Supporting information for FlywayNet: A hidden semi-Markov model for inferring the structure of migratory bird flyway networks

S1. Model construction

Birds trajectories and populations

Under our HSMM model, the probability of a trajectory π_n conditional on the initial state is equal to

$$P_{R,\lambda}(\pi_n \mid i_0^n) = \prod_{k=0}^{F_n-1} R\left(i_k^n, i_{k+1}^n\right) P_{\lambda_{i_k^n}}\left(\tau = t_{k+1}^n - t_k^n - f_{i_k^n, i_{k+1}^n}\right).$$

Let $\Pi = \{\pi_1, \ldots, \pi_N\}$ denote the set of trajectories of the N birds.

Since trajectories are independent, their joint probability distribution is equal to

$$\begin{aligned} P_{R,\lambda}\left(\Pi \mid \{i_0^n\}_{n=1}^N\right) &= \prod_{n=1}^{N} P_{R,\lambda}\left(\pi_n \mid i_0^n\right) \\ &= \prod_{n=1}^{N} \prod_{k=0}^{F_n-1} R\left(i_k^n, i_{k+1}^n\right) P_{\lambda_{i_k^n}}\left(\tau = t_{k+1}^n - t_k^n - f_{i_k^n, i_{k+1}^n}\right) \end{aligned} \quad (1)$$

Given the set of trajectories Π, we can determine $N_i^t(\Pi)$, the number of birds located in site i at time $t \leq T$. $N_i^t(\Pi)$ can be computed from Π by summing the individuals present in the site at the desired time. Formally:

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

where, by definition we let $t_{F_n+1}^n = T + 1$ and the notation $|x|$ denotes the cardinality, i.e. the number of elements in the set x.

Note that bird n is visible in site i_k^n only until it starts its flight towards site i_{k+1}^n, which occurs at time $t_{k+1}^n - f_{i_k^n, i_{k+1}^n}$.

S2. MCEM algorithm

Expectation-Maximization algorithm

We consider the maximum likelihood estimator, Λ^*, of the model's parameters, i.e. the values of parameters which maximize the likelihood of observations

O. The Expectation-Maximization algorithm is an iterative algorithm that seeks to find Λ^* by iterating updating current estimates of Λ^* as follows:

$$\begin{aligned} \Lambda_{new} &= arg\max_{\Lambda} E_{\Pi\sim P_{\Lambda_{old}}(\cdot|O)}\Big[\log\left(P_\Lambda(\Pi,O)\right) \mid O\Big] \\ &= \arg\max_{\Lambda}\sum_{\Pi}\log\left(P_\Lambda(\Pi,O)\right)P_{\Lambda_{old}}(\Pi|O). \end{aligned} \tag{2}$$

Let us denote

$$Q(\Lambda|\Lambda_{old}) = \sum_{\Pi}\log\left(P_\Lambda(\Pi,O)\right)P_{\Lambda_{old}}(\Pi|O).$$

The computation required to evaluate the expectation in Q is intractable since the domain of sets of trajectories is too large. However it can be approximated using simulations.

Monte-Carlo Expectation Maximization algorithm

Provided that we are able to simulate $P_{\Lambda_{old}}(\Pi|O)$, the sum in (2) can be approximated, leading to a Monte Carlo version of EM. One iteration is as follows:

- **MC step.** Generate M samples of the N bird trajectories, $\left\{\Pi^{(1)},\ldots,\Pi^{(M)}\right\}$ from $P_{\Lambda_{old}}(\Pi|O)$.
- **M step.** Approximate $Q(\Lambda|\Lambda_{old})$ with

$$\hat{Q}(\Lambda|\Lambda_{old}) = \frac{1}{M}\sum_{m=1}^{M}\log\left(P_\Lambda(\Pi^{(m)},O)\right). \tag{3}$$

  and compute Λ_{new} as:

$$\Lambda_{new} = \arg\max_{\Lambda}\hat{Q}(\Lambda|\Lambda_{old}). \tag{4}$$

In practice, steps (3) and (4) are grouped together, in order to build a single optimization program defining the update $\Lambda_{old}\to\Lambda_{new}$.

M step: Update formula for the transition probabilities

Let us first derive the update formula for the M step. For any (Π,O), the expression of $P_\Lambda(\Pi,O)$ decomposes into:

$$P_\Lambda(\Pi,O) = P_\Lambda(\Pi)P_\Lambda(O|\Pi) = F_R(\Pi)F_\lambda(\Pi)P_\Lambda(O|\Pi),$$

where the two first terms can be easily obtained from Eq (1):

$$F_R(\Pi) \stackrel{\text{def}}{=\!=} \prod_{n=1}^{N} \prod_{k=0}^{F_n-1} R(i_k^n, i_{k+1}^n) \text{ and}$$
$$F_\lambda(\Pi) \stackrel{\text{def}}{=\!=} \prod_{n=1}^{N} \prod_{k=0}^{F_n-1} P_{\lambda_{i_k^n}}\left(\tau = t_{k+1}^n - t_k^n - f_{i_k^n, i_{k+1}^n}\right).$$

Given these expressions, the optimization problem (3-4) nicely decomposes, for a given sample $\{\Pi^{(m)}\}_{m=1..M}$. The transition probabilities $R(i,j)^{new}$ are simply the empirical frequencies of observed transitions[1] ($i \to j$), while parameter λ_i^{new} is simply the empirical mean sojourn time in node i.

$$R(i,j)^{new} = \frac{\sum_{m=1}^{M} |\{\pi_n \in \Pi^{(m)} \text{s.t. } \exists k, i_k^n = i \text{ and } i_{k+1}^n = j\}|}{\sum_{m=1}^{M} |\{\pi_n \in \Pi^{(m)} \text{ s.t. } \exists k, i_k^n = i\}|}. \qquad (5)$$

M step: Update formula for the sojourn times

Since the Poisson distribution is shifted by 1, λ_i^{new} is the average of observed sojourn times minus one, that is:

$$\lambda_i^{new} = Mean\left\{\left(t_{k+1}^n - t_k^n - f_{i_k^n, i_{k+1}^n}\right), \exists m \in \{1..M\}, \pi_n \in \Pi^{(m)} \text{ s.t. } \exists k \; i_k^n = i\right\} - 1,$$

i.e.

$$\lambda_i^{new} = \frac{\sum_{m=1}^{M} \sum_{\pi_n \in \Pi^{(m)}} \sum_{k<F_n} (t_{k+1}^n - t_k^n - f_{i_k^n i_{k+1}^n}) \mathbb{1}_{\{i_k^n = i\}}}{\sum_{m=1}^{M} \sum_{\pi_n \in \Pi^{(m)}} \sum_{k<F_n} \mathbb{1}_{\{i_k^n = i\}}} - 1. \qquad (6)$$

Note that the parameter of the sojourn time distribution in the arrival node is not estimated.

M step: Update formula for the observation model

We present update formulae for three possible observation models, i.e., Poisson, Binomial and negative Binomial distributions. In the main study we use the negative Binomial observation model.

[1] Transitions to the 'death' state should not be forgotten, so that $\sum_{j \in I} R(i,j)^{new}$ may be strictly less than 1.

• Poisson distribution. If underestimation and over estimation are possible, the most parsimonious model for $P(O \mid N_i^{t(m)})$ is a Poisson distribution of parameter $N_i^{t(m)}$. In this case, there is no parameter to estimate.

• Binomial distribution. In the case where only underestimation of the real count N_i^t is possible, it is natural to use a Binomial distribution for the observation model. If p_i is the parameter of the Binomial distribution at site i, i.e. the reporting probability, the update formula is

$$p_i^{new} = \frac{M \sum_{t \ s.t. \ (i,t)\in\Omega} O_i^t}{\sum_{m=1}^{M} \sum_{t \ s.t. \ (i,t)\in\Omega} N_i^{t(m)}},$$

with $N_i^{t(m)}$ being the real count at site i and time t associated to the set $\Pi^{(m)}$.

• Negative binomial distribution. Let us consider now the case of a Negative Binomial distribution for the observation model, with parameters r and p (with the convention that r is the fixed number of successes and p is the probability of success). This model is useful to take into account overdispersion in the data (i.e. where the variance of the observed data is substantially larger than expected). The expectation of the negative binomial is $r(1-p)/p$. If we denote by δ the probability of report of a bird, then in our case we want the expectation to be equal to $\delta N_i^{t(m)}$. This implies that for the observation model at time t and for site i we have $r_i^{t(m)} = \delta N_i^{t(m)} p/(1-p)$. Since p is estimated in a pre-processing step, before running MCEM (see Supporting Information S5), in the M step we need only to update the estimation of δ. The expression of $P_\Lambda(O_i^t \mid N_i^{t(m)})$ is $C^{O_i^t}_{O_i^t + r_i^{t(m)} - 1} (1-p)^{O_i^t} p^{r_i^{t(m)}}$. To update δ we maximise the following function:

$$\begin{aligned}
\frac{1}{M} \sum_{m=1}^{M} \log P_\Lambda(O \mid \pi^{(m)}) &= \frac{1}{M} \sum_{m=1}^{M} \sum_{(i,t)\in\Omega} \log P_\Lambda(O_i^t \mid N_i^{t(m)}) \\
&= \sum_{m=1}^{M} \sum_{(i,t)} [\log C^{O_i^t}_{O_i^t + r_i^{t(m)} - 1} + O_i^t \log(1-p) + r_i^{t(m)} \log(p)
\end{aligned}$$

We optimise this expression using the Brent option of the 'optim' function in R.

MC step: Metropolis-Hasting algorithm

The generation of M samples of the N bird trajectories, $\left\{\Pi^{(1)}, \ldots, \Pi^{(M)}\right\}$ from $P_{\Lambda_{old}}(\Pi|O)$ is done using a Metropolis Hastings procedure.

A first sample $\Pi^{(1)}$ is simulated using $P_{\Lambda_{old}}(\Pi)$.

Then, for m in $\{1, \ldots, M-1\}$, we apply the following procedure:
A candidate set of N trajectories, Π, is generated using a proposal distribution $g(.|\Pi^{(m)})$. The proposal distribution consists of, for every $i \in [1; N]$, either simulating according to distribution $P_{\Lambda_{old}}(\pi)$ (with probability q, which is a parameter of the method) or keeping the current one.

