University of the Highlands and Islands
Oilthigh na Gàidhealtachd agus nan Eilean

UHI Research Database pdf download summary

A generic approach for the development of short-term predictions of E. coli and biotoxins in shellfish

schmidt, wiebke; Evers-King, Hayley; Campos, Carlos; Jones, Darren; Miller, Peter; Davidson, Keith; Shutler, Jamie D.

Published in:
AQUACULTURE ENVIRONMENT INTERACTIONS

Publication date:
2018

Publisher rights:
©2018 The Authors

The re-use license for this item is:
CC BY

The Document Version you have downloaded here is:
Peer reviewed version

The final published version is available direct from the publisher website at:
10.3354/aei00265

Link to author version on UHI Research Database

Citation for published version (APA):
schmidt, W., Evers-King, H., Campos, C., Jones, D., Miller, P., Davidson, K., & Shutler, J. D. (2018). A generic approach for the development of short-term predictions of E. coli and biotoxins in shellfish. *AQUACULTURE ENVIRONMENT INTERACTIONS*, *10*, 173-185. https://doi.org/10.3354/aei00265

General rights
Copyright and moral rights for the publications made accessible in the UHI Research Database are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights:

1) Users may download and print one copy of any publication from the UHI Research Database for the purpose of private study or research.
2) You may not further distribute the material or use it for any profit-making activity or commercial gain
3) You may freely distribute the URL identifying the publication in the UHI Research Database

Take down policy
If you believe that this document breaches copyright please contact us at RO@uhi.ac.uk providing details; we will remove access to the work immediately and investigate your claim.

Download date: 24. Jan. 2025

A generic approach for the development of short-term predictions of *E. coli* and biotoxins in shellfish

Wiebke Schmidt [1, 5] *, Hayley L. Evers-King [2], Carlos J. A. Campos [3], Darren B. Jones [1], Peter I. Miller [2], Keith Davidson [4] and Jamie D. Shutler [1]

[1] Centre for Geography, Environment and Society, University of Exeter, Penryn Campus, Penryn, TR10 9FE, United Kingdom

[2] Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, United Kingdom

[3] Centre for Environment, Fisheries & Aquaculture Science (Cefas), Weymouth Laboratory, Barrack Road, Weymouth, DT4 8UB, United Kingdom

[4] Scottish Association for Marine Science, Oban, Argyll, PA37 1QA, United Kingdom

[5] Renewable Energy Group, University of Exeter, Penryn Campus, Penryn, TR10 9FE, United Kingdom

* Corresponding author: Tel.: +44 1326 259291; E-mail: schmidt.wiebke@gmail.com (W. Schmidt)

Abstract: Microbiological contamination or elevated marine biotoxin concentrations within shellfish can result in temporary closure of shellfish aquaculture harvesting, leading to financial loss for the aquaculture business and a potential reduction in consumer confidence in shellfish products. We present a method for predicting short-term variations in shellfish concentrations of *Escherichia coli* (*E. coli)* and biotoxin (okadaic acid, its derivates dinophysistoxins and pectenotoxins). The approach is evaluated for two contrasting shellfish harvesting areas. Through a meta-data analysis and using environmental data (*in situ*, satellite observations and meteorological nowcasts and forecasts) key environmental drivers were identified and used to develop models to predict *E. coli* and biotoxin concentrations within

shellfish. Models were trained and evaluated using independent datasets and the best models were identified based on the model exhibiting the lowest root mean squared error (RMSE). The best biotoxin model was able to provide one-week forecasts with an accuracy of 86 %, a 0% false positive rate and a 0 % false discovery rate (n = 78 observations) when used to predict the closure of shellfish beds due to biotoxin. The best *E. coli* models were used to predict the European hygiene classification of the shellfish beds to an accuracy of 99 % (n = 107 observations) and 98% (n = 63 observations) for a bay (St Austell Bay) and an estuary (Turnaware Bar), respectively. This generic approach enables high accuracy short-term farm-specific forecasts, based on readily accessible environmental data and observations.

Keywords: aquaculture; modelling; forecast; shellfish; water quality

1. Introduction

Aquaculture plays a major role in meeting the demand in seafood production and with the decline of wild fish stocks (FAO 2005), production is expected to grow further (Kobayashi et al. 2015). In Europe, shellfish aquaculture, produced 632,000 tonnes of bivalves in 2014 and this constituted 25 % of the total European marine and coastal aquaculture production (FAO 2016). The sustainability of shellfish farming businesses can be compromised by events of poor water quality due to either microbiological contamination or naturally occurring marine phytoplankton producing biotoxins, both of which can cause temporary closures of shellfish aquaculture harvesting. Frequent or sustained events can often determine the success or failure of an aquaculture business. Furthermore, closure due to poor water quality has been shown to influence the confidence of consumers in shellfish products (Lucas &

Southgate 2012). In the European Union shellfish harvesting areas are required to be classified according to their sanitary quality, on the basis of *Escherichia coli* (*E. coli*) monitoring in shellfish flesh (EU 2015). The classification determines the likely contamination with viral and bacterial pathogens and determines the extent of post-harvest processing required, before shellfish can be placed on the market for human consumption. Similar monitoring occurs in many other parts of the world (Rees et al. 2010). Increased *E. coli* concentrations in coastal or estuarine water bodies are often related to direct water run-off from urban and agricultural land, or due to sewage overflow entering the water body (Defra 2011). As a result, environmental factors such as rainfall, river flow and solar radiation can amplify or modulate the abundance and distribution of *E. coli* in shellfish waters (Kelsey et al. 2004, Coulliette et al. 2009, Kay et al. 2010).

Some naturally occurring phytoplankton can produce a range of marine biotoxins (Hinder et al. 2011) and once filtered by shellfish the biotoxins are retained and hence pose a health risk to humans when the shellfish are consumed (Anderson 2014). Therefore, farmed shellfish are also monitored for biotoxins and once a threshold concentration within the shellfish is exceeded, the harvesting area is closed (EU 2011). One important phytoplankton genus known to produce the biotoxin okadaic acid, its derivates dinophysistoxins and pectenotoxins (OA/DTX/PTX) is *Dinophysis* (Reguera et al. 2012). This group of toxins can cause gastrointestinal illness (Diarrhetic Shellfish Poisoning) in humans even when the density of causative organisms is low (Reguera et al. 2012). Production of toxin in *Dinophysis* cells is influenced by intrinsic and genetic factors as well as by the response of the species cells to environmental conditions, e.g. physical, chemical and biological factors (Reguera et al. 2012, Anderson 2014, Whyte et al. 2014, Gobler et al. 2017).

Whilst elevated *E. coli* or biotoxin concentrations within shellfish can cause temporary closure of the farm and restrictions on sales, the shellfish themselves are unharmed. Furthermore, leaving the shellfish living in the water allows the contaminants to depurate and dissipate naturally (Egmond et al. 2004, Davidson et al. 2011). Once this has occurred the shellfish farm and harvesting is re-opened and all stock can be safely sold and consumed. However, once harvested it is often not economically viable, or dependent upon the farm type even feasible, to return the live shellfish to the farmed waters (Morgan et al. 2009, Berdalet et al. 2015). Therefore, any information to guide farm decisions about when to harvest, when not to harvest and when to sell existing harvested stock at wholesale prices can be beneficial to the farm.

Whilst previous studies related environmental conditions, such as rainfall to *E. coli* or biotoxin concentrations (e.g. Kay et al. (2010)), the use of models to predict concentrations within shellfish has been limited (e.g. Bougeard et al. (2011)). This study identifies key environmental drivers of *E. coli* and biotoxin concentrations within the shellfish in two different shellfish harvesting areas. This information is then used to create short-term (one week) forecast models, with the intention that the forecast could be used to inform and support farm management decisions.

