1 **Initial skill assessment of the California Harmful Algae Risk Mapping (C-HARM) system**
2
3

4 Anderson, Clarissa R.[a*], Kudela, Raphael M.[b], Kahru, Mati[c], Chao, Yi[d], Rosenfeld, Leslie K.[e],

5 Bahr, Frederick L.[e], Anderson, David M.[e], and Norris, Tenaya A.[f]

6

7 [a] Institute of Marine Sciences, University of California, Santa Cruz, 1156 High St., Santa Cruz,

8 CA 95064, U.S.A.

9 [b] Ocean Sciences Department, University of California, Santa Cruz, 1156 High St., Santa Cruz,

10 CA 95064, U.S.A.

11 [c] Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive #

12 0218, La Jolla CA, 92093, U.S.A.

13 [d] Joint Institute for Regional Earth System Science and Engineering University of California, Los

14 Angeles, 607 Charles E Young Drive, East Young Hall, Room 4242, Los Angeles, CA 90095,

15 U.S.A. Also affiliated with Remote Sensing Solutions, Inc., Pasadena, CA 91107, U.S.A.

16 [e] Central and Northern California Ocean Observing System, Monterey Bay Aquarium Research

17 Institute, 7700 Sandholdt Rd., Moss Landing, CA 95039, U.S.A.

18 [f] The Marine Mammal Center, 2000 Bunker Road, Fort Cronkite, Sausalito, CA 94965, U.S.A.

19 ***Corresponding author: clrander@ucsc.edu**

22
23
24 **Abstract**
25

26   Toxic algal events are an annual burden on aquaculture and coastal ecosystems of

27 California. The threat of domoic acid (DA) toxicity to human and wildlife health is the dominant

harmful algal bloom (HAB) concern for the region, leading to a strong focus on prediction and

mitigation of these blooms and their toxic effects. This paper describes the initial development of

the California Harmful Algae Risk Mapping (C-HARM) system that predicts the spatial

likelihood of blooms and dangerous levels of DA using a unique blend of numerical models,

ecological forecast models of the target group, *Pseudo-nitzschia*, and satellite ocean color

imagery. Data interpolating empirical orthogonal functions (DINEOF) are applied to ocean

color imagery to fill in missing data and then used in a multivariate mode with other modeled

variables to forecast biogeochemical parameters. Daily predictions (nowcast and forecast maps)

are run routinely at the Central and Northern California Ocean Observing System (CeNCOOS)

and posted on its public website. Skill assessment of model output for the nowcast data is

restricted to nearshore pixels that overlap with routine pier monitoring of HABs in California

from 2014-2015. Model lead times are best correlated with DA measured with solid phase

adsorption toxin tracking (SPATT) and marine mammal strandings from DA toxicosis,

suggesting long-term benefits of the HAB predictions to decision-making. Over the next three

years, the C-HARM application system will be incorporated into the NOAA operational HAB

forecasting system and HAB Bulletin.

1. Introduction

        Previously considered rare in phytoplankton communities, toxic and/or extremely high-

biomass blooms are now recognized as high impact events reported to frequently affect

aquaculture operations, recreational zones, marine mammal and bird populations, and even open

1   ocean ecosystems (D. M. Anderson et al., 2012). A call for predictive capabilities for integrated

2   environmental assessments, early warning systems, action plans, and mitigation strategies was

3   articulated at the West Coast Governors Alliance Harmful Algal Bloom Summit (Lewitus et al.,

4   2012) and is a major component of the strategic vision for a national Ecological Forecasting

5   Roadmap presented by NOAA (http://oceanservice.noaa.gov/ecoforecasting/noaa-

6   ecoforecasting-roadmap.pdf). Domoic acid (DA) poisoning in marine mammal/bird populations

7   and the threat of Amnesic Shellfish Poisoning (ASP) in humans is now considered to be the

8   leading HAB and conservation issue for much of the U.S. west coast (Kudela et al., 2008;

9   Lewitus et al., 2012; Trainer et al., 2012) since these events have the potential to be devastating

10   to aquatic life via bioaccumulation in the food web (Fritz et al., 1992; Lefebvre et al., 1999;

11   Scholin et al., 2000). There are many as-yet unquantified indirect effects, such as those from

12   reduced demand for seafood following a publicized HAB event (Jin et al., 2008) and negative

13   ecosystem consequences in the water column (Sekula-Wood et al., 2010). While DA events are

14   unlikely to be reported until a highly visible impact occurs, such as DA poisoning of California

15   sea lions or the closing of commercial shellfish grounds, with each passing year since 1998,

16   these reports have become more numerous. Clearly, there are significant societal and scientific

17   benefits to predicting when and where these DA events occur relative to susceptible populations

18   of marine mammals, birds, and consumable shellfish.

19   Due to the fact that optical signals of *Pseudo-nitzschia* abundance and more importantly,

20   DA, have not yet been successfully isolated from the bulk chlorophyll signal using our current

21   constellation of multi-spectral ocean color satellite sensors, other methods have been examined.

22   McKibben et al. (2012) assessed the use of a chlorophyll anomaly product for isolating likely

23   *Pseudo-nitzschia* spp. blooms in Oregon coastal waters, akin to that applied operationally to

3

Karenia brevis blooms in the Gulf of Mexico (Stumpf et al., 2009). With an optimized eight-day

window to smooth gaps from cloudy pixels, this anomaly method has been successfully applied

to HAB detection in Oregon and been included in the NOAA CoastWatch product list

(http://www.ngdc.noaa.gov). Several empirical model studies on HABs in the California Current

System (CCS) have also improved our ability to apply near real-time information to the

prediction of *Pseudo-nitzschia* blooms as well as the presence of DA in surface waters of

California. Anderson et al. (2011) extended a previous study from the Santa Barbara Channel at

the northern extreme of the Southern California Bight (C. R. Anderson et al., 2009) that

identified several variables to be good predictors of *Pseudo-nitzschia* abundance and high DA

concentrations in surface waters south and east of Pt. Conception (Fig. 1). The first study

employed stepwise multiple linear regression to correlate hydrographic and chemical shipboard

data with a broad range of *Pseudo-nitzschia* cell densities and DA loads from inshore and

offshore sites within the Santa Barbara Channel from 2004-2006 (C. R. Anderson et al., 2009).

In the follow-on study, Anderson et al. (2011) created similar models to predict the likelihood of

a *Pseudo-nitzschia* bloom or DA event from a longer dataset (Table 1), which included new

observations from 2009-2010 and used the more flexible generalized linear model (GLM)

method in place of the optimal least squares approach (C. R. Anderson et al., 2010; Lane et al.,

2009). The significant predictor variables in the new GLMs agree well with previous models (C.

R. Anderson et al., 2009; Lane et al., 2009), indicating a consistent set of environmental controls

on *Pseudo-nitzschia* blooms over time. An initial goal of the modeling exercise was the creation

of explicit "remote-sensing" (RS) models that would exclude nutrients since their estimation in a

forecasting scenario would require sampling platforms other than satellite remote sensing. These

sampling systems are not routinely deployed and are subject to considerable delays in data

1   availability. The RS models therefore predict toxic events solely from remote-sensing reflectance

2   values (measured with an in-water radiometer), sea-surface temperature, and salinity (Table 1),

3   whereby surface temperature and salinity values can be estimated from either satellite sensors or

4   numerical model output produced by regional circulation models (Anderson et al., 2011).

5   Satellite observations of sea surface temperature (SST) and ocean color have the potential

6   to detect long-term trends due to their high temporal frequency and almost global coverage but

7   are limited by their relatively short duration and cloud cover (Frolov et al., 2013), as well as

8   being restricted to the upper third of the euphotic zone (one optical depth). Thus, a major

9   challenge to developing an operational predictive capability for *Pseudo-nitzschia* blooms has

10   been the acquisition of temporally and spatially coherent, real-time satellite data with which to

11   compute the empirical HAB models. Temporal and spatial averaging is routinely used to reduce

12   the gaps due to cloud cover; however, averaging over scales larger than the decorrelation scale

13   blurs the mesoscale features and limits the utility of satellite data in decision making or for

14   interpretation of underlying bio-physical processes. To quantify the combined effect of cloud

15   cover and satellite coverage on Moderate Resolution Imaging Spectroradiometer (MODIS)

16   imagery, Frolov et al. (2013) compared the statistics of temporal gaps between two consecutive

17   pixels using a fluorescence line height (FLH) time series from 2002-2008 and also estimated

18   spatial decorrelation scales for the same dataset. The shortest gaps between two consecutive

19   pixels were off central California and in the Santa Barbara Channel (3-4 days), while for many

20   places along the California coast (north of San Diego, off Point Sur, and between Cape

21   Mendocino and Point Arena), satellite coverage was exceptionally poor in the first 4-10 km

22   distance from the coast (mean gap of 8 to 10 days). Thus, from the perspective of operational or

1   semi-operational use of ocean color to inform decisions about extreme blooms or toxic events,

2   historical and existing ocean color sensors are inadequate.

3   One method for addressing these satellite data gaps employs Data Interpolating Empirical

4   Orthogonal Function (DINEOF) routines (Alvera-Azcárate et al., 2005; Beckers and Rixen,

5   2003) to solve spatial and temporal EOFs using either single data sets (e.g. chlorophyll) or

6   covarying bio/geophysical data (e.g., chlorophyll, SST), thereby statistically reconstructing the

7   full dataset after convergence on an optimized number of EOFs (Beckers and Rixen, 2003). This

8   method has been used to reconstruct SST fields of satellite data (Alvera-Azcárate et al., 2005;

9   Beckers et al., 2006), and there has been increasing interest and application of this method for

10   other variables, including chlorophyll (Alvera-Azcárate et al., 2007; Mauri et al., 2007; Miles et

11   al., 2009) and total suspended matter (Nechad et al., 2011; Sirjacobs et al., 2008). Here, we

12   describe application of DINEOF to MODIS-Aqua (MODISA) ocean color data to create

13   continuous imagery for use in operational forecasting of HABs in the CCS. The approach

14   capitalizes on the ability to use DINEOF to forecast biogeochemical fields forward in time

15   without the use of a coupled physical-biological model. Instead, the forecast method suggested

16   here (Alvera-Azcárate et al., 2007) uses a multivariate approach to estimate missing satellite data

17   by combining numerical forecasts of physical fields and near real-time estimates of remote

18   sensing parameters to forecast the ocean color fields up to three days in advance. A prototype

19   prediction system for *Pseudo-nitzschia* blooms and DA events in the CCS (Fig. 2) has been

20   running routinely since February 2014 and is now available online via the Central and Northern

21   California Ocean Observing System (CeNCOOS) website (Fig. 3). In this paper, the first year of

22   California Harmful Algae Risk Mapping (C-HARM) model-observation matchups are evaluated

1   for coherence in the temporal patterns and performance metrics commonly used in decision-

2   making.

3   **2. Methods**
4
5   *2.1. ROMS forecasts*
6
7   A state-of-the-art physical circulation model is crucial to forecasting HABs in the coastal

8   zone. As with the Anderson et al. (2011) study, the model employed here is based on the

9   Regional Ocean Modeling System (ROMS), a community-based hydrodynamic model designed

10   for regional applications (Shchepetkin and McWilliams, 2005) with a vertical, coordinate-

11   following bottom topography (Song and Haidvogel, 1994). The ROMS configuration consists of

12   a single domain covering the entire California coastal ocean from north of Crescent City, CA to

13   Ensenada, Mexico and extending approximately 1000 km offshore at a resolution of 3.3 km.

14   There are 40 unevenly-spaced sigma vertical layers with the majority of these clustered near the

15   surface to better resolve processes in the mixed layer. The non-tidal lateral boundary conditions

16   for the ROMS domain are provided by the real-time operational nowcast/forecast from the global

17   $1/12^{o}$ HYCOM (see http://hycom.org). The tidal forcing is added through lateral boundary

18   conditions that are obtained from a global barotropic tidal model TPXO.6 (Egbert and Erofeeva,

19   2002). The atmospheric forcing required by the ROMS model is derived from hourly output

20   from operational forecasts performed with the NCEP NAM 5-km North American model. Every

21   six hours, both *in situ* and remotely sensed (from both satellite and land-based high-frequency

22   radar) observations (e.g., SST, sea surface height, surface current, vertical profiles of temperature

23   and salinity) are assimilated into ROMS using the multi-scale (MS) three-dimensional

24   variational (3DVar) algorithm (Li et al., 2008, 2009) to produce nowcast fields. Every day, a 72-

25   hour forecast is produced using the nowcast at 03 UTC as the initial condition. One of the

1   earliest ROMS configurations is centered on Monterey Bay and has been used to support a well-

2   described 2003 field experiment (Chao et al., 2008; 2009; Doyle et al., 2009; Wang et al., 2009).

3   The Integrated Ocean Observing System (U.S. IOOS) regional associations for central

4   (CeNCOOS) and southern California (SCCOOS) now host the 3-km CA ROMS nowcast and 72-

5   hour forecast output. From the data portal that is jointly hosted by UCLA and Remote Sensing

6   Solutions, Inc. (RSSI) (http://west.rssoffice.com/ca_roms), six hourly runs of SST and sea

7   surface salinity (SSS) and daily three-day forecasts are acquired to compute the HAB empirical

8   models with the same forecasting horizon (Fig. 2).

9   *2.2. Continuous ocean color imagery from DINEOF*
10
11   Moderate Resolution Imaging Spectroradiometer on Aqua (MODISA) data are acquired

12   from NASA's Ocean Biology Processing Group (http://oceancolor.gsfc.nasa.gov/) for the U.S.

13   West Coast region, processed from Level 1 to Level 2 with the standard *l2gen* routine in

14   SeaDAS (v. 7.0.2), and further subscened to the CA-ROMS domain. The DINEOF routine

15   (compiled from http://modb.oce.ulg.ac.be/mediawiki/index.php/DINEOF) is applied to a time

16   series of MODISA imagery. In order to maximize the variability captured by the EOF analysis

17   and minimize the overall percentage of cloud coverage, one month of MODISA imagery was

18   chosen as a starting point (t_0) for seeding DINEOF, and this has been routinely increased at each

19   time step to include the most recent180 days, which runs in reasonable time. Short-term (3-day)

20   forecasts of chlorophyll (Chl) and remote sensing reflectance (R_{rs}) at 488 and 555 nm are

21   generated by solving multivariate EOFs that simultaneously include satellite ocean color fields

22   and forecasted physical fields (SSS, SST) from ROMS. The reconstructed dataset from this

23   analysis provides the forecasted fields (Chl, $R_{rs}(0^+, \lambda)$, SSS, SST) that are then used to calculate

24   the empirical HAB models at each grid point in the model domain (351 x 391 x t) with a three-

1   day lead time. A very simple advection scheme is also employed and may prove to be a more

2   pragmatic alternative to the time-intensive multivariate EOFs. Pixels of Chl and $R_{rs}(0^+, \lambda)$ are

3   "advected" by multiplying by the daily change in the east-west and north-south components

4   [u(dt) and v(dt)] of the 3-km surface current velocity projected by ROMS after first mapping u

5   and v from a model native grid to a latitude-longitude grid (WGS84 coordinate reference system).

