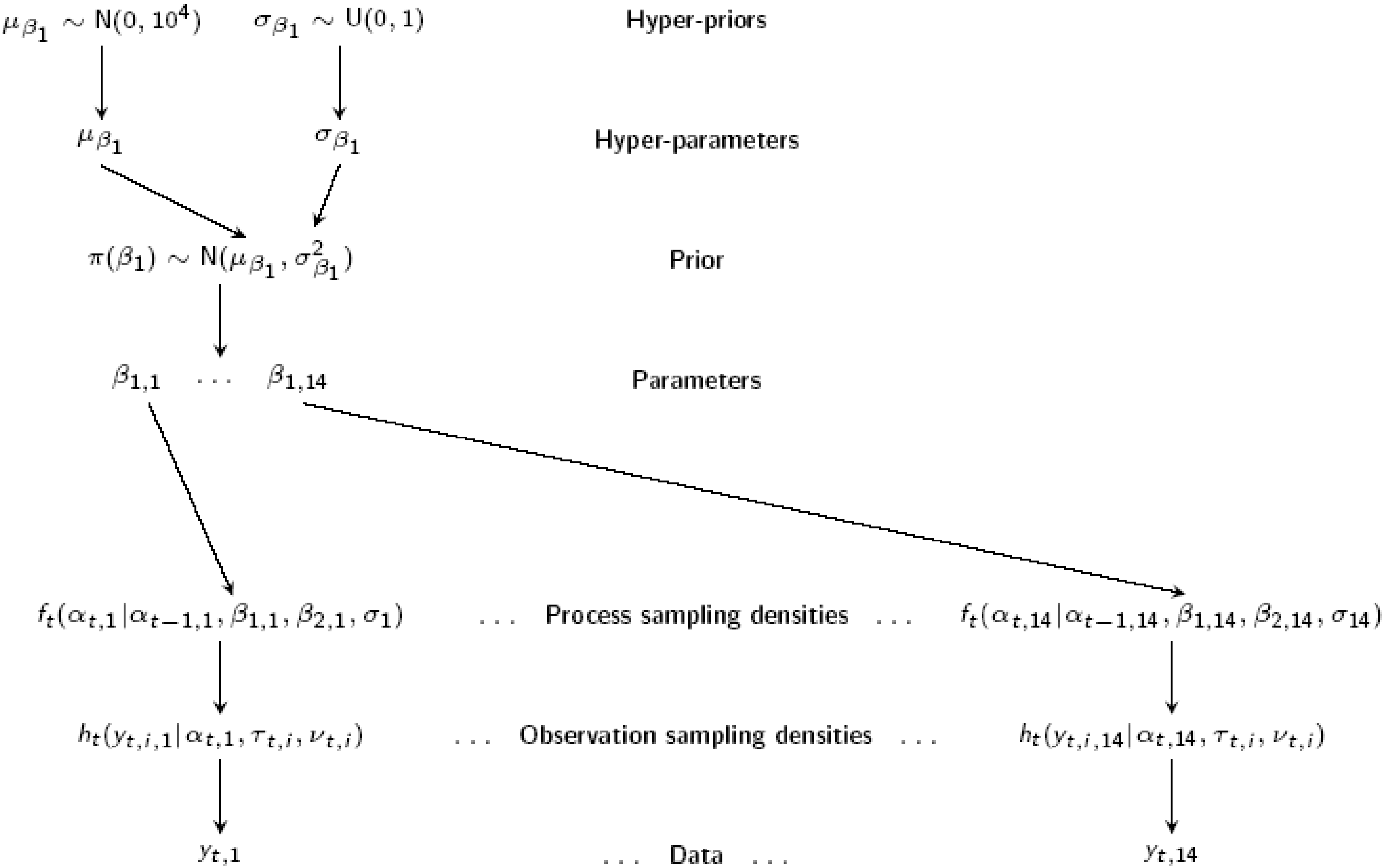
Examining Diel Migration Behaviour in Leatherbacks



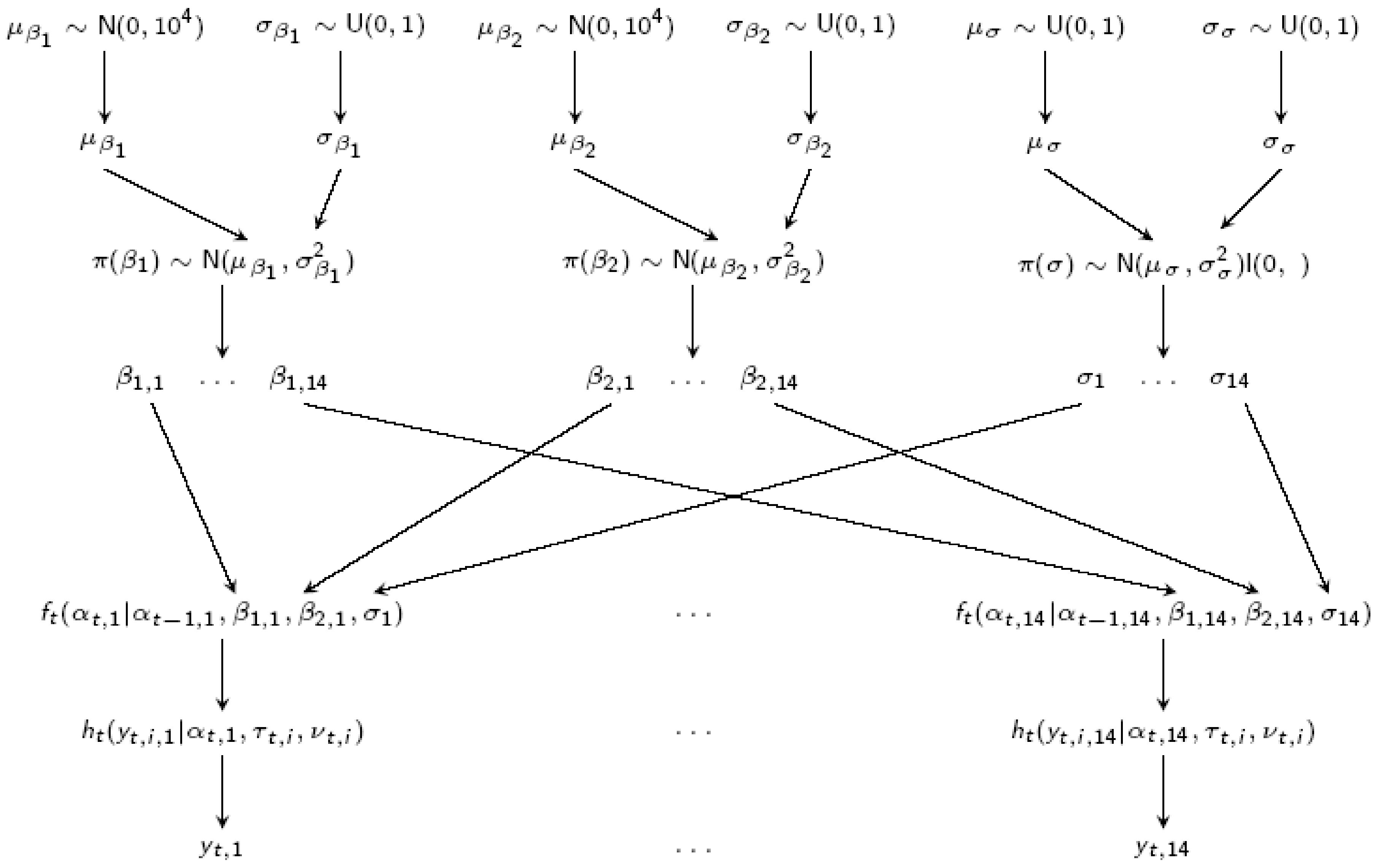
Jonsen, James Myers. in press. Journal of Animal Ecology


40°
30°
20°
10°
280°
290°
300°
310°
320°

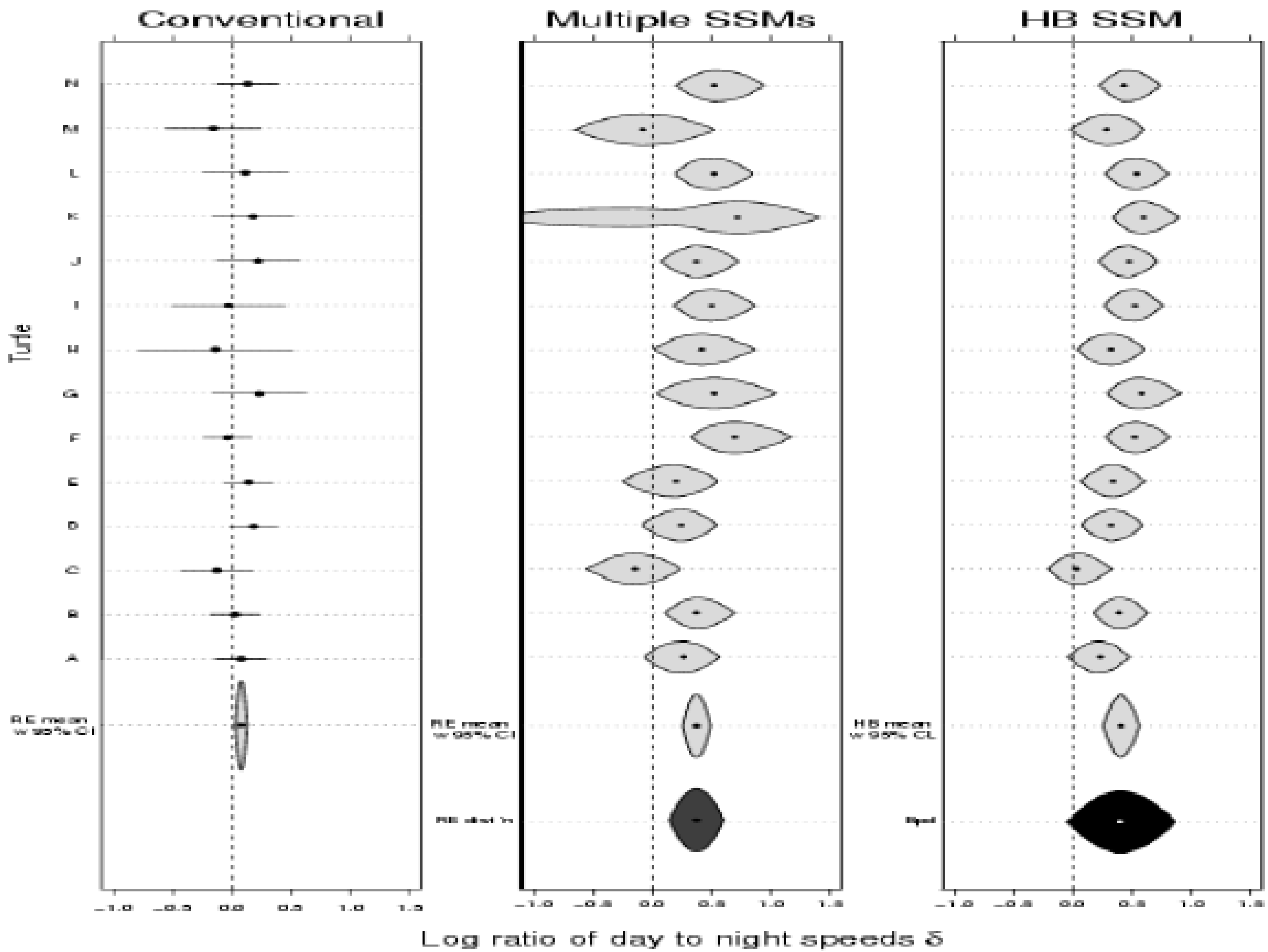
Hierarchical Bayes State-Space Model (HB SSM)