Then, according the Metropolis Hastings method, an acceptance ratio $\alpha(\Pi^{(m)}, \Pi)$ (which is the probability with which a generated sample Π should be accepted, given that $\Pi^{(m)}$ was the previously generated sample) is computed using the joint distribution of observations:

$$\alpha(\Pi^{(m)}, \Pi) = min\{1, \frac{P_\Lambda(O|\Pi)}{P_\Lambda(O|\Pi^{(m)})}\}$$

Indeed, even if the proposal distribution is not symmetric, we have the following property:

$$A = \frac{P_\Lambda(O|\Pi)P_\Lambda(\Pi)g(\Pi^{(m)}|\Pi)}{P_\Lambda(O|\Pi^{(m)})P_\Lambda(\Pi^{(m)})g(\Pi|\Pi^{(m)})} = \frac{P_\Lambda(O|\Pi)}{P_\Lambda(O|\Pi^{(m)})}$$

To prove it, let us partition the set of indices of trajectories $\{1, \ldots, N\}$ into two sets S and D as follows

$$\begin{aligned} S &= \{i \in \{1, \cdots, N\} | \pi_i = \pi_i^{(m)}\} \\ D &= \{j \in \{1, \cdots, N\} | \pi_j \neq \pi_j^{(m)}\} = \{1, \cdots, N\} \setminus S \end{aligned}$$

S is the subset of indices of trajectories that are identical in Π and $\Pi^{(m)}$ and D the subset of indices of trajectories which differ in Π and $\Pi^{(m)}$. We can decompose the two distributions P_Λ and g using S and D:

$$\begin{aligned} P_\Lambda(\Pi) &= \prod_{j \in D} P_\Lambda(\pi_j) \prod_{i \in S} P_\Lambda(\pi_i^{(m)}) \\ P_\Lambda(\Pi^{(m)}) &= \prod_{j \in D} P_\Lambda(\pi_j^{(m)}) \prod_{i \in S} P_\Lambda(\pi_i^{(m)}) \end{aligned}$$

and

$$
\begin{aligned}
g(\Pi^{(m)}|\Pi) &= \prod_{i\in S}(qP_\Lambda(\pi_i^{(m)}) + (1-q)) \prod_{j\in D} qP_\Lambda(\pi_j^{(m)}) \\
g(\Pi|\Pi^{(m)}) &= \prod_{i\in S}(qP_\Lambda(\pi_i^{(m)}) + (1-q)) \prod_{j\in D} qP_\Lambda(\pi_j)))
\end{aligned}
$$

$$
\begin{aligned}
A &= \frac{P_\Lambda(O|\Pi)P_\Lambda(\Pi)g(\Pi^{(m)}|\Pi)}{P_\Lambda(O|\Pi^{(m)})P_\Lambda(\Pi^{(m)})g(\Pi|\Pi^{(m)})} \\
&= \frac{P_\Lambda(O|\Pi)\prod_{j\in D}P_\Lambda(\pi_j)\prod_{j\in D}qP_\Lambda(\pi_j^{(m)})}{P_\Lambda(O|\Pi^{(m)})\prod_{j\in D}P_\Lambda(\pi_j^{(m)})\prod_{j\in D}qP_\Lambda(\pi_j)},
\end{aligned}
$$

since the products over indices in S are equal. Finally

$$
\begin{aligned}
A &= \frac{q^{|D|}P_\Lambda(O|\Pi)\prod_{j\in D}P_\Lambda(\pi_j)\prod_{j\in D}P_\Lambda(\pi_j^{(m)})}{q^{|D|}P_\Lambda(O|\Pi^{(m)})\prod_{j\in D}P_\Lambda(\pi_j^{(m)})\prod_{j\in D}P_\Lambda(\pi_j)} \\
&= \frac{P_\Lambda(O|\Pi)}{P_\Lambda(O|\Pi^{(m)})}
\end{aligned}
$$

Rejection or acceptance of the candidate Π relies on the value $\alpha\left(\Pi^{(m)},\Pi\right)$:

$$
\begin{aligned}
\text{Select } \Pi^{(m+1)} &\leftarrow \Pi \text{ with probability } \alpha\left(\Pi^{(m)},\Pi\right) \\
&\leftarrow \Pi^{(m)} \text{ with probability } 1-\alpha\left(\Pi^{(m)},\Pi\right)
\end{aligned}
$$

Note that not all individual bird trajectories are re-simulated with the symmetric proposal distribution g, since this would lead to very few accepted proposals. The arbitrary ratio of re-simulation, q, is tuned in order to achieve a reasonable rate of accepted proposals ($\sim 30\%$). In practice, a value of $q = 0.1$ was chosen. For benchmarking, convergence of the MCEM was considered reached when difference in log likelihood is below 2 during 10 consecutive iterations. Since MCEM is a local optimization method, 5 optimizations were performed with different initial values of Λ and we kept the estimated parameters provided by the optimization that led to the best log likelihood.

S3. ABC algorithm

We use a modified version of the Adaptive Population Monte Carlo (APMC) algorithm to solve our problem. APMC uses a particle filtering approach to

select the next set of candidate parameter values. Particles are assigned weights that are used to iteratively sample new new particles in the next step of the algorithm. The newly generated particles in a sequential procedure are no longer drawn from the prior distribution but from a specific probability density that depends on the particles selected at the previous step and on the chosen kernel. This introduces a bias in the procedure which can be removed by correcting the weight attached to the newly generated particle. The mathematical formulation of the corrected weight for APMC can be found in Lenormand et al. (2013).

We need to modify this approach because APMC expects independent draws from uniform distributions, but the sum of transition probabilities from a node must sum to 1 (i.e. a categorial distribution). We modify the APMC algorithm to ensure that this property is respected. In the remainder of this section, we outline the APMC algorithm and detail the modifications that we made for our problem.

First, APMC relies on the simulation of observed count $O(\Lambda) = \{O(\Lambda)_i^t\}_{(i,t)\in\Omega}$ given the parameters Λ (transition probabilities, mean sojourn times and reporting probability). In practice, the simulation model consists of:

1 The generation of N bird trajectories using the parameters $\Lambda : \Pi \sim P_\Lambda(.)$

2 The generation for all $(i,t) \in \Omega$, according to the negative binomial distribution, of $O(\Lambda)_i^t$ from the bird counts N_i^t derived from the trajectories $\Pi : O(\Lambda)_i^t \sim \mathcal{NB}_{(r_i^t,p)}(.)$

The initialization of APMC consists of generating an initial sample of model parameters, called particles, $S^{(0)} = (\Lambda_k)_{k=1,\cdots,K\alpha}$, where K is the number of proposals and α is the proportion of particles kept amongst these proposals for the next population/iteration of the algorithm. This initialization is also used to define the criterion of distance between the simulated observations and the true observations. To do so, APMC first generates a population of K particles from independent uniform distributions using a Latin Hypercube sample into the model parameters space, which constitutes the proposal population: $S_p^{(0)} = (\Lambda_k)_{k=1,\cdots,K}$. Using the model, observations $O(\Lambda_k)$ are simulated and

the variance of each statistic $v_i^{(t)}$ is computed:

$$v_i^{(t)} = \frac{1}{K}\sum_{k=1}^{k=K}(O(\Lambda_k)_i^t - \frac{1}{K}\sum_{k=1}^{k=K}O(\Lambda_k)_i^t)^2$$

The distance criterion between the simulated observations $O(\Lambda_k)$ and the true observation is defined as the following weighted sum:

$$\rho(O, O(\Lambda_k)) = \sum_{(i,t)\in\Omega}\frac{(O(\Lambda_k)_i^t - O_i^t)^2}{v_i^{(t)}} \tag{7}$$

The lower the variance $v_i^{(t)}$ of simulated observations, the more the difference to the true observations O_i^t will be accounted in this weighted sum. Only the $K\alpha$ best particles are kept in the initial population, ie.: $S^{(0)} = \{\Lambda_k \in S_p^{(0)} | \rho(O, O(\Lambda_k)) \leq \mathcal{Q}^{(0)}(\alpha)\}$ where $\mathcal{Q}^{(0)}(\alpha)$ is the quantile of order α of the ρ values computed for particles in $S_p^{(0)}$. The threshold $\epsilon^{(0)} = \mathcal{Q}^{(0)}(\alpha)$ satisfies by construction: $\forall \Lambda_k \in S^{(0)}, \rho(O, O(\Lambda_k)) < \epsilon^{(0)}$

A following iteration of the algorithm consists of deriving a sample of model parameters $S^{(d)} = (\Lambda_k)_{k=1,\cdots,K\alpha}$ from the previous sample $S^{(d-1)}$ (with $d > 0$) using a particle filter methodology. First, a set of particles $S_p^{(d)}$ of size $K - K\alpha$ is generated as new proposals. The generation of a particle in $S_p^{(d)}$ consists of randomly drawing a particle $\Lambda \in S^{(d-1)}$ (with a weight that is proportional to the inverse of the importance of Λ in $S^{(d-1)}$) and then applying a Markov transition kernel to Λ. Finally, $S^{(d)}$ is defined as the $K\alpha$ parameters values from the K particles $S^{(d-1)} \cup S_p^{(d)}$ for which the simulation of the model leads to the lower distance ρ, i.e. : $S^{(d)} = \{\Lambda_k \in S^{(d-1)} \cup S_p^{(d)} | \rho(O, O(\Lambda_k)) \leq \mathcal{Q}^{(d)}(\alpha)\}$ where $\mathcal{Q}^{(d)}(\alpha)$ is the quantile of order α of the ρ values computed in $S^{(d-1)} \cup S_p^{(d)}$. The current value of threshold is calculated as : $\epsilon^{(d)} = \mathcal{Q}^{(d)}(\alpha)$. The proportion of newly introduced particles in $S^{(d)}$ is noted $p_{acc}^{(d)}$.

Authors of the algorithm have shown that the proportion $p_{acc}^{(d)}$ converges in probability towards 0 when $d \to \infty$. Thus, in order to specify convergence, the user provides a threshold $p_{acc_{min}}$ on $p_{acc}^{(d)}$. Let D be the iteration at which the algorithm reached convergence, then by construction $\epsilon^{(0)} < \epsilon^{(r)} < \epsilon^{(D)}$.

However slight modifications of the algorithm had to be carried out in order

to adapt this algorithm to our problem. First, APMC relies on independent uniform distributions but since we aim at estimating transition probabilities (which corresponds to categorial distributions), the assumption of independent uniform distributions does not hold. We solved this issue by defining a bijection between multivariate independent uniform distributions and a categorial distribution. Since the mortality μ_i of a given site i is known and the transition probability matrix R to estimate is upper triangular, we have: $\mu_i + \sum_{j=i+1}^{I} R(i,j) = 1$. From independent draws in $[0;1]$ uniform distribution $\{Q_{i+1}, ..., Q_{I-1}\} \in U^{I-i-1}[0,1]$, we compute the transition probabilities recursively by:

$$R(i,j) = \begin{cases} Q_j * (1 - \mu_i) & \text{if } j = i+1 \\ Q_j * (1 - \mu_i - \sum_{i+1<l<j} R(i,l)) & \text{if } i+1 < j < I \\ 1 - \mu_i - \sum_{i+1<l<I} R(i,l) & \text{if } i = I \end{cases}$$

We used APMC to estimate the Q values and we derived the R matrix from these equations.

