

Individual based population inference using tagging data

Martin W. Pedersen, Uffe H. Thygesen, Henrik Baktoft and Henrik Madsen
Fourth International Science Symposium on Bio-logging
15 March 2011



Technical University of Denmark

Motivation/Objective

The purpose of many bio-logging experiments is to gain knowledge about populations rather than about specific individuals.

Motivation/Objective

The purpose of many bio-logging experiments is to gain knowledge about populations rather than about specific individuals.

However, many bio-logging data are individual based.

Motivation/Objective

The purpose of many bio-logging experiments is to gain knowledge about populations rather than about specific individuals.

However, many bio-logging data are individual based.

Therefore, techniques for analysing multiple individuals are often needed.

Motivation/Objective

The purpose of many bio-logging experiments is to gain knowledge about populations rather than about specific individuals.

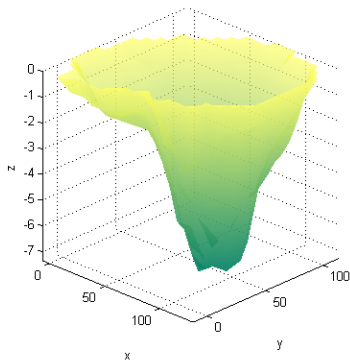
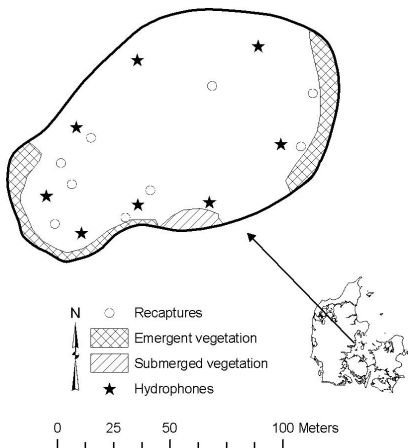
However, many bio-logging data are individual based.

Therefore, techniques for analysing multiple individuals are often needed.

Goal: Develop a simple, flexible, and fast method for inference in groups using data from individuals.

Case: Analysis of acoustic telemetry data

Lake Gosmer with listening grid



Acoustic data logging

Setup

Acoustic data logging

Setup

- ▶ Acoustic transmitters on $M = 20$ pike (*Esox lucius* L.).



Acoustic data logging

Setup

- ▶ Acoustic transmitters on $M = 20$ pike (*Esox lucius* L.).
- ▶ Retrieved data: horizontal location (via BioMap software), depth (via pressure).



Acoustic data logging

Setup

- ▶ Acoustic transmitters on $M = 20$ pike (*Esox lucius* L.).
- ▶ Retrieved data: horizontal location (via BioMap software), depth (via pressure).
- ▶ Sample interval 45 sec. Total recording time: 2 months (~ 100000 obs per individual).



Acoustic data logging

Setup

- ▶ Acoustic transmitters on $M = 20$ pike (*Esox lucius* L.).
- ▶ Retrieved data: horizontal location (via BioMap software), depth (via pressure).
- ▶ Sample interval 45 sec. Total recording time: 2 months (~ 100000 obs per individual).
- ▶ Filtering to reduce the effect of outliers.



Acoustic data logging

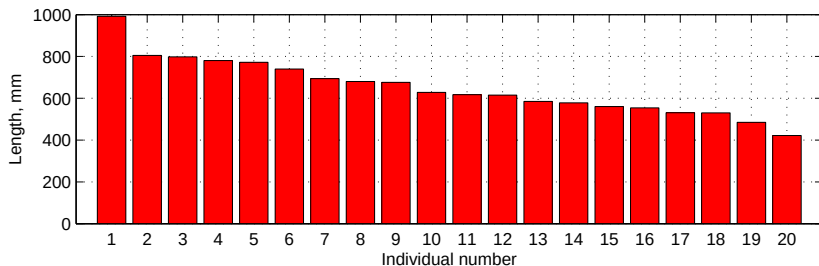
Setup

- ▶ Acoustic transmitters on $M = 20$ pike (*Esox lucius* L.).
- ▶ Retrieved data: horizontal location (via BioMap software), depth (via pressure).
- ▶ Sample interval 45 sec. Total recording time: 2 months (~ 100000 obs per individual).
- ▶ Filtering to reduce the effect of outliers.
- ▶ Location uncertainty is negligible after filtering.



The population of tagged pike

Numbered and sorted by length:



Case: Acoustic data
ooo●

Individual analysis
oo

Population analysis
oo

Population inference
ooooo

Conclusion
o

Population inference

Model for individual analysis

Hidden Markov model (HMM) for movement behaviour

Setup a 2-state hidden Markov model for the unobservable movement behaviour, B_t , of an individual pike.

