Author(s): Hänninen, Jari; Mäkinen, Katja; Nordhausen, Klaus; Laaksonlaita, Jussi; Loisa, Olli; Virta, Joni

Title: The "Seili-index" for the Prediction of Chlorophyll-α Levels in the Archipelago Sea of the northern Baltic Sea, southwest Finland

Year: 2022

Version: Published version

Please cite the original version:

Hänninen, J., Mäkinen, K., Nordhausen, K., Laaksonlaita, J., Loisa, O., & Virta, J. (2022). The "Seili-index" for the Prediction of Chlorophyll-α Levels in the Archipelago Sea of the northern Baltic Sea, southwest Finland. Environmental Modeling and Assessment, 27(4), 571-584. https://doi.org/10.1007/s10666-022-09822-9

The "Seili-index" for the Prediction of Chlorophyll-*a* Levels in the Archipelago Sea of the northern Baltic Sea, southwest Finland

Jari Hänninen[1] · Katja Mäkinen[1] · Klaus Nordhausen[2,3] · Jussi Laaksonlaita[4] · Olli Loisa[4] · Joni Virta[5]

Abstract
To build a forecasting tool for the state of eutrophication in the Archipelago Sea, we fitted a *Generalized Additive Mixed Model* (GAMM) to marine environmental monitoring data, which were collected over the years 2011–2019 by an automated profiling buoy at the Seili ODAS-station. The resulting "Seili-index" can be used to predict the chlorophyll-α (chl-a) concentration in the seawater a number of days ahead by using the temperature forecast as a covariate. An array of test predictions with two separate models on the 2019 data set showed that the index is adept at predicting the amount of chl-a especially in the upper water layer. The visualization with 10 days of chl-a level predictions is presented online at https://saaristomeri.utu.fi/seili-index/. We also applied GAMMs to predict abrupt blooms of cyanobacteria on the basis of temperature and wind conditions and found the model to be feasible for short-term predictions. The use of automated monitoring data and the presented GAMM model in assessing the effects of natural resource management and pollution risks is discussed.

1 Introduction

In recent decades, perhaps the most important consequence of the global climate change in aquatic, terrestrial, and marine environments has been the alterations in seasonal cycles, resulting in, for example, habitat loss and fragmentation (e.g., [1, 2]). The changing environment has further altered species existence and life cycle processes that strongly promote the ongoing biodiversity decline (e.g., [3]). The relevance of biodiversity loss is a crucial question especially in coastal areas that are extremely valuable for the economy and, in general, for human life. Impoverishment of the aquatic habitats lessens also their recreational use and is mainly due to changes in land management and practices, which accelerate especially eutrophication processes caused by increased anthropogenic nutrient flux into the sea (e.g., [4, 5]). Hence, there is a common understanding that conservation measures, such as the establishment of protected areas and biosphere reserves, should be undertaken primarily in areas least affected by human activities [6].

The main environmental issue impacting the water quality of coastal areas during the last five decades, especially in the northern hemisphere, has been eutrophication (e.g., [6–9]). As in many other coastal areas, in the Finnish Archipelago Sea, the most important regional sources of nutrient surplus have been industrial and domestic wastewaters, runoffs from forestry, animal production, and land cultivation as well as salmon aquaculture at the sea area [10]. However, due to marine conservation measures, the relative importance of these nutrient contributions has generally declined during the last 15 years, and currently the main nutrient source is agricultural runoffs, which provides an average of 71% of nitrogen and 83% of phosphorus loadings [11]. The most

✉ Jari Hänninen
jari.hanninen@utu.fi

1 Archipelago Research Institute, Biodiversity Unit, University of Turku, 20 014 Turku, Finland

2 Institute of Statistics & Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria

3 Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland

4 Water and Environmental Engineering RG, Turku University of Applied Sciences, Turku, Finland

5 Department of Mathematics and Statistics, University of Turku, Turku, Finland

predominant symptom of the eutrophication is increased levels of microscopic phytoplankton or microalgae in the seawater, expressed as chlorophyll-α (chl-a) concentration, and resulting in, for instance, increased turbidity and decreased visibility in seawater and increasingly frequent cyanobacteria blooms during the summertime.

Since the nutrient sources for the Archipelago Sea eutrophication processes are defined mainly by their location being in the inner-zone agricultural nutrient runoff from the catchment area transported by rivers [12], in the middle-zone fish farming and agriculture runoff within the archipelago and in the outer zone, where remote nutrients are transported by sea currents [10], there exists an opportunity to model statistically local conditions in the marine habitats within the Archipelago Sea. In the present study, we aimed to check whether the most evident sign of eutrophication, i.e., an increased rate of microalgae primary production, could be modelled to build short-term predictions of its levels on the basis of seawater chl-a concentrations in the middle Archipelago Sea. Since temperature is the most important booster to control the rate of biological processes in seawater, such as the efficiency of the primary production (e.g., [13]), we selected it as an explanatory variable for modelling of chl-a levels predictions. As material for the models, we utilized data from an automated profiling buoy located in 2006 to the ODAS (*Oceanographic Data Acquisition System*) monitoring station of the University of Turku field station in the vicinity of the island of Seili. Our goal was to compose a visualized online service called the "Seili-index" based on advanced statistical modelling and for this service to be available online to anyone interested in water quality information and trends.

2 Materials and Methods

2.1 Study Area

The Archipelago Sea is a semi-enclosed archipelago at the southwest coast of Finland between the Baltic proper and the Bothnian Sea (59° 45′–60° 45′ N and 21° 00′–23° 00′ E) in the northern Baltic Sea (Fig. 1). Depending on the definition of an island, the area is estimated to contain up to 60,000 islands, of which, some 41,000 are named in regional charts [11]. In this respect, it is the biggest archipelago in the world with a complex and variable topography and has mainly wind-driven water mass movement patterns. The total area of this brackish water sea is 9436 km^2 with a water volume of 213 km^3 and a salinity content of 4 to 6 PSU [14]. The total catchment area of the Archipelago Sea is approximately 8900 km^2 (with a lake area of less than 2% and an arable land area of 28%) [10]. The average water depth is only 23 m as the deepest hollows reach 140 m. The wind-caused sea level variation is generally low with mostly ± 0.5 m variation compared to the theoretical mean level with the insignificant tidal fluctuation [14]. The sea is characterized by strong seasonality with the summer temperature of seawater reaching 20 °C and with a more than 90% probability of an annual 1–5-month ice cover during the winter [15].

The special topographical characteristic of the Archipelago Sea is its zonation in accordance with the relative shares in the land or seascape areas. Originally Häyrén [16], by using this biogeographical criterion, divided the archipelago into three zones. He showed that zonation ranges from the sheltered inner archipelago, where the landscape dominates, to the intermediate zone (even proportions) and finally to the more open, seascape dominating outer archipelago.