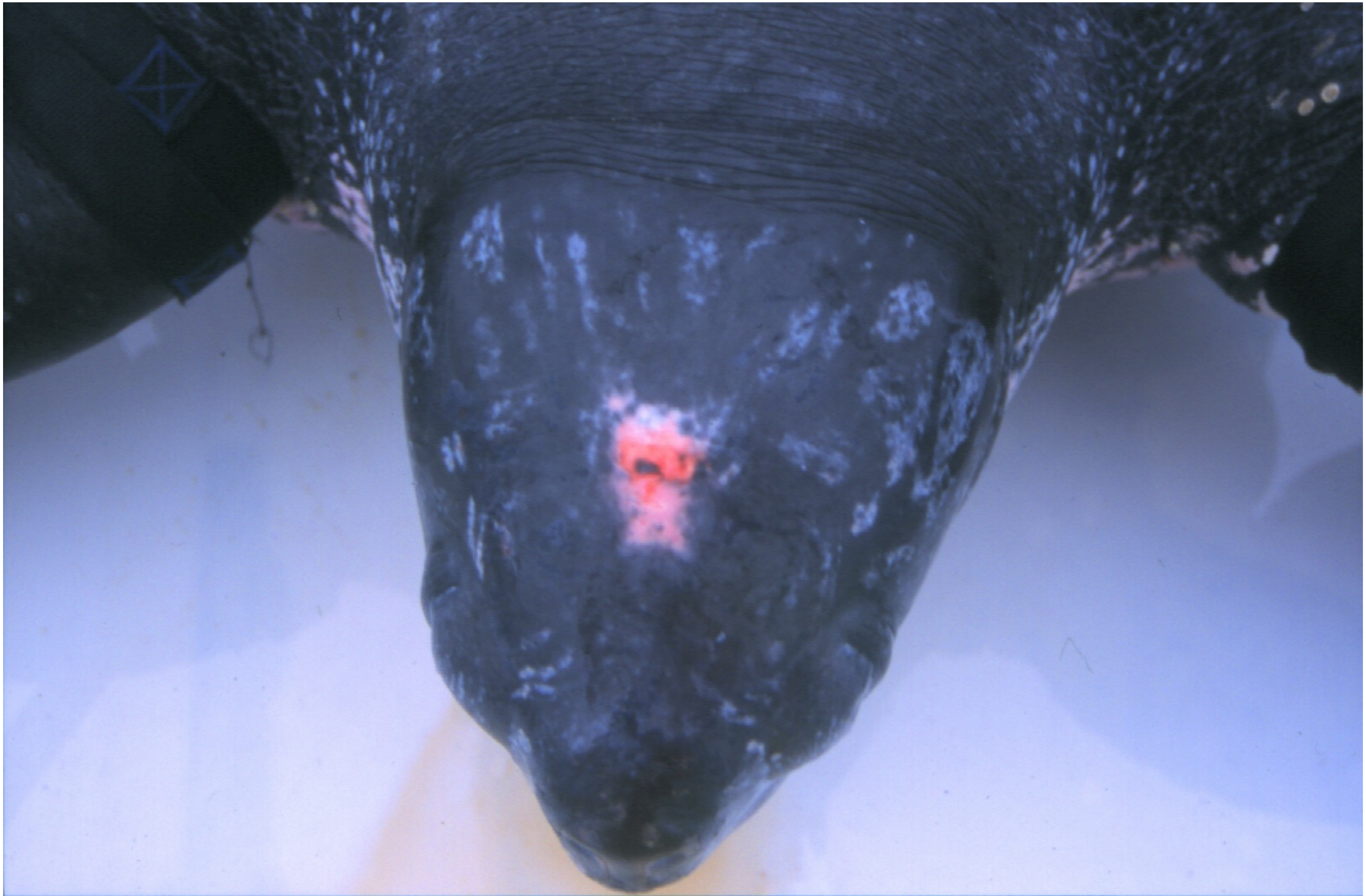
HB SSM


$\mu_{\beta_1} \sim N(0, 10^4)$
$\sigma_{\beta_1} \sim U(0, 1)$
$\mu_{\beta_2} \sim N(0, 10^4)$
$\sigma_{\beta_2} \sim U(0, 1)$
$\mu_\sigma \sim U(0, 1)$
$\sigma_\sigma \sim U(0, 1)$
μ_{β_1}
σ_{β_1}
μ_{β_2}
σ_{β_2}
μ_σ
σ_σ
$\pi(\beta_1) \sim N(\mu_{\beta_1}, \sigma^2_{\beta_1})$
$\pi(\beta_2) \sim N(\mu_{\beta_2}, \sigma^2_{\beta_2})$
$\pi(\sigma) \sim N(\mu_\sigma, \sigma^2_\sigma)I(0,)$
$\beta_{1,1} \cdots \beta_{1,14}$
$\beta_{2,1} \cdots \beta_{2,14}$
$\sigma_1 \cdots \sigma_{14}$
$f_t(\alpha_{t,1} | \alpha_{t-1,1}, \beta_{1,1}, \beta_{2,1}, \sigma_1)$
$\cdots$
$f_t(\alpha_{t,14} | \alpha_{t-1,14}, \beta_{1,14}, \beta_{2,14}, \sigma_{14})$
$h_t(y_{t,i,1} | \alpha_{t,1}, \tau_{t,i}, \nu_{t,i})$
$\cdots$
$h_t(y_{t,i,14} | \alpha_{t,14}, \tau_{t,i}, \nu_{t,i})$
$y_{t,1}$
$\cdots$
$y_{t,14}$

Conventional Approaches Do Not Work



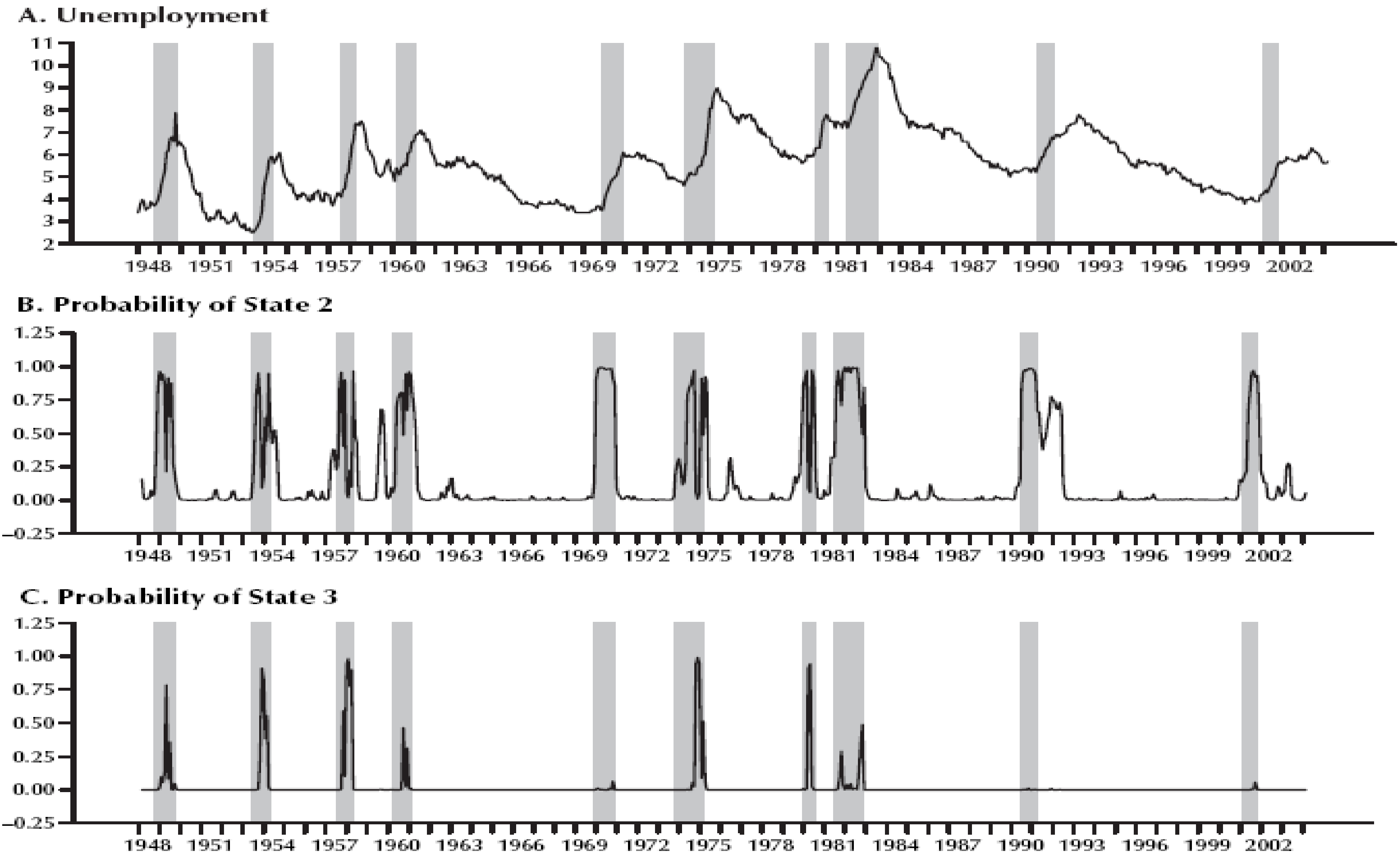
Results are consistent with the hypothesis that the pineal spot improves navigation.



Dynamics of behavior is very nonlinear, to determine hot spots and foraging

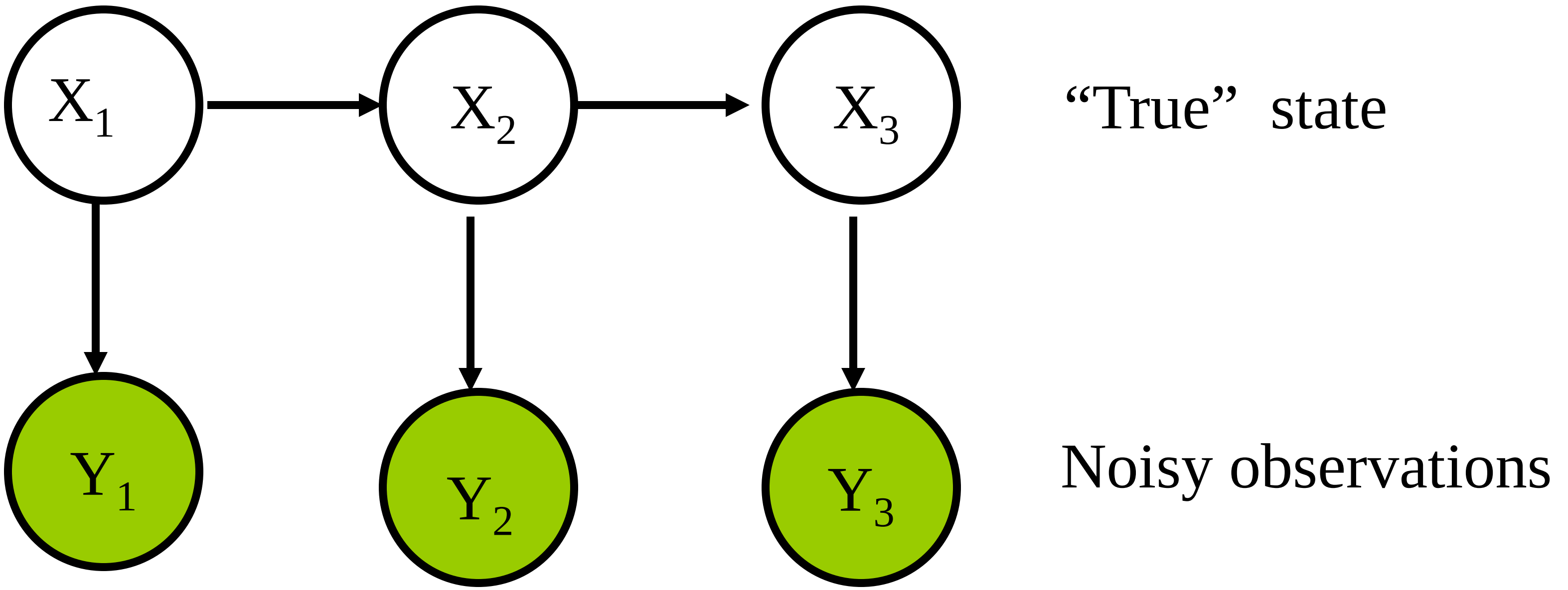
- Solution: Markov switching models between behavioral modes
- Dynamics within a behavioral mode is linear

State-Space Switching Models



Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 435-52.

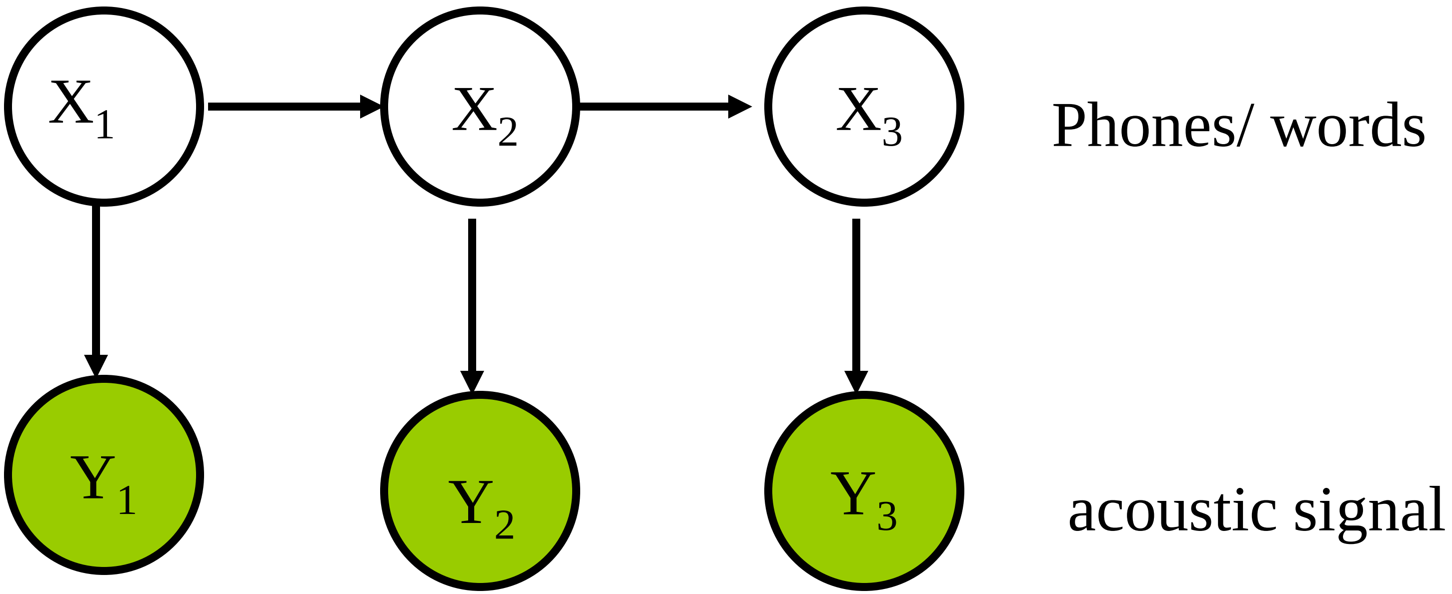
State-space model (SSM)



$$p(X_t | X_{t-1}) = \mathcal{N}(X_t; AX_{t-1}, Q)$$

$$p(Y_t | X_t) = \mathcal{N}(Y_t; BX_t, R)$$

Hidden Markov model (HMM)

