Satellite derived indices of red tide severity for input for Gulf of Mexico Gag grouper stock assessment

John F. Walter, Mary C. Christman, Jan Landsberg, Brian Linton, Karen Steidinger, Richard Stumpf, and Jacon Tustison

SEDAR33-DW08

20 May 2013

Please cite as:

Walter, J.F., M.C. Christman, J. Landsberg, B. Linton, K. Steidinger, R. Stumpf, and J. Tustison. 2013. Satellite derived indices of red tide severity for input for Gulf of Mexico Gag grouper stock assessment. SEDAR33-DW08. SEDAR, North Charleston, SC. 43 pp.

Satellite derived indices of red tide severity for input for Gulf of Mexico gag grouper stock assessment

John Walter[1], Mary Christman[3], Jan Landsberg[2,] Brian Linton[1], Karen Steidinger[2], Richard Stumpf[4] and Jacob Tustison[2]

[1] Southeast Fisheries Science Center. 75 Virginia Beach Drive, Miami, FL 33149. John.f.walter@noaa.gov

[2] Fish and Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, 100 Eighth Ave SE, St. Petersburg, FL 33701

[3] MCC Statistical Consulting LLC. 2219 NW 23rd Ter. Gainesville, FL 32605

[4] NOAA, National Ocean Service, 1305 East-West Highway, rm 9115, Silver Spring, MD, 20910.

Introduction

In 2005 a severe red tide (*Karenia brevis*) event affected a vast area of the West Florida shelf. This event triggered a record number of fish kill reports, beach closures, reports of respiratory distress as a result of noxious aerosols as well as economic disturbance to the West Florida region (http://research.MyFWC.com). At the time, fish kill observations indicated that large numbers of fish likely died as a result of the red tide, however it was unknown how these observations may have translated into population level impacts.

Red tide events have been recorded in the Gulf of Mexico for hundreds of years (Ingersoll, Steidinger & Joyce 1973) and have been implicated in several catastrophic mortality events of marine animals (Gunter et al. 1947; Rounsefell and Nelson 1966; Landsberg & Steidinger 1998). Red tide poses a threat to living organisms by acute exposure to brevetoxin, a neurotoxin which inhibits breathing; chronic exposure to brevetoxin through bioaccumulation; and, upon death, by creating large areas of hypoxic water (Landsberg 2002). These three threats, coupled with the high concentrations and large affected area, make it a severe ecological problem in years of large blooms (Steidinger & Vargo 1988). While red tide has been implicated in mass kills of inshore fishes, blooms also occur in offshore waters and have caused near complete extirpation of reef biota of which red grouper appear to be particularly susceptible (Smith 1975).

In 2009, update assessm`ents of red and gag grouper (Anon, 2009a, 2009b) indicated a severe decline in all indices of abundance between 2005 and 2006 which caused the models to fit poorly and to fail to be able to recreate historical population patterns. In response to the lack of fit, an additional natural mortality term was estimate``ed for 2005. This produced much improved fits to the indices and the resulting M was 0.32 and 0.27 for red and gag grouper, respectively. As the baseline M for the two species was approximately 0.14 and 0.15, the episodic M estimates for 2005 were twice the baseline M and approximately equal to the sum of natural and target fishing mortality rates. These translated to an estimated 21 and 18% of the population having died as a result of some event in 2005. The Gulf of Mexico Fisheries Management Council's Scientific and Statistical Committee accepted the results of both assessments with the additional red tide mortality term as the basis for management advice. With this action, and with the evidence, both historical and from the model fits, that red tide may severely impact groupers, it is important that future assessments explicitly consider red tide events in grouper population dynamics.

Substantial research efforts have been expended to detecting these blooms and determining their extent and concentration because of the high potential for human health impacts (Landsberg et al 2002) as well as adverse effects on biota (Gunter 1948). Satellite remotely sensed imagery is useful for detecting and evaluating the extent of blooms but is prone to false positive detections and often hampered by cloud cover. Ground-truthed water quality data provides accurate detection and quantification of HAB presence, species and concentration but is extremely expensive to conduct on the routine and systematic basis necessary for comprehensive monitoring. Use of satellite imagery for positive identification of HABs has focused on obtaining a positive or negative answer (Carvahlo et al 2011) rather than a probabilistic prediction. Few proposed algorithms have performed exceptionally well under all circumstances. We posit that a more comprehensive approach will be to consider the use of all available remotely sensed data to develop a predictive model for red tide presence. A probabilistic prediction allows for the use and evaluation of multiple factors that might determine the presence of a HAB. Factor evaluation within the model testing framework broadens the scope of potential predictors, rather than focusing on simply evaluating the performance of a single predictor or algorithm. Rather than simply seeking a positive or negative result for HAB presence, a model-based prediction approach provides a mapping of the probability of a HAB occurring and associated prediction error.