Fig. 1 (**A**) Illustration of the main structure and (**B**) location (marked with star) of the Seili automated monitoring buoy. 1 = weather station; 2 = the platform with solar panels, winch and GSM telemetry equipment; 3 CTD-probe; and 4 = anchors. The buoy illustration is reproduced by permission of *YSI Inc*. The map is derived and modified from *Ortho* images by National Land Survey of Finland 2004–2018, license CCBY-4.0
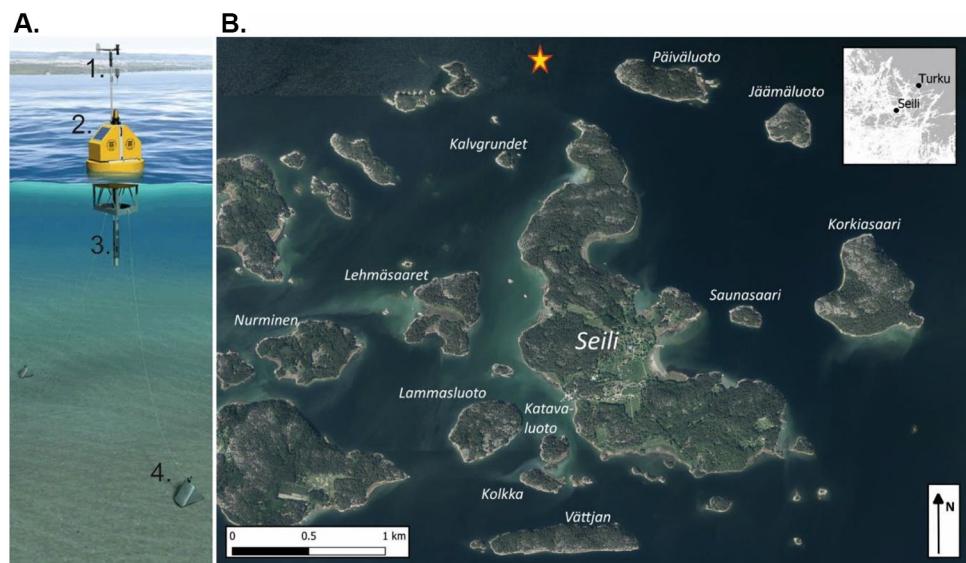
The existence of these zones is caused by the slow post-glacial uplift of the tilting coastal plain, around 0.5 cm annually, and this uplift has been estimated to be proceed some 10,000 years, still [17]. Hänninen et al. [10] proved that zonation could also be traced in water quality, i.e., in nutrient concentrations in seawater, which form hydrographical zonation comparable to those found in biogeographic studies. Kirkkala et al. [18] and Erkkilä and Kirkkala [19] showed that similar gradients could also be revealed through seawater Secchi depth transparency (i.e., a method to estimate the intensity or rate of biological primary production in the open sea environment) and microalgae concentration (chl-a). Later, Hänninen et al. [5] confirmed that other pelagial marine biodiversity patterns and also the biodiversity of benthic biota generally followed transitional zones corresponding to presented hydrographical gradients.

2.2 Profiling Buoy

The technical description of the profiling buoy and examples of produced time-series visualizations are presented in detail on the Seili environmental monitoring program webpages at: https://saaristomeri.utu.fi/home/, which is administered by the current authors from the Archipelago Research Institute, University of Turku and Turku University of Applied Sciences. Therefore, the text of this chapter is an amended update of captions presented therein under license CCBY-4.0. [20].

At-sea monitoring is important for the comprehension of the oceanographic processes and the further development of the environmental state of the sea, and it could provide essential information for environmental management purposes. The database provider in the present study is an automated monitoring buoy located in the vicinity of island of Seili, in the northern Baltic Sea, SW, Finland (Fig. 1). The water depth in the study location is ca. 50 m. The intention of the buoy monitoring is to produce detailed data on seawater quality and vertical stratification as well as seasonal and inter-annual oceanographic changes occurring in the sea. Combined short-term (e.g., hours, days, weeks) and long-term observations (e.g., months, years) also reveal trends and patterns that can help interpret experimental results or provide new research hypotheses. Since 2006, the buoy has been in operation annually during the growing seasons from the early spring to the late autumn. This Seili environmental monitoring program is carried out by the Turku University of Applied Sciences in close collaboration with the University of Turku (the Archipelago Research Institute), other Finnish Marine Research Infrastructure (FINMARI) consortium partners (the Finnish Environment Institute and the Finnish Meteorological Institute), and the visualizations of the buoy measurements are produced by the Archipelago Research Institute [20].

The monitoring station is composed of a *YSI 6952* buoy platform with a *YSI 6000* multiparameter probe [21], which measures seawater salinity (PSU), temperature (°C), dissolved oxygen (O_2 mg/l and saturation %), turbidity (NTU), chl-a (fluorescence, expressed as µg/l), and cyanobacteria or blue-green algae (BGA; phycocyanin fluorescence, expressed as cells/ml) concentrations (Fig. 1). The probe is operated vertically in the 2–40 water column by winch, and the station produces four vertical profiles per day with a measuring interval of every 0.5 s, when the winch is active. The data are sent and stored to an external server twice a day via GSM, and profiles can be visualized and downloaded online at https://saaristomeri.utu.fi/odas_en/. The buoy is deployed to operate in the spring as soon as the ice breaks and is removed in the early winter before the sea ice forms [20]. Detailed technical information for the profiling buoy operations are given in Loisa et al. [22].

In 2015, the station was equipped with a *Vaisala weather station* (*WXT-520*), which is intended to provide information of the closely linked interaction between the weather conditions just above the sea surface and uppermost sea-surface water layer. The weather station measures air temperature (°C), air pressure (mbar), humidity (%), precipitation (mm), wind speed (m/s), and wind direction (°). This is the first weather station in Finland to be installed on a floating automated monitoring platform. Similar to water quality measurements, the weather station data are sent and stored to an external server twice a day, and it can be visualized and downloaded online at https://saaristomeri.utu.fi/buoyweather/ [23].

2.3 Data and Modelling

2.3.1 Data

For this study, we focused on two variables: chl-a and seawater temperature (Fig. 2). The temperature (hereafter, temp) is a reasonable and practical candidate predictor of, for example, chl-a concentration (e.g., [24, 25]), as its increase in spring is the main trigger that launches biological activities and accelerates the speed and rate of the biological processes in the Baltic Sea seawater after the cold and icy winter. The seawater temperature is closely connected by the air temperature above the sea surface, as when we have a warm period, the heat energy is very rapidly transferred to the upper water column and mixed by the local winds, and during the cool period, the dynamics will go in the same but opposite way (e.g., [26]). Moreover, the temperature is one of the key variables in all environmental monitoring programs, and its future predicted values can be obtained reliably from various weather information services.

To avoid having to model the water depth as a continuous variable, we discretized it into two classes, being surface
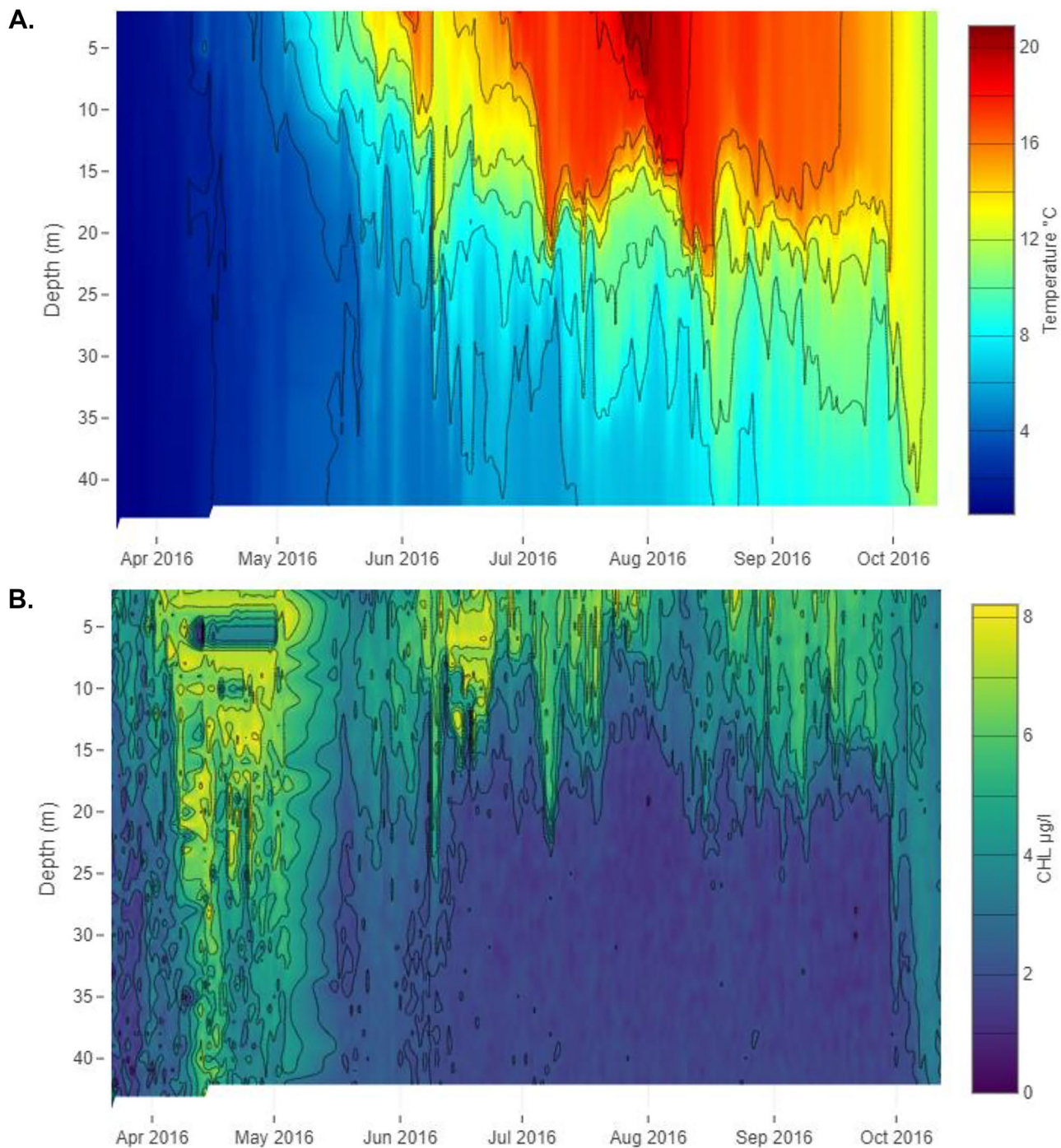
Fig. 2 The seasonal variation of (**A**) seawater temperature (°C) and (**B**) chlorophyll-a (CHL, µg/l) profiles at the monitoring station from the surface to 40 m depth in April–October 2016. The temperature profile shows the development of thermal stratification, whereas the chlorophyll-a profile reflects the microscopic algal biomass in the water column (see text for more detail). The figures are derived from the website of Seili environmental monitoring program 2020a