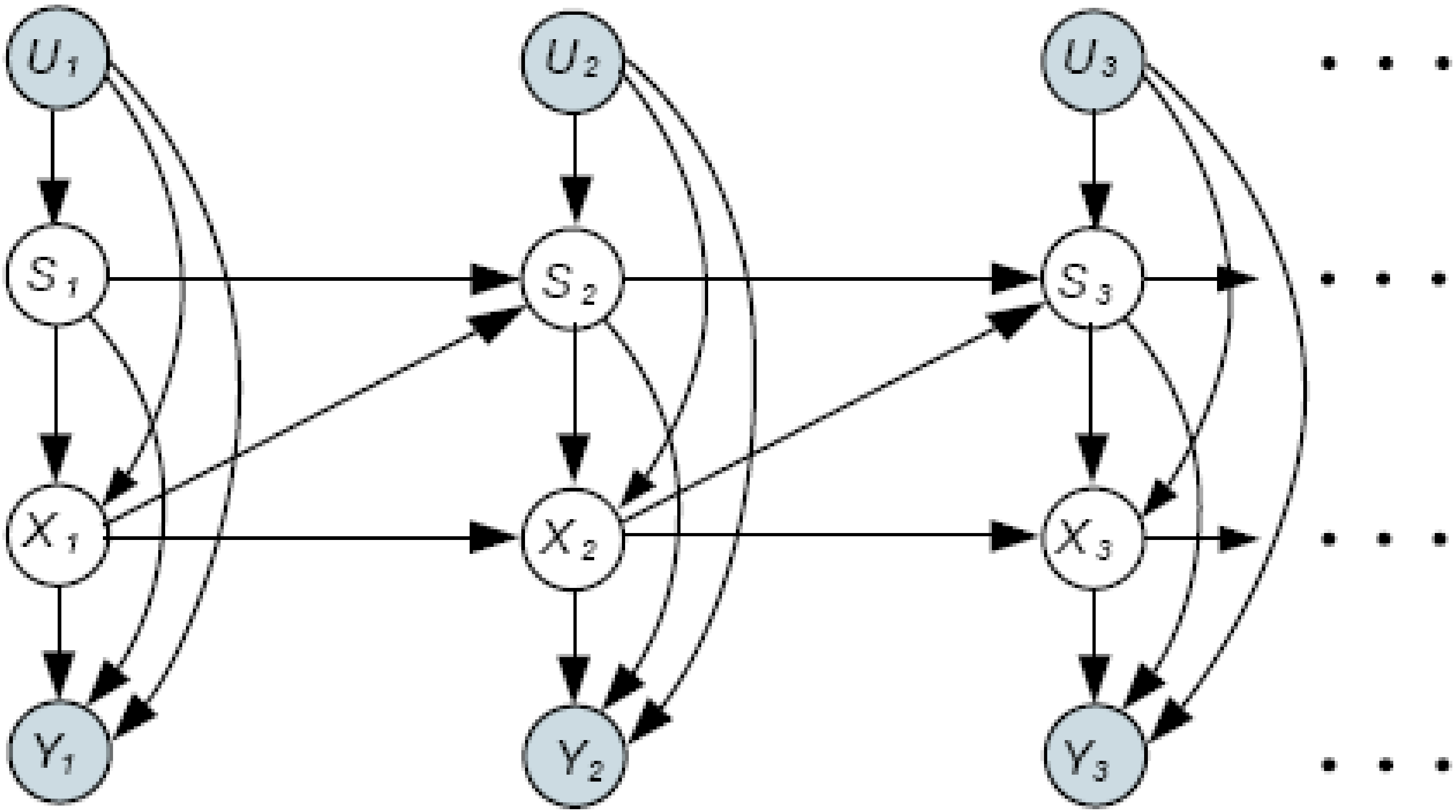


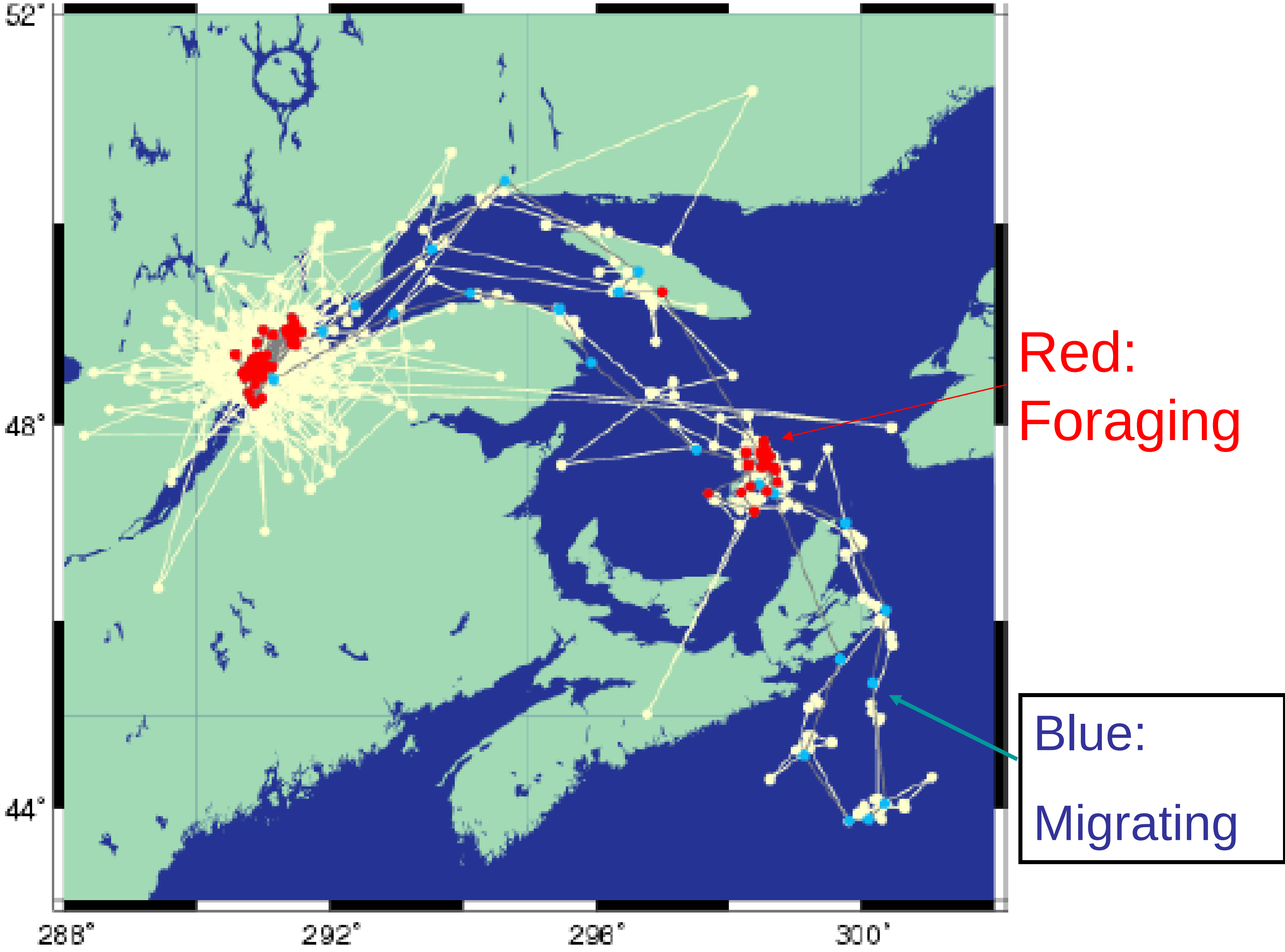
Sparse transition matrix)/sparse graph

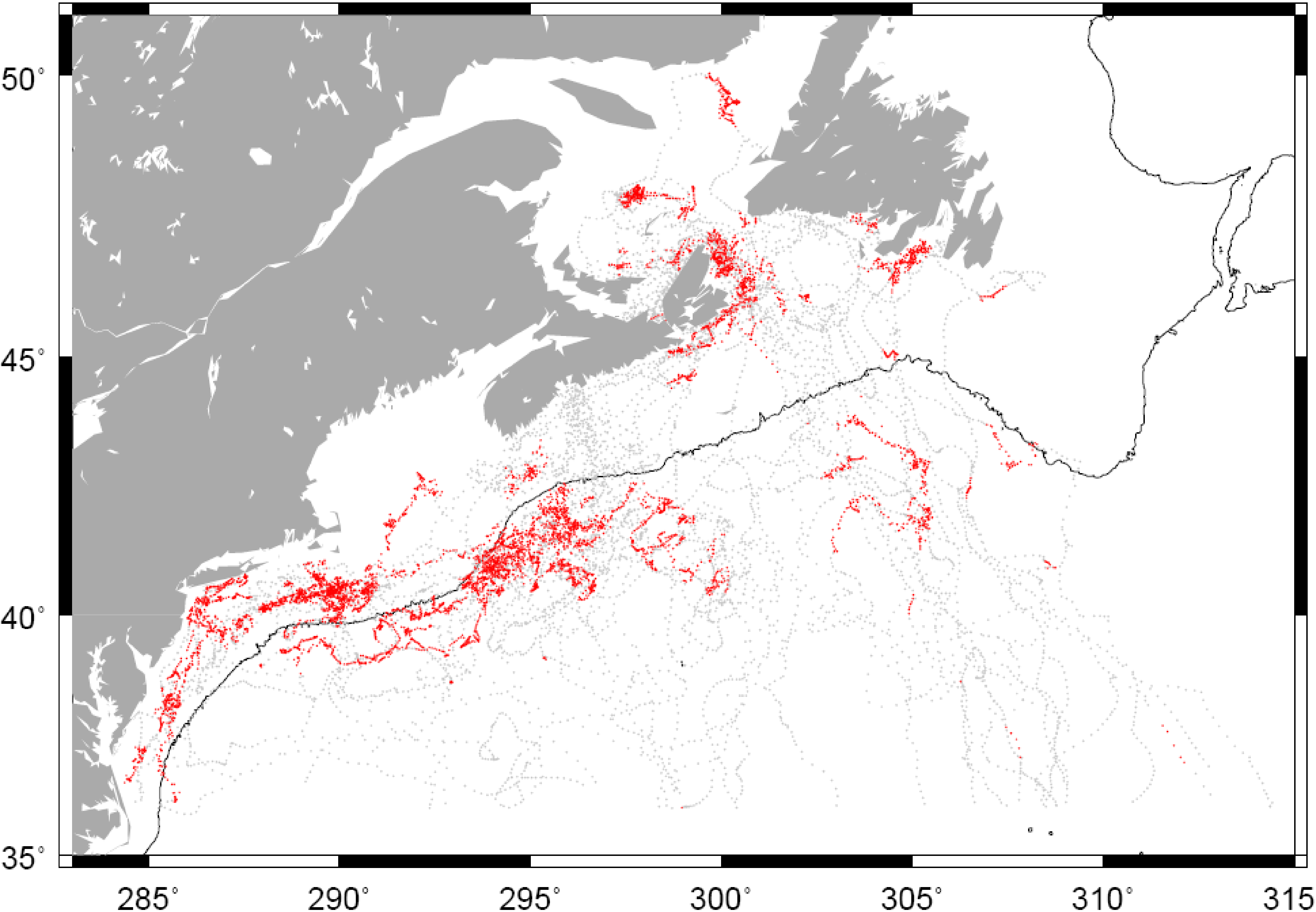
$$P(X_t = j | X_{t-1} = i) = A(i, j)$$ transition matrix

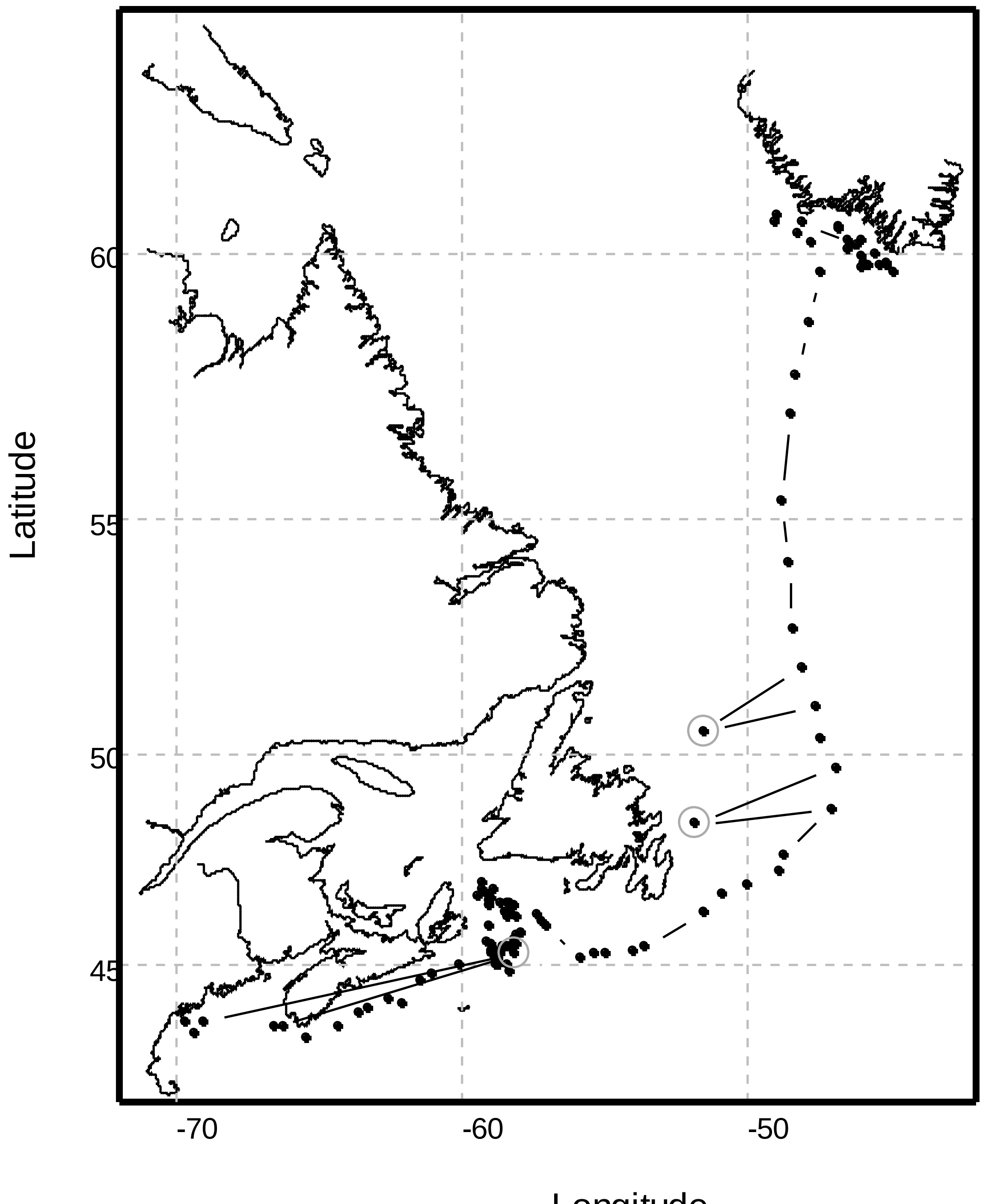
$$p(Y_t = y | X_t = i) = \mathcal{N}(y; \mu_i, \Sigma_i)$$ Gaussian observations