FWRI maintains an extensive harmful algal bloom (HAB) database of water samples going back over 50 years (FWRI). The samples contained in this database come from a combination of systematic, opportunistic and event-response sampling where sampling is only conducted in response to an event. For these reasons it is very difficult to make inferences on trends in red tide severity from this data (Young and Christman 2010). In contrast satellite remote sensing data provides information on ambient conditions generally free from sampling biases and only prone to lapses in coverage due to cloud coverage or sensor failure.

In this paper we develop indices of red tide severity from a generalized additive model (GAM) that predicts the probability of a red tide bloom using a suite of satellite derived remote sensing products and the FWRI HAB database. These indices will then be incorporated as environmental covariates (Deriso et al. 2008) into stock assessment models. Several derived indices constituting different spatial and temporal partitions are created based upon hypothesis regarding the spatial and temporal overlap of grouper populations with red tide blooms. Then the 'best' candidate index is determined on the basis of providing the best fit to the original 2009 gag and red grouper stock assessment models.

Methods

Satellite data processing

Satellite data was obtained from Sea-viewing Wide Field-of-view Sensor (SeaWIFS) daily or twice daily imagery for the years 1998-2010 starting in July of 1998 (Figures 1 and 2 show example images). Daily or twice daily imagery was processed onto a 1.1 km grid over the Eastern Gulf of Mexico, resulting in 12.2 million data points. Seven derived products potentially considered to be indicative of *K. brevis* blooms (Tomlinson et al 2009) were evaluated as candidate factors for the predictive models:

1. Chlorophyl concentration obtained by the NASA SeaWiFS OC4 algorithm.
2. Chorophyl anomaly is calculated as the anomaly against a 60 day average chlorophyl value lagged by 15 days (Stumpf et al. 2003. This measure has been used to identify anomalous phytoplankton levels, and hence likely *K. brevis* blooms, from baseline chlorophyl levels.
3. Morel- backscatter for algae (Morel 1988).

4. Carder- the ratio of total backscatter from the Carder method (Cannizaro 2009) to Morel hypothetical backscatter from normal phytoplankton. Low values are generally indicative of *K brevis*.
5. CMbbp- is the ratio of observed backscatter to that expected for marine algae.
6. Rrs670- remote sensing reflectance at 670 nm.
7. ssnLw490- spectral shape at 490 nm that has been observed to identify *K. brevis* blooms from non-*K. brevis* features and was applied. The Wynne et al. (2008) spectral shape algorithm was used.
8. HAB_ensemble is a three-part algorithm that uses chorophyl anomaly,CMbbp, and ssnLw490
9. SST: satellite derived sea surface temperatures were also obtained from http://www.ncdc.noaa.gov/oa/climate/research/sst/oi-daily-information.php and assigned to the same grid locations.
10. Easting and Westing- Locations converted from latitude and longitude into Universal Transverse Mercator units.
11. Water depth- depth in meters at the location of each satellite record
12. Week of year and month of a satellite record.

Water quality monitoring data processing

Water quality monitoring data were provided by FWRI. The dataset consists of over 100,000 water samples collected in the Gulf of Mexico spanning the years 1950-2011 from over 50 agencies. Appendix figure 1 is included for the SEDAR 33 WG. It provides a concise summary of the documented red tide events derived from this dataset, fish kill reports and historical evidence. As satellite data are available only since 1998, we confine our study to the time frame 1998 – 2010.

HAB Data for 1998 – 2010 (n = 51200) were processed as follows. First, any records where the water sample was collected at a depth greater than 1.5 m was removed. Next, any records in which one or more predictor variables had missing values were removed from the dataset. Using this dataset, we then averaged the harmful algae counts within each satellite grid cell for each date. This partially removes the effect of clustered sampling by ships, a partial effort to remove bias caused by the event-response type sampling that is common to this database. To further remove this bias, we averaged these each grid cell values over each week of the year for each year. This left us with n = 9050 records. To check that the averaging did not cause a bias, such as underestimating the presence of a HAB, we compared the grid cell averages to the maximum concentration within a grid cell and week combination for a large subset of the data (4535 records). In only 44 (0.86%) instances did we observe a maximum concentration > 100K but a mean concentration < 100K. Hence we felt confident that the data were not biased by our summarization. The final dataset had 741 records with algal concentration greater than 100,000 and 8309 records without a bloom event. For analyses, we created the indicator variable PA100K which equals 1 when the algal concentration is greater than or equal to 100,000 cells/liter and 0 otherwise. Of interest is modeling the probability of occurrence (PA100K = 1) based on a set of potential covariates: easting, northing, date, water depth, sea surface temperature, and several of the satellite-derived variables described above.