2. Study areas and data collection

2.1. Coastal bay - St Austell Bay, UK

St Austell Bay (Figure 1) is located on the south coast of Cornwall, United Kingdom and covers approximately an area of ~ 21 km^2, with mean depth ranges of 5 m near the shore to 20 m at the mouth (Sherwin & Jonas 1994, Sherwin et al. 1997). The bay is characterised by very small tidal currents (about 0.024 m s^{-1}) and the

99 circulation within the bay is driven by wind and density effects, with a mean
100 eastwards circulation of ~ 0.06 m s^{-1} (Sherwin & Jonas 1994). Thermal stratification
101 occurs during calm wind conditions (less than 5 m s^{-1}) (Sherwin et al. 1997). The
102 river Par and other smaller streams entering the bay have been noted in the past to
103 influence the near-shore dynamics of the bay (Sherwin et al. 1997). Within the bay
104 two shellfish farms cultivate blue mussels (*Mytilus* spp.) on ropes and both are
105 classified as class B under European hygiene classification (EU 2004a).

106 *2.2. Estuary – Turnaware Bar within the lower Fal estuary, UK*

107 The sheltered Fal estuary is situated at the south coast of Cornwall, United Kingdom
108 and covers an intertidal area of 0.46 km^2 (ABPmer & Wallingford 2007). The estuary
109 can be distinguished into two geographical areas (Upper and Lower Fal) and this
110 study focused on Turnaware Bar (Figure 1), which is within the lower Fal estuary. For
111 more than two centuries, native oysters (*Ostrea edulis*), mussels (*Mytilus* spp.) and
112 Pacific oysters (*Magallana gigas*) have been commercially harvested from this
113 estuary. Under European hygiene classification the bivalve molluscs production area
114 at Turnaware Bar is classified as class B (EU 2004a).

115 *2.3. Data collection and processing*

116 *2.3.1. Escherichia coli* (*E. coli*)

117 Following European legislation concentrations of β-glucuronidase-positive
118 *Escherichia coli* (*E. coli*) in shellfish (mussels and oysters) are routinely monitored in
119 both study sides. The *E. coli* data were collated for St Austell Bay (sampling site
120 Ropehaven) and Turnaware Bar, for the period 2008 – 2016 and 2011 – 2016,
121 respectively (table 1). *E. coli* is determined using the enumeration method ISO

16649-3 (ISO 2015) and concentrations are reported as most probable number (MPN) of *E. coli* 100 g^{-1} of flesh. Therefore concentration values at the lower limit of quantification of the MPN method (< 20 MPN 100 g^{-1} and from 2015 onwards < 18 MPN 100 g^{-1}) were adjusted to 19 MPN 100 g^{-1} and 17 MPN 100 g^{-1}, respectively as recommended by the US National Shellfish Sanitation Program (USFDA & ISSC 2013, 2015). The expanded uncertainty of the *E. coli* MPN method has been estimated at 0.66 (of the log_{10} MPN 100 g^{-1} transformed data), which is calculated as twice the measured standard deviation (SD) (Baker-Austin et al. unpubl. data).

2.3.2. Biotoxin

Shellfish samples for St Austell Bay are routinely analysed for the group biotoxin, okadaic acid (OA) and its derivates, dinophysistoxin (DTX) and pectenotoxin (PTX) using liquid chromatography coupled with mass spectrometry as described by Cefas (2011). This biotoxin group is reported in µg OA equivalent (eq.) kg^{-1} shellfish flesh and the Food Standard Agency (FSA) regulatory monitoring data were obtained from Cefas for the time period from 2014 to 2016 (table 1). The minimum reporting limit of the analysis method is stated as 16 µg OA eq. kg^{-1} shellfish flesh (EU 2004b, c) and so values within the dataset below this reporting limit were adjusted to 15 µg OA eq. kg^{-1} shellfish flesh.

2.3.3. Environmental datasets

A metadata analysis identified that the following variables were important for controlling *E. coli* and biotoxin concentrations within shellfish (rainfall, river flow, solar radiation, sea surface temperature (SST), wind speed and direction (Šolić & Krstulović 1992, Catalao Dionisio et al. 2000, Izbicki et al. 2009, Raine et al. 2010, Defra 2011, Reguera et al. 2012, Campos et al. 2013). Therefore, a suite of

146 environmental data was collated to allow the linkages between the environmental
147 conditions and the *E. coli* and biotoxin concentrations in shellfish to be investigated.
148 Daily rainfall measurements (mm day^{-1}) at Luxulyan (for St Austell Bay) and St
149 Mawes (for Turnaware Bar) were obtained from the UK Environment Agency (EA) for
150 2008 - 2016. The rainfall on the day prior to shellfish sampling (lag rainfall) was
151 included as an additional variable. River flow measurements (m s^{-1}) of the major
152 rivers (Par river for St Austell Bay; Fal, Carnon, Kennall, Kenwyn and Allen river for
153 Turnaware Bar, Figure 1) were obtained from the EA for 2008 - 2016. Due to the
154 multiple river sources, the sum of the river flow of the rivers entering the lower Fal
155 estuary was calculated and used for the statistical analysis for Turnaware Bar.
156 Reanalysed meteorological forecast data (10 m U wind, 10 m V wind in m s^{-1} and
157 downwards surface solar radiation in J m^{-2}) were downloaded from the ERA-Interim
158 archive for 2008 to 2015 from the European Centre for Medium-Range Weather
159 Forecast website (http://apps.ecmwf.int/datasets/) (Berrisford et al. 2011). The data
160 for the two study sites were then extracted from the nearest model grid point (St
161 Austell Bay: latitude 49.50; longitude 356.25 and Turnaware Bar: latitude 49.50;
162 longitude 354.75). Daily 10 m U wind and 10 m V wind instantaneous data at 00, 06,
163 12, 18 Coordinated Universal Time (UTC) were used to calculate the daily mean
164 wind speed and direction. Daily surface solar radiation at 00 and 12 UTC were
165 extracted and added to determine total surface solar radiation per day. Satellite SST
166 observations were obtained for both study sites for 2008 - 2016. The Multi-scale
167 Ultra-high Resolution (MUR) dataset from NASA Jet Propulsion Laboratory
168 (http://mur.jpl.nasa.gov/) were used to ensure that data were always available (even
169 in cloudy conditions). These data are a daily, 1 km resolution dataset consisting of a
170 combination of infrared and passive microwave based sensor observations. The

171 mean values between latitude 50.28 and 50.31, and longitude -4.69 and -4.75 were
172 extracted for St Austell Bay, and between latitude 50.12 and 50.13, and longitude
173 -5.02 and -5.13 for Turnaware Bar (Figure 1). Any pixels defined as land within the
174 MUR land mask were excluded from the analysis.
175 For the independent evaluation of models and to predict *E. coli* and biotoxin
176 concentrations, near-real time and now cast environmental data for 2016 for both
177 sites were obtained by the EA, extracted via the MARS ECMWF website
178 (http://apps.ecmwf.int/mars-catalogue/) and satellite data were extracted as
179 described above. The selected dates were chosen to coincide with the official control
180 monitoring results for *E. coli* and biotoxin concentrations tested in shellfish by the
181 FSA and Cefas.

182 **3. Environmental drivers for *E. coli* and biotoxin concentrations**

183 *3.1. Statistical approaches for a suite of models*

184 All data exploration and modelling analyses were conducted using the R statistical
185 software (version 3.3.0 on Mac ODS X 10.10.3, (R Core Team 2016)).

186 The collated datasets (section 2) were all split into two, based on two time periods.
187 The first dataset was used for model development and initial characterisation of the
188 models. The second dataset was used to provide an independent evaluation of
189 model performance. The details of the dataset splitting are provided in table 1.