Switching State Space Model



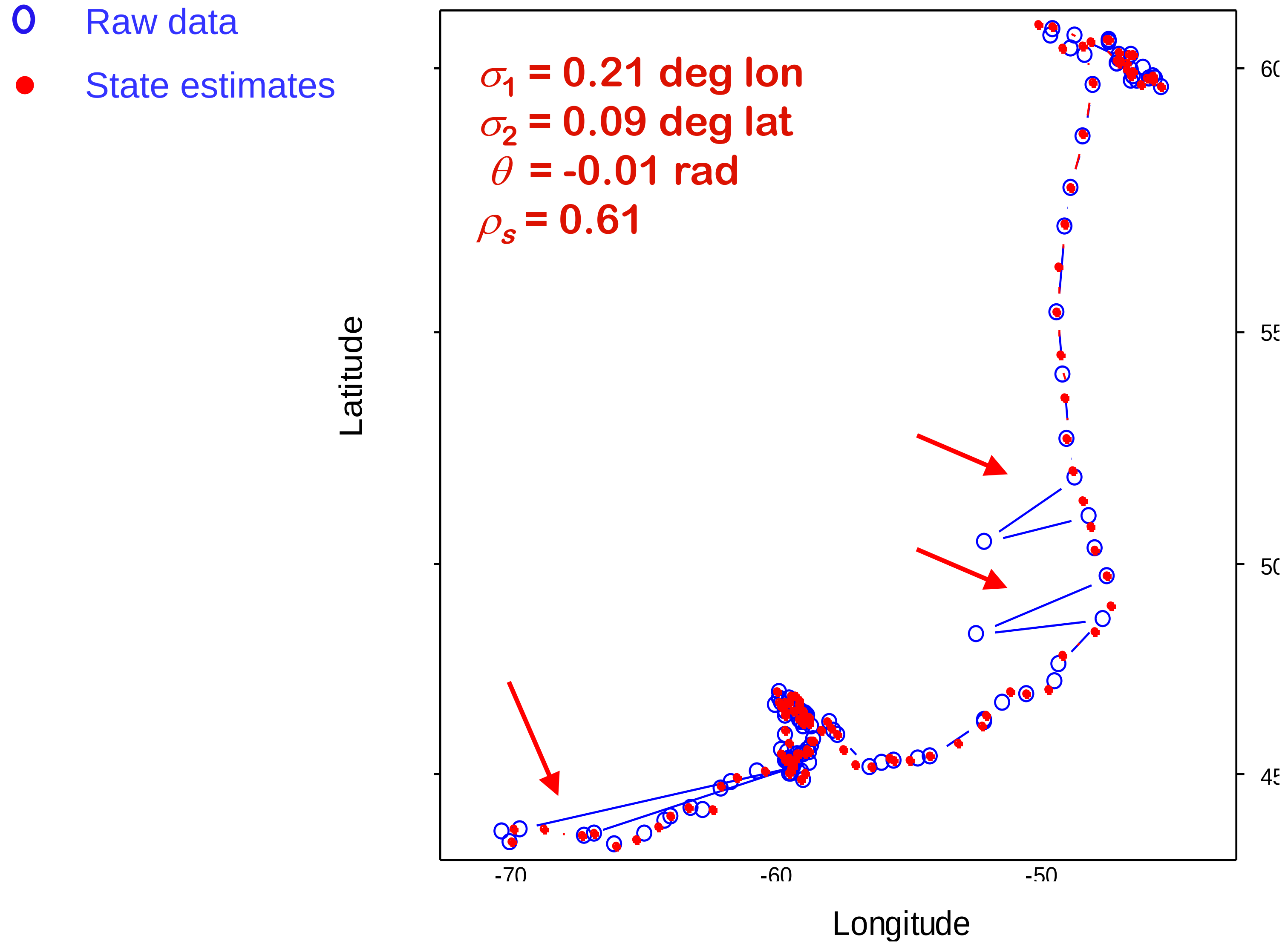

52°
48°
44°
288°
292°
296°
300°
Red: Foraging
Blue: Migrating


50°
45°
40°
35°
285°
290°
295°
300°
305°
310°
315°

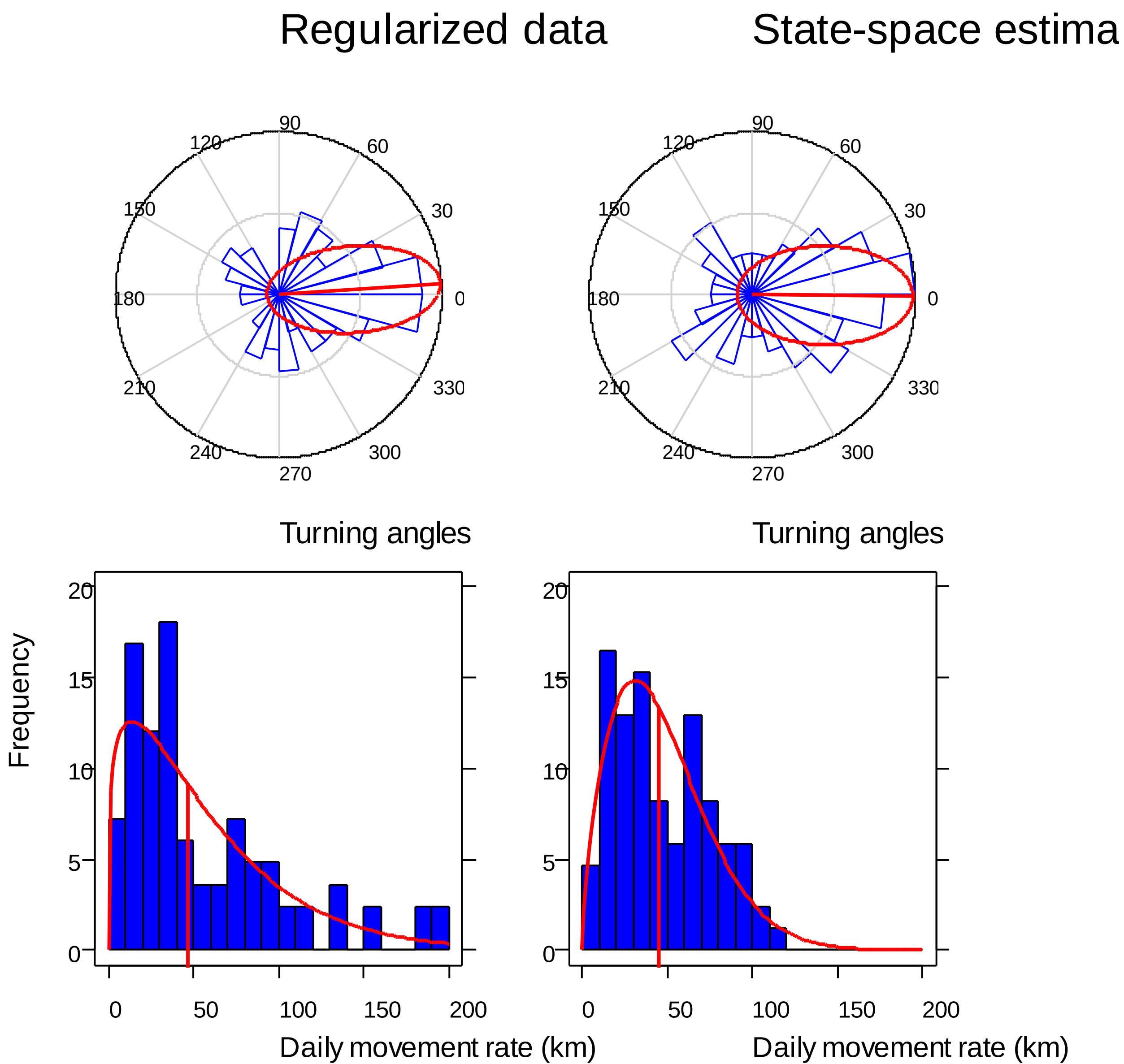


Latitude
60
55
50
45
-70
-60
-50
Longitude

Filtered Data

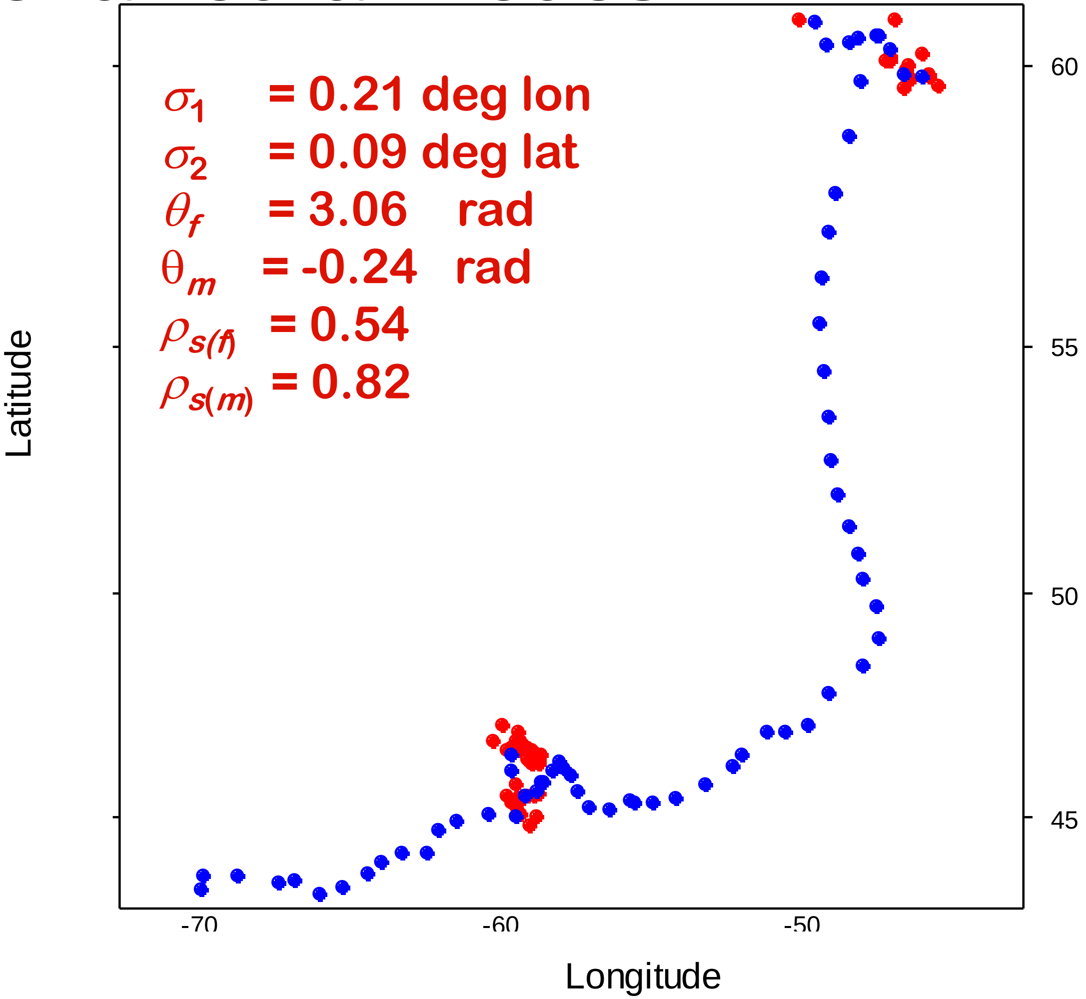


Derived Variables



A Switching SSM

Switching model, estimates switches b/w 2 behavioural modes



Summary

- State-space models allow you to think about problems which have no conventional solution
- Fundamentally different approach to analysis of complex, error-prone data emphasis on estimation of “true” states, biological parameters and uncertainty
- Models can be fit to other types of sequential movement data (GPS, Archival tags)

The Future

- Better incorporation of oceanographic data
- Model testing, statisticians do NOT know how to compare models with non-Gaussian errors
- More “user friendly” (i.e. less “user angry” methods). This would include an easy to use library with a variety of possible behaviour.
- Include more than 2 states in the the switching models.