Second, modifications were required for the most complicated cases of our benchmark experiments. Indeed, when the number of parameters to estimate was too large (ie. upper than 19), the random walk to generate new candidates, which is based on a Markov transition kernel with infinite support, led systematically to candidates outside the uniform prior which is given with a finite support. Originally, the implementation of the algorithm in EasyABC raised an error after 10000 drawings from the Markov transition kernel that lead to candidates outside the prior. Rather than stopping the algorithm, we furnished, after 10000 unsuccessful tests, a new candidate from the uniform prior. Doing this, the algorithm ended without error but it took sometimes much more time. This is why we stopped the algorithm after 10 days of computation time if convergence was not reached. In our benchmark results, we identified both cases where the use of the uniform prior in the random walk was needed and the p_{acc} values of cases where the early stop was required.

Finally, in our experiments, parameter K was set to 5000, $p_{acc_{min}}$ was set

to 0.01 and we kept the default value $\alpha = 0.5$.

S4. Benchmark evaluation criterion

Results of estimation were assessed by the computation of log likelihood and the mean absolute error of estimated parameters. The first one aims at quantifying the quality of estimation, especially for MCEM, and the second one aims at quantifying the identification of parameters.

Logarithmic form of likelihood

To compute the log-likelihood, we rely on a Monte Carlo approach and use the $MAX*$ operator to avoid underflow issues. First, the likelihood of parameters can be integrated over all possible trajectories :

$$\begin{aligned} P(O|\Lambda) &= \int_{\Pi} P(O, \Pi|\Lambda) \mathrm{d}\Pi \\ &= \int_{\Pi} P(\Pi|\Lambda) * P(O|\Pi, \Lambda) \mathrm{d}\Pi \end{aligned} \tag{8}$$

We then used a Monte Carlo approach by performing M simulations $\Pi^i \sim P(.|\Lambda)$ for $i \in [1, M]$ in order to approximate the integral. Since the probabilities $P(O|\Pi^i, \Lambda)$ are very small, we compute the log probabilities.

$$\begin{aligned} P(O|\Lambda) &\approx \frac{1}{M} \sum_{i=1}^{M} P(O|\Pi^i, \Lambda) \\ log(P(O|\Lambda)) &\approx log(\frac{1}{M} \sum_{i=1}^{M} P(O|\Pi^i, \Lambda)) \\ &= log(\frac{1}{M}) + log(\sum_{i=1}^{M} P(O|\Pi^i, \Lambda)) \end{aligned} \tag{9}$$

As suggested in Yu (2015), to avoid underflow issues, we rely on the MAX^* operator and its generalization to an arbitrary number of arguments:

$$\begin{aligned} MAX^*(x, y) &\stackrel{\text{def}}{=} log(e^x + e^y) \\ &= max(x, y) + log(1 + e^{-|x-y|}) \end{aligned} \tag{10}$$

$$
\begin{aligned}
MAX^*(x,y,z) &\stackrel{\text{def}}{=} log(e^x + e^y + e^z) \\
&= log(e^{MAX^*(x,y)} + e^z)
\end{aligned}
\tag{11}
$$

This way we can compute the log likelihood without handling the very low probabilities $P(O|\Pi^i)$:

$$
\begin{aligned}
log(P(O|\Lambda)) &= log(\frac{1}{M}) + log(\sum_{i=1}^{M} P(O|\Pi^i, \Lambda)) \\
&= log(\frac{1}{M}) + log(\sum_{i=1}^{M} e^{log(P(O|\Pi^i,\Lambda))}) \\
&= log(\frac{1}{M}) + MAX^*(log(P(O|\Pi^1,\Lambda)), ..., log(P(O|\Pi^M,\Lambda)))
\end{aligned}
\tag{12}
$$

Mean absolute error

Given N real parameters to estimate, taking values in $[0,1]$ (or rescaled in $[0,1]$) we define the mean absolute error $meanAE$ defined as

$$
meanAE = \frac{1}{N} \sum_{i=1}^{N} |\tilde{v}_i - v_i| \tag{13}
$$

Where $\tilde{v}_i$ is the estimated value and v_i is the true value. This criterion is only available when we know the true values. For the curlew application, it is not available.

Supporting Information S5: Eastern Curlew case study—data extraction from eBird and network definition

157

This supporting information contains the methods used to create the Eastern Curlew case study network for the main text. The document describes how data was obtained from eBird, how nodes and edges for the network were defined, and basic exploration of the data contained in each network node.

Data source and preliminary filtering

Eastern curlew count data was obtained from the June 2020 version of the eBird basic dataset (World; available from https://ebird.org/data/download). eBird data consists of checklists. A "checklist" is a data type that records all the birds sighted by an observer during a sampling event. This section describes how the checklists were used to estimate weekly node abundance in our case study.

A custom dataset containing all Eastern Curlew records was downloaded and then filtered to obtain relevant Eastern Curlew records using the auk package in R. Filters applied were:

1) *Species*: Eastern Curlew records were extracted from all bird records (via the custom download—it was not necessary to include a species filter in the call to auk).
2) *Country:* The following countries and regions were included in the filtered dataset, reflecting the expected distribution of the Eastern Curlew within the EAAF: USA (Alaska only), Russia, Mongolia, China, Hong Kong, Taiwan, Macao, South Korea, North Korea, Japan, Phillippines, Vietnam, Cambodia, Laos, Thailand, Myanmar, Bangladesh, India, Malaysia, Singapore, Brunei, Indonesia, Timor Leste, Papua New Guinea, Australia, Cocos (Keeling) Islands, New Zealand, Federated states of Micronesia, Northern Mariana Islands, Pulau, Solomon Islands, Vanuatu, Samoa.
3) *Longitude*: a longitudinal filter was applied by removing any data from the contiguous USA (i.e. all except Alaska; defined by the US-Canada border at the 141W meridian), and by removing all data from western Russia (i.e. anything west of the 80E meridian). This was done to remove data from Russia and the USA that are outside the curlew's known distribution and reduce the size of the dataset.
4) *Survey length and distance*: To even out the count data, we removed checklists with a duration longer than 5 hours and a distance travelled of greater than 5km (Strimas-Mackey et al. 2020).

Data was combined with the sampling event data to obtain presence-absence data as described in the auk package vignette (https://cran.r-project.org/web/packages/auk/vignettes/auk.html).

The following eBird fields were retained for further analysis:

- "checklist_id", "country", "country_code", "iba_code","state", "state_code","locality_id","locality_type", "latitude", "longitude", "observation_date", "sampling_event_identifier", "duration_minutes", "effort_distance_km", "effort_area_ha", "number_observers", "scientific_name", "observation_count", "species_observed"

The R script (getData_curlew.R) used to filter data is available from the manuscript GitHub. Note that to run the script, users will first need to download the eBird basic dataset and sampling event data from the eBird website. The output of the script is called "eastern_curlew_all_eBird2020.Rds"

Data was then formatted further in the R script “format_weekly_data_BirdlifeDist.R”. This script does the following steps:

1) Remove all records that don’t explicitly record a count for the species of interest (i.e. eBird allows users to record presence with an “X”, but not record a count. These data were removed)[1].
2) Remove all old data (data older than 1980 were removed). For our analysis we selected only data from one year at a time (2019 and 2018) to ensure that the migration signals were not blurred by averaging over multiple years. Results shown in this document are for the 2019 data.
3) Remove all data with a missing location field.
4) Assign each datapoint to a bounding box polygon roughly corresponding to the expert-elicited Eastern Curlew network in Iwamura et al. (2013). Coordinates of these polygons were defined manually; see the section “Node definitions” for further details on this process.
5) To remove records from outside the expected habitat of the curlew, we intersected curlew records with distribution maps produced by BirdLife International (BirdLife International and Handbook of the Birds of the World 2019). Records that intersected the distribution maps were kept and other records were discarded. Note that zero counts that lay inside the distribution maps were maintained in the dataset.
6) eBird sample data are incomplete and spatially biased. To obtain crude estimates of total abundance within each node, we completed the following steps:
    a. Clipped the area of each node to the intersection of the node boundaries (defined in step 4) and the Birdlife species distribution maps (defined in step 5). This ensures that records within each node are within the known distribution of the species.
    b. Overlaid these intersected areas with a hexagonal grid (cell size 100km^2, which roughly coincides with the 10x10km grid cell sizes used to extrapolate populations in Hansen et al. (2016)).
    c. Within each hexagon, computed the average count in each week by taking the mean count observed in each hexagon in each week. This step was intended to remove the impacts of double counting and spatially variable survey effort. The mean count per hexagon was assumed to be an estimate of density for extrapolation.
    d. Obtained an extrapolated count estimate for the node by assigning the mean count per hexagon (from hexagons with observed records) to all hexagons. The sum of

[1] Removing checklists reporting an X for curlew induces a bias unless all checklists reporting a 0 that ‘would have’ reported an X had the observer detected a curlew are also removed. This could be adjusted for by removing all checklists with an X for any species, however the eBird package *auk* does not provide functionality to search for and remove checklists that have an ‘X’ recorded anywhere in the dataset. In our problem, the number of checklists that recorded ‘X’ for curlews was relatively small (1039 out of 900317 checklists; 0.1%), so only a small proportion of the data was removed by this step. Eastern curlews are also an easily identifiable species, so we expect that relatively few observers would fail to identify them if they were present. For these reasons we postulate that the bias introduced by our removal of ‘X’ values should be low.

counts over all hexagons was assumed to be an estimate of the weekly count in the node.

For further information on the node abundance estimation, see the section "From eBird checklists to an observed count per node" at the end of this document.

After filtering[169], a total of 222006 checklists remained in our dataset. A total of 71140 hexagons were generated to represent the distribution of curlew within the predicted distribution.

Node definitions

Bounding box polygons to assign eBird count lists to network nodes were created manually by examining the literature and the eBird count data. The primary information source for nodes was data from Iwamura et al. (2013), which assigns 65 Eastern Curlew Internationally Important Sites (IISs; areas which support at least 1% of the global population; as defined in Bamford et al. (2008)) to 9 geographic nodes. Iwamura et al's "Yellow Sea" node (Block 2 in Figure 1) was defined for the southward migration only. Since our analysis considers only the northward migration, this node was dropped from our analysis, leaving us with 8 nodes for our problem.

The Iwamura et al. (2013) study provides a rough guide to node boundaries but it does not provide exact geographic boundaries, which are necessary to assign precisely located eBird counts to nodes. Node boundaries were therefore defined using: (i) the available published literature on species movements, particularly Driscoll and Ueta (2002), Minton et al. (2011a), Minton et al. (2011b); (ii) by plotting the locations of Eastern Curlew eBird checklists and attempting to draw polygons to capture sensible proportions of recorded sightings (Figure 2). Note that not all recorded locations are captured in the polygons. We varied the boundaries manually to capture a reasonable proportion of sightings (Table 3) while maintaining the intent of the expert-derived boundaries defined in Iwamura et al. (2013). The logic for assigning our selected boundaries is contained in

Table 1.