6   The DINEOF routine is then used to fill in gaps left by the movement of advected pixels, most

7   notably around the Channel Islands and the northwestern corner of the domain as water is

8   advected south. Gap-filled Chl imagery is included on the HAB forecasting front page as an

9   additional end user product and as a visual aid for interpreting the HAB model imagery (Fig. 4).

10   While forecasts and nowcasts are both produced, the skill assessment discussed in the remainder

11   of this contribution is focused on the nowcast predictions.

12   *2.3. Empirical HAB models*
13
14   The remote-sensing-based empirical models employed in this study are the same as those

15   reported in Anderson et al. (2011) where the methodology for developing the GLMs from *in situ*

16   radiometer and hydrographic observations is well described. Three independent models predict

17   1) *Pseudo-nitzschia* blooms (threshold = 10^4 cells L^{-1}), 2) particulate DA events (pDA threshold

18   = 500 ng L^{-1}), and 3) cellular DA events (cDA threshold 10 pg cell^{-1}) using some combination of

19   $R_{rs}(0^+, \lambda)$ at various wavelengths (λ), MODISA chlorophyll (Chl), salinity, temperature, or time

20   of year (Table 1). At each grid point, these models are computed from the gap-filled MODISA

21   ocean color retrievals and the ROMS physical estimates, leading to spatial/temporal resolution of

22   3km and daily. Given that these are logistic regression models, the value is a probability or

23   likelihood of encountering a bloom and/or toxic state at a given grid point/pixel in the model

24   domain. A current estimate of HAB state (Fig. 3) is *routinely* generated using gap-filled ocean

1 color imagery (Fig. 4a) and the most recent ROMS simulation (Fig. 4b) before the

2 biogeochemical fields are forward projected via advection and/or DINEOF. Predictions up to

3 three days forward (i.e. forecasts) are displayed separately, while archived predictions are

4 available as image files in a set of directories (Fig. 3). The gap-filled ocean color imagery for Chl

5 and the ROMS SST images are also displayed alongside HAB predictions (Fig. 4).

6 *2.4. Routine Ocean Observing Data*

7   The task of verifying forecasts is essential to communicating model accuracy, utility to

8 decision makers, and informing model improvement. The fact that there are appropriate

9 biological field observations with which to compare our model results is due to the diligent

10 efforts of the California Harmful Algal Bloom Monitoring and Alert Program (HABMAP).

11 Originally a grassroots collaboration of academics and government resource managers,

12 HABMAP is now a component of the regional IOOS program that greatly increases our ability to

13 evaluate model output and compare nearshore observations with potential offshore HAB activity

14 in near real-time (Frolov et al., 2013; Kudela et al., 2015). Seven stations from Santa Cruz, CA

15 to La Jolla, CA are sampled weekly for phytoplankton community composition, and at a select

16 few, for key HAB phycotoxin concentrations (saxitoxin and DA). Data are shared with the

17 public in near real-time via the HABMAP (http://www.habmap.info/) and SCCOOS websites

18 (http://www.sccoos.org/data/habs/index.php). The shore stations used for the bulk of the analyses

19 reported here are the Santa Cruz Municipal Wharf (SCMW, 36° 57.48' N, 122° 1.02' W) in the

20 northern part of Monterey Bay and Stearns Wharf (34° 24.48' N, 119° 41.10' W) in the Santa

21 Barbara Channel (Fig. 1). Weekly particulate DA measurements (LC-MS or ELISA) and light

22 microscopy counts of *Pseudo-nitzschia* spp. (total of *seriata* and *delicatissima* size classes) from

23 surface phytoplankton bucket samples are matched to daily model output from 15 Feb 2014 – 10

1   Feb 2015. The longest record of particulate DA for the 2014 to 2015 model demonstration year

2   is at the SCMW shore station where we also compare time series of DA from mussel tissue and

3   from Solid Phase Adsorption Toxin Tracking (SPATT) mesh bags (LC-MS, Lane et al., 2010b)

4   with HAB model predictions. In addition to phytoplankton and mussel data, records of central

5   CA marine mammal strandings (mostly pinnipeds) associated with DA toxicosis from Jan 2007 –

6   Sep 2014 are associated with DA measurements over that time period. Given the uncertainty

7   surrounding when a given animal actually stranded before it was admitted, monthly means of

8   mammal strandings are compared with shore station and model data at the SCMW to assess the

9   relationship between stranding episodes and the variability in nearshore DA levels and predicted

10   DA events. For model/observation match-ups, a time series of daily modeled probabilities from

11   the 3x3 model grid point that best approximates the shore station coordinates is extracted from

12   the netCDF gridded data and compared to these measurements. For all match-ups, we use the

13   nowcasted netCDF data rather than any of the forecast data to simplify the analyses.

14   *2.5. Trend Analysis*

15   At present, it is challenging to compare modeled probabilities to binary outcomes in the

16   observations without a clear understanding of the relationship between the two over the CCS

17   model domain. As model and observation matchups become more abundant through continued

18   daily model runs and HABMAP monitoring in CA, model calibration will become a useful

19   option for creating regionally relevant prediction points. Some of the available methods for

20   comparing patterns in the predicted values with discrete weekly time series of *Pseudo-nitzschia*

21   abundance and DA concentration observed at the HABMAP shores stations are described below.

22   Cross Correlation Functions – Cross-correlation functions (CCF) were computed between time

23   series of daily, modeled values and weekly *in situ* measurements at select shore stations to assess

11

1   correlation and lead-lag relationships between discrete observations and predicted probability

2   values. With CCF analysis, it is important not to violate assumptions of stationarity. In lieu of a

3   formal test for stationarity (e.g. unit root test), stationarity in the time series was diagnosed by

4   plotting the autocorrelation function (ACF) and assessing the steepness of the slope; a gradual

5   decrease in the correlogram that does not taper to zero is a first order measure of a non-stationary

6   process. If a linear trend exists in the time series, it is generally sufficient to detrend the data by

7   simply removing the linear trend. In this case, the residuals of the linear regression of the data

8   are used in the CCF analysis in place of the original time series. In cases of non-linear trends,

9   stationarity can be achieved either by pre-whitening or by applying an Autoregressive Integrated

10   Moving Average (ARIMA) model. An ARIMA model was applied to those data that fit this

11   category by first differencing the time series and estimating the autoregressive, AR (p),

12   integrated (d), and moving average, MA (q) parameters. The autocorrelation (ACF) and partial

13   autocorrelation (PACF) plots of the differenced time series allow us to determine the type of

14   ARIMA (either AR or MA) and number of terms that are needed for fitting the ARIMA (p,d,q)

15   model (Box, Jenkins, and Reinsel,1994). The residuals from the ARIMA fit are then cross-

16   correlated with the second (observational) time series to attain the CCF.

17   <u>Kolmogorov-Smirnov (K-S) Test</u> - A K-S test is a non-parametric statistical test to determine the

18   likelihood that two time series were drawn from the same continuous probability distribution

19   (Lehmann and Romano, 2006) and tests for deviations from the null hypothesis in terms of

20   median, variance, and distribution. It is particularly powerful in detecting differences in

21   distribution shape (or pattern) so we apply it here to compare patterns in observed and modeled

22   time series. A two-sample (or two-sided) K-S test was conducted on the normalized time series

23   of modeled and observed data for each set of matchups (R v 2.15.3), generating a D statistic

12

1 which is evaluated with respect to a critical D value (D_{crit}) at an alpha level of 0.05 (p=α). If the

2 D statistic is greater than the critical value, the null hypothesis that the samples are from a

3 population with the same distribution is rejected.

4 *2.6. Skill Assessment*

5 A number of verification methods commonly used in the meteorological community for

6 dichotomous forecasts (Jolliffe and Stephenson, 2012; 2003) have already been applied during

7 the development and testing of the logistic toxic *Pseudo-nitzschia* bloom GLMs, and these

8 include measures of sensitivity and specificity of the models along with cross-validation and

9 metrics of model skill (C. R. Anderson et al., 2011). The same bloom thresholds used during

10 model development are applied here to transform 2014-2015 observations to a binary value

11 (10,000 cells L^{-1} for *Pseudo-nitzschia* spp. blooms and 500 ng L^{-1} pDA). For simplicity, the

12 cellular DA model is left out of the analysis and assumed to be secondary to the performance of

13 the pDA model.

14 There is no single skill score or metric that can capture all aspects of model performance,

15 and it is a key focus of our research to identify those that are most relevant to assessing the

16 usefulness of a HAB forecast model. Some common metrics are the probability of detection

17 (POD), false alarm ratio (FAR), and probability of false detection (POFD) or false alarm rate. All

18 provide some measure of sensitivity or specificity in the model predictions. Their definitions and

19 use in evaluating HAB predictions are well described in Anderson et al. (2010) and are repeated

20 here in Table 2. To this list we have added *accuracy*, which provides a measure of the

21 proportion of correct forecasts. A popular method of describing model reliability for a

22 dichotomous solution is the calculation of a *bias score* that indicates the frequency with which

23 the model predicts a particular outcome (i.e. yes/no); the bias score is not based on how well the

1   model matches observations and thus is not to be confused with a bias function. A good bias

2   score equals one, but it can range from zero to infinity (Table 2). For all the metrics listed above,

3   calculations are based on a contingency table that categorizes "hits" as correctly predicted

4   blooms (or DA events), "false alarms" as false positives, "misses" as false negatives, and

5   "correct negatives" as correctly predicted non-blooms (Table 2). When the metric scores are

6   plotted across all possible thresholds for determining the relevant probability at which HAB

7   bulletins or other alerts should be issued (i.e. the "prediction point" for issuing a bloom forecast),

8   an optimized prediction point can be achieved with respect to any value of interest to resource

9   managers or to maximize or minimize a particular performance metric (Anderson et al., 2010;

10   Lane et al., 2009). It is important to note that the chosen prediction point must meet the needs of

11   the end-user; just as there are multiple metrics that can be optimized, the prediction point must

12   also be chosen to optimize the metric of interest, possibly at the expense of another metric. For

13   example, a manager may prefer a higher rate of false positives to minimize the chance of a false

14   negative.

15   The receiver operating characteristic (ROC) describes POD (sensitivity) as a function of

16   POFD (1-specificity) over the full range of prediction points, i.e. the cumulative distribution

17   function, and is used in many fields to translate binary classifications into decision-making

18   criteria. It quantifies the relative accuracy of a model to separate true positive and true negative

19   outcomes (Fawcett, 2004). The area under the ROC curve (AUC) reduces the two-dimensional

20   space to a scalar value that is a well-used metric of model performance (e.g. Lane et al., 2009)

21   and is essentially equivalent to a Wilcoxon rank test (Fawcett, 2004). The AUC is calculated

22   using Simpson's rule to approximate the integral of the function (*sintegral* in R, Bolstad2

23   library). When values on the ROC curve fall above the 1:1 line, they are in the realm of good

1 predictions, while those at or below the 1:1 line (particularly in the lower right quadrant) are

2 deemed as good or worse than random. Similarly, AUC scores above 0.5 suggest a model that

3 performs better than random chance. Both the ROC curve and its AUC are reported as a measure

4 of the usefulness of the HAB models, as best as can be ascertained at the shore station pixel(s).

5 **3. Results**

6 *3.1. Summary of Major 2014 DA Events*

7
8 Interestingly, 2014 was a very high DA year, particularly in Monterey Bay (MB), with

9 record fishery alerts and marine mammal strandings. Modeled pDA probabilities from the

10 SCMW pixel were somewhat elevated in mid-March (> 0.5) but broadly high for MB as the

11 region saw a spike in California sea lion (*Zalophus californianus*) strandings from MB (Fig. 5).

12 A week later, predictions targeted MB as a hot spot of HAB/DA likelihood (mapped in Fig. 5)

13 just as measured pDA at the SCMW shore station rose to 0.84 ng mL^{-1} and modeled pDA

14 probabilities at the SCMW rose to 0.8. This instance corroborates hypotheses that marine

15 mammals are good sentinels of the onset of DA events (Bargu et al., 2012), particularly since

16 these first DA-related strandings of the year were all from the MB region. We should note the

17 discrepancy in the magnitude of the hot spot in the posted imagery (maps in Fig. 5) versus the

18 time series extracted from the netCDF file (pDAmod in Fig. 5) since the latter reflects a year of

19 MODIS variability analyzed with DINEOF, whereas the imagery posted in March 2014 was

20 influenced by a fewer number of positive excursions in the shorter DINEOF time series. In the

21 week following the hot spot imagery (on 1 April 2014), the CA Dept. of Public Health (CDPH)

22 issued closures of recreational shellfish harvests and a historical warning regarding fishing

23 anchovies and sardines in the MB due to dangerous levels of DA toxicity. The 1-week lag time is

1   consistent with past observations of delayed toxicity in shellfish (Lane et al., 2010a) and

2   corroborates the timing of the predicted rise of DA.

3   The fall HAB event in 2014 was focused on the Central-Southern California transition

4   zone where the CDPH issued a warning on October 11 not to eat crabs, lobsters, and bivalves

5   harvested in the Santa Barbara Channel region, another well-known hot spot for DA (C. R.

6   Anderson et al., 2006; Trainer et al., 2000). Forecast maps published on the CeNCOOS website

7   illustrated the increased risk in this region (Fig. 6). In Monterey Bay, on the other hand, samples

8   taken within the Bay from a cruise of opportunity on 15 -17 October verified that DA levels were

9   extremely low while *Pseudo-nitzschia* abundance was at moderate to bloom levels $\sim 10^3$-10^4

10   cells L^{-1} (pers. comm., M. Blakely Peacock), as seen in the imagery for 15 Oct (Fig. 6). DA

11   event probabilities at Stearns Wharf in the Santa Barbara Channel remained elevated relative to

12   background levels for another week (Fig. 6). The CDPH only lifted the advisory for the Ventura

13   County coastline on 21 February 2015, which may reflect the persistence of DA in benthic

14   environments (C. R. Anderson et al., 2011; Kvitek et al., 2008; Sekula-Wood et al., 2009),

15   particularly if the low pDA probabilities are reflective of low surface pDA values.