Hidden Markov model (HMM) for movement behaviour

Setup a 2-state hidden Markov model for the unobservable movement behaviour, B_t , of an individual pike.

- Movement behaviours: $B_t \in \{\text{resting}, \text{moving}\}$.

Hidden Markov model (HMM) for movement behaviour

Setup a 2-state hidden Markov model for the unobservable movement behaviour, B_t , of an individual pike.

- ▶ Movement behaviours: $B_t \in \{\text{resting}, \text{moving}\}$.
- ▶ Observations: Average speed (body length per second) between locations.

Hidden Markov model (HMM) for movement behaviour

Setup a 2-state hidden Markov model for the unobservable movement behaviour, B_t , of an individual pike.

- ▶ Movement behaviours: $B_t \in \{\text{resting}, \text{moving}\}$.
- ▶ Observations: Average speed (body length per second) between locations.
- ▶ Covariate information: Time of day, $\tau \in \{\text{day}, \text{night}\}$.

Hidden Markov model (HMM) for movement behaviour

Setup a 2-state hidden Markov model for the unobservable movement behaviour, B_t , of an individual pike.

- ▶ Movement behaviours: $B_t \in \{\text{resting}, \text{moving}\}$.
- ▶ Observations: Average speed (body length per second) between locations.
- ▶ Covariate information: Time of day, $\tau \in \{\text{day}, \text{night}\}$.
- ▶ Estimate model parameters with maximum likelihood.

Hidden Markov model (HMM) for movement behaviour

Setup a 2-state hidden Markov model for the unobservable movement behaviour, B_t , of an individual pike.

- ▶ Movement behaviours: $B_t \in \{\text{resting}, \text{moving}\}$.
- ▶ Observations: Average speed (body length per second) between locations.
- ▶ Covariate information: Time of day, $\tau \in \{\text{day}, \text{night}\}$.
- ▶ Estimate model parameters with maximum likelihood.

Summarize estimates via the stationary distribution of B_t for individual i .

Hidden Markov model (HMM) for movement behaviour

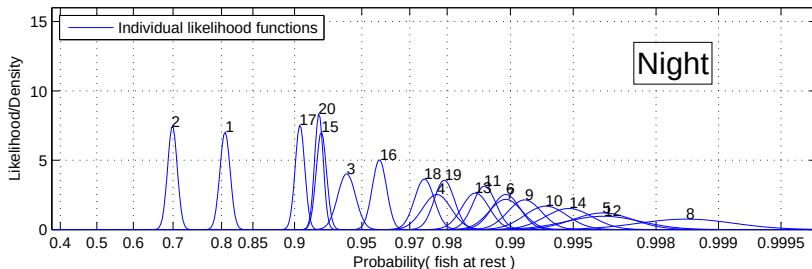
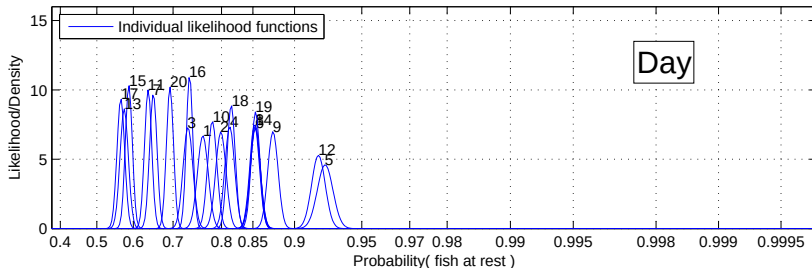
Setup a 2-state hidden Markov model for the unobservable movement behaviour, B_t , of an individual pike.

- ▶ Movement behaviours: $B_t \in \{\text{resting}, \text{moving}\}$.
- ▶ Observations: Average speed (body length per second) between locations.
- ▶ Covariate information: Time of day, $\tau \in \{\text{day}, \text{night}\}$.
- ▶ Estimate model parameters with maximum likelihood.

Summarize estimates via the stationary distribution of B_t for individual i .

So, $\theta_i = [\theta^{(\text{day})}, \theta^{(\text{night})}]_i$ is the probability that i is resting at day and night time respectively.

