Turun yliopisto
University of Turku

UNIVERSITY OF TURKU | Biodiversity

UNIVERSITY OF TURKU | Biodiversity
ARCHIPELAGO RESEARCH INSTITUTE
TURKU ARCHIPELAGO - SEILI

Seili-index for the prediction of chlorophyll in the Turku archipelago

To build a forecasting tool for the state of eutrophication in the Archipelago Sea, we fitted a generalized additive mixed model to the marine biological data, collected over the years 2011-2017 by an automated monitoring buoy at the Seili station. The resulting "Seili-index" can be used to predict chlorophyll content of the sea water a number of days ahead using the temperature forecast as a covariate. An array of test predictions on the 2017 data set shows that the index is adept at predicting the amount of chlorophyll especially at the greater depths.

DATA SPECIFICATION AND PREPROCESSING

The automated monitoring buoy measures temperature (temp; °C), salinity (PSU), turbidity (NTU), chlorofyll (Chl-A; ug/L), blue-green algae (BGA; cells/mL), and oxygen (mg/L) of the water column (0-42 m) at depth stages of one meter. The data, measured 4 times a day, is sent to an external server twice a day. The station is in operation from early spring (as soon as the ice allows) and removed in early winter before ice formation.

For this preliminary example of the Seili-index we will focus on the two variables **Chl-A** and **temperature**. The water temperature (temp) is a reasonable predictor for e.g. Chl-A content as its future values can be obtained reliably from various weather forecasts.

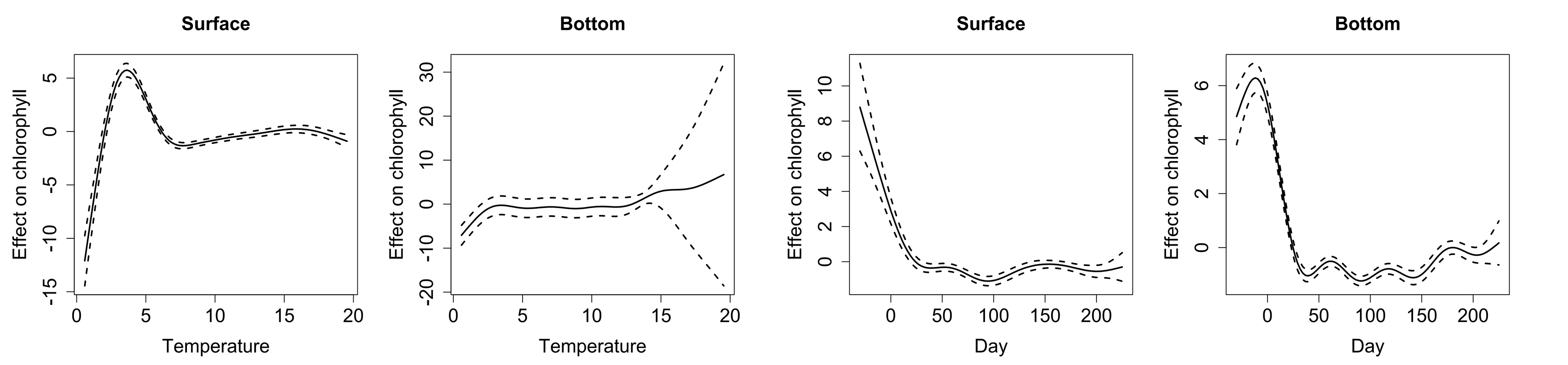
To avoid having to model the water depth as a continuous variable we discretized it into two classes, *surface* (depth < 20m) and *bottom* (depth ≥ 20m) and worked with the daily means of the observed values of Chl-A and temp over the two depth categories. Finally, as the biological year does not necessarily coincide with the calendar year we needed to align the data from the different years based on some landmark in the yearly observations of the two variables. A useful choice for this is the blooming of the Chl-A during the spring months and for each year we located the peak value of Chl-A and aligned the data so that the peak always occurs on day 0. For the year 2017 the blooming had occurred already before the buoy measurements started, and we aligned its data using the blooming of 2016 as a landmark.

