Development of a logistic regression model for the prediction of toxigenic *Pseudo-nitzschia* blooms in Monterey Bay, California.

Running head: Predictive modeling of toxigenic *Pseudo-nitzschia* blooms

Jenny Q. Lane[1,*] , Peter T. Raimondi[2], Raphael M. Kudela[1]

[1]Department of Ocean Sciences, University of California, 1156 High Street, Santa Cruz, California 95064, USA

[2]Department of Ecology and Evolutionary Biology, University of California, Center for Ocean Health, 100 Shaffer Road, Santa Cruz, California 95064, USA

ABSTRACT: Blooms of the diatom genus *Pseudo-nitzschia* have been recognized as a public health issue in California since 1991 when domoic acid, the neurotoxin produced by toxigenic species of *Pseudo-nitzschia*, was first detected in local shellfish. Although these blooms are recurring and recognized hazards, the factors driving bloom proliferation remain poorly understood. Contributing to this, and hindering the development of robust forecasting tools, has been a lack of long-term field studies and/or deficiencies in the scope of environmental data included with field observations. For this study, we successfully developed predictive logistic models of toxigenic *Pseudo-nitzschia* blooms in Monterey Bay, California from a multi-project dataset representing 8.3 years of sampling effort. Models were developed for year-round use (Annual model) or for seasonal use (Spring model and Fall-Winter model). Chlorophyll a and silicic acid emerged as significant variables common to all models, while model-specific inclusions of temperature, upwelling index, river discharge, and/or nitrate provided further model refinement. Sensitivity (specificity) of the Annual, Spring, and Fall-Winter models was 77% (78%), 75% (75%), and 89% (89%) respectively. Predictive power for 'unknown' bloom cases (i.e. future blooms) was demonstrated at ≥75% for all models. Our models out-performed chlorophyll a anomaly models applied to our dataset and performed comparably to, or better than, previously described statistical models for *Pseudo-nitzschia* blooms or domoic acid. The models presented here are the first to have been developed from long-term (>1.5y) monitoring efforts, and the first to have been developed for bloom prediction of toxigenic *Pseudo-nitzschia* species.

Keywords: *Pseudo-nitzschia* · predictive model · logistic regression · harmful algal bloom · phytoplankton monitoring · domoic acid

*Email: jquay@ucsc.edu

INTRODUCTION

Harmful algal blooms are a multi-faceted threat in that they can have severe deleterious consequences for local industry (e.g. shellfish, tourism), public health, and ecosystem health. In addition, the incidence of harmful algal blooms appears to be increasing in both frequency and intensity (Hallegraeff 1993, Anderson et al. 2002, Glibert et al. 2005). This trend, and its potential to inflict rising economic and societal costs, have motivated numerous monitoring and forecasting efforts in recent years (Schofield et al. 1999, Johnsen & Sakshuag 2000, Fisher et al. 2003). Since the first confirmed report of domoic acid poisoning on the U.S. west coast in 1989, blooms of toxigenic *Pseudo-nitzschia* have occurred in Monterey Bay, California with regularity and with harmful effects (Buck et al. 1992, Fritz et al. 1992, Work et al. 1993, Scholin et al. 2000, Trainer et al. 2001), but there has been little to no development of predictive models. Researchers have urged long-term monitoring approaches with *Pseudo-nitzschia* in particular (Trainer 2000), recognizing the prevalence of short-term and inconsistent research efforts within the bloom research literature and the consequential unlikelihood of robust model development. The generation of the models presented here was, indeed, enabled by the use of datasets generated by long-term monitoring projects.

Here we develop logistic regression (LOGIT) models of toxigenic *Pseudo-nitzschia* blooms in Monterey Bay, California. This modeling exercise was undertaken with a three-fold purpose: (1) develop *Pseudo-nitzschia* bloom models that are straight-forward and useful in their application towards bloom monitoring, (2) through model development, identify environmental variables that are significant factors in bloom

incidence, and (3) test the recurrence of these significant environmental variables in previous *Pseudo-nitzschia* models described by Anderson et al. (in press) and Blum et al. (2006) for other geographic regions. The previous modeling efforts and our efforts are not wholly consistent with each other in terms of scope, evaluated variables, or specific aim: Anderson et al. (in press) developed models of 'generic' *Pseudo-nitzschia* blooms (i.e. blooms of toxigenic and non-toxic species) and domoic acid concentration (cellular and particulate) from a 1.5-year dataset collected from the Santa Barbara Channel, while Blum et al. (2006) developed models of particulate domoic acid (only) from a mixture of experimental and field data. These previous models and our efforts clearly differ in their region of interest and specific model subject. Here, these disparities are an advantage, in that they allow for inter-model comparison capable of identifying factors that are likely to be universally significant to *Pseudo-nitzschia* bloom incidence and to the introduction of domoic acid into the marine environment through bloom proliferation.

In this study, we present three logistic regression models of toxigenic *Pseudo-nitzschia* blooms in Monterey Bay, California as they occur throughout the year (Annual model) and seasonally (Spring and Fall-Winter models). Thirty-one environmental variables were evaluated, and a total of six variables were identified as statistically significant for bloom prediction. Prediction success is assessed and reported according to the maximization of both non-bloom and bloom case prediction and the minimization of false negative responses (the rate at which blooms elude prediction).

METHODS

Compilation of the model dataset. We compiled a dataset from published literature that included *Pseudo-nitzschia* cell counts for Monterey Bay, including MS and PhD theses (Walz 1995, Villac 1996, Goldberg 2003) and journal articles (Buck et al. 1992, Walz et al. 1994, Scholin et al. 2000). Additional unpublished datasets were provided by Moss Landing Marine Laboratories (MLML), and internally generated through the Center for Integrated Marine Technologies (CIMT) and through the California Program for Regional Enhanced Monitoring for PhycoToxins (Cal-PReEMPT). Details on sampling and analytical methods for internally generated datasets are provided.

We obtained 2099 discrete cases from the above sources, 1156 of which were from surface waters (depth ≤ 5m). All of the data were assessed to ensure methodological consistency. The dataset extracted from Walz (1995) was removed entirely due to a disparity in sample collection (i.e. environmental data and cell counts had not been collected concurrently nor from within suitably close spatial proximity). Of the remaining 1071 cases, 576 contained cell counts of toxigenic *Pseudo-nitzschia*.

Model development required concurrent macronutrient, chlorophyll a, and temperature variables in combination to achieve a model with 'very good' accuracy, i.e. a model with ROC area > 0.8 (Table 1; see 'Results' section for more on ROC area). Inclusion of cases for model development was therefore based on the following: (1) sample collection from Monterey Bay surface waters, and (2) toxigenic *Pseudo-nitzschia* cell counts (*P. multiseries* plus *P. australis*) with concurrent environmental measurements of seawater temperature, chlorophyll a, and macronutrients.

Internal data: sample collection. Samples were collected monthly June 2002 – November 2007 from eleven stations throughout Monterey Bay as part of the CIMT project. Ten-liter PVC Niskin bottles (refitted with silicone rubber band strings) mounted on an instrumented rosette were used to collect water from near-surface depth (5m). Surface samples were collected from two stations by PVC bucket. Temperature data was obtained from a Seabird SBE-19 CTD deployed concurrently with water sampling.

Shore-based samples were collected weekly May 2005 – April 2008 from the Santa Cruz Municipal Wharf (36° 57.48´ N, 122° 1.02´ W) as part of the Cal-PReEMPT project. Shore-based samples were collected from the surface with a PVC bucket or were mixtures of water samples collected from three discrete depths (0m, 1.5m, 3m) with a FieldMaster 1.75L basic water bottle. Temperature was measured in the field by digital hand thermometer immediately following sample retrieval.

River discharge rates for the Salinas, San Lorenzo, Soquel and Pajaro Rivers were obtained from the United States Geological Survey National Water Information System (http://waterdata.usgs.gov/nwis/). Bakun daily upwelling index values for the Monterey Bay region (36°N, 122° W) were obtained from the National Oceanographic and Atmospheric Administration Pacific Environmental Research Division (http://www.pfeg.noaa.gov/products/PFEL/).

Internal data: analytical methods. Samples for chlorophyll a were collected in duplicate and filtered onto uncombusted glass-fiber filters (Whatman GF/F; nominal pore size 0.7μm) and processed using the non-acidification method (Welschmeyer 1994). Macronutrients [Nitrate plus nitrite (hereafter referred to as nitrate), silicic acid

and *ortho*-phosphate] were stored frozen prior to analysis with a Lachat Quick Chem 8000 Flow Injection Analysis system using standard colorimetric techniques (Knepel & Bogren 2001, Smith & Bogren 2001a, b). *Pseudo-nitzschia* species identification and enumeration utilized species-specific large subunit rRNA-targeted probes following standard protocols (Miller & Scholin 1998). Samples were enumerated with a Zeiss Standard 18 compound microscope equipped with a fluorescence Illuminator 100 (Zeiss, Thornwood, NY, USA). Duplicate filters were prepared for each species and the entire surface area of each filter was counted.

