Predicting potentially toxigenic *Pseudo-nitzschia* blooms in the Chesapeake Bay

Authors: Clarissa R. Anderson[1*], Mathew R.P. Sapiano[2], M. Bala Krishna Prasad[2], Wen Long[3], Peter J. Tango[4], Christopher W. Brown[5], Raghu Murtugudde[2]

1 Ocean Sciences Department, University of California, Santa Cruz, 1156 High St.,
Santa Cruz, CA 95064

2 Earth System Science Interdisciplinary Center, University of Maryland, M Square
Office Building # 950, Suite 4001, 5825 University Research Court, College Park, MD
20740

3 Horn Point Laboratory, University of Maryland Center for Environmental Science, P.O.
Box 775, Cambridge, MD 21613

4 U. S. Geological Survey @ U.S. Environmental Protection Agency Chesapeake Bay
Program Office, 410 Severn Ave. Suite 109, Annapolis MD 21403

5 National Oceanic and Atmospheric Administration, Cooperative Institute of Climate
Studies, M Square Office Building # 950, Suite 4001, 5825 University Research Court,
College Park, MD 20740

*Corresponding Author: clrander@ucsc.edu, clarissa@umd.edu, 831-426-5074

Running Head: Predicting *Pseudo-nitzschia* blooms

Abstract

Harmful algal blooms are now recognized as a significant threat to the Chesapeake Bay as they can severely compromise the economic viability of important recreational and commercial fisheries in United State’s largest estuary. This study describes the development of an empirical model for the domoic acid-producing *Pseudo-nitzschia* species complex that will provides a first attempt at predicting the probability of potentially toxigenic blooms in the Chesapeake Bay using long term observations of cell abundance and concurrent measurements of hydrographic and chemical properties. Using a logistic Generalized Linear Model approach, it is shown that the *Pseudo-nitzschia* bloom threshold response defined as 100 cells mL^{-1} is best explained by surface values of phosphate, salinity, temperature, dissolved organic carbon, silicic acid, time of year, and the N:P ratio, where N represents nitrate plus nitrite and P is orthophosphate. These predictors are similar to other models for *Pseudo-nitzschia* blooms on the west coast, suggesting commonalities across ecosystems, including the possibility of nutrient limitation as a factor for bloom initiation. While salinity tolerance limits for this group restrict the extent of blooms in freshwater reaches of tributaries where nutrient enrichment may be greatest, the implication of possible Baywide eutrophication effects on *Pseudo-nitzschia* blooms are discussed. Hindcasts of modeled probabilities for past bloom and non-bloom events illustrate the effects of the overall model skill that is optimized at a 10% probability threshold according to a Heidke Skill Score of 30%, a Probability of Detection ~ 69%, a False Alarm Ratio of ~76%, and a Probability of False Detection ~ 10%.

Keywords: algal blooms, domoic acid, amnesic shellfish poisoning, diatoms, prediction, regression analysis, Chesapeake Bay

1. Introduction

A subset of species in the cosmopolitan diatom genus *Pseudo-nitzschia* has emerged as a major player in the global theatre of harmful algal blooms (HABs), occurring on both east and west coasts of the United States (Trainer *et al.*, 2000; Thessen & Stoecker, 2008), the Gulf of Mexico (Dortch *et al.*, 2000; Pan *et al.*, 2001), and throughout Europe and parts of Asia (Bates & Trainer, 2006). In response to a coarsely understood set of environmental conditions, toxigenic *Pseudo-nitzschia* species are known to produce domoic acid, a potent neurotoxin that can be devastating to aquatic life (Fritz *et al.*, 1992; Scholin *et al.*, 2000) via trophic transfer in the food web. In humans, domoic acid (DA) exposure manifests as amnesic shellfish poisoning (ASP) following the consumption of DA-contaminated filter-feeding mollusks. Samples of *Pseudo-nitzschia* spp. taken from tributaries of the largest estuary in the USA, the Chesapeake Bay, have recently been shown capable of DA production (Thessen & Stoecker, 2008 ;Thessen *et al.*, 2009) yet to date there has been no reported incidence of major bird or marine mammal strandings nor ASP in the Bay. However, emerging toxicological and epidemiological research suggests that chronic, sub-acute doses of domoic acid that do not cause any outward signs of intoxication may cause neuropathic injury to vertebrates (Levin, 2006; Ramsdell & Zabka, 2008) but that neuropathic injury may result from subacute doses (Lefebvre *et al.*, 2007). For humans, this implies the potential for serious neurological health risks associated with repeated exposure to low levels of domoic acid in coastal communities with an established culture of shellfish consumption (Grattan *et al.*, 2007). Furthermore, the accumulation of domoic acid in sediments after the demise of a surface bloom may reverberate throughout the ecosystem via accumulation of toxins in benthic food webs and turbulent re-mixing events (Thessen *et* *al.*, 2009). Clearly, there are significant benefits to extending HAB predictions from ecosystem models into the public health arena (Dyble *et al.*, 2008).

Human-driven nutrient enrichment, or cultural eutrophication, is a contentious factor driving the apparent global increase in HAB frequency and intensity (Hallegraeff, 1993; VanDolah, 2000; Anderson *et al.*, 2002; Glibert *et al.*, 2005; Heisler *et al.*, 2008; Anderson *et al.*, 2008). A number of questions are relevant to systems where eutrophication has been shown to be a major influence on HABs. These include the expected effects of nutrient stoichiometry (reviewed in Anderson *et al.* 2002; Glibert *et al.*, this issue), potential changes in ratios of dissolved organic carbon to dissolved organic nitrogen (DOC: DON; Paerl, 1988; Anderson *et al.*, 2002) and resultant increases in the probability of toxin production by certain species, such as those within the *Pseudo-nitzschia* genus (Davidson and Fehling, 2006; Pan *et al.*, 1996). With increased nutrient inputs from runoff come reduced Si:N and Si:P ratios (reviewed in Anderson *et al.* 2002), a factor shown to contribute to and/or be associated with increases of *Pseudo-nitzschia* abundance and domoic acid production (Pan *et al.*, 1996; Fehling *et al.*, 2004; Parsons & Dortch, 2002; Anderson *et al.*, 2006). Whereas most diatoms only thrive under Si-replete conditions, some *Pseudo-nitzschia* species appear to be favored by Si-limitation and certainly tend towards toxicity under the most extreme Si-limitation (Pan *et al*., 1996, 1998).

In the Chesapeake Bay, where HABs are common, diverse, and may be increasing in frequency, Glibert *et al.* (2001) documented a direct relationship between nitrogenous fertilizer use in the Bay watershed in spring and the onset of potentially toxic *Prorocentrum minimum* blooms. *Pseudo-nitzschia* spp. on the west coast generally bloom in response to surface macronutrient increases after natural, coastal upwelling events (Trainer *et al.*, 2000; Anderson *et al.*, 2006; Kudela *et al.*, 2005; Garcia-Mendoza *et al.*, 2009; Lane *et al.*, accepted) and could respond to similar pulses from spring runoff in the Chesapeake Bay (Heisler *et al.*, 2008). Furthermore, recent laboratory and field experiments link organic forms of nitrogen, such as glutamine and urea, to sustained bloom events in an upwelling region and even enhanced production of DA by toxigenic species of *Pseudo-nitzschia* (Howard *et al.*, 2007; Kudela *et al.*, 2008). In the Gulf of Mexico, time series of *Pseudo-nitzschia* spp. from sediment cores document increasing cell abundance in association with increased nutrient input from Mississippi River runoff (Parsons et al. 2006). Nonetheless it remains unclear if highly eutrophied bays and estuaries experience significantly more *Pseudo-nitzschia* blooms or DA events than coastal upwelling systems (Dortch *et al.*, 1997; Cochlan *et al.*, 2008). Long-term hindcasts and forecasts from predictive models of HABs will be extremely useful for understanding climate and land-use change effects on the frequency and magnitude of potentially deleterious algal blooms in the eutrophied Chesapeake Bay region (Kemp *et al.*, 2005).

We present here our efforts to develop a threshold-based, empirical model that aims to reliably predict the probability of *Pseudo-nitzchia* spp. bloom occurrence (and by extension the potential for domoic acid production) from a long-term monitoring dataset for the Chesapeake Bay. Currently, no such model exists for *Pseudo-nitzschia* on the east coast, despite the obvious importance of having such predictive capabilities in place. While the methods used to develop statistical models for estimating the distribution of harmful algae and pathogens are not new in scientific application, they are now more widely employed and accepted in ecological forecasting. A fully operational alert system that combines satellite imagery, numerical models, and a rules-based biological model is running in the Gulf of Mexico for the aerosolized toxin-producing, toxic *Karenia brevis* blooms (Stumpf *et al.*, 2003; Tomlinson *et al.*, 2004). Accompanying this trend are rapid advances in the development of dynamically-downscaled regional products for forecasting the physical and biogeochemical states of coastal ecosystems (e.g. Moore *et al.*, 2002; Chai *et al.*, 2002; Fennel *et al.*, 2006; Chao *et al.*, 2008; Allen *et al.*, 2008; Xu & Hood, 2006). A system has been implemented that predicts the likely distribution of sea nettles (*Chrysaora quinquecirrha*), the toxic dinoflagellate *Karlodinium veneficum*, and *Vibrio cholerae* in the Chesapeake Bay based on habitat suitability and hydrodynamic models (Decker *et al.*, 2007; Constantin de Magny, 2009). On the west coast, investigators are building upon earlier statistical models (Blum *et al.*, 2006) for predicting the likelihood and toxicity of *Pseudo-nitzschia* blooms from a suite of hydrographic and nutrient inputs in the Monterey Bay (Lane *et al.*, 2009) and the Santa Barbara Channel (Anderson *et al.*, 2009). The goal of the present study is to 1) identify the physico-chemical predictors of regional *Pseudo-nitzschia* spp. blooms, 2) develop and validate a predictive habitat model based on these environmental predictors that can be employed to predict the probability of occurrence of these blooms in the Chesapeake Bay, and 3) present hindcasts of past *Pseudo-nitzschia* blooms generated from coupling the habitat model with a regional hydrodynamic and nested ecosystem model.

2. Materials and Methods

Long- term data on *surface* phytoplankton abundance and *surface* water quality were acquired from the Maryland Department of Natural Resources (MD DNR) and US EPA Chesapeake Bay Program (US EPA-CBP) from the tidal waters of Chesapeake Bay and its tributaries. There were historical data for 30 stations within the mainstem Bay and 12 tidal estuarine tributaries covering the period 1985-2007 (N=6,989; Fig. 1), though somewhat patchy both spatially and temporally due to the often event-response nature of data collection and the requirement of daily matchups between water quality and cell count data. Because these data are derived from a monitoring program and were not collected with the purpose of creating predictive models, biases in space and time make it a less than ideal dataset despite the long time period and effort represented. The data subset used for model development included the monthly suite of water quality parameters collected within one day of generic *Pseudo-nitzschia* spp. cell abundance records at 30 sites across the broad range of salinities represented within the Chesapeake Bay. The poly- and oligo- haline regions are the most comprehensively represented salinity regimes in the pre-selected dataset, with fewer observations available for the mesohaline, or middle portion of the Bay due to difficulties in acquiring a high-confidence dataset with (Fig. 1). While Virginia and Maryland taxonomists both used the Utermöhl method for the inverted light microscope (Hasle, 1978), they also used different counting methodologies prior to 2005 (US EPA Chesapeake Bay Program, 2007). However, we do not expect this to significantly affect abundance counts of large cells (~10 μm), such as *Pseudo-nitzschia* spp. Additionally, due to changes in taxonomic status for the *Pseudo-nitzschia* group over the study period and difficulties with species identification using classic light microscopy, model development (Section 3.1) is focused on *Pseudo-nitzschia* spp. with no distinction between the various individual toxigenic and non-toxic species recorded for the Chesapeake Bay (Marshall *et al.*, 2005); (Thessen & Stoecker, 2008).

Monthly freshwater discharge data (ft^3 sec-^{1}) were acquired from USGS River Input Monitoring Program (http://va.water.usgs.gov/chesbay/RIMP/index.html) and geo-referenced with the phytoplankton monitoring stations for accurate match-ups. Note that there are no discharge data for 1984, only 1985 forward. For some stations, negative discharge values are reported, in which case, we eliminated those observations. Discharge data for stations CB6.4 and CB7.3E were not available, but due to their proximity to the Mattaponni River, those discharge data have been substituted for those observation matchups.

3. Modeling Approach

3.1 Model Development

A suite of 21 physical and chemical variables made publically available by the MD DNR and US EPA-CBP long-term water quality monitoring programs was chosen to examine conditions associated with regional variability in *Pseudo-nitzschia* spp. abundance (Table 1). Past studies for the *Pseudo- nitzschia* genus used Ordinary Least Squares (OLS) regression to model cell abundance and toxin concentration (Blum *et al.*, 2006; Anderson *et al.*, 2009) or logistic regression to model cell abundance (Lane *et al*., 2009) as a function of physico-chemical and other environmental properties. In this case, as in Lane *et al*. (2009), we are interested in only the presence or absence of ‘bloom’ levels of *Pseudo-nitzschia* spp. (hereafter referred to only as *Pseudo-nitzschia*). The response variable for cell abundance was transformed to a binary variable using an operational ‘bloom’ threshold value of 100 cells mL^{-1} for *Pseudo-nitzschia*, a value that is likely associated with increases in ecosystem effects on the west coast and which allows us to avoid further difficulties when attempting to predict absolute cell abundance (Westberry & Siegel, 2005; Anderson *et al.*, 2009; Lane *et al.*, 2009).

Bloom occurrence is a binary variable which rarely follows a Gaussian (Normal) distribution and renders OLS linear regression an unsuitable model. Instead, we used the Generalized Linear Model (GLM; McCullagh & Nelder, 1989; Fox, 2002; Agresti, 2007), a regression-based approach that allows for several non-Gaussian distributions in addition to the OLS linear regression. The GLM for binary data is often referred to as logistic regression and assumes that the binary response follows a Bernoulli distribution or the closely-related binomial distribution. The GLM relates a transformation of the response (‘bloom’ occurrence) to a linear combination of the explanatory variables. The transformation is referred to as the logit link function for binary variables and transforms the expectation of the response to the linear predictor as:

$$\log_e [p/(1-p)] = \beta_0 + \beta_1 x_{i1} + \ldots. + \beta_k x_{ik}, \tag{1}$$

where the term p/(1-p) represents the odds of a ‘bloom’ event, β_0 is the intercept, and β_i represents the regression coefficient for each explanatory variable x_i. Thus, the binary GLM simply models the log odds of an event as a linear combination of the explanatory variables. Solving for p, the expected value or probability of a bloom is then described by:

$$p = P_{bloom} = e^{(logit)}/[e^{(logit)} + 1]. \qquad (2)$$

A forward stepwise approach was used to determine the final model with reduction in deviance used to assess the relative importance of parameters in the model. Explanatory variables were added to the model individually, with the parameter leading to the largest statistically significant (at the 5% level, □ = 0.05) reduction in deviance being added at each successive step. To assess the influence of the annual cycle on environmental variables, hypothesis testing was performed twice, once with raw data values, and again with anomalies of all variables whereby monthly climatological means for each station are removed from the observations. Explanatory variables were also lagged from one to three months relative to the response variable in the GLM to test for significant lead/lag relationships.