MODEL FORMULATION

Our model of choice is a Generalized Additive Mixed Model (GAMM) (Wood, 2017) in which the conditional expected value of a normally distributed response Y is modelled as a linear combination of random effects and smooth functions of fixed effects:

$$E(Y|\boldsymbol{X}) = \beta_0 + f_1(X_1) + \cdots + f_p(X_p) + \beta_1 Z_1 + \cdots + \beta_q Z_q,$$

where $X_1, \ldots, X_p$ are the fixed effects, $f_1, \ldots, f_p$ are unknown smooth functions estimated using splines and $Z_1, \ldots, Z_p$ are the random effects. We used the function *gamm* in the R-package *mgcv* (Wood, 2017) to fit the model with depth as a categorical variable, depth-wise temperature and depth-wise day of the year as smoothed fixed effects and year as a random effect. All smoothed fixed effects were highlystatistically significant in predicting Chl-A and of the fixed effects only depth did not carry any predictive power. The model has the adjusted R^2-value of 0.765, indicating a good fit.



Figs. 3. The estimated smooth functions and their 95% pointwise confidence intervals for temp and day of the year on the two categories of depth. To interpret the plots, we note that the model is *additive*. That is, e.g. if the surface water temperature drops from 10 degrees to 3 degrees we can expect the chlorophyll content to increase approximately 5 units, assuming all other covariates are held fixed. Similarly, when we go 25 days forward from the day 100 of a year (with respect to the new blooming-centered time scale) we can expect the Chl-A content to increase roughly 0.5 unit, assuming all other covariates are held fixed.

CONCLUSION

We propose the preliminary form of the Seili-index, used to predict the Chl-A content of the sea water a number of days ahead in time. To make the predictions the index requires a forecast of the sea water temperature for the prediction period which can be obtained e.g. from online weather services. Our test predictions revealed that the index was able to predict the Chl-A especially well in deeper depths (depth ≥ 20m) but also for the autumn part of the surface depths (depth < 20m).

Future work includes the implementation of the index as an online service, trying to improve the prediction accuracy by using more sophisticated modelling techniques and obtaining predictions also for other variables measured by the Seili buoy, for example, **the BGA content (cells/mL)**.

Fig.1. Location of the automated monitoring buoy. The monitoring is carried out by the Turku University of Applied Sciences in collaboration with University of Turku, Finnish Meteorological Institute and other FINMARÍ partners.

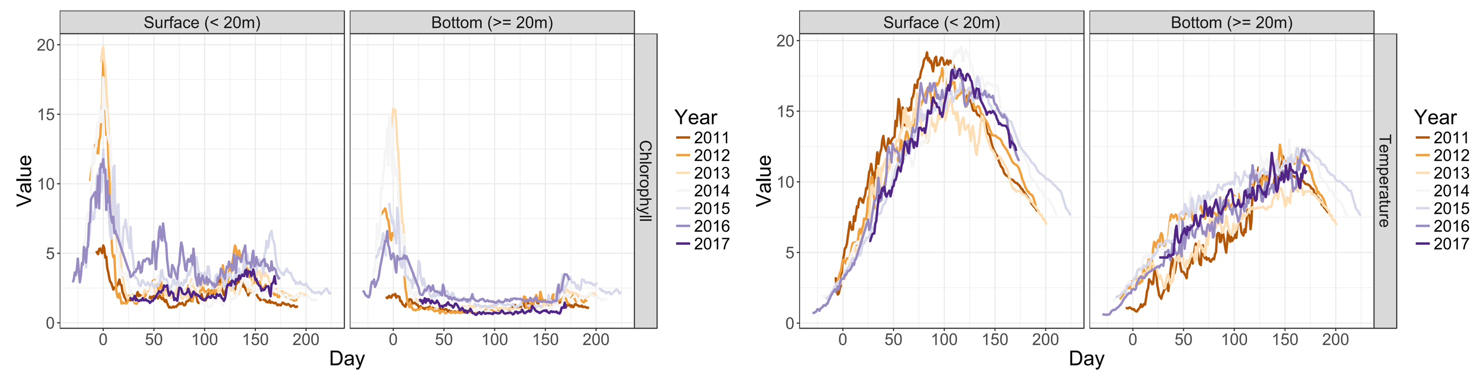


Fig. 2. The time courses of the aligned Chl-A and temp for the two categories of the depth. Both plots of Chl-A show the blooming peak during the day 0 and also a second peak is visible around day 150 (mid-autumn). There are clear differences in the Chl-A-curves of the seven years, the years 2012, 2013 and 2016 especially standing out.

PREDICTIONS

The data for the years 2011-2016 along with first 26 measurement days of the year 2017 are used to fit the GAMM model, where the latter set of data is needed to estimate the effect of the year 2017. The fitted model is then used to predict Chl-A for the next 10 days using the true temperature data of the days. In practice, weather forecasts would be used to obtain the values of temp for the prediction period. Next, the true data for the 10 days are added to the model and we do the previous procedure again, again using the obtained model to predict the next ten day interval based on the true temp values of the interval. This cycle is continued for a total of 12 times, until the end of the measurement period of the year 2017.

