Predicting Harmful Algal Blooms and Impacts on Shellfish Mariculture using Novel Data-Driven Approaches

Oliver Stoner
University of Glasgow https://orcid.org/0000-0003-0612-4306
Theo Economou
University of Exeter
Ricardo Torres
Plymouth Marine Laboratory
Ian Ashton
University of Exeter
Andrew Brown (ross.brown@exeter.ac.uk)
University of Exeter

Article

Keywords: Phycotoxin Concentrations, Dinophysis Species, Spatiotemporal Statistical Modelling Framework, Optimal Siting, Safe Harvesting Periods

Posted Date: July 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-668820/v1

Predicting Harmful Algal Blooms and Impacts on Shellfish Mariculture using Novel Data-Driven Approaches

Oliver Stoner[1], Theo Economou[2], Ricardo Torres[3], Ian Ashton[4], A Ross Brown[5]

[1]School of Mathematics and Statistics, University of Glasgow, UK; [2]Department of Mathematics, University of Exeter, UK; [3]Plymouth Marine Laboratory, Prospect Place, Plymouth, UK; [4]Department of Engineering, University of Exeter, UK; [5]Department of Biosciences, University of Exeter, UK.

Address correspondence to ross.brown@exeter.ac.uk

Harmful algal blooms (HABs) intoxicate and asphyxiate marine life, causing devastating environmental and socio-economic impacts costing at least $8bn/yr globally. Accumulation of phycotoxins from HAB phytoplankton in filter-feeding shellfish can poison human consumers, prompting site harvesting closures if concentrations in shellfish exceed safe levels. To better quantify both long- and short-term HAB risks, we developed novel data-driven approaches to predict phycotoxin concentrations in bivalve shellfish associated with HAB forming *Dinophysis* species. Our spatiotemporal statistical modelling framework assesses long-term HAB risks for different shellfish species in both data-rich and data-poor locations. This can revolutionise mariculture management by more confidently informing optimal siting of new shellfish operations and safe harvesting periods for businesses. Meanwhile, our machine learning framework forecasts phycotoxin concentrations further into the future than previously possible. Across 6 coastal, estuarine and loch sites, we achieve 87% overall accuracy in predicting future harvesting shutdowns 0-8 weeks ahead.

Aquaculture will play a key role in future food security. Global aquaculture production has recently overtaken capture fisheries by producing 82 million tonnes of food fish per year globally (worth US$232 billion), with projections rising to 109 million tonnes by 2030[1] . Marine aquaculture (mariculture) shows enormous potential for sustainable food production[2]. Expanding the farming of marine bivalve shellfish (currently amounting to 17.3 million tonnes, worth over $24 billion per year globally) is particularly attractive, since these filter feeders derive their food from freely available marine planktonic microalgae; shellfish also perform a wide range of ecosystem services, including nutrient regulation and carbon capture[1,3].

However, future expansion of bivalve mariculture depends on several socio-economic and environmental constraints relating to competing demands from other marine economic activities[4], environmental carrying capacities[5] and environmental hazards, including adverse climatic conditions, pollution, and Harmful Algal Blooms (HABs)[6].The increasingly frequent and widespread occurrence of HABs, proliferations of harmful planktonic algae causing intoxication and/or asphyxiation of marine life, is a major constraint on bivalve shellfish mariculture in NW European

shelf seas and other HAB hotspots around the globe[7-10]. Global annual economic impacts of HABs exceed $8 billion, approximately half due to enforced seafood harvesting bans or product recalls and half due to unavoided HAB toxin-related human health costs[11,12]. In NW Europe, dinoflagellate HAB species from the genus *Dinophysis* (*Dinophysis acuminata* and *Dinophysis acuta*) are particularly problematic, since low biomass blooms of these species >100 cells/L are sufficient to intoxicate bivalves and cause diarrhetic shellfish poisoning in human consumers[13,14]. These low biomass blooms are undetectable by satellite surveillance and, as with several other dinoflagellate HAB species, they are likely to increase in prevalence with global warming, thermal stratification, and stabilisation of the water column[7,9,15,16].

There is an immediate need for new tools to better quantify long-term and short-term changes in HAB risk, to design optimal mitigation strategies. Contemporaneous HAB surveillance, including for *Dinophysis* spp., relies on regular weekly or bi-weekly *in situ* sampling and subsequent chemical analysis of phycotoxins in shellfish, alongside microscopic quantification of HAB species abundance at routine monitoring points across regulatory networks e.g. in the UK and EU[17]. Meanwhile, predicting HAB events is highly complex, due to multiple environmental drivers varying temporally from seconds to decades; and spatially, from the microscopic scale (mm) to the mesoscale (100 km). These include physical factors driving water stratification and nutrient depletion, as well as mixing and nutrient replenishment. Ecological interactions between HAB species, other planktonic organisms and their physico-chemical environments can also be highly influential, notably resource competition, predation, and parasitism – which can invoke biochemical (allelopathic) defence mechanisms involving phycotoxins[10,18]. Phycotoxins, including *Dinophysis* toxins (Okadaic Acid (OA), *Dinophysis* toxin (DTX), Pectenotoxin (PTX)), are heat stable and not affected by cooking, which contribute to their risk to human health[13].

Existing HAB models account for spatial and temporal variation either: explicitly in the case of dispersion models[19] (Davidson et al., 2016) and mechanistic models incorporating key ecological life-history processes[20], or implicitly in the case of data models (i.e. statistical or machine learning models) that incorporate short-term and long-term trends[19,21]. Data models are simpler and less subject to structural errors at the expense of overlooking mechanistic/process detail. Moreover, coupled with probabilistic elements, data models can quantify the uncertainty associated with any estimates or predictions. Provided they are updated regularly (with new data), data models can be developed for generating short-term predictions[22], with the potential to inform short-term operational planning decisions for mariculture businesses[10]. Recent approaches have been shown to have increasing skill in making more reliable short-term predictions, notably via increasingly sophisticated smoothing functions[23] or threshold functions accounting for sudden shifts, e.g. when temperature falls below tolerable or optimal physiological limits for a HAB species[24]. Until now, accurate HAB forecasting has generally been limited to 1-2 weeks, corresponding to the general extent and accuracy of meteorological forecasting[19,23,25,26]. These existing approaches, which mostly utilise meteorological and hydrographical variables as predictors, tend to ‘over-predict’ HAB duration[19].

While reassuring for human safety, this is not so appealing to businesses waiting for harvesting bans to be lifted[6].

Here we present novel data modelling approaches to quantify HAB risk from *Dinophysis* spp., based on 10-20 years of Official Control monitoring data for a range of shellfish species at coastal sites around Scotland, SW England, and Northern France. We leverage the direct connection between these data and regulatory action levels for *Dinophysis* toxins (i.e. 160 μg/kg Okadaic Acid equivalents [OA eq.] in shellfish flesh) to better understand impacts on shellfish mariculture. To quantify long-term HAB risks, we present a statistical framework built on advanced smoothing techniques capable of evaluating spatiotemporal trends in average *Dinophysis* toxin concentrations in shellfish. Then, we present a machine learning approach combined with statistical pre- and post-processing to forecast toxin concentrations in shellfish up to 8 weeks into the future.

1 Long-term risk assessment and spatial planning

Using advanced smoothing methods based on penalised regression splines[27], we developed data models implemented on data from individual study regions incorporating medium- to long-term temporal (seasonal and interannual), spatial, and shellfish species-specific variability in *Dinophysis* toxin concentrations. The models can capture structured interactions between spatial and temporal variability in toxin concentration to quantify HAB risk in new and data poor sites, while also quantifying the relative susceptibility of different shellfish species for accumulating *Dinophysis* toxins within each national dataset. The models can therefore aid site selection through spatial estimates of toxins (Section 1.1), harvest planning through spatiotemporal estimates of toxins (Section 1.2) and species selection for different sites (Section 1.3).

1.1 Overall spatial distribution of Dinophysis toxins

Estimates of long-term average HAB toxicity can be used to indicate persistent or frequently recurring blooms and can therefore inform decisions about suitable locations for shellfish farming. Here we show the estimated effect of location on average *Dinophysis* toxin concentrations in shellfish (at the log-scale, e.g. a location with a plotted value of 2 has an $\exp(2) \approx 7.4$ times higher average toxin concentration than a location with a plotted value of 0), for South-West England (Figure 1), Scotland (Figure 2a) and France (Figure 2b).