3.2 Model Validation

The resulting logistic GLM was tested using cross-validation. For each year, predictions were made from the model fitted to all years excluding the one for which data were predicted. This process was repeated for each year of data to produce cross-validation model predictions. Predictions from the model come in the form of probabilities, such that a probability threshold is required for determining bloom from non-bloom conditions. Rather than setting this probability (or “prediction point” as expressed in Lane *et al.*, 2009) to an arbitrary value of 50%, it is appropriate here for extreme, episodic blooms to optimize the threshold in order to determine a relevant probability for defining a possible bloom event. Specifically, we optimized this value relative to the probability of detection (POD), the false alarm ratio (FAR), the probability of false detection (POFD) and the Heidke Skill Score (HSS; Heidke, 1926; Sohn & Park, 2008). These are defined as follows:

$$\text{POD} = (\text{correct hits})/(\text{correct hits} + \text{false negatives}), \quad (3)$$

where *correct hits* are the equivalent of correct bloom predictions and *false negatives* where bloom observations were predicted by the model to be non-blooms. POD is a component of the widely-used Relative Operating Characteristic (ROC; e.g. Lane *et al.*, 2009) :

$$\text{FAR} = (\text{false positives})/(\text{correct hits} + \text{false positives}), \quad (4)$$

where *false positives* are non-bloom observations over-predicted as blooms by the GLM; and

$$\text{POFD} = (\text{false positives})/(\text{correct negatives} + \text{false positives}). \quad (5)$$

where *correct negatives* are correctly predicted non-bloom observations. POFD is also an important aspect of the ROC and describes the proportion of non-bloom events that were falsely predicted to be blooms.

HSS is a generalized skill score that takes into account artificial predictability based on chance statistical relationships between parameters in the model and is an alternative to the more general ‘true skill’ for describing model performance but more stringent in terms of not giving credit for predicting climatological values (Davis, 1976; Siegel & Dickey, 1986; Song & Haidvogel, 1994; Anderson *et al.*, 2009). HSS values are functionally equivalent to Cohen’s □, range from –∞ to 1, and can be represented by the formula:

$$\text{HSS} = [(\text{correct hits} + \text{correct negatives}) - (\text{expected correct})_{\text{random}}] / [\text{N} - (\text{expected correct})_{\text{random}}]$$

(6)

where the expected number of correct predictions is expressed as:

$$(\text{expected correct})_{\text{random}} = 1/\text{N}\ [(\text{correct hits} + \text{false negatives})(\text{correct hits} + \text{false positives}) + (\text{correct negatives} + \text{false negatives})(\text{correct negatives} + \text{false positives})]$$

(7).

For each year in the cross-validation, the probability threshold is optimized for the POD, FAR, POFD, and HSS (R, version 2.7.0). Overall model performance is expressed as the mean of these iterated results.

4. RESULTS

4.1 Observations

Over the 22-year period for which we acquired *Pseudo-nitzschia* abundance data, very large bloom events are highly episodic in the region (Fig. 2). Mean abundance was 53 cells mL^{-1} over the range 0 - 1.9 x 10^4 cells mL^{-1} for all 30 stations. This maximum is consistent with other *Pseudo-nitzschia* observations (Thessen & Stoecker, 2008) and the recently noted increases in diatom abundance observed for the Bay (Marshall *et al.*, 2005). The largest bloom on record for the three stations shown in Figure 3 occurred in 1998 and reached over 5.0 x 10^3 cells mL^{-1} and 1.8 x 10^4 cells mL^{-1} at stations CB5.1 and CB6.4, respectively (Fig. 2, Table 2), far exceeding the 'bloom' threshold of 100 cells mL^{-1} employed in the predictive logistic model. These concentrations of 10^3 to 10^4 cells mL^{-1} are comparable to the largest blooms observed on the west coast where domoic acid poisoning is a recurring threat to living resources and human health. When transformed to a binary bloom variable, the *Pseudo-nitzschia* abundance data set contains 6,522 non-bloom and 733 bloom samples (Table 2). The majority of blooms in the record occur in the middle and lower bay- mouth region (stations CB5.1, CB6.1, CB6.4, CB7.3E, CB7.4) with notable hotspots for bloom development in tidal estuarine zones (EE3.0, EE3.3, LE1.1, LE3.1, LE2.2, RET1.1; Fig. 1, Table 2). Locations with the largest variability in bloom occurrence are the middle and lower bay stations as well as the downstream Patuxent River sites (LE1.1, LE1.3, RET1.1), the downstream Potomac River sites (LE2.2, RET2.1, RET2.2, RET2.4), the eastern bay estuarine stations EE3.0 and EE3.3, and the Patapsco River mouth in the north of the Bay (WT5.1; Table 2).

Although several large blooms were recorded for the lower Bay during the 1980s and 1990s, blooms at the mouths of major tributaries such as the Potomac (LE2.2), Choptank (ET5.2), and Patuxent Rivers (LE1.3) are relatively small and infrequent prior to 1997 at which point the lower bay and tidal estuaries appear to become synchronized in terms of bloom timing, but not necessarily intensity (Fig. 2). This potential lag between the lower bay stations (CB6.1, CB6.4, CB7.3E, CB7.4) and the middle/upper bay over the study period could signify the introduction of oceanic seeding populations transported through the mouth of the Chesapeake Bay northward into the estuary proper (Marshall *et al.*, 2005; McGillicuddy Jr. *et al.*, 2005) where they may persist at depth until an upwelling or mixing event occurs and brings them to the surface (Mengelt, 2006). Alternatively, the observed relationship could suggest an evolution of conditions desirable for *Pseudo-nitzschia* blooms, such as eutrophication, in the mesohaline middle reaches of the bay beginning in the late 1990s (Glibert *et al.*, 2001; Marshall *et al.*, 2005). Any likely relation to the persistent cold conditions observed since 1997 in the tropical Pacific will need further exploration (Wang *et al*., 2009*.)*

Data collected at stations along a broad salinity gradient from the North Bay near the Susquehanna River to the South Bay mouth to the Atlantic Ocean demonstrate a positive trend of increasing bloom frequency and intensity with increasing salinity (Fig. 3). No *Pseudo-nitzschia* cells were recorded at the most oligohaline station (Fig. 3) just south of the Susquehanna River delta (0.5-5 ppt; CB2.1), whereas the typically mesohaline (5-18 ppt; CB5.1) and polyhaline (>18 ppt; CB6.4) stations experienced seasonal blooms with large interannual variability and a greater frequency and intensity of blooms at the most saline station (Fig. 3). These patterns are consistent with expected salinity requirements for a marine diatom (Hasle & Syvertsen, 1996; Thessen *et al.*, 2005) and corroborate observational data on the low temperature and high salinities associated with the presence of *Pseudo-nitzschia* populations in the Chesapeake Bay (Thessen & Stoecker, 2008).

4.2 *Pseudo-nitzschia* Bloom Likelihood Model

Correlations between environmental variables and the response variable *Pseudo-nitzschia* cell abundance indicate a weakly significant relationship between temperature (r = -0.10), salinity (r = 0.12), orthophosphate (PO_4; r = -0.10), nitrate (NO_3; r = -0.07), and silicic acid ($Si(OH)_4$; r = -0.08; α = 0.05 level; Table 1, Table 3, Fig. 4). The correlation between *Pseudo-nitzchia* cell abundance and river discharge is not significant (r = -0.03). However, it is expected that nutrient concentrations and salinity are biologically-relevant proxies for the effects of freshwater runoff into the Bay that incorporate the effects of precipitation and discharge. Cell abundance is generally highest from 5-27°C with an apparent temperature optimum near 10°C, 5-28 psu with an apparent salinity optimum near 15 psu, and at low nutrient concentrations (Fig. 4). However, it should be noted that the nutrient concentration range (Fig. 4) indicates a highly-eutrophied system where concentrations are rarely limiting to phytoplankton. These correlation values are generally consistent with the results of the logistic GLM that significantly associates ‘blooms’ of *Pseudo-nitzschia* (threshold = 100 cells mL^{-1}) with PO_4, salinity, temperature, dissolved organic carbon (DOC), $Si(OH)_4$, month of the year, and the NO_3 + NO_2 to PO_4 molar ratio (abbreviated N:P; Table 1, Fig. 4, Table 4). Despite the fact that climatological anomalies of all variables were tested during model development to account for seasonality in the parametric effects, the final model that best captures the environmental thresholds that drive *Pseudo-nitzschia* blooms is the model built from the raw data, retaining a seasonal signal within the parameters and including month as a significant predictor. Likewise, lag effects were not found to be significant for any of the explanatory variables tested.

In agreement with a previous study (Thessen & Stoecker, 2008) and explaining the negative relationship between blooms and month of the year, the majority of bloom events occurred in winter and spring months (76% for Jan-May) when surface nutrient enrichment from storm-induced mixing and runoff should be greatest. The nutrient parameter PO_4 is the most significant predictor in the model and is negatively associated with the *Pseudo-nitzschia* bloom index, indicating a connection between *Pseudo-nitzschia* blooms and P-limitation or simply the biological drawdown of PO_4 in the presence of blooms. A significant positive relationship with salinity and negative relationship with temperature (Table 4) again implies physical tolerance limits and seasonality that are expected for this diatom group and have been observed for *Pseudo-nitzschia* in the Chesapeake Bay (Pan *et al.*, 1993; Thessen & Stoecker, 2008). The inclusion of DOC as a significant predictor in the GLM is particularly interesting given its connection to riverine inputs (e.g. Doering *et al.*, 1994; McKenna, 2004) and the previously observed associations between *Pseudo-nitzschia* blooms and river discharge or potential proxies of runoff on the west coast (Kudela *et al.*, 2004; Anderson *et al.*, 2009; Lane *et al.*, 2009). For $Si(OH)_4$, there is an inverse relationship with blooms as was seen with PO_4, resembling the previously observed association between *Pseudo-nitzschia* blooms and the onset of Si-depletion in other regions and experiments (Sommer, 1994; Marchetti *et al.*, 2004; Anderson *et al.*, 2006; Lane *et al.*, 2009). The least significant explanatory variable is the N:P ratio that is included in the model only when nitrite (NO_2) is accounted for in the nitrogen pool, because variability in nitrate alone is not significantly associated with *Pseudo-nitzschia* blooms in the GLM. A negative relationship with NO_3+NO_2 availability relative to PO_4 may point to N-limitation as a weak bloom driver or could indicate that N-uptake by surface blooms leads to depletion of both the NO_3 and NO_2 pools, rather than to NO_2 regeneration, which is expected during more N-replete conditions (Bode *et al.*, 2002; Lomas *et al.*, 2000).

4.3 Model Performance

The boxplot in Figure 5 shows the probability of a bloom as predicted by the logistic GLM split into bloom (median = 16%) and non-bloom (median = 0.08%) events, again defined by transforming raw cell abundance to a binary bloom variable using the 100 cells mL^{-1} threshold. The low predicted probability of blooms reflects the rarity of a bloom event (the observed probability of a bloom event is 6.7%) and the relatively weak relationship between the explanatory variables as well as the choice of threshold. Points falling outside of the 95th percentile whiskers in the boxplot denote outlier points, or in the case of the non-bloom observations, the false positive predictions (i.e. model-predicted blooms where no blooms were observed). Model fits for a logistic regression cannot be assessed using a coefficient of determination (R^2) as is often done in the case of OLS linear regression, however 'pseudo R^2' measures are available such as the one suggested by Nagelkerke (1991) that can be applied to the logistic GLM and very roughly translates to the proportion of variability explained by the model. The pseudo-R^2 for the *Pseudo-nitzschia* GLM is 9.7%.

If the GLM were applied in a nowcast or forecast mode, it would be necessary to further define a bloom based on the probabilities generated by incorporating real-time, near real-time, or forecasted environmental parameters. If a default 'bloom' probability threshold of 0.50 is employed to define a *Pseudo-nitzschia* bloom, then the POFD (Table 5) is very low (0.003) but the POD is also low (0.01). While a low POFD is desirable, a low POD is highly undesirable, and it is clear from this scenario that a bloom threshold probability of 0.5 is far too high and leads to a severe under-estimation of bloom events. Since our goal is to maximize the model predictive skill (in this case, HSS) and consequently maximize the POD while minimizing the FAR and the POFD, we have optimized the probability threshold according to these metrics (Fig. 6). By doing this, we balance the need to predict potentially harmful blooms with the inconvenience of incorrectly predicting a bloom. Using this method, we chose the maximum POD and minimum FAR that correspond to the top 10% of HSS values, resulting in an optimized bloom probability threshold of 0.10 to achieve an HSS of ~0.30, POD of ~ 0.69, FAR at ~0.76, and POFD at ~0.10 (Fig. 6, Table 5).

4.4 Model Cross-Validation

In order to maximize skill in the GLM, the model presented above was developed using all available data (n=5,121) rather than split into training and validation groups as is often done in the absence of an independent validation data set. Cross-validation of the GLM was then performed by excluding a single year of data for the model fitting and making predictions for the excluded year from this model. This process was iterated for each year in the time series and model performance for each iteration was calculated according to the optimization method described earlier for a 0.10 bloom probability threshold. Cross-validation of the GLM yields the following performance metrics: HSS ~0.31, POD ~ 0.69, FAR ~0.76, and POFD ~0.10, and Nagelkerke's R^2 ~0.10 (Table 5). These cross-validation results are consistent with the performance metrics given above, indicating consistency in the hindcasts across years, where the most variability explained by the model is ~10%, with an approximately 69% chance of correctly predicting a bloom.

4.5 *Pseudo-nitzschia* Bloom Prediction Maps

Hindcast maps of Baywide bloom probabilities were created using spatial interpolation of physico-chemical observations relevant to the logistic GLM for *Pseudo-nitzschia* blooms for the period 1990-2007 (Prasad *et al.*, 2009). The maps illustrate bloom probabilities interpolated according to the Chesapeake Bay Regional Ocean Model System (ChesROMS) grid over that time period (Fig. 7). Ideally, modeled hindcasts of surface PO_4, salinity, temperature, DOC, and $Si(OH)_4$, are paired with month index to calculate bloom probabilities at 4.5-km^2 horizontal resolution according to the empirical GLM algorithm:

$$\text{logit} = -1.84 - 69.5*PO_4 + 0.101*Sal - 0.063*Temp + 0.306*DOC - 0.373*Si(OH)_4 - 0.078*Month - 0.753*N{:}P, \quad (8)$$

where the probability of a bloom (P_{bloom}) follows eq. (2). Unfortunately, DOC measurements are not available for many of the tributary monitoring sites, precluding computation of the full GLM from eq. (8) for most months. To illustrate the difference in outcomes between hindcasts that incorporate DOC (and thus have limited tributary coverage) verses hindcasts that employ a GLM with the DOC parameter removed, both hindcast scenarios for a data-replete month (November 1993) are presented in Figure 7 using eq. (8) and the alternative logit algorithm,

$$\text{logit} = 0.016 - 106*PO_4 + 0.083*Sal - 0.058*Temp - 0.106*Si(OH)_4 - 0.082*Mo - 1.05*N{:}P.$$

(9)

It appears that the inclusion of DOC is important for predictive skill based on differences in the probability range generated by the logit with and without DOC (Fig. 7a) but may not have profound impacts on the overall patterns of bloom prediction. The month of November falls within a non-bloom period, as is shown for November 1993 by by both models, which assign low probabilities (i.e. at or below the optimized 10% probability threshold) to the oligohaline and polyhaline regions where observations are in the 0-50 cells mL^{-1} range. For the mesohaline region, on the other hand, both GLM's predict the potential for blooms by assigning bloom probabilities of 10-16% and 10-14% when using the GLM with DOC (eq. 8) and the GLM without DOC (eq. 9), respectively. Predicted bloom likelihood for this region, however, cannot be adequately evaluated using the available suite of observations for November 1993. An alternative may be the use of river discharge data as a proxy for DOC since it emerged as a significant predictor in GLM development but was outcompeted by DOC in the final model.

