

A hidden semi-Markov model for inferring the structure of migratory bird flyway networks

Nathalie Peyrard, Régis Sabbadin, Marie-Josée Cros,
Ronan Trépos, Sam Nicol

Netbio, 2021



For the conservation of shorebirds

Which stopover sites should be protected in priority?

- 50 millions of shorebirds migrate from Siberia to Australasia
- Migration is threatened by human developments along the migratory route

Coastal Development

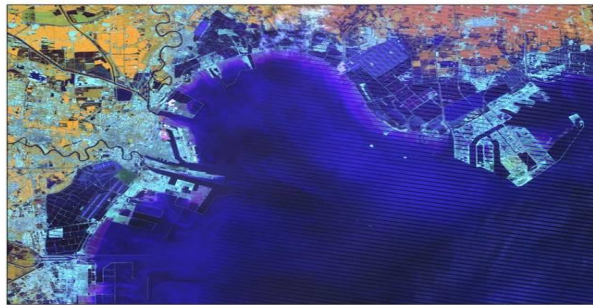
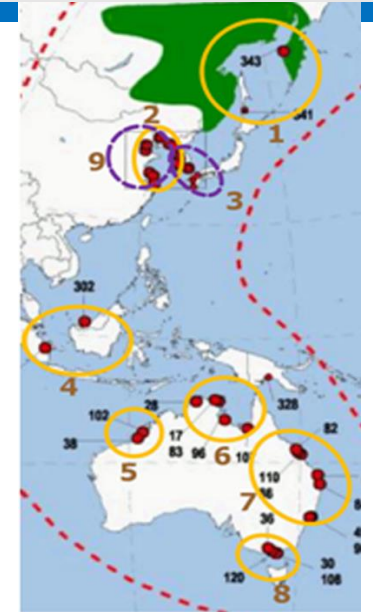


Image: Nick Murray, University of Queensland

Sea Level Rise



East Asian Australasian flyway



From Iwamura et al. 2013

But ... prioritising is difficult because only partial knowledge on the main routes

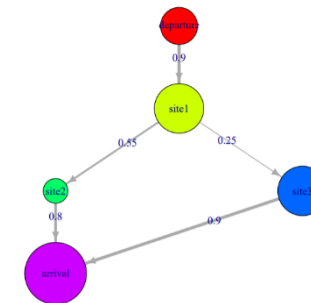
How to infer the most likely migratory routes ?



- Given that
 - Monitoring flows is difficult
 - Observed count data are imperfect (detection error)

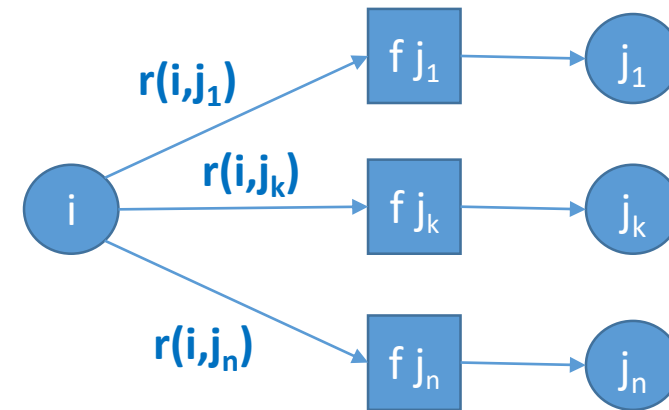
➤ We propose to combine networks and statistical inference

- A representation of possible migratory routes as a **directed and weighted graph**
- A **Hidden Semi Markov Model** of birds trajectories and observed counts
- Adapted estimation **algorithms**



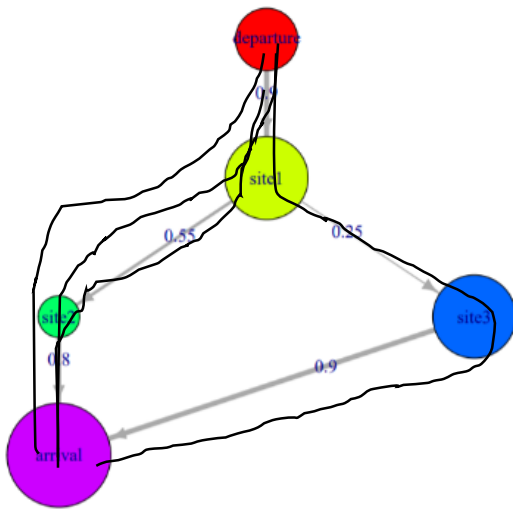
The BIRDNET HSMM model formally: a discrete time multidimensional HSMM