190 *3.1.1. Generalized Linear Model development*

191 Due to non-Gaussian nature of the response variables (concentrations of
192 OA/DTX/PTX and *E. coli*), they were log transformed prior to use (Kay et al. 2008).
193 Four Generalized Linear Models (GLMs) were developed using the function

194 bestglm() in R. Four different cross-validation methods were used to determine the
195 optimal model (McLeod & Xu 2014). These were i) R's default option deleted-d cross-
196 validation with random subsamples using the delete-d algorithm d=ceil(n(1-1/(log n -
197 1))) and t=10 repetitions, where the parameter d is chosen using the formula
198 recommended by Shao (1997); ii) K-fold cross validation (Hastie et al. 2009); iii)
199 adjusted K-fold cross validation (Davison & Hinkley 1997) and iv) leave-one-out
200 cross-validation approach. To compare the performance of the four GLMs to each
201 other, the root mean square error (RMSE) was calculated. The output from R
202 provides a list of the significant explanatory (predictor) variables.

203 *3.1.2. Averaged GLM development*

204 The model-averaging GLM based on the information-theoretic approach by Burnham
205 and Anderson (2002) was calculated. Firstly, a 'global model' was built, using the
206 glm() function with all environmental factors as dependent variables. In order to
207 directly compare the global model's independent variables, the model was then
208 standardised (mean = 0 and SD=0.5). The dredge() function within the R package
209 'MuMIn' (Bartoń 2015) enabled the generation and comparison of all possible models
210 and the best performing models, used to construct the model-averaged result, were
211 selected using delta Akaike Information Criterion < 2 as the decision metric. Using
212 the variance inflation factors, the averaged model was tested for potential co-linearity
213 between covariates. Where correlated covariates existed, only one variable was
214 retained (Zuur et al. 2010).

215 *3.1.3. Generalized Additive Model development*

216 The gam() function in the R package 'mgcv' (Wood 2006) was used to develop the
217 Generalized Additive Model (GAM). Initially all explanatory variables were set as

smooth terms and knots were set manually to four as the number of observations were < 100 (Thomas et al. 2015). The estimated degrees of freedom of the smoothed term was used to check the GAM model assumptions and the model terms were then adjusted to be linear terms in cases where estimated degrees of freedom equalled 1 (Thomas et al. 2015).

3.2. Independent evaluation of all models

An independent comparison is recommended for model evaluation (Verbyla & Litvaitis 1989, Fielding & Bell 1997). Therefore, the accuracy of each model (the best GLM based on its lowest cross-validation RMSE, the best averaged GLM, and the GAM) was evaluated using their respective evaluation datasets (table 1) by calculating the RMSE and bias between the predicted and the actual observed *E. coli* and biotoxin concentrations. The limits and uncertainties in their analytical detection have been described in section 2.3.1 and 2.3.2. The best model was identified as the model that produced the lowest RMSE. A confusion matrix analysis was then used to understand how the model accuracy (RMSE and bias) translates to the ability to answer specific management questions. This analysis allows the determination of the accuracy, precision, false positive rate and false discovery rate, which could have significant impacts on an aquaculture business (Stehman 1997). The confusion matrix was used to determine i) the best *E. coli* model's capability to predict the EU shellfish bed classification (which is class B for both sites, see section 2.1. and 2.2.) and ii) the best biotoxin model's ability to predict the closure and re-opening of shellfish harvesting based on the regulatory threshold of 160 µg OA eq. kg^{-1} shellfish flesh.

4. Results and discussion

242 *4.1. Models for E. coli concentrations in St Austell Bay and Turnaware Bar*

243 The observed and modelled *E. coli* concentrations can be seen in figure 2. Both
244 study sites displayed low observed *E. coli* concentrations over the year 2016, with
245 mean of 1.54 $\log_{10}$ *E. coli* 100 g^{-1} of flesh for St Austell Bay and of 2.23 $\log_{10}$ *E. coli*
246 100 g^{-1} of flesh for Turnaware Bar (table 2). Observed *E. coli* concentrations were
247 generally below class A limit in St Austell Bay (class A requirements are 80% of
248 samples must not exceed 230 *E. coli* 100 g^{-1} of flesh and the remaining 20% of
249 samples must not exceed 700 *E. coli* 100 g^{-1} of flesh). Only one result during the
250 winter of 2016 exceeds the class A limit of *E. coli* 100 g^{-1} of flesh (figure 2; on 2nd Feb
251 2016). A slightly larger variability in observed *E. coli* concentrations is apparent at
252 Turnaware Bar (figure 2).

253 All three model outputs (GLM, average GLM and GAM) fitted the observed *E. coli*
254 concentrations reasonably well. Statistical measures for observed (mean, standard
255 deviation) and modelled (RMSE, bias) *E. coli* concentrations in St Austell Bay and
256 Turnaware Bar are listed in table 2.
257 For both study areas, the GLM model showed the lowest RMSE (St Austell Bay: 0.48
258 $\log_{10}$ *E. coli* 100 g^{-1} shellfish flesh; Turnaware Bar: 0.68 $\log_{10}$ *E. coli* 100 g^{-1} shellfish
259 flesh (table 2)).

260

261 *4.2. Environmental drivers for E. coli concentrations in St Austell Bay and Turnaware*
262 *Bar*

263 Significant explanatory variables used for predicting *E. coli* concentrations for each
264 model and both study sites are summarised in table 3. The majority of models

identified rainfall and/or lag rainfall as one of the key explanatory variables for predicting *E. coli* concentrations (table 3).

Rainfall has been previously described as a common environmental factor associated with controlling *E. coli* concentrations in receiving waters (Kelsey et al. 2004, Coulliette et al. 2009, Kay et al. 2010, Campos et al. 2013). However, the degree of response of *E. coli* concentrations to rainfall can vary considerably between sampling points in a given sampling area (Defra 2011), which can be attributed to differences in land use (e.g. agriculture run-off versus urban sewage) and soil conditions surrounding the sampling site (Kay et al. 2010, Defra 2011). In all models for Turnaware Bar, river flow was identified as significant and this catchment has been previously described as highly responsive to rainfall (Cefas 2012). River flow may also provide a proxy for other events contributing to faecal contamination (e.g. combined sewer overflows, application of slurry to fields, etc.) and therefore it is not possible to explain the direct cause for the varying *E. coli* concentrations. Turnaware Bar, the study point within the lower Fal estuary, is exposed to a lower 'flushing' time due to tides and thus resulting in a potentially higher retention time of microbial contamination from nearby sources (Uncles et al. 2002, Langston et al. 2006).

For St Austell Bay, solar radiation was identified as an important explanatory variable for predicting *E. coli* concentrations (table 3). All models displayed the solar radiation term as negative and this is consistent with previous work by Campos et al. (2013) that showed that solar radiation can influence the die-off of *E. coli* in seawater. Additionally, Šolić and Krstulović (1992) suggested that solar radiation may be more important than seawater temperature in altering the number of the faecal indicator. However, for Turnaware Bar all three models identified SST as a key explanatory

290 variable. Although *E. coli* optimal growth temperature lies at around 37 °C, it has
291 been previously shown that the optimal temperature for survival is not necessarily the
292 same as that needed for optimal growth (Rozen & Belkin 2001) as *E. coli* can survive
293 and stabilise at lower temperatures within estuaries (Rhodes & Kator 1988).

294 *4.3 Ability to predict shellfish bed classification*

295 Using easily attainable environmental measurements to predict the shellfish *E. coli*
296 concentrations would enable local authorities to estimate the likely variation in *E. coli*
297 within a shellfish area, in the absence of regular monitoring e.g. if shellfish beds are
298 currently not used, but are available for lease. Therefore, the confusion matrix
299 allowed the overall performance (termed accuracy) and precision (if predicted class
300 B, how often was the model correct) to be calculated for both sites, using the best *E.*
301 *coli* model (the model with the lowest RMSE).
302 The overall accuracy for St Austell Bay was 99 % determined using data from 2008
303 to 2016 (n = 107); 100% for data from 2016 only (n = 13)). Additionally, the precision
304 was 100 % for data from 2008 to 2016 (n = 107) and 100% for data from 2016 (n=
305 13)).
306 Similarly, results for Turnaware Bar showed a high accuracy of 98% for data from
307 2011 to 2016 (n = 63) and 100 % using data from 2016 only (n = 10)), with a
308 precision of 100 % (2011 - 2016, n=63) and 100 % for just 2016 (n = 10).
309 It is noted, that despite the use of long-term datasets (St Austell Bay: 9 years;
310 Turnaware Bar: 5 years) relatively limited variations in *E. coli* are apparent.

