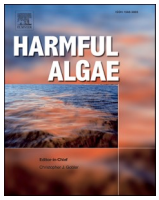
Harmful Algae 126 (2023) 102442

Contents lists available at ScienceDirect

Harmful Algae

journal homepage: www.elsevier.com/locate/hal



Original Article

Forecasting harmful algae blooms: Application to *Dinophysis acuminata* in northern Norway

Edson Silva [a,*], François Counillon [a], Julien Brajard [b], Lasse H. Pettersson [b], Lars Naustvoll [c]

[a] Nansen Environmental and Remote Sensing Center, and Bjerknes Centre for Climate Research, Jahnebakken 3, Bergen, N-5007, Vestland, Norway
[b] Nansen Environmental and Remote Sensing Center, Jahnebakken 3, Bergen, N-5007, Vestland, Norway
[c] Institute of Marine Research, Nye Flødevigveien 20, Arendal, NO-4817, Agder, Norway

ARTICLE INFO

Edited by Dr. C. Gobler

Keywords:
Harmful algae bloom
Forecast model
Dinophysis
Arctic

ABSTRACT

Dinophysis acuminata produces Diarrhetic Shellfish Toxins (DST) that contaminate natural and farmed shellfish, leading to public health risks and economically impacting mussel farms. For this reason, there is a high interest in understanding and predicting *D. acuminata* blooms. This study assesses the environmental conditions and develops a sub-seasonal (7 - 28 days) forecast model to predict *D. acuminata* cells abundance in the Lyngen fjord located in northern Norway. A Support Vector Machine (SVM) model is trained to predict future *D. acuminata* cells abundance by using the past cell concentration, sea surface temperature (SST), Photosynthetic Active Radiation (PAR), and wind speed. Cells concentration of *Dinophysis* spp. are measured in-situ from 2006 to 2019, and SST, PAR, and surface wind speed are obtained by satellite remote sensing. *D. acuminata* only explains 40% of DST variability from 2006 to 2011, but it changes to 65% after 2011 when *D. acuta* prevalence reduced. The *D. acuminata* blooms can reach concentration up to 3954 cells l^{-1} and are restricted to the summer during warmer waters, varying from 7.8 to 12.7 °C. The forecast model predicts with fair accuracy the seasonal development of the blooms and the blooms amplitude, showing a coefficient of determination varying from 0.46 to 0.55. SST has been found to be a useful predictor for the seasonal development of the blooms, while the past cells abundance is needed for updating the current status and adjusting the blooms timing and amplitude. The calibrated model should be tested operationally in the future to provide an early warning of *D. acuminata* blooms in the Lyngen fjord. The approach can be generalized to other regions by recalibrating the model with local observations of *D. acuminata* blooms and remote sensing data.

1. Introduction

The *Dinophysis* spp. are cosmopolitan algae present in coastal waters of tropical, sub-tropical, sub-arctic, and arctic regions, which can produce diarrhetic shellfish toxins (DST) and pectenotoxins (PTX) (Reguera et al., 2012). The algae toxins can be accumulated in both wild and aquaculture shellfish and further be consumed by humans, leading to diarrhetic shellfish poisoning outbreaks. Monitoring DST and PTX in shellfish farms helps prevent poisoning, but the loss of production and economic impact on the farms still occurs (Fernandes-Salvador et al., 2021; Martino et al., 2020). For this reason, there is a high interest in forecasting harmful algae blooms (HAB) of *Dinophysis* spp. since it may provide early warning and support the development of mitigation actions (Pettersson and Pozdnyakov, 2013).

Dinophysis spp. are mixotrophic and can feed on other algae - e.g., *Mesodinium rubrum* (Lohmann) Leegard 1908 - and retain their plastids to make photosynthesis, requiring prey and light for a higher growth rate and long-term survival (Kim et al., 2008). When prey decreases and is no longer available, *Dinophysis* spp. continues increasing depending on the light availability, and it fails to grow in complete darkness even in the presence of prey. Little is known about the importance of grazing in controlling *Dinophysis* spp. abundance, but it has been reported in experiments that it may be preyed on by copepods (Jansen et al., 2006; Setälä et al., 2009) and other dinoflagellates (Park and Kim, 2010). In general, the *Dinophysis* spp. HAB have been associated with seasonal variability of thermal stratification and water column stability (Karlson et al., 2021; Reguera et al., 2012).

Three main toxic *Dinophysis* spp. are commonly reported in Scandinavian coastal waters, *D. acuminata* Claparede & Lachmann, *D. acuta* Ehrenberg, and *D. norvegica* Claparede & Lachmann (Karlson et al.,

* Corresponding author.
E-mail address: edson.silva@nersc.no (E. Silva).

https://doi.org/10.1016/j.hal.2023.102442
Received 24 October 2022; Received in revised form 24 April 2023; Accepted 30 April 2023
Available online 5 May 2023

2021). A few studies have assessed the variability of those toxic species in the region. In the Sognefjord, on the west coast of Norway, *D. acuminata* was detected between later spring and early summer, while *D. acuta* and *D. norvegica* were noticed during autumn (Séchet et al., 1990). In the Fløvedigen Bay in the south of Norway, *D. acuminata* and *D. norvegica* were found in the highest abundance in the surface layer from March to December, while *D. acuta* was found from mid-August to December (Dahl and Johannessen, 2001; Naustvoll et al., 2012). In the Gullmar and Koljo fjords on the Swedish west coast (Lindahl et al., 2007), *D. acuminata* hazardous concentrations (cells $l^{-1} > 1000$) were found from August to October in the surface layer, in the pycnocline, and below the pycnocline down to 20 m depth. *D. acuta* was also found in hazardous concentrations (cells $l^{-1} > 100$) for the same period but was more restricted in the surface layer. In the high latitudes, *Dinophysis* spp. HAB are not expected during winter, probably because of the lack of light. The HAB presents a threat mainly during the productivity season from spring to autumn.

To ensure the healthy consumption of natural and farmed mussels in Norway, the Norwegian Food and Safety Authority (NSFA) monitors toxic algae and their respective toxins in mussels for the entire Norwegian coastal waters (https://www.matportalen.no/verktoy/blaskjell-varsel/). This public service, which has been in operation since 1992, is based on weekly sampling of mussels and seawater at 34 locations, forming the basis for dietetic advice on the potential risk associated with the consumption of mussels. One of the genera monitored is the *Dinophysis* spp. (Karlson et al., 2021; Naustvoll et al., 2012; Pettersson and Pozdnyakov, 2013; Reguera et al., 2012). Although the current monitoring system provides quality observations limiting possible human poisoning, a forecast method has yet to be developed in the region.

Several approaches for forecasting *Dinophysis* spp. HAB have been proposed lately (Cruz et al., 2021). Lagrangian particle tracking for dynamic modeling the dispersion of *Dinophysis* spp. cells abundance from contaminated farm locations to the surrounding areas (Ruiz-Villarreal et al., 2016). Generalized linear model fed with past environmental data records to forecast toxin concentration in shellfish flesh (Schmidt et al., 2018). Decision tree model fed with environmental data for predicting the risk *Dinophysis* spp. cell abundance and toxins above a hazardous concentration (Bouquet et al., 2022). Machine learning model fed with past *Dinophysis* spp. cells concentration to forecast the abundance evolution over time (Velo-Suárez and Gutiérrez-Estrada, 2007). Among all those methods, the advantage of forecasting *Dinophysis* spp. cell abundance oriented to a single farm location should be emphasized. A model tailored to a single farm does not rely on surrounding locations being contaminated beforehand, and forecasting cell abundance can anticipate future toxin accumulation in shellfish.

In this study, we assess the *D. acuminata* bloom variability and environmental conditions, and use them to calibrate a statistical forecast model for the Lyngen fjord located in northern Norway. Located in the Arctic, water temperature and daylight hours have large seasonal variability (Giesen et al., 2014; Jakowczyk and Stramska, 2014), and little is known of *Dinophysis* spp. in this region. Besides, long-lasting aquaculture activities at the location have contributed to more than a decade of observational data record, which machine learning methods can explore. A range of environmental drivers provided by remote sensing is used to calibrate a machine learning model — including sea surface temperature (SST), Photosynthetically Active Radiation (PAR), and surface winds. The prediction system is based on a support vector machines (SVM) for predicting the *D. acuminata* cells abundance on a sub-seasonal time scale — at 7, 14, 21, and 28 lead days. The prediction skill is validated with data from 2014 to 2019 and compared to trivial predictors, such as climatological and persistence forecasts.

