Harmful Algae 59 (2016) 1–18

Contents lists available at ScienceDirect

Harmful Algae

journal homepage: www.elsevier.com/locate/hal

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

Clarissa R. Anderson [a,*], Raphael M. Kudela [b], Mati Kahru [c], Yi Chao [d,e], Leslie K. Rosenfeld [f], Frederick L. Bahr [f], David M. Anderson [f], Tenaya A. Norris [g]

[a] Institute of Marine Sciences, University of California, Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA
[b] Ocean Sciences Department, University of California, Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA
[c] Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive # 0218, La Jolla, CA, 92093, USA
[d] Joint Institute for Regional Earth System Science and Engineering University of California, Los Angeles, 607 Charles E Young Drive, Los Angeles, CA 90095, USA
[e] Remote Sensing Solution, Monrovia, CA 91016, USA
[f] Central and Northern California Ocean Observing System, Monterey Bay Aquarium Research Institute, 7700 Sandholdt Rd., Moss Landing, CA 95039, USA
[g] The Marine Mammal Center, 2000 Bunker Road, Fort Cronkite, Sausalito, CA 94965, USA

ARTICLE INFO

Article history:
Received 17 September 2015
Received in revised form 22 August 2016
Accepted 27 August 2016

Keywords:
Domoic acid
DINEOF
Ocean color
ROMS
Ecological forecasting
Ocean observing

ABSTRACT

Toxic algal events are an annual burden on aquaculture and coastal ecosystems of 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 to 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 ocean ecosystems (D.M. Anderson et al., 2012). A call for predictive capabilities for integrated environmental assessments, early warning systems, action plans, and mitigation strategies was articulated at the West Coast Governors Alliance Harmful Algal Bloom Summit (Lewitus et al., 2012) and is a major component of the strategic vision for a national Ecological Forecasting Roadmap presented by NOAA (http://oceanservice.noaa.gov/ecoforecasting/noaa-ecoforecasting-roadmap.pdf). Domoic acid (DA) poisoning in marine mammal/bird populations and the threat of Amnesic Shellfish Poisoning (ASP) in humans is now considered to be the leading HAB and conservation issue for much of the U.S. west coast (Kudela et al., 2008; Lewitus et al., 2012; Trainer et al., 2012) since these events have the potential to be devastating to aquatic life via bioaccumulation in the food web (Fritz et al., 1992; Lefebvre et al., 1999; Scholin et al., 2000). There are many as-yet unquantified indirect effects, such as those from reduced demand for seafood following a publicized HAB event (Jin et al., 2008) and negative ecosystem consequences in the water column (Sekula-Wood et al., 2010). While DA events are unlikely to be reported until a highly visible impact occurs, such as DA

* Corresponding author.
E-mail address: clrander@ucsc.edu (C.R. Anderson).

http://dx.doi.org/10.1016/j.hal.2016.08.006
1568-9883/

poisoning of California sea lions or the closing of commercial shellfish grounds, with each passing year since 1998, these reports have become more numerous. Clearly, there are significant societal and scientific benefits to predicting when and where these DA events occur relative to susceptible populations of marine mammals, birds, and consumable shellfish.

Due to the fact that optical signals of *Pseudo-nitzschia* abundance and more importantly, DA, have not yet been successfully isolated from the bulk chlorophyll signal using our current constellation of multi-spectral ocean color satellite sensors, other methods have been examined. McKibben et al. (2012) assessed the use of a chlorophyll anomaly product for isolating likely *Pseudo-nitzschia* spp. blooms in Oregon coastal waters, akin to that applied operationally to *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 to 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 to 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 availability. The RS models therefore predict toxic events solely from remote-sensing reflectance values (measured with an in-water radiometer), sea-surface temperature, and salinity (Table 1), whereby surface temperature and salinity values can be estimated from either satellite sensors or numerical model output produced by regional circulation models (Anderson et al., 2011).

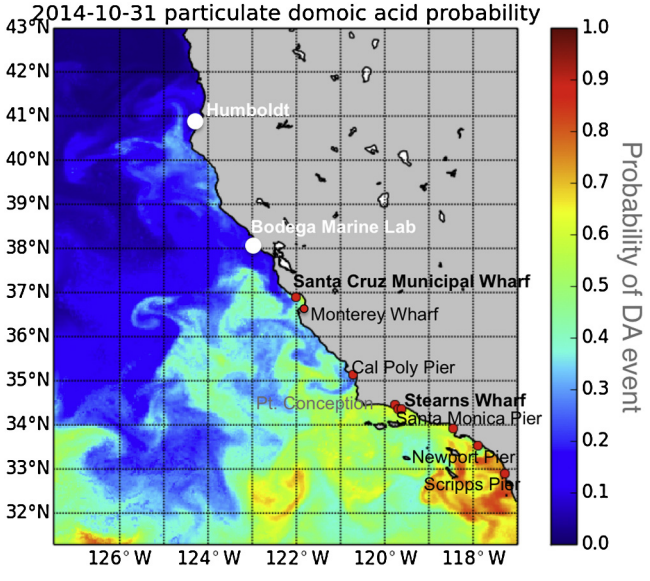

Fig. 1. Map of California coastal and offshore DA event probabilities for 31 October 2014. Red dots are HABMAP sampling locations where weekly HAB monitoring provides phytoplankton community and phycotoxin data. The two sites in bold (SCMW and Stearns Wharf) are the focus of the model performance analysis,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

Table 1
Generalized linear models described in Anderson et al. (2011) that are employed in the routine prediction of *Pseudo-nitzschia* blooms and domoic acid (DA) events for the domain covering the California coast. Key variables estimated from either MODIS-Aqua satellite data or ROMS hydrodynamic model output are remote-sensing reflectance (R_{rs}) at various wavebands, month of the year (month), chlorophyll (Chl), sea surface salinity (SSS), and sea surface temperature (SST).

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]

Satellite observations of sea surface temperature (SST) and ocean color have the potential to detect long-term trends due to their high temporal frequency and almost global coverage but are limited by their relatively short duration and cloud cover (Frolov et al., 2013), as well as being restricted to the upper third of the euphotic zone (one optical depth). Thus, a major challenge to developing an operational predictive capability for *Pseudo-nitzschia* blooms has been the acquisition of temporally and spatially coherent, real-time satellite data with which to compute the empirical HAB models. Temporal and spatial averaging is routinely used to reduce the gaps due to cloud cover; however, averaging over scales larger than the decorrelation scale blurs the mesoscale features and limits the utility of satellite data in decision making or for interpretation of underlying bio-physical processes. To quantify the combined effect of cloud cover and satellite coverage on Moderate Resolution Imaging Spectroradiometer (MODIS) imagery, Frolov et al. (2013) compared the statistics of temporal gaps between two consecutive pixels using a fluorescence line height (FLH) time series from 2002 to 2008 and also estimated spatial decorrelation scales for the same dataset. The shortest gaps between two consecutive pixels were off central California and in the Santa Barbara Channel (3–4 days), while for many places along the California coast (north of San Diego, off Point Sur, and between Cape Mendocino and Point Arena), satellite coverage was exceptionally poor in the first 4–10 km distance from the coast (mean gap of 8 to 10 days). Thus, from the perspective of operational or semi-operational use of ocean color to inform decisions about extreme blooms or toxic events, historical and existing ocean color sensors are inadequate.

One method for addressing these satellite data gaps employs Data Interpolating Empirical Orthogonal Function (DINEOF) routines (Alvera-Azcárate et al., 2005; Beckers and Rixen, 2003) to solve spatial and temporal EOFs using either single data sets (e.g. chlorophyll) or covarying bio/geophysical data (e.g., chlorophyll, SST), thereby statistically reconstructing the full dataset after convergence on an optimized number of EOFs (Beckers and Rixen, 2003). This method has been used to reconstruct SST fields of

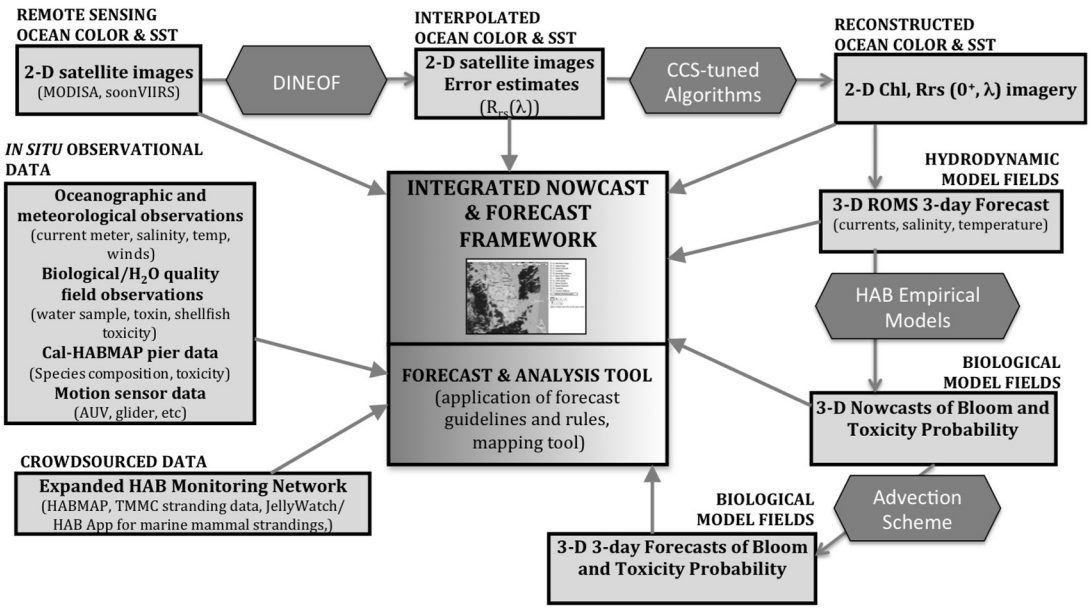

Fig. 2. C-HARM framework. Primary data streams or validation datasets for the HAB empirical models are shown in squares. Polygons highlight the computational steps required to statistically reconstruct gaps in satellite imagery (DINEOF), merge data to compute nowcasts of HABs (HAB empirical models),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

satellite data (Alvera-Azcárate et al., 2005; Beckers et al., 2006), and there has been increasing interest and application of this method for other variables, including chlorophyll (Alvera-Azcárate et al., 2007; Mauri et al., 2007; Miles et al., 2009) and total suspended matter (Nechad et al., 2011; Sirjacobs et al., 2008). Here, we describe application of DINEOF to MODIS-Aqua (MODISA) ocean color data to create continuous imagery for use in operational forecasting of HABs in the CCS. The approach capitalizes on the ability to use DINEOF to forecast biogeochemical fields forward in time without the use of a coupled physical-biological model. Instead, the forecast method suggested here (Alvera-Azcárate et al., 2007) uses a multivariate approach to estimate missing satellite data by combining numerical forecasts of physical fields and near real-time estimates of remote sensing parameters to forecast the ocean color fields up to three days in advance. A prototype prediction system for *Pseudo-nitzschia* blooms and DA events in the CCS (Fig. 2) has been running routinely since February 2014 and is now available online via the Central and Northern California Ocean Observing System (CeNCOOS) website (Fig. 3). In this paper, the first year of California Harmful Algae Risk Mapping (C-HARM) model-observation matchups are evaluated for coherence in the temporal patterns and performance metrics commonly used in decision-making.

2. Methods

2.1. ROMS forecasts

A state-of-the-art physical circulation model is crucial to forecasting HABs in the coastal zone. As with the Anderson et al. (2011) study, the model employed here is based on the Regional Ocean Modeling System (ROMS), a community-based hydrodynamic model designed for regional applications (Shchepetkin and McWilliams, 2005) with a vertical, coordinate-following bottom topography (Song and Haidvogel, 1994). The ROMS configuration consists of a single domain covering the entire California coastal ocean from north of Crescent City, CA to Ensenada, Mexico and extending approximately 1000 km offshore at a resolution of 3.3 km. There are 40 unevenly-spaced sigma vertical layers with the majority of these clustered near the surface to better resolve processes in the mixed layer. The non-tidal lateral boundary conditions for the ROMS domain are provided by the real-time operational nowcast/forecast from the global 1/12° HYCOM (see http://hycom.org). The tidal forcing is added through lateral boundary conditions that are obtained from a global barotropic tidal model TPXO.6 (Egbert and Erofeeva, 2002). The atmospheric forcing required by the ROMS model is derived from hourly output from operational forecasts performed with the NCEP NAM 5-km North American model. Every six hours, both *in situ* and remotely sensed (from both satellite and land-based high-frequency radar) observations (e.g., SST, sea surface height, surface current, vertical profiles of temperature and salinity) are assimilated into ROMS using the multi-scale (MS) three-dimensional variational (3DVar) algorithm (Li et al., 2008, 2009) to produce nowcast fields. Every day, a 72-hour forecast is produced using the nowcast at 03 UTC as the initial condition. One of the earliest ROMS configurations is centered on Monterey Bay and has been used to support a well-described 2003 field experiment (Chao et al., 2008; Chao et al., 2009; Doyle et al., 2009; Wang et al., 2009). The Integrated Ocean Observing System (U.S. IOOS) regional associations for central (CeNCOOS) and southern California (SCCOOS) now host the 3-km CA ROMS nowcast and 72-hour forecast output. From the data portal that is jointly hosted by UCLA and Remote Sensing Solutions, Inc. (RSSI) (http://west.rssoffice.com/ca_roms), six hourly runs of SST and sea surface salinity (SSS) and daily three-day forecasts are acquired to compute the HAB empirical models with the same forecasting horizon (Fig. 2).