Individual likelihood functions



Population analysis

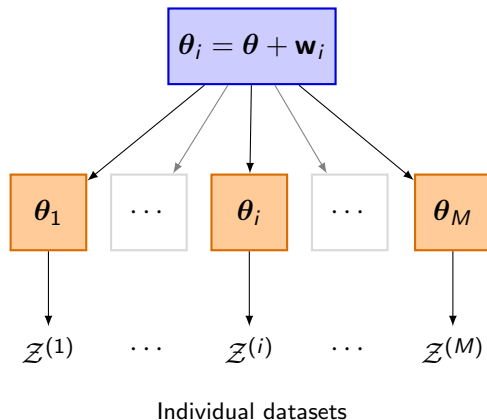
combining individual analyses

Combining individual analyses

Mixed effects model:

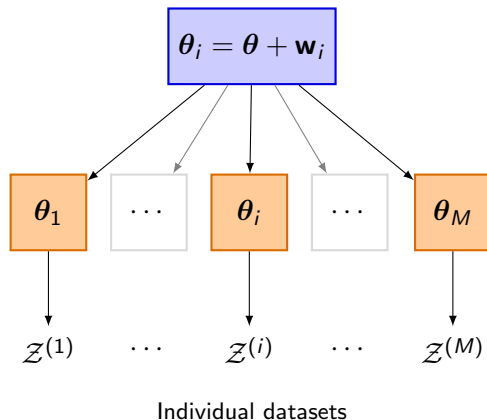
Combining individual analyses

Mixed effects model:



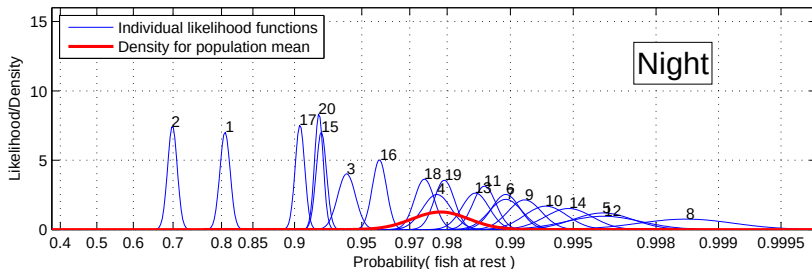
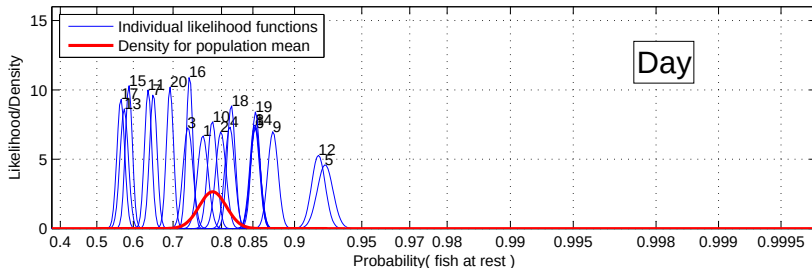
Combining individual analyses

Mixed effects model:



Random effects: $\mathbf{w}_i \sim N(\mathbf{0}, \mathbf{W})$, Population parameters: θ and $\mathbf{W}$.

Individuals and population



Population inference

Difference between day and night?

Population parameter estimates (θ is probability fish is at rest):

$$\theta^{(\text{day})} = 0.7838, \quad \theta^{(\text{night})} = 0.9785.$$

Difference between day and night?

Population parameter estimates (θ is probability fish is at rest):

$$\theta^{(\text{day})} = 0.7838, \quad \theta^{(\text{night})} = 0.9785.$$

Standard statistical tests conclude that day and night time behaviour are significantly different (p-value 3.73×10^{-13})

Deviating individuals?

Leave-one-out procedure (for night or day):

Deviating individuals?

Leave-one-out procedure (for night or day):

Loop over $j \in \{1, \dots, M\}$

- ▶ Leave out individual j from the population.

Deviating individuals?

Leave-one-out procedure (for night or day):

Loop over $j \in \{1, \dots, M\}$

- ▶ Leave out individual j from the population.
- ▶ For the reduced population calculate the ML estimates: $\hat{\theta}$, $\hat{W}$.

Deviating individuals?

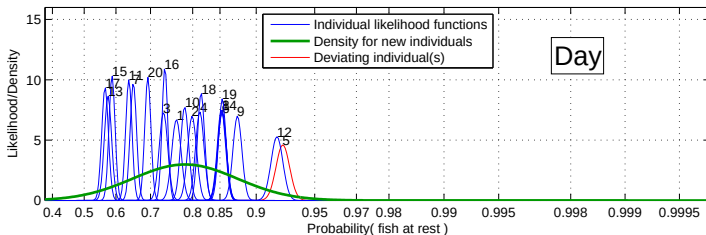
Leave-one-out procedure (for night or day):

Loop over $j \in \{1, \dots, M\}$

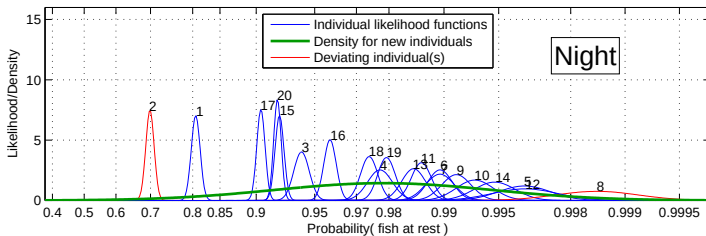
- ▶ Leave out individual j from the population.
- ▶ For the reduced population calculate the ML estimates: $\hat{\theta}$, $\hat{W}$.
- ▶ Test in a χ^2 -distribution if individual j is significantly different from the reduced population.

