Regulators Pioneer Fund Project - Cornwall Port Health Authority (Cornwall Council)

Artificial Intelligence approaches for predicting Harmful Algal Blooms (HABs)

Final Report to Department for Business, Energy and Industrial Strategy

"This project was made possible by a grant from the £3.7 million Regulators' Pioneer Fund launched by The Department for Business, Energy and Industrial Strategy (BEIS). "The fund enables UK regulators and local authorities to help create a UK regulatory environment that unleashes innovation and makes the UK the best place to start and grow a business"

Introduction

In July 2021 Cornwall Port Health Authority (CPHA) with a consortium of partners successfully bid for £199,444.00 from the Department for Business, Energy and Industrial Strategy (BEIS) to conduct a discrete local study into Algal Blooms (HABs). HABs can produce toxins, which accumulate in filter-feeding shellfish and intoxicate human consumers. The toxins are heat stable and so can't be destroyed by freezing and/or cooking.

Under current regulations shellfish toxin monitoring is effectively retrospective: regulators sample, await results, and if the regulatory threshold is breached there is an investigation into the amount of shellfish harvested since the sample was taken, which might then result in a full-scale food chain product recall. By gathering high resolution field monitoring data using novel qPCR and lateral-flow (LF) techniques, we planned to refine and validate a computer model for predicting HABs caused by Dinophysis species. One of the aims of the project was to use the higher resolution data collected as part of the project to train the model towards a more accurate forecast in respect of breaches in the Dinophysis toxin threshold up to 6-8 weeks ahead. The model would then aid planning decisions for harvesting and will save costly recalls and protect human health (in this case from Diarrhetic Shellfish Poisoning - DSP).

The other strands of the project consisted of use of a Novel monitoring tools, a qPCR for quantifying HAB cell abundance in seawater, and a Lateral Flow testing for quantifying *Dinophysis* toxins in shellfish, directly in the field. Field data from these novel methods will be validated by an accredited light microscopy technique which enables the cell densities to be quantified in water and by liquid chromatography with tandem mass spectrometry (LC-MS/MS) for validating the shellfish flesh test results from the field.

CPHA developed a sampling plan and deployed a vessel to sea twice weekly, at two sites to sample water and bivalve flesh for in-field LF/qPCR and laboratory analysis.

Key motivations for the project included:

- Developing tools that are in early – mid stage development (existing capability) that can be used by the industry to make harvesting judgements; and
- Further developing and training an existing model to forecast breaches in HAB (Dinophysis) toxin thresholds in shellfish.

Commission Implementing Regulation (EU) 2019/627 stipulates that biotoxin sampling must be undertaken weekly, unless there is a formal risk assessment in place. A risk assessment has been published by the Food Standards Agency (FSA) it is provided to each Local Authority with shellfish official control duties.

In the case of Cornwall, a HAB hotspot, it has increased the frequency of sampling and hence extra pressure on Local Authority staff and budgets. The FSA has determined that the sampling frequency at all production areas should ensure that the risk of missing a biotoxin event should be no more than 2% in any given month. This whole approach is predicated on the fact that the system of sampling is retrospective. LAs sample, send to the lab, the lab tests and reports the results, and if failure occurs there is a necessity to close the shellfish beds, and instigate a full product recall of short shelf-life shellfish product, which may have already been consumed in some cases. This is a risk for shellfish harvesters, a burden for business, and risks public health and consumer confidence.

Combining a HAB forecasting model with accurate field testing by harvesters, will allow regulatory sampling to focus on critical periods when HAB toxin concentrations in shellfish are indicated to increase above and decline below the regulatory threshold.

This targeted approach could reduce the overall frequency of regulatory sampling, whilst giving the industry the tools they require to place food upon the market in a safe manner, without the worry of a retrospective recall.

Project Overview

Strand 1 – Predictive Modelling (Partners: University of Glasgow and University of Exeter)

Strand Overview

The ultimate aim of Strand 1 is to develop and apply data-driven models to provide a reliable HAB early warning system for shellfish aquaculture sites in SW England. Reliable forecasting of HAB events and intoxication of shellfish is important for protecting public (shellfish consumer) health by identifying safe harvest periods for shellfish operators and high-risk periods when increased regulatory (Official Control) monitoring is required.

The goal was to create a HAB early warning system (the intended gain) for shellfish businesses and regulators, with the capability to forecast breaches of HAB toxin thresholds in shellfish up to 8 weeks ahead.

Coastal waters in SW England to the west of Start Point constitute a known HAB hotspot, particularly for dinoflagellate HAB species belonging to the genus *Dinophysis*, which produce Dinophysis toxins that accumulate in shellfish and cause diarrhetic shellfish poisoning in human consumers (Brown et al., 2022). Currently HAB monitoring by Cornwall Port Health Authority is conducted on a weekly basis throughout the potential HAB season (May to end of September) and reduces to a fortnightly or monthly frequency thereafter. Reliable HAB forecasting will enable a more informed risk-based approach aiding forward planning of harvesting by shellfish producers and monitoring by regulators. For example, monitoring can be ramped up when the risk of HABs is predicted to be high (shellfish toxins exceed regulatory action levels) and ramped down when HAB risk is predicted to be low.

Data-driven forecasting models for Dinophysis toxins in shellfish have previously been based on long-term (~10 year) HAB monitoring programmes in SW England, NW Scotland and Northern

France. These models incorporate simple statistical and machine learning methods, providing a useful starting point for work in Strand 1. To increase forecasting accuracy and forecasting horizon, we aim to exploit additional monitoring data gathered during the life of the Project (September 2021 to end of March 2022) for HAB toxins in shellfish, HAB cell abundance in seawater, and associated environmental parameters (e.g. temperature, salinity, wave action). Additional data include more frequent (twice weekly) assessments of HAB cell abundance and HAB toxins in shellfish.

The forecasting model we have developed in this Project focuses on predicting Dinophysis toxins (comprising Dinophysis toxins - DTX, Pectenotoxins - PTX and Okadaic Acid – OA) collectively expressed as OA equivalents measured in μg/kg shellfish flesh. The reason for focusing on toxin concentrations is that this is the basis for the regulatory threshold above which shellfish harvesting is prohibited (i.e. 160μg OA eq./kg shellfish flesh). Thus, we aim to use the model predict when a shellfish production area is likely to close and when its likely to reopen for harvesting. Monitoring can be intensified during these periods.

Key elements for refining predictive modelling work in Strand 1 include:

- Improving the integration of monitoring and modelling approaches for use in HAB early warning systems:
    - identifying data bottlenecks and quantifying the extent to which increasing monitoring frequency can improve model accuracy (by reducing uncertainty around data extrapolation and statistical smoothing);
    - exploring how modelling can be used to inform optimal monitoring schedules and frequencies.
- Quantifying and visualising time lags between changes in *Dinophysis spp.* cell abundance and sea surface temperature and changes in toxin concentrations in shellfish.
- Evaluating the potential benefits for modelling of *in situ* monitoring data (Strand 2) obtained from novel lateral flow devices for measuring HAB toxins in shellfish and portable qPCR devices for measuring HAB cell abundance.

Methodology for Prediction

General Approach

Our goal is to predict the Okadaic Acid (OA) concentration $y_i(t,s)$ in shellfish sample i on a given day t and from a given harvesting site s. To achieve this, we assume y can be modelled using some probability distribution. The concentration level$y_i(t,s)$ is always non-negative and its distribution is very heavy tailed, with small values in the winter and measurements exceeding 1000μg/L$y_i(t,s)$ can be modelled effectively using the Negative-Binomial family of distributions:

$$y_i(t,s) \sim \text{Negative} - \text{Binomial}(\mu(t,s), \theta).$$

Here $\mu(t,s)$ is the average (mean) OA level we would expect to measure if many samples were taken on day t from site s, and θ is a parameter which controls the variance of the Negative-Binomial distribution.

To develop predictions of future OA levels, we need to estimate the relationship between $\mu(t,s)$ and input variables. Here we consider sea surface temperature $temp(t,s)$ and *Dinophysis* spp. abundance counts $abun(t,s)$ as potential inputs. However, we might suppose that changes in these variables take time to change to change the OA levels in shellfish. Consider that shellfish take time to absorb harmful toxins produced by phytoplankton, and that more time still might be needed for changes is sea temperatures to influence the phytoplankton population and production of toxins.

Therefore, we estimate the impact of past input variable values (e.g. $temp(t-7,s)$) on current or future OA levels, as captured by $\mu(t,s)$. We call these past values the "lagged" values (e.g. $temp(t-l,s)$), measured at some lag l (days) up to a maximum lag L. Unfortunately, we often don't know in advance which lags are most important or how the past values measured at different lags might work together to achieve a certain OA level. To utilise lagged information effectively, we should consider lag models.

Unconstrained lag models

An unconstrained lag model includes all lags of input variables and treats them as independent variables, e.g.

$$\log\big(\mu(t,s)\big) = \beta_0 + \beta_1 temp(t,s) + \beta_2 temp(t-1,s) + \cdots \beta_{L+1} temp(t-L,s).$$

However, when considering a large number of lags (e.g. every day in the past year), we would very likely end up with a very overly complex (over-parametrised) model, which can fit data it has seen well, but not unseen future values. At the same time, we might expect the values of input variables (e.g. sea surface temperature - SST) at more proximal lag time points to be more highly correlated than more distant time points, leading to a loss of precision when treating all timepoints as independent.

Distributed lag models (DLMs)

Instead, we can impose some sensible constraints of the coefficients (β) of the lagged values. This is known as a distributed lag model. For instance, we might suppose that the effect of past values might be more similar for nearby lags (e.g. 14 days ago and 15 days ago), than for very different lags (e.g. 14 days ago and 70 days ago). One approach to this is to constrain the effect (coefficients) of past values to change "smoothly" as the lag increases, by estimating the coefficients as some smooth function f of lag $f(l)$, i.e.

$$\log\big(\mu(t,s)\big) = \beta_0 + f(0)temp(t,s) + f(1)temp(t-1,s) + \cdots f(L)temp(t-L,s).$$

Distributed non-linear lag models (DLNMs)

We might also want to consider that the effect of lagged input variables on OA concentration might be non-linear. For instance, we could capture the effect of sea surface temperature using some estimated smooth function g:

$$\log\big(\mu(t,s)\big) = \beta_0 + f(0)g(temp(t,s)) + f(1)g(temp(t-1,s)) + \cdots f(L)g(temp(t-L,s)).$$

However, instead of estimating f and g as independent functions, a more general approach is to define some 2-dimensional smooth function h of the input variable and lag, e.g.

$$\log\big(\mu(t,s)\big) = \beta_0 + h(temp(t,s),0) + h(temp(t-1,s),1) + \cdots h(temp(t-L,s),L).$$

This model, known as a distributed non-linear lag model, allows the non-linear effect of the input variable to be different depending on the lag. We can develop and fit Negative-Binomial models defined in this way using the *mgcv* package for the R programming language.

Model A: using phytoplankton abundance as the main predictor variable

For this project, we developed two families of models for predicting future OA levels. The first, called Model A, uses lagged phytoplankton abundance counts as the main input variable. Denote $d(t)$ the calendar day within the year the toxin sample was collected. Including $d(t)$ as an additional input to $\mu(t,s)$ allows the model to estimate seasonal structures in the mean OA levels. In Model A, the mean OA level $\mu(t,s)$ is characterised by:

$$\log\big(\mu(t,s)\big) = a(site(s)) + f\big(\boldsymbol{abun}(s), l, d(t)\big) + g(t, site(s)).$$

First, $a(site(s))$ is a factor variable for each harvesting site s, which captures the overall differences in average OA levels between the study sites. Then, $f\big(\boldsymbol{abun}(s), l, d(t)\big)$ is a 3-dimensional smooth distributed non-linear lag term. The first two variables $\boldsymbol{abun}(s)$ and l, capture the effect of past abundance values on mean OA levels now. The third dimension, $d(t)$, allows the structure of the lagged abundance effect to vary smoothly depending on the time of year the OA sample is collected. For instance, we might imagine the lag between abundance and OA levels might be different depending on whether the shellfish are absorbing toxins in the onset of the bloom or depurating after *Dinophysis* spp. populations have diminished in the water. This function (f) also captures the average seasonal cycle of blooms across all sites.

Through f, the model has great flexibility to make predictions of future OA levels based on lagged abundance values and learned seasonal patterns. However, we should always expect some degree of error when making predictions. For example, a bloom may be more or less severe than expected. The final term $g(t, site(s))$ is an independent spline of time for each site, to learn structured variability on this error over time and develop more reliable predictions.

Lagged abundance values are included up to 90 days in the past, relative to the date the shellfish sample is collected. Although through our experimentation it is apparent that the most important lags are considerably less than 90 days, we may wish include as many lags as possible and allow the model to estimate their importance. However, this comes at the expense of potentially excluding data at the beginning of data collection, because abundance is not known far into the past. Some sites studied here (e.g. Mevagissey Bay) have relatively short time-series for both *Dinophysis* spp. abundance in water and Dinophysis toxin levels in shellfish, hence why we did not consider lags of abundance beyond 90 days.

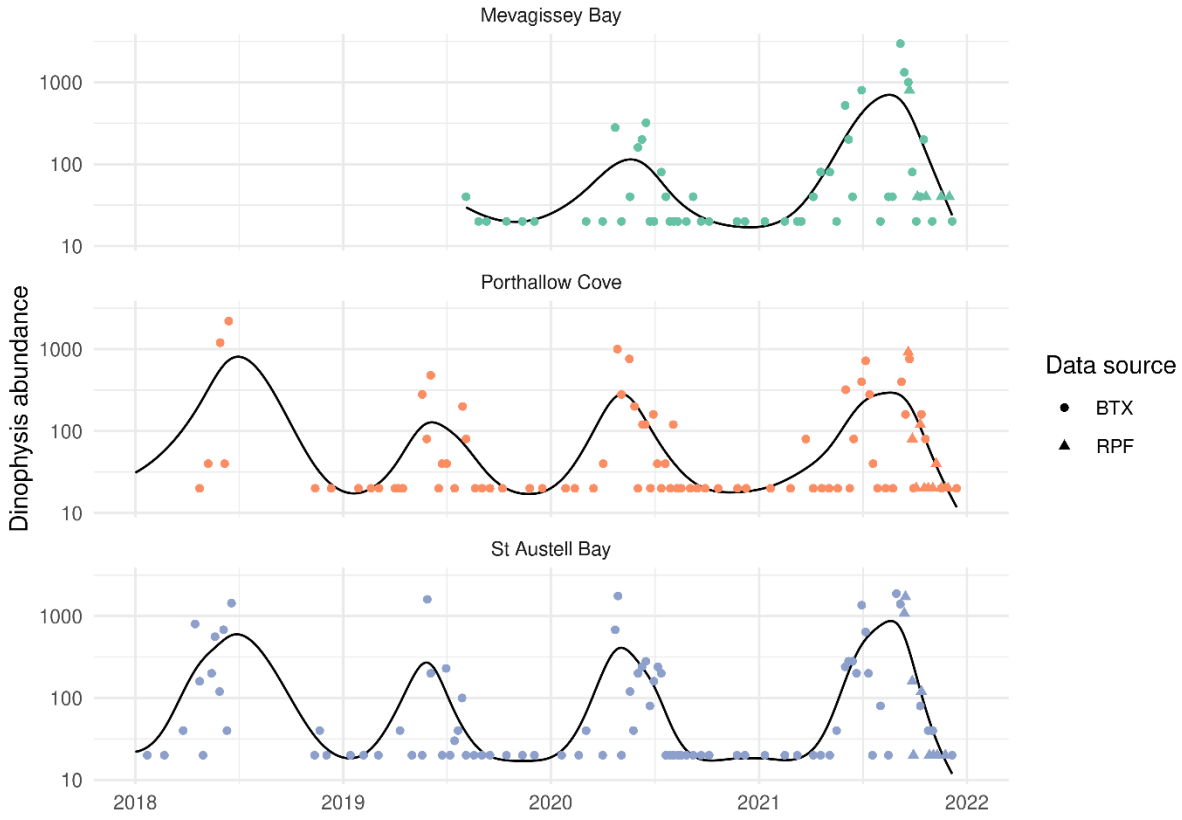
Pre-processing of abundance data

Phytoplankton abundance is not measured at a daily frequency in any of the sites considered here, yet our model requires that our lagged variable $abun(t-l, s)$ is observed for all lags $l = 0,1,\dots,90$. Therefore, we choose to construct a smooth interpolation of the abundance measurements using a Generalize Additive Model, consisting of a single flexible regression spline of time. The result of this interpolation is illustrated by Figure 1.

This is not the only method we could have chosen, e.g. we could have used a moving average or a linear interpolation, and the choice of method will inevitably introduce some form of bias in the predictions. The advantage of using a smooth regression spline is that we can “smooth-out” individual noisy measurements, and therefore more robustly capture the “signal”. This comes at the expense of smoothing-out some important features, for instance the distinctive “M” shape the bloom displayed in 2021, with a dip to low levels of toxicity towards the end of the summer, before a second wave of higher toxin levels. The M-shape pattern can be seen in the data points but is not captured well by the regression splines.