- **Hidden variable:** the state of each bird at each time step, among
 - Being at site 1, ... I, flying towards a site, dead
 - **Sojourn time** at site i : **Poisson**(λ_i)
 - **Transition probabilities**
 - from 'site i ' to 'flying towards j ' : **$r(i,j)$**
 - from 'site i ' to 'death': $r(i,\text{death})$, assumed known
 - from 'flying towards j ' to 'site j ' : 1
 - **Observed variable** : noisy bird counts for some time steps and some sites : **Poisson**($N(i,t)$)
-
- **Inferring the flyway network amounts to estimate the transition rates $r(i,j)$**
 - **By product: estimation of the mean sojourn times λ_i**

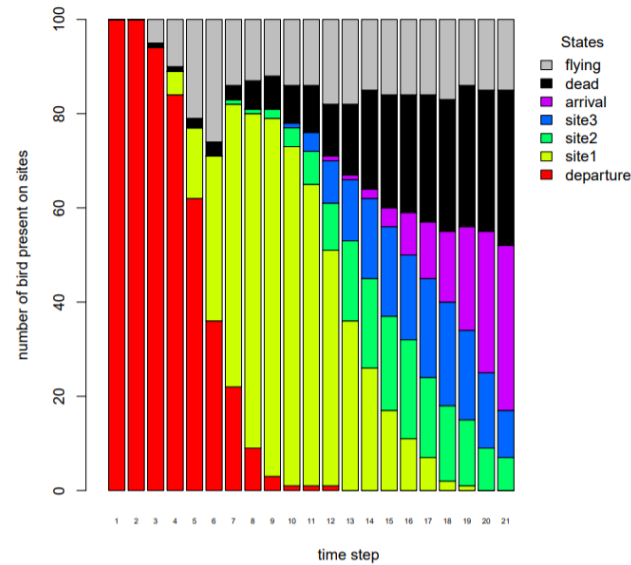


The BIRDNET HSMM model as a simulator

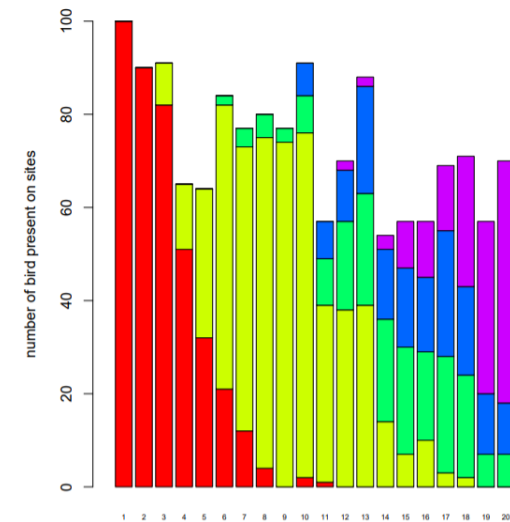
1. Simulate the hidden trajectories using
 - Sojourn time
 - Transition rates



