Machine Learning Methods for Spatio-Temporal Modeling of Environmental Data

Federico Montesino Pouzols  
<federico.pouzols@aalto.fi>

Environmental and Industrial Machine Learning (EIML) Group

Adaptive Informatics Center of Excellence  
Department of Information and Computer Science, Aalto University
Outline

Topics and Methods

Spatio-temporal models: Maps

Predictive modeling of time series

Papers, results and discussion
Environmental part of EIML: Scope and Topics

Focus on topics relevant to the Baltic Region

Current application areas:

▶ Long-term management of lakes.
▶ Wastewater treatment.
▶ Models for Marine Biology.
▶ Hydrology: Runoff and impact on the Baltic Sea.

From the machine learning viewpoint, this involves:

▶ Time series prediction (evolving dynamics).
▶ Spatial modeling.
▶ Irregular sampling, missing chunks.
▶ Fusion of multivariate data sources.
▶ Many sources of uncertainty, outliers, noise, etc.
EIML: People

- Ph.D.: 2010–2014:
  - Emil Eirola (FICS):
    “Causality and ensemble techniques for environmental modeling of the Baltic Sea.”
  - Dušan Sovilj (HECSE)
    “Machine learning methods for environmental modeling of the Baltic Sea.”
- M.Sc.: Ajay Ramaseshan, spatio-temporal models.
- Summer student: Tommi Pesu, Computational Geometry methods for spatial modeling.

More information: http://research.ics.tkk.fi/eiml
Machine learning methods and tools:

Data sources
- HELCOM
- Nest Inst.
- MARBEF
- etc.

Missing values

Spatial / temporal aggregation

Trends
Predictive models
Causal chains
Salinity maps: 7 PSU border

Average salinity 0–20 m, Jun-Sep

psal, June-September, 1977

psal, June-September, 2001
Salinity maps: 7 PSU border

Average salinity 0–20 m, Jun-Sep

psal, June-September, 1977

psal, June-September, 2001

Salinity < 7 area

% of total Baltic Sea area

Year

Example: Horohalinicum 5–8 PSU

→ maps and time series for different species...
**Example: Laomedea Lovéni: 6.3–7.4 PSU range**

Average salinity 0–20 m, Jun-Sep

<table>
<thead>
<tr>
<th>Year</th>
<th>Percentage of Total Baltic Sea Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td></td>
</tr>
<tr>
<td>1964</td>
<td></td>
</tr>
<tr>
<td>1968</td>
<td></td>
</tr>
<tr>
<td>1972</td>
<td></td>
</tr>
<tr>
<td>1976</td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
</tr>
</tbody>
</table>

**Equivalent latitudes (for a straight horizontal line)**

<table>
<thead>
<tr>
<th>Latitude °</th>
<th>psal, June-September, 1980</th>
<th>psal, June-September, 2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>64.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Laomedea loveni: northern (blue) and southern (black) latitudes**
Example: *Laomedea Lovéni*:

6.3–7.4 PSU range

Average salinity 0–20 m, Jun-Sep

psal, June-September, 1980

psal, June-September, 2003

Laomedea loveni area

% of total Baltic Sea area

Year

Example: Laomedea Lovéni: 6.3–7.4 PSU range

Average salinity 0–20 m, Jun-Sep
Predictive Modeling of Time Series

Time series prediction and missing value imputation.
Problem: Fill in gaps in zooplankton time series.
Visit of Dušan to IOW Warnemünde, June–July 2010.

Acartia adults, quarterly series, prediction using BSI

(subtractive clustering and neuro-fuzzy system)
Zooplankton Gap Filling: Results

Spring values using OP-ELM with corrected Akaike inf. criterion:

Prediction using BSI

Prediction using AO+BSI
Zooplankton Gap Filling: Thoughts

- Prediction heavily influenced by time period (measurements until 1999) → climate shifts important but black-box models cannot capture such difficult dynamics
- Single climate index (BSI) can provide solid fit/training and prediction
- Additional climate index (AO) can lower training error, but does not provide better generalization → the need to properly validate models
- To accurately predict future values, a lot of \textit{a priori} information must be included (factors revolving around zooplankton)
Papers and other results (1/2)

Some recent posters/papers:

▷ F. M. Pouzols “Predictive modeling of global and regional runoff to the Baltic Sea using machine learning methods,” Hydroinformatics: computational intelligence and systems analysis, EGU meeting 2011, Vienna, Apr.

Papers and other results (2/2)

Forthcoming journal papers:

▶ Alternative Food Webs in the Baltic Sea - Biodiversity and Salinity Reconsidered together with Climate Change or Moving Horohalinicum will Crash the Biodiversity.

▶ A system for analyzing maps of water bodies (and its application to the Baltic Sea), Environmental Modeling & Software.

▶ Gap filling for zooplankton time series?

Software tools

▶ WBMG tool to generate maps for restrospective analysis.
Thanks and Discussion

- Do predictions based on climate change scenarios agree with data-driven machine learning predictions?
- Network of factors (climate, physical and ecological data).
- Emphasis on regional models. Local indices/time series?