We also experimented with linear interpolation. This ensured that all of the “signal” in the abundance measurements was captured, at the expense of the result being quite “spiky”, and very sensitive individual measurements. As a consequence, we found that the distributed lag term in the OA model $f\big(\boldsymbol{abun}(s), l, d(t)\big)$ had to spread out over a wider range of lags, to “smooth-out” the spikes and uncover the signal. Therefore, it was not as precise in identifying the lagged relationship between *Dinophysis* spp. abundance and OA.

Interpolated dinophysis abundance



Interpolated *Dinophysis* abundance

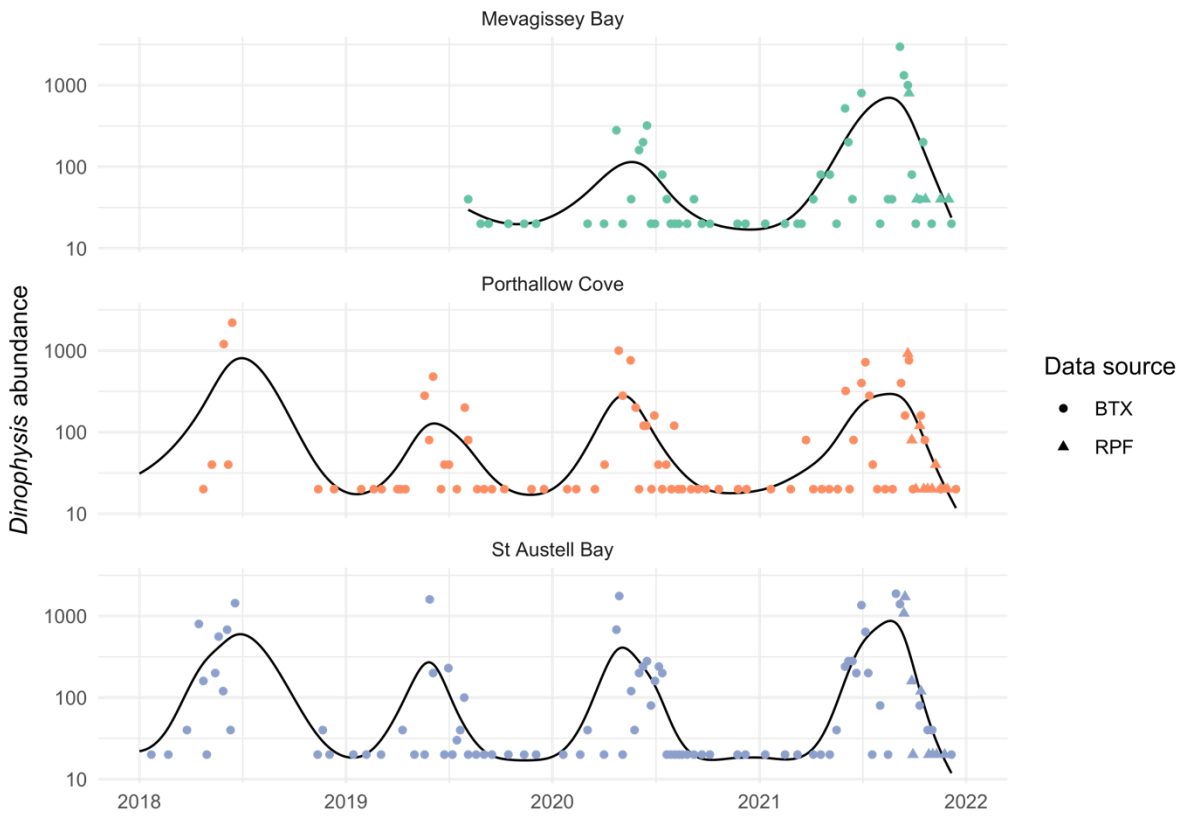


Figure 1: Illustration of interpolation method for forming a continuous (daily) time series of Dinophysis abundance measurements (black lines). Circles show abundance counts collected through the regulatory control scheme (BTX) and triangles show measurements collected through the Regulator's Pioneer Fund (RPF).

Model T: using sea surface temperature as the main predictor variable

Model T is the same as Model, except that past abundance values are replaced with past sea surface temperature values $\boldsymbol{temp}(s)$:

$$\log\big(\mu(t,s)\big) = a(s) + f\big(\boldsymbol{temp}(s), l, d(t)\big) + g(t, site(s)).$$

We also consider temperature lags up to 365 days in the past. This is important, to ensure that longer-term processes determining the relationship between SST and OA levels can be captured. We are able to do this as time-series of SST data are available much further into the past than abundance and shellfish sample data for these sites.

Sea surface temperature data

We downloaded sea surface temperature data from the Copernicus Marine Service. For dates prior to 2021, we used hourly mean SST data from the Atlantic-European North West Shelf- Ocean Physics Reanalysis product, which has a spatial resolution of 0.111° × 0.067°, and computed daily mean values. For dates in 2021 and 2022, we used daily mean values from the European North West Shelf/Iberia Biscay Irish Seas – High Resolution ODYSSEA L4 Sea Surface Temperature Analysis product, which has a spatial resolution of 0.02° × 0.02°. For each site in this study, we matched both sets of SST data by finding the closest grid cell to the harvesting location with a complete time series of SST values (meaning no missing values, e.g. due to the grid cell being land instead of water).

Model B: baseline model with no abundance or sea surface temperature

To assess the effectiveness of past *Dinophysis* spp. abundance or past sea surface temperature values, models A and T are compared to a "baseline" model, Model B. In this model, we replace the 3-dimension seasonally varying distributed lag term with two new terms:

$$\log\big(\mu(t,s)\big) = a(site(s)) + f\big(d(t)\big) + g(t, site(s)).$$

First, the smooth term $f\big(d(t)\big)$ captures the overall seasonal pattern in mean OA levels. Then, $g(t, site(s))$ remaining temporal structure, including differences in how early or late the bloom season is, or how severe or mild the bloom is. With no covariate inputs included, the model must rely heavily on $g(t, site(s))$ to learn the characteristics of the new bloom.

Empirical predictions

Finally, predictions from all model-based approaches are compared to predictions from a purely data-based approach, which we call the "empirical" predictions. For a given site and date, the empirical predicted toxin concentration is simply assumed equal to latest available measurement. For instance, if we want to predict 4 weeks into the future, we use the latest available shellfish sample measurement at that site relative to present day. This approach is equivalent to predicting future weather based only on recent conditions, but it nonetheless serves as an additional useful baseline against which we can compare the models.

Prediction Experiment 1

To compare the forecasting capabilities of Models A, T, and B, we carried out a rolling forecasting experiment for 2016-2021 inclusive. This involves predicting mean OA levels ($\mu(t,s)$) for each day with data in order. When preparing to forecast $\mu(t,s)$, only shellfish sample data which would actually be available in reality when making that forecast is supplied to the models. For example, if

forecasting 4-weeks in the future, we don't supply any data collected in those 4 weeks between the pretend present day and the forecasting day.

At the same time, future values of abundance (and temperature) will be unknown in practice. Specifically, values of our predictors for lags less than or equal to the forecasting lead time (defined as how far in the future we are forecasting) will be unknown. We therefore fit a distinct model for each of the specific lead times, where every model includes only lagged predictors which would be observed if the model were realistically used for that lead time. We specified models with a lead time of 0 weeks (nowcasting), 1 week, 2 weeks, 4 weeks, 6 weeks, 8 weeks, 10 weeks, 12 weeks, and 16 weeks. The significance of forecasting with a lead-time of 0 is that all lags are treated as observed. This represents how well the model might be able to predict toxin concentrations in shellfish if accurate forecasts of abundance and temperature were available. As 12 weeks almost exceeds the number of lags available in Model A (90 days), predictions 12 and 16 weeks ahead are only generated using Model T and Model B.

To maintain a consistent level of training data throughout the experiment, we provide all models with shellfish toxin sample data starting 365 days in the future, relative to the forecasting day, in addition to past data. We argue this does not compromise the validity of forecasting results, since all our models treat years as exchangeable, i.e. they do not have any structures which recognise adjacent years as being closer to each other in time.

Prediction Experiment 2

We also carried out a second experiment, to assess the impact on forecasting performance of the increased data frequency enabled by the RPF. The experiment involved repeating the same rolling forecasting scenario as Prediction Experiment 1 for 2021 but *excluding both abundance counts and shellfish samples collected through the RPF*. We can then compare the performance of Model A and T with and without this additional data, focussing on the period when RPF data were collected.

Insight and Results from Predictive Modelling

Insights from Model A

Key insights can be obtained about the relationship between lagged abundance and OA levels by looking at the estimated function $f\left(\boldsymbol{abun}(s), l, d(t)\right)$ from a fit of Model A that has seen all currently available data.

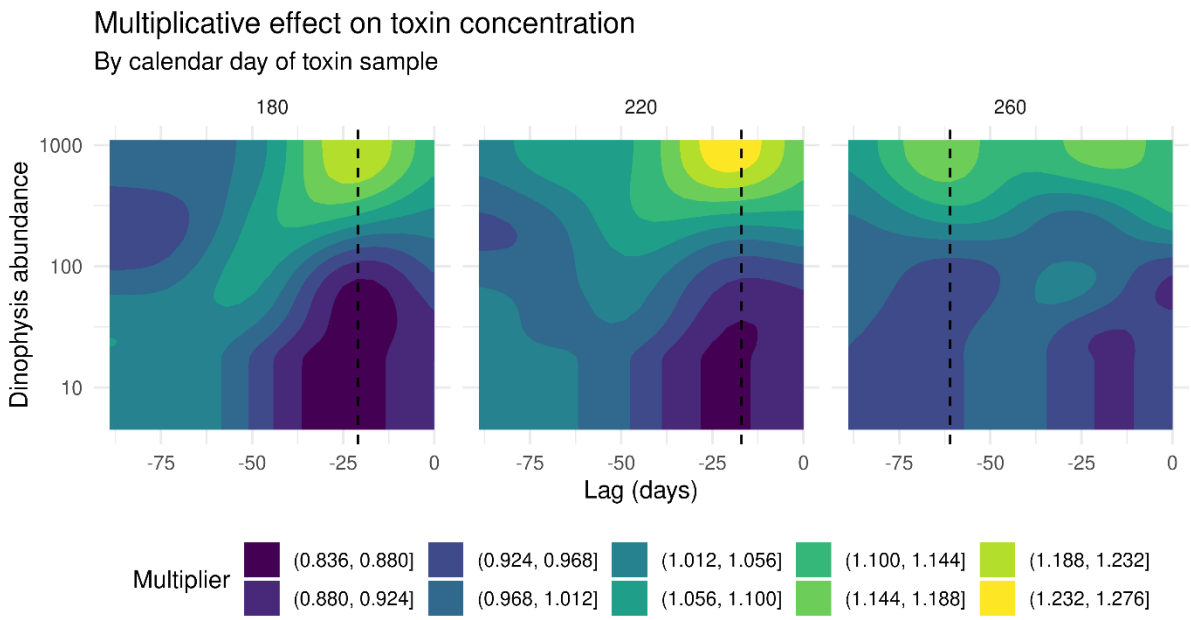


Figure 2: Multiplicative effect of past Dinophysis abundance counts on typical toxin concentrations taken at different calendar days (220 is early August), from Model A (the model with lagged abundance only). The vertical dashed lines show the lag which contributes most to the predicted toxin level at each time of year (the "principal lag").

Figure 2 shows the effect of past abundance values for a typical toxin sample taken at three different times of the year when blooms might be present (left: mid-June; middle: early-August; right: mid-September). Here, brighter green-yellow contours boost the summer OA levels if the past abundance values fall within those bands, whereas darker blue-purple contours dampen the summer toxin concentration. We can also define the "principal lag" as the lag which contributes the most overall to predicted OA levels. These are shown as the vertical dashed lines. In early-August (centre panel of Figure 2), when OA levels are high on average, the model suggests that higher abundance counts impact OA level 2-4 weeks in the future.

Using the fitted model, we can examine how the importance of different lags changes depending on the calendar day a shellfish sample is collected on. Figure 3 show the overall contribution of different lags (y-axis) for different calendar days (x-axis). Brighter green-yellow contours show important lags and darker blue-purple contours show less important ones. The dashed line tracks the trajectory of the "principal lag". We can characterise the trajectory of the principal lag as three-states. In the first state, up to about calendar day 230 (August), the principal lag gradually decreases from about 4 weeks to about 2 weeks. Beyond about calendar day 280 (October), the principal lag is about 10 weeks in the past. However, in between these dates, there is a very rapid shift in the principal lag towards higher values (from about 2 weeks to about 10 weeks). This is notable since it is during this period that most blooms would typically be decaying. We could speculate then, that this shift might be a representation of mussels accumulating toxins more quickly than depurating them.

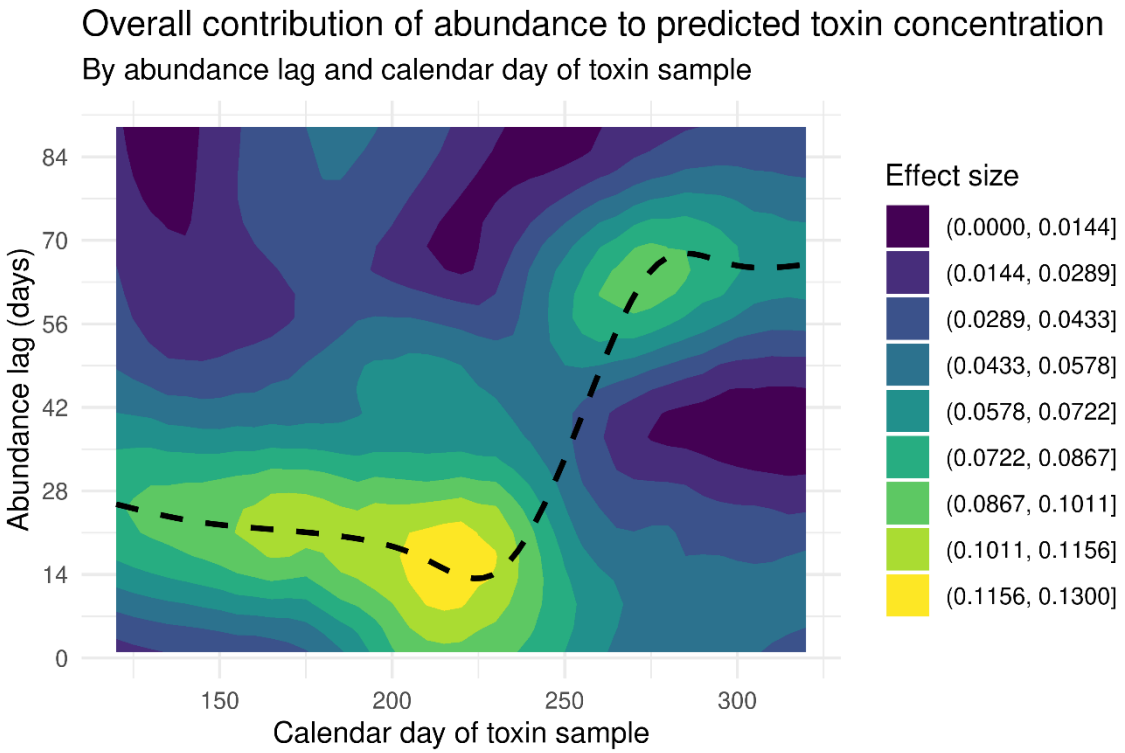


Figure 3: The overall contribution of each lag of Dinyophysis abundance for shellfish samples taken at different calendar days. The dashed line shows the trajectory of the "principal lag".

Figure 4 shows fitted values from Model A for the 3 sites studied in this project, for 2016 to 2022 (where data are available). The line shows how the model can flexibly capture the different peaks and troughs seen in the data (points). Note these are not predictions of unseen values, those will be seen in the results from the prediction experiment.