We experimented with including two additional nodes representing Taiwan/southern Chinese mainland and the Philippines, since there are a number of IISs on Taiwan, the southern Chinese mainland and the Philippines, and studies (Driscoll and Ueta 2002, Minton et al. 2011a, Minton et al. 2011b) suggest that these areas are important short-term resting areas for birds on the northern migration. However these nodes were removed after examining the eBird count data, which indicated low total numbers and densities of birds in these locations (Table 2; Table 3), as well as commentary in Driscoll and Ueta (2002) that suggested that these sites are likely to be short stopovers in contrast to the much longer staging undertaken at the other nodes.

Once the node boundaries were defined, eBird checklists were assigned to each node if the location of the checklist fell within the boundaries of the node.

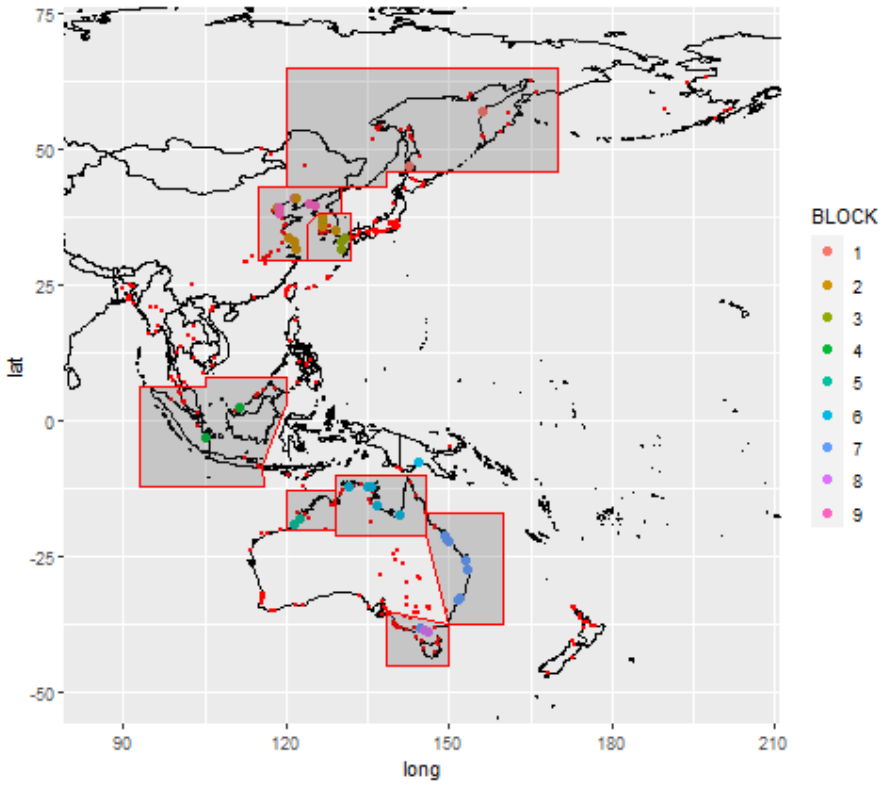


Figure 1: Locations of Important Bird Areas for Eastern Curlew. Red dots show the locations of Internationally Important Sites (IIS) for shorebird species in the EAAF (from Bamford et al 2008). Larger coloured dots show the IISs for Eastern Curlew (Bamford et al 2008). The colours (BLOCK) show how IISs were allocated to nodes in the expert-derived network model of Iwamura et al (2013). Our boundaries used these node assignments as a guide: red polygons are the node boundaries selected for this analysis. Note that one node (Yellow Sea; Block 2) from the Iwamura analysis is only relevant to the southward migration and is not included in this analysis.

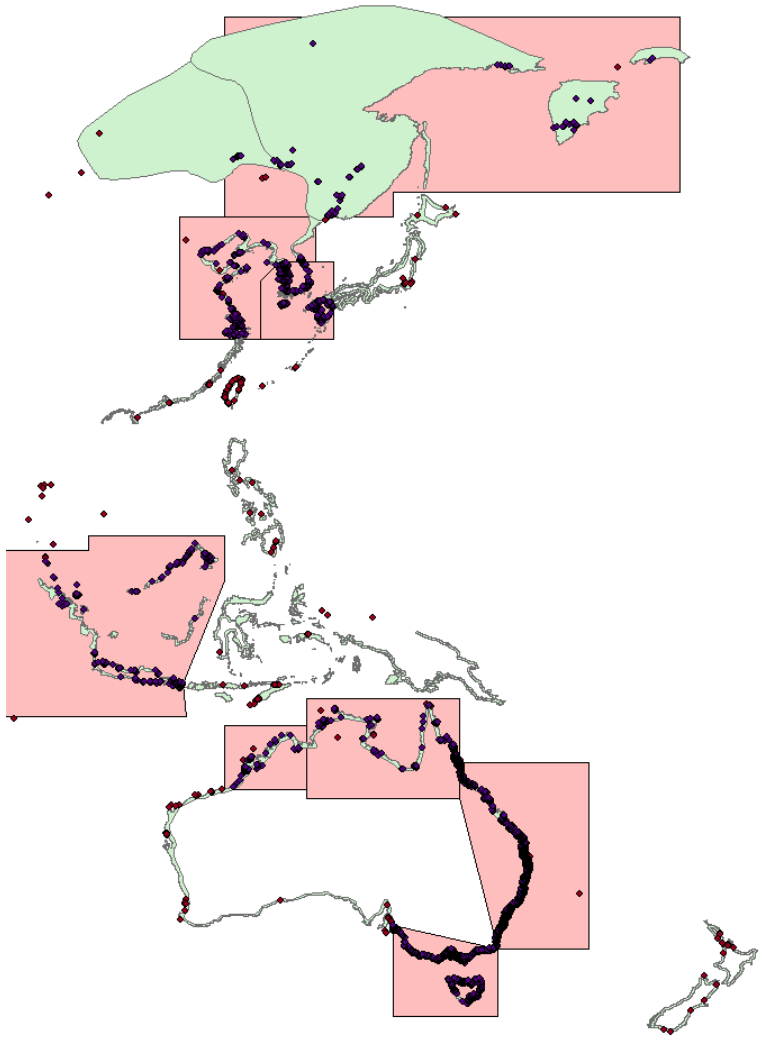


Figure 2: Locations of species observations relative to node boundaries (pink polygons) and BirdLife distribution maps for the Eastern Curlew (green polygons) in 2019. Purple dots show the locations of Eastern Curlew checklists that lie inside both node boundaries and the predicted distribution (these include both presence and absence checklists). Red dots show the locations of checklists where Eastern Curlews were recorded in eBird (presence only) but fall outside our node or distribution boundaries.

Table 1: Logic used to define nodes for the Eastern Curlew case study network

Node	**Latitude and longitude of bounding box**	**Logic for bounding box**
Southern Australia	Intersection of Victorian border at the coast; connected by a straight line to the city of Adelaide to capture curlew sites	A large number of curlews use the southern Australian coast during the non-breeding season. Classified as a distinct node in Iwamura et al (2013).

	throughout Victoria and eastern South Australia. Tasmania included in box.	Many tracking studies based in Victoria.
South-eastern Australia [162]	Intersection of Victorian border at the coast to the city of Cairns in north Queensland. Includes all sites in New South Wales and southern Queensland.	Somewhat arbitrary distinction at Northern boundary (Cairns). Satellite tracking (Driscoll and Ueta 2002) suggests that some birds from Moreton Bay stop migrating in North-Eastern Qld and New Guinea, suggesting that these two places should be grouped in a similar node. The exact location of this cutoff (i.e. at Cairns) may not be critical other than to separate northern sites (which may have resident birds or act as stopover sites) from the population hubs in southern Queensland/NSW (e.g. Moreton Bay, Great Sandy Strait).
Northern Australia	Northern Territory border to Cairns (QLD; see justification for SE Australia above). Excludes Papua New Guinea; although PNG is known as a stopover/resident site for aborted migrants (Driscoll and Ueta 2002), observed counts in PNG were very low in the eBird data, so the inclusion of the extra node area was not warranted by the data.	Some evidence that Northern Australia is a staging site for birds from Victoria/QLD (Driscoll and Ueta 2002). Important bird areas and high counts in the Northern Territory. Considered a distinct node by experts in Iwamura et al (2013).
North-Western Australia	Western Australian border to 120E (approximate location of Western Australia's Kimberley region).	Includes tropical Western Australia, which encompasses most of the observed Eastern Curlews counted in eBird. Very few birds were recorded in eBird in the Pilbara region of Western Australia in the 2019 observation period; although sporadic sightings were recorded, extrapolating counts over this large area was not warranted with so few records.
Malaysia and Indonesia	Malaysia and western Indonesia (to Wallace line)	Distinct node in Iwamura et al (2013). Intended to capture resident birds that spend the non-breeding season in Malaysia and Indonesia rather than passage

		migrants passing through. Cut at Wallace line arbitrarily to reduce number of lists with low counts through eastern Indonesia.
Southern Japan and South Korea[163]	South Korean border and southern Japan (Kyushu (to 126E) to Korean border latitude at 38N)	Southern Japan to 38N—tracking data suggests that a large majority of Australian birds fly non-stop to southern Japan or Korea (Minton et al 2011b); whole country not included as we aimed to capture the main southern aggregation sites. Very few records in Shikoku Honshu and Hokkaido so we restricted Japanese node area to Kyushu to minimise over-extrapolation. Northern Japanese sites are probably passage sites to the breeding area. 38N was chosen because it is the latitude of the South-North Korean border.
Yellow Sea and North Korea	Approximate mouth of Yellow Sea (29.5N) including North Korea (northernmost boundary at 43N); excluding Russian sites.	Chinese and North Korean Yellow Sea stopover sites.
Breeding node	Eastern Mongolian border (120E), including Kamchatka to 170E. North Korean border northward to 65N.	Based on approximate Birdlife distribution map for breeding area (Birdlife International 2020); boundaries set to include locations of known eBird records. This node boundary is unlikely to be highly accurate due to low list density and high number of extrapolation polygons, however we assume that the breeding node is a final destination and therefore the accuracy of this count has lower importance than other nodes (since all birds are assumed to arrive at the breeding site eventually).

Nodes investigated but not included in the network are outlined in Table 2.

