State space model for estimating geographic positions from light measurements

Anders Nielsen & John Sibert

anders.nielsen@hawaii.edu
Geolocation by light level

- Often a two step process is applied
  1. Get “raw daily geolocations” from tag manufacturer or via their software
  2. Try to reconstruct “most probable track” from raw geolocations
Why combine?

- With two step approach there is no feedback and not all is carried through to kftrack
- Kftrack assumes a fairly complex annual variance structure
- Kftrack estimates several (≈ 4) variance parameters
- Raw geolocations can sometimes contain extreme outliers, as the method has no knowledge of the entire track
- Extreme outliers can obstruct the final most probable track
- Frustrating to see patterns that can only be explained as artifacts of the unknown details of the raw geolocation

We aim to get one model that is transparent, statistically sound, and without ad-hoc solutions.
Focus on the data that matters

- State space model with 10 minute time steps

- We see that the track is only influenced by data in certain periods
- This is also useful information if storage or transmission capacity is limited
The scanning algorithm we apply

1. Divide the light record into 24-hour periods

2. Within each, select the two observation times such that the mean of the light measurements between the two times ($\tilde{t}_i$ and $\tilde{t}_{i+1}$) and the mean of the remaining light measurements times differ the most. In other words:

$$\tilde{t}_i, \tilde{t}_{i+1} = \arg\max_{a,b} \left| \text{mean}(\{l_j : a < t_j \leq b\}) - \text{mean}(\{l_j : t_j \leq a \lor b < t_j\}) \right|$$

3. Around each $\tilde{t}_i$ use observations from ca. 15 min. towards night to ca. 75 min. towards day

4. Gives two observation intervals per day around the mean observation time in each $\tilde{t}_i$. 
Model - 1/3: Movement model

- A state space model with underlying random walk (like in kftrack), and observation equation that includes all the astronomical stuff.

- Underlying movement model:

\[
\begin{pmatrix}
    x_{\bar{t}_i} \\
    y_{\bar{t}_i}
\end{pmatrix} =
\begin{pmatrix}
    x_{\bar{t}_{i-1}} \\
    y_{\bar{t}_{i-1}}
\end{pmatrix} +
\begin{pmatrix}
    u \\
    v
\end{pmatrix} \Delta \bar{t}_i + \eta_i
\]

where \( \eta_i \sim \mathcal{N}(0, 2D \Delta \bar{t}_i I_{2 \times 2}) \)

- \((x, y)\) is the position measured in nautical miles from a fix-point.

- Notation: \( \alpha_i = (x_{\bar{t}_i}, y_{\bar{t}_i})' \)

- The times \( \bar{t}_1, \bar{t}_2, \ldots, \bar{t}_{2N} \) semi-daily, and each corresponds approximately to a sunrise or a sunset.
Model - 2/3: Predicting light at given position

- The model predicts two positions every day. One near sunrise and one near sunset.
- Light observations at times $\{\tau_1^{(i)}, \ldots, \tau_{m_i}^{(i)}\}$ near each solar event must be predicted.
- In order to do so we need:
  a) Function $z$ to translate position coordinates into longitude and latitude
  b) Function $\theta$ to calculate the altitude angle from position $(\text{lon}_i, \text{lat}_i)$ and times $\tau^{(i)}$.
  c) Function $\phi$ connecting the altitude angle to light level (increasing spline estimated).
  d) A depth correction: $\gamma(d) = -k_1 d - k_2 (d - 50) 1_{\{d>50m\}}$

- It all connects as:
  \[
  \hat{l}^{(i)} = \Lambda(\alpha_i, \tau^{(i)}, d^{(i)}) \overset{\text{def}}{=} \phi(\theta(z(\alpha_i), \tau^{(i)})) + \gamma(d^{(i)})
  \]
Model - 3/3: Correlation structure

- Clearly not reasonable to assume independent light measurements
- The following covariance structure is assumed

\[ l^{(i)} = \hat{l}^{(i)} + \varepsilon^{(i)} \quad \text{where} \quad \varepsilon^{(i)} \sim \mathcal{N}(0, \Sigma^{(i)}) \]

\[ \Sigma_{j,k}^{(i)} = \begin{cases} 
\sigma_1^2 + \sigma_2^2 + \sigma_3^2 & \text{if } j = k \\
\sigma_1^2 + \sigma_2^2 \exp(-|\tau_{j}^{(i)} - \tau_{k}^{(i)}|/\rho) & \text{if } j \neq k 
\end{cases} \]

- Here \( \sigma_1, \sigma_2, \sigma_3, \) and \( \rho \) are model parameters.
- Observations from same solar event are correlated
- Close observations are more correlated than observations far apart
Summary of model

- Underlying random walk

\[ \alpha_i = \alpha_{i-1} + \begin{pmatrix} u \\ v \end{pmatrix} \Delta \bar{t}_i + \eta_i \]

- Observation equation

\[ l^{(i)} = \Lambda(\alpha_i, \tau^{(i)}, d^{(i)}) + \varepsilon^{(i)} \]

- Here \( \eta_i \) and \( \varepsilon^{(i)} \) are independent and normally distributed with mean zero and relevant covariance matrices.