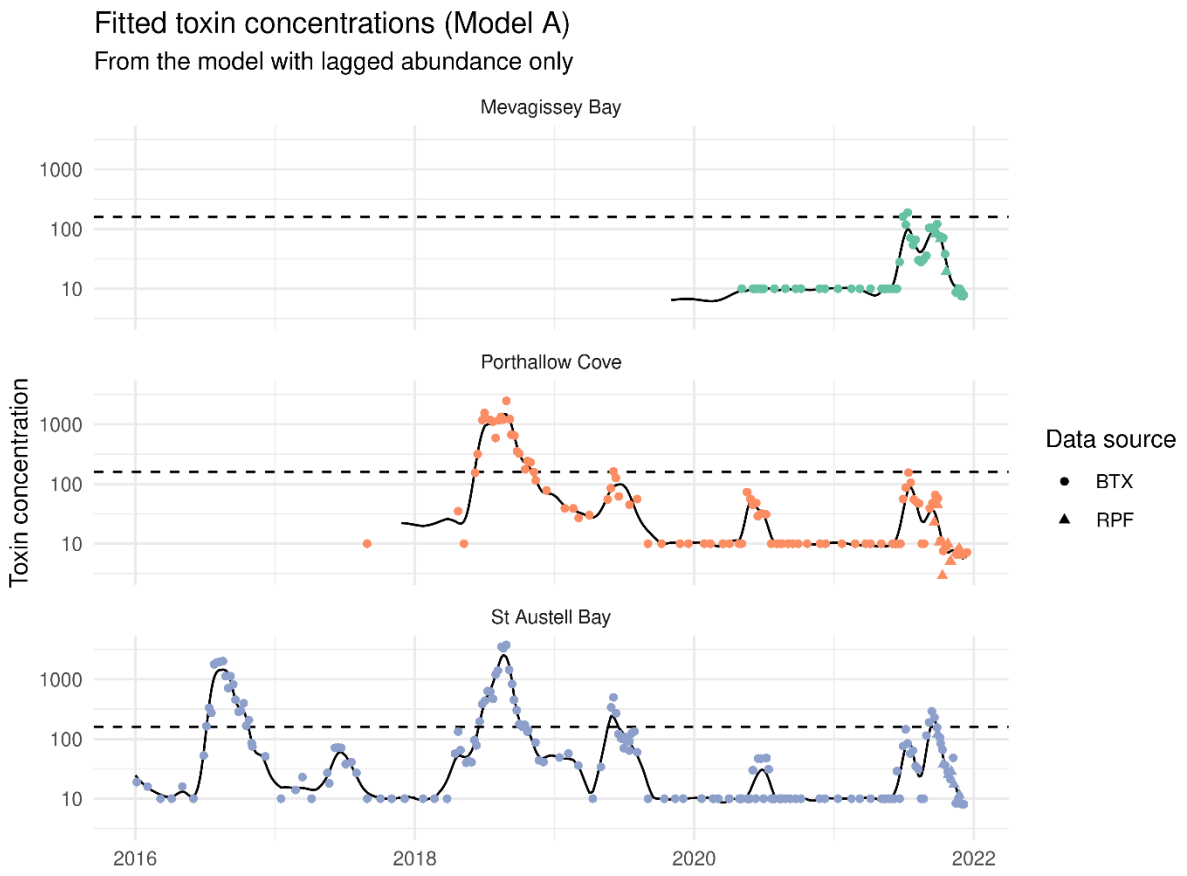


Figure 4: Lines show fitted (mean) toxin concentrations from the model with lagged abundance only (Model A). Points show observed toxin concentrations. The horizontal dashed lines show the regulatory action level for harvesting closure (160 micrograms/kg shellfish flesh). BTX is Official Control data, RPF is Regulators Pionner Fund data (from this Project).

Insights from Model T

Key insights can be obtained about the relationship between lagged sea surface temperature and OA levels by looking at the estimated function $f\left(\boldsymbol{temp}(s), l, d(t)\right)$ from a fit of Model T that has seen all currently available data.

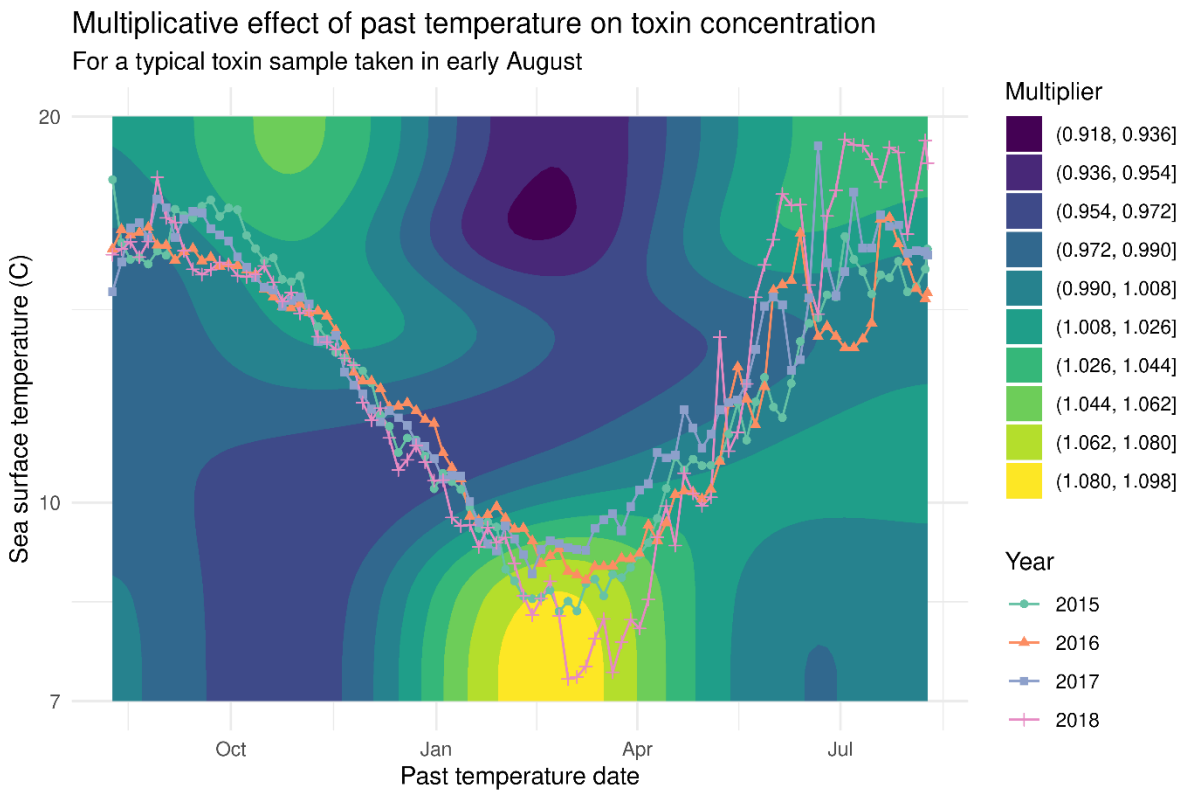


Figure 5: Multiplicative effect of past sea surface temperature values on typical toxin concentrations in mussels from St Austell Bay in early August, from Model T (the model with lagged sea surface temperature only). The lines show sea surface temperature values for different years at St Austell Bay.

Figure 5 shows the effect of past sea surface temperature values for a typical toxin sample taken in early August, and sea surface temperature values for St Austell Bay in each year from 2015 to 2018. Here, brighter green-yellow contours boost the summer OA levels if the SST trajectory passes through them, whereas darker blue-purple contours dampen the summer toxin concentration. The plot suggests that late-winter sea surface temperatures have the greatest impact, with colder temperatures being associated with higher summer toxin levels. Looking at the actual trajectory of sea surface temperatures previous years, there is a clear separation around March. In order of the coldest March to the warmest, the order is 2018, 2015, 2016, 2017. This is the same as the order of highest summer toxin levels to lowest, and this is the relationship the model is able capture.

The fact that the estimated summer toxin levels are substantially driven by winter SSTs suggests that the model may potentially be able to make credible predictions several months ahead.

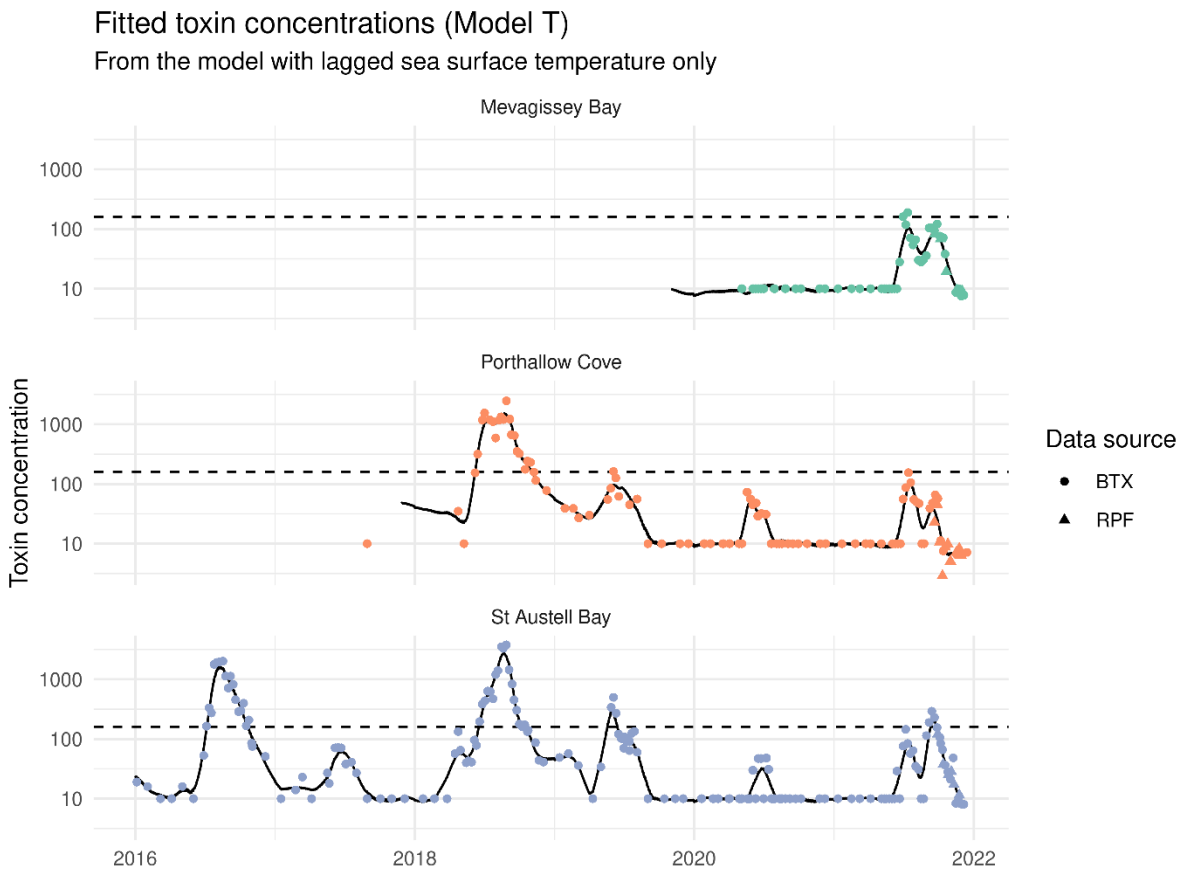


Figure 6: Lines show fitted (mean) toxin concentrations from the model with lagged abundance only (Model A). Points show observed toxin concentrations. The horizontal dashed lines show the regulatory action level for harvesting closure (160 micrograms/kg shellfish flesh).

Relative strength of abundance and sea surface temperatures as predictors (Prediction Experiment 1)

The goal of Prediction Experiment 1 was to quantify the relative effectiveness of lagged abundance and lagged temperature as predictors of toxin levels in future shellfish samples. To do this, the models are deployed in a rolling forecasting experiment, mimicking realistic operational use over a long period of time. Figure 7 shows rolling predictions for St Austell Bay, using the various prediction approaches and for different lead times (0 weeks, 2 weeks, 4 weeks). Generally, predictions get worse the longer the lead-time. Shellfish samples collected through this project are clearly marked as triangles (RPF), and it is clear how much denser the data is in the latter part of 2021 when these samples were taken.

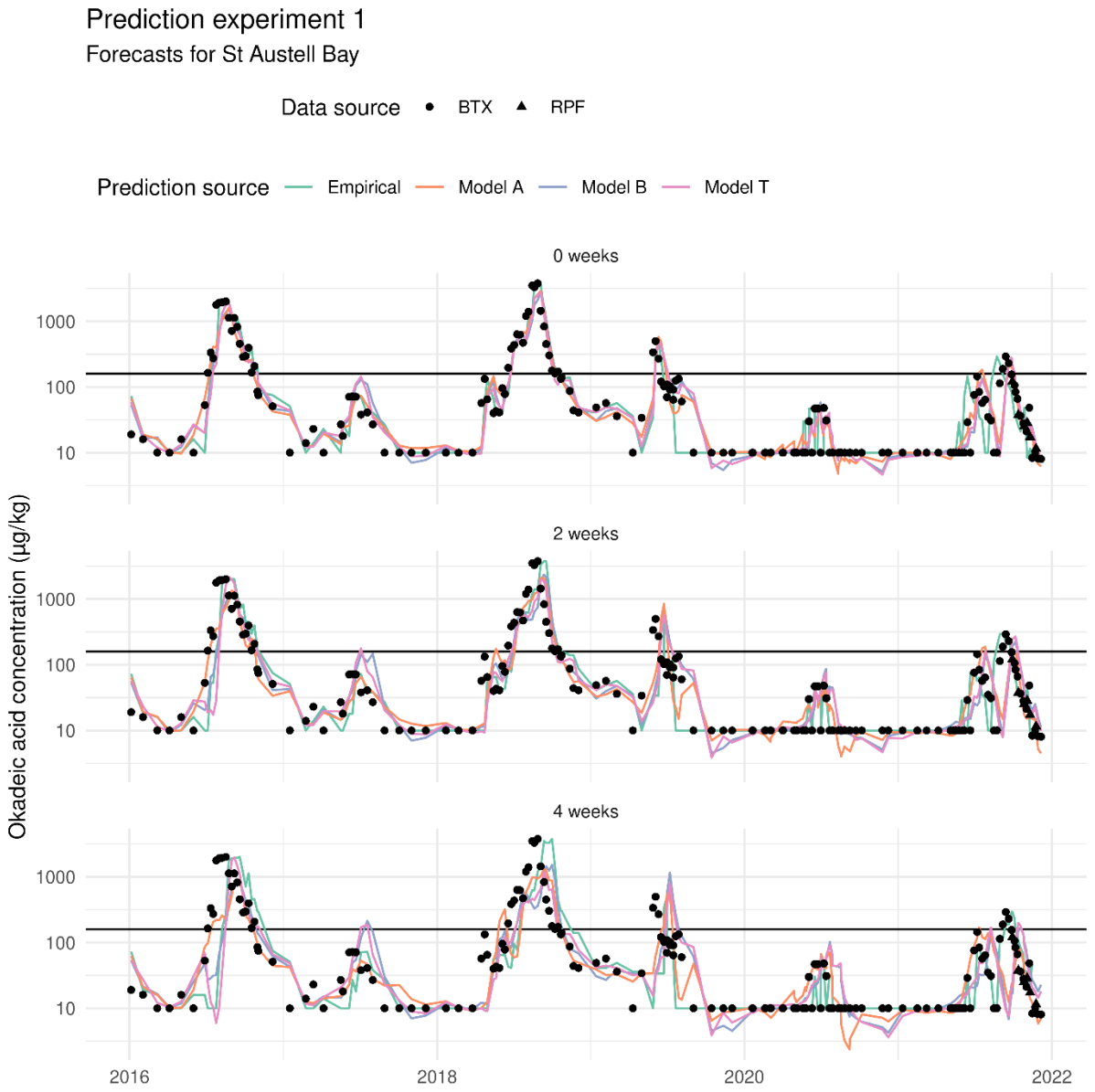


Figure 7: Rolling forecasts for St Austell Bay at different lead times (top: 0 weeks; middle: 2 weeks; bottom: 4 weeks). The horizontal black lines show the regulatory action level for harvesting closure (160 micrograms/kg shellfish flesh).

However, it is challenging to determine the overall performance of the different approaches using a purely visual approach. To quantify the performance, we computed the prediction errors (at the log-scale) and then computed two performance metrics for each prediction source (e.g. Model A) and forecasting lead time:

1. Mean absolute error – this represents how wrong the predictions are on average.
2. Root-mean squared error – how wrong the squared predictions are on average, then take the square-root to return the metric to the original scale of the predictions.

The mean absolute error is most intuitive, but the root-mean squared error is also insightful as it exaggerates big errors which may be particularly costly to stakeholders, e.g. by forecasting very severe future toxin values that turn out to be mild.