Associating Satellite Imagery with Monitoring Data

The satellite data were assigned to an 1100 x 1100 meter grid spanning the study area (Figure 1). Each WQ record was converted from latitude and longitude to UTM zone 17 and then assigned to the SD grid cell that encompassed it (Figure 3). As the satellite data was provided for every 1-3 days throughout the year we allowed a data window of 1 day prior and 1 day after the date of the WQ record and then took the first record available within this record to assign the most proximate satellite data products.

Modeling

We started the analyses by reviewing the frequency distributions of the predictor variables and running logistic regressions of the indicator variable PA100K on each predictor variable. Exploratory analyses of a multiple regression approach using all of the predictors on the response variable included classification trees and random forests (Hothorn et al 2006; Strobl et al 2009). The stopping rules we used to grow a classification tree were: each terminal node must contain at least 25 observations, the minimum number of observations required in a node before splitting was 50, and the p-value of the test for identifying the best predictor for splitting had to be less than 0.15 with a Bonferroni adjustment for multiple testing. Since some of the predictors are correlated with each other, we also performed a random forest analysis to determine the relative importance of each predictor variable. The random forest analysis was run 2000 times using 5 randomly selected covariates each time. Analyses were done using the package "party" in R (R Core Team, 2012: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.) Based on the preliminary analyses, we identified potential interactions among those variables for predicting the probability of a HAB. Generalized additive models (GAM; Hastie & Tibshirani, 1990; Wood, 2006), with tensor products for spline fitting, cross-validation for knot selection and smoothing parameters chosen to minimize the REML score, were used to obtain predictive equations of the probability of a bloom occurring. Tensor product smoothers often perform better than other smoothers when the covariates are not on the same scale. We assumed a binary distribution for the response variable with a logit link. Several models were fit, using various combinations of predictor variables and interactions, to find a "best" model among those compared. The criteria for comparing alternative models included, the adjusted R^2, deviance explained, area under the curve for receiver operating characteristic (ROC) curves, the overall misclassification rate, and the false positive and false negative rates. If prediction of an occurrence of a bloom is desired rather than a probability, the ROC curve can be used to identify the optimal cutoff whereby an estimated probability is converted to a predicted presence or absence of HAB. We identified the cutoff using the value associated with the maximum sum of the sensitivity and specificity and used it in the calculation of misclassification rates. We used the package "mgcv" for the GAM fitting and "pROC" for calculating the ROC curve in R (R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org) . Once the final model was selected, the model's predictive capability was determined by running a 5-fold cross-validation, in which a random 20% of the dataset is reserved and the model fitted to the remaining data. This model is then used to predict the probability of presence for the reserved records. Using the cutoff identified from the ROC for the model using the complete dataset, we converted the predicted probabilities to presence or absence for the subset of data kept aside. The misclassifications (absent identified as present and vice versa) were calculated for each reserved subset. Finally, using these values we calculated the false positive and false negative rates for the cross-validation. These were compared to the same rates calculated from the model based on the complete dataset.

For predictions two separate models were developed. One model, applicable to the entire region but likely poorer predictive ability was constructed. A second model, restricted to waters deeper than 10 meters was constructed to develop predictions in deeper waters, with the understanding that this model likely has greater predictive capacity due to the greater water clarity.

Prediction

Using the model developed above we can predict the probability of occurrence at any location or date for which we have values for the explanatory variables. The only caveat is that prediction should only be done within the range of the explanatory variables and not for locations of environmental conditions outside of what we have modeled.

Index Development

Indices can be developed for any sub-region and over any time frame of interest. For example, annual indices could be developed using only inshore locations (water depths less than 10 m say) for the months of October – December. So, given a set of dates and locations with associated explanatory variable values, we calculated indices as follows:

$$I_{D,L} = \sum_{dates\, \in\, D}\ \sum_{locations\, \in\, L} \hat{p}_{date,location}$$

where $\hat{p}_{date,location}$ is the predicted probability of occurrence of a bloom from the model. The predicted probability is the back-transformation based on the linear estimate, $\hat{\eta}$, from the logit link used in the GAM:

$$\hat{p}_{date,location} = \frac{\exp(\hat{\eta}_{date,location})}{1 + \exp(\hat{\eta}_{date,location})}.$$

Hence the index values can be interpreted as the probability of a bloom averaged over all prediction points within the spatial domain of the index for the entire year.

