Oikos

o19044
Pedersen, M. W., Patterson, T. A., Thygesen, U. H. and Madsen, H. 2011. Estimating animal behavior and residency from movement data. – Oikos 120: 1281–1290.

Appendix 1

Computing the generator and the probability transition matrices

The movement dynamics of the animal is described by a Brownian motion. As shown in Okubo (1980) the time evolution of the probability density of the animal's location is described by the advection-diffusion equation

$$\frac{\partial \varphi_i}{\partial t} = -\nabla \cdot (u_i \varphi_i - D_i \nabla \varphi_i). \quad (1)$$

To derive the entries of the movement generator matrix we write Eq. 1 on its finite difference form (Mitchell and Griffiths 1980) in one dimension:

$$\varphi(t_k, x) = (\mu - v)dt\varphi(t_{k-1}, x + dx) + (1 - 2\mu\mu dt\varphi(t_{k-1}, x) + (\mu + v)dt\varphi(t_{k-1}, x - dx)$$

where $\mu = \frac{D}{dx^2}$, $v = \frac{u}{2dx}$ and $dt = t_k - t_{k-1}$. From this we gather that the probability of a step of length ±dx approaches $(\mu \pm v)dt$ as $dt \downarrow 0$. A stable finite difference scheme has the conditions that $dt < \frac{dx^2}{2D}$ and $|u| < \frac{2D}{dx}$ where $|\cdot|$ means absolute value.

This leads to the one dimensional generator of the movement process having the entries

$$g_{ij} = \begin{cases} \mu - v & \text{for} \quad j = i - 1 \\ \mu + v & \text{for} \quad j = i + 1 \\ -2\mu & \text{for} \quad j = i \\ 0 & \text{otherwise} \end{cases}.$$

This generalizes to two dimensions with the generator having two additional entries in each row except at boundary locations.

The dynamics of the behavior and movement in behavior state i are described in continuous time by their generators $\mathbf{G}^b$ and G_i^m respectively. For a given time interval Δ_k, probability transition matrices can be computed for the behavior process by $P^b(t_k) = \exp(G^b \Delta_k)$, where exp( ) means the matrix-exponential operation (Grimmett and Stirzaker 2001). Calculating transition matrices for large state-spaces requires a matrix exponential implementation that utilizes the so-called uniformization algorithm (Grassmann 1977) which exploits the sparsity of the generator matrix.

We write

$$P^b(t_k) = \begin{bmatrix} p_1^b(t_k) \\ \vdots \\ p_n^b(t_k) \end{bmatrix} \tag{2}$$

where $p_i^b(t_k)$ are row vectors containing transition probabilities conditional on state i. The probability transition matrices of the movement processes are analogously given by $P_i^m(t_k) = \exp(G_i^m \Delta_k)$.

We can assemble $P^b(t_k)$ and $P_i^m(t_k)$ into a probability transition matrix that describes the joint process of behavior and movement

$$P_k = \begin{bmatrix} p_1^b(t_k) \otimes P_1^m(t_k) \\ \vdots \\ p_n^b(t_k) \otimes P_n^m(t_k) \end{bmatrix} \tag{3}$$

where $\otimes$ is the Kronecker product operator. The matrix $\mathbf{P}_k$ has a block structure which is illustrated for a simple case in Fig. 1 in the main text.

Appendix 2

State and parameter estimation

As described above, the $\mathbf{P}_k$ matrices merge the two processes of movement and behavior into a single Markov process in which all spatial and behavioral dynamics are captured. For this process the symbol $\alpha = (x, y, i)$ is used to represent a state, where x and y refer to position in the two-dimensional space and i indexes the n behavioral states. We have partitioned the longitudinal and latitudinal directions into n_x and n_y cells and therefore the total number of spatial states is $n_{xy} = n_x n_y - n_u$, where n_u is the number of cells inside the grid that are inaccessible to the animal (such as land areas for marine animals). The probability distribution of the position and behavior states at time t_k is therefore a column vector $\varphi(t_k \mid Z_k)$ of length $n_{xy}n$ since n vectors of length n_{xy} are concatenated. The vector

$\varphi(t_k \mid Z_k)$ has elements $\varphi_\alpha(t_k \mid Z_k)$.