For the entire 17-year period, we compared gridded hindcast maps generated from the alternative GLM from eq. (9) with the distribution of *Pseudo-nitzschia* abundance observations. Figures 7b, c present two of these hindcasts: a large bloom event in March 1998 (see also Fig. 2) and a non-bloom event in August 2005 when there were no *Pseudo-nitzschia* recorded in the phytoplankton dataset. The distribution of observations above the 100 cells mL^{-1} bloom threshold is generally well-captured by the alternative GLM which assigns probabilities between 10-20% to the cluster of observations at or above 1,000 cells mL^{-1} in the lower bay (Fig. 7b). For one bloom-level observation in the middle mesohaline region of the Bay, the GLM assigned a probability just below the optimized 10% probability threshold (Fig. 7b), thus underestimating the potential for a bloom in this zone where few observational data were available for GLM construction. Conversely, very low probabilities were predicted throughout the Bay during the non-bloom event of August 2005, reflecting the absence of *Pseudo-nitzschia* in the cell abundance records for that month (Fig. 7c). What is most evident for all hindcasts (not shown) in the study period is the general ability for the GLM to capture the seasonal and interannual variability in bloom presence and intensity. However, the GLM also tends to underestimate bloom presence in portions of the Bay typified by the lower end of the mesohaline salinity range (5-10 psu), particularly where the salinity gradient is strong near the mouths of the Potomac, Choptank, and Patuxent Rivers during the spring bloom months. This is expected given the fairly high FAR presented in the cross-validation results and the fact that the GLM is strongly driven by a positive relationship between salinity and blooms. In other words, the GLM may not be sensitive enough to predict all blooms in the fresher, tidal tributaries, a factor that is certainly contingent on placement of the probability threshold (Fig. 6). Since the GLM does better at capturing major bloom events in the lower and eastern Bay by assigning higher bloom probabilities (> 10%) to those areas in the meso- and polyhaline portions of the Bay most often affected by episodic *Pseudo-nitzschia* blooms (Thessen & Stoecker, 2008), the 10% probability threshold may be most appropriate for this region. While development of separate models for the upper and lower portions of the Bay may help reduce false positives, the best, publically-available dataset for *Pseudo-nitzschia* is limited (and biased) by a disproportionate number of stations in the meso- and oligo-haline portions of the Bay.

5. Discussion

Model results presented here for *Pseudo-nitzschia* blooms in the Chesapeake Bay join the growing effort to create and validate empirical habitat suitability models for species or groups of biogeochemically important or potentially deleterious pathogens, microbes, zoonotic diseases, and planktonic organisms (e.g. Patz *et al.*, 2008; Decker *et al.*, 2007; Constantin de Magny *et al.*, 2008; Iglesias-Rodriguez *et al.*, 2002; Blauw *et al.*, 2006). An inherent constraint of these ecological forecasting models is the explanatory power of the available predictor variables, which in this study were limited to a standard array of hydrographic data and chemical constituents from long-term water quality monitoring programs in an ecologically complex estuarine system. This complexity presents challenges for forecasting biological processes in the absence of strong physiological data to support mechanistic modeling. Great strides have been made in this direction for at least one group of HABs in the mid-Atlantic region (McGillicuddy Jr. *et al.*, 2003; McGillicuddy Jr. *et al.*, 2005). However, pressing human health issues drive the need for statistical forecasting methods that take advantage of data currently in existence to glean process and predictive understanding and build models for a variety of HABs and toxic agents (Moore *et al.*, 2008). To that end, the approach described herein marks the first such attempt for the *Pseudo-nitzschia* genus on the U.S. eastern seaboard, utilizing one of the few long-term HAB datasets in the world.

5.1 Performance of the *Pseudo-nitzschia* Bloom GLM

Predictive success for the GLM depends in part on our chosen threshold for defining a 'bloom,' both in creating the binary variable for the logistic regression and determining a relevant probability threshold for forecasting blooms from model output. The optimization method for choosing a probability threshold (Fig. 5) allows the user to adjust the operational probability according to the relative weightings of the various performance metrics. A variety of metrics aided in the evaluation of GLM performance: HSS, POD, FAR, and POFD. Optimizing according to the skill score results in a 69% success rate (POD), a 76% false positive rate (FAR), and 10% false negative rate (POFD) for the cross-validated results (Fig. 6, Table 5). A Nagelkerke's R^2 of 9.7%

reflects the overall weak relationship between the environmental variables and bloom events seen in Table 3 and indicates the limitations that this model may present without future tuning and validation. It further points to the potential influence of driving factors that are from this analysis, in particular the inclusion of organic nutrients and trace metals.

Relative to *Pseudo-nitzschia*-specific predictive models for the west coast, this GLM slightly underperforms in its ability to correctly predict blooms for the Chesapeake Bay with a 69% probability of correct bloom prediction (cf. 75-89%; Anderson *et al.*, 2009; “sensitivity” in Lane *et al.*, 2009) and a high FAR of 76%. In terms of error types, minimizing POFD over FAR is arguably more desirable in the context of protecting human health since the perceived or actual damage from missing a potential health threat would be worse than a false alarm scenario. In that sense, this model does perform well at the optimized bloom probability threshold of 10%. However, management use of such threshold models ultimately requires a keen understanding of this conditionality (see Fig. 6) along with an adaptive strategy for implementation that takes into account model and threshold sensitivity within different zones of the Bay as well as emerging data on the risks of both acute and chronic exposure to domoic acid toxins (Erdner *et al.*, 2008). In any case, it is expected the predictions would be paired, at least initially, with microscopic examination of water sampled from the predicted bloom site for confirmation.

In addition to the explanatory content of model predictors, developing the GLM for *Pseudo-nitzschia* blooms requires several key assumptions about the data that may reduce predictive success. The first is the taxonomic fidelity of the microscopic identifications in the available monitoring observations that can skew results of a predictive model should incorrect assignments exist. To minimize this problem, we limited the majority of observations to those collected by a single taxonomist whenever possible (W. Butler, pers. comm.); this excludes data from the four lower bay sites that were required for full bay coverage. A second assumption is that all species of *Pseudo-nitzschia* bloom in response to similar ecological conditions or occupy the same niche space. There are nine species capable of producing domoic acid in the *Pseudo-nitzschia* genus, and of the eight species present in the Chesapeake Bay (Marshall, 1980, 1994; Marshall *et al.*, 2005), six are known toxin-producers (Thessen & Stoecker, 2008). Since such a large proportion of resident species are potentially toxigenic and were only included in analysis when collected from surface samples, it is reasonable to assume they are similarly adapted or that a given sample is dominated by only one or two *Pseudo-nitzschia* species most adapted to the present conditions (Bidigare *et al.*, 1990; Smayda, 1963; Margalef, 1978). This further implies wider utility in a model that predicts generic *Pseudo-nitzschia* blooms since the likelihood of domoic acid production may be high when members of the indigenous *Pseudo-nitzschia* assemblage are present at high densities. However, the important caveat remains that genetic diversity at the species and strain level can strongly regulate toxicity and cannot be accounted for in the GLM (Orsini *et al.*, 2004; Erdner *et al.*, 2008; Evans *et al.*, 2005; Thessen *et al.*, 2009) and that the niche-space flexibility offered by such diversity introduces further complications for predicting *Pseudo-nitzschia* distributions based on habitat suitability models.

5.2 Forecasting *Pseudo-nitzschia* Blooms

The habitat suitability model for *Pseudo-nitzschia* spp. was accomplished in conjunction with and for two on-going projects: (1) a NOAA-sponsored project to predict the probability of key harmful species for the region, and (2) the Chesapeake Bay Forecasting System (CBFS) pilot project for dynamically-downscaling seasonal to interannual climate forecasts and IPCC projections (Mehl *et al*., 2007) with a regional Earth System model consisting of atmospheric, watershed, and estuarine components at present (Murtugudde, 2009). The CBFS creates a 14-day outlook every three days for the Regional Earth System for the Chesapeake Bay and its watershed (http://www.climateneeds.umd.edu/). Dynamically downscaling with the Weather Research and Forecast Model (WRF; http://www.wrf-model.org/wrfadmin/publications.php) provides high resolution forcing for the Regional Ocean Model System (ChesROMS) ocean model and Soil and Water Assessment Tool (SWAT) river basin-scale model. The forecast is generated using a six to eight member ensemble forced with the Global Ensemble System (GENS) every three days. The ChesROMS model currently includes a nested Nitrogen-Phytoplankton-Zooplankton-Detritus (NPZD) ecosystem model for generating forecasts of macronutrient and chlorophyll concentrations, among other parameters (Fennel et al 2006).

Implementing a *Pseudo-nitzschia* bloom nowcast-forecast system will require assimilation of environmental data from these predictive models into the diagnostic GLM computed for raw data values. Ideally, the ChesROMS and NPZD models would predict all seven explanatory variables, however, values for $Si(OH)_4$ and DOC are not predicted from the ecosystem model currently coupled to ChesROMS. In the case of

DOC, Figure 7 illustrates the ramifications of removing this parameter from the GLM, and additional removal of $Si(OH)_4$ would reduce predictive skill even further, particularly since this nutrient is critical to diatom growth. Present research involves diagnostic coupling of alternative ecosystem models to ChesROMS that include $Si(OH)_4$ in their parameterization (e.g. Moore *et al.*, 2002; Chai et al. 2002). Unfortunately, it may not be possible to generate accurate DOC predictions from existing ecosystem models which at best include a “small detritus” carbon fraction (Christian *et al*., 2001; Moore *et al*., 2002), in which case, the alternative logit (8) would be employed for making *Pseudo-nitzschia* bloom nowcasts and forecasts.

In any operational or routine monitoring context, probability maps of *Pseudo-nitzschia* blooms should be presented with annotated instructions for interpretation of these uncertainties. With the optimized bloom threshold of 10%, there is a higher likelihood of bloom overprediction in the more oligohaline regions (Fig. 7) suggesting that forecasts for the meso- and poly- haline domains are more trustworthy at this operational threshold. We expect these forecasts to serve more as cautious warnings of the potential for bloom development rather than as absolute predictions of bloom intensity while forecast validation and predictive skill continue to be assessed, much in the way that predictions of stinging sea nettles and the toxic dinoflagellate *Karlodinium veneficum* are currently presented for the Chesapeake Bay (Brown *et al.*, 2002; Decker *et al*. 2007, http://155.206.18.162/seanettles/).

5.3 Eutrophication and *Pseudo-nitzschia* Blooms

Much deserved attention has been given in recent decades to studying the potential influences of cultural eutrophication on HAB ecology (e.g. reviews by

Anderson *et al.*, 2002; Anderson *et al.*, 2008; Heisler *et al.*, 2008; Glibert *et al*, this issue). While most HAB species fall within various flagellate and cyanobacterial phylogenies, *Pseudo-nitzschia* is unique in being a toxin-forming diatom, the group classically used as the default counterpoint in descriptions of HAB/flagellate ecophysiology (Smayda, 1997). Due to fundamental differences in adaptive strategies and life histories between diatoms and flagellates, our understanding of how *Pseudo-nitzschia* would respond to eutrophication is in its nascent stage. In an evaluation of Chesapeake Bay phytoplankton community composition, Marshall et al. (2005) emphasize a post-European settlement shift from diatom assemblages dominated by benthic, pennate diatoms to more planktonic, centric species (Cooper & Brush, 1991), a general decrease in diatoms associated with increased anoxia and P loading (Brush & Davis, 1984), and a shift towards more lightly silicified diatom species. However, the most modern surveys also document a dramatic increase in the magnitude of surface diatom abundance that coincides with an overall biomass enhancement and seasonal fluctuations of potentially toxic phytoplankton in the Bay and its tributaries (Marshall *et al.*, 2005).

In the present study, while freshwater discharge was not selected as one of the most significant predictors of *Pseudo-nitzschia* blooms in the GLM, it did retain significance on a par with DOC until finally being outcompeted by those potentially covarying factors, DOC and N:P. Moreover, there seems to be an increase in both the abundance and frequency of *Pseudo-nitzschia* blooms over the time series, particularly since the 1990s (Fig. 2), though sampling at many stations also increased during that decade. Perhaps in part due to their lightly silicified frustule, *Pseudo-nitzschia* cells seem better adapted to low-Si environments than many other diatoms, often outcompeting other groups towards the end stages of a spring diatom bloom when surface waters have been fairly depleted of Si (Anderson *et al.*, 2006; Kudela *et al.*, 2004; Marchetti *et al.*, 2004). As with other *Pseudo-nitzschia* habitat models for the west coast (Anderson et al 2009; Lane *et al*., 2009), this pattern is supported by the Chesapeake Bay GLM results where a significant negative relationship between *Pseudo-nitzschia* blooms and $Si(OH)_4$ concentrations could indicate a lower Si-requirement for this diatom group and may further suggest a link between runoff-induced reductions to Si:N and Si:P ratios (Smayda, 1990; Smayda, 1997) and the occurrence of these blooms. However, like all diatoms, *Pseudo-nitzschia* still require Si, N, and P for maximal growth and will bloom in response to these inputs, particularly during winter and spring months when turbulence and reduced light levels in the water column favors diatoms over flagellates.

The positive association between DOC and *Pseudo-nitzschia* can be interpreted as a link between runoff and blooms through the delivery of organic carbon to the Bay and possible direct effects on the phytoplankton population. Glibert et al (2001) found strong correlations between DOC and DOC:DON in relation to *Prorocentrum minimum* populations in the Bay and underscore the difficulty in ascribing direct or indirect effects to this relationship. One possible indirect effect is the potential for covariation between DOC and DON if *Pseudo-nitzschia* are directly taking up organic nitrogen species during blooms. There is some evidence for the uptake of organic substrates by *Pseudo-nitzschia* after dark adaptation (Mengelt, 2006), but the possibility of interference by extra-cellular or epiphytic bacteria remains (Bates *et al.*, 2004; (Stewart, 2008). Studies on the west coast also implicate some kind of allochthonous component to *Pseudo-nitzschia* blooms from either direct observation between blooms and rain events (Trainer *et al.*, 2000) or via the inclusion of chromophoric dissolved organic matter and precipitation/nitrate relationships in their habitat models (Anderson *et al.*, 2009; Lane *et al.*, 2009). Providing there is a causal relationship here, a positive influence of DOC on *Pseudo-nitzschia* blooms implies the potential for future problems as precipitation events over the Bay increase in frequency and/or intensity according to current IPCC projections (Meehl *et al.*, 2007).

The same inverse relationship with nutrients is again manifest in the association with PO_4 and N:P in the *Pseudo-nitzschia* GLM. It is important to note that these relationships to nutrient concentrations are not any more robust when lagged relative to bloom occurrence. However, a meaningful lag correlation for phytoplankton responses to nutrient supply is expected to be less than the average monthly sampling frequency available here because of the relatively short timescales for nutrient uptake. Thus, in the GLM, the inverse relationships between nutrient parameters and the *Pseudo-nitzschia* bloom index invoke both concurrent surface nutrient uptake and the potential onset of nutrient limitation. In spring, when the majority of these blooms occur, the Chesapeake Bay is indeed more limited by PO_4 relative to nitrogen (Prasad *et al.*, 2009). However, in the absence of direct measurements using nutrient limitation assays, it is not possible to assess the physiological status of the phytoplankton community at the time of sampling. P-limitation has not been linked to the onset or magnitude of natural *Pseudo-nitzschia* blooms per se but has been shown to stimulate domoic acid production in laboratory cultures (Pan *et al.*, 1998).

