

Model complexity and model choice for animal movement models

Ben Bolker, McMaster University

Departments of Mathematics & Statistics and Biology

ISEC

29 June 2016

Outline

- 1 Florida panthers
- 2 Hidden Markov models
- 3 Basic analysis (van de Kerk et al., 2015)
- 4 Incorporating diurnal variation (Li, 2015)
- 5 Broader issues/outlook

Acknowledgements

People Michael Li, Madelon van de Kerk,
Dave Onorato, Madan Oli; many unnamed field
biologists

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

Funding NSERC Discovery grant, NSF IGERT program

Outline

- 1 Florida panthers
- 2 Hidden Markov models
- 3 Basic analysis (van de Kerk et al., 2015)
- 4 Incorporating diurnal variation (Li, 2015)
- 5 Broader issues/outlook

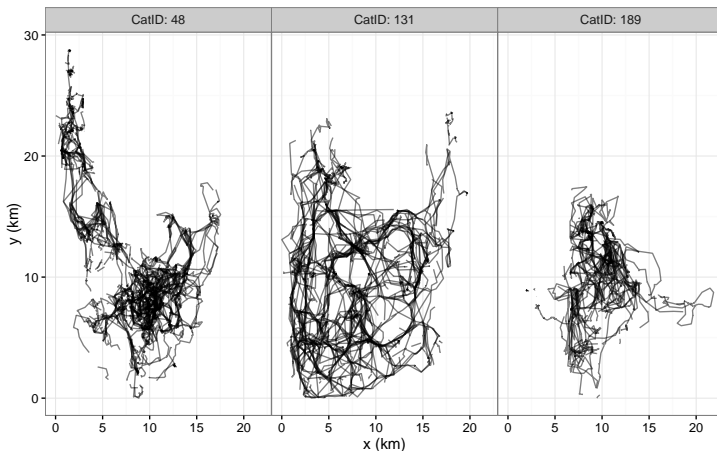
Florida panther movement

- endangered subspecies, *Puma concolor coryi*
- movement variation by sex and life history stage (juvenile, adult, mom with kittens . . .)
- threats
- predicting the effects of changes in habitat, etc.



panther movement tracks

18 GPS-collared animals, 3200 panther days;
49000 locations (hourly); 1-15K observations per animal



Outline

- 1 Florida panthers
- 2 Hidden Markov models**
- 3 Basic analysis (van de Kerk et al., 2015)
- 4 Incorporating diurnal variation (Li, 2015)
- 5 Broader issues/outlook

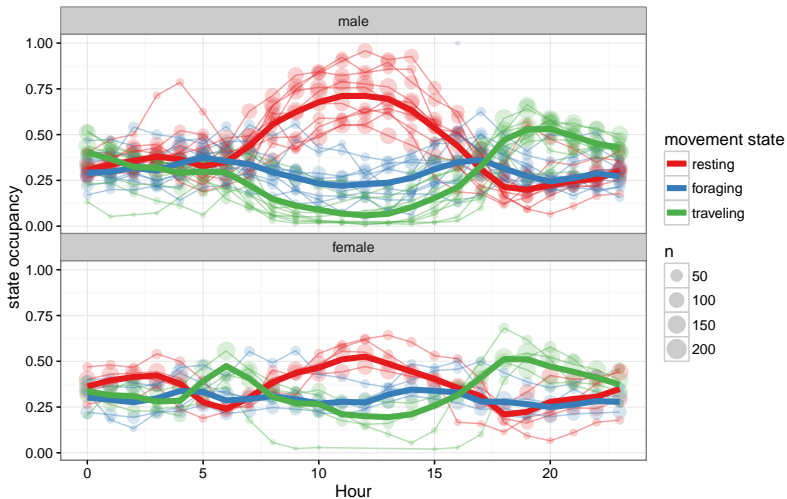
Hidden Markov models in practice

- *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 time distributions* (Langrock, 2011; Augustine, 2016): `move.HMM`

Outline

- 1 Florida panthers
- 2 Hidden Markov models
- 3 Basic analysis (van de Kerk et al., 2015)**
- 4 Incorporating diurnal variation (Li, 2015)
- 5 Broader issues/outlook

Diurnal variation: Viterbi results



what can we conclude so far?

good news

- basic biology: males move faster, farther
- three states are identifiable, sensible
- dwell distributions approximately geometric (HSMM \rightarrow HMM)

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?