2.2. Continuous ocean color imagery from DINEOF

Moderate Resolution Imaging Spectroradiometer on Aqua (MODISA) data are acquired from NASA's Ocean Biology Processing Group (http://oceancolor.gsfc.nasa.gov/) for the U.S. West Coast region, processed from Level 1 to Level 2 with the standard *l2gen* routine in SeaDAS (v. 7.0.2), and further subscened to the CA-ROMS domain. The DINEOF routine (compiled from http://modb.oce.ulg.ac.be/mediawiki/index.php/DINEOF) is applied to a time series of MODISA imagery. In order to maximize the variability captured by

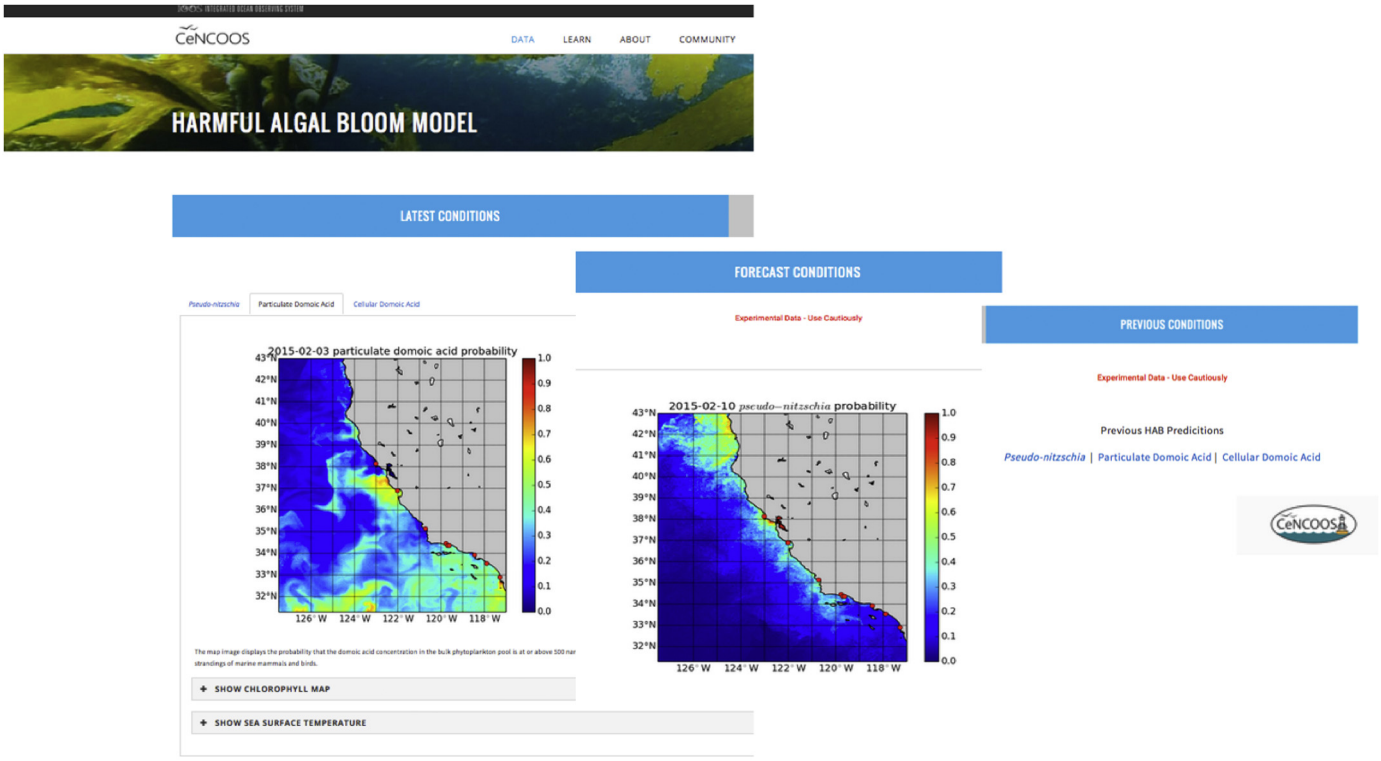

Fig. 3. C-HARM web display. Routine nowcasts and forecasts of *Pseudo-nitzschia* blooms and DA events have been produced since February 2014 and broadcast to the public on the CeNCOOS website http://www.cencoos.org/data/models/habs. Additional links to information and tutorial pages offer detailed descriptions 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.

the EOF analysis and minimize the overall percentage of cloud coverage, one month of MODISA imagery was chosen as a starting point (t_0) for seeding DINEOF, and this has been routinely increased at each time step to include the most recent180 days, which runs in reasonable time. Short-term (3-day) forecasts of chlorophyll (Chl) and remote sensing reflectance (R_{rs}) at 488 and 555 nm are generated by solving multivariate EOFs that simultaneously include satellite ocean color fields and forecasted physical fields (SSS, SST) from ROMS. The reconstructed dataset from this analysis provides the forecasted fields (Chl, $R_{rs}(0^+, \lambda)$, SSS, SST) that are then used to calculate the empirical HAB models at each grid point in the model domain ($351 \times 391 \times t$) with a three-day lead time. A very simple advection scheme is also employed and may prove to be a more pragmatic alternative to the time-intensive multivariate EOFs. Pixels of Chl and $R_{rs}(0^+, \lambda)$ are "advected" by multiplying by the daily change in the east-west and north-south components [u(dt) and v(dt)] of the 3-km surface current velocity projected by ROMS after first mapping u and v from a model native grid to a latitude-longitude grid (WGS84 coordinate reference system). The DINEOF routine is then used to fill in gaps left by the movement of advected pixels, most notably around the Channel Islands and the northwestern corner of the domain as water is advected south. Gap-filled Chl imagery is included on the HAB forecasting front page as an additional end user product and as a visual aid for interpreting the HAB model imagery (Fig. 4). While forecasts and nowcasts are both produced, the skill assessment discussed in the

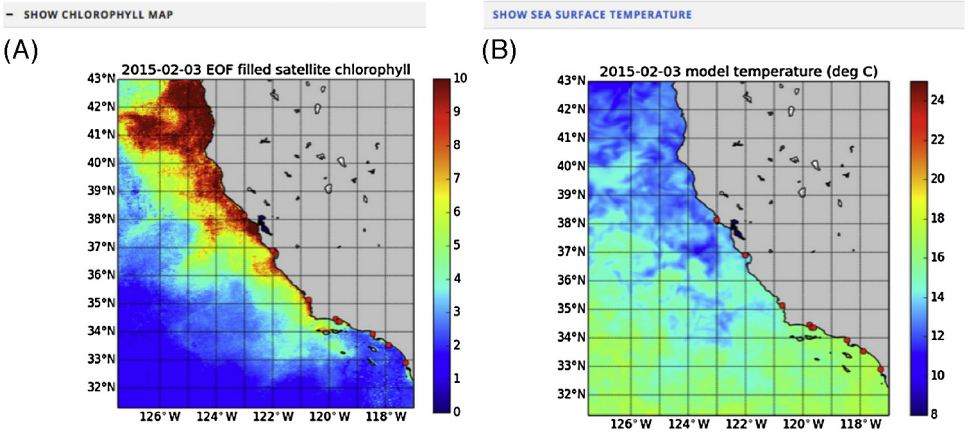

Fig. 4. C-HARM web display. The Latest Conditions page contains expansion links to (A) gap-filled Chl imagery from DINEOF run on MODISA retrievals (mg m^{-3}), and (B) modeled SST (°C) from the 3-km CA ROMS. Note that the pDA HAB probability model requires SSS forecasts (not displayed), which are also provided by ROMS.

remainder of this contribution is focused on the nowcast predictions.

2.3. Empirical HAB models

The remote-sensing-based empirical models employed in this study are the same as those reported in Anderson et al. (2011) where the methodology for developing the GLMs from *in situ* radiometer and hydrographic observations is well described. Three independent models predict (1) *Pseudo-nitzschia* blooms (threshold = 10^4 cells L^{-1}), (2) particulate DA events (pDA threshold = 500 ng L^{-1}), and (3) cellular DA events (cDA threshold 10 pg $cell^{-1}$) using some combination of $R_{rs}(0^+, \lambda)$ at various wavelengths (λ), MODISA chlorophyll (Chl), salinity, temperature, or time of year (Table 1). At each grid point, these models are computed from the gap-filled MODISA ocean color retrievals and the ROMS physical estimates, leading to spatial/temporal resolution of 3 km and daily. Given that these are logistic regression models, the value is a probability or likelihood of encountering a bloom and/or toxic state at a given grid point/pixel in the model domain. A current estimate of HAB state (Fig. 3) is *routinely* generated using gap-filled ocean color imagery (Fig. 4a) and the most recent ROMS simulation (Fig. 4b) before the biogeochemical fields are forward projected via advection and/or DINEOF. Predictions up to three days forward (i.e. forecasts) are displayed separately, while archived predictions are available as image files in a set of directories (Fig. 3). The gap-filled ocean color imagery for Chl and the ROMS SST images are also displayed alongside HAB predictions (Fig. 4).

2.4. Routine ocean observing data

The task of verifying forecasts is essential to communicating model accuracy, utility to decision makers, and informing model improvement. The fact that there are appropriate biological field observations with which to compare our model results is due to the diligent efforts of the California Harmful Algal Bloom Monitoring and Alert Program (HABMAP). Originally a grassroots collaboration of academics and government resource managers, HABMAP is now a component of the regional IOOS program that greatly increases our ability to evaluate model output and compare nearshore observations with potential offshore HAB activity in near real-time (Frolov et al., 2013; Kudela et al., 2015). Seven stations from Santa Cruz, CA to La Jolla, CA are sampled weekly for phytoplankton community composition, and at a select few, for key HAB phycotoxin concentrations (saxitoxin and DA). Data are shared with the public in near real-time via the HABMAP (http://www.habmap.info/) and SCCOOS websites (http://www.sccoos.org/data/habs/index.php). The shore stations used for the bulk of the analyses reported here are the Santa Cruz Municipal Wharf (SCMW, 36°57.48′ N, 122°1.02′ W) in the northern part of Monterey Bay and Stearns Wharf (34°24.48′ N, 119°41.10′ W) in the Santa Barbara Channel (Fig. 1). Weekly particulate DA measurements (LC-MS or ELISA) and light microscopy counts of *Pseudo-nitzschia* spp. (total of *seriata* and *delicatissima* size classes) from surface phytoplankton bucket samples are matched to daily model output from 15 Feb 2014–10 Feb 2015. The longest record of particulate DA for the 2014 to 2015 model demonstration year is at the SCMW shore station where we also compare time series of DA from mussel tissue and from Solid Phase Adsorption Toxin Tracking (SPATT) mesh bags (LC-MS, Lane et al., 2010b) with HAB model predictions. In addition to phytoplankton and mussel data, records of central CA marine mammal strandings (mostly pinnipeds) associated with DA toxicosis from Jan 2007 to Sep 2014 are associated with DA measurements over that time period. Given the uncertainty surrounding when a given animal actually stranded before it was admitted, monthly means of mammal strandings are compared with shore station and model data at the SCMW to assess the relationship between stranding episodes and the variability in nearshore DA levels and predicted DA events. For model/observation match-ups, a time series of daily modeled probabilities from the 3×3 model grid point that best approximates the shore station coordinates is extracted from the netCDF gridded data and compared to these measurements. For all match-ups, we use the nowcasted netCDF data rather than any of the forecast data to simplify the analyses.

2.5. Trend analysis

At present, it is challenging to compare modeled probabilities to binary outcomes in the observations without a clear understanding of the relationship between the two over the CCS model domain. As model and observation matchups become more abundant through continued daily model runs and HABMAP monitoring in CA, model calibration will become a useful option for creating regionally relevant prediction points. Some of the available methods for comparing patterns in the predicted values with discrete weekly time series of *Pseudo-nitzschia* abundance and DA concentration observed at the HABMAP shores stations are described below.

2.5.1. Cross correlation functions

Cross-correlation functions (CCF) were computed between time series of daily, modeled values and weekly *in situ* measurements at select shore stations to assess correlation and lead-lag relationships between discrete observations and predicted probability values. With CCF analysis, it is important not to violate assumptions of stationarity. In lieu of a formal test for stationarity (e.g. unit root test), stationarity in the time series was diagnosed by plotting the autocorrelation function (ACF) and assessing the steepness of the slope; a gradual decrease in the correlogram that does not taper to zero is a first order measure of a non-stationary process. If a linear trend exists in the time series, it is generally sufficient to detrend the data by simply removing the linear trend. In this case, the residuals of the linear regression of the data are used in the CCF analysis in place of the original time series. In cases of non-linear trends, stationarity can be achieved either by pre-whitening or by applying an Autoregressive Integrated Moving Average (ARIMA) model. An ARIMA model was applied to those data that fit this category by first differencing the time series and estimating the autoregressive, AR (p), integrated (d), and moving average, MA (q) parameters. The autocorrelation (ACF) and partial autocorrelation (PACF) plots of the differenced time series allow us to determine the type of ARIMA (either AR or MA) and number of terms that are needed for fitting the ARIMA (p,d,q) model (Box, Jenkins, and Reinsel, 1994). The residuals from the ARIMA fit are then cross-correlated with the second (observational) time series to attain the CCF.

2.5.2. Kolmogorov–Smirnov (K-S) Test

A K-S test is a non-parametric statistical test to determine the likelihood that two time series were drawn from the same continuous probability distribution (Lehmann and Romano, 2006) and tests for deviations from the null hypothesis in terms of median, variance, and distribution. It is particularly powerful in detecting differences in distribution shape (or pattern) so we apply it here to compare patterns in observed and modeled time series. A two-sample (or two-sided) K-S test was conducted on the normalized time series of modeled and observed data for each set of matchups (R v 2.15.3), generating a D statistic which is evaluated with respect to a critical D value (D_{crit}) at an alpha level of 0.05 ($p = \alpha$). If the D statistic is greater than the critical value, the

null hypothesis that the samples are from a population with the same distribution is rejected.

2.6. Skill assessment

A number of verification methods commonly used in the meteorological community for dichotomous forecasts (Jolliffe and Stephenson, 2012; Jolliffe and Stephenson, 2003) have already been applied during the development and testing of the logistic toxic *Pseudo-nitzschia* bloom GLMs, and these include measures of sensitivity and specificity of the models along with cross-validation and metrics of model skill (C.R. Anderson et al., 2011). The same bloom thresholds used during model development are applied here to transform 2014–2015 observations to a binary value (10,000 cells L^{-1} for *Pseudo-nitzschia* spp. blooms and 500 ng L^{-1} pDA). For simplicity, the cellular DA model is left out of the analysis and assumed to be secondary to the performance of the pDA model.

There is no single skill score or metric that can capture all aspects of model performance, and it is a key focus of our research to identify those that are most relevant to assessing the usefulness of a HAB forecast model. Some common metrics are the probability of detection (POD), false alarm ratio (FAR), and probability of false detection (POFD) or false alarm rate. All provide some measure of sensitivity or specificity in the model predictions. Their definitions and use in evaluating HAB predictions are well described in Anderson et al. (2010) and are repeated here in Table 2. To this list we have added *accuracy*, which provides a measure of the proportion of correct forecasts. A popular method of describing model reliability for a dichotomous solution is the calculation of a *bias score* that indicates the frequency with which the model predicts a particular outcome (i.e. yes/no); the bias score is not based on how well the model matches observations and thus is not to be confused with a bias function. A good bias score equals one, but it can range from zero to infinity (Table 2). For all the metrics listed above, calculations are based on a contingency table that categorizes "hits" as correctly predicted blooms (or DA events), "false alarms" as false positives, "misses" as false negatives, and "correct negatives" as correctly predicted non-blooms (Table 2). When the metric scores are plotted across all possible thresholds for determining the relevant probability at which HAB bulletins or other alerts should be issued (i.e. the "prediction point" for issuing a bloom forecast), an optimized prediction point can be achieved with respect to any value of interest to resource managers or to maximize or minimize a particular performance metric (Anderson et al., 2010; Lane et al., 2009). It is important to note that the chosen prediction point must meet the needs of the end-user; just as there are multiple metrics that can be optimized, the prediction point must also be chosen to optimize the metric of interest, possibly at the expense of another metric. For example, a manager may prefer a higher rate of false positives to minimize the chance of a false negative.