Structural Equation Modelling (SEM)

Minimise the difference between the observed (S) and implied (Σ) covariances by adjusting the path coefficients (B)

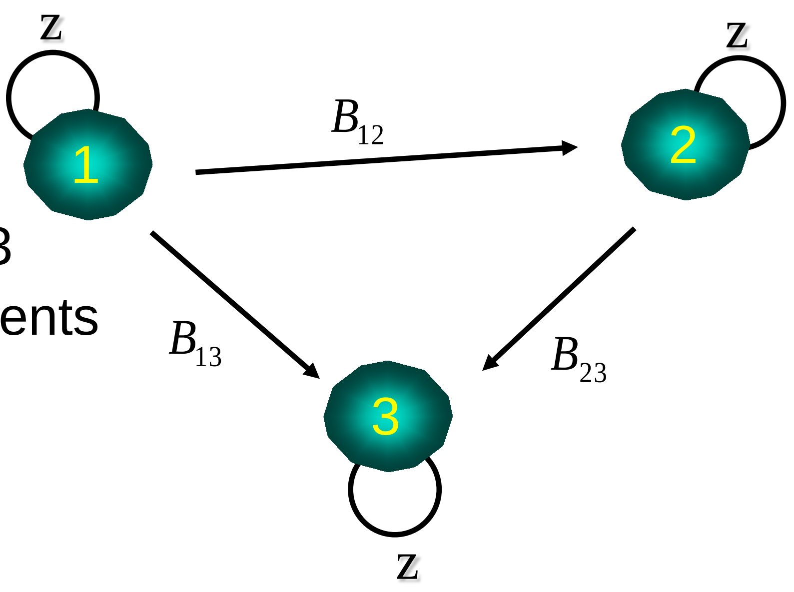
The implied covariance structure:

$x = x.B + z$

$x = z.(I - B)^{-1}$

x : matrix of time-series of Regions 1-3

B: matrix of unidirectional path coefficients



Variance-covariance structure:

$x^T . x = \Sigma = (I-B)^{-T}. C.(I-B)^{-1}$

where $C = z^T z$

$x^T.x$ is the implied variance covariance structure Σ

C contains the residual variances (u,v,w) and covariances

The free parameters are estimated by minimising a [maximum likelihood]

Modeling Sequential Data

- Sequential data arises in many areas of science & engineering
- Types of data sources:
  - Time series, generated by a dynamical system
  - Sequence generated by one-dimensional spatial process
- On- line analysis vs. Off-line analysis

Classical Solutions

- Classic approaches to time-series prediction
  - Linear models: ARIMA(auto-regressive integrated moving average), ARMAX(autoregressive moving average exogenous variables model)
  - Nonlinear models: neural networks, decision trees

- Problems with classic approaches
  - prediction of the future is based on only a finite window
  - it's difficult to incorporate prior knowledge
  - difficulties with multi-dimensional inputs and/or outputs

State-Space Models

- Assumptions:
  - There is some underlying hidden state of the world **(query)** that generates the observations **(evidence)**, and evolves in time, possibly as a function of our inputs
  - Models are first-order Markov, i.e.,

    $$P(X_t | X_{1:t-1}) = P(X_t | X_{t-1})$$
  - observations are conditional first-order Markov

    $$P(Y_t | X_t , Y_{t-1}) = P(Y_t | X_t)$$
  - Time-invariant or homogeneous

- The goal: computing of the belief state:

  The belief on the hidden state of the world given the observations up to the current time $y_{1:t}$ and inputs $u_{1:t}$ to the system, $\mathbf{P(X | yS_{1:t}, u_{1:t})}$

- State-space model must define a prior $P(X_1)$, a state-transition function, $P(X_t | X_{t-1})$, and an observation function, $P(Y_t | X_t)$

SSM: Representation

Hidden Markov Models (HMMs):

X_t is a discrete random variables

Kalman Filter Models (KFMs):

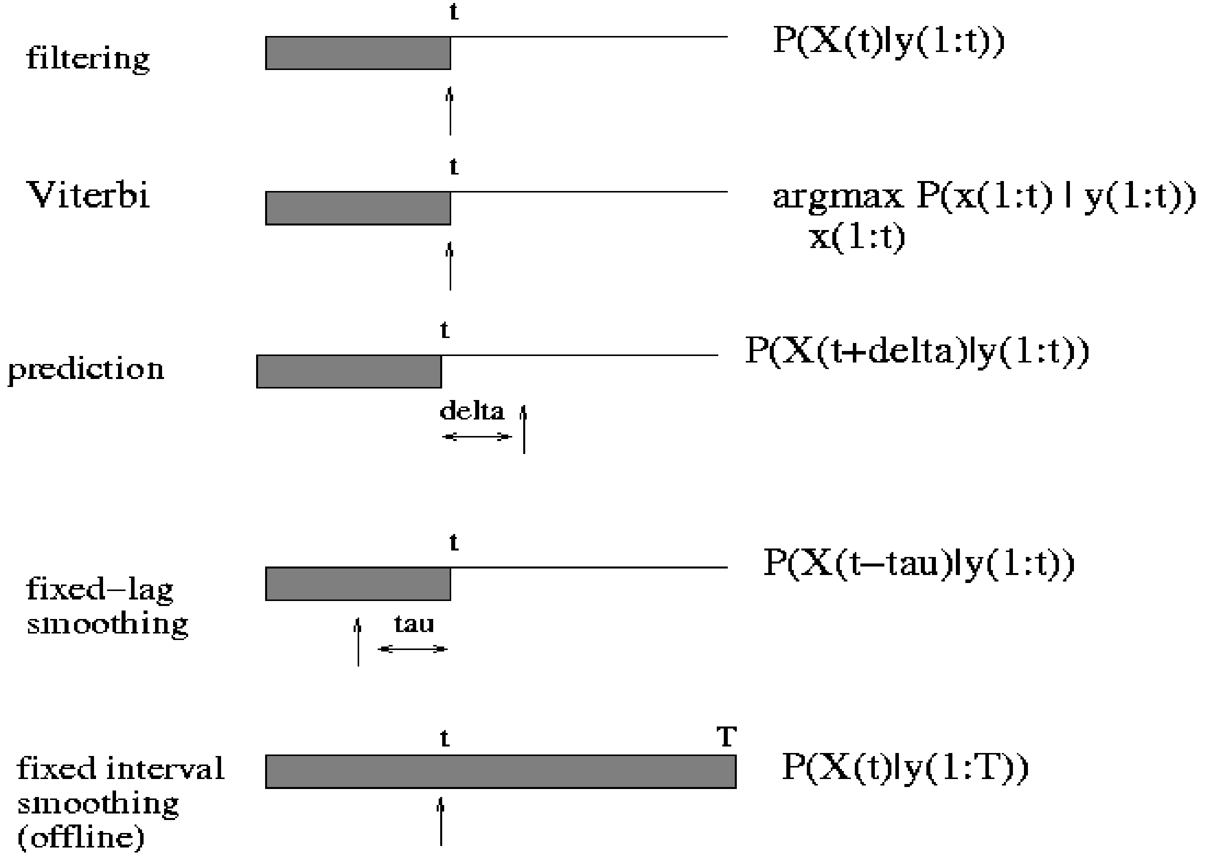
X_t is a vector of continuous random variables

Dynamic Bayesian Networks (DBNs):

more general and expressive language for representing state-space models

SSM: Inference

- A state-space model defines how X_t generates Y_t and $X_{t.}$
- The goal of inference is to infer the hidden states (query) $X_{1:t}$ given the observations (evidence) $Y_{1:t}$.



SSM: Inference (cont.)

- Inference tasks:
  - Filtering (monitoring): recursively estimate the belief state using Bayes' rule
    - prediction: computing $P(X_t \mid y_{1:t-1})$
    - updating: computing $P(X_t \mid y_{1:t})$
    - throw away the old belief state once we have computed the prediction ("rollup")
  - Smoothing: estimate the state of the past, given all the evidence up to the current time
    - Fixed-lag smoothing (hindsight): computing $P(X_{t-l} \mid y_{1:t})$ where $l > 0$ is the lag
    - Fixed-interval smoothing (offline): computing $P(X_t \mid y_{1:T})$ for all $1 \le t \le T$
  - Prediction: predict the future
    - Lookahead: computing $P(X_{t+h} \mid y_{1:t})$ where $h > 0$ is how far we want to look ahead
  - Viterbi decoding: compute the most likely sequence of hidden states given the data
    - MPE (abduction): $x^*_{1:t} = \operatorname{argmax} P(x_{1:t} \mid y_{1:t})$

SSM: Learning

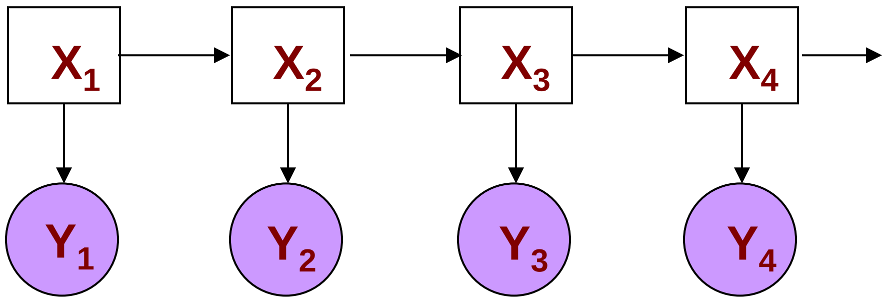
- Parameters learning (system identification) means estimating from data these parameters that are used to define the transition model $P(X_t | X_{t-1})$, the observation model $P(Y_t | X_t)$ & the prior $P(X_1)$
- The usual criterion is maximum-likelihood(ML)
- The goal of parameter learning is to compute
  - $\theta^*_{ML} = \text{argmax}_{\theta} P(Y| \theta) = \text{argmax}_{\theta} \log P(Y| \theta)$, where

$$\log P(Y|\theta) = \log \prod_{m=1}^{N_{train}} P(y_{1:T}^m|\theta) = \sum_{m=1}^{N_{train}} \log P(y_{1:T}^m|\theta)$$

  - Or $\theta^*_{MAP} = \text{argmax}_{\theta} \log P(Y| \theta) + \log P(\theta)$ if we include a prior on the parameters
  - Two standard approaches: gradient ascent and EM(Expectation Maximization)

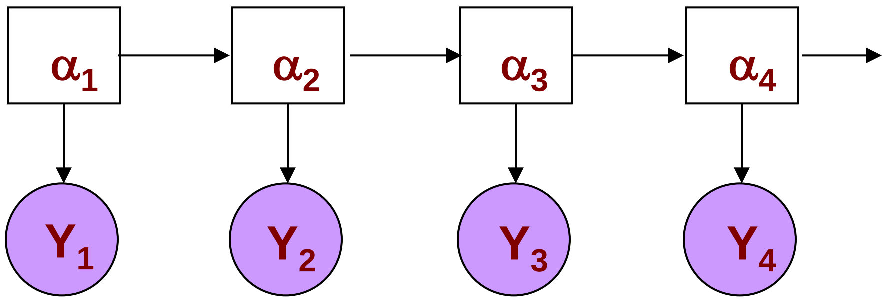
- Problem: Hidden variables complicate finding of the globally optimal parameters
- Structure learning: more ambitious

HMM: Hidden Markov Model



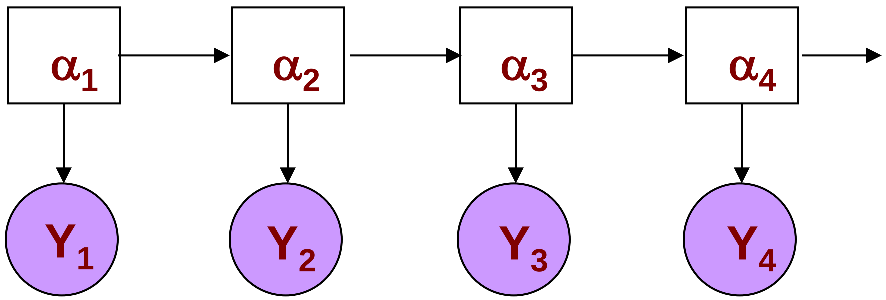
- one discrete hidden node and one discrete or continuous observed node per time slice.
- X: hidden variables
- Y: observations
- Structures and parameters remain same over time
- Three parameters in a HMM:
  - The initial state distribution $P(X_1)$
  - The transition model $P(X_t \mid X_{t-1})$
  - The observation model $P(Y_t \mid X_t)$

HMM: Hidden Markov Model



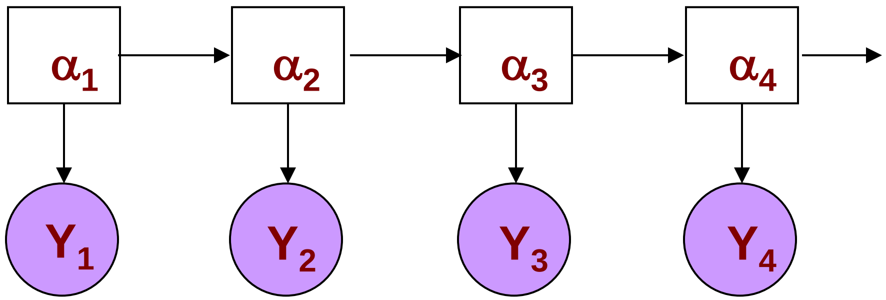
- one discrete hidden node and one discrete or continuous observed node per time slice.
- X: hidden variables
- Y: observations
- Structures and parameters remain same over time
- Three parameters in a HMM:
  - The initial state distribution $P(X_1)$
  - The transition model $P(X_t \mid X_{t-1})$
  - The observation model $P(Y_t \mid X_t)$
- HMM is the simplest DBN
  - a discrete state variable with arbitrary dynamics and arbitrary measurements

