A mosaic of models for light-based geolocation: How to choose, what to be careful about, and future directions

Anders Nielsen & John Sibert

anders.nielsen@hawaii.edu
Light based geolocation

- We got:
  - Light, depth, and temperature
  - Measured for instance every minute
  - From archival tags the entire record can be retrieved
  - From satellite transmitting tags only a summary

- We want:
  - A track of geographic positions (geolocations)
  - Some idea about the uncertainties
  - Perhaps some quantitative movement parameters

- Problems:
  - Indirect measurements: Light $\rightarrow$ solar altitude $\rightarrow$ geolocation
  - High and correlated uncertainties from changing weather and incomplete depth corrections
This talk

- Will talk about:
  - Raw geolocations
  - Kftrack
  - Kfsst/ukfsst
  - Trackit (with and without SST)

- Will not talk about:
  - Satellite methods
  - Tidal location models
  - Sunrise/sunset times models
  - SST matching algorithms
  - EASy-FishTracker
  - ...
Similarities of kftrack, kfsst, ukfsst, and trackit

- Underlying movement model
  - Assume the same random walk model
  - Includes drift and diffusion

- At any given position the observation model
  - Predicts the observation
  - Describes the observation error

- All model parameters are maximum likelihood estimated

- Any point on the most probable track is a weighted average of:
  - What is learned from the current observation
  - What is learned from the entire track
The differences are

- What they take as data
  - Raw geolocations (lon,lat) used by kftrack
  - Raw geolocations and SST (lon,lat,SST) used by kfsst and ukfsst
  - Light readings and SST used by trackit

- Technical details in handling non-linearities
  - Extended Kalman filter used by kftrack and kfsst
  - Unscented Kalman filter used by ukfsst and trackit

<table>
<thead>
<tr>
<th></th>
<th>(lon,lat)</th>
<th>SST</th>
<th>Light</th>
<th>EKF</th>
<th>UKF</th>
</tr>
</thead>
<tbody>
<tr>
<td>kftrack</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>kfsst</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>ukfsst</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trackit</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If possible, running more than one can often be very instructive ...
• Algorithms (partly) proprietary, but in essence:
  – Associate a certain light level crossing with a solar angle (problematic)
  – Calculate position from two of those

• These raw geolocations are often imprecise and biased

• Especially latitude around equinox
library(kftrack)
track<-read.table('datafile.dat', header=TRUE)
fit<-kftrack(track)
plot(fit, map=TRUE, ci=TRUE)
library(kfsst) or library(ukfsst)

track <- read.table('datafile.dat', header=TRUE)
sst.path <- get.sst.from.server(track)
sst.file <- write.sst.field(sst.path) delete line in ukfsst
fit <- kfsst(track)
plot(fit, map=TRUE, ci=TRUE)
library(trackit)
track<-read.table('datafile.dat', header=TRUE)
sst<-read.table('sstdatafile.dat', header=TRUE)
sst.path<-get.sst.from.server(track,150,250,0,40)
prep.track<-prepit(track,
            sst=sst,
            fix.first=c(198.55,22.85,2002,9,10,0,0,0),
            fix.last=c(200.13,21.95,2003,5,21,0,0,0))
fit<-trackit(prep.track)
plot(fit)
fitmap(fit)
Be careful about trusting raw geolocations
Be careful about trusting raw geolocations - 2
Be careful about convergence

# R-KFtrack fit
# Number of observations: 76
# Negative log likelihood: 322.056
# The convergence criteria was met

- Convergence should be obtained

- Ways to help the optimizer if a track is problematic
  - Simplify model (especially for short tracks)
    (e.g. `fit<-kftrack(track, bx.a=FALSE, by.a=FALSE)`)  
  - Supply better initial values
    (e.g. `fit<-kftrack(track, D.i=500)`)  
  - Remove extreme outliers
    (e.g. `track<-track[abs(track$lat)<90,]`)  
  - A combination
  - Also check data
Be careful about selecting satellite SST data

- In open ocean coarse resolution is fine
- Near the coast a fine resolution is needed
- In areas with frequent cloud cover consider the blended source
- See the options in the documentation
  - `?get.sst.from.server`
  - `?get.blended.sst`

Remember to report back

- Like to hear when it is working
- Need to hear when it is not
Future directions

- Grid based methods interesting
  - allows other distributions than Gaussian
  - allows strong non-linearities (land areas)
  - very computational demanding

- Numerous tracks in one model
  - The right thing to do if some (or all) parameters are common
  - More confidence in estimated parameters
  - Possible to allow more flexible movement patterns

- Conventional tracks and archival tags in one model
  - First step in using all tagging data in fish stock assessment models

- All packages can be downloaded from:
  
  http://www.soest.hawaii.edu/tag-data/software/

Thank you for listening!
Combining individual and population based models

An appealing idea

- The parameters are the same (drift and diffusion)
- All tagged fish from the same population should be equal representatives, no matter what type of tag
- Might get better individual tracks when parameter estimates get better
- How much more is learned from an (expensive) archival tag

Simulation study

- 100 data sets are simulated each with 5100 simulated individuals
- 5000 with conventional tags 100 with archival tags (randomly assigned)
- Realistic effort pattern, fishing mortality, and natural mortality are applied

Parameter estimation

- A–D–R model is used for the conventional tags
- The Kalman filter likelihood was extended to include survival and recapture probabilities, and the individuals that were not recaptured
Results

Histogram of 100 estimations from 100 archival tags vs 5000 conventional tags.
<table>
<thead>
<tr>
<th>Parameter name</th>
<th>True value</th>
<th>Conventional tags</th>
<th>Archival tags</th>
<th>All tags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>bias</td>
<td>std. dev.</td>
<td>bias</td>
</tr>
<tr>
<td>u</td>
<td>0.02</td>
<td>-0.00143</td>
<td>0.00259</td>
<td>-0.00364</td>
</tr>
<tr>
<td>v</td>
<td>0.00</td>
<td>-0.00232</td>
<td>0.00302</td>
<td>0.00059</td>
</tr>
<tr>
<td>D</td>
<td>0.50</td>
<td>-0.02078</td>
<td>0.02403</td>
<td>-0.00244</td>
</tr>
<tr>
<td>Q</td>
<td>0.50</td>
<td>-0.00228</td>
<td>0.01221</td>
<td>0.02835</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>100.00</td>
<td>-0.13608</td>
<td>1.52063</td>
<td>0.09419</td>
</tr>
</tbody>
</table>

- Best results from combined model
- In this setting we get almost the same amount of information about drift $(u, v)'$ from one conventional tag as from one archival tag. This will likely change in a more complex setting.
- Archival tags provide more information per tag about diffusion $D$ than conventional tags