The receiver operating characteristic (ROC) describes POD (sensitivity) as a function of POFD (1-specificity) over the full range of prediction points, i.e. the cumulative distribution function, and is used in many fields to translate binary classifications into decision-making criteria. It quantifies the relative accuracy of a model to separate true positive and true negative outcomes (Fawcett, 2004). The area under the ROC curve (AUC) reduces the two-dimensional space to a scalar value that is a well-used metric of model performance (e.g. Lane et al., 2009) and is essentially equivalent to a Wilcoxon rank test (Fawcett, 2004). The AUC is calculated using Simpson's rule to approximate the integral of the function (*sintegral* in R, Bolstad2 library). When values on the ROC curve fall above the 1:1 line, they are in the realm of good predictions, while those at or below the 1:1 line (particularly in the lower right quadrant) are deemed as good or worse than random. Similarly, AUC scores above 0.5 suggest a model that performs better than random chance. Both the ROC curve and its AUC are reported as a measure of the usefulness of the HAB models, as best as can be ascertained at the shore station pixel(s).

Table 2
A description of the common performance metrics applied to the results of routine HAB predictions for California in 2014–2015 made using the models in Table 1.

Performance Metric	Calculation	Range
Accuracy	(hits + correct negatives)/total	0–1
Probability of Detection (POD)	hits/(hits + misses)	0–1
False Alarm Ratio (FAR)	false alarms/(hits + false alarms)	0–1
Probability of False Detection (POFD)	false alarms/(correct negatives + false alarms)	0–1
Bias Score (BS)	(hits + false alarms)/(hits + misses)	0–∞

3. Results

3.1. Summary of major 2014 DA events

Interestingly, 2014 was a very high DA year, particularly in Monterey Bay (MB), with record fishery alerts and marine mammal strandings. Modeled pDA probabilities from the SCMW pixel were somewhat elevated in mid-March (>0.5) but broadly high for MB as the region saw a spike in California sea lion (*Zalophus californianus*) strandings from MB (Fig. 5). A week later, predictions targeted MB as a hot spot of HAB/DA likelihood (mapped in Fig. 5) just as measured pDA at the SCMW shore station rose to 0.84 ng mL^{-1} and modeled pDA probabilities at the SCMW rose to 0.8. This instance corroborates hypotheses that marine mammals are good sentinels of the onset of DA events (Bargu et al., 2012), particularly since these first DA-related strandings of the year were all from the MB region. We should note the discrepancy in the magnitude of the hot spot in the posted imagery (maps in Fig. 5) versus the time series extracted from the netCDF file (pDAmod in Fig. 5) since the latter reflects a year of MODIS variability analyzed with DINEOF, whereas the imagery posted in March 2014 was influenced by a fewer number of positive excursions in the shorter DINEOF time series. In the week following the hot spot imagery (on 1 April 2014), the CA Dept. of Public Health (CDPH) issued closures of recreational shellfish harvests and a historical warning regarding fishing anchovies and sardines in the MB due to dangerous levels of DA toxicity. The 1-week lag time is consistent with past observations of delayed toxicity in shellfish (Lane et al., 2010a) and corroborates the timing of the predicted rise of DA.

The fall HAB event in 2014 was focused on the Central-Southern California transition zone where the CDPH issued a warning on October 11 not to eat crabs, lobsters, and bivalves harvested in the Santa Barbara Channel region, another well-known hot spot for DA (C.R. Anderson et al., 2006; Trainer et al., 2000). Forecast maps published on the CeNCOOS website illustrated the increased risk in this region (Fig. 6). In Monterey Bay, on the other hand, samples taken within the Bay from a cruise of opportunity on 15–17 October verified that DA levels were extremely low while *Pseudo-nitzschia* abundance was at moderate to bloom levels $\sim 10^3$–10^4 cells L^{-1} (pers. comm., M. Blakely Peacock), as seen in the imagery for 15 Oct (Fig. 6). DA event probabilities at Stearns Wharf in the Santa Barbara Channel remained elevated relative to background levels for another week (Fig. 6). The CDPH only lifted the advisory for the Ventura County coastline on 21 February 2015, which may reflect the persistence of DA in benthic environments (C.R. Anderson et al., 2011; Kvitek et al., 2008; Sekula-Wood et al.,

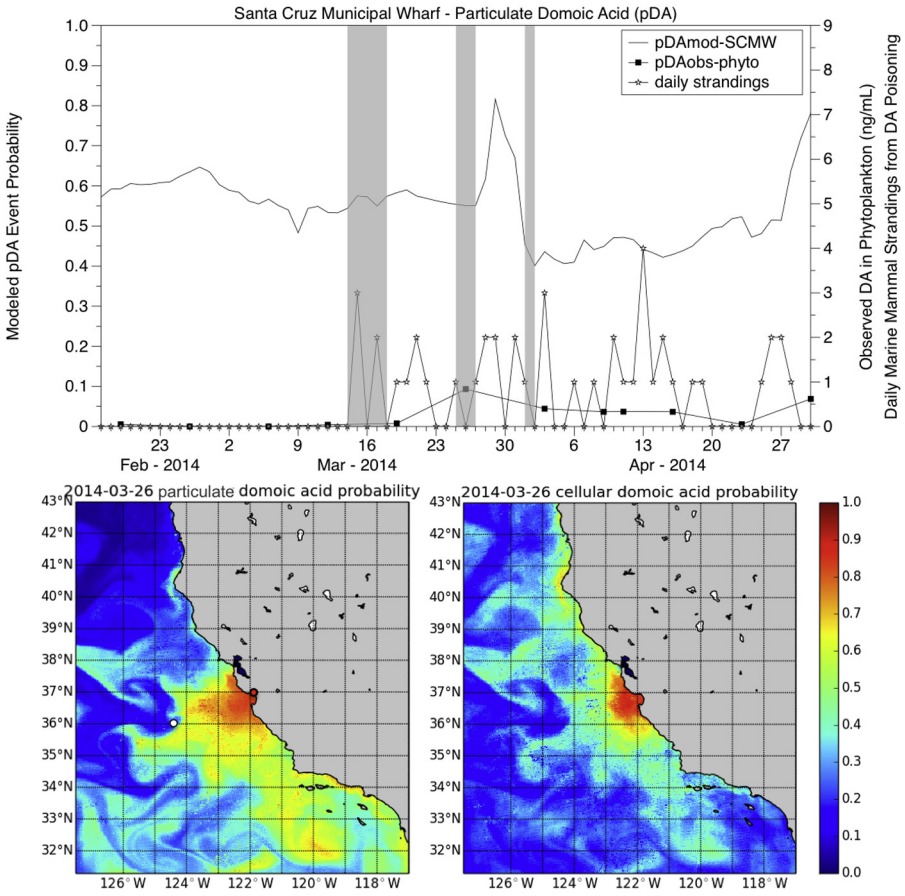

Fig. 5. Spring 2014 DA event in the Monterey Bay region. The upper panel describes the time evolution of marine mammal strandings from DA toxicosis, particulate DA (pDA) at the SCMW, and modeled pDA at the SCMW pixel (36.7 °N, 122.4 °W, red circle) and a pixel much further offshore in the CCS (36.0° N, 124.5° W, white circle). Grey bars correspond to (1) the first reports of mammal strandings from The Marine Mammal Center (16–17 Mar), (2) the rise in particulate DA (pDA) at the SCMW on 26 Mar, and 3) the CDPH advisory against harvesting anchovies or sardines from the Monterey Bay on 1 Apr. The bottom two images are particulate and cellular DA nowcast imagery displayed on the CeNCOOS website for 26 Mar.

2009), particularly if the low pDA probabilities are reflective of low surface pDA values.

3.1.1. Santa Cruz Municipal Wharf

We first discuss the temporal trend in model and observational time series for the SCMW where there is the most comprehensive record of *Pseudo-nitzschia* abundance and DA from both phytoplankton and mussel tissue for 2014–2015 (Fig. 7). Predictions of *Pseudo-nitzshia* blooms at that SCMW pixel over the course of 2014 were dynamic with rapid fluctuations between low and high probabilities with a significant upward linear trend ($r = 0.37$; $p < 0.001$; Fig. 7a), potentially related to the increasing bloom conditions observed in the CCS later in 2015 (McCabe et al., accepted). The dramatic rise and fall in *Pseudo-nitzschia* abundance between April and May at the SCMW is reflected in the modeled probabilities, with predictions leading the rise in abundance at the SCMW. The *Pseudo-nitzschia* group disappeared from the SCMW samples in mid-July (Fig. 7a). After July, however, the modeled time series maintains an elevated mean value, albeit highly variable (Fig. 7a). The model correctly captures the drop in *Pseudo-nitzschia* abundance between April and May, and given that the GLM for *Pseudo-nitzschia* is driven by month and $R_{rs}(488/555)$, this large variation is most likely due to large fluctuations in phytoplankton pigments during this spring transition. The model does not capture the disappearance of *Pseudo-nitzschia* at the SCMW after July. A K-S test quantifies this mismatch between the patterns in the modeled and observed time series at the SCMW ($D_{crit} = 0.26$; $D = 0.44$, $p < 0.01$). The resulting CCF [ARIMA (1,1,0)] of the predictions with respect to SCMW abundance measurements shows a weakly significant correlation at zero lag and one-day lead time as well as a negative correlation at an eight-day lead (Fig. 8a).

Particulate domoic acid (pDA) at the SCMW follows a similar pattern to *Pseudo-nitzshia* in that all the peaks occur in spring, and modeled probabilities at this nearshore pixel are better aligned with pDA observations in the early part of the year than after July (Fig. 7b). The time series trend is negative ($r = -0.51$; $p < 0.001$), however. The K-S test also reflects the diverging patterns of modeled and observed time series at the SCMW over the course of the year ($D_{crit} = 0.26$; $D = 0.41$, $p < 0.01$). The inability of the models to capture the summer-fall variability at the SCMW is possibly due to nearshore conditions in summer being especially uncoupled from the regional signal, a pattern that emerged from a separate analysis of crosshore *Pseudo-nitzschia* and DA in the Santa Barbara Channel from 2009 to 2013 (Umhau et al., submitted for publication). The CCF analysis for the nearshore pixel relative to the SCMW observations indicates several significant lead times (e.g. 5, 19, and 26 days) even after applying ARIMA (0,1,0) (Fig. 8b).

The relationship between mussel DA and phytoplankton pDA at the SCMW is fairly close, as indicated by a two-sided Kendall rank correlation test that yields significant agreement ($\tau = 0.41$, $p < 0.05$). There is a significant five-day lead between the pDA model at the SCMW pixel and mussel DA (Fig. 9a). A K-S test also reflects a stronger relationship between the mussel DA

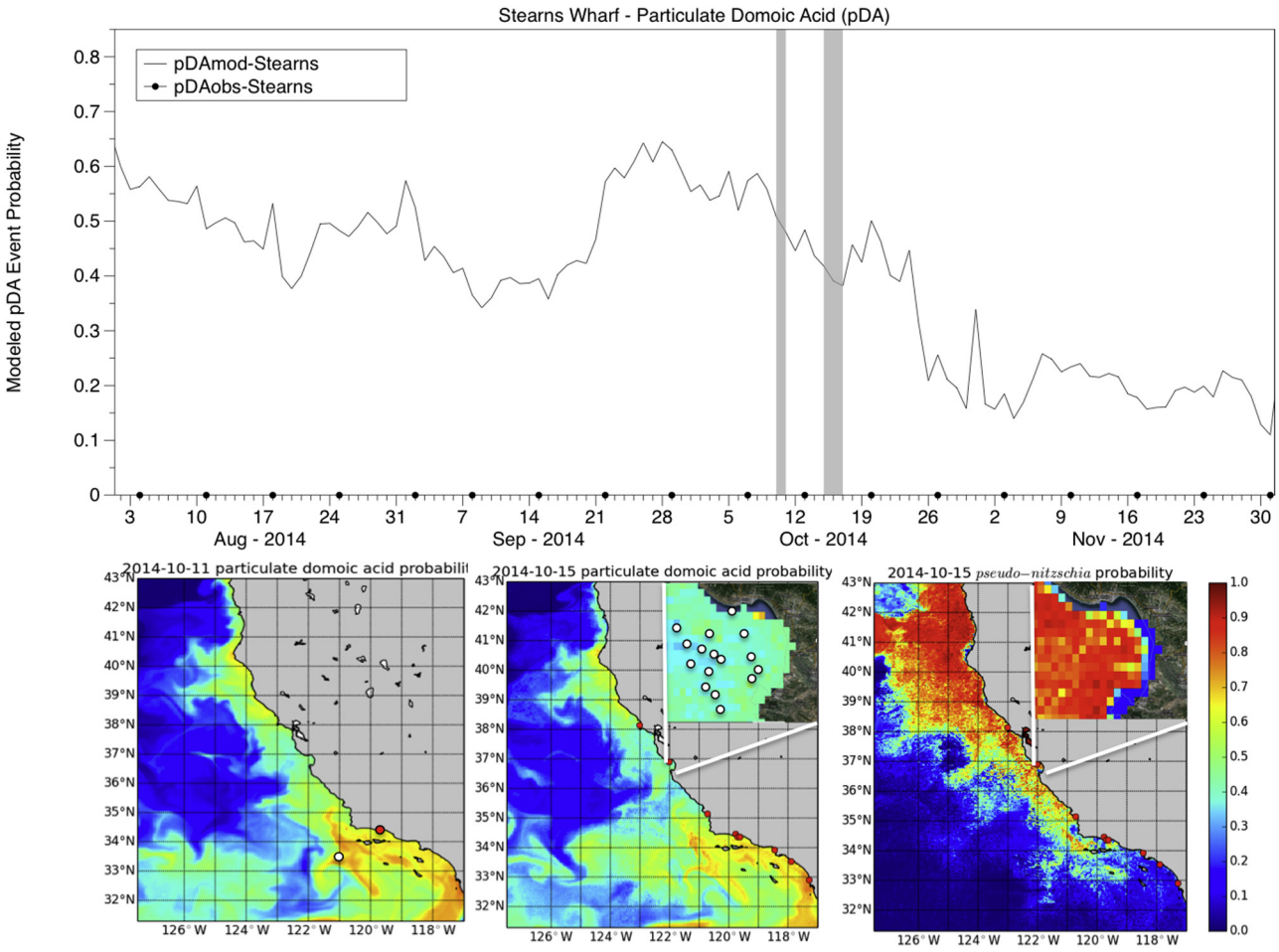

Fig. 6. Fall 2014 DA event in the Santa Barbara Channel region. The upper panel describes the time evolution of modeled pDA at the Stearns Wharf pixel (red circle in the 11 Oct image, 34.4° N, 119.6° W) and a pixel southwest of the Santa Barbara Channel (white circle in the 11 Oct image, 33.5° N, 121.0° W). Grey bars correspond to (1) the CDPH advisory against harvesting benthic invertebrates such as crabs and lobster on 11 Oct, and (2) the timing of an offshore cruise of opportunity in Monterey Bay 15–17 Oct that validated the low DA and relatively high *Pseudo-nitzschia* in that region at the time (black and white circles on inset map). The bottom three images are pDA nowcast imagery for the 11 Oct and pDA and *Pseudo-nitzschia* for the 15 Oct displayed on the CeNCOOS website (inset maps expanded for Monterey Bay).

observations and modeled pDA at the SCMW pixel for this shorter time period (D_{crit} = 0.34; D = 0.36, p = 0.006). An alternative record of DA that has been shown to be a better tool for monitoring phycotoxins than sentinel shellfish sampling due to the incredibly high depuration rates of bivalves such as mussels is SPATT resin beads (Lane et al., 2010b), which have been deployed in mesh bags and recovered for DA extraction at one to two week intervals since 2008 at the SCMW (Fig. 7b). In this case, the mean modeled pDA was used during the two week SPATT deployment and matched to the midpoint date of the SPATT deployment. There is a significant 9 to 12-day lead shown by the CCF ($r \sim 0.40$) between the mean pDA time series (data untransformed in this case due to the sine-cosine wave decay in the ACF) at the SCMW pixel and SPATT-measured DA (HP20-type resin; Fig. 9b). The K-S test indicates a significant similarity in the time series patterns (D_{crit} = 0.37; D = 0.36, p = 0.013).