311 *4.3. Models for biotoxin concentrations in St Austell Bay*

312 The observed and modelled biotoxin concentrations can be seen in figure 3. During
313 2016 the observed biotoxin concentrations ranged from below the reporting limit up

to 2013 µg OA eq. kg^{-1} shellfish flesh. Between mid July and end of October, the observed biotoxin concentrations were above the maximum permitted level of 160 µg OA eq. kg^{-1} shellfish flesh (EU 2011), indicated as the red horizontal line within figure 3, and therefore the shellfish farm was closed for this period.

Table 4 lists statistical measures for observed (mean, standard deviation) and modelled (RMSE, bias) biotoxin concentrations. The averaged GLM produced the lowest RMSE (582.3 µg OA eq. kg^{-1} shellfish flesh), whereas the GAM showed the lowest bias of -71.01 µg OA eq. kg^{-1} shellfish flesh.

All three models were consistent to capture the increase and reduction of biotoxin concentrations within the shellfish (i.e. the points at which concentrations increase and decrease in relation to the legal limit, figure 3). However, all models showed differences in their response when modelling the period between the 'accumulation' and 'depuration' phases. The depuration of biotoxin concentrations within the shellfish depends upon the species and the physiological conditions of the shellfish e.g. fat content, respiration state, growth, food availability etc. (Hallegraeff 1995). No physiological parameters, describing the depuration of the group biotoxin OA/DTXs/PTXs in blue mussels, were found in the published literature and thus it was not possible to identify specific depuration parameters for the model development. Future studies describing the toxicological profile of OA/DTXs/PTXs in blue mussels would therefore be beneficial.

4.5. Modelling biotoxin accumulation and depuration phases in St Austell Bay

To investigate the difference between accumulation and depuration of the biotoxin, the models were retrained using either only accumulation data (n = 27 observations from 2014 & 2015) or only depuration data (n = 27 observations from 2014 & 2015).

The ‘accumulation’ phase was defined as the beginning of the year to the highest observed biotoxin concentrations, while the ‘depuration’ phase was defined as from the highest observed concentration to the end of the year.

The accumulation models showed higher RMSE (GLM = 1745.73; averaged GLM = 858.33; GAM = 1151.99) than the models trained on the full dataset (section 4.2) when evaluated using the full evaluation dataset (table 4). Similarly, the depuration models also showed a higher RMSE (RMSE for GLM = 868.39; averaged GLM = 110.29; GAM = 17,965.03) and both models were unable to capture the appropriate response of the biotoxin concentration. This result supports the usage of the model trained on the full dataset in order to forecast biotoxin concentrations and highlights that different environmental and physiological factors are controlling the accumulation and depuration phase.

4.6. Ability to predict farm closure and reopening due to biotoxins

The ability for a farm to identify the potential closure due to accumulation of biotoxins in mussels is advantageous for supporting decisions on harvesting and sales (e.g. when to sell existing stock at wholesale prices). Therefore, the performance of the model with the lowest RMSE (averaged GLM) trained on the full dataset (table 1 & 4) to accurately forecast the closure of a shellfish farm was evaluated. The overall performance (accuracy), false positive rate (false prediction of an open farm, when it should actually be closed) and false discovery rate (error in predicting the opening and closure of the farm) were calculated.

The nowcast predictions (n = 79 observations for 2014 - 2016) of the confusion matrix produced an accuracy of 84 % with a false positive rate of 2 % and false discovery rate of 4 %. Results using data from just 2016 showed a higher accuracy

(92 %) with a false positive rate of 6 % and a false discovery rate of 13 % (n = 25 observations).

For a one-week forecast (n = 78 observations for 2014 - 2016) the accuracy was 86 % with a false positive rate of 0 % and false discovery rate of 0 %; in comparison results for the one-week forecast using just 2016 data showed an accuracy of 96 %, with a false positive rate of 0 % and false discovery rate of 0 % (n = 24 observations).

4.5. Environmental drivers for biotoxin concentrations in St Austell Bay

Environmental variables such as SST, solar radiation and wind speed were identified as key drivers for the different biotoxin models (table 5). Additionally, other environmental variables, including lag rainfall and wind direction contributed to some of the models.

Previously, it has been shown that the ambient temperature influences the filtration rate and pumping activity in blue mussels (Jørgensen et al. 1990). The distribution and occurrence of the dinoflagellate genus *Dinophysis* spp. in the water column in temperate regions can be related to stratification of water column (Raine & McMahon 1998). Seawater temperature, salinity and dissolved oxygen concentrations were recorded in close proximity of the shellfish farm from autumn 2015 until summer 2016 using a moored buoy (Schmidt et al. 2018). Figure 4 shows that the SST just outside of the farm increased by around of 2°C at 1.1 m depth from 2nd to 8th July 2016, indicating that thermal stratification could have taken place within the farm in July (e.g. as the farm itself is likely to accelerate stratification by dampening vertical mixing). In addition, step increased of dissolved oxygen concentrations (green line within Figure 4) from 5th to 7th July 2016 indicates an increase of biological activity close to the farm site. This period would coincide with the closure of the farm on the

386 5th July 2016 due to high biotoxin concentrations (indicated as red vertical line within
387 Figure 4). These observations support the hypothesis (Farrell et al. 2012) that
388 increased abundance of *Dinophysis* spp. in the water column can be related to
389 thermal stratification. However, a future study would need to confirm this by placing
390 the instruments within the mussel farm ropes.

391 A further important factor influencing the distribution of *Dinophysis* spp. can be its
392 transport by ocean currents from offshore locations into coastal bays (Escalera et al.
393 2010). Raine et al. (2010) suggested that wind driven water exchange between the
394 south coast of Ireland and the continental shelf is responsible for an influx of
395 *Dinophysis* spp. into a shellfish harvesting area. Large scale wind driven advective
396 processes were also proposed by Whyte et al. (2014) to explain large coastal blooms
397 of *Dinophysis* ssp. in the Shetland Islands. All three biotoxin models identified wind
398 speed as a key environmental factor for the shellfish biotoxin concentration.
399 However, no significant correlation was found between increased biotoxin
400 concentrations and wind speed or direction. As a third key environmental driver, solar
401 radiation, was identified as an important factor by the models. Toxicological profiles
402 of DSP toxin reported that its synthesis requires light and is coupled to the cell
403 division cycle (Pan et al. 1999). This again supports the hypothesis that stratification
404 is important, as stratification will imply reduced turbidity and therefore an improved
405 light field within the water column.
406 This study chose to focus on OA/DTXs/PTXs toxin as these were a significant issue
407 for the shellfish farmer in St Austell Bay. Mussels were the focus shellfish as these
408 dominate the global bivalve production and they can be used as indicator species for
409 monitoring other bivalve production. Clearly the generic nature of the approach
410 means that it could be applied to other toxin groups (e.g. PSP, ASP, AZA) if sufficient

411 data to train and evaluate the models exists. However for our study site and temporal
412 period (2014 - 2016) neither PSP nor ASP have exceeded the maximum permitted
413 limit and there was only three instances of AZA exceeding the permitted limit.
414 Therefore, we were unable to test the applicability to other toxin groups.

