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:

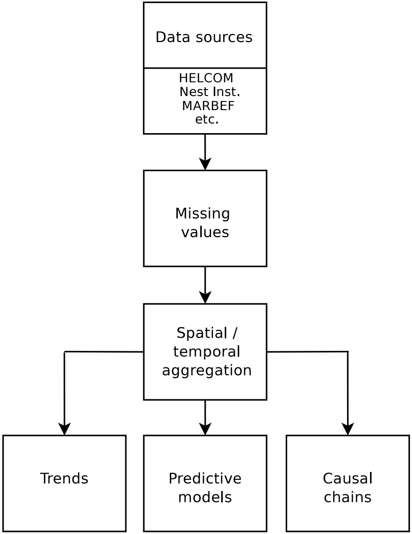
- Time series prediction (evolving dynamics).
- Spatial modeling.
- Irregular sampling, missing chunks.
- Fusion of multivariate data sources.
- Many sources of uncertainty, outliers, noise, etc.

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.

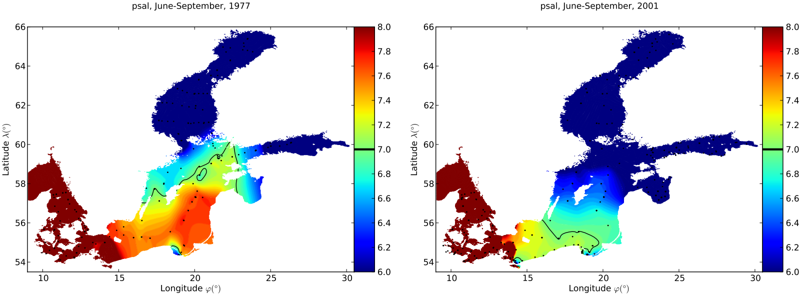
More information: `http://research.ics.tkk.fi/eiml`

Machine learning methods and tools:



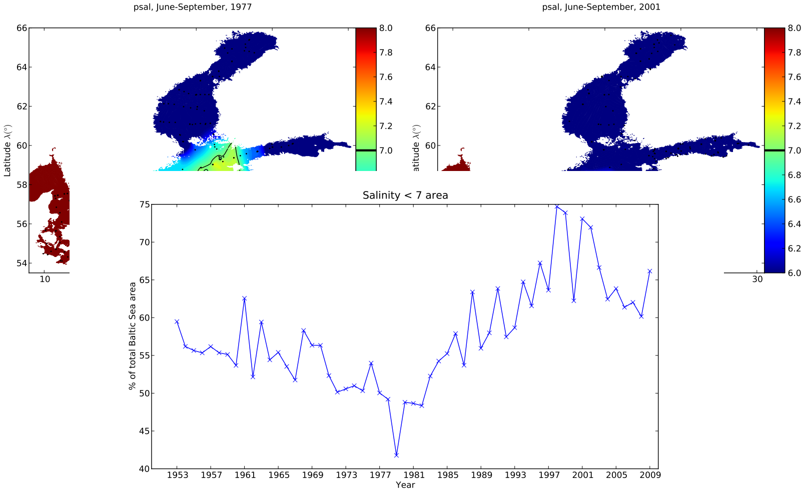
Salinity maps: 7 PSU border

Average salinity 0–20 m, Jun-Sep

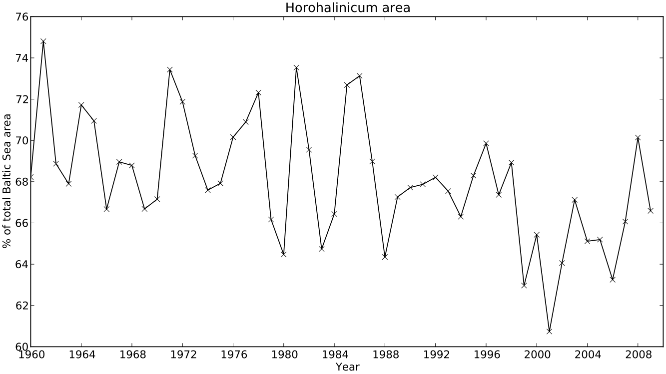


Salinity maps: 7 PSU border

Average salinity 0–20 m, Jun-Sep



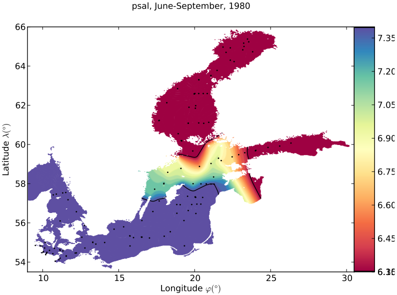
Example: Horohalinicum 5–8 PSU

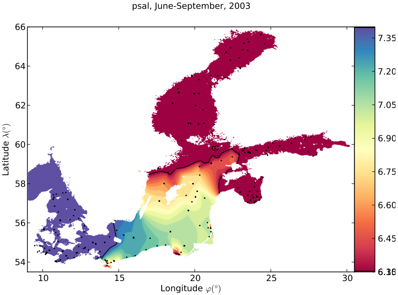


→ maps and time series for different species...