The CCF for marine mammal strandings averaged over a month suggests that pDA predictions lead the peaks in strandings by five days at the SCMW pixel (r = 0.48, Fig. 9c) and are negatively associated with the model at a five-day lag. The patterns fail to indicate similarity with a K-S test (D_{crit} = 0.15; D = 0.28, $p < 0.01$). This outcome suggests the model does potentially capture time and space scales relevant to California sea lion foraging patterns and may provide adequate lead-time to assist marine mammal rescue efforts.

3.1.2. Stearns Wharf

Further south in the Santa Barbara Channel, predicted *Pseudo-nitzschia* bloom probabilities from the Stearns Wharf pixel are highly variable but with a flatter mean over the study period than at the SCMW (Fig. 10a). The CCF shows that the strongest relationship between modeled [ARIMA(1,1,0)] and observed data occur when the model either leads by 20 days or lags by two days (r = 0.25), and that at lag zero, they are negatively correlated (Fig. 11a).

Spring peaks in pDA levels at Stearns Wharf do not coincide with elevated pDA probabilities except for the mid-June increase when probabilities rise abruptly (Fig. 10b). Significant correlations in the CCF occur when the model [ARIMA (1,1,0)] leads observations by 8 and 15 days (Fig. 11b) and lags by roughly 30 days. The K-S test suggests similarities in the shape of nearshore observations and the model at Stearns Wharf (D_{crit} = 0.26; D = 0.33, $p < 0.05$).

3.1.3. All California HAB stations

Routine measurements of *Pseudo-nitzschia* spp. abundance in southern and central California are combined into one analysis for a California-wide assessment (Fig. 12). There is a significantly increasing trend (r = 0.37; $p < 0.001$) in the modeled probabilities over the 2014–2015 timeframe as the ecosystem moved towards the massive HAB event of spring-summer 2015 all along the

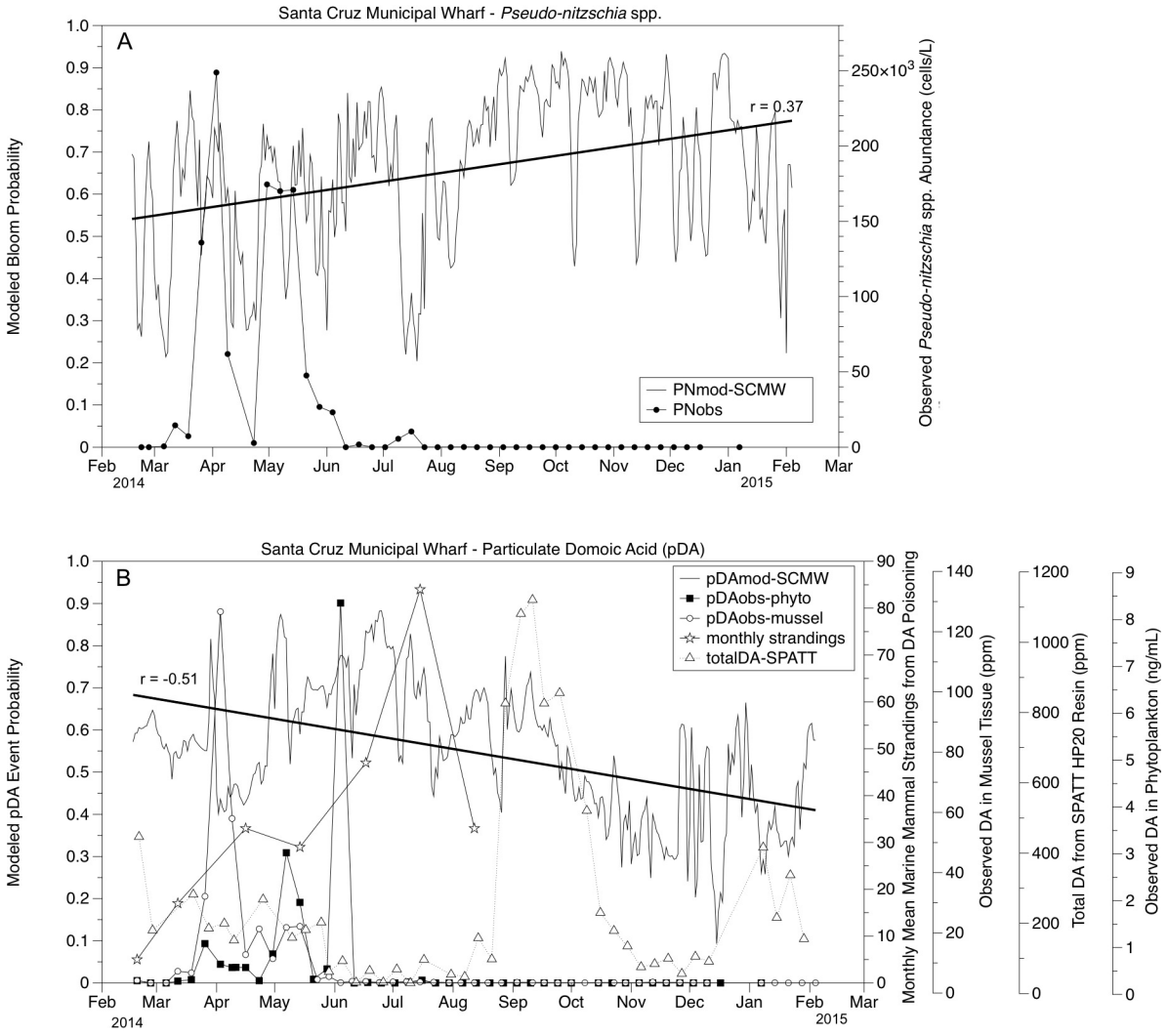

Fig. 7. Modeled bloom/event probabilities at the SCMW. (A) Weekly microscopic counts of *Pseudo-nitzschia seriata* and *delicatissima* size classes (combined; cells $\times 10^3$ L^{-1}; K. Negrey) from 19 February 2014 to 7 January 2015 are represented as discrete points (closed circles) in comparison with daily modeled values at the SCMW pixel (solid line). The solid dark line is the linear trend for the PN model time series ($r = -0.51$, $p < 0.001$). (B) Weekly measurements of particulate DA (pDA) by LC–MS (ng mL^{-1}; R. Kudela and K. Negrey) from 19 February 2014 to 7 January 2015 are represented as discrete points (closed squares) in comparison with daily modeled values (solid line). In addition to phytoplankton pDA, measurements of DA from mussel tissue collected at the SCMW (open circles), DA recovered from SPATT-HP20 mesh bags (open triangles), and monthly mean counts of marine mammals admitted to The Marine Mammal Center with DA toxicosis (open stars) are compared with pDA and modeled DA events at the SCMW (solid line). Gaps in data collection are shown as empty space. The solid dark line is the linear trend for the pDA model time series ($r = -0.51$; $p < 0.001$).

U.S. west coast (McCabe et al., accepted). Departures from the mean model state are shown in the upper panel of Fig. 12. Scripps Pier and Cal Poly Pier (in San Luis Obispo) diverge the most from the mean state, suggesting that the timing of conditions supportive to HAB development at these sites may be uncoupled from the timing in other parts of southern and central California. There are many weakly significant spikes in the CCF [ARIMA (0,1,0)] for all nearshore stations, but none point to a near-term lead between model and bloom observations (Fig. 13a). Abundance of *Pseudo-nitzschia* spp. does significantly lag the nowcasts by three days (Fig. 13a).

3.2. Contingency plots of skill assessment

Figs. 13–15 show the outcome of each selected metric (Table 2) for comparing a binary classification with modeled/forecasted probabilities over all possible prediction points. The *Pseudo-nitzschia* model yields a high rate of false positives across all HAB monitoring stations. For this reason, the optimized prediction point (0.52) for California is a minimization of the false alarm rate (FAR) with respect to the probability of detection (POD). Their optimized value (i.e. where the two lines cross) is 67% and total accuracy is 43% (Fig. 13b). The maximal accuracy of the *Pseudo-nitzschia* model at all HAB stations is 67%. At the SCMW site specifically, the prediction point was optimized for POD and accuracy since the false positive rate was lower, yielding an optimal prediction point of 0.68 where accuracy and POD = 38% (Fig. 14a). The pDA model at SCMW appears to be more constrained than the *Pseudo-nitzschia* model in that it yields a lower false positive rate. Given this, we optimized relative to accuracy and POD to maximize the sensitivity. This shared value is 68% at an optimal prediction point of 0.6 (Fig. 14b). The maximum bias score when the prediction point is zero is four for the *Pseudo-nitzschia* model and nine for the pDA model at both sites (maxima not shown on plots). If the prediction points were optimized with respect to the BS instead of the POD, higher accuracies and reliability would be achieved using more conservative prediction points. At Stearns Wharf, the *Pseudo-nitzschia* model was optimized relative to FAR and POD as before, yielded an optimal threshold of 0.55 with POD and FAR = 57% (Fig. 15a). There are

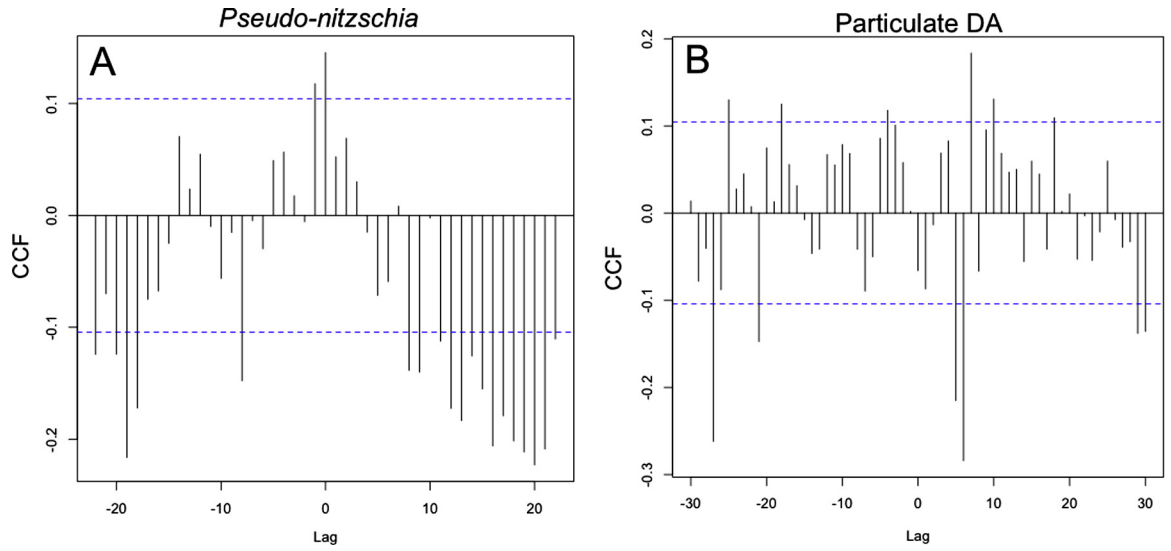

Fig. 8. Santa Cruz Municipal Wharf (SCMW). Cross-correlation functions (CCF) describe the correlational relationship between the SCMW observations and the nearest SCMW model pixel over a range of lead-lag time scales. Negative lags occur when the model leads observations. The dashed lines show the 95% confidence interval.

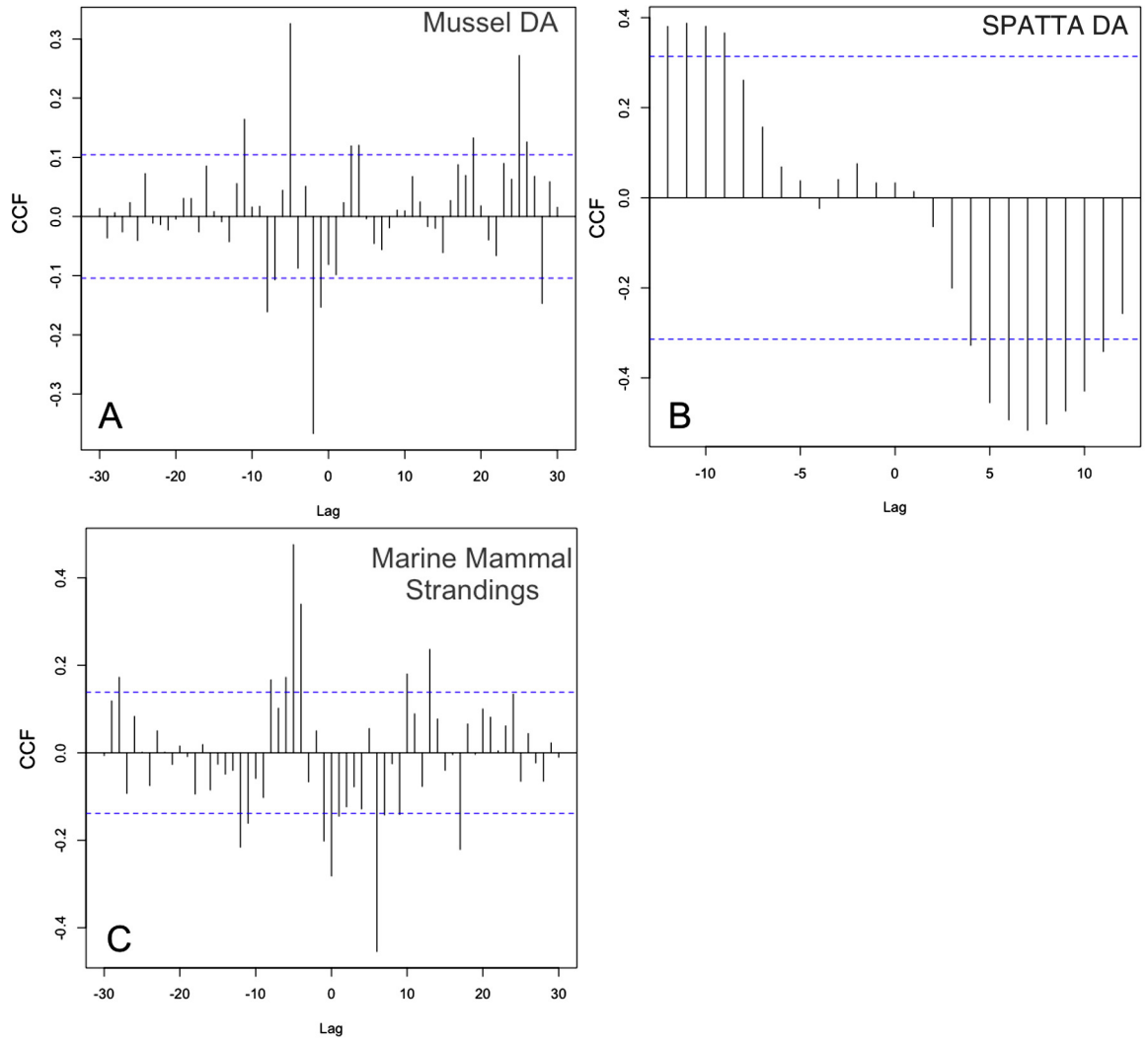

Fig. 9. Santa Cruz Municipal Wharf (SCMW): Particulate DA model-observation matchups for mussel tissue and SPATT sampling. Cross-correlation functions (CCF) describe the correlational relationship between the pDA model at the SCMW pixel and (A) mussel tissue DA from the SCMW, (B) Solid Phase Adsorption Toxin Tracking (SPATT) HP20 resin bags deployed at the SCMW DA from the SCMW, and (C) monthly mammal strandings over a range of lead-lag times. Negative lags occur when the model leads observations. The dashed lines show the 95% confidence interval.

