Model complexity and model choice for animal movement models

Ben Bolker, McMaster University

Departments of Mathematics & Statistics and Biology

Guelph Biomathematics & Biostatistics Symposium

9 June 2016

Outline

Acknowledgements

People **Michael Li**, **Madelon van de Kerk**, Dave Onorato, Madan Oli

Agencies US Fish and Wildlife, US Geological Survey, US National Park Service

Funding NSERC Discovery grant, NSF IGERT program

To do

- pix!
- general REFS: Turchin, Morales et al, Langrock ... ?

Animal movement: data

- observations:
  e.g. mass mark-recapture, longitudinal density, direct observation, telemetry (VHF, GPS)
- most methods provide a sequence of times and locations for each individual
- summaries:
  - home range
    (convex hull, kernel density estimate, etc.)
  - root-mean-squared displacement
  - step length and turning angle
- covariates:
  e.g. habitat map,
  individual characteristics (sex, age, weight . . .)

Animal movement: questions

- simple description
- how do animals' movements change as a function of their (internal or external) environment?
  what does that tell us about their biology?
- how might animals' distributions, etc. change when conditions (density, habitat, ...) change?

Biological/conservation issues

- Florida panther: *Felix ...*
- severely reduced habitat, endangered
- small, isolated population
- currently recovering

(cute panther pictures here)