The N:P parameter in the GLM points to rapid NO_3 and NO_2 utilization and possibly even an N-limitation which is somewhat inconsistent with evidence of a positive association between nitrate, possibly from runoff, and fall *Pseudo-nitzschia* blooms in the Monterey Bay, CA (Lane *et al.*, 2009). However, the same caveat remains that a negative relationship between a nutrient parameter and the bloom index may only reflect the effects of nutrient drawdown by actively growing phytoplankton or could also suggest the possible onset of limiting conditions that may favor one phytoplankton group over another or alter a toxigenic group's ability to produce toxin. Several studies have now effectively demonstrated the uptake of eutrophic or regenerated forms of nitrogen -- ammonium (NH_4), and urea -- by *Pseudo-nitzschia* species (Howard *et al.*, 2007; Kudela *et al.*, 2008; Cochlan *et al.*, 2008; Thessen *et al.*, 2009), and an increase in the potential for domoic acid production when incorporating organic N or NH_4 in place of nitrate (Howard *et al.*, 2007; Thessen *et al.*, 2009). While NH_4 is not a significant explanatory variable for blooms in the Chesapeake Bay GLM, it is significantly associated with *Pseudo-nitzschia* abundance in this dataset (Table 3) and could be a concern for the future odds of toxic events. Because of the salinity tolerance limits for *Pseudo-nitzschia* shown in this study and elsewhere (Thessen *et al.*, 2005; Thessen and Stoecker, 2008), blooms are not expected to migrate or occur in the upper, fresh reaches of tributaries where nutrient enrichment from runoff is high. However, this enrichment is manifest in the meso-haline and poly-haline portions of the Bay as well (Prasad *et al*., 2009). Given that total N from the Susquehanna River, a major tributary of the Chesapeake Bay, increased 2.5-fold from 1945 to 1990 (Kemp *et al.*, 2005) and that urea usage in agriculture and industry is increasing worldwide (Glibert et al 2006), the potential role of NO_3 and urban forms of nitrogen runoff in driving *Pseudo-nitzschia* blooms and even toxin production should not be ignored, particularly in light of the projected increases in precipitation and population indices for the region (Meehl *et al.*, 2007).

Eutrophication is not only manifest in the bottom-up direction with respect to HABs but is expected to influence the entire ecosystem with the net result being more favorable conditions for HAB development. The timing of nutrient enrichment, for instance, may be de-coupled from predator population increases, resulting in the release of grazing pressure which can be further enhanced by a build-up in anoxia as the bloom decomposes (e.g. Buskey *et al.*, 1997). The present study, however, found no significant relationship between chlorophyll and *Pseudo-nitzschia* abundance for the Chesapeake Bay, suggesting a de-coupling of these blooms from the overall rise and fall of phytoplankton biomass and again suggesting a late-succession role for *Pseudo-nitzschia* species in the spring bloom time period. Top down controls such as grazing could, in part, be regulating the demise of *Pseudo-nitzschia* blooms and the extent to which they develop into toxic blooms.

6. Concluding Remarks

A predictive model for potentially toxigenic *Pseudo-nitzschia* blooms in the Chesapeake Bay serves two major purposes: 1) to identify possible drivers of initiation and presence of these blooms in both the temporal and spatial domains, and 2) to establish statistical relationships that could be used for real-time, ecological forecasting of this species in a critical U.S. ecosystem. The GLM developed in this study identifies several environmental variables associated with these blooms – PO_4, salinity, temperature, DOC, $Si(OH)_4$, time of year, and the N:P ratio – from a unique, coincident long-term water quality and phytoplankton monitoring dataset.   Unfortunately, it cannot reveal environmental indicators of the presence of its associated toxic agent, domoic acid. Indeed, defining the distinction between factors that initiate bloom development and those that trigger cellular toxin production is a recurring theme in HAB ecology, particularly for *Pseudo-nitzschia*, and is highly relevant in a region such as the Chesapeake Bay which to date has experienced no known outbreaks of domoic acid poisoning but whose populations may still be affected by chronic exposure to moderate toxin levels (Thessen and Stoecker, 2008; Grattan *et al.*, 2007). Forecasting endeavors that follow should not only relate these uncertainties to end-users but also continually re-define bloom thresholds with the addition of new validation data and input from resource managers and shellfish harvesters alike. Lastly, future efforts to model HABs must look closely at the combined effects of inorganic and organic nutrient uptake, trace metal utilization (e.g. Maldonado *et al.*, 2002; Rue and Bruland, 2001; Wells *et al.*, 2005), allelopathic competition (e.g. Adolf *et al.*, 2008), and grazing (e.g. Bargu *et al.*, 2003) to move beyond statistics-based habitat models and incorporate a mechanistic understanding of the complex processes regulating HAB development and toxicity.

7. Acknowledgements

This study was funded by a National Research Council postdoctoral fellowship to C.R.A., sponsored by the NOAA National Environmental Satellite, Data, and Information Service (NESDIS) with additional support from NOAA MERHAB Award (PI: C.W.B.) and the Chesapeake Bay Forecasting System project (PI: R. M.). The authors gratefully acknowledge the Maryland Department of Natural Resources and USEPA Chesapeake Bay Program for contributing monitoring data, and we are especially appreciative of the technical assistance provided by J. Johnson, B. Cole, M. Naylor, W. Butler, B. Wind, M. Baith, and J. Beauchamps. J. Lane and R.Kudela provided comments which greatly improved the paper. We give special thanks to A. Thessen for her innovative and inspiring work on *Pseudo-nitzschia* ecophysiology in the Chesapeake Bay and for significant input to this manuscript.

Table and Figure Captions

Table 1. A list of the available environmental parameters tested in model development of the response variable *Pseudo-nitzschia* spp. abundance (in bold).

Parameter	Relevant Abbreviations	Units
Latitude	Lat	deg
Longitude	Long	deg
Month		***
Chlorophyll-*a*	Chl*a*	$\mu g\ L^{-1}$
Temperature	Temp	deg C
Salinity	Sal	psu
Freshwater Discharge	Dis	$ft^3\ sec^{-1}$
Nitrate	NO_3	$mg\ L^{-1}$
Nitrite	NO_2	$mg\ L^{-1}$
Ammonium	NH_4	$mg\ L^{-1}$
Orthophosphate	PO_4	$mg\ L^{-1}$
Silicic Acid	$Si(OH)_4$	$mg\ L^{-1}$
Nitrate+Nitrite:Phosphate	N:P	$mg\ L^{-1}$ / $mg\ L^{-1}$
Silicid Acid: Nitrate	Si:N	$mg\ L^{-1}$ / $mg\ L^{-1}$
Silicic Acid: Phosphate	Si:P	$mg\ L^{-1}$ / $mg\ L^{-1}$
Dissolved Oxygen	O_2	$mg\ L^{-1}$
Dissolved Organic Carbon	DOC	$mg\ L^{-1}$
Particulate Carbon		$mg\ L^{-1}$
Total Organic Carbon		$mg\ L^{-1}$
Total Phosphorous		$mg\ L^{-1}$
Total Suspended Solids		$mg\ L^{-1}$
Secchi Depth		m
***Pseudo-nitzschia* spp. Abundance**	*Pseudo-nitzschia* Abundance	cells L^{-1}

Table 2. *Pseudo-nitzschia* bloom statistics for each monitoring station shown in Fig. 1.

STATION	**LATITUDE**	**LONGITUDE**	**MEAN (cells mL^{-1})**	**ST DEV**	**MAX**	**N**	**#NON-BLOOMS**	**#BLOOMS**
CB2.1	39.44	-76.03	0	0	0	278	278	0
CB5.1	38.32	-76.29	93	489	5691	386	332	54
CB6.1	37.59	-76.16	440	1882	19000	125	80	45
CB6.4	37.24	-76.21	207	644	6670	229	167	62
CB7.3E	37.23	-76.05	178	536	4330	240	176	64
CB7.4	36.99	-76.02	176	746	10576	523	191	332
EE3.0	38.28	-76.01	364	877	4373	49	32	17
EE3.3	38.2	-76	474	1148	6254	49	33	16
ET3.1	38.58	-76.06	0	0	0	73	73	0
ET5.2	38.52	-76.06	11	102	1749	361	351	10
ET6.2	38.3	-75.89	24	117	795	48	45	3
ET7.1	38.27	-75.79	6	41	265	42	41	1
ET8.1	38.14	-75.81	286	1060	7155	49	37	12
LE1.1	39.42	-76.60	62	540	9699	426	402	24
LE1.3	38.34	-76.49	70	487	7950	426	387	39
LE2.2	38.16	-76.6	33	461	9699	394	368	26
RET1.1	38.49	-76.66	27	202	3127	419	404	15
RET2.1	38.40	-77.27	3	34	542	317	314	3
RET2.2	38.35	-77.20	1	19	318	320	322	2
RET2.4	38.36	-76.99	7	73	1217	314	309	5
TF1.5	38.71	-76.70	0	0	0	395	395	0
TF2.1	38.71	-77.05	0	0	0	319	319	0
TF2.2	38.69	-77.11	0	0	0	318	318	0
TF2.3	38.61	-77.17	0	0	0	328	328	0
TF2.4	38.53	-77.26	0	0	0	324	324	0
WT3.1	39.30	-76.40	0	0	0	42	42	0
WT5.1	39.21	-76.52	3	34	530	374	371	3
WT6.1	39.08	-76.51	0	0	0	28	28	0
WT7.1	39.01	-76.55	0	0	0	28	28	0
WT8.1	38.95	-76.55	0	0	0	27	27	0

Table 3. Correlation coefficients (r) for select variables; correlations were calculated for raw values before detrending. Significant correlations at the □ = 0.05 level are shown in bold for DF > 6,000.

	P-n Abund	Chl-*a*	Temp	Sal	NO_3	NH_4	PO_4	$Si(OH)_4$	N:P	Si:P	Si:N	DOC	Discharge
P-n Abund.	****	-0.001	**-0.10**	**0.12**	**-0.07**	-0.05	**-0.10**	**-0.08**	-0.04	**0.06**	-0.02	0.01	-0.03
Chl-*a*		****	**0.17**	**-0.10**	**-0.20**	-0.17	**-0.06**	**-0.10**	0.004	-0.04	**0.04**	**0.21**	-0.02
Temp			****	-0.05	**-0.20**	**-0.24**	**0.21**	-0.05	**-0.25**	**-0.07**	**0.25**	**0.22**	**-0.17**
Sal				****	**-0.71**	**-0.26**	**-0.31**	**-0.18**	**-0.36**	**0.22**	**0.21**	**-0.28**	**-0.24**
NO_3					****	**0.42**	**0.11**	**0.40**	**0.43**	-0.04	**-0.28**	**-0.15**	**0.25**
NH_4						****	**0.20**	**0.22**	**0.08**	**-0.11**	**-0.17**	**-0.10**	**0.14**
PO_4							****	**0.24**	**-0.20**	**-0.25**	-0.05	**0.15**	0.002
$Si(OH)_4$								****	-0.03	**0.37**	**0.16**	-0.01	**0.08**
N:P									****	0.03	**-0.18**	**-0.08**	**0.19**
Si:P										****	**0.16**	**-0.13**	-0.03
Si:N											****	**0.06**	-0.03
DOC												***	**-0.08**
Discharge													***

Table 4. Model Coefficients and Analysis of Deviance/Likelihood Ratio Test with raw data (N = 5,121) after hypothesis testing; chi-square distribution ANOVA.

Predictor	Coefficients	Standard Error	p-val (Chi)
Intercept	-1.84	0.643	
PO_4	-69.5	13.6	<<0.001
Salinity	0.101	0.015	<<0.001
Temp	-0.063	0.011	<<0.001
DOC	0.306	0.054	<<0.001
$Si(OH)_4$	-0.373	0.096	<<0.001
Month	-0.079	0.025	0.002
N:P	-0.753	0.301	0.008

Table 5. Contingency table for the outcomes shown in Fig. 5 based on a default bloom threshold of 50% verses an optimized bloom probability threshold of 10%. Chosen performance metrics are the Heidke Skill Score (HSS), Probability of Detection (POD), False Alarm Ratio (FAR), and the Probability of False Detection (POFD); see text for definitions. Cross-validation performance metrics represent total model performance after resampling.

	HSS	**POD**	**FAR**	**POFD**	**Nagelkerke's R^2**
Absent if $P_{bloom} < 0.50$	0.01	0.01	0.83	0.003	0.10
Absent if $P_{bloom} < 0.10$	0.30	0.69	0.76	0.10	0.10
Cross-Validation $P_{bloom} = 0.10$	0.31	0.69	0.76	0.10	0.10

List of Figures:

Figure 1. Map of the Chesapeake Bay and its tributaries; dark circles denote the 30 stations for which data was retrieved from the MD DNR and EPA Chesapeake Bay Program monitoring projects from 1985-2007 (n= 6,988).

Figure 2. Time series of *Pseudo-nitzschia* spp. abundance (cells mL^{-1}) for the stations shown in Fig. 1. Note the large baywide bloom in 1998.

Figure 3. *Pseudo-nitzschia* abundance at select stations across a strong salinity gradient (oligohaline = 0.5-5 ppt; mesohaline = 5-18 ppt; oligohaline > 18 ppt) in the mainstem Chesapeake Bay, demonstrating the general increase in bloom magnitude with salinity.

Figure 4. Plots of the relationship between *Pseudo-nitzschia* abundance and the most significantly correlated physical and chemical variables from Table 3: temperature (C), salinity (ppt), NO_3 (mg L^{-1}), PO_4 (mg L^{-1}), and $Si(OH)_4$ (mg L^{-1}). Cell abundance is generally highest from 5-27 °C, 5-28 ppt, and at relatively low nutrient concentrations. Nutrient scales indicate a highly-eutrophied system.

Figure 5. Model performance using all bloom and non-bloom observations plotted against modeled bloom probabilities, where the threshold for a “bloom” is 100 cells mL^{-1}. Horizontal lines are median probabilities; boxes are 25th and 75th percentiles and whiskers are 5th and 95th percentiles. Individual open circles beyond the whiskers represent probabilities outside that range, i.e. outliers.

Figure 6. Optimization of the probability threshold for determining a “bloom” forecast where the Heidke Skill Score (HSS) is optimized with respect to the Probability of

Detection (POD), False Alarm Ratio (FAR), and the Probability of False Detection (POFD). The vertical line indicates the optimized threshold of 10% (see Table 5).

Figure 7. Monthly means of model-predicted probabilities (colorbar) and observations (open circles) of *Pseudo-nitzscha* blooms and *Pseudo-nitzschia* abundance, respectively, for three different time points in the Chesapeake Bay over the study period: a) November 1993, b) March 1998, and c) August 2005. For (a) the left panel represents model predictions using the full GLM that incorporates DOC (7) an the right panel illustrates the difference in predictive outcome when the non-DOC GLM (8) is employed.

References

Adolf J.E., Bachvaroff T., Place A.R.(2008) Cryptophyte abundance drives blooms of mixotrophic harmful aglae: a hypothesis based on *Karlodinium veneficum* as a model system. Harmful Algae 8: 3-13.

Agresti A., (2007) An Introduction to Categorical Data Analysis. Wiley-Interscience,John Wiley and Sons, Inc.

Allen J.I., Smyth T.J., Siddorn J.R., Holt M. (2008) How well can we forecast high biomass algal events in a eutrophic coastal sea? Harmful Algae 8: 70-76.