2. Material and methods

2.1. Study region

The sampling station at a mussel aquaculture farm (20.6005°E; 69.7918°N) is located in the Lyngen fjord in northern Norway and 60 km from Tromsø city (Fig. 1). The fjord orientation has an opening to the ocean at its north and steep terrain on the side. The area around the sampling station reaches 340 m depth and is sheltered by Uløya island (Hegstad, 2014; Olsen, 2015). Northern Norway presents strong seasonal variability in the incoming solar radiation, SST, and phytoplankton blooms. At this latitude, the sun does not rise above the horizon between the 18th November and the 23rd January and there is sun all day between the 20th May and the 24th July, reaching up to 300 Wm^{-2} day^{-1} of solar radiation (Giesen et al., al.,2014). In the adjacent sea, SST is on average 7 °C, varying from 1 °C to 15 °C between winter and summer (Jakowczyk and Stramska, 2014). Spring and summer phytoplankton blooms are recurrent and their variability is mainly driven by nutrient supply, light availability, SST, water stratification, and wind speed (Silva et al., 2021; Sverdrup, 1953).

2.2. Measurements of Dinophysis spp. and diarrhetic shellfish toxins

D. acuminata, D. acuta, and *D. norvegica*, as well as DST, were provided by the monitoring program of algae toxins in mussels and dietetic advice to the public, and can be provided on demand (www.matportalen.no) by the Norwegian Food Safety Authority (NFSA). Algae have

Fig. 1. Study region. Subplot a) shows the overall location of northern Norway and the bounding box highlighted by the red rectangle show the zoom area of subplot b). The aquaculture farm (sampling station) where *Dinophysis* spp. and DST are measured and geographic locations mentioned in the paper are tagged.

been collected at the aquaculture mussel site weekly (every Monday) between week 11 and 46, which is the prime algae growth season. The monitoring program is run continuously, and the data used in this study covers from 2006 to 2019, except in 2011, when data was unavailable. An integrated (0–3 m) water sample was collected by lowering a tube from the surface to 3 m depth. A subsample of 25 ml was taken from the integrated sample and preserved with acidic Lugol's iodine before being transported to the laboratory for analysis. The sub-sample (25 ml) was filtered on a membrane filter, and the three *Dinophysis* spp. were identified and counted on the whole filter under a light microscope at 200x magnification (Dahl and Naustvoll, 2010). The detection limit for *Dinophysis* spp. was 40 cells l^{-1}. Samples of blue mussels (*Mytilus edulis*) were collected monthly between depths of 0.5 to 1.5 m. According to the EU regulations 853/2004, 854/2004, 2074/2005, 15/2011, the mussel samples were analyzed and DST was estimated at the Norwegian School of Veterinary Science (NMBU) using high-performance liquid chromatography (HPLC) (European Union Reference Laboratory for marine biotoxins, 2015).

The raw time series of *Dinophysis* spp. does not correlate well with DST because the low detection limit of 40 cells l^{-1} adds noise to the time series. For this reason, the time series of *Dinophysis* spp. were smoothed. Missing data during winter is filled with 0. We apply a maximum moving window filter and functional data analysis (Ramsay, 2006) to remove the noise while preserving the maximum values. The time series is smoothed in a β-spline function. The smoothness is controlled by knot spacing (or the number of knots) and the degree of the function. Here, we tested several knots spacing alongside the filter window size from 30 to 90 days with a degree of 4. The optimal choice was 60 days for window filter and knot spacing as it preserved the maximum values and maximized the correlation with DST.

2.3. Satellite data

Satellite SST (K) from 2006 to 2019 was obtained from the ESA CCI SST and C3S global SST Reprocessed product level 4, available on the Copernicus Marine Environment Monitoring Service (CMEMS). The product uses the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system (Good et al., 2020) that combines satellite AATSR, ATSR, SLSTR, and AVHRR sensors and in-situ observations to produce daily average SST at 0.05° spatial resolution (Merchant et al., 2019).

PAR ($Em^{-2}d^{-1}$) from 2006 to 2019 was retrieved from the GlobColour project, which is estimated based on MODIS, SeaWiFS, and VIIRS sensors (Frouin et al., 2003), and binned at an 8-day interval at a 4 km of spatial resolution.

Surface wind speed (ms^{-1}) from 2006 to 2019 was accessed from the IFREMER CERSAT Global Blended Mean Wind Fields reprocessed product retrieved from the CMEMS. Northward and Eastward surface wind speed is derived from scatterometers (ASCAT-A and ASCAT-B satellites), the SSMIS radiometers (F16, F17, F18, and F19 satellites) and the WindSat radiometer onboard the Coriolis satellite. All satellite observations are binned into a single product with a 6-hourly wind field at 0.25° spatial resolution.

All satellite data were reprojected to stereographic projection centered at 65°N, 7°E and 4 km spatial resolution using the nearest neighbor interpolation method. Because of the coarse spatial resolution, we have averaged all unmasked grid cells within the 7 × 7 grid around the station. Therefore, the environmental data assessed should not be interpreted as the conditions on the aquaculture farm location but rather the conditions of the surrounding area.

We have compared the satellite-averaged data with very few ($n<20$) in-situ observations of SST and surface wind speed and a few hundred observations of PAR (Fig. 2). These match-ups are far too few and spread over the entire observational period to justify any vicarious validation of the data quality used but indicate the relevance of the data selected for use in this study. For this validation exercise, the PAR in-situ data is measured at the meteorological station Holt located in Tromsø (60 km away), and it is obtained from the Landbruksmeteorologisk Tjeneste (LMT) from the Norwegian Institute of Bioeconomy Research (NIBIO). The in-situ SST was estimated as a 10 m deep temperature average from CTD casts in the Lyngen fjord, provided by the Norwegian Institute of Marine Research (IMR). Simultaneously with the CTD casts, in-situ wind speed is measured. In summary, satellite SST matches well with in-situ data, showing a Pearson correlation (R) of 0.96 and a mean absolute error (MAE) of 0.8 °C. Satellite PAR shows an $R = 0.65$ and an MAE = 12.7 $Em^{-2}d^{-1}$, tending to overestimate in-situ data. The eastward wind shows an $R = 0.7$ and MAE = 4.9 ms^{-1}, while the Northward wind shows no significant correlation and an MAE = 7.9 ms^{-1}. These differences can

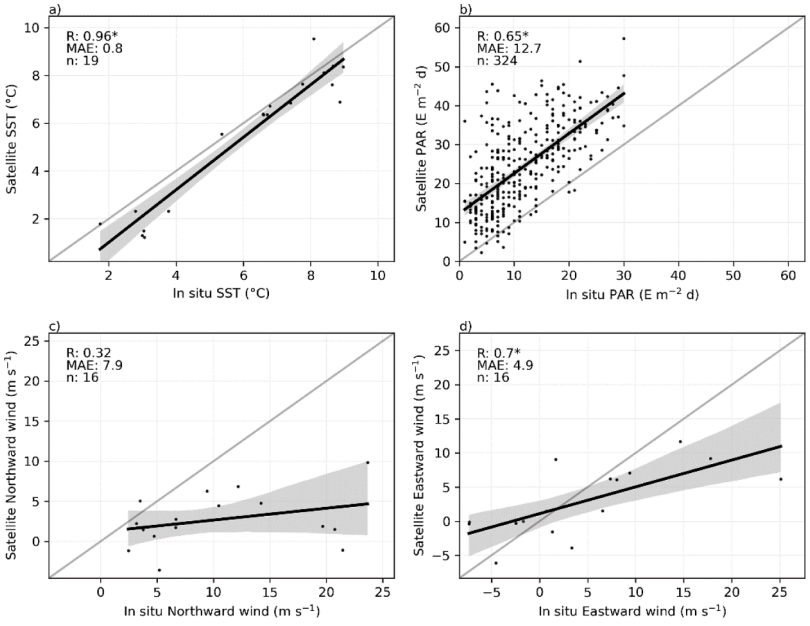


Fig. 2. Match-up between satellite estimates with in-situ data. Comparisons are shown for a) SST, b) PAR, c) Northward winds, and d) Eastward winds. Black line is the linear fit and the shaded area is the confidence interval. The R, MAE, and n are shown in the upper-left corner. The * denotes significant R at a 5% significance level.

be partly explained by the orography and orientation of the fjord, as presented above.