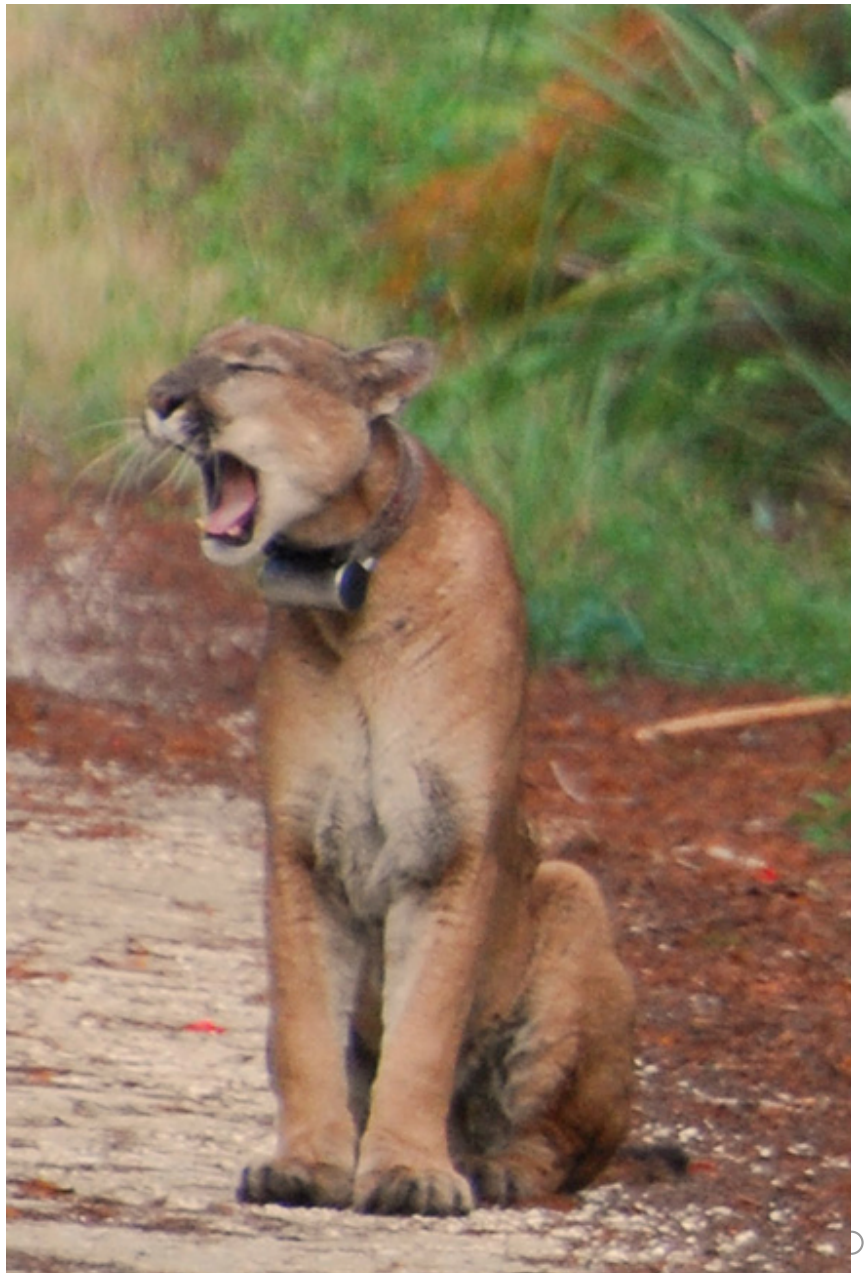
Panther movement questions

- how does panther movement vary by sex and life history stage (juvenile, adult, mom with kittens . . .)
- how does movement mediate exposure to threats (intraspecific aggression, roadkill) ?
- can we predict the effects of future changes in population density / population structure / habitat?

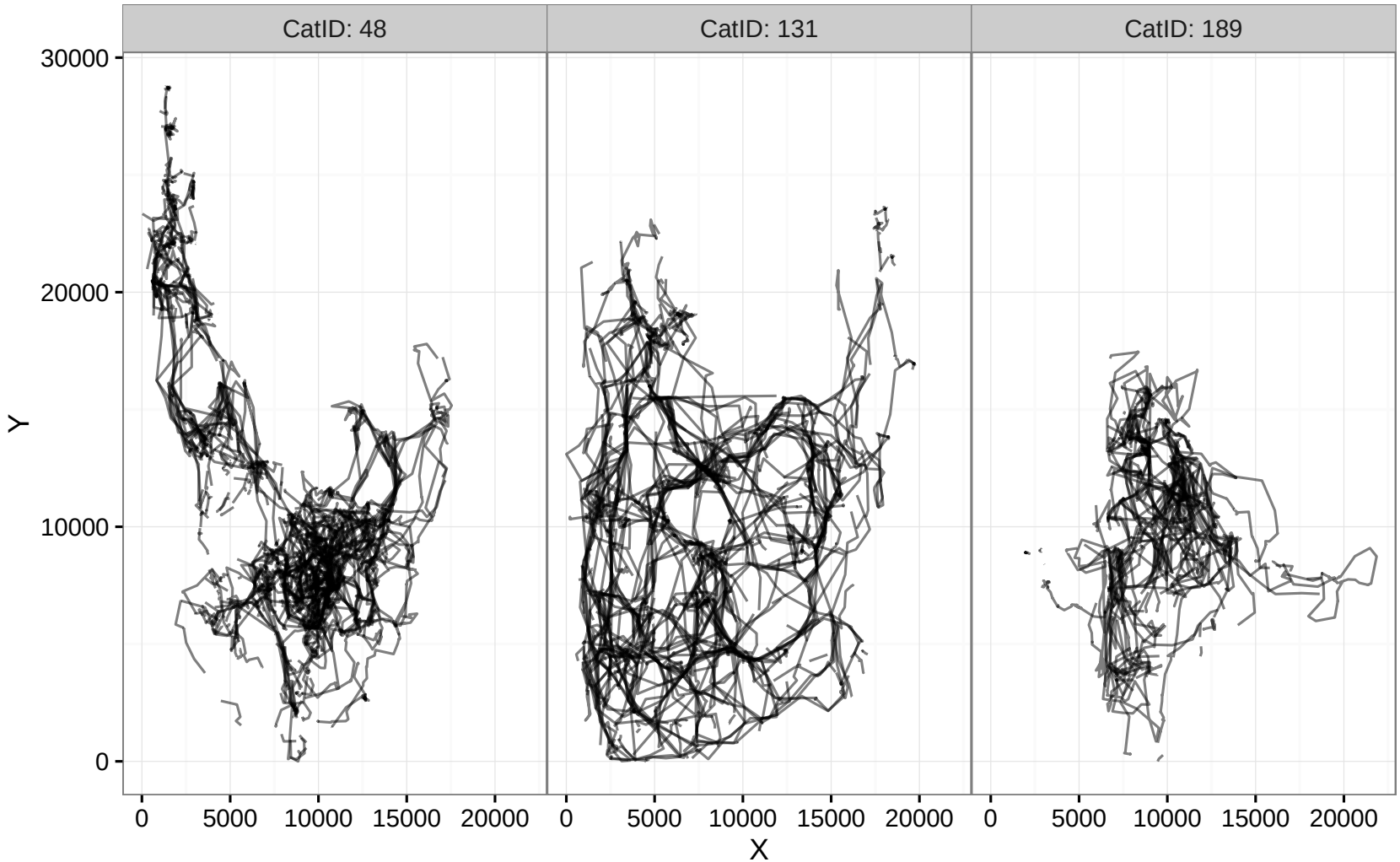
(more cute pictures here ... or roadkill?)