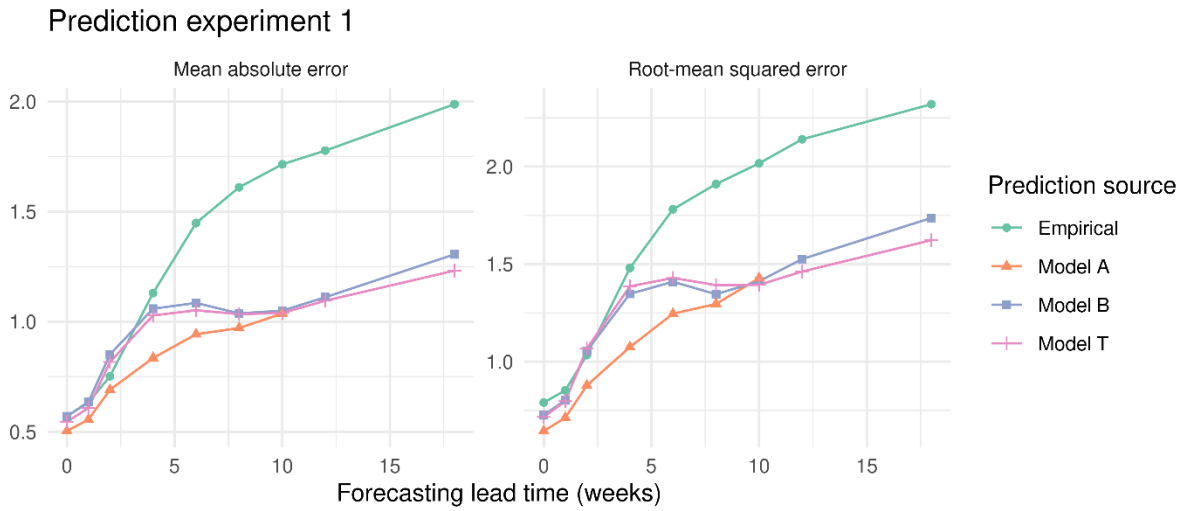


Figure 8: Mean absolute error (left) and root-mean squared error (right) of toxin forecasts, from the different prediction approaches and for different lead times. Errors were computed at the log-scale.

Figure 8 shows how these error metrics change with forecasting lead time and vary between the prediction approaches. Both metrics tell a consistent story, so we will focus on the root-mean squared error (RMSE). When forecasting up to 4 weeks in the future, model with lagged abundance (Model A) is the clear winner. The baseline model with no lagged variables (Model B) and the model with lagged temperature (Model T) don't fare noticeably better than the empirical approach (predicting the future toxin value as equal to the latest available measurement). Beyond 4 weeks, Models B and T do begin to show considerably more skill than the empirical approach, though Model A maintains its lead when forecasting up to 8 weeks ahead. Beyond this point, Model A runs out of informative abundance lags and effectively turns into Model B, relying exclusively on temporal and seasonal structures to make predictions. Finally, when predicting far into the future, Model T does outperform the baseline Model B, suggesting there is some information to be exploited in long-past sea surface temperature values.

Impact of increased data collection frequency (Prediction Experiment 2)

The goal of Prediction Experiment 2 was to quantify the impact of additional data collection – enabled through the RPF Project funding – on the ability of our models to forecast toxin levels. This involved running the rolling forecasting experiment as in Prediction Experiment 1, first with the additional data from the RPF, and second with only the official monitoring data.

We summarise prediction performance from the two scenarios by computing the root-mean squared error values in both cases, for each forecasting lead time. However, we only use prediction errors for official monitoring samples (present in both sets of model runs) and consider errors from samples taken on or after the 13th September 2021 (until the end of 2021), as this is when the first RPF sample was taken. This ensures the results from this experiment focus as much as possible on the period of time when data collection was taking place for the RPF.

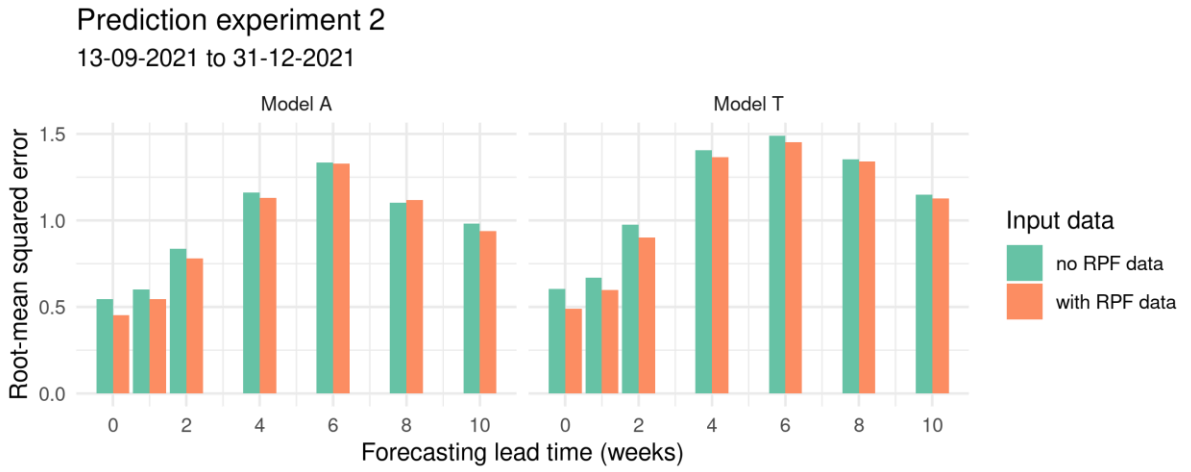


Figure 9: Root-mean squared errors for official monitoring shellfish sample toxin concentrations taken on or after the 13th September 2021, until the end of 2021. Green bars show the error value when no RPF abundance or shellfish sample data are included in the training data. Orange bars show errors with this data included.

Figure 9 shows error values with and without the RPF data included, for different forecasting lead times. The RMSE values are almost always lower when the RPF data is included, and the reduction in the RMSE is very sizeable when forecasting the short-medium term. This clearly demonstrates the potency of a higher frequency of data collection, but also hints that model-based prediction approaches might perform better still with even higher data collection frequencies, e.g. daily *Dinophysis* spp. abundance measurements.

Comparison between qPCR and Laboratory Data

To investigate potential for future integration of qPCR into regulatory monitoring of HABs, we carried out a statistical comparison between qPCR and laboratory results (the latter based on microscopic analysis) for *Dinophysis acuminata*.

The laboratory data had a high proportion of non-detects (74%) over the project time period. To enable quantitative comparison with the qPCR results, all non-detects were replaced with the value 10 cells/L. We could have chosen a different value, meaning all results in this comparison are conditional on that choice.

Figure 10 shows *Dinophysis acuminata* cell counts from both methodologies over the project duration (September 2021 to April 2022), for the three study sites. Across the project duration, the qPCR measurements were around one order of magnitude higher than the laboratory measurements. Both lab and qPCR measurements capture the decrease in cell abundance in the weeks and months following the end of the 2021 bloom season. In Porthallow Cove and St Austell Bay, the decrease in abundance appears to be much steeper in the laboratory measurements. Towards the end of the project in Mevagissey Bay, there is a noteworthy increase in the qPCR results not captured by the non-detects in the laboratory results, which may indicate the very beginnings of a bloom in 2022.

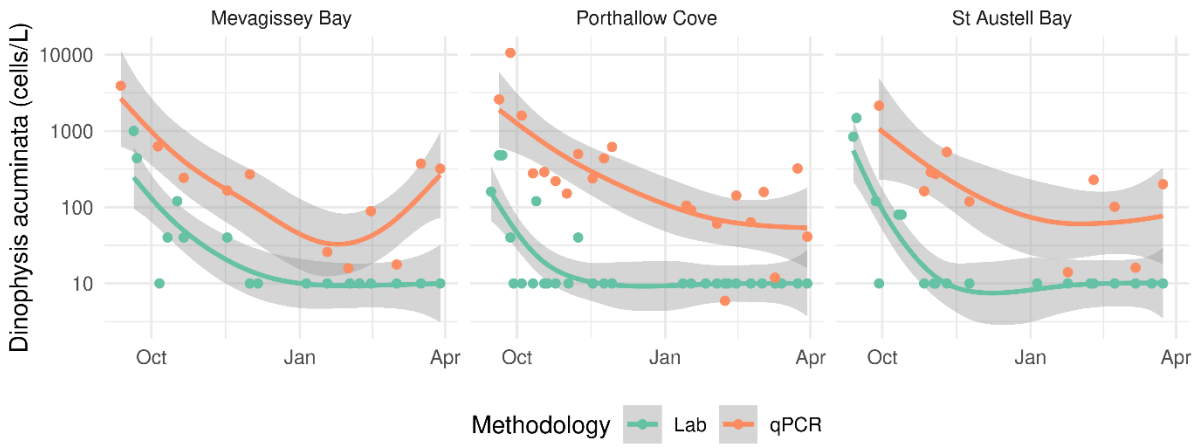


Figure 10: Comparison of laboratory and qPCR measurements of Dinophysis acuminata cell abundance for September 2021 – April 2022. Points show measurements and lines are smooth "lines of best fit" with 95% confidence intervals.

Regression analysis

We developed a simple regression approach to quantify the difference between lab and qPCR measurements more formally. We fitted Generalized Additive Models to data from each site independently of other sites. Here we model *Dinophysis acuminata* cell count y_i taken on day t at the log-scale through the combination of smooth functions of time, the effect of the cell count methodology used, and a Gaussian (Normal) error term:

$$\log\big(y_i(t)\big) = \beta_0 + f(t) + g(t)x_i + \epsilon_i;$$

$$\epsilon_i \sim \text{Normal}(0, \sigma^2).$$

The term $f(t)$ captures smooth variation over time in the laboratory measurements, tracking increases and decreases in cell abundance. The variable x_i equals 0 if measurement i is a laboratory result and equals 1 if measurement i is a qPCR result. The term $g(t)x_i$ therefore captures smooth variation over time in the difference (at the log-scale) between laboratory measurements and qPCR measurements. We call $g(t)$ the "qPCR coefficient" as for any given t it tells you what should divide a qPCR result by to obtain a result on the same scale of contemporary laboratory results.

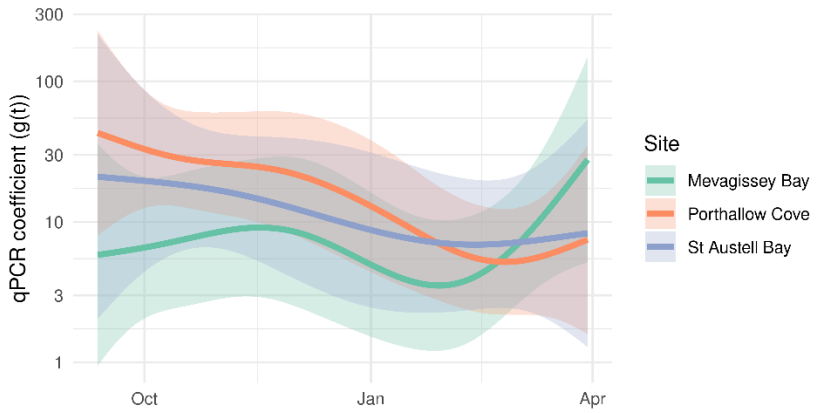


Figure 11: Estimates of the qPCR coefficient $g(t)$ (the value you should divide a qPCR result by to obtain a result on the same scale of contemporary laboratory results), with 95% confidence intervals.

Figure 11 shows estimates and 95% confidence intervals of the qPCR coefficient over the project duration. The coefficient fluctuates considerably over the time period, for instance in Porthallow

Cove qPCR measurements were on average over 30x higher than laboratory results in September, reducing to under 10x in September.

Conclusions

The comparison has two main limitations. The first is the high proportion of non-detects in the laboratory data (74%) over the project time period. The second is that the project period mostly coincided with periods of low cell abundance. Taken together, these limitations make it challenging to draw firm conclusions on the discrepancies seen between laboratory and qPCR results. For example, the uptick in cell abundance seen in Mevagissey Bay might suggest increased sensitivity of the qPCR methods in detecting the early stages of the bloom, which would significantly improve prospects for regulatory monitoring using predictive modelling. This is because there may only be a brief time period between cell abundance reaching detection levels for laboratory measurement and toxin concentrations in shellfish breaching safe harvesting levels. Meanwhile, qPCR provides quantitative measurements even when HAB cell counts are low, potentially providing early warnings. An extended time series of both qPCR and laboratory results would help to clarify the validity of the uptick.

Strand Conclusion

Strand 1 aimed to further refine predictive models for harmful algal blooms to facilitate early warnings driven by data and ultimately reduce impacts on shellfish industry and human health. A new methodology based on distributed non-linear lag models and advanced Generalized Additive Model structures allowed the relationship between past sea surface temperature or *Dinophysis* spp. abundance information and shellfish toxin concentrations to be estimated and visualised. Contemporary HAB monitoring data gathered during the Project were supplemented with long-term HAB monitoring records obtained from the FSA (Food Standards Agency, 2021). This enabled the evaluation of long-term environmental signals for forecasting *Dinophysis* HAB events, which have been recorded repeatedly, with different magnitudes and durations in the St Austell Bay area.

Analysis of the lagged effect of *Dinophysis* spp. abundance on shellfish toxin concentrations (Figure 2) suggest a lag of about 2 weeks for the study area, at the time of year when blooms are usually at their peak (during the summer). This lag is beneficial for short-term prediction of HAB impacts, as it means the footprint of a severe toxic event can be potentially determined using the cell counts a few weeks before it happens. This new knowledge (of the lags) could also be used to inform optimal scheduling of abundance measurements, to best predict upcoming shellfish sample results. Meanwhile, analysis of the lagged effect of sea surface temperature (Figure 5) suggested colder winter temperatures could play a key role in driving higher and more persistent summer toxin concentrations in shellfish. This novel insight for *Dinophysis* spp. matches observations for other encysting dinoflagellate HAB species (Fischer et al., 2018) and is very significant for regulators and the shellfish industry because it establishes the principle for very long-term predictions of HABs. Further study should focus on long-term predictability based on colder winter temperatures, but also other environmental variables which may determine the long-term trajectory of phytoplankton populations.

A first forecasting experiment was carried out to evaluate the relative effectiveness of predictive models on past *Dinophysis* spp. abundance, versus predictive models based on past sea surface temperature. Models based on abundance (Model A) had the best performance when predicting less than 10-weeks ahead, suggesting abundance provides richer contemporary information about HABs than sea surface temperature (Figure 8). However, when forecasting beyond 10 weeks into the

future, models based on sea surface temperature performed best, highlighting the predictive power of winter and spring temperatures.

A second forecasting experiment sought to quantify the implications for predictive modelling of more frequent data measurements of both *Dinophysis* spp. abundance and shellfish toxin concentrations, enabled by the RPF. Increasing the frequency of toxin measurements in shellfish from once to twice per week during the Project picked up a previously undetected second peak in toxin levels in September 2021, following an earlier (conventional) summer bloom from June to August. This unusual bimodal bloom profile helped to highlight the advantages of increasing the frequency of HAB cell and toxin monitoring data for use in more accurate HAB forecasting.[1] The experiment demonstrated consistent improvements in forecasting accuracy attributable to the additional data (Figure 9), suggesting frequency of data collection is one major bottleneck for reliable monitoring and forecasting of HAB events. Further data bottlenecks include:

- The low signal-to-noise ratio seen in abundance measurements compared to toxin measurements, which mask the trajectory of the bloom from week to week. This suggests toxin concentrations provide a more integrative and reliable measure of HAB risk, with one explanation being that shellfish integrate HAB cells which are unevenly distributed in the water as it circulates around them, producing a moving average. To improve forecasting efforts based on abundance, a more reliable signal could be achieved by increased sampling frequency or producing repeated measurements on sampling days.
- The high proportion of non-detects in the Official Monitoring data for both *Dinophysis* spp. abundance and OA concentrations in shellfish samples, which might obscure the early stages of a bloom and make forecasting impacts more challenging. For toxin measurements, non-detects are generally limited to only periods of time when toxin levels are low (e.g. see Figure 4). For *Dinophysis* spp. abundance, however, non-detects are seen amongst periods of apparently high levels of abundance (see Figure 1), which is problematic for modelling efforts. This problem could be mitigated through the changes in measurement practice, or through the use of alternative methods like those based on qPCR.

The Project period (September 2021 to end of March 2022) was not optimal for testing HAB forecasting models as an operational tool for regulatory decision-making, since the majority of this period corresponds with a quiescent period for dinoflagellate HAB species, including *Dinophysis* species. *Dinophysis* spp. cell counts and sea surface temperature (a key environmental predictor of *Dinophysis* spp. cell counts) declined during September/October and subsequently remained low. On the other hand, the low phytoplankton population levels provided the opportunity to compare the sensitivity of traditional microscopic analysis methods versus modern molecular (qPCR) methods for quantifying low levels of HAB cell abundance in seawater. This comparison showed that qPCR methods can provide quantitative data below the detection levels of microscopic analysis (Figure 10). Notably, the qPCR analysis was able to detect increasing *Dinophysis* spp. abundance during the early spring, while microscopic analysis generated non-detects, which offer no indication of a trend.

[1] Other notable HAB species that were observed to fluctuate in abundance (cell counts in seawater) during the Project period included *Pseudo-nitzschia* spp., but no associated toxins (Domoic acid) were detected in shellfish at any point. Toxin production in Pseudo-nitzschia spp. is often not correlated with cell abundance, and is understood to be influenced by variations in environmental cues, including temperature, irradiance, nutrient levels, and abundances of predators (Anderson, Cembella, & Hallegraeff, 2011; Wells et al., 2015).

This suggests that qPCR data could be especially helpful at times of year when *Dinophysis* spp. populations are low, but expected to increase imminently, to support predictive models in identifying the early warning signs of a potential HAB event.