Table 2: Nodes investigated but not included in the network

Node name	**Latitude and longitude of bounding box**	**Logic for bounding box**
Southern China and Taiwan	Taiwan, Okinawa islands, Chinese mainland east coast from Vietnamese border to	Satellite tracking (Driscoll and Ueta 2002, Minton et al 2011b) suggests that Taiwan in

[164]	29.5N (approximate mouth of Yellow Sea; southern side of Hangzhou Bay) Exact latitude of bounding box: 17.5,29.5,29.5,21, 21, 17.5 Exact longitude of bounding box: 108,108,130,130,115,115	particular is a stopover target for Australian birds, but that migrants also use the Chinese coastline. Birds likely to use these sites but fairly briefly compared to other nodes such as Yellow Sea (Driscoll and Ueta 2002). This was not included as a node in Iwamura et al 2013. Node was not included because eBird data suggested that weekly average densities are very low (0.005 birds/list/week). Total of 192 birds observed over 38172 lists. Relatively low densities and good survey coverage suggest that this is likely to be a short-term stopover for birds *en route* between non-breeding and key staging sites in Japan, ROK and China.
Philippines	Philippines border; designed to skirt other bounding boxes. Exact latitude of bounding box: 7.9, 21,21,5,5, 7.9 Exact longitude of bounding box: 115,115,130,130,120,120	Satellite tracking suggests Philippines are an important stopover site for Eastern Curlews, but only briefly (Driscoll & Ueta 2002). Only 59 birds observed in this region across 2317 lists in 2019; average density 0.02 birds/list/week). Relatively low average weekly densities (0-3 birds per list) except for a short peak in a single week on northward migration, i.e. 21 birds/list in early March/week 11). Very short durations of higher densities mean that this node is likely to be a temporary stopover for birds observed en route between non-breeding and key staging sites in Japan, ROK and China. Not included as a distinct node due to short duration of observing birds at high density and evidence from tracking in Driscoll and Ueta (2002).

As a preliminary investigation of the coverage of our chosen node boundaries, we analysed the proportion of eBird observations captured by each of the nodes (after clipping to the predicted distribution; Table 3). We also computed the number of observations and lists that were not captured by our nodes (i.e. data that lies outside the geographic boundaries of our nodes). 7.8% of observed birds and 68.0% of checklists (this includes checklist absences, i.e. checklists where eastern curlew were not recorded) are not captured by the proposed polygons. The high proportion of birds captured within the nodes (92.8% of observed birds) suggests that our node boundaries are reasonably effective at capturing most of the observed birds. Note that the missing checklists are largely recorded absences outside the distribution maps of the curlew, so this statistic has limited use as a measure of the inclusiveness of the node boundaries, but it is included as a crude proxy for the extent of surveying in each node.

Table 3: Analysis of raw 2019 Bird data observations captured by selected nodes for Eastern Curlew

Site Name	Total birds observed from all 2019 checklists	Number of checklists (includes absences)
BREED	70	154
SK-JPN	7948	1659
MSIA-IND	150	5254
NWAUS	1209	587
NAUS	1320	6441
SEAUS	22148	34554
SAUS	1987	21039
NK-YS	336	1452
Lists not included in polygons	2958	161296
TOTAL	**38126**	**150886**

Edge definitions

Connections between nodes were initially based on the expert networks in Iwamura et al. (2013). Information was also incorporated from Driscoll and Ueta (2002), Minton et al. (2011a), Minton et al. (2011b). Edges were added to the Iwamura network to investigate key areas of uncertainty about Eastern Curlew migration. In particular, we added edges to investigate: (i) the relative proportion of birds visiting the east and west sides of the Yellow Sea from the key non-breeding areas in Eastern and Northern Australia, and (ii) migration of birds within Australian regions. See Table 5 for additional justification for the inclusion of each edge. Figure 3 illustrates the edges for the network on a global map.

To compute travel time between locations, we assume a migration flight speed of 50km/h (ground speed) consistent with Driscoll and Ueta (2002) (estimated flight speed of 50km/h); Minton et al. (2011a) and Minton et al. (2013) (median tracked speed 50.2km/h). Travel times are computed using the distance between key known aggregation sites (Table 4) and rounded to the nearest day. To match the timestep of our statistical modelling, in the HSMM modelling travel times are rounded to the nearest week, with a minimum travel time of 1 week imposed on short duration flights (i.e. <3.5 days). Aggregation site locations in Table 4 were selected using the location of the IIS with highest count according to Bamford et al 2008; table 4.12) Where possible these are confirmed with known flight times from the literature.

Note that the total migration duration is estimated to be 6-8 weeks from Southern Australia and 5-6 weeks from SE and NW Australia (Minton et al. 2011a). This duration includes sojourn times, so it will be longer than the estimated travel times in the table below.

Table 4: Key aggregation sites used to define inter-node distances (from Bamford et al, 2008).

Node name	**Key aggregation site**	**Location**
Southern Australia	Corner inlet	-38.73, 146.22
Southeastern Australia	Great Sandy Strait	-25.67, 152.93
Northern Australia/Papua New Guinea	SE Gulf of Carpentaria	-17.47, 140.76
North Western Australia	Roebuck Bay	-18.07, 122.33
Malaysia and Indonesia	Pulau Bruit	2.57, 111.35
Southern Japan and South Korea	Kanghwa Island	37.58, 126.5
Yellow Sea and North Korea	Yalu Jiang National Nature reserve	39.82, 124.11
Breeding node	Moroshechnaya River Estuary	56.83, 156.17

Table 5: Definition and justification of edges, including estimated travel times along edges (days). Justification for flight time estimates are provided where literature provides useable information on flight times.

Source node	**Destination node**	Justification	Distance	**Flight time (days)**	Justification for flight time (observed times)
Southern Australia	South-Eastern Australia	This edge model is consistent with Iwamura et al (2013), with the addition of a route between	1581 km	**1.3**	
	Japan/South Korea		8716 km	**7.3**	6.8 days (Minton, Gosbell et al

[167]		Southern Australia and the Yellow Sea (Minton, Gosbell et al 2013). Minton et al 2011b shows band recoveries of birds visiting SE Australia before undertaking the long migration across the Pacific. Band recovery data also shows birds moving to both Japan/ROK and the Yellow Sea/North Korea (Minton et al 2011b).			2013 Table 5); 5-7 days estimated (Minton et al 2011a)
	Yellow Sea		9013 km	**7.5**	7.0 days (Minton, Gosbell et al 2013 Table 5); 5-7 days estimated (Minton et al 2011a)
Southeastern Australia	Northern Australia/PNG	Multiple sources confirm the route between SE Australia and Japan (Minton et al 2011a,b; Driscoll & Ueta 2002; Iwamura et al 2013; Minton, Gosbell et al 2013) and the Yellow Sea (Minton et al 2011a,b; Driscoll & Ueta 2002; Minton, Gosbell et al 2013). Driscoll and Ueta 2002 also show evidence that some SE Australian birds visit Northern Australia and the southern coast of PNG, either aborting migration there or using it as a staging site for the Pacific flight.	1551 km	**1.3**	1-3 days (Driscoll and Ueta 2002, Figure 1)
	Japan/South Korea		7550 km	**6.3**	
	Yellow Sea		7866 km	**6.6**	
Northern Australia/Papua New Guinea	Japan/South Korea	Iwamura et al (2013) and Minton et al 2011b show birds migrating from Northern Australia to Japan/ROK; Driscoll and Ueta 2002	6296 km	**5.2**	
	Yellow Sea		6596 km	**5.5**	

		illustrates some birds staging in PNG en route to the Yellow Sea.			
North-Western Australia[168]	Yellow Sea	Minton et al 2011a,b note that no birds from NW Australia have been observed taking the eastern (Japan/ROK) migratory route; all birds observed have followed the western Yellow Sea route. Iwamura et al 2013 reflect this observation in their model.	6435 km	**5.4**	5-7 days estimated (Minton et al 2011a)
Malaysia and Indonesia	Yellow Sea	Iwamura et al 2013 show Malaysian/Indonesian birds exclusively using the western Yellow sea migratory route.	4335 km	**3.6**	
Southern Japan and South Korea	Breeding node	Iwamura et al (2013)	3053 km	**2.5**	
Yellow Sea and North Korea	Breeding node	Iwamura et al (2013)	2984 km	**2.5**	
Breeding node	Final/destination node	Not applicable	0 km	**0**	NA

Distances computed from: https://www.nhc.noaa.gov/gccalc.shtml

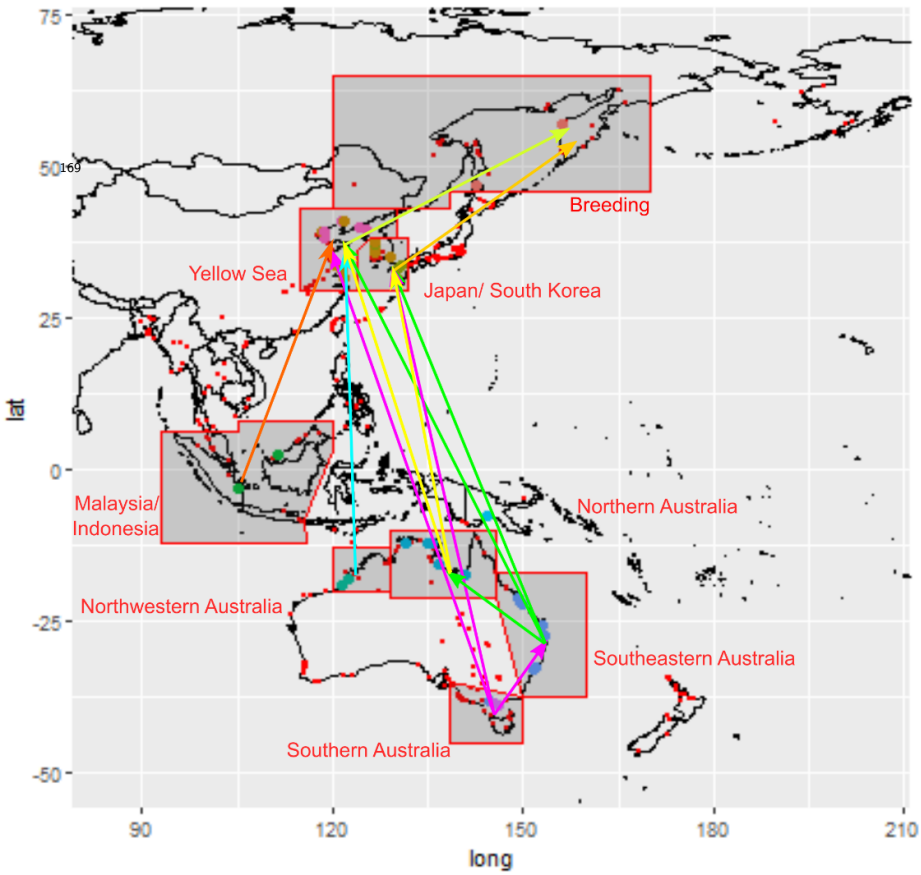


Figure 3: Location of nodes with labels. Arrows depict the edges defined in Table 5.