16

17   Santa Cruz Municipal Wharf - We first discuss the temporal trend in model and observational

18   time series for the SCMW where there is the most comprehensive record of *Pseudo-nitzschia*

19   abundance and DA from both phytoplankton and mussel tissue for 2014-2015 (Fig. 7).

20   Predictions of *Pseudo-nitzshia* blooms at that SCMW pixel over the course of 2014 were

21   dynamic with rapid fluctuations between low and high probabilities with a significant upward

22   linear trend (r = 0.37; p<0.001; Fig. 7a), potentially related to the increasing bloom conditions

23   observed in the CCS later in 2015 (McCabe et al. in review). The dramatic rise and fall in

1   *Pseudo-nitzschia* abundance between April and May at the SCMW is reflected in the modeled

2   probabilities, with predictions leading the rise in abundance at the SCMW. The *Pseudo-nitzschia*

3   group disappeared from the SCMW samples in mid-July (Fig. 7a). After July, however, the

4   modeled time series maintains an elevated mean value, albeit highly variable (Fig. 7a). The

5   model correctly captures the drop in *Pseudo-nitzschia* abundance between April and May, and

6   given that the GLM for *Pseudo-nitzschia* is driven by month and R_{rs}(488/555), this large

7   variation is most likely due to large fluctuations in phytoplankton pigments during this spring

8   transition. The model does not capture the disappearance of *Pseudo-nitzschia* at the SCMW after

9   July. A K-S test quantifies this mismatch between the patterns in the modeled and observed time

10   series at the SCMW (D_{crit} = 0.26; D = 0.44, p<0.01). The resulting CCF [ARIMA (1,1,0)] of the

11   predictions with respect to SCMW abundance measurements shows a weakly significant

12   correlation at zero lag and one-day lead time as well as a negative correlation at an eight-day

13   lead (Fig 8a).

14   Particulate domoic acid (pDA) at the SCMW follows a similar pattern to *Pseudo-nitzshia*

15   in that all the peaks occur in spring, and modeled probabilities at this nearshore pixel are better

16   aligned with pDA observations in the early part of the year than after July (Fig. 7b). The time

17   series trend is negative (r= -0.51; p<0.001), however. The K-S test also reflects the diverging

18   patterns of modeled and observed time series at the SCMW over the course of the year (D_{crit} =

19   0.26; D = 0.41, p<0.01). The inability of the models to capture the summer-fall variability at the

20   SCMW is possibly due to nearshore conditions in summer being especially uncoupled from the

21   regional signal, a pattern that emerged from a separate analysis of crosshore *Pseudo-nitzschia*

22   and DA in the Santa Barbara Channel from 2009-2013 (Umhau et al. in prep). The CCF analysis

17

1   for the nearshore pixel relative to the SCMW observations indicates several significant lead

2   times (e.g. 5, 19, and 26 days) even after applying ARIMA (0,1,0) (Fig. 8b).

3   The relationship between mussel DA and phytoplankton pDA at the SCMW is fairly

4   close, as indicated by a two-sided Kendall rank correlation test that yields significant agreement

5   (τ=0.41, p<0.05). There is a significant five-day lead between the pDA model at the SCMW

6   pixel and mussel DA (Fig. 9a). A K-S test also reflects a stronger relationship between the

7   mussel DA observations and modeled pDA at the SCMW pixel for this shorter time period (D_{crit}

8   = 0.34; D = 0.36, p = 0.006). An alternative record of DA that has been shown to be a better tool

9   for monitoring phycotoxins than sentinel shellfish sampling due to the incredibly high depuration

10   rates of bivalves such as mussels is SPATT resin beads (Lane et al., 2010b), which have been

11   deployed in mesh bags and recovered for DA extraction at one to two week intervals since 2008

12   at the SCMW (Fig 6b). In this case, the mean modeled pDA was used during the two week

13   SPATT deployment and matched to the midpoint date of the SPATT deployment. There is a

14   significant 9 to 12-day lead shown by the CCF (r ~ 0.40) between the mean pDA time series

15   (data untransformed in this case due to the sine-cosine wave decay in the ACF) at the SCMW

16   pixel and SPATT-measured DA (HP20-type resin; Fig. 8b). The K-S test indicates a significant

17   similarity in the time series patterns (D_{crit} = 0.37; D = 0.36, p = 0.013).

18   The CCF for marine mammal strandings averaged over a month suggests that pDA

19   predictions lead the peaks in strandings by five days at the SCMW pixel (r = 0.48, Fig. 9c) and

20   are negatively associated with the model at a five-day lag. The patterns fail to indicate similarity

21   with a K-S test (D_{crit} = 0.15; D = 0.28, p<0.01). This outcome suggests the model does

22   potentially capture time and space scales relevant to California sea lion foraging patterns and

23   may provide adequate lead-time to assist marine mammal rescue efforts.

1   Stearns Wharf – Further south in the Santa Barbara Channel, predicted *Pseudo-nitzschia* bloom

2   probabilities from the Stearns Wharf pixel are highly variable but with a flatter mean over the

3   study period than at the SCMW (Fig 10a). The CCF shows that the strongest relationship

4   between modeled [ARIMA(1,1,0)] and observed data occur when the model either leads by 20

5   days or lags by two days (r = 0.25), and that at lag zero, they are negatively correlated (Fig. 11a).

6   Spring peaks in pDA levels at Stearns Wharf do not coincide with elevated pDA

7   probabilities except for the mid-June increase when probabilities rise abruptly (Fig. 10b).

8   Significant correlations in the CCF occur when the model [ARIMA (1,1,0)] leads observations

9   by 8 and 15 days (Fig. 11b) and lags by roughly 30 days. The K-S test suggests similarities in the

10   shape of nearshore observations and the model at Stearns Wharf (D_{crit} = 0.26; D = 0.33, p<0.05).

11   All California HAB Stations –Routine measurements of *Pseudo-nitzschia* spp. abundance in

12   southern and central California are combined into one analysis for a California-wide assessment

13   (Fig. 12). There is a significantly increasing trend (r = 0.37; p<0.001) in the modeled

14   probabilities over the 2014-2015 timeframe as the ecosystem moved towards the massive HAB

15   event of spring-summer 2015 all along the U.S. west coast (McCabe et al. in review). Departures

16   from the mean model state are shown in the upper panel of Figure 12. Scripps Pier and Cal Poly

17   Pier (in San Luis Obispo) diverge the most from the mean state, suggesting that the timing of

18   conditions supportive to HAB development at these sites may be uncoupled from the timing in

19   other parts of southern and central California. There are many weakly significant spikes in the

20   CCF [ARIMA (0,1,0)] for all nearshore stations, but none point to a near-term lead between

21   model and bloom observations (Fig. 13a). Abundance of *Pseudo-nitzschia* spp. does significantly

22   lag the nowcasts by three days (Fig. 13a).

23   *3.3. Contingency Plots of Skill Assessment*

1   Figures 13-15 show the outcome of each selected metric (Table 2) for comparing a binary

2   classification with modeled/forecasted probabilities over all possible prediction points. The

3   *Pseudo-nitzschia* model yields a high rate of false positives across all HAB monitoring stations.

4   For this reason, the optimized prediction point (0.52) for California is a minimization of the false

5   alarm rate (FAR) with respect to the probability of detection (POD). Their optimized value (i.e.

6   where the two lines cross) is 67% and total accuracy is 43% (Fig. 13b). The maximal accuracy of

7   the *Pseudo-nitzschia* model at all HAB stations is 67%. At the SCMW site specifically, the

8   prediction point was optimized for POD and accuracy since the false positive rate was lower,

9   yielding an optimal prediction point of 0.68 where accuracy and POD = 38% (Fig. 14a). The

10   pDA model at SCMW appears to be more constrained than the *Pseudo-nitzschia* model in that it

11   yields a lower false positive rate. Given this, we optimized relative to accuracy and POD to

12   maximize the sensitivity. This shared value is 68% at an optimal prediction point of 0.6 (Fig 14b).

13   The maximum bias score when the prediction point is zero is four for the *Pseudo-nitzschia*

14   model and nine for the pDA model at both sites (maxima not shown on plots). If the prediction

15   points were optimized with respect to the BS instead of the POD, higher accuracies and

16   reliability would be achieved using more conservative prediction points. At Stearns Wharf, the

17   *Pseudo-nitzschia* model was optimized relative to FAR and POD as before, yielded an optimal

18   threshold of 0.55 with POD and FAR = 57% (Fig. 15a). There are relatively few actual DA

19   events at Stearns Wharf over the time period, and there is a high percentage of correct negatives.

20   When optimized relative to accuracy and specificity (each at 92%), the optimal threshold is 0.63.

21   The results of plotting the ROC suggest that the pDA model at the SCMW performs

22   better than random chance because the AUC is 0.77 (Fig. 14d). The optimized prediction point is

23   solidly in the upper middle to left quadrant, indicative of a moderately "liberal" classifier

1   (Fawcett, 2004). In other words, it is considered a *useful* model since the points on the curve fall

2   almost entirely in the upper triangle region above the 1:1 line, and the AUC is greater than 0.5.

3   The ROC for the *Pseudo-nitzschia* model across all stations almost overlies the 1:1 line (Fig.

4   13c), i.e. it is a borderline case. At the SCMW and Stearns Wharf, the ROC for the *Pseudo-*

5   *nitzschia* model is almost entirely below the 1:1 line (Fig. 14c, 15c), meaning it is performing

6   worse than random (AUC = 0.33). Even the optimized prediction point falls below the 1:1 line,

7   although if it were optimized relative to the BS, this score would fall in the upper right quadrant

8   where the model performs somewhat better than chance. For the pDA model at Stearns Wharf,

9   the ROC is well below the 1:1 line with an AUC of 0.04, reflecting the poor performance of the

10   model at this site as well as the very few DA events that took place over the study period.

11

12   *3.4. Partitioning sources of error between DINEOF and ROMS*

13   Lastly, we evaluate error generated by the input variables used to compute the empirical

14   HAB models. The DINEOF reconstruction of near-daily MODISA data introduces error to the

15   ocean color imagery, and these error fields for the reconstruction are obtained from cross-

16   validation of the DINEOF fit (standard in the DINEOF routine), thereby producing both the fully

17   reconstructed data and information about the error covariance for the same data. Given that the

18   expected error for the reconstructed Chl and $R_{rs}(0^+, \lambda)$ fields rarely exceeds one standard

19   deviation from the mean in our routine runs (data not shown), the error introduced by the

20   DINEOF routine is likely not a significant error source for our predictions.

21   A second source of error from DINEOF could be introduced by the length of time

22   interval used to run the reconstruction. It is not entirely clear if the smaller degree of variability

23   present in a shorter time series (particularly one with a high percentage of gaps due to clouds or

21

1   other flagged data) skews the EOFs towards spurious patterns or if the opposite is true. Figure 16

2   compares the DINEOF Chl data reconstruction performed in 180-day intervals, where for the last

3   10-day period DINEOF was run on the previous 180 days (days 1-180, 2-181, etc.), versus a

4   365-day interval (days 1-365). This is tested using either a 3x3 box around the SCMW or

5   extracting the nearest 1x1 pixel relative to the SCMW. Sequentially adding days for ten days

6   does very little to change the output. Log-transformed Chl values for this shorter interval (red

7   lines in Fig. 16a) generally track the 365-day runs (solid and dashed blue lines in Fig. 16a) until

8   the end of the record. Shifting from a 3x3 window to the nearest pixel approach (red lines in Fig.

9   16b) shows convergence of the 180-day intervals with the 365-day interval (solid blue line) in

10   that same pixel space. Several of those 180-day intervals within the ten-day period that was

11   tested exhibit no variability (i.e. flat line). This can occur when data close to land are included,

12   since small errors in the satellite navigation or in viewing angle result in the pixels being

13   classified as land (NaN values), resulting in a flat line.

14   The 3-km ROMS temperature and salinity fields can be evaluated with physical

15   measurements from the SCMW in the same manner as the HAB empirical models. Direct

16   comparison of the closest ROMS grid point with the SCMW time-series shows good agreement

17   for salinity (Fig. 16c) and temperature (Fig. 16d), although many of the large drops in salinity

18   below 30 psu from winter/spring storms are missed by ROMS, likely due to the lack of terrestrial

19   freshwater input to the model. Using as much *in situ* data from the SCMW as possible, we

20   calculate the pDA model ("observed) and compare those outcomes with our routine pDA

21   predictions that use ROMS salinity and DINEOF chlorophyll ("model"). We plot the outcomes

22   of alternately substituting the DINEOF chlorophyll with observed chlorophyll ("pDA CHL) and

23   the ROMS salinity field with observed salinity ("pDA Salt"). The results suggest that the pDA

1   model is more sensitive to substitutions of ROMS salinity ($r^2 = 0.37$, slope = 0.71) than

2   DINEOF-filled ocean color values ($r^2 = 0.16$, slope = 0.58). This sensitivity analysis

3   demonstrates the disproportionate effect that salinity has on the pDA model outcomes relative to

4   chlorophyll (Fig. 17).

5

6   **4. Discussion**

7   *4.1. Skill Assessment at Shore Stations*

8   Model skill was evaluated using a suite of performance metrics for dichotomous

9   predictions using the best available observational data from shore stations in several important

10   locations. This analysis demonstrates that the model provides information that is in relative

11   agreement with shore station data but with significant differences that require continual

12   evaluation and offshore sampling in order to ascertain the scales of variability captured by the

13   model versus the nearshore monitoring. In other words, while these HABMAP monitoring data

14   represent the most consistent source of HAB observations in the CCS, there are obvious

15   problems with employing them in model evaluation given the limitations in spatial (and

16   temporal) scales. For instance, the spatial resolution of the final HAB model predictions is

17   constrained by the lowest resolution input product, which is the 3-km CA ROMS model, such

18   that these moderate-resolution modeled matchups (nearest pixels) are only rough approximations

19   of HAB activity in the nearshore zone. Spatial mismatch between *in situ* observations and

20   interpolated 3-km satellite imagery is indeed an impediment to validation studies of this sort

21   (Carvalho et al., 2011). Unfortunately, there is no consistent source of offshore data for

22   comprehensive skill assessment. Nevertheless, we must assess the model relative to the HAB

1   shore station data that are the baseline for monitoring ecosystem health in coastal California

2   (Kudela et al., 2015).