The comparison between qPCR and microscopic data also highlighted that a statistical approach could be used to integrate the two types of measurement into a more data-rich time series. In practice, qPCR results could be divided by a contemporary estimate of the "qPCR coefficient" (Figure 11) to indicate what cell abundance might be measured using microscopic methods. Given more comparative data (e.g. spanning one or two annual cycles), it may also be possible to back-calculate qPCR-equivalent cell counts from historical microscopic analysis data (i.e. Official Control data gathered over the last 10 years). However, combining these data types would only make sense if non-detects in the microscopic data were handled in a more rigorous way, for instance by treating them as left-censored values in a more complex statistical model.

Strand 2: Nucleic Acid Sequence Amplification-Based Testing (Partners: The National Oceanography Centre, Southampton).

Strand Overview:

The National Oceanography Centre (NOC) were involved in the project to undertake DNA-based testing of water samples to detect and enumerate key phytoplankton genera associated with the onset of toxic harmful algal blooms (HABs) in UK waters. These were *Alexandrium* spp., *Dinophysis* spp. and *Pseudo-nitzschia* spp., of which several species within each genus are known to synthesise potent toxins respectively implicated in Paralytic Shellfish Poisoning (PSP), Diarrheic Shellfish Poisoning (DSP) and Amnesic Shellfish Poisoning (ASP). The DNA-based testing was carried out using established Quantitative Polymerase Chain Reaction (qPCR) assays, including those developed at the NOC (*Alexandrium* spp. and *Pseudo-nitzschia* spp.) and the University of Exeter (*Dinophysis* spp. including *D. acuminata*). The qPCR testing was carried out on genetic material recovered from filter membranes, which were prepared by Cornwall Port Health Authority during boat-based weekly sampling expeditions. The assays were completed using NOC's bench-top qPCR apparatus and optimised workflows. This was the main activity carried out by the NOC during the project.

Additionally, the project was used to demonstrate a proof of concept for a ruggedised, battery powered and field-portable DNA analysis system that could be used to undertake the testing at the point of sample, including on-vessel. Portable qPCR technology has been developed by the NOC over the last decade, culminating in the current system/concept, which utilises a 'tape' featuring spatially separated assay 'pockets', containing complete, dry preserved DNA amplification reagents. The tape is passed over an 'analytical head' module, featuring P.I.D.-controlled heaters and specialist optics to measure fluorescence from fluorometric DNA probes. This enables DNA- or RNA-sequence amplification using qPCR or isothermal assay chemistries. In the current study/project, the prototype tape-based analysis system was fabricated within a robust, water-tight housing with a rechargeable battery and a power supply unit in order to demonstrate and evaluate how the system might be used during the routine monitoring undertaken by CPHA. The primary goal was to obtain experience and feedback from CPHA (as potential end-users of the technology) towards further development and refinement of the concept. The system was originally developed with the assistance of NERC funding (grant number NE/R013721/1) and with support from BioSystems Assure (BSA) Ltd, a technology start-up and innovation partner of this initial research. This strand of the project culminated in a field test

of the DNA analyser on the Cornwall Fire Service R.I.B. during a visit to a shellfish farming location, just off the Cornish coastline.

Methodology:

Sampling:

CPHA were provided with disposable 50mL plastic syringes and Sterivex filtration units (Millipore, 0.22-micron pore size), each with a Luer Lock style connector. Sterivex cartridges are made from a filter membrane, encased within a polycarbonate housing and are used for a variety of microbiological water sampling purposes. They are used as the disposable filtration units on the NOC's Robotic Cartridge Sampling Instrument (RoCSI) which, although not featured in the current project, would ultimately provide the 'front end' water sampling capability in an integrated DNA-sensor. In the current project, water filtration and DNA extraction were done manually. During sampling excursions, seawater was collected with the syringe and pushed through the Sterivex filter unit until approximately 300-400mL of water had been filtered. The volume of water filtered, together with the date, location and water temperature on-site was recorded. The filter units were immediately returned to shore and kept frozen at -80^{o}C until later shipping (on dry ice) and processing at the NOC.

Sample DNA Extraction:

DNA was recovered from the Sterivex filter units using the Power Water DNA isolation Kit, available from Qiagen Ltd. This method has been optimised by the NOC, and is able to recover total DNA from the filter membrane without the need to crack-open the filter housing. During this procedure, the filter housing was filled with a buffered lysis solution before being mechanically agitated and then heated to dislodge and break open biological cells trapped on or within the filter membrane. The lysate was recovered to a sterile tube containing sterile micro-beads, and agitated at high speed to ensure complete lysis of the recovered cells. DNA was purified using solid phase extraction with solvent washes, and finally eluted into Tris-EDTA (pH 8.0) buffer for qPCR analysis. The final volume of eluant from each Sterivex unit was 100µL. Prior to analysis, each extracted DNA sample was analysed using a Nanodrop spectrophotometer, estimating the concentration of recovered DNA and indicating the level of purity. The DNA samples were stored at -20^{o}C until use.

qPCR:

Each DNA sample was analysed using 3 different qPCR assays, targeting *Alexandrium* spp., *Pseudo-nitzschia* spp. or *Dinophysis accuminata*. Assays were previously optimised by the University of Exeter (*D. accuminata*) and the NOC, and summary information is provided in Table 1 below.

Table 1. Oligonucleotide Sequences used During the qPCR Testing

Species	Target	Forward Primer	Reverse Primer
Alexandrium spp.	ITS1-5.8S-ITS2	YGATGAAGAATGCAGCAAMATG	CAAGCAHACCTTCAAGMATATCC
Dinophysis accuminata	5.8S	GCATGCTGTATGTATCACA	AATGAGGCCATACAGACA
Pseudo-nitzschia spp.	18S	CTGTGTAGTGCTTCTTAGAGG	AGGTAGAACTCGTTGAATGC

Each qPCR reaction was prepared in a 'PCR Hood' using nuclease- and DNA-free plastic consumables and reagents. Each qPCR reaction was set-up to contain 12.5µL of Sso Fast EvaGreen

Supermix (Biorad Ltd), 1µL of each primer, and 9µL of PCR-grade water; template DNA (1µL) was added to a total, final reaction volume of 25µL. The qPCR reactions underwent thermal cycling using either a Stratagene MxPro3005P or Roche LightCycler 96 real-time PCR instrument. Each reaction was run for 40 cycles, followed by a dissociation (melting) analysis to ensure product specificity. The thermal cycling parameters were as follows. For *Alexandrium* spp. 95°C for 10 min, then 95°C for 15 sec and 60°C for 60 sec per cycle. For *D. accuminata* 95°C for 2 min, then 95°C for 15 sec, 57°C for 15 sec and 72°C for 45 sec per cycle. Finally, for *Pseudo-nitzschia* spp. 95°C for 2 min, then 95°C for 30 sec, 61°C for 30 sec and 72°C for 30 sec per cycle. Each reaction was run in triplicate for 40 cycles, and the mean cycle threshold (C_t) value for each replicate was used to estimate cell number as described below. Cell number was calculated as cells per 100mL of sampled water after considering the amount of DNA sample added to each reaction, and the quantity of water originally passed through the filter unit.

Quantification of *Alexandrium* spp. and *Dinophysis accuminata* was achieved by preparing standard curves using a 10-fold dilution series of synthetically produced DNA duplex sequences containing the assay target sequence. The DNA standards were synthesised by Integrated DNA Technologies Ltd and purified by HPLC, before being diluted to the appropriate concentrations using PCR-grade water. The mean C_t value obtained from triplicate qPCR reactions for each standard and dilution were used to plot standard curves, from which the number of template sequence copies in each analysis was extrapolated using a simple linear regression of the standard curve data points. The number of target sequence copies in the genome was considered when estimating the number of cells on each filter.

Quantification of *Pseudo-nitzschia* spp. was achieved using cell number standards. These were prepared from a culture of *Pseudo-nitzschia multistriata* (SZN-B954, originally recovered from the Gulf of Naples, Italy) in f/2 growth medium, maintained at 18°C with a 12-hour photoperiod. An exponentially dividing culture was enumerated using a Sedgwick–Rafter cell counting slide (PYSER-SGI) and a compound light microscope (Carl-Zeiss), and dilutions were prepared in f/2 medium. DNA was extracted using the DNeasy Power Water kit and the DNA samples were used as template for a series of qPCR reactions, which were used to construct the standard curve.

Portable DNA Analysis System Field test:

A prototype DNA analysis system, which is capable of performing qPCR and isothermal DNA amplification assays, was installed into a water-tight, robust housing to undertake a field test of the instrument. The system was originally developed as part of NERC grant NE/R013721/1, and with support from BSA Ltd (www.biosysa.com), and the purpose of the field test was to demonstrate potential applications of the system for HAB surveillance on-board a small vessel during a routine sampling expedition. The prototype uses a specialised polycarbonate ‘tape’ featuring pre-prepared and packaged assay reagents, distributed into discreet ‘pockets’ which can be stored without cold-chain and activated in a single step by the addition of an aqueous DNA sample. To prepare the system for field testing, a 3D-printed assembly was fabricated in ABS using a FDM 3Dprinter, and the assembly was bolted onto a 2mm thick steel sheet, secured to a fibreglass housing using Stainless Steel bolts. The lid of the housing was secured using plastic locking bolts with a rubber O-ring gasket to make the whole assembly water tight. A control panel and power supply unit (PSU) interface (including master power switch, fuse and 12V DC Charging Jack) were also 3D printed, and all components and joints were assembled and secured with Stainless Steel bolts and a silicone sealant. The mechanical system comprised a stepper-motor-driven tape positioning mechanism with a reflective object sensor for

alignment. It uses solenoid 'latches' to position/hold the tape against aluminium plates, which are heated using P.I.D-controlled, ceramic cartridge heaters. The system was controlled using an Arduino MEGA2560 microcontroller board, and powered using a 14V Li-ion, rechargeable battery. The outer-shell of the system can be seen in Figure 12, shown below. Further information on the system can be obtained by contacting BSA Ltd (info@biosysa.com).

Figure 12: A photograph of the portable DNA analysis system prototype outer shell, which was used to demonstrate the concept of in situ DNA/qPCR analysis. The system was developed by collaboration between the National Oceanography Centre and BioSysA Ltd.

The instrument was transported to CPHA headquarters in Falmouth, Cornwall on 30th March 2022, and taken on-board a R.I.B. vessel, operated by the Cornwall Fire Service. The boat was used to visit a Shellfish farm, close to the coastline, where a demonstration of how the system could operate for routine HAB surveillance was undertaken with representatives from the NOC, University of Exeter, CPHA and BSA Ltd present. A water sample was filtered, and used to prepare a crude lysate, which could be used for DNA amplification. In this particular demonstration, the lysate was applied directly to a pre-prepared assay mixture within the analytical 'tape', which was subsequently loaded into the instrument for analysis.

Findings:

The results of the filter membrane DNA extractions are summarised in Table 2 Below. DNA was recovered from all of the filter membranes prepared by CPHA, with a typical yield in the order of a few hundred nanograms per filter. Spectrophotometric analysis of the purified DNA samples generally showed a single peak absorbance at 280 nm, with A260/280 ratio of >1.7. The measured A260/30 values were generally low (below 1.7) consistent with some contamination; this is typical of DNA recovered from seawater using the described method and did not inhibit amplification of the target DNA sequences by qPCR.

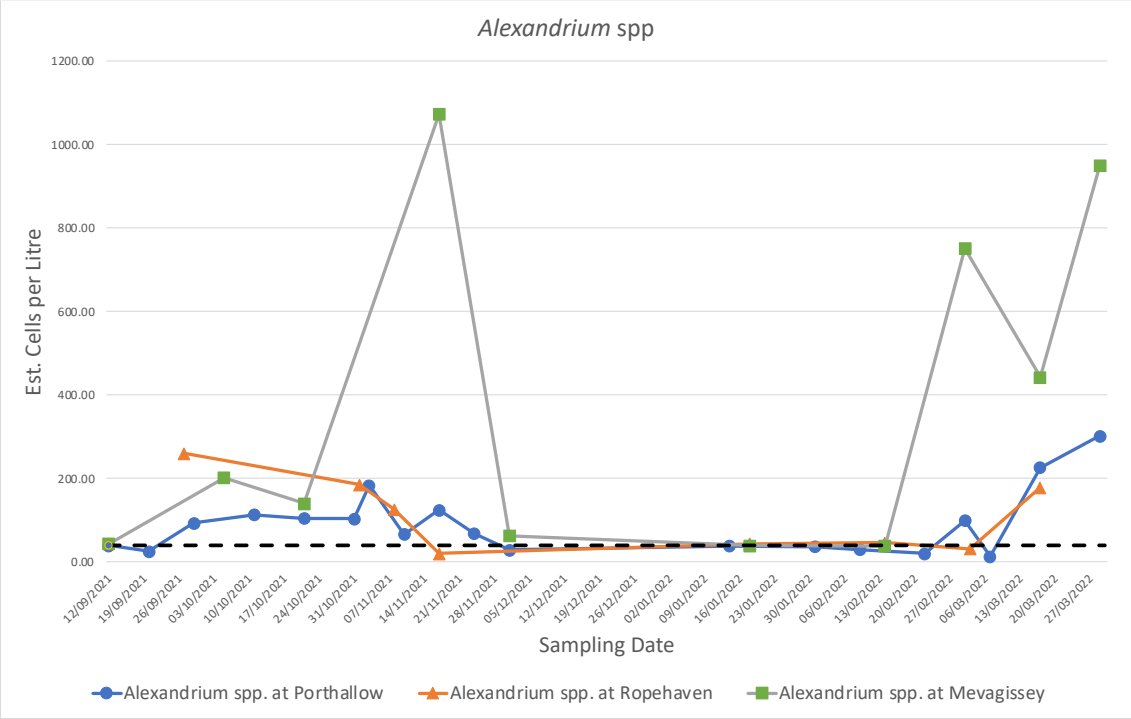
Table 2: Filtered Seawater DNA Extractions Results

Sampling Date	Sample Location	Temperature at Site (Celcius)	DNA Concentration (ng/uL)	A260/80	A260/30
12/09/2021	South Mevagissey (B7OAK)	17.3	3.01	2.31	0.51
20/09/2021	Porthallow (B34AA)	16.9	3.47	2.18	0.48
27/09/2021	Porthallow (B34AA)	16.2	2.26	1.7	-12.02
29/09/2021	Rophaven Outer (B70AE)	16	10.56	2.2	0.2
04/10/2021	Porthallow (B34AA)	14.8	8.99	2.05	0.44
05/10/2021	South Mevagissey (B7OAK)	14.8	7.5	1.96	0.91
11/10/2021	Porthallow (B34AA)	15.1	2.72	1.46	0.61
18/10/2021	Porthallow (B34AA)	15.3	3.03	13.32	0.21
21/10/2021	South Mevagissey (B7OAK)	14.9	8.45	1.59	0.11
25/10/2021	Porthallow (B34AA)	14.7	6.51	1.82	0.2
27/10/2021	Rophaven Outer (B70AE)	15	3.22	1.24	0.04
31/10/2021	Rophaven Outer (B70AE)	8	3.97	1.31	0.68
01/11/2021	Porthallow (B34AA)	14.1	3.51	1.5	0.56
03/11/2021	Rophaven Outer (B70AE)	13.8	7.98	1.49	1.34
08/11/2021	Porthallow (B34AA)	13.9	6.13	1.51	1.01
10/11/2021	Rophaven Outer (B70AE)	14.4	5.6	2.42	0.94
17/11/2021	South Mevagissey (B7OAK)	12.8	3.43	2.44	0.56
17/11/2021	Porthallow (B34AA)	13.5	4.04	2.05	0.3
24/11/2021	Porthallow (B34AA)	12.4	4.45	1.57	1.67
24/11/2021	Rophaven Outer (B70AE)	12.8	2.01	3.34	2.99
29/11/2021	Porthallow (B34AA)	10.9	3.4	1.7	0.02
01/12/2021	South Mevagissey (B7OAK)	12.4	3.21	4.3	0.03
14/01/2022	Porthallow (B34AA)	10	4.03	1.78	0.71
18/01/2022	South Mevagissey (B7OAK)	9.7	1.89	2.42	0.33
24/01/2022	Rophaven Outer (B70AE)	8.9	2.43	2.77	0.25
26/01/2022	Porthallow (B34AA)	9.7	5.95	1.93	0.68
31/01/2022	South Mevagissey (B7OAK)	10.2	1.32	2.68	0.42
02/02/2022	Porthallow (B34AA)	10	2.98	1.44	0.64
07/02/2022	Porthallow (B34AA)	10.6	3.19	1.24	0.49
09/02/2022	Rophaven Outer (B70AE)	10.6	4.31	1.34	0.25
14/02/2022	Porthallow (B34AA)	10.2	2.59	1.42	0.26
14/02/2022	South Mevagissey (B7OAK)	10.4	2.41	1.62	0.29