(depth < 20 m) and hypolimnion (depth ≥ 20 m) water columns and worked with the daily means of the observed values of chl-a and temp over the two depth categories. The reason for the choosing of these categories comes from the biological events occurring differently in the layers (see later in the text), and by the choosing of two separate layers, we wanted to, firstly, make difference between these layers, and, secondly, decrease the variation in data due to different kind and speed of biological processes occurring in these layers. We selected daily means to compensate for occasional

missing values and to decrease internal variation within data, as we used daily weather forecasts as predictors for the chl-a changes in the latter analysis. However, daily weather forecasts were used only in final predictions for the "Seili-index" and its visualization for the webpages (the index is illustrated detailed in *Discussion*) but not in the actual model construction. There, we used air temperature data from the *Vaisala* weather station, which was mounted on the buoy platform.

The biological grounds for the selected two water columns are the formation of thermocline during the summer season between the upper and lower water layers at the depth of around 20 m. Thermocline is a real physical boundary, in which the vertical seawater column has a strong temperature gradient within 1–2 m of depth difference. Its formation in early summer is due to increased solar energy due to the onset of the thermal growing season that warms up seawater in the upper layer, which is constantly mixed by local winds. Biologically, the thermocline is essential for the pelagic ecosystem functioning to support the majority of the primary production of microalgae during their growing season, which occurs within this photic zone above the thermocline (Fig. 2) (e.g., [10, 14]).

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 these two variables. A useful choice for this is the vernal blooming of the diatom microalgae, called the "the spring bloom", which occurs regularly at different intensities during the spring months of every year. Blooming is triggered not only by increased light, but also by fast-melting snow and ice in the catchment area of the Archipelago Sea resulting in nutrient-rich freshwater runoffs into the sea [27]. We located the blooming peak value of chl-a for each year and aligned the data so that the peak always occurs on day 0. For years 2017 and 2018, the blooming had occurred already before the buoy measurements started, and we aligned these data using the blooming of year 2016 as a benchmark because of the similar spring conditions in the terms of snow and ice melting (Fig. 3).

2.3.2 Modelling

Our model of choice is the *Generalized Additive Mixed Model* (GAMM) [28], 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:

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

where $X_1, \ldots, X_p$ are the fixed effects, $f_1, \ldots, f_p$ are unknown smooth functions estimated using ten thin plate regression
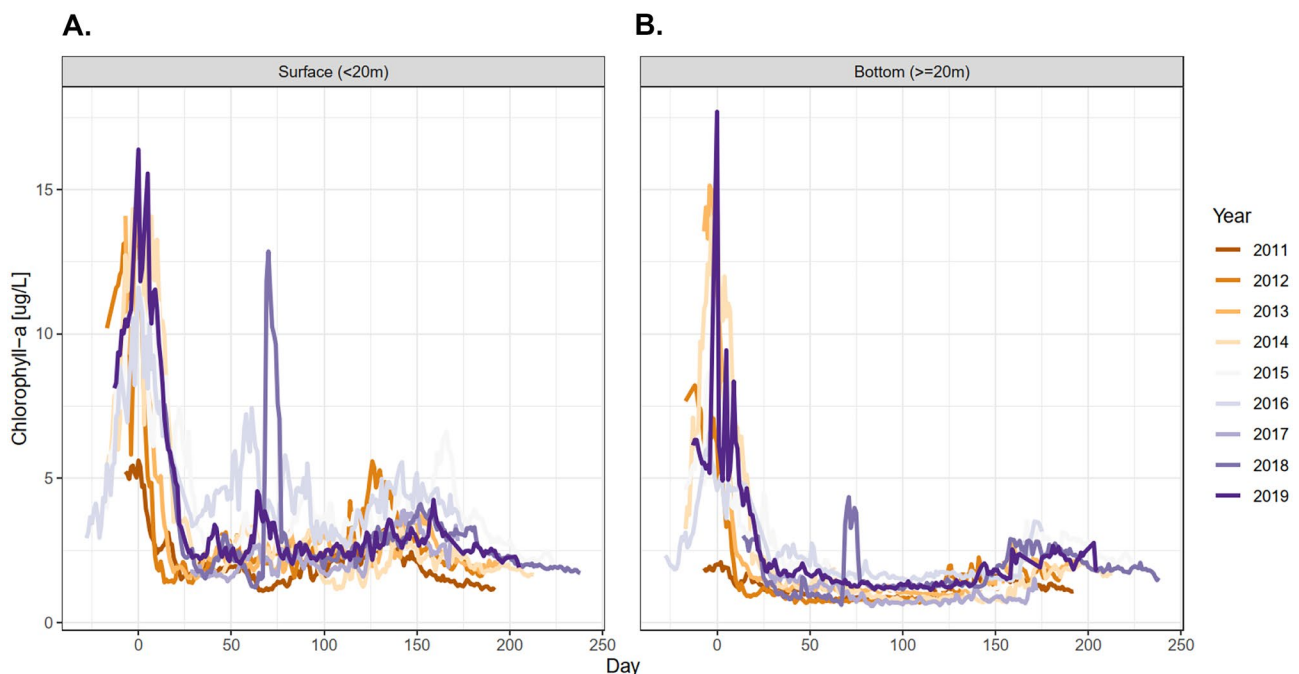

Fig. 3 The time courses of the aligned chl-a and seawater temperature for the two categories (**A** = surface water (< 20 m); **B** = bottom water (≥ 20 m)) of the depth in 2011–2019. In both plots, the time axis is relative to the spring diatom blooming peak, which is set to occur on day 0 for all years. Additionally, also a second peak, being the so-called autumn bloom, is visible around day 150 (around mid-September)

splines via restricted maximum likelihood (REML), and Z_1, …, Z_q are the random effects. We used the function *gamm* in the R-package *mgcv* [28] 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 (shared by all observations of a given year, avoiding the oversaturation of the model with a large number of year and parameter interactions). We would next like to comment briefly on the model assumptions. First, while appearing unfit for our non-negative-valued response variable, our experiments (not shown here) revealed that the normal distribution provided, in fact, better results in the current scenario than the Gamma or Tweedie distribution (which are distributions taking only non-negative values) combined with the log-link. This observation is in line with the long history of using the normal distribution to model biological data. Second, the validity of the defining assumption of a generalized additive model, i.e., additivity in the smooth effects, is usually verified with residual diagnostics, and these have been provided in the supplementary material. Finally, we note that, while some other modelling strategies, such as Gaussian process regression, could have been used with the current non-linear data, we ultimately chose GAMMs for their ability to keep a balance between flexibility and interpretability, the latter of which is a particularly desirable feature in a study such as this. Some alternative modelling approaches have also been detailed in the *Supplementary material*.