2. Compute the exact number of birds at each site and each time step



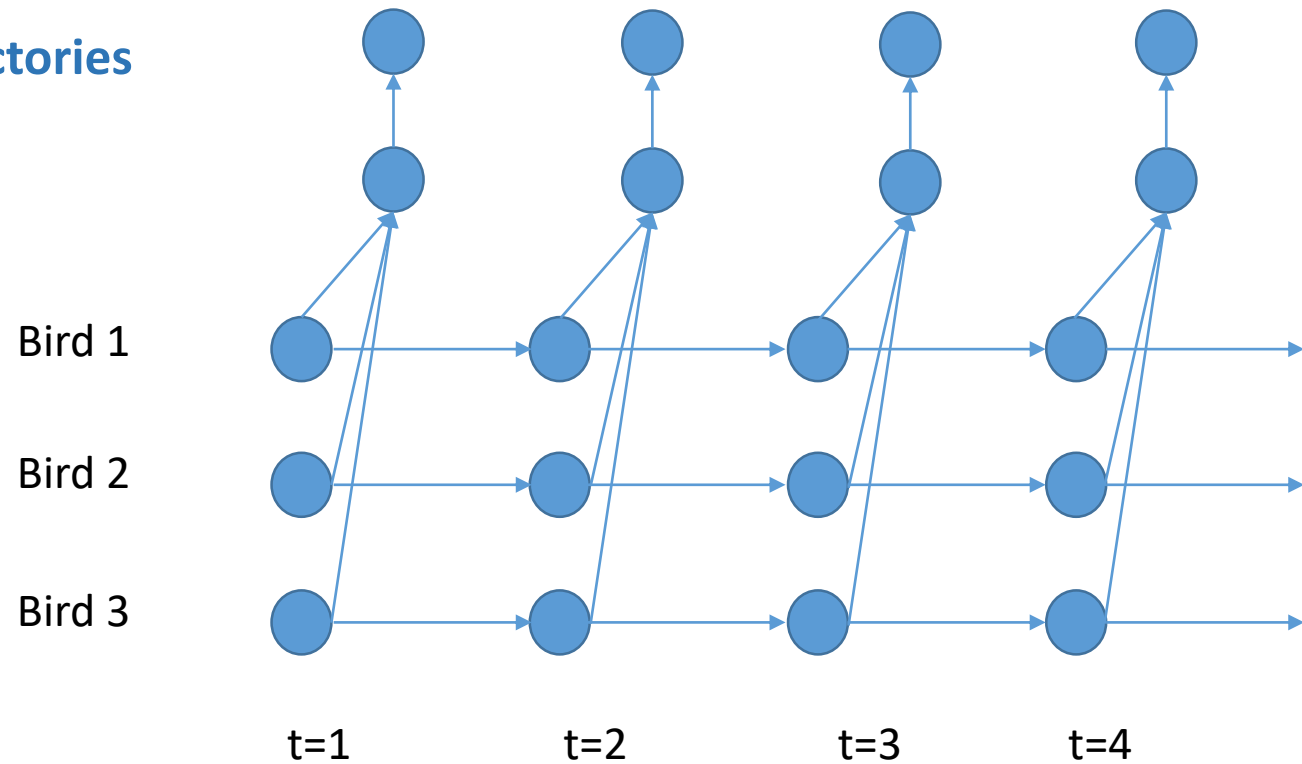
3. Simulate a noisy version of the exact counts



How to estimate transition probabilities (and sojourn times)?

BIRDNET model is **easy to simulate** but
likelihood is **complex to optimize**:

- hidden variables
- **observations make bird trajectories dependent !**



How to estimate transition probabilities (and sojourn times)?

BIRDNET model is **easy to simulate** but likelihood is **complex to optimize**:

- hidden variables
- **observations make bird trajectories dependent !**

- **Existing estimation tools reach their limits**
- We explore several options for approximate estimation

	Deterministic algorithm	Simulation-based algorithm
Frequentist estimation	EM	Monte Carlo EM (MCEM)
Bayesian estimation	Variational Bayes EM (VBEM)	ABC



Outline

- ▶ A factorial Hidden semi markov model for modeling the Australasian migration of birds
- ▶ MCEM Metropolis Hastings algorithm
- ▶ An ABC algorithm
- ▶ Experiments: Benchmarks + real data



Single bird semi-Markov trajectory model

- ▶ l sites, $T + 1$ time steps ($H = \{0, \dots, T\}$).
- ▶ Bird n 's ($1 \leq n \leq N$) trajectory uniquely defined by:

$$\pi_n = \left((i_0^n, t_0^n), (i_1^n, t_1^n), \dots, (i_{F_n}^n, t_{F_n}^n) \right). \quad (1)$$

given that flight times (f_{ij}) are fixed and known.

t_k^n is the arrival time at site $i_k^n \in I \cup \{\text{death}\}$.

$t_0^n = 0 \leq t_1^n \leq \dots \leq t_{F_n}^n \leq T$.

- ▶ Sojourn time model: $t_{k+1}^n - t_k^n - f_{i_k i_{k+1}} = \tau_{i_k}$ where:

$$P_{\lambda_{i_k}}(\tau_{i_k} = d) = \frac{(\lambda_{i_k})^{d-1}}{(d-1)!} e^{-\lambda_{i_k}}, \forall d = 1, 2, \dots \quad (2)$$

- ▶ Transition probabilities: $R(i, j) > 0$ iff $i < j$ (no way back) and $\mu_i = 1 - \sum_{j>i} R(i, j)$ is the mortality rate in site i .
- ▶ Model parameters: $(\lambda_{i_k})_{1 \leq i_k \leq l}$ and $(R(i, j))_{1 \leq i < j \leq l}$.