Visualising the count data

For each node, we plotted the average total number of birds observed per week (Figure 4). We also plotted the same data on a common scale with the y-axis scaled to a recent estimate of the total curlew population (~32000 individuals; Figure 5). Aggregation formulae to create these plots are in the following paragraph. Migration signals are clear from total count data in all nodes, although there is clearly some noise in the data, particularly with surprisingly low bird numbers observed in SAUS (Southern Australia), unusually high counts in NWAUS (Northwestern Australia), and by the "three peaks" in the data for SK-JPN (South Korea/Japan). All sites have at least a few lists each week, with the exception of the breeding node, which is poorly surveyed outside the migration period (probably due to remoteness, dispersed populations, and the harsh winter).

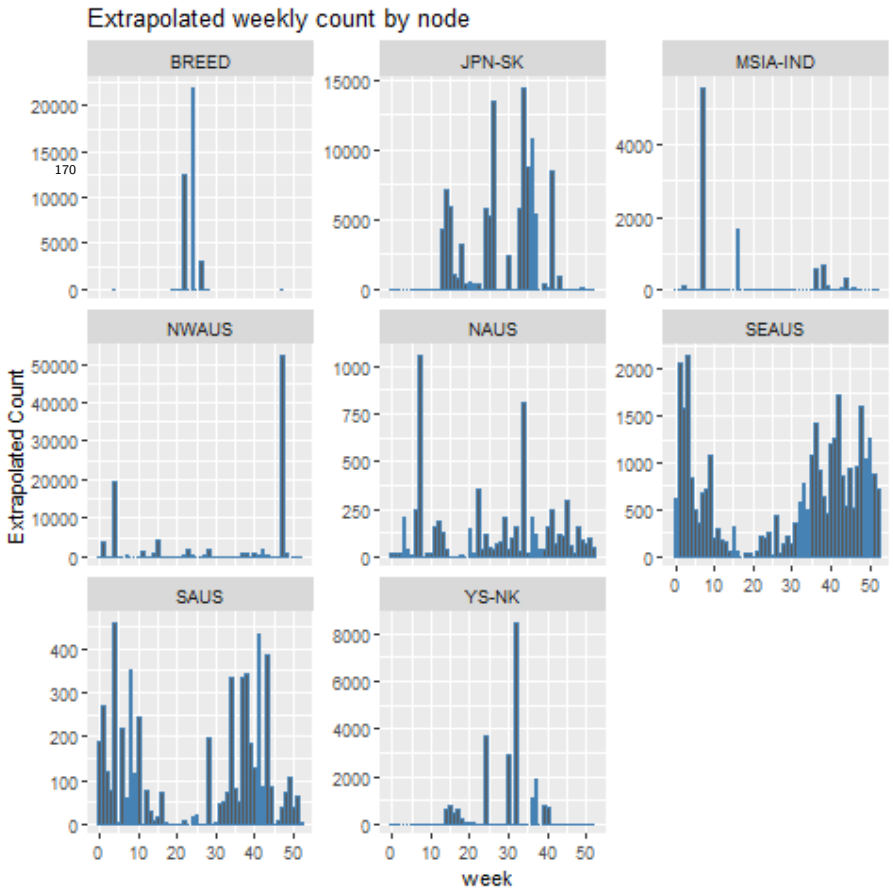


Figure 4: Extrapolated weekly counts for each node. Note that the y-axis has different scales for each node.

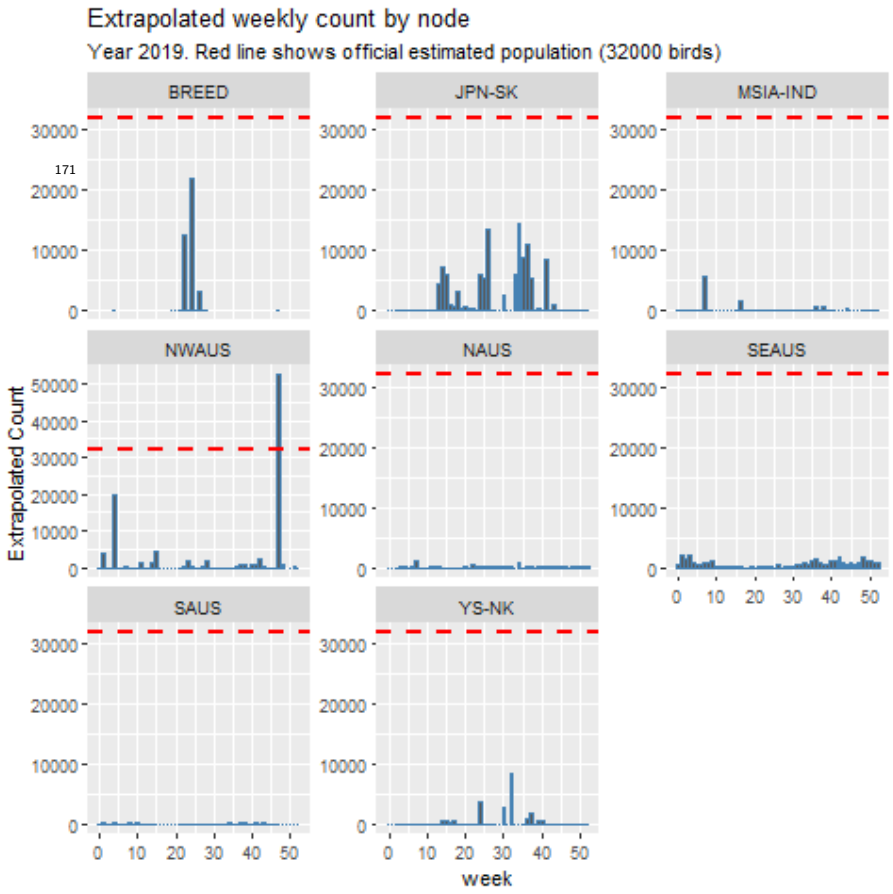


Figure 5: Extrapolated weekly counts for each node scaled relative to a recent total curlew population estimate (32000 birds).

From eBird checklists to an observed count per node

Here we explain formally how we build the variable $O = O(i,t,y)$ in the article, where i is one of the migration network nodes and t is a week during the migration period of a given year[2].

First we need to compute the area of habitat in node i. As explained above, we take the intersection surface between the node boundary (see Figure 3) and the map of Eastern Curlew spatial distribution from BirdLife (Figure 6). We name this area $A(i)$.

We divide $A(i)$ into regular spatial units, one spatial unit being a hexagon of area 100km^2. For a given year, y, and a given week, t, in some of these hexagons there is no checklist. We compute the mean count across all hexagons containing at least one checklist. This mean count is extrapolated over the remaining hexagons to obtain a total abundance estimate $O = O(i,t,y)$ for the node.

Formally, let:

[2] In the article, we do not explicit include the year y in the notation for O, but here we include it for completeness.

$A(i)$ be the area of node i. $A(i)$ is divided into hexagons of area 100km^2. For a given year, y, and a given week, t, in some of these hexagons there is no checklist.

$H(i)$ be the set of hexagons that lie within node i. Let $n_{Hex}(i)$ be the total number of 100km^2 hexagons in node i, such that $n_{Hex}(i) = |H(i)|$.

172

$H^+(i,t,y)$ be the set of hexagons in node i that contain at least one checklist in week w of year y. Let $n^+_{Hex}(i,t,y)$ be the number of hexagons in node i that contain at least one checklist in week t of year y, s.t. $n^+_{Hex}(i,t,y) = |H^+(i,t,y)|$.

$C(h,t,y)$ is the mean of the counts over the lists in hexagon h in week t and year y. $C(h,t,y)$ is only defined over $H^+(i,t,y)$ and the counts in each hexagon are indexed by $h \in \{1, \dots, n^+_{Hex}(i,t,y)\}$.

In the hexagons with at least one list, the mean number of birds observed per hexagon is:

$$\frac{\sum_{h \in H^+(i,t,y)} C(h,t,y)}{n^+_{Hex}(i,t,y)}$$

Extrapolating this mean number of birds across all hexagons in the node, we obtain an estimate of $O(i,t,y)$, the observed number of birds in node i in week t and year y:

$$O(i,t,y) = n_{Hex}(i)\left(\frac{\sum_{h \in H^+(i,t,y)} C(h,t,y)}{n^+_{Hex}(i,t,y)}\right)$$

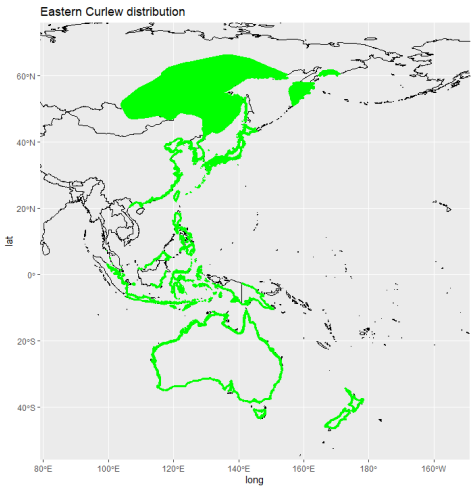


Figure 6: Map of spatial distribution of the Eastern Curlew (Birdlife International. 2020)

Initial node population estimates

Our model requires an estimate of the initial state of the population but does not account for observation error for the initial timestep. Consequently, rather than rely on noisy estimates of initial counts based on the eBird checklists analysis described in the previous sections, we instead seeded our model with initial node counts based on expert estimates of the proportion of birds in each node during the non-breeding season (proportions are derived from figure A3 of Iwamura et al 2013).

To compute initial node counts (Table 6), we multiplied the non-breeding proportions of the population at each node by an assumed Eastern Curlew population size of 32000 (Hansen et al 2016).

Table 6: Initial node population estimates based on expert estimates from Iwamura et al., 2013.

[173]

NODE	Non-breeding proportion of population	Initial node population estimate (t=0)
MSIA-IND	0.27	8640
NWAUS	0.06	1920
NAUS	0.05	1600
SEAUS	0.45	14400
SAUS	0.17	5440
JPN-SK	0	0
YS-NK	0	0
BREED	0	0

Estimating the reporting probability in the negative binomial model

Since the eastern curlew data were overdispersed (i.e., the variability in the data are much greater than the mean counts), we used a negative binomial distribution to model the observations. The negative binomial model requires two parameters, but to reduce the computational burden of the estimation, we chose to fix one of the parameters (see Supporting Information S2 for details). We fixed the parameter p, the probability of success of each trial.

To estimate p for the eastern curlew case study, we used the MASS package in R to fit a negative binomial distribution to the total number of observed curlews in all nodes for each week (i.e. $\sum_i O(i,t,y)$; Figure 7). The resulting estimates for p were 2.95×10^{-4} in 2018 and 1.46×10^{-4} in 2019. We then estimated the other parameter using the approach detailed in Supporting Information S2.