Anderson C.R., Brzezinski M.A., Washburn L., Kudela R. (2006) Circulation and environmental conditions during a toxigenic *Pseudo-nitzschia australis* bloom in the Santa Barbara Channel, California. Mar Ecol Prog Ser 327: 119-133.

Anderson C.R., Seigel D.A., Kudela R., Brzezinski M.A. (2009) Empirical models of toxigenic *Pseudo-nitzschia* blooms: Potential use as a remote detection tool in the Santa Barbara Channel. Harmful Algae 8: 478-492.

Anderson D.M., Burkholder J.M., Cochlan W.P., Glibert P.M., Gobler C.J., Heil C.A., Kudela R.M., Parsons M.L., Rensel J.E.J., Townsend D.W., Trainer V.L., Vargo G.A. (2008) Harmful algal blooms and eutrophication: Examining linkages from selected coastal regions of the United States. Harmful Algae 8: 39-53.

Anderson D.M., Glibert P.M., Burkholder J.M. (2002) Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. Estuaries 25 (4B): 704-726.

Bargu S, Marinovic B, Mansergh S, Silver MW (2003) Feeding responses of krill to the toxin-producing diatom *Pseudo-nitzschia*. J Exp Mar Biol Ecol 284: 87-104.

Bates S.S., Gaudet J., Kaczmarska I., Ehrman J.M. (2004) Interaction between bacteria and the domoic-acid producing diatom *Pseudo-nitzschia multiseries* (Hasle) Hasle: can bacteria produce domoic acid autonomously? Harmful Algae 3: 11-20.

Bates S.S., Trainer V.L (2006) The ecology of harmful diatoms. In: Graneli, E.,Turner, J.T.(Ed.), Ecology of Harmful Algae. Springer-Verlag, Berlin, p. 440.

Bidigare R.R., Marra J., Dickey T.D., Iturriaga R., Baker K.S., Smith RC, Pak H. (1990) Evidence for phytoplankton succession and chromatic adaptation in the Sargasso Sea during spring 1985. Mar Ecol Prog Ser 60: 113-122.

Blauw A.N., Anderson P., Estrada M., Johansen M., Laanemets J., Peperzak L., Purdie

D., Raine R., Vahtra E. (2006) The use of fuzzy logic models for data analysis and modelling of European harmful algal bloomos: results of the HABES project. Afr J Mar Sci 28 (2): 365-369.

Blum I., Subba Rao D.V., Pan Y., Swaminathan S., Adams N.G. (2006) Development of statistical models for prediction of the neurotoxin domoic acid levels in the pennate diatom *Pseudo-nitzschia pungens f. multiseries* utilizing data from cultures and natural blooms. In: Rao, D.V.S. (Ed.), Algal Cultures: Analogues of Blooms and Applications. Science Publishers, Enfield, NH, pp. 891-930.

Bode A., Castro C.G., Doval M.D., Varela M. (2002) New and regenerated production and ammonium regeneration in teh western Bransfield Strait region (Antarctica) during phytoplankton bloom conditions in summer. Deep-Sea Res Part II 49: 787-804.

Brush G.A., Davis F.W. (1984) Stratigraphic evidence of human disturbance in an estuary. Quaternary Research 22: 91-108.

Buskey E.J., Montagna P.A., Amos A.F., Whitledge T.E.Q. (1997) Disruption of grazer populations as a contributing factor to the intitiation of the Texas brown tide algal bloom. Limnol Oceanogr 42 (5, part 2): 1215-1222.

Chai F., Dugdale R.C., Peng T.H., Wilkerson F.P., Barber R.T (2002) One-dimensional ecosystem model of the equatorial Pacific upwelling system. Part I: model development and silicon and nitrogen cycle. Deep-Sea Res Part II: Topical Studies in Oceanography 49 (13-14): 2713-2745.

Chao Y., Li Z., Farrara J.D., Moline M., Schofield O.M.E., Majumdar S.J. (2008) Synergistic applications of autonomous underwater vehicles and regional ocean modeling system in coastal ocean forecasting. Limnol Oceanogr 53 (5): 2251–2263.

Cochlan W.P., Herndon J., Kudela, R.M. (2008) Inorganic and organic nitrogen uptake by the toxigenic diatom *Pseudo-nitzschia australis* (Bacillariophyceae). Harmful Algae 8: 111-118.

Constantin de Magny G., Long W., Brown C.W., Hood R.R., Huq A., Murtugudde R., Colwell R.R. (2009) Predicting the distribution of *Vibrio cholerae* in the Chesapeake Bay, in prep. To be submitted to the *Interdisciplinary Society of Microbial Ecology Journal.*

Constantin de Magny G., Murtugudde R., Sapiano M.R.P., Nizam A., Brown C.W., Busalacchi A.J., Yunus M., Nair G.B., Gil A.I., Lanata C.F., Calkins J., Manna B., Rajendran K., Bhattacharya M.K., Hug A., Sack R.B., Colwell R.R. (2008) Environmental signatures associated with cholera epidemics. PNAS 105 (46): 17676-17681.

Cooper S.R., Brush G.S. (1991) Long-term history of Chesapeake Bay anoxia. Science 15 (5034): 992-996.

Christian J.M., Verschell M.A., Murtugudde R., Busalacchi A., McClain C. (2002) Biogeochemical modeling of the tropical Pacific Ocean I: Seasonal and interannual variability. Deep-Sea Res Part II 49: 509-543.

Davidson K., Fehling J. (2006) Modelling the influence of silicon and phosphorus limitation on the growth and toxicity of Pseudo-nitzschia seriata. Afr J Mar Sci 28 (2): 357-360.

Davis R.E. (1976) Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. Journal of Physical Oceanography 6: 249-266.

Decker M.B., Brown C.W., Hood R.R., Purcell J.E., Gross T.F., Matanoski J.C., Bannon R.O., Setzler-Hamilton E.M. (2007) Predicting the distribution of the scyphomedusa *Chrysaora quinquecirrha* in Chesapeake Bay. Mar Ecol Prog Ser 329: 99-113.

Doering P.H., Oviatt C.A., McKenna J.H., Reed L.W. (1994) Mixing behavior of dissolved organic carbon and its potential biologial significance in the Pawcatuk River Estuary. Estuaries 17 (3): 521-536.

Dortch Q., Parsons M.L., Doucette G.J., Fryxell G.A., Maier A., Thessen A., Powell C.L., Soniat TM (2000) *Pseudo-nitzschia* spp. in the northern Gulf of Mexico: Overview and response to increasing eutrophication., Symposium on Harmful Marine Algae, Marine Biological Laboratory, Woods Hole, Massachusetts, p. p. 27.

Dortch Q., Robichaux R., Pool S., Milsted D., Mire G., Rabalais N.N., Soniat T.M., Fryxell G.A., Turner J.E. Parsons M.L. (1997) Abundance and vertical flux of *Pseudo-nitzschia* in the northern Gulf of Mexico. Mar Ecol Prog Ser 146: 249-264.

Dyble J., Bienfang P., Dusek E., Hitchcock G., Holland F., Laws E., Lerczak J., McGillicuddy D., Minnett P., Moore S., O'Kelly C., Solo-Gabriele H., Wang J. (2008) Environmental controls, oceanography and population dynamics of pathogens and harmful algal blooms: connecting sources to human exposure. Environ Health 7 (Suppl 2): S5.

Erdner D.L., Dyble J., Parsons M.L., Stevens R.C., Hubbard K.A., Wrabel M.L., Moore S.K., Lefebvre K., Anderson D.M., Bienfang P., Bidigare R.R., Parker M.S., Moeller P.D., Brand L.E., Trainer V.L. (2008) Centers for oceans and human health: a unified approach to the challenge of harmful algal blooms. Environ Health 7 (Suppl. 2).

Evan, K.M., Kuhn S.F., Hayes P.K. (2005) High levels of genetic diversity and low levels of genetic differentiation in North Sea *Pseudo-nitzschia pungens* (Bacillariophyceae) populations. J Phycol 41 (3): 506-514.

Fehling J., Davidson K., Bolch C.J., Bates S.S. (2004) Growth and domoic acid production by *Pseudo-nitzschia seriata* (Bacillariophyceae) under phosphate and silicate limitation J Phycol 40 (4): 674-683.

Fennel K., Wilkin J., Levin J., Moisan J., O'Reilly J., Haidvogel D. (2006) Nitrogen cycling in the Middle Atlantic Bight: Results from a three-dimensional model and implications for the North Atlantic nitrogen budget. Global Geochemical Cycles 20: GB3007.

Fox J. (2002) An R and S-Plus Companion to Applied Regression. Sage Publications, Inc.,Thousand Oaks, London, New Dehli.

Fritz L., Qulliam M.A., Wright J.L.C., Beale A.M., Work T.M. (1992) An outbreak of domoic acid poisoning attributed to the pennate diatom *Pseudonitzschia australis*. J Phycol 28: 439-442.

Garcia-Mendoza E., Rivas D., Olivos-Ortiz A., Almazan-Becerril A., Castaneda-Vega C., Pena-Manjarrez J.L. (2009) A toxic *Pseudo-nitzschia* bloom in Todos Santos Bay, northwestern Baja California, Mexico. Harmful Algae 8: 493-503.

Glibert P.M., Magnien R., Lomas M.W., Alexander J., Fan C., Haramoto E., Trice M., Kana T.M. (2001) Harmful algal blooms in the Chesapeake and Coastal Bays of Maryland, USA: comparisons of 1997, 1998, and 1999 events. Estuaries 24: 875-883.

Glibert P.M., Seitzinger S., Heil C.A., Burkholder J.M., Parrow M.W., Codispoti L.A., Kelly V. (2005) The role of eutrophication in the global proliferation of harmful algal blooms: new perspectives and new approaches. Oceanography 18 (2): 198-209.

Glibert, P.M., Icarus Allen, J., Bouwman, A.F., Brown, C.W., Flynn, K.J., Lewitus, A.J., Madden, C.J. (2009) Modeling of HABs and Eutrophication: Status, Advances, Challenges, this issue.

Grattan L.M., Roberts S., Trainer V., Boushey C., Burbacher T., Grant K., Tracy K., Morris J.G. (2007) Domoic acid neurotoxicity in native americans in the Pacific Northwest: Human health project methods and update. Fourth Symposium on Harmful Algae in the U.S., Woods Hole, MA.

Hallegraeff G.M. (1993) A review of harmful algal blooms and their apparent global increase. Phycologia 32 (2): 79-99.

Hasle G.R. (1978) The inverted-microscope method. In: Sournia, A. (Ed.), Phytoplankton Manual. UNESCO, Paris, pp. 88-96.

Hasle G.R., Syvertsen E.E. (1996) Marine Diatoms. In: Tomas, C.R. (Ed.), Identifying marine phytoplankton. Academic Press, Inc., New York, pp. 5-385.

Heidke P. (1926) Berechnung des Erfolges und der Gote der Windstarkevorhersagen im Sturmwarnungsdienst. Geografiska Annaler 8: 301349.

Heisler J., Glibert P.M., Burkholder J.M., Anderson D.M., Cochlan W., Dennison W.C., Dortch Q., Gobler C.J., Heil C.A., Humphries E., Lewitus A., Magnien R., Marshall H.G., Sellner K., Stockwell D.A., Stoecker D.K., Suddleson M. (2008.) Eutrophication and harmful algal blooms: A scientific consensus. Harmful Algae 8: 3-13.

Howard M.D.A., Cochlan W.P., Ladizinksy N., Kudela R.M .(2007) Nitrogenous preference of toxigenic *Pseudo-nitzschia australis* (Bacillariophyceae) from field and laboratory experiments. Harmful Algae 6: 206-217.

Iglesias-Rodriguez M.D., Brown C.W., Doney S.C., Kleypas J., Kolber D., Kolber Z., Hayes P.K., Falkowski P.G. (2002) Representing key phytoplankton functional groups in ocean carbon cycle models; coccolithophorids. Global Biogeochem Cycles 16 (4): 1100.

Kemp W.M., Boynton W.R., Adolf J.E., Boesch D.F., Boicourt W.C., Brush G., Cornwell J.C., Fisher T.R., Glibert P.M., Hagy J.D., Harding L.W., Houde E.D., Kimmel D.G., Miller W.D., Newell R.I.E., Roman M.R., Smith E.M., Stevenson J.C. (2005) Eutrophication of Chesapeake Bay: historical trends and ecological interactions. Mar Ecol Prog Ser 303: 1-29.

Kudela R., Cochlan W., Roberts A. (2004) Spatial and temporal patterns of Pseudo nitzschia spp. in central California related regional oceanography. In: Steidenger, K.A., J.H. Landsberg, C.R. Tomas, and G.A. Vargo (Ed.), Harmful Algae 2002. Florida and Wildlife Conservation Commission, Florida Institute of Oceanography, and Intergovernmental Oceanographic Commission of UNESCO.

Kudela R., Pitcher G., Probyn T., Figueiras F., Moita T., Trainer V.L. (2005) Harmful algal blooms in coastal upwelling systems. Oceanography 18 (2): 185-197.

Kudela R.M., Lane J.Q., Cochlan W.P. (2008) The potential role of anthropogenically derived nitrogen in the growth of harmful algae in California, USA. Harmful Algae 8: 103-110.

Lane J.Q., Raimondi P.T., Kudela R.M. (2009) Development of a logistic regression model for the prediction of toxigenic *Pseudo-nitzschia* blooms in Monterey Bay, California. Mar Ecol Prog Ser 383: 37-51.

Lefebvre K.A., Tilton S.C., BammlerT.K., Beyer R.P., Srinouanprachan S., Stapleton P.L., Farin F.M., Gallagher E.P. (2007) Gene expression pProfiles in zebrafish brain after acute exposure to domoicaAcid at symptomatic and asymptomatic doses. Toxicological Sciences 107 (1): 65-77.

Levin E.D., Pang W.G., Harrison J., Williams P., Petro A., Ramsdell J.S. (2006) Persistent neurobehavioral effects of early postnatal domoic acid exposure in rats. Neurotoxicology and Teratology 28: 673-680.

Lomas M.W., Rumbley C.J., Glibert P.M. (2000) Ammonium release by nitrogen sufficient diatoms in response to rapid increases in irradiance. J Plankton Res 22 (12): 2351-2366.

Maldonado M.T., Hughes M.P., Rue E.L., Wells M.L. (2002) The effect of Fe and Cu on growth and domoic acid production by *Pseudo-nitzschia multiseries* and *Pseudo-nitzschia australis*. Limnol Oceanogr 47 (2): 515-526.

Marchetti A., Trainer V.L., Harrison P.J. (2004) Environmental conditions and phytoplankton dynamics associated with *Pseudo-nitzschia* abundance and domoic acid in the Juan de Fuca eddy. Mar Ecol Prog Ser 281: 1-12.

Margalef R. (1978) Life-forms of phytoplankton as survival alternatives in an unstable environment. Oceanol. Acta 1: 493-509.

Marshall H.G. (1980) Seasonal phytoplankton composition in the lower Chesapeake Bay and Old Plantation Creek, Cape Charles, Virginia. Estuaries 3: 207-216. Marshall HG (1994) Chesapeake Bay phytoplankton. I. Composition. Proc Biol 107: 573-585.

Marshall H.G., Burchardt L., Lacouture R. (2005) A review of phytoplanktn composition within Chesapeake Bay and its tidal estuaries. J Plankton Res 27 (11): 1083-1102.

McCullagh P., Nelder J.A. (1989) Generalized Linear Models. Chapman and Hall, London.