415 **5. Conclusion**

416 This study has demonstrated that a generic approach and a suite of readily available
417 environmental data (*in situ*, satellite observations and meteorological nowcasts and
418 forecasts) can be used to successfully model the *E. coli* concentrations within
419 shellfish living in an estuarine and coastal shellfish site, despite the sites exhibiting
420 contrasting hydrography. The same methodology has then been shown to
421 successfully model the biotoxin concentrations within shellfish living in a coastal
422 shellfish site. The accuracy of models have been evaluated and characterised using
423 data independent to the training data, and indirectly verified using measurements
424 from an *in situ* monitoring buoy located close to one of the sites. The inputs to these
425 models were identified by a metadata analysis as being important for influencing the
426 *E. coli* and biotoxin concentrations within shellfish and the parameters identified as
427 significant by each model analysis are consistent with previous studies.

428 Whilst the models were less able to accurately predict the absolute values of the
429 concentrations within the shellfish, the modelled variations in *E. coli* and biotoxin
430 concentrations have been demonstrated to be reliable for supporting farm decision
431 marking. Accurate classifications of a change in shellfish bed class and forecasting
432 closure due to biotoxin accumulation were both possible. The biotoxin models can be
433 used to provide a one-week forecast. Such an early warning can provide and support
434 the shellfish farmers in their management by guiding harvesting decisions and pricing

435 strategies. Using these forecasting approaches is also likely to help increase
436 customer confidence in shellfish products, and give farmers increased confidence
437 when selling their product. After initial model development for a harvesting site, the
438 shellfish farmer could use this approach and the subsequent models to forecast
439 changes within their farm shellfish stock, towards supporting farm management
440 decisions. However, it is likely that a yearly update of the model parameters would be
441 needed to account for temporal changes in the catchment, regional climate, changes
442 to farm composition and size, and influences that the farm itself will have upon the
443 local ecosystem.

444 **Acknowledgment:** The authors wish to thank the shellfish farmers (Marina Rawle
445 and Gary Rawle) for their support during the project '*ShellEye*'. This work was carried
446 out as part of the UK Biotechnology and Biological Science Research Council
447 (BBRSC) and UK National Environmental Research Council (NERC) funded
448 '*ShellEye'* project (BB/M026698/1). Due to the confidential nature of some of the
449 research materials supporting this publication, not all of the data can be made
450 accessible to other researchers. Please contact W. Schmidt for more information. For
451 weekly provision of *in situ* river flow and rainfall data, the authors would like to thank
452 the UK Environment Agency (Paul Blacker and Suzanne Long).

453 **References**

454 ABPmer, Wallingford H (2007) The Estuary-Guide: A website based overview of how
455 to identify and predict morphological change within estuaries. Website prepared
456 for the joint Defra/EA Flood and Coastal Erosion Risk Management R&D
457 Programme

458 Anderson D (2014) HABs in a changing world: a perspective on harmful algal
459 blooms, their impacts, and research and management in a dynamic era of
460 climactic and environmental change. Harmful algae 2012 : proceedings of the 15th
461 International Conference on Harmful Algae : October 29 - November 2, 2012,
462 CECO, Changwon, Gyeongnam, Korea / editors, Hak Gyoon Kim, Beatriz
463 Reguera, Gustaaf M Hallegraeff, Chang Kyu Lee, M 2012:3-17

464 Bartoń K (2015) Multi-Model Inference. In. R package version 1.15.6

465 Berdalet E, Fleming LE, Gowen R, Davidson K, Hess P, Backer LC, Moore SK,
466 Hoagland P, Enevoldsen H (2015) Marine harmful algal blooms, human health and
467 wellbeing: challenges and opportunities in the 21st century. Journal of the Marine
468 Biological Association of the United Kingdom 96:61-91

469 Berrisford P, Dee DP, Poli P, Brugge R, Fielding K, Fuente M, Kållberg PW,
470 Kobayashi S, Uppala S, Simmons A (2011) The ERA-Interim archive. Version 2.0.
471 In: ERA Report Series, ECMWF, Reading, United Kingdom

472 Bougeard M, Le Saux J-C, Pérenne N, Baffaut C, Robin M, Pommepuy M (2011)
473 Modeling of Escherichia coli Fluxes on a Catchment and the Impact on Coastal
474 Water and Shellfish Quality1. --> 47:350-366

475 Burnham KP, Anderson DR (2002) Model Selection and Multimodel Inference: A
476 Practical Information-Theoretic Approach, Vol. Springer,, Secaucus, NJ, USA

477 Campos CJA, Kershaw SR, Lee RJ (2013) Environmental Influences on Faecal
478 Indicator Organisms in Coastal Waters and Their Accumulation in Bivalve
479 Shellfish. Estuaries and Coasts 36:834-853

480 Catalao Dionisio LP, Rheinheimer G, Borrego JJ (2000) Microbiological Pollution of
481 Ria Formosa (South of Portugal). Marine Pollution Bulletin 40:186-193

482 Cefas (2011) In-house validation of an LC-MS/MS method for the determination of
483 lipophilic toxins in shellfish species typically tested in the United Kingdom. In:

484 Cefas (2012) Sanitary Survey of the lower Fal Estuary and Percuil River (Cornwall).
485 In: Cefas report on behalf of the Food Standards Agency, to demostrate
486 compliance with requirements for classification of bivalve mollusc production areas
487 in England and Wales under of Regulation (EC) no 854/2004

488 Coulliette AD, Money ES, Serre ML, Noble RT (2009) Space/Time Analysis of Fecal
489 Pollution and Rainfall in an Eastern North Carolina Estuary. Environmental
490 Science & Technology 43:3728-3735

491 Davidson K, Tett P, Gowen RJ (2011) Harmful Algal Blooms. In: Harrison, Hester
492 (eds) Marine Pollution & Human Health. RSC publishing

493 Davison AC, Hinkley DV (1997) Bootstrap Methods and their Application, Vol.
494 Cambridge University Press

495 Defra (2011) Factors affecting the microbiological quality of shellfish. In: EVID4
496 Evidence Project Final Report, Book Rev 06/11. Defra

497 Egmond HP, van Apeldoorn ME, Speijers GJA (2004) Marine Biotoxins. In: Issue 80
498 Food and Nutrition Paper. Food and Agriculture Organization of the United
499 Nations, Rome

Escalera L, Reguera B, Moita T, Pazos Y, Cerejo M, Cabanas JM, Ruiz-Villarreal M (2010) Bloom dynamics of Dinophysis acuta in an upwelling system: In situ growth versus transport. Harmful Algae 9:312-322

EU (2004a) Corrigendum to Regulation (EC) No 854/2004 of the European Parliament and of the Council of 29 April 2004 laying down specific rules for the organisation of official controls on products of animal origin intended for human consumption. In. Official Journal of the European Union

EU (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. In: Union OJotE (ed) L139 Section VII, Chapter V 60-61

EU (2004c) Regulation (EC) No 882/2004 of the European Parliament and of the Council of 29 April 2004 on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules. In: Official Journal of the European Union OJotEU (ed) L165 Annex III, p 120

EU (2011) Commission Regulation (EU) No 15/2011 of 10 January 2011 amending Regulation (EC) No 2074/2005 as regards recognised testing methods for detecting marine biotoxins in live bivalve molluscs Text with EEA relevance. Official Journal of the European Communities

EU (2015) Commission Regulation (EU) 2015/2285 of 8 December 2015 amending Annex II to Regulation (EC) No 854/2004 of the European Parliament and of the Council laying down specific rules for the organisation of official controls on products of animal origin intended for human consumption as regards certain requirements for live bivalve molluscs, echinoderms, tunicates and marine

523 gastropods and Annex I to Regulation (EC) No 2073/2005 on microbiological
524 criteria for foodstuffs. Official Journal of the European Union 58

525 FAO (2005) Review of the state of world marine fishery resources. FAO FISHERIES
526 TECHNICAL PAPER 457

527 FAO (2016) The State of World Fisheries and Aquaculture 2016. Contributing to food
528 security and nutrition for all. In, Rome, Italy

529 Farrell H, Gentien P, Fernand L, Lunven M, Reguera B, González-Gil S, Raine R
530 (2012) Scales characterising a high density thin layer of *Dinophysis acuta*
531 Ehrenberg and its transport within a coastal jet. Harmful Algae 15:36-46