Panther movement data

- panthers tracked, captured
- GPS collars
- 18 males (13 male, 5 female, 1-15 years old)
- 3200 panther days, hourly/bihourly; 49000 locations
- ?? per panther

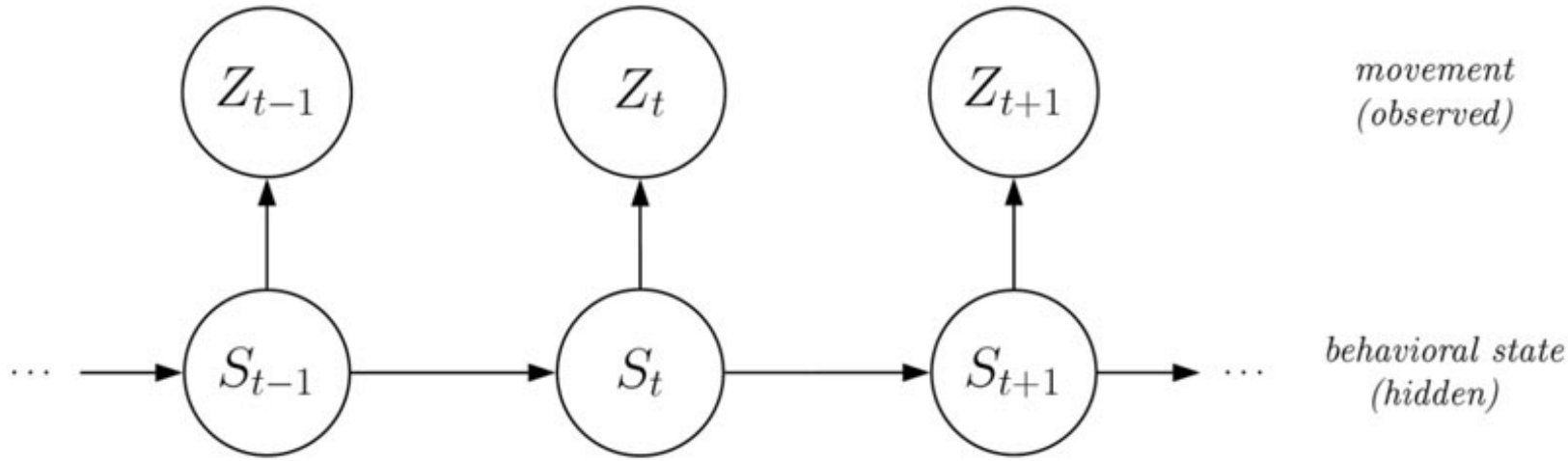


example movement tracks



Hidden Markov models

- finite mixture model with temporal dependence
- discrete time steps
- discrete latent state; *transition matrix*
- observations from *emission distributions*
  (continuous or discrete, univariate or multivariate)
- **multiphasic movement** (Fryxell et al., 2008; Langrock et al., 2012)



Hidden Markov models (cont.)