- The model parameters are

\[ \vartheta = (u, v, D, \bar{\varphi}_1, \ldots, \bar{\varphi}_{15}, \sigma_1^2, \sigma_2^2, \sigma_3^2, \rho) \]

- Likelihood computed by unscented Kalman filter
  - First position is known without error (release position)
  - Predict next via model; calculate difference \( w \) and \( \text{var}(w) \); update position; ...

- All parameters are maximum likelihood estimated
Simulations

- Data is simulated from the model to verify model and implementation
- Real relationship between altitude and angle used
- Real covariance parameters ($\sigma_1^2, \sigma_2^2, \sigma_3^2, \rho$) used
- Different scenarios tested:
  - Are model parameters identifiable from data?
  - Can slow/fast track be reconstructed?
  - How sensitive to sampling rate?
  - ...
<table>
<thead>
<tr>
<th>Parameter name</th>
<th>True value</th>
<th>Mean of estimates</th>
<th>Std. dev. of estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>0.000</td>
<td>-0.138</td>
<td>1.366</td>
</tr>
<tr>
<td>$v$</td>
<td>0.000</td>
<td>0.123</td>
<td>1.310</td>
</tr>
<tr>
<td>$D$</td>
<td>300.000</td>
<td>295.021</td>
<td>34.916</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.032</td>
<td>0.034</td>
<td>0.006</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>83.000</td>
<td>81.829</td>
<td>5.041</td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>9.900</td>
<td>10.424</td>
<td>1.281</td>
</tr>
<tr>
<td>$\sigma_3^2$</td>
<td>2.600</td>
<td>2.606</td>
<td>0.075</td>
</tr>
<tr>
<td>$\varphi_{-5}$</td>
<td>98.843</td>
<td>98.793</td>
<td>0.550</td>
</tr>
<tr>
<td>$\varphi_0$</td>
<td>132.255</td>
<td>132.242</td>
<td>0.407</td>
</tr>
<tr>
<td>$\varphi_5$</td>
<td>145.571</td>
<td>145.487</td>
<td>0.353</td>
</tr>
</tbody>
</table>

Table 1. True parameter values, mean, and standard deviation of estimates from 100 simulated tracks.

- All model parameters are identifiable
- The variance structure is trading a bit of the overall correlation $\sigma_1^2$ for serial correlation $\sigma_2^2$ and $\rho$, but this is sensible as the model assumes no movement in the 90-minute observation window.
Drifter

- PAT 2
Drifter sd

- PAT 2

- Notice the clear pattern in latitude standard deviation
- Highest standard deviation not exactly at equinox
- No assumptions about seasonal pattern in model
Real Fish

- Five tracks will be presented here.
  - Three track from Mike Musyl, PAT 2
  - Two Mako tracks from David Holts and Suzanne Kohin at SWFSC, PAT 3

- The added difficulties are:
  - Depth correction applied within tag
  - Relying on an algorithm within the tag to select points
    (12 per solar event for PAT 2, and 9 for PAT 3).
Marlin

Estimated track of tag20546
Marlin

Longitude

Days at liberty

[Graph showing estimated track of tag21158]

Latitude

Date

Days at liberty

Longitude

Latitude

SST

Estimated track of tag21158
Marlin

Days at liberty

Longitude

Latitude

Days at liberty

Estimated track of tag21118

Date

Latitude

Longitude

SST

Days at liberty

Longitude

Latitude

Days at liberty

Longitude

Latitude

Estimated track of tag21118
Mako

Days at liberty

Latitude

Longitude

14 64 114 164

Days at liberty

●

●●

●● ●●

●●

●● ●● ●● ●● ● ●● ● ● ●● ●● ●● ●● ●

●

● ●● ●● ● ●● ●● ● ● ● ● ● ●

●

●

● ●● ●● ●

●

●

● ●● ●● ●

●

●

●

14Jul2003

2Sep2003

22Oct2003

11Dec2003
Mako

Days at liberty

Longitude

Latitude

Date

14 Jul 2003
23 Oct 2003
21 Dec 2003
30 Jan 2004
20 Mar 2004

230 235 240 245

20 25 30 35 40 45

10 15 20 25 30 35 40

11 12 16 21 26

130 125 120 115 110 105 100

50 40 30 20 10 0

112 162 212 262

14 Jul 2003
23 Oct 2003
21 Dec 2003
30 Jan 2004
20 Mar 2004

230 235 240 245

20 25 30 35 40 45

10 15 20 25 30 35 40

11 12 16 21 26

130 125 120 115 110 105 100

50 40 30 20 10 0
Conclusions and future developments

- The model is transparent, statistically sound, and without ad-hoc solutions
- Uncertainties are propagated all the way through the model (from the light observations to the uncertainties on the reconstructed track)
- The fact that light measurements are correlated (not independent) is accounted for.
- The geolocations are a lot more accurate compared to other purely light based methods
- Other methods assume that the fish is not moving in between sunrise and sunset. This model only ignores movement in a 90 minute window around each solar event
- Two positions are estimated each day
- SST can later be added to the model similar to the way it is done in kfsst
- R-package is under development, so this will be as simple to apply as kftrack and kfsst.

Thank you for listening!