Prediction of Harmful Algae Blooms Impacting Shellfish Farms in Norway

Edson Filisbino Freire da Silva

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2023

UNIVERSITY OF BERGEN

Prediction of Harmful Algae Blooms Impacting Shellfish Farms in Norway

Edson Filisbino Freire da Silva

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 30.11.2023

Year: 2023

Title: Prediction of Harmful Algae Blooms Impacting Shellfish Farms in Norway

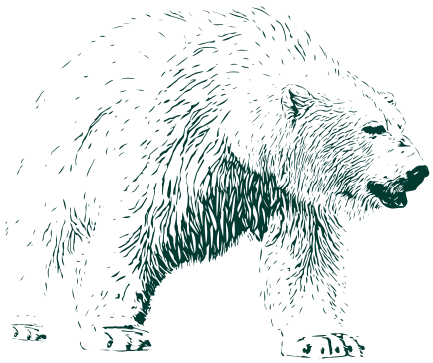
Name: Edson Filisbino Freire da Silva

Print: Skipnes Kommunikasjon / University of Bergen

Scientific environment

I have carried out this thesis at the Nansen Environmental and Remote Sensing Center (NERSC) in Bergen, Norway, where I was part of the Climate Dynamics and Prediction Group (CDPG). The work was supported by an institute research fellowship (INST-STIP), funded by the basic institutional funding through the Research Council of Norway (#318085). I have been involved in the Bjerknes Centre for Climate Research, the Bjerknes Climate Prediction Unit (BCPU), and the center for research innovation Climate Futures, contributing to the ocean sustainable food node. I have attended courses provided by the University of Bergen, the University of Svalbard, and the Research School on Changing Climates in the Coupled Earth System (CHESS). I have also participated in workshop training, including the TRIATLAS – TAPIOCA – CANEMS – UFPE Summer School and Workshop on the functioning of Tropical and South Atlantic marine ecosystems – from physics to top predators and fisheries, Brazil, 2022; and the ESA/NERSC Advanced Ocean Synergy Training Course, Mozambique - South Arica, 2023. I have participated in several symposiums and conferences. A poster presentation in Workshop: Multi-annual to Decadal Climate Predictability in the North Atlantic-Arctic Sector, Denmark - Norway, 2022; a poster presentation in the Arctic Science Summit Week (ASSW), Norway, 2022; an oral presentation in The Symposium on Decadal Variability of the North Atlantic and its Marine Ecosystems: 2010-2019, Norway, 2022; a poster presentation in The TREVOR symposium, England, 2023. I have presented my work in popular outreach programs, including a meeting with potential users of harmful algae prediction, such as Eide Fjordbruk SA; I have also presented my work in the Bjerknes Climate Podcast: Predicting algae blooms - a new tool in our arsenal.

≡



Acknowledgements

I am deeply grateful for the guidance and mentorship of my advisors François Counillon and Julien Brajard provided during these 3 years of Ph.D., including in writing scientific papers, learning about climate and machine learning, and supporting applications to projects and symposiums. My sincere appreciation also goes to my co-advisor Noel Keenlyside for helping with the 1^{st} paper and giving network through workshops and meetings.

This work would not have been possible without the assistance of NERSC researchers. I am thankful to Lasse H. Pettersson for providing valuable insights on harmful algae in Norway; Anton A. Korosov and Annette Samuelsen for assisting with the 1^{st} paper; and Jiping Xie for making TOPAZ data available. I would like to offer my most profound appreciation to the NERSC administrative and technical staff. In particular, I would like to express my heartfelt thanks to Marian Tvedt and Kristiyana Lolova. In summary, I am deeply grateful to NERSC for supporting the development of my research.

I am also grateful to Frode Vikebø for giving feedback on needed improvements in harmful algae monitoring and prediction; Lars Naustvoll for helping with the methods and discussion of the 2^{nd} and 3^{rd} papers; Merete Hestdal for facilitating and providing harmful algae data.

Finally, I extend my deepest gratitude to my family and friends. Especially, I am thankful to my beloved wife, Olenka Monteiro, for her love, patience, and sacrifices.

Edson Filisbino Freire da Silva
Bergen, 15.08.2023

"I love fools' experiments. I am always making them"

Charles Darwin

Abstract

Harmful algae blooms (HABs) cause severe damage to the ecosystem and human health, and have significant economic impacts on shellfish farms. HAB prediction models have become increasingly popular because they can help stakeholder to take mitigation actions and reduce economic loss. Few studies have attempted to predict toxic algae species related to shellfish contamination because the time extent of data is limited and modeling the environmental response of specific taxa is complex. However, toxic algae monitoring programs have now been running for several years and have produced large datasets of toxic algae. Combined with long-time series observations by satellites and model reanalysis, we can now calibrate prediction models for toxic algae affecting shellfish farms. This thesis calibrates machine learning models to predict toxic algae impacting shellfish farms in Norwegian coastal waters for the first time. It is conducted by combining toxic algae data from the Norwegian Food Safety Authority with satellite observations of Chl-a concentration, Suspended Particulate Matter (SPM), Sea Surface Temperature (SST), Photosynthetically Active Radiation (PAR), and wind speed, as well as model reanalysis data of Mixed Layer Depth (MLD) and Sea Surface Salinity (SSS). Paper I demonstrates that the blooms phenology has a strong interannual variability in the North, Norwegian, and Barents Seas, which is related to the variability of the environmental ocean and atmospheric factors (SST, MLD, SPM, and winds). It implies that these variables are potential predictors for blooms in the region. Paper II exhibit that a Support Vector Machine (SVM) model can predict the presence probability of eight toxic algae on the Norwegian coast using SST, PAR, SSS, and MLD. The models can also predict the probability of harmful levels for *Alexandrium* spp., *Alexandrium tamarense*, *Dinophysis acuta*, and *Azadinium spinosum*. It can produce a climatological overview of the HABs along the Norwegian coast and provide monitoring and prediction applications. Paper III extends the SVM application to the prediction of *D. acuminata* abundance in a sub-seasonal range (7 -28 days) when fed with the current and past *D. acuminata* abundance, SST, PAR, and wind speed. The sub-seasonal forecast model is developed for the Lyngenfjord in northern Norway as a proof of concept. The probability estimates in Paper II and the sub-seasonal forecast of *D. acuminata* abundance in Paper III are two complementary approaches. The first is employable in the entire coast even where algae monitoring is unavailable, while the latter requires tuning to specific aquaculture farms and can achieve refined prediction. Since the SVM models are fed with data commonly available worldwide, they are portable to other regions where data from harmful algae monitoring programs are available.

Sammendrag

Skadelig algeoppblomstring (HAB) forårsaker alvorlig skade på økosystemet og menneskers helse, samt økonomiske konsekvenser for skalldyr og oppdrettsanlegg. Utviklingen av HABs prediksjonsmodeller har blitt økenende populart fordi det kan hjelpe næringsliv til å håndtere risko bedre. Få har prøvde å varsle alger som fører skalldyrforurensning, på grunn av det er begrenset data av individuelt taxa og at faktors som påvirker det varieres strort. Overvåkingsprogrammer for giftige alger har nå blitt etablert i flere år, som gjøre at det er nå mullig å lykkes. Denne avhandlingen kalibrere for først gang maskinlæring models som kan brukes til å varlse HABs som påvirker skalldyroppdrett i det norske kystvannene. Studien bruker giftige algedata fra Norskehysten, satellittobservasjoner av Chl-a-konsentrasjon, suspendert partikulært materiale (SPM), havoverflatetemperatur (SST), fotosyntetisk aktiv stråling (PAR), og vindhastighet, og modellreanalysedata av blandingsdybde (MLD) og havoverflatesaltholdighet (SSS). Paper I- viser at våroppblomstringer har en stor år-til-år variasjon i Nord-Norge og i Barentshavet som er påvirket av miljø (vann og vind). Det fører til at det er prediktabilitet i vår fokus område. Paper II- viser at sannsynligheten får å måle en av det åtte giftige alger som fines i Norge kan varsles og er knyttet til variasjonen av SST, PAR, SSS og MLD. Modellen kan også varsle sannsynligheten å måle skadelige konsentrasjon for *Alexandrium* spp., *Alexandrium tamarense*, *Dinophysis acuta* og *Azadinium spinosum*. Den Support Vector Machines (SVM) kan kartlegge risiko langs det Norskekysten og kan bli brukt både på overvåkning og varsling. Paper III vises at den SVM metode kan varlse *D. acuminata* konsentrasjon ved å bruke nåværende og tidligere *D. acuminata* konsentrasjon, SST, PAR og vindhastighet. Den pilot studie fokuserer for Lyngen i Nord-Norge. Sesongvarlsing utvilet i Paper III og sansynlighet model utviklet i Paper II er svært komplementært. Det første kan fører til mer nøyaktige predisjkon men krever at overvåkning data er tilgjengelig i lokasjonet, mens den andre kan brukes selv om ingen algaer data finnes i nærethen. Metoder kan lett tilpasses for andre område i verden så langt en algær overvåkning er etablert.

Outline

This thesis consists of an introductory part, data and methods, and three scientific papers. Chapter 1 gives an introduction to harmful algae blooms in Norway, prediction methods for harmful algae blooms, motivation and objective. Chapter 2 describes the study region, data used, and the Support Vector Machine method. Chapter 3 introduces the three papers. Chapter 4 summarizes and concludes the three papers. Chapter 5 includes the papers:

1. Edson Silva, François Counillon, Julien Brajard, Anton Korosov, Lasse H. Pettersson, Annette Samuelsen, Noel Keenlyside, (2021) *Twenty-One Years of Phytoplankton Bloom Phenology in the Barents, Norwegian, and North Seas*, Frontiers in Marine Science **8**

2. Edson Silva, Julien Brajard, François Counillon, Lasse H. Pettersson, Lars Naustvoll, (2023) *Probabilistic Models for Harmful Algae: Application to the Norwegian Coast*, (In review, Environmental Data Science, July 2023)

3. Edson Silva, François Counillon, Julien Brajard, Lasse H. Pettersson, Lars Naustvoll, (2023) *Forecasting Harmful Algae Blooms: Application to Dinophysis acuminata in Northern Norway*, Harmful Algae **126**

Contents

1 Introduction

Harmful algae blooms (HABs) cause severe damage to the ecosystem and human health, as well as economic loss to shellfish farms. For this reason, algae biomass, chlorophyll-a concentration, cell abundance of toxic species, and toxins are constantly monitored in coastal waters. Even though such monitoring helps prevent human poisoning incidents with some degree of success, the economic impact on aquaculture farms continues to be significant. Prediction models of HABs have become increasingly popular as they can allow end-users to plan and undertake mitigation actions. The following sub-sections introduce the definitions of algae and harmful blooms, summarize the taxa of HABs impacting shellfish farms in Norway, describe recent developments in HAB predictions, and state the motivation and objective of the current thesis.

1.1 Harmful Algae Blooms

Algae are eukaryotic cell organisms capable of photosynthesis and are primarily composed of single-cell organisms, with a few multi-cellular species reaching up to 60 m long. Prokaryotic cyanobacteria are also commonly called blue-green algae due to their importance as the basis of the eukaryotic algae evolution (*Brodie and Lewis*, 2007). When aquatic environments go through rapid changes in their limiting factors, such as an increase in nutrients (e.g., NO_3^- and PO_4^{3-}) in nutrient-depleted waters or an increase in light in high-latitude regions, single-cell algae biomass increases quickly and leads to what is called a bloom. HABs refer to blooms that cause harm to human health and wildlife or cause significant economic loss. For example, HABs can kill valuable farmed seafood in aquaculture farms or lead to extensive closure periods when farms cannot sell their products to prevent human poisoning.

Some algae species produce toxins that can cause poisoning when high concentrations are generated during blooms. A common category of HABs is those related to poisoning incidents involving shellfish consumption. These filter-feeding organisms can retain toxins when filtering waters contaminated with a high abundance of toxic algae. Common shellfish poisoning syndromes and toxins are Diarrhetic Shellfish Toxins (DST), Paralytic Shellfish Toxins (PST), Amnesic Shellfish Toxins (AST), Azaspiracid shellfish toxins (AZA), and yessotixins (YTX) (*Pettersson et al.*, 2001). These toxins are produced by several dinoflagellates and diatoms in northern Europe (*Karlson et al.*, 2021). Most of the toxins comprising the DST are produced by dinoflagellates such as *Dinophysis* Ehrenberg, 1839 and *Prorocentrum* Ehrenberg, 1834. PSTs are produced by dinoflagellates as *Alexandrium* Halim, 1960, *Gymnodinium catenatum* H.W.Graham, 1943 and *Pyrodinium bahamense* Plate, 1906. AST producers include some members of diatoms *Pseudo-nitzschia* H. Peragallo in H. Peragallo & M. Peragallo, 1900 and *Nitzschia* A.H. Hassall, 1845. AZA is produced by dinoflagellates of the genus *Azadinium* Elbrächter & Tillmann, 2009 and *Amphidoma* Stein, 1883. YTX producers are dinoflagellates as *Lingulodinium polyedra* (F.Stein) J.D.Dodge, 1989, *Protoceratium reticulatum* (Claparède & Lachmann) Bütschli, 1885, and *Gonyaulax spinifera* (Claparède & Lachmann) Diesing, 1866.

Several studies have estimated the economic impact of HABs related to shellfish farming and consumption. Public health care is one of the main impacts. For example, in France, between 1996 and 2010, there were 561 reported outbreaks of diseases attributed to shellfish consumption, of which 26% were caused by DST, resulting in 179 hospitalizations (*Baron et al.*, 2012). In a study conducted in Canada in 1995, the total cost of medical expenses and income loss due to 525 reported PST, DST, and ciguatera poisoning amounted to approximately USD $670,000 (*Mardones et al.*, 2020). In Chile, the cost of hospitalizations due to consuming contaminated mussels (mainly PST and DST) has sharply increased from 2004 to 2018, reaching USD $93,119 in 2018 (*Mardones et al.*, 2020). Note that the costs vary depending on the study method and estimated period.

Monitoring programs are commonly employed to prevent human poisoning. It includes estimating the abundance of toxic algae in shellfish farms and toxin measurements in shellfish flesh. Consequently, the requirement for monitoring increases the costs of operating shellfish farms. For example, the cost of the Olympic Regional Harmful Algal Bloom (ORHAB) program on the Washington Pacific coast was estimated at USD $150,000 per year (*Huppert and Trainer*, 2014). In the United States of America, from 1987 to 1992, the average cost of the monitoring and management of HABs was estimated at USD $2,088,885 per year (*Anderson et al.*, 2000). In France during the 1990s, the French Phytoplankton Monitoring Network (REPHY) cost was estimated at USD $1,200,000 per year (*Anderson et al.*, 2001). In the same period in Denmark, monitoring cost was estimated at USD $400,000 per year, of which USD $112,000 were for algae analysis and USD $288,000 for toxin estimations (*Anderson et al.*, 2001).

When the monitoring programs detect dangerous levels of toxic algae or toxins, the farms are not allowed to sell their products. Loss of revenue occurs when the shellfish cannot be commercialized. While most shellfish eventually depurate the toxins naturally and reach safe conditions once the HAB ends, the depuration is impossible when the shellfish are harvested and processed. Furthermore, some commercial contract stipulates supplying shellfish at a designated time. Any unharvested product during a HAB will become mismatched with market demand and probably remain unsold (*Martino et al.*, 2020). In Maine, USA, a two-month closure period is estimated to cause an economic loss of up to USD $10.4 million (*Jin et al.*, 2020). In coastal Massachusetts in 2005, USA, an *Alexandrium catenella* (Whedon & Kofoid) Balech, 1985 bloom resulted in a loss of shellfish fishery revenues of USD $18 million (*Jin et al.*, 2008). In Scotland, on average, DST is estimated to cause a 15% loss in total annual production (*Martino et al.*, 2020). While these estimate only accounts for revenue loss, the overall cost is expected to be higher as social welfare, including consumer and producer surpluses, should also be accounted for (*Jin et al.*, 2020).

1.2 Shellfish-related HABs in Norway

Shellfish farms in Norway mainly comprise blue mussels farming (*Mytilus edulis* Linnaeus, 1758). Data available through the Norwegian Directorate of Fisheries (www.fiskeridir.no) reveals that in 2021, farmed shellfish reached up to 2,200 metric tons, of which the majority (2,163) were blue mussels, equivalent to a gross sale of NKr $24,474,000. The 2021 production accounted for 43 companies in operation, employing 114 people.

Recent human poisoning incidents caused by HABs are rare in Norway (*Karlson*

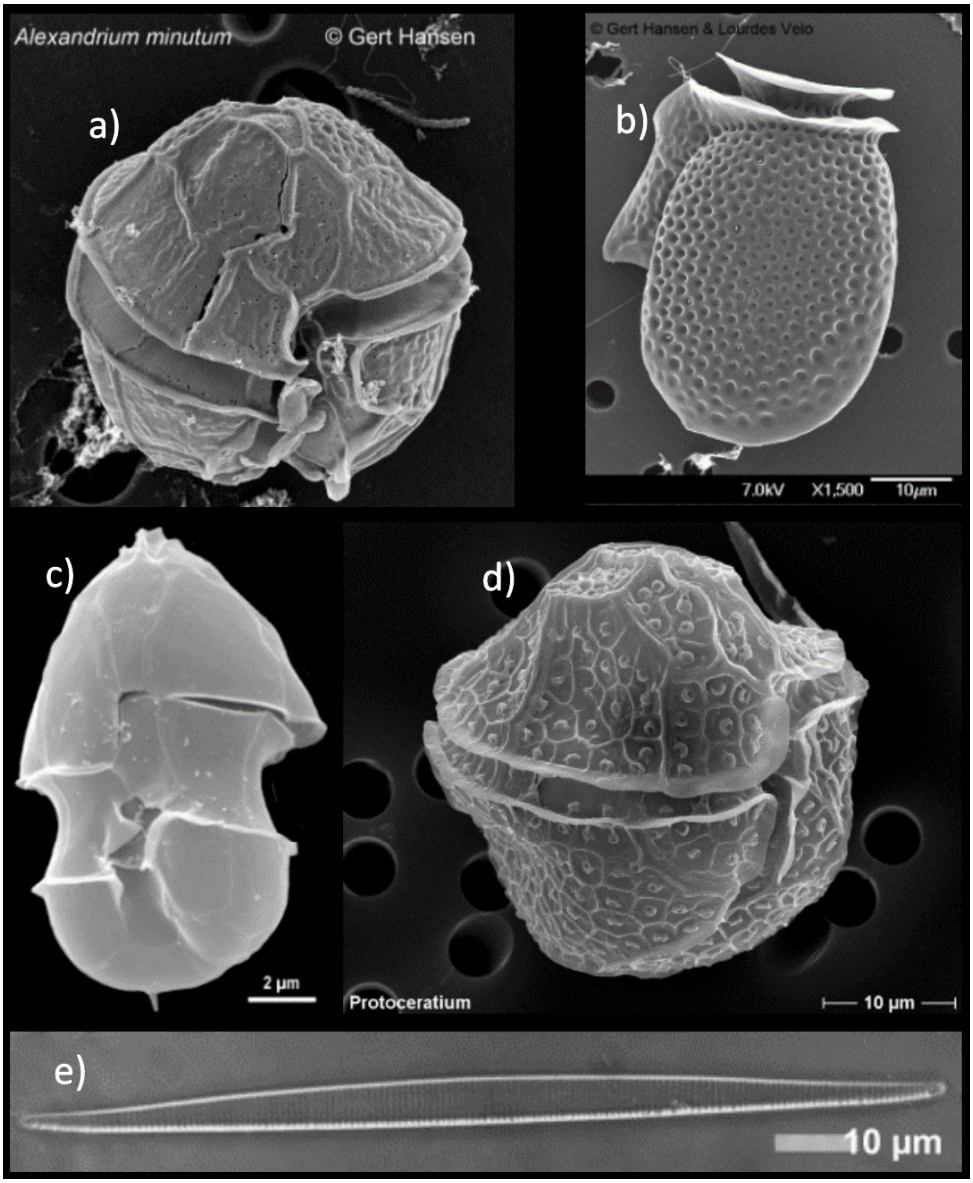


Figure 1.1: Main toxic genera monitored in blue mussels farms in Norway. The genera represented are a) Alexandrium spp., b) Dinophysis spp., c) Azadinium spp., d) Protoceratium spp., and e) Pseudo-nitzschia spp. Note it is a representation of the genus and the species may not exactly match the ones described in (Table 1.1) The authors of the pictures are a-b-c) Gert Hansen, d) Mona Hoppenrath, and e) Nina Lundholm. Picture e) has been cropped from the original file. All pictures were downloaded from www.marinespecies.org and are licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License (https://creativecommons.org/licenses/by-nc-sa/4.0).

et al., 2021), summing up to only three confirmed incidents since the 1980s. In 1981, eight people were intoxicated by PST after eating shellfish (*Langeland et al.*, 1984). In the two remaining events, poisoning was related to the consumption of brown crabs (*Cancer pagurus* Linnaeus, 1758) probably contaminated after feeding on shellfish. In 2002, from 200 to 300 people were poisoned by DST after consuming contaminated brown crabs (*Pettersson and Pozdnyakov*, 2013). In 2005, two people were hospitalized after eating brown crabs contaminated with AZA (*Karlson et al.*, 2021).

The number of poisoning incidents is small thanks to the regional monitoring done by the Norwegian Food and Safety Authority (NFSA). The NFSA monitoring estimates the abundance of toxic species in the surface waters of shellfish farms every week and the concentration of toxins in the farmed blue mussels' flesh every month. The toxic algae and their respective toxins include *Dinophysis acuminata* Claparède & Lachmann, 1859, *Dinophysis acuta* Ehrenberg, 1839, and *Dinophysis norvegica* Claparède & Lachmann, 1859, and their DST; *Alexandrium* spp. and *Alexandrium tamarense* (Lebour) Balech, 1995, related to PST; *Pseudo-nitzschia* spp. and AST; *Azadinium spinosum* Elbrächter & Tillmann, 2009 and AZA, and *P. reticulatum* and its YTX (Figure 1.1). By establishing safe concentration levels of both cells and toxins (Table 1.1), the NFSA provides public advice and restricts the sale of contaminated shellfish through closure periods.

*Table 1.1: Maximum safe concentrations of algae abundance and toxins. Algae are estimated in the water column and toxins are estimated in the blue mussels flesh. The thresholds are defined by NFSA to ban selling and to advise the public to avoid eating blue mussels. The term "to be considered" means that alga is monitored, but no specific threshold is established for banning the sale. * Means the threshold is for hygienic assessment only. ** Means the respective threshold is only for ban selling when detected in three consecutive weeks.*

Taxa or Toxin	Maximum Concentration
A. tamarense	200 Cells.L^{-1}
Alexandrium spp	to be considered
Pseudo-nitzschia spp.	1,000,000* Cells.L^{-1}
P. reticulatum	1,000* Cells.L^{-1}
D. acuta	200/100** Cells.L^{-1}
D. acuminata	1,000 Cells.L^{-1}
D. norvegica	4,000 Cells.L^{-1}
Azadinium spp.	to be considered
DST	160 µg.kg^{-1}
PST	400 µgSTX.kg^{-1}
YTX	3,750 µg.kg^{-1}
AST	20,000 µg.kg^{-1}
AZA	160 µg.kg^{-1}

Data on algae abundance and toxin concentration from 2006 to 2020 (courtesy of NFSA) reveal that the samples regularly exceed safety levels along the Norwegian coast. The relative number of samples containing harmful levels of algae abundance reaches up to 5% per year (Figure 1.2a), and the relative number of samples exceeding safe levels of toxin concentration up to 20% per year (Figure1.2b). Each sample with exceeded safe levels means selling the blue mussels was banned for a certain period until safe conditions were regained. Although human intoxication caused by HABs is rare in

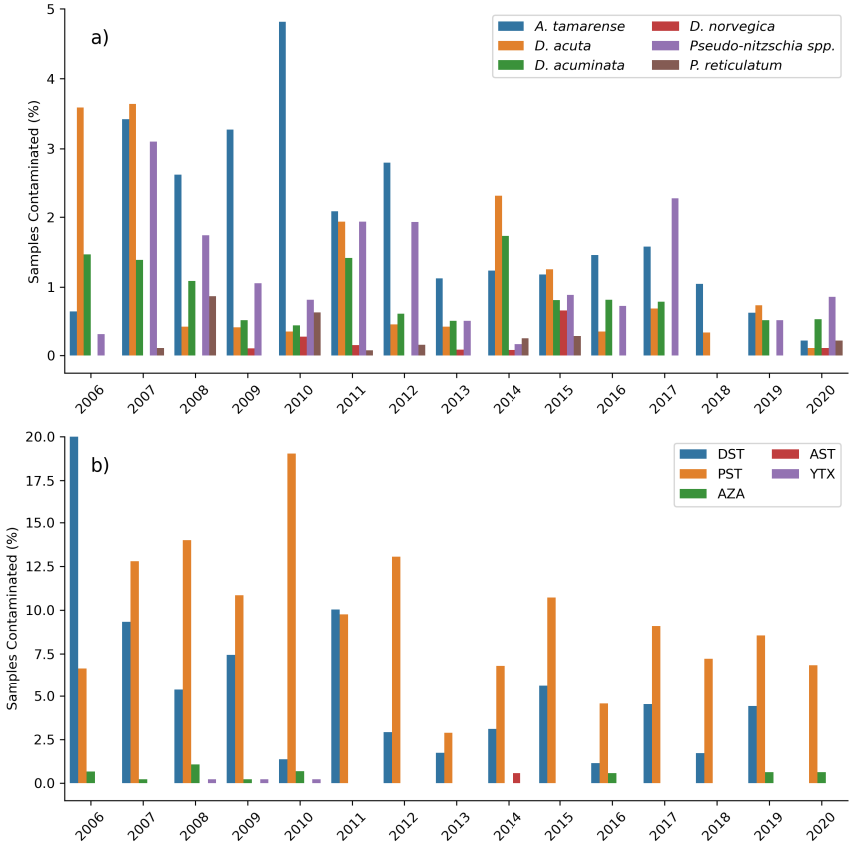


Figure 1.2: The relative number of samples containing harmful algae (a) and toxins (b) exceeding safe levels from 2006 to 2020. Raw data is provided by NFSA. Note that interannual variability is susceptible to changes in the farms' location.