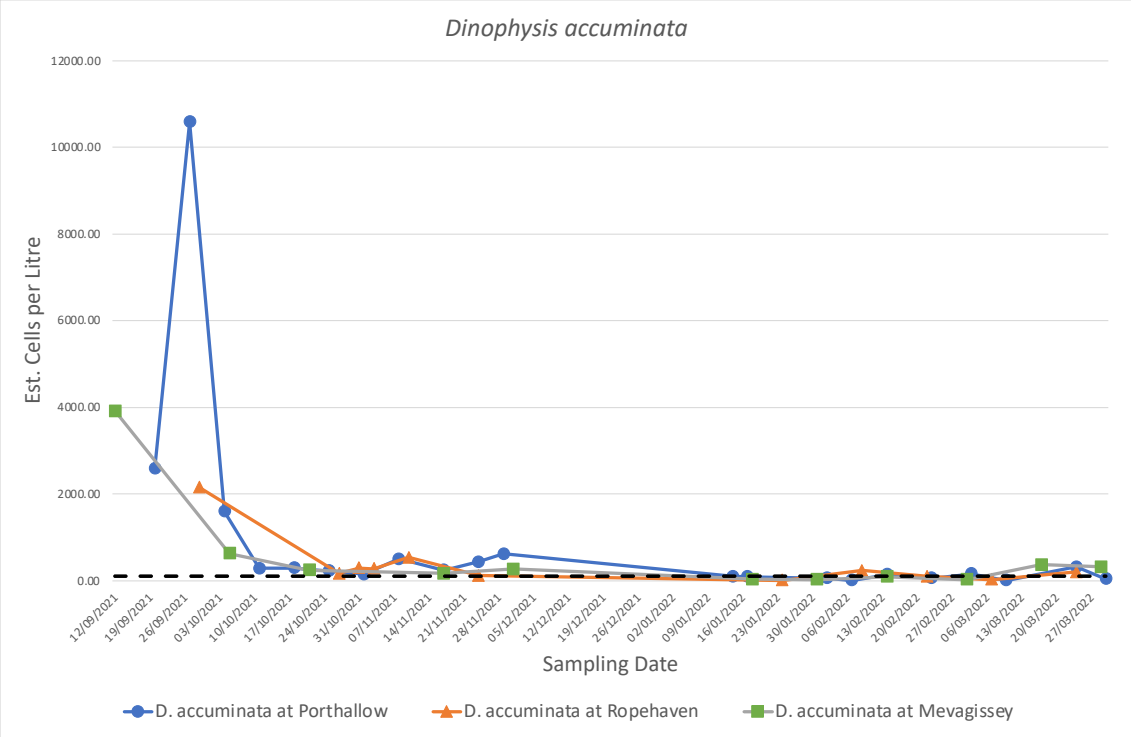
22/02/2022	Rophaven Outer (B70AE)	10.7	3.39	1.97	0.35
23/02/2022	Porthallow (B34AA)	10.6	1.88	1.81	0.16
02/03/2022	South Mevagissey (B7OAK)	10.2	0.66	1.72	1.28
03/03/2022	Porthallow (B34AA)	10.1	3.54	1.65	1.66
07/03/2022	Rophaven Outer (B70AE)	10.3	2.72	2.66	0.68
10/03/2022	Porthallow (B34AA)	10	1.53	1.93	1.21
17/3/2022	Porthallow (B34AA)	10.2	3.57	1.78	0.36
17/03/2022	South Mevagissey (B7OAK)	10.8	3.57	1.94	0.24
24/03/2022	Rophaven Outer (B70AE)	10.9	1.99	1.69	1.02
24/03/2022	Porthallow (B34AA)	11.5	2.07	1.95	1.03
29/03/2022	South Mevagissey (B7OAK)	11.3	4.57	1.47	1.41
30/03/2022	Porthallow (B34AA)	11.5	2.7	1.15	0.47

The extracted DNA was analysed by qPCR to estimate the number of cells for each target species/group on the CPHA-prepared filter samples. The qPCR results were prepared for each species/group to show the number of cells per 100mL of water, at each sampling location over the 6-month time-course. For *Alexandrium* spp. and *D. accuminata*, the results were obtained by measuring the number of target sequence copies, and from this estimating the number of algal cells based upon the number of target sequence copies per genome/cell. These are presented below (Figure 13).

Figure 13: qPCR Estimation of *Alexandrium* spp. and *D. accuminata*



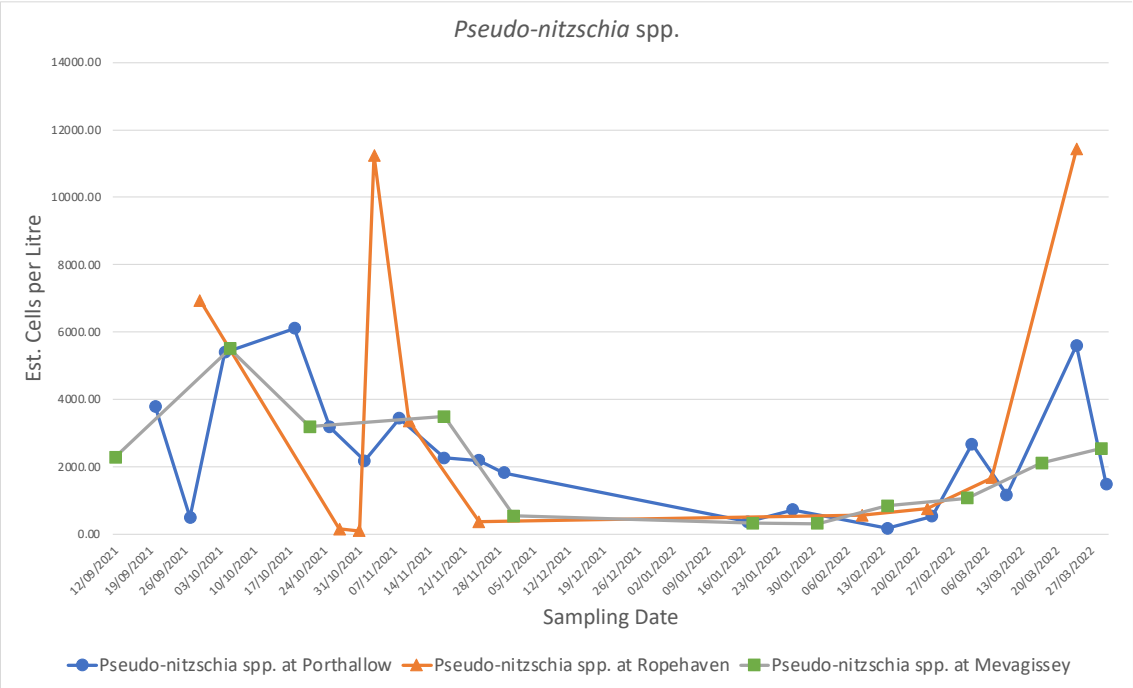
Dashed line represents suggested 'trigger' level of 40 *Alexandrium* cells/Litre in seawater



Dashed line represents suggested 'trigger' level of 100 *Dinophysis* cells/Litre in seawater

For *Pseudo-nitzschia* spp. the results were obtained by comparing the qPCR threshold cycle (C_t) value against those obtained from amplifying DNA extracted from 'cell standards', which were cultured *P. multistriata* cells, diluted to a known cell concentration. These results are shown below in Figure 14.

Figure 14: qPCR Estimation of *Pseudo-nitzschia* spp.



Trigger level of 150,000 cells/Litre in seawater (not shown on graph)

Portable DNA Analysis Prototype Test:

A prototype DNA analysis system was built in order to carry out a demonstration to CPHA (and the wider project team) of how qPCR results could be generated on-site by minimally trained personnel, without the need to send collected filter samples back to an analytical laboratory. A fully-functional system could not be achieved within the very limited time-frame and budget of this particular project, however the primary purpose the test was (i) for the development team to experience the conditions, workflow and requirements associated with the routine sampling and testing undertaken by the CPHA in order to guide further development of the system, and (ii) for the end-user (CPHA) to understand how such as system could be utilised to address the shortfalls of current testing methods; ultimately (in further work) the system could be placed with CPHA for a longer-term trial.

Analysis and Conclusion

The qPCR analysis produced an estimation of the populations of *Alexandrium* spp. *Dinophysis* spp. and *Pseudo-nitzschia* spp. in Cornish shellfisheries waters between September 2021 and March 2022. In general, there were low levels of HAB cells over the duration of the study, which was expected due to the season and associated low temperatures. *Alexandrium* spp. cells exceeded the 40 cell per litre 'trigger' level only once during mid-November at the Mevagissey site. At this concentration, additional sampling is required/enforced to monitor the levels of PSP toxin in shellfish flesh. At around the same time of year the other two sampling locations also generated elevated levels of *Alexandrium* spp. cells, albeit this was relatively slight. Towards the end of the project, from late February into march, levels of *Alexandrium* spp. cells increased at each site, which may reflect an early-Spring increase in temperature and sunlight, and this again was most pronounced at the Mevagissey sampling site. In contrast, *D. accuminata* cells were, in general, measured at very low levels throughout the project at all sites, with the exception that the first few samples, collected in early to mid-September. In these samples the levels of *D. accuminata* were significantly elevated, reaching more than 10,000 estimated cells per litre of filtered water at the Porthallow site; the effect was less pronounced at the other locations. There was evidence of an increase in *D. accuminata* populations in the early spring (March) at Mevagissey. For *Pseudo-nitzschia* spp. there was a generally more variable trend in the estimated cell numbers over the project duration, with elevated levels detected across all the sampling locations during the early and late periods, and minimal levels detected during the winter. The estimated levels of *Pseudo-nitzschia* cells were significantly greater than those for *Alexandrium* spp. and *D. accuminata* throughout, however did not reach the 150,000 cell/litre 'trigger' level at which additional shellfish sampling would be enforced.

The qPCR assays employed for this project have been developed and optimised over many years, with well-established and streamlined assay workflows. In work undertaken prior to this project, qPCR estimates of various HAB species have been shown to correlate well with direct counting methods (microscopy and flow-cytometry). In this work, for *D. accuminata*, the estimated cell numbers as determined by qPCR followed the same trend as those determined by direct counting (microscopy), which can be seen in Figure 10. However, the qPCR estimations were consistently higher, across all sites, and also identified periods where *D. accuminata* levels were significantly higher than indicated by microscopy, specifically at the Mevagissey site towards the end of the project (early spring). The qPCR also indicated that, at all sites, the decline in *D. accuminata* cell numbers into the winter was more gradual than indicated by microscopy. The reason for these differences may

reflect that, although qPCR is a particularly sensitive method, this can often lead to an over-estimation of cell numbers, particularly where there may be an excess of genetic material that is not associated with living, viable cells. Furthermore, whilst the qPCR target sequences were selected based upon their inclusivity and selectivity for the target cell type, there is likely to be some co-amplification of non-target species, which may also lead to over-estimation. It is important to note, however that other techniques such as microscopy suffer from their own unique limitations, for example sub-sampling and overlapping morphometric characteristics between different HAB genera can lead to over- or under-estimation. The principal advantages of using qPCR are the sensitivity (up to less than single cell sensitivity in some cases due to multiple repeats of target sequences in HAB genomes), the speed of analysis and the ease with which qPCR can be adapted for portable or deployable field-based analysis set-ups. The latter is particularly important because on-site analysis removes the effect of bottling and transporting a living sample to a centralised lab, improving accuracy and reducing cost. Accordingly, a second objective of this project strand was to demonstrate the potential for an on-site qPCR instrument and workflow by conducting a field test of a prototype system during a routine sample collection from shellfish waters.

The trial of the portable DNA analysis prototype was intended to investigate the concept of using an *in situ* DNA analysis instrument instead of the usual routine of sample collection and analysis in a centralised lab. Although the prototype was partially functional (could undertake basic DNA analysis, but did not at the time have full real-time PCR capability), the general form and workflow could be demonstrated whilst on-board a small vessel at one of the sampling sites. This experience has been essential to the scientific and engineering personnel undertaking further development of the system, and provided CPHA as a potential end-user of the system to understand the methodology, and to contribute to discussions on improvements and additions. During the deployment the prototype was subjected to significant vibration/inertial force, etc as we journeyed to the sampling site at speed, and also suffered some splashes with seawater, but remained operational throughout due to the sturdy construction. In future work (see section on general conclusions at the end of this report), this experience will directly contribute to the fabrication of a next generation prototype, which will be fully functional and could be placed with CPHA to undergo independent evaluation over months/years as a new tool to support routine HAB monitoring. This, unfortunately, was not possible within the scope, budget and duration of this particular project.

Strand 3 – CEFAS validating Labs (Partners: Centre for Environment, Fisheries and Aquaculture Science)

Strand Overview

Cefas were contracted through this project to undertake the analysis of water for the presence of *Dinophysis* phytoplankton species and shellfish samples for levels of harmful marine biotoxins. The geographical area of interest, St. Austell Bay in Cornwall, is monitored on a routine basis as part of the official control (OC) monitoring programme on behalf of the Food Standards Agency. However, monitoring frequencies are intermittent, particularly during autumn and winter, so in this study Cefas laboratories were tasked to undertake additional testing activities to provide additional phytoplankton and toxin data to supplement the OC data and increase the data available for statistical assessment. The aim of this strand therefore was to conduct testing of both water samples and shellfish (mussel) samples for phytoplankton and toxins so as to deliver a set of data with weekly testing results. Samples were taken across three different sites in St. Austell Bay, sent to the Cefas

laboratories in Lowestoft and Weymouth and processed using validated and UKAS-accredited testing methods. Once results were generated, data was shared with all project participants and utilised for predictive modelling purposes.

Methodology

Reagents and chemicals

Methanol utilised for sample extraction was of HPLC grade. Water and acetonitrile used for LC–MS/MS mobile phases and instrument wash reagents were of LC–MS grade. Sodium hydroxide (VWR International Ltd, UK) and hydrochloric acid (Fisher Scientific UK Ltd) were analytical grade. Mobile phase additive, 25–31% ammonium hydroxide, was of LC–MS grade (Sigma-Aldrich, Poole, England). Certified reference materials for OA, DTX1, DTX2 and PTX2 were purchased from the Institute of Biotoxin Metrology, National Research Council Canada (NRCC, Halifax, Nova Scotia, Canada). Primary toxin standards were diluted in 100% methanol to form concentrated stock standard solutions prior to further dilution for production of calibration standards.

Sampling and transportation

Three geographical sampling sites were incorporated into the St. Austell Bay study:

1. Ropehaven Outer, Bed Id B70AE, Grid ref SX05744972
2. Porthallow North, Bed Id B34AA, Grid ref SW80212383
3. South Mevagissey Bottom, Bed Id B70AK, Grid ref SX05214698

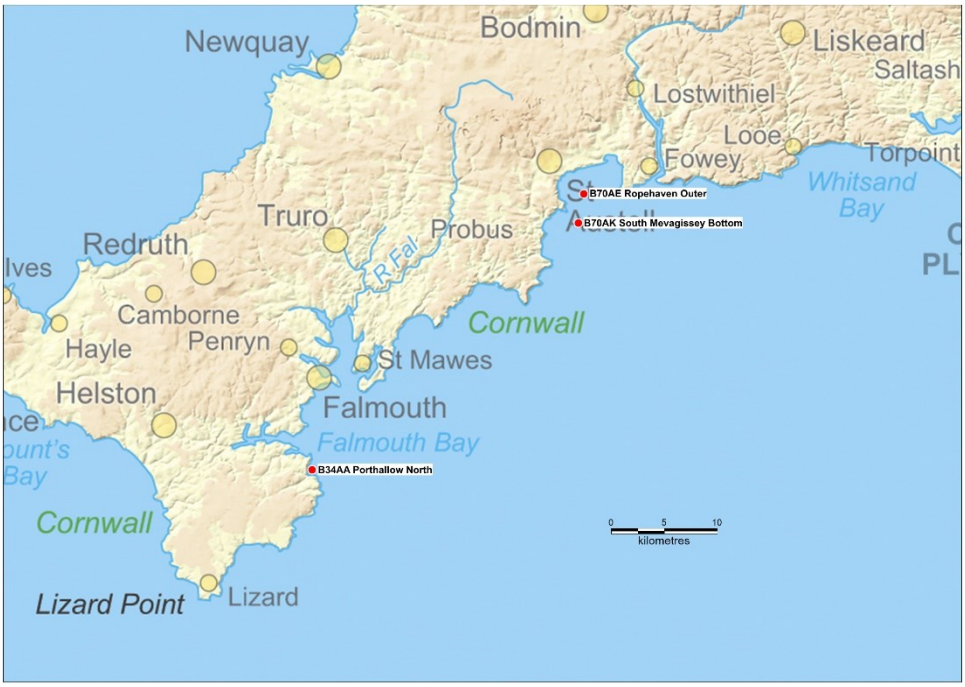


Figure 15. Map of SW Cornwall highlighting three shellfish sampling areas

Water samples were taken by Cornwall Port Health officers between 13th September 2021 and 30th March 2022. Pole collection was used for the majority of samples, with surface samples taken on just two occasions. 500 mL water samples were taken at the representative monitoring points for each of the three classified harvesting areas. Once a sample had been taken, it was fixed using

Lugol's iodine solution, before being transported to the Cefas Lowestoft laboratory under temperature-controlled conditions.