Model development. Models were developed by logistic regression modeling (LOGIT) using MYSTAT Version 12.02.11. LOGIT model development is appropriate when the dependent variable (the variable being predicted) is dichotomous (e.g. 0 versus 1), and since they serve to classify samples into one of two groups, LOGIT models are commonly referred to as classification models. Our goal was to develop a LOGIT model for the prediction of toxigenic *Pseudo-nitzschia* blooms; our dependent variable was therefore non-bloom/bloom. Since our dataset contained continuous data for *Pseudo-nitzschia* abundance, LOGIT modeling required concatenation into a new dichotomous dependent variable (*bloom_nobloom*) using a bloom threshold of 10,000 toxigenic *Pseudo-nitzschia* cells l^{-1}. For the sake of simplicity, binary values (0 or 1) were used to represent non-bloom and bloom conditions, with ≥10,000 cells l^{-1} considered a bloom (binary value of 1).

Independent variables evaluated during model development are provided in Table 2. The subset of predictor variables that produced the most efficient and accurate

159 model was identified using an automatic stepwise approach (forward, backward, and
160 bidirectional). Variable selections were refined according to optimization of sensitivity
161 (sensitivity = % of blooms that are successfully predicted), minimization of false
162 response (false response = % of non-bloom predictions that are incorrect), and
163 maximization of ROC area, while controlling for covariance among the independent
164 variables. Covariance diagnostics included Variance Inflation Factors and Condition
165 Indices. Variables exhibiting severe covariation were considered mutually exclusive.
166 Only significant variables ($p < 0.05$) were included in the final models. Once the subset
167 of predictor variables had been finalized for each LOGIT model, predictive power with
168 unknown (future) samples was assessed for each model using a script (supplied by
169 SYSTAT) for jackknife cross-validation of LOGIT models. For the development of the
170 two seasonal models, the data was partitioned according to the oceanic periods
171 previously described for Monterey Bay (Pennington & Chavez 2000), while the entire
172 dataset was used for development of the Annual model. The final three models have
173 been developed for application: (1) year-round (Annual model), (2) February 14 - June
174 30 (Spring model), and (3) July 1 - February 13 (Fall-Winter model).
175
176 **RESULTS**
177 After removing those cases from the original (2099) dataset that did not fulfill the
178 quality criteria, 506 samples from 2002-2008 remained, 74 of which were classified as a
179 bloom (toxigenic *Pseudo-nitzschia* ≥ 10,000 cells l^{-1}), There was clear seasonality in
180 these data: the rate of bloom incidence was 28% during the Spring model time period,

compared to a rate of 9% for the remainder of the year (i.e. the Fall-Winter model time period).

All LOGIT models have the form:

$$\text{LOGIT}(p) = \ln[p/(1-p)] = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_k z_k$$

where p is the probability of a given condition and LOGIT(p) is the natural log of that probability. In our case, p represents the probability that a given sample will be a bloom sample. β_0 is a constant, and β_1, β_2, … β_k are the regression coefficients of z_1, z_2, … z_k, respectively. The LOGIT models (Annual, Spring, and Fall-Winter) are presented in Table 3. The regression curve for the Spring model is also presented (Figure 1).

The Annual model, Spring model and Fall-Winter model all had sufficient ROC curve area to qualify them as having 'very good' to 'excellent' model fit accuracy (Table 4). As a plot of sensitivity versus 1-specificity (specificity = the % of non-blooms that are successfully predicted), the ROC demonstrates how severely specificity is decreased as adjustments are made to increase sensitivity. Receiver Operating Characteristic curve area is therefore a 0-to-1 scaled measure of model fit, and is assessed by a scale similar to that of a traditional academic point system (<0.6 = poor; 0.6 – 0.7 = fair; 0.7 – 0.8 = good; 0.8 – 0.9 = very good; >0.9 = excellent). Other demonstrations of model accuracy and fit include log-likelihood-ratio tests, McFadden's ρ^2 and Naglekerke's R^2, for which values >0.2 are indicative of very good model fit (Hensher & Johnson 1981) (Table 4).

All of the models achieved this high level of accuracy with four or fewer predictive variables. Two predictive variables ln(*chl a*), ln(*silicic acid*) were included in all three models. The set of predictive variables used in the Annual model and Spring model

were most similar, sharing the variables ln(*chl a*), ln(*silicic acid*) and *temp* and differing only in the inclusion of *upwelling* in the Annual model. The Fall-Winter model is the most disparate of the three models with the variable set ln(*Pajaro River*), *nitrate*, ln(*silicic acid*) and ln(*chl a*). Of particular note is the omission of *temp* and *upwelling* from the Fall-Winter model, and the inclusion of a river discharge variable [ln(*Pajaro River*)] and *nitrate*.

The inclusion of non-significant variables in a model can inflate the standard error of the coefficient estimates, thereby reducing the efficiency of the model (Menard 1995). In addition, the inclusion of non-significant variables in a model reduces its applicability for general use. All variables included in the models had the greatest significance possible (p=0.000), except for *temp* (Spring model, p=0.015) and *upwelling* (Annual model, p=0.003). The inclusion of more than four variables resulted in models with intolerable levels of covariance among the independent variables (Condition Indices > 30). The Spring model was especially prone to covariance among the independent variables, likely because the Monterey Bay system is dominated by a single and specific environmental forcing (upwelling) during the Spring model time period.

Model performance at the default cut-point (0.500) and model-optimized cut-points is summarized in Table 5 (see 'Cut-point selection' for more on cut-points and cut-point optimization). The use of seasonal models significantly enhanced predictive ability at the default cut-point, allowing the sensitivity of the Spring and Fall-Winter Models to increase by 16% and 19%, respectively. Rates of false reference (false reference = % of bloom predictions that were incorrect) were 24%, 20% and 19% for the Annual, Spring and Fall-Winter models, respectively. Rates of false response (false

response = % of non-bloom predictions that were incorrect) were 9%, 14% and 4% for the Annual, Spring and Fall-Winter models, respectively. While the default cut-point was useful in model development, it was a true default value and not appropriate for optimized model performance. A side-by-side comparison of model performances at the 0.500 cut-point is useful, however, by allowing for comparison of performances under the equalizing (but unrealistic) assumption that blooms are evenly distributed throughout the year and expected to occur with as much frequency as non-blooms.

Optimized cut-points were determined by generating model prediction failure rates over the full range of cut-points at 0.05 increments (Figure 2). Bloom prediction failure is defined as the % of blooms the model fails to predict (1-sensitivity) and non-bloom prediction failure is defined as the % of non-blooms the model fails to predict (1-specificity). The overall failure rate is minimized when the failure rates of both types are simultaneously minimized and equal. The optimized cut-point values assigned to each model can be visualized graphically, therefore, as the cut-points where the failure rate curves intersect (Figure 2). At optimized cut-points, the Fall-Winter model demonstrated the highest predictive success (sensitivity = 89%, specificity = 89%), followed by the Annual model (sensitivity = 77%, specificity = 78%) and the Spring model (sensitivity = 75%, specificity = 75%). False reference rates were 62%, 46% and 55% for the Annual, Spring, and Fall-Winter models, respectively. False response rates were 5%, 11% and 1% for the Annual, Spring, and Fall-Winter models, respectively. The inflated false reference rates for the models are driven by a relatively low frequency of non-bloom predictions (which is in part a product of cut-point optimization; see ‘Cut-point selection’ for more details).

A jackknife cross-validation module suitable for use with LOGIT was supplied by SYSTAT and used to validate model performance with respect to unknown (future) cases. At each iteration, the models were developed from N-1 cases and cross-validated with the 'unknown' (omitted) case. This method is similar to the training data/validation data (i.e. bootstrap validation) approach taken by Blum et al. (2006) except that (1) it does not reduce the dataset that can be used for initial model development, and (2) it is an iterative process that allows for N instances of cross-validation against 'unknown' single cases. The SYSTAT jackknife cross-validation module was run with the model-optimized cut-points (Table 6a) and with cut-points set according to bloom prior probabilities ("priors") for Monterey Bay (2002-2005), as calculated from an independent California Department of Public Health (CDPH) *Pseudo-nitzschia* bloom monitoring dataset (Table 6b; see 'Cut-point selection' below for more details on priors and their application as cut-points). At the model-optimized cut-points, model sensitivity (specificity) was 77% (77%), 75%, (76%) and 89% (89%) for the Annual, Spring and Fall-Winter models, respectively. False responses, the error rate of most concern from a regulatory standpoint, were 5%, 11% and 1% for the Annual, Spring and Fall-Winter models, respectively. Bloom success indices, representing predictive improvement over a constant-only model, are highly favorable for both the seasonal models (Spring model 48%; Fall-Winter model 80%) and for the Annual model (61%) when applied with the model-optimized cut-points. Non-bloom success indices ranged from -7% to 3%, indicating that non-bloom predictive performance is only slightly better, or slightly worse, than that achieved by application of a constant-only model. The large discrepancies between bloom and non-bloom success

index values reflect the fact that the models were developed specifically for the prediction of bloom cases. The Pearson's chi-square-test statistic for each model indicates extreme significance in the association of modeled predictions and true outcome of unknown (future) cases (p=0.000). In Table 6b, the application of inadvertently 'low' cut-points (such as the CDPH priors), i.e. those that give advantage to sensitivity over specificity, offers, effectively, a demonstration of cut-point administration under circumstances in which future blooms occur with unexpectedly high frequency. The apparently conservative modeling response towards bloom prediction under these circumstances is in part an artifact of LOGIT modeling: LOGIT models generally guard against the misclassification of cases belonging to the under-represented case group, a quality that makes them especially attractive for application in high-risk predictive contexts such as environmental regulation and clinical health (Fan & Wang 1998). The use of CDPH priors is therefore an appropriate but conservative approach, resolving with enhanced rates of sensitivity (≥ 91% for all models) and reduced rates of false response (≤ 3% for all models), but at cost to non-bloom parameters (specificity, rates of false reference and non-bloom success indices). Similarly, inflated frequencies of bloom prediction forced a large denominator into the calculation of bloom predictive values (and vice-versa, in the calculation of non-bloom predictive values).