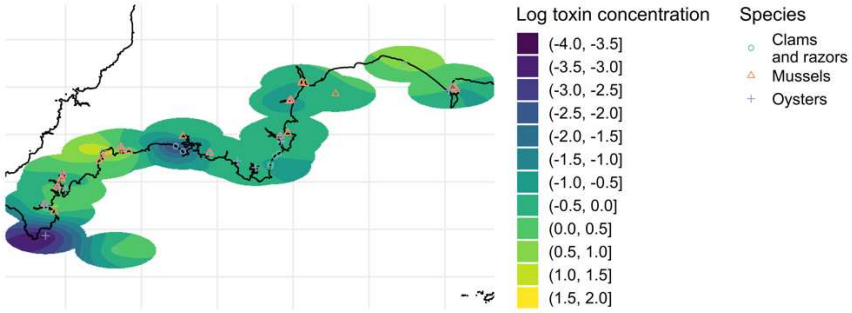

Figure 1: Estimated spatial variation in log Dinophysis toxin concentrations in South-West England. Symbols indicate location of shellfish sites and species monitored for Dinophysis toxins.

In SW England (Figure 1), *Dinophysis* HAB hotspots (highlighted in yellow) include Lyme Bay to the east and Falmouth Bay and St Austell Bay to the west. The western areas are particularly at risk, since they are sheltered to some extent from prevailing SW winds and undergo rapid sea surface warming and prolonged thermal stratification during the summer[28,29].

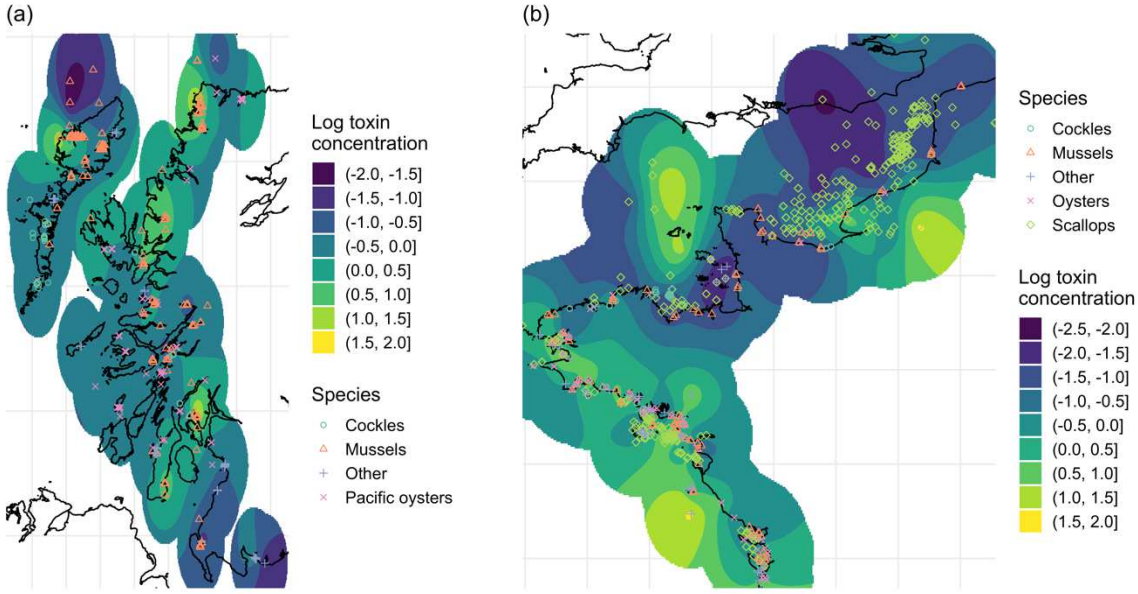

Figure 2: Estimated spatial variation in log Dinophysis toxin concentrations in (a) the West coast of Scotland and (b) the north and west coasts of France. Symbols indicate location of shellfish sites and species monitored for Dinophysis toxins.

The majority of Scottish shellfish sites (Figure 2a) are located on Scotland's west coast, and mainly harvest blue mussels (*Mytilus edulis*) and cockles (*Cerastoderma edule*). These sites are warmed by the North Atlantic Current, extending from the Gulf Stream. Regularly recurring HAB hotspots are in the relatively sheltered Clyde Sea and along the indented NW coast.

Shellfish production sites in France are depicted for the north and west coasts (Figure 2b); HAB hotspots comprise wild beds, including offshore king scallop (*Pecten maximus*) beds in the English Channel, as well as embayments along the north and west coast containing wild and cultivated scallops, Pacific oysters and blue mussels.

1.2 Seasonal variation in Dinophysis toxin concentrations

A clear understanding of systematic seasonal variability, or in other words what we expect the average toxin levels in shellfish to be at different times of the year, is essential for developing a sustainable maricultural industry, i.e. for site selection and mitigation of HAB impacts through optimal harvest scheduling (whilst also accounting for market conditions). Our models for France, Scotland and South-West England estimate overall temporal and seasonal signals in *Dinophysis* toxin concentrations at a national level, as illustrated in Figure 3.

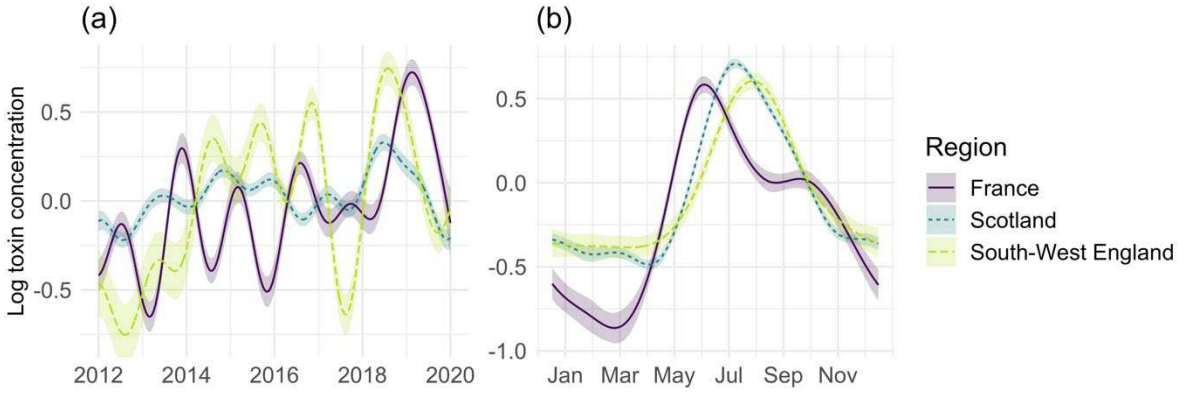

Figure 3: Estimated (a) temporal and (b) seasonal signals in the mean (log) toxin concentration for each study region (from the spatial/main effects model), with 95% confidence intervals.

In SW England, toxin concentrations indicated an increasing trend over the last decade (most visible in the peaks), interrupted temporarily in 2017, when concentrations fell substantially (Figure 3a). Consecutive seasonal (summer) peak toxin concentrations in SW England and Scotland corresponded closely in timing, while toxin concentrations in France did not synchronise with them. Furthermore, in France there was some evidence of a repeating pattern of a larger followed by a smaller toxin spike in consecutive years. Nevertheless, maximum toxin concentrations were recorded from all three data series in 2018, one of the warmest summers on record in NW Europe[30].

The seasonally averaged signals for Scotland and SW England were similar, displaying an approximately symmetrical shape with single peaks occurring around July-August, while the highest toxin concentrations for France occur around June, followed by a further small increase/inflection in the Autumn (Figure 3b). Examining space-season interactions (Figure 4), our model showed that peak toxicity occurs in July along France's Atlantic coast, while on the north coast the peak occurs later in the summer and early autumn.