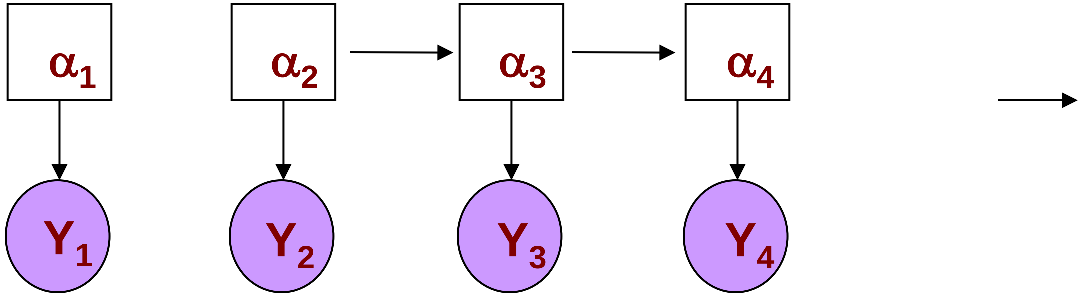
Special Case of State-space Models: Hidden Markov Model



- one **discrete** hidden node and one discrete or continuous observed node per time slice.
- α: hidden variables
- Y: observations
- Structures and parameters remain same over time
- Three parameters in a HMM:
  - The initial state distribution $P(\alpha_1)$
  - The transition model $P(\alpha_t \mid \alpha_{t-1})$
  - The observation model $P(Y_t \mid \alpha_t)$

HMM: Hidden Markov Model





KFM: Kalman Filter Model

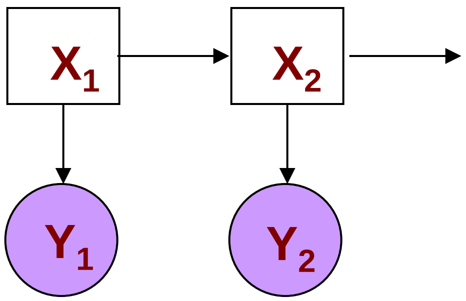
- KFM has the same topology as an HMM
- all the nodes are assumed to have linear-Gaussian distributions

x(t+1) = A*x(t) + v(t),

v ~ N(0, Q) : process noise, x(0) ~ N(X(0), V(0))

y(t) = C*x(t) + w(t),

w ~ N(0, R) : measurement noise



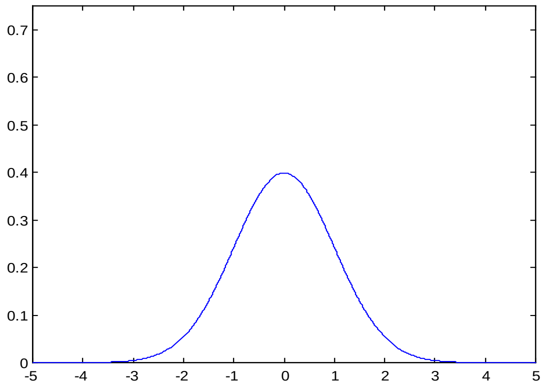
- Also known as Linear Dynamic Systems (LDSs)
  - a partially observed stochastic process
  - with linear dynamics and linear observations: f(a + b) = f(a) + f(b)
  - both subject to Gaussian noise
- KFM is the simplest continuous DBN
  - a continuous state variable with linear-Gaussian dynamics and measurements

All Roads Lead From Gauss 1809

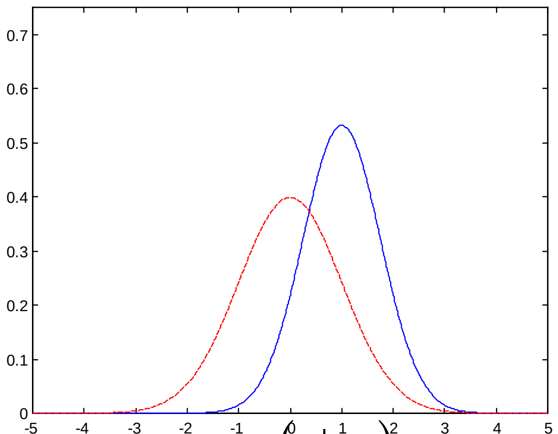
- “ … since all our measurements and observations are nothing more
- than approximations to the truth, the same must be true of all
- calculations resting upon them, and the highest aim of all
- computations made concerning concrete phenomenon must be to
- approximate, as nearly as practicable, to the truth. But this can be
- accomplished in no other way than by suitable combination of more
- observations than the number absolutely requisite for the determination of
- the unknown quantities. This problem can only be properly undertaken
- when an approximate knowledge of the orbit has been already attained,
- which is afterwards to be corrected so as to satisfy all the observations
- in the most accurate manner possible.”
-
- - From Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections, Gauss, 1809
-

What does a Kalman filter do ?

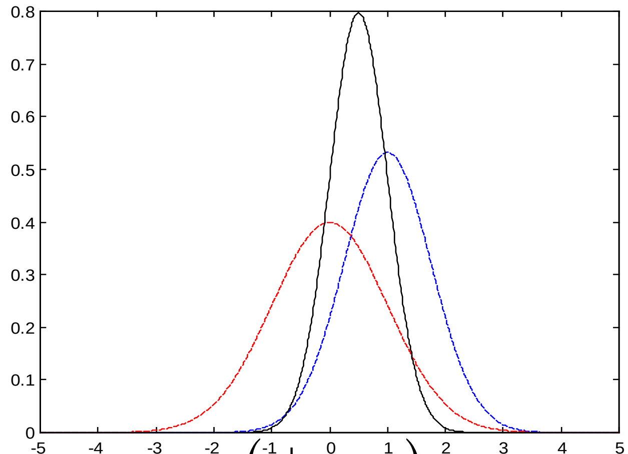
- The Kalman filter propagates the conditional density in time.



$f\left(y|x_1\right)$



$f\left(y|x_2\right)$



$f\left(y|x_1, x_2\right)$

How does it do it ?

- The Kalman filter iterates between two steps
  - Time Update (Predict)
    - Project current state and covariance forward to the next time step, that is, compute the next a priori estimates.
  - Measurement Update (Correct)
    - Update the a priori quantities using noisy measurements, that is, compute the a posteriori estimates.

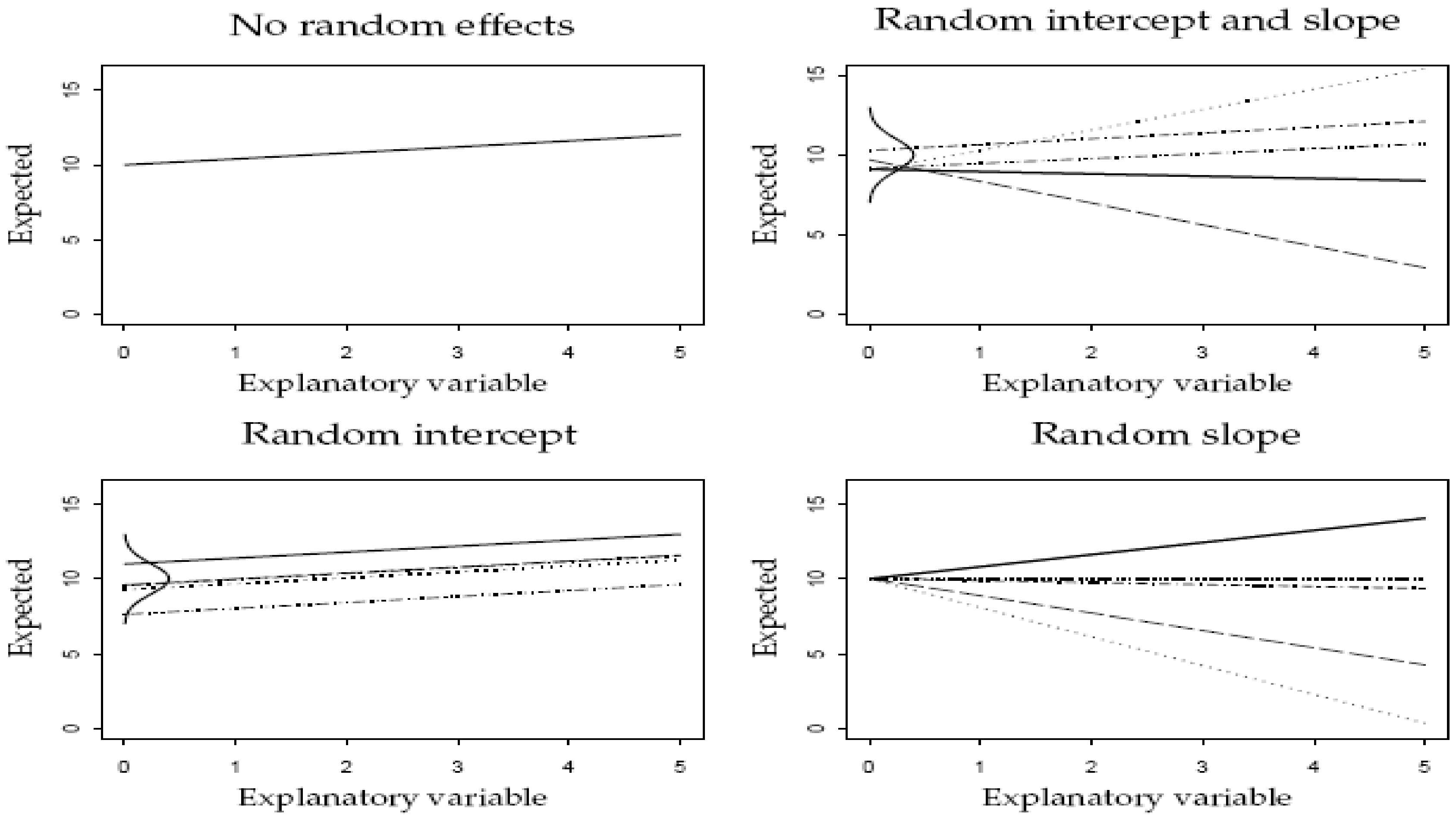
$$\hat{y}_k = \hat{y}_k^- + K_k\left(x_k - M_k \hat{x}_k^-\right)$$

- Choose K_k to minimize error covariance

You can ask questions and think about questions you can not otherwise.

- Circle of confusion
- Turtle speed at night
- Are there modes, or “behavioural states” in their behaviour.