Norway, the economic impact caused by loss of revenue and continuous monitoring is probably significant. Unfortunately, no estimates of economic losses caused by HAB in shellfish farms in Norway were found.

The most common algae exceeding safe levels in Norway are *A. tamarense*, *D. acuta*, and *D. acuminata* (Figure 1.2). Consequently, PST and DST are the most common toxins exceeding safe levels. AZA is the third most common toxin reaching hazardous levels as it has been detected in almost all years from 2006 to 2019. YTX has only exceeded safe levels in a few samples from 2008 to 2010, and its producer - *P. reticulatum* - has been detected exceeding safe levels in a few years. AST hazardous concentrations were only detected in 2014, although *Pseudo-nitzschia* spp. exceeding safe levels occur every year.

The phenology of toxic algae along the Norwegian coast has been the subject of few studies. On the west coast, the highest abundance levels of *D. acuta* and *D. norvegica* were observed during the autumn season, while *D. acuminata* highest abundance levels were found in spring and early summer (*Séchet et al.*, 1990). On the south coast, *D.*

acuminata and *D. norvegica* has been predominantly abundant from March to December, and *D. acuta* has occurred typically in late summer and autumn. The *Pseudo-nitzschia* spp. are commonly found along all the Norwegian coastal waters during spring and autumn blooms (*Hasle et al.*, 1996). High cells abundance of *P. reticulatum* have been detected on the west coast from June to August (*Braarud*, 1976) and during early summer in the southern region (*Aasen et al.*, 2005). *A. spinosum* has been detected from the southern to western Norwegian coast (*Tillmann et al.*, 2018).

Empirical evidence of environmental factors influencing shellfish-related HABs in Norway is lacking. Still, laboratory experiments can indicate how HABs respond to environmental factors in isolation:

- Light: it is utilized for photosynthesis and strongly influences algae growth. From 0 to 359 μmol.m^{-2}s^{-1}, higher levels of light cause increased growth rate in *Pseudo-nitzschia australis* Frenguelli, 1939 and *Pseudo-nitzschia turgidula* (Hustedt) G.R.Hasle (*Bill et al.*, 2012), both species belonging the genus *Pseudo-nitzschia*. *A. spinosum* has stable growth rate from 100 to 250 μmol.m^{-2}s^{-1}, but it can significantly reduce when light is below 50 μmol.m^{-2}s^{-1} (*Jauffrais et al.*, 2013). *D. acuta* can grow with minimum 10 μmol.m^{-2}s^{-1}, but levels higher than 370 μmol.m^{-2}s^{-1} cause photodamage and reduces its growth rate (*García-Portela et al.*, 2018).

- Nutritional strategies: relatively low phosphate concentration can lead to low abundance but high cell toxicity of *P. australis*, *Pseudo-nitzschia pungens* (Grunow ex Cleve) G.R.Hasle, 1993 and *Pseudo-nitzschia fraudulenta* (Cleve) Hasle, 1993 (*Lema et al.*, 2017). Phosphorus and nitrogen limited environments reduce the growth rate of *P. reticulatum* (*Guerrini et al.*, 2007), while selenium and iron can increase its growth rate (*Mitrovic et al.*, 2004). *Alexandrium* spp. and *Dinophysis* spp. exhibits more complex nutritional strategies. *Alexadrium* spp. can utilize various inorganic and organic nutrient sources, as well as feed other organisms (*Anderson et al.*, 2012). *Dinophysis* spp. can grow by feeding on other algae, such as *Mesodinium rubrum* (Lohmann, 1908), and appropriate their plastids for photosynthesis (*Kim et al.*, 2008).

- Salinity: Variations in salinity affect algae by inducing osmotic stress, creating ion stress through the unavoidable absorption or loss of ions, and altering the cellular ionic ratios due to selective mechanisms (*Kirst*, 1990). Since all taxa monitored along the Norwegian coast are marine, most are adapted to more saline waters. *D. acuminata* and *D. acuta* can grow at salinity higher than 10 PSU and demonstrate reduced grow in fresher waters (*Rial et al.*, 2023). *Pseudo-nitzschia seriata* (Cleve) H.Peragallo, 1899 grow only at salinity higher than 15 PSU, yielding maximum growth rates at 30 PSU (*Weber et al.*, 2021). *P. reticulatum* grows in salinity ranging from 22 to 42 PSU – with an optimal growth between 25 to 30 PSU. *A. spinosum* grow in salinity from 30 to 40 PSU (*Jauffrais et al.*, 2013). The salinity tolerance among species of *Alexandrium* spp. varies widely, including species adapted to fresh and saline waters (*Klemm et al.*, 2022). Regarding *A. tamarense*, higher encystment is observed at 30 PSU (*Nagai et al.*, 2004).

- Temperature: Algae present distinct temperature-related traits that allow them favorable growth rates to different temperature ranges (*Thomas et al.*, 2012). *D.*

acuminata and *D. acuta* can grow at temperatures ranging from 8 to 32° C (*Basti et al.*, 2018) and from 12 to 25.5° C (*Rial et al.*, 2023), respectivelly. *A. tamarense* grow in temperatures from 5 to 24° C, reaching maximum values at 14° C (*Nagai et al.*, 2004). Although *P. seriata* is typically considered a psychrophilic species, some strains demonstrated successful growth at temperatures higher than 15° C (*Fehling et al.*, 2004). *P. reticulatum* highest growth rates are around 15° C and it fails to grow in temperatures higher than 26° C (*Guerrini et al.*, 2007; *Röder et al.*, 2012). *A. spinosum* grow from 10 to 26° C, reaching maximum growth rates between 18° C and 22° C.

1.3 HAB Prediction Applications

Developing methods for predicting HABs is in increasing demand by stakeholders in aquaculture farming. A recent update of the aquaculture farms stakeholders' demands – produced jointly by the Food and Agriculture Organization of the United Nations (FAO), UNESCO's Intergovernmental Oceanographic Commission (IOC), and International Atomic Energy Agency (IAEA) (*FAO et al.*, 2023) – advocate the need for HAB forecasting (of at least 2 to 3 days) that can serve as an early warning system. HAB prediction can reduce the related costs of HABs for the aquaculture industry by allowing farmers to harvest their product before toxin outbreaks, reduce HAB monitoring and management costs, improve business planning and investment decisions, and increase consumer confidence (*Jin et al.*, 2020). For example, farmers can adjust their supply demands accordingly to the risk of HAB and avoid harvesting contaminated shellfish, helping reduce the loss of revenue. Specifically in Norway, assessing the likelihood of HABs is a demand for optimizing the monitoring program (Frode Vikebø, personal communication). Monitoring resources can be relocated to periods when the risk of HAB increases, helping safeguard public health.

A standard method to predict the evolution of HABs is particle-tracking models based on Lagrangian dispersion models (*Fernandes-Salvador et al.*, 2021). Particle-tracking models are run once a HAB is detected by satellite or in-situ observations, which feeds the model to predict the HAB dispersion from simulated currents. Particle tracking models have been operational in several regions, including Ireland, Scotland, Spain, and Portugal (*Fernandes-Salvador et al.*, 2021). In Norway, particle tracking models have been employed to support aquaculture farms for the 2019 *Chrysochromulina leadbeateri* Estep, Davis, Hargreaves & Sieburth, 1984 bloom (*John et al.*, 2022), as well as for lice dispersion (*Asplin et al.*, 2020). However, particle-tracking models can only infer the spatial evolution of the HAB and cannot predict their start. As such, their applicability depends heavily on an accurate HAB detection program. Their success also depends on the accuracy of simulated ocean currents, which are often poor in coastal regions (*Mourre et al.*, 2018). Finally, particle-tracking models only consider the physical transport, and not the underlying environmental mechanisms triggering the bloom.

Another method for predicting HABs is using machine learning models (*Fernandes-Salvador et al.*, 2021). It consists of training models to predict the HABs using a set of environmental conditions that can trigger blooms. Unlike particle-tracking models, machine learning models aim to forecast HABs without their prior detection. Many machine learning techniques have been tested (*Cruz et al.*, 2021). Common techniques em-

ployed for HAB predictions include auto-regressive integrated moving average (ARIMA), Support Vector Machine (SVM), Random Forest (RF), Bayesian Networks (BN), and Artificial Neural Networks (ANN). These methods are run in nowcast mode – using current environmental conditions – to predict the current HAB status, or in forecast mode – e.g., using present, past or future forecasted environmental conditions – to predict the future of HAB.

Although there is a growing number of studies on machine learning models for HAB prediction, few studies have targeted HABs impacting shellfish farms. In a review on machine learning applications for HAB prediction *Cruz et al.* (2021), 27 studies have been published, and most (18) target high biomass blooms (HBB) using Chl-a concentration or cyanobacteria abundance as targets. HBBs are most relevant for fish farms, while for shellfish farms the forecast of specific toxic taxa is necessary as they can be harmful even at low concentrations (Table 1.1). Here is a summary of the studies that developed machine learning predictions of HAB impacting shellfish farms:

- *González Vilas et al.* (2014) calibrate SVM models to predict the presence ($>100\,\text{Cells.L}^{-1}$) and bloom conditions ($>100{,}000\,\text{Cells.L}^{-1}$) of *Pseudo-nitzschia* spp. in the Galician rias, using as predictors the day of the year, temperature, salinity, upwelling indices, and bloom occurrence in previously weeks.

- *Guallar et al.* (2016) calibrates an ANN model to predict the presence and abundance of *Pseudo-nitzschia* spp. in Alfacs Bay from one to two weeks ahead. The predictors are river runoff, bottom temperature, salinity, wind speed, atmospheric pressure, and *Pseudo-nitzschia* spp. abundance of previous weeks.

- *Velo-Suárez and Gutiérrez-Estrada* (2007) calibrates an ANN model to predict *D. acuminata* abundance one week ahead using the past five weeks' estimations of *D. acuminata* abundance as predictors.

- Two recent studies – not included in the *Cruz et al.* (2021) review – were developed for the Canadian east coast (*Boivin-Rioux et al.*, 2021, 2022). The studies target *D. acuminata*, *D. norvegica*, *P. seriata*, and *A. catenella* using generalized additive mixed models (GAMMs) fed with temperature, salinity, and wind speed. However, they focus on the evolution of the HABs in a future climate rather than forecasts to support shellfish farms.

A key challenge for forecasting shellfish-related HABs is that it requires a large volume of observations for each toxic algae. Unlike Chl-a concentration, the estimation of algae abundance to a genus level cannot be acquired by remote sensing or automated in-situ sensors. The algae abundance data for specific taxa come from survey campaigns, including water sampling and laboratory analysis. Consequently, data on algae abundance for specific taxa is often too few to calibrate machine learning models. A secondary challenge is that the algae abundance estimations are usually not combined with measurements of key predictors collocated during overlapping periods for training the models.

Thanks to the sustained effort by public agencies, we are now in a configuration where both specific taxa abundance and environmental data are long enough to train such machine learning models. National programs for monitoring shellfish toxins have

become a common practice and produced a large volume of toxic algae data (often publicly available) covering extensive areas for long periods. Additionally, the monitoring programs are often regulated and coordinated, meaning one can relate variability in the abundance of toxic algae to environmental conditions without facing the risk of artifacts caused by changes in the monitoring program standards. For example, the Norwegian monitoring of toxins has been operational since 1992, and NFSA hold consistent weekly records of all shellfish farms along the coast since 2006. Besides, some environmental conditions can be provided by remote sensing or model reanalysis in long enough time series to feed the machine learning models (see Sections 2.3 and 2.4),

1.4 Motivation

HABs related to shellfish poisoning are a risk to human health and potentially cause severe economic impacts to shellfish farmers in Norway. Predicting algae related to these HABs can guide stakeholders to execute mitigation plans and avoid economic losses. Forecasting methods must be accurate, reliable, and flexible enough to be employed or adapted to several regions for broad applicability.

1.5 Objective

The present thesis aims to calibrate prediction models of HABs impacting shellfish farms in Norway. The objective is divided into the following sub-objectives:

- Produce prediction models ready to be employed and tested in the shellfish farms in Norway.
- Evaluate the suitability of environmental data from remote sensing and model reanalysis to calibrate machine learning models for HAB prediction.
- Develop calibration methodology that can be applicable to other areas.

2 Data and Methods

2.1 Study Region

The Norwegian coastline is encompassed by several water bodies: the Skagerrak Strait, the North Sea, the Norwegian Sea, and the Barents Sea (Figure 2.1). The circulation in this area is mainly influenced by two significant current systems: the Norwegian Atlantic Current (NwAC) and the Norwegian Coastal Current (NCC). The NwAC is an extension of the North Atlantic Current, which flows between the Faroe Islands and Scotland. It continues its northward path along the Norwegian Continental Shelf break, reaching the Barents Sea (*Eldevik et al.*, 2009; *Furevik et al.*, 2002). In contrast, the NCC flows from the Skagerrak strait, running along the coast and heading northward towards the Barents Sea. The NCC differs significantly from the NwAC as it carries fresher waters from the land inflows, the Baltic Sea, and the North Sea.

Norwegian coastal waters extend from the sub-Arctic to Arctic regions and cover various environmental conditions. In northern Norway, the polar night lasts from the 18^{th} of November to the 23^{rd} of January and continuous daylight from the 20^{th} of May to the 24^{th} of July (*Giesen et al.*, 2014). Sea surface temperatures (SST) vary significantly across the region, ranging from 5° C in winter to 20° C in summer in the North Sea, and from 1° C to 15° C in the Barents Sea opening (*Chen et al.*, 2021; *Jakowczyk and Stramska*, 2014). The Skagerrak Strait receives fresher waters from the Baltic Sea and river runoff, resulting in salinity levels varying from 10 PSU in summer to 30 PSU in winter (*Frigstad et al.*, 2020a; *Hordoir et al.*, 2013). Northern Norway generally experiences salinity levels above 34 PSU, occasionally decreasing to 20 PSU during episodic freshwater input during the summer (*Frigstad et al.*, 2020b). This region's Mixed Layer Depth (MLD) fluctuates throughout the year. In winter, it can reach depths of more than 50 meters, while in summer, it becomes shallower, going below 30 meters due to the input of fresh waters and surface heating (*Peralta-Ferriz and Woodgate*, 2015).

2.2 In-situ Data: Algae Abundance and Toxins

Abundance (Cells.L^{-1}) of *Alexandrium* spp., *A. tamarense*, *D. acuta*, *D. acuminata*, *D. norvegica*, *Pseudo-nitzschia* spp., *P. reticulatum*, and *A. spinosum*, was provided by the monitoring program of algae toxins in mussels and dietetic advice to the public from the NFSA. The monitoring program consists in assessing algae abundance at several aquaculture mussel sites weekly (every Monday) and toxin concentration in the blue mussels monthly (Table 2.1). For algae abundance, the surface waters are sampled by lowering a tube from the surface to a 3 m depth, and 25 ml is sub-sampled and preserved with acidic Lugol's iodine before being transported to the laboratory for analysis. The sub-sample (25 ml) is filtered on a membrane filter, and the genus and species are identified and counted on the whole filter under a light microscope at 200x magnification. For toxin concentration, samples of blue mussels are collected between depths of 0.5 to 1.5 m. According to the EU regulations 853/2004, 854/2004, 2074/2005, and 15/2011, the mussel samples are analyzed, and toxins are estimated at the Norwegian School of Veterinary

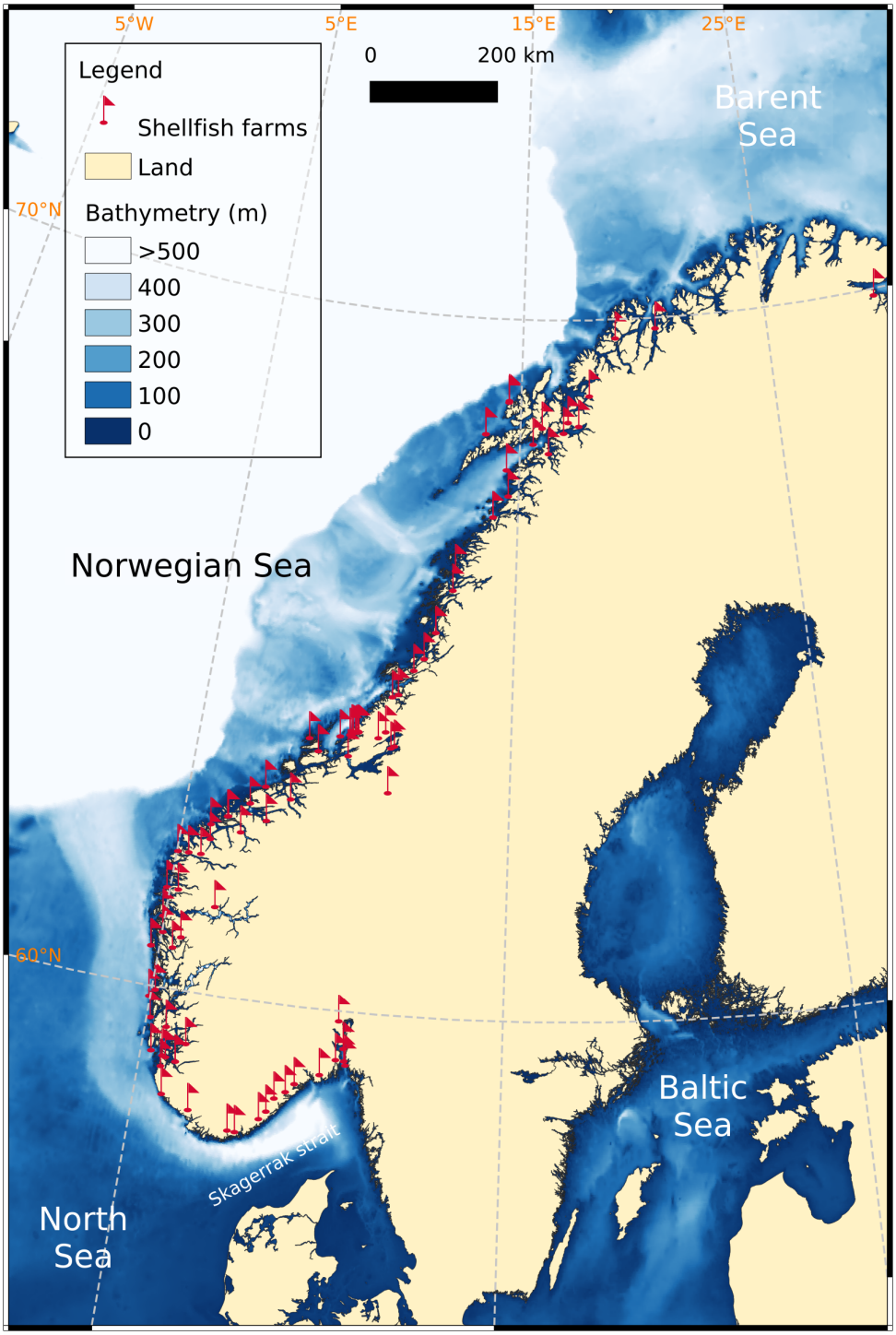


Figure 2.1: Study region. Sampling stations with algae and toxin data from 2006 to 2020 are shown in red.

Science (NMBU) using high-performance liquid chromatography (HPLC).

The monitoring program is run routinely, and the NFSA data record extends from 2006 to 2020. The data was not structured in a single database, and data cleansing was necessary. For each year, an Excel file was provided where each sheet represents one aquaculture farm, containing the table with the weekly and monthly time series of algae abundance and toxin concentration. Auxiliary Excel files containing the location coordinates of each farm were provided for 2010, 2015, and 2020. The coordinates of the auxiliary file with the farm names were matched with the algae abundance and toxin time series. For the years without farms with coordinates, the locations in 2010, 2015, and 2020 files were extrapolated to the other years to farms with similar names. After matching the time series with the location, the following data cleansing was performed: removing samples with inconsistent values (e.g., "?" and "#"), standardizing measured samples without algae detected, such as replacing "i.p." (ikke påvist, meaning not detected in Norwegian) by 0 values and standardizing taxa names. Finally, a new CSV file with standardized and cleaned data for each location was created containing the whole time series from 2006 to 2020, totaling 118 different stations, of which 93 have identified coordinates (Figure 2.1). The total number of samples containing at least one measurement of algae abundance or toxin is 9,599.

2.3 Satellite Observations

2.3.1 Chl-a concentration

Chl-a concentration (μg.L^{-1}) indicates phytoplankton biomass and has been regularly retrieved through satellite remote sensing since 1998. The satellite Chl-a is accessed from the European Space Agency (ESA) Ocean Colour Climate Change Initiative (OC-CCI) project (*Sathyendranath et al.*, 2019), specifically using product version 5 (*Sathyendranath et al.*, 2021). The product is available at https://dx.doi.org/10.5285/1dbe7a109c0244aaad713e078fd3059a. The OC-CCI data merges observations from several satellite sensors, including MODIS, MERIS, OLCI, SeaWiFS, and VIIRS. The product corrects and minimizes the inter-sensor bias that operates in slightly different wavebands to avoid artifacts in the time series. Product version 5 blends remote-sensing reflectance data through band shifting, and the selected bio-optical algorithms to estimate Chl-a concentration are applied to the merged data to generate maps of Chl-a concentration. In this product, several algorithms for atmospheric correction and Chl-a estimation were tested, and the best suited for climate studies were employed. The Chl-a is estimated using the OCI, OCI2 OC2 and OCx algorithms, weighted and blended by their relative skill in specific optical water types (*Calton*, 2020). Comparison between the OC-CCI Chl-a concentration with surface in-situ measurements (n=17,901) exhibit a good correlation coefficient (R) of 0.78, a Root Mean Square Difference (RMSD) of 0.3 μg.L^{-1}, and a bias of 0.003 μg.L^{-1} (*Calton*, 2020). The OC-CCI Chl-a is provided at a spatial resolution of 4 km and has been re-binned in 8-day averages for 2000–2020 to assess the phytoplankton blooms phenology in Paper I. The 8-day average binning approach minimizes the impact of cloud contamination, allowing continuous estimation of the Chl-a time series. The daily resolution is also accessed in Paper I to compare the satellite product with in-situ Chl-a in the study region.

The OC-CCI product's algorithm is optimized explicitly for open oceans, where phy-

Table 2.1: Data summary.

Variable	Source	Spatial resolution	Time resolution	Time coverage
Algae abundance (Cells.L^{-1})	NFSA	Point data	Weekly	2006-2020
Toxin concentration (μg.kg^{-1})	NFSA	Point data	Monthly	2006-2020
Chl-a (mg.m^{-3})	OC-CCI	4 km	8-days average	2000-2020
Chl-a (mg.m^{-3})	OC-CCI	4 km	Daily	2000-2020
PAR (E.m^{-2}day^{-1})	GlobColour	4 km	8-days average	2006-2019
SPM (g.m^{-3})	GlobColour	4 km	8-days average	2000-2020
Wind speed (m.s^{-1})	IFREMER CERSAT	0.25°	6-hours	2000-2019
SST(K)	CCI/C3S	0.05°	Daily	2000-2020
MLD(m)	TOPAZ	12.5 km	Daily	2000-2020
SSS(PSU)	TOPAZ	12.5 km	Daily	2000-2020

toplankton abundance primarily controls the optical properties of the water, commonly referred to as case-1 waters (*Morel and Prieur*, 1977). The OC-CCI Chl-a product is unsuitable for coastal waters, where factors such as land input and resuspension of Suspended Particulate Matter (SPM) and colored dissolved organic matter (CDOM) significantly contribute to the optical properties, commonly referred to as case-2 waters. Neither validation of satellite Chl-a products nor calibration of regional bio-optical algorithms are available in the Norwegian coastal waters. Furthermore, the optical absorption in the blue to green bands, commonly applied to estimate Chl-a, is dominated by the absorption of CDOM in the region (*Nima et al.*, 2016). For these reasons, in Paper I, the grid cells within 30 km of the coastline are excluded from the analysis. For Paper II and Paper III, the Chl-a concentration from the satellite is not employed, as it targets the coastal domain.