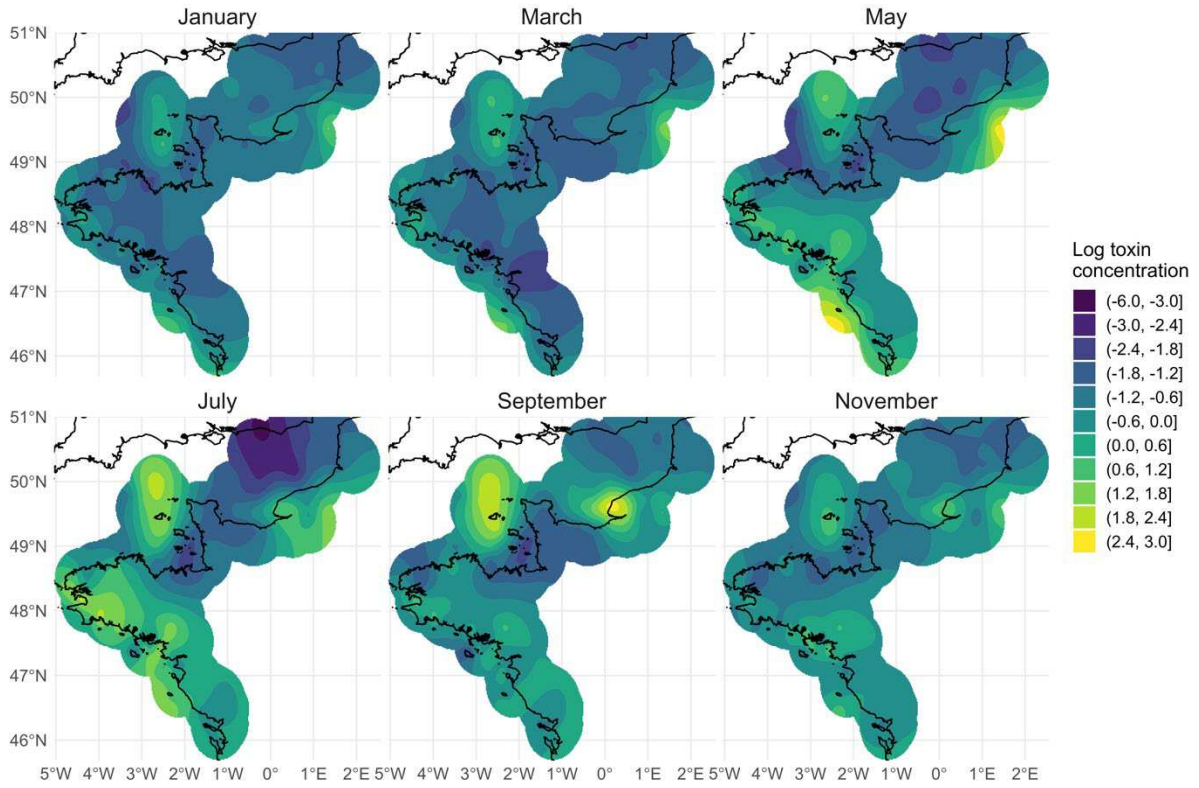

Figure 4: Estimated average (relative) toxin levels in France at different times of the year, from the spatiotemporal model for French Dinophysis toxin data.

1.3 Vulnerability of shellfish species and spatial planning

Our models also estimated the average *Dinophysis* toxin concentrations for each study region in different shellfish species relative to the average toxin concentration in blue mussels (Figure 5). By considering both the point estimates and uncertainty (95% confidence intervals, shown as error bars), we can draw conclusions about the risks associated with the farming of different shellfish species. We only compared estimates for species with at least 200 data points in total across all regions. Several species exhibited similar average levels of intoxication to mussels, including Dog cockles, King scallops and Queen scallops. Meanwhile, Manila clams, Pacific oysters and common cockles in France exhibited substantially lower average toxin concentrations than blue mussels. Both native oysters in Scotland and abrupt wedge shell clams (*Donax trunculus*) in France exhibited substantially higher average toxin levels compared to mussels.

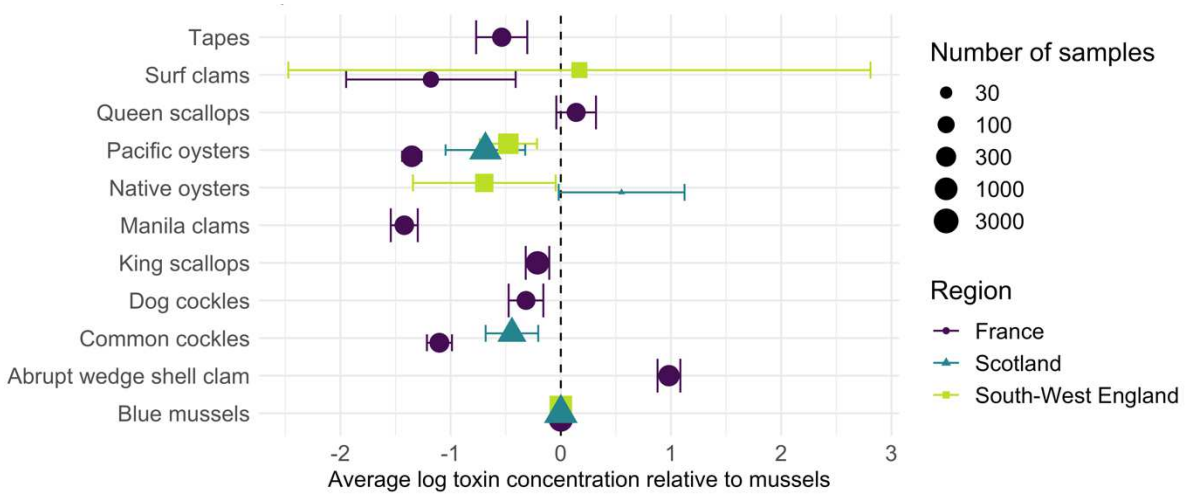

Figure 5: Estimated species effect (variation in Dinophysis toxin concentrations across shellfish species) from the long-term spatial risk models for France, Scotland, and South-West England. Effects are shown relative to the point estimates for blue mussels (Mytilus edulis) for each region. Points show point estimates and error bars show 95% uncertainty intervals. The size of the point relates to the number of unique observations for each species, separately for each region.

Estimating the relative *Dinophysis* toxin concentrations for different shellfish species provides the opportunity to assess the potential change in HAB risk associated with the farming of alternative species at existing sites. For example, for Antifer ponton pêche on France's north coast, *Dinophysis* toxin concentrations in blue mussels, consistently exceeded the action level (160 μg/kg OA eq., requiring harvesting closure) between 2010 and 2018, often for a month at a time (Figure 6). Our model suggests that, if it were possible for Pacific oysters or common cockles to be farmed here instead, the toxin concentration, and consequently the risk of disruption to harvesting, could be substantially reduced (Figure 6). Meanwhile, harvesting of abrupt wedge shell clams here could suffer from longer periods of enforced closure on average.

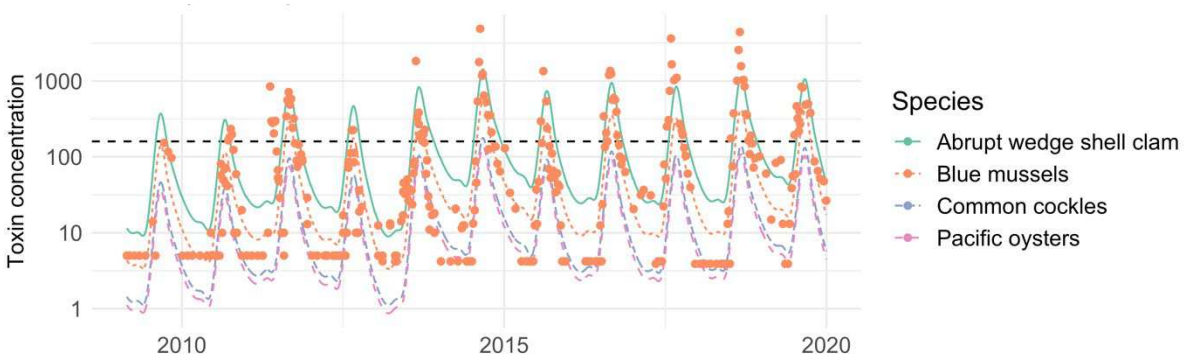

*Figure 6: Estimated mean toxin concentration at **Antifer ponton pêche** for different species (lines). Points show the observed concentrations from blue mussel samples and the horizontal dashed line shows the toxin threshold above which harvesting is shut down.*

2 Forecasting harvesting closures

Using only short- to medium-term histories of sea surface temperature and *Dinophysis* abundance (cell counts in water), lagged at intervals of 0-12 weeks as inputs to a series of machine learning (random forest) models, we predicted future *Dinophysis* toxin concentrations for different forecasting lead times (0 to 8 weeks in the future). Models were trained for prediction at each of the following geographically

diverse sites in SW England: St Austell Bay (coastal); Fowey (Ria estuary); Lyme Bay (offshore); and Scotland: Loch Eishort (inland sea loch); Loch Laxford (coastal sea loch); and Sandsound Voe.