3   Given the spatial mismatch just mentioned and the many sources of error that contribute

4   to the HAB model signal, it is somewhat surprising that there is any skill between the model-

5   observation matchups generated by our approach. However, we do see some degree of *sensitivity*

6   and *specificity* as well as integration of several scales of variability, even at the individual pixel

7   level. The *Pseudo-nitzschia* model is less accurate than the pDA model due to its high false-

8   positive rate. Model sensitivity (or POD in Figs. 13-15) describes the fraction of DA events

9   correctly predicted and is above 50% at the optimized prediction point shown in Figure 14b. The

10   overall problems with model performance when assessed at the pixel level relative to very

11   nearshore, pier-sampling matchups is corroborated by *in situ* observations from the Santa

12   Barbara Channel where *Pseudo-nitzschia* and pDA at Stearns Wharf rarely align with offshore

13   levels in the middle of the channel (Umhau et al., in prep). Future analyses of all offshore

14   datasets in the CA region will allow us to fully evaluate alongshore variations in model

15   performance and refine the GLMs.

16   The SCMW appears to be capturing a mammal-stranding signal at a two-week lead time

17   despite the fact that the stranding data are grouped for the entire central/northern CA region (Fig.

18   7b). The peaks in strandings generally correspond with times when pDA is present at the SCMW,

19   although there are many instances when strandings occur in the absence of nearshore pDA

20   measured at SCMW. This suggests a possible offshore (or even subsurface) source of DA to

21   these stranded animals that is not explicitly captured by the inputs to the empirical models, but

22   might be driving the broader spatial signal in the low to moderate-resolution input fields. It may

23   also point to failures of the model. This is consistent with studies examining the subsurface

1   initiation or seeding of *Pseudo-nitzschia* spp. blooms in coastal California (Seegers et al. 2015).

2   It may also point to failures of the model. The 5-day lead-time between predicted pDA and

3   mussel tissue at the SCMW (Fig. 9a) also signifies that C-HARM could provide a useful warning

4   tool to aquaculture and recreational shellfish growers who stand to benefit from advanced

5   warning of DA contamination.

6   Since SPATT represents a time-integrative (but spatially discrete) detection method of

7   dissolved toxin (Lane et al., 2010b; MacKenzie, 2010), it is not surprising that model variability

8   more resembles SPATT variability than that of the sentinel mussel species, *Mytilus californianus*.

9   Lane et al. (2010) showed that "DA-signaling" by SPATT preceded CDPH detection of

10   dangerous levels of DA in mussels by eight weeks and is often uncoupled from temporally and

11   spatially discrete measures of pDA at the SCMW. However, the exact lead-lag relationship

12   between particulate DA, SPATT DA, mussel DA, and modeled DA is more difficult to discern in

13   the present study given the complicated results of the cross correlation functions (Fig. 9). Both

14   SPATT and the model appear to capture a larger, regional signal than traditional sampling

15   methods and may be more tightly coupled with one another than either of them is with other

16   discrete sampling methods. Ultimately, all sources of information, from models to *in situ*

17   sampling need to be considered as a whole for sound decision-making.

18   An additional and important consideration is that our choice of threshold or prediction

19   point clearly alters model form and sensitivity/specificity (Fig. 15-17) and is always subject to

20   future adjustment to assess its relevancy to ecosystem response. As observational data are

21   constantly assimilated into our statistical models for re-tuning, there will be no single, static

22   optimized probability threshold but rather a dynamic assessment built into the protocol for

23   issuing HAB alerts and bulletins across the different CA subregions.

1   *4.2. Sources of Error*

2   Model skill is ultimately dependent on several factors, including error in the input

3   variables, error in the statistical habitat models and mismatch (spatial, temporal, and scale)

4   between model results and observations models (GLMs), as already discussed. Our method of

5   merging DINEOF-filled MODIS imagery and ROMS fields introduces several sources of error

6   before the HAB empirical models are even computed. First, the ocean color imagery itself

7   represents an imperfect estimation of $R_{rs}(\lambda, 0^+)$ in the coastal zone that varies depending on the

8   satellite sensor. Kahru et al. (2014) recently compared Chl retrievals for the CCS from SeaWiFS,

9   MODIS-Terra, MODIS-Aqua, MERIS, and VIIRS sensors relative to available *in situ* Chl

10   observations for the region. All sensors, with the exception of MERIS, displayed low accuracy at

11   estimating Chl concentrations above 1 mg m^{-3}, a value that often distinguishes coastal

12   environments from pelagic/oligotrophic zones. In the case of MODISA, Chl retrievals and *in situ*

13   measurements were related by an R^2 of 0.32 and a 28% underestimation rate in the medium to

14   high Chl range (Kahru et al., 2014). While the imagery is excellent for its large spatial coverage

15   and operational reliability, we must contend with the downstream influence on HAB model

16   predictions.

17   The application of DINEOF introduces error to the ocean color data in that the gap-filling

18   procedure is an imperfect interpolation method but one that can be quantified in terms of an

19   uncertainty estimate generated from the iterative EOF solution (Beckers et al., 2006). The

20   uncertainty for both the reconstructed Chl and remote sensing reflectance fields is within a

21   standard deviation of the mean value and well below the error exhibited for the actual MODISA

22   retrievals (Kahru et al., 2014). The same DINEOF methods compared well with a coupled

23   physical-biological model in the Pacific Northwest (Giddings et al. 2015), providing further

26

1   indirect evidence that the interpolation scheme does not introduce substantial error. Another

2   factor influencing variability in the HAB predictions is the *time window* used during DINEOF

3   data reconstruction. There is some evidence that a long time domain containing some heavily

4   clouded imagery can produce spurious temporal variability, but that the data reconstruction can

5   be improved by filtering out spikes in the temporal covariance matrix before the EOF

6   decomposition (Alvera-Azcárate et al., 2009). In our sensitivity analysis, the shorter (180-days;

7   red lines in Fig. 16) and longer (365-days; blue lines in Fig. 16) time windows did not

8   significantly diverge from one another until the end of the 180 day record, suggesting that a

9   shorter time domain preserves the majority of the temporal variability captured by the longer

10   time domain while also allowing improved computational efficiency. In terms of the influence on

11   spatial variability, Beckers et al. (2006) suggest a nested DINEOF method in which a large-scale

12   (up to global) reconstruction is first created, and then the anomaly (reconstruction minus

13   observed values) is reconstructed at higher spatial and temporal resolution, followed by

14   reconstruction at even higher resolution, etc. Since each reconstruction is influenced by the scale

15   of patterns inherent in the raw data, this method would sequentially include more small-scale

16   features. Exploration of these methods is part of our on-going investigation into the appropriate

17   use of DINEOF for operational modeling. This ultimately involves trade-space constraints

18   between improving model output and (potentially dramatically) increasing computational cost.

19   Lastly, the ROMS nowcast and forecast fields also have uncertainties. This uncertainty

20   has been gradually reduced over the past few years with the improvement of data assimilation

21   algorithms and increase in observational data sets. The RMS difference of SST between the

22   ROMS nowcast and satellite measurements is about 0.4 $^{\circ}$C during the winter seasons but

23   increased to 0.8 $^{\circ}$C during the summer due to the lack of satellite-retrieved SST when there are

1   clouds. The model uncertainty of temperature below the ocean surface is significantly larger

2   simply because of the lack of subsurface data. Using the independent data from CalCOFI that are

3   not assimilated into ROMS, we have estimated the ROMS nowcast uncertainty to be on the order

4   of 1 ^{o}C for temperature and 0.2 psu for salinity (data not shown). It is expected that a more

5   sophisticated data assimilation scheme (e.g., ensemble Kalman Filter or 4DVAR) would further

6   reduce model uncertainty. Improvements to ROMS are underway, and these include coupling the

7   hydrodynamic model with a biogeochemical/ecosystem model (e.g., CoSiNE) with the goal of

8   providing nowcast and forecast fields for biogeochemical variables to the HAB forecasts.

9   *4.3. The Decision-Making Context*

10   Many methods of determining model skill categorize a model as poor quality if it does

11   not capture event scale activity in a particular location despite the veracity of the regional signal.

12   However, the model is likely quite valuable to a decision-maker, particularly if it provides

13   offshore, spatial information that would otherwise not be available to managers and provides

14   complementary information to nearshore monitoring data. According to Murphy (1993), in

15   addition to consistency and quality, the "goodness" of a model is measured by its *value*, or its

16   ability to aid the decision-maker in achieving some benchmark improvement. Without a high-

17   resolution, observational record with which to evaluate the true spatial and temporal skill of our

18   C-HARM system, it is difficult to quantify the added value brought to bear by having such a

19   regionally comprehensive, predictive capability. For instance, the newly launched Catalina Sea

20   Ranch located in the San Pedro Basin is the first offshore aquaculture operation within the

21   United States exclusive economic zone (http://www.catalinasearanch.com/). It stands to reason

22   that a model with the ability to predict DA events in the waters surrounding this cultivation site

23   will be an asset to business decisions, in this context. At the state level, the CDPH, which is

1   mandated to monitor biotoxins in the commercial and recreational shellfish supply, will

2   hopefully be able to better allocate sampling resources to regions with high alert levels,

3   particularly if the model signals an event outside of the normal targeted sampling zones. Marine

4   mammal resource managers already report extensive use of the routine model product to guide

5   their efforts and assess past risk likelihoods for stranded animals at the regional level. It will

6   behoove us to continue to assess the skill of the predictions in both nowcast and forecast mode to

7   demonstrate its performance over a variety of time and space scales.

8   *4.4. Towards an operational forecasting system*

9   With the completion of a yearlong pilot study to establish a HAB prediction and

10   forecasting system for coastal California, we have demonstrated the capacity to routinely

11   produce spatially explicit outlooks of the HAB horizon for neurotoxic *Pseudo-nitzschia* blooms.

12   In fact, truly operational HAB forecasts in the United States only exist for *Karenia brevis* in the

13   Gulf of Mexico at the time of this writing. The NOAA Harmful Algal Bloom Operational

14   Forecast System issues routine *K. brevis* and brevetoxin conditions reports for coastal Florida

15   and Texas (http://tidesandcurrents.noaa.gov/hab/) along with metadata in HAB Bulletins that

16   collate the observational, satellite, model, and public health information relevant to making

17   forecasts of bloom and respiratory irritation risk (Stumpf et al., 2009). While some of the

18   methods differ for the CCS and Gulf of Mexico, the most notable being the use of a Chl anomaly

19   product for detecting *K. brevis* biomass (Stumpf et al., 2003), the fundamental approach of

20   merging available information from multiple platforms to create a forecast is similar to that

21   described in this paper. The Ecological Forecasting Roadmap was established by NOAA to guide

22   the next generation of ecological forecasting services and places a strong emphasis on HABs

23   (NOAA, 2012). The goal is to include California in the suite of HAB forecasting programs on

1   the verge of transition from *Research and Development* to *Operations* in regions such as the

2   Pacific Northwest (Trainer and Suddleson, 2005), Lake Erie (Wynne et al., 2013), Chesapeake

3   Bay (Brown et al., 2012), and Gulf of Maine (Kleindinst et al., 2014; McGillicuddy et al., 2011).

4   As with these efforts, assessing model skill relative to the current baseline is a fundamental task

5   for initiating a robust early warning system for HABs.

6

21
22
23

5. References

Alvera-Azcarate, A., Barth, A., Rixen, M., Beckers, J.M., 2005. Reconstruction of incomplete oceanographic data sets using Empirical Orthogonal Functions: Application to the Adriatic Sea. Ocean Modelling 9, 325–346. doi:10.1016/j.ocemod.2004.08.001

Alvera-Azcárate, A., Barth, A., Beckers, J.M., Weisberg, R.H., 2007. Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. Journal of Geophysical Research: Oceans (1978–2012) 112.

Alvera-Azcárate, A., Barth, A., Sirjacobs, D., Beckers, J.M., 2009. Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF. Ocean Science 5, 475–485.

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. Marine Ecology Progress Series 327, 119–133.

Anderson, C.R., Kudela, R.M., Benitez-Nelson, C., Sekula-Wood, E., Burrell, C.T., Chao, Y., Langlois, G., Goodman, J., Siegel, D.A., 2011. Detecting toxc diatom blooms from ocean color and a regional ocean model. Geophys. Res. Lett. 38. doi:10.1029/2010GL045858

Anderson, C.R., Sapiano, M.R.P., Prasad, M.B.K., Long, W., Tango, P.J., Brown, C.W., Murtugudde, R., 2010. Predicting potentially toxigenic diatom blooms in the Chesapeake Bay. Journal of Marine Systems 83, 127–140. doi:10.1016/j.jmarsys2010.04.003

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. doi:10.1016/j.hal.2008.10.005

Anderson, D.M., Cembella, A.D., Hallegraeff, G.M., 2012. Progress in understanding harmful

1   algal blooms: paradigm shifts and new technologies for research, monitoring, and

2   management. Annual Review of Marine Science 4, 143–176.

3   Bargu, S., Goldstein, T., Roberts, K., Chunyan, L., Gulland, F.M.D., 2012. *Pseudo-nitzschia*

4   blooms, domoic acid, and related California sea lion strandings in Monterey Bay, California.

5   Marine Mammal Science 28, 237–253. doi:10.1111/j.1748-7692.2011.00480.x

6   Beckers, J.M., Barth, A., Alvera-Azcárate, A., 2006. DINEOF reconstruction of clouded images

7   including error maps? application to the Sea-Surface Temperature around Corsican Island.

8   Ocean Science 2, 183–199.

9   Beckers, J.M., Rixen, M., 2003. EOF calculations and data filling from incomplete

10   oceanographic data sets. Journal of Atmospheric and Oceanic Technology 20, 1839–1856.

11   Breaker, L.C., 2007. A closer look at regime shifts based on coastal observations along the

12   eastern boundary of the North Pacific. Continental Shelf Research 27, 2250–2277.

13   Breaker, L.C., Flora, S.J., 2009. Expressions of 1976-1977 and 1988-1989 Regime Shifts in Sea-

14   Surface Temperature off Southern California and Hawai'i 1. Pacific Science 63, 39–60.

15   Brown, C.W., Hood, R.R., Long, W., Jacobs, J., Ramers, D.L., Wazniak, C., Wiggert, J.D.,

16   Wood, R., Xu, J., 2012. Ecological Forecasting in Chesapeake Bay: Using a Mechanistic-

17   Empirical Modelling Approach. Journal of Marine Systems.

18   Carvalho, G.A., Minnett, P.J., Banzon, V.F., Baringer, W., Heil, C.A., 2011. Long-term

19   evaluation of three satellite ocean color algorithms for identifying harmful algal blooms

20   (*Karenia brevis*) along the west coast of Florida: A matchup assessment. Remote Sensing of

21   Environment 115, 1–18.

22   Chao, Y., Li, Z., Farrara, J.D., Moline, M., Schofield, O.M.E., Majumdar, S.J., 2008. Synergistic

23   applications of autonomous underwater vehicles and regional ocean modeling system in

1   coastal ocean forecasting. Limnology and Oceanography 53, 2251–2263.