McGillicuddy Jr D.J., Anderson D.M., Lynch D.R., Townsend D.W. (2005) Mechanisms regulating large-scale seasonal fluctuations in *Alexandrium fundyense* populations in the Gulf of Maine: Results from a physical-biological model. Deep-Sea Res Part II 52: 2698-2714.

McGillicuddy Jr D.J., Stock C.A., Anderson D.M., Signell R.P. (2003) Hindcasting blooms of the toxic dinoflagellate *Alexandrium* spp. in the western Gulf of Maine. In: Valette Silver, S.a. (Ed.), Ecological Forecasting Report, pp. 49-54.

McKenna J.H. (20040 DOC dynamics in a small temperatue estuary: Simultaneous addition and removal processes and implications on observed nonconservative behavior. Estuaries 27 (4): 604-616.

Meehl G.A., Stocker T.F., Collins W.D., Friedlingstein P., Gaye A.T., Gregory J.M., Kitoh A., Knutti R., Murphy J.M., Noda A., Raper S.C.B., Watterson I.G., Weaver A.J., Zhao Z.C. (2007) Global Climate Projections. In: Solomon, S., Quin, D., Manning, M., Marquis, M., Avery, K., Tignor, M.M.B., Miller, H.L.J., Chen, Z. (Eds.), Climate change 2007: the physical science basis. Cambridge Univ. Press, Cambridge, UK and New York.

Mengelt C. (2006) How two species of the diatom genus *Pseudo-nitzschia* respond to adverse conditions: *P. australis* and *P. multiseries* UV-photoecology, dark survival, and seasonal abundance at two coastal sites in Central California. Ph.D., University of California, Santa Barbara, Santa Barbara.

Moore J.K., Doney S.C., Kleypas J.A., Glover D.M., Fung I.Y. (2002) An intermediate complexity marine ecosystem model for the global domain. Deep-Sea Res Part II 49: 403-462.

Moore S.K., Trainer V.L., Mantua N.J., Parker M.S., Laws E., Backer L.C., Fleming L.E. (2008) Impacts of climate variability and future climate change on harmful algal blooms and human health. Environ Health 7 (Suppl. 2).

Murtugudde R. (2009) Regional Earth System prediction: A decision-making tool for sustainability? Curr O Environ Sust 1: 37-45, doi: 10.1016/j.cosust.2009.07.004.

Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. Biometrika 78: 691-692.

Orsini L., Procaccini G., Sarno D., Montresor M. (2004) Multiple rDNA ITS-types within the diatom *Pseudo-nitzschia delicatissima* (Bacillariophysceae) and their relative abundances across a spring bloom in the Gulf of Naples. Mar Ecol Prog Ser 271: 87-98.

Paerl H.W. (1988) Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. Limnol Oceanogr 33 (4, part 2): 823-847.

Pan Y., Bates S.S., Cembella A.D. (1998) Environmental stress and domoic acid production by *Pseudo-nitzschia*: a physiological perspective. Natural Toxins 6: 127-135.

Pan Y., Parsons M.L., Busman M., Moeller P.D., Dortch Q., Powell C.L., Doucette G.J. (2001) *Pseudo-nitzschia* sp. cf. *pseudodelicatissima*- a confirmed producer of domoic acid from the northern Gulf of Mexico. Mar Ecol Prog Ser 220: 83-92.

Pan Y., Subba Rao D.V., Mann K.H., Brown R.G., Pocklington R. (1996) Effects of silicate limitation on production of domoic acid, a neurotoxin, by the diatom *Pseudo-nitzschia multiseries*. I. Batch culture studies. Mar Ecol Prog Ser 131 (2-3): 225-233.

Pan Y., Subba Rao D.V., Mann K.H., Li W.K.W., Warnock R.E. (1993) Temperature dependence of growth and carbon assimilation in *Nitzschia pungens* f. *multiseries*, the causative diatom of domoic acid poisoning. In: Smayda, T.J., Shimizu, Y. (Eds.), Toxic Phytoplankton Blooms in the Sea. Elsevier Science Publishers B.V., pp. 619-624.

Parsons M.L., Dortch Q. (2002) Sedimentological evidence of an increase in *Pseudo nitzschia* (Bacillariophycea) abundance in response to coastal eutrophication. Limnol Oceanogr 47 (2): 551-558.

Patz J.A., Vavrus S.J., Uejio C.K., McLellan S.L. (2008) Climate change and waterborne disease risk in the Great Lakes Region of the U.S. Am J Prev Med 35 (5): 451-458.

Prasad M.B.K., Sapiano M.R.P., Long W., Anderson C.R., Murtugudde R. (2009) Long-term variability of nutrients in the Chesapeake Bay: A Retrospective Analysis, 1990-2006. to be submitted to Estuaries and Coasts on 1 June 2009.

Program E.C.B. (2007) Technical Support Document for Phytoplankton Monitoring Data. In: Agency, U.E.P. (Ed.), Washington, D.C.

Ramsdell J.S., Zabka T.S. (2008) In utero domoic acid toxicity: A fetal basis to adult disease in the California sea lion (*Zalophus californianus*). Marine Drugs 6 (2): 262-290.

Rue E.L., Bruland K.W. (2001) Domoic acid binds iron and copper: a possible role for the toxin produced by the marine diatom *Pseudo-nitzschia*. Mar Chemistry 76: 127-134.

Scholin C.A., Gulland F., Doucette G.J., Benson S., Busman M., Chavez F.P., Cordaro J., DeLong R., De Vogelaere A., Harvey J., Haulena M., Lefebvre K.A., Lipscomb T., Loscutoff S., Lowenstine L.J., Marin III R., Miller P.E., McLellan W.A., Moeller P.D.R., Powell C.L., Rowles T., Silvagni P., Silver M.W., Spraker T., Trainer V.L., Van Dolah F.M. (2000) Mortality of sea lions along the central California coast linked to a toxic diatom bloom. Nature 403: 80-83.

Siegel D.A., Dickey T.D. (1986) Variability of net longwave radiation over the Eastern North Pacific Ocean. J Geophys Res C 91 (C6): 7657-7666.

Smayda T.J. (1963) Succession of phytoplankton, and the ocean as an holocoenotic environment. In: Oppenheimer, C.H. (Ed.), Symposium on marine microbiology. Thomas, Springfield, pp. 260-274.

Smayda T.J. (1990) Novel and nuisance phytoplankton blooms in the sea: evidence for a global epidemic In: Graneli, E.e.a. (Ed.), Toxic Marine Phytoplankton: Fourth International Conference, Lund, Sweden, June 26-30, 1989. Elsevier Science Publishing Co., Inc., New York, New York, pp. 29-40.

Smayda T.J. (1997) Harmful algal blooms: Their ecophysiology and general relevance to phytoplankton blooms in the sea. Limnol Oceanogr 42 (5, part 2): 1137-1153.

Sohn K.T., Park S.M. (2008) Guidance on the Choice of Threshold for Binary Forecast Modeling. Advances in Atmospheric Sciences 25 (1); 8388.

Sommer U. (1994) Are marine diatoms favoured by high Si:N ratios? Marine Ecology Progress Series 115: 309-315.

Song Y., Haidvogel D.B. (1994) A semi-implicit ocean circulation model using a generalized topography-following coordinate system. Journal of Computational Physics 115 (1): 228-244.

Stewart J.E.(2008) Bacterial involvlement in determining domoic acid levels in *Pseudo nitzschia multiseries* cultures. Aquatic Microbial Ecology 50 (2): 135-144.

Stumpf R.P., Culver M.E., Tester P.A., Tomlinson M., Kirkpatrick G.J., Pederson B.A., Truby E., Ransibrahmanakul V., Soracco M. (2003) Monitoring *Karenia brevis* blooms in the Gulf of Mexico using satellite ocean color imagery and other data. Harmful Algae 2: 147-160.

Thessen A.E., Bowers H.A., Stoecker D.K. (2009) Intra- and interspecies differences in growth and toxicity of *Pseudo-nitzsch*ia while usin different nitrogen sources. Harmful Algae 8: 792-810.

Thessen A.E., Dortch Q., Parsons M.L., Morrison W. (2005) Effect of salinity on *Pseudo nitzschia* species (Bacillariphyceae) growth distribution. J Phycol 41: 21-29.

Thessen A.E., Stoecker D.K. (2008) Distribution, abundance, and domoic acid analysis of the toxic diatom genus *Pseudo-nitzschia* in the Chesapeake Bay. Estuaries and Coasts 31 (4): 664-672.

Tomlinson M.C, Stumpf R.P., Ransibrahmanakul V., Truby E.W., Kirkpatrick G.J., Pederson B.A., Vargo G.A., Heil C.A. (2004) Evaluation of the use of SeaWiFS imagery for detecting *Karenia brevis* harmful algal blooms in the eastern Gulf of Mexico. Remote Sensing of Environment 91 (3-4): 293-303.

Trainer V.L., Adams N.G., Bill B.D., Stehr C.M., Wekell J.C., Moeller P.D.R., Busman M, Woodruff (2000) Domoic acid production near California coastal upwelling zones, June 1998. Limnol Oceanogr 45 (8): 1818-1833.

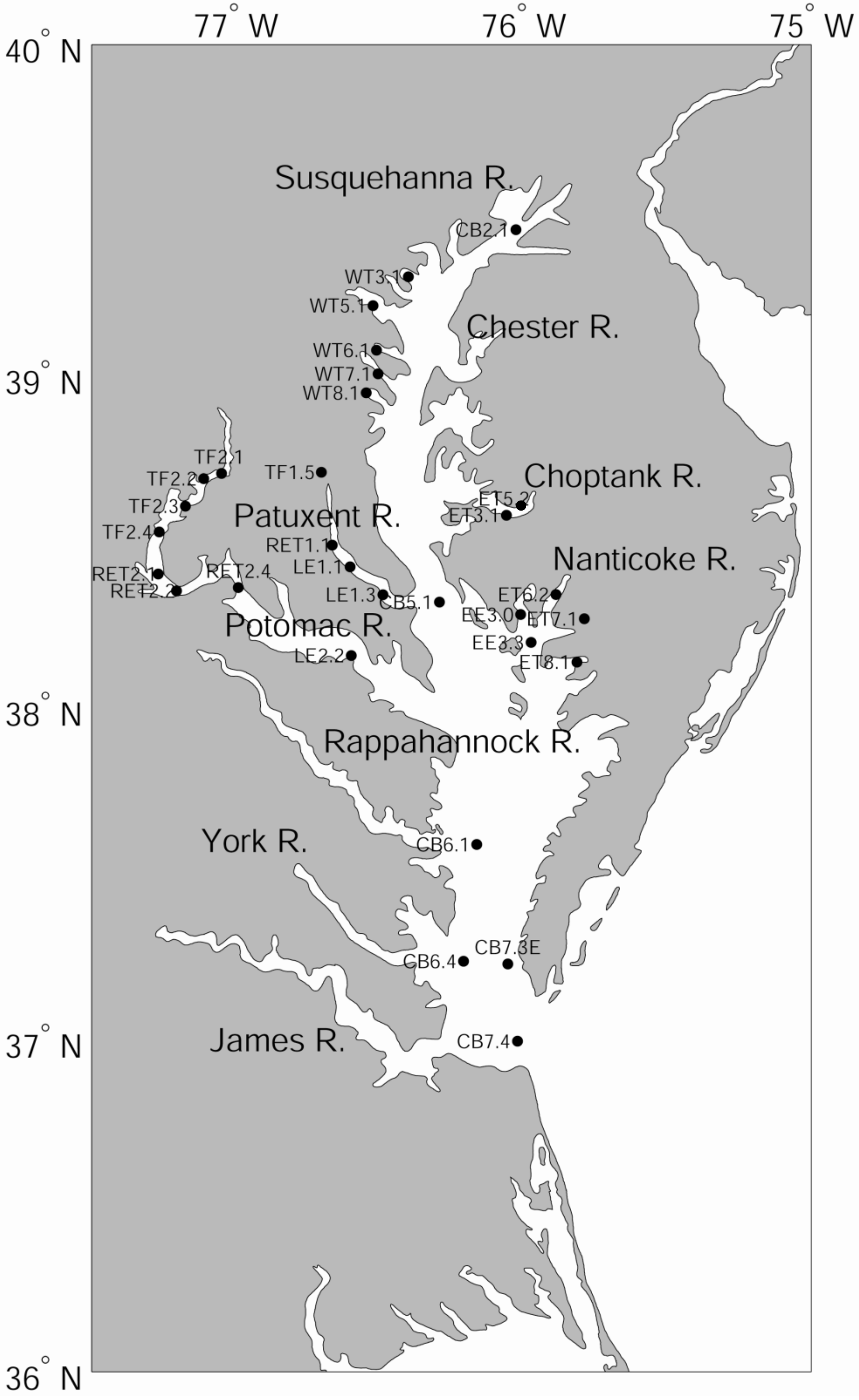
VanDolah F.M. (2000) Marine algal toxins: origins, health effects, and their increased occurrence. Environ Health Perspectives 108 (Suppl 1): 133-141.

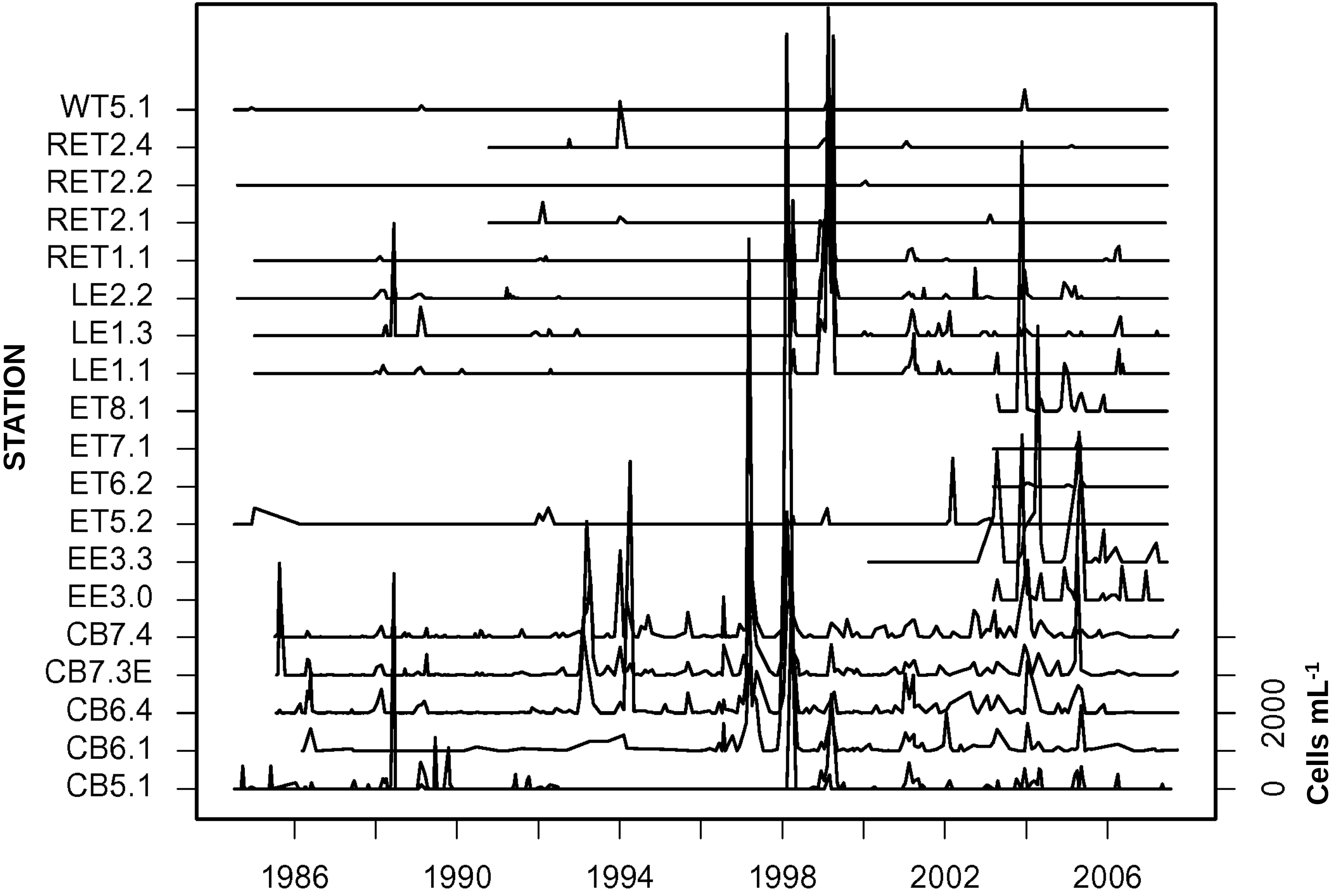
Wang X., Murtugudde R., Busalacchi A.J. (2009) Enhanced biological activity in the western and central equatorial Pacific since the late 1990s. Geophys. Res. Letters, in press.