Two separate analyses supporting each other were performed for the daily mean data. The first analysis, named as the *raw values model*, is conducted with the original, untransformed data, depicted in Fig. 4, as the response variable, *Y*, and was used to study temporal changes and dynamical features of chl-a concentrations in seawater during the season. The second, the *difference-day model*, used instead the daily differences of the original values as the response. The first analysis was included due to its capability to accurately infer the effect of temperature on chl-a concentrations in seawater, which gave us valuable insight on the temporal differences of microalgae dynamics between the years. The second analysis, while more difficult to interpret due to the modelling of differences instead of the raw values, was included from the viewpoint of local prediction. Namely, unlike the *raw values model*, the *difference-day model* allowed for aligning the predicted chl-a concentrations with the true observed values in a local time frame (see the "Results" section for a more detailed explanation). However, in order to compare the two models' capabilities, we still considered prediction also with the *raw values model*. Note that, whereas the GAMM model itself already allows capturing the trend and large-scale behavior of the response variable in time through the smoothed day of the year predictor, the

difference-day model further takes the time dependency in the data into account through the differenced values.

The prediction study was conducted in a moving-window fashion. First, "training data" comprising of the measurements for the years 2011–2018 along with first 25 measurement days of the year 2019 were formed, as the latter set of data was needed to estimate the effect of the year 2019. These data were then used to fit the GAMM model described above, and the fitted model was then used to predict chl-a for the next 10 days of 2019 using the true air temperature data of the days obtained from *Vaisala* weather station. To make the predictions, the index required a forecast of the sea surface air temperature for the prediction period, which we obtained from Norwegian online weather services (https://www.yr.no/) (note that this "prediction using predicted values" naturally accumulates more error in the results than would be the case if the predictor values were fully accurate). Next, the days 26–35 were appended to the training data, and the process was repeated, which again predicted the next 10 days of 2019. This cycle was continued for a total of 18 times until the end of the measurement period of the year 2019. The reason for predicting only 10 days ahead was partly due to the fact that this is typically the maximum length given in public weather forecasts and partly because, after some time point, the uncertainty in our predictions and the uncertainty of the weather predictions start to accumulate to the extent of making the predictions too unreliable to be useful in practice.

3 Results

In the *raw values model,* all smoothed fixed effects were significant in predicting chl-a, and for the fixed effects, only depth itself did not carry any predictive power (Table 1). The model had the adjusted R^2-value of 0.734 with the *scale parameter estimation* of 1.146, indicating a good model fit (a more extensive study of the model diagnostics, both for this model and the ones to come, is conducted in the supplement). The estimated depth-wise smooth functions of temperature and day, which play the role of the straight line from a linear regression model in GAMMs, are given in Fig. 4 revealing the effects of the two covariates on the chl-a for the two depth categories.

Several landmarks are clearly visible in the plots, including the increased amount of chl-a around both day 0 (the spring blooming) and day 150 (the autumn bloom) (Fig. 3). Similarly, the water surface temperature of around 4 °C is most favorable for the increased production of chl-a (Fig. 4), which is usually explained by the post-winter release of nutrients from the Airisto Inlet catchment area that triggers the spring diatom blooming in sea water. Moreover, we found clear differences in the chl-a curves for the 9 years
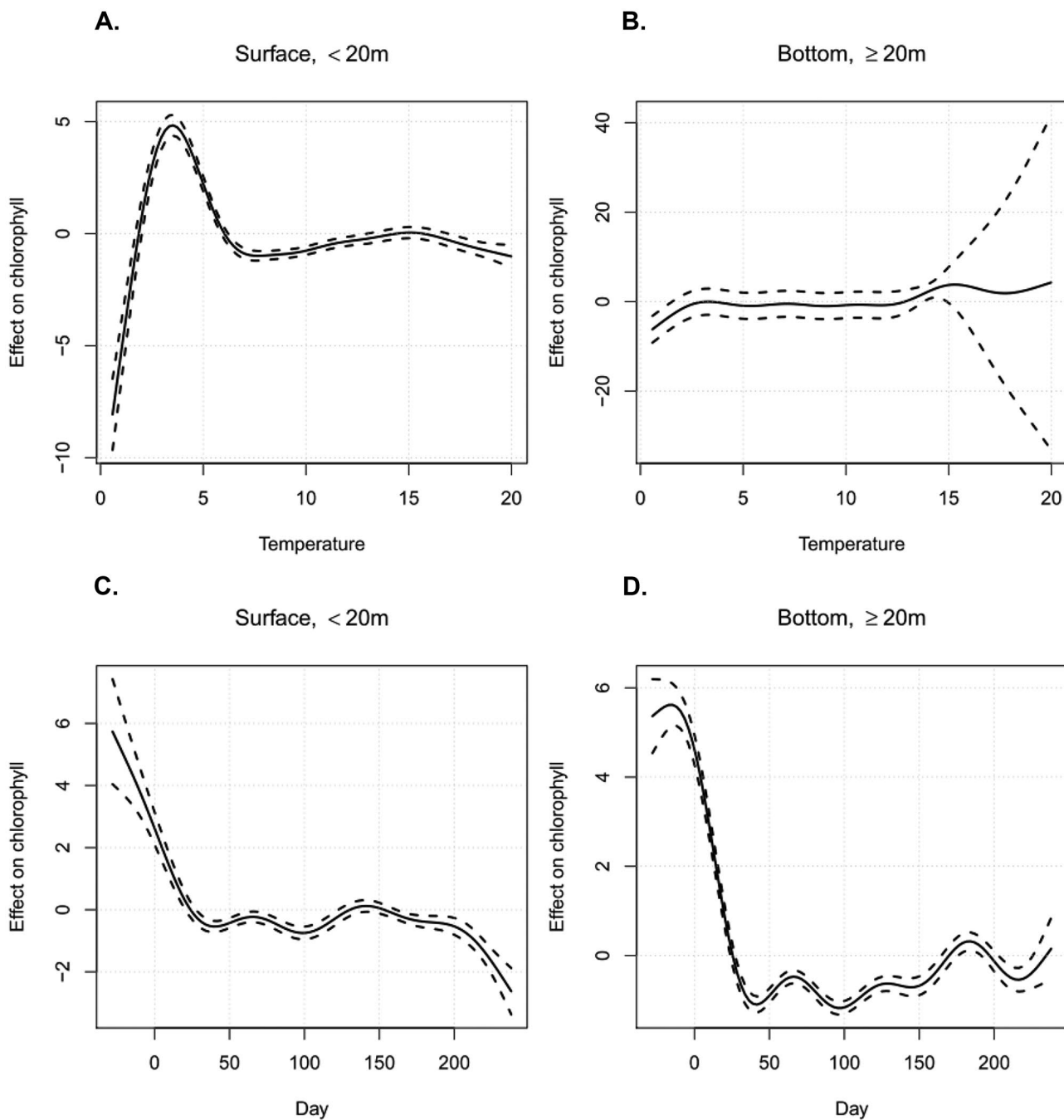
A.

Surface, < 20m

B.

Bottom, ≥ 20m

C.

Surface, < 20m

D.