DISCUSSION

Cut-point selection. As with linear regression models, LOGIT models must be developed and applied while remaining mindful of the system under investigation. In

296 particular, the model should be developed and implemented with consideration of (1)
297 prior probabilities (here, the general frequency at which blooms occur), (2) the
298 prediction error rates that are inherent to the model being used, and (3) the cost of
299 prediction error to the model user. The first of these is taken into account by the
300 designation of an optimized cut-point. Consideration of (2) and (3) is left to the
301 discretion of the model user, since it is only necessary when the risk of a *specific type* of
302 predictive error, rather than *overall* predictive error, needs to be reduced.
303 When the regression equation is solved, the model user is presented with the
304 probability of a bloom occurrence. However, the degree of probability that can be
305 tolerated (the 'cut-point') must be defined. For LOGIT solutions where the probability of
306 a bloom is above the cut-point, the probability is considered high enough to warrant a
307 bloom prediction. Similarly, for LOGIT solutions where the probability of a bloom is
308 below the cut-point, the probability is considered sufficiently low to warrant a 'non-
309 bloom' prediction. Reducing the cut-point minimizes the chance of failure to predict a
310 bloom by relaxing the criteria deemed sufficient for a 'bloom' prediction and forcing the
311 prediction of more blooms overall. Figure 3 offers a schematic diagram of this
312 relationship, illustrating the trade-off between minimizing the number of blooms that the
313 model fails to predict (the number of open symbols along the '1' vertical axis) and
314 minimizing the number of non-blooms that are incorrectly predicted as blooms (the
315 number of solid symbols along the '0' horizontal vertical axis) as the cut-point is
316 adjusted.
317 The value of a model's optimized cut-point is largely influenced by the relative
318 sizes of the groups being modeled (here, bloom and non-bloom). At present, blooms of

319 *Pseudo-nitzschia* are relatively rare occurrences and the proportion of bloom samples to
320 non-bloom samples is small. Considering these proportions, the probability of a bloom
321 occurrence is generally never 50% and a default cut-point of 50% cannot provide
322 optimized predictive capability. Use of an optimized cut-point, rather than the default
323 cut-point, should be implemented whenever the probability of the outcomes being
324 modeled is significantly unequal, as is the case here (Neter et al. 1989).
325 An optimized cut-point is designated according to either (1) the optimized cut-
326 point inherent to the model, or (2) according to the prior probabilities and costs of
327 incorrect predictions within the system to which the model is being applied (Neter et al.
328 1989). If a system demonstrates priors that are overwhelmingly different from those
329 encountered in our modeling dataset (Annual model time period: 15%, Spring model
330 time period: 28%; Fall-Winter model time period: 9%), the cut-point can be reduced (or
331 inflated) to account for the more infrequent (or more frequent) occurrence, and therefore
332 likelihood, of blooms. Similarly, if the cost of a certain type of incorrect prediction is
333 disproportionately high, the cut-point should be adjusted to protect from the type of
334 predictive failure associated with that exaggerated cost. In the case of the former, the
335 cut-point should be set according to the system priors when they are available. In the
336 case of the latter, Figure 2 may be used as a guide in the appointment of a cut-point
337 that sufficiently reduces the failure rate of greatest concern.
338
339 **Inter-model comparison: model performance.** In Table 7, we present the
340 predictive performance of our annual and seasonal models in conjunction with those
341 demonstrated by (1) *Pseudo-nitzschia* bloom prediction based solely upon chlorophyll a

anomaly, (2) linear hindcasting models developed for *Pseudo-nitzschia* blooms, cellular domoic acid (cDA), and particulate domoic acid (pDA) in the Santa Barbara Channel (Anderson et al. in press), and (3) LOGIT models developed for pDA from a combination of field and experimental data (Blum et al. 2006). To assess the performance of *Pseudo-nitzschia* bloom prediction based solely upon chlorophyll a anomalies, chlorophyll a data from June 2004 – July 2008 was obtained from the LOBOVIZ website (http://www.mbari.org/lobo/loboviz.htm) for a nearshore mooring in Monterey Bay (M0) and a 30-day median chlorophyll a anomaly was calculated according to methods previously described for 60-day mean-anomalies (Tomlinson et al. 2004, Wynne et al. 2006). The LOBOVIZ website was selected as a data source due to its ease of access and applicability. A median was employed in lieu of a mean since it has been recently recognized as generally more appropriate (Stumpf 2008).

The LOGIT regression models presented here were developed from the largest dataset to-date, and demonstrated a relatively high level of predictive skill. Our models out-performed the chlorophyll a anomaly model throughout the year and on a seasonal basis, although the sensitivity of the chlorophyll a anomaly model was surprisingly comparable during the Spring model time period (50% versus 75%). Interestingly, the chlorophyll a anomaly model completely failed to predict blooms during the Fall-Winter time period (sensitivity = 0%). It should be noted, however, that the chlorophyll a anomaly model is generally applicable to all potential HABs and in particular for true "red tides" [e.g. (Kudela et al. 2008, Ryan et al. 2008)], and may therefore be a better model choice when not applied specifically for *Pseudo-nitzschia* bloom prediction.

Inter-model comparison: model variables. While the areas of interest and, in some cases, the dependent variable, differ between the models developed by Blum et al. (2006), Anderson et al. (in press), and herein, the similarities and differences shared between the models can provide insight into the global patterns of *Pseudo-nitzschia* ecophysiology. Figure 4 illustrates the significant variables that are shared and not shared between the studies, as well as the significant variables that are shared and not shared between the Annual and Seasonal models presented in this study.

In the Annual, Spring and Fall-Winter models, silicic acid exhibits a statistically significant negative regression coefficient, indicating that low silicic acid conditions are associated with blooms of toxigenic *Pseudo-nitzschia* in Monterey Bay. Nitrate exhibits a positive coefficient in the Fall-Winter model, indicating that nitrate is positively correlated with toxigenic blooms of *Pseudo-nitzschia* during the majority of the year (July 1 – February 14). The *Pseudo-nitzschia* bloom model developed by Anderson et al. (in press) indicated an association between *Pseudo-nitzschia* blooms and low silicic acid conditions, but only when nitrate concentrations were more replete. Our results are similar but more specific, suggesting that low silicic acid concentrations are associated with toxigenic blooms of *Pseudo-nitzschia* throughout the year, but that high nitrate concentrations and low silicic acid conditions are associated with *Pseudo-nitzschia* blooms especially during the Fall-Winter time period (July 1 – February 13).

Interestingly, a strong negative coefficient (-5.026) is associated with the Pajaro River discharge variable in the Fall-Winter model, suggesting that toxigenic blooms of *Pseudo-nitzschia* are associated with reduced river flow during a period of the year most predominantly associated with discharge-induced eutrophication events (Kudela et

al. in press). A time series plot of Pajaro River discharge overlain with *Pseudo-nitzschia* bloom events demonstrates that *Pseudo-nitzschia* blooms generally occur after peak river flow as discharge rates are falling, but, in the case of Fall-Winter time period blooms, when flow discharge rates are at their minimum (Figure 5) while relatively high nitrate levels are sustained. The association of high river discharge with low phytoplankton growth (presumably due to concurrently high levels of turbidity) and with dilute but annually predominant nutrient loading to the ocean has been reported, with particular attention to high nitrogen loading (in the form of urea, ammonium and/or nitrate) associated with runoff from urban and agricultural areas (Warrick et al. 2005). In an assessment of 7 major river discharges into Monterey Bay from 2001-2005, the Pajaro River, draining an area of urban expansion and intense agricultural activity, consistently demonstrated moderate levels of annual discharge and disproportionately high annual loads of nitrate (CCLEAN 2006). While the inclusion of Pajaro River discharge in the Fall-Winter model is significant and noteworthy, the inclusion of a river discharge factor in a *Pseudo-nitzschia* bloom model is not unprecedented: Anderson et al. (in press) chose not to evaluate river discharges as model variables, but presented a *Pseudo-nitzschia* bloom model ("Best Full Model") and a cDA model ("Best Full Model") that both include variables generally associated with recently significant river discharge events (particle absorption and CDOM absorption) (Warrick et al. 2004, Warrick et al. 2007). Similarly, although not through modeling exercises, others have suggested a correlation between river discharge and toxic blooms [e.g. (Zou et al. 1993, Qi et al. 1994, Subba Rao et al. 1998, Scholin et al. 2000, Parsons et al. 2002, Trainer et al. 2007)].