2   Chao, Y., Z, L., Farrara, J., McWillliams, J.C., Bellingham, J., Capet, X., Chavez, F., Choi, J.K.,

3   Davis, R., Doyle, J., Frantaoni, D., Li, P.P., Marchesiello, P., Moline, M.A., Paduan, J.,

4   Ramp, S., 2009. Development, implementation and evaluation of a data-assimilative ocean

5   forecasting system off the central. Deep-Sea Research II.

6   Doyle, J.D., Yiang, Q., Chao, Y., Farrara, J., 2009. High-resolution real-time modeling of the

7   marine atmospheric boundary layer in support of the AOSNII field campaign. Deep-Sea

8   Research II.

9   Edwards, M., Johns, D.G., Leterme, S.C., Svendsen, E., Richardson, A.J., 2006. Regional

10   climate change and harmful algal blooms in the northeast Atlantic. Limnology and

11   Oceanography 51, 820–829.

12   Egbert, G.D., Erofeeva, S.Y., 2002. Efficient inverse modeling of barotropic ocean tides. Journal

13   of Atmospheric and Oceanic Technology 19, 183–204.

14   Fawcett, T., 2004. ROC graphs: Notes and practical considerations for researchers. Machine

15   learning 31, 1–38.

16   Fritz, L., Qulliam, M.A., Wright, J.L.C., Beale, A.M., Work, T.M., 1992. An outbreak of domoic

17   acid poisoning attributed to the pennate diatom *Pseudonitzschia australis*. Journal of

18   Phycology 28, 439–442.

19   Frolov, S., Kudela, R.M., Bellingham, J.G., 2013. Monitoring of harmful algal blooms in the era

20   of diminishing resources: a case study of the US West Coast. Harmful Algae 21, 1–12.

21   Glibert, P.M., Allen, J., Bouwman, A.F., Brown, C.W., Flynn, K.J., Lewitus, A.J., Madden, C.J.,

22   2010. Modeling of HABs and eutrophication: status, advances, challenges. Journal of Marine

23   Systems 83, 262–275.

1   Jin, D., Thunberg, E., Hoagland, P., 2008. Economic impact of the 2005 red tide event on

2   commercial shellfish fisheries in New England. Ocean & Coastal Management 51, 420–429.

3   Jolliffe, I.T., Stephenson, D.B., 2003. Forecast Verification. Chichester, England and Hoboken.

4   Jolliffe, I.T., Stephenson, D.B., 2012. Forecast verification: a practitioner's guide in atmospheric

5   science. John Wiley & Sons.

6   Kahru, M., Kudela, R.M., Anderson, C.R., Manzano-Sarabia, M., Mitchell, B.G., 2014.

7   Evaluation of satellite retrievals of ocean chlorophyll-a in the California Current. Remote

8   Sensing 6, 8524–8540.

9   Kleindinst, J.L., Anderson, D.M., McGillicuddy, D.J., Stumpf, R.P., Fisher, K.M., Couture, D.A.,

10   Hickey, J.M., Nash, C., 2014. Categorizing the severity of paralytic shellfish poisoning

11   outbreaks in the Gulf of Maine for forecasting and management. Deep Sea Research Part II:

12   Topical Studies in Oceanography 103, 277–287.

13   Kudela, R.M., Banas, N.S., Barth, J.A., Frame, E.R., Jay, D.A., Largier, J.L., Lessard, E.J.,

14   Peterson, T.D., Woude, A.J.V., 2008. New insights into the controls and mechanisms of

15   plankton productivity in coastal upwelling waters of the northern California Current System.

16   Oceanography 21, 46–59.

17   Kudela, R. M., A. Bickel, M. L. Carter, M. D. Howard, and L. K. Rosenfeld. 2015. The

18   Monitoring of Harmful Algal Blooms through Ocean Observing:The Development of the

19   California Harmful Algal Bloom Monitoring and Alert Program, p. 58–75. *In* Y.Y. Liu, H.A.

20   Kerkering, and R.H. Weisberg [eds.], Coastal Ocean Observing Systems. Elsevier Inc.

21   Kvitek, R.G., Goldberg, J.D., Smith, G.J., Doucette, G.J., Silver, M.W., 2008. Domoic acid

22   contamination within eight representative species from the benthic food web of Monterey

23   Bay, California, USA. Marine Ecology Progress Series 367, 35–47.

1   Lane, J.Q., Raimondi, P.T., Kudela, R.M., 2009. Development of a logistic regression model for

2   the prediction of toxigenic *Pseudo-nitzschia* blooms in Monterey Bay, California. Marine

3   Ecology Progress Series 383, 37–51.

4   Lane, J.Q., Roddam, C., Meiling, C., Langlois, G.W., Kudela, R.M., 2010. Application of Solid

5   Phase Adsorption Toxin Tracking (SPATT) for field detection of the hydrophilic

6   phycotoxins domoic acid and saxitoxin in coastal California. Limnology and Oceanography-

7   Methods 8, 645–660. doi:10.4319/lom.2010.8.0645

8   Lefebvre, K.A., Powell, C.L., Busman, M., Doucette, G.J., Moeller, P.D.R., Silver, J.B., Miller,

9   P.E., Hughes, M.P., Singaram, S., Silver, M.W., Tjeerdema, R.S., 1999. Detection of domoic

10   acid in northern anchovies and California sea lions associated with an unusal mortality event.

11   Natural Toxins 7, 85–92.

12 Lehmann, E.L., Romano, J.P., 2006. Testing statistical hypotheses. Springer Science & Business

13   Media.

14 Lewitus, A.J., Horner, R.A., Caron, D.A., Garcia-Mendoza, E., Hickey, B.M., Hunter, M.,

15   Huppert, D.D., Kudela, R.M., Langlois, G.W., Largier, J.L., 2012. Harmful algal blooms

16   along the North American west coast region: History, trends, causes, and impacts. Harmful

17   Algae 19, 133–159.

18 Li, Z., Chao, Y., McWilliams, J.C., Ide, K., 2008. A three-dimensional variational data

19   assimilation scheme for the Regional Ocean Modeling System: Implementation and basic

20   experiments. Journal of Geophysical Research 113, C05002. doi:10.1029/2006JC004042

21 Li, Z., **Chao**, Y., McWilliams, J.C., Ide, K., 2009. A Three-Dimensional Variational Data

22   Assimilation Scheme for the Regional Ocean Modeling System. *Journal of Atmospheric and*

23   *Oceanic Technology*, 25, 2074-2090.

MacKenzie, L.A., 2010. In situ passive solid-phase adsorption of micro-algal biotoxins as a
monitoring tool. Current opinion in biotechnology 21, 326–331.

Mauri, E., Poulain, P.-M., Južnič Zonta, Ž., 2007. MODIS chlorophyll variability in the northern
Adriatic Sea and relationship with forcing parameters. Journal of Geophysical Research:
Oceans (1978–2012) 112.

McCabe , R.M., B.M. Hickey, R.M. Kudela, K.A. Lefebvre, N.G. Adams, B.D. Bill, F.M.D.
Gulland, R.E. Thomson, W.P. Cochlan, and V.L. Trainer. An unprecedented coastwide toxic
algal bloom linked to anomalous ocean conditions, Geophysical Research Letters *in review*.

McGillicuddy, D.J., Jr, Townsend, D.W., He, R., Keafer, B.A., Kleindinst, J.L., Li, Y., Manning,
J.P., Mountain, D.G., Thomas, M.A., Anderson, D.M., 2011. Suppression of the 2010
Alexandrium fundyense bloom by changes in physical, biological, and chemical properties
of the Gulf of Maine. Limnology and Oceanography 56, 2411.

McKibben, S.M., Strutton, P.G., Foley, D.G., Peterson, T.D., White, A.E., 2012. Satellite‐based
detection and monitoring of phytoplankton blooms along the Oregon coast. Journal of
Geophysical Research: Oceans (1978–2012) 117.

Miles, T.N., He, R., Li, M., 2009. Characterizing the South Atlantic Bight seasonal variability
and cold‐water event in 2003 using a daily cloud‐free SST and chlorophyll analysis.
Geophys. Res. Lett. 36.

Murphy, A.H., 1993. What is a good forecast? An essay on the nature of goodness in weather
forecasting. Weather and forecasting 8, 281–293.

Nechad, B., Alvera-Azcárate, A., Ruddick, K., Greenwood, N., 2011. Reconstruction of MODIS
total suspended matter time series maps by DINEOF and validation with autonomous
platform data. Ocean Dynamics 61, 1205–1214.

1   NOAA, 2012. A strategic vision for NOAA's ecological forecasting roadmap: 2015-2019.

2   National Oceanic and Atmospheric Administration, Office of the Chief Administrative

3   Officer.

4   Page, E.S., 1961. Cumulative sum charts. Technometrics 3, 1–9.

5   Scholin, C.A., Gulland, F., Doucette, G.J., Benson, S., Busman, M., Chavez, F.P., Cordaro, J.,

6   DeLong, R., De Vogelaere, A., Harvey, J., Haulena, M., Lefebvre, K.A., Lipscomb, T.,

7   Loscutoff, S., Lowenstine, L.J., Marin, R., III, Miller, P.E., McLellan, W.A., Moeller, P.D.R.,

8   Powell, C.L., Rowles, T., Silvagni, P., Silver, M.W., Spraker, T., Trainer, V.L., Van Dolah,

9   F.M., 2000. Mortality of sea lions along the central California coast linked to a toxic diatom

10   bloom. Nature 403, 80–83.

11 Seegers, B.N., Birch, J.M., Marin III, R., C.A. Scholin, D.A. Caron, E.L. Seubert, M.D.A.

12   Howard, G.L. Robertson, B.H. Jones, 2015. Subsurface seeding of surface harmful algal

13   blooms observed through the integration of autonomous gliders, moored environmental

14   sample processors, and satellite remote sensing in southern California, Limnology and

15   Oceanography 60, 754-764.

16 Sekula-Wood, E., Benitez-Nelson, C., Morton, S.L., Anderson, C.R., Thunell, R., 2010. *Pseudo-*

17   *nitzschia* and domoic acid fluxes in Santa Barbara Basin (CA) from 1993 to 2008. Harmful

18   Algae 10, 567-575.

19 Sekula-Wood, E., Benitez-Nelson, C., Morton, S.L., Anderson, C.R., Thunell, R., 2010. *Pseudo-*

20   *nitzschia* and domoic acid fluxes in Santa Barbara Basin (CA) from 1993 to 2008. Harmful

21   Algae submitted.

22 Sekula-Wood, E., Schnezer, A., Benitez-Nelson, C., Anderson, C.R., Berelson, W.M.,

23   Brzezinski, M.A., Burns, J.M., Caron, D.A., Cetinic, I., Ferry, J.L., Fitzpatrick, E., Jones,

B.H., Miller, P.E., Morton, S.L., Schaffner, R.A., Siegel, D.A., Thunell, R., 2009. Rapid downward transport of the neurtoxin domoic acid in coastal waters. Nature Geoscience 2, 342–350. doi:10.1038/NGEO472

Shchepetkin, A.F., McWilliams, J.C., 2005. The Regional Ocean Modeling System: A split-explicit, free-surface, topography following coordinates ocean model. Ocean Modelling 9, 347–404.

Sirjacobs, D., Alvera-Azcárate, A., Barth, A., Lacroix, G., Park, Y., Nechad, B., Ruddick, K., Beckers, J.-M., 2008. Reconstruction of missing satellite total suspended matter data over the Southern North Sea and English Channel using empirical orthogonal function decomposition of satellite imagery and hydrodynamical modelling, in:. Presented at the Ocean Optics 2008.

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

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.

Stumpf, R.P., Tomlinson, M.C., Calkins, J.A., Kirkpatrick, B., Fisher, K., Nierenberg, K., Currier, R., Wynne, T.T., 2009. Skill assessment for an operational algal bloom forecast system. Journal of Marine Systems 76, 151–161.

Trainer, V.L., Adams, N.G., Bill, B.D., Stehr, C.M., Wekell, J.C., Moeller, P.D.R., Busman, M., Woodruff, D., 2000. Domoic acid production near California coastal upwelling zones, June 1998. Limnology and Oceanography 45, 1818–1833.

1   Trainer, V.L., Bates, S.S., Lundholm, N., Thessen, A.E., Cochlan, W.P., Adams, N.G., Trick,

2   C.G., 2012. *Pseudo-nitzschia* physiological ecology, phylogeny, toxicity, monitoring, and

3   impacts on ecosystem health. Harmful Algae 14, 271–300.

4   Trainer, V.L., Suddleson, M., 2005. Monitoring approaches for early warning of DA events in

5   Washington State. Oceanography 18, 228–237.

6   Wang, X., Chao, Y., Dong, C., Farrara, J., Li, Z., McWilliams, C., Paduan, J.D., Rosenfeld, L.K.,

7   2009. Modeling tides in Monterey Bay, California. Deep-Sea Research II in press.

8   Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Fahnenstiel, G.L., Dyble, J., Schwab, D.J., Joshi,

9   S.J., 2013. Evolution of a cyanobacterial bloom forecast system in western Lake Erie:

10   Development and initial evaluation. Journal of Great Lakes Research 39, 90–99.

11   Umhau, B., C.B. Benitez-Nelson, C.R. Anderson, R. C. Thunell, A. Burns, K. McCabe, and C.

12   Burrell. Water column distributions, and sinking and sediment records of *Pseudo-nitzschia*

13   and domoic acid in the Santa Barbara Basin, California, *in preparation*.

14

15

1   **TABLES**

2
3   Table 1. Generalized linear models described in Anderson et al. (2011) that are employed in the

4   routine prediction of *Pseudo-nitzschia* blooms and domoic acid (DA) events for the domain

5   covering the California coast. Key variables estimated from either MODIS-Aqua satellite data or

6   ROMS hydrodynamic model output are remote-sensing reflectance (R_{rs}) at various wavebands,

7   month of the year (month), chlorophyll (Chl), sea surface salinity (SSS), and sea surface

8   temperature (SST).