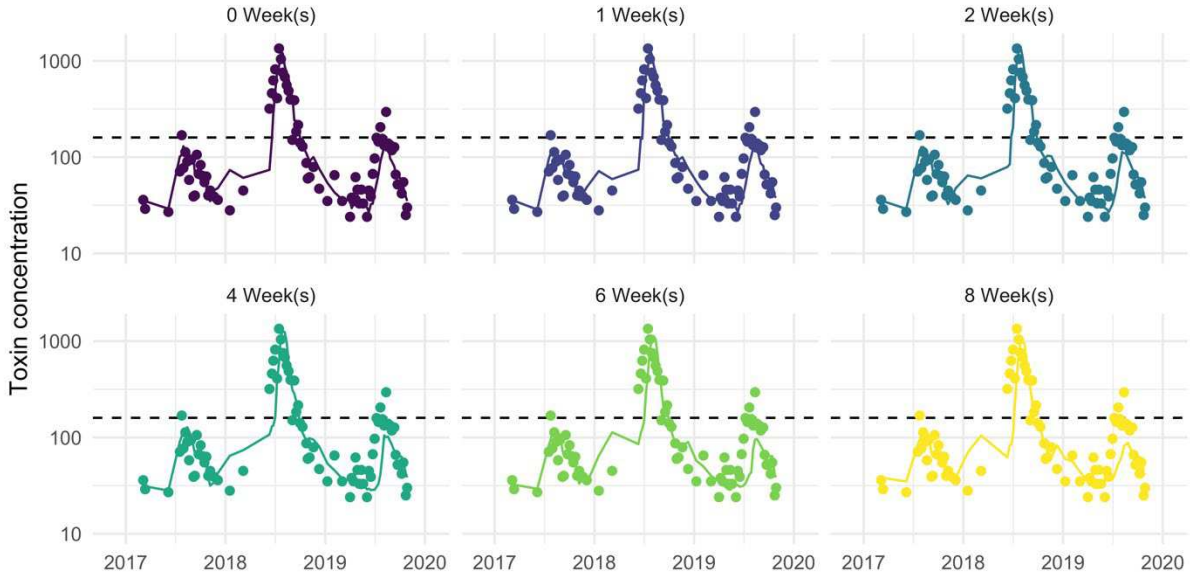

*Figure 7: Predicted toxin concentrations from the machine learning algorithm (lines) for **Sandsound Voe**, where dots represent measured data. Each panel displays predictions from a specific forecasting lead time (e.g. 2 weeks). The horizontal dashed lines show the regulatory harvesting closure threshold (160 µg/kg OA equivalents).*

Predicted *Dinophysis* toxin concentrations are plotted for Sansound Voe, Shetland Figure 7over the 2018 and 2019 HAB seasons in Figure 7. In 2018, the measured toxin concentration reached a peak of 1346 µg/kg OA eq in July and exceeded the harvesting closure threshold for over 3 months. In contrast, the bloom in 2019 was far less intense, reaching a peak of only 295 µg/kg OA eq in August. Similar plots are available for other study sites in the Supporting Information. By combining the information from the temperature and *Dinophysis* spp. abundance, the machine learning algorithm can predict future patterns of *Dinophysis* toxin concentrations in shellfish in relation to bloom parameters, including: i) the 'growth' versus 'decay' phase for temperature and cell count; ii) bloom intensity versus threshold cell count (trigger level = 100 cells/L) and versus threshold toxin concentration (action level = 160 µg/kg OA eq); and iii) seasonal trends. For all lead times, the models for Sandsound Voe were able to capture the differences in intensities between the two annual HAB events. Predictions of the toxin concentrations were particularly impressive in 2018, with predictions closely matching the unknown future observations, even when predicting 8 weeks into the future (Supplementary Information).

However, some short-term discrepancies remained between predicted and observed *Dinophysis* toxin concentrations. For example, the models for Sandsound Voe systematically under-predicted the toxin level in the 2019 HAB season (Figure 7), such that they failed to foresee a harvesting closure. Minimising these discrepancies is vital to be able to accurately predict whether toxin concentrations will exceed or fall short of the harvest closure threshold.

2.1 Postprocessing and measures of uncertainty

Figure 7A statistical “post-processing” algorithm was employed to obtain more accurate predictions of *Dinophysis* toxin concentrations in shellfish. Specifically, we fit a statistical model capturing systematic changes over time in the prediction errors from the machine learning algorithm. Using this model, we predicted the error of a new forecast and obtained a corrected prediction of intoxication. Finally, we measured the uncertainty by computing the standard error of these corrected predictions (see Supporting Information). Figure 8 shows the close concordance between mean predicted (line) and measured (dots) log *Dinophysis* toxin concentrations in blue mussels farmed in Sandsound Voe in NW Scotland. Importantly, the under-prediction in 2019 was eliminated by the correction, such that the risk of harvesting closure was then correctly identified. For this site, we also note that the uncertainty in predictions increased marginally from 0 to 1 week but did not sizeably increase from 1 to 8 weeks into the future.

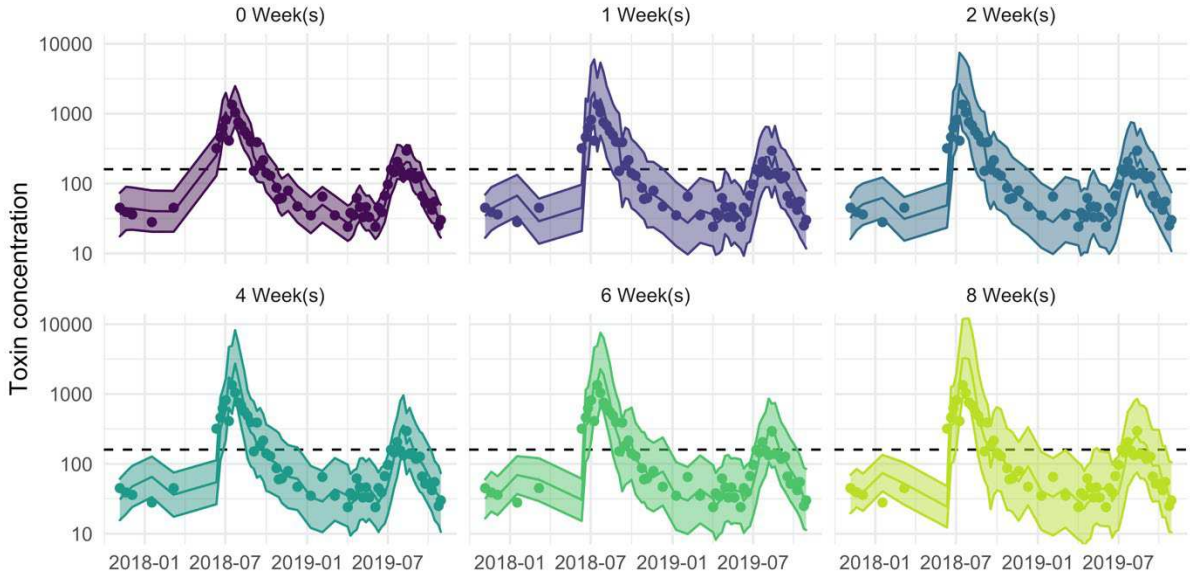

*Figure 8: Corrected predictions of toxin concentrations in **Sandsound Voe**, obtained using the machine learning algorithm and the statistical postprocessing algorithm, with each panel displaying predictions from a specific forecasting lead time (e.g. 2 weeks). Upper/lower bounds are the corrected predictions plus/minus twice the estimated standard error. The horizontal dashed lines show the regulatory harvesting closure threshold (160 µg/kg OA equivalents).*

2.2 Accuracy and specificity of predictions

We assessed the out-of-sample (nowcasting/forecasting) performance of the “raw” predictions from the machine learning algorithm and the corrected predictions using a suite of quantitative metrics. These are: the root mean-squared error (RMSE); R^2; the true positive prediction rate (proportion of predictions which correctly identify the toxin concentration as being greater than the harvesting closure threshold); and the false positive prediction rate (proportion of predictions which incorrectly identify the toxin concentration as being greater than the threshold).