Bloom modeling versus toxin modeling. The monitoring of domoic acid for public health purposes is carried out continually by the CDPH and focuses, quite appropriately, on the protection of human health from domoic acid intoxication. This monitoring effort is more accurately described as the monitoring of domoic acid bioaccumulation in shellfishery products, and, specifically, domoic acid bioaccumulation in sentinel shellfish supplies (*Mytilus californianus*). Modeling or monitoring efforts which are focused on toxin load alone, while useful and appropriate for regulatory purposes, obviously do not allow for the estimation or monitoring of the toxin source itself, *Pseudo-nitzschia* blooms, which can be highly variable in their toxicity (Trainer et al. 2002, Marchetti et al. 2004, Anderson et al. 2006). Variability in bloom toxicity translates into a weak relationship between toxin bioaccumulation and toxigenic *Pseudo-nitzschia* concentration, evidenced here by CDPH/Cal-PReEMPT project data compiled from northern, central, and southern California study sites (Point Reyes, Santa Cruz, Avila Beach) over a three-year period (2005 – 2007) (Figure 6). Note that there were cases where extreme bloom concentrations of toxigenic *Pseudo-nitzschia* (>>10,000 cells/L) were associated with sub-regulatory toxin levels (<20 μg/g), but no observed cases where sub-bloom concentrations of toxigenic *Pseudo-nitzschia* (<10,000 cells/L) were associated with toxin levels approaching the regulatory limit in shellfish. LOGIT models developed for toxigenic *Pseudo-nitzschia* blooms can therefore be used for detection of both acute and sub-acute toxic bloom events, while models developed for domoic acid alone will fail to address the injection of toxin into the system via sub-acute bloom events. This is a significant failure inherent to all toxin models, since chronic or early life

stage exposure to sub-lethal levels of domoic acid are increasingly being recognized as an emerging threat to both human health and wildlife (Kreuder et al. 2005, Miller 2007, Goldstein et al. 2008, Grattan et al. 2008, Ramsdell & Zabka 2008). By providing estimations of all toxigenic *Pseudo-nitzschia* bloom events, whether low or high in toxicity, *Pseudo-nitzschia* bloom models have the unique ability to address this emerging threat. Ideally, models would be developed for both cell abundance (as we report) and for toxin production (Blum et al. 2006, Anderson et al. in press). While the domoic acid data associated with the cases used herein were insufficient for inclusion of a toxin component, a two-step model would maximize both regulatory monitoring (likely best served by bloom prediction) and better understanding of the ecophysiological conditions associated with toxin production.

CONCLUSION

All models demonstrated a toxigenic bloom classification rate of 75% or greater. These represent success rates that are comparable to, or improved over, those reported for previous models of toxicity and generic *Pseudo-nitzschia* blooms. The models consistently and significantly out-performed toxigenic *Pseudo-nitzschia* models derived from chlorophyll a anomaly alone. Although we report the largest modeling data set to date, consisting of 506 cases from 2002-2008, the removal of approximately 75% of the full (2099 cases) dataset highlights the need for more consistent data collection. We also note that several parameters identified as important [e.g. urea (Howard et al. 2007, Kudela et al. in press); lithium (Subba Rao et al. 1998); ammonium (Trainer et al. 2007); iron and copper (Rue & Bruland 2001, Maldonado et al. 2002, Wells et al. 2005)] are not

457 included as potential variables and could potentially further improve model prediction.
458 The parameters common to the three regional models developed for the U.S. west
459 coast, provide initial guidance for ongoing and future HAB monitoring efforts. All models
460 contained macronutrient variables, indicating the influence of upwelling and possibly
461 eutrophication on toxigenic *Pseudo-nitzschia* bloom proliferation. Bloom proliferation
462 was associated with low silicic acid and high chlorophyll a throughout the year and
463 seasonally. Other factors were associated with blooms depending on whether their
464 effect was considered for a specific season or over the entire year, specifically: low sea
465 surface temperatures (annual and springtime), high upwelling index (entire year), high
466 nitrate and low river discharge (summer, fall and wintertime). Seasonal significance of
467 river discharge, and, more specifically, river discharge from the Pajaro River (a river
468 heavily influenced by agricultural and urban land-use and bearing disproportionately
469 high nitrate concentrations) suggests that the influence of eutrophication on toxigenic
470 *Pseudo-nitzschia* bloom dynamics extends to both natural and cultural eutrophication
471 processes.

REFERENCES

Anderson CR, Brzezinski MA, Washburn L, Kudela R (2006) Circulation and environmental conditions during a toxigenic *Pseudo-nitzschia* bloom. Marine Ecology Progress Series 327:119-133

Anderson CR, Siegel DA, Kudela RM, Brzezinski MA (in press) Empirical models of toxigenic *Pseudo-nitzschia* blooms: potential use as a remote detection tool in the Santa Barbara Channel. Harmful Algae

Anderson DM, Glibert PM, Burkholder JM (2002) Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. Estuaries 25:704-726

Blum I, Subba Rao DV, Pan Y, Swaminathan S, Adams NG (2006) Development of statistical models for prediction of the neurotoxin domoic acid levels in the pennate diatom *Pseudonitzschia pungens f. multiseries* utilizing data from cultures and natural blooms. In: Subba Rao DV (ed) Algal cultures, analogues of blooms and applications. Science Publishers Inc. , p 891-916

Buck KR, Uttal-Cooke L, Pilskaln CH, Roelke DL, Villac MC, Fryxell GA, Cifuentes L, Chavez FP (1992) Autecology of the diatom *Pseudonitzschia austalis*, a domoic acid producer, from Monterey Bay, California. Marine Ecology Progress Series 84:293-302

CCLEAN (2006) CCLEAN Annual Report, 2004-2005, Santa Cruz, CA

Fan X, Wang L (1998) Comparing linear discriminant function with logistic regression for the two-group classification problem. Annual Meeting of the American Educational Research Association, San Diego, CA

Fisher WS, Malone TC, Giattina JD (2003) A pilot project to detect and forecast harmful algal blooms in the northern Gulf of Mexico. Environmental Monitoring and Assessment 81:373-381

Fritz L, Quilliam MA, Wright JLC, Beale AM, Work TM (1992) An outbreak of domoic acid poisoning attributed to the pennate diatom *Pseudonitzschia australis* Journal of Phycology 28:439-442

Glibert PM, Anderson DM, Gentien P, Granéli E, Sellner KG (2005) The global, complex phenomena of harmful algal blooms. Oceanography 18:136-147

Goldberg JD (2003) Domoic acid in the benthic food web of Monterey Bay, California. Master's Thesis, California State University Monterey Bay, CA

Goldstein T, Mazet JAK, Zabka TS, Langlois G, Colegrove KM, Silver M, Bargu S, Dolah FV, Leighfield T, Conrad PA, Barakos J, Williams DC, Dennison S, Haulena M, Gulland FMD (2008) Novel symptomatology and changing epidemiology of domoic acid toxicosis in California sea lions (*Zalophus californianus*): an increasing risk to marine mammal health. Proceedings of the Royal Society B 275:267-276

Grattan LM, Roberts S, Trainer V, Boushey C, Burbacher T, Grant K, Tracy K, Morris JG (2008) Domoic acid neurotoxocity in Native Americans in the Pacific Northwest: human health project methods and update. Fourth Symposium on Harmful Algae in the US Woods Hole, MA

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

Hensher DA, Johnson LW (1981) Applied discrete-choice modeling, Vol. Croom Helm, London

Howard MA, Cochlan WP, Ladizinsky N, Kudela RM (2007) Nitrogenous preference of toxigenic *Pseudo-nitzschia australis* (Bacillariophyceae) from field and laboratory experiments Harmful Algae 6:206-217

Johnsen G, Sakshuag E (2000) Monitoring of harmful algal blooms along the Norwegian coast using bio-optical methods. South African Journal of Marine Science 22:309-321

Knepel K, Bogren K (2001) Determination of orthophosphorous by flow injection analysis in seawaters: QuickChem Method 31-113-01-1-H. In: Saline Methods of Analysis. Lachat Instruments, Milwaukee, WI

Kreuder C, Miller MA, Lowenstine LJ, Conrad PA, Carpenter TE, Jessup DA, Mazet JA (2005) Evaluation of cardiac lesions and risk factors associated with myocarditis and dilated cardiomyopathy in southern sea otters (*Enhydra lutris nereis*). American Journal of Veterinary Research 66:289-299

Kudela RM, Lane JQ, Cochlan WP (in press) The potential role of anthropogenically derived nitrogen in the growth of harmful algae in California, USA Harmful Algae