2.4. Environmental conditions assessment

We compare the probability density function of the environmental conditions during bloom (above 40 cells l^{-1}) and no bloom periods (below 40 cells $^{l-1}$). This threshold corresponds to the detection limit of *D. acuminata* cells abundance. Furthermore, we only consider values between the 2 and 98 percentiles to avoid outliers.

2.5. Forecast model calibration

The forecast model has a 2-step pre-processing: i) scaling of predictors between -1 and 1, and ii) extracting the principal components using principal components analysis (PCA). The principal components are then used to calibrate a SVM model. This study does not focus on an exhaustive evaluation of different machine learning algorithms, although a few initial tests were performed with Random Forest and AdaBoost without much success (not shown). The SVM is chosen for two reasons. First, the SVM has a powerful generalization and works well with small databases, showing better accuracy than Random Forest, Nearest Neighbor, Neural Network, and CART (Shao and Lunetta, 2012; Thanh Noi and Kappas, 2017). Second, the SVM is simple to calibrate as it relies only on three hyperparameters, the kernel function (and its parameters such as γ or degree), the penalty factor, and the ε. An explanation of these hyperparameters can be found in Mountrakis et al. (2011). The simplicity should allow a more straightforward adaptation of the prediction method to other locations in the future.

The smoothed cells abundance of *D. acuminata* on a log scale is used as the target to be predicted. The log scale is used to improve the prediction accuracy of the high values (e.g., cells $l^{-1} > 200$). We performed several tests without log scale, which could predict the seasonal development of the blooms but not the amplitude. The lack of skill in predicting the high values could be caused by the skewed distribution where the high values are few. SVM has high incidences of false negatives in unbalanced datasets for classification approaches (Wu and Chang, 2005; Yang et al., 2007), and it should also happen to regression applications.

The past measured cells abundance, SST, PAR, northward winds, and eastward winds, are used as predictors. All predictors are tagged and averaged as lag 0 from day 0 to day −13, lag 14 from day −14 to day −27, and lag 28 from day −28 to day −41, summing up 15 predictors (Fig. 3). Three strategies are used for training the forecast models: i) auto-regressive prediction, where only the past cells abundance is used as predictors; ii) environmental prediction, where only the environmental data is used as predictors; iii) combined prediction, where past cells abundance and environmental data are combined as predictors. The models are trained for predictions with lead times of 7, 14, 21, and 28 days.

The database was split into training (2006 - 2013) and testing (2014 - 2019) databases. Because this period is relatively short, we have used an incremental database approach to calibrate the SVM models and evaluate the accuracy of the predictions. For example, the SVM model was calibrated from 2006 to 2013 and tested in 2014. Then, 2014 is added to the training data, and the SVM model is calibrated again and tested in 2015. The process continues until 2019. This approach is more rigorous than randomly splitting the data because it is closely related to an operational level, where past data is used for training a model to predict future data.

For each SVM model and lead time, the initial training data (2006 - 2013) is used for tuning the model parameters, including the number of principal components used in the pre-processing and the SVM hyperparameters. The tuning is performed in a grid-search cross-validation using time series split and the coefficient of determination (R^2) as scoring criteria. Number of components tested were from 1 to the total number of predictors (e.g., 15 to the combined prediction). SVM hyperparameters tested were the linear and radial basis function (RBF), penalty factor from 0.1 to 100, ε from 0.1 to 10, and γ from 0.1 to 10. In most approaches and lead time, the linear kernel showed better results than the RBF kernel, so we opted for using only the linear kernel in all predictions.

2.6. Metric for assessment of forecast accuracy

Accuracy is measured by the R^2 and the MAE between the predicted and the reference (smoothed *D. acuminata*) values. They are estimated as follows:

$$R^2_{X,Y} = 1 - \frac{RSS}{TSS} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{t=0}^{n-1}|X_t - Y_t| \tag{2}$$

where RSS is the residual sum of squares, TSS is the total sum of squares, X and Y are the pairwise vectors of true and predicted values, t is the sampling date, and n is the sample size. For evaluating the errors of the interannual blooms amplitude, we computed MAE only using the data at the peak of the bloom, referred to as MAE_p. The peak date obtained from the smoothed *D. acuminata* is used for taking the reference and predicted values for computing the MAE_p.

Climatological and persistence forecasts are used as the benchmark for evaluating the usefulness of the forecast models in beating trivial predictors. The climatological forecast corresponds to the mean value of the smoothed *D. acuminata* estimated in the training dataset on the day of the year. Because the training data set is increasing with the years, the climatological forecast changes slightly for each year. The persistence

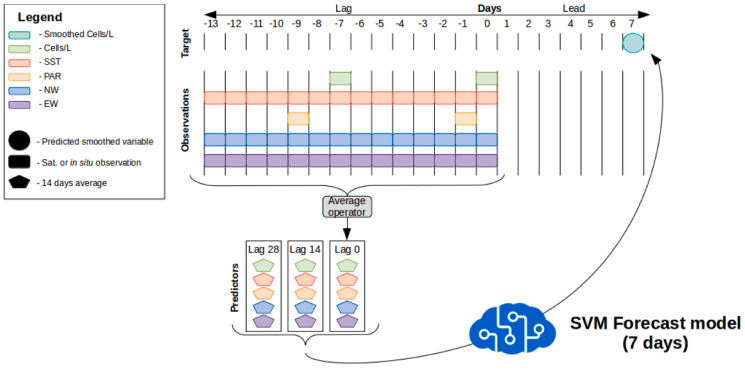


Fig. 3. Forecast model diagram. An example is given to the 7-day forecast of the combined model. The average from day 0 to lag −13 is called lag 0, and lag 14 and 28 follow the same method by averaging from −14 to −27 and from −28 to −41, respectively.

forecast is the last measured value at the start of the prediction. We present the skill score with respect to climatological (SS_c) and persistence (SS_p) forecasts (Murphy, 1992):

$$SS_c = 1 - \left(\frac{MAE_f}{MAE_{climatology}}\right) \quad (3)$$

$$SS_p = 1 - \left(\frac{MAE_f}{MAE_{persistence}}\right) \quad (4)$$

where MAE_f is computed using the forecast model, $MAE_{climatology}$ is computed using the climatological predictions, and $MAE_{persistence}$ is computed using the persistence predictions. Positive values of SS_c and SS_p indicate that the forecast model errors are lower than the climatology and persistence forecast.

For evaluating the forecast model skill in anticipating cell abundance above a standard level of hazardous concentration, we estimate the true positive rate (TPR) and false alarm rate (FAR) as follows:

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

$$FAR = \frac{FP}{TP + FP} \quad (6)$$

where TP is the true positive for values above 100 cells l^{-1}, FN is the false negative, and FP is the false positive. A TPR=1 means that all values above 100 cells l^{-1} were correctly detected, while a TPR=0 means that none of those values were detected. An FAR=1 means that all forecast values above 100 cells l^{-1} were incorrect, while a value of 0 means all forecast values were correct. Note that NFSA currently considers 1000 cells l^{-1} as the standard level of hazardous concentration for banning shellfish consumption. Since we are not able to estimate a robust TPR and FAR for 1000 cells l^{-1} because the test dataset only has 5 of 111 samples above this level, we resorted to estimating TPR and FAR for 100 cells l^{-1}.

The importance of all predictors was estimated by permutation (McGovern et al., 2019) during the testing process. Each predictor is separately permuted 100 times. The permutation shuffles the values so they correspond to the wrong dates, creating a distorted predictor version. The R^2 of the distorted version is computed and compared to the R^2 without being distorted. The R^2 decrease after distortion shows the importance of the feature permuted.