Bottom, ≥ 20m

Fig. 4 The estimated smooth functions and their 95% point-wise confidence intervals for seawater temperature (above) and day of the year (below) on the two categories of depth for the raw values model for the full data for the years 2011–2019. To interpret the plots, one must note that the model is *additive* (see text)

with the years 2012–2016 being prominent (Fig. 3). A high peak at around day 75 in the 2018 curve showed an exceptionally high blue-green algae (BGA) blooming period in July in the mentioned year. To interpret the Fig. 4 plots more accurately, one must note that the model is *additive*. That is, for example, if the surface water temperature drops from 10 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 units, assuming all other covariates are held fixed. One must note that splines are notoriously inefficient at extrapolating, which explains the expansion of the

Table 1 The model output for the fixed effects. The upper table shows the estimated parameters for the standard linear effects for the smooth effects, and the lower table shows the smooth effects. R-sq. (adj) = 0.734, scale est. = 1.146, n = 3640

Parametric coefficients

	Estimate	Std. Error	t	p-value
(Intercept)	3.432	0.160	21.480	< 0.001
Depth ≥ 20	-0.827	1.477	−0.560	0.575
Approximate significance of smooth terms				
	Estimated d.f.	F		p-value
s(temp):depth < 20	8.955	107.060		< 0.001
s(temp):depth ≥ 20	6.906	57.100		< 0.001
s(day):depth < 20	8.394	30.770		< 0.001
s(day):depth ≥ 20	8.909	163.860		< 0.001

confidence intervals near the ends of the range, especially for chl-a in the bottom-depth category.

The results of the prediction study are shown in Fig. 5. The main difference between the results of the two models is that the *raw values model* has no guarantee that the predicted values in a window are close to the true observed values at the beginning of the window, which are used as part of the training data. This is clearly visible in Fig. 5, panel A, where the true values and predictions are sometimes quite far off already at the left end of the prediction window, whereas the *difference-day model* produces predictions by adding the predicted differences directly to the true observed value at the beginning of the window meaning that the predictions always start at the true value. Still, regardless of this difference, both models perform rather well in prediction and are able to follow the global patterns of the true curve (however, for the *difference-day model*, this is partially a consequence of forcing the predictions to match the true values at the window borders). Finally, note that the high accuracy of the results is to some extent also due to the used short window length of only 10 days. If the window length was increased, the results would naturally be more prone to error and uncertainty.

4 Discussion

Our aim was to create an online "Seili-index" that could be visualized easily. We managed to propose a reliable GAMM model that predicted the chl-a content of the seawater a number of days ahead. Both presented models, the *raw values model* and the *difference-day model,* supported this finding and suggested a similar kind of relation between temperature and chl-a concentration in seawater. However, the relation was more evident in the upper water column (< 20 m).

Moreover, our test predictions for the year 2019 revealed that the index was able to predict chl-a adequately in both depth categories throughout the year.

In the beginning, our original plan was to use also nutrient levels, especially phosphorus and nitrogen, as predictors for the chl-a concentrations in the seawater, but very soon, we realized that those were not possible to include into the model. The practical reason for this was missing or inadequate nutrient data. The present-day technical solutions to include nutrient measuring sensors into the profiling buoy probe are not yet sophisticated enough to make reliable measurements without too expensive or too laborious efforts to run an automated monitoring program like this. Thus, we need to wait for future innovations to make this a reality. However, when possible in later applications, we will certainly be able to produce longer and more accurate predictions and to deal more precisely with inter-annual variations in seawater quality changes. Then again, the solution for sufficient data availability cannot be found from the conventional monitoring programs either, as environmental sea area monitoring, operated by regional authors and institutes with day- or week-long sampling intervals, is too scarce to fulfill the data presumptions and further modelling requirements. Therefore, we were pleasantly surprised with the findings that the seawater temperature alone had enough statistical power to produce adequate chl-a level predictions.

Similarly, as the energy to the primary production is based on solar radiation, we originally considered also involving daylight dose in to the model, but this turned out not to be feasible. The main practical reason for leaving out the daylight dose was that the buoys' weather station is not equipped with the sensor that measures solar radiation intensity. Here we pondered using data of our another solar sensor equipped weather station, which is located at our research institute at the island of Seili, some 5 km south-east from the buoy monitoring station. However, as positioned in the middle of the island on dry land, the station is at times shaded by the trees and surrounding buildings, at least partly. Thus, the station does not measure the same natural solar phenomena that occur at sea level in the open sea area. Therefore, as the present study could possibly act as an example to some other setup in later implementations, we decided to restrict the data to that produced only by the buoy itself. Later, if the buoy will be equipped with the solar sensor, the solar dose is rather easy to incorporate in to the model, if seen as necessary.

4.1 Relevance and Future Prospects

The relevance of this kind of study has a great importance not only for the public but also for research and especially for environmental monitoring. For the public or for individuals, the "Seili-index" gives an opportunity, for instance, to
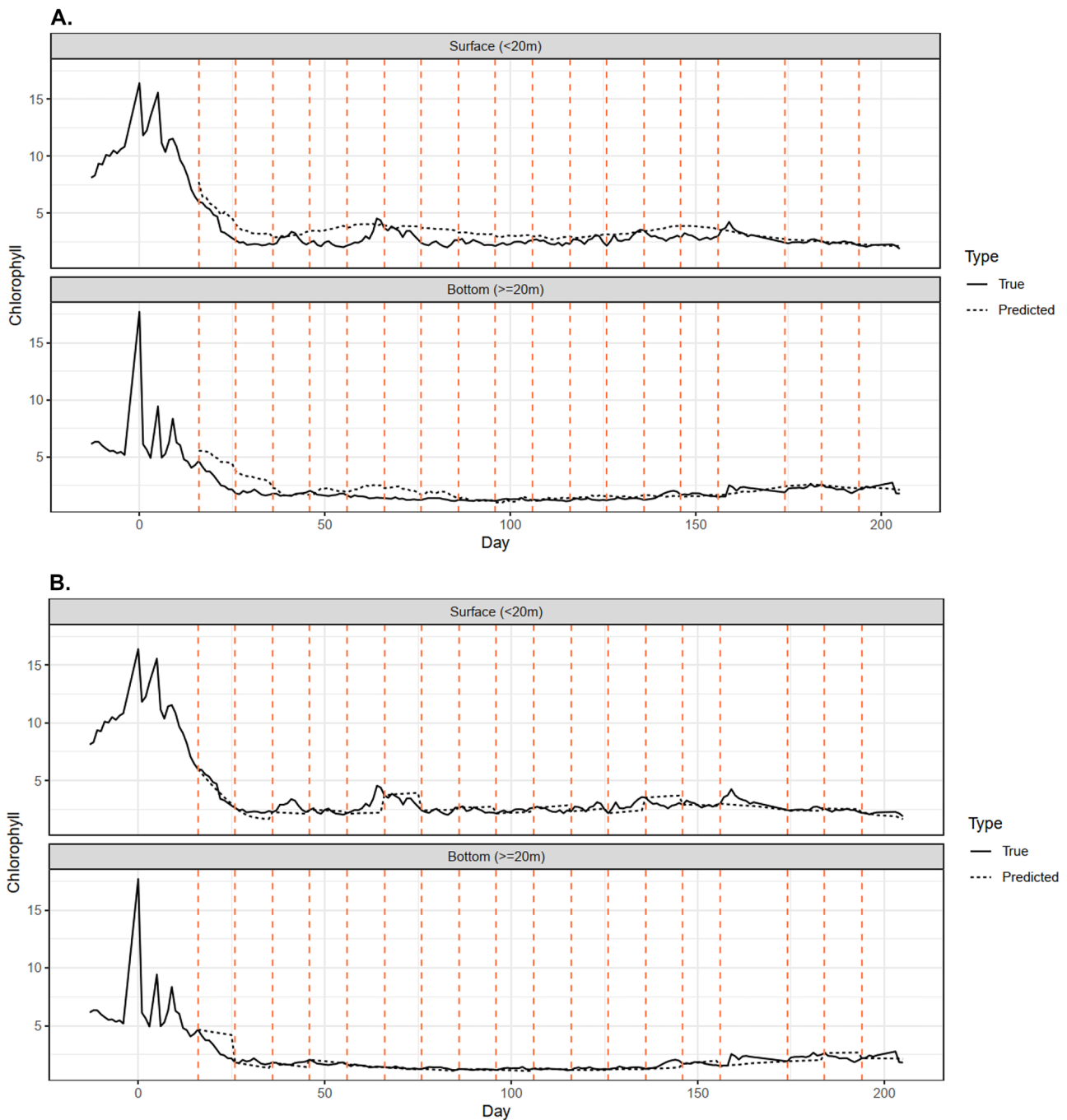
A.

B.