► Birds population's trajectory model.

Single bird's trajectory π_n semi-Markov model:

$$P_{r,\lambda}(\pi_n) = \prod_{k=0}^{F_n-1} R(i_k^n, i_{k+1}^n) P_{\lambda_{i_k^n}}(\tau = t_{k+1}^n - t_k^n - f_{i_k^n i_{k+1}^n}). \quad (3)$$

Let $\Pi = \{\pi_1, \dots, \pi_N\}$ denote the set of trajectories of the N birds. Since trajectories are independent, we have:

$$P_{r,\lambda}(\Pi) = \prod_{n=1}^N \prod_{k=0}^{F_n-1} R(i_k^n, i_{k+1}^n) P_{\lambda_{i_k^n}}(\tau = t_{k+1}^n - t_k^n - f_{i_k^n i_{k+1}^n}).$$



Birds population observation model

$N_i^t(\Pi)$ is the number of birds located in site i at time t :

$$N_i^t(\Pi) = \left| \left\{ \pi_n \in \Pi, \exists k \in \{0, \dots, F_n - 1\}, i_k^n = i \right. \right. \\ \left. \left. \text{and } t_k^n \leq t < t_{k+1}^n - f_{i_k^n i_{k+1}^n} \right\} \right|.$$

For a set $\Omega \subseteq I \times H$ of observed site-times, the set of observations is represented by a matrix O of observed bird counts, of size $I \times T$:

$$O(i, t) = O_i^t \text{ if } (i, t) \in \Omega \text{ and NA else} \quad (4)$$

The joint distribution of all the observations $O = \{O_i^t\}$ given the trajectories is

$$P(O|\Pi) = \prod_{(i,t) \in \Omega} \mathcal{P}_{N_i^t}(O_i^t). \quad (5)$$

where $\mathcal{P}_{N_i^t}$ is a Poisson distribution of mean-variance N_i^t .



Outline

- ▶ A factorial Hidden semi markov model for modeling the Australasian migration of birds
- ▶ MCEM Metropolis Hastings algorithm
- ▶ An ABC algorithm
- ▶ Experiments: Benchmarks + real data



➤ Expectation Maximization algorithm (1)

Our objective is to compute Λ^* , the values of parameters which *maximize* the likelihood of observations O (in fact, it is more convenient to optimize the log-likelihood, which is equivalent):

$$\Lambda^* = \arg \max_{\Lambda} \log (P_{\Lambda}(O)),$$

$$\Lambda^* = \arg \max_{\Lambda} \log \int_{\Pi} P_{\Lambda}(O, \Pi) d\Pi.$$



> Expectation Maximization algorithm (2)

The *Expectation-Maximization* algorithm is an iterative algorithm that seeks to find an approximation of Λ^* by iterating updates of estimates of Λ^* as follows:

$$\begin{aligned}\Lambda_{new} &= \arg \max_{\Lambda} E_{\Pi}[\log(P_{\Lambda}(\Pi, O)) \mid \Lambda_{old}, O] \\ &= \arg \max_{\Lambda} \int_{\Pi} \log(P_{\Lambda}(\Pi, O)) P_{\Lambda_{old}}(\Pi \mid O) d\Pi.\end{aligned}$$



➤ Expectation Maximization algorithm (3)

The following convenient notation is used for computations:

$$Q(\Lambda|\Lambda_{old}) = \int_{\Pi} \log(P_{\Lambda}(\Pi, O)) P_{\Lambda_{old}}(\Pi|O) d\Pi.$$

This gives the usual E-M representation of the algorithm:

- ▶ Expectation. Compute the quantities involving Λ_{old} in $Q(\Lambda|\Lambda_{old})$.
- ▶ Maximization. Compute

$$\Lambda_{new} = \arg \max_{\Lambda} Q(\Lambda|\Lambda_{old}).$$



▶ Monte-Carlo Expectation Maximization

The Expectation step is intractable. Indeed, the domain of sets of trajectories is too large.