The standard error of $I_{D,L}$ is obtained using a parametric Monte Carlo approach. The model coefficients are assumed to be asymptotically multivariate Normal with variance-covariance matrix as estimated by the modeling. The Monte Carlo approach generates random vectors of coefficients from the multivariate Normal distribution. Each vector is used to calculate the predicted values for the date and location ranges. These are then summed and the index stored. We repeated the random generation of the coefficient vector 1000 times. The standard deviation of the 1000 estimates of the index is an approximation to the standard error of the index.

Index configurations

Seven separate indices that span different hypotheses regarding the effects of red tide on grouper mortality were constructed. The indices differ in their spatial and temporal domain (Figure 4) and are described below:

1. Greater than 10m and 75% Minimum convex polygon (10m 75MCP)– This index is calculated by calculating a minimum convex polygon that encompasses 75% of the HAB data and then using only prediction locations within this area and greater than 10 meters in water depth. This index represents compromise between minimizing the error in predicting outside of the spatial range of much of the HAB data and of predicting in shallow turbid waters where the predictive capacity of the model substantially degraded.
2. Greater than 10m and 75% Minimum convex polygon, August-December (10m 75MCP Aug-Dec)- This index is the same as (1) but predictions are restricted to the months August-December when the severity of red tides are often higher and water column stratification is

stronger, potentially increasing the negative effects of red tides on grouper. The water column stratification increases the potential for hypoxic water to harm fish.

3. Inshore index (Inshore)- This index is calculated only for regions with water depths less than 10 meters. It is unlikely to be a well-estimated index due to the diminished predictive ability of the satellite data for shallow-turbid waters.
4. Inshore index, August-December (Inshore Aug-Dec)- this is the same as (3) but only for August to December. It is also likely to have a high variance due to poor predictive ability in areas outside of the range of the HAB data.
5. All prediction areas- This index covers the entire prediction area.
6. Grouper index-This index is constructed exclusively for the West Florida shelf area that encompasses much of the habitat for red and gag grouper and is greater than 10 meters in water depth since the 10m 75MCP index does not cover the southern or northern portions of the Florida shelf known to be important grouper habitat.
7. Threshold index- This index is calculated as an indicator index where the 10m 75MCP index is either 0 or 1, depending upon if the calculated value is above the ROC cutoff for predicting the probability of a bloom. This logic for this index is that the negative effects on grouper may only occur under conditions where a red tide is above a threshold. Hence if the index value *averaged* over all locations and times is greater than the ROC cutoff, then this year has a high probability that the bloom was extensive in duration and concentration. This index does not have an associated variance.

Results

Statistical Modeling

A preliminary review of the data (Table 1) indicated that all of the predictors were pair-wise informative of HAB presence. The classification tree (Figure 5a) showed contributions to partitioning the variation of PA100K from all variables except easting, northing, and ssnLw490. The order of importance of the covariates for predicting presence of a HAB based on the random forest (Figure 5b) was: month, water depth, chlorophyll a, chlorophyll anomaly, Morel, and CMbbp.

The best GAM model included month (as a categorical variable), week of year interacting pairwise with chlorophyll anomaly, chlorophyll a, HAB ensemble and rrs670, Carder interacting with Morel, SST interacting with water depth, water depth, and the interaction of easting with northing(Figure 6). The smooth components indicate that the probability of presence of HAB (PA100K = 1) is higher at shallower depths between 10 and 35 m, higher values of sea surface temperature for deeper water depths, for values of rrs670 between 50 and 150 but that ranges moves higher as the week of the year increases, for high values of chlorophyll a but mostly late or early in the year, for intermediate chlorophyll anomalies during the latter part of the year or for high values of the anomaly late in the year, and for high values of both Carder and Morel

The model explained 36.6% of the total deviance, had an adjusted R^2 of 0.304, and an area under the ROC curve of 0.9046. The cutoff identified by the ROC curve was 0.0827, very close to the observed proportion of occurrences (0.0819). Using this cutoff to convert the predicted probabilities to indicator values, the 5-fold cross-validation overall misclassification rate was 19.7% with a false positive rate of 19.8% and a false negative rate of 19.3%. The model based on the full dataset had equivalent rates of 17.4%, 19.4% and 15.4%, respectively.