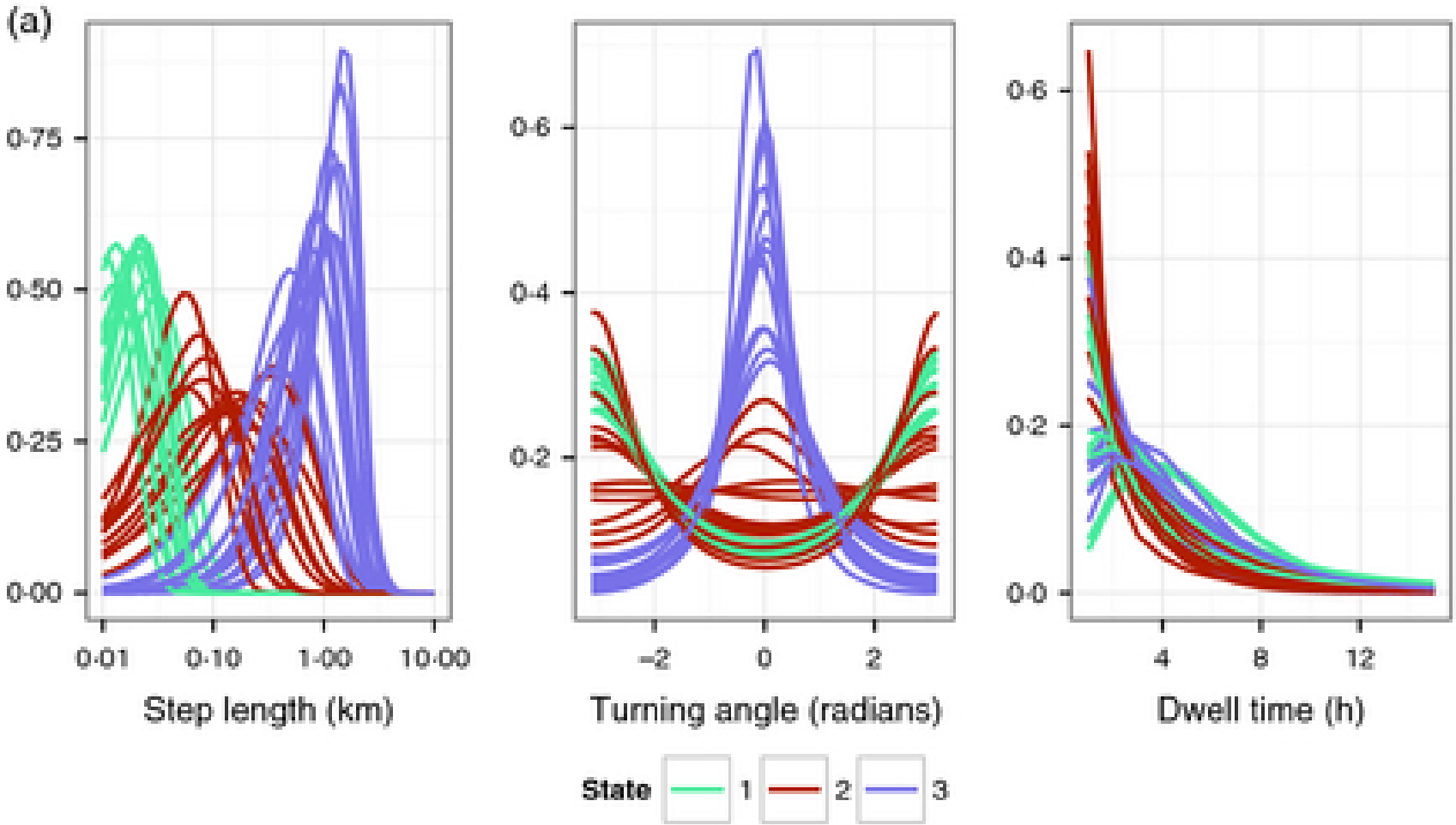
notation, underbraces

$$P(\mathbf{Y}_{1:T}, \mathbf{Z}_{1:T} | \boldsymbol{\theta}, \mathbf{X}_{1:T}) = P(z_1 \mid \mathbf{x}_1) P(\mathbf{y}_1 | z_1, \mathbf{x}_1) \cdot \prod_{k=2}^{T} P(z_k | z_{k-1}, \mathbf{x}_k) P(\mathbf{y}_k | z_k, \mathbf{x}_k) \quad (1)$$