Example: Laomedea Lovéni: 6.3–7.4 PSU range

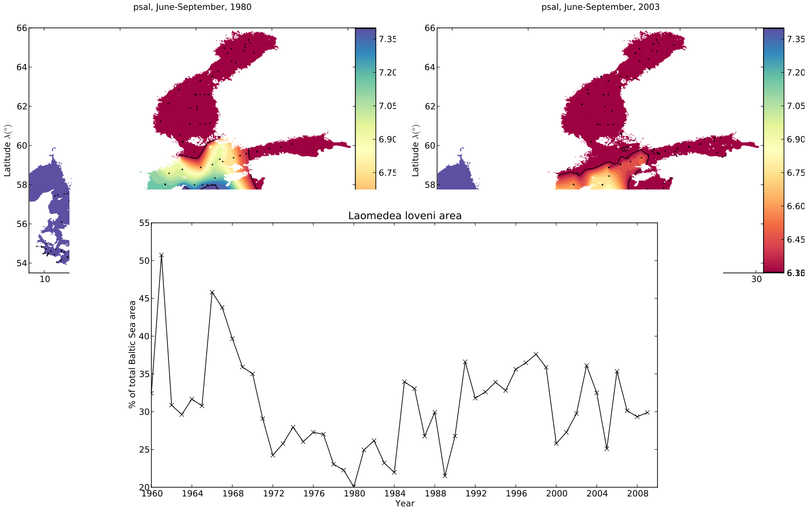
Average salinity 0–20 m, Jun-Sep





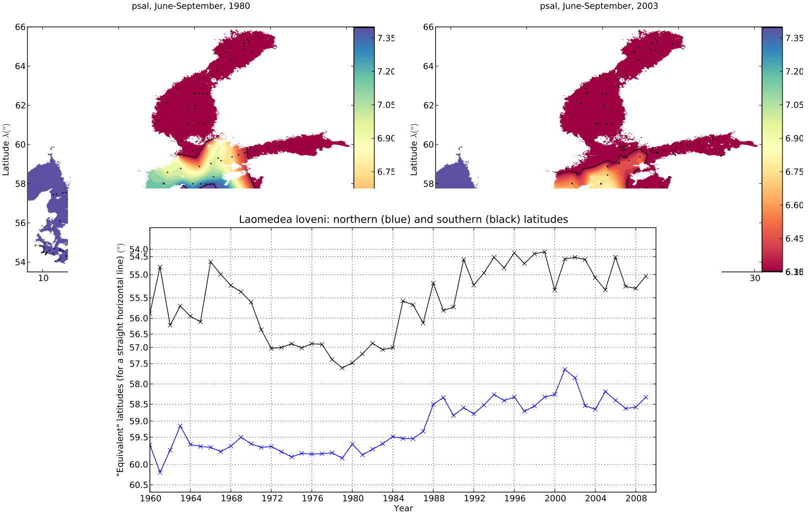
Example: Laomedea Lovéni: 6.3–7.4 PSU range

Average salinity 0–20 m, Jun-Sep



Example: Laomedea Lovéni: 6.3–7.4 PSU range

Average salinity 0–20 m, Jun-Sep



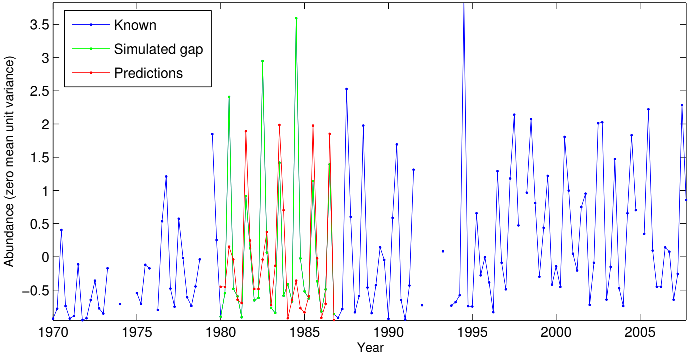
Predictive Modeling of Time Series

Time series prediction and missing value imputation.

Problem: Fill in gaps in zooplankton time series.

Visit of Dušan to IOW Warnemünde, June–July 2010.

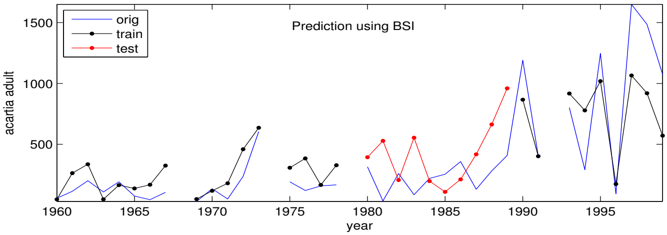
Acartia adults, quarterly series, prediction using BSI

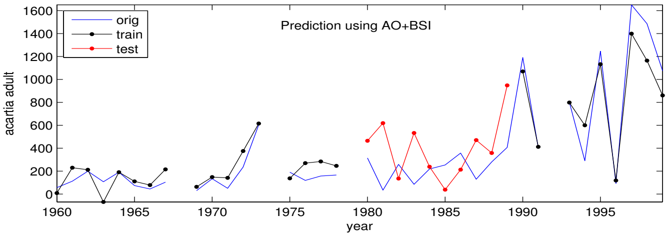


(subtractive clustering and neuro-fuzzy system)

Zooplankton Gap Filling: Results

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





Zooplankton Gap Filling: Thoughts

- Prediction heavily influenced by time period (measurements until 1999) $\rightarrow$ 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 $\rightarrow$ the need to properly validate models
- To accurately predict future values, a lot of *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.
- F. M. Pouzols and A. Lendasse, "Adaptive Kernel Smoothing Regression for Spatio-Temporal Environmental Datasets," in *ESANN 2011*, Bruges, 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?