Mussel samples were taken from the same monitoring points as the water samples. For each sample enough live mussels were taken to provide a minimum of 100g of mussel meat once shucked. The mussels chosen for the samples were those representing the entire population in terms of size. Once the mussels were removed from their location, they were placed into a polythene sample bag, and packed into a cool box containing ice packs and foam protection. These boxes were then shipped to the Cefas Weymouth laboratory for toxin testing.

Water testing

The 500 mL water sample bottle was slowly and gently inverted 10 times to homogenously mix the water containing the phytoplankton cells. Immediately on completion of the 10 inversions, the sample was poured into a 25 mL settling chamber until the chamber just begins to overflow. A glass chamber cover was immediately slid onto the top of the tower, cutting off the meniscus before it had time to contribute significant additional suspended particles. After a few hours, the 25 mL chamber was scanned on the microscope to determine if a smaller subsample should be settled. The chambers were left to settle as follows:

- A 5 mL chamber must have settled for a minimum of 4 hours.
- A 10 mL chamber must have settled for a minimum of 8 hours.
- A 25 mL chamber must have settled for a minimum of 12 hours.

The settling chamber was carefully mounted into a Perspex stage holder on a high power, inverted microscope. The microscope was focussed, and light level adjusted to ensure maximum resolution for resolving morphological features of the phytoplankton cells. This enabled cell identification to the lowest taxonomic level. A quick scan of the sample was performed to determine the method of analysis to be used. Two enumeration methods were used, depending on the cell density of reportable taxa. The first method was employed when the cell concentration of a reportable taxon was low (less than approximately 4 cells per field of view (FOV)). This is the preferred method of enumeration. In this case the whole base of the chamber was scanned for the presence of these cells at a magnification of x200. This allowed a minimum detection level of 40 cells per litre, if a 25 mL chamber were used. However, species identification may require higher magnification. The second method was used when cells of a reportable taxon were too numerous to ensure accurate counting over the whole base of the chamber (greater than approximately 4 cells per FOV). In this case random FOV are counted. The whole base plate was systematically viewed to allow the total number of *Dinophysiceae spp* observed to be recorded, speciating each individual as much as possible. Empty cells were not included in the analysis. Where a cell could be identified and contained visible cell contents (whether a half cell or damaged), it was included in the analysis count. The concentrations of algal cells in terms of cells per litre was subsequently calculated by multiplying the cell count by the raising factor (1000/subsample size).

Shellfish testing

Mussel samples were rinsed to remove debris, allowed to drain and the edible flesh removed and collected in a pot. Once a minimum of ten animals and 50 g flesh had been collected, the tissue was thoroughly homogenised using a high-speed blender, prior to extraction and analysis following the EU Reference Laboratory (EURL) reference method (EURLMB, 2015). 2.0 ± 0.01 g of each mussel tissue homogenate was then weighed into separate polypropylene 50 mL centrifuge tubes and a unique sample id applied to each sample.

9 mL methanol was added and vortex mixed for 3 min before centrifuging for 8 min (3500 rpm). After the supernatants were collected, the remaining pellets were subjected to a second extraction step whereby 9 mL of methanol was added and the content was vortex-mixed for a further 3 min before centrifuging again for 8 min (3500 rpm). A final third extraction and centrifugation was conducted prior to combining the supernatants from the three extractions in 20 mL volumetric flasks and diluted to 20 mL with pure methanol prior to filtering with 0.2 µm nylon filters (Phenomenex, Manchester, UK). For each batch of samples analysed, a negative control procedural blank and positive control laboratory reference material (LRM) was co-extracted for quality control assessment. Hydrolysis of methanolic extracts was conducted based on the procedure described by Mountfort et al. (2001). 125 µL of 2.5M sodium hydroxide was added to 1 mL of extract, vortex mixed for 5 s and heated for 40 min at 76 °C using the block heater. Samples were cooled to ambient temperature, neutralised using 125 μ L 2.5M hydrochloric acid and vortex mixed for 5 s. Hydrolysed samples were analysed alongside a crude aliquot of unhydrolysed extract using LC–MS/MS.

Two UHPLC systems (Acquity and Acquity I-class) were coupled to a Xevo TQ and Xevo TQ-S triple quadrupole mass spectrometer respectively (Waters Ltd., Manchester, UK). The alkaline (pH 11) LC method described by Gerssen et al. (2009) was adopted with modifications and subjected to single laboratory validation prior to implementation into the routine biotoxin monitoring programmes. Mobile phase A comprised of deionised water adjusted to pH 10.7 +/- 0.2 with 0.1% ammonium hydroxide. Mobile phase B was 90% acetonitrile with 0.1% ammonium hydroxide. Chromatographic and mass spectrometry conditions were as detailed by Dhanji-Rapkova et al., 2018. Instrument data were analysed using MassLynx^{TM} v.4.1 (Waters Ltd.). LC-MS/MS performance was checked applying quality control (QC) criteria outlined in internal SOPs and the EURLMB SOP (EURLMB, 2015). External calibrations generated from analysis of calibrant solutions were used for quantitation of all target compounds. Individual OA-group toxin concentrations were calculated with post-hydrolysis concentrations equating to concentrations of free toxins and esterified toxins combined. Toxicity Equivalent Factors (TEFs) recommended by EFSA (2009b) were applied and toxin concentrations were subsequently summed into representative groups as stipulated by EU legislation (Anon, 2004b). The OA group included OA, DTX1, DTX2, PTX1 and PTX2 and the combined concentration was expressed in µg OA eq./kg. Unhydrolysed extracts were used for quantitation of all toxins, including OA, DTX1 and DTX2 in their free form. Total (free form plus fatty acid esters) concentrations of these toxins were assessed using hydrolysed samples only.

Results

Phytoplankton analysis of preserved water samples revealed the presence of a range of *Dinophysis* species including most commonly *D. acuminata* complex, *D. ovum, D. fortii,* together with lower amounts of *D. sacculus, D. acuta, D. dens, D. caudata,* as well as *Phalacroma rotunda* and *Phalacroma nasturtium.* Notably, higher cell densities were enumerated in the first month of samples, between 13th September 2021 and 17th October 2021.Table 3 summarises the results obtained from the samples between these dates.

Table 4 summarises the DSP toxin concentrations quantified in the mussel samples received throughout the project between 20th September 2021 and 30th March 2022. Concentrations are tabulated for both freely extractable DSP toxins (OA, DTX1 and DTX2) as well as total DSP toxins quantified after alkaline hydrolysis (incorporating both free and esterified toxins). The LC-MS/MS method also includes PTX toxins, but none were detected, so are not included in the table. Other lipophilic toxins such as Azaspriacids and Yessotoxins were also acquired in the method, but not reported here as have no relevance to *Dinophysis* sp. presence and the predictive modelling being developed. However, overall, only low/trace levels of these toxins were occasionally observed.

In terms of DSP toxin presence, the dominant toxin detected and quantified in the mussel tissues was OA, with only trace levels of DTX2 and very occasionally trace DTX1. On average 97% ± 8% of the total DSP toxin concentrations consisted of OA. The results also show the high proportion of OA to be present in ester form, with on average only 15% ± 17% freely-extractable. This means that the alkaline hydrolysis step is essential to avoid significantly underestimating total OA-group toxicity in mussel samples from this region.

Table 3. Summary of *Dinophysis* cell densities (cells/L) enumerated from Cornish water samples during project. ND = not detected. Cells shaded red are results where cell densities have exceeded trigger threshold levels.

Sample	Production Area	Bed ID	Grid Reference	Sampling Point	Sample Collection Method	Date Sample Collected	Total Dinophysiaceae cells	Dinophysis sp	Dinophysis acuminata complex	Dinophysis ovum	Dinophysis fortii	Dinophysis sacculus	Dinophysis acuta
1	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	13/09/2021	**1080**	**40**	**840**	**80**	**120**	ND	ND
2	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	15/09/2021	**1720**	**40**	**1480**	**120**	**80**	ND	ND
745/21	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	15/09/2021	**160**	ND	**160**	ND	ND	ND	ND
3	Porthallow Cove	B34AA	SW80212383	Porthallow North	SURFACE	20/09/2021	**920**	**120**	**480**	**160**	**120**	**40**	ND
752/21	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	20/09/2021	**1000**	ND	**1000**	ND	ND	ND	ND
4	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	22/09/2021	**800**	**80**	**440**	**200**	ND	ND	ND
767/21	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	22/09/2021	**760**	**80**	**480**	**120**	ND	ND	**40**
5	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	27/09/2021	**80**	**40**	**40**	ND	ND	ND	ND
6	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	27/09/2021	**160**	**40**	**120**	ND	ND	ND	ND
7	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	29/09/2021	0	ND	ND	ND	ND	ND	ND
790/21	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	29/09/2021	0	ND	ND	ND	ND	ND	ND
8	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	04/10/2021	0	ND	ND	ND	ND	ND	ND
9	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	06/10/2021	**40**	ND	ND	**40**	ND	ND	ND
814/21	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	11/10/2021	**40**	ND	**40**	ND	ND	ND	ND
10	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	11/10/2021	**120**	**40**	ND	ND	ND	ND	**40**
815/21	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	11/10/2021	**80**	ND	**80**	ND	ND	ND	ND
11	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	13/10/2021	**120**	ND	**80**	ND	ND	ND	**40**

831/21	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	10/13/2021	**160**	ND	**120**	ND	ND	**40**	ND
12	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	18/10/2021	0	ND	ND	ND	ND	ND	ND
842/21	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	17/10/2021	**200**	**40**	**120**	ND	ND	ND	ND
13	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	21/10/2021	**40**	ND	**40**	ND	ND	ND	ND
14	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	25/10/2021	0	ND	ND	ND	ND	ND	ND
850/21	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	10/20/2021	**80**	ND	ND	ND	ND	ND	**40**
15	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	27/10/2021	0	ND	ND	ND	ND	ND	ND
16	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	02/11/2021	0	ND	ND	ND	ND	ND	ND
861/21	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	02/11/2021	**40**	ND	ND	ND	ND	ND	ND
17	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	03/11/2021	0	ND	ND	ND	ND	ND	ND
18	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	08/11/2021	**40**	ND	**40**	ND	ND	ND	ND
19	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	10/11/2021	0	ND	ND	ND	ND	ND	ND
20	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	17/11/2021	0	ND	ND	ND	ND	ND	ND
21	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	17/11/2021	**40**	ND	**40**	ND	ND	ND	ND
911/21	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	17/11/2021	0	ND	ND	ND	ND	ND	ND
22	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	24/11/2021	0	ND	ND	ND	ND	ND	ND
23	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	24/11/2021	0	ND	ND	ND	ND	ND	ND
24	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	01/12/2021	**40**	ND	ND	ND	**40**	ND	ND
25	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	29/11/2021	0	ND	ND	ND	ND	ND	ND
923/21	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	06/12/2021	0	ND	ND	ND	ND	ND	ND
26	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	17/01/2022	0	ND	ND	ND	ND	ND	ND
27	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	18/01/2022	0	ND	ND	ND	ND	ND	ND
28	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	24/01/2022	0	ND	ND	ND	ND	ND	ND
29	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	26/01/2022	0	ND	ND	ND	ND	ND	ND
30	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	01/02/2022	0	ND	ND	ND	ND	ND	ND
31	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	02/02/2022	0	ND	ND	ND	ND	ND	ND
32	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	02/07/2022	0	ND	ND	ND	ND	ND	ND

33	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	02/09/2022	0	ND	ND	ND	ND	ND	ND
34	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	14/02/2022	0	ND	ND	ND	ND	ND	ND
35	Mevagissey Bay	B70AK	SX05214698	South Mevagissey Bottom	POLE	14/02/2022	0	ND	ND	ND	ND	ND	ND
36	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	22/02/2022	0	ND	ND	ND	ND	ND	ND
37	Porthallow Cove	B34AA	SW80212383	porthallow North	POLE	23/02/2022	0	ND	ND	ND	ND	ND	ND
38	South Mevagissey	B70AK	SX05214698	South Mevagissey Bottom	POLE	03/02/2022	0	ND	ND	ND	ND	ND	ND
39	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	02/03/2022	0	ND	ND	ND	ND	ND	ND
40	St Austell Bay	B70AE	SX05744972	Ropehaven Outer	POLE	07/03/2022	0	ND	ND	ND	ND	ND	ND
41	Porthallow Cove	B34AA	SW80212383	Porthallow North	SURFACE	10/03/2022	0	ND	ND	ND	ND	ND	ND
42	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	14/03/2022	0	ND	ND	ND	ND	ND	ND
43	Porthallow Cove	B34AA	SW80212383	Porthallow North		24/03/2022	0	ND	ND	ND	ND	ND	ND
44	St Austell Bay	B70AE	SX05744972	Ropehaven Outer		24/03/2022	0	ND	ND	ND	ND	ND	ND
45	South Mevagissey	B70AK	SX05214698	South Mevagissey Bottom	POLE	29/03/2022	0	ND	ND	ND	ND	ND	ND
46	Porthallow Cove	B34AA	SW80212383	Porthallow North	POLE	30/03/2022	0	ND	ND	ND	ND	ND	ND

Table 4. Summary of shellfish DSP toxin concentrations (µg/kg) quantified in study samples, highlighting both freely extractable OA-group toxins and total OA-group toxins including toxin esters. Total toxicity values for OA/DTXs shown in µg OA eq/kg, giving actual summed values together with low and high values calculated from measurement uncertainty (<RL = less than reporting limit of 16 µg OA eq/kg). No PTX toxins were detected in any of the samples.

Sample	Collected	Bed ID	Sample location	Free OA	Free DTX1	Free DTX2	OA total	DTX1 total	DTX2 total	Total OA/DTXs/PTXs Low value	Total OA/DTXs/PTXs Actual value	Total OA/DTXs/PTXs High value
BTX/2021/2493	20/09/2021	B70AE	Ropehaven outer	65.5			188.8			107	189	271
BTX/2021/2494	20/09/2021	B70AK	Mevagissey	26.1			77.3			44	77	111
RPF1	20/09/2021	B34AA	Porthallow	2.4			22.8			13	**23**	**33**
BTX/2021/2540	22/09/2021	B34AA	Porthallow	18.3		1.9	54		1.9	30	54	78
RPF2	22/09/2021	B70AK	Mevagissey	25.4			83.5			47	84	120
BTX/2021/2554	27/09/2021	B70AE	Ropehaven outer	42.5		1.6	127.7		2.9	72	128	183
BTX/2021/2555	27/09/2021	B70AK	Mevagissey	30.9			99.3		1.2	56	99	143
RPF3	27/09/2021	B34AA	Porthallow	14.4			45.3			26	45	65
BTX/2021/2610	29/09/2021	B34AA	Porthallow North	15			46.1			23	46	69
RPF4	29/09/2021	B70AE	Ropehaven	42.2			121.1			68	121	174
BTX/2021/2630	04/10/2021	B70AK	South Mevagissey Bottom	26.4		1.4	61		2.5	31	61	91
BTX/2021/2631	04/10/2021	B70AE	Ropehaven Outer	35.5		0.9	85.2		2.2	43	85	127
RPF5	04/10/2021	B34AA	Porthallow	1.4			10.7			<RL	<RL	<RL
BTX/2021/2686	06/10/2021	B70AE	Ropehaven Outer	39.9	1.8		67.6	1.8	3.6	34	68	101
BTX/2021/2687	06/10/2021	B34AA	Porthallow North	4	2.1		12.5	3.6		<RL	<RL	<RL
RPF6	06/10/2021	B70AK	Mevagissey	35			68.3			39	68	98
BTX/2021/2720	11/10/2021	B70AE	Ropehaven Outer	22.1			54.3			31	54	78
RPF7	11/10/2021	B34AA	Porthallow				2.9			<RL	<RL	<RL
BTX/2021/2776	13/10/2021	B70AK	South Mevagissey Bottom	23.9			58.7			33	59	84
BTX/2021/2777	13/10/2021	B34AA	Porthallow North				5.1			<RL	<RL	<RL