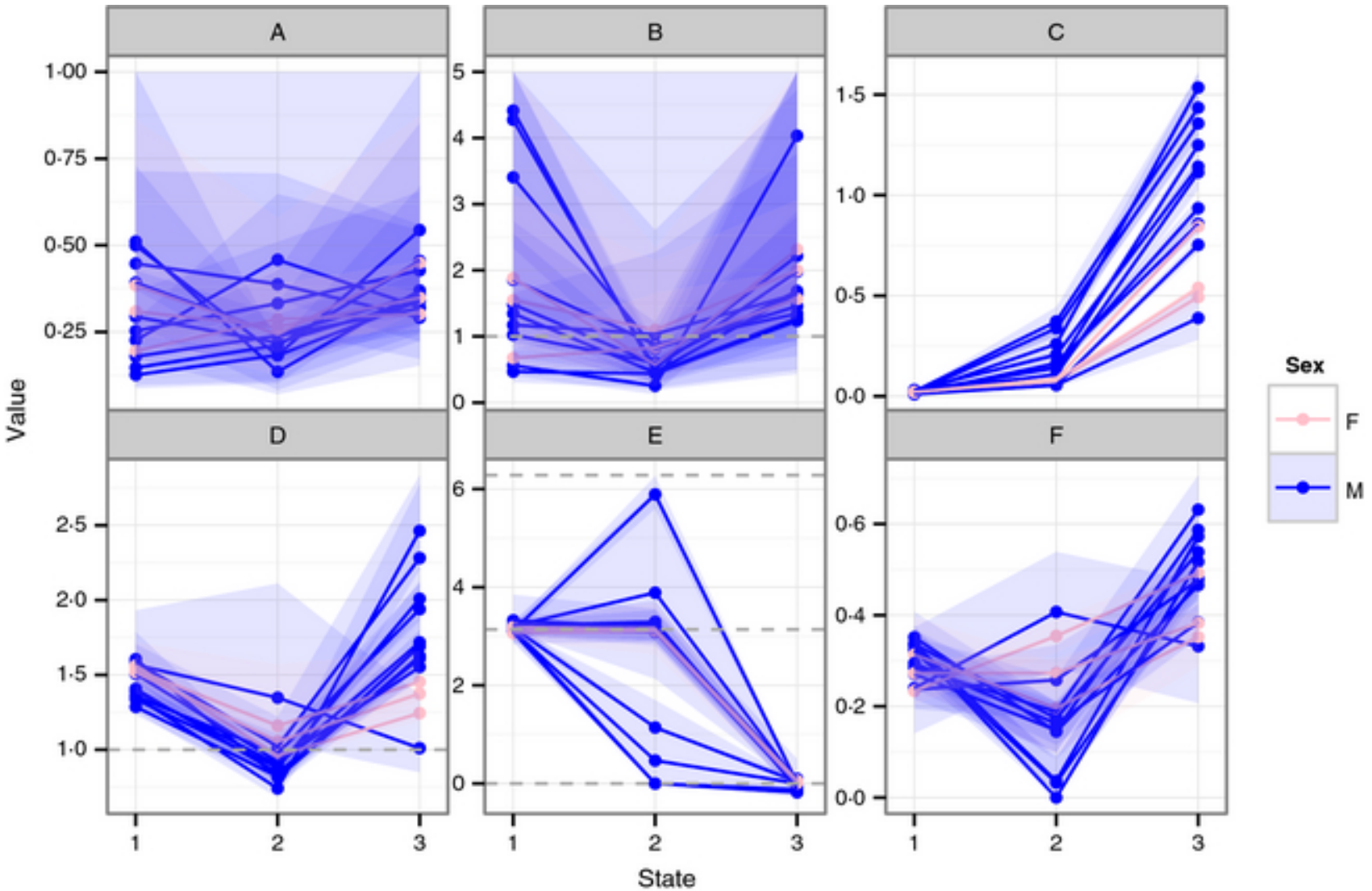
Hidden Markov models (part 3)

- *forward-backward algorithm* for estimating parameters
- *Viterbi algorithm* for estimating most probable state sequences
- `depmixS4` package (Visser and Speekenbrink, 2010) (also `moveHMM` (Michelot et al., 2016))
- hidden *semi-Markov* models: allow for non-geometric *dwell distributions* (Langrock, 2011; Augustine, 2016): `move.HMM`

