Incorporation of environmental drivers in the prediction of pelagic stocks recruitment in the Baltic Sea using random forest algorithms

Szymon Smoliński
ssmolinski@mir.gdynia.pl
National Marine Fisheries Research Institute
Department of Fisheries Resources
Fish recruitment

Recruitment of Central Baltic herring (a) and Baltic sprat (b).
Fish recruitment

Which factors have impact on pelagic fish recruitment?

- Spawning stock biomass
- Spawners characteristic
- Interactions with other species
- Environmental factors (hydrological conditions, climate)

Köster et al., 2003
Cardinale et al., 2009
Margonski et al., 2010
Bartolino et al., 2014
Gröger et al., 2014

Ricker model of Baltic sprat recruitment.
Environmental data in recruitment modeling

Examples of nonlinear relationships and interactions between variables
Random forests


---

Smoliński & Radtke, 2017

*ICES Journal of Marine Science*
Boruta algorithm

• an extension of the random forest method which utilizes the importance measure generated by the original algorithm
• compares, in the iterative fashion, the importance of original attributes with importance of their randomized copies (shadow attributes)
• used for feature selection in the systems with unclear mechanisms of investigated processes

Kursa & Rudnicki, 2010
Variables in the temporal scale

Examples of hydroclimatic time series

Bailey & van de Pol, 2016

van de Pol et al., 2016
Aims of the work

• Investigation of hydroclimatic factors, that may have influence on recruitment of:
  • herring (*Clupea harengus*) in ICES Subdivisions 25–29, 32, excluding Gulf of Riga
  • sprat (*Sprattus sprattus*) in ICES Subdivisions 22–32

• Hypotheses:
  • Hydroclimatic conditions have impact on the recruitment success
  • Signals from different time windows may have different effects

• Incorporation of random forests in the „sliding window“ method

Map of the Baltic Sea with indicated ICES subdivisions.
List of variables used in the modelling of Baltic herring and sprat recruitment.

<table>
<thead>
<tr>
<th>Variables abbreviation</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>her/sprR</td>
<td>Herring or sprat recruitment at age 1</td>
<td>(ICES, 2016)</td>
</tr>
<tr>
<td>her/sprSSB</td>
<td>Herring or sprat spawning stock biomass</td>
<td>(ICES, 2016)</td>
</tr>
<tr>
<td>her/sprWAAx</td>
<td>Herring or sprat weight at age x</td>
<td>(ICES, 2016)</td>
</tr>
<tr>
<td>codTSB</td>
<td>Total stock biomass of cod in subdivisions 25-32</td>
<td>(ICES, 2013)</td>
</tr>
<tr>
<td>Hydroclimatic data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SST</td>
<td>Mean sea surface temperature</td>
<td>Monthly resolution</td>
</tr>
<tr>
<td>BSI</td>
<td>Mean Baltic Sea Index</td>
<td>(Lehmann et al., 2002)</td>
</tr>
</tbody>
</table>
Data analysis

Separate runs for each stock and hydroclimatic variable (SST and BSI)

For each time window (range of 36 months):
- Calculate the value of investigated variable for each observation using aggregate statistic (e.g. mean)
- Add new variable to the data matrix
- Select variables using Boruta algorithm (test of variable relevance)
- Cross-validate random forest grown on selected variables:
  - Randomly split the data into the 5 subsets
  - For each subset: train model on 4 subsets and test model on remaining subset

Variable relevance
- RMSE (root mean square error)

Identification of optimal signal

Further analysis
Results of random forest “sliding window” analysis for the environmental effects on herring recruitment. Outcome of first and second step of optimal signals identification for BSI (a, b) and SST (c, d) were shown on the plots.
Results of random forest “sliding window” analysis for the environmental effects on sprat recruitment. Outcome of first and second step of optimal signals identification for BSI (a, b) and SST (c, d) were shown on the plots.
Results

Relevance of variables in the random forest model of herring (a) and sprat (b) recruitment according to the results of Boruta.
Results

Partial dependence plots for hydroclimatic variables for random forest predictions of herring (a, b) and sprat (c, d) recruitment. Effects of different BSI signals (according to results of 1st and 2nd step of “sliding window” analysis) were presented.

<table>
<thead>
<tr>
<th></th>
<th>Window open</th>
<th>Window close</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Herring</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSI</td>
<td>24</td>
<td>14</td>
<td>-</td>
</tr>
<tr>
<td>BSI2</td>
<td>16</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>SST</td>
<td>9</td>
<td>5</td>
<td>+</td>
</tr>
<tr>
<td><strong>Sprat</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSI</td>
<td>11</td>
<td>9</td>
<td>+</td>
</tr>
<tr>
<td>BSI2</td>
<td>19</td>
<td>19</td>
<td>-</td>
</tr>
<tr>
<td>SST</td>
<td>4</td>
<td>3</td>
<td>+</td>
</tr>
</tbody>
</table>
Results

Plots of two first principal component scores derived by PCA based on proximity matrix of recruitment years of herring (a) and sprat (b).
Results

Prediction accuracy ($R^2\pm sd$) of models derived from cross-validation.

<table>
<thead>
<tr>
<th>Variables included</th>
<th>Herring</th>
<th>Sprat</th>
</tr>
</thead>
<tbody>
<tr>
<td>All relevant variables</td>
<td>0.59±0.211</td>
<td>0.50±0.229</td>
</tr>
<tr>
<td>Biological variables</td>
<td>0.44±0.237</td>
<td>0.17±0.191</td>
</tr>
<tr>
<td>SSB</td>
<td>0.11±0.127</td>
<td>0.13±0.159</td>
</tr>
</tbody>
</table>

Plots of observed (points) and predicted (open triangles) recruitment of herring (a) and sprat (b).
Conclusions

• Incorporation of environmental data improves accuracy of Baltic pelagic fish recruitment prediction

• SST and BSI have significant impact on recruitment processes

• Observed relationships of pelagic fish recruitment and hydroclimatic conditions were nonlinear

• The same environmental variable from different time windows may have counteracting effects (Kruuk et al. 2015, GCB)

• Data mining methods and random forests may be used to obtained new knowledge from ecological data
References