A HMM filter recursion consists of a time and a data update step. The time update gives the predicted distribution $\varphi(t_{k+1} \mid Z_k)$ and the data update gives the estimated distribution $\varphi(t_k \mid Z_k)$. The time update of the probability distribution is a simple multiplication of the state probability vector with the transition matrix

$$\varphi(t_{k+1} \mid Z_k)^T = \varphi(t_k \mid Z_k)^T P_k. \tag{4}$$

In the data update step the predicted distribution $\varphi(t_k \mid Z_{k-1})$ is adapted to the observation $\mathbf{z}_k$ by applying Bayes' rule

$$\varphi(t_k \mid Z_k) = \psi_k^{-1} \varphi(t_k \mid Z_{k-1}) \bullet L(z_k \mid \alpha) \tag{5}$$

where $\bullet$ denotes elementwise multiplication, $L(z_k \mid \alpha)$ is the likelihood of $\mathbf{z}_k$ given the state and $\psi_k = \sum_\alpha \varphi(t_k \mid Z_{k-1}) \bullet L(z_k \mid \alpha)$. The data likelihood $L(z_k \mid \alpha)$ has a value for all α and is computed by comparing the observed data to the data expected to be generated in a given state. The way to compute $L(z_k \mid \alpha)$ depends on the form of the mapping function h in the observation equation (Eq. 2 in the main text). For example if $\mathbf{z}_k$ are noisy positions e.g. from satellite tags, h is linear which makes computations simple

$$L(z_k \mid \alpha) = N_{pdf}(z_k, \alpha, \Sigma_w), \quad \text{for all} \alpha$$

where N_{pdf} is a Gaussian probability density function with mean α and covariance matrix Σ_w evaluated at $\mathbf{z}_k$. For outlier-prone observations such as Argos positions, a heavy tailed t-distribution may be applied instead of a Gaussian. Even more complex and non-linear links may be implemented if needed, see for example Pedersen et al. (2008).

The likelihood value of a given parameter set, θ, is a product of the one-step prediction errors

$$L(\theta \mid Z_N) = \prod_{k=1}^{N} \psi_k. \tag{6}$$

Maximum likelihood estimates are obtained by optimizing $L(\theta \mid Z_N)$ with respect to θ.

To incorporate all observations in each state estimate, i.e. to get $\varphi(t_k \mid Z_N)$, the so-called smoothing step is required. The smoothed state estimates are therefore not only conditioned on data observed by t_k but also on future measurements and are more accurate and appear 'smoother' than $\varphi(t_k \mid Z_k)$.

We state the smoothing recursions, but omit derivation details. For supplements on the smoothing step see Thygesen et al. (2009). The recursions are

1. Compute the vector

$$\Psi(t_{k+1}) = \varphi(t_{k+1} \mid Z_N) // \varphi //_{k+1} \mid Z_k)$$

where // is elementwise division.

2. Right multiply with the transition matrix to step backwards in time

$$\Lambda(t_k) = P_k \Psi(t_{k+1})$$

3. Get the smoothed estimate at t_k by

$$\varphi(t_k \mid Z_N) = \varphi(t_k \mid Z_k) \bullet \Lambda(t_k)$$

where $\bullet$ denotes elementwise multiplication.

The recursion is initiated with the last estimated distribution from the final iteration of the forward filter, $\varphi(t_N \mid Z_N)$ which is also a smoothed estimate.

Appendix 3

Finding the most probable track

The likelihood of a track (an outcome of the posterior distribution), $a = (\alpha_1, \ldots, \alpha_N)$, can be computed as

$$L(a) = L(z_1 \mid \alpha_1) \prod_{k=2}^{N} p_{\alpha_{k-1}, \alpha_k} L(z_k \mid \alpha_k)$$

where $p_{\alpha_{k-1}, \alpha_k}$ is the entry in P_{k-1} corresponding to the transition from α_{k-1} to α_k. For summarizing movement and behavior switching we use the track that maximizes L(**a**) denoted $\hat{a}$. For HMMs, estimating $\hat{a}$ is an often occurring problem that can be solved by the Viterbi algorithm (Viterbi 2006) which relies on principles from dynamic programming and is proved to be a maximum likelihood estimator (Forney 1973).

We define the branch metric

$$B_{\alpha_{k-1}, \alpha_k}(t_k) = \log p_{\alpha_{k-1}, \alpha_k} + \log L_{\alpha_k}(z_k \mid \alpha_k)$$

An intermediate step in the maximization algorithm uses the state metric, $S_{\alpha_k}(t_k)$, which is the log-likelihood of the most likely of all possible tracks leading from the initial state to state α_k. The state metric is given by

$$S_{\alpha_k}(t_k) = \max_{\alpha_1, \ldots, \alpha_k} \left\{ \log L_{\alpha_1}(z_1 \mid \alpha_1) + \sum_{l=2}^{k} B_{\alpha_{l-1}, \alpha_l}(t_l) \right\}$$

This maximization problem can be solved recursively forward in time when it is realized that

$$S_{\alpha_k}(t_k) = \max_{\alpha_k} \left\{ S_{\alpha_{k-1}}(t_{k-1}) + B_{\alpha_{k-1}, \alpha_k}(t_k) \right\}$$

The procedure exploits the Markov property of the HMM to reject all but the most likely paths after each recursion.