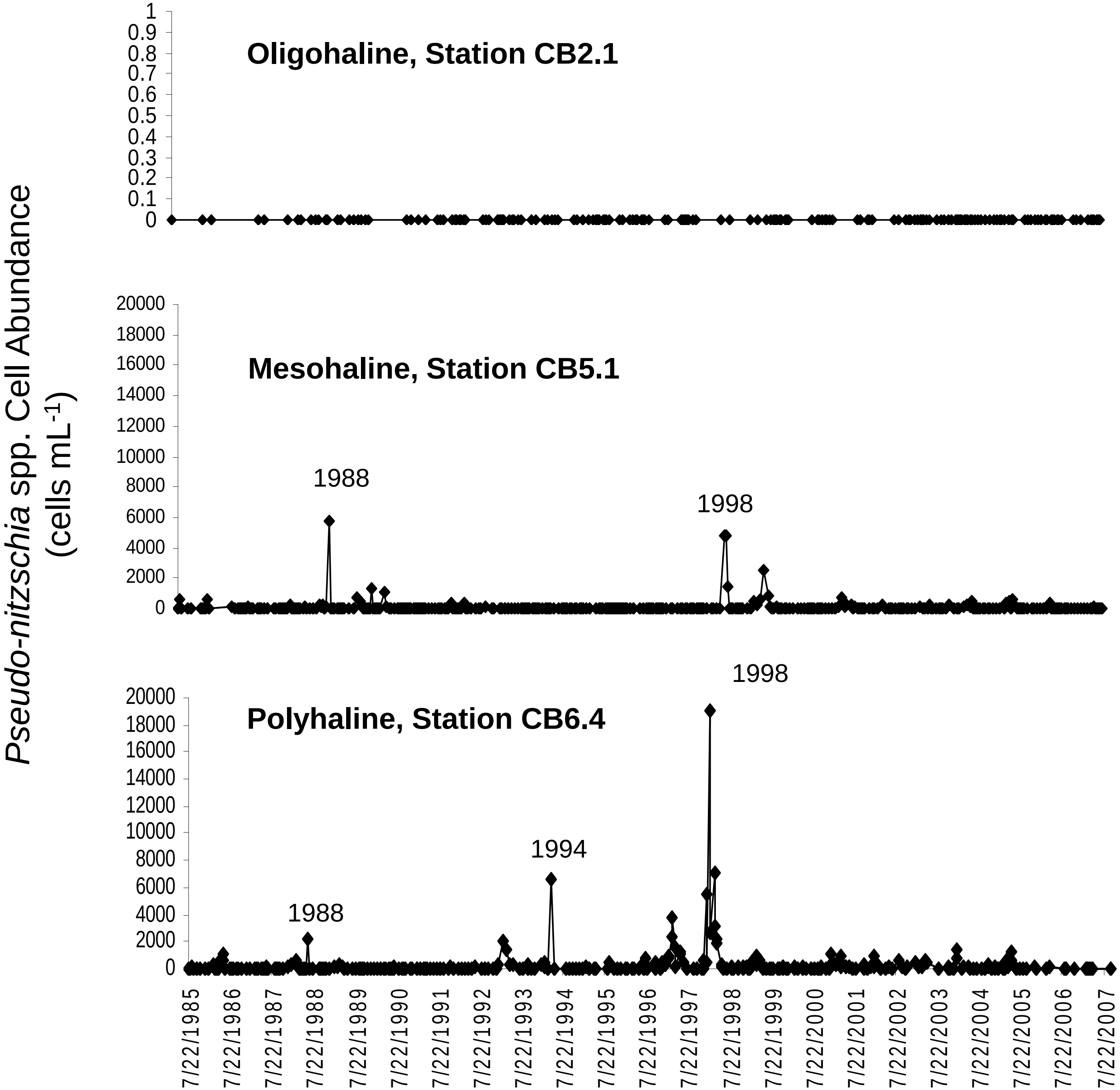
Wells M.L., Trick C.Q., Cochlan W.P., Hughues M.P., Trainer V.L. (2005) Domoic acid the synergy of iron, copper, and the toxicity of diatoms. Limnol Oceanogr 50 (6): 1908-1917.

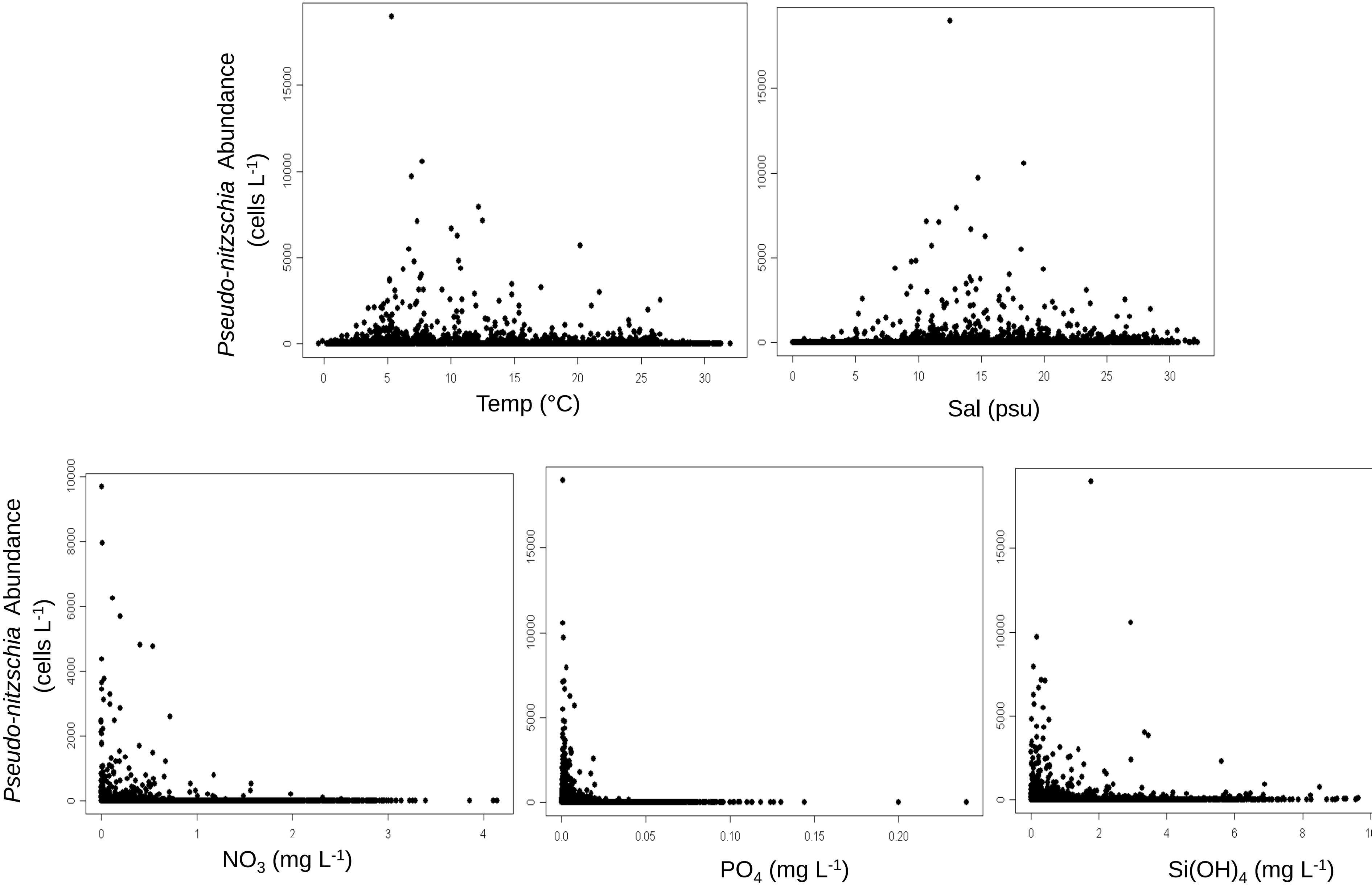
Westberry T., Siegel D.A. (2005) An improved bio-optical model for the remote sensing of *Trichodesmium* spp. blooms. J Geophys Res C 110 (C06012).

Xu J., Hood R.R. (2006) Modeling biogeochemical cycles in Chesapeake Bay with a coupled physical-biological model. Est Coast Shelf Sci 69(1-2): 19-46.


77° W
76° W
75° W
40° N
39° N
38° N
37° N
36° N
Susquehanna R.
CB2.1
WT3.1
WT5.1
Chester R.
WT6.1
WT7.1
WT8.1
TF2.1
TF2.2
TF1.5
Choptank R.
TF2.3
ET5.2
ET3.1
Patuxent R.
TF2.4
RET1.1
Nanticoke R.
LE1.1
RET2.1
RET2.4
ET6.2
RET2.2
LE1.3
CB5.1
EE3.0
ET7.1
Potomac R.
EE3.3
LE2.2
ET8.1
Rappahannock R.
CB6.1
York R.
CB7.3E
CB6.4
James R.
CB7.4


STATION
WT5.1
RET2.4
RET2.2
RET2.1
RET1.1
LE2.2
LE1.3
LE1.1
ET8.1
ET7.1
ET6.2
ET5.2
EE3.3
EE3.0
CB7.4
CB7.3E
CB6.4
CB6.1
CB5.1
1986
1990
1994
1998
2002
2006
0
2000
Cells mL^{-1}


Pseudo-nitzschia spp. Cell Abundance (cells mL^{-1})
Oligohaline, Station CB2.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Mesohaline, Station CB5.1
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
1988
1998
Polyhaline, Station CB6.4
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
1988
1994
1998
7/22/1985
7/22/1986
7/22/1987
7/22/1988
7/22/1989
7/22/1990
7/22/1991
7/22/1992
7/22/1993
7/22/1994
7/22/1995
7/22/1996
7/22/1997
7/22/1998
7/22/1999
7/22/2000
7/22/2001
7/22/2002
7/22/2003
7/22/2004
7/22/2005
7/22/2006
7/22/2007


Pseudo-nitzschia Abundance (cells L^{-1})
0
5000
10000
15000
0
5
10
15
20
25
30
Temp (°C)
Sal (psu)
0
2000
4000
6000
8000
10000
0
1
2
3
4
NO_3 (mg L^{-1})
0.0
0.05
0.10
0.15
0.20
PO_4 (mg L^{-1})
0
2
4
6
8
$Si(OH)_4$ (mg L^{-1})