Provided that we are able to simulate $P_{\Lambda_{old}}(\Pi|O)$, the integral can be approximated, leading to a Monte-Carlo version of EM:

- ▶ Generate M samples of the N bird trajectories, $\{\Pi^{(1)}, \dots, \Pi^{(M)}\}$ from $P_{\Lambda_{old}}(\Pi|O)$.
- ▶ Approximate $Q(\Lambda|\Lambda_{old})$ with

$$\hat{Q}(\Lambda|\Lambda_{old}) = \frac{1}{M} \sum_{m=1}^M \log \left(P_{\Lambda}(\Pi^{(m)}, O) \right). \quad (6)$$

- ▶ Compute Λ_{new} as:

$$\Lambda_{new} = \arg \max_{\Lambda} \hat{Q}(\Lambda|\Lambda_{old}). \quad (7)$$

In practice, steps (6) and (7) are grouped together, in order to build a single optimization program defining update $\Lambda_{old} \rightarrow \Lambda_{new}$.



➤ MCEM updates in Birdnet

- ▶ Generate $\{\Pi^{(1)}, \dots, \Pi^{(M)}\}$ from $P_{\Lambda_{old}}(\Pi|O)$.
- ▶ The r_{ij}^{new} are empirical frequencies:

$$r_{ij}^{new} = \frac{\sum_{m=1}^M |\{\pi_n \in \Pi^{(m)} \text{ s.t. } \exists k, i_k^n = i \text{ and } i_{k+1}^n = j\}|}{\sum_{m=1}^M |\{\pi_n \in \Pi^{(m)} \text{ s.t. } \exists k, i_k^n = i\}|}.$$

- ▶ The λ_i^{new} are obtained from empirical sojourn times:

$$\lambda_i^{new} = \frac{\sum_{m=1}^M \sum_{\pi_n \in \Pi^{(m)}} \sum_{k < F_n} (t_{k+1}^n - t_k^n - f_{i_k^n i_{k+1}^n}) \mathbb{1}_{\{i_k^n = i\}}}{\sum_{m=1}^M \sum_{\pi_n \in \Pi^{(m)}} \sum_{k < F_n} \mathbb{1}_{\{i_k^n = i\}}} - 1.$$

So, the hard part is to generate sample populations' trajectories from $P_{\Lambda_{old}}(\Pi|O)$.



Metropolis-Hastings algorithm

Generation of M samples of the N bird trajectories, $\{\Pi^{(1)}, \dots, \Pi^{(M)}\}$ from $P_{\Lambda_{old}}(\Pi|O)$.

- ▶ A first sample $\Pi^{(1)}$ is simulated using $P_{\Lambda_{old}}(\Pi)$.
- ▶ For m in $\{1, \dots, M-1\}$:
Every $\pi_i \in \Pi^{(m)}$ is kept (with probability $1-p$) or re-simulated according to $P_{\Lambda_{old}}(\cdot)$ (with probability p).
The result is Π .
- ▶ Rejection or acceptance of Π according to $\alpha(\Pi^{(m)}, \Pi)$:

$$\alpha(\Pi^{(m)}, \Pi) = \min\left\{1, \frac{P(O|\Pi)}{P(O|\Pi^{(m)})}\right\}$$

Select $\Pi^{(m+1)} \leftarrow \Pi$ with probability $\alpha(\Pi^{(m)}, \Pi)$

$\leftarrow \Pi^{(m)}$ with probability $1 - \alpha(\Pi^{(m)}, \Pi)$

p is the only parameter of the MH algorithm. $p = 0.1$ was chosen.

➤ Take-away points of the MCEM approach

- ▶ MCEM updates given $\{\Pi^{(1)}, \dots, \Pi^{(M)}\}$ are easy (averaging)
- ▶ Getting simulation samples from $P_{\Lambda_{old}}(\Pi|0)$ is hard \Rightarrow Metropolis Hastings
- ▶ The algorithm is slow
- ▶ it is prone to degeneracy (if edge $i \rightarrow j$ unobserved in $\Pi^{(m)}$ it will never be observed again \Rightarrow hard coding of transition probabilities lower bounds)

We also tried ABC to overcome the degeneracy problem (and spare computation time).