One outlier was removed from the 2018 count data to compute these estimates. Specifically, the YS-NK count in week 15 was removed from the 2018 count due to an abnormally large count that led to a large overestimate in the total number of birds.

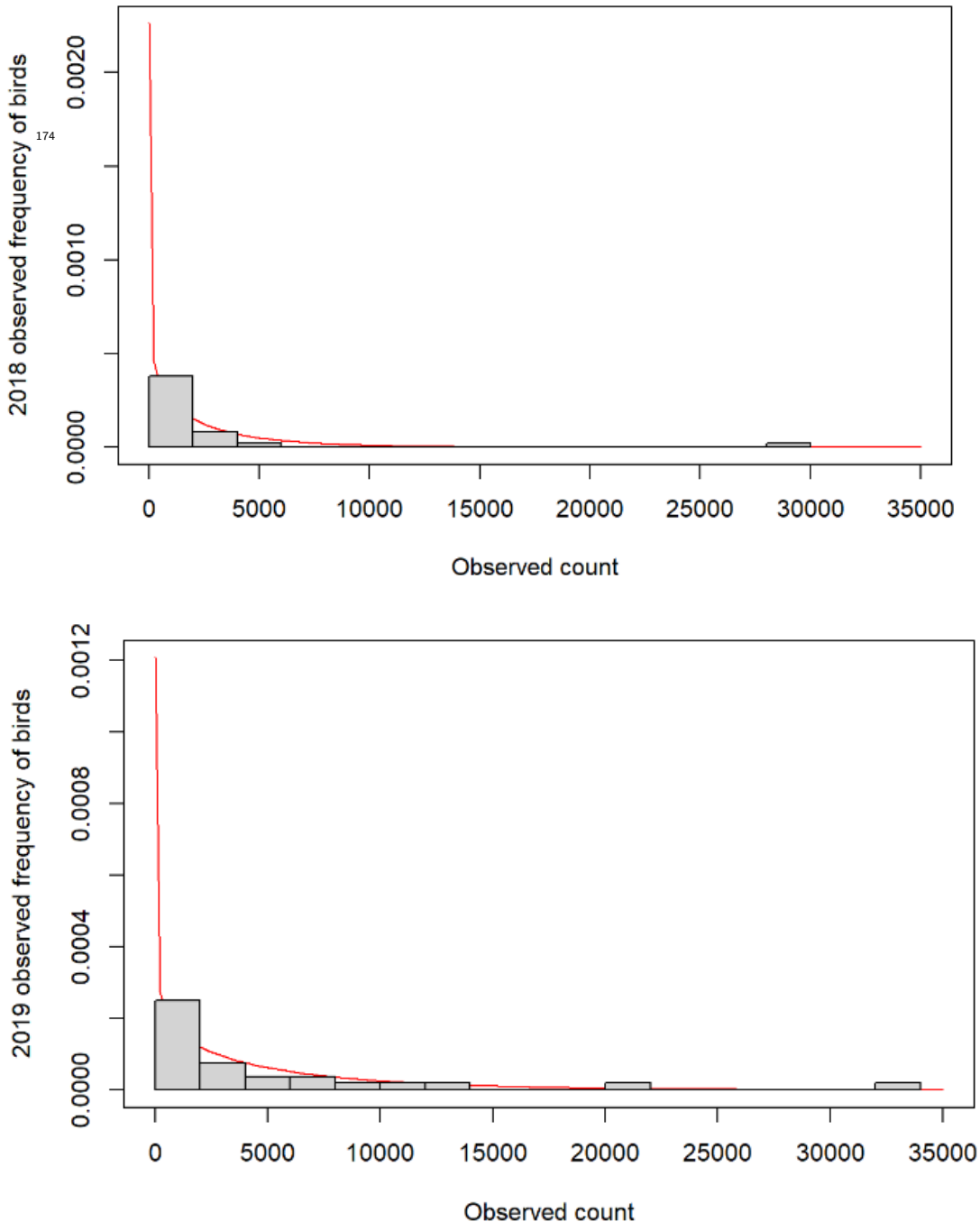


Figure 7: Best-fit negative binomial distribution to summed weekly observed curlew data for 2018 and 2019 respectively. These distributions were used to fix the parameter p in the curlew case study experiments.

Since the value of parameter *p* is estimated in the curlew case study (but known in the benchmark experiments), we undertook a sensitivity analysis to investigate the impact of mis-specifying the value of *p* (Figure 8). To do this, we varied the value of *p* estimated from the 2018 data to observe the resulting changes to log-likelihood. Our experiments showed that using the empirical estimate of *p* from the observed data resulted in the highest log-likelihood. Where the parameter p is unknown, using the empirical estimate of *p* is likely to provide a good fit to the data.

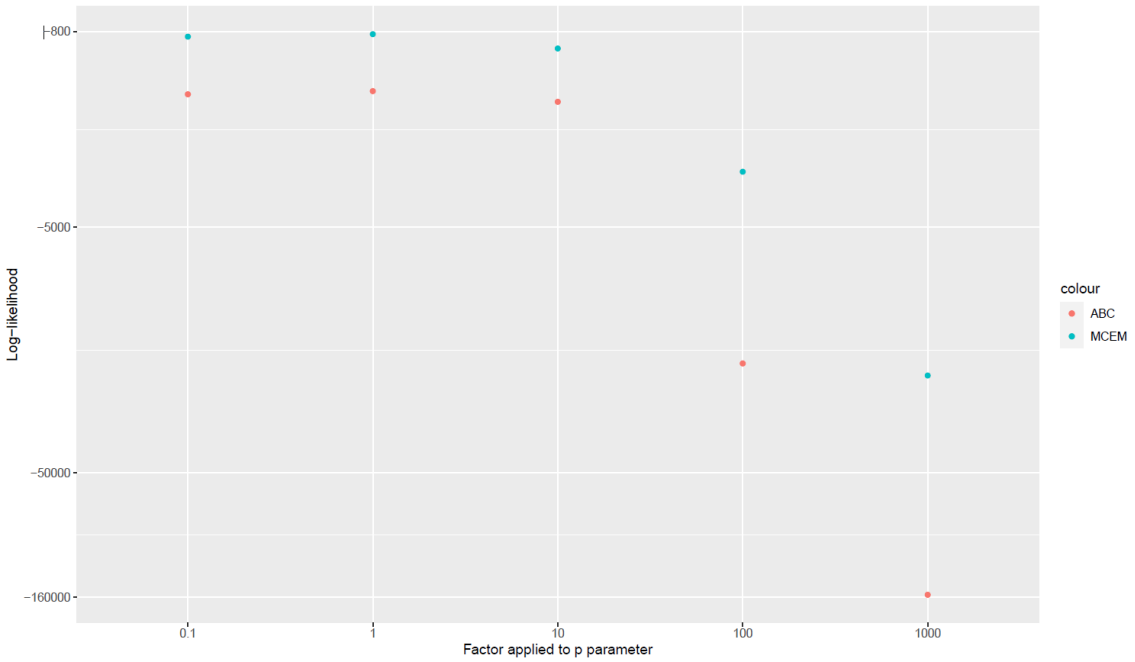


Figure 8: Sensitivity of log-likelihood to variability in the parameter p. X-axis values depict the size of a factor applied to the value of p estimated from the 2018 curlew data (i.e., p= 2.95×10^{-4}). Y-axis values show the resulting estimates of the log-likelihood, with larger values corresponding to better fits to the data.

As well as affecting the likelihood, mis-specifying *p* can affect the resulting estimated parameters. For example, the mean absolute difference[3] between parameters estimated with *p* =0.1 and *p* =1000 with MCEM in this case is 0.38. Combined with our experiments about the log-likelihood which suggested that using the empirical value of *p* gives a good fit to the data, this result suggests that using other values of *p* will result in worse parameter estimates than using the empirical estimate of *p*.

[3] The mean absolute difference has the same formula as the mean absolute error in the benchmark experiments, but is used here because we compare two networks without knowing the true one. Its value lies in [0;1] and is larger when the estimated parameters from the two networks being compared are more different from each other.

References

Bamford M, Watkins D, Bancroft W, Tischler G, Wahl J. 2008. Migratory shorebirds of the East Asia-Australasia flyway: population estimates and internationally important sites. Canberra, Australia: Wetlands International-- Oceania. Report no.

Birdlife International. 2020. Species factsheet: Numenius madagascariensis. Downloaded from http://www.birdlife.org on 10/07/2020. . Report no.

BirdLife International and Handbook of the Birds of the World. 2019. Bird species distribution maps of the world. Version 2019.1. Available at http://datazone.birdlife.org/species/requestdis.

Driscoll PV, Ueta M. 2002. The migration route and behaviour of Eastern Curlews Numenius madagascariensis. Ibis 144: E119-E130.

Hansen BD, Fuller RA, Watkins D, Rogers DI, Clemens RS, Newman M, Woehler EJ, Weller DR. 2016. Revision of the East Asian-Australasian Flyway Population Estimates for 37 listed Migratory Shorebird Species. Unpublished report for the Department of the Environment. Melbourne: BirdLife Australia. Report no.

Iwamura T, Possingham HP, Chadès I, Minton C, Murray NJ, Rogers DI, Treml EA, Fuller RA. 2013. Migratory connectivity magnifies the consequences of habitat loss from sea-level rise for shorebird populations. Proceedings of the Royal Society B: Biological Sciences 280: 20130325.

Minton C, Jessop R, Collings P, Standen R. 2011a. The migration of Eastern Curlew Numenius Madagascariensis to and from Australia. Stilt 59: 6-16.

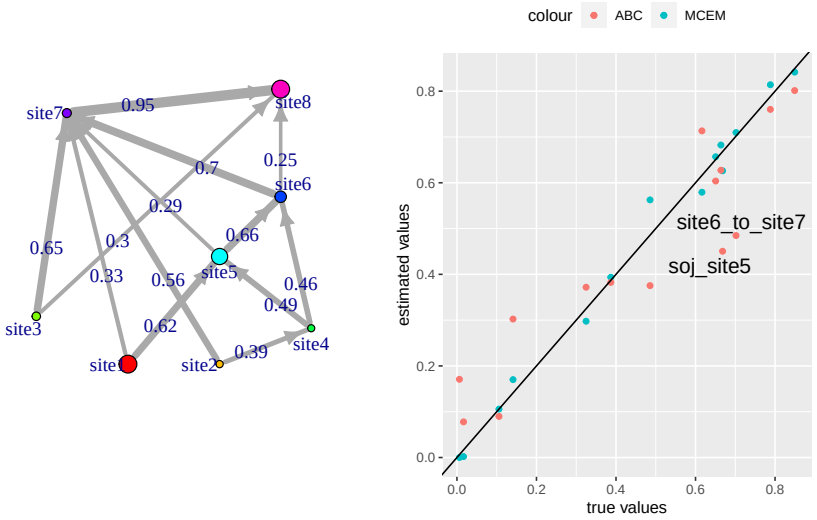
Minton C, Wahl J, Gibbs H, Jessop R, Hassell C, Boyle A. 2011b. Recoveries and flag sightings of waders which spend the non-breeding season in Australia. Stilt 59: 17-43.