Kudela RM, Ryan JP, Blakely MD, Lane JQ, Peterson TD (2008) Linking the physiology and ecology of *Cochlodinium* to better understand harmful algal bloom events: A comparative approach. Harmful Algae 7:278-292

Maldonado MT, Hughes MP, Rue EL, Wells ML (2002) The effect of Fe and Cu on growth and domoic acid production by *Pseudo-nitzschia multiseries* and *Pseudo-nitzschia australis*. Limnology & Oceanography 47:515-526

Marchetti A, Trainer VL, Harrison PJ (2004) Environmental conditions and phytoplankton dynamics associated with *Pseudo-nitzschia* abundance and domoic acid in the Juan de Fuca eddy. Marine Ecology Progress Series 281:1-12

Menard S (1995) Applied logistic regression analysis, Vol. Sage, Thousand Oaks

Miller MA (2007)

Miller PE, Scholin CA (1998) Identification and enumeration of cultured and wild *Psuedo-nitzschia* (Bacillariophyceae) using species-specific LSU rRNA-targeted fluorescent probes and filter-based whole cell hybridization. Journal of Phycology 32:371-382

Neter J, Wasserman W, Kutner M (1989) Applied linear regression models, Vol. Irwin

Parsons ML, Dortch Q, Turner RE (2002) Sedimentological evidence of an increase in *Pseudo-nitzschia* (Bacillariophyceae) abundance in response to coastal eutrophication. Limnology & Oceanography 47:551-558

Pennington JT, Chavez FP (2000) Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station H3/M1 over 1989-1996 in Monterey Bay, California. Deep-Sea Research II 47:947-973

Qi Y, Ju W, Lei Z (1994) The taxonomy and bloom ecology of *Pseudo-nitzschia* on the coasts of China. IOC-WESTPAC Third International Scientific Symposium, Bali, Indonesia, p 88-95

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

562 Rue E, Bruland K (2001) Domoic acid binds iron and copper: a possible role for the
563 toxin produced by the marine diatom *Pseudo-nitzschia*. Marine Chemistry
564 76:127-134
565 Ryan JP, Gower JFR, King SA, Bissett WP, Fischer AM, Kudela RM, Kolber Z, Mazzillo
566 F, Rienecker EV (2008) A coastal ocean extreme bloom incubator. Geophysical
567 Research Letters 35
568 Schofield O, Grzymski J, Bissett WP, Kirkpatrick GJ, Millie DF, Moline MA, Roesler CS
569 (1999) Optical monitoring and forecasting systems for harmful algal blooms:
570 possibility of pipe dream? Journal of Phycology 35:1477-1496
571 Scholin CA, Gulland F, Doucette GJ, Benson S, Busman M, Chavez FP, Cordaro J,
572 DeLong R, De Vogelaere A, Harvey J, Haulena M, Lefebvre K, Lipscomb T,
573 Loscutoff S, Lowenstine LJ, Marin III R, Miller PE, McLellan WA, Moeller PD,
574 Powell CL, Rowles T, Silvagnl P, Silver M, Spraker T, Trainer V, Van Dolah FM
575 (2000) Mortality of sea lions along the central California coast linked to a toxic
576 diatom bloom. Nature 403:80-84
577 Smith P, Bogren K (2001a) Determination of nitrate and/or nitrite in brackish or
578 seawater by flow injection analysis colorimeter: QuickChem method 31-107-04-1-
579 E. In: Saline Methods of Analysis. Lachat Instruments, Milwaukee, WI
580 Smith P, Bogren K (2001b) Determination of silicate in brackish or seawater by flow
581 injection analysis colorimeter: QuickChem Method 31-114-27-1-C. In: Saline
582 Methods of Analysis. Lachat Instruments, Milwaukee, WI
583 Stumpf RP (2008)
584 Subba Rao DVS, Pan Y, Mukhida K (1998) Production of domoic acid by *Pseudo-*
585 *nitzschia multiseries* Hasle, affected by lithium. Marine Ecology 19:31-36
586 Tomlinson MC, Stumpf RP, Ranisbrahmanakul V, Truby EW, Kirkpatrick GJ, Pederson
587 BA, Vargo GA, Heil CA (2004) Evaluation of the use of SeaWiFS imagery for
588 detecting *Karenia brevis* harmful algal blooms in the eastern Gulf of Mexico.
589 Remote Sensing of the Environment 91:293-303
590 Trainer VL, Adams NG, Wekell JC (2001) Domoic acid producing *Pseudo-nitzschia*
591 species off the U.S. west coast associated with toxification events. In: Hallegraeff
592 GM, Blackburn SI, Bolch CJ, Lewis RJ (eds) Harmful Algal Blooms 2000.
593 Intergovernmental Oceanographic Commission of UNESCO, Paris, p 46-49
594 Trainer VL, Adams, N.G., Bill, B.D., Stehr, C.M., Wekell, J.C., Moeller, P., Busman, M.,
595 and Woodruff, D. (2000) Domoic acid production near California coastal
596 upwelling zones, June 1998. Limnology & Oceanography 45:1818-1833
597 Trainer VL, Cochlan WP, Erickson A, Bill BD, Cox FH, Borchert JA, Lefebvre KA (2007)
598 Recent domoic acid closures of shellfish harvest areas in Washington State
599 inland waterways. Harmful Algae 6:449-459
600 Trainer VL, Hickey BM, Horner RA (2002) Biological and physical dynamics of domoic
601 acid production off the Washington coast. Limnology & Oceanography 47:1438-
602 1446
603 Villac MC (1996) Synecology of the genus *Pseudo-nitzschia* H. Peragallo from
604 Monterey Bay, California, USA. Texas A&M University
605 Walz PM (1995) *Pseudo-nitzschia* species and domoic acid in Monterey Bay, CA.
606 University of Santa Cruz

Walz PM, Garrison DL, Graham WM, Cattey MA, Tjeerdema RS, Silver MW (1994) Domoic acid-producing diatom blooms in the Monterey Bay, California: 1991-1993. Natural Toxins 2:271-279

Warrick JA, DiGiacomo PM, Weisberg SB, Nezlin NP, Mengel M, Jones BH, Ohlmann JC, Washburn L, Terrill EJ, Farnsworth KL (2007) River plume patterns and dynamics within the Southern California Bight. Continental Shelf Research 27:2427-2448

Warrick JA, Mertes LAK, Siegel DA, Mackenzie C (2004) Estimating suspended sediment concentrations in turbid coastal waters of the Santa Barbara Channel with SeaWiFS. International Journal of Remote Sensing 25:1995-2002

Warrick JA, Washburn L, Brzezinski MA, Siegel DA (2005) Nutrient contributions to the Santa Barbara Channel, California, from the ephemeral Santa Clara River. Estuarine, Coastal and Shelf Science 62:559-574

Wells ML, Trick CG, Cochlan WP, Hughes MP, Trainer VL (2005) Domoic acid: the synergy of iron, copper, and the toxicity of diatoms. Limnology & Oceanography 50:1908-1917

Welschmeyer NA (1994) Fluorometric analysis of chlorophyll *a* in the presence of chlorophyll *b* and pheopigments. Limnology & Oceanography 39:1985-1992

Work TM, Beale AM, Fritz L, Quilliam MA, Silver M, Buck K, Wright JLC (1993) Domoic acid intoxication of brown pelicans and cormorants in Santa Cruz, California. In: Smayda TJ, Shimizu Y (eds) Toxic phytoplankton blooms in the sea. Elsevier Science Publication B.V., Amsterdam, p 643-650

Wynne T, Stumpf R, Richardson A (2006) Discerning resuspended chlorophyll concentrations from ocean color satellite imagery. Continental Shelf Research 26:2583-2597

Zou JZ, Zhou MJ, Zhang C (1993) Ecological features of toxic *Nitzschia pungens* Grunow in Chinese coastal waters. In: Smayda TJ, Shimizu Y (eds) Toxic phytoplankton blooms in the sea. Elsevier Scientific Publishing Company B.V., Amsterdam, p 651-657

Table 1. Record omission rates and model performance (area under ROC curve) according to independent variable requirements as applied to all compiled literature and field data with counts of toxic *Pseudo-nitzschia* in Monterey Bay where depth ≤5m (N = 576). Area under ROC curve is a measure of model fit accuracy, where <0.6 = poor, 0.6 – 0.7 = fair, 0.7 – 0.8 = good, 0.8 – 0.9 = very good, and >0.9 is considered excellent. A key to variable names is provided in Table 2.

Independent Variables	N	Records omitted	ROC curve
salinity	427	149	0.462
temp	493	83	0.573
ln(*silicic acid*)	516	60	0.614
ln(*chl a*)	497	79	0.618
ln(*chl a*), *temp*	473	103	0.638
ln(*chl a*), *upwelling*	492	84	0.638
ln(*silicic acid*), *upwelling*	516	60	0.713
ln(*silicic acid*), ln(*chl a*)	444	132	0.757
ln(*nitrate*), *temp*, ln(*chl a*)	419	157	0.766
ln(*silicic acid*), *temp*	438	138	0.785
ln(*silicic acid*), *temp*, ln(*chl a*)	422	154	0.848

Table 2. Complete list of the variables evaluated as independent variables in the LOGIT models.