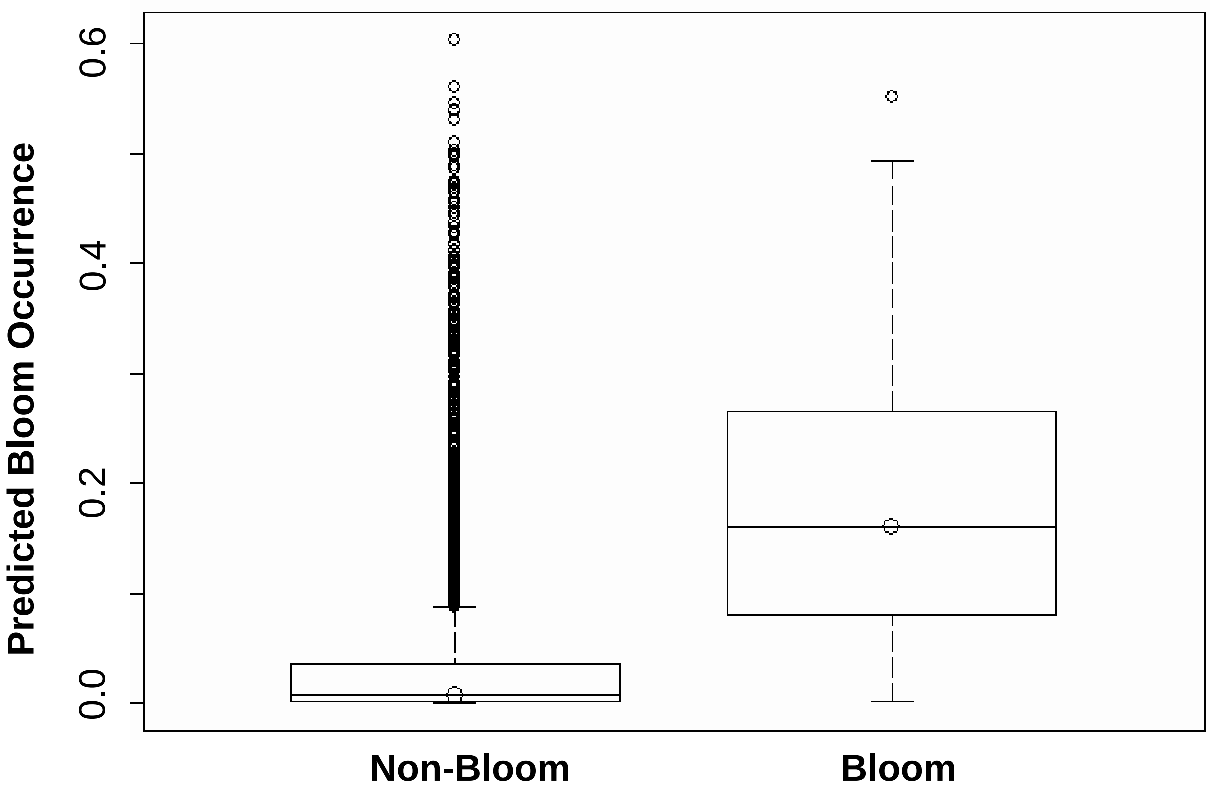
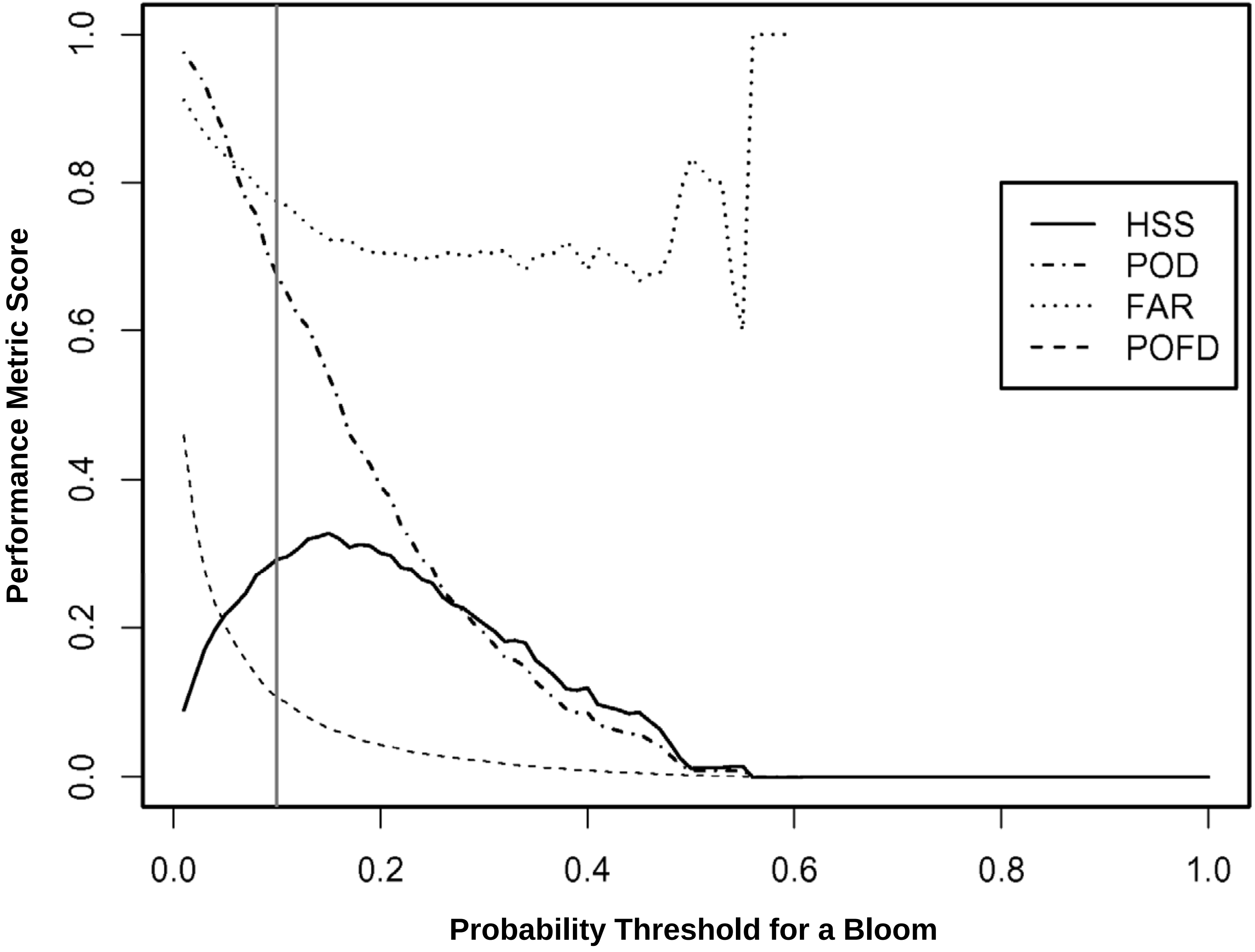
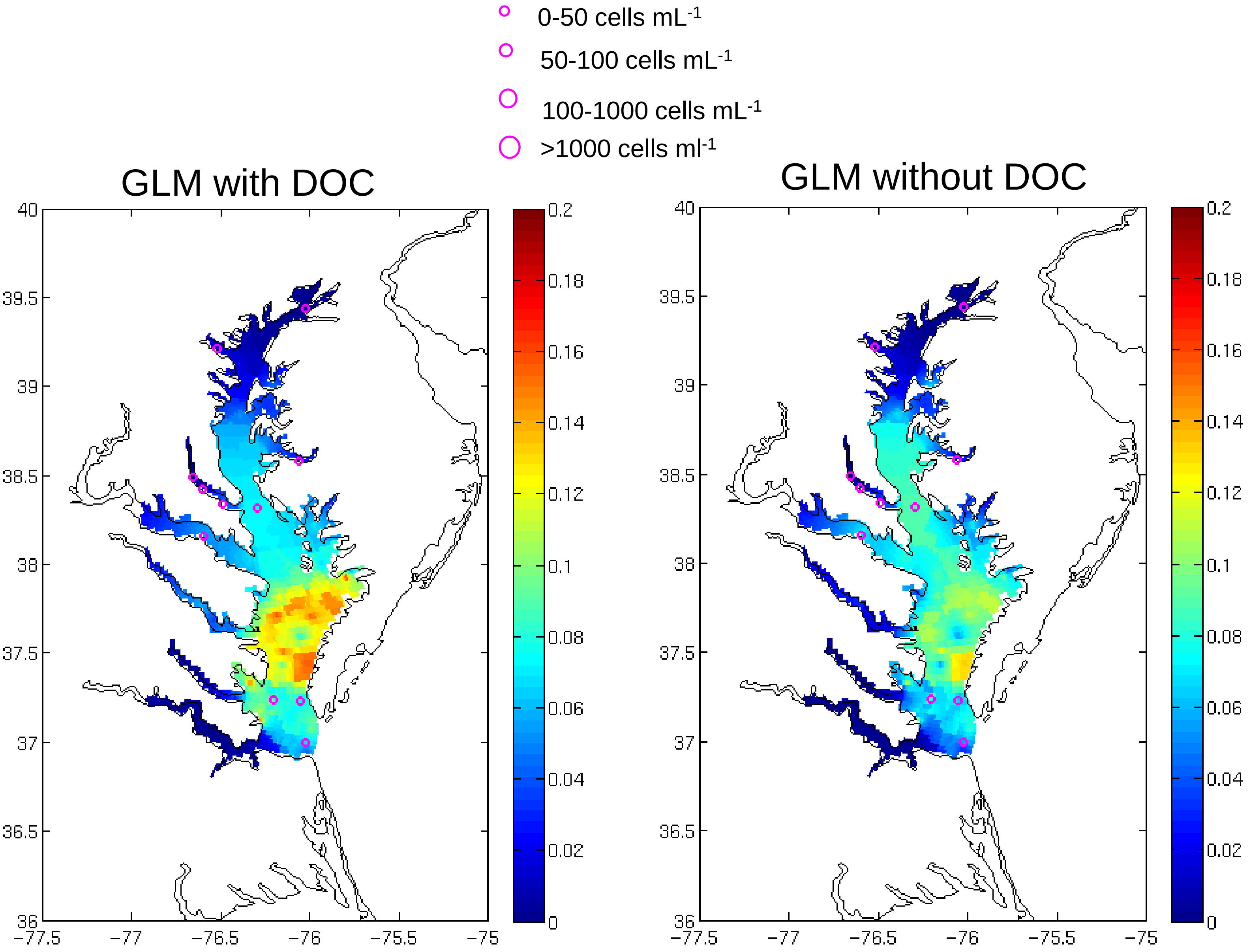

Predicted Bloom Occurrence
0.0
0.2
0.4
0.6
Non-Bloom
Bloom

FIGURE 6

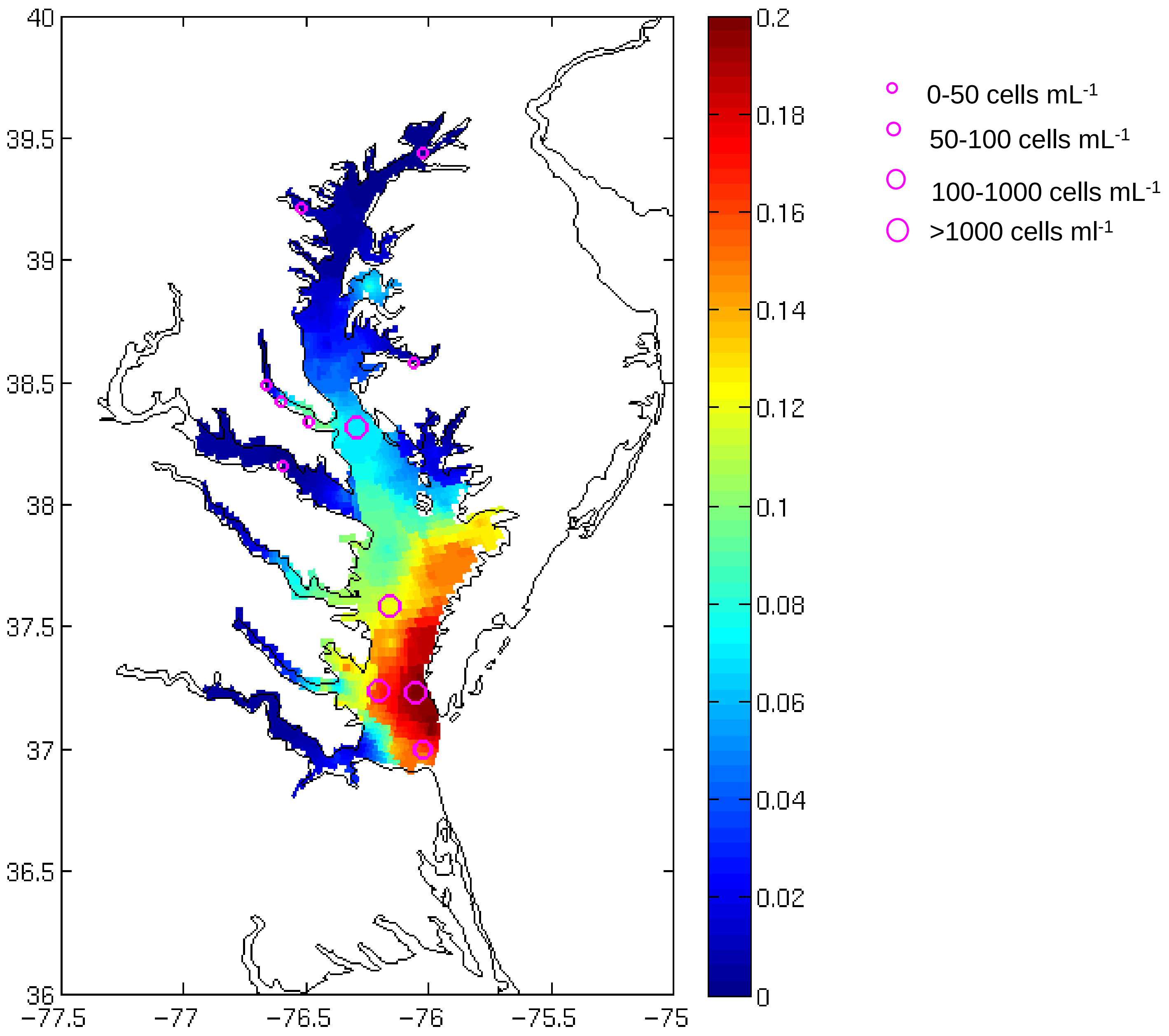


Performance Metric Score
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Probability Threshold for a Bloom
HSS
POD
FAR
POFD

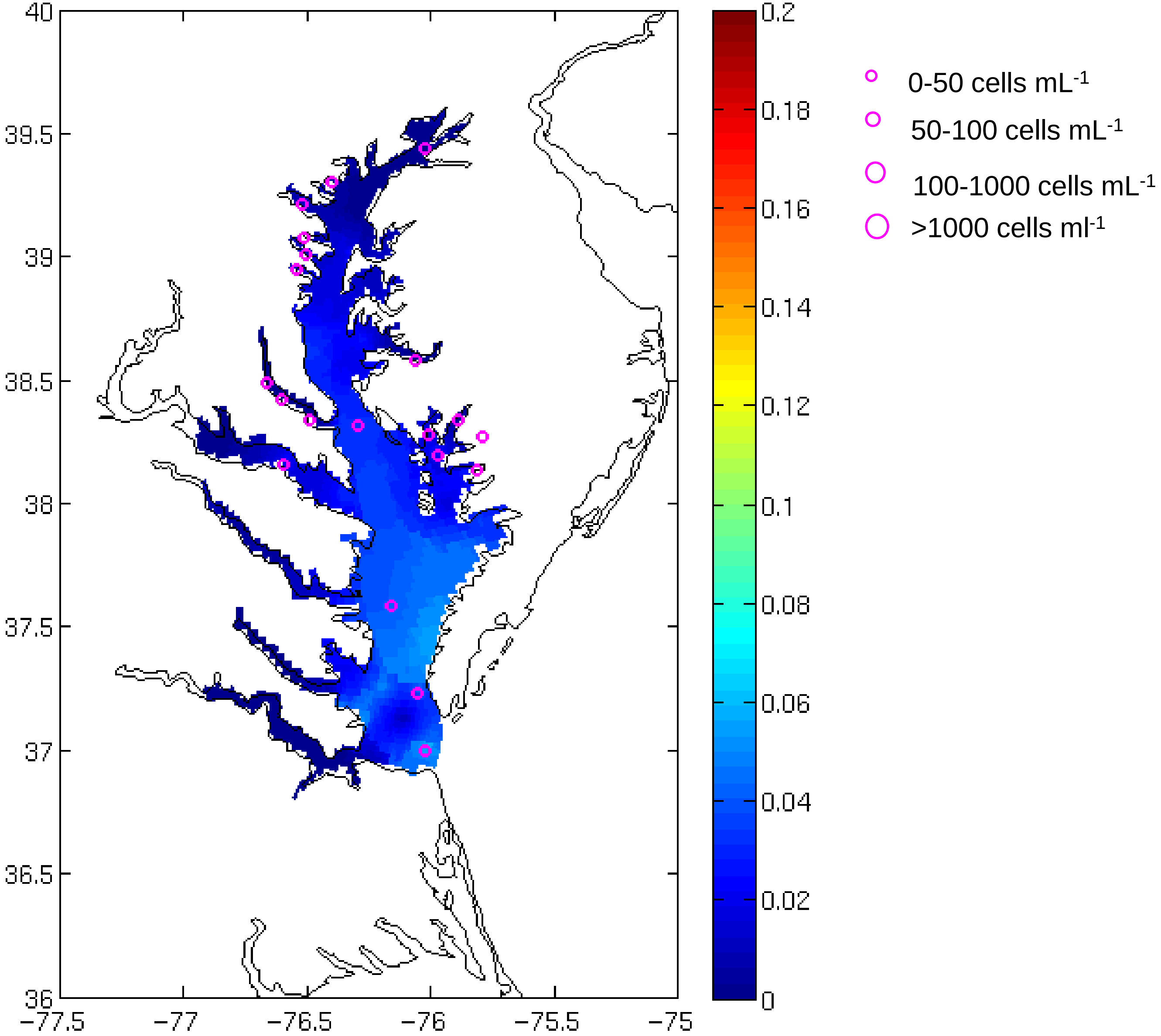
A)


0-50 cells mL^{-1}
50-100 cells mL^{-1}
100-1000 cells mL^{-1}
>1000 cells ml^{-1}
GLM with DOC
GLM without DOC
40
39.5
39
38.5
38
37.5
37
36.5
36
−77.5
−77
−76.5
−76
−75.5
−75
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0

B)


40
39.5
39
38.5
38
37.5
37
36.5
36
-77.5
-77
-76.5
-76
-75.5
-75
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0-50 cells mL^{-1}
50-100 cells mL^{-1}
100-1000 cells mL^{-1}
>1000 cells ml^{-1}

C)


40
39.5
39
38.5
38
37.5
37
36.5
36
−77.5
−77
−76.5
−76
−75.5
−75
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
○ 0-50 cells mL-1
○ 50-100 cells mL-1
○ 100-1000 cells mL-1
○ >1000 cells ml-1