2.3.2 Photosynthetically active radiation

Photosynthetically Active Radiation (PAR) is the mean daily photon flux density in the visible range (400 to 700 nm) available for photosynthesis. PAR data is accessed from the GlobColour project (`www.globcolour.info`), which merges data from MODIS, SeaWiFS, and VIIRS sensors (*Frouin et al.*, 2003). The PAR algorithm calculates the daily average PAR irradiance reaching the ocean surface, represented as the quantum energy flux from the Sun within the wavelength range of 400-700 nm and is expressed in units of E.m^{-2}day^{-1}. The algorithm adopts plane-parallel theory, assuming that the impacts of clouds and a clear atmosphere can be treated separately. The planetary atmosphere is defined as a clear sky atmosphere positioned above a cloud layer. The solar flux reaching the ocean surface (E) is expressed by:

$$E = E_{clear}(1 - A)(1 - A_s)^{-1}(1 - S_a A)^{-1} \tag{2.1}$$

where $E_{clear} = E_0 cos(\theta_s) T_d T_g$ is the solar flux that would reach the surface if the cloud and surface system were non-reflecting and non-absorbing, $E_0 cos(\theta_s)$ is the incoming

solar flux at the top of the atmosphere, E_0 is the extra-terrestrial solar irradiance, θ_s is the solar incidence angle, T_d represents the clear sky diffuse transmittance, T_g denotes the gaseous transmittance, A is the cloud/surface system albedo, A_s is the albedo of the ocean surface, and S_a is the spherical albedo. To estimate PAR, the albedo A is expressed as a function of the radiance measured by the satellite instrument within the PAR spectral range, following the procedure described in *Frouin et al.* (2003). The algorithm was initially developed for SeaWiFS, comprising bands 1 (402 - 422 nm), 2 (433 - 453 nm), 3 (480 - 500 nm), 4 (500 - 520 nm), 5 (545 - 565 nm), and 6 (660 - 680 nm). It displays an accuracy of R=0.88 and 13.4% of error compared to in-situ measurements (*Frouin et al.*, 2003). In the GlobColour product, the PAR algorithm is applied separately to each sensor, and all estimates are a weighted average considering each sensor error (*GlobColour*, 2020). The current study assesses the 8-day bin average product and a spatial resolution of 4 km to avoid grid cells contaminated by cloud coverage. PAR is applied as a predictor for feeding the SVM models in Paper II and Paper III.

2.3.3 Suspended particulate matter concentration

Suspended Particulate Matter (SPM) concentration is essential in controlling the scattered PAR that penetrates the water, influencing light availability to algae. The SPM data is obtained through satellite observations and accessed from the GlobColour project (www.globcolour.info), which merges data from MODIS, MERIS, OLCI, SeaWiFS, and VIIRS. The SPM is estimated using the algorithm developed in *Gohin* (2011), which considers solely the non-algal particulate (NAP) as SPM. The algorithm also assumes that CDOM absorption can be neglected at wavelengths longer than 550 nm. The algorithm expresses the reflectance based on the absorption and backscattering coefficients of pure seawater, phytoplankton, and NAP:

$$a = a_w + a_\phi + a_{\text{NAP}} = a_w + a^*_\phi \times [\text{Chl-a}] + a^*_{\text{NAP}} \times [\text{NAP}] \tag{2.2}$$

and

$$b_b = b_{bw} + b_{b\phi} + b_{b\text{NAP}} = b_{bw} + b^*_\phi \times [\text{Chl-a}] + b^*_{\text{NAP}} \times [\text{NAP}] \tag{2.3}$$

where a is the absorption coefficient, w refers to pure waters, ϕ refers to phytoplankton, [NAP] refers to non-algal particulates concentration, [Chl-a] refers to Chl-a concentration, * denotes specificity for a known normalized concentration, b is the scattering coefficient, and b_b is the back scattering coefficient.

Next, a linear relationship is defined between $R^*(550)$, a variable associated with reflectance, and the satellite remote-sensing reflectance (R_{rs}):

$$R^*(550) = \frac{b_b}{a + b_b} = \alpha + \beta R_{\text{rs}}(550) \tag{2.4}$$

where α and β are obtained through minimization using in-situ observations of Chl-a and NAP. The $R^*(550)$ is derived from Chl-a and NAP using the absorption and backscattering coefficients obtained in Equation 2.2 and 2.3. Finally, using the estimated Chl-a from the OC5 algorithm, $R^*(550)$ is inverted to determine the concentration of NAP (*GlobColour*, 2020). Similar to PAR, the SPMs of all sensors are merged using the weighted average. The SPM data is binned at an 8-day interval with a spatial resolution of 4 km to avoid grid cells contaminated by cloud coverage. The SPM is correlated with interannual variability of spring and summer blooms in Paper I.

2.3.4 Surface wind speed

Surface wind speed data from 2006 to 2019 is obtained from the IFREMER CERSAT Global Blended Mean Wind Fields reprocessed product available on Copernicus Marine Environment Monitoring Service (CMEMS). The wind speed dataset includes northward and eastward components and is derived from scatterometers (ASCAT-A and ASCAT-B satellites), SSMIS radiometers (F16, F17, F18, and F19 satellites), and the WindSat radiometer on the Coriolis satellite. The satellite observations are combined into a single product, provided at a 6-hour frequency and a spatial resolution of 0.25°. This product is correlated with interannual variability of phytoplankton blooms in Paper I, and fed in SVM models to forecast *D. acuminata* abundance in Paper III. While the product has been discontinued, the archived reference can be found at https://doi.org/10.48670/moi-00184.

2.3.5 Sea surface temperature

The Sea Surface Temperature (SST) product comprises the SST from ESA SST CCI for the period 2000-2016 followed by an extension from C3S global Sea Surface Temperature Reprocessed for the period 2017-2019 (*Good et al.*, 2020; *Lavergne et al.*, 2019; *Merchant et al.*, 2019). Both datasets are flagged as the same product in the CMEMS portal (doi.org/10.48670/moi-00169). They utilize the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system (*Good et al.*, 2020), which combines data from satellite sensors, such as AATSR, ATSR, SLSTR, and AVHRR, along with in-situ observations, to generate daily average SST at a spatial resolution of 0.05°(*Merchant et al.*, 2019).

SST retrieval relies on the top-of-atmosphere radiance's sensitivity to the Planck emission originating from the sea surface. For SST retrieval, specific "window channels" are commonly employed, namely, the 11 µm, 12 µm, and 3.7 µm for nighttime scenes. The differences between SST and brightness temperature for different thermal channels and view angles exhibit certain relationships, allowing the inversion of multi-channel observations (and multi-angle observations where available) to estimate the SST. Various inversion algorithms are utilized depending on the sensor, and further information about each algorithm can be found in *Petrenko et al.* (2013a) and *Merchant and Embury* (2014).

The in-situ SST assimilated is from multiple platforms, including drifting and mooring buoys and ships. The satellites offer SST data approximately twice a day, although data gaps may be caused by cloud coverage. In OSTIA, satellite SST data that exhibit a significant diurnal signal are excluded. Satellite SST potentially influenced by a diurnal signal is identified using accompanying wind information (when available). Daytime data with wind speeds below $6\,\mathrm{m.s^{-1}}$ are classified as affected by a diurnal signal. For nighttime and daytime data with wind speeds above $6\,\mathrm{m.s^{-1}}$ threshold, the SST below the surface is expected to approximate the skin SST. The outcome is the generation of daily average SST estimates at a depth of 20 cm. Considering the Nordic seas, the uncertainty of these estimations is below 0.4° C (*Good et al.*, 2020). The SST is correlated with the spring and summer blooms in Paper I and fed to SVM models in Paper II and Paper III.

2.4 Model Reanalysis: Mixed Layer Depth and Sea Surface Salinity

Mixed Layer Depth (MLD) – in meters – and Sea Surface Salinity (SSS) – in PSU – data are provided by the CMEMS (doi.org/10.48670/moi-00007) Arctic TOPAZ4 system, as described by *Sakov et al.* (2012) and *Xie et al.* (2017). The TOPAZ4 system is a coupled ocean-sea ice data assimilation system designed for the North Atlantic and the Arctic Ocean. The system combines a Hybrid Coordinate Ocean Model (*Bleck*, 2002) with an elasto-viscous-plastic sea ice model (*Hunke and Dukowicz*, 1997). The TOPAZ4 system assimilates weekly available ocean and sea ice data using the Ensemble Kalman filter (*Evensen*, 2003). The assimilated observations are the satellite SST, sea level anomalies, in-situ temperature and salinity profiles, sea ice concentration and low-resolution sea ice drift data from satellites. The MLD is calculated using a density criterion with a threshold of $0.01\,\mathrm{kg.m^{-3}}$, following the approach of *Petrenko et al.* (2013b) and *Ferreira et al.* (2015). This means the MLD is the shallowest depth where the potential density is $0.01\,\mathrm{kg.m^{-3}}$ higher than the surface waters. SSS is estimated as the average from 0 to 3 m depth (the thickness of the first model layer).

An assessment of TOPAZ4 reanalysis data demonstrates the accuracies for salinities and temperature in the Norwegian and Barents Sea (*Xie et al.*, 2017; *Lien et al.*, 2016). TOPAZ4 performs well concerning ocean variables near the surface. In the first 200 m, salinity RMSD and bias are below 0.3 PSU and between -0.05 and 0.05 PSU. The temperature RMSD and bias are below 1° C and between -0.5 and 0.5 ° C, respectively. Note that MLD uncertainties are not provided, but since it is computed using density, it can be derived from the uncertainties of temperature and salinity.

2.5 Support Vector Machines for HAB Prediction

The Support Vector Machines (SVM) machine learning model is employed in Paper II for probabilistic estimates and in Paper III for regression. Among the many existing machine learning methods for HAB prediction (*Cruz et al.*, 2021), the SVM method offers several advantages. SVM demonstrates strong generalization and requires only a small amount of training data to find an optimal solution. Considering the limited amount of data with identified toxic algae taxa, producing reliable models using small datasets is an optimal skill for calibrating HAB prediction models. Some studies using SVM for HAB prediction have demonstrated this advantage. *Ribeiro and Torgo* (2008) compared the SVM, ANN, and RF to forecast seven algae groups (e.g., cyanobacteria) using biweekly data from 1998 to 2003 in the river Douro, Portugal. The SVM model outperformed the other approaches for most groups analyzed. *Li et al.* (2014) employed monthly/biweekly water quality data from 1997 to 2004 in Tolo Harbor, Hong Kong, to forecast Chl-a concentration using SVM and several ANNs techniques. For one and two weeks, SVM demonstrated superior performance to all ANNs. Finally, SVM is also a versatile tool for regressions, classifications, and probabilistic applications.

The SVM classification – which is further employed for probabilistic estimates – consists in optimizing a margin that best separates two classes by a set of features. These features can be n-dimensional, and the maximized margin refers to a hyperplane. The closest samples between both classes optimize this hyperplane and are referred to as the nearest support vectors (*Cortes and Vapnik*, 1995). The decision function for a

linear SVM classifier is expressed by:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b \tag{2.5}$$

where $\mathbf{x}$ is the input feature vector, $\mathbf{w}$ is the weight vector perpendicular to the hyperplane, and b is the bias term, constrained to:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \tag{2.6}$$

where y_i is the class label (+1 or -1) for data point $\mathbf{x}_i$. Since this constraint is not always satisfied because classes may not be perfectly separated, a soft margin concept is commonly introduced. In this regard, the decision function is constrained to:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \tag{2.7}$$

where ξ_i denotes the soft margin width. Then, the SVM cost function is defined as:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + C(\sum_{i=1}^{n} \xi_i)^2 \tag{2.8}$$

where $\min_{\mathbf{w},b}$ is the objective of the cost function to minimize the $\mathbf{w}$ and b. An optimization algorithm is employed to iteratively update the parameters $\mathbf{w}$ and b to minimize the cost function. The term $\frac{1}{2}\|\mathbf{w}\|^2$ represents the regularization term, which aims to control the complexity of the model and prevent overfitting. The penalty factor (C) controls the trade-off between maximizing the margin and minimizing the training errors. A larger C emphasizes minimizing training errors, while a smaller value allows for a larger margin and more missed classification. In practical terms, C decides when the model should stop optimizing the margin.

Note that equation 2.8 returns a linear separation and only depends on C, which should be tuned through cross-validation. Non-linear separations are also possible using kernel functions, such as polynomial and radial basis functions (RBF). In addition, non-linear kernels require new hyperparameters. For example, the polynomial kernel introduces the degree, and the RBF kernel introduces γ. The kernel function converts the predictors' values to a feature space where the hyperplane is computed. Note that the implementation of cost functions may vary, and the equation 2.8 is for simplifying the explanation. For example, the hyperplane can be computed without converting values to the feature space by using the kernel trick. Deeper details in the mathematical expressions involved in SVM are described in *Cortes and Vapnik* (1995).

We now explore idealized examples to explain the sensitivity of the SVM models to hyperparameters in the HAB prediction context. While the datasets produced are purely artificial, they mimic configurations faced during the thesis.

We assume first a dataset containing two distinct populations, one with *D. acuminata* detected (class=1), and another one without *D. acuminata* (class=-1). Assuming that *D. acuminata* grows better in more saline and warmer waters (*Basti et al.*, 2018; *Rial et al.*, 2023), class 1 has warmer and more saline waters than class -1. An SVM model is calibrated to separate classes 1 and -1 using C of 0.001, 0.01, 0.1, and 1 (Figure 2.2). For C=0.001 (2.2a), a relatively high number of samples fall within the soft margin and serve as support vectors. As the values C increase, leading to fewer support vectors within

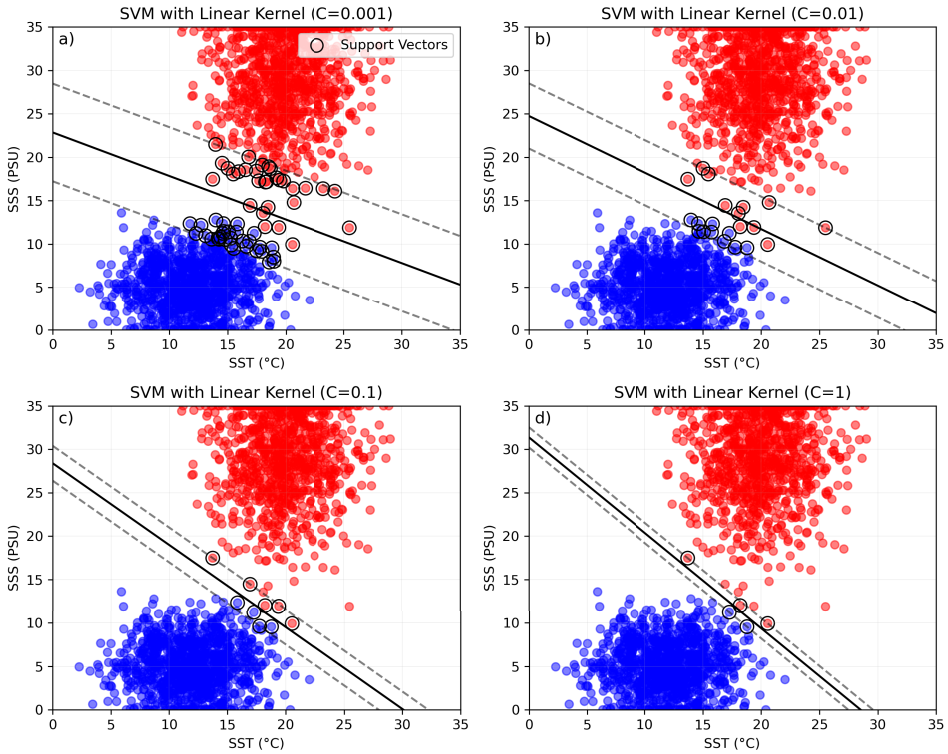


Figure 2.2: SVM calibrated to an idealized dataset of D. acuminata detected (class 1, red) and not detected (class -1, blue). The model is calibrated with a linear kernel, using C of a) 0.001, b) 0.01, c) 0.1, and d) 1. The black circles refer to the support vectors (samples within the margin boundary). The hyperplane is a solid black line, and the soft margin corresponds to the dashed grey area.

the soft margin. Note that the computed hyperplane separating both classes changes depending on C.

We now consider The margin decreases asother idealized situation where *D. acuta* is detected (class=1) and another one without *D. acuta* (class=-1). Class 1 is associated with a wide range of salinity values and moderate light intensity, while class -1 is associated with a wide range of salinity but relatively low and high light conditions. This typically occurs in real configurations where a minimum value of PAR is needed for photosynthesis, but elevated values can cause photodamage (*Rial et al.*, 2023; *García-Portela et al.*, 2018). We consider two SVM models, one using a linear kernel and the other using an RBF kernel (Figure 2.3). The linear kernel cannot effectively separate both classes due to the nonlinearity of the problem in PAR (Figure 2.3a), and class 1 and class -1 related to low PAR conditions are mixed. The RBF kernel otherwise can adjust well to the nonlinearity of the problem (Figure 2.3b). The model adjusts the border areas separating class -1 and 1, isolating the class -1 related to low and high PAR values.

With the RBF kernel method, the hyperparameter γ is introduced:

$$K(x, y) = \exp(-\gamma \cdot ||x - y||^2) \tag{2.9}$$

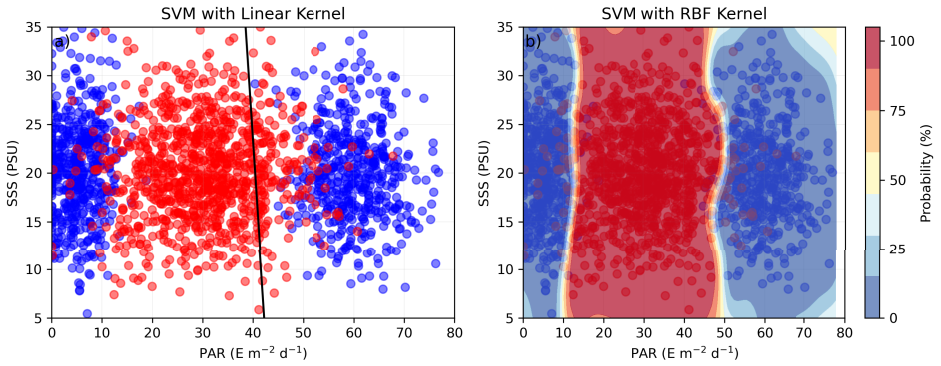


Figure 2.3: SVM calibrated to an idealised configuration of D. acuta detected (class 1, red) and not detected (class -1, blue). The models are calibrated with a) linear and b) RBF kernel, both using a C=1. For the linear kernel, the hyperplane is the solid black line. For RBF kernel, the probability of being class=1 is provided using equation 2.11, with probabilities larger than 50% meaning the predicted class is 1, and values lower than 50% meaning the predicted class is -1.

where x and y are input feature vectors, $||x - y||^2$ is the squared Euclidean distance between x and y, and γ is a parameter that determines the width of the Gaussian curve. Therefore, the γ parameter determines the influence of a single training sample on the decision boundary of the SVM. A small γ implies a larger similarity radius around each training example, resulting in a smoother decision boundary. On the other hand, a larger gamma reduces the similarity radius, causing the decision boundary to fit the training data more tighly. The γ is commonly obtained by tuning hyperparameters empirically on cross-validation datasets. However, if the training dataset is not large enough to exhaustively search, a smooth γ can be estimated as a function of the data variance:

$$\gamma = \frac{1}{n_{\text{features}} \times \sigma^2} \tag{2.10}$$

where n_{features} is the number of features (predictors), and σ^2 denotes the variance of the whole dataset.

Supposing the same situation where the dataset contains two distinct groups, one where *D. acuminata* is detected and is related to warmer and more saline waters (class=1), another without *D. acuminata* detected and associated with colder and fresh waters (class=-1). The SVM model is calibrated using the RBF kernel with varying γ values (Figure 2.4). The smallest γ of 0.001 (Figure 2.4a) produces a slight linear separation between both classes due to the meager single sample influence on the training, resulting in a too-smooth decision boundary. When γ is increased to 0.1 (Figure 2.4b), the sample influencing the decision boundary increases and it fits the cloud of samples more tightly, generating ambiguous areas of 50% probability. A high γ value of 10 (Figure 2.4c) fits the data too tightly and only new samples matching the training samples would be classified as one of the classes 1 or -1. This is typically characteristic of overfitting. Finally, the γ estimated by equation 2.10 produces a smooth and almost linear separation between both classes, which might be enough for this idealized case.

In Paper II the SVM classification algorithm is employed for probabilistic estimations

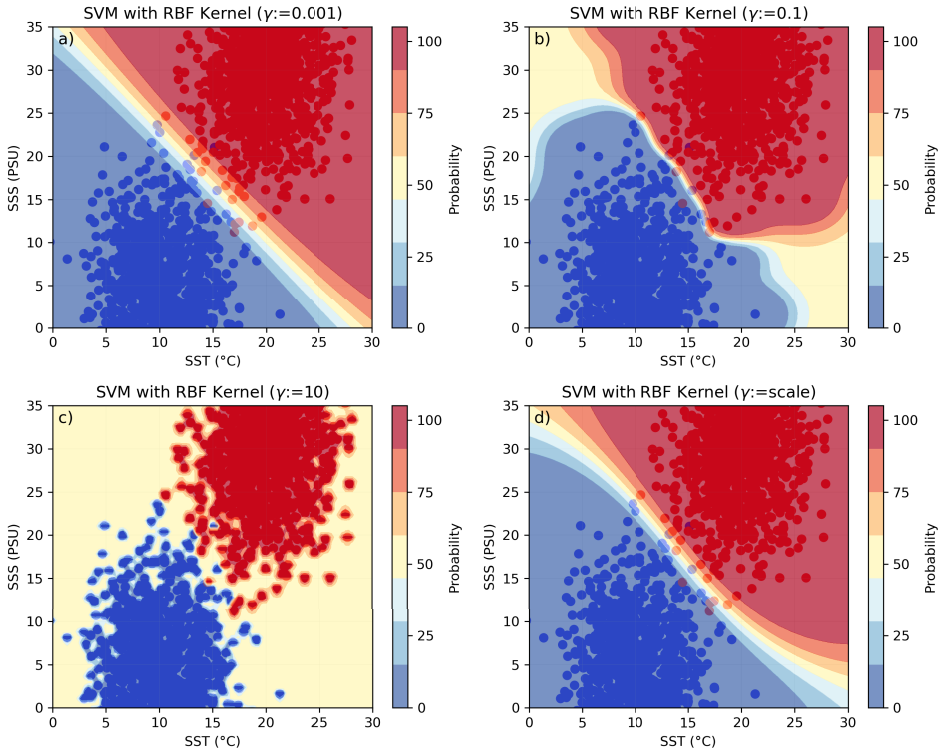


Figure 2.4: SVM calibrated to an idealised case where D. acuminata is detected (1, in red) and not detected (-1, in blue). The models are calibrated with a C=1 and RBF kernel, with varying γ values of a) 0.001, b) 0.1, c) 10, and d) scale. The latter refers to the γ obtained by equation 2.10. Probability is estimated using equation 2.11.

using the *Platt* (1999) method:

$$P(x = 1|f) = \frac{1}{1 + \exp(Af + B)} \tag{2.11}$$

where P is the probability of sample x being class 1, the input f is the SVM output of each predicted sample corresponding to its orthogonal distance from the hyperplane, scaled proportionally from -1 to 1 defined between the support vectors distance, and A and B are the parameters fitted using the maximum likelihood in the training dataset.

In Paper III the SVM is employed for regression, which is constrained to:

$$|y_i - (\mathbf{w} \cdot \mathbf{x}_i + b)| \leq \epsilon \tag{2.12}$$

where ϵ is a new hyperparameter introduced and corresponds to the width of the tolerance region (referred to as epsilon-tube). Therefore, the SVM regression aims to compute a hyperplane that best fits the data points while allowing some errors within the epsilon-tube. The ϵ parameter defines the error tolerance and influences the flexibility of the model. A higher ϵ allows for a higher tolerance to errors. To demonstrate the effect of ϵ on the model during training, we fit a model that estimates the *D. acuminata* growth

rate as a response to temperature (*Basti et al.*, 2018). The model calibrated with ϵ=0.01 adjusts well to the training data points (Figure 2.5) as the epsilon-tube is narrower. For increased $\epsilon = 0.1$, the epsilon-tube is larger and allows for more errors, consequently leading to a model less adjusted to the data points. With an ϵ=0.3, the epsilon-tube is too wide and the model is nearly insensitive to temperature.