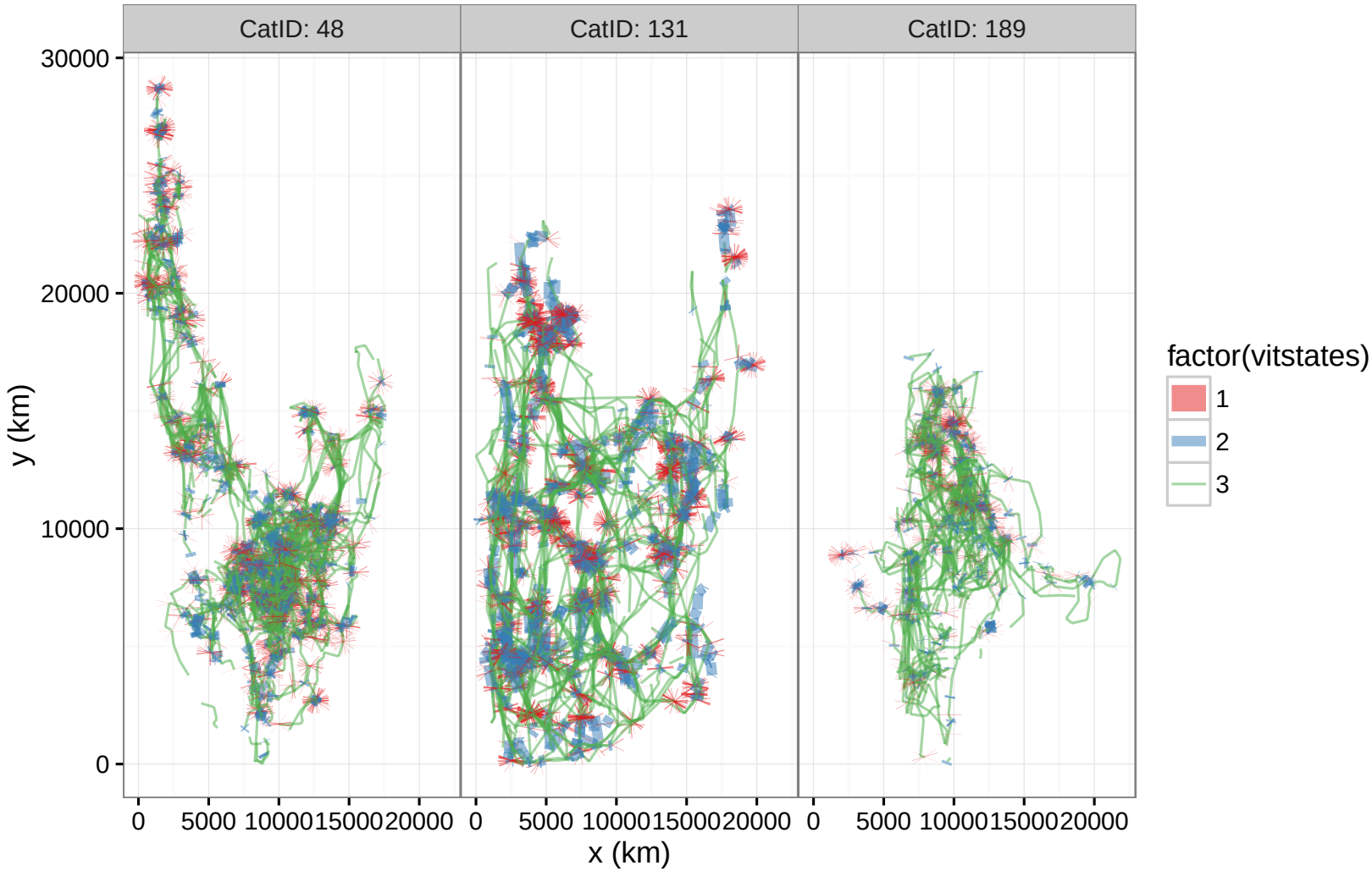
State distributions



Parameter estimates



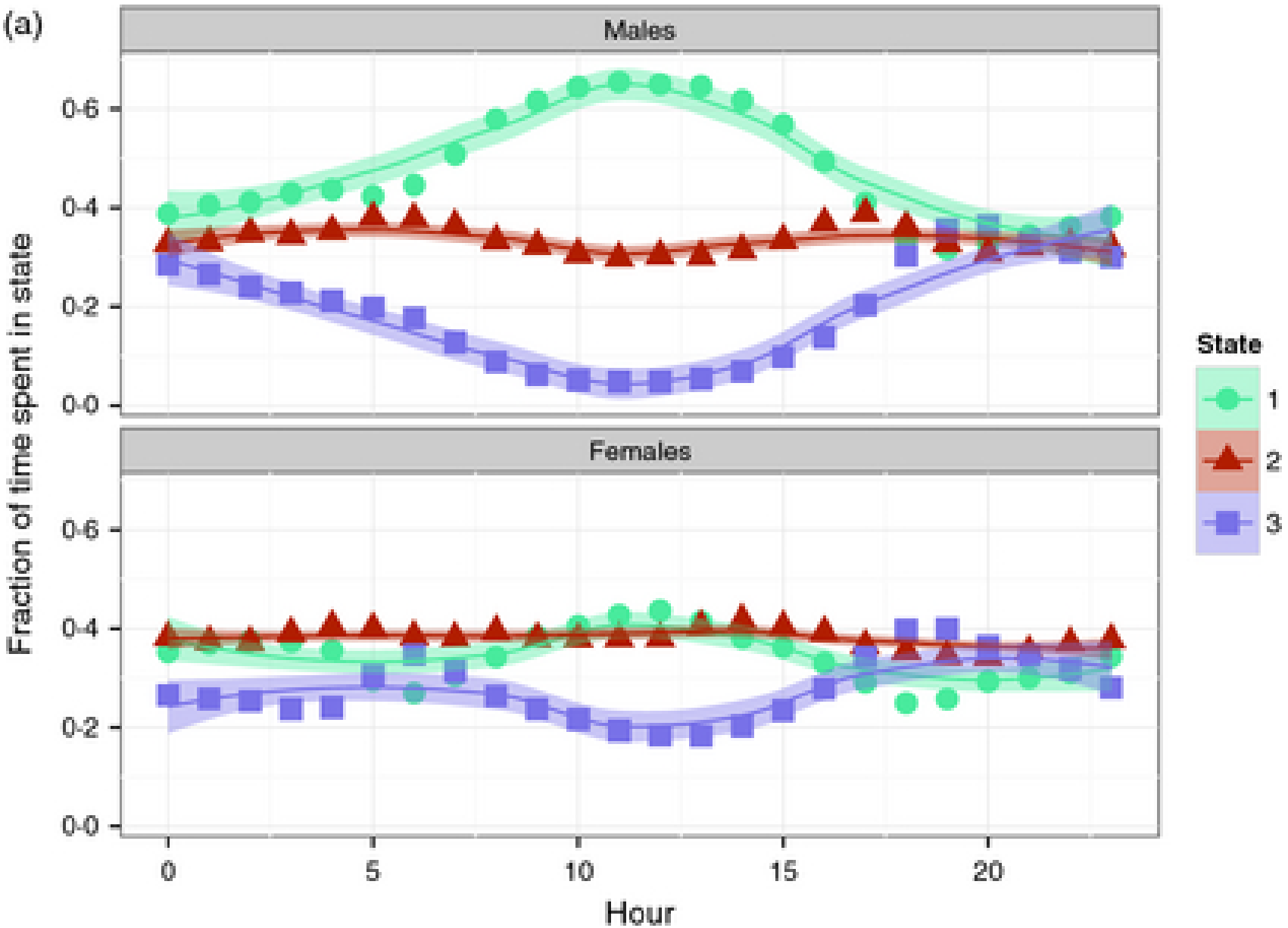
Tracks with Viterbi estimates



Transition parameters

picture/table here of transition parameters (network diagram??)

Diurnal variation



what can we conclude so far?

good news

- basic biology: males move faster, farther
- three states are identifiable, sensible
- ...

bad news

- diurnal variation in Viterbi results - but it's not in the model!
- estimates of model complexity are too high

what can we conclude so far?