RPF8	13/10/2021	B70AE	Ropehaven	8.6			37.6			21	38	54
BTX/2021/2809	18/10/2021	B70AE	Ropehaven Outer	15.2			29.8			17	30	43
BTX/2021/2810	18/10/2021	B70AK	South Mevagissey Bottom	8.8		0.5	31.3		1.6	18	31	45
RPF9	18/10/2021	B34AA	Porthallow				8.8			<RL	<RL	<RL
BTX/2021/2855	20/10/2021	B34AA	Porthallow North	2.1			7.9		1.5	<RL	<RL	<RL
RPF10	21/10/2021	B70AK	Mevagissey	1.7			19.3			11	19	28
BTX/2021/2876	25/10/2021	B70AE	Ropehaven Outer	10.2		2.5	23.2		3.5	13	23	33
RPF11	25/10/2021	B34AA	Porthallow	1.6			9.9			<RL	<RL	<RL
RPF12	27/10/2021	B70AE	Ropehaven	8		0.7	25.3		1.3	13	25	38
BTX/2021/2929	01/11/2021	B70AE	Ropehaven Outer	7.1		1.1	17.3		2.2	10	17	25
RPF13	01/11/2021	B34AA	Porthallow	1.4			5.1			<RL	<RL	<RL
RPF14	03/11/2021	B70AE	Ropehaven	10.3		0.8	28.5		0.8	14	29	43
BTX/2021/2993	08/11/2021	B70AE	Ropehaven Outer	15.4			38.7			20	39	58
RPF15	10/11/2021	B70AE	Ropehaven	3.9		0.4	17.3		0.4	9	17	26
BTX/2021/3080	15/11/2021	B70AK	South Mevagissey Bottom			2.3	7.3		2.3	<RL	<RL	<RL
BTX/2021/3082	15/11/2021	B70AE	Ropehaven Outer				6.7			<RL	<RL	<RL
BTX/2021/3125	17/11/2021	B34AA	Porthallow North				3.1			<RL	<RL	<RL
RPF16	17/11/2021	B34AA	Porthallow	1.8		0.2	7.3		0.2	<RL	<RL	<RL
RPF17	17/11/2021	B70AK	Mevagissey	2.6		0.5	9.5		1.9	<RL	<RL	<RL
BTX/2021/3143	22/11/2021	B70AK	Mevagissey Bay							<RL	<RL	<RL
BTX/2021/3144	22/11/2021	B70AE	St. Austell Bay							<RL	<RL	<RL
RPF18	24/11/2021	B70AE	Ropehaven	2		0.3	11.5		0.2	<RL	<RL	<RL
RPF19	24/11/2021	B34AA	Porthallow	1.2			8.3			<RL	<RL	<RL
BTX/2021/3186	29/11/2021	B70AE	Ropehaven Outer				6.3		4	<RL	<RL	<RL
BTX/2021/3187	29/11/2021	B70AK	South Mevagissey Bottom				5		3.8	<RL	<RL	<RL
RPF20	29/11/2021	B34AA	Porthallow				6.3			<RL	<RL	<RL

RPF21	01/12/2021	B70AK	Mevagissey	2.4		0.3	9.2		0.1	<RL	<RL	<RL
BTX/2021/3238	06/12/2021	B70AK	South Mevagissey Bottom				5.6			<RL	<RL	<RL
BTX/2021/3239	06/12/2021	B70AE	Ropehaven Outer				6.1		0.8	<RL	<RL	<RL
BTX/2021/3339	14/12/2021	B34AA	Porthallow North				4.3			<RL	<RL	<RL
BTX/2022/0002	05/01/2022	B70AE	Ropehaven Outer	0.9		0.4	8.6		0.4	<RL	<RL	<RL
BTX/2022/0003	05/01/2022	B70AK	South Mevagissey Bottom	0.3			4.6			<RL	<RL	<RL
BTX/2022/0074	13/01/2022	B34AA	Porthallow North				0.1			<RL	<RL	<RL
RPF22	17/01/2022	B34AA	Porthallow				3.4			<RL	<RL	<RL
RPF23	18/01/2022	B70AK	Mevagissey				4.9			<RL	<RL	<RL
RPF25	22/01/2022	B34AA	Porthallow				4.5			<RL	<RL	<RL
RPF24	24/01/2022	B70AE	Ropehaven				4.8			<RL	<RL	<RL
RPF26	31/01/2022	B70AK	Mevagissey				5			<RL	<RL	<RL
RPF27	02/02/2022	B34AA	Porthallow				4.8			<RL	<RL	<RL
BTX/2022/0197	07/02/2022	B70AK	South Mevagissey Bottom				8.2			<RL	<RL	<RL
BTX/2022/0198	07/02/2022	B70AE	Ropehaven Outer				5			<RL	<RL	<RL
RPF28	07/02/2022	B34AA	Porthallow				4.8			<RL	<RL	<RL
RPF29	09/02/2022	B70AE	Ropehaven				7			<RL	<RL	<RL
RPF30	14/02/2022	B70AK	Mevagissey				6.8			<RL	<RL	<RL
RPF31	14/02/2022	B34AA	Porthallow				4.2			<RL	<RL	<RL
RPF32	22/02/2022	B70AE	Ropehaven				6.7			<RL	<RL	<RL
BTX/2022/0325	23/02/2022	B34AA	Porthallow North							<RL	<RL	<RL
RPF33	23/02/2022	B34AA	Porthallow				6.5			<RL	<RL	<RL
RPF34	02/03/2022	B70AK	Mevagissey				10.4			<RL	<RL	<RL
RPF35	03/03/2022	B34AA	Porthallow				4.1			<RL	<RL	<RL
RPF36	07/03/2022	B70AE	Ropehaven				7			<RL	<RL	<RL

RPF37	10/03/2022	B34AA	Porthallow				5.5			<RL	<RL	<RL
BTX/2022/0443	11/03/2022	B34AA	Porthallow North				0.8			<RL	<RL	<RL
RPF38	14/03/2022	B34AA	Porthallow				5.4			<RL	<RL	<RL
RPF39	17/03/2022	B70AK	Mevagissey				6.3			<RL	<RL	<RL
BTX/2022/0478	18/03/2022	B70AE	Ropehaven Outer				6.7			<RL	<RL	<RL
BTX/2022/0479	18/03/2022	B70AK	South Mevagissey Bottom				4			<RL	<RL	<RL
RPF40	24/03/2022	B70AE	Ropehaven				6.7			<RL	<RL	<RL
RPF41	24/03/2022	B34AA	Porthallow				6.5			<RL	<RL	<RL
RPF42	29/03/2022	B70AK	Mevagissey				7			<RL	<RL	<RL
RPF43	30/03/2022	B34AA	Porthallow				8.8			<RL	<RL	<RL

Strand Conclusion

Work at Cefas involved both the detection and enumeration of preserved water samples for *Dinophysis* species and the LC-MS/MS analysis of mussel samples for DSP toxins. Validated and accredited methods were used for both tests following UKAS-accredited protocols. Results showed the highest cell densities of *Dinophysis* sp. at the beginning of the project, around September 2021. This coincided, as would be expected, with the highest concentrations of toxins in the mussel flesh. After October, plankton cells and toxin concentrations dropped and remained low or non-detected for the rest of the project.

Strand 4 – Lateral Flow Testing (Partners: Cornwall Port Health Authority)

Strand Overview

The main elements of the project were the first and second strands, which were backed up by CEFAS validated labs for phytoplankton and Diarrhetic Shellfish Poisoning (DSP) toxin (DTX3 / Free OA) in shellfish flesh. Since the second strand focused on harmful algal cell abundance within the water and this was used to predict DSP toxins (a.k.a. Dinophysis toxins) in shellfish, Cornwall Port Health Authority (CPHA) proposed to use a commercially available tool which allows food business operators to detect DTX3 / Free OA in shellfish flesh on a pass / fail basis (set around the regulatory limit). We wished to compare the results to the validated CEFAS labs and provide some additional yet more limited data into the predictive model, if possible.

Methodology

Officers at CPHA followed a standard procedure to obtain extracts of okadaic acid group toxins (OA and DTXs) both in free and esterified form, from the flesh of the shellfish. The process was such that with all the consumables it would be conducted by a person with reasonable knowledge, training and access to an appropriate space.

Unfortunately, the first device supplied had an error which meant that we obtained a series of invalid results, the fault possibly being a misaligned camera inside the device which aims to read the result from a strip inserted in the rear of the device that the clear liquid tracks up to provide a result.

All regulatory exceedances for phytoplankton and shellfish toxin were experienced in September and October 2021, therefore sample testing focused on those samples obtained from Porthallow during those months. Tests for the other sites to the east of the county remain at CEFAS, with the intention of testing, as soon as reasonably able to.

Table 5: Lateral flow test results for Diarrhetic Shellfish Poisoning (DSP) toxin (free and esterified OA-group toxins)

Sample Location	Date of Sample	Time	Temp at sampling	Date of Analysis	DSP Result
Porthallow	20/09/2021	10.15am	16.9°C	05/01/2022	Negative
Porthallow	27/09/2021	10.30 am	16.2°C	05/01/2022	Invalid
Porthallow	04/10/2021	10.20 am	14.2°C	06/01/2022	Invalid
Porthallow	11/10/2021	10.25am	15.1°C	06/01/2022	Positive
Porthallow	18/10/2021	08.00am	15.3°C	03/02/2022	Negative
Porthallow	25/10/2021	09.45am	14.7°C	03/02/2022	Negative

Analysis

Due to issues in the testing process, CPHA tested a relatively small number of samples within the higher risk period. There was a finite amount of shellfish flesh available, which limited the ability to re-test after invalid results were received. There was one exceedance on the lateral flow method, however we have no corresponding official control data, the closest sample due to be obtained was unable to be collected due to a missing mooring, otherwise samples were negative or invalid.

Further testing of samples is required at the St. Austell Bay sites in the high-risk period using the lateral flow test equipment.

Strand Conclusion

At this stage it has not been possible to comment on the viability of the lateral flow test equipment against the Predictive Model and/or the qPCR testing data, for testing shellfish flesh for DTX3/Free OA.

Overall Project Conclusions

- There is a need for a more proactive regulatory environment, where HAB impacts are predicted and mitigated. This can be supported by new technologies and data-driven approaches.
- Strand 1 shows that more frequent data collection improves the accuracy of detecting and predicting HABs and harmful toxin levels in shellfish, but this data collection is limited by human resource and associated cost:
    - Quantitative PCR (Strand 2) is less human resource demanding than conventional microscopic analysis of HAB cell counts and has the potential to be deployed at sea and also provide increased sensitivity for detecting increasing trends at low levels of abundance (<40 cells per L).
    - Lateral flow devices (Strand 4) are also likely to be less resource demanding than conventional LC-MS methods, with the potential for deployment at sea, but their sensitivity is lower than LC-MS.
    - In a world of finite data, we need to be able to combine and utilise different sources and types of data, taking into account different sampling and analytical methods, sampling locations and time points.
- Strand 2 shows there is considerable scope for using qPCR to quantify HAB cell abundance at lower levels than are detectable by conventional microscopic methods. Further work is required to cross-validate these different methods and identify factors which contribute most to differences in apparent sensitivity (e.g. sampling error, co-amplification of DNA from non- viable cells and non-target HAB species during qPCR analysis, and/or failure to detect HAB cells at low abundance under a microscope).
- The inclusion of molecular (e.g. qPCR) and biochemical (e.g. lateral flow) tools and statistical models for detecting and forecasting HABs will benefit shellfish businesses and regulators by better informing decision making around the scheduling of harvesting and monitoring
    - Monitoring and modelling of contemporary trends in HAB cell abundance and HAB toxin levels in shellfish can provide businesses and regulators with up to ~10 weeks warning of an impending HAB event.
    - Statistical analysis of the available data from St Austell Bay provides some evidence of longer-term predictability of HAB events (i.e. colder winter temperatures are associated with more intense blooms of *Dinophysis* spp. in the following spring and summer.
- Obtaining more frequent higher sensitivity molecular (qPCR) data for Dinophysis toxin concentrations in shellfish (as well as *Dinophysis* cell counts) is likely to bring further benefits in terms of improving HAB forecasting and better targeting and reducing monitoring effort.

Recommendations for further work

qPCR as an Alternative 'Tool' for HAB Surveillance:

The inclusion of qPCR-based analysis in this study demonstrated how this 'molecular' technique could be used to monitor 3 different groups of harmful algae over a 6-month time-course, with results generally agreeing with those obtained using the traditional methodology (microscopy). One key benefit of qPCR is the very high-sensitivity. This can be both an advantage, where low-level fluctuations in HAB populations, including those that may precede a 'bloom' may be missed using microscopy (e.g. early spring at Mevagissey there is a marked increase in *D. accuminata* in the qPCR data, but this was not detected by microscopy), and also a disadvantage if overestimation of HAB cell number leads to unnecessary concern. To address this, it is possible to augment the standard qPCR workflow by including sample-pre-treatment processes that effectively eliminate non-cellular DNA, and it has been shown in numerous studies that this reduces or removes the 'over-estimation problem'. For the purposes of predicting/modelling HABs, the over-estimation is most significant if the degree of error is unstable; in this study the direct comparison between microscopy and qPCR indicates this was not the case, and the qPCR estimated cell number increase was more or less consistent over the 6-month period.

The promising correlation between the conventional (microscopy) and new (qPCR) approaches, together with feedback provided to the NOC from CPHA on issues with the LF analysis indicates a real need/potential/scope for an integrated qPCR analysis platform, but one important consideration is cost. In the course of this project samples were collected by hand, requiring vessel time, and then delivered to the NOC by cold-chain. At the NOC, samples were processed using a state-of-the-art laboratory and by a qPCR specialist. This is laborious and expensive. However, the project also considered the potential for adopting an *in situ* analysis approach using new instrumentation. Portable qPCR instruments have been in development in various forms for many decades, however the recent global pandemic has seen record investment in biotechnology methods and instrumentation, and these systems have become highly disruptive technologies in a variety of sectors from healthcare (*in vitro* diagnostics) to environmental science and public health protection. These advances have brought about the expected size and cost reduction that has widened the range of potential applications. The NOC recently partnered with a biotechnology start-up, BioSysA Ltd to capitalise on these recent trends to prepare a low-cost qPCR-based analyser that could be operated at the point of sample, and by a non-specialist operator. An early prototype was demonstrated to CPHA as part of this study. On-going and future work will continue to refine this system, incorporating feedback gained from the project, with the intention of being able to place a fully-functional system with CPHA for a long-term trial. In this scenario, CPHA would be able to carry the system on-board during regular shellfisheries visits, and make on-site qPCR measurements. This would generate the evidence to indicate how qPCR could be adopted by those undertaking routine surveillance, and how it could complement those methods already in use.

Authorship

1. Dr. Ross Brown, University of Exeter.
2. Dr. Oliver Stoner, University of Glasgow.
3. Ms. Suzanne Kay, University of Exeter.
4. Dr. Andrew Turner, Centre of Environment, Fisheries and Aquaculture Science.
5. Dr. Jonathan Mcquillan, National Oceanography Centre.
6. Mr. Timothy Bage, Cornwall Port Health Authority

References

Anderson, D. M., Cembella, A. D., & Hallegraeff, G. M. (2011). Progress in Understanding Harmful Algal Blooms: Paradigm Shifts and New Technologies for Research, Monitoring, and Management. *Annual Review of Marine Science, 4*(1), 143-176. doi:10.1146/annurev-marine-120308-081121

Anon, 2004b. Regulation (EC) No 853/2004 of the European Parliament and of the Council of 29 April 2004 laying down specific hygiene rules for food of animal origin. Off. J. Eur. Union L139, 55–205

Dhanji-Rapkova, M., O'Neill, A., Maskrey, B.H., Coates, L., Alves, M.T., Kelly, R.J., Hatfield, R.G., Rowland-Pilgrim, S., Lewis, A.D., Algoet, M. and Turner, A.D. (2018) Variability and profiles of lipophilic toxins in bivalves from Great Britain during five and a half years of monitoring: okadaic acid, dinophysis toxins and pectenotoxins. *Harmful Algae.* 77, 66-80

EFSA, 2009b. Marine biotoxins in shellfish - summary on regulated marine biotoxins. EFSA J. 7, 1306. http://dx.doi.org/10.2903/j.efsa.2009.1306EURLMB, 2015. Harmonised Standard Operating Procedure for Determination of Lipophilic Marine Biotoxins in Molluscs by LC-MS-MS. Version 5. https://doi.org/ www.aesan.msps.es/en/CRLMB/web/home.shtml

Fischer, A.D., Brosnahan, M. L., Anderson, D. M. (2018). Quantitative Response of *Alexandrium catenella* Cyst Dormancy to Cold Exposure, Protist 169(5):645-661, https://doi.org/10.1016/j.protis.2018.06.001.

Food Standards Agency. (2021). Biotoxin and phytoplankton monitoring. Retrieved from https://www.food.gov.uk/business-guidance/biotoxin-and-phytoplankton-monitoring

Gerssen, A., Mulder, P.P., McElhinney, M.A., de Boer, J., 2009. Liquid chromatography–tandem mass spectrometry method for the detection of marine lipophilic toxins under alkaline conditions. J. Chrom. A 1216 (9), 1421–1430

Stoner, O., Economou, T., Torres, R., Ashton, I., & Brown, R. (2021). Predicting Harmful Algal Blooms and Impacts on Shellfish Mariculture using Novel Data-Driven Approaches. doi:10.21203/rs.3.rs-668820/v1

Wells, M. L., Trainer, V. L., Smayda, T. J., Karlson, B. S. O., Trick, C. G., Kudela, R. M., . . . Cochlan, W. P. (2015). 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