3. Results

3.1. Dinophysis spp. blooms and DST

In the records of *Dinophysis* spp. cell concentrations, the *D. acuminata* blooms once per year (Fig. 4) between the 22nd July and the 28th September, bearing concentrations up to 3954 cells l^{-1} (Table 1). The *D. acuminata* only reaches indicative hazard limit (cells $l^{-1} = 1000$) in 2006 and 2017. The *D. acuta* blooms are less frequent and occur in 7 out of the 13 years assessed. The *D. acuta* annual peak happens between 14th August and the 19th of October, and concentrations higher than the indicative hazard limit (cells $l^{-1} = 100$) occur in 2006, 2007, and 2019. After 2008 the *D. acuta* is less present and merely appears in low concentrations. The *D. norvegica* blooms once per year and typically between 10th August and the 5th of November, with one exception in 2014 when a small rise is detected in 9th of April. For all the study period, the *D. norvegica* has not reached concentrations above the indicative hazard limit (cells $l^{-1} = 4000$).

The DST time series exhibit two distinct periods. One before 2011 when DST is constantly detected and reaches concentrations above the hazard limit (μg $kg^{-1} = 160$) five times. Another period after 2011 when DST is less detected and concentration beyond the hazard limit is observed only in 2017. Before 2011, the three *Dinophysis* spp. can significantly explain DST, where *D. acuta* appears to be the most

Table 1
Date of the peak and the maximum cells abundance of all *Dinophysis* spp. blooms detected from 2006 to 2019 in the Lyngen fjord.

Year	*D. acuminata*		*D. acuta*		*D. norvegica*	
	Date	Cells l^{-1}	Date	Cells l^{-1}	Date	Cells l^{-1}
2006	07–22	3954	09–08	1325	09–15	664
2007	08–12	201	09–24	350	09–09	345
2008	08–20	261	09–04	59	09–18	400
2009	08–04	119			08–31	145
2010	08–08	899	10–11	31	08–25	755
2012	07–30	447			08–18	151
2013	08–24	85			08–22	87
2014a					04–09	60
2014b	09–11	307			08–10	17
2015	09–09	113			08–25	115
2016	08–24	163	08–14	16	11–05	36
2017	08–27	1868	10–19	21	09–01	76
2018	09–28	155			08–28	321
2019	08–28	549	09–27	126	08–26	110

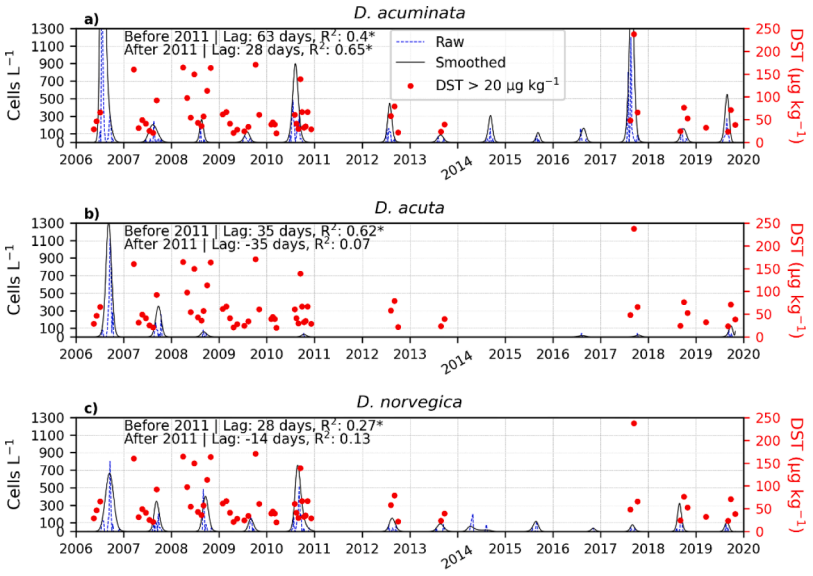


Fig. 4. The *D. acuminata* (a), *D. acuta* (b), *D. norvegica* (c), and DST (a, b, and c) time series. The Blue dashed line is the raw cells abundance, the black line is the smoothed cells abundance, and the red dots are the DST concentration (> 20 μg kg^{-1}). The R^2 is computed between each species smoothed time series and DST in a log scale from −90 to +90 lag days, and the lag with the highest R^2 is shown in the top left corner of each subplot. The * denotes significant R^2 at a 5% significance level. Note that R^2 is computed exclusively for the summer and autumn periods because we are interested in the Dinophysis spp. and DST association during their growth. DST may still be present in the winter and spring due to the DST accumulation in the previous year, but such variability is subjected to dilution processes rather than the Dinophysis spp. variability.

meaningful species as it explains 62% of DST variability. After 2011, *D. acuta* is present in low cells concentrations and cannot explain DST variability. In this period, *D. acuminata* is the only species explaining DST variability, showing an R^2 of 0.63 with the DST 28 days after the *D. acuminata*.

D. acuta has been reported as the main organism causing DST risks in Norway (Naustvoll et al., 2012), and this is consistent with our observations between 2006 and 2011. However, a substantial decline of *D. acuta* happens after this period. The same decline was also observed in southern Norway, although for the period between 1985 and 2009 (Naustvoll et al., 2012). Since *D. acuminata* is one of the three studies species showing a significant variability (R^2) with DST throughout the entire observational period, we have chosen to focus on assessing the environmental conditions and prediction skill of *D. acuminata* cells concentration.

3.2. Environmental conditions during D. acuminata blooms

SST exhibits the highest difference between bloom and no bloom periods compared to the other selected variables in this study (Fig. 5a). Blooms occur during higher SST, showing an average of 10.1 °C and ranging from 7.8 to 12.7 °C. Periods without bloom are, on average, at 5 °C and are highly widespread from 0.2 to 10.8 °C.

PAR shows slight (albeit significant) differences between bloom and no bloom periods (Fig. 5b). During bloom, PAR is on average 22.4 $Em^{-2}d^{-1}$ and ranges from 3.7 to 41.1 $Em^{-2}d^{-1}$, while periods without blooms, PAR is on average 17.3 $Em^{-2}d^{-1}$ and ranges from 4.3 to 42.3 $Em^{-2}d^{-1}$.

The winds show minor (albeit significant) discrepancies during bloom and no bloom periods, particularly with the northward component having weaker winds during high cells abundance (Fig. 5c, d).

3.3. Prediction skill of D. acuminata variability

The auto-regressive models have poor performances (Table 2), showing an R^2 varying from −0.12 to 0.3, MAE varying from 1.83 to 1.43 ln (cells l^{-1}), and SS_p varying from 0.23 to 0.29. Although the results are better than persistence, the model prediction is worse than the climatological forecast (SS_c is negative). The results are particularly poor at the start of the seasonal development of the bloom (Fig. 6a), when the predicted cells abundance is constant while the reference grows. The models show the worst results to estimate the amplitude of the blooms with a MAE_p varying from 2.85 to 1.2 ln(cells l^{-1}). The model shows a FAR inferior of 0.14 but a very low TPR.

The environmental forecast model (Fig. 6b) shows reasonable skill in predicting the cells abundance (Table 2). The R^2 varies from 0.37 to 0.5, MAE from 1.35 to 1.19 ln(cells l^{-1}), SS_p from 0.35 to 0.46, and SS_c from 0.05 to 0.16. However, the model presents two limitations. First, in the

Table 2
Accuracy results of auto-regressive (AR), environmental (ENV.), and combined (COM.) predictions with respectively lead of 7, 14, 21, and 28 days.