9
10

HAB Variable (Threshold)	Best-fit Logistic GLM $P_{bloom} = e^{(logit)}/[e^{(logit)} + 1]$
Pseudo-nitzschia (10^4 cells mL^{-1})	logit = 5.32 − 2.87*[R_{rs} (488/555)] − 0.165*[Month]
pDA (500 ng L^{-1})	logit = -134.3 + 0.253[Chl] + 4.0*[SSS] − 502*[R_{rs} (555)]
cDA (10 pg cell^{-1})	logit = -90.0 - 0.35*[SST] − 666*[R_{rs} (555)] + 2.87*[SSS]

11
12
13

14

15

16

17

18

19

1   Table 2. A description of the common performance metrics applied to the results of

2   routine HAB predictions for California in 2014-2015 made using the models in Table 1.

3

4

Performance Metric	Calculation	Range
Accuracy	$\dfrac{\text{hits + correct negatives}}{\text{total}}$	0 - 1
Probability of Detection (POD)	$\dfrac{\text{hits}}{\text{hits + misses}}$	0 - 1
False Alarm Ratio (FAR)	$\dfrac{\text{false alarms}}{\text{hits + false alarms}}$	0 - 1
Probability of False Detection (POFD)	$\dfrac{\text{false alarms}}{\text{correct negatives + false alarms}}$	0 - 1
Bias Score (BS)	$\dfrac{\text{hits + false alarms}}{\text{hits + misses}}$	$0 - \infty$

5

FIGURE CAPTIONS

Figure 1. **Map of California coastal and offshore DA event probabilities for 31 October 2014.** Red dots are HABMAP sampling locations where weekly HAB monitoring provides phytoplankton community and phycotoxin data. The two sites in bold (SCMW and Stearns Wharf) are the focus of the model performance analysis, while the white dots are additional CeNCOOS shore stations for which there are archived HAB data and on-going water quality observations.

Figure 2. **C-HARM Framework.** Primary data streams or validation datasets for the HAB empirical models are shown in squares. Polygons highlight the computational steps required to statistically reconstruct gaps in satellite imagery (DINEOF), merge data to compute nowcasts of HABs (HAB empirical models), and project the biological HAB data forward in time (advection scheme) to generate 3-day forecasts. *In situ* and crowdsourced data streams are integrated into the forecast and analysis tool for near real-time quality control and hindcast validation.

Figure 3. **C-HARM Web Display**. Routine nowcasts and forecasts of *Pseudo-nitzschia* blooms and DA events have been produced since February 2014 and broadcast to the public on the CeNCOOS website http://www.cencoos.org/data/models/habs. Additional links to information and tutorial pages offer detailed descriptions of model construct and interpretation, and the Previous Conditions page points to directories of archived imagery. Transition to the CeNCOOS interactive data portal took place in Spring 2015, and time series of the predicted values can be obtained and downloaded for any pixel selected on the map.

1

2   Figure 4. **C-HARM Web Display**. The Latest Conditions page contains expansion links to (a)

3   gap-filled Chl imagery from DINEOF run on MODISA retrievals (mg m^{-3}), and (b) modeled

4   SST (°C) from the 3-km CA ROMS. Note that the pDA HAB probability model requires SSS

5   forecasts (not displayed), which are also provided by ROMS.

6

7   Figure 5. **Spring 2014 DA event in the Monterey Bay region.** The upper panel describes the

8   time evolution of marine mammal strandings from DA toxicosis, particulate DA (pDA) at the

9   SCMW, and modeled pDA at the SCMW pixel (36.7 °N, 122.4 °W, red circle) and a pixel much

10   further offshore in the CCS (36.0 °N , 124.5°W, white circle). Grey bars correspond to 1) the

11   first reports of mammal strandings from The Marine Mammal Center (16-17 Mar), 2) the rise in

12   particulate DA (pDA) at the SCMW on 26 Mar, and 3) the CDPH advisory against harvesting

13   anchovies or sardines from the Monterey Bay on 1 Apr. The bottom two images are particulate

14   and cellular DA nowcast imagery displayed on the CeNCOOS website for 26 Mar.

15

16   Figure 6. **Fall 2014 DA event in the Santa Barbara Channel region.** The upper panel

17   describes the time evolution of modeled pDA at the Stearns Wharf pixel (red circle in the 11 Oct

18   image, 34.4 °N, 119.6 °W) and a pixel southwest of the Santa Barbara Channel (white circle in

19   the 11 Oct image, 33.5 °N , 121.0°W). Grey bars correspond to 1) the CDPH advisory against

20   harvesting benthic invertebrates such as crabs and lobster on 11 Oct, and 2) the timing of an

21   offshore cruise of opportunity in Monterey Bay 15-17 Oct that validated the low DA and

22   relatively high *Pseudo-nitzschia* in that region at the time (black and white circles on inset map).

23   The bottom three images are pDA nowcast imagery for the 11 Oct and pDA and *Pseudo-*

1   *nitzschia* for the 15 Oct displayed on the CeNCOOS website (inset maps expanded for Monterey

2   Bay).

3

4   Figure 7. Modeled bloom/event probabilities at the **SCMW**. (A) Weekly microscopic counts of

5   *Pseudo-nitzschia seriata* and *delicatissima* size classes (combined; cells x 10^3 L^{-1}; K. Negrey)

6   from 19 February 2014 to 7 January 2015 are represented as discrete points (closed circles) in

7   comparison with daily modeled values at the SCMW pixel (solid line). The solid dark line is the

8   linear trend for the PN model time series (r = -0.51, p<0.001). (B) Weekly measurements of

9   particulate DA (pDA) by LC-MS (ng mL^{-1}; R. Kudela and K. Negrey) from 19 February 2014 to

10   7 January 2015 are represented as discrete points (closed squares) in comparison with daily

11   modeled values (solid line). In addition to phytoplankton pDA, measurements of DA from

12   mussel tissue collected at the SCMW (open circles), DA recovered from SPATT-HP20 mesh

13   bags (open triangles), and monthly mean counts of marine mammals admitted to The Marine

14   Mammal Center with DA toxicosis (open stars) are compared with pDA and modeled DA events

15   at the SCMW (solid line). Gaps in data collection are shown as empty space. The solid dark line

16   is the linear trend for the pDA model time series (r = -0.51; p<0.001).

17

18   Figure 8. **Santa Cruz Municipal Wharf (SCMW).** Cross-correlation functions (CCF) describe

19   the correlational relationship between the SCMW observations and the nearest SCMW model

20   pixel over a range of lead-lag time scales. Negative lags occur when the model leads

21   observations. The dashed lines show the 95% confidence interval.

22

1   Figure 9. **Santa Cruz Municipal Wharf (SCMW)** : **Particulate DA model-observation**

2   **matchups for mussel tissue and SPATT sampling**. Cross-correlation functions (CCF) describe

3   the correlational relationship between the pDA model at the SCMW pixel and (A) mussel tissue

4   DA from the SCMW, (B) Solid Phase Adsorption Toxin Tracking (SPATT) HP20 resin bags

5   deployed at the SCMW DA from the SCMW, and (C) monthly mammal strandings over a range

6   of lead-lag times. Negative lags occur when the model leads observations. The dashed lines show

7   the 95% confidence interval.

8

9   Figure 10. **Modeled bloom/event probabilities at Stearns Wharf.** (A) Weekly microscopic

10   counts of *Pseudo-nitzschia seriata* and *delicatissima* size classes (combined; cells x 10^5 L^{-1}; H.

11   McNair) from 17 February 2014 to 12 January 2015 are represented as discrete points (closed

12   circles) in comparison with daily modeled values at the wharf (solid line) and at an offshore pixel

13   (dashed line) at 33.5 °N, 121 °W. Bloom level is considered to be 0.1 x 10^5 cells L^{-1}. The solid

14   dark line is the linear trend for the PN model time series (r = 0.04; ns). (B) Weekly

15   measurements of particulate DA (pDA) by ELISA (ng mL^{-1}; D. Caron) from 17 February 2014

16   to 4 August 2014 are represented as discrete points (closed squares) in comparison with daily

17   modeled values at the wharf (solid line) and at an offshore pixel (dashed line) at 33.5 °N, 121 °W.

18   Gaps in data collection are shown as empty space. DA concentrations above 500 ng L^{-1} are

19   considered alert level. The solid dark line is the linear trend for the pDA model time series (r =

20   0.10; ns).

21

22   Figure 11. **Stearns Wharf.** Cross-correlation functions (CCF) describe the correlational

23   relationship between the Stearns Wharf observations and the nearest model pixel over a range of

1   lead-lag time scales. Negative lags occur when the model leads observations. The dashed lines

2   show the 95% confidence interval.

3

4   Figure 12. **Modeled bloom/event probabilities at All Stations.** Weekly microscopic counts of

5   *Pseudo-nitzschia seriata* and *delicatissima* size classes (combined; cells x 10^5 L^{-1}; H. McNair)

6   from 17 February 2014 to 12 January 2015 are represented as discrete points (closed circles) in

7   comparison with daily modeled values (solid line). Bloom level is considered to be 0.1 x 10^5

8   cells L^{-1}. The solid line in the lower panel is the linear trend line (r = 0.37; p<0.001).

9

10   Figure 13. **Contingency plots, ROC Curves for All Stations** compare *Pseudo-nitzschia* binary

11   observations (transformed using bloom threshold) with modeled probabilities at the nearest pixel

12   for each monitoring station. (A) Contingency plot metrics are accuracy, probability of detection

13   (POD), false alarm rate (FAR), probability of false detection (POFD), and bias score (BS); the

14   vertical line is the optimized prediction point (0.52). (B) ROC curve; the value in the box is the

15   area under the curve (AUC). A value near 0.5 suggests a model that performs comparable to

16   random chance. (C) Cross-correlation functions (CCF) describe the correlational relationship

17   between the *Pseudo-ntizschia* observations at all stations and the nearest model pixel over a

18   range of lead-lag time scales. Negative lags occur when the model leads observations. The

19   dashed lines show the 95% confidence interval.

20

21   Figure 14. **Contingency plots and ROC Curves to assess model performance at the SCMW.**

22   (A) *Pseudo-nitzschia* and (B) pDA binary observations (transformed using bloom/event

23   thresholds) are compared with modeled probabilities at the SCMW pixel. Metrics used are

1   accuracy, probability of detection (POD), false alarm rate (FAR), probability of false detection

2   (POFD), and bias score (BS); the vertical line is the prediction point – 0.68 for *Pseudo-nitzschia*

3   and 0.6 for pDA. ROC curves compare (C) *Pseudo-nitzschia* and (D) pDA binary observations

4   (transformed using bloom/event thresholds) with modeled probabilities at the SCMW pixel.

5   Values in boxes are the area under the curve (AUC), and values along the ROC curve show

6   positions of various prediction points along the range 0-1 used to model POFD vs. POD.

7

8   Figure 15. **Contingency plots and ROC Curves to assess model performance at Stearns**

9   **Wharf**. (A) *Pseudo-nitzschia* and (B) pDA binary observations (transformed using bloom/event

10   thresholds) are compared with modeled probabilities at the SCMW pixel. Metrics used are

11   accuracy, probability of detection (POD), false alarm rate (FAR), probability of false detection

12   (POFD), and bias score (BS); the vertical line is the optimized prediction point– 0.55 for *Pseudo-*

13   *nitzschia* and 0.63 for pDA. ROC curves compare (C) *Pseudo-nitzschia* and (D) pDA binary

14   observations (transformed using bloom/event thresholds) with modeled probabilities at the

15   SCMW pixel. Values in boxes are the area under the curve (AUC), and values along the ROC

16   curve show positions of various prediction points along the range 0-1 used to model POFD vs.

17   POD.

18

19   Figure 16. **DINEOF-reconstructed chlorophyll** (A) extracted from a 3x3 box around the

20   location of the SCMW using consecutive 180-day intervals for 10 days (red lines); extracted

21   from a 1x1 pixel closest to the SCW using 365 days to run DINEOF (solid blue lines); and

22   extracted from a 3x3 box around the SCMW using 365 days of data but truncated to the first 180

23   days (dashed blue lines). In (B), the blue lines remain the same, but the red lines are now the

1   closest 1x1 pixel to the SCW and more closely track the solid blue line. The right panel shows

2   matchups between SCMW and ROMS (C) salinity and (D) temperature, where blue = SCMW

3   (calibrated thermometer; YSI-3100), green = SCW (YSI Sonde-6600), and red = ROMS (nearest

4   3km grid point).

5

6   Figure 17. **Sensitivity of the HAB statistical model to estimated or simulated variables**.

7   Particulate DA probability at the SCMW for the study period is calculated with observed

8   parameters (x-axis) measured at the SCW (salinity, chlorophyll) and DINEOF-filled R_{rs}(555)

9   values. These values are plotted against pDA probability calculated using ROMS salinity and

10   DINEOF-filled ocean color values (filled square, "model"), as shown for previous figures;

11   substituting ROMS salinity with laboratory-measured salinity from a YSI-3100 at the SCW

12   (open square, "SALT"); and substituting DINEOF-filled Chl values with filter-pad measured Chl

13   from the SCW (asterisk, "CHL"). Dashed and solid lines are linear fits to the data, and R^2 values

14   are reported in the legend; slopes (y/x) are reported on the line itself. The model is most sensitive

15   to substitution of observed salinity in place of simulated salinity as compared with Chl.