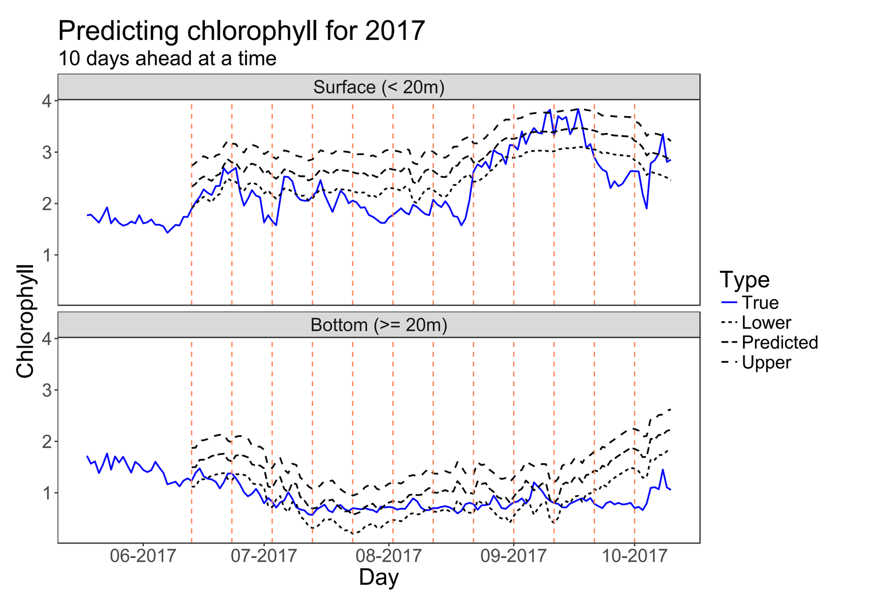


Fig. 4. The predictions along with pointwise 95% confidence intervals (the three dashed black lines) shown together with the true values (solid blue lines) for two depth categories. The red vertical lines mark the 10 day periods described previously. That is, the predictions in any given interval between two red lines are obtained by fitting a model with all data to the left of that interval and using the true temp values of the interval to obtain the predictions.

VISUALISATION

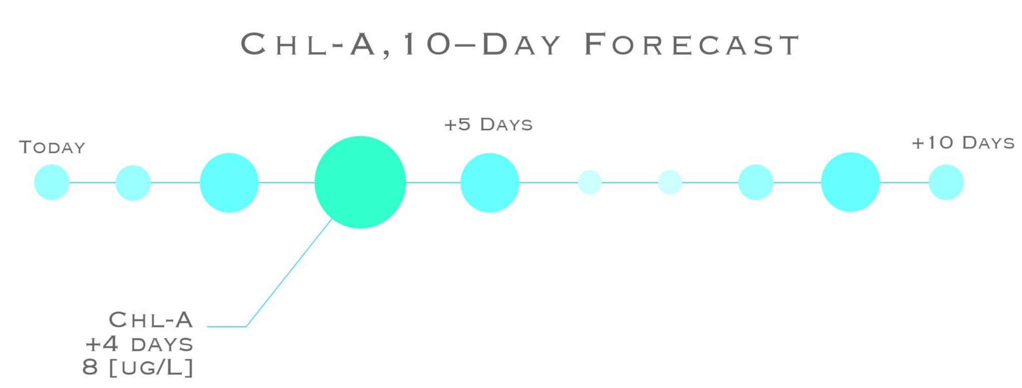


Fig. 5. An example of the visualisation of the 10-day forecast.
The Seili-index will be published in spring 2018

References: Wood, S.N. (2017) Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

In collaboration with:
TURKU AMK
TURKU UNIVERSITY OF APPLIED SCIENCES

Hänninen J[1], Löwendahl T[1], Mäkinen K[1], Nordhausen K[2] & Virta J[3]

[1]Archipelago Research Institute, Biodiversity Unit, University of Turku
[2]Institute of Statistics & Mathematical Methods in Economics, Vienna University of Technology
[3]Department of Mathematics and Systems Analysis, Aalto University

Contact information:
Jari Hänninen
Email: jarhan@utu.fi
Archipelago Research Institute
Biodiversity Unit ,University of Turku