Fig. 5 A The predicted values of chl-a (the dashed black lines) shown together with the true values (solid lines) for the two depth categories. The red vertical lines mark the 10-day prediction 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 temperature values of the interval to obtain the predictions. **B** Similarly, the prediction results using the differences between daily measurements as a response (the dashed black lines) shown again together with the true values (solid lines) for the two depth categories

plan practical recreational seaside activities based on accurate information about the present seawater quality and, moreover, for the next 10 days. This study could later provide some additional value when interpreting environmental monitoring data on public health guidelines and underscore

the consequences of recreational usage or potential environmental pollution to the sea. In terms of monitoring, a special case here is the concept of biodiversity or rather, a loss of it. The *Marine Strategy Framework Directive* [29] defines that biodiversity is one of the main depictions for

a balanced environmental status. In general, marine biodiversity and all ecological production are based on primary production, which in the pelagial habitat, only occurs in the photic surface water layer (0–20 m). In the shallow Archipelago Sea, with an average depth of only 23 m, this means that basically the vast majority of all biodiversity is promoted by the upper water column. Therefore, Vuorinen et al. [30] have stressed supportively that direct monitoring of chl-a above the seasonal thermocline (< 20 m) or any other biological monitoring program of seawater therein can be seen as a priority when implementing MSFD in the Baltic Sea. MSFD [29] also highlighted that eutrophication-induced biodiversity loss at the trophic levels of the marine food web does not necessarily, in the terms of MSFD and other biodiversity monitoring programs, indicate parallel deterioration of environmental quality. The loss of marine biodiversity, which could be resulting from other reasons than environmental damage, has not been included in any monitoring program yet. Therefore, to get the best possible understanding about occurring processes and driving forces contributing to the ecological state of the Archipelago Sea, we strongly suggest that, apart from chl-a, the inclusion of certain chemical parameters of seawater (e.g., salinity, nutrient concentrations) and one or two environmentally sensitive indicatory species groups, like the reproduction success of the Baltic herring (*Clupea harengus membras*, pelagial key species) and the zooplankton community ratio (*cladocera/copepod*, the first trophic level above primary producers indicated by chl-a) in the intensive and simultaneously occurring monitoring program, would more effectively reflect the state of the pelagic environment. The reason for this is that, economically, it would be impossible to include everything under surveillance; thus, we must concentrate on only the best revealing indicators. This would also underscore the importance of basic biological research and modelling for eutrophication tolerance, food webs, and cascading effects therein as described previously by MSFD [29] as well. Moreover, when considering future modelling studies, biological ecosystem models built for the assessment of the Baltic Sea environmental development must take into account changes in species composition during the gradual transition of the ecosystem from marine species to freshwater species due to a climate change-induced increase of rainfall in the area resulting little-by-little in a more lake-type marine environment.

Considering our future modelling studies, the work will include the better implementation of the index as an online service for improving the prediction accuracy by using more sophisticated modelling techniques and obtaining predictions also for other variables measured by the Seili buoy such as the cyanobacteria concentration (BGA cells/ml). Moreover, in the southern Finland coastline, three similarly operating monitoring stations will be moored in the Hanko

Fig. 6 A The time courses of cyanobacteria concentration (BGA ▶ cells/ml) for the surface water column (< 20 m) in 2015–2019. The curves have been adjusted in a similar manner as in for chl-a, i.e., the spring blooming peak has been set as the day 0. There are clear differences in the BGA curves among the five years with the year 2018 being prominent. The end-peak in the 2019 curve is due to missing values at the end of validation series. Long-standing high peaks in the 2015 and 2018 curves signify exceptional summers in terms of high cyanobacteria mass blooming events in the mentioned years. **B** The validation results of BGA post hoc GAMM model. The plot shows the predicted values of BGA (the dashed black lines) together with the true values (solid lines) for surface seawater layer (< 20 m). The red vertical lines mark the 10-day periods described previously, i.e., 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 temperature values of the interval to obtain the predictions

Peninsula (owned by Tvärminne Zoological Station/Helsinki University) and in the Porkkala Peninsula and at Utö island (both owned by Finnish Meteorological Institute), which will provide a possibility to create predictive GAMMs for a larger geographical area in the future. Nevertheless, the presented GAMM modelling of profiling buoy data with graphical visualization software can be considered as an important first step or initiative for a similar type of marine environment analytical tool, not only in Finland but also in other countries running coastal environmental monitoring programs as well.

4.2 Next-Step Training Test

Due to workable GAMM models with seawater chl-a levels, we were encouraged to conduct a separate *next-step training test* also on cyanobacteria blooming events. We wanted to test whether the effect of environmental parameters of surface water (< 20 m), temperature (°C), and local wind speed (m/s, obtained from the buoy's weather station) could be used as predictors of blue-green algae (BGA cells/ml) in seawater using a similar type of GAMM model as for chl-a. In the Archipelago Sea, it is typical that during warm midsummer conditions, intensive BGA blooming events occur only within a couple of hours, if the specific environmental preconditions are met. Intensive BGA blooming requires warm seawater (~ 20 °C), intensive sunshine, and calm wind conditions (wind speed ~ 0 m/s) to burst out suddenly. Therefore, we modified the GAMM model parametrization to include also the average daily wind speed along with the temperature as a predictor keeping the BGA model otherwise similar to the chl-a *difference-day model*. This model was chosen out of the two based on the results of the diagnostic study conducted in a supplementary material. Note that we had to adjust the observed 2018 values of BGA to the same level as other years due to a calibration difference in that year. Again, we tested our model with the year 2019 real values based on the model built with the data from years 2015 to 2019 until the beginning of the prediction window.

A.

B.
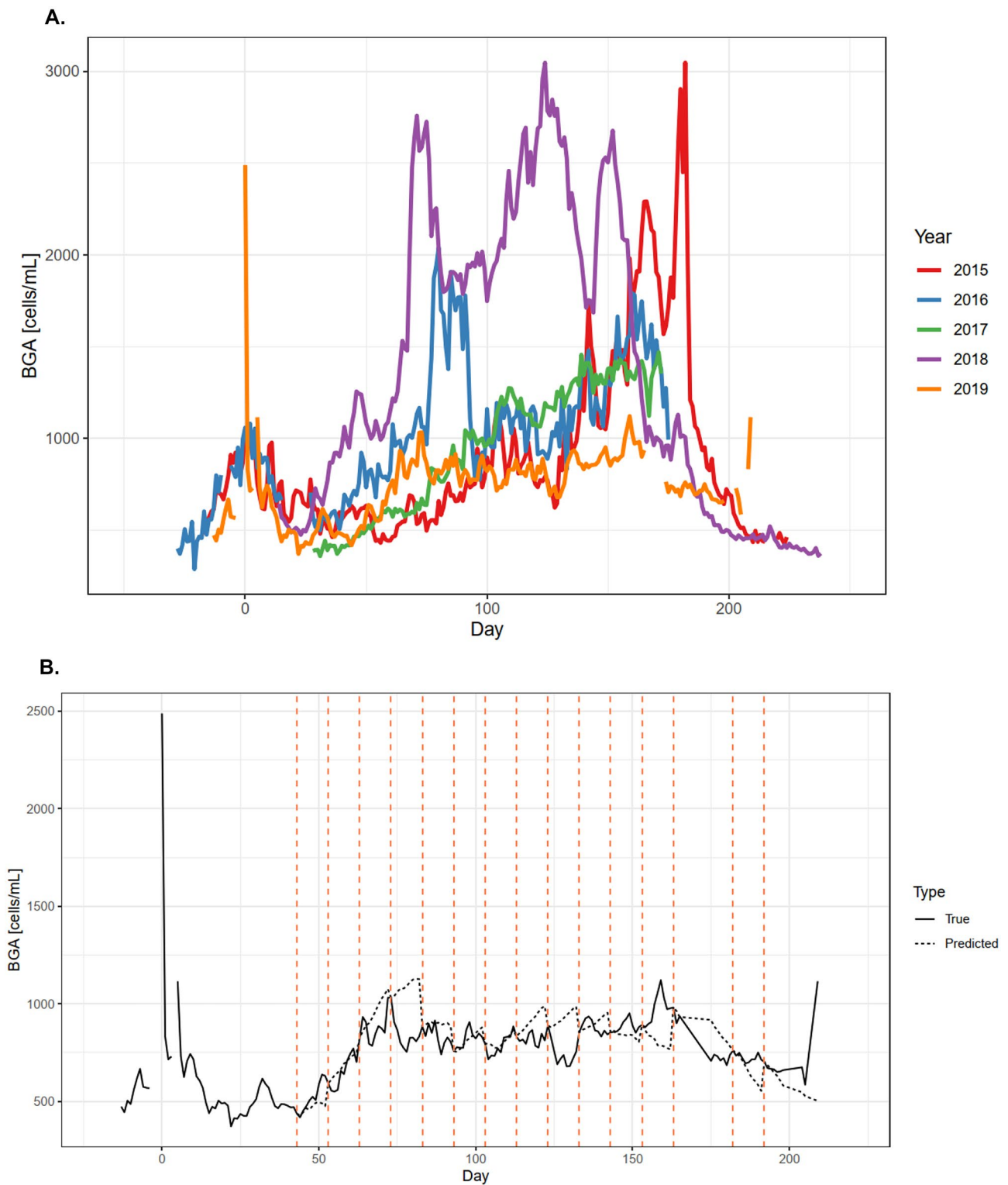
The BGA model showed a rather weak model fit between BGA blooming events and the predictors (adjusted R^2-value of 0.0104 with *scale parameter estimate* of 17,445). Both the wind conditions and air temperature proved to be almost significant predictors (p-values = 0.0917, 0.0895, respectively), whereas the day of the year did not (p-value = 0.3898). We interpret that the possible reason for the statistically indicative p-value for "wind" parameter in the model could initially be a result from the wrong kind of predictor selection. Average wind speed does not necessarily reveal accurately