We also fit the model to a subset of the total dataset, namely to those observations taken offshore at water depths greater than 10 m (n = 4294). The model explained 44.2% of the total deviance, had an adjusted R^2 of 0.39, and an area under the ROC curve of 0.9295. The cutoff identified by the ROC curve was 0.0541, smaller than the observed proportion of occurrences (0.0785). Another common approach to choosing a cutoff is to use the observed proportion of presences, which in this case is higher than the cutoff chosen by the ROC curve method. Choosing a higher cutoff increases the false negative rate and lowers the false positive rate, so that worse prediction of presence of a HAB is accompanied by a better prediction of absence. The final choice should be based on the determination of which is a more serious error, misclassifying absences as presences or the converse. Using the cutoff identified by the ROC to convert the predicted probabilities to indicator values, the 5-fold cross-validation overall misclassification rate was 22.1% with a false positive rate of 23.1% and a false negative rate of 10.4%. The model based on using all of the offshore data had equivalent rates of 20.9%, 22.1% and 6.2%, respectively.

Prediction

Predictions for any particular day were often sparse due to cloud coverage. Monthly aggregate predictions obtained by averaging the probability of red tide over all prediction grids (Figures 7 and 8, but see also Appendix 2). Approximate standard errors were obtained by averaging the standard errors for each prediction point over the month. These approximate SEs provide a measure of relative precision of the predictions but should not be taken as exact estimates. We show an example of a three month time period for the of a documented severe red tide (Figure 7, October-December 2005) and a three month time period where no red tide was detected or documented (April-May 2010).

It is important to note several key points with the predictions. First these prediction surfaces are shown for the entire prediction domain and were obtained with model A. All areas. Hence the prediction errors are often high, particularly for inshore areas less than 10m and for areas to the north and west of the bulk of the HAB samples (eg. The relatively high probability for April and May 2010, off of the Panhandle of Florida and in the Big Bend region, which are commensurate with high standard errors). Hence while the predictions might indicate the presence of red tide, these predictions should considered in context with the standard errors and the poor performance of the model in these areas.

In contrast the well-documented bloom directly west of Tampa Bay in fall of 2005 shows extremely high probabilities with low standard errors, indicating that the model clearly and unambiguously documented this bloom. The model also predicts high probabilities for the Big Bend region but these predictions also have low precision.

Indices

The six calculated indices show relatively high similarity (Figure 9) but differ in both absolute magnitude and in precision. All indicate that 2005 was the highest year and all have relatively high precision for this prediction indicating that 2005 clearly stands out from the other years, except for the inshore index construction where 2002 also is high. The differences in absolute magnitude are indicative of mainly of the temporal domain of the index, e.g. the probability of a bloom is higher for August-December than for the entire year. The spatial domain also affects the absolute magnitude in that the area defined by the MCP75% has a higher bloom probability than other areas. However, the differences in absolute magnitude are uninformative to the assessment models as the indices are input only as relative values.

The different index constructions should be considered with respect to the ability of the model to predict. The best predictive model is Model B, constructed for offshore waters and the indices constructed with have the best performance as measured by the coefficients of variation. This model is less likely to give false negative, e.g. predict no red tide when it does exist than the all areas model but both have a high probability (~20%) of predicting red tide when none exists.

Discussion

This paper presents probabilistic satellite-derived indices of the probability of a *K. brevis* bloom. It combines a substantial body of work dedicated to using satellite remote sensing information to detect harmful algal blooms (see Carder & Steward 1985; Stumpf et al. 2000 , Stumpf 2001; Stumpf et al. 2003; Tomlinson 2004; Hu et al 2005; Wynne et al 2005; Cannizzaro et al 2008; Wynne et al 2008; Amin et al 2009; Carvahlo et al 2011 for an extensive but not exhaustive list) and an vast compilation of sea-based HAB concentration samples (FWRI). Our unique contribution is to develop a model that incorporates a suite of documented or hypothesized predictors of *K. brevis* blooms to be able to predict probabilities in time and space. We then to combine these predictions to develop annual indices of red tide severity which provide precisely the critical input for an environmental covariate to fisheries stock assessment models (Deriso et al 2008). This represents a substantial advancement in combining both the FWRI HAB database and remotely-sensed satellite data to provide indices that overcome some biases that might be associated with the HAB data alone (Young and Christman 2010). Furthermore, the probabilistic modeling approach also provides a measure of uncertainty both in the index values for a given year but also for any particular prediction in time and space. This probabilistic predictive model greatly advances our capacity to predict red tide events from satellite data and has many applications.