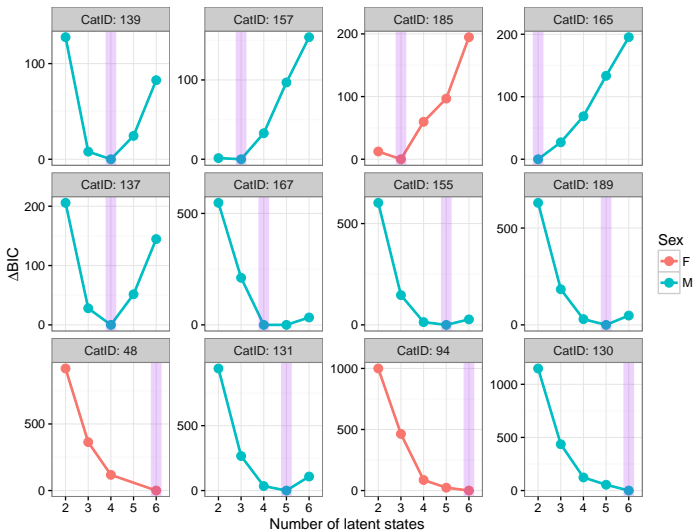
good news

- basic biology: males move faster, farther
- three states are identifiable, sensible
- dwell distributions approximately geometric (HSMM \rightarrow HMM)

bad news

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

Model complexity (bad news)



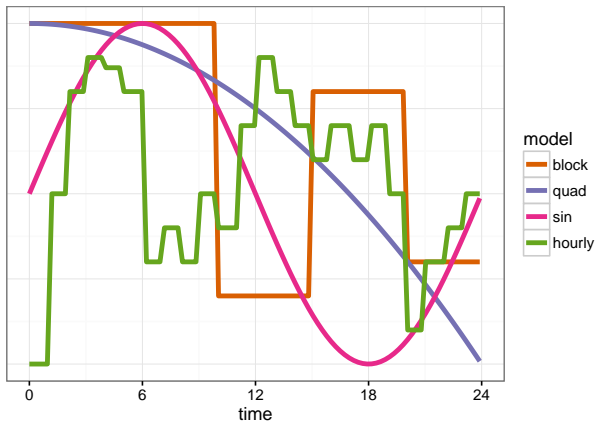
Outline

- 1 Florida panthers
- 2 Hidden Markov models
- 3 Basic analysis (van de Kerk et al., 2015)
- 4 Incorporating diurnal variation (Li, 2015)**
- 5 Broader issues/outlook

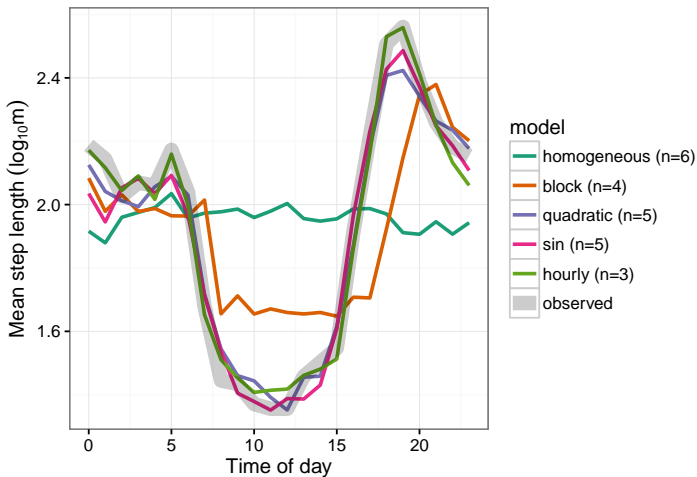
Expanding the model

- allow for diurnal variation:
time-varying transition parameters
- (also try *finite mixture models*
with time-varying occupancy)