Outline

- ▶ A factorial Hidden semi markov model for modeling the Australasian migration of birds
- ▶ MCEM Metropolis Hastings algorithm
- ▶ An ABC algorithm
- ▶ Experiments: Benchmarks + real data



➤ Approximate Bayesian Computation

- ▶ Adaptive Population Monte Carlo Approximate Bayesian Computation (Lenormand et al. 2012)
- ▶ Provides a distribution of parameter values, starting from uniform priors
- ▶ Iterative accept-reject algorithm, particle filter approach
- ▶ No computation of likelihood, relies on a distance ρ (default behavior)

$$\rho(x, O) = \sum_{(s,t) \in \Omega} \frac{(x_{(s,t)} - O_{(s,t)})^2}{V_{(s,t)}} \quad \text{with } x \sim f(x|\Lambda) \quad (8)$$



ABC initialization

- 1 Generate $j = 1, \dots, N$ particles from uniform prior (optimized LHS) and evaluate them:
 - 1a $\Lambda_j \sim \pi(\Lambda)$ (generation)
 - 1b $x^{(j)} \sim f(x|\Lambda_j)$ (simulation)
 - 1c Using $\{x^{(j)}\}_{1 \leq j \leq N}$ compute the variance of statistics $v_{(s,t)}$ (statistics ponderation, definition of ρ)
 - 1d $\rho_j = \rho(x^{(j)}, O)$ (evaluation)
 - 1e $\omega_j = 1$ (particle weight)
- 2 Select subset of particles
 - 2a Let ϵ the α -quantile of $\{\rho_j\}_{1 \leq j \leq N}$
 - 2b Let $\{(\Lambda_i, \omega_i, \rho_i)\}_{1 \leq i \leq N_\alpha} = \{(\Lambda_j, \omega_j, \rho_j) | \rho_j \leq \epsilon\}_{1 \leq j \leq N}$
 - 2c Take σ^2 as twice the empirical variance of $\{(\Lambda_i, \omega_i)\}_{1 \leq i \leq N_\alpha}$



ABC iteration

- 1 Let current population be : $\{(\Lambda_i, \omega_i, \rho_i)\}_{1 \leq i \leq N\alpha}$
- 2 Generate new particles : $\{(\Lambda_j, \omega_j, \rho_j)\}_{N\alpha+1 \leq j \leq N}$
 - 2a $(\Lambda_j | \Lambda_k) \sim \mathcal{N}(\Lambda_k, \sigma^2)$ with $k \in \{1, \dots, N\alpha\}$ according probs. $\{\omega_1, \dots, \omega_{N\alpha}\}$
 - 2b $\rho_j = \rho(O, x)$ with $x \sim f(x, \Lambda_j)$
 - 2c $\omega_j = \frac{\pi(\Lambda_j)}{\sum_{n=1}^{N\alpha} (\omega_n / \sum_{m=1}^{N\alpha} \omega_m) \sigma^{-1} \psi(\sigma^{-1}(\Lambda_j - \Lambda_n))}$
- 3 Update population:
 - 3a $p_{acc} = \frac{1}{N - N\alpha} \sum_{j=N\alpha+1}^N \mathbb{1}_{\rho_j < \epsilon}$ (stopping criterion, proved to converge towards 0)
 - 3b Update ϵ the α -quantile of $\{\rho_p\}_{1 \leq p \leq N}$
 - 3c Update $\{(\Lambda_i, \omega_i, \rho_i)\}_{1 \leq i \leq N\alpha} = \{(\Lambda_p, \omega_p, \rho_p) | \rho_p \leq \epsilon\}_{1 \leq p \leq N}$
 - 3d Take σ^2 as twice the empirical variance of $\{(\Lambda_i, \omega_i)\}_{1 \leq i \leq N\alpha}$



Outline

- ▶ A factorial Hidden semi markov model for modeling the Australasian migration of birds
- ▶ MCEM Metropolis Hastings algorithm
- ▶ An ABC algorithm
- ▶ Experiments: Benchmarks + real data





Benchmark experiments settings

- ▶ Migration graph structure with 4 to 10 nodes.
 - ▶ A maximal number of destination nodes per node ranging from 2 to 4.
 - ▶ Mean sojourn time range is $[1, 3]$ weeks. Five replicates by structure settings (different mean sojourn time and transition probabilities). The first replicate is designed with equiprobable transitions and mean sojourn times of 2 weeks per node.
 - ▶ For each problem replicate, an observation matrix is simulated and only a subset of week-site observations are kept for estimation (100%, 70%, 50%).
- 300 estimations for both MCEM and ABC.
- 4 to 30 parameters to estimate by problem to solve.