Model	Lead days	R^2	MAE	MAE_p	SS_p	SS_c	TPR	FAR
AR	7	0.3	1.43	1.2	0.29	-0.01	0.39	0
	14	0.14	1.59	1.52	0.24	-0.12	0.32	0.07
	21	0.01	1.71	2.01	0.23	-0.2	0.34	0.07
	28	-0.12	1.83	2.85	0.27	-0.29	0.29	0.14
ENV.	7	0.4	1.31	0.99	0.35	0.08	0.54	0.35
	14	0.5	1.19	1.02	0.43	0.16	0.51	0.3
	21	0.38	1.31	1.15	0.41	0.08	0.54	0.35
	28	0.37	1.35	1.34	0.46	0.05	0.46	0.39
COM.	7	0.55	1.13	0.62	0.44	0.2	0.56	0.04
	14	0.49	1.25	0.86	0.4	0.12	0.44	0.18
	21	0.46	1.26	0.87	0.43	0.11	0.54	0.21
	28	0.47	1.23	1.19	0.51	0.13	0.44	0.31

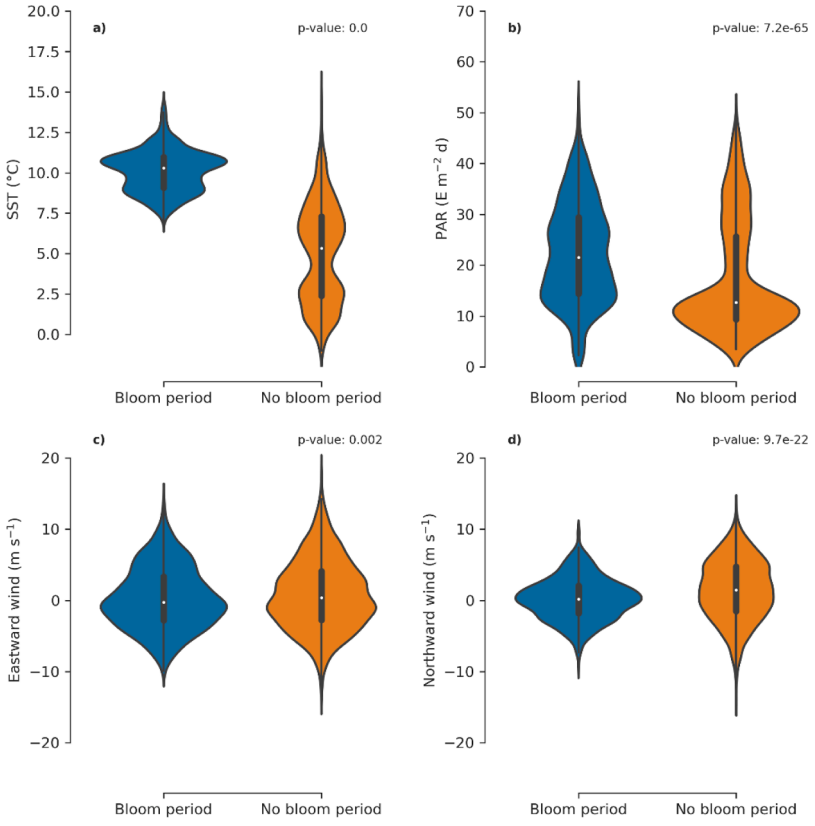


Fig. 5. SST (a), PAR (b), eastward wind (c), and northward wind (d) during *D. acuminata* bloom and no bloom periods. The p-value of a two-sided Mann-Whitney U rank test between bloom and no bloom periods is shown in the top-right corner of each subplot.

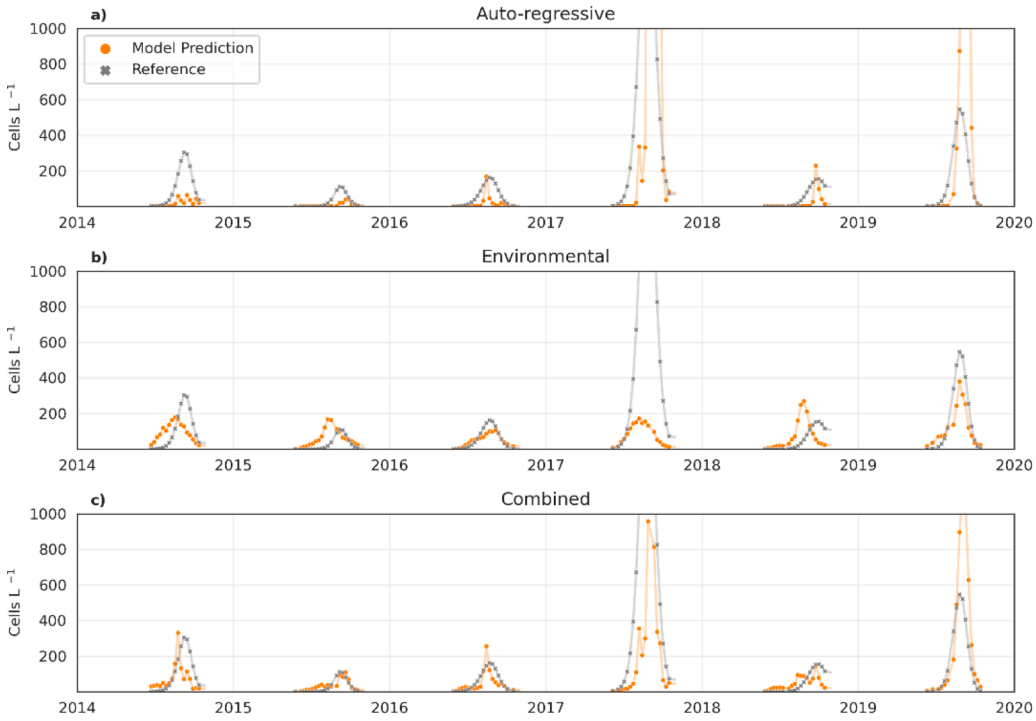


Fig. 6. SVM model forecast of *D. acuminata* smoothed cells abundance with 7 lead days for the a) Auto-regressive, b) Environmental, and c) Combined model. The gray line is the reference and the orange line is the forecast model prediction. Predictions with 14, 21, and 28 days lead time can be found in the supplementary material.

years of late *D. acuminata* blooms, (e.g., in 2014, 2015, and 2018; see Table 1), the environmental model predicts the beginning of *D. acuminata* bloom too early. Second, the environmental model has poor accuracy in predicting the amplitude of the peak of the bloom (i.e., the amplitude of the bloom predicted is similar for all years). For example, when *D. acuminata* reached hazardous concentrations in 2017, the environmental model predicted the peak concentration as less than 200 cells l^{-1}. Compared to the auto-regressive model, the environmental model shows a higher TPR up to 0.54. However, the better TPR comes with the cost of a higher FAR of at least 0.3.

The combined forecast model shows the best results. The R^2 varies from 0.46 to 0.55, MAE from 1.26 to 1.13 ln (cells l^{-1}), SS_p from 0.4 to 0.51, and SS_c from 0.11 to 0.2 (Table 2). Unlike in the environmental forecast, the combined model shows good skill in predicting the seasonal development by modulating reasonably well the bloom onset and amplitude (Fig. 6c). For example, the later blooms of 2014, 2015, and

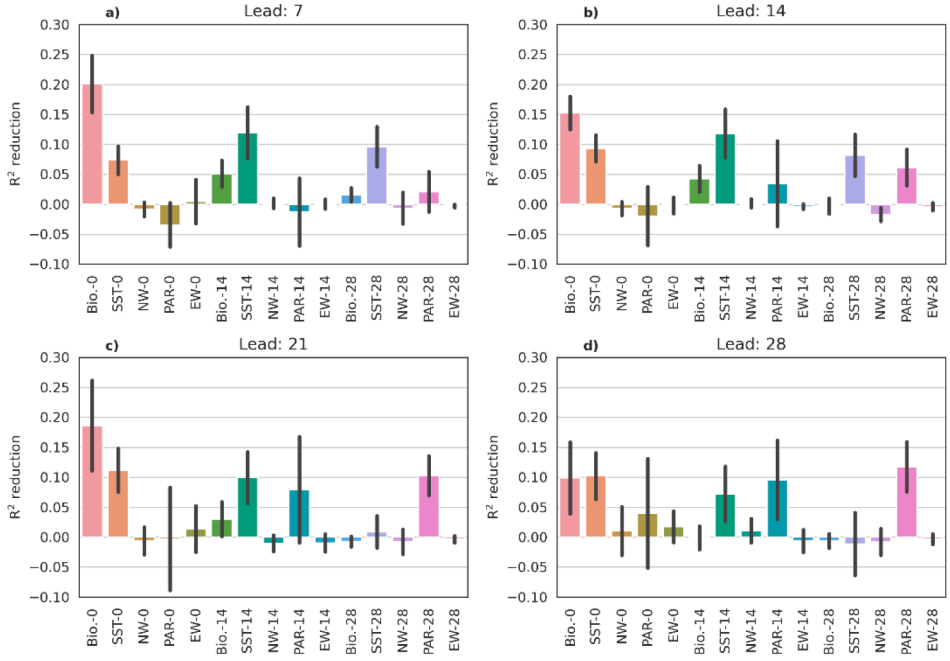


Fig. 7. The R^2 reduction after permutation of the combined model predictors in 7 (a), 14 (b), 21 (c), and 28(d) lead days. The x-axis is to the features names followed by the lag. The NW and EW denotes the northward and eastward wind components. The black lines in each bar indicate the standard deviation over the test years.