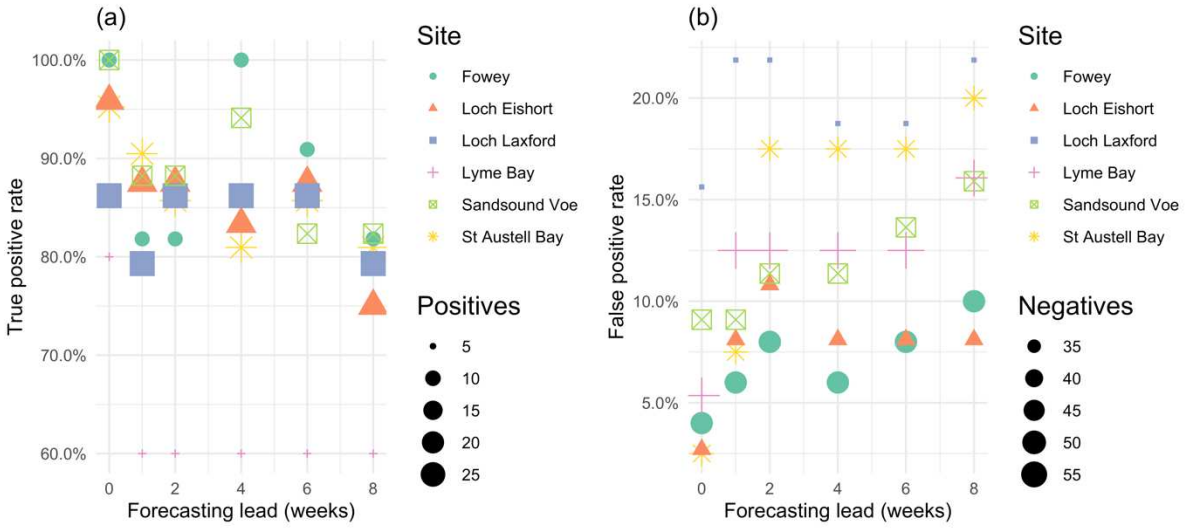

Figure 9: True positive prediction rate (a) and false positive prediction rate (b) (proportion of predictions which correctly/falsely identify the toxin concentration as being greater than the harvesting closure threshold). Diameter of symbol represents in (a) the number of threshold exceedances (positives) and in (b) the number of threshold deceedances (negatives).

For these study sites, excluding Lyme Bay where *Dinophysis* toxin concentrations rarely exceeded harvesting thresholds, true positive prediction rates ranged from about 80% to 100% when forecasting up to 6 weeks ahead, and from about 75% to 85% 8 weeks ahead (Figure 9a). The overall true positive rate for all sites and lead time was 87%. Meanwhile, false positive prediction rates ranged from about 0% to 15% when nowcasting and about 5% to 20% when forecasting up to 8 weeks ahead (Figure 9b).

3 Discussion

3.1 A step change in HAB risk assessment and forecasting

Better quantifying and predicting the risk of HABs is vital for sustaining the future operation and growth of shellfish mariculture. Addressing a need for more advanced risk assessment tools, we have developed novel data-driven approaches to exploit standardised HAB monitoring data for shellfish production. Models based on our proposed frameworks could be deployed to benefit HAB and biotoxin alert systems in Scotland[26], SW England, Northern France and elsewhere in Northern Europe, including Spain and Portugal and other key shellfish harvesting areas in the Atlantic ARC and Mediterranean, which are plagued by *Dinophysis* HABs.

Using advanced multidimensional smoothing structures available within the Generalized Additive Model framework, we have developed models which explicitly account for spatiotemporal variability and interactions (e.g. space-season) in *Dinophysis* toxin concentrations in shellfish. These models can predict average annual HAB toxin risk profiles at new and data-poor sites, for a variety of shellfish species, offering the potential to revolutionise marine spatial planning by identifying: a) recurring HAB hotspots versus lower risk areas and b) opportunities for optimisation of existing mariculture, through changes in farmed shellfish species (informed by the weight of evidence across regions) and seasonal harvesting schedules. For example, the French dataset showed earlier onset of *Dinophysis* blooms and elevated toxin levels in shellfish on the west coast and highlighted higher

toxin concentrations in certain shellfish species, such as the abrupt wedge shell clam (*Donax trunculus*). Conversely our model indicated lower HAB risk associated with harvesting of Pacific oysters or common cockles, with the potential to strategically target the cultivation and collection of alternative species, as we illustrated for the north coast of France (Figure 6).

Furthermore, using a combination of machine learning and statistical models we have demonstrated previously unseen levels of accuracy when forecasting HAB toxins in shellfish up to 8 weeks into the future, for a geographically diverse set of 6 study sites in SW England and Scotland. Our longer-term forecast modelling provides new opportunities for mitigation activities requiring more advanced warning, notably targeting *in situ* end-product testing and informing site-specific harvest scheduling, e.g. directing harvesting activities to alternative locations until toxins have purged (depurated) from shellfish. Crucially for practicality of daily operational use, the computational run time for these algorithms is just a few seconds per site, while requiring very basic CPU, memory, and storage resources.

The power and wide applicability of our models stems from the integration of substantial (10-20 year) EU-standardised, national shellfish monitoring data. These publicly available data are updated on a weekly (or bi-weekly) basis, which allows for continual model recalibration, assuring the validity of spatial interpolations and short- to long-term future predictions of HAB toxin accumulation in shellfish. While we acknowledge that multi-decadal data are required to discern long-term trends and confirm HAB drivers (causal factors), holding out for extended time-series data (>10-20 years) won't necessarily help anticipate unforeseen future changes. For example, there have been notable increases in the occurrence of HABs following major hydroclimatic regime shifts, such as North Atlantic Oscillation switching abruptly (NAO+) and generating stronger westerly winds and milder, wetter winters since the mid-1980s in the North-East Atlantic and North Sea[16,30,31]. Here we take the pragmatic approach of constantly updating contemporary signals, as well as reviewing of longer-term signals, based on *Dinophysis* spp. abundance and other readily available proximate predictor variables - here sea surface temperature (obtained daily) - resulting in compelling prediction accuracy.

3.2 Scope to improve HAB risk assessment and forecasting

Predicting low biomass HABs using models is a key priority because their progression is difficult to track by traditional microscopic methods or by high resolution Earth observation systems[32]. Furthermore, low biomass HAB species including *Dinophysis* spp. can bloom 'under the radar' for protracted periods with limited effects from density-dependent predation, parasitism and competition for resources (compared to high biomass blooms)[15]. Nutrient availability is not a major limitation for mixotrophs like *Dinophysis* spp. since they are able to switch between autotrophy (photosynthesis) and heterotrophy (predating other microalgae)[33,34]. Given the small ecological footprint and trophic plasticity of *Dinophysis spp.*, we were able to predict HAB risk with ~87% accuracy (on average) based on histories of only two 'essential' input variables: HAB cell count and sea surface temperature. However, we can foresee situations where more input variables may be needed: i) At sites where availability of HAB cell count data is not available in sufficient quantities, we could include additional variables capturing the conditions that lead to

changes in cell abundance, e.g. physical advection, related to wind speed and direction and upwelling[23,35,36];
ii) Other low biomass HAB species might also be predictable from a limited number of variables, but it is likely that accurate prediction of high biomass HABs with larger "ecological footprints" may require a wider range of environmental predictor variables[19];
iii) For some HAB species (including both low and high biomass species), cell abundance may not be directly proportional to HAB toxin accumulation in shellfish. This disconnect may occur for example for *Pseudo-nitzschia*, *Alexandrium* and *Prorocentrum* species, and has been attributed to variations in environmental cues stimulating toxin production/release, including temperature, irradiance, nutrient levels, and abundances of predators[15,18].
iv) Different shellfish species have different propensities for accumulating HAB toxins, which may be exacerbated by variation in site-specific environmental conditions affecting toxin production, as well as uptake and depuration by shellfish e.g. temperature, chlorophyll and turbidity levels, which can affect ventilation, filtration, egestion, metabolic and excretion rates in shellfish[6,37].

Moreover, while we succeeded in separately tackling the aims of spatially quantifying long-term risk from HABs and generating more accurate forecasts of HAB impacts on shellfish quality at relatively data-rich individual sites, future modelling should combine these efforts to generate forecasts for data-poor sites, by pooling information about recent HAB events and the effects of predictor variables across sites. Importantly for achieving this goal and for addressing situations i) to iv), the flexibility and generality of the approaches presented here mean it is very straightforward to include additional inputs. For example, site-specific data, including exposure and water depth, which govern mixing versus stratification, may significantly increase predictive power for several HAB species[15].