Variable	Abbreviation	Units
Seawater temperature	*temp*	°C
Total chlorophyll a	*chl a*	$\mu g\ l^{-1}$
Nitrate	*nitrate*	μM
Silicic acid	*silicic acid*	μM
Ortho-phosphate	*phosphate*	μM
Silicic acid/Nitrate	*silicic acid:nitrate*	
Nitrate/Silicic acid	*nitrate:silicic acid*	
Ortho-phosphate/Nitrate	*phosphate:nitrate*	
Nitrate/*Ortho*-phosphate	*nitrate:phosphate*	
Ortho-phosphate/Silicic acid	*phosphate:silicic acid*	
Silicic acid/*Ortho*-phosphate	*silicic acid:phosphate*	
Pajaro River discharge	*Pajaro River*	$ft^3\ s^{-1}$
San Lorenzo River discharge	*San Lorenzo River*	$ft^3\ s^{-1}$
Soquel River discharge	*Soquel River*	$ft^3\ s^{-1}$
Salinas River discharge	*Salinas River*	$ft^3\ s^{-1}$
Bakun Upwelling index	*upwelling*	$m^3\ s^{-1}$
$\ln(X^{\dagger} +1)$	$\ln(X^{¥})$	

¥ "X" represents all environmental variables and ratios, excluding TEMP.

Table 3. The three LOGIT models (Annual, Spring and Fall-Winter). The Annual model is refined for year-round use. The Spring model is refined for use from February 14 – June 30. The Fall-Winter model is refined for use from July 1 – February 13. A complete key to variables is provided in Table 2.

Name	Usage	Model
Annual model	Year-round	LOGIT(p) = 9.763 – 1.700[ln(*silicic acid*)] + 1.132[ln(*chl a*)] – 0.800(*temp*) + 0.006(*upwelling*)
Spring model	February 14 – June 30	LOGIT(p) = 5.853 + 1.398[ln(*chl a*)] – 1.135[ln(*silicic acid*] – 0.549(*temp*)
Fall-Winter model	July 1 – February 13	LOGIT(p) = 10.832 – 5.026[ln(*Pajaro River*)] – 3.893[ln(*silicic acid*)] + 1.972[ln(*chl a*)] + 0.652(*nitrate)*

Table 4. Model specifications and diagnostics for the LOGIT models presented in this study. A complete key to variables is provided in Table 2.

Model	Variables	N (cases)	N (bloom cases)	ROC Area	2*[LL(N)-LL(0)]	p-value	McFadden's ρ^2	Naglekerke's R^2
Annual	[ln(*silicic acid*)], [ln(*chl a*)], *temp*, *upwelling*	422	64	0.860	102.377	0.000	0.285	0.376
Spring	[ln(*silicic acid*)], [ln(*chl a*)], *temp*	144	40	0.848	45.885	0.000	0.270	0.394
Fall-Winter	[ln(*silicic acid*)], [ln(*chl a*)], [ln(*Pajaro River*)], *nitrate*	289	27	0.943	96.859	0.000	0.540	0.616

Table 5. Model sensitivity (% of blooms predicted) and specificity (% of non-blooms predicted) at the default cut-point (A) and at model-specific optimized cut-points (B). Optimized cut-points are defined as the cut-point where overall model prediction error is minimized. Models are refined for seasonal and year-round use as follows: Annual = year-round; Spring = February 14 – June 30; Fall-Winter = July 1 – February 13.

A. Default cut-point (0.500)

Model	Cut-point	Sensitivity	False res.	Specificity	False ref.
Annual	0.500	44%	9%	98%	24%
Spring	0.500	60%	14%	94%	20%
Fall-Winter	0.500	63%	4%	99%	19%

B. Optimized cut-point

Model	Cut-point	Sensitivity	False res.	Specificity	False ref.
Annual	0.145	77%	5%	78%	62%
Spring	0.275	75%	11%	75%	46%
Fall-Winter	0.110	89%	1%	89%	55%

Table 6. Model sensitivity (% of blooms predicted) and specificity (% of non-blooms predicted) at model-specific optimized cut-points (A) and where cut-points are set equal to the priors of an independent California Department of Public Health (CDPH) bloom monitoring dataset (Monterey Bay region, 2002-2005) (B). Optimized cut-points are defined as the cut-point where model sensitivity and specificity are simultaneously maximized. Priors are defined as the proportion of total observations that are bloom observations for each model time-period (Annual = year-round; Spring = February 14 – June 30; Fall-Winter = July 1 – February 13). χ^2 Pearson's chi-square statistic. Parentheses denote negative values.

A. Jackknife results with optimized cut-points

						Predictive value		Success Index			
Model	Cut-point	Sensitivity	False res.	Specificity	False ref.	Bloom	Non-bloom	Bloom	Non-bloom	χ^2	p-value
Annual	0.145	77%	5%	77%	62%	37%	95%	62%	(8%)	74.10	0.000
Spring	0.275	75%	11%	76%	45%	55%	89%	47%	4%	31.78	0.000
Fall-Winter	0.110	89%	1%	89%	55%	45%	99%	80%	(2%)	98.98	0.000

B. Jackknife results with CDPH priors (Monterey Bay, 2002-2005) as cut-points

						Predictive value		Success Index			
Model	Cut-point	Sensitivity	False res.	Specificity	False ref.	Bloom	Non-bloom	Bloom	Non-bloom	χ^2	p-value
Annual	0.081	91%	3%	60%	71%	29%	97%	76%	(25%)	54.63	0.000
Spring	0.101	98%	2%	42%	61%	39%	98%	70%	(30%)	21.31	0.000
Fall-Winter	0.066	93%	1%	82%	65%	35%	99%	84%	(9%)	74.37	0.000

Table 7. Performance comparisons among the annual and seasonal models developed in the present study and in previous *Pseudo-nitzschia* modeling publications. Model sensitivity ("Sens") is defined as the % of blooms successfully predicted. Model specificity ("Spec") is defined as the % of non-blooms successfully predicted. Parentheses indicate negative scores.

						Predictive value		Success Index	
Study	Time-period	Dep. Variable	N	Sens.	Spec.	Bloom	Non-bloom	Bloom	Non-bloom
Present study	Annual	tox. *Pseudo-nitzschia* bloom	422	77%	78%	38%	95%	61%	(7%)
Blum et al.	Annual - LOGIT	*Pseudo-nitzschia* toxicity	139	77%	75%	-	-	-	-
Anderson et al.	Annual - "Full Model"	*Pseudo-nitzschia* bloom	75	75%	93%	-	-	-	-
Anderson et al.	Annual - "Remote sensing"	*Pseudo-nitzschia* bloom	89	53%	96%	-	-	-	-
Chl a anomaly	Annual	tox. *Pseudo-nitzschia* bloom	182	39%	72%	12%	95%	31%	(28%)
Present study	Spring	tox. *Pseudo-nitzschia* bloom	144	75%	75%	54%	89%	47%	3%
Chl a anomaly	Spring	tox. *Pseudo-nitzschia* bloom	65	50%	62%	29%	90%	35%	(38%)
Present study	Fall-Winter	tox. *Pseudo-nitzschia* bloom	289	89%	89%	45%	99%	80%	(2%)
Chl a anomaly	Fall-Winter	tox. *Pseudo-nitzschia* bloom	117	0%	77%	0%	97%	(2.6%)	(23%)