The final state of the overall most probable track is given by

$$\hat{\alpha}_N = \arg\max_{\alpha_N} S_{\alpha_N}(t_N)$$

By continuously storing the most probable tracks for each iteration of the recursion the overall most probable track, $\hat{a}$, is simply given by extracting the track related to $\hat{\alpha}_N$ when iterations are finalized.

Appendix 4

Simulation study with results

To better understand the performance of the HMM approach with respect to estimation and model selection, we first applied the method to synthetic data sets. We examined if a relatively complex model could be reliably differentiated from simpler, candidate models. The synthetic data sets were generated with a two-state switching model comprised of a resident state with low diffusivity and no advection and a migratory state with a higher diffusivity and advection (i.e. the SDA model, main text Table 1).

The simulation was intended to mimic the natural behavior of southern bluefin tuna (SBT) *Thunnus maccoyii*. These fish make long distance migrations into the Indian Ocean from the Great Australian Bight (Bestley et al. 2008). The parameter values of the data generating movement model were $(D_1,D_2,u_x,u_y,p_{11},p_{22})$ =(300,1000,–50,0,0.95,0.95) and the initial location of the fish (35.5°S, 126.7°E) was considered as known. The synthetic data of sea surface temperature (SST) and longitude were collected daily for 183 days from the simulated horizontal movements of the fish (see main text for details on error model).

The quality of the estimated most probable tracks was quantified using the root mean square (RMS) error of the residuals, σ_{MPT} (as compared to the true track). To evaluate the model's ability to correctly estimate the behavioral state, the average number of misclassified states avg(n_{mis}) was calculated. A total of 50 synthetic data sets were generated. For each data set maximum likelihood parameter estimation, model selection between the four models listed in the main text Table 1, state estimation and estimation of the most probable track was undertaken.

Results

AIC based model selection of the synthetic data sets resulted in 50 out of the 50 analyses arriving at the correct model (SDA) as the final model. The significant advection term made the data generating model easily distinguishable from the simpler switching model, SD, for which the average of the maximum likelihood estimates of D_2 was 2358 km^2 day^{-1}. It was clear that the estimation process compensated for the lack of advection by inflating the diffusivity estimate, thus decreasing the likelihood value of this model. Generally estimation of models D, DA and SD did not result in correct parameter values. This was expected since their model structures deviate from the data-generating model. The correct model however, did provide parameter estimates that were identifiable and consistent with the true parameters (Table 1).

The average numbers of misclassified states, avg(n_{mis}), showed that state switching was significantly better estimated by the SDA model as compared to the SD model, and in the optimal situation when the filter model is equal to the data generating model 87% of the behavior states were correctly estimated. The average value of σ_{MPT} for the SDA model was not significantly different to the three other models, which indicates that reasonable track estimates was still obtained even when the applied model differed from the data-generating model.

Table 1. Simulation results. Empirical averages (avg) and standard deviations of the averages (sd) of ML estimates and statistics from the 50 synthetic data sets. σ_{MPT} is the root mean square of the residuals of the estimated most probable track compared to the true track with unit km. n_{mis} is the average number of misclassified state estimates. Unit for D_i is km^2 day^{-1}, unit for u_x and u_y is km day^{-1}.

		D_1	D_2	u_x	u_y	p_{11}	p_{22}	σ_{MPT}	n_{mis}
D	avg	1353	–	–	–	–		88.9	–
	sd	58.9	–	–	–	–		2.9	–
DA	avg	1004	–	–19.4	–0.1	–		84.1	–
	sd	31.1	–	1.15	0.37	–		2.8	–
SD	avg	276	2358	–	–	0.94	0.94	84.7	39.6
	sd	17.2	85.0	–	–	0.006	0.009	2.8	2.4
SDA	avg	307	1060	–48.4	–1.2	0.94	0.92	79.2	23.5
	sd	15.1	36.0	1.23	0.85	0.008	0.009	2.3	1.7

Appendix 5

Estimation of data error variances

Using diagnostics data transmitted by the PSAT we were able to examine the relationship between the remotely-sensed surface temperature and that measured by the PSAT. While there were a few departures this relationship was strongly linear (Fig. 1a). The residuals from the fit were approximately Gaussian distributed (Fig. 1b) and the residual variance, $\sigma^2_L=(0.71°C)^2$, was therefore used as estimate for the temperature error variance in the model.
An average of the empirical root mean square estimates determined in Musyl et al. (2001) was used as longitude error variance, $\sigma^2_L=(35\ km)^2$.