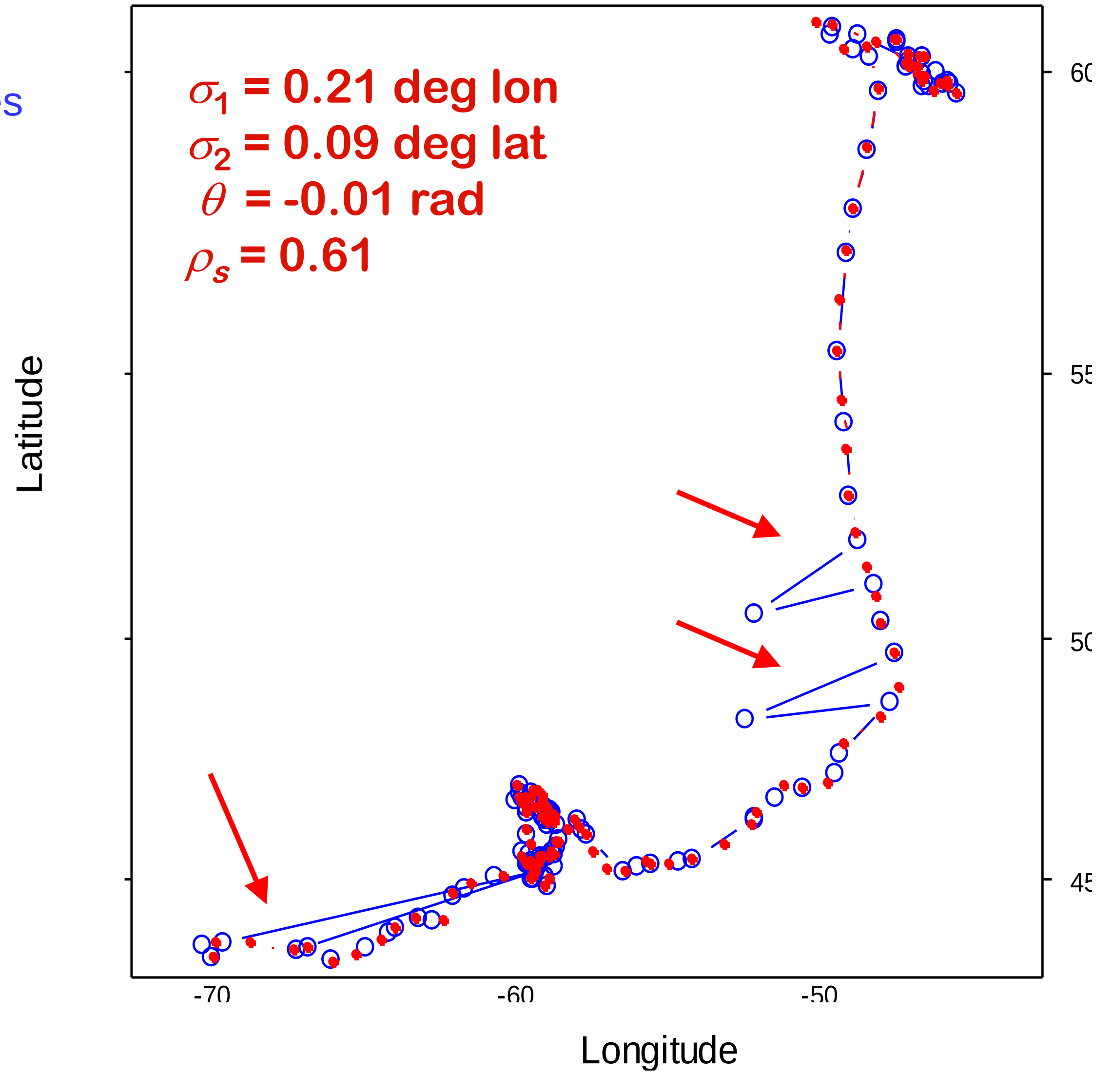
Random Effect Model



Filtered Data

O Raw data

● State estimates

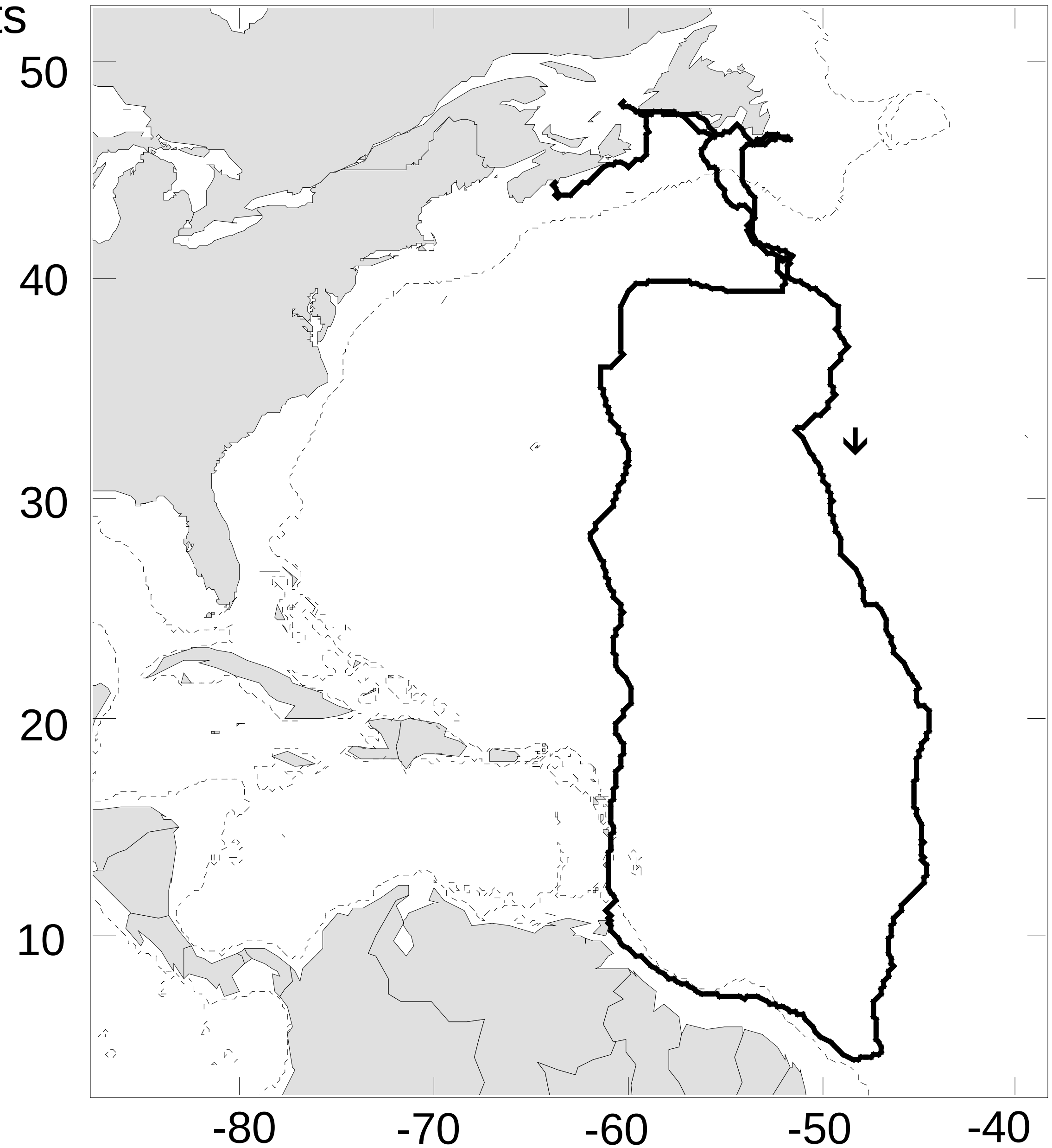


Male leatherback movements

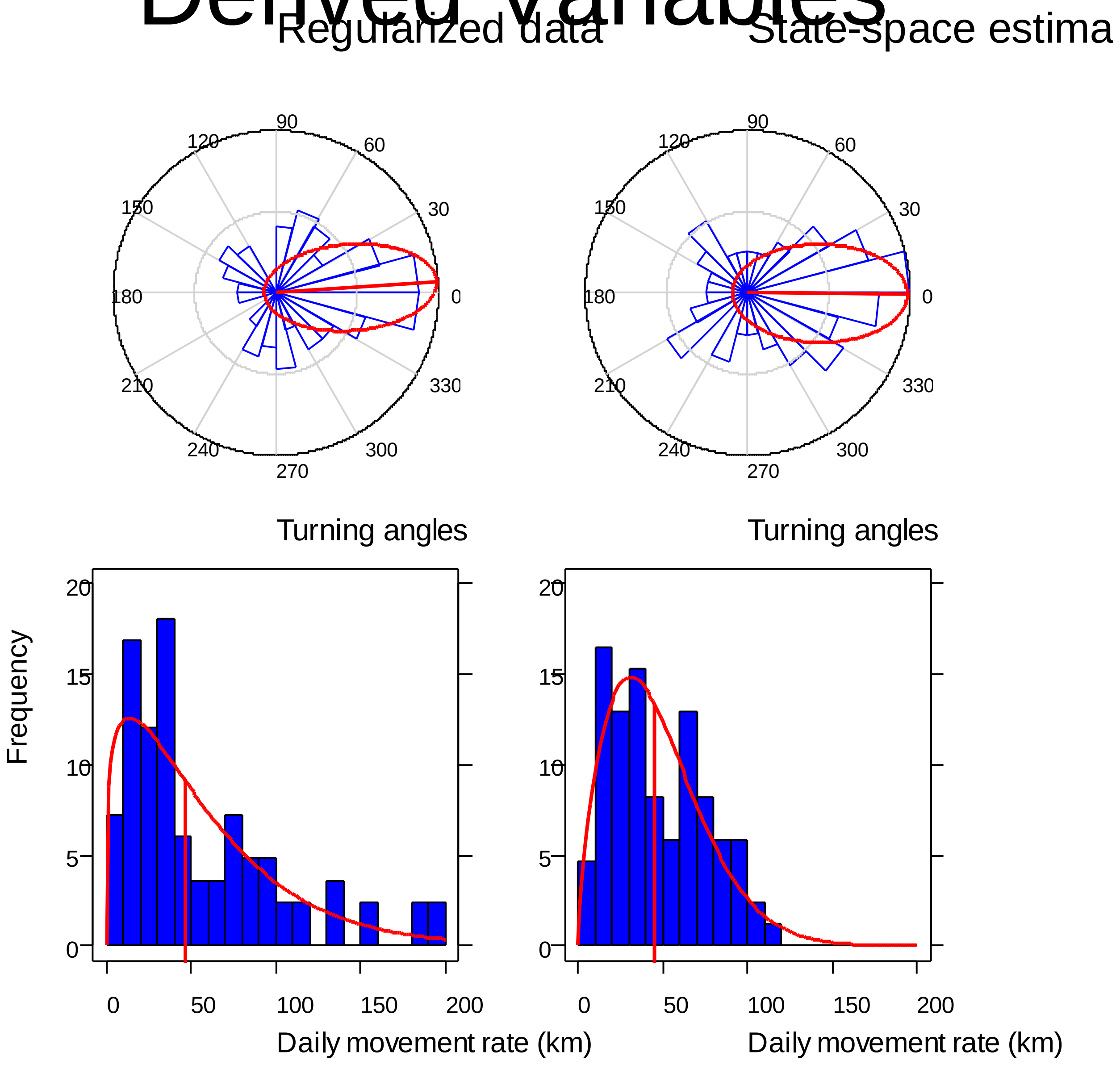
- not previously described

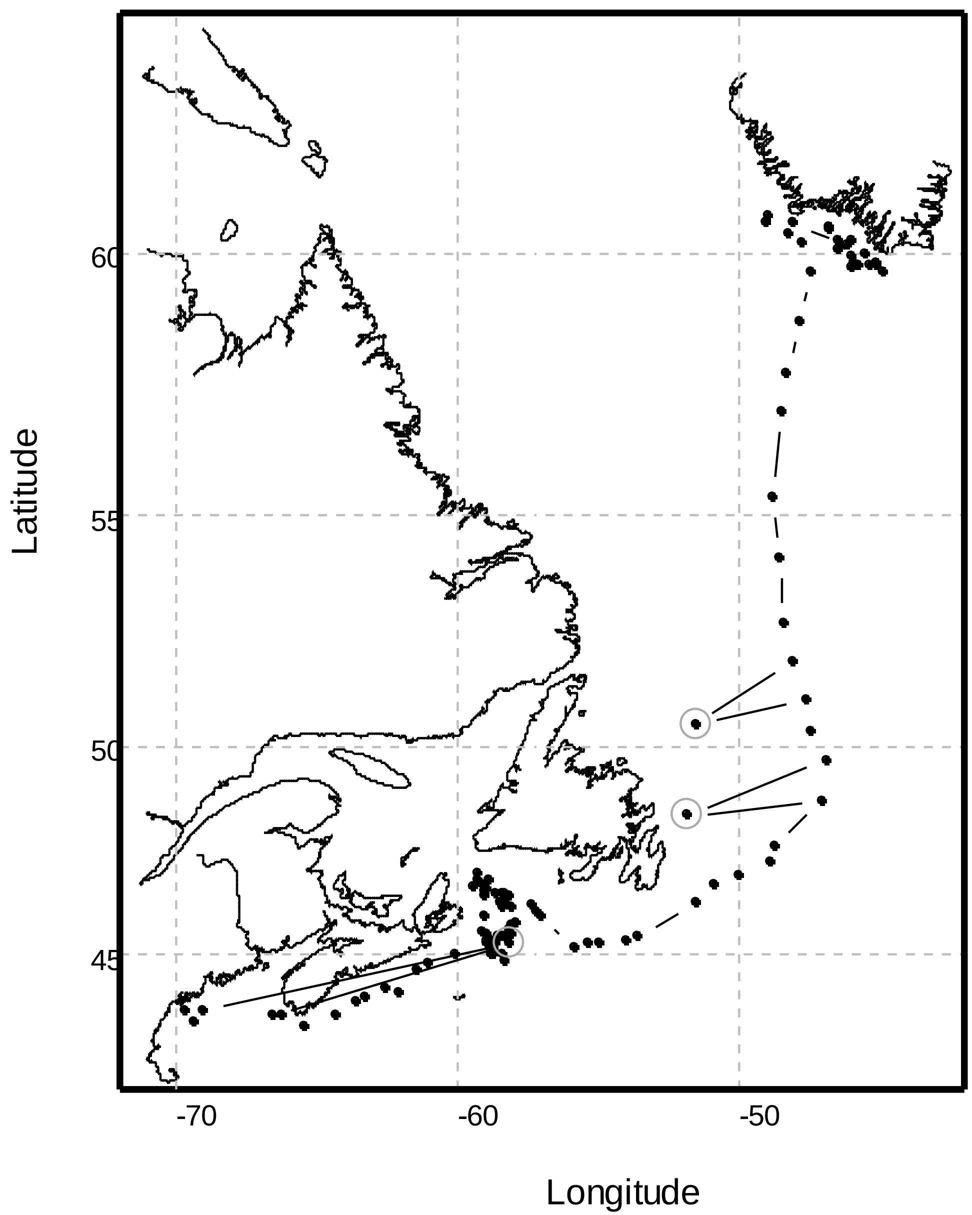
- annual migratory cycle that includes movement between temperate foraging areas and tropical breeding areas

James, Eckert and Myers
Marine Biology (*in press*)