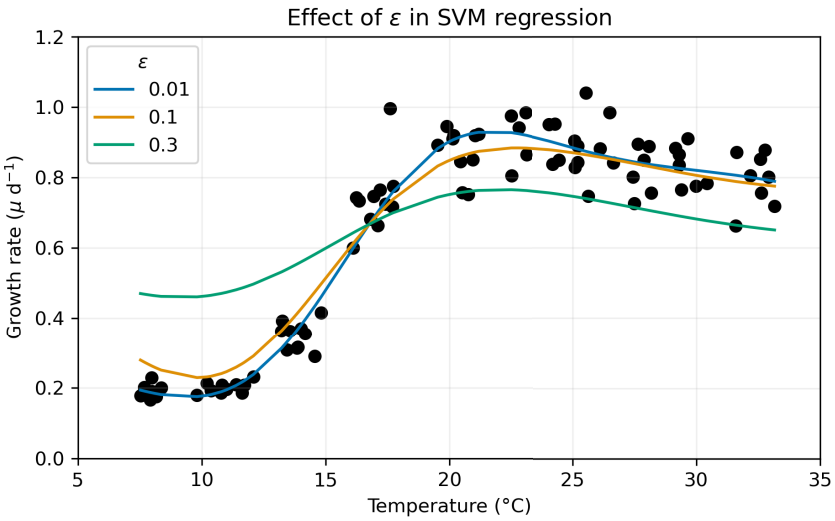


Figure 2.5: SVM regression calibrated to the D. acuminata growth rate as a response to temperature (Basti et al., 2018). Noise is artificially added as a 10% standard deviation of the values. The SVM model is calibrated with kernel=RBF, C=1, gamma=scale, and ϵ of 0.01 (blue), 0.1 (orange), and 0.3 (green).

In the above-idealized examples, two characteristics of the SVM models should be noted. First, the cost function and the RBF kernel in equations 2.8 and 2.9 utilize the feature space constituted by all predictor values. Some predictors might differ in scale. For example, SPM may vary from 0 to 5 $g.m^{-3}$, while SSS can range from 0 to 35 PSU. The discrepancy in the scale requires an a priori rescaling (or normalization step) of all predictors to avoid the predictor with the largest amplitudes prevailing in the cost function. Second, the SVM cost function is also sensitive to unbalanced classes. A class with more samples would outweigh the predictors in the cost function (equation 2.8). Therefore, pre-processing is needed for SVM training, including scaling and weighting the classes accordingly to the number of samples. More details of pre-processing are described in Paper II and Paper III.

3 Introduction to the Papers

Paper I: **Twenty-One Years of Phytoplankton Bloom Phenology in the Barents, Norwegian, and North Seas**

Edson Silva, François Counillon, Julien Brajard , Anton Korosov, Lasse H. Pettersson, Annette Samuelsen, Noel Keenlyside (2021), Frontiers in Marine Science, volume (8)

Paper I is an initial assessment of the environmental factors – either from remote sensing or modeling reanalysis – influencing the blooms phenology and that can feed HAB prediction models. Paper I assess the spring and summer bloom phenology in the seas around Norway from 2000 to 2020 and estimate their trends and interannual variability. The interannual variability of the blooms phenology – earlier or later, weaker or stronger – is correlated to SST, SPM, MLD, and surface winds. Earlier and shallower MLD driven by weaker winds can lead to the blooms developing faster and reaching higher concentrations during the spring. Stronger spring blooms are also correlated with SST but with opposite responses in different regions. Warmer waters correlate with stronger spring blooms between the Faroe Islands and Scotland, while colder waters correlate with stronger spring blooms along the west Norwegian coast. Finally, the summer blooms onset, peak day and duration are rapidly delaying over the last two decades.

Paper II: **Probabilistic Models for Harmful Algae: Application to the Norwegian Coast**

Edson Silva, Julien Brajard, François Counillon, Lasse H. Pettersson, Lars Naustvoll, in review, Environmental Data Science

Paper II calibrates probabilistic models from presence to harmful concentrations (Table 1.1) of toxic algae monitored in the shellfish farms in Norway. Models are based on the SVM machine learning algorithm and are calibrated for *Alexandrium* spp., *A. tamarense*, *D. acuta*, *D. acuminata*, *D. norvegica*, *Pseudo-nitzschia* spp., *P. reticulatum*, and *A. spinosum*. The SST, PAR, MLD and SSS are the input for training the models. The models can predict (nowcasting) the probability of presence for all taxa ($R>0.69$) and remain skillful in predicting harmful concentrations for *Alexandrium* spp., *A. tamarense*, *D. acuta*, and *A. spinosum*. The SVM model skill for the remaining taxa reduces for increasing thresholds, mainly because the samples for training and testing are too few with increasing thresholds. The probabilistic models provide a climatological estimate of the risk of detecting and reaching harmful concentrations (when skillful) along the entire coast of Norway.

Paper III: **Forecasting Harmful Algae Blooms: Application to *Dinophysis acuminata* in Northern Norway**

Edson Silva, François Counillon, Julien Brajard, Lasse H. Pettersson, Lars Naustvoll (2023), Harmful Algae, volume (126)

Paper III calibrates a forecasting machine learning model to predict *D. acuminata* abundance in the Lyngenfjord from 7 to 28 days ahead. The Lyngenfjord and *D. acuminata* are chosen because the monitoring is long enough (extending back to 2006) and there is a pronounced interannual variability. These conditions are optimal for exploring the skill of the SVM machine learning method for forecasting. Considering the results of Paper I, SST and winds are fed to the SVM model. PAR is also fed due to its relevance in sub-seasonal variability. MLD from model reanalysis is not available at the sampling station – in the inner fjord – and is therefore discarded. SST is found to be the most crucial predictor among the physical variables considered as it controls the seasonal period when *D. acuminata* can start blooming. However, environmental drivers cannot predict the interannual variability of *D. acuminata* blooms amplitude alone. Complementing the model with the past *D. acuminata* abundance improved the model's skill in that aspect. The combined models can predict *D. acuminata* up to 28 days ahead (R^2>0.46), beating the skill of trivial predictors such as persistence and climatology.

4 Summary and Conclusions

This thesis calibrates prediction models for harmful algae occurring on the Norwegian coast, specifically focusing on algae related to shellfish poisoning. The study demonstrates that local toxic algae monitoring, remote sensing data, and model reanalysis can calibrate skillful prediction models based on machine learning.

Paper I exhibits that stronger spring blooms are predictable in the North, Norwegian, and Barents seas. Earlier water stratification during spring is found to lead to more intense blooms. This is particularly relevant to predict *Pseudo-nitzschia* spp. blooms that commonly occur during the spring and are only harmful in high abundance. Predictions of HBB – important to fish farms – may also be produced by considering water stratification observations. Nevertheless, the Paper I results are restricted to waters at least 30 km off the coast where no aquaculture farms are located. Empirical evidence relating water stratification and spring bloom intensity has yet to be provided in the Norwegian inner shelves.

The models developed in Paper II can skillfully predict the probability of harmful concentrations of *Alexandrium* spp., *A. tamarense*, *D. acuta*, and *A. spinosum*, and presence probability of *D. acuminata*, *D. norvegica*, *Pseudo-nitzschia* spp., *P. reticulatum*. The models can infer regions and periods with increased risks of HABs, and operate as a general tool for the Norwegian coast on a wide range of time scales. They can provide HAB risk in nowcasting to 10 days using CMEMS and European Center for Medium-range Weather Forecast data and in sub-seasonal-to-decadal using the Copernicus Climate Change Service (C3S). Both time scales can help stakeholders plan mitigation actions and relocations. Furthermore, the probabilistic models can infer harmful algae's future and support aquaculture's evolution in Norway by using climate projections such as the World Climate Research Program Coupled Model Intercomparison Project (CMIP).

The sub-seasonal SVM model of Paper III can provide more refined predictions for shellfish farms with long-time series available. In other words, predicting toxic algae abundance instead of probability, and in a lead time from 7 to 28 days. Paper III model is already skillful with remote sensing observations of SST, PAR, and winds, as well as cell counts from the local farm. The method's accuracy on other species and locations is not evaluated. Still, we can replicate the technique on other species and stations with increasing data availability (such as from 2020 to 2023) in the future.

The SVM demonstrates good skill in producing probabilistic and regression models for predicting HABs. Long enough in-situ observations of toxic algae containing well-represented bloom conditions – weaker and stronger blooms – are necessary to train the models successfully. It can be tempting to train machine learning models only using cross-validation techniques without a final test period in a configuration with limited data. However, only using cross-validation may lead to poorer performance in real configuration and a breach of trust from potential users. Detecting overfitting is only possible by allocating a fraction of the data to the testing period. For instance, models using the RBF kernel and high γ values exhibited high skill during cross-validation (2006–2013) in both Paper II and Paper III but performed poorly during the testing period (2014–2019). Setting the kernel to linear in Paper III and estimating the γ value using equation 2.10 in Paper II prevented overfitting.

We demonstrate that coarse remote sensing observation and model reanalysis are useful for calibrating HAB prediction models. Products in coarse spatial and time resolutions could correlate to spring and summer blooms (Paper I) and calibrate harmful algae probabilistic models (Paper II) and regression model for *D. acuminata* abundance (Paper III). The advantage of using these data rather than more accurate regional models or in-situ observations is that the method can be expandable to other regions. HABs are a worldwide issue and require generic solutions that can be employed elsewhere. The satellite observations are available anywhere and can feed HAB models in other regions. The TOPAZ4 model reanalysis utilized for MLD and SSS is available in the North Atlantic and therefore is more useful for this region. Nevertheless, similar products, such as Global Ocean Physics Reanalysis (doi.org/10.48670/moi-00021), are available globally and may be employed in other regions to replace TOPAZ4.

This study is the first to produce HAB prediction models for the Norwegian coastal waters, and the next logical steps should be presenting these models to stakeholders and conducting a testing period to gather feedback on their usefulness and identify areas for improvement. We can already anticipate that the accuracy of the HAB model can improve by enhancing the quality of the predictors fed to the SVM models. For example, regional models such as NorKyst800 (*Albretsen et al.*, 2011) and NorKyst-DA (*Sperrevik et al.*, 2017) have been developed to improve the representation of coastal dynamics in Norway and could replace the satellite and model reanalysis currently fed into the SVM. These models would improve the accuracy of the environmental quantity and provide estimates inside the fjords that were not available in Paper II and Paper III. The HAB models' accuracy can also improve by including new environmental factors related to nutritional strategies as predictors, such as nitrate, phosphate, selenium, iron, organic matter, prey availability, as well as zooplankton grazing. Some of these variables (e.g., selenium and iron) are primarily accessible by in-situ observations, limiting the calibration of new models to a few farms and only using the Paper III method. Some, such as nitrate, phosphates, and zooplankton, can be retrieved by model reanalysis with moderate accuracy (*Simon et al.*, 2015). Ocean Color is another important candidate. Parameters such as Chl-a can relate to prey availability and CDOM to organic matter, both obtainable by ocean color remote sensing. Satellite Chl-a was not assessed in Paper II and Paper III as no product calibrated for the Norwegian coast exists. Developing ocean-color products for the Norwegian coast would provide new predictors available in long and continuous periods necessary to calibrate machine learning models.

5 Scientific results

Paper I

Twenty-One Years of Phytoplankton Bloom Phenology in the Barents, Norwegian, and North Seas

Edson Silva, François Counillon, Julien Brajard, Anton Korosov, Lasse H. Pettersson, Annette Samuelsen, Noel Keenlyside
Frontiers in Marine Science, **8** (2021)

frontiers in Marine Science

ORIGINAL RESEARCH
published: 10 November 2021
doi: 10.3389/fmars.2021.746327

Twenty-One Years of Phytoplankton Bloom Phenology in the Barents, Norwegian, and North Seas

Edson Silva[1], François Counillon[2], Julien Brajard[1], Anton Korosov[1], Lasse H. Pettersson[1], Annette Samuelsen[2] and Noel Keenlyside[2,3]*

[1] Nansen Environmental and Remote Sensing Center, Bergen, Norway, [2] Nansen Environmental and Remote Sensing Center, and Bjerknes Centre for Climate Research, Bergen, Norway, [3] Geophysical Institute, University of Bergen, Bergen, Norway

Phytoplankton blooms provide biomass to the marine trophic web, contribute to the carbon removal from the atmosphere and can be deadly when associated with harmful species. This points to the need to understand the phenology of the blooms in the Barents, Norwegian, and North seas. We use satellite chlorophyll-a from 2000 to 2020 to assess robust climatological and the interannual trends of spring and summer blooms onset, peak day, duration and intensity. Further, we also correlate the interannual variability of the blooms with mixed layer depth (MLD), sea surface temperature (SST), wind speed and suspended particulate matter (SPM) retrieved from models and remote sensing. The climatological spring blooms start on March 10th and end on June 19th. The climatological summer blooms begin on July 13th and end on September 17th. In the Barents Sea, years of shallower mixed layer (ML) driven by both calm waters and higher freshwaters input keeps the phytoplankton in the euphotic zone, causing the spring bloom to start earlier and reach higher biomass but end sooner due to the lack of nutrients upwelling from the deep. In the Norwegian Sea, a correlation between SST and the spring blooms is found. Here, warmer waters are correlated to earlier and stronger blooms in most regions but with later and weaker blooms in the eastern Norwegian Sea. In the North Sea, years of shallower ML reduces the phytoplankton sinking below the euphotic zone and limits the SPM increase from the bed shear stress, creating an ideal environment of stratified and clear waters to develop stronger spring blooms. Last, the summer blooms onset, peak day and duration have been rapidly delaying at a rate of 1.25-day year^{-1}, but with inconclusive causes based on the parameters assessed in this study.

Keywords: phytoplankton, bloom phenology, Barents Sea, Norwegian Sea, North Sea, remote sensing, spring algae bloom, summer algae bloom

OPEN ACCESS

Edited by:
Hjálmar Hátún, Faroe Marine Research Institute (FAMRI), Faroe Islands

Reviewed by:
Rajdeep Roy, Indian Space Research Organisation, India
Tomonori Isada, Hokkaido University, Japan

****Correspondence:***
Edson Silva
edson.silva@nersc.no

Specialty section:
This article was submitted to Marine Biogeochemistry, a section of the journal Frontiers in Marine Science

Received: *23 July 2021*
Accepted: *12 October 2021*
Published: *10 November 2021*

Citation:
Silva E, Counillon F, Brajard J, Korosov A, Pettersson LH, Samuelsen A and Keenlyside N (2021) Twenty-One Years of Phytoplankton Bloom Phenology in the Barents, Norwegian, and North Seas. Front. Mar. Sci. 8:746327. doi: 10.3389/fmars.2021.746327

INTRODUCTION

Phytoplankton blooms play a crucial role in the marine trophic web and global climate. By assimilating the sunlight, carbon dioxide and nutrients, the algae produce high biomass concentration blooms that feed zooplankton and the higher trophic levels or sink below the euphotic zone. Thus, the algae blooms can support the development of fish larvae

(Townsend and Cammen, 1987; Platt et al., 2003; Vikebø et al., 2012; Asch et al., 2019), provide biomass to the benthic fauna (Zhang et al., 2015) and also contribute to the carbon removal from the atmosphere (Legendre, 1990; Leblanc et al., 2018). Conversely, harmful algae blooms (HABs) can be devastating and lethal (Pettersson and Pozdnyakov, 2013; Gobler, 2020). The high organic matter concentration generated during the bloom can damage or clog fish gills (Chang et al., 1990; Kent et al., 1995) and increase bacteria activity, depleting the dissolved oxygen and causing hypoxia in fishes (Harrison et al., 2017; Mohd-Din et al., 2020). Besides, some algal species can produce toxins, leading to the mortality of fish or even humans when contaminated mussels are consumed (Tangen, 1977; Kaartvedt et al., 1991; Landsberg et al., 2020). The influence of algae blooms on natural living resources and global climate points to the need of assessing the phenology of algae blooms, such as the date of onset, duration, date of the bloom reaching its maximum biomass and the maximum biomass (intensity).

In the Barents, Norwegian, and North seas, two well-known seasonal blooms are the spring and later summer/autumn blooms. During the spring, the sunlight increases and allows the phytoplankton to consume the nutrients upwelled to the upper layers during the winter storms and through terrestrial river discharges (in the coastal waters). At the same time, the mixed layer (ML) shallows above the Sverdrup critical depth and makes the phytoplankton biomass production exceed respiration losses, triggering the spring bloom (Sverdrup, 1953). During summer or autumn, the surface waters are depleted of nutrients due to the algae consumption during the spring bloom. Remineralization, upwelling and river runoff refresh the surface waters with nutrients again, leading to a secondary bloom called summer or autumn bloom (Sverdrup, 1953; Glen Harrison et al., 2013; Sundby et al., 2016).

Although the main processes of spring and summer blooms are well understood in the Barents, Norwegian, and North seas, there is a need to understand mechanisms that can influence the phytoplankton bloom phenology on interannual variability as well as its response to climate change. During a survey campaign in 2013 in the North Atlantic, Naustvoll et al. (2020) observed that regions with earlier shallowing of ML are related to earlier spring blooms. Using satellite data from 2003 to 2017 along the Norwegian coast, Vikebø et al. (2019) found that years with strong winds delay the spring bloom onset. Using a water column model, Opdal et al. (2019) suggested that reducing the water transparency could delay the spring bloom onset in the North Sea. While these studies provided valuable insight on the variability of spring blooms onset and its potential driving mechanisms, still little is known about the duration, date of the peak and intensity of the spring bloom. Besides, the phenology of summer blooms has not been addressed to our knowledge in this region. Thanks to the available extended period of optical remote sensing data (from 1998 to present), one can now more robustly assess the spatial distribution, climatology, trend and potential drivers of interannual variability of both spring and summer blooms.

Here we take advantage of ocean color remote sensing data to provide a comprehensive assessment of both spring and summer blooms phenology in the Barents, Norwegian, and North seas. We use pattern recognition tools to cluster the bloom phenology in regions of comparable statistical behavior, which reduces small-scale noise and allows for a coherent visualization of the properties of the blooms phenology–onset, duration, peak day and intensity. Furthermore, the extended data set used (2000–2020) allows for providing a primary analysis of the trend of the property of the bloom phenology and of potential drivers that can influence the interannual variability.

MATERIALS AND METHODS

Study Region

This study focuses on the Norwegian shelf seas, including the North Sea, the Norwegian Sea, and the Barents Sea entrance (**Figure 1**). The North Sea is shallow, with depth varying from 20 m in the southern region to 700 m in the south of Norway (Eisma et al., 1987). In the Norwegian Sea, the shallowest waters are on the Norwegian Continental Shelf, varying from 100 to 400 m, and the deepest waters are in the Norway Basin, ranging from 3600 to 3800 m. The Barents Sea has a wide continental shelf varying from 100 to 300 m deep (Perry, 1986).

Two significant currents dominate the circulation in the study region, the Norwegian Atlantic Current (NwAC) and the Norwegian Coastal Current (NCC). The NwAC originates from the North Atlantic Current as it flows between the Faroe Islands and Scotland and continues northward along with the Norwegian Continental Shelf break. The NwAC splits into two branches close to the Barents Sea. One branch flows eastward into the Barents Sea, while the other flows northward into the Fram Strait (Furevik et al., 2002; Eldevik et al., 2009). The NCC flows from the south of Norway and along the Norwegian coast up the Barents Sea. It is substantially fresher than the Atlantic Water as it transports fresh waters from the land inflow, the Baltic Sea and the North Sea (Mork, 1981).

Satellite Chlorophyll-a Measurements

Chlorophyll-a concentration (mg m^{-3}) is used as a proxy to phytoplankton biomass and has been retrieved regularly by satellite remote sensing since 1998. We accessed satellite data from the European Space Agency (ESA) Ocean Colour Climate Change Initiative (OC-CCI) project (Sathyendranath et al., 2019), product version 5, which has a spatial resolution of 4 km at a fixed geographical grid and binned in an 8-day average between 1998 and 2020 (**Table 1**). The OC-CCI data is intended for climate studies and merges Chl-a concentration estimated from MODIS, MERIS, OLCI, SeaWiFS, and VIIRS sensors. We have used the data set in the 8-day average bin so that cloud contaminated grid cells are reduced and allow a continuous estimation of the Chl-a time series. From 1998 to 1999, there is only SeaWiFS data available and many grid cells with gaps in the high latitudes of the Norwegian Sea. Thus, we have excluded these 2 years from our analysis.

Global validation of the OC-CCI satellite data with *in situ* Chl-a measurements (n = 17901) showed good agreements between the two data sets–with a correlation coefficient (R)

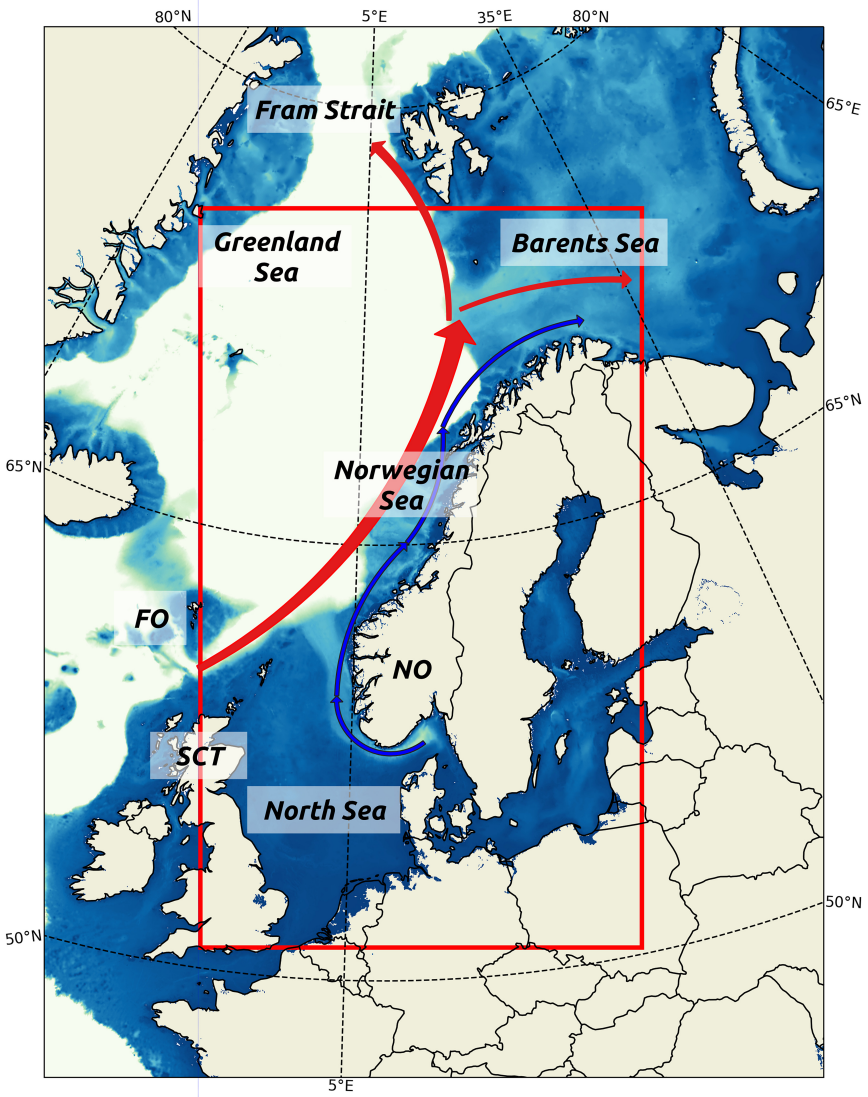


FIGURE 1 | The study region. The red box delimits the area where the blooms phenology is assessed. Red arrow is a coarse representation of the NwAC and blue arrow is the NCC. Abbreviations are Faroe Islands (FO), Norway (NO), and Scotland (SCT).

of 0.78, a root mean square difference of 0.3 mg m^{-3} and a bias of 0.003 mg m^{-3} (Calton, 2020). Note that the OC-CCI product uses an algorithm tuned to perform best in open ocean case-1 waters, where the phytoplankton abundance controls both direct and indirect the optical properties of water (Morel and Prieur, 1977). The OC-CCI Chl-a product has limited validity for coastal case-2 waters where the land input or resuspension of suspended particulate matter (SPM) and yellow substance contributes significantly to the optical properties. Since case-2 waters occur mainly close to the coastline and standard case-1 water algorithms perform poorly in such regions (Blondeau-Patissier et al., 2004; Folkestad et al., 2007), we have discarded grid cells within 30 km off the coastline from the analysis.

TABLE 1 | Summary of the data used in this study.