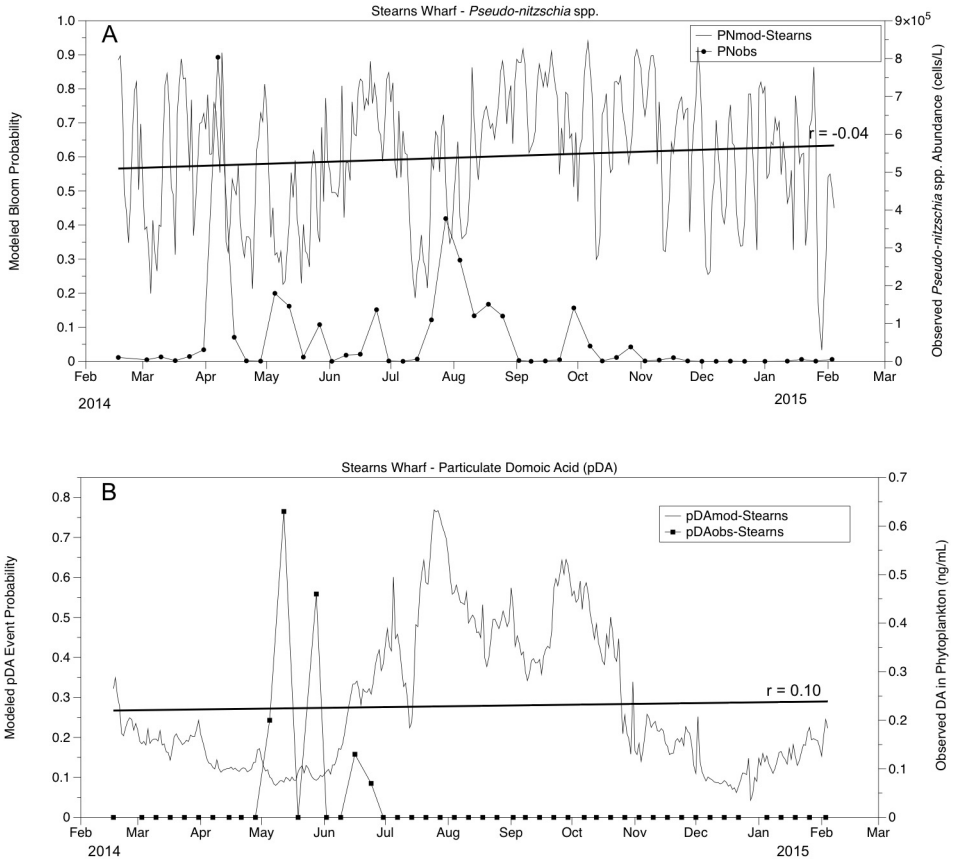

Fig. 10. Modeled bloom/event probabilities at Stearns Wharf. (A) Weekly microscopic counts of *Pseudo-nitzschia seriata* and *delicatissima* size classes (combined; cells $\times 10^5$ L^{-1}; H. McNair) from 17 February 2014 to 12 January 2015 are represented as discrete points (closed circles) in comparison with daily modeled values at the wharf (solid line) and at an offshore pixel (dashed line) at 33.5° N, 121° W. Bloom level is considered to be 0.1 $\times 10^5$ cells L^{-1}. The solid dark line is the linear trend for the PN model time series ($r = 0.04$; ns). (B) Weekly measurements of particulate DA (pDA) by ELISA (ng mL^{-1}; D. Caron) from 17 February 2014 to 4 August 2014 are represented as discrete points (closed squares) in comparison with daily modeled values at the wharf (solid line) and at an offshore pixel (dashed line) at 33.5° N, 121° W. Gaps in data collection are shown as empty space. DA concentrations above 500 ng L^{-1} are considered alert level. The solid dark line is the linear trend for the pDA model time series ($r = 0.10$; ns).

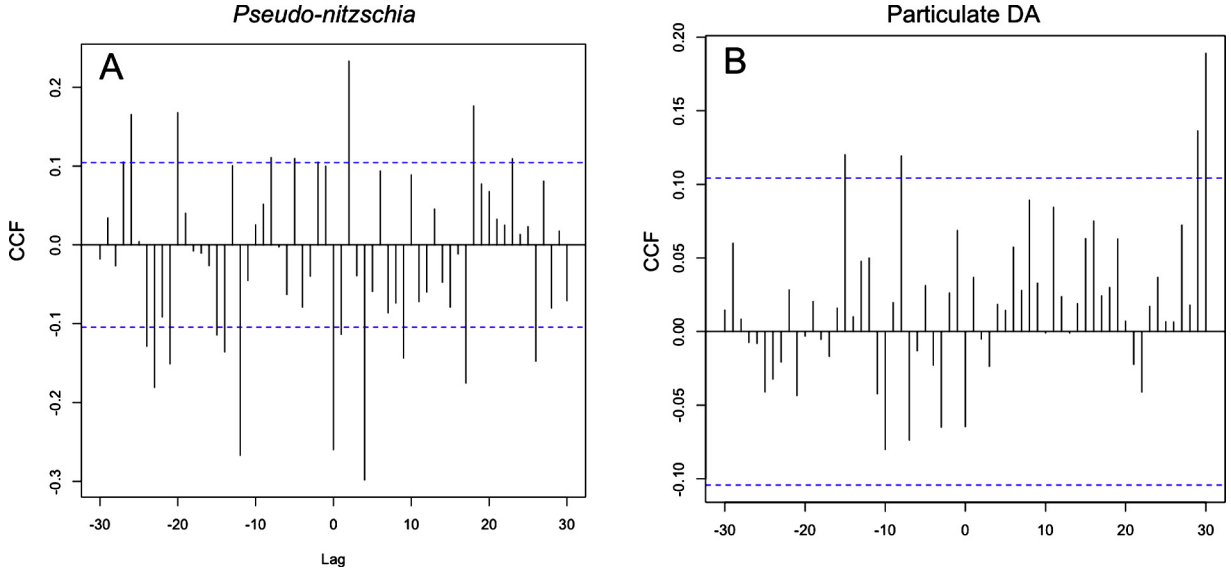

Fig. 11. Stearns Wharf. Cross-correlation functions (CCF) describe the correlational relationship between the Stearns Wharf observations and the nearest model pixel over a range of lead-lag time scales. Negative lags occur when the model leads observations. The dashed lines show the 95% confidence interval.

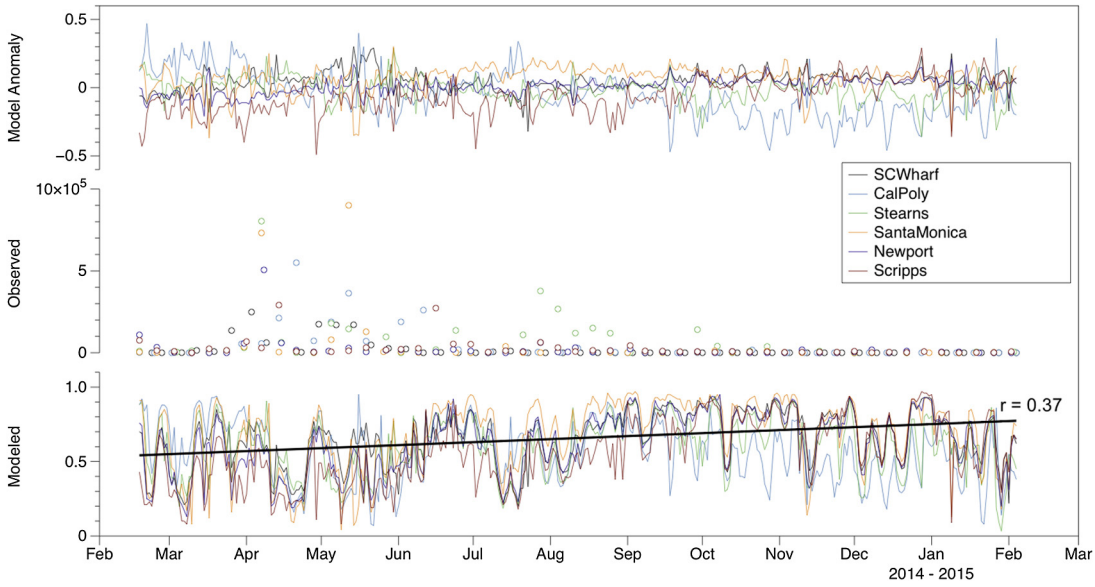

Fig. 12. Modeled bloom/event probabilities at All Stations. Weekly microscopic counts of *Pseudo-nitzschia seriata* and *delicatissima* size classes (combined; cells $\times 10^5$ L^{-1}; H. McNair) from 17 February 2014 to 12 January 2015 are represented as discrete points (closed circles) in comparison with daily modeled values (solid line). Bloom level is considered to be 0.1 $\times 10^5$ cells L^{-1}. The solid line in the lower panel is the linear trend line ($r = 0.37$; $p < 0.001$).

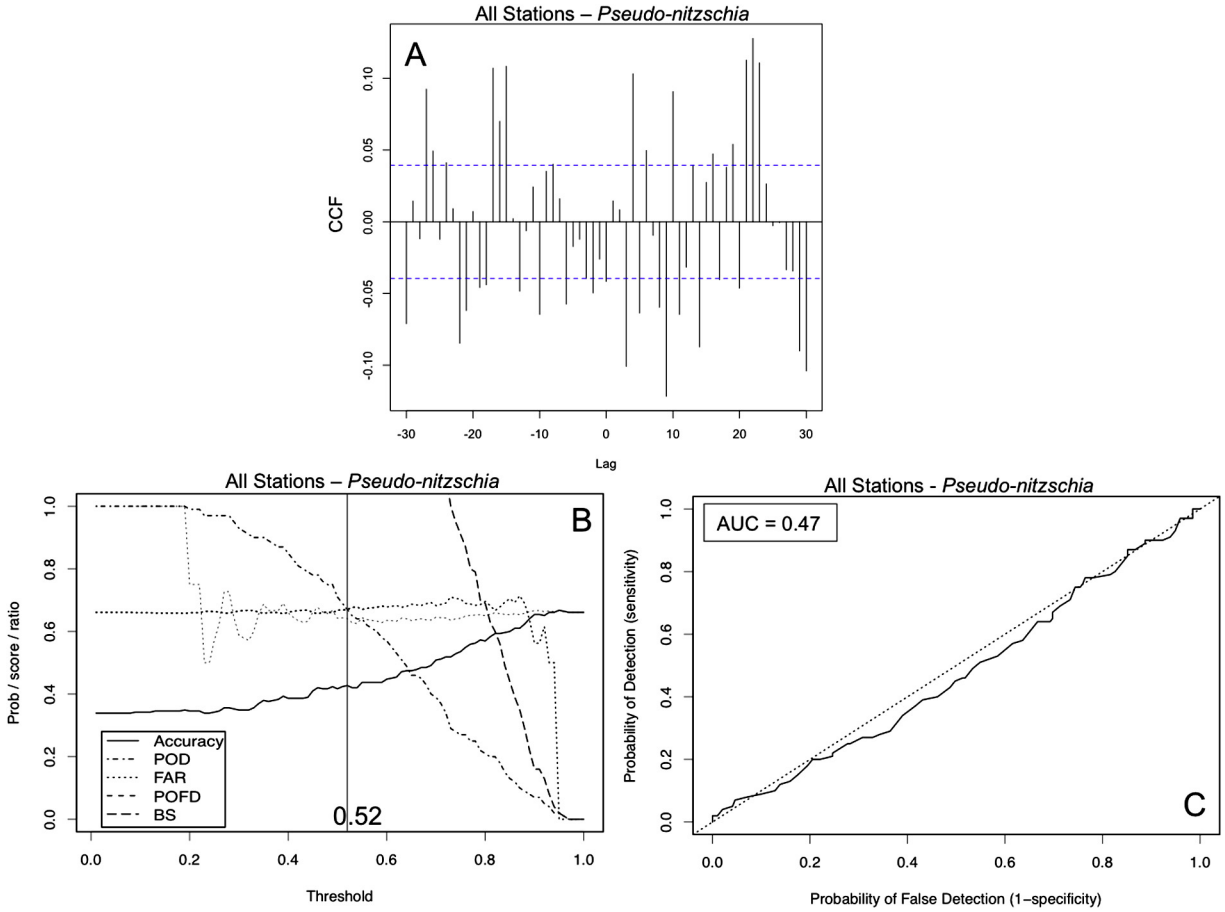

Fig. 13. Contingency plots, ROC Curves for All Stations compare *Pseudo-nitzschia* binary observations (transformed using bloom threshold) with modeled probabilities at the nearest pixel for each monitoring station. (A) Contingency plot metrics are accuracy, probability of detection (POD), false alarm rate (FAR), probability of false detection (POFD), and bias score (BS); the vertical line is the optimized prediction point (0.52). (B) ROC curve; the value in the box is the area under the curve (AUC). A value near 0.5 suggests a model that performs comparable to random chance. (C) Cross-correlation functions (CCF) describe the correlational relationship between the *Pseudo-nitzschia* observations at all stations and the nearest model pixel over a range of lead-lag time scales. Negative lags occur when the model leads observations. The dashed lines show the 95% confidence interval.