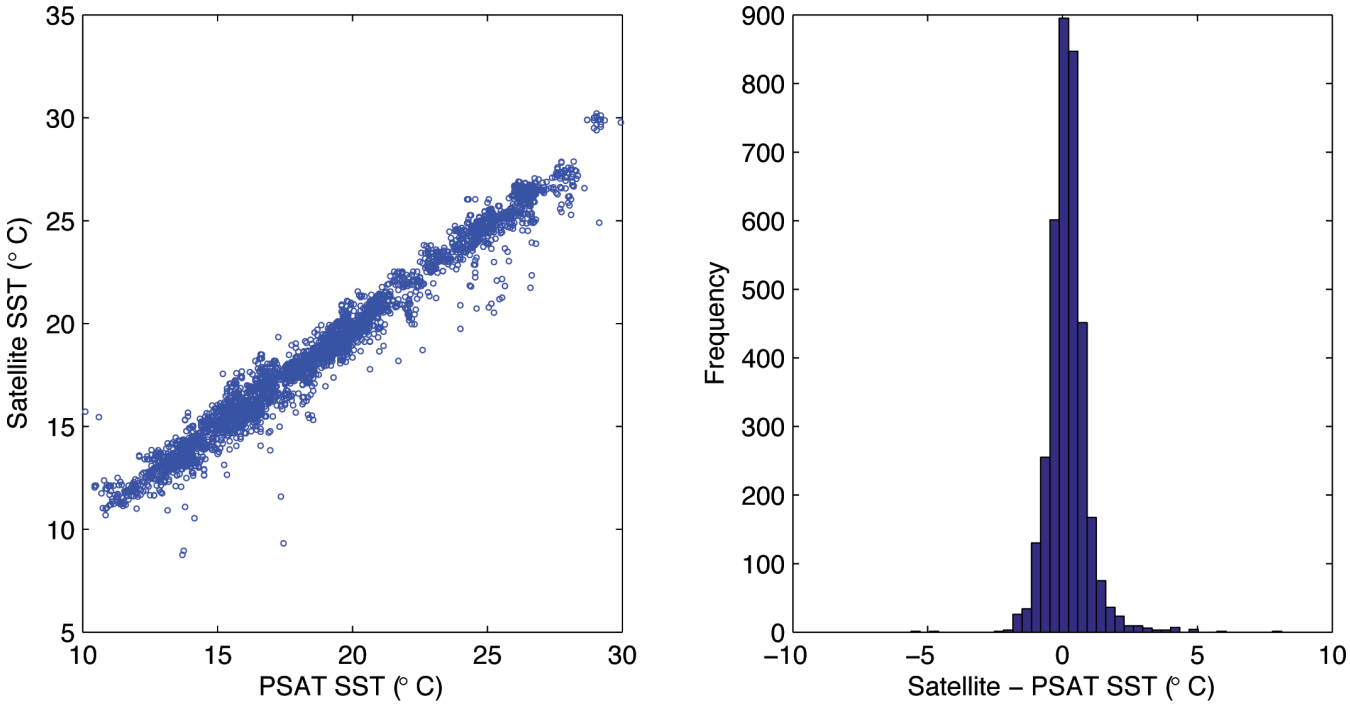

Figure 1. Left: plot of SST as measured by PSAT versus remotely sensed SST by satellite. A strong linear relation is clear. Right: histogram of the residuals of a linear regression model for the SST data in the left pane. The errors are approximately Gaussian distributed with zero mean.

Video Appendix

Animation of time marginals

The animation is found via this link: <www2.imm.dtu.dk/~mwp/VA1_animated_time_marginals.avi>

References

Bestley, S. et al. 2008. Feeding ecology of wild migratory tunas revealed by archival tag records of visceral warming. – J. Anim. Ecol. 77: 1223–1233.

Forney, G. D. 1973. The Viterbi algorithm. – Proc. IEEE 61: 268–278.

Grassmann, W. 1977. Transient solutions in Markovian queueing systems. – Comput. Oper. Res. 4: 47–53.

Grimmett, G. and Stirzaker, D. 2001. Probability and random processes. – Oxford Univ. Press.

Mitchell, A. and Griffiths, D. 1980. The finite difference method in partial differential equations. – Wiley.

Musyl, M. et al. 2001. Ability of archival tags to provide estimates of geographical position based on light intensity. – In: Sibert, J. R. and Nielsen, J. L. (eds), Electronic tagging and tracking in marine fisheries. Kluwer, pp. 343–367.

Okubo, A. 1980. Diffusion and ecological problems: mathematical models. – Springer.

Pedersen, M. W. et al. 2008. Geolocation of North Sea cod (*Gadus morhua*) using hidden Markov models and behavioural switching. – Can. J. Fish. Aquat. Sci. 65: 2367–2377.

Thygesen, U. H. et al. 2009. Geolocating fish using hidden Markov models and data storage tags. – In: Nielsen, J. et al. (eds), Tagging and tracking of marine animals with electronic devices. Springer, pp. 277–293.

Viterbi, A. J. 2006. A personal history of the Viterbi algorithm. – IEEE Signal Proc. Mag. 23: 120–142.