Source	Variables	Spatial resolution	Temporal resolution	Temporal range
OC-CCI	Chl-a (mg m^{-3})	4 km	8 days	2000–2020
	Chl-a (mg m^{-3})	4 km	Daily	2000–2020
CCI/C3S	SST (°K)	0.05°	Daily	2000–2020
GlobColour	SPM (g m^{-3})	4 km	8 days	2000–2020
TOPAZ	MLD (m)	12.5 km	Daily	2000–2020
IFREMER CERSAT	Wind Speed (m s^{-1})	0.25°	6 h	2000–2019
NMDC	*In situ* Chl-a (mg m^{-3})	–	–	2000–2017
PRG	*In situ* Chl-a (mg m^{-3})	–	–	2000–2016
NVE	River flow (m^3 s^{-1})	–	Daily	2000–2020

This study relies entirely on exploiting information on the phytoplankton dynamics derived from satellite remote sensing sensors. Since passive sensors depend on the solar light scattering, there is no optical satellite data in the winter darkness. Reliable data are only available from February to November in the southern region of the study region and from March to October in the northern region. Although the low sun angle and light availability during the winter may not be sufficient to trigger algae blooms, the bloom onset may occur just before the satellite measurements have sufficient quality when the light starts increasing. This issue was observed in the North Sea (see section "Sub-regions and Chlorophyll-a Validation").

Auxiliary Data

Auxiliary data is used to correlate the interannual variability of bloom phenology (onset, peak day, duration, and intensity) with potential key drivers. The parameters assessed are the mixed layer depth (MLD), sea surface temperature (SST), wind speed, SPM concentration, and river runoff (**Table 1**).

We accessed SST (K) from the ESA SST CCI and C3S global SST Reprocessed product level 4, available on the Copernicus Marine Environment Monitoring Service (CMEMS)[1]. The product is created by running the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system (Good et al., 2020), which combines satellite (AATSR, ATSR, SLSTR, and AVHRR) and *in situ* observations to produce gap-free maps of daily average SST at 0.05° of spatial resolution (Merchant et al., 2019).

Suspended particulate matter (g m^{-3}) data was obtained by satellite observations and accessed from the GlobColour project[2]. SPM is estimated using Gohin (2011) algorithm on MODIS, MERIS, OLCI, SeaWiFS, and VIIRS sensors, and binned at an 8-day interval at a 4 km of spatial resolution. Note that SPM is estimated by radiometric measurements from the same optical sensors used for estimating Chl-a concentration, and they may share a common bias.

The MLD (m) is provided by the CMEMS Arctic MFC TOPAZ system (Sakov et al., 2012; Xie et al., 2017). The TOPAZ system is a coupled ocean–sea ice data assimilation system for the North Atlantic and the Arctic Ocean. The model couples a Hybrid Coordinate Ocean Model (Bleck, 2002) with an elasto-viscous-plastic sea ice model (Hunke and Dukowicz, 1997). TOPAZ assimilates available ocean and sea ice data with the Ensemble Kalman filter (Evensen, 2003) every week. The MLD is calculated using a density criterion with a threshold of 0.01 kg m^{-1}, as in Petrenko et al. (2013) and Ferreira et al. (2015).

Surface wind speed (m s^{-1}) was obtained from the IFREMER CERSAT Global Blended Mean Wind Fields reprocessed product accessed from the CMEMS. 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 a 0.25° of spatial resolution.

We use the daily flow data from the Tana (70.070°N, 28.016°E) and Målselva (69.035°N, 18.658°E) rivers accessed from The Norwegian Water Resources and Energy Directorate[3]. The river flow was used as a proxy of coastal waters freshening to discuss the MLD variability in sections "Barents Sea: Stronger, Earlier and Shorter Spring Blooms Driven by Shallower Mixed Layer" and "Mixed Layer Depth and Sea Surface Temperature Influence on the Spring Bloom Phenology in the Norwegian Sea."

Clustering Sub-Regions

We have used cluster analysis to objectively identify 20 sub-regions of similar bloom phenology using the 21 years of Chl-a time series. In the pre-processing, we subset the time series between the Julian days 60 and 300 to exclude the winter when there is a lack of data. Although we have accessed Chl-a data binned in 8-day average, 30.8% of the data cube (latitude × longitude × time) is still missing–mainly in the Greenland Sea–due to the cloud contamination. We interpolate those missing values as we need them for clustering the sub-regions. For each year, we use linear interpolation (limit = 10-time intervals) for filling the missing values over time, reducing the missing data to 5.2%. The remaining missing values are in the beginning and end of the time series in the higher latitudes. Since we cannot interpolate values in the time series borders, we filled them with 0.01, assuming that the Chl-a concentration is virtually null but still present during the beginning and end of the winter. We emphasize that the filling value with 0.01 was only used for clustering and not further used to assess the blooms phenology.

We use principal component analysis to reduce the dimension of 630-time intervals to 300 components, accounting for 95% of the Chl-a time variability. Then, we use the k-means algorithm (MacQueen, 1967) fed with the components as features and the grid cells as observations. We chose the optimal number of sub-regions (k) as the maximum value where coherent (spatially continuous) sub-regions were still obtained. With more than 20 clusters, we got noisy sub-regions composed of a few grid cells. We use the k-means++ algorithm (Arthur and Vassilvitskii, 2007) to define the initial seeds and avoid changing the sub-regions areas each time the k-means is reproduced.

[1]https://marine.copernicus.eu/

[2]https://www.globcolour.info/

[3]https://www.nve.no/

In the post-processing, we observed that four sub-regions in the Greenland Sea were heavily contaminated by clouds and probably sea ice. We set a high interpolation limit of 10-time intervals in the pre-processing to address this issue during the clustering. However, this wide range of interpolation made the Chl-a time series in the Greenland Sea too linear and unrealistic to assess the blooms phenology. For this reason, we have excluded those four sub-regions, resulting in 16 sub-regions. Furthermore, a few grid cells along the sub-regions boundaries were overlapping each other sub-region. We have smoothed the boundaries by removing those grid cells and interpolating with the nearest sub-region.

For assessing the clustering performance, we computed the Silhouette score (Rousseeuw, 1987) using the 300 components of every grid cell:

$$S(i) = \frac{b(i) - a(i)}{max\left(a(i), b(i)\right)} \quad (1)$$

where S is the Silhouette score for a given grid cell i, b is the mean nearest-cluster Euclidean distance, and a is the mean intra-cluster Euclidean distance. A Silhouette score higher than 0 means that the Chl-a time series of one individual grid cell is more similar–in the Euclidean space–to the cells of the same sub-region than the remaining regions.

Satellite Chlorophyll-a Validation

We assess the performance of the OC-CCI Chl-a product in the 16 sub-regions by comparing the daily OC-CCI Chl-a with *in situ* Chl-a collected on the same day (**Table 1**). The *in situ* Chl-a data was provided by the Norwegian Institute of Marine Research (IMR) from two different sources. One was retrieved from the Norwegian Marine Data Centre[4] with the registered id imr_11. The other was provided by the Plankton Research Group (PRG) (courtesy of Dr. Kjell Gundersen) and was analyzed by the Plankton Chemistry Laboratory at IMR. For measuring *in situ* Chl-a, a standard volume (265 mL) is collected and filtered onto a 25 mm GFF filter and stored frozen (−20°C) until analysis in the land-based laboratory. Samples are transported in specially designed coolers, with an internal temperature recorder rated for −20°C for a minimum of 3 days. The samples are thawed in 90% acetone in the laboratory and stored at +4°C overnight before analysis on a Turner Design 10AU fluorometer. The fluorometer is regularly calibrated using a solid standard with known fluorescence following Holm-Hansen and Riemann (1978). Measurements of the top 10 m have been averaged to compare well with the remote sensing data. Note that the satellite data is calibrated with Chl-a estimated by high-performance liquid chromatography (HPLC), which slightly differs from Chl-a estimated by fluorometers available in this study (Neveux et al., 1990; Reynolds et al., 2001; Giannini et al., 2021).

To match the *in situ* and satellite Chl-a data, we followed the protocols described by Bailey and Werdell (2006) with some slight adjustments. As the satellite product is a merged output of different sensors and has more than one measurement time, we matched the satellite data with *in situ* data collected on the same day and between 09:00 and 15:00 UTC+2. For each match-up, we extracted the average of a 3 × 3 window from the satellite product and compared it with the *in situ* measurements. Besides, we computed the coefficient of variation in the same window for assessing spatial homogeneity in the satellite product. Windows with a coefficient of variation higher than 30% were considered spatial heterogeneous and unsuitable for assessing the satellite data accuracy, so we removed them. Last, we removed a few *in situ* samples with values lower than 0.1 mg m^{-3}, which were substantially overestimating the mean absolute percentage error (MAPE).

The satellite Chl-a validation includes linear correlation (R), MAPE, and the root mean squared error (RMSE):

$$R_{x,y} = \frac{cov(X, Y)}{\sigma X \sigma Y} \quad (2)$$

$$MAPE(\%) = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{X_t - Y_t}{X_t} \right| \quad (3)$$

$$RMSE\left(mg\ m^{-3}\right) = \sqrt{\frac{\sum_{t=1}^{n} (X_t - Y_t)^2}{n}} \quad (4)$$

where X and Y are the independent and dependent variables, respectively; cov is the covariance function; σ is the standard deviation; n is the number of samples; t is the sample. For the satellite and *in situ* match-ups, X is the *in situ* Chl-a, and Y is the satellite Chl-a. Furthermore, we also fit a linear regression between satellite and *in situ* match-ups.

Bloom Phenology Estimates

This study only focuses on the seasonal spring and summer blooms that last a couple of weeks. As mentioned in section "Satellite Chlorophyll-a Measurements," the Chl-a ocean color data is averaged into 8-day bins with the frequency of Chl-a variations in each bin reaching as high as 1/8 day^{-1}. We use functional data analysis (Ramsay and Silverman, 2005) to smooth the seasonal Chl-a time series and remove high frequencies from Chl-a variability. We smooth each year time series of each grid cell using a Fourier series with five basis coefficients. The number of basis coefficients can significantly impact the smoothness of the time series. On one hand, having too many basis coefficients will fail to smooth the time series enough and contains the high frequencies of Chl-a variability. On the other hand, too few basis coefficients can exaggeratedly smooth the time series and miss the phytoplankton blooms. We tested the number of basis coefficients varying from 3 to 10 on 1 year of data, and five basis coefficients were found to be the most suitable for representing the seasonality of spring and summer blooms (not shown). With less than four basis coefficients, some summer blooms in the Barents Sea were not detected, and with more than six basis coefficients, the weekly variability of Chl-a was included for spring bloom in the North Sea.

We estimate the bloom onset, peak day, duration and intensity using the smoothed Chl-a time series. The peak day corresponds

[4]https://nmdc.no/

to the local Chl-a maxima, and the bloom intensity is the Chl-a concentration at the peak. Since there is high uncertainty in the ocean color derived Chl-a concentration estimate below 0.5 mg m^{-3} (see section "Sub-regions and Chlorophyll-a Validation"), the local Chl-a maximum is only computed when values reach more than 0.5 mg m^{-3}.

Several methods were considered for estimating the bloom onset from the time series of Chl-a observations, such as the threshold method, the rate of change method and the cumulative sum method (Brody et al., 2013). The threshold method computes the climatological median of a Chl-a time series, defines a threshold value above the median, and the onset is estimated when the Chl-a concentration reaches this threshold. The rate of change method estimates the onset as the date with the highest increase of Chl-a before the peak. The cumulative sum method computes the cumulative sum from the beginning of the time series to the peak, and the onset is estimated when the sum reaches a percentual threshold of the cumulative sum (e.g., 15%). The threshold method cannot be applied in our case because there is no data available during the winter, and the climatological median of the Chl-a time series cannot be estimated. We found the rate of change method to delay the onset by a few weeks for blooms of longer duration (not shown). The cumulative sum method estimates the onset when the Chl-a starts to increase and can be applied despite the lack of winter data, and it is retained for the following analysis. We use the 15% threshold to estimate the bloom onset as recommended in Brody et al. (2013) for subpolar regions.

Since the annual Chl-a time series usually has two peaks of bloom intensity, in spring and summer, and the ocean color data is not available at the beginning of the year during the winter, we had to define the beginning of the time series for computing each bloom onset. For the spring bloom onset, the beginning of the time series was defined as the first local minima of the annual Chl-a time series. For estimating the summer bloom onset, the beginning of the time series was defined as the first local minima between the first and second peaks. Last, the bloom duration is estimated as the time between the onset and peak day.

Climatology, Trends and Detrended Correlation Assessment

We estimate the spring and summer blooms phenology for each grid cell of each year. The climatological spring and summer blooms onset, peak day, duration and intensity are estimated by averaging the grid cells belonging to each of the 16 sub-regions over the entire period. We also estimate the average value of the auxiliary parameters during the climatological bloom period (from onset to peak day) of each sub-region and year. For computing the trends of all parameters assessed, we use the sub-regions averaged values of each year and fit a linear least-squares regression. In order to relate the interannual variability of blooms phenology characteristics with key potential drivers, we subtract the trends from the interannual variability. Then, we use the Pearson correlation between the blooms phenology and the auxiliary parameters.

The correlation significance level is estimated using a probability density function (Student, 1908). For a given two random datasets with zero correlation, a probability density function of R is drawn for a given n (number of years). Then, we use a two-sided Wald test between the estimated R and the probability density function, and the null hypothesis is that the R is zero for an α threshold of 0.05. In practical terms, for a given $n = 21$ the $-0.42 < R < 0.42$ is insignificant. Note that it was ensured that there is no autocorrelation in the interannual variability. To estimate the trend significance, we test the slope of the linear least-squares regression using a two-sided Wald test with an α threshold of 0.05.

RESULTS

Sub-Regions and Chlorophyll-a Validation

The Barents, Norwegian, and North seas have been clustered into 16 sub-regions using the Chl-a time series (**Figure 2**). Four sub-regions (1–4) are in the Barents Sea, six sub-regions are in the Norwegian Sea (5–10), and the other six regions are in the North Sea (11–16). All 16 sub-regions are well localized in specific geographical regions except for sub-region 11, which starts in the central North Sea, extends throughout Skagerrak and covers the southern part of the NCC. Based on earlier investigations (Pozdnyakov et al., 2017), we initially intended to split the sub-region 11. However, all further clustering analyses gave nearly identical results for splitting the sub-region further, so we decided to keep it as one sub-region.

The 16 sub-regions showed heterogeneous annual Chl-a time series patterns and, consequently, different spring and summer blooms phenology (**Figure 2**). The sub-regions heterogeneity is also supported by the Silhouette score, where 84% of the grid cells showed values higher than 0 (**Supplementary Figure 1**). Most grid cells with a Silhouette score lower than 0 are concentrated on the sub-regions boundaries. The Chl-a time-series patterns are expected to vary among the sub-regions smoothly rather abruptly. Thus, Silhouette scores lower than 0 in the boundaries represent transitional areas among the Chl-a time-series patterns shown in **Figure 2**.

As explained in section "Satellite Chlorophyll-a Measurements," satellite data is only of sufficient quality after the beginning of February to the end of October, and there is no data available during the winter. This restriction limited the assessment of the spring bloom phenology in the southernmost sub-regions 11 and 14. There, Chl-a concentration only decreases after the middle of February, and there is no spring peak. Besides, the secondary bloom seems to reach its peak after October. Without detecting the bloom peaks, it is not possible to assess the bloom phenology in those regions. Another limitation occurs in sub-region 2 and 10, where the summer bloom developed only in a few grid cells and in some years. Since summer blooms seem rare in those sub-regions, it is unfeasible to compute a robust summer bloom climatology based on the 21 years dataset.

The OC-CCI product shows a fair agreement with *in situ* Chl-a concentration in sub-regions from 1 to 12 (**Figure 3**). The

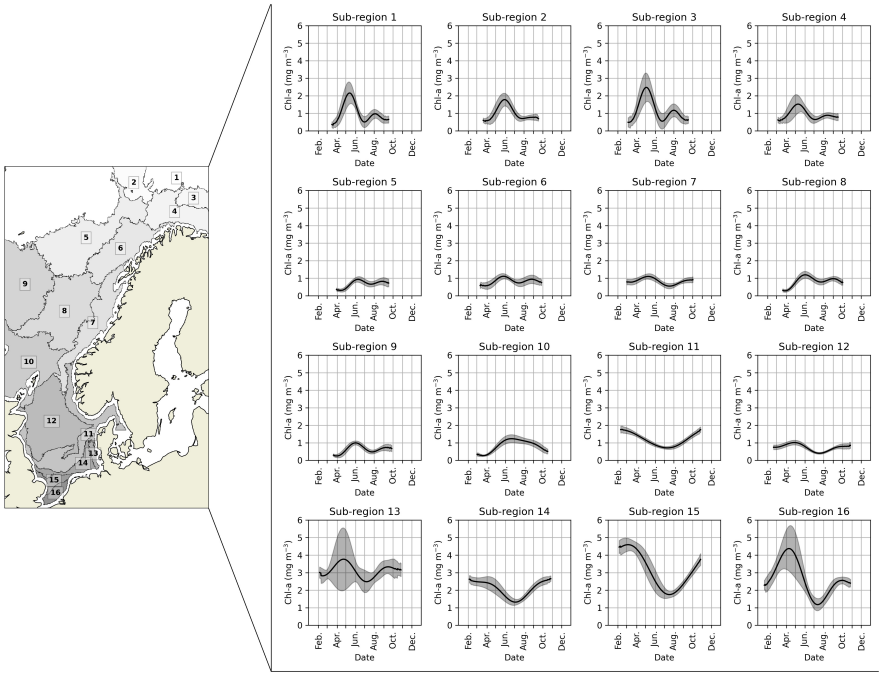


FIGURE 2 | The clustered sub-regions (left) and their annual Chl-a time series (right). The solid black line is the average over the period 2000–2020 and the shaded area is one standard deviation of the interannual variability.

R varies from 0.37 to 0.77, MAPE varies from 33 to 63% and RMSE varies from 0.3 to 0.8 mg m^{-3}. Most regions show results comparable to the OC-CCI overall validation for case-1 waters, which showed an R of 0.78 and RMSE of 0.23 for Chl-a logged data (Calton, 2020). Besides, the match-up also shows that the main source of the misfits relates to a satellite overestimation of Chl-a concentrations for values lower than 0.5 mg m^{-3}. This overestimation for the lower-level concentrations is not of great concern for assessing the seasonal bloom phenology because the bloom is only computed when the peak reaches values above 0.5 mg m^{-3}. Furthermore, we do not have access to enough *in situ* samples ($n < 8$) matching satellite data in sub-regions from 13 to 16. Since other studies show a need to use case-2 water algorithms for retrieving Chl-a concentration in the south of the North Sea (De Cauwer et al., 2004; Tilstone et al., 2012), we decided to exclude those sub-regions in the assessment of bloom phenology.

Bloom Climatology

The average spring bloom in the North Sea (sub-region 12) starts on March 10th and lasts until April 26th for 46 days (**Figure 4**), reaching 1.2 mg m^{-3} at the peak during the studied period. In the Norwegian Sea (sub-regions 6–10), the timing (onset, peak day, and duration) of the spring blooms has high longitudinal variability. The average bloom starts on March 30th in the eastern side and on May 3rd in the western sub-regions, reaching its peak between May 10th and June 19th, respectively. The spring bloom lasts between 39 and 45 days in most sub-regions. However, the bloom lasts for 58 days in the sub-region 10 in the southern Norwegian Sea, which is the longest spring bloom in the study region. The spring bloom intensity varies from 1.1 to 1.4 mg m^{-3} in the Norwegian Sea. In the Barents Sea (sub-regions 1–4), the spring bloom starts between April 11th and 27th and ends between May 7th and 28th, lasting for up to 33 days, which is the shortest spring bloom of the study region. Nevertheless, the Barents Sea shows the strongest intensity of the spring blooms that vary on average from 1.8 to 2.7 mg m^{-3}.

The timing and intensity of summer blooms have lower spatial discrepancies than for the spring blooms. The summer bloom begins on average from July 8th in the Barents Sea to August 1st in the Norwegian Sea and it ends from August 3rd in the Barents Sea to September 17th in the North Sea. The blooms last from 23 days in the Barents Sea to 48 days in the North Sea. Finally, the summer blooms intensity varies between 0.9 and 1.2 mg m^{-3}.

Bloom Phenology Trends

The spring bloom timing does not significantly change in the North and the Norwegian seas over the 21 years study period, but the spring bloom onset advances and the duration increases

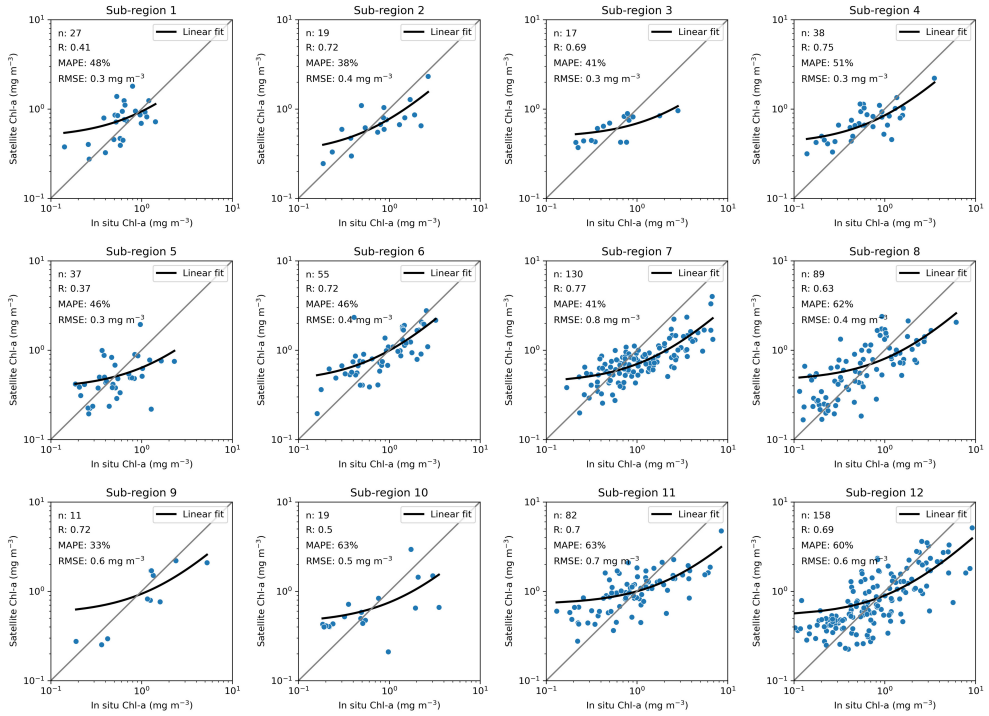


FIGURE 3 | Comparison between *in situ* and satellite Chl-a concentration for each sub-region. The black line is a linear model fitted using least squared regression, and it is curved due to the axes log scale.

in the Barents Sea (**Figure 5** and **Supplementary Figure 2**). In sub-regions 1 and 3, the blooms onset advances by −0.74-day year^{-1}, and in sub-regions 3 and 4, the blooms duration increases by 0.49-day year^{-1}. Furthermore, the intensity of spring blooms increases in the Norwegian Sea (**Supplementary Figure 3**) in sub-regions 5 and 7, with a rate of 0.02 and 0.01 mg m^{-3} year^{-1}, respectively.

The summer blooms significantly change from the North Sea and the Barents Sea. The summer blooms onset and peak day is delayed by 1.25-day year^{-1} in sub-regions 5, 6, 7, 8, 9, and 11. The duration of the blooms also increases at a rate up to 0.35-day year^{-1} in sub-regions 4, 6, 7, 8, 9, and 12. Last, the intensity of blooms increases in sub-regions 7, 9, and 12 with a rate of 0.025 mg m^{-3} year^{-1}.

Interannual Detrended Correlation

The MLD interannual variability correlates with the spring bloom timing and intensity in the North, Norwegian, and Barents seas (**Figure 6**). The correlation varies between 0.51 and 0.74 with the bloom onset and peak day and between −0.54 and −0.62 with the intensity. Thus, despite a study arguing that the spring bloom is not triggered by the shallowing of the ML and it starts in winter when the MLD is maximum (Behrenfeld, 2010), our results show that MLD seems to regulate the timing of spring blooms. The only exceptions are in sub-regions 2, 7, and 10, where MLD does not correlate with the spring bloom phenology.

Sea surface temperature correlates with the spring blooms timing and intensity in the Norwegian Sea but with inverse relationships depending on the region. On one hand, warmer waters are correlated with later and weaker spring blooms in sub-region 7. On the other hand, warmer waters are correlated with earlier spring blooms in sub-regions 5, 8, and 9, and stronger spring blooms in sub-region 10. Wind speed and SPM are correlated with the spring blooms just in a few regions. Wind speed is significantly correlated with the Barents Sea spring bloom timing and intensity, where weaker winds relate to earlier onsets and peak days and more intense blooms. SPM is correlated to spring bloom timing in sub-regions 2, 8, and 9, and with the intensity only in the North Sea, where low SPM concentrations relate to stronger spring blooms.