2018, and the largest event in 2017. Consequently, the combined model shows the lowest MAE_p among the three models, varying from 1.19 to 0.62 ln(cells l^{-1}). Furthermore, the combined model has the best balance between TPR and FAR, where TPR can reach up to 0.56 with an FAR = 0.04.

Among all predictors used in the combined model, the past *D. acuminata* cells abundance (referred to as Bio.-lag) has the highest importance (Fig. 7). After permutation, the Bio.−0 causes an R^2 drop from 0.1 to 0.2, while Bio.−14 and Bio.−28 have little importance and causes a drop in R^2 lower than 0.05. SST is the second most important predictor, causing an R^2 drop higher than 0.05 in most of the lag and lead times. The only exception is the SST-28 at leads 21 and 28. PAR influence increases with lead days and show high influence on leads 21 and 28, where PAR-14 and PAR-28 cause a drop in R^2 of 0.09 and 0.12. Last, both wind components have not shown any importance to the predictions. The permutation importance analysis confirms that the model used effectively both environmental factors and past cells abundance to compute its prediction.

4. Discussion

4.1. SST influence on D. acuminata and prediction skill

From the Skagerrak strait to the west coast of southern Norway, the *D. acuminata* has been characterized as a species that can grow for the whole productive season from March to early December (Dahl and Johannessen, 2001; Lindahl et al., 2007; Naustvoll et al., 2012; Séchet et al., 1990). Although *D. acuminata* is mixotrophic and can grow by preying on other organisms, it cannot grow in darkness (Kim et al., 2008). In the Lyngen fjord, the polar night (i.e., without any daylight) is from the 18th November to the 23rd January, which would make the theoretical growing season last 10 months. However, from our observations, the *D. acuminata* blooms have been restricted to a shorter period from July to September in the region over the past 14 years. This may relate to water temperature. *D. acuminata* only grows in waters warmer than 8 °C, both in laboratory experiments (Basti et al., 2018) and field observations (Alves-de-Souza et al., 2019; Boivin-Rioux et al., 2022; Hattenrath-Lehmann et al., 2015; Hoshiai et al., 2003). In the Lyngen fjord, waters warmer than 8 °C typically occur from early June to mid-October. Seasonal temperature variability entails the favorable growth period of *D. acuminata* blooms and modulates its seasonal cycle. As a result, the SST variability shows the highest importance among all environmental factors in the combined forecasting model.

4.2. Using past cell abundance to mitigate the lack of critical predictors

Although the statistical model fed with environmental data (i.e., SST, PAR, and surface wind speed) outperforms climatological predictors (see Table 2), it fails to predict the interannual variability of the bloom onset and its amplitude. No environmental parameters assessed correlate to the amplitude of interannual variability of *D. acuminata* (e.g., higher summer SST does not correlate significantly to more intense blooms of *D. acuminata*). Other environmental factors not assessed in this study (because good quality time series are unavailable) may help condition the interannual variability of bloom onset and amplitude.

D. acuminata is often associated with water stratification (Reguera et al., 2012). The surface wind speed controls the mixed layer depth and the water stratification, so we use satellite winds. However, the winds show no importance in forecasting *D. acuminata* cells abundance. It may be that *D. acuminata* is influenced by local wind occurring in the closed, north-south oriented narrow fjord system rather than by the large-scale flow estimated in the averaged 7 × 7 grid cells. The poor relationship between the satellite and in-situ wind speeds indicates the disagreement between local and large-scale data (see Section 2.3). Furthermore, river runoff can influence the local water stratification in the fjord, but data is unavailable near the study region. Therefore, water stratification could be a critical predictor for *D. acuminata* in the Lyngen fjord, and the accuracy of the forecast should improve if local observations were available.

Another essential predictor is predator and prey interactions, for which observations are also lacking. On the one hand, prey availability can allow *D. acuminata* to grow 5 times faster than in the absence of prey (Kim et al., 2008). On the other hand, *D. acuminata* growth can be reduced by grazing, such as by copepods that can prey on *D. acuminata* at a rate of up to 47 cells female^{-1} h^{-1} (Jansen et al., 2006). In this regard, prey availability and grazing pressure could explain part of the interannual variability (onset and amplitude) of the blooms. We have no information about predator and prey interactions of *D. acuminata* in the Lyngen fjord, and such estimations may enhance predictions of cell abundance.

While the lack of aforementioned potential predictors reduces the environmental model accuracy, combining the past cells abundance with the available environmental factors may alleviate this issue. A few studies have demonstrated that auto-regressive models of HAB (e.g., chl-a or cells l^{-1}) can have acceptable accuracy (Chen et al., 2015; Cruz et al., 2021; Velo-Suárez and Gutiérrez-Estrada, 2007). Although our auto-regressive model is insufficient to forecast *D. acuminata* in the Lyngen fjord, we could improve the forecast of bloom onset and amplitude, as well as R^2, MAE, MAE_p, SS_c, and SS_p, when the measured cell abundance is combined with SST, PAR, and winds observations (the combined model). The combined model gives the past cell abundance (Bio.−0) the highest importance among all predictors. The Bio.−0 should correct the present cell abundance conditions and help predict the bloom onset and amplitude.

4.3. Making the D. acuminata forecast operational

A forecast model of *D. acuminata* can provide improved public recommendations and early warning to the end users. Here we propose a forecast model of *D. acuminata* cell abundance that could be integrated into a so called “ready-set-go” framework (Vitart and Robertson, 2018). For example, the forecast model can allow monitoring authorities to tailor the observational strategy, such as frequency and eventual more rapid analysis turnover to detect and follow hazardous conditions earlier.

For using our model operationally, a few considerations regarding the predictors variables should be addressed. The SST data used are reprocessed and unavailable in near real-time, making it necessary to change the dataset. We foresee that as a minor limitation because satellite SST data is available with a one-day lag, and the quantity evolves slowly at the spatial scale used. Furthermore, the cost of installing a weather station that measures in-situ SST is marginal compared to the potential gain.

The main limitation of the combined model is the dependence on in-situ cell abundance being available to be used in the predictions. Counting cells includes the time for water sampling, transport to the laboratory, filtration, species identification, actual counting and quality control. The whole process may take a few days and make the sub-seasonal forecasts less feasible. The current logistic process of the NFSA monitoring is sampling and sending it to specialized laboratories on Mondays. The laboratories report the data back on the following Thursdays, creating a delay of 3 days in an eventual operational forecasting service. This delay is not substantial to the forecasts from 14 to 28 days, but it decreases the 7 days forecast to 4 days. In order to provide operational forecasts for the coming week, the time for analyzing algae cell abundance should be improved. Depending on methods for

analyzing, available expertise, and equipment, the results of *D. acuminata* can be available in at least 12 h.

4.4. Adapting the forecast model to other regions

The statistical forecasting model is only valid for the Lyngen fjord aquaculture site since it is calibrated to the local measurements of *D. acuminata* and environmental conditions. The behavior of the variables used can vary strongly from one location to the other. Nonetheless, our method can be re-calibrated to other regions, which would be of great interest due to the widespread presence of *D. acuminata* and its severe impact on shellfish (Reguera et al., 2012). Thanks to the continuous public monitoring of HAB and an increasing number of aquaculture farm sites in northern Norway, we anticipate that the data needed to make the method applicable to several other locations will increase in the coming years.