532 Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction
533 errors in conservation presence/absence models. Environmental Conservation
534 24:38-49

535 Gobler CJ, Doherty OM, Hattenrath-Lehmann TK, Griffith AW, Kang Y, Litaker RW
536 (2017) Ocean warming since 1982 has expanded the niche of toxic algal blooms in
537 the North Atlantic and North Pacific oceans. Proceedings of the National Academy
538 of Sciences 114:4975-4980

539 Hallegraeff GM (1995) 1. Harmful algal blooms: a global overview. In: Hallegraeff
540 GM, Anderson DM, Cembella AD (eds) Manual on Harmful Marine Microalgae.
541 UNESCO

542 Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning, Vol.
543 2nd Ed. Springer-Verlag

544 Hinder SL, Hays GC, Brooks CJ, Davies AP, Edwards M, Walne AW, Gravenor MB
545 (2011) Toxic marine microalgae and shellfish poisoning in the British isles: history,
546 review of epidemiology, and future implications. Environmental Health 10:54

547 ISO (2015) Microbiology of Food and Animal Feeding Stuffs – Horizontal method for
548 the Enumeration of β-glucuronidase positive *Escherichia coli* - Part 3: Most
549 Probable Number technique using 5-bromo-4-cloro-3-indolyl-β-D-glucuronide Acid.
550 In: Standardization IOf (ed) ISO TS 16649–3:2015, Geneva

551 Izbicki JA, Swarenski PW, Reich CD, Rollins C, Holden PA (2009) Sources of Fecal
552 Indicator Bacteria in Urban Streams and Ocean Beaches, Santa Barbara,
553 California. Annals of Environmental Science 3:139-178

554 Jørgensen CB, Larsen PS, Riisgård HU (1990) Effects of temperature on the mussel
555 pump. Marine Ecology Progress Series 64:89-97

556 Kay D, Anthony S, Crowther J, Chambers BJ, Nicholson FA, Chadwick D, Stapleton
557 CM, Wyer MD (2010) Microbial water pollution: A screening tool for initial
558 catchment-scale assessment and source apportionment. Science of The Total
559 Environment 408:5649-5656

560 Kay D, Crowther J, Stapleton CM, Wyer MD, Fewtrell L, Anthony S, Bradford M,
561 Edwards A, Francis CA, Hopkins M, Kay C, McDonald AT, Watkins J, Wilkinson J
562 (2008) Faecal indicator organism concentrations and catchment export coefficients
563 in the UK. Water Research 42:2649-2661

564 Kelsey H, Porter DE, Scott G, Neet M, White D (2004) Using geographic information
565 systems and regression analysis to evaluate relationships between land use and

566 fecal coliform bacterial pollution. Journal of Experimental Marine Biology and
567 Ecology 298:197-209

568 Kobayashi M, Msangi S, Batka M, Vannuccini S, Dey MM, Anderson JL (2015) Fish
569 to 2030: The Role and Opportunity for Aquaculture. Aquaculture Economics &
570 Management 19:282-300

571 Langston WJ, Chesman BS, Burt GR, Taylor M, Covey R, Cunningham N, Jonas P,
572 Hawkins SJ (2006) Characterisation of the European Marine Sites in South West
573 England: The Fal and Helford Candidate Special Area of Conservation (cSAC).
574 Hydrobiologia 555:321-333

575 Lucas JS, Southgate PC (2012) Aquaculture: Farming Aquatic Animals and Plants,
576 Vol. Wiley-Blackwell

577 McLeod AI, Xu C (2014) bestglm: Best Subset GLM. In. R package version 0.34

578 Morgan KL, Larkin SL, Adams CM (2009) Firm-level economic effects of HABS: A
579 tool for business loss assessment. --> 8:212-218

580 Pan Y, Cembella AD, Quilliam MA (1999) Cell cycle and toxin production in the
581 benthic dinoflagellate Prorocentrum lima. Marine Biology 134:541-549

582 R Core Team (2016) R: A language and environment for statistical computing. In:
583 Computing RFfS (ed), Vienna, Austria

584 Raine R, McDermott G, Silke J, Lyons K, Nolan G, Cusack C (2010) A simple short
585 range model for the prediction of harmful algal events in the bays of southwestern
586 Ireland. Journal of Marine Systems 83:150-157

587 Raine R, McMahon T (1998) Physical dynamics on the continental shelf off
588 southwestern Ireland and their influence on coastal phytoplankton blooms.
589 Continental Shelf Research 18:883-914

590 Rees G, Pond K, Kay D, Bartram J, Santo Domingo J (2010) Safe Management of
591 Shellfish and Harvest Waters. In: Organization WH (ed). IWA Publishing, London,
592 New York

593 Reguera B, Velo-Suárez L, Raine R, Park MG (2012) Harmful *Dinophysis* species: A
594 review. Harmful Algae 14:87-106

595 Rhodes MW, Kator H (1988) Survival of *Escherichia coli* and *Salmonella* spp. in
596 estuarine environments. Applied and Environmental Microbiology 54:2902-2907

597 Rozen Y, Belkin S (2001) Survival of enteric bacteria in seawater. FEMS
598 Microbiology Reviews 25:513-529

599 Schmidt W, Raymond D, Parish D, Ashton IGC, Miller PI, Campos CJA, Shutler JD
600 (2018) Design and operation of a low-cost and compact autonomous buoy system
601 for use in coastal aquaculture and water quality monitoring. Aquaculture
602 Engineering 80C:28-36

603 Shao J (1997) An Asymptotic Theory for Linear Model Selection. Statistica Sinica
604 7:221-264

605 Sherwin TJ, Jonas PJC (1994) The impact of ambient stratification on marine outfall
606 studies in British waters. Marine Pollution Bulletin 28:527-533

607 Sherwin TJ, Jonas PJC, Sharp C (1997) Subduction and dispersion of a buoyant
608 effluent plume in a stratified English Bay. Marine Pollution Bulletin 34:827-839

609 Šolić M, Krstulović N (1992) Separate and combined effects of solar radiation,
610 temperature, salinity, and pH on the survival of faecal coliforms in seawater.
611 Marine Pollution Bulletin 24:411-416

612 Stehman SV (1997) Selecting and interpreting measures of thematic classification
613 accuracy. Remote Sensing of Environment 62:77-89

614 Thomas R, Lello J, Medeiros R, SPollard A, Seward A, Smith J, Vafidis J, Vaughn I
615 (2015) Data Analysis with R Statistical Software, Vol. Eco-explore

616 Uncles RJ, Stephens JA, Smith RE (2002) The dependence of estuarine turbidity on
617 tidal intrusion length, tidal range and residence time. Continental Shelf Research
618 22:1835-1856

619 USFDA, ISSC (2013) National shellfish sanitation program - Guide for the control of
620 molluscan shellfish. 2011 Revision. In: US Department of Health and Human
621 Services PHS, Food and Drug Administration (ed)

622 USFDA, ISSC (2015) National shellfish sanitation program - Guide for the control of
623 molluscan shellfish. 2015 Revision. In: US Department of Health and Human
624 Services PHS, Food and Drug Administration (ed)

625 Verbyla DL, Litvaitis JA (1989) Resampling methods for evaluating classification
626 accuracy of wildlife habitat models. Environmental Management 13:783-787

627 Whyte C, Swan S, Davidson K (2014) Changing wind patterns linked to unusually
628 high Dinophysis blooms around the Shetland Islands, Scotland. Harmful Algae
629 39:365-373

630 Wood SN (2006) Generalized Additive Models: An Introduction with R, Vol. Chapman
631 and Hall/CRC

632 Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid
633 common statistical problems. Methods in Ecology and Evolution 1:3-14

634 **Figures and Tables**

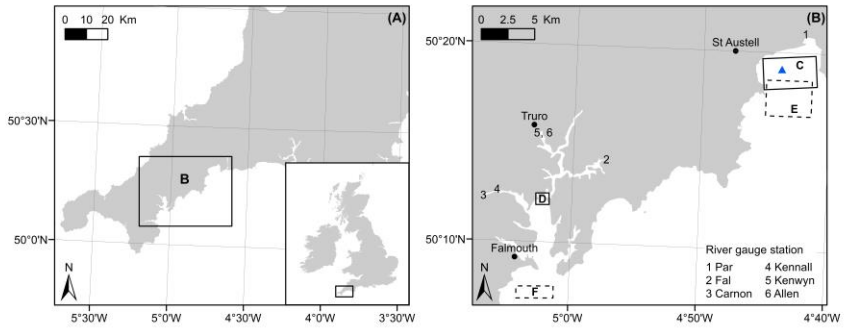

Figure 1. Location of St Austell Bay (C), Turnaware Bar (D), river gauge stations, in situ buoy within St Austell Bay (blue triangle) and data extraction areas for satellite sea surface temperature (dotted boxes, St Austell Bay: E; Turnaware Bar: F).