Finally, we acknowledge there are several avenues for improving data collection to benefit HAB risk assessment. First, resolution of *Dinophysis* bloom dynamics could be improved by acquiring real-time data from *in situ* sensors capable of near real-time quantification of changing HAB toxin concentrations and HAB species abundance via cytometric or molecular-based methods[38-40]. Prediction of *Dinophysis* toxin concentrations in shellfish could also potentially be improved by discerning cell counts for prominent species, which have different toxin profiles – including *Dinophysis acuminata* and *Dinophysis acuta*, which are most regularly associated with intoxication of bivalve shellfish in key shellfish growing regions from NW Scotland and Scandinavia to the Iberian Peninsula and the Mediterranean[35,36,41].

So far, only a limited number of real-time monitoring systems have been implemented in national or regional HAB surveillance programmes: NOAA's HAB operational forecast system for the Gulf of Mexico[42,43]; the Autonomous Ocean Sampling Network including Monterey Bay[44], the Gulf of Maine[45], and the Hong Kong coast[46]. There is huge scope to expand and advance HAB monitoring systems, but this work should be undertaken in tandem with the advancement of data-driven models that can help direct data acquisition to where this is most needed (i.e. HAB hotspots) and to ultimately exploit these data in terms of HAB risk forecasting.

4 Supplementary Information

Spreadsheet **Data Summary.xlsx** quantifies the extent of the phycotoxin data used in this analysis and spreadsheet **Performance.xlsx** contains all predictive performance metrics for the 6 trial sites (as detailed in Sections 2.2 and 5.2). Plots of smoothed temperature and *Dinophysis* abundance input variables are (**Smoothed.pdf**) and plots of raw and corrected predictions for the 6 trial sites (**Predictions.pdf**) are provided for download. Finally, all code and data used in this analysis are provided for download in a single ZIP file **Code and Data.zip**.

5 Methods

Standardised HAB monitoring data used in this analysis span 12 years (2008-2020) and include a total of 302 sites and at least 19 shellfish species farmed or collected in Scotland, SW England and Northern France (Data Summary). At routine monitoring points for each of these sites, HAB toxins are measured in shellfish meat each week during blooms, in conjunction with measurements of HAB species abundance according to official control methods stipulated in EU Hygiene Regulation (EC) No 853/2004 (EC, 2004). In the following subsections we leverage this data to develop models for assessing the long-term risks of HABs (Section 5.1) and models for nowcasting and forecasting toxin concentration, in combination with temperature data and *Dinophysis* spp. abundance data (Section 5.2).

5.1 Spatiotemporal risk assessment models

For assessment of long-term average toxin levels from HABs, we have developed statistical/probabilistic models within the Generalized Additive Model[27] (GAM) framework. First, we assume that toxin samples y can be modelled using a log-Normal distribution (i.e. $\log(y) \sim \text{Normal}(\mu, \sigma^2)$). We can then characterise systematic variability in average toxin levels by including spatio-temporal structures and other important variables in the mean of this distribution. These structures are defined as linear combinations of basis functions, typically smooth spline functions of one or more variables. The functions are estimated under a smoothness constraint to prevent overfitting, which is optimised using a measure of out-of-sample predictive power. Analysis is based on the estimated functions and associated measures of uncertainty (e.g. 95% confidence intervals), and predictions are based on estimates of μ.

The power of the GAM framework for this application is that by specifying functions of time and space (longitude and latitude) in the model we can account for spatiotemporal variation in phycotoxin levels. For instance, seasonal cycles in HAB occurrence may vary substantially over a large geographical region, or the risk from HABs may increase or decrease at different locations due to changes in ocean regimes or climate. Taking these spatiotemporal structures into account can better inform decisions about geographical expansion/diversification of mariculture. Specifically, we can more reliably estimate HAB risk at data-poor sites because: a) the model is more parametrically constrained compared to models where temporal variability is assumed independent for each site; and b) it is possible for data-poor sites to borrow information on HAB risk from nearby data-rich sites.

For a toxin measurement $y_{t,s,i,j}$ observed on day $t = 1, \ldots, \mathrm{N}$ (with day in the year $d = 1, \ldots, 365$), at geographical location $s \in S$, in species i (e.g. mussels, oysters) and in production\harvesting area $j = 1, \ldots \mathrm{J}$, our spatial risk model is given by:

$$\log\left(y_{t,s,i,j}\right) \sim \mathrm{Normal}\left(\mu_{t,s,i,j}, \sigma_y^2\right)$$

$$\mu_{t,s,i,j} = \mathrm{intercept} + \mathrm{species} + f_1(\mathrm{time}) + f_2(\mathrm{season}) + f_3(\mathrm{space}) + f_4(\mathrm{space}, \mathrm{season}) + f_5(\mathrm{space}, \mathrm{time})$$

First, the species factor effect captures overall differences in the uptake of HAB toxins by different shellfish species, with blue mussels acting as the baseline (intercept) species for all regions (Figure 5). Then, $f_1(\mathrm{time})$ is the overall temporal trend (Figure 3a) for the region of interest, intended to capture the overall inter- and intra-annual signal for each study region. For example, it is this function we could examine for any possible evidence of long-term trends in HAB intensity associated with climate change. Here we define $f_1(\mathrm{time})$ as a thin-plate spline of time (in days) with 50 knots spread evenly across the time series, which is more than adequate to capture general 'good' and 'bad' periods, as well as longer term trends. The specific basis representation chosen imitates a 1D Gaussian process with a Matérn covariance function[27,47]. The next function $f_2(\mathrm{season})$ is the overall 'seasonal' signal (Figure Figure 3b), i.e. what happens on average in the region depending on the time of year. This is intended to capture the average timing and intensity of HABs in the calendar year. We define $f_2(\mathrm{season})$ as a cyclic cubic spline of calendar day (cyclic means the value is the same for day 0 and day 365), with 10 knots spread evenly across the calendar year. Spatial effect $f_3(\mathrm{space})$ is specified as a 2D thin-plate spline of longitude and latitude. The specific basis representation imitates a 2D Gaussian process in space with an isotropic Exponential function covariance function[27]. For this function, we specified varying numbers of knots for SW England (50), Scotland (100) and France (100), reflecting differences in spatial extent. It is this function we examine in Figure 1 and Figure 2 to draw conclusions about the geographical distribution of HAB risk.

In the GAM framework it is possible to capture nonlinear interactions between 2 or more variables. One option is to specify isotropic smooth functions (i.e. the smoothness is invariant to rotation of the coordinate axes[27]), however this isotropy may be inappropriate for interactions involving variables with very different scales e.g. space and time. A section option which we will use here is to use tensor products which interacts smooths across multiple variables[48], resulting in different smoothness penalties for each dimension. First, $f_4(\mathrm{space}, \mathrm{season})$ interacts a 2D thin plate spline (50 knots for France and Scotland, 25 for SW England) of latitude and longitude with a 1D cubic cyclic spline of calendar day (10 knots), resulting in a 3D function which captures spatial variability in the seasonal signal (e.g. Figure 4). Similarly, $f_5(\mathrm{space}, \mathrm{time})$ combines a 2D thin-plate spline of longitude and latitude (10 knots) with a 1D thin-plate spline (10 knots) of time (days), to capture spatial variability in long-term temporal trends.

To assess the relative importance of the different functions/effects, we can decompose fitted values μ into each term and compute the percentage of the variance of the toxin measurements y explained by each term. For France, the most important term was the spatial function f_3 (22%), followed by the species effect

(16%) – reflecting the wide variety of species being cultivated in France – and the space-season interaction f_4 (12%). In SW England, the spatial function f_3 was also most important (19%), followed by the function of time f_1 (16%) – reflecting the significant interannual variability for SW England as seen in Figure 3 – followed by the space-season interaction f_4 (15%). Finally, in Scotland the spatial and space-season terms were tied as the most important (19% each), followed by the seasonal function f_2(17%).

5.2 Nowcasting and forecasting

The general concept of our approach to nowcasting and forecasting is to model *Dinophysis* toxicity in shellfish tissue as an estimated function of two inputs: sea surface temperature and *Dinophysis* spp. abundance (total cell counts measured per litre of seawater). The daily averaged surface sea temperature was extracted from the Copernicus Marine Environment Monitoring Service reanalysis ocean product NWSHELF_MULTIYEAR_PHY_004_009 version 5 produced by the UK Met Office. The system is the Forecasting Ocean Assimilation Model 7 km Atlantic Margin model (FOAM AMM7), based on version 3.6 of the Nucleus for European Modelling of the Ocean (NEMO) ocean model code[49]. The model domain resolves all shelf seas surrounding the UK and includes all the sites considered here at 7km resolution in the horizontal and with the vertical dimension discretised into 51 vertical layers. The reanalysis includes daily assimilation of satellite derived sea surface temperature (SST) and profiles of temperature and salinity assimilated using version 6 of NEMOVar[50]. For each shellfish site, daily temperature values were extracted for the nearest grid cell for which temperature data is available.