calm wind conditions coinciding with exceptionally warm weather at the same time. We think that a better modelling result could be achieved by using interaction terms or a dummy variable, for example, one based on principal components of wind conditions and air temperature (also showed indicative statistical significance) and combining these conditions simultaneously together. Moreover, we found clear differences in the BGA curves within the 5 years with the year 2018 especially prominent (Fig. 6). The highest one-time blooming peak was found during year 2015 and occurred during August/September. However, perhaps the most obvious indication of the effect of present global climate warming to the marine ecosystem was exposed in the 2016, 2017, and 2018 curves, which showed long-standing high levels of BGA that signified exceptional summers in terms of high BGA blooming in the mentioned years. The prediction results of the BGA model are shown in panel B of Fig. 6 and imply that, despite the earlier lack of fit, the model performed rather well in prediction, apart from a small number of time intervals where it tried to overestimate the amount of BGA.

4.3 The "Seili-Index" Online Visualization

The text of this chapter is an amended update of captions presented therein under license CCBY-4.0. [31].

The "Seili-index" graphical presentation of chl-a with the described GAMM model is illustrated online at https://saaristomeri.utu.fi/seili-index/ (presented also in Fig. 7). The values (which are discretized to the scale "*Excellent*" = 0.0–2.0 µg/l; "*Good*" = 2.1–2.5 µg/l; "*Fair*" = 2.6–5.8 µg/l; "*Poor*" = 5.9 + µg/l) represent the total amount of phytoplankton at a permanent monitoring station in the vicinity of Seili Island and reflect the general level of eutrophication of the sea. Together with real present-day and estimated values, a larger sphere and a greener color in the visualization indicate a higher level of eutrophication and turbid seawater and a smaller and bluer sphere for cleaner and more transparent seawater. The water quality classification of excellent, good, fair, and poor is based on Aroviita et al. [32]. Thus, an increase in index values may indicate, for example, the onset of phytoplankton blooms as well as increasing/decreasing changes in the state of the
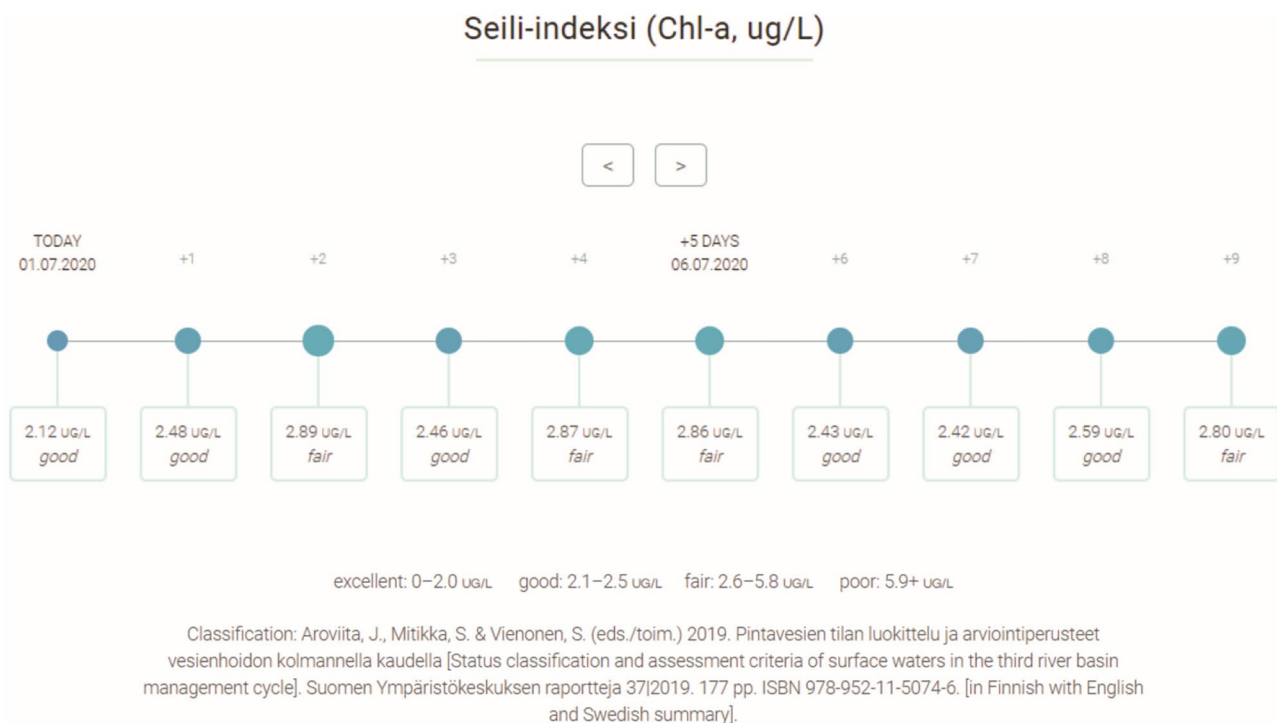

Fig. 7 The illustration of real-time "Seili-index" visualization 1.–10.7.2020 describes the present surface water chlorophyll-α concentration (chl-a, µg/l) in surface sea water (2–20 m) and predicts it 10 days ahead. Together with real present-day and estimated values, a larger sphere and a greener color indicate a higher level of eutrophication and turbid seawater and a smaller and bluer sphere for cleaner seawater. The changes in index values may indicate, for example, the onset of phytoplankton blooms as well as an improving/worsen-ing state of seawater quality in general. The chl-a data is based on data from the Seili automatic profiling monitoring buoy and temperature data from a 10-day, long-term forecast from the *Norwegian Meteorological Institute* https://www.yr.no/ (see text). The figure is derived from the website of Seili environmental monitoring program [31], and the text is an amended update of captions presented therein, license CCBY-4.0

general seawater quality. The prediction manifests a statistically proven relationship between chl-a concentration and surface water/surface air temperatures. The prediction is based on chl-a data (2–20 m) derived from the Seili automatic profiling monitoring buoy and temperature data from a 10-day, long-term forecast from the Norwegian Meteorological Institute https://www.yr.no/ [30].

Declarations

Ethics Approval Not applicable in this MS.

Consent to Participate Not applicable in this MS.

Consent for Publication Not applicable in this MS.

Conflict of Interest/Competing Interests The authors declare no competing interests.