This study shows that coarse remote sensing observations can be used to forecast the abundance of *D. acuminata* cells. Satellite remote sensing observations are quasi-homogeneous worldwide and could be considered in other regions where in-situ observations are limited. Which remote sensing variables should be used may depend on the region. For example, we foresee SST being of great relevance in high-latitude regions where it can limit the growing season of *D. acuminata*. However, its relevance may diminish in regions where temperatures are constantly above 8 °C or have low variability. Different regions may consider other features, such as salinity (Godhe et al., 2002) and chl-a (Hattenrath-Lehmann et al., 2013). It could also be considered data from models such as water stratification and nutrients (Ajani et al., 2016).

Statistical modeling methods such as machine learning are data-driven, which limits their general applicability. Even though SVM is known for having good performance with small databases (Shao and Lunetta, 2012; Thanh Noi and Kappas, 2017), the model calibration still needs enough data representing different ranges of *D. acuminata* cell abundance in both training and test datasets. For example, Lyngen fjord has years of weak and strong blooms in both datasets, so the model can learn how to forecast both conditions and have its performance evaluated in both situations. We have adapted the method to three other farms in northern Norway (not shown). While one farm showed similar results to the Lyngen fjord, the other two locations could not be well calibrated because high cell abundance was only present in the most recent years (the test dataset). Thus, our proposed method is most useful for farm locations with long time series undergoing different *D. acuminata* bloom conditions.

5. Conclusion

This study shows that DST in the Lyngen fjord in northern Norway is controlled mostly by *D. acuminata*, which peaks every year during the summer when the surface temperature is above 7.8 °C. A forecast machine learning model, based on SVM, is trained with in-situ *D. acuminata* data and environmental data derived from satellite measurements (SST, PAR, and wind speed). The model can predict reasonable well the *D. acuminata* blooms' seasonal development and amplitude up to 28 days ahead. In the future, the calibrated model could contribute to a ready-set-go framework of *D. acuminata* HAB in the aquaculture farm located in the Lyngen fjord. For example, the model could be used for delineating periods of increased observation frequency to monitor the development of local HAB events. Finally, how the model can be adapted and calibrated to other aquaculture locations remains to be tested. Such adaptation will depend on the local *D. acuminata* variability and the environmental parameters available from satellite observations.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The codes and input data are available on GitHub: https://github.com/nansencenter/Forecasting-harmful-algae-blooms–application-to-Dinophysis-acuminata-in-northern-Norway.

Acknowledgements

The algae toxin measurement data were obtained with permission from the monitoring program of algae toxins in mussels and dietetic advice to the public (https://www.matportalen.no/verktoy/blaskjellvarsel/), operated by the Norwegian Food Safety Authority (NFSA). The in-situ PAR data is provided by the Landbruksmeteorologisk Tjeneste (LMT) from the Norwegian Institute of Bioeconomy Research (NIBIO). ES is a holder of an institute research fellowship (INSTSTIP) funded by the basic institutional funding through Norwegian Research Council (#318085). FC acknowledges the Trond Mohn Foundation under project number: BFS2018TMT01. JB acknowledges the NFR Climate Futures (309562). We thank Stephen Outten for the English review.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.hal.2023.102442.

References

Ajani, P., Larsson, M.E., Rubio, A., Bush, S., Brett, S., Farrell, H., 2016. Modelling bloom formation of the toxic dinoflagellates Dinophysis acuminata and Dinophysis caudata in a highly modified estuary, south eastern Australia. Estuar. Coast Shelf Sci. 183, 95–106. https://doi.org/10.1016/j.ecss.2016.10.020.

Alves-de-Souza, C., Iriarte, J.L., Mardones, J.I., 2019. Interannual Variability of dinophysis acuminata and protoceratium reticulatum in a Chilean Fjord: insights from the realized niche analysis. Toxins (Basel) 11 (19). https://doi.org/10.3390/toxins11010019.

Basti, L., Suzuki, T., Uchida, H., Kamiyama, T., Nagai, S., 2018. Thermal acclimation affects growth and lipophilic toxin production in a strain of cosmopolitan harmful alga Dinophysis acuminata. Harmful Algae 73, 119–128. https://doi.org/10.1016/j.hal.2018.02.004.

Boivin-Rioux, A., Starr, M., Chassé, J., Scarratt, M., Perrie, W., Long, Z., Lavoie, D., 2022. Harmful algae and climate change on the Canadian East Coast: exploring occurrence predictions of Dinophysis acuminata, D. norvegica, and Pseudo-nitzschia seriata. Harmful Algae 112, 102183. https://doi.org/10.1016/j.hal.2022.102183.

Bouquet, A., Laabir, M., Rolland, J.L., Chomérat, N., Reynes, C., Sabatier, R., Felix, C., Berteau, T., Chiantella, C., Abadie, E., 2022. Prediction of Alexandrium and Dinophysis algal blooms and shellfish contamination in French Mediterranean Lagoons using decision trees and linear regression: a result of 10 years of sanitary monitoring. Harmful Algae 115, 102234. https://doi.org/10.1016/j.hal.2022.102234.

Chen, Q., Guan, T., Yun, L., Li, R., Recknagel, F., 2015. Online forecasting chlorophyll a concentrations by an auto-regressive integrated moving average model: feasibilities and potentials. Harmful Algae 43, 58–65. https://doi.org/10.1016/j.hal.2015.01.002.

Cruz, R.C., Reis Costa, P., Vinga, S., Krippahl, L., Lopes, M.B., 2021. A review of recent machine learning advances for forecasting harmful algal blooms and shellfish contamination. J. Mar. Sci. Eng. 9, 283. https://doi.org/10.3390/jmse9030283.

Dahl, E., Johannessen, T., 2001. Relationship between occurrence of Dinophysis species (Dinophyceae) and shellfish toxicity. Phycologia 40, 223–227. https://doi.org/10.2216/i0031-8884-40-3-223.1.

Dahl, E., Naustvoll, L., 2010. Filtering – semitransparent filters for quantitative phytoplankton analysis, in: Karlson, B., Cusack, C., Bresnan, E. (Eds.), Microscopic and Molecular Methods For Quantitative Phytoplankton Analysis. IOC Manuals and Guides 55, Paris, pp. 37–39.

European Union Reference Laboratory for marine biotoxins, 2015. EU-harmonised standard operating procedure for determination of Lipophilic marine biotoxins in molluscs by LC-MS/MS.

Fernandes-Salvador, J.A., Davidson, K., Sourisseau, M., Revilla, M., Schmidt, W., Clarke, D., Miller, P.I., Arce, P., Fernández, R., Maman, L., Silva, A., Whyte, C., Mateo, M., Neira, P., Mateus, M., Ruiz-Villarreal, M., Ferrer, L., Silke, J., 2021. Current status of forecasting toxic harmful algae for the north-east Atlantic shellfish aquaculture industry. Front. Mar. Sci. 8 https://doi.org/10.3389/fmars.2021.666583.

Frouin, R., Franz, B., Wang, M., 2003. Algorithm to estimate PAR from SeaWiFS data Version 1.2-Documentation. NASA Tech Memo 206892, 46–50.

Giesen, R.H., Andreassen, L.M., Oerlemans, J., Van Den Broeke, M.R., 2014. Surface energy balance in the ablation zone of Langfjordjokelen, an arctic, maritime glacier in northern Norway. J. Glaciol. 60, 57–70. https://doi.org/10.3189/2014JoG13J063.

Godhe, A., Svensson, S., Rehnstam-Holm, A.S., 2002. Oceanographic settings explain fluctuations in Dinophysis spp. and concentrations of diarrhetic shellfish toxin in the plankton community within a mussel farm area on the Swedish west coast. Mar. Ecol. Prog. Ser. 240, 71–83. https://doi.org/10.3354/meps240071.

Good, S., Fiedler, E., Mao, C., Martin, M.J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J., Worsfold, M., 2020. The current configuration of the OSTIA system for operational production of foundation sea surface temperature and ice concentration analyses. Remote Sens. (Basel) 12, 1–20. https://doi.org/10.3390/rs12040720.

Hattenrath-Lehmann, T.K., Marcoval, M.A., Berry, D.L., Fire, S., Wang, Z., Morton, S.L., Gobler, C.J., 2013. The emergence of Dinophysis acuminata blooms and DSP toxins in shellfish in New York waters. Harmful Algae 26, 33–44. https://doi.org/10.1016/j.hal.2013.03.005.