The 3 Scottish trial sites (Loch Eishort, Loch Laxford and Sandsound Voe) were selected by first quantifying data availability for all sites by ranking the product of the number of toxin samples and the number of abundance samples, and then selecting the 3 sites with the highest data availability. The 3 trial sites in SW England were selected by considering both data availability and diversity of site location and type (i.e. offshore vs coastal vs estuarine). For nowcasting and forecasting, all sites were modelled separately, so the remainder of this section details the approach for an individual site.

Both short-term and long-term histories of sea surface temperature and abundance have the potential to be informative. However, linking toxin observations with previous *Dinophysis* cell counts is challenging due to large gaps and significant noise in the time series of abundance measurements. To tackle both issues, we fit the (log) abundance values using a simple Generalized Additive Model[27], consisting of smooth functions of time and time of year. The same approach is applied to the temperature values to smooth out any short-term “spikiness”, which was occasionally exploited by some models to overfit the data. This results in continuous “smooth” time series of temperature and abundance, which do not strictly interpolate the observed values but instead aim to reflect the underlying “average” trend (reference supporting information). Starting 12 weeks after the earliest abundance observations, we then create lagged versions of the smoothed temperature and abundance variables, with the lag increasing up to 12 weeks in steps of 1 week. For example, the 2-week lagged temperature variable is equal to the value of the smoothed temperature variable 2 weeks ago. The result of this procedure is 13 temperature variables and 13 abundance variables.

To accurately predict toxin values using (up to) all 26 of these variables as inputs, we require a modelling framework which can account for complex interactions between inputs. One example of a possible interaction effect is if the differences between the lagged abundance variables can predict the rate of change in the toxin. Random forests are a very straightforward option to achieve capture complex interactions, while also being robust to correlated inputs of potentially different scales. Random forests optimally discretise the multi-dimensional space of the inputs, into contiguous "regions" where the predicted quantity is homogeneous. In its most basic form, it can be thought of as a moving window approach, albeit a robust one by being applied to different subsamples of the data. To fit random forests in R using the *ranger*[51] package, we need only specify the output variable, here the logarithm of the toxin values, and the inputs, here the lagged variables. There is no need to specify any interactions when defining the model. The only argument we modified was the number of trees, which we increased from 500 to 1000, which has the effect of slightly decreasing the error of the random forest at the expense of computation time.

When making future predictions we must consider that future values of abundance and temperature (where site-specific forecasts of sea surface temperature are unavailable) will be unknown. Specifically, values of our lagged variables for lags less than the forecasting lead time will be unknown. Therefore, the forecasting experiment presented in Section 2 was constructed by specifying distinct random forest models for each lead time, where each model includes only lagged variables which would be observed for that lead time. We specified models with a lead time of 0 weeks (nowcasting), 1 week, 2 weeks, 4 weeks, 6 weeks, and 8 weeks. A final model was fit with all lags treated as observed, to investigate how well the model might be able to predict toxicity if accurate forecasts of abundance and temperature were available. These models are implemented in a loop to generate predictions for the last 80 toxin observations in the time series.

Out-of-sample prediction errors (defined as the observed value minus the prediction) for each lead time are stored as the prediction experiment progresses through the time series. Using a further Generalized Additive Model with a spline function of time to be estimated (one knot for every 2 observations), we can model the errors to estimate any short-term systematic bias in the forecasts. Then, we can predict the error of the next forecast and add together the predicted error and the predicted toxin value, to obtain a "corrected" prediction. This approach is similar to how statistical post-processing methods are used to correct numerical weather predictions. Again, we consider the lead time, so we only provide the GAM with errors which would be known when making the next prediction. The post-processing model is only implemented after at least 20 predictions have been made by the machine learning algorithm, so that enough data is available to begin predicting new errors. Finally, we compute a basic measure of out-of-sample uncertainty by calculating the sample standard deviation of the new errors, defined as the observed value minus the corrected prediction.

5.3 Handling of missing values

For the French phycotoxin data, we assume that the entire distribution of toxin concentrations has been observed, noting that around half of measurements are low

values in the single figures. In both the UK phycotoxin datasets, particularly for SW England, a high proportion of observations were recorded as '<RL' – around 80% of the Scottish data and around 90% of the SW English data – meaning the toxin concentration was too low to be detected. Nonetheless, these observations can provide valuable information on the periods of the time and locations where the toxin level was low. To accommodate such data in our spatiotemporal models for SW England and Scotland, as well as in our forecasting models for sites in SW England, we replaced all <RL entries with a fixed value below the lowest measured value, 10 µg/kg OA. For the Scottish 3 Scottish trial sites (Loch Eishort, Loch Laxford and Sandsound Voe), on the other hand, we obtained the best forecasting performance when all '<RL' observations were discarded. Prediction performance metrics RMSE and R^2 were calculated using only detected concentrations. All other missing values were discarded.

For the Scottish *Dinophysis* spp. abundance data, we discarded values equal to 0. For the SW English *Dinophysis* spp. abundance data, we replaced values below the detection threshold (recorded as 'ND') with a fixed value below the lowest measured value, which we chose to be 20 cells per litre. All other missing values were discarded.

Author Contributions

Stoner undertook data-driven modelling and preparation of the manuscript. The following authors also contributed to the manuscript: Economou provided expert input on the study and statistical framework design; Ashton, Torres and Brown provided expert input on climatic, oceanographic, and biogeochemical covariates. Brown conceived and coordinated the study.

Competing Interests

The authors declare no competing interests.

Acknowledgements

The authors gratefully acknowledge funding from the European Maritime and Fisheries Fund (ENG3103); Turing Pilot Research Grant (260320) and IIB Open Innovation Project Fund (115717).

1 Food and Agriculture Organization. The State of World Fisheries and Aquaculture 2020. Sustainability in action. (Rome, 2020).

2 Costello, C. *et al.* The future of food from the sea. *Nature* **588**, 95-100, doi:10.1038/s41586-020-2616-y (2020).

3 van der Schatte Olivier, A. *et al.* A global review of the ecosystem services provided by bivalve aquaculture. *Reviews in Aquaculture* **12**, 3-25, doi:https://doi.org/10.1111/raq.12301 (2020).

4 Daniels, C. *et al.* Supporting Mariculture Development: Evidence for Informed Regulation. Policy Brief. (2020).

5 Ferreira, J. G. *et al.* Analysis of coastal and offshore aquaculture: Application of the FARM model to multiple systems and shellfish species. *Aquaculture* **292**, 129-138, doi:https://doi.org/10.1016/j.aquaculture.2009.03.039 (2009).