The primary contribution of this paper is the development of the predictive GAM models that incorporate a suite of predictors and their interactions and also provides prediction errors. Other papers have focused on developing methods and algorithms for the detection of a bloom as a presence/absence phenomenon (e.g., Stumpf 2001, Stumpf et al 2003; Tomlinson et al 2004; Hu et al 2005; Wynne et al 2005; Cannizzaro et al 2008; Wynne et al 2008; Tomlinson et al 2008, Amin et al 2009; Carvahlo et al 2011). Given the substantial body of literature regarding potential predictors evaluated in the above paper, we had the benefit of taking a more comprehensive approach to evaluate all potential predictors in a statistical modeling framework. Previous papers have focused on developing deterministic algorithms based upon one or a suite of predictors to determine whether red tide was present or not. The GAM modeling approach we employ uses a suite of available predictors and their interactions rather than focusing on one or several predictors. The reality is that none of the variables are perfect predictors, many interact or co-vary with each other and a modeling approach that evaluates many variables to select an adequate model and which provides probabilistic predictions better represents the state of our ability to predict red tides. We also feel that this framework more accurately represents the uncertainty inherent in the predictions and might overcome some of the limitations of any single indicator.

Overall model performance was good but declined as predictions extended into inshore areas where the satellite observations were occluded by turbid, shallow waters and into spatial areas outside of the range of the HAB data. The limitations of satellite data for predicting *K brevis* are well known (Amin et al. 2009). In turbid, inshore waters, suspended particulate matter can make detection of *K brevis* difficult (Carder et al., 1999; Hu et al., 2003). Even in clear waters satellite measures of chlorophyll and other derived products cannot always separate *K brevis* from other algal species (primarily *Trichodesmium* species) with *K brevis*. Both factors result in a tendency for satellite data to over-predict red tide resulting in a high false-positive rate. This known tendency indicates that the indices are likely to over-predict the probability of a red tide, particularly for areas outside of the bulk of the data used to build the prediction or for shallow inshore areas. It is, however, unlikely that this bias would change over time so that the constructed indices would still reflect an unbiased *relative* measure of red tide severity.

The empirically estimated ROC threshold provides an objective criterion for determining presence/absence. We use this to define the threshold index as either a vector of 0/1 values which might be useful for predicting mortality effects in groupers. The rationale for using this cut-off is that if the probability of a red tide bloom, averaged over all prediction points for a year, is above the threshold for bloom prediction it is likely that this year represents and extraordinary bloom event. With respect to grouper mortality, as red tides occur in almost all years but fish kills due to red tides only occur in a few years, it is likely that there is threshold of intensity which causes extra mortality. The threshold index may also be useful if one considers that any mortality due to baseline levels of red tide may already be included in estimates of natural mortality derived from empirical data and it may only be desirable to estimate episodic mortality (SEDAR 2012) extraordinary to this baseline.

The ROC cutoff also allows for comparison of model performance with the approaches in Carvahlo (2011). Both predictive models from this study perform as high or higher on specificity (1-false negative rate~78-80%) or the ability of the model to identify non-blooms from blooms and on sensitivity (1-false negative rate ~85-94%) or the ability detect a bloom when present (Table 4). However, we feel that such a focus on bloom detection being a yes/no situation misses key information contained in prediction standard errors as well as key information that can be obtained by viewing a prediction surface. Predictions with low standard errors are likely to be well determined, hence the clear bloom/non bloom events in Figures 7 and 8. Equally important, spatial and temporal predictions clearly illustrate the development, movement and persistence of large clusters of high probability conditions which could be identified as a bloom. Hence the probabilistic models give analysts an effective tool for bloom detection.

The differences in predictive capacity can be seen in the comparison of the two (*A. Offshore*) and (*B. All area*) where the offshore model explains more deviance and has a lower overall prediction error. Furthermore there is less contrast in the inshore indices likely due to the relatively high baseline predictions due to contamination with other sources of chlorophyll or suspended particles. Hence we would not recommend using the: *3. Inshore*, *4. Inshore (Aug-Dec)* of *5. All areas* indices for inclusion in the SEDAR 33 gag grouper models as they are likely to give poor predictions and be more affected by the diminished ability of the satellite data to predict in shallow waters. However, the remaining indices are all potentially valid indicators of the intensity of red tides in Eastern Gulf of Mexico waters > 10 m deep.