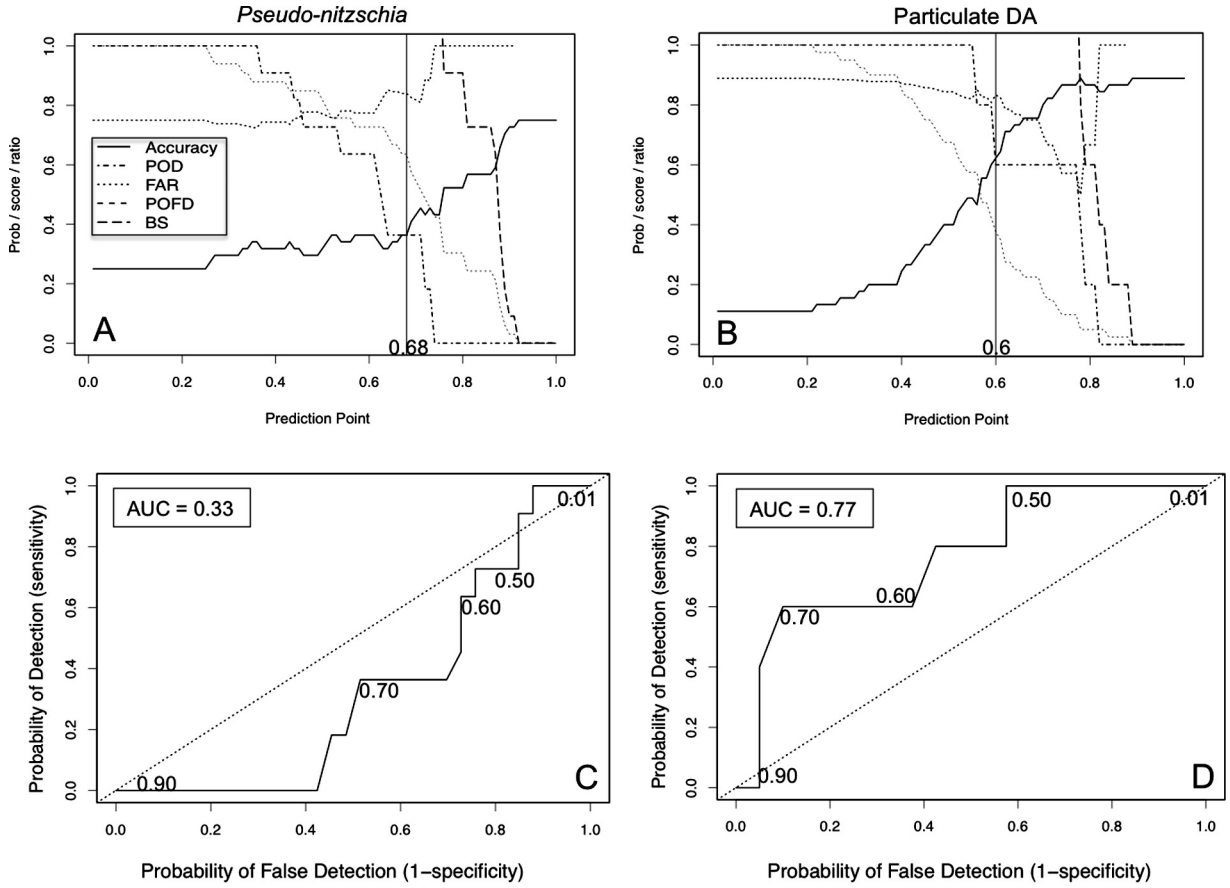

Fig. 14. Contingency plots and ROC Curves to assess model performance at the SCMW. (A) *Pseudo-nitzschia* and (B) pDA binary observations (transformed using bloom/event thresholds) are compared with modeled probabilities at the SCMW pixel. Metrics used are accuracy, probability of detection (POD), false alarm rate (FAR), probability of false detection (POFD), and bias score (BS); the vertical line is the prediction point −0.68 for *Pseudo-nitzschia* and 0.6 for pDA. ROC curves compare (C) *Pseudo-nitzschia* and (D) pDA binary observations (transformed using bloom/event thresholds) with modeled probabilities at the SCMW pixel. Values in boxes are the area under the curve (AUC), and values along the ROC curve show positions of various prediction points along the range 0–1 used to model POFD vs. POD.

relatively few actual DA events at Stearns Wharf over the time period, and there is a high percentage of correct negatives. When optimized relative to accuracy and specificity (each at 92%), the optimal threshold is 0.63.

The results of plotting the ROC suggest that the pDA model at the SCMW performs better than random chance because the AUC is 0.77 (Fig. 14d). The optimized prediction point is solidly in the upper middle to left quadrant, indicative of a moderately "liberal" classifier (Fawcett, 2004). In other words, it is considered a *useful* model since the points on the curve fall almost entirely in the upper triangle region above the 1:1 line, and the AUC is greater than 0.5. The ROC for the *Pseudo-nitzschia* model across all stations almost overlies the 1:1 line (Fig. 13c), i.e. it is a borderline case. At the SCMW and Stearns Wharf, the ROC for the *Pseudo-nitzschia* model is almost entirely below the 1:1 line (Figs. 14 and 15c), meaning it is performing worse than random (AUC = 0.33). Even the optimized prediction point falls below the 1:1 line, although if it were optimized relative to the BS, this score would fall in the upper right quadrant where the model performs somewhat better than chance. For the pDA model at Stearns Wharf, the ROC is well below the 1:1 line with an AUC of 0.04, reflecting the poor performance of the model at this site as well as the very few DA events that took place over the study period.

3.3. Partitioning sources of error between DINEOF and ROMS

Lastly, we evaluate error generated by the input variables used to compute the empirical HAB models. The DINEOF reconstruction of near-daily MODISA data introduces error to the ocean color imagery, and these error fields for the reconstruction are obtained from cross-validation of the DINEOF fit (standard in the DINEOF routine), thereby producing both the fully reconstructed data and information about the error covariance for the same data. Given that the expected error for the reconstructed Chl and $R_{rs}(0^+, \lambda)$ fields rarely exceeds one standard deviation from the mean in our routine runs (data not shown), the error introduced by the DINEOF routine is likely not a significant error source for our predictions.

A second source of error from DINEOF could be introduced by the length of time interval used to run the reconstruction. It is not entirely clear if the smaller degree of variability present in a shorter time series (particularly one with a high percentage of gaps due to clouds or other flagged data) skews the EOFs towards spurious patterns or if the opposite is true. Fig. 16 compares the DINEOF Chl data reconstruction performed in 180-day intervals, where for the last 10-day period DINEOF was run on the previous 180 days (days 1–180, 2–181, etc.), versus a 365-day interval (days 1–365). This is tested using either a 3 × 3 box around the SCMW or extracting the nearest 1 × 1 pixel relative to the SCMW. Sequentially adding days for ten days does very little to change the output. Log-transformed Chl values for this shorter interval (red lines in Fig. 16a) generally track the 365-day runs (solid and dashed blue lines in Fig. 16a) until the end of the record. Shifting from a 3 × 3 window to the nearest pixel approach (red lines in Fig. 16b) shows convergence of the 180-day intervals with the 365-day interval (solid blue line) in that same pixel space. Several of those 180-day intervals within the ten-day period that was tested exhibit no variability (i.e. flat line).

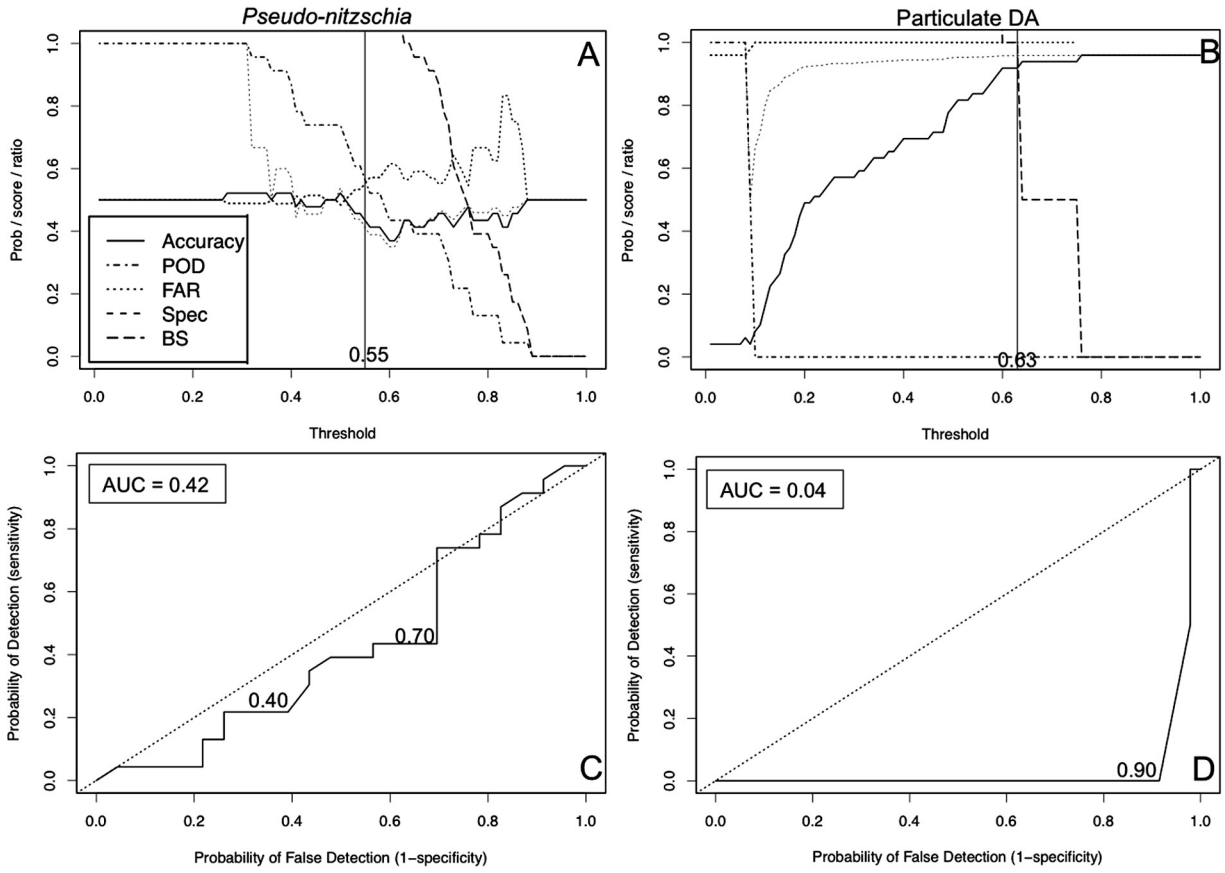

Fig. 15. Contingency plots and ROC Curves to assess model performance at Stearns Wharf. (A) *Pseudo-nitzschia* and (B) pDA binary observations (transformed using bloom/event thresholds) are compared with modeled probabilities at the SCMW pixel. Metrics used are accuracy, probability of detection (POD), false alarm rate (FAR), probability of false detection (POFD), and bias score (BS); the vertical line is the optimized prediction point–0.55 for *Pseudo-nitzschia* and 0.63 for pDA. ROC curves compare (C) *Pseudo-nitzschia* and (D) pDA binary observations (transformed using bloom/event thresholds) with modeled probabilities at the SCMW pixel. Values in boxes are the area under the curve (AUC), and values along the ROC curve show positions of various prediction points along the range 0–1 used to model POFD vs. POD.

This can occur when data close to land are included, since small errors in the satellite navigation or in viewing angle result in the pixels being classified as land (NaN values), resulting in a flat line.

The 3-km ROMS temperature and salinity fields can be evaluated with physical measurements from the SCMW in the same manner as the HAB empirical models. Direct comparison of the closest ROMS grid point with the SCMW time-series shows reasonable agreement for salinity (Fig. 16c) and temperature (Fig. 16d), although many of the large drops in salinity below 30 psu from winter/spring storms are missed by ROMS, likely due to the lack of terrestrial freshwater input to the model. Using as much *in situ* data from the SCMW as possible, we calculate the pDA model ("observed") and compare those outcomes with our routine pDA predictions that use ROMS salinity and DINEOF chlorophyll ("model"). We plot the outcomes of alternately substituting the DINEOF chlorophyll with observed chlorophyll ("pDA CHL) and the ROMS salinity field with observed salinity ("pDA Salt"). The results suggest that the pDA model is more sensitive to substitutions of ROMS salinity ($r^2 = 0.37$, slope = 0.71) than DINEOF-filled ocean color values ($r^2 = 0.16$, slope = 0.58). This sensitivity analysis demonstrates the disproportionate effect that salinity has on the pDA model outcomes relative to chlorophyll (Fig. 17).

4. Discussion

4.1. Skill assessment at shore stations

Model skill was evaluated using a suite of performance metrics for dichotomous predictions using the best available observational data from shore stations in several important locations. This analysis demonstrates that the model provides information that is in relative agreement with shore station data but with significant differences that require continual evaluation and offshore sampling in order to ascertain the scales of variability captured by the model versus the nearshore monitoring. In other words, while these HABMAP monitoring data represent the most consistent source of HAB observations in the CCS, there are obvious problems with employing them in model evaluation given the limitations in spatial (and temporal) scales. For instance, the spatial resolution of the final HAB model predictions is constrained by the lowest resolution input product, which is the 3-km CA ROMS model, such that these moderate-resolution modeled matchups (nearest pixels) are only rough approximations of HAB activity in the nearshore zone. Spatial mismatch between *in situ* observations and interpolated 3-km satellite imagery is indeed an impediment to validation studies of this sort (Carvalho et al., 2011). Unfortunately, there is no consistent source of offshore data for comprehensive skill assessment. Nevertheless, we must assess the model relative to the HAB shore station data that are the baseline for monitoring ecosystem health in coastal California (Kudela et al., 2015).