16

17

18

19

20

21

22

23
24

2014-10-31 particulate domoic acid probability

Probability of DA event

Labels on map: Humboldt, Bodega Marine Lab, Santa Cruz Municipal Wharf, Monterey Wharf, Cal Poly Pier, Pt. Conception, Stearns Wharf, Santa Monica Pier, Newport Pier, Scripps Pier

REMOTE SENSING OCEAN COLOR & SST

2-D satellite images
(MODISA, soonVIIRS)

DINEOF

INTERPOLATED OCEAN COLOR & SST

2-D satellite images
Error estimates
($R_{rs}(\lambda)$)

CCS-tuned Algorithms

RECONSTRUCTED OCEAN COLOR & SST

2-D Chl, Rrs (0^+, λ) imagery

IN SITU OBSERVATIONAL DATA

Oceanographic and meteorological observations
(current meter, salinity, temp, winds)
Biological/H_2O quality field observations
(water sample, toxin, shellfish toxicity)
Cal-HABMAP pier data
(Species composition, toxicity)
Motion sensor data
(AUV, glider, etc)

INTEGRATED NOWCAST & FORECAST FRAMEWORK

FORECAST & ANALYSIS TOOL
(application of forecast guidelines and rules, mapping tool)

HYDRODYNAMIC MODEL FIELDS

3-D ROMS 3-day Forecast
(currents, salinity, temperature)

HAB Empirical Models

BIOLOGICAL MODEL FIELDS

3-D Nowcasts of Bloom and Toxicity Probability

CROWDSOURCED DATA

Expanded HAB Monitoring Network
(HABMAP, TMMC stranding data, JellyWatch/ HAB App for marine mammal strandings,)

BIOLOGICAL MODEL FIELDS

3-D 3-day Forecasts of Bloom and Toxicity Probability

Advection Scheme

Figure 3

2015-02-03 EOF filled satellite chlorophyll

A

2015-02-03 model temperature (deg C)

B

Santa Cruz Municipal Wharf - Particulate Domoic Acid (pDA)

2014-03-26 particulate domoic acid probability
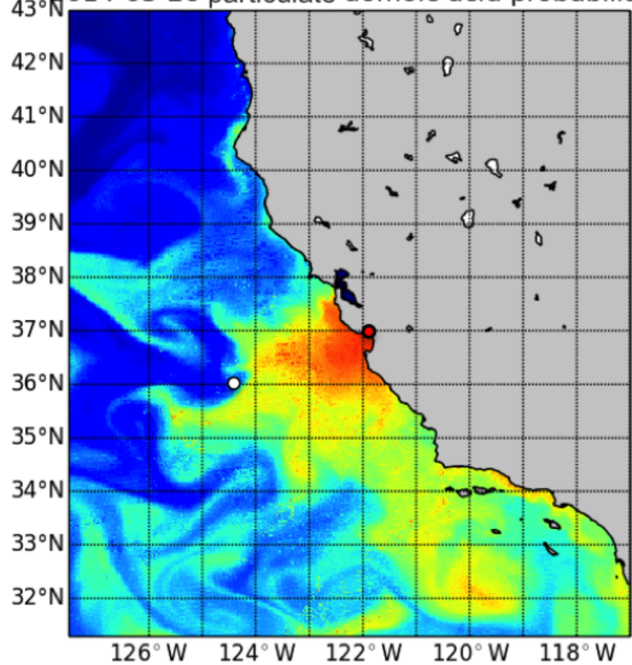
2014-03-26 cellular domoic acid probability

Stearns Wharf - Particulate Carbon Hort (cDS)
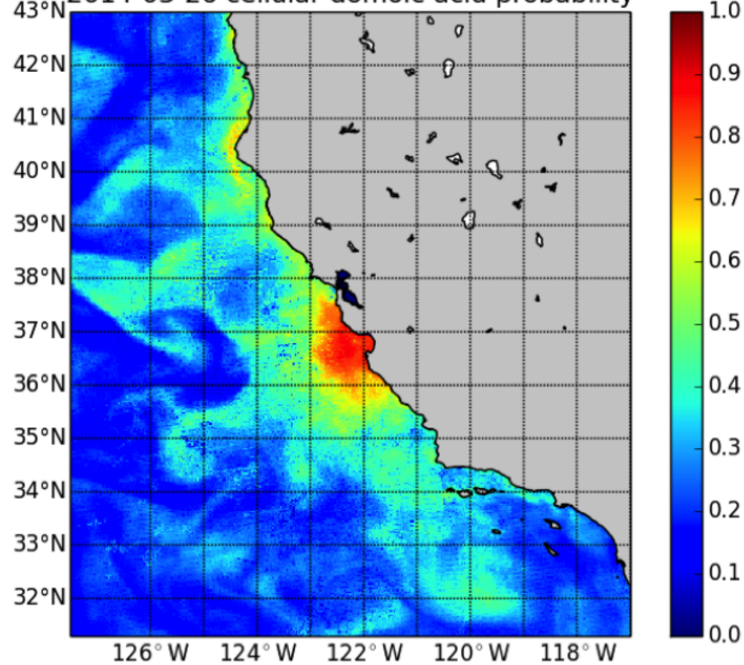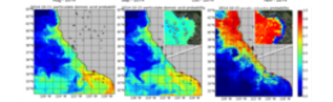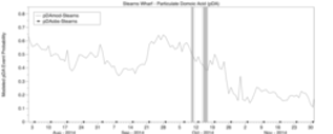
Santa Cruz Municipal Wharf - *Pseudo-nitzschia* spp.

A

r = 0.37

PNmod-SCMW
PNobs

Santa Cruz Municipal Wharf - Particulate Domoic Acid (pDA)

Legend:
- pDAmod-SCMW
- pDAobs-phyto
- pDAobs-mussel
- monthly strandings
- totalDA-SPATT

r = -0.51

Y-axis (left): Modeled pDA Event Probability

Y-axis (right): Monthly Mean Marine Mammal Strandings from DA Poisoning; Observed DA in Mussel Tissue (ppm); Total DA from SPATT HP20 Resin (ppm); Observed DA in Phytoplankton (ng/mL)

X-axis: Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, Jan, Feb, Mar — 2014 to 2015

Stearns Wharf - *Pseudo-nitzschia* spp.
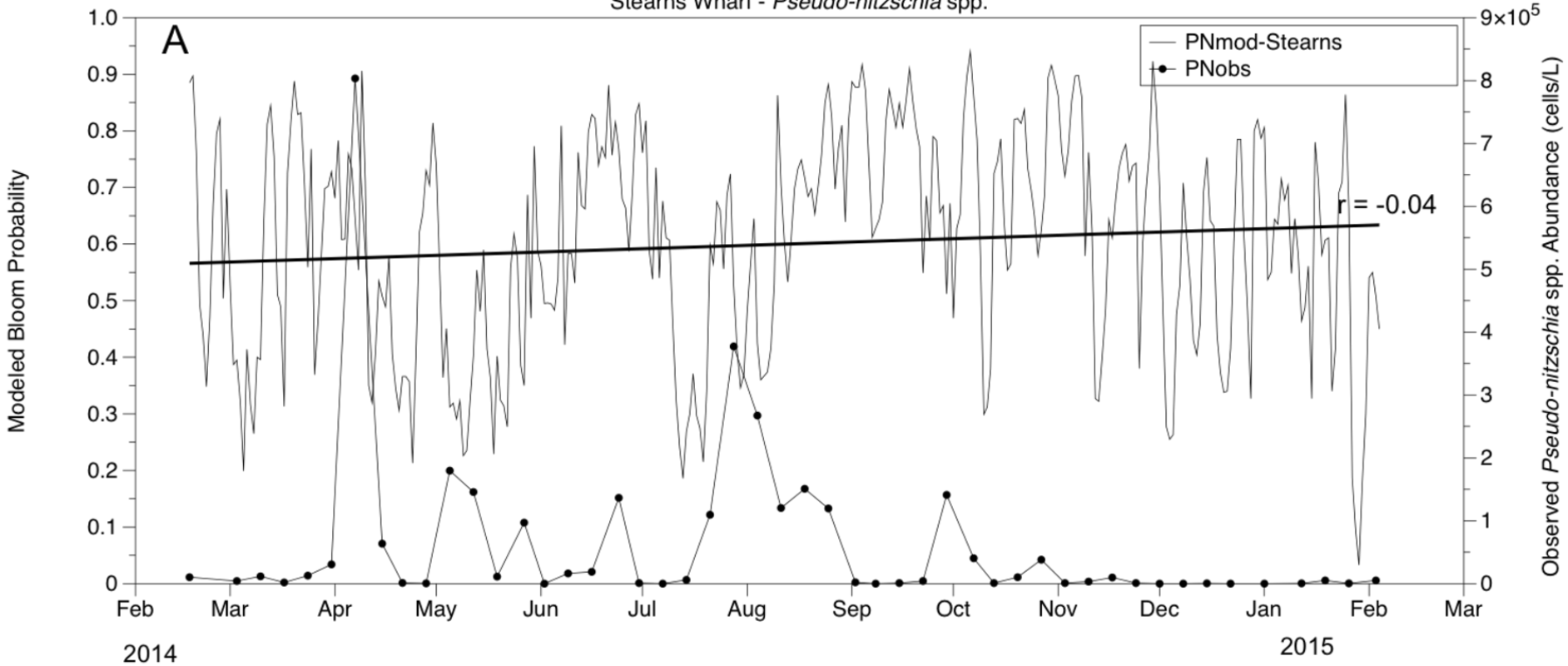
Stearns Wharf - Particulate Domoic Acid (pDA)
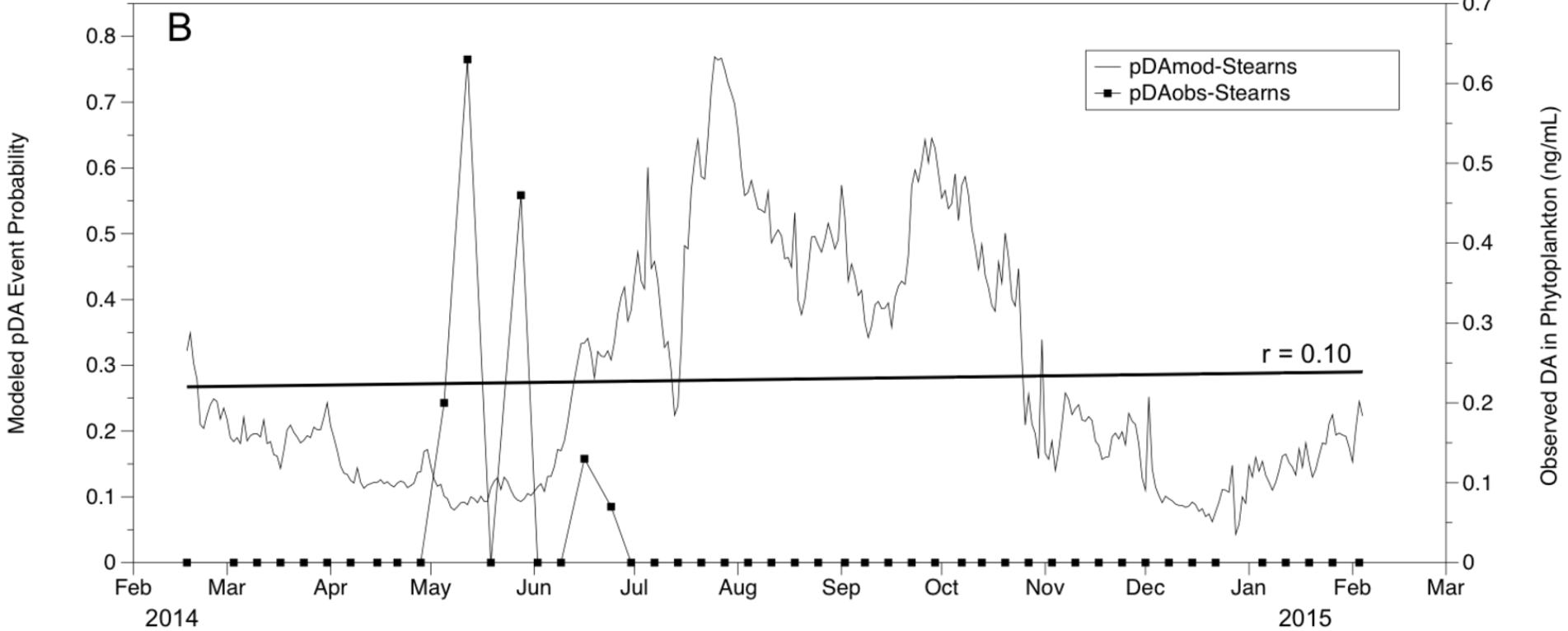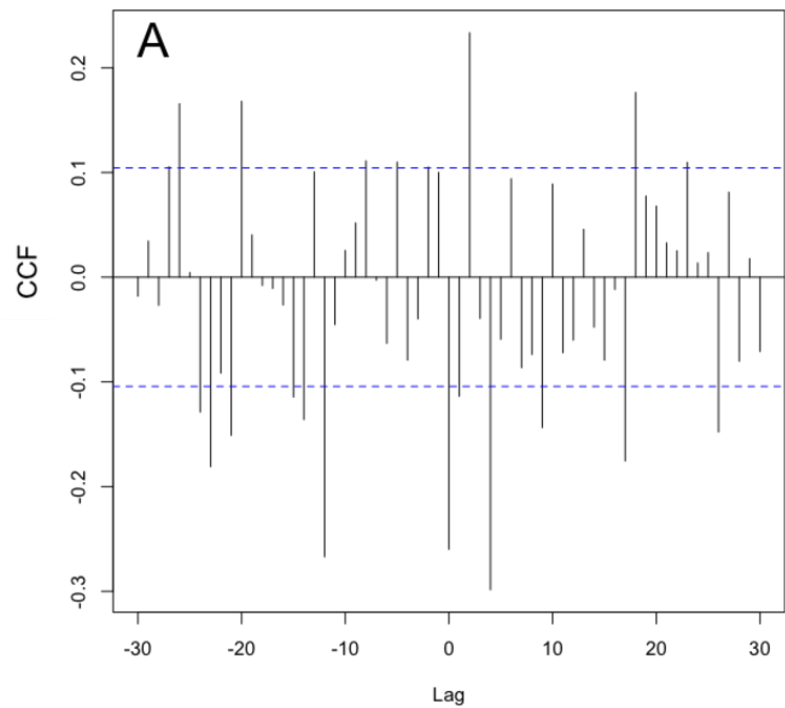
Pseudo-nitzschia   Particulate DA
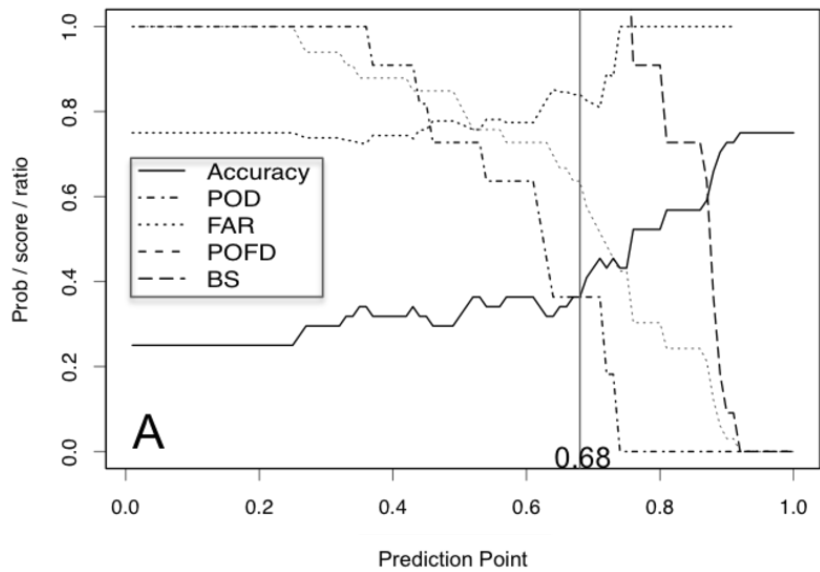
Pseudo-nitzschia