635

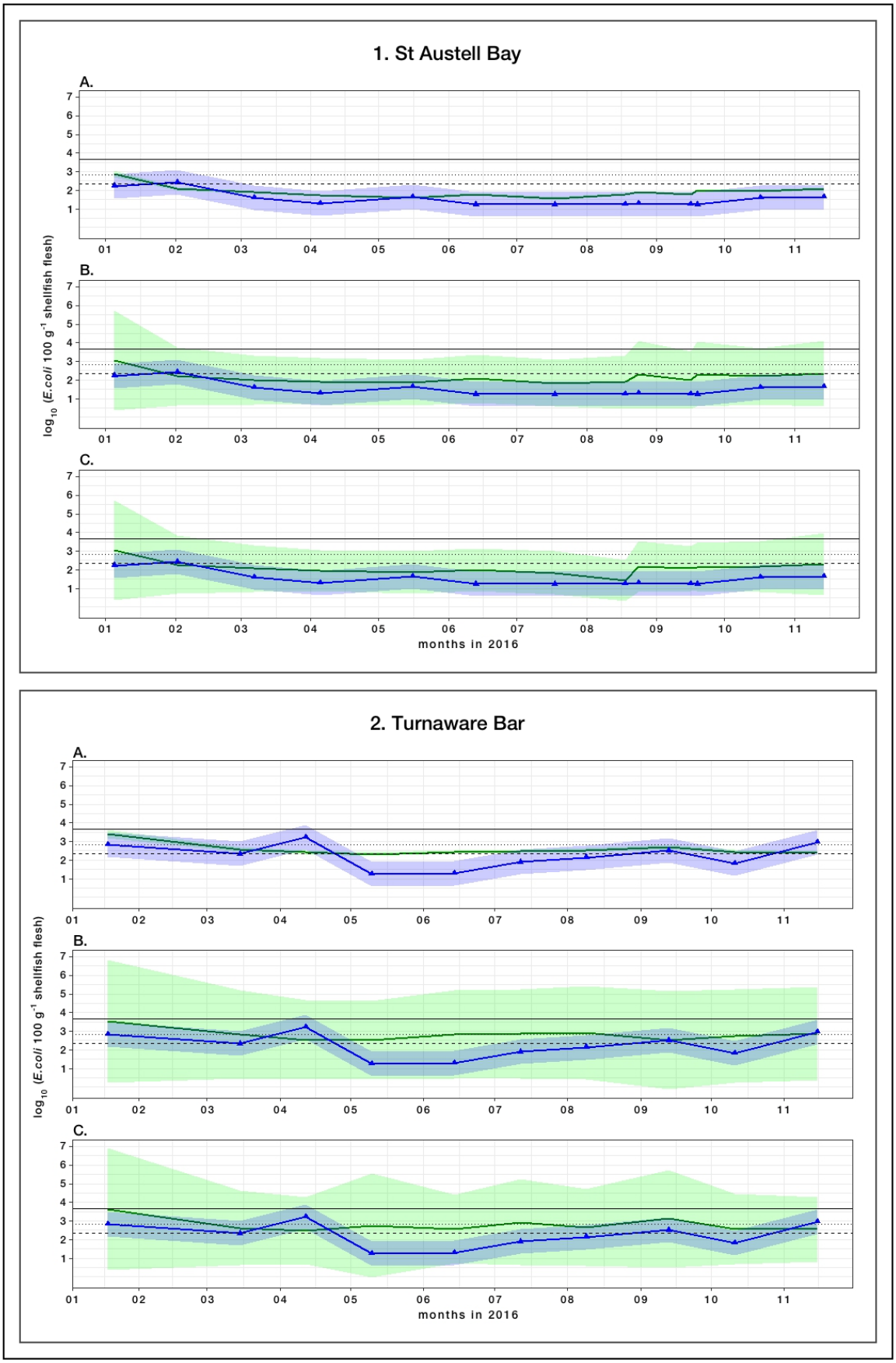
1. St Austell Bay
A.
B.
C.
log10 (E.coli/ 100 g-1 shellfish flesh)
months in 2016
2. Turnaware Bar
A.
B.
C.
log10 (E.coli/ 100 g-1 shellfish flesh)
months in 2016

Figure 2.

Modelled and observed concentrations of *E. coli* in 1. St Austell Bay and 2. Turnaware Bar for the year 2016. All *E. coli* concentrations are displayed as $\log_{10}$ *E. coli* 100 g^{-1} shellfish flesh. Modelled concentrations (green line ± model standard error in shaded green) were obtained by A) GLM, B) averaged GLM and C) GAM. Observed *E. coli* concentrations are shown with a line with triangles (blue line ± uncertainty of 0.66 (2 x SD) in shaded blue). The dashed and dotted black line indicates the limit of class A classification for shellfish bed (dashed = $\log_{10}$ 230 *E. coli* 100 g^{-1} shellfish flesh for 80% samples; dotted = $\log_{10}$ 700 *E. coli* 100 g^{-1} shellfish flesh for 20% samples), whereas the solid black line represents the upper limit of class B classification ($\log_{10}$ 4,600 *E. coli* 100 g^{-1} shellfish flesh).