Given the spatial mismatch just mentioned and the many sources of error that contribute to the HAB model signal, it is somewhat surprising that there is any skill between the model-observation matchups generated by our approach. However, we do see some degree of *sensitivity* and *specificity* as well as integration of several scales of variability, even at the individual pixel level. The *Pseudo-nitzschia* model is less accurate than the pDA model due to its high false-positive rate. Model sensitivity (or POD in

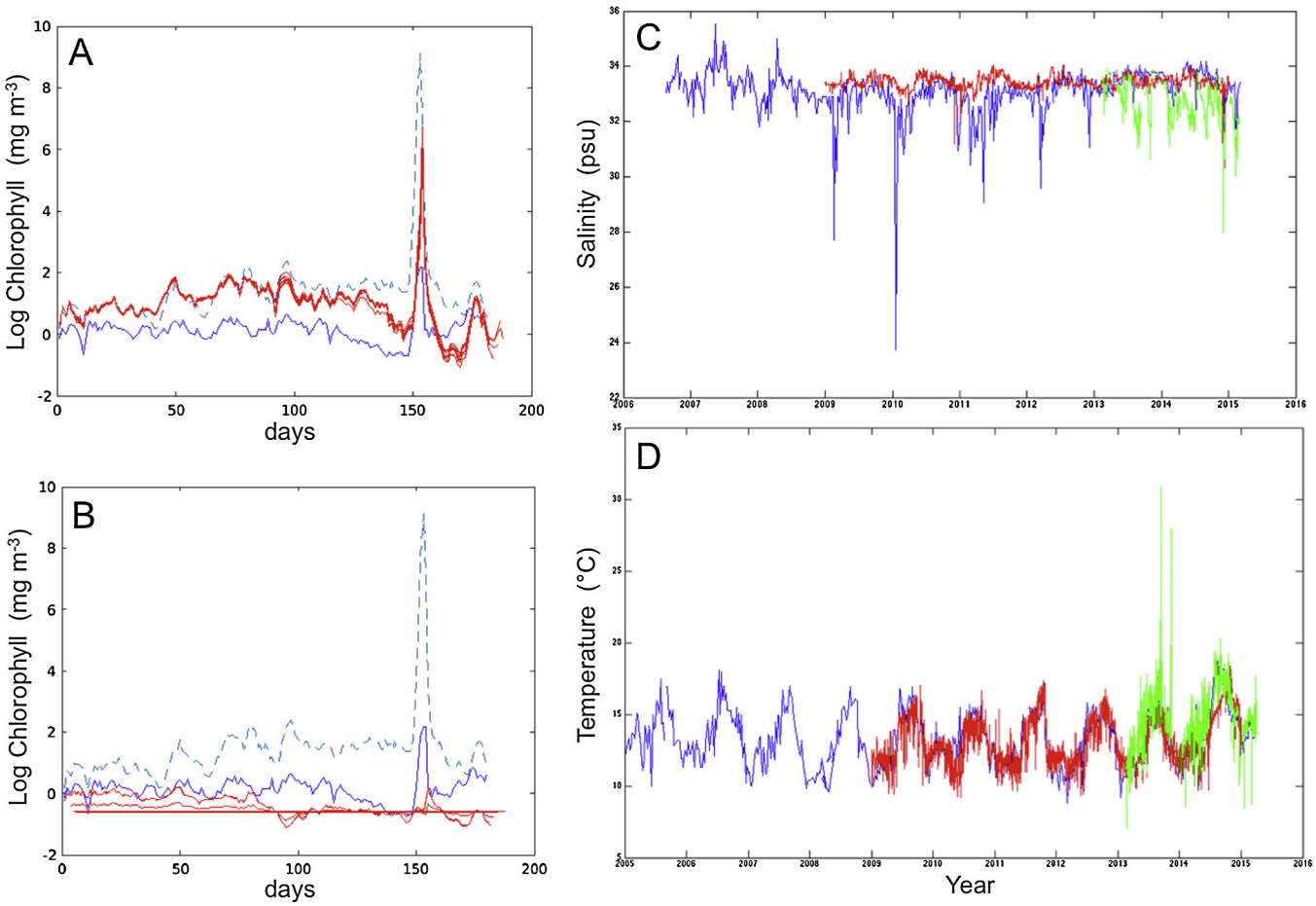

Fig. 16. DINEOF-reconstructed chlorophyll (A) extracted from a 3 × 3 box around the location of the SCMW using consecutive 180-day intervals for 10 days (red lines); extracted from a 1 × 1 pixel closest to the SCW using 365 days to run DINEOF (solid blue lines); and extracted from a 3 × 3 box around the SCMW using 365 days of data but truncated to the first 180 days (dashed blue lines). In (B), the blue lines remain the same, but the red lines are now the closest 1 × 1 pixel to the SCW and more closely track the solid blue line. The right panel shows matchups between SCMW and ROMS (C) salinity ($r = 0.25$, $p \ll 0.001$) and (D) temperature ($r = 0.78$, $p \ll 0.001$), where blue = SCMW (calibrated thermometer; YSI-3100), green = SCW (YSI Sonde-6600), and red = ROMS (nearest 3 km grid point).

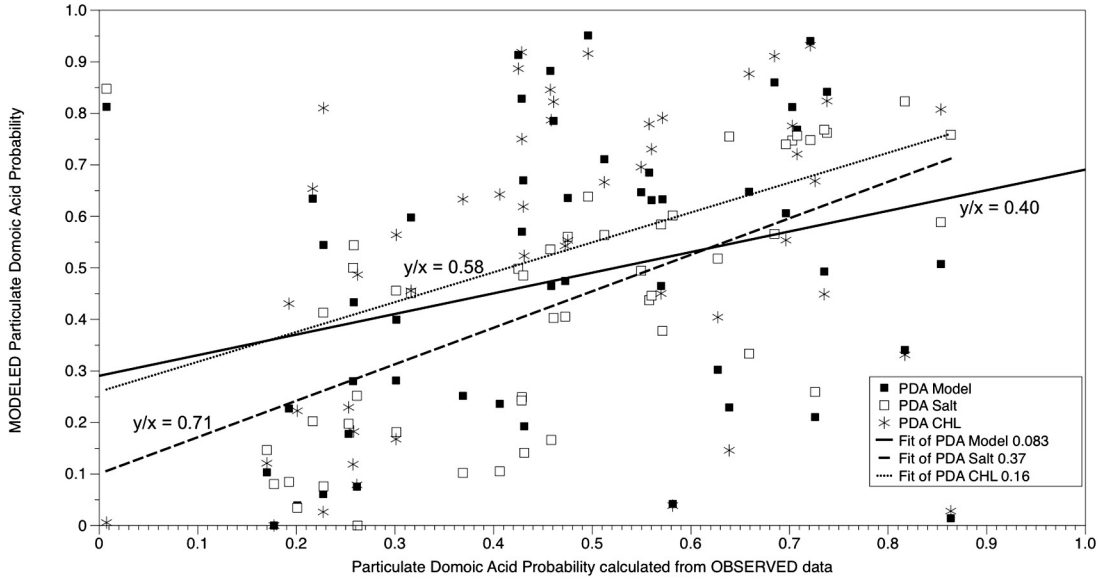

Fig. 17. Sensitivity of the HAB statistical model to estimated or simulated variables. Particulate DA probability at the SCMW for the study period is calculated with observed parameters (x-axis) measured at the SCW (salinity, chlorophyll) and DINEOF-filled $R_{rs}(555)$ values. These values are plotted against pDA probability calculated using ROMS salinity and DINEOF-filled ocean color values (filled square, "model"), as shown for previous figures; substituting ROMS salinity with laboratory-measured salinity from a YSI-3100 at the SCW (open square, "SALT"); and substituting DINEOF-filled Chl values with filter-pad measured Chl from the SCW (asterisk, "CHL"). Dashed and solid lines are linear fits to the data, and R^2 values are reported in the legend; slopes (y/x) are reported on the line itself. The model is most sensitive to substitution of observed salinity in place of simulated salinity as compared with Chl.

Figs. 13–15) describes the fraction of DA events correctly predicted and is above 50% at the optimized prediction point shown in Fig. 14b. The overall problems with model performance when assessed at the pixel level relative to very nearshore, pier-sampling matchups is corroborated by *in situ* observations from the Santa Barbara Channel where *Pseudo-nitzschia* and pDA at Stearns Wharf rarely align with offshore levels in the middle of the channel (Umhau et al., submitted for publication). Future analyses of all offshore datasets in the CA region will allow us to fully evaluate alongshore variations in model performance and refine the GLMs.

The SCMW appears to be capturing a mammal-stranding signal at a two-week lead time despite the fact that the stranding data are grouped for the entire central/northern CA region (Fig. 7b). The peaks in strandings generally correspond with times when pDA is present at the SCMW, although there are many instances when strandings occur in the absence of nearshore pDA measured at SCMW. This suggests a possible offshore (or even subsurface) source of DA to these stranded animals that is not explicitly captured by the inputs to the empirical models, but might be driving the broader spatial signal in the low to moderate-resolution input fields. It may also point to failures of the model. This is consistent with studies examining the subsurface initiation or seeding of *Pseudo-nitzschia* spp. blooms in coastal California (Seegers et al., 2015). It may also point to failures of the model. The 5-day lead-time between predicted pDA and mussel tissue at the SCMW (Fig. 9a) also signifies that C-HARM could provide a useful warning tool to aquaculture and recreational shellfish growers who stand to benefit from advanced warning of DA contamination.

Since SPATT represents a time-integrative (but spatially discrete) detection method of dissolved toxin (Lane et al., 2010; MacKenzie, 2010), it is not surprising that model variability more resembles SPATT variability than that of the sentinel mussel species, *Mytilus californianus*. Lane et al. (2010) showed that "DA-signaling" by SPATT preceded CDPH detection of dangerous levels of DA in mussels by eight weeks and is often uncoupled from temporally and spatially discrete measures of pDA at the SCMW. However, the exact lead-lag relationship between particulate DA, SPATT DA, mussel DA, and modeled DA is more difficult to discern in the present study given the complicated results of the cross correlation functions (Fig. 9). Both SPATT and the model appear to capture a larger, regional signal than traditional sampling methods and may be more tightly coupled with one another than either of them is with other discrete sampling methods. Ultimately, all sources of information, from models to *in situ* sampling need to be considered as a whole for sound decision-making.

An additional and important consideration is that our choice of threshold or prediction point clearly alters model form and sensitivity/specificity (Figs. 15–17) and is always subject to future adjustment to assess its relevancy to ecosystem response. As observational data are constantly assimilated into our statistical models for re-tuning, there will be no single, static optimized probability threshold but rather a dynamic assessment built into the protocol for issuing HAB alerts and bulletins across the different CA subregions.

4.2. Sources of error

Model skill is ultimately dependent on several factors, including error in the input variables, error in the statistical habitat models and mismatch (spatial, temporal, and scale) between model results and observations models (GLMs), as already discussed. Our method of merging DINEOF-filled MODIS imagery and ROMS fields introduces several sources of error before the HAB empirical models are even computed. First, the ocean color imagery itself represents an imperfect estimation of $R_{rs}(\lambda, 0^+)$ in the coastal zone that varies depending on the satellite sensor. Kahru et al. (2014) recently compared Chl retrievals for the CCS from SeaWiFS, MODIS-Terra, MODIS-Aqua, MERIS, and VIIRS sensors relative to available *in situ* Chl observations for the region. All sensors, with the exception of MERIS, displayed low accuracy at estimating Chl concentrations above 1 mg m^{-3}, a value that often distinguishes coastal environments from pelagic/oligotrophic zones. In the case of MODISA, Chl retrievals and *in situ* measurements were related by an R^2 of 0.32 and a 28% underestimation rate in the medium to high Chl range (Kahru et al., 2014). While the imagery is excellent for its large spatial coverage and operational reliability, we must contend with the downstream influence on HAB model predictions.

The application of DINEOF introduces error to the ocean color data in that the gap-filling procedure is an imperfect interpolation method but one that can be quantified in terms of an uncertainty estimate generated from the iterative EOF solution (Beckers et al., 2006). The uncertainty for both the reconstructed Chl and remote sensing reflectance fields is within a standard deviation of the mean value and well below the error exhibited for the actual MODISA retrievals (Kahru et al., 2014). The same DINEOF methods compared well with a coupled physical-biological model in the Pacific Northwest (Giddings et al., 2014), providing further indirect evidence that the interpolation scheme does not introduce substantial error. Another factor influencing variability in the HAB predictions is the *time window* used during DINEOF data reconstruction. There is some evidence that a long time domain containing some heavily clouded imagery can produce spurious temporal variability, but that the data reconstruction can be improved by filtering out spikes in the temporal covariance matrix before the EOF decomposition (Alvera-Azcárate et al., 2009). In our sensitivity analysis, the shorter (180-days; red lines in Fig. 16) and longer (365-days; blue lines in Fig. 16) time windows did not significantly diverge from one another until the end of the 180 day record, suggesting that a shorter time domain preserves the majority of the temporal variability captured by the longer time domain while also allowing improved computational efficiency. In terms of the influence on spatial variability, Beckers et al. (2006) suggest a nested DINEOF method in which a large-scale (up to global) reconstruction is first created, and then the anomaly (reconstruction minus observed values) is reconstructed at higher spatial and temporal resolution, followed by reconstruction at even higher resolution, etc. Since each reconstruction is influenced by the scale of patterns inherent in the raw data, this method would sequentially include more small-scale features. Exploration of these methods is part of our on-going investigation into the appropriate use of DINEOF for operational modeling. This ultimately involves trade-space constraints between improving model output and (potentially dramatically) increasing computational cost.

Lastly, the ROMS nowcast and forecast fields also have uncertainties. This uncertainty has been gradually reduced over the past few years with the improvement of data assimilation algorithms and increase in observational data sets. The RMS difference of SST between the ROMS nowcast and satellite measurements is about 0.4 °C during the winter seasons but increased to 0.8 °C during the summer due to the lack of satellite-retrieved SST when there are clouds. The model uncertainty of temperature below the ocean surface is significantly larger simply because of the lack of subsurface data. Using the independent data from CalCOFI that are not assimilated into ROMS, we have estimated the ROMS nowcast uncertainty to be on the order of 1 °C for temperature and 0.2 psu for salinity (data not shown). It is expected that a more sophisticated data assimilation scheme (e.g., ensemble Kalman Filter or 4DVAR) would further reduce model uncertainty. Improvements to ROMS are underway, and these include coupling the hydrodynamic model with a

biogeochemical/ecosystem model (e.g., CoSiNE) with the goal of providing nowcast and forecast fields for biogeochemical variables to the HAB forecasts.

4.3. The Decision-Making Context

Many methods of determining model skill categorize a model as poor quality if it does not capture event scale activity in a particular location despite the veracity of the regional signal. However, the model is likely quite valuable to a decision-maker, particularly if it provides offshore, spatial information that would otherwise not be available to managers and provides complementary information to nearshore monitoring data. According to Murphy (1993), in addition to consistency and quality, the "goodness" of a model is measured by its *value*, or its ability to aid the decision-maker in achieving some benchmark improvement. Without a high-resolution, observational record with which to evaluate the true spatial and temporal skill of our C-HARM system, it is difficult to quantify the added value brought to bear by having such a regionally comprehensive, predictive capability. For instance, the newly launched Catalina Sea Ranch located in the San Pedro Basin is the first offshore aquaculture operation within the United States exclusive economic zone (http://www.catalinasearanch.com/). It stands to reason that a model with the ability to predict DA events in the waters surrounding this cultivation site will be an asset to business decisions, in this context. At the state level, the CDPH, which is mandated to monitor biotoxins in the commercial and recreational shellfish supply, will hopefully be able to better allocate sampling resources to regions with high alert levels, particularly if the model signals an event outside of the normal targeted sampling zones. Marine mammal resource managers already report extensive use of the routine model product to guide their efforts and assess past risk likelihoods for stranded animals at the regional level. It will behoove us to continue to assess the skill of the predictions in both nowcast and forecast mode to demonstrate its performance over a variety of time and space scales.