Quantitative evaluation

- Likelihood evaluation:

$$P(O|\Lambda) = \int_{\Pi} P(O|\Pi, \Lambda) * P(\Pi|\Lambda) d\Pi \quad (9)$$

$$= \int_{\Pi} P(O|\Pi) * P(\Pi|\Lambda) d\Pi \quad (10)$$

$$\approx \frac{1}{M} \sum_{i=1}^{i=M} P(O|\Pi^i) \text{ with } \Pi^i \sim P(\cdot|\Lambda) \quad (11)$$

- Mean absolute error for N parameters; λ_i (true value) and $\tilde{\lambda}_i$ (est. value) rescaled in $[0; 1]$

$$MeanAE = \frac{1}{N} \sum_{i=1}^{i=N} |\tilde{\lambda}_i - \lambda_i| \quad (12)$$

Note: ABC estimator is the Venter mode of the marginals of ABC posterior distribution



Illustration on a problem with 8 nodes

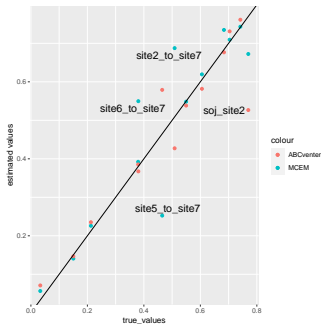
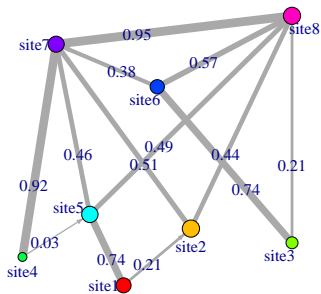
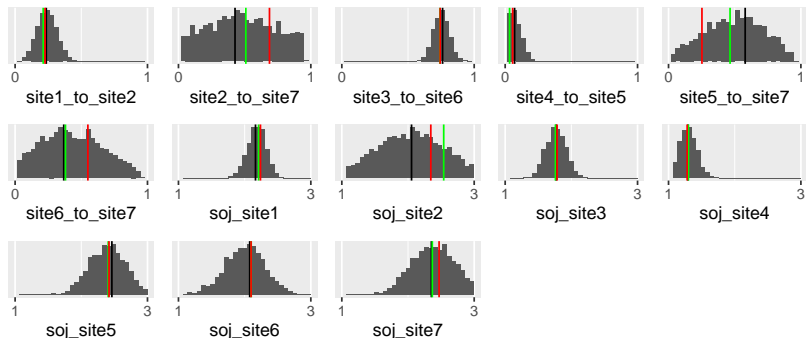


Figure: The true network to estimate (left) and the estimated parameters for both ABC and MCEM (right)

➤ Marginals of ABC posterior distribution



- ▶ ABC mode Venter (black), MCEM (red), true values (green)
- ▶ MeanAE : 0.0469 (ABC mode), 0.0608 (MCEM)
- ▶ Likelihood : -631 (ABC mode), -544 (MCEM), -582 (true)

> Comparisons of methods

	MCEM		true parameters	
	meanAE	log-likelihood	meanAE	log-likelihood
ABCventer	0.01	-551.50	0.08	-523.13
MCEM			0.06	28.38

Table: Comparison of meanAE and log-likelihood for the two estimation methods. Differences between *meanAE* and log-likelihood are all significant ($p < 0.01$ using a paired Wilcoxon test).



Impact of the number of nodes

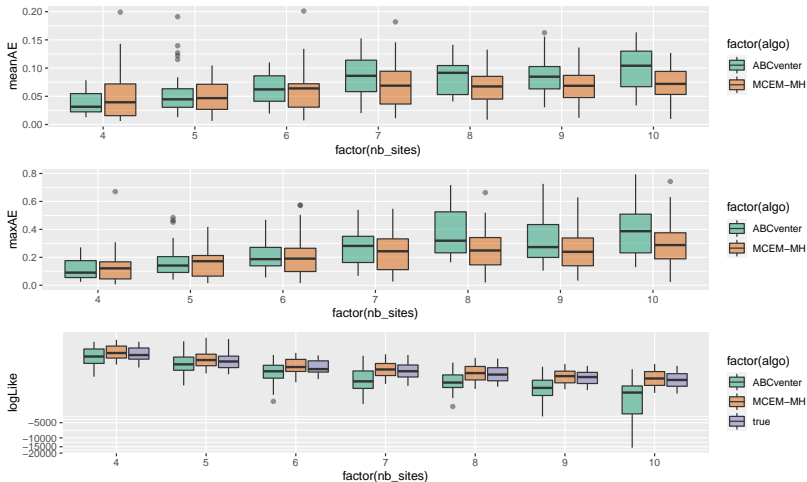


Figure: Mean absolute error (above), maximal absolute error (center) and log likelihood (below).