636

637

638

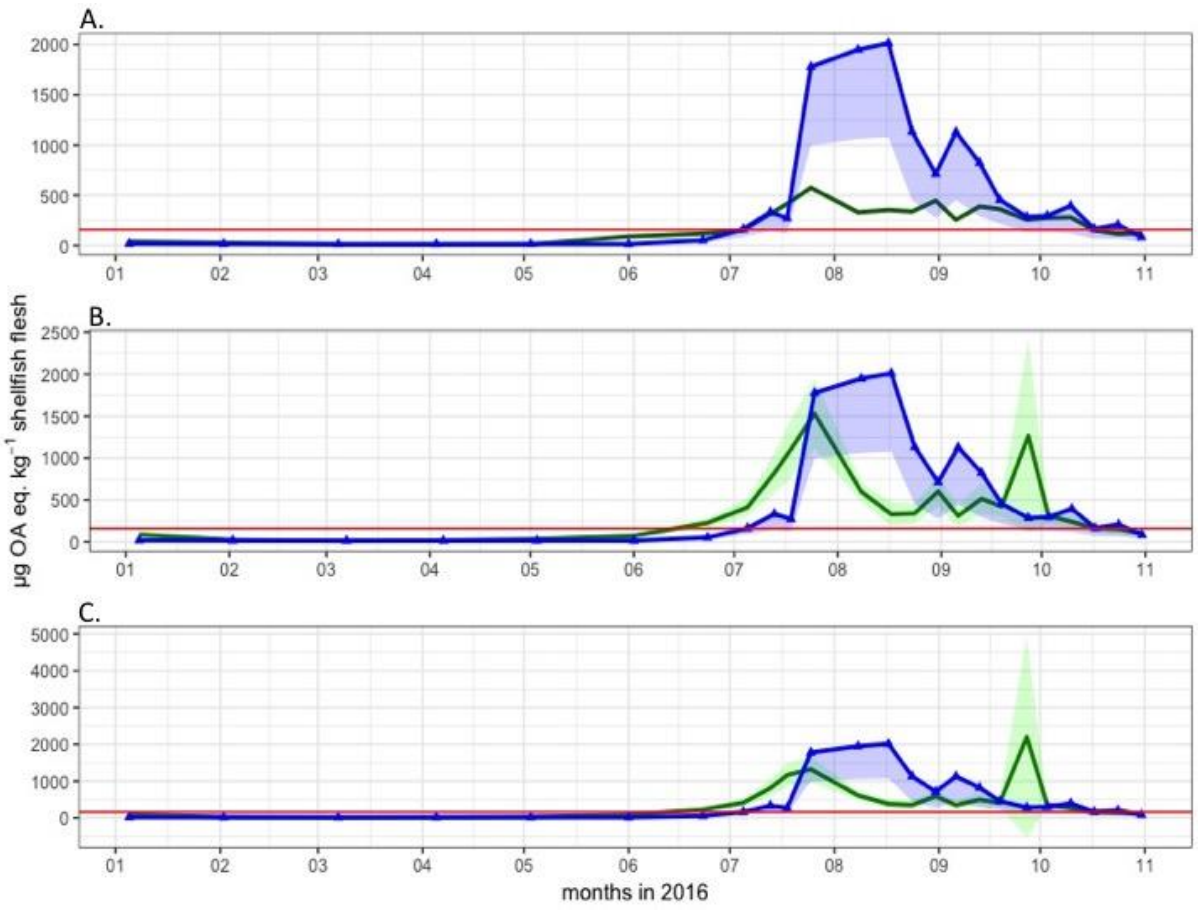

Figure 3. Modelled and observed biotoxin concentrations St Austell Bay for the year 2016. All biotoxin concentrations are displayed as µg OA eq. kg^{-1} shellfish flesh. Modelled concentrations (green line ± model standard error) were obtained by A) GLM, B) averaged GLM and C) GAM. Observed biotoxin concentrations are shown in line with triangle (blue line = high measurement uncertainty (MU) values, blue shade = low MU values). The red line indicates the maximum permitted level of 160 µg OA eq. kg^{-1} shellfish flesh at which point the farm is closed.

639

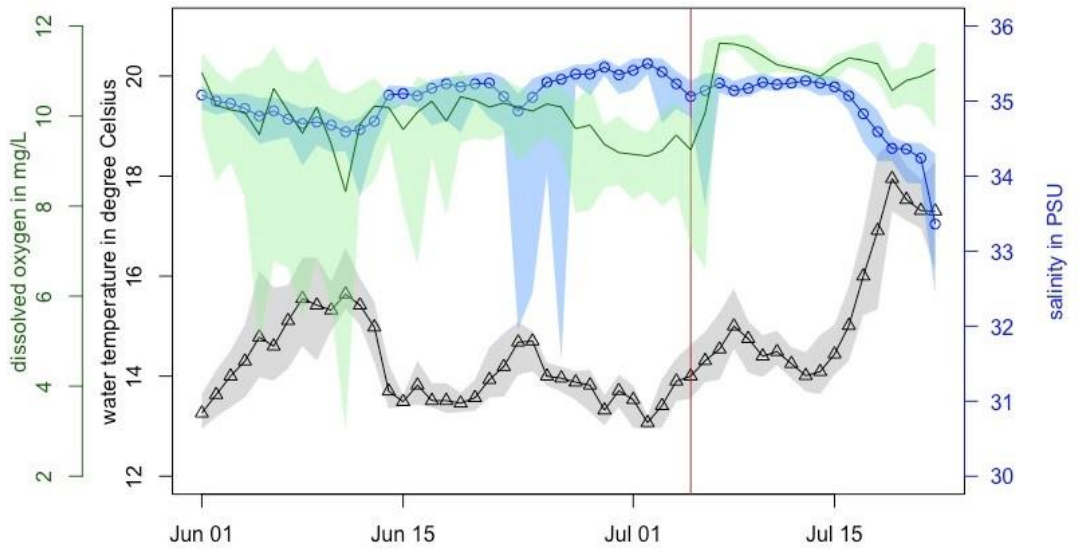

Figure 4. *In situ* data from monitoring buoy from 1st June to 22nd July 2016; a SST sensor at 1.10 m, a salinity sensor at 1.25 m (blue line with circle) and an oxygen sensor at 1.65 m (green line). The monitoring buoy was located in close proximity to the shellfish farm and its design is described in (Schmidt et al. unpubl. data) . Vertical red line indicates the date (5th of July 2016), when the shellfish farm was closed due to high biotoxin concentrations (> 160 µg OA eq. kg^{-1} shellfish flesh). The mean SST is shown as a black line with triangles (grey shading represents the minimum and maximum SST). The mean salinity is shown as a blue line with circles (light blue shading represents the minimum and maximum salinity) and the mean dissolved oxygen concentration is shown as a green line (light green shading represents the minimum and maximum concentrations).

640

641

Table 1. Overview of datasets used in the model development (Dev. dataset) and evaluation (Eval. dataset) for both study sites.

	St Austell Bay		**Turnaware Bar**	
	Dev. dataset	Eval. dataset	Dev. dataset	Eval. dataset
Time period	2008 - 2015	2016	2011 – 2015	2016
E. coli observations	94	13	53	10
Time period	2014 - 2015	2016		
Biotoxin observations	54	24		

642

643

Table 2. Comparison between modelled and observed *E. coli* for 2016 for both study sites. All values are presented as log_{10} *E. coli* 100 g^{-1} shellfish flesh. * = Best performing model

	St Austell Bay				**Turnaware Bar**			
	Observations	GLM*	Averaged GLM	GAM	Observations	GLM*	Averaged GLM	GAM
Mean	1.54	1.92	2.15	2.1	2.23	2.57	2.82	2.79
Standard deviation	0.39	0.33	0.32	0.37	0.68	0.31	0.3	0.36
RMSE		0.48	0.7	0.65		0.68	0.87	0.87
Bias		-1.54	0.61	0.57		0.34	0.58	0.56

644

Table 3. Explanatory variables identified for *E. coli* models for St Austell Bay and Turnaware Bar.

Model	**Explanatory variables for *E. coli* in St Austell Bay**	**Explanatory variables for *E. coli* in Turnaware Bar**
GLM	Lag rainfall* and solar radiation	River flow* and SST*
Averaged GLM	Lag rainfall*, solar radiation*, rainfall, wind speed and SST	River flow*, SST*, wind direction, solar radiation and rainfall
GAM	Lag rainfall, solar radiation and as smoothed term: rainfall	River flow*, SST* and as smoothed term: lag rainfall*

* Marked variables contributed significantly ($p<0.05$) to the model.

645

Table 4. Statistical measures for modelled and observed biotoxin concentrations in St Austell Bay for the year 2016, all values are presented as µg OA eq. kg^{-1} shellfish flesh. *= Best performing model

Observation/ Model	**Statistical parameters**			
	Mean	**Standard deviation**	**RMSE**	**Bias**
Observed data	514.71	633.46		
GLM	227.61	159.48	596.14	-287.1
Averaged GLM*	404.07	409.7	582.3	-110.64
GAM	443.7	507.24	671.59	-71.01

646

Table 5. Explanatory variables identified for modelled biotoxin concentrations in St Austell Bay.

Model	Explanatory variables for biotoxin in St Austell Bay
GLM	SST* and wind speed
Averaged GLM	SST*, solar radiation*, wind speed, lag rainfall and wind direction
GAM	SST*, solar radiation*, wind speed* and as smoothed term: lag rainfall

* Marked variables contributed significantly ($p<0.05$) to the model.

647