The summer bloom correlates with physical and biogeochemical parameters only in a few sub-regions (**Figure 7**). Shallower MLD correlates with the earlier summer blooms onset and peak day in sub-regions 8 and 9. Warmer SST correlates with later and longer summer blooms in sub-regions 6, 7, 8, and 12, and with less intense

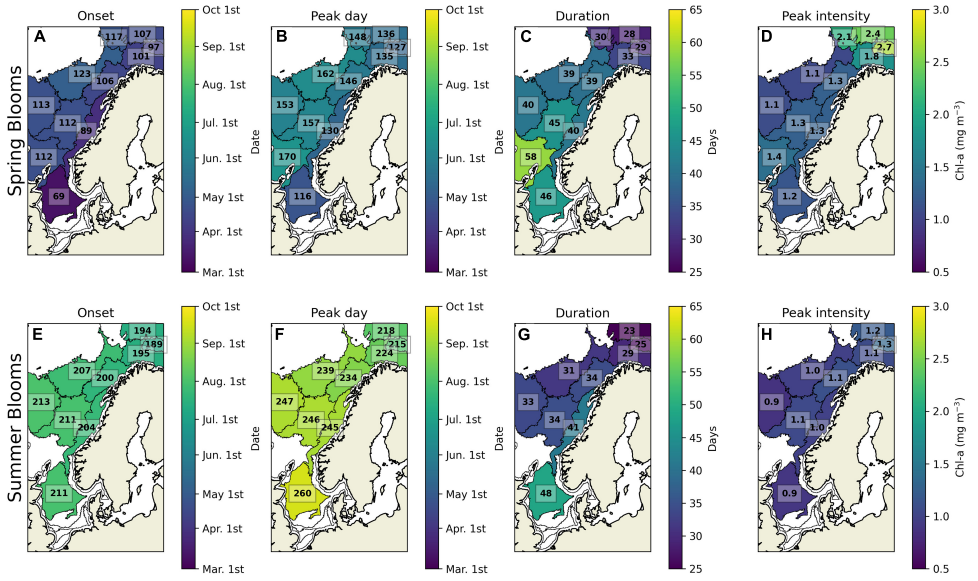


FIGURE 4 | Climatology of spring and summer blooms phenology. The subplots are the spring bloom **(A)** onset, **(B)** peak day, **(C)** duration, and **(D)** intensity; the summer blooms **(E)** onset, **(F)** peak day, **(G)** duration, and **(H)** intensity. The numbers in the onset and peak day are Julian days; in the duration are in days; in the peak intensity are in Chl-a (mg m^{-3}). Sub-regions where quantity is untrustworthy are masked in white.

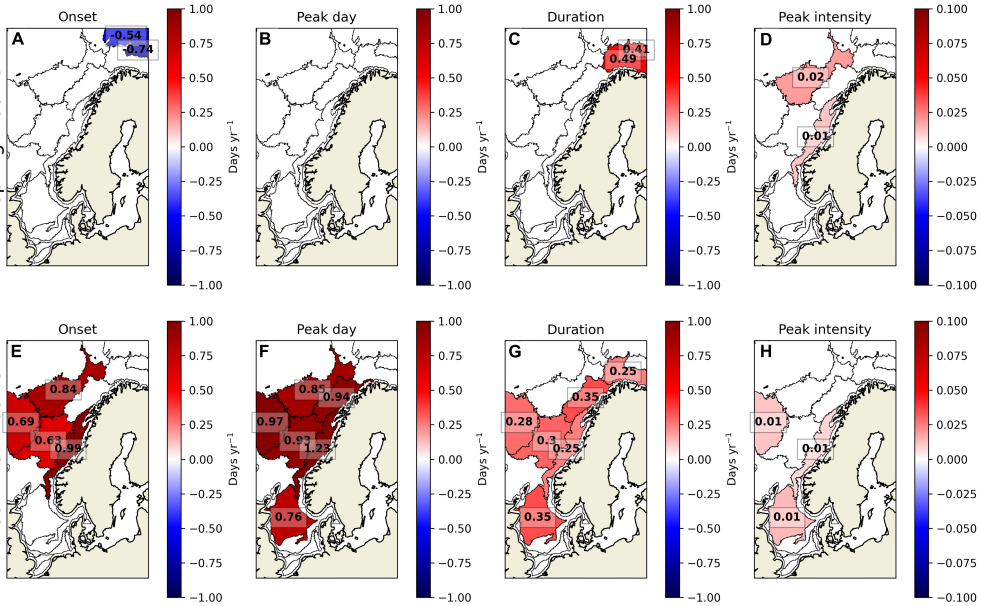


FIGURE 5 | The trends of spring and summer blooms phenology from 2000 to 2020. The subplots are the spring bloom **(A)** onset, **(B)** peak day, **(C)** duration, and **(D)** intensity; the summer blooms **(E)** onset, **(F)** peak day, **(G)** duration, and **(H)** intensity. Sub-regions where trends are insignificant ($\alpha > 0.05$) are masked in white.

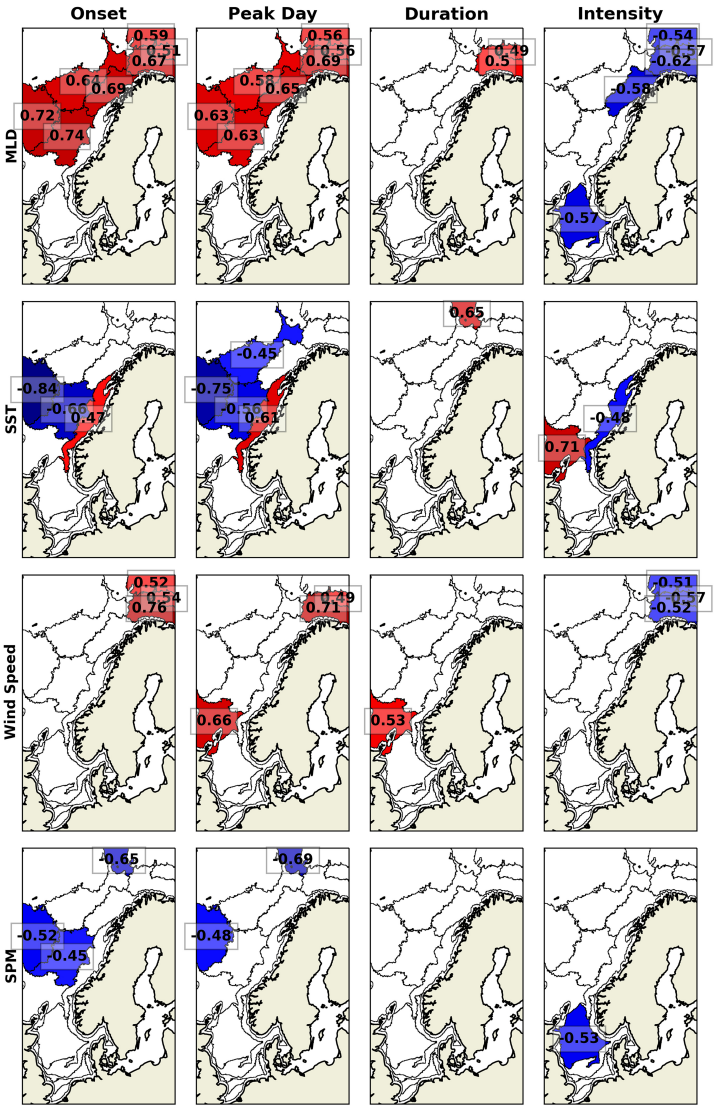


FIGURE 6 | Detrended interannual correlation between the spring bloom phenology and MLD, SST, wind speed, and SPM. Red colors are positive correlations, blue colors are negative correlations and white colors are used for insignificant correlations ($\alpha > 0.05$).

blooms in sub-regions 3 and 8. Last, wind speed only correlates with summer blooms intensity in sub-region 6, and SPM only correlates with summer blooms intensity in sub-regions 4, 5, and 6.

It should be noted that other remotely sensed and modeled parameters have been considered but were not presented to keep the paper concise. For spring bloom, photosynthetically available radiation (PAR) has significant correlations in sub-region 4 in a similar manner to MLD–albeit weaker. For summer blooms, the correlations with PAR are at the edge of significance and somewhat counter-intuitive–with higher PAR linked to shorter and weaker summer blooms. Euphotic depth (Z_{eu}) and light attenuation coefficient (K_d) were also considered. These are inherently related to the algae blooms–i.e.,

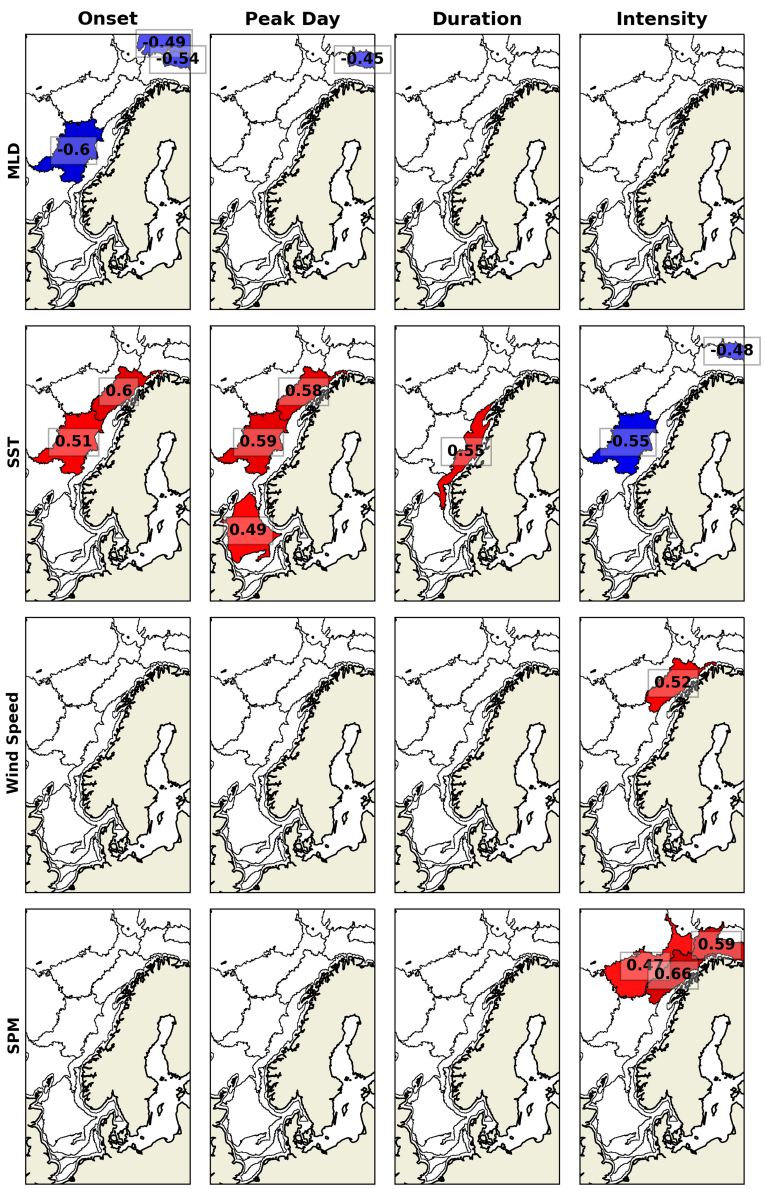


FIGURE 7 | Detrended interannual correlation between the summer bloom phenology and MLD, SST, wind speed, and SPM. Red colors are positive correlations, blue colors are negative correlations and white colors are used for insignificant correlations ($\alpha > 0.05$).

Z_{eu} and K_d change the blooms (Sverdrup, 1953) at the same time the Chl-a change the Z_{eu} and K_d (Kirk, 2011). Thus, we retrieved high correlations. However, the correlations showed that higher biomass blooms are correlated to higher K_d and lower Z_{eu}, meaning that the blooms probably control the interannual variability of K_d and Z_{eu} during the bloom rather than the contrary. For this reason, PAR, K_d, and Z_{eu} are not presented in this study.

DISCUSSION

Summer Bloom Delay

In the trends analysis, it appears that summer blooms are getting delayed from 2000 to 2020. However, the possible causes of this delay are unclear to us. The spring bloom exhausts the nutrients on the surface, and associated with other factors such as zooplankton grazing, leads to the ending of spring bloom. In the poor nutrient waters of summer, the increase of MLD results in an upwell of nutrients required for a summer or autumn bloom (Glen Harrison et al., 2013; Wihsgott et al., 2019). Therefore, we would expect that changes in the summer bloom phenology are related to the MLD. For example, Martinez et al. (2011) suggested that a possible cause for the weaker autumn bloom in the North Atlantic (30–50°N) was a delayed increase of MLD during the autumn of the 2000s. We computed the MLD trend during the summer bloom and down to 4 weeks before the onset. Our results showed that there had been no significant trend in the MLD from 2000 to 2020. Besides, we also computed the trends of SST, SPM and wind speed, and none showed a significant trend.

The lack of trends in MLD, SST, SPM and wind speed suggests that other factors that are not assessed in this study may be influencing the delay of summer blooms. For example, bio-advection of phytoplankton and zooplankton pressure. A recent study showed that bio-advection caused by a faster intrusion of the North Atlantic Current is causing the phytoplankton *Emiliania huxleyi* to increase in the Barents Sea (Oziel et al., 2020). In the North Atlantic, the predominant zooplankton species has been changing (Beaugrand, 2002), and the new species biomass increases during autumn (Planque and Fromentin, 1996). Martinez et al. (2011) suggested that the new predominant zooplankton species increased the grazing pressure in the autumn, probably leading to a weaker bloom in the 2000s. Likewise, the delay of summer blooms found in this study could be related to the bio-advection of phytoplankton or changes in the zooplankton community structure. Still, no firm conclusion can be held yet on the reason of the delayed summer blooms. This will be the topic of a follow-up study.

Barents Sea: Stronger, Earlier and Shorter Spring Blooms Driven by Shallower Mixed Layer

The intensity of spring blooms in the Barents Sea (sub-regions 1, 3, and 4) showed the highest interannual variability of the study area (**Supplementary Figure 3**). For example, varying from 1.2 mg m^{-3} in 2008 to 4.1 mg m^{-3} in 2002 in sub-region 3. Besides, the intensity of spring blooms is negatively correlated with the onset, peak day and duration ($R < -0.44$), meaning that the spring blooms in the Barents Sea are typically either early, short and strong or late, long and weak.

Two dynamical processes contribute to the nutrient loads in the surface waters of the southern and central Barents Sea: the lateral inflow of Atlantic waters that come from NwAC and the vertical mixing when the ML is deeper during winter (Wassmann et al., 2006). In March, the deep ML makes the phosphate, nitrate and silicate evenly distributed in the water column, reaching 0.85, 11.2, and 4.5 μM, respectively (Reigstad et al., 2002). It was shown that springs with a deeper ML are correlated with higher nitrate concentrations (Olsen et al., 2003). However, deep ML and high nitrate concentrations do not necessarily lead to stronger spring blooms. Our results suggest that a shallower ML is correlated to stronger spring blooms. Stratified waters reduce the phytoplankton sinking below the euphotic zone, which may be critical in the Barents Sea, where one of the predominant phytoplankton taxa is Diatom that quickly sinks due to its dense cell walls of silica (Degerlund and Eilertsen, 2010). The high concentration of nutrients available at the beginning of the bloom in March (Reigstad et al., 2002) could be sufficient for developing the spring blooms, while interannual variability of MLD would control the intensity and the timing of the blooms. Nevertheless, a reduced level of nutrients available in years of shallower ML can explain why stronger spring blooms end sooner.

The seasonal ML shallowing during spring is driven by the SST increase from solar radiation and the freshening of water coming from sea ice melting, interactions with the fresher fjord water systems and inflow of the NCC (Drinkwater et al., 2003; Olsen et al., 2003; Loeng and Drinkwater, 2007). In the sub-regions 1, 3, and 4, we found that the interannual variability of MLD is significantly correlated with wind speed ($R = 0.58$) and freshwaters discharge of Tana River ($R = -0.65$) during the bloom. This indicates that calm waters and high discharge of freshwaters may lead to shallower ML, resulting in stronger spring blooms. This also explains why weaker winds correlate with earlier and stronger spring blooms in the Barents Sea.

Mixed Layer Depth and Sea Surface Temperature Influence on the Spring Bloom Phenology in the Norwegian Sea

Surface heating from solar radiation and freshening of the coastal waters leads to the shallowing ML during the spring in the Norwegian Sea, and the spring bloom starts when the MLD reaches a lower depth than the Sverdrup critical depth (Sverdrup, 1953). The MLD has been shown to influence the spatial variability of spring blooms onset (Naustvoll et al., 2020). Our results show that MLD is also a possible driver of the interannual variability of bloom onset and peak day. The positive correlation between MLD and the spring blooms onset and peak day in the sub-regions 5, 6, 8, and 9 ($R > 0.58$) suggests that a shallower ML favor earlier spring blooms. In sub-regions 5, 8, and 9, interannual SST is correlated with MLD ($R < -0.46$), and consequently, SST is also correlated with the spring bloom timing. In sub-region 6, only the Målselva River inflow has a significant correlation with MLD ($R = -0.55$). Therefore, warmer years during the spring could lead to shallower ML and earlier blooms from the middle to the northwestern Norwegian Sea, while the spring bloom in sub-region 6 behaves similarly to spring blooms in the Barents Sea.

The SST is correlated with the spring blooms in sub-regions 7 and 10 without the influence of MLD. In both sub-regions, MLD has not correlated with SST and the spring blooms as observed in the remaining Norwegian Sea. Besides, the spring bloom intensity of both sub-regions shows an inverse relationship with SST. Warmer waters are correlated with stronger blooms in sub-region 10, whereas colder waters are correlated with earlier and stronger blooms in sub-region 7. Without the influence of SST on MLD interannual variability, two hypotheses could explain why warmer waters are correlated with stronger spring blooms in sub-region 10. First, warmer waters increase the growth rate of phytoplankton (Eppley, 1972; Moisan et al., 2002; Bissinger et al., 2008), and during the spring bloom, a higher growth rate could support the bloom to reach higher Chl-a biomass. Second, warmer waters could indicate a higher intrusion of Atlantic waters from the North Atlantic Current. If Atlantic waters are associated with the input of nutrients in sub-region 10, more nutrients could lead the spring bloom to higher Chl-a biomass.

Regarding sub-region 7, warm Atlantic waters trapped in mesoscale anticyclonic eddies were found to delay the spring blooms in the Norwegian Shelf (Hansen et al., 2010). However, the spring bloom delay in the eddy was probably caused by the delayed shallowing of the ML, and our results have not shown a relationship between SST and MLD in sub-region 7. Thus, we have not found a plausible explanation why colder waters could relate to earlier and stronger spring blooms in this region.

Shallower Mixed Layer Drives Clearer Waters and Stronger Spring Blooms in the North Sea

The average spring bloom intensity in the sub-region 12 ranges from 0.9 to 1.5 mg m^{-3}, and it is significant correlated with MLD and SPM. A shallower ML reduces the phytoplankton sinking to deeper waters, and a lower SPM concentration increases the euphotic zone and the light available to the algae photosynthesis. Moreover, we also found that SPM is correlated with MLD during the spring bloom ($R = -0.44$). A previous study showed that surface SPM variability in the North Sea is correlated with the bed shear stress caused by oceanic waves, but only for months with a deep ML (Wilson and Heath, 2019). In March, deep ML allows the bed shear stress to increase the surface SPM concentration ($R^2 = 0.47$). In April, the MLD decreases and the bead shear stress is no longer correlated with the surface SPM variability. The climatological spring bloom in sub-region 12 starts on March 10th and ends on April 26th, and this implies that the relation between MLD, bed shear stress and SPM are strongly correlated for almost half of the bloom period. Therefore, the compound influence of SPM and MLD was found to explain the intensity of spring blooms. Years of shallower ML reduces the phytoplankton sinking and could lead to more transparent waters, resulting in more sunlight available to photosynthesis and, hence, stronger spring blooms.

CONCLUSION

This study presents an exhaustive analysis of the regional spring and summer blooms phenology for a region that extends from the Barents, Norwegian to the North seas using an extended and novel clustering analysis. The regional climatology and trend of the phytoplankton blooms phenology, as well as a primary analysis of the co-variability of potential drivers with interannual variability, are presented and discussed. In the Barents Sea, we found that low wind amplitude and increased freshening lead to springs with shallower ML, and more stratified waters are related to spring blooms starting earlier and reaching higher biomass during the peak. In the Norwegian Sea, we found that SST has an opposed correlation with spring blooms. There, warmer waters are correlated with earlier and higher biomass blooms from the north of Scotland to the off-shore central parts of the Norwegian Sea, but lower temperatures are correlated with earlier and higher biomass blooms in the eastern Norwegian Sea. In the North Sea, springs with shallower ML limits the SPM increase by bottom shear stress, favoring an environment of more stratified and clearer waters, correlating well with blooms of higher biomass. Last, we also found a rapid delay in the onset and increased biomass of the summer blooms during the past 21 years in almost the entire study region. The summer blooms are starting later, ending later, getting longer and reaching higher Chl-a concentrations.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: OC:CCI: https://www.oceancolour.org/, CMEMS: https://marine.copernicus.eu/, GlobColour: https://www.globcolour.info/, and NMDC: https://nmdc.no/.

AUTHOR CONTRIBUTIONS

ES, FC, JB, AK, and LP contributed to the conception and design of the study. ES organized the database, performed the statistical analysis, and wrote the first draft of the manuscript. ES and FC performed the data processing. FC and JB supervised the data processing and statistical analysis. All authors contributed to the interpretation of the results of this work, manuscript revision, read, and approved the submitted version.

FUNDING

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). NK acknowledges the Nansen Legacy project (276730). This work has also received a grant for computer time from the Norwegian

Program for super computing (NOTUR2, project number nn9039k) and a storage grant (NORSTORE, NS9039k).

ACKNOWLEDGMENTS

We want to thank Jiping Xie for making the TOPAZ4 MLD data available to us. We also want to thank the Norwegian Institute of Marine Research for making the *in situ* Chl-a data available. JB is member of Sorbonne University as associate professor.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.746327/full#supplementary-material

REFERENCES

Arthur, D., and Vassilvitskii, S. (2007). "k-means++: the advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, (New Orleans), 11.

Asch, R. G., Stock, C. A., and Sarmiento, J. L. (2019). Climate change impacts on mismatches between phytoplankton blooms and fish spawning phenology. *Glob. Chang. Biol.* 25, 2544–2559. doi: 10.1111/gcb.14650

Bailey, S. W., and Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.* 102, 12–23. doi: 10.1016/j.rse.2006.01.015

Beaugrand, G. (2002). Reorganization of North Atlantic marine copepod biodiversity and climate. *Science (80-.).* 296, 1692–1694. doi: 10.1126/science.1071329

Behrenfeld, M. J. (2010). Abandoning sverdrup's critical depth hypothesis on phytoplankton blooms. *Ecology* 91, 977–989. doi: 10.1890/09-1207.1

Bissinger, J. E., Montagnes, D. J. S., Sharples, J., and Atkinson, D. (2008). Predicting marine phytoplankton maximum growth rates from temperature: improving on the Eppley curve using quantile regression. *Limnol. Oceanogr.* 53, 487–493. doi: 10.4319/lo.2008.53.2.0487

Bleck, R. (2002). An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Model.* 4, 55–88. doi: 10.1016/S1463-5003(01)00012-9

Blondeau-Patissier, D., Tilstone, G. H., Martinez-Vicente, V., and Moore, G. F. (2004). Comparison of bio-physical marine products from SeaWiFS, MODIS and a bio-optical model with in situ measurements from Northern European waters. *J. Opt. A Pure Appl. Opt.* 6, 875–889. doi: 10.1088/1464-4258/6/9/010

Brody, S. R., Lozier, M. S., and Dunne, J. P. (2013). A comparison of methods to determine phytoplankton bloom initiation. *J. Geophys. Res. Ocean.* 118, 2345–2357. doi: 10.1002/jgrc.20167

Calton, B. (2020). *ESA Ocean Colour Climate Change Initiative – Phase 3 Product User Guide for v5*. 0 Dataset. Paris: ESA.

Chang, F. H., Anderson, C., and Boustead, N. C. (1990). First record of a Heterosigma (Raphidophyceae) bloom with associated mortality of cage-reared salmon in Big Glory Bay, New Zealand. *New Zeal. J. Mar. Freshw. Res.* 24, 461–469. doi: 10.1080/00288330.1990.9516437

De Cauwer, V., Ruddick, K., Park, Y.-J., Nechad, B., and Kyramarios, M. (2004). *Optical Remote Sensing in Support of Eutrophication Monitoring in the Southern North Sea*. European Association of Remote Sensing Laboratories (EARSeL), 208–221.