Other impacts of benchmark settings

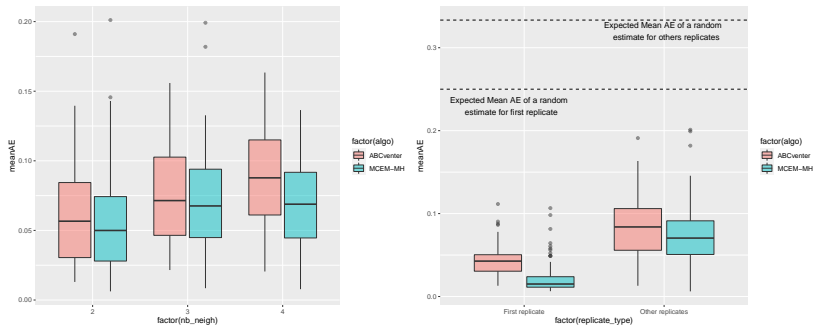


Figure: Mean absolute error according to the maximal number of destination sites per site (left) and mean absolute error according the replicate type (right).



Time of computation

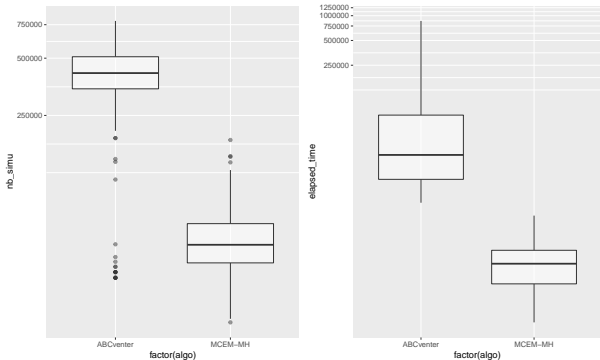


Figure: Number of simulations (left) and elapsed time in seconds (right) for the two methods.



Far Eastern Curlew case study

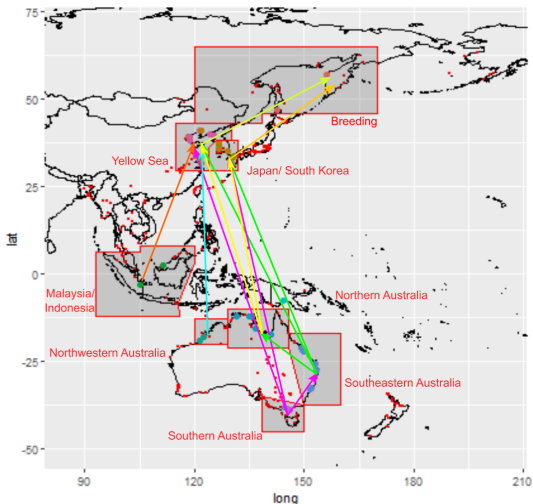


Figure: Departure and stopover sites of FE Curlew northward migration



Far Eastern Curlew case study

- ▶ Migration settings: 8 sites/nodes, 12 parameters
- ▶ Count data from citizen science data base eBird :
<https://ebird.org/>
- ▶ Unbiasing extracted data (temporal and spatial biases)
- ▶ Aggregation/extrapolation to larger study sites
- ▶ Transition probabilities and sojourn mean time estimation for years 2018 and 2019 (ongoing work)





Conclusion

- ▶ Birds migration formalized as Factorial Hidden Semi Markov Model.
- ▶ Two contrasted methods for estimating parameters (transition probabilities and mean sojourn times) :
 - ▶ Monte Carlo Expectation Maximization : frequentist approach, based on likelihood maximization, adhoc
 - ▶ Approximate Bayesian Computation: Bayesian approach, based on empirical statistics, blax-box generic method (package R EasyABC)
- ▶ Satisfying results on benchmark simulated experiments (up to 10 nodes and 30 parameters)
- ▶ Count data extracted from eBird databases.





Future work

- ▶ Provide final results for Far Eastern Curlew case study
- ▶ Variational Based Expectation Maximization method (without simulation)
- ▶ Integrate birds trajectories from GPS tracking systems