FIGURE CAPTIONS

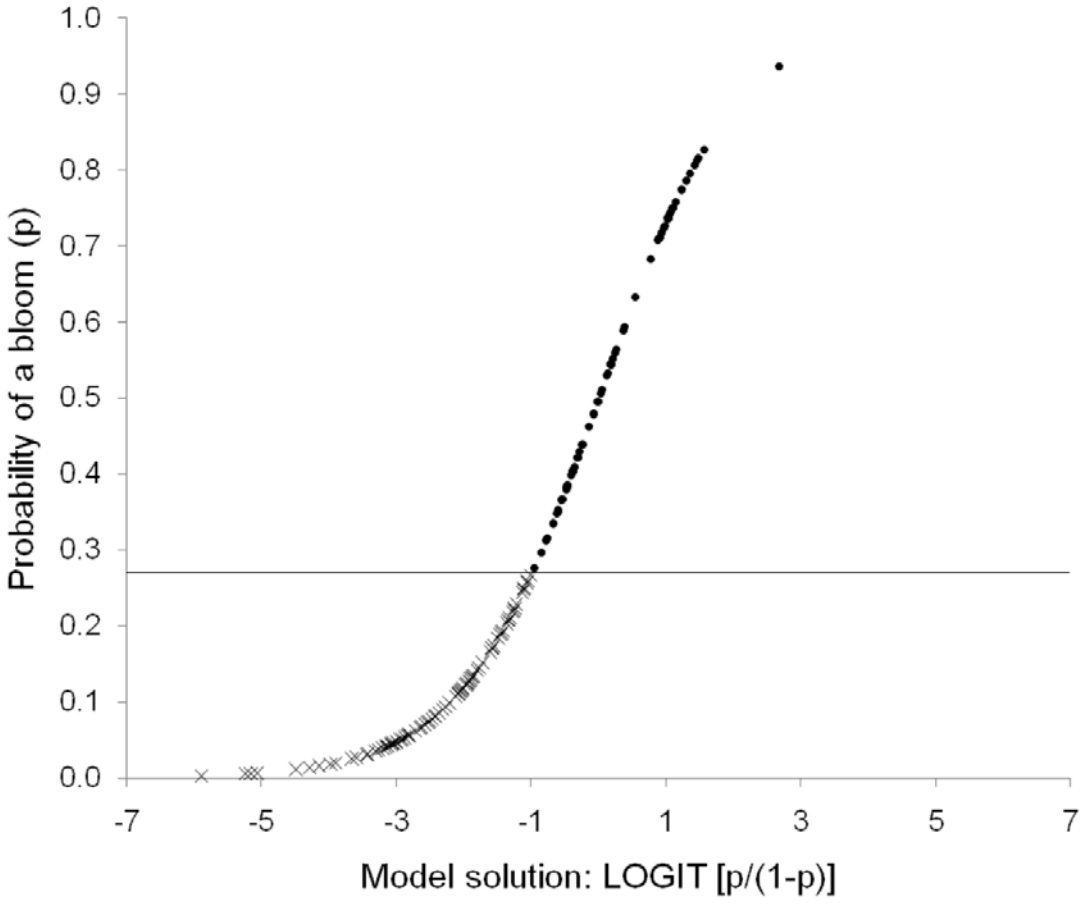
Figure 1. Spring model logistic regression curve case estimations. Cases lying above the optimized cut-point of probability = 0.275 (solid circles) are predicted as blooms; cases lying below the optimized cut-point (X symbols) are predicted as non-blooms.

Figure 2. Prediction failure rates for blooms (solid line) and non-blooms (dashed line) along the range of possible cut-points. The overall prediction failure rate is minimized at the optimized cut-point, where the two lines cross.

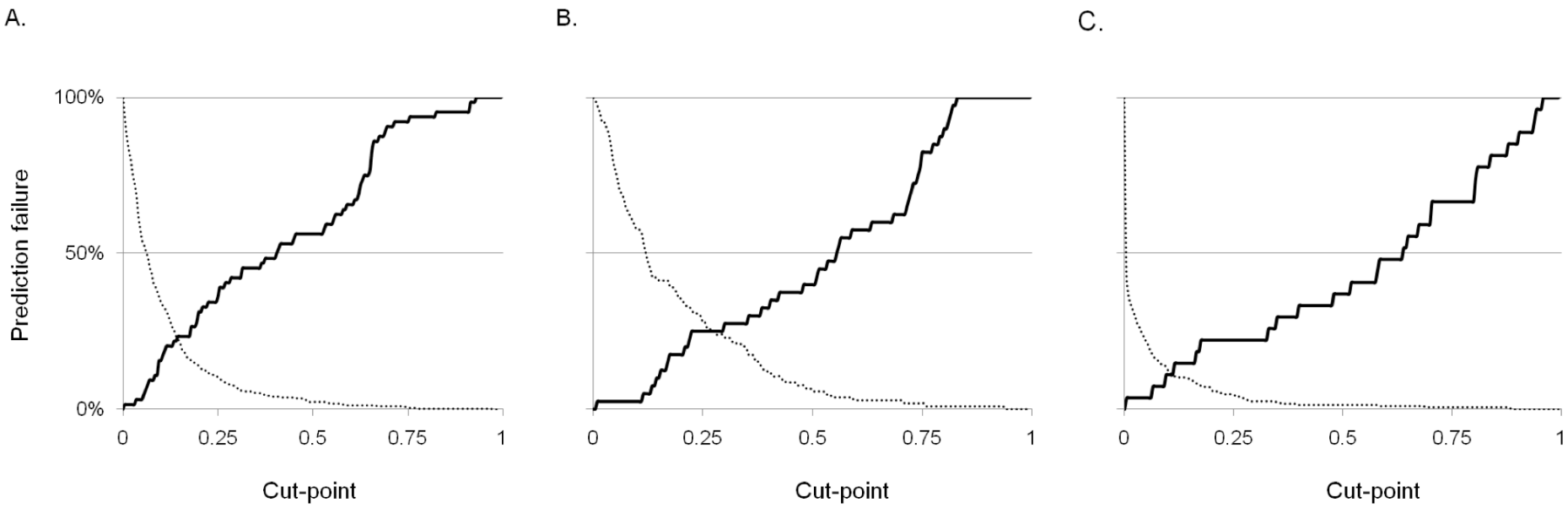
Figure 3. Schematic diagram of a logistic regression with cut-point = 0.25 (A), cut-point = 0.5 (B) and cut-point = 0.75 (C). Cases lying on the “0” vertical axis are actual non-bloom cases; cases lying on the “1” vertical axis are actual bloom cases. Filled symbols lie above the cut-point and are therefore predicted as blooms; open symbols lie below the cut-point and are therefore predicted as non-blooms. Reducing the cut-point (A) increases the number of cases that are predicted as blooms overall, maximizing the % of actual blooms that are successfully predicted (model sensitivity) but reducing the % of actual non-blooms that are successfully predicted (model specificity). Increasing the cut-point has the opposite effect (C).

Figure 4. Venn diagram showing shared and unique variables included in (A) models of *Pseudo-nitzschia* ecophysiology presented within this study, in Anderson et al. (in press), and in Blum et al. (2006) and (B) the Annual model, Spring model, and Fall-Winter model presented in the current study.

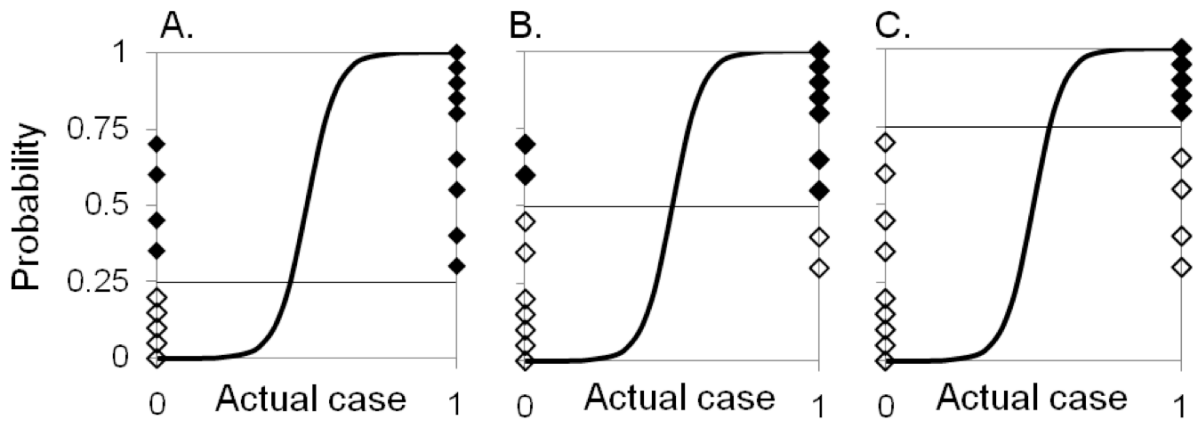
Figure 5. Time series of Pajaro River discharge ($ft^3\ s^{-1}$) (solid line) and toxigenic *Pseudo-nitzschia* bloom events (vertical bars) from the full dataset. Partial vertical bars (hatches) represent non-bloom events.