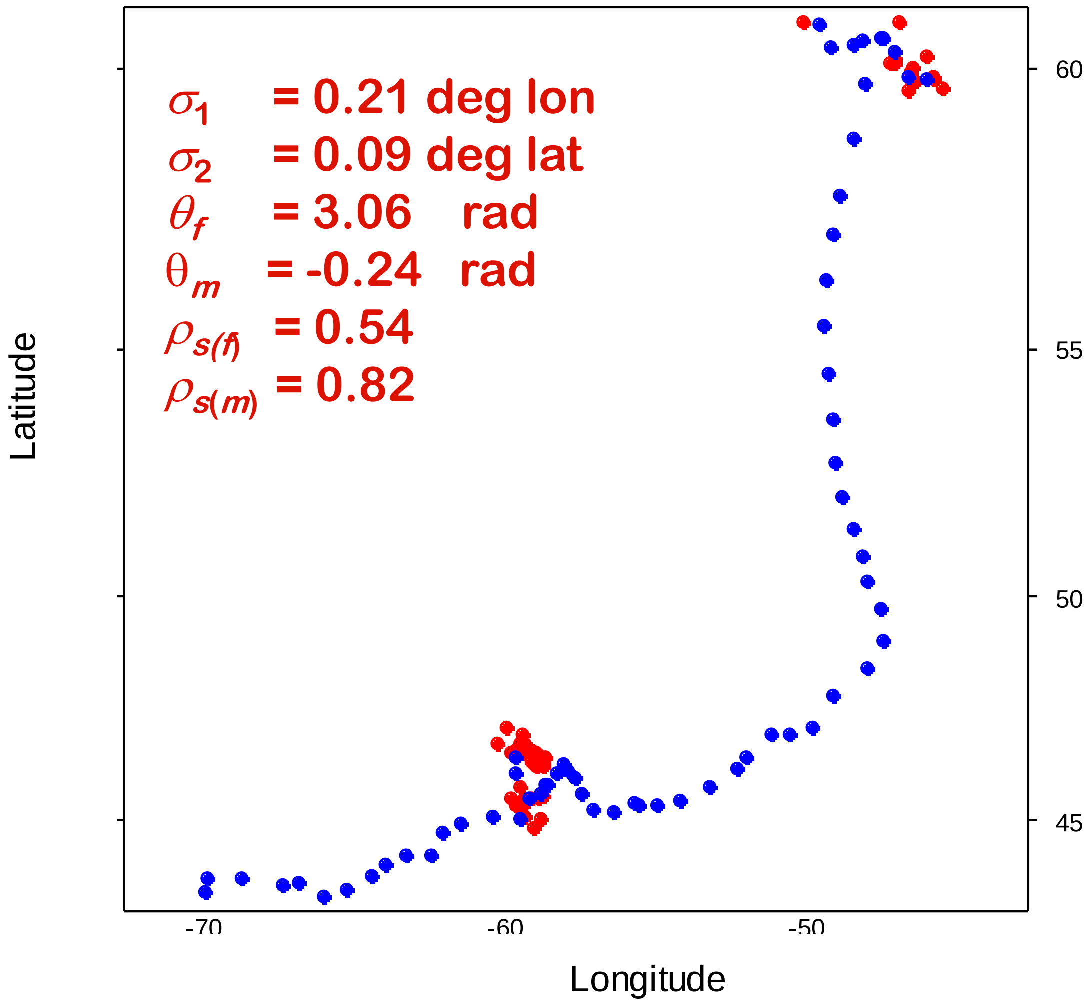
Derived Variables



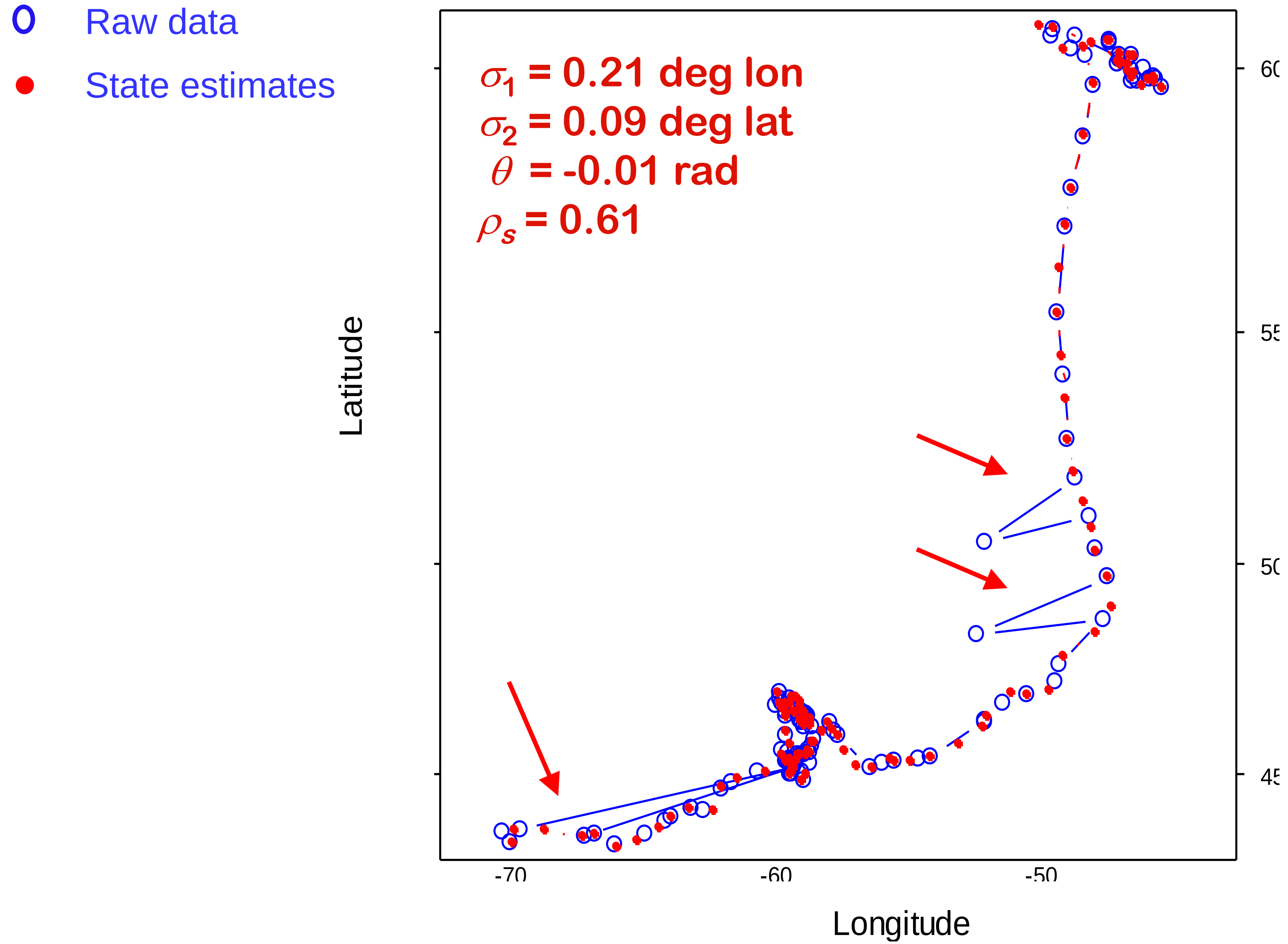

Latitude
60
55
50
45
-70
-60
-50
Longitude

A Switching SSM

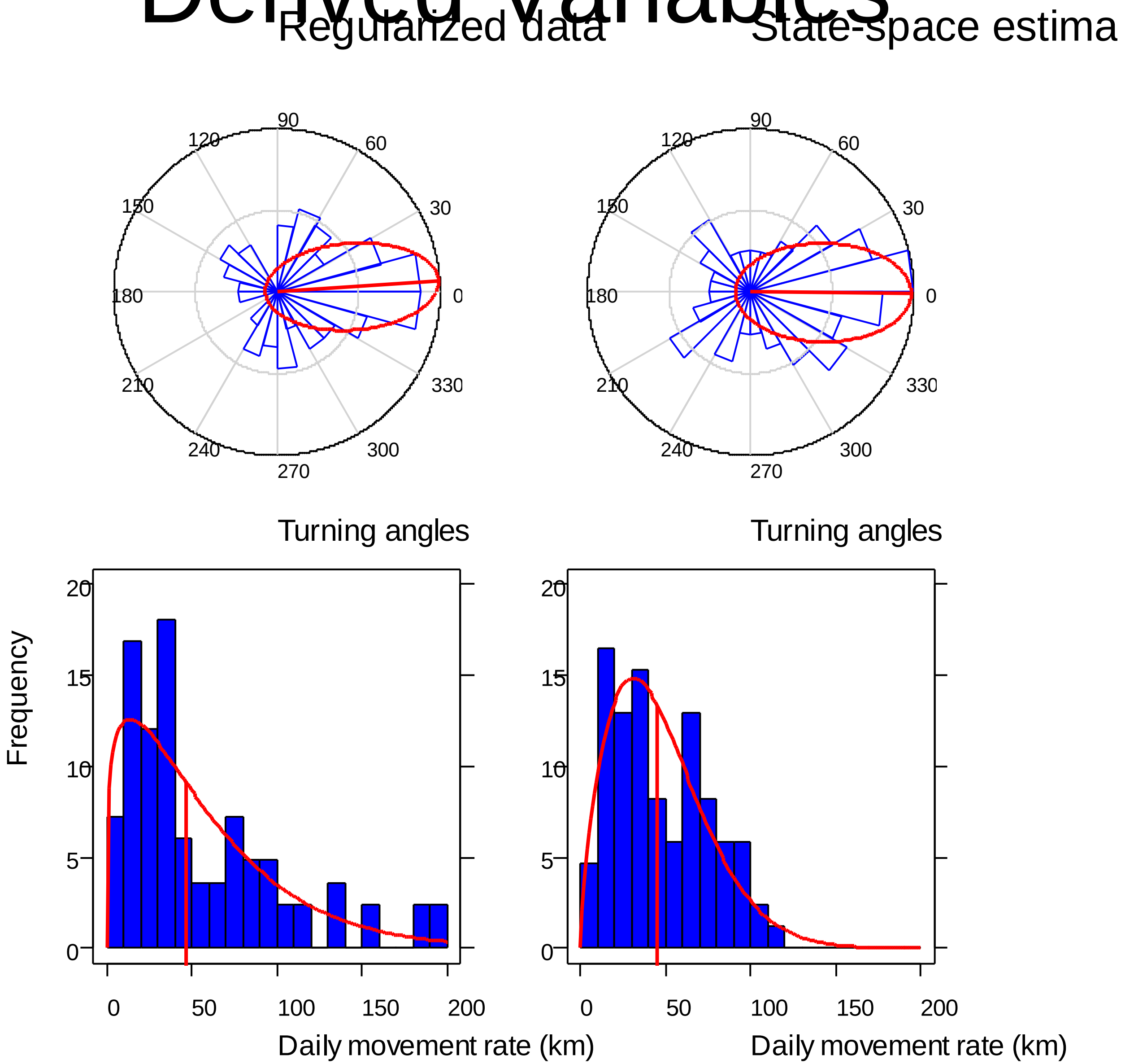
Switching model, estimates switches b/w 2 behavioural states

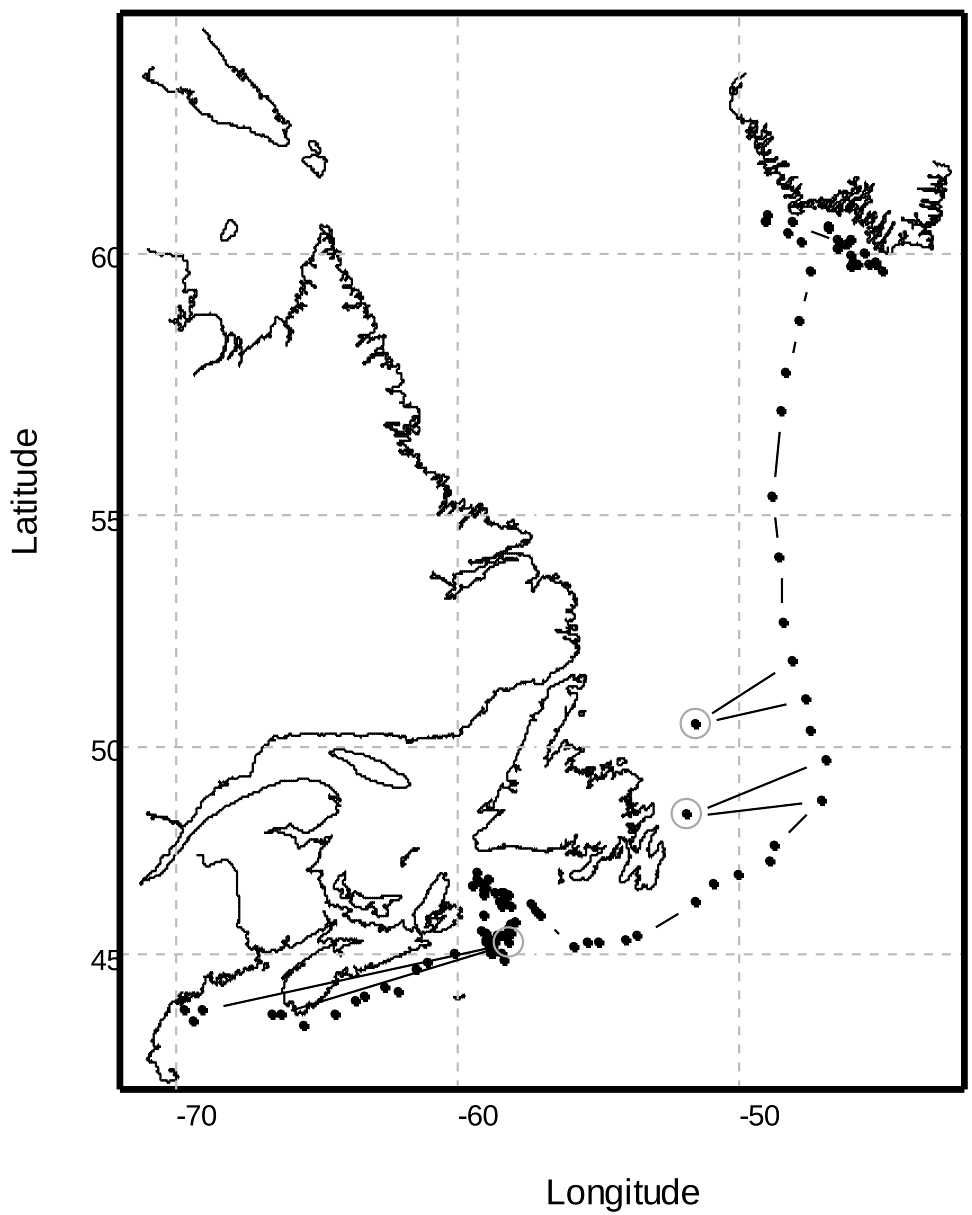


Filtered Data



Derived Variables





60
55
50
45
-70
-60
-50
Latitude
Longitude

A Switching SSM

Switching model, estimates switches b/w 2 behavioural modes