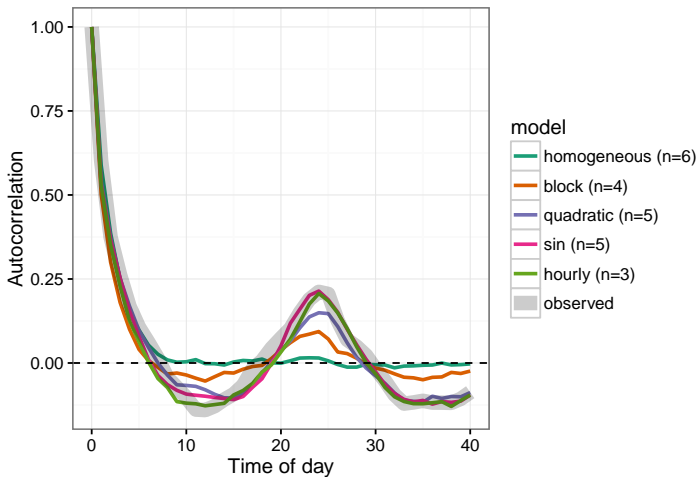
Temporal models



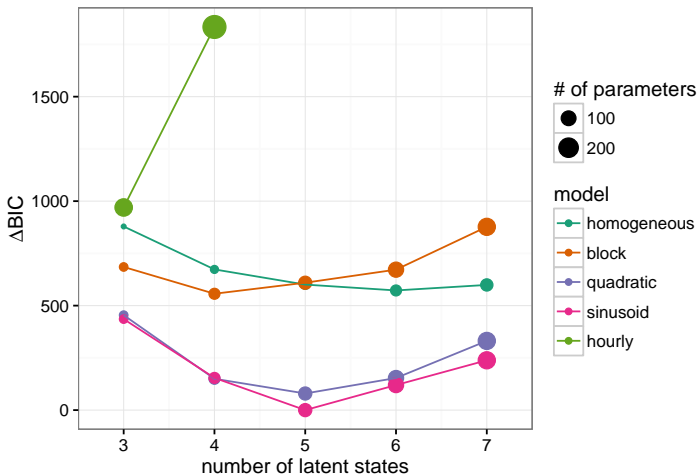
Temporal patterns (step length)



Temporal patterns (autocorrelation)

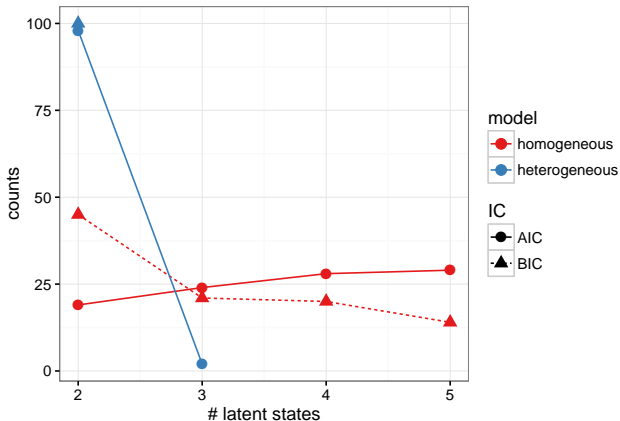


Goodness of fit/model complexity



Model complexity: simulation

(2-state heterogeneous)



Diurnal model: conclusions

- diurnal structure greatly improves fit ($\Delta\text{BIC} \approx 500$)
- slightly improves latent-state issue ($n = 6 \rightarrow 5$)
- be careful using Viterbi to assess model fit (double-dipping)
- lots left to do!
 - seasonal variation
 - incorporate habitat, home range behaviour
 - etc. etc. etc.

Outline

- 1 Florida panthers
- 2 Hidden Markov models
- 3 Basic analysis (van de Kerk et al., 2015)
- 4 Incorporating diurnal variation (Li, 2015)
- 5 Broader issues/outlook**

Big data and small models

- simple model families +
large data →
model misspecification →
overparameterization

- Gelman: “Sample sizes are never large”: ([blog post](#))

N is never enough because if it were “enough” you’d already be on to the next problem for which you need more data.

...i.e. should add *appropriate* complexity to soak up variation

Tools needed

- convenient tools for model checking:
graphical & quantitative (Potts et al., 2014)
- cross-validation (Wenger and Olden, 2012) ?
- flexible computational frameworks
(ecologists can't afford consultants/
there are too many species)



<http://tinyurl.com/panthermoves2>

Panthers
○○

HMM
○○

Basic analysis
○○○

Diurnal model
○○○○○○○

Broader issues/outlook
○○○

References

extras

Hidden Markov models (cont.)

state:

$$S_t \sim \text{Multinomial}(S_{t-1}, \mu_{S,t})$$

$$\mu_{S,t} = \text{multi-logistic}(\mathbf{X}_{S,t} \beta_S)$$

emission:

$$\mathbf{Z}_t \sim \{\text{Dist}_1(\mu_{Z_1,S_t}), \dots, \text{Dist}_n(\mu_{Z_n,S_t})\}$$

$$\mu_{Z_i,S_t} = g^{-1}(\mathbf{X}_{Z_i,t} \beta_{Z_i,S_t})$$

An aside on AIC vs BIC

- “should I use AIC or BIC? I heard that AIC is inconsistent ...”
- complexity penalty = 2 (AIC) vs $\log(n)$ (BIC)
- best prediction vs. model identification (Yang, 2005)
- *effect size spectrum*: tapering or discrete?