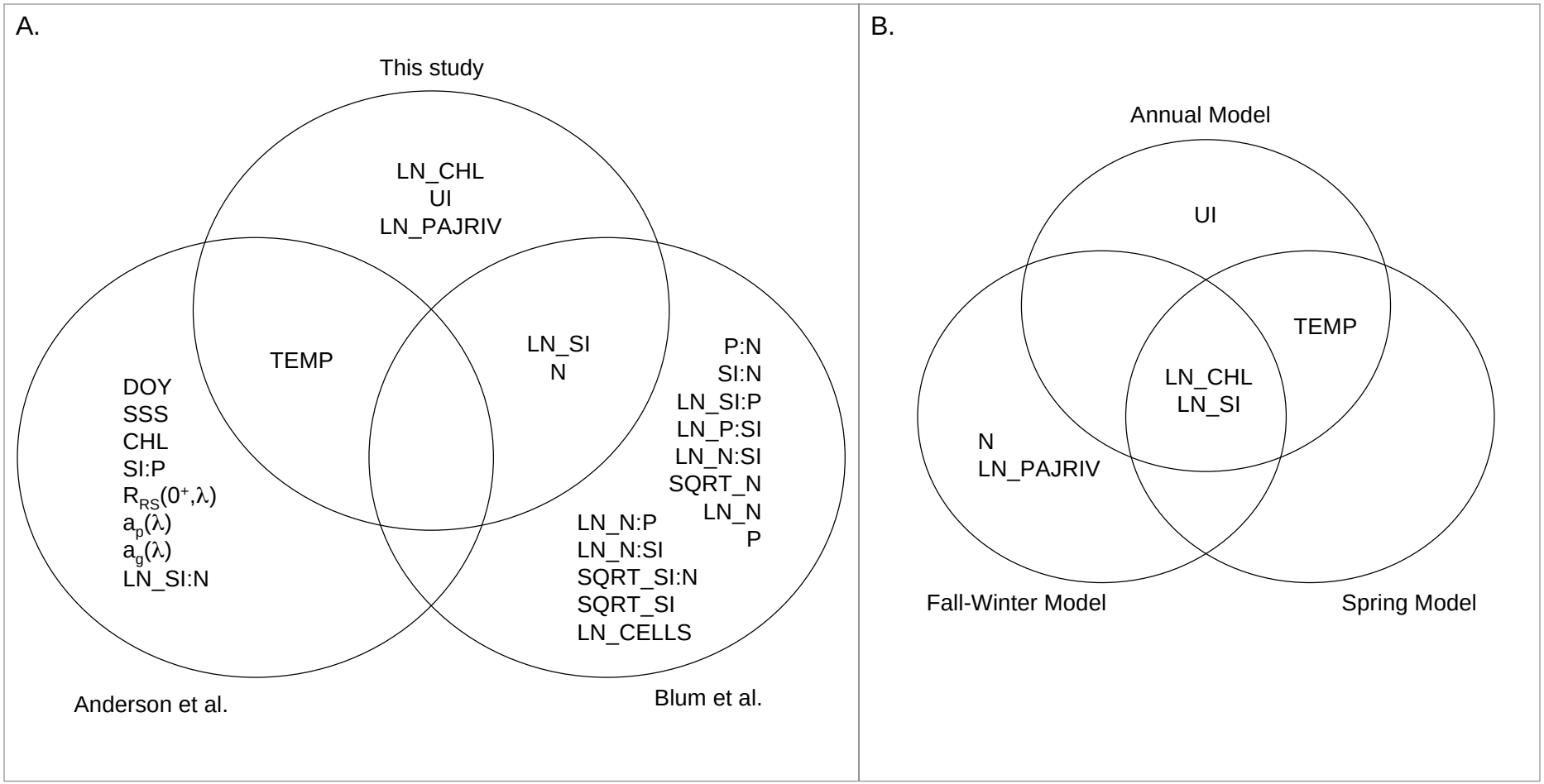
JLANE2008.MEPS.Figure 1



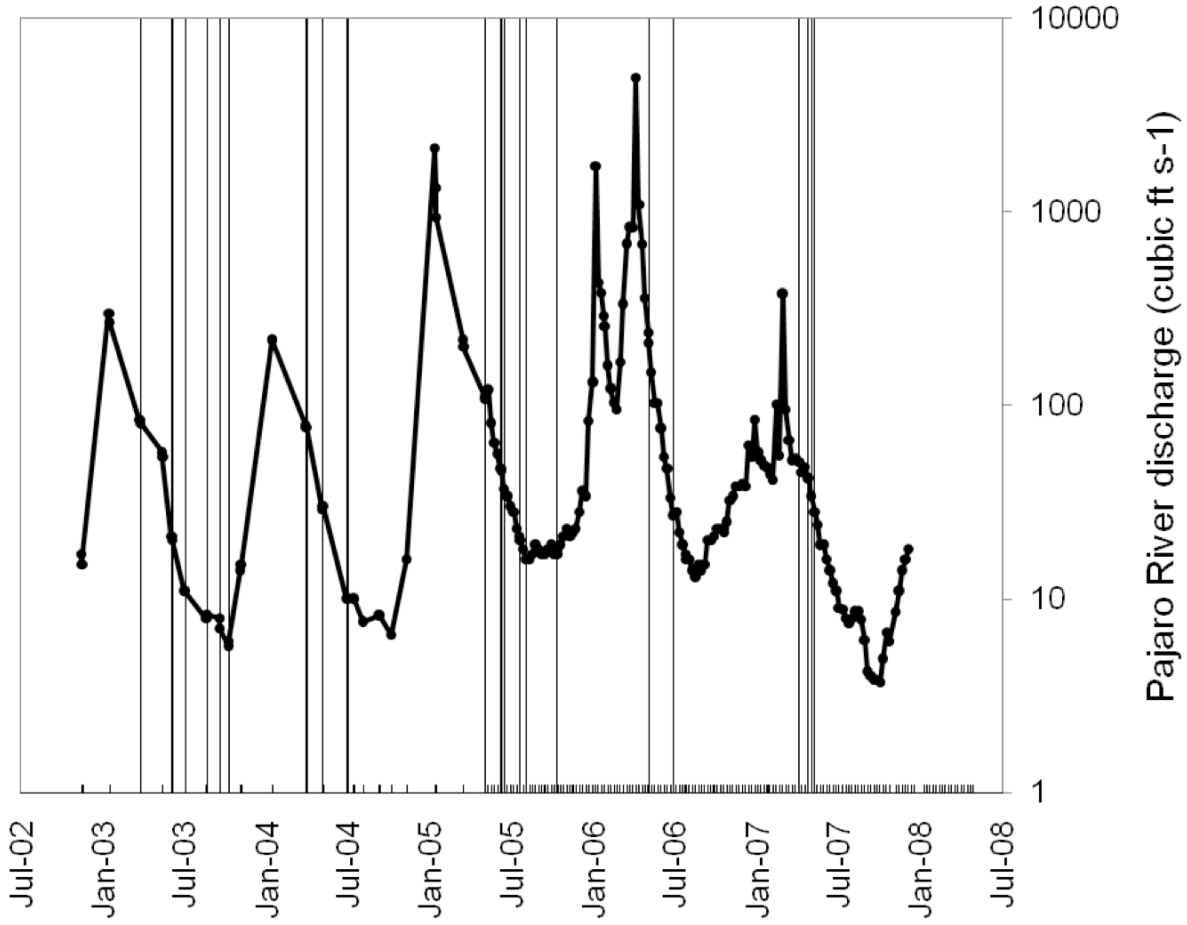
JLANE2008.MEPS.Figure 2



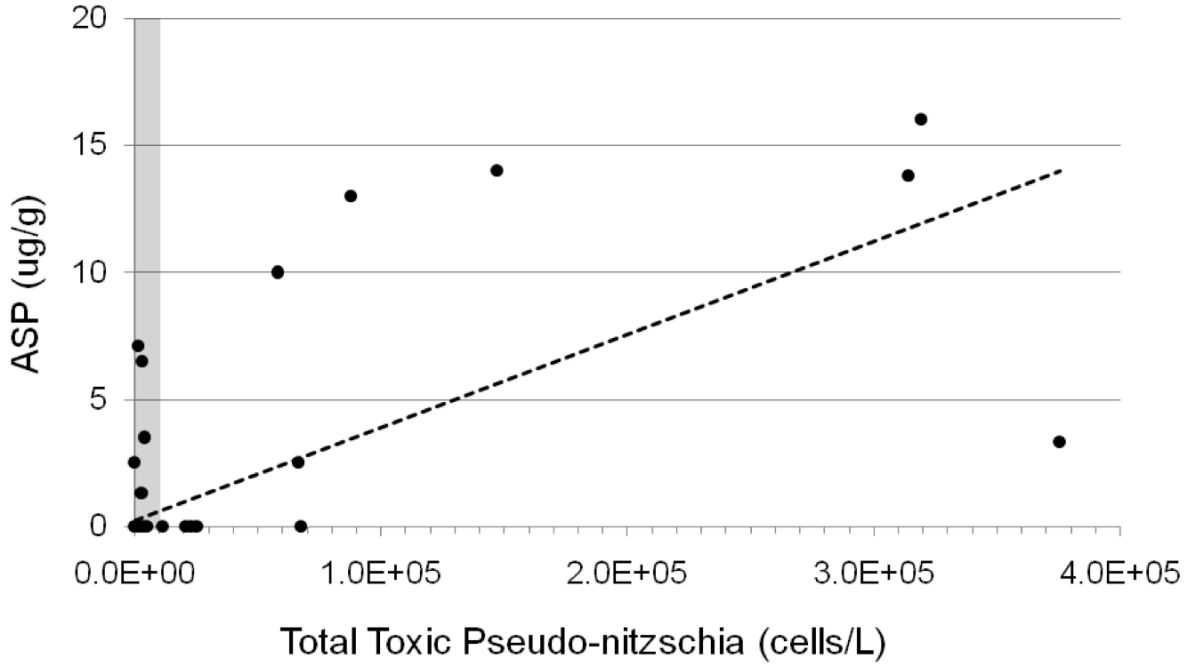
JLANE2008.MEPS.Figure 3



JLANE2008.MEPS.Figure 4



JLANE2008.MEPS.Figure 5



JLANE2008.MEPS.Figure 6