6 Brown, A. R. *et al.* Assessing risks and mitigating impacts of harmful algal blooms on mariculture and marine fisheries. *Reviews in Aquaculture* **12**, 1663-1688, doi:https://doi.org/10.1111/raq.12403 (2020).
7 Glibert, P. M. *et al.* Vulnerability of coastal ecosystems to changes in harmful algal bloom distribution in response to climate change: projections based on model analysis. *Global Change Biology* **20**, 3845-3858, doi:https://doi.org/10.1111/gcb.12662 (2014).
8 Weisberg, R. H. *et al.* The Coastal Ocean Circulation Influence on the 2018 West Florida Shelf K. brevis Red Tide Bloom. *Journal of Geophysical Research: Oceans* **124**, 2501-2512, doi:https://doi.org/10.1029/2018JC014887 (2019).
9 Trainer, V. L. *et al.* GlobalHAB: Evaluating, reducing and mitigating the cost of Harmful Algal Blooms: A Compendium of case studies. *PICES Press* **28**, 30-32 (2020).
10 Wells, M. L. *et al.* Future HAB science: Directions and challenges in a changing climate. *Harmful Algae* **91**, 101632, doi:https://doi.org/10.1016/j.hal.2019.101632 (2020).
11 Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP). Protecting the Oceans from Land-based Activities: Land-based sources and activities affecting the quality and uses of the marine, coastal and associated freshwater environment., (2001).
12 Berdalet, E. *et al.* Marine harmful algal blooms, human health and wellbeing: challenges and opportunities in the 21st century. *Journal of the Marine Biological Association of the United Kingdom* **96**, 61-91, doi:10.1017/S0025315415001733 (2016).
13 Reguera, B. *et al.* Dinophysis Toxins: Causative Organisms, Distribution and Fate in Shellfish. *Marine Drugs* **12**, 394-461 (2014).
14 Manfrin, C. *et al.* Physiological and molecular responses of bivalves to toxic dinoflagellates. *Invertebrate Survival Journal* **9** (2012).
15 Wells, M. L. *et al.* Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful Algae* **49**, 68-93, doi:https://doi.org/10.1016/j.hal.2015.07.009 (2015).
16 Gobler, C. J. *et al.* Ocean warming since 1982 has expanded the niche of toxic algal blooms in the North Atlantic and North Pacific oceans. *Proceedings of the National Academy of Sciences* **114**, 4975, doi:10.1073/pnas.1619575114 (2017).
17 European Union. (Official Journal of the European Union, 2004).
18 Anderson, D. M., Cembella, A. D. & Hallegraeff, G. M. Progress in Understanding Harmful Algal Blooms: Paradigm Shifts and New Technologies for Research, Monitoring, and Management. *Annual Review of Marine Science* **4**, 143-176, doi:10.1146/annurev-marine-120308-081121 (2011).
19 Davidson, K. *et al.* Forecasting the risk of harmful algal blooms. *Harmful algae* **53**, 1-7, doi:10.1016/j.hal.2015.11.005 (2016).
20 Gillibrand, P. A., Siemering, B., Miller, P. I. & Davidson, K. Individual-based modelling of the development and transport of a Karenia mikimotoi bloom on the North-west European continental shelf. *Harmful Algae* **53**, 118-134, doi:https://doi.org/10.1016/j.hal.2015.11.011 (2016).
21 Karasiewicz, S., Chapelle, A., Bacher, C. & Soudant, D. Harmful algae niche responses to environmental and community variation along the French coast. *Harmful Algae* **93**, 101785, doi:https://doi.org/10.1016/j.hal.2020.101785 (2020).

22 Janssen, A. B. G. *et al.* How to model algal blooms in any lake on earth. *Current Opinion in Environmental Sustainability* **36**, 1-10, doi:https://doi.org/10.1016/j.cosust.2018.09.001 (2019).
23 Schmidt, W. *et al.* A generic approach for the development of short-term predictions of Escherichia coli and biotoxins in shellfish. *Aquac Environ Interact* **10**, 173-185, doi:10.3354/aei00265 (2018).
24 Taranu, Z. E., Gregory-Eaves, I., Steele, R. J., Beaulieu, M. & Legendre, P. Predicting microcystin concentrations in lakes and reservoirs at a continental scale: A new framework for modelling an important health risk factor. *Global Ecology and Biogeography* **26**, 625-637, doi:https://doi.org/10.1111/geb.12569 (2017).
25 Cusack, C. *et al.* Harmful algal bloom forecast system for SW Ireland. Part II: Are operational oceanographic models useful in a HAB warning system. *Harmful Algae* **53**, 86-101, doi:https://doi.org/10.1016/j.hal.2015.11.013 (2016).
26 Davidson, K. *et al.* HABreports: Online Early Warning of Harmful Algal and Biotoxin Risk for the Scottish Shellfish and Finfish Aquaculture Industries. *Frontiers in Marine Science* **8**, doi:10.3389/fmars.2021.631732 (2021).
27 Wood, S. N. *Generalized Additive Models: An Introduction with R, Second Edition*. (CRC Press, 2017).
28 Pingree, R. D., Pugh, P. R., Holligan, P. M. & Forster, G. R. Summer phytoplankton blooms and red tides along tidal fronts in the approaches to the English Channel. *Nature* **258**, 672-677, doi:10.1038/258672a0 (1975).
29 Pingree, R. D., Mardell, G. T. & Maddock, L. A marginal front in Lyme Bay. *Journal of the Marine Biological Association of the United Kingdom* **63**, 9-15, doi:10.1017/S0025315400049754 (1983).
30 Drouard, M., Kornhuber, K. & Woollings, T. Disentangling Dynamic Contributions to Summer 2018 Anomalous Weather Over Europe. *Geophysical Research Letters* **46**, 12537-12546, doi:https://doi.org/10.1029/2019GL084601 (2019).
31 Raine, R., Edwards, M., Reid, P., Bresnan, E. & Fernand, L. Harmful Algal Blooms in Marine Climate Change Impacts Annual Report Card 2007–2008. (2008).
32 Caballero, I., Fernández, R., Escalante, O. M., Mamán, L. & Navarro, G. New capabilities of Sentinel-2A/B satellites combined with in situ data for monitoring small harmful algal blooms in complex coastal waters. *Scientific Reports* **10**, 8743, doi:10.1038/s41598-020-65600-1 (2020).
33 Park, M. *et al.* First successful culture of the marine dinoflagellate Dinophysis acuminata. *Aquatic Microbial Ecology* **45**, 101-106 (2006).
34 Flynn, K. J. & McGillicuddy, D. J. in *Harmful Algal Blooms* 115-134 (2018).
35 Díaz, P. A. *et al.* Mesoscale Dynamics and Niche Segregation of Two Dinophysis Species in Galician-Portuguese Coastal Waters. *Toxins* **11**, 37 (2019).
36 Fernández, R. *et al.* Dinophysis Species and Diarrhetic Shellfish Toxins: 20 Years of Monitoring Program in Andalusia, South of Spain. *Toxins* **11**, 189 (2019).
37 Basti, L. *et al.* Harmful algal blooms and shellfish aquaculture in changing environment. *Bulletin of Japan Fisheries Research and Education Agency* (2019).
38 Bickman, S. R. *et al.* An Innovative Portable Biosensor System for the Rapid Detection of Freshwater Cyanobacterial Algal Bloom Toxins. *Environmental Science & Technology* **52**, 11691-11698, doi:10.1021/acs.est.8b02769 (2018).

39 Scholin, C. *et al.* The Quest to Develop Ecogenomic Sensors: A 25-Year History of the Environmental Sample Processor (ESP) as a Case Study. *Oceanography* **30**, 100-113, doi:https://doi.org/10.5670/oceanog.2017.427 (2018).

40 McPartlin, D. A. *et al.* Biosensors for the monitoring of harmful algal blooms. *Current Opinion in Biotechnology* **45**, 164-169, doi:https://doi.org/10.1016/j.copbio.2017.02.018 (2017).

41 Swan, S. C. *et al.* Dinophysis acuta in Scottish Coastal Waters and Its Influence on Diarrhetic Shellfish Toxin Profiles. *Toxins* **10**, 399, doi:10.3390/toxins10100399 (2018).

42 Campbell, L., Henrichs, D. W., Olson, R. J. & Sosik, H. M. Continuous automated imaging-in-flow cytometry for detection and early warning of Karenia brevis blooms in the Gulf of Mexico. *Environ Sci Pollut Res Int* **20**, 6896-6902, doi:10.1007/s11356-012-1437-4 (2013).

43 National Oceanographic and Atmospheric Administration. *Gulf of Mexico Harmful Algal Bloom Forecast, Tides and Currents*, <https://tidesandcurrents.noaa.gov/hab/gomx.html> (2021).

44 Thompson, A. *et al.* Satellites to seafloor: Toward fully autonomous ocean sampling. *Oceanography* **30**, 160-168 (2017).

45 Anderson, C. R. *et al.* Scaling Up From Regional Case Studies to a Global Harmful Algal Bloom Observing System. *Frontiers in Marine Science* **6**, doi:10.3389/fmars.2019.00250 (2019).

46 Yamahara, K. M. *et al.* In situ Autonomous Acquisition and Preservation of Marine Environmental DNA Using an Autonomous Underwater Vehicle. *Frontiers in Marine Science* **6**, doi:10.3389/fmars.2019.00373 (2019).

47 Kammann, E. E. & Wand, M. P. Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **52**, 1-18, doi:https://doi.org/10.1111/1467-9876.00385 (2003).

48 Wood, S. N. Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. *Biometrics* **62**, 1025-1036, doi:10.1111/j.1541-0420.2006.00574.x (2006).

49 NEMO Ocean Engine (Zenodo, 2019).

50 Kristian, M. & M. Alonso Balmaseda, A. W. (ECMWF, 2012).

51 Wright, M. N. & Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *2017* **77**, 17, doi:10.18637/jss.v077.i01 (2017).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DataSummary.xlsx
- Performance.xlsx
- Predictions.pdf
- Smoothed.pdf