References

1. BACC. (2008). *Assessment of climate change for the Baltic Sea Basin*. Springer, Berlin, 474 pp. https://doi.org/10.1007/978-3-540-72786-6

2. Sintayehu, D. W. (2018). Impact of climate change on biodiversity and associated key ecosystem services in Africa: A systematic review. *Ecosystem Health and Sustainability, 4*(9), 225–239. https://doi.org/10.1080/20964129.2018.1530054

3. Fardila, D., Kelly, T. K., Moore, J. L., & Mccarthy, M. A. (2017). A systematic review reveals changes in where and how we have studied habitat loss and fragmentation over 20 years. *Biological Conservation, 212*, 130–138. https://doi.org/10.1016/j.biocon.2017.04.031

4. Leppäkoski, E., Helminen, H., Hänninen, J., & Tallqvist, M. (1999). Aquatic biodiversity under anthropogenic stress: An insight from the Archipelago Sea (SW Finland). *Biodiversity and Conservation, 8*(1), 55–70. https://doi.org/10.1023/A:1008805007339

5. Hänninen, J., Toivonen, R., Vahteri, P., Vuorinen, I., & Helminen, H. (2007). Environmental factors shaping the littoral biodiversity in the Finnish archipelago, northern Baltic, and the value of low biodiversity. *SEILI Archipelago Research Institute Publications, 4*, 1–19.

6. HELCOM. (2018). *Sources and pathways of nutrients to the Baltic Sea*. Baltic Sea Environment Proceedings 153ISSN 0357–2994. 48 pp.

7. Nixon, S. W. (1995). Coastal marine eutrophication: A definition, social causes, and future concerns. *Ophelia, 41*(1), 199–219. https://doi.org/10.1080/00785236.1995.10422044

8. Cloern, J. E. (2001). Our evolving conceptual model of the coastal eutrophication problem. *Marine Ecology Progress Series, 210*, 223–253.

9. Malone, T. C., & Newton, A. (2020). The globalization of cultural eutrophication in the coastal ocean: Causes and consequences. *Frontiers in Marine Science, 7*, 670. https://doi.org/10.3389/fmars.2020.00670

10. Hänninen, J., Vuorinen, I., Helminen, H., Kirkkala, T., & Lehtilä, K. (2000). Trends and gradients in nutrient concentrations and loading in the Archipelago Sea, Northern Baltic, in 1970–1997. *Estuarine, Coastal and Shelf Science, 50*, 153–171. https://doi.org/10.1006/ecss.1999.0568

11. Väänänen, K., Lappalainen, M., & Högmander, J. (2020). *The Blue Book of Archipelago Sea of Finland*. The unique Archipelago Sea. Veijonen Consulting Oy. ISBN: 9789526944241. 228 p.

12. Hänninen, J., & Vuorinen, I. (2015). Riverine tot-P loading and seawater concentrations in the Baltic Sea 1970s to 2000 - Transfer function modeling on the basis of the total runoff. *Environmental Monitoring and Assessment, 187*, 343. https://doi.org/10.1007/s10661-015-4538-y

13. Boscolo-Galazzo, F., Crichton, K. A., Barker, S., & Pearson, P. N. (2018). Temperature dependency of metabolic rates in the upper ocean: A positive feedback to global climate change? *Global and Planetary Change, 170*, 201–212. https://doi.org/10.1016/j.gloplacha.2018.08.017

14. Voipio, A. (1981). *The Baltic Sea*. Elsevier Oceanography Series, vol. 30. Elsevier, Amsterdam. eBook ISBN: 9780080870687. 417 pp.

15. Leppäkoski, E., & Bonsdorff, E. (1989). *Ecosystem variability and gradients. Examples from the Baltic Sea as a background for hazard assessment*. In: L. Landner (Ed.) *Chemicals in the Aquatic Environment*. *Advanced Hazard Assessment*, pp. 6–58. Springer-Verlag, Berlin.

16. Häyrén, E. (1900). Längszonerna i Ekenäs skärgård. *Geografiska föreningens tidsskrift* 12.

17. Johansson, J. M., Davis, J. L., Scherneck, H.-G., Milne, G. A., Vermeer, M., Mitrovica, J. X., Bennett, R. A., Jonsson, B., Elgered, G., Elósegui, P., Koivula, H., Poutanen, M., Rönnäng, B. O., & Shapiro, I. I. (2002). Continuous GPS measurements of postglacial adjustment in Fennoscandia. 1. Geodetic Results. *Journal of Geophysical Research, 107*, 2157–2184. https://doi.org/10.1029/2001JB000400

18. Kirkkala, T., Helminen, H., & Erkkilä, A. (1998). Variability of nutrient limitation in the Archipelago Sea, SW Finland. *Hydrobiologia, 363*, 117–126.

19. Erkkilä, A., & Kirkkala, T. (2000). Zones of surface water in the Archipelago Sea. *Regional Environmental Publications, 149*, 1–34. [In Finnish].

20. Seili Environmental monitoring programme. (2020a). Odas profiling buoy datasets. Turku University/ Turku University of Applied sciences, Turku. https://saaristomeri.utu.fi/odas_en/. Accessed 1 January 2022.

21. YSI inc. (2012). *6-series multiparameter water quality Sondes*. User manual, Rev. J, 2012. 379 p.

22. Loisa, O., Körber, J., Laaksonlaita, J., & Kääriä, J. (2017). High-resolution monitoring of stratification patterns in the Archipelago Sea, Northern Baltic Sea, using an autonomous moored vertical profiling system. *Proceedings of IEEE/MTS OCEANS 2017 – Anchorage, Alaska, USA*.

23. Seili Environmental monitoring programme. (2020b). Odas buoy weather station datasets. Turku University/ Turku University of Applied sciences, Turku. https://saaristomeri.utu.fi/buoyweather/. Accessed 1 January 2022.

24. Fox, M. F., Kester, D. R., & Yoder, J. A. (2005). Spatial and temporal distributions of surface temperature and chlorophyll in the Gulf of Maine during 1998 using SeaWiFS and AVHRR imagery. *Marine Chemistry, 97*, 104–123.

25. Kavak, M. T., & Karadoğan, S. (2012). The relationship between sea surface temperature and chlorophyll concentration of phytoplanktons in the Black Sea using remote sensing techniques. *Journal of Environmental Biology, 33*(2 Suppl), 493–498.

26. Hall, S. (2016). How sea surface temperatures affect an atmospheric phenomenon. *Eos*. https://doi.org/10.1029/2016EO057821

27. Hänninen, J., & Vuorinen, I. (2011). Time varying parameter analysis of the Baltic Sea freshwater runoff. *Environmental Modelling and Assessment., 16*, 53–60. https://doi.org/10.1007/s10666-010-9231-5

28. Wood, S. N. (2017). Generalized additive models: An introduction with R (2nd Edition). Chapman & Hall/CRC, Boca Raton. ISBN 978-1-4987-2833-1. 476 pp.

29. MSFD. (2008). Directive 2008/56/ec of the European parliament and of the council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive).

30. Vuorinen, I., Hänninen, J., Rajasilta, M., Laine, P., Eklund, J., Montesino-Pouzols, F., Corona, F., Junker, K., Meier, M. H. E., & Dippner, J. W. (2015). Scenario simulations of future salinity and ecological consequences in the Baltic Sea and adjacent North Sea areas - implications for environmental monitoring. *Ecological Indicators, 50*, 196–205. https://doi.org/10.1016/j.ecolind.2014.10.019

31. Seili Environmental monitoring programme. (2020c). Seili-index portal. Turku University/ Turku University of Applied sciences, Turku. https://saaristomeri.utu.fi/seili-index/. Accessed 1 January 2022.

32. Aroviita, J., Mitikka, S., & Vienonen, S. (eds./toim.). (2019). *Pintavesien tilan luokittelu ja arviointiperusteet vesienhoidon kolmannella kaudella [Status classification and assessment criteria of surface waters in the third river basin management cycle]*. Suomen Ympäristökeskuksen raportteja 37|2019. ISBN 978–952–11–5074–6. [in Finnish with English and Swedish summary]. 177 pp.