Degerlund, M., and Eilertsen, H. C. (2010). Main species characteristics of Phytoplankton spring blooms in NE Atlantic and Arctic Waters (68–80° N). *Estuaries Coasts* 33, 242–269. doi: 10.1007/s12237-009-9167-7

Drinkwater, K. F., Belgrano, A., Borja, A., Conversi, A., Edwards, M., Greene, C. H., et al. (2003). "The response of marine ecosystems to climate variability associated with the north atlantic oscillation," in *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, eds J. W. Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck (Washington, DC: American Geophysical Union), 211–234.

Eisma, D., Johnston, R., and Cadogan, J. I. G. (1987). The North Sea: an overview. *Philos. Trans. R. Soc. London. B, Biol. Sci.* 316, 461–485. doi: 10.1098/rstb.1987.0032

Eldevik, T., Nilsen, J. E., Iovino, D., Anders Olsson, K., Sandø, A. B., and Drange, H. (2009). Observed sources and variability of Nordic seas overflow. *Nat. Geosci.* 2, 406–410. doi: 10.1038/ngeo518

Eppley, R. W. (1972). Temperature and Phytoplankton Growth in the Sea. *Fish. Bull.* 70, 1063–1085.

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.* 53, 343–367. doi: 10.1007/s10236-003-0036-9

Ferreira, A. S. A., Hátún, H., Counillon, F., Payne, M. R., and Visser, A. W. (2015). Synoptic-scale analysis of mechanisms driving surface chlorophyll dynamics in the North Atlantic. *Biogeosciences* 12, 3641–3653. doi: 10.5194/bg-12-3641-2015

Folkestad, A., Pettersson, L. H., and Durand, D. D. (2007). Inter-comparison of ocean colour data products during algal blooms in the Skagerrak. *Int. J. Remote Sens.* 28, 569–592. doi: 10.1080/01431160600821044

Furevik, T., Bentsen, M., Drange, H., Johannessen, J. A., and Korablev, A. (2002). Temporal and spatial variability of the sea surface salinity in the Nordic Seas. *J. Geophys. Res. Ocean.* 107, 1–16. doi: 10.1029/2001jc001118

Giannini, F., Hunt, B. P. V., Jacoby, D., and Costa, M. (2021). Performance of OLCI Sentinel-3A satellite in the Northeast Pacific coastal waters. *Remote Sens. Environ.* 256:112317. doi: 10.1016/j.rse.2021.112317

Glen Harrison, W., Yngve Børsheim, K., Li, W. K. W., Maillet, G. L., Pepin, P., Sakshaug, E., et al. (2013). Phytoplankton production and growth regulation in the Subarctic North Atlantic: a comparative study of the Labrador Sea-Labrador/Newfoundland shelves and Barents/Norwegian/Greenland seas and shelves. *Prog. Oceanogr.* 114, 26–45. doi: 10.1016/j.pocean.2013.05.003

Gobler, C. J. (2020). Climate change and harmful algal blooms: insights and perspective. *Harmful Algae* 91:101731. doi: 10.1016/j.hal.2019.101731

Gohin, F. (2011). Annual cycles of chlorophyll-a, non-algal suspended particulate matter, and turbidity observed from space and in-situ in coastal waters. *Ocean Sci.* 7, 705–732. doi: 10.5194/os-7-705-2011

Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., et al. (2020). The current configuration of the OSTIA system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sens.* 12, 1–20. doi: 10.3390/rs12040720

Hansen, C., Kvaleberg, E., and Samuelsen, A. (2010). Anticyclonic eddies in the Norwegian Sea; their generation, evolution and impact on primary production. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 57, 1079–1091. doi: 10.1016/j.dsr.2010.05.013

Harrison, P. J., Piontkovski, S., and Al-Hashmi, K. (2017). Understanding how physical-biological coupling influences harmful algal blooms, low oxygen and fish kills in the Sea of Oman and the Western Arabian Sea. *Mar. Pollut. Bull.* 114, 25–34. doi: 10.1016/j.marpolbul.2016.11.008

Holm-Hansen, O., and Riemann, B. (1978). Chlorophyll a Determination: improvements in Methodology. *Oikos* 30:438. doi: 10.2307/3543338

Hunke, E. C., and Dukowicz, J. K. (1997). An Elastic–Viscous–Plastic model for sea ice dynamics. *J. Phys. Oceanogr.* 27, 1849–1867. doi: 10.1175/1520-04851997027<1849:AEVPMF<2.0.CO;2

Kaartvedt, S., Johnsen, T. M., Aksnes, D. L., Lie, U., and Svendsen, H. (1991). Occurrence of the toxic phytoflagellate Prymnesium parvum and associated fish mortality in a Norwegian fjord system. *Can. J. Fish. Aquat. Sci.* 48, 2316–2323. doi: 10.1139/f91-272

Kent, M., Whyte, J., and LaTrace, C. (1995). Gill lesions and mortality in seawater pen-reared Atlantic salmon Salmo salar associated with a dense bloom of Skeletonema costatum and Thalassiosira species. *Dis. Aquat. Organ.* 22, 77–81. doi: 10.3354/dao022077

Kirk, J. T. O. (2011). *Light and Photosynthesis in Aquatic Ecosystems*. Cambridge: Cambridge University Press.

Landsberg, J. H., Hendrickson, J., Tabuchi, M., Kiryu, Y., Williams, B. J., and Tomlinson, M. C. (2020). A large-scale sustained fish kill in the St. Johns River, Florida: a complex consequence of cyanobacteria blooms. *Harmful Algae* 92:101771. doi: 10.1016/j.hal.2020.101771

Leblanc, K., Quéguiner, B., Diaz, F., Cornet, V., Michel-Rodriguez, M., Durrieu de Madron, X., et al. (2018). Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. *Nat. Commun.* 9:953. doi: 10.1038/s41467-018-03376-9

Legendre, L. (1990). The significance of microalgal blooms for fisheries and for the export of particulate organic carbon in oceans. *J. Plankton Res.* 12, 681–699. doi: 10.1093/plankt/12.4.681

Loeng, H., and Drinkwater, K. (2007). An overview of the ecosystems of the Barents and Norwegian Seas and their response to climate variability. *Deep. Res. Part II Top. Stud. Oceanogr.* 54, 2478–2500. doi: 10.1016/j.dsr2.2007.08.013

MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations," in *Proceeding of the Fifth Berkeley Symposium* (Berkeley), 281–297.

Martinez, E., Antoine, D., D'Ortenzio, F., and De Boyer Montégut, C. (2011). Phytoplankton spring and fall blooms in the North Atlantic in the 1980s and 2000s. *J. Geophys. Res. Ocean.* 116, 1–11. doi: 10.1029/2010JC006836

Merchant, C. J., Embury, O., Bulgin, C. E., Block, T., Corlett, G. K., Fiedler, E., et al. (2019). Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Sci. Data* 6, 1–18. doi: 10.1038/s41597-019-0236-x

Mohd-Din, M., Abdul-Wahab, F., Mohamad, S. E., Jamaluddin, H., Shahir, S., Ibrahim, Z., et al. (2020). Prolonged high biomass diatom blooms induced formation of hypoxic-anoxic zones in the inner part of Johor Strait. *Environ. Sci. Pollut. Res.* 27, 42948–42959. doi: 10.1007/s11356-020-10184-6

Moisan, J. R., Moisan, T. A., and Abbott, M. R. (2002). Modelling the effect of temperature on the maximum growth rates of phytoplankton populations. *Ecol. Modell.* 153, 197–215. doi: 10.1016/S0304-3800(02)00008-X

Morel, A., and Prieur, L. (1977). Analysis of variations in ocean color1. *Limnol. Oceanogr.* 22, 709–722. doi: 10.4319/lo.1977.22.4.0709

Mork, M. (1981). Circulation phenomena and frontal dynamics of the Norwegian coastal current. *Philos. Trans. R. Soc. London. Ser. A, Math. Phys. Sci.* 302, 635–647. doi: 10.1098/rsta.1981.0188

Naustvoll, L. J., Melle, W., Klevjer, T., Drinkwater, K. F., Strand, E., and Knutsen, T. (2020). Dynamics of phytoplankton species composition, biomass and nutrients in the North Atlantic during spring and summer - A trans-Atlantic study. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 180:104890. doi: 10.1016/j.dsr2.2020.104890

Neveux, J., Delmas, D., Romano, J., Algarra, P., Ignatiades, L., Herbland, A., et al. (1990). Comparison of chlorophyll and phaeopigment determinations by spectrophotometric, fluorometric, spectrofluorometric and HPLC methods. *Mar. Microb. Food Webs* 4, 217–238.

Olsen, A., Johannessen, T., and Rey, F. (2003). On the nature of the factors that control spring bloom development at the entrance to the Barents Sea and their interannual variability. *Sarsia* 88, 379–393. doi: 10.1080/00364820310003145

Opdal, A. F., Lindemann, C., and Aksnes, D. L. (2019). Centennial decline in North Sea water clarity causes strong delay in phytoplankton bloom timing. *Glob. Chang. Biol.* 25, 3946–3953. doi: 10.1111/gcb.14810

Oziel, L., Baudena, A., Ardyna, M., Massicotte, P., Randelhoff, A., Sallée, J. B., et al. (2020). Faster Atlantic currents drive poleward expansion of temperate phytoplankton in the Arctic Ocean. *Nat. Commun.* 11, 1–8. doi: 10.1038/s41467-020-15485-5

Perry, R. K. (1986). "Bathymetry," in *The Nordic Seas*, ed. B. G. Hurdle (New York, NY: Springer New York), 211–236.

Petrenko, D., Pozdnyakov, D., Johannessen, J., Counillon, F., and Sychov, V. (2013). Satellite-derived multi-year trend in primary production in the Arctic Ocean. *Int. J. Remote Sens.* 34, 3903–3937. doi: 10.1080/01431161.2012.762698

Pettersson, L. H., and Pozdnyakov, D. (2013). *Monitoring of Harmful Algal Blooms*. Berlin: Springer Berlin Heidelberg.

Planque, B., and Fromentin, J. (1996). Calanus and environment in the eastern North Atlantic. I. Spatial and temporal patterns of C. finmarchicus and C. helgolandicus. *Mar. Ecol. Prog. Ser.* 134, 101–109. doi: 10.3354/meps134101

Platt, T., Fuentes-Yaco, C., and Frank, K. T. (2003). Spring algal bloom and larval fish survival off Nova Scotia. *Nature* 423, 398–399.

Pozdnyakov, D. V., Pettersson, L. H., and Korosov, A. A. (2017). *Exploring the Marine Ecology From Space*. Cham: Springer International Publishing.

Ramsay, J. O., and Silverman, B. W. (2005). *Functional Data Analysis*. New York, NY: Springer New York.

Reigstad, M., Wassmann, P., Wexels Riser, C., Øygarden, S., and Rey, F. (2002). Variations in hydrography, nutrients and chlorophyll a in the marginal ice-zone and the central Barents Sea. *J. Mar. Syst.* 38, 9–29. doi: 10.1016/S0924-7963(02)00167-7

Reynolds, R. A., Stramski, D., and Mitchell, B. G. (2001). A chlorophyll-dependent semianalytical reflectance model derived from field measurements of absorption and backscattering coefficients within the Southern Ocean. *J. Geophys. Res. Ocean.* 106, 7125–7138. doi: 10.1029/1999JC000311

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi: 10.1016/0377-0427(87)90125-7

Sakov, P., Counillon, F., Bertino, L., Lisæter, K. A., Oke, P. R., and Korablev, A. (2012). TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci.* 8, 633–656. doi: 10.5194/os-8-633-2012

Sathyendranath, S., Brewin, R. J. W., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., et al. (2019). An ocean-colour time series for use in climate studies: the experience of the ocean-colour climate change initiative (OC-CCI). *Sensors (Switzerland)* 19:4285. doi: 10.3390/s19194285

Student. (1908). Probable error of a Correlation Coefficient. *Biometrika* 6, 302–310. doi: 10.2307/2331474

Sundby, S., Drinkwater, K. F., and Kjesbu, O. S. (2016). The North Atlantic spring-bloom system-where the changing climate meets the winter dark. *Front. Mar. Sci.* 3:28. doi: 10.3389/fmars.2016.00028

Sverdrup, H. U. (1953). On conditions for the vernal blooming of phytoplankton. *ICES J. Mar. Sci.* 18, 287–295. doi: 10.1093/icesjms/18.3.287

Tangen, K. (1977). Blooms of gyrodinium aureolum (dinophygeae) in north european waters, accompanied by mortality in marine organisms. *Sarsia* 63, 123–133. doi: 10.1080/00364827.1977.10411330

Tilstone, G. H., Peters, S. W. M., van der Woerd, H. J., Eleveld, M. A., Ruddick, K., Schönfeld, W., et al. (2012). Variability in specific-absorption properties and their use in a semi-analytical ocean colour algorithm for MERIS in North Sea and Western English Channel Coastal Waters. *Remote Sens. Environ.* 118, 320–338. doi: 10.1016/j.rse.2011.11.019

Townsend, D. W., and Cammen, L. M. (1987). Potential importance of the timing of spring plankton blooms to benthic-pelagic coupling and recruitment of juvenile demersal fishes. *Biol. Oceanogr.* 5, 215–228. doi: 10.1080/01965581.1987.10749514

Vikebø, F. B., Korosov, A., Stenevik, E. K., Husebø, A., and Slotte, A. (2012). Spatio-temporal overlap of hatching in Norwegian spring-spawning herring and the spring phytoplankton bloom at available spawning substrata. *ICES J. Mar. Sci.* 69, 1298–1302. doi: 10.1038/278097a0

Vikebø, F. B., Strand, K. O., and Sundby, S. (2019). Wind intensity is key to phytoplankton spring bloom under climate change. *Front. Mar. Sci.* 6:518. doi: 10.3389/fmars.2019.00518

Wassmann, P., Reigstad, M., Haug, T., Rudels, B., Carroll, M. L., Hop, H., et al. (2006). Food webs and carbon flux in the Barents Sea. *Prog. Oceanogr.* 71, 232–287. doi: 10.1016/j.pocean.2006.10.003

Wihsgott, J. U., Sharples, J., Hopkins, J. E., Woodward, E. M. S., Hull, T., Greenwood, N., et al. (2019). Observations of vertical mixing in autumn and its effect on the autumn phytoplankton bloom. *Prog. Oceanogr.* 177:102059. doi: 10.1016/j.pocean.2019.01.001

Wilson, R. J., and Heath, M. R. (2019). Increasing turbidity in the North Sea during the 20th century due to changing wave climate. *Ocean Sci.* 15, 1615–1625. doi: 10.5194/os-15-1615-2019

Xie, J., Bertino, L., Counillon, F., Lisæter, K. A., and Sakov, P. (2017). Quality assessment of the TOPAZ4 reanalysis in the Arctic over the period 1991–2013. *Ocean Sci.* 13, 123–144. doi: 10.5194/os-13-123-2017

Zhang, Q., Warwick, R. M., McNeill, C. L., Widdicombe, C. E., Sheehan, A., and Widdicombe, S. (2015). An unusually large phytoplankton spring bloom drives rapid changes in benthic diversity and ecosystem function. *Prog. Oceanogr.* 137, 533–545. doi: 10.1016/j.pocean.2015.04.029

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Paper II

Probabilistic Models for Harmful Algae: Application to the Norwegian Coast

Edson Silva, Julien Brajard, François Counillon, Lasse H. Pettersson, Lars Naustvoll
Environmental Data Science, **In review** (2023)

II

Paper III

Forecasting Harmful Algae Blooms: Application to *Dinophysis acuminata* in Northern Norway

Edson Silva, François Counillon, Julien Brajard, Lasse H. Pettersson, Lars Naustvoll
Harmful Algae, **126** (2023)

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 Fløddevigen 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/blaskjellvarsel/). 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

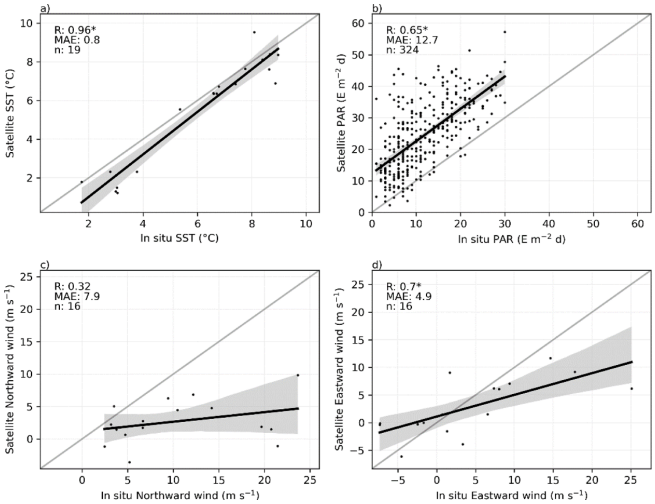


Fig. 2. Match-up between satellite estimates with in-situ data. Comparisons are shown for a) SST, b) PAR, c) Northward winds, and d) Eastward winds. Black line is the linear fit and the shaded area is the confidence interval. The R, MAE, and n are shown in the upper-left corner. The * denotes significant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

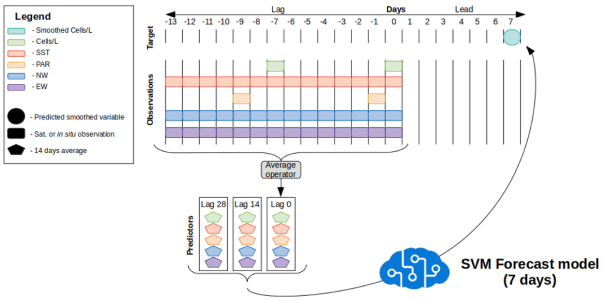


Fig. 3. Forecast model diagram. An example is given to the 7-day forecast of the combined model. The average from day 0 to lag −13 is called lag 0, and lag 14 and 28 follow the same method by averaging from −14 to −27 and from −28 to −41, respectively.

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

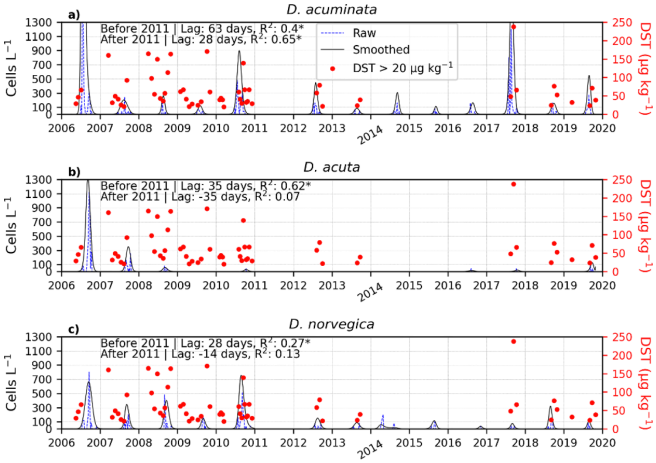


Fig. 4. The *D. acuminata* (a), *D. acuta* (b), *D. norvegica* (c), and DST (a, b, and c) time series. The Blue dashed line is the raw cells abundance, the black line is the smoothed cells abundance, and the red dots are the DST concentration (> 20 μg kg^{-1}). The R^2 is computed between each species 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

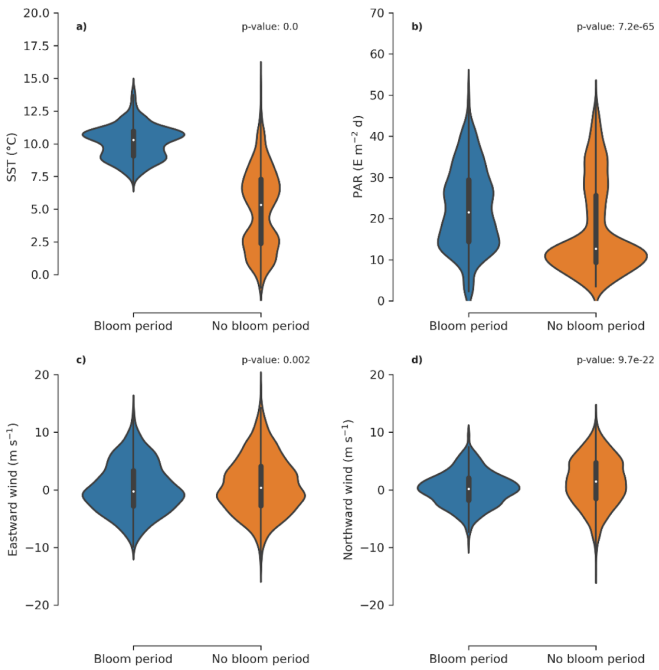


Fig. 5. SST (a), PAR (b), eastward wind (c), and northward wind (d) during *D. acuminata* bloom and no bloom periods. The p-value of a two-sided Mann-Whitney U rank test between bloom and no bloom periods is shown in the top-right corner of each subplot.

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

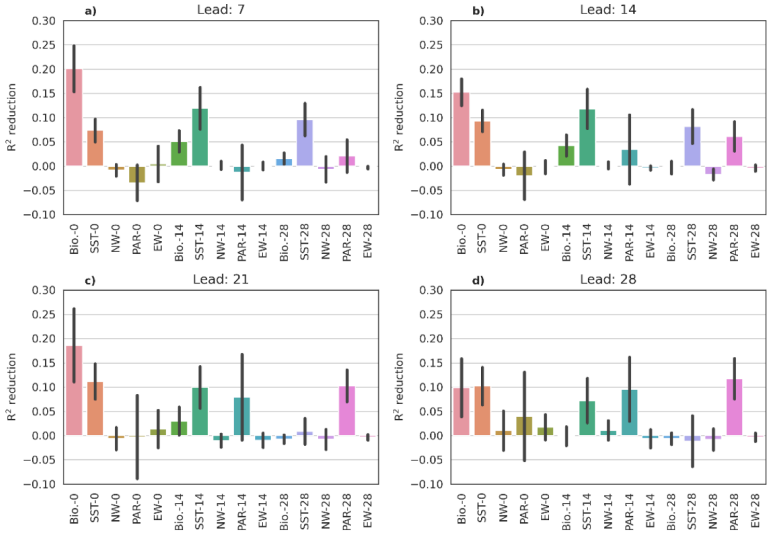


Fig. 7. The R^2 reduction after permutation of the combined model predictors in 7 (a), 14 (b), 21 (c), and 28(d) lead days. The x-axis is to the features names followed by the lag. The NW and EW denotes the northward and eastward wind components. The black lines in each bar indicate the standard deviation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

Bibliography

Aasen, J., I. A. Samdal, C. O. Miles, E. Dahl, L. R. Briggs, and T. Aune (2005), Yessotoxins in Norwegian blue mussels (Mytilus edulis): Uptake from Protoceratium reticulatum, metabolism and depuration, *Toxicon*, *45*(3), 265–272, doi:10.1016/j.toxicon.2004.10.012. 6

Albretsen, J., A. K. Sperrevik, A. Staalstrøm, A. D. Sandvik, F. Vikebø, and L. Asplin (2011), NorKyst-800 Rapport nr. 1 : Brukermanual og tekniske beskrivelser, *Tech. Rep. 2*, Institute of Marine Research. 26

Anderson, D. M., P. Hoagland, Y. Kaoru, and A. W. White (2000), Economic impacts from harmful algal blooms (HABs) in the United States, *Tech. Rep. September*, Woods Hole Oceanographic Institution. 2

Anderson, D. M., P. Andersen, V. M. Bricelj, J. J. Cullen, and J. E. J. Rensel (2001), Monitoring and Management Strategies for Harmful Algal Blooms in Coastal Waters, *Tech. rep.*, Woods Hole Océanographie Institution, Paris. 2

Anderson, D. M., T. J. Alpermann, A. D. Cembella, Y. Collos, E. Masseret, and M. Montresor (2012), The globally distributed genus Alexandrium: Multifaceted roles in marine ecosystems and impacts on human health, *Harmful Algae*, *14*, 10–35, doi:10.1016/j.hal.2011.10.012. 6

Asplin, L., J. Albretsen, I. A. Johnsen, and A. D. Sandvik (2020), The hydrodynamic foundation for salmon lice dispersion modeling along the Norwegian coast, *Ocean Dynamics*, *70*(8), 1151–1167, doi:10.1007/s10236-020-01378-0. 7

Baron, S., J. Benrekassa, D. Calavas, H. De Valk, C. Delorme, L. Goué-Benadiba, V. Henry, C. Marcé, J. Santolini, and V. Vaillant (2012), *Surveillance des infections à Campylobacter chez l'homme (entre 2003 et 2010)*, Bulletin épidémiologique hebdomadaire, Hors-Série, doi:10.1016/S0992-5945(12)71359-4. 2