Minton C, Gosbell K, Johns P, Christie M, Klaassen M, Hassell C, Boyle A, Jessop R, Fox J. 2013. New insights from geolocators deployed on waders in Australia. Wader study group bulletin 120: 37-46.

Strimas-Mackey M, Hochachka W, Ruiz-Gutierrez V, Robinson O, Miller E, Auer T, Kelling S, Fink D, Johnston A. 2020. Best Practices for Using eBird Data. Version 1.0 https://cornelllabofornithology.github.io/ebird-best-practices/. Ithaca, New York. : Cornell Lab of Ornithology.

S6. Benchmark problem illustration

To go further into the analysis of the estimations results, we identified a case which is similar to the case study of the Eastern Curlew bird migration. We selected a benchmark problem with 8 sites, 2 destination sites per site and 0% of missing observations.



(a) (b)

Figure 1: (a) True network to estimate; the labels on edges stand for the transition probability, the size of the site is proportional to the mean sojourn time and the *delta* parameter is 0.85. (b) Estimated parameters with ABC (red points) and MCEM (blue points).

In Figure 1, the network structure of this case is plotted and estimations of parameters are compared to true parameters. There are three departure sites (sites 1, 2 and 3) and site 5 is an important stopover site. Log-likelihoods are respectively -585 for the true parameters, -1070 for the mode value of the marginals of the ABC posterior distribution and -582 for the MCEM estimated parameters. The mean absolute error is 0.0218 and 0.0897 for the ABC estimation and the MCEM estimation respectively.

In this case, MCEM succeeds in inferring as likely parameter values as the

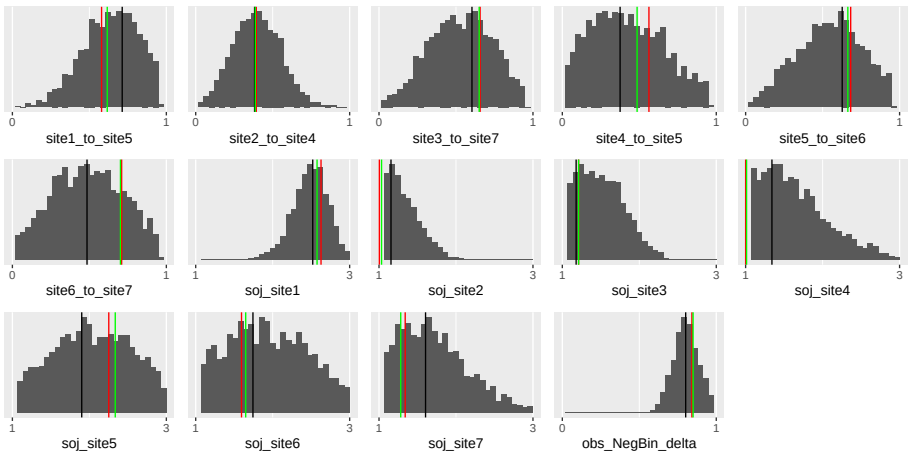


Figure 2: The marginals of the posterior distribution for each parameter estimated with ABC; the green vertical lines represent the true values, the black lines represent the Venter mode of the marginals of the posterior distribution and the red lines represent the estimation by MCEM.

true parameters. ABC estimated parameters are very close to the true parameters. However ABC fails to correctly infer the transition probability from site 6 to site 7 and the mean sojourn time on site 5, with an error greater than 20%. Looking at the full posterior distribution computed with ABC (see Figure 2), the marginal distributions are rather flat, which underlines a large uncertainty for these two parameters. For this case, the shape of the marginal posterior distributions provides a way to diagnose the quality of the estimations.

S7. Venter mode values for estimated parameters in the eastern curlew case study

Table 2 contains Venter mode values for the estimated transition (top) and sojourn time (bottom) parameters in the curlew case study.

Transition parameters	SAUS to SEAUS	SAUS to JPN-SK	SAUS to YS-NK	SEAUS to NAUS	SEAUS to JPN-SK	SEAUS to YS-NK	NAUS to JPN-SK	NAUS to YS-NK
2019 observations								
ABC	0.32	0.03	0.65	0.36	0.03	0.61	0.42	0.58
MCEM	0.41	0.42	0.07	0.89	0.05	0.06	0.50	0.50
2018 observations								
ABC	0.62	0.03	0.35	0.46	0.03	0.50	0.61	0.39
MCEM	0.19	0.35	0.46	0.35	0.50	0.25	0.53	0.47

Sojourn parameters	SAUS	SEAUS	NWAUS	NAUS	MSIA-IND	JPN-SK	YS-NK	Observation parameter δ
2019 observations								
ABC	8.2	9.0	8.3	7.8	10.9	2.6	2.1	0.04
MCEM	10.0	10.8	13.0	4.0	18.3	1.05	0.15	0.23
2018 observations								
ABC	7.8	7.7	9.6	7.2	8.0	1.9	2.2	0.009
MCEM	13.0	9.9	12.0	4.9	18.4	2.05	0.37	0.11

Table 1: Venter modes of the estimated parameters for the Curlew network. Sojourn times are reported in weeks.

S8. ABC based on loglikelihood

Using MCEM, the estimation seeks to maximize the log likelihood of the parameters (see Supporting Information S2), whereas, using ABC, the estimation seeks to provide a set of parameter values for which the simulation of site-week bird counts satisfies a criterion of proximity to observations (see Supporting Information S3). One hypothesis to explain the differences between MCEM and ABC estimates in the case study is the difference in this criterion. To investigate it, we used a modified version of ABC that relies on only one statistic (compared to all site-week observations) which is the log-likelihood of parameters. By doing this, we make the objective of ABC closer that of to MCEM.

Formally, in equation (7), instead of using the set of statistics $\{O(\Lambda_k)_i^t, (i,t) \in \Omega\}$, we used the statistics $\sqrt{-log(P(O|\Lambda_k))}$, for which evaluation is given in equation (12). As a target we used 0 instead of $\{O_i^t, (i,t) \in \Omega\}$. Thus the proximity criterion used in ABC is $\rho(O, O(\Lambda_k)) = \frac{-log(P(O|\Lambda_k))}{v_{\mathcal{L}}}$ where $v_{\mathcal{L}}$ is the variance of log likelihood computed into the preliminary Latin Hypercube Sample. Note that the cost in terms of number of simulations to evaluate the log likelihood of Λ_k is much more substantial than the simulation of one matrix of site-week observations.

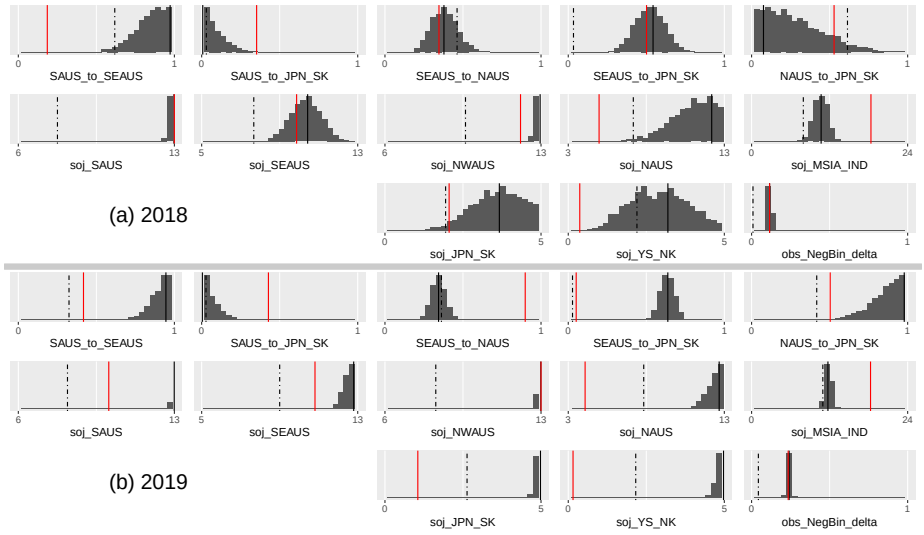


Figure 3: Case study: posterior distributions provided by ABC when we use only the log likelihood of the parameters as statistic, for year 2018. Black dashed vertical lines represent the Venter mode of the marginals. Red vertical lines represent the estimated parameters computed by MCEM. Dashed lines represent the Venter mode of the ABC posterior distribution shown in Figure 5.

As expected, the modified "ABC likelihood" experiment resulted in a log-likelihood that was much closer to the MCEM likelihood (Table 2). This demonstrates the difference in objectives is a likely reason for differences in the final parameter values, since the parameters chosen by ABC and "ABC likelihood" are heavily affected by the change in objective. However, the similarity in likelihood values did not obviously lead to more similar parameter estimates between

"ABC likelihood" and MCEM (Figure 3). This is likely due to the fact that there may be many sets of parameters that give similar likelihood values, so there is no guarantee that they will provide similar estimates when approaching the same problem in different ways. In particular, if the likelihood surface is relatively flat (as we suspect for the curlew problem), then small changes in the likelihood can result in considerable changes in the parameter values. Our experiments demonstrate the importance of the objective in determining in the final parameter set, and suggest that this may be one likely reason why our ABC and MCEM results differ.

Year	ABC		ABC likelihood		MCEM
	Venter	Best	Venter	Best	
2018	-1397 ± 112	-1054 ± 146	-795 ± 0.3	-800 ± 7	-819 ± 0.2
2019	-2776 ± 1722	-1066 ± 1.1	-913 ± 0.2	-914 ± 0.2	-983 ± 0.3

Table 2: Log likelihood of different set of parameters and their uncertainty computed by bootstrap. Here, 'ABC' stands for the experiment with ABC using the full matrix of observations as statistics. 'ABC likelihood' stands for the experiment with ABC where the only statistics is the loglikelihood (this section). 'Venter' stands for the Venter mode of the marginals and 'Best' stands for the set of parameters of the ABC posterior which simulated statistic is the closest to the observation. For 'ABC likelihood', the 'Best' particle can be interpreted as the MAP (Maximum As Posteriori). Finally, 'MCEM' stands for the set of parameters estimated with MCEM.

References

Lenormand, M., Jabot, F. & Deffuant, G. (2013), 'Adaptive approximate Bayesian computation for complex models', Computational Statistics **28**(6), 2777–2796.

URL: *http://dx.doi.org/10.1007/s00180-013-0428-3*

Yu, S.-Z. (2015), Hidden Semi-Markov Models: Theory, Algorithms and Applications, 1st edn, Elsevier Science Publishers, Amsterdam.