Legend:
- Accuracy
- POD
- FAR
- POFD
- BS

0.68

Particulate DA

0.6

AUC = 0.33

0.01
0.50
0.60
0.70
0.90

AUC = 0.77

0.50
0.01
0.60
0.70
0.90

C

D

Pseudo-nitzschia

Accuracy
POD
FAR
Spec
BS

0.55

Particulate DA

0.63

AUC = 0.42

0.40

0.70

AUC = 0.04

0.90

y/x = 0.40

y/x = 0.58

y/x = 0.71
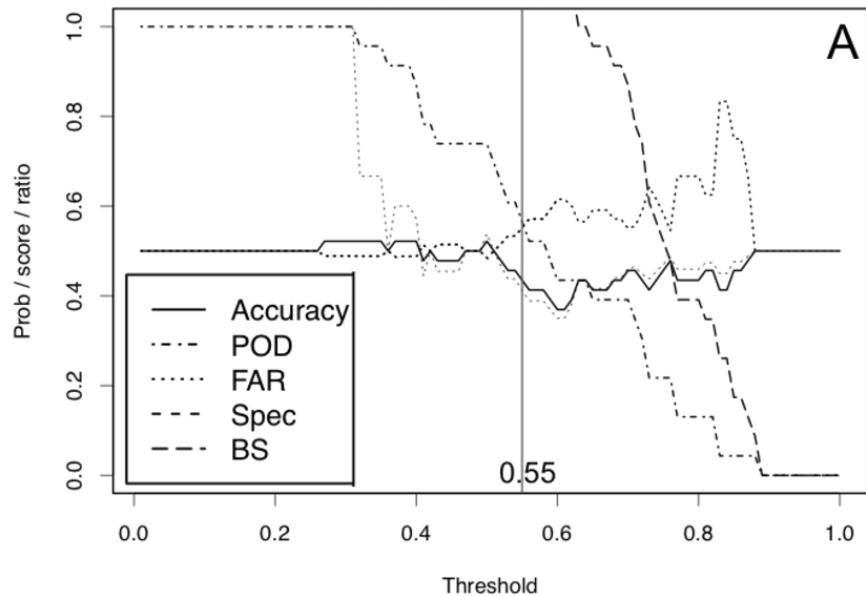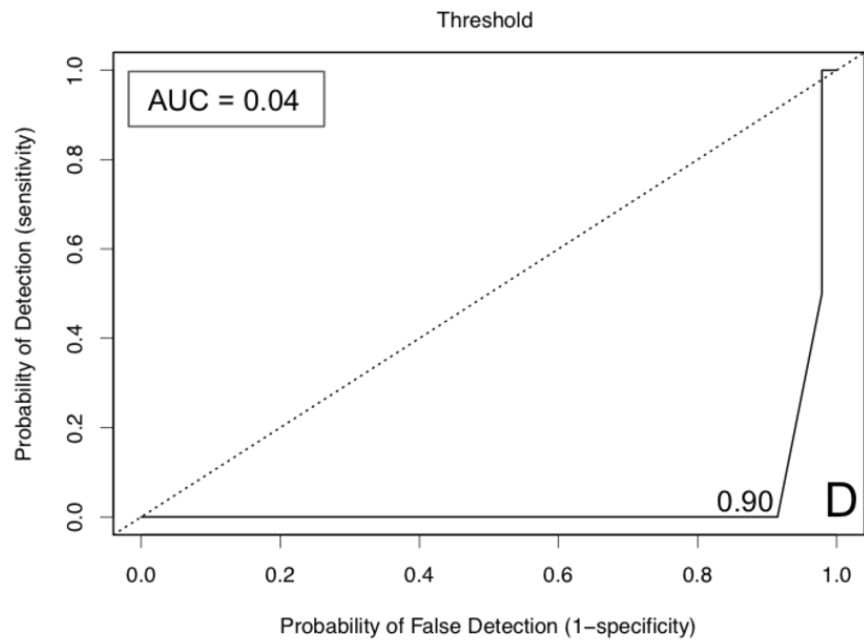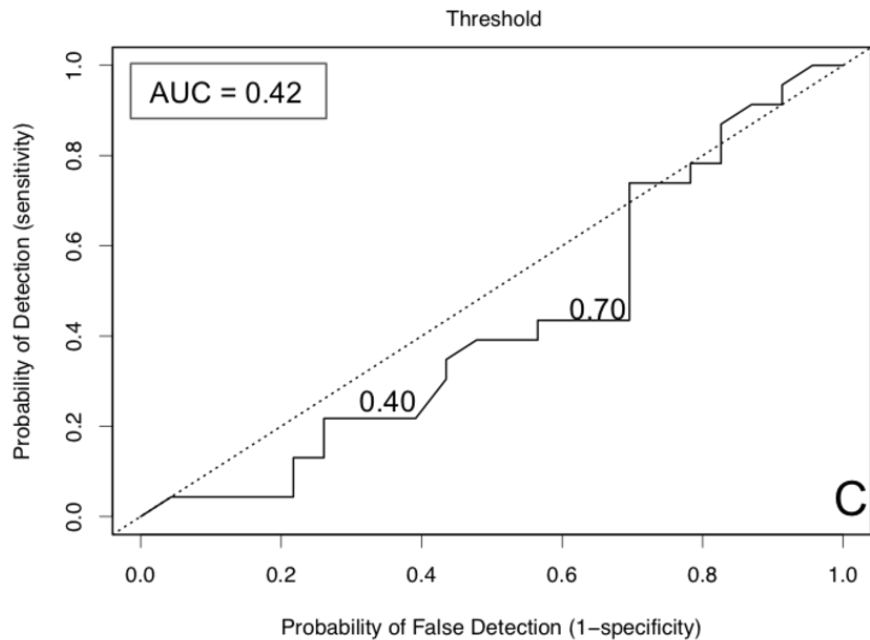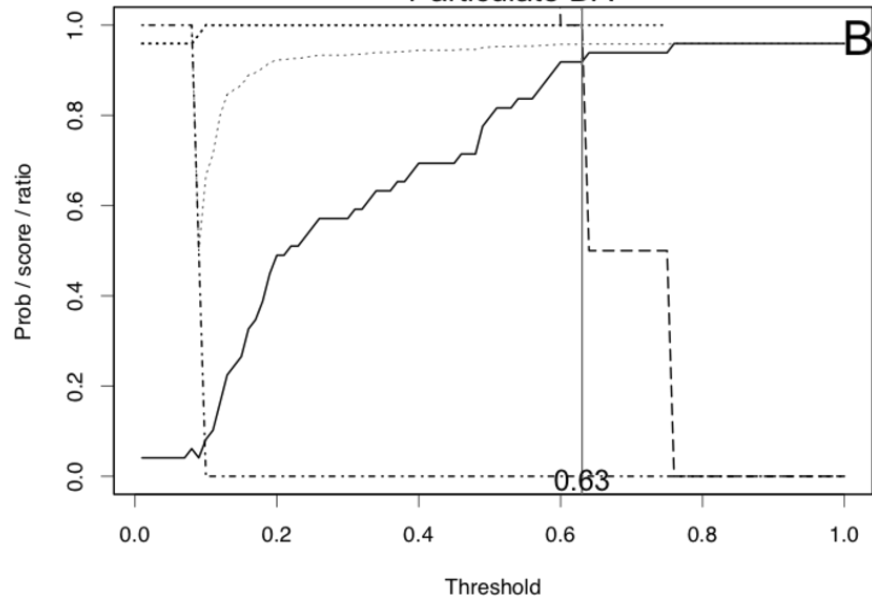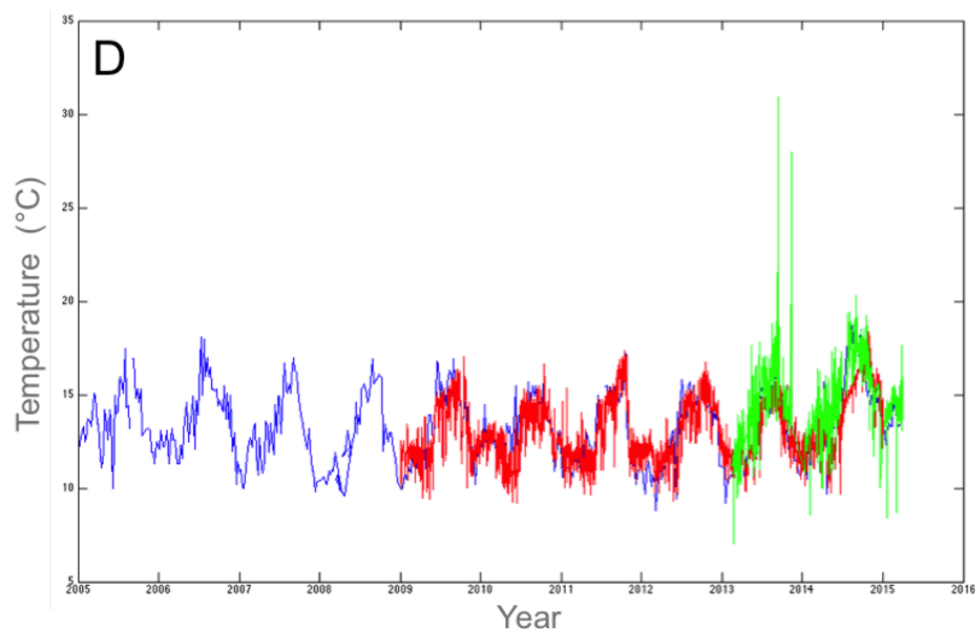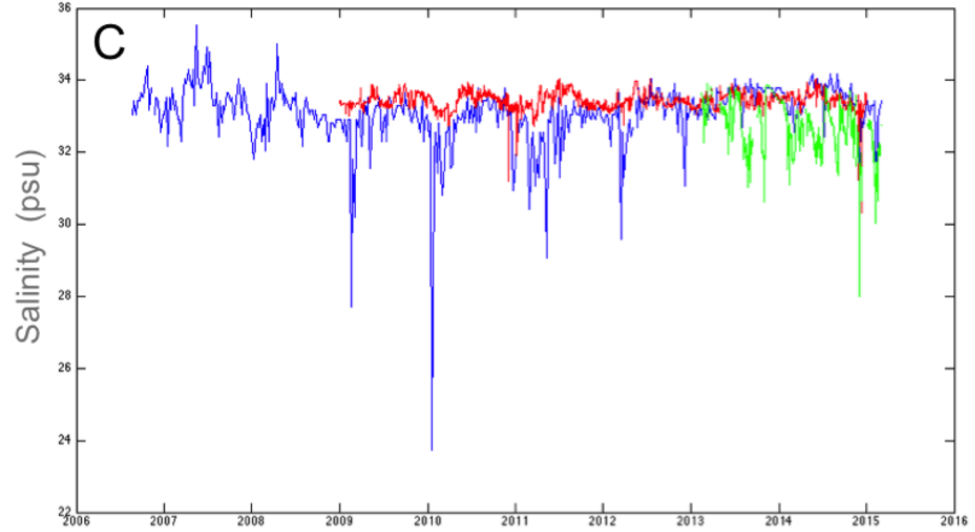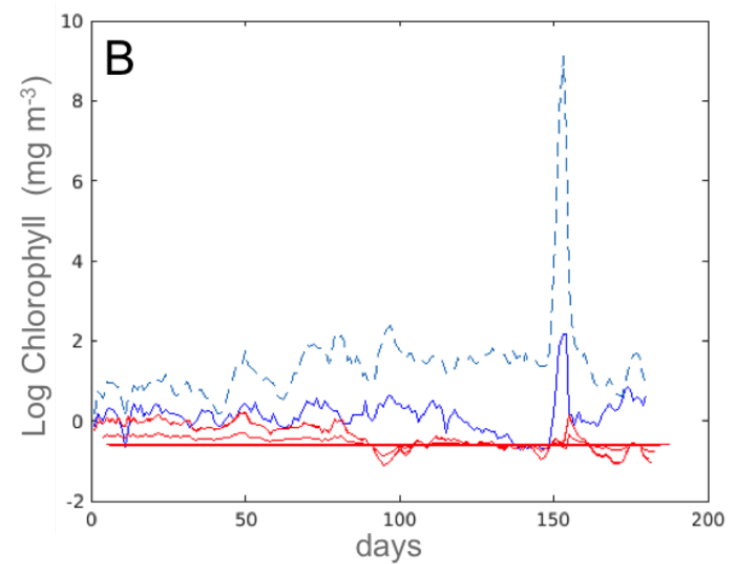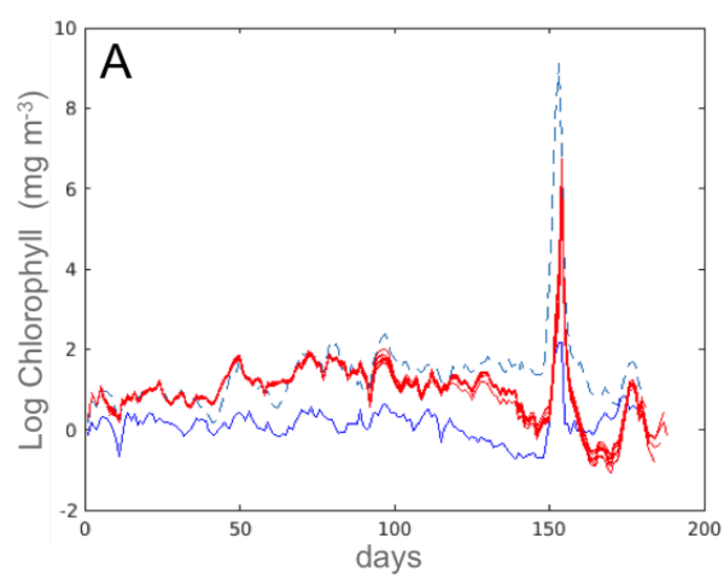
MODELED Particulate Domoic Acid Probability

Particulate Domoic Acid Probability calculated from OBSERVED data

- PDA Model
- PDA Salt
- PDA CHL
- Fit of PDA Model 0.083
- Fit of PDA Salt 0.37
- Fit of PDA CHL 0.16