Basti, L., T. Suzuki, H. Uchida, T. Kamiyama, and S. Nagai (2018), Thermal acclimation affects growth and lipophilic toxin production in a strain of cosmopolitan harmful alga Dinophysis acuminata, *Harmful Algae*, *73*, 119–128, doi:10.1016/j.hal.2018.02.004. 7, 18, 22

Bill, B., W. P. Cochlan, and V. L. Trainer (2012), The effect of light on growth rate and primary productivity in Pseudo-nitzschia australis and Pseudo-nitzschia turgidula, in *Proceedings of the 14th International Conference on Harmful Algae. International Society for the Study of Harmful Algae and Intergovernmental Oceanographic Commission of UNESCO*, pp. 78–80. 6

Bleck, R. (2002), An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates, *Ocean Modelling*, *4*(1), 55–88, doi:10.1016/S1463-5003(01)00012-9. 17

Boivin-Rioux, A., M. Starr, J. Chassé, M. Scarratt, W. Perrie, and Z. Long (2021), Predicting the Effects of Climate Change on the Occurrence of the Toxic Dinoflagellate Alexandrium catenella Along Canada's East Coast, *Frontiers in Marine Science*, *7*(January), 1–20, doi:10.3389/fmars.2020.608021. 8

Boivin-Rioux, A., M. Starr, J. Chassé, M. Scarratt, W. Perrie, Z. Long, and D. Lavoie (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*, 102,183, doi:10.1016/j.hal.2022.102183. 8

Braarud, T. (1976), The natural history of the hardangerfjord: 13. The ecology of taxonomic groups and species of phytoplankton related to their distribution patterns in a fjord area, *Sarsia*, *60*(1), 41–62, doi:10.1080/00364827.1976.10411295. 6

Brodie, J., and L. Lewis (Eds.) (2007), *Unravelling the algae: the past, present, and future of algal systematics*, 414 pp., CRC Press Taylor & Francis Group. 1

Calton, B. (2020), ESA Ocean Colour Climate Change Initiative – Phase 3 Product User Guide for v5 . 0 Dataset, *Tech. rep.*, European Space Agency. 13

Chen, W., J. Schulz-Stellenfleth, S. Grayek, and J. Staneva (2021), Impacts of the Assimilation of Satellite Sea Surface Temperature Data on Volume and Heat Budget Estimates for the North Sea, *Journal of Geophysical Research: Oceans*, *126*(5), doi:10.1029/2020JC017059. 11

Cortes, C., and V. Vapnik (1995), Support-vector networks, *Machine Learning*, *20*(3), 273–297, doi:10.1007/BF00994018. 17, 18

Cruz, R. C., P. Reis Costa, S. Vinga, L. Krippahl, and M. B. Lopes (2021), A Review of Recent Machine Learning Advances for Forecasting Harmful Algal Blooms and Shellfish Contamination, *Journal of Marine Science and Engineering*, *9*(3), 283, doi:10.3390/jmse9030283. 7, 8, 17

Eldevik, T., J. E. Nilsen, D. Iovino, K. Anders Olsson, A. B. Sandø, and H. Drange (2009), Observed sources and variability of Nordic seas overflow, *Nature Geoscience*, *2*(6), 406–410, doi:10.1038/ngeo518. 11

Evensen, G. (2003), The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics*, *53*(4), 343–367, doi:10.1007/s10236-003-0036-9. 17

FAO, IOC, and IAEA (2023), *Joint FAO-IOC-IAEA technical guidance for the implementation of early warning systems for harmful algal blooms*, FAO, Rome, doi:10.4060/cc4794en. 7

Fehling, J., D. H. Green, K. Davidson, C. J. Bolch, and S. S. Bates (2004), Domoic Acid Production by Pseudo-nitzschia seriata (Bacillariophyceae) in Scottish Waters, *Journal of Phycology*, *40*(4), 622–630, doi:10.1111/j.1529-8817.2004.03200.x. 7

Fernandes-Salvador, J. A., K. Davidson, M. Sourisseau, M. Revilla, W. Schmidt, D. Clarke, P. I. Miller, P. Arce, R. Fernández, L. Maman, A. Silva, C. Whyte, M. Mateo, P. Neira, M. Mateus, M. Ruiz-Villarreal, L. Ferrer, and J. Silke (2021), Current Status of Forecasting Toxic Harmful Algae for the North-East Atlantic Shellfish

Aquaculture Industry, *Frontiers in Marine Science*, *8*(June), doi:10.3389/fmars.2021.666583. 7

Ferreira, A. S., H. Hátún, F. Counillon, M. R. Payne, and A. W. Visser (2015), Synoptic-scale analysis of mechanisms driving surface chlorophyll dynamics in the North Atlantic, *Biogeosciences*, *12*(11), 3641–3653, doi:10.5194/bg-12-3641-2015. 17

Frigstad, H., T. Harvey, A. Deininger, and A. Poste (2020a), Increased light attenuation in Norwegian coastal waters - A literature review, *Tech. rep.*, Norwegian Institute for Water Research. 11

Frigstad, H., Kaste, A. Deininger, K. Kvalsund, G. Christensen, R. G. Bellerby, K. Sørensen, M. Norli, and A. L. King (2020b), Influence of Riverine Input on Norwegian Coastal Systems, *Frontiers in Marine Science*, *7*(June), doi:10.3389/fmars.2020.00332. 11

Frouin, R., B. Franz, and M. Wang (2003), Algorithm to estimate PAR from SeaWiFS data Version 1.2-Documentation, *NASA Tech Memo*, *206892*(22), 46–50. 14, 15

Furevik, T., M. Bentsen, H. Drange, J. A. Johannessen, and A. Korablev (2002), Temporal and spatial variability of the sea surface salinity in the Nordic Seas, *Journal of Geophysical Research: Oceans*, *107*(12), 1–16, doi:10.1029/2001jc001118. 11

García-Portela, M., P. Riobó, B. Reguera, J. L. Garrido, J. Blanco, and F. Rodríguez (2018), Comparative ecophysiology of Dinophysis acuminata and D. acuta (DINOPHYCEAE, DINOPHYSIALES): effect of light intensity and quality on growth, cellular toxin content, and photosynthesis, *Journal of Phycology*, *54*(6), 899–917, doi:10.1111/jpy.12794. 6, 19

Giesen, R. H., L. M. Andreassen, J. Oerlemans, and M. R. Van Den Broeke (2014), Surface energy balance in the ablation zone of Langfjordjokelen, an arctic, maritime glacier in northern Norway, *Journal of Glaciology*, *60*(219), 57–70, doi:10.3189/2014JoG13J063. 11

GlobColour (2020), GlobColour Product User Guide, *Tech. Rep. version 4.2.1*, GlobColour. 15

Gohin, F. (2011), Annual cycles of chlorophyll-a, non-algal suspended particulate matter, and turbidity observed from space and in-situ in coastal waters, *Ocean Science*, *7*(5), 705–732, doi:10.5194/os-7-705-2011. 15

González Vilas, L., E. Spyrakos, J. M. Torres Palenzuela, and Y. Pazos (2014), Support Vector Machine-based method for predicting Pseudo-nitzschia spp. blooms in coastal waters (Galician rias, NW Spain), *Progress in Oceanography*, *124*, 66–77, doi:10.1016/j.pocean.2014.03.003. 8

Good, S., E. Fiedler, C. Mao, M. J. Martin, A. Maycock, R. Reid, J. Roberts-Jones, T. Searle, J. Waters, J. While, and M. Worsfold (2020), The current configuration of the OSTIA system for operational production of foundation sea surface temperature and ice concentration analyses, *Remote Sensing*, *12*(4), 1–20, doi:10.3390/rs12040720. 16

Guallar, C., M. Delgado, J. Diogène, and M. Fernández-Tejedor (2016), Artificial neural network approach to population dynamics of harmful algal blooms in Alfacs Bay (NW Mediterranean): Case studies of Karlodinium and Pseudo-nitzschia, *Ecological Modelling*, *338*, 37–50, doi:10.1016/j.ecolmodel.2016.07.009. 8

Guerrini, F., P. Ciminiello, C. Dell'Aversano, L. Tartaglione, E. Fattorusso, L. Boni, and R. Pistocchi (2007), Influence of temperature, salinity and nutrient limitation on yessotoxin production and release by the dinoflagellate Protoceratium reticulatum in batch-cultures, *Harmful Algae*, *6*(5), 707–717, doi:10.1016/j.hal.2007.02.006. 6, 7

Hasle, G. R., C. B. Lange, and E. E. Syvertsen (1996), A review of Pseudo-nitzschia, with special reference to the Skagerrak, North Atlantic, and adjacent waters, *Helgoländer Meeresuntersuchungen*, *50*(2), 131–175, doi:10.1007/BF02367149. 6

Hordoir, R., C. Dieterich, C. Basu, H. Dietze, and H. Meier (2013), Freshwater outflow of the Baltic Sea and transport in the Norwegian current: A statistical correlation analysis based on a numerical experiment, *Continental Shelf Research*, *64*, 1–9, doi:10.1016/j.csr.2013.05.006. 11

Hunke, E. C., and J. K. Dukowicz (1997), An Elastic–Viscous–Plastic Model for Sea Ice Dynamics, *Journal of Physical Oceanography*, *27*(9), 1849–1867, doi:10.1175/1520-0485(1997)027<1849:AEVPMF>2.0.CO;2. 17

Huppert, D. D., and V. L. Trainer (2014), Economics of razor clam fishery closures due to harmful algal blooms in Washington State, in *Proceedings of the Workshop on Economic Impacts of Harmful Algal Blooms on Fisheries and Aquaculture. PICES Scientific Report No. 47*, pp. 59–71. 2

Jakowczyk, M., and M. Stramska (2014), Spatial and temporal variability of satellite-derived sea surface temperature in the Barents Sea, *International Journal of Remote Sensing*, *35*(17), 6545–6560, doi:10.1080/01431161.2014.958247. 11

Jauffrais, T., V. Séchet, C. Herrenknecht, P. Truquet, S. Véronique, U. Tillmann, and P. Hess (2013), Effect of environmental and nutritional factors on growth and azaspiracid production of the dinoflagellate Azadinium spinosum, *Harmful Algae*, *27*, 138–148, doi:10.1016/j.hal.2013.05.009. 6

Jin, D., E. Thunberg, and P. Hoagland (2008), Economic impact of the 2005 red tide event on commercial shellfish fisheries in New England, *Ocean and Coastal Management*, *51*(5), 420–429, doi:10.1016/j.ocecoaman.2008.01.004. 2

Jin, D., S. Moore, D. Holland, L. Anderson, W.-A. Lim, D. Kim, S. Jardine, S. Martino, F. Gianella, and K. Davidson (2020), Evaluating the Economic Impacts of HABs 2 Evaluating the Economic Impacts of Harmful Algal Blooms: Issues, Methods, and Examples, in *GlobalHAB: Evaluating, Reducing and Mitigating the Cost of Harmful Algal Blooms: A Compendium of Case Studies*, edited by V. L. Trainer, no. 59 ed., chap. 2, p. 5, PICES Sci. Rep. 2, 7

John, U., L. Šupraha, S. Gran-Stadniczeñko, C. Bunse, A. Cembella, W. Eikrem, J. Janouškovec, K. Klemm, N. Kühne, L. Naustvoll, D. Voss, S. Wohlrab, and B. Edvardsen (2022), Spatial and biological oceanographic insights into the massive fish-killing bloom of the haptophyte Chrysochromulina leadbeateri in northern Norway, *Harmful Algae*, *118*(November 2021), 102,287, doi:10.1016/j.hal.2022.102287. 7

Karlson, B., P. Andersen, L. Arneborg, A. Cembella, W. Eikrem, U. John, J. J. West, K. Klemm, J. Kobos, S. Lehtinen, N. Lundholm, H. Mazur-Marzec, L. Naustvoll, M. Poelman, P. Provoost, M. De Rijcke, and S. Suikkanen (2021), Harmful algal blooms and their effects in coastal seas of Northern Europe, *Harmful Algae*, *102*, 101,989, doi:10.1016/j.hal.2021.101989. 1, 2, 4

Kim, S., Y. Kang, H. Kim, W. Yih, D. Coats, and M. Park (2008), Growth and grazing responses of the mixotrophic dinoflagellate Dinophysis acuminata as functions of light intensity and prey concentration, *Aquatic Microbial Ecology*, *51*(3), 301–310, doi:10.3354/ame01203. 6

Kirst, G. O. (1990), Salinity Tolerance of Eukaryotic Marine Algae, *Annual Review of Plant Physiology and Plant Molecular Biology*, *41*(1), 21–53, doi:10.1146/annurev.pp.41.060190.000321. 6

Klemm, K., A. Cembella, D. Clarke, C. Cusack, L. Arneborg, B. Karlson, Y. Liu, L. Naustvoll, R. Siano, S. Gran-Stadniczeñko, and U. John (2022), Apparent biogeographical trends in Alexandrium blooms for northern Europe: identifying links to climate change and effective adaptive actions, *Harmful Algae*, *119*(September), 102,335, doi:10.1016/j.hal.2022.102335. 6

Langeland, G., T. Hasselgård, K. Tangen, O. M. Skulberg, and A. Hjelle (1984), An outbreak of paralytic shellfish poisoning in western Norway, *Sarsia*, *69*(3-4), 185–193, doi:10.1080/00364827.1984.10420605. 4

Lavergne, T., A. Macdonald Sørensen, S. Kern, R. Tonboe, D. Notz, S. Aaboe, L. Bell, G. Dybkjær, S. Eastwood, C. Gabarro, G. Heygster, M. Anne Killie, M. Brandt Kreiner, J. Lavelle, R. Saldo, S. Sandven, and L. T. Pedersen (2019), Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, *Cryosphere*, *13*(1), 49–78, doi:10.5194/tc-13-49-2019. 16

Lema, K. A., M. Latimier, Nézan, J. Fauchot, and M. Le Gac (2017), Inter and intraspecific growth and domoic acid production in relation to nutrient ratios and concentrations in Pseudo-nitzschia : phosphate an important factor, *Harmful Algae*, *64*, 11–19, doi:10.1016/j.hal.2017.03.001. 6

Li, X., J. Yu, Z. Jia, and J. Song (2014), Harmful algal blooms prediction with machine learning models in Tolo Harbour, in *2014 International Conference on Smart Computing*, pp. 245–250, IEEE, doi:10.1109/SMARTCOMP.2014.7043865. 17

Lien, V. S., S. S. Hjøllo, M. D. Skogen, E. Svendsen, H. Wehde, L. Bertino, F. Counillon, M. Chevallier, and G. Garric (2016), An assessment of the added value from data assimilation on modelled Nordic Seas hydrography and ocean transports, *Ocean Modelling*, *99*, 43–59, doi:10.1016/j.ocemod.2015.12.010. 17

Mardones, J. I., D. S. Holland, L. Anderson, V. Le Bihan, F. Gianella, A. Clément, K. Davidson, S. Sakamoto, T. Yoshida, and V. L. Trainer (2020), Estimating and Mitigating the Economic Costs of Harmful Algal Blooms on Commercial and Recreational Shellfish Harvesters, in *GlobalHAB: Evaluating, Reducing and Mitigating the Cost of Harmful Algal Blooms: A Compendium of Case Studies*, November, p. 66, PICES Sci. Rep. 2

Martino, S., F. Gianella, and K. Davidson (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*(February), 101,912, doi:10.1016/j.hal.2020.101912. 2

Merchant, C. J., and O. Embury (2014), Simulation and Inversion of Satellite Thermal Measurements, in *Experimental Methods in the Physical Sciences*, vol. 47, pp. 489–526, Elsevier, doi:10.1016/B978-0-12-417011-7.00015-5. 16

Merchant, C. J., O. Embury, C. E. Bulgin, T. Block, G. K. Corlett, E. Fiedler, S. A. Good, J. Mittaz, N. A. Rayner, D. Berry, S. Eastwood, M. Taylor, Y. Tsushima, A. Waterfall, R. Wilson, and C. Donlon (2019), Satellite-based time-series of sea-surface temperature since 1981 for climate applications, *Scientific Data*, *6*(1), 1–18, doi:10.1038/s41597-019-0236-x. 16

Mitrovic, S. M., M. Fernández Amandi, L. McKenzie, A. Furey, and K. J. James (2004), Effects of selenium, iron and cobalt addition to growth and yessotoxin production of the toxic marine dinoflagellate Protoceratium reticulatum in culture, *Journal of Experimental Marine Biology and Ecology*, *313*(2), 337–351, doi:10.1016/j.jembe.2004.08.014. 6

Morel, A., and L. Prieur (1977), Analysis of variations in ocean color1, *Limnology and Oceanography*, *22*(4), 709–722, doi:10.4319/lo.1977.22.4.0709. 14

Mourre, B., E. Aguiar, M. Juza, J. Hernandez-Lasheras, E. Reyes, E. Heslop, R. Escudier, E. Cutolo, S. Ruiz, E. Mason, A. Pascual, and J. Tintoré (2018), Assessment of High-Resolution Regional Ocean Prediction Systems Using Multi-Platform Observations: Illustrations in the Western Mediterranean Sea, in *New Frontiers in Operational Oceanography*, pp. 663–694, GODAE OceanView, doi:10.17125/gov2018.ch24. 7

Nagai, S., Y. Matsuyama, S. J. Oh, and S. Itakura (2004), Effect of nutrients and temperature on encystment of the toxic dinoflagellate Alexandrium tamarense (Dinophyceae) isolated from Hiroshima Bay, Japan, *Plankton Biology and Ecology*, *51*(2), 103–109. 6, 7

Nima, C., Frette, B. Hamre, S. R. Erga, Y.-C. Chen, L. Zhao, K. Sørensen, M. Norli, K. Stamnes, and J. J. Stamnes (2016), Absorption properties of high-latitude Norwegian coastal water: The impact of CDOM and particulate matter, *Estuarine, Coastal and Shelf Science*, *178*, 158–167, doi:10.1016/j.ecss.2016.05.012. 14

Peralta-Ferriz, C., and R. A. Woodgate (2015), Seasonal and interannual variability of pan-Arctic surface mixed layer properties from 1979 to 2012 from hydrographic data, and the dominance of stratification for multiyear mixed layer depth shoaling, *Progress in Oceanography*, *134*, 19–53, doi:10.1016/j.pocean.2014.12.005. 11

Petrenko, B., A. Ignatov, and Y. Kihai (2013a), Evaluation and selection of SST regression algorithms for S-NPP VIIRS, in *Ocean Sensing and Monitoring V*, vol. 8724, edited by W. W. Hou and R. A. Arnone, p. 15, doi:10.1117/12.2017454. 16

Petrenko, D., D. Pozdnyakov, J. Johannessen, F. Counillon, and V. Sychov (2013b), Satellite-derived multi-year trend in primary production in the Arctic Ocean, *International Journal of Remote Sensing*, *34*(11), 3903–3937, doi:10.1080/01431161.2012.762698. 17

Pettersson, L. H., and D. Pozdnyakov (2013), *Monitoring of Harmful Algal Blooms*, 9–25 pp., Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-540-68209-7. 4

Pettersson, L. H., D. Durand, O. M. Johannessen, E. Svendsen, T. Noji, H. Søiland, S. Groom, and P. Regner (2001), Monitoring and model predictions of harmful algae blooms in Norwegian waters, *International Geoscience and Remote Sensing Symposium (IGARSS)*, *3*(C), 1146–1148, doi:10.1109/igarss.2001.976773. 1

Platt, J. (1999), Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, *Advances in large margin classifiers*, *10*(3), 61–74. 21

Rial, P., M. Sixto, J. Vázquez, B. Reguera, R. Figueroa, P. Riobó, and F. Rodríguez (2023), Interaction between temperature and salinity stress on the physiology of Dinophysis spp. and Alexandrium minutum: implications for niche range and blooming patterns, *Aquatic Microbial Ecology*, *89*, 1–22, doi:10.3354/ame01994. 6, 7, 18, 19

Ribeiro, R., and L. Torgo (2008), A comparative study on predicting algae blooms in Douro River, Portugal, *Ecological Modelling*, *212*(1-2), 86–91, doi:10.1016/j.ecolmodel.2007.10.018. 17

Röder, K., F. M. Hantzsche, C. Gebühr, C. Miene, T. Helbig, B. Krock, M. Hoppenrath, B. Luckas, and G. Gerdts (2012), Effects of salinity, temperature and nutrients on growth, cellular characteristics and yessotoxin production of Protoceratium reticulatum, *Harmful Algae*, *15*, 59–70, doi:10.1016/j.hal.2011.11.006. 7

Sakov, P., F. Counillon, L. Bertino, K. A. Lisæter, P. R. Oke, and A. Korablev (2012), TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, *Ocean Science*, *8*(4), 633–656, doi:10.5194/os-8-633-2012. 17

Sathyendranath, S., R. J. Brewin, C. Brockmann, V. Brotas, B. Calton, A. Chuprin, P. Cipollini, A. B. Couto, J. Dingle, R. Doerffer, C. Donlon, M. Dowell, A. Farman, M. Grant, S. Groom, A. Horseman, T. Jackson, H. Krasemann, S. Lavender, V. Martinez-Vicente, C. Mazeran, F. Mélin, T. S. Moore, D. Müller, P. Regner, S. Roy, C. J. Steele, F. Steinmetz, J. Swinton, M. Taberner, A. Thompson, A. Valente, M. Zühlke, V. E. Brando, H. Feng, G. Feldman, B. A. Franz, R. Frouin, R. W. Gould, S. B. Hooker, M. Kahru, S. Kratzer, B. G. Mitchell, F. E. Muller-Karger, H. M. Sosik, K. J. Voss, J. Werdell, and T. Platt (2019), An ocean-colour time series for use in climate studies: The experience of the ocean-colour climate change initiative (OC-CCI), *Sensors (Switzerland)*, *19*(19), doi:10.3390/s19194285. 13

Sathyendranath, S., T. Jackson, C. Brockmann, V. Brotas, B. Calton, A. Chuprin, O. Clements, P. Cipollini, O. Danne, J. Dingle, C. Donlon, M. Grant, S. Groom, H. Krasemann, S. Lavender, C. Mazeran, F. Mélin, D. Müller, F. Steinmetz, A. Valente, M. Zühlke, G. Feldman, B. Franz, R. Frouin, J. Werdell, and T. Platt (2021), ESA Ocean Colour Climate Change Initiative (Ocean_Colour_cci): Version 5.0 Data. NERC EDS Centre for Environmental Data Analysis, doi:10.5285/1dbe7a109c0244aaad713e078fd3059a. 13

Séchet, V., P. Safran, P. Hovgaard, and T. Yasumoto (1990), Causative species of diarrhetic shellfish poisoning (DSP) in Norway, *Marine Biology*, *105*(2), 269–274, doi:10.1007/BF01344296. 5

Simon, E., A. Samuelsen, L. Bertino, and S. Mouysset (2015), Experiences in multiyear combined state–parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using the Ensemble Kalman Filter, *Journal of Marine Systems*, *152*, 1–17, doi:10.1016/j.jmarsys.2015.07.004. 26

Sperrevik, A. K., J. Röhrs, and K. H. Christensen (2017), Impact of data assimilation on Eulerian versus Lagrangian estimates of upper ocean transport, *Journal of Geophysical Research: Oceans*, *122*(7), 5445–5457, doi:10.1002/2016JC012640. 26

Thomas, M. K., C. T. Kremer, C. A. Klausmeier, and E. Litchman (2012), A Global Pattern of Thermal Adaptation in Marine Phytoplankton, *Science*, *338*(6110), 1085–1088, doi:10.1126/science.1224836. 6

Tillmann, U., B. Edvardsen, B. Krock, K. F. Smith, R. F. Paterson, and D. Voß (2018), Diversity, distribution, and azaspiracids of Amphidomataceae (Dinophyceae) along the Norwegian coast, *Harmful Algae*, *80*(August), 15–34, doi:10.1016/j.hal.2018.08.011. 6

Velo-Suárez, L., and J. Gutiérrez-Estrada (2007), Artificial neural network approaches to one-step weekly prediction of Dinophysis acuminata blooms in Huelva (Western Andalucía, Spain), *Harmful Algae*, *6*(3), 361–371, doi:10.1016/j.hal.2006.11.002. 8

Weber, C., A. K. J. Olesen, B. Krock, and N. Lundholm (2021), Salinity, a climate-change factor affecting growth, domoic acid and isodomoic acid C content in the diatom Pseudo-nitzschia seriata (Bacillariophyceae), *Phycologia*, *60*(6), 619–630, doi:10.1080/00318884.2021.1973789. 6

Xie, J., L. Bertino, F. Counillon, K. A. Lisæter, and P. Sakov (2017), Quality assessment of the TOPAZ4 reanalysis in the Arctic over the period 1991–2013, *Ocean Science*, *13*(1), 123–144, doi:10.5194/os-13-123-2017. 17

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230860182 (print)
9788230841327 (PDF)