The strong predictive capacity of the models for offshore waters make these models quite valuable for determining potential effects upon grouper populations for which adults are generally found in the areas defined by polygon for the grouper index (Figure 4). Furthermore, since red tide blooms often start in offshore waters and then advect inshore (Steidinger & Vargo 1988), the offshore predictions may correlate with high inshore blooms and potential human health hazards but with a time lag. It should also be noted that the blooms can also start at depth and occur prior to their detection in surface waters (Steidinger & Vargo 1988), a point that we discus below. But, in most cases when a bloom is severe enough it eventually reaches surface waters and is detected in water samples and, presumably, in the satellite predictions (Steidinger *pers comm*).

A minor limitation of the current derived indices is missing data early in the time series as the satellite was just coming online (2008) and later in the time series due to sensor failure (2008, 2009, 2010). In some situations, the missing data is simply because a single calculated variable was not available so the overall model could not be estimated. In the future it may be possible to evaluate whether a simpler model which would not need that particular derived product may give similar predictions. For many sets of daily due to satellite observations, there were substantial gaps due to clouds. For the predictions we did not average the raw satellite-derived information; when no data was available for a spatial cell we simply had no predictions. For the purposes of this paper we assume that missing data is missing at random and does not affect the overall trends. This assumption may not hold for 2008 and one might consider removing 2008 predictions from the time series as only July and August had all complete information for all derived satellite data products.

These models may also be valuable for incorporating into HAB bulletins that provide notification of HAB conditions. Since satellite data collection is continuous and covers the entire spatial domain, except for clouds, it provides a valuable spatial and temporal prediction. Given the quantified propensity (~20%) of these models to give false positive indications it is likely that both the standard error of any predictions and the persistence of the prediction in space and time be used for making any determination of bloom status. But from the perspective of assessing HAB conditions, the low probability of false negatives (~6-19%) is probably quite useful as the model does fairly well at determining when a bloom is not occurring. And, as satellite-remote sensing will never supplant the need for ground-truthed water samples, model predictions of high probability areas can be used for targeted water quality sampling to determining the true presence of a bloom.

When considering the potential harmful effects of red tides, there is considerable historical information in the published literature for groupers (Ingersoll 1882, Gunter 1947, Smith 1971) and for other fauna (Landsberg & Steidinger 1998; Flewelling et al. 2005; Landsberg 2002) as well as substantial anecdotal evidence. One of particular importance to note is taken from the field log for the NMFS bottom longline station log, August 24, 2005:

"4 or 5 dolphins (frontalis) close to ship at haul. During steam for 3 or so hours broad area of dead fish floating at surface, what looked like a 50 lb warsaw grouper spotted at haulback end."

The cruise track and the station location is plotted on the red tide prediction for that month which clearly shows the vessel steaming through an area of high probability of red tide. Subsequent predictions for September-November indicate that the red tide event was large, sustained and if grouper mortalities occurred in August, were most likely to have continued for several weeks. It is also important to note that the mortalities appeared to have occurred prior to the development of the large surface bloom and it may be that grouper mortality precedes the surface blooms perhaps due to the tendency of the blooms to start at depth (Steidinger et al 2008). Note, however, that no *gag* grouper were reported by scientists on the longline vessel and that many of the anecdotally reported mortalities of grouper were of red grouper.

In conclusion, these predictions provide a suite of indices for consideration of the effects of red tide on grouper mortality. High values in year 2005 match numerous observations of the severity of red tide and its effect upon groupers in that year (Figure 10). We recommend that indices *1 > 10m MCP75%* , *6 Grouper* or *7. Threshold* be considered as candidates for inclusion in the assessments. Further work by the authors on this paper will test the 7 indices in the 2009 update assessment models to determine which model provides the best fit to the model, which might be used to indicate which index construction best predicts grouper mortality. This work will likely be available as a later data workshop paper.

References

Amin, R., Jing Zhou, Alex Gilerson, Barry Gross, Fred Moshary, and Samir Ahmed. 2009. Novel optical techniques for detecting and classifying toxic dinoflagellate *Karenia brevis* blooms using satellite imagery. OPTICS EXPRESS: Vol. 17, No. 11 .

Anon, 2009a. http://www.gulfcouncil.org/Beta/GMFMCWeb/downloads/SEDAR%20GAG%20 AND%20RED%202009/Red_Grouper_2009_Assessment_Update_Report-FINAL.pdf

Anon, 2009b.http://www.gulfcouncil.org/Beta/GMFMCWeb/downloads/SEDAR%20GAG%20 AND%20RED%202009/Gag_2009_Assessment_Update_Report-FINAL.pdf

Cannizzaro, J. P., Carder, K. L., Chen, F. R., Heil, C. A., & Vargo, G. A. 2008. A novel technique for detection of the toxic dinoflagellate *Karenia brevis* in the Gulf of Mexico from remotely sensed ocean color data. Continental Shelf Research, 28, 137−158.