4.4. Towards an operational forecasting system

With the completion of a yearlong pilot study to establish a HAB prediction and forecasting system for coastal California, we have demonstrated the capacity to routinely produce spatially explicit outlooks of the HAB horizon for neurotoxic *Pseudo-nitzschia* blooms. In fact, truly operational HAB forecasts in the United States only exist for *Karenia brevis* in the Gulf of Mexico at the time of this writing. The NOAA Harmful Algal Bloom Operational Forecast System issues routine *K. brevis* and brevetoxin conditions reports for coastal Florida and Texas (http://tidesandcurrents.noaa.gov/hab/) along with metadata in HAB Bulletins that collate the observational, satellite, model, and public health information relevant to making forecasts of bloom and respiratory irritation risk (Stumpf et al., 2009). While some of the methods differ for the CCS and Gulf of Mexico, the most notable being the use of a Chl anomaly product for detecting *K. brevis* biomass (Stumpf et al., 2003), the fundamental approach of merging available information from multiple platforms to create a forecast is similar to that described in this paper. The Ecological Forecasting Roadmap was established by NOAA to guide the next generation of ecological forecasting services and places a strong emphasis on HABs (NOAA, 2012). The goal is to include C-HARM in the suite of HAB forecasting programs on the verge of transition from *Research and Development* to *Operations* in regions such as those in the Pacific Northwest (Trainer and Suddleson, 2005), Lake Erie (Wynne et al., 2013), Chesapeake Bay (Brown et al., 2012), and Gulf of Maine (Kleindinst et al., 2014; McGillicuddy et al., 2011). As with these efforts, assessing model skill relative to the current baseline is a fundamental task for initiating a robust early warning system for HABs.

Acknowledgements

Funding was provided by the National Aeronautics and Space Administration (awards NNX13AL28G and NNX14AC42G to Anderson and Kudela) and to MBARI/CeNCOOS by the National Oceanic and Atmospheric Administration IOOS (award NA11NOS0120032).

Long-term monitoring has been supported by NOAA ECOHAB (award NA11NOS4780030 to Kudela) and California Sea Grant/Ocean Protection Council (award R/OPCCONT-12A to Kudela and Anderson). Significant contributions by Jennifer Patterson and Aric Bickel at CeNCOOS and the entire team at Axiom Consulting and Design have helped advance visualization and communication of the operational forecasting effort. We are extremely grateful to the dedication of the HABMAP monitoring team without whom most of the model validation would not be possible: Meredith Howard, David Caron, Kendra Negrey, G. Jason Smith, Kelsey Loupy, Heather McNair, Jayme Smith, Rebecca Shipe, Meghan Wilson, Melissa Carter, and SCCOOS. Lastly, we sincerely thank several anonymous reviewers whose comments greatly improved the manuscript.[SS]

References

Alvera-Azcárate, 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 Model. 9, 325–346, http://dx.doi.org/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. J. Geophys. Res.: 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 Sci. 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. Mar Ecol Prog. Ser. 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 , http://dx.doi.org/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. J. Mar. Syst. 83, 127–140, http://dx.doi.org/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, http://dx.doi.org/10.1016/j.hal.2008.10.005.

Anderson, D.M., Cembella, A.D., Hallegraeff, G.M., 2012. Progress in understanding harmful algal blooms: paradigm shifts and new technologies for research, monitoring, and management. Annu. Rev. Mar. Sci. 4, 143–176.

Bargu, S., Goldstein, T., Roberts, K., Chunyan, L., Gulland, F.M.D., 2012. *Pseudo-nitzschia* blooms, domoic acid, and related California sea lion strandings in Monterey Bay, California. Mar. Mammal Sci. 28, 237–253, http://dx.doi.org/10.1111/j.1748-7692.2011.00480.x.

Beckers, J.M., Barth, A., Alvera-Azcárate, A., 2006. DINEOF reconstruction of clouded images including error maps? Application to the Sea-Surface Temperature around Corsican Island. Ocean Sci. 2, 183–199.

Beckers, J.M., Rixen, M., 2003. EOF calculations and data filling from incomplete oceanographic data sets. J. Atmos. Oceanic Technol. 20, 1839–1856.

Brown, C.W., Hood, R.R., Long, W., Jacobs, J., Ramers, D.L., Wazniak, C., Wiggert, J.D., Wood, R., Xu, J., 2012. Ecological forecasting in Chesapeake Bay: using a mechanistic-empirical modelling approach. J. Mar. Syst..

Carvalho, G.A., Minnett, P.J., Banzon, V.F., Baringer, W., Heil, C.A., 2011. Long-term evaluation of three satellite ocean color algorithms for identifying harmful algal blooms (*Karenia brevis*) along the west coast of Florida: a matchup assessment. Remote Sens. Environ. 115, 1–18.

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

Chao, Y.Z.L., Farrara, J., McWillliams, J.C., Bellingham, J., Capet, X., Chavez, F., Choi, J.K., Davis, R., Doyle, J., Frantaoni, D., Li, P.P., Marchesiello, P., Moline, M.A.,

Paduan, J., Ramp, S., 2009. Development, implementation and evaluation of a data-assimilative ocean forecasting system off the central. Deep-Sea Res. II.
Doyle, J.D., Yiang, Q., Chao, Y., Farrara, J., 2009. High-resolution real-time modeling of the marine atmospheric boundary layer in support of the AOSNII field campaign. Deep-Sea Res. II.
Egbert, G.D., Erofeeva, S.Y., 2002. Efficient inverse modeling of barotropic ocean tides. J. Atmos. Oceanic Technol. 19, 183–204.
Fawcett, T., 2004. ROC graphs: notes and practical considerations for researchers. Mach. Learn. 31, 1–38.
Fritz, L., Qulliam, M.A., Wright, J.L.C., Beale, A.M., Work, T.M., 1992. An outbreak of domoic acid poisoning attributed to the pennate diatom *Pseudonitzschia australis*. J. Phycol. 28, 439–442.
Frolov, S., Kudela, R.M., Bellingham, J.G., 2013. Monitoring of harmful algal blooms in the era of diminishing resources: a case study of the US West Coast. Harmful Algae 21, 1–12.
Giddings, S.N., MacCready, P., Banas, N.S., Davis, K.A., Siedlecki, S.A., Hickey, B.M., Trainer, V.L., Kudela, R.M., Pelland, N., 2014. Hindcasts of harmful algal bloom transport on the Pacific Northwest coast. J. Geophys. Res., in review.
Jin, D., Thunberg, E., Hoagland, P., 2008. Economic impact of the 2005 red tide event on commercial shellfish fisheries in New England. Ocean Coast. Manage. 51, 420–429.
Jolliffe, I.T., Stephenson, D.B., 2003. Forecast Verification. England and Hoboken, Chichester.
Jolliffe, I.T., Stephenson, D.B., 2012. Forecast Verification: A Practitioner's Guide in Atmospheric Science. John Wiley & Sons.
Kahru, M., Kudela, R.M., Anderson, C.R., Manzano-Sarabia, M., Mitchell, B.G., 2014. Evaluation of satellite retrievals of ocean chlorophyll-a in the California Current. Remote Sens. 6, 8524–8540.
Kleindinst, J.L., Anderson, D.M., McGillicuddy, D.J., Stumpf, R.P., Fisher, K.M., Couture, D.A., Hickey, J.M., Nash, C., 2014. Categorizing the severity of paralytic shellfish poisoning outbreaks in the Gulf of Maine for forecasting and management. Deep Sea Res. II: Top. Stud. Oceanogr. 103, 277–287.
Kudela, R.M., Banas, N.S., Barth, J.A., Frame, E.R., Jay, D.A., Largier, J.L., Lessard, E.J., Peterson, T.D., Woude, A.J.V., 2008. New insights into the controls and mechanisms of plankton productivity in coastal upwelling waters of the northern California Current System. Oceanography 21, 46–59.
Kudela, R.M., Bickel, A., Carter, M.L., Howard, M.D., Rosenfeld, L.K., 2015. The monitoring of harmful algal blooms through ocean observing: the development of the California Harmful Algal Bloom Monitoring and Alert Program. In: Liu, Y.Y., Kerkering, H.A., Weisberg, R.H. (Eds.), Coastal Ocean Observing Systems. Elsevier Inc, pp. 58–75.
Kvitek, R.G., Goldberg, J.D., Smith, G.J., Doucette, G.J., Silver, M.W., 2008. Domoic acid contamination within eight representative species from the benthic food web of Monterey Bay, California, USA. Mar. Ecol. Prog. Ser. 367, 35–47.
Lane, J.Q., Raimondi, P.T., Kudela, R.M., 2009. Development of a logistic regression model for the prediction of toxigenic *Pseudo-nitzschia* blooms in Monterey Bay, California. Mar. Ecol. Prog. Ser. 383, 37–51.
Lane, J.Q., Roddam, C., Meiling, C., Langlois, G.W., Kudela, R.M., 2010. Application of Solid Phase Adsorption Toxin Tracking (SPATT) for field detection of the hydrophilic phycotoxins domoic acid and saxitoxin in coastal California. Limnol. Oceanogr.-Methods 8, 645–660, http://dx.doi.org/10.4319/lom.2010.8.0645.
Lefebvre, K.A., Powell, C.L., Busman, M., Doucette, G.J., Moeller, P.D.R., Silver, J.B., Miller, P.E., Hughes, M.P., Singaram, S., Silver, M.W., Tjeerdema, R.S., 1999. Detection of domoic acid in northern anchovies and California sea lions associated with an unusual mortality event. Nat. Toxins 7, 85–92.
Lehmann, E.L., Romano, J.P., 2006. Testing Statistical Hypotheses. Springer Science & Business Media.
Lewitus, A.J., Horner, R.A., Caron, D.A., Garcia-Mendoza, E., Hickey, B.M., Hunter, M., Huppert, D.D., Kudela, R.M., Langlois, G.W., Largier, J.L., 2012. Harmful algal blooms along the North American west coast region: history, trends, causes, and impacts. Harmful Algae 19, 133–159.
Li, Z., Chao, Y., McWilliams, J.C., Ide, K., 2008. A three-dimensional variational data assimilation scheme for the Regional Ocean Modeling System: implementation and basic experiments. J. Geophys. Res. 113, C05002, http://dx.doi.org/10.1029/2006JC004042.
Li, Z., Chao, Y., McWilliams, J.C., Ide, K., 2009. A three-dimensional variational data assimilation scheme for the regional ocean modeling system. J. Atmos. Ocean. Technol. 25, 2074–2090.
MacKenzie, L.A., 2010. In situ passive solid-phase adsorption of micro-algal biotoxins as a monitoring tool. Curr. Opin. Biotechnol. 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. J. Geophys. Res.: Oceans (1978–2012) 112.
McCabe, R.M., Hickey, B.M., Kudela, R.M., Lefebvre, K.A., Adams, N.G., Bill, B.D., Gulland, F.M.D., Thomson, R.E., Cochlan, W.P., Trainer, V.L., submitted for publication. An unprecedented coastwide toxic algal bloom linked to anomalous ocean conditions. Geophys. Res. Lett., accepted.
McGillicuddy Jr., D.J., 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. Limnol. Oceanogr. 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. J. Geophys. Res.: 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 Forecast. 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 Dynam. 61, 1205–1214.
NOAA, 2012. A Strategic Vision for NOAA's Ecological Forecasting Roadmap: 2015–2019. National Oceanic and Atmospheric Administration, Office of the Chief Administrative Officer.
Scholin, C.A., Gulland, F., Doucette, G.J., Benson, S., Busman, M., Chavez, F.P., Cordaro, J., DeLong, R., De Vogelaere, A., Harvey, J., Haulena, M., Lefebvre, K.A., Lipscomb, T., Losctutoff, S., Lowenstine, L.J., Marin III, R., Miller, P.E., McLellan, W.A., Moeller, P.D.R., Powell, C.L., Rowles, T., Silvagni, P., Silver, M.W., Spraker, T., Trainer, V.L., Van Dolah, F.M., 2000. Mortality of sea lions along the central California coast linked to a toxic diatom bloom. Nature 403, 80–83.
Seegers, B.N., Birch, J.M., Marin III, R., Scholin, C.A., Caron, D.A., Seubert, E.L., Howard, M.D.A., Robertson, G.L., Jones, B.H., 2015. Subsurface seeding of surface harmful algal blooms observed through the integration of autonomous gliders, moored environmental sample processors, and satellite remote sensing in southern California. Limnol. Oceanogr. 60, 754–764.
Sekula-Wood, E., Benitez-Nelson, C., Morton, S.L., Anderson, C.R., Thunell, R., 2010. Pseudo-nitzschia and domoic acid fluxes in Santa Barbara Basin (CA) from 1993 to 2008. Harmful Algae 10, 567–575.
Sekula-Wood, E., Schnezer, A., Benitez-Nelson, C., Anderson, C.R., Berelson, W.M., 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. Nat. Geosci. 2, 342–350, http://dx.doi.org/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 Modell. 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. p. 2008.
Song, Y., Haidvogel, D.B., 1994. A semi-implicit ocean circulation model using a generalized topography-following coordinate system. J. Comput. Phys. 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. J. Mar. Syst. 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. Limnol. Oceanogr. 45, 1818–1833.
Trainer, V.L., Bates, S.S., Lundholm, N., Thessen, A.E., Cochlan, W.P., Adams, N.G., Trick, C.G., 2012. *Pseudo-nitzschia* physiological ecology, phylogeny, toxicity, monitoring, and impacts on ecosystem health. Harmful Algae 14, 271–300.
Trainer, V.L., Suddleson, M., 2005. Monitoring approaches for early warning of DA events in Washington State. Oceanography 18, 228–237.
Umhau, B., Benitez-Nelson, C.B., Anderson, C.R., Thunell, R.C., Burns, A., McCabe, K., Burrell, C., 2016. Water column distributions, and sinking and sediment records of *Pseudo-nitzschia* and domoic acid in the Santa Barbara Basin, California. Limnol. Oceanogr. (submitted for publication).
Wang, X., Chao, Y., Dong, C., Farrara, J., Li, Z., McWilliams, C., Paduan, J.D., Rosenfeld, L.K., 2009. Modeling tides in Monterey Bay, California. Deep-Sea Res. II.
Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Fahnenstiel, G.L., Dyble, J., Schwab, D.J., Joshi, S.J., 2013. Evolution of a cyanobacterial bloom forecast system in western Lake Erie: development and initial evaluation. J. Great Lakes Res. 39, 90–99.