Hattenrath-Lehmann, T.K., Marcoval, M.A., Mittlesdorf, H., Goleski, J.A., Wang, Z., Haynes, B., Morton, S.L., Gobler, C.J., 2015. Nitrogenous nutrients promote the growth and toxicity of Dinophysis acuminata during estuarine bloom events. PLoS ONE 10, e0124148. https://doi.org/10.1371/journal.pone.0124148.

Hegstad, S.M.K., 2014. Post-glacial Sedimentary Processes and Slope Instabilities Off Nordnesfjellet, Lyngenfjorden, Northern Norway. The Arctic University of Norway.

Hoshiai, G., Suzuki, T., Kamiyama, T., Yamasaki, M., Ichimi, K., 2003. Water temperature and salinity during the occurrence of Dinophysis fortii and *D. acuminata* in Kesennuma Bay, northern Japan. Fisheries Sci. 69, 1303–1305. https://doi.org/10.1111/j.0919-9268.2003.00760.x.

Jakowczyk, M., Stramska, M., 2014. Spatial and temporal variability of satellite-derived sea surface temperature in the Barents Sea. Int. J. Remote Sens. 35, 6545–6560. https://doi.org/10.1080/01431161.2014.958247.

Jansen, S., Riser, C.W., Wassmann, P., Bathmann, U., 2006. Copepod feeding behaviour and egg production during a dinoflagellate bloom in the North Sea. Harmful Algae 5, 102–112. https://doi.org/10.1016/j.hal.2005.06.006.

Karlson, B., Andersen, P., Arneborg, L., Cembella, A., Eikrem, W., John, U., West, J.J., Klemm, K., Kobos, J., Lehtinen, S., Lundholm, N., Mazur-Marzec, H., Naustvoll, L., Poelman, M., Provoost, P., De Rijcke, M., Suikkanen, S., 2021. Harmful algal blooms and their effects in coastal seas of Northern Europe. Harmful Algae 102, 101989. https://doi.org/10.1016/j.hal.2021.101989.

Kim, S., Kang, Y., Kim, H., Yih, W., Coats, D., Park, M., 2008. Growth and grazing responses of the mixotrophic dinoflagellate Dinophysis acuminata as functions of light intensity and prey concentration. Aquat. Microb. Ecol. 51, 301–310. https://doi.org/10.3354/ame01203.

Lindahl, O., Lundve, B., Johansen, M., 2007. Toxicity of Dinophysis spp. in relation to population density and environmental conditions on the Swedish west coast. Harmful Algae 6, 218–231. https://doi.org/10.1016/j.hal.2006.08.007.

Martino, S., Gianella, F., Davidson, K., 2020. An approach for evaluating the economic impacts of harmful algal blooms: the effects of blooms of toxic Dinophysis spp. on the productivity of Scottish shellfish farms. Harmful Algae 99, 101912. https://doi.org/10.1016/j.hal.2020.101912.

McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G.E., Elmore, K.L., Homeyer, C. R., Smith, T., 2019. Making the black box more transparent: understanding the physical implications of machine learning. Bull. Am. Meteorol. Soc. 100, 2175–2199. https://doi.org/10.1175/BAMS-D-18-0195.1.

Merchant, C.J., Embury, O., Bulgin, C.E., Block, T., Corlett, G.K., Fiedler, E., Good, S.A., Mittaz, J., Rayner, N.A., Berry, D., Eastwood, S., Taylor, M., Tsushima, Y., Waterfall, A., Wilson, R., Donlon, C., 2019. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. Sci. Data 6, 1–18. https://doi.org/10.1038/s41597-019-0236-x.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66, 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001.

Murphy, A.H., 1992. Climatology, persistence, and their linear combination as standards of reference in skill scores. Weather Forecast 7, 692–698. https://doi.org/10.1175/1520-0434(1992)007<0692:CPATLC>2.0.CO;2.

Naustvoll, L.-J., Gustad, E., Dahl, E., 2012. Monitoring of Dinophysis species and diarrhetic shellfish toxins in Flødevigen Bay, Norway: inter-annual variability over a 25-year time-series. Food Addit. Contamin. Part A 29, 1605–1615. https://doi.org/10.1080/19440049.2012.714908.

Olsen, J.A., 2015. Sedimentære Avsetningsmiljøer Og Deglasiasjonshistorie i Ersfjorden, Kvaløya, Troms fylke (Master Thesis). The Arctic University of Norway.

Park, M.G., Kim, M., 2010. Prey specificity and feeding of the thecate mixotrophic dinoflagellate fragilidium duplocampanaeforme. J. Phycol. 46, 424–432. https://doi.org/10.1111/j.1529-8817.2010.00824.x.

Pettersson, L.H., Pozdnyakov, D., 2013. Monitoring of Harmful Algal Blooms. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68209-7.

Ramsay, J.O., 2006. Encyclopedia of Statistical Sciences. Functional Data Analysis. John Wiley & Sons, Inc, Hoboken, NJ, USA, pp. 1–8. https://doi.org/10.1002/0471667196.ess3138.

Reguera, B., Velo-Suárez, L., Raine, R., Park, M.G., 2012. Harmful Dinophysis species: a review. Harmful Algae 14, 87–106. https://doi.org/10.1016/j.hal.2011.10.016.

Ruiz-Villarreal, M., García-García, L.M., Cobas, M., Díaz, P.A., Reguera, B., 2016. Modelling the hydrodynamic conditions associated with Dinophysis blooms in Galicia (NW Spain). Harmful Algae 53, 40–52. https://doi.org/10.1016/j.hal.2015.12.003.

Schmidt, W., Evers-King, H., Campos, C., Jones, D., Miller, P., Davidson, K., Shutler, J., 2018. A generic approach for the development of short-term predictions of Escherichia coli and biotoxins in shellfish. Aquac. Environ. Interact. 10, 173–185. https://doi.org/10.3354/aei00265.

Séchet, V., Safran, P., Hovgaard, P., Yasumoto, T., 1990. Causative species of diarrhetic shellfish poisoning (DSP) in Norway. Mar. Biol. 105, 269–274. https://doi.org/10.1007/BF01344296.

Setälä, O., Sopanen, S., Autio, R., Erler, K., 2009. Grazing and food selection of the calanoid copepods Eurytemora affinis and Acartia bifilosa feeding on plankton assemblages containing Dinophysis spp. Boreal Environ. Res. 14, 837–849.

Shao, Y., Lunetta, R.S., 2012. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS J. Photogramm. Remote Sens. 70, 78–87. https://doi.org/10.1016/j.isprsjprs.2012.04.001.

Silva, E., Counillon, F., Brajard, J., Korosov, A., Pettersson, L.H., Samuelsen, A., Keenlyside, N., 2021. Twenty-One Years of Phytoplankton Bloom Phenology in the Barents, Norwegian, and North Seas. Front. Mar. Sci. 8, 1–16. https://doi.org/10.3389/fmars.2021.746327.

Sverdrup, H.U., 1953. On Conditions for the Vernal Blooming of Phytoplankton. ICES J. Mar. Sci. 18, 287–295. https://doi.org/10.1093/icesjms/18.3.287.

Thanh Noi, P., Kappas, M., 2017. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors 18, 18. https://doi.org/10.3390/s18010018.

Velo-Suárez, L., Gutiérrez-Estrada, J.C., 2007. Artificial neural network approaches to one-step weekly prediction of Dinophysis acuminata blooms in Huelva (Western Andalucía, Spain). Harmful Algae 6, 361–371. https://doi.org/10.1016/j.hal.2006.11.002.

Vitart, F., Robertson, A.W., 2018. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. NPJ Clim. Atmos. Sci. 1, 3. https://doi.org/10.1038/s41612-018-0013-0.

Wu, G., Chang, E.Y., 2005. KBA: kernel boundary alignment considering imbalanced data distribution. IEEE Trans. Knowl. Data Eng.

Yang, J., Yan, R., Hauptmann, A.G., 2007. Adapting SVM classifiers to data with shifted distributions. Seventh IEEE Int. Confer. Data Mining Workshops (ICDMW 2007). IEEE 69–76. https://doi.org/10.1109/ICDMW.2007.37.