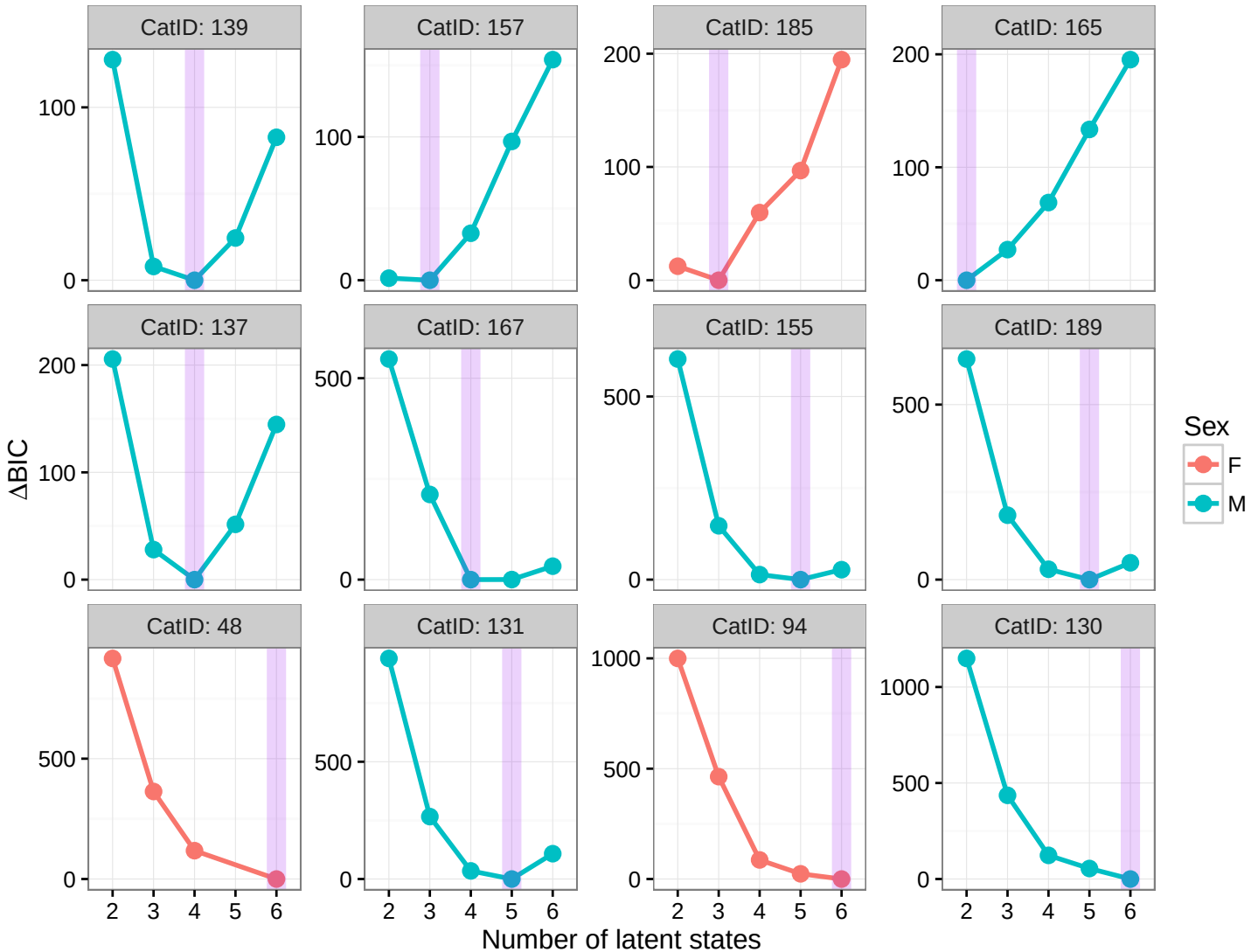
good news

- basic biology: males move faster, farther
- three states are identifiable, sensible
- ...

bad news

- diurnal variation in Viterbi results - but it's not in the model!
- estimates of model complexity are too high

Model complexity



Broadening the model

Attempting to fix these problems:

- extend the model to allow covariates
- specifically, allow for diurnal variation
  - simplify model (log-Normal step length only)
  - *fixed* state-specific emissions parameters (step length mean and std dev)
  - time-varying transition parameters
  - also try *finite mixture models* (independent occupancy)
- how much does this help?

Temporal models

figure showing alternative temporal models

Comparison

Figure 1 from paper

Goodness of fit

Figure 2 from paper

Temporal patterns

Figures 3 (and 4?) from paper

Model complexity

Figure showing BIC plots for all cats tried

Model complexity

Simulation results

Big data and small models

- simple model families + model misspecification $\rightarrow$ overparameterization

an aside on AIC vs BIC

Tools needed

- cross-validation (Wenger and Olden, 2012)
- diagnostic plots
- score tests?

References

Augustine, B., 2016. Flexible, user-friendly hidden (semi) Markov models for animal movement data.

Fryxell, J.M., Hazell, M., et al., 2008. *Proceedings of the National Academy of Sciences*, 105(49):19114–19119. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.0801737105.

Langrock, R., 2011. *Computational Statistics and Data Analysis*, 55(1):715–724. ISSN 01679473.

Langrock, R., King, R., et al., 2012. *Ecology*, 93(11):2336–2342. ISSN 0012-9658. doi:10.1890/11-2241.1.

Michelot, T., Langrock, R., and Patterson, T.A., 2016. *Methods in Ecology and Evolution*, in press. doi:10.1111/2041-210X.12578.

van de Kerk, M., Onorato, D.P., et al., 2015. *Journal of Animal Ecology*, 84(2):576–585.

Visser, I. and Speekenbrink, M., 2010. *Journal of Statistical Software*, 36(7):1–21.

Wenger, S.J. and Olden, J.D., 2012. *Methods in Ecology and Evolution*, 3(2):260–267. ISSN 2041210X. doi:10.1111/j.2041-210X.2011.00170.x.