Leave-one-out (1/3)

Day, left out:

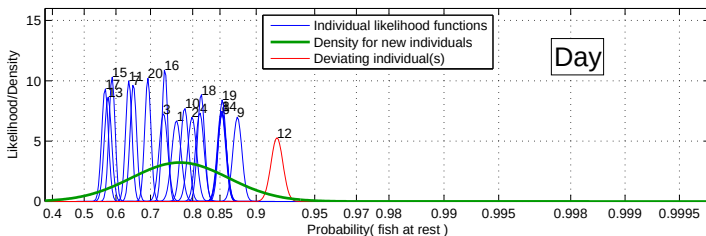


Night, left out:

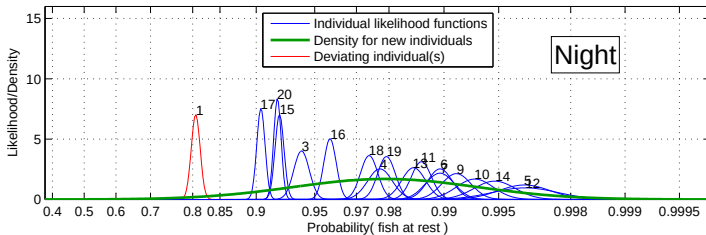


Leave-one-out (2/3)

Day, left out: 5

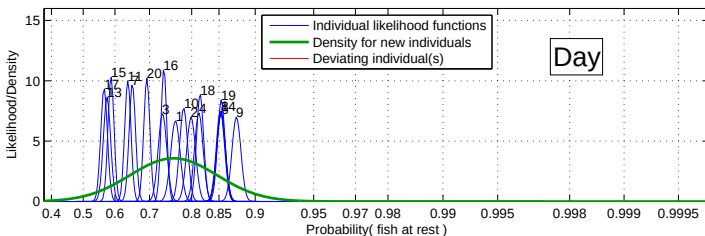


Night, left out: 2,8

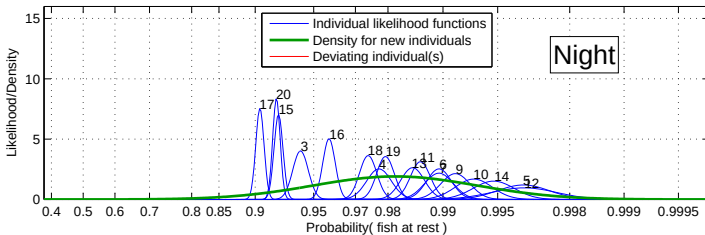


Leave-one-out (3/3)

Day, left out: 5,12



Night, left out: 2,8,1



Conclusion

Conclusion

Key assumption: Individual likelihood functions must be well approximated by a Gaussian.

Conclusion

Key assumption: Individual likelihood functions must be well approximated by a Gaussian.

Important properties of the method:

Conclusion

Key assumption: Individual likelihood functions must be well approximated by a Gaussian.

Important properties of the method:

- ▶ The technique is compatible with any type of individual model estimated with ML.

Conclusion

Key assumption: Individual likelihood functions must be well approximated by a Gaussian.

Important properties of the method:

- ▶ The technique is compatible with any type of individual model estimated with ML.
- ▶ Individuals are estimated independently, easy to include new data.

Conclusion

Key assumption: Individual likelihood functions must be well approximated by a Gaussian.

Important properties of the method:

- ▶ The technique is compatible with any type of individual model estimated with ML.
- ▶ Individuals are estimated independently, easy to include new data.
- ▶ Population estimation is fast (seconds).

Conclusion

Key assumption: Individual likelihood functions must be well approximated by a Gaussian.

Important properties of the method:

- ▶ The technique is compatible with any type of individual model estimated with ML.
- ▶ Individuals are estimated independently, easy to include new data.
- ▶ Population estimation is fast (seconds).
- ▶ Individuals deviating from the population can be identified.

Conclusion

Key assumption: Individual likelihood functions must be well approximated by a Gaussian.

Important properties of the method:

- ▶ The technique is compatible with any type of individual model estimated with ML.
- ▶ Individuals are estimated independently, easy to include new data.
- ▶ Population estimation is fast (seconds).
- ▶ Individuals deviating from the population can be identified.

Thank you for listening!