Cannizzaro, J.P., C. Hu. D.C. English , K. L. Carder, C. A. Heil, & F. E. Muller-Karger. 2008. Detection of *Karenia brevis* blooms on the west Florida shelf using in situ backscattering and fluorescence data. Harmful Algae 8: 898–909.

Carder, K. L., Chen, F. R., Lee, Z. P., Hawes, S. K., & Kamykowski, D. 1999. Semi-analytic Moderate-Resolution Imaging Spectrometer algorithms for chlorophyll a and absorption with bio-optical domains based on nitrate-depletion temperatures. Journal of Geophysical Research, 104, 5403−5422.

Carder, K. L., & Steward, R. G. 1985. A remote-sensing reflectance model of a red-tide dinoflagellate off West Florida. Limnology and Oceanography, 30(2), 286−298.

Carvahlo, G. A. P. J. Minnett V. F. Banzon W Baringer, C.A. Heil 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 Sensing of Environment 115 1–18.

Deriso, R., M. Maunder & W. H. Pearson. 2008. Incorporating covariates into fisheries stock assessment models with application to Pacific Herring. Ecological Applications, 18(5):1270–1286.

Flewelling, L.J., Naar, J.P., Abbott, J.P., Baden, D.G., Barros, N.B., Bossart, G.D. 2005. Brevetoxicosis: red tides and marine mammal mortalities. Nature 435, 755–756.

FWRI, 2008. Florida Fish and Wildlife Research Institute, http://research.myfwc.com.

Gunter, G. F. G. W. Smith & R. H. Williams. 1947. Mass mortality of animals on the lower west coast of Florida, November 1946- January 1947. Science. 105:256-257.

Hastie, T. J. and Tibshirani, R. J. 1990. *Generalized Additive Models*. Chapman & Hall/CRC

Hothorn T, Hornik K, Zeileis A 2006. “Unbiased Recursive Partitioning: A Conditional Inference Framework.” *Journal of Computational and Graphical Statistics*.

Hu, C., F. E. Muller-Karger, C. Taylor, K. L. Carder, C. Kelble, E. Johns, and C. A. Heil, 2005. Red tide detection and tracing using MODIS fluorescence data: A regional example in SW Florida coastal waters. Remote Sens. Environ. **97**(3), 311–321.

Ingersoll, E. 1882. On the fish mortality in the Gulf of Mexico. Proc. US. Nat. Mus., 4:74-80.

Landsberg, J.H., Steidinger, K.A., 1998. A historical review of *Gymnodinium breve* red tides implicated in mass mortalities of the manatee (*Trichechus manatus Latrirostris*) in Florida, USA. In: Reguera, B., Blanco, J., Fernandez, M.L., Wyatt, T. (Eds.), Harmful Algae. Intergovernmental Oceanographic Commission of UNESCO, pp. 97–100.

Landsberg, J. 2002. The effects of harmful algal blooms on aquatic organisms. Reviews in Fisheries Science. 10(2): 113-390.

Morel, A. 1988. Optical modeling of the upper ocean in relation to its biogenous matter content (case I waters). Journal of Geophysical Research, 93, 10,749−10,768.

Rounsefell, G.A., Nelson, W.R. 1966. Red-tide research summarized to 1964 including an annotated bibliography. U.S. Fish Wildl. Serv. Spec. Sci. Rep. Fish. 535, 1–85.

SEDAR 2012. Southeast Data, Assessment, and Review SEDAR Episodic Events Workshop for Gulf of Mexico Fisheries: Workshop Summary Report. SEDAR, North Charleston, SC, USA. 26 pp.

Smith, G.B. 1975. The 1971 red tide and its impact on certain reef communities in the mid-eastern Gulf of Mexico. *Environmental Letters* 9(2):141-152.

Steidinger, K.A. & Ingle, R.M. 1972. Observations on the 1971 red tide in Tampa Bay, Florida. *Environmental Letters* 3:271-278.

Steidinger, K.A., & Joyce, E.A., 1973. Florida red tides. Educ. Ser. No. 17. Florida Dept. Nat. Res., St. Petersburg, 26 pp.

Steidinger, K.A., Vargo, G.A., 1988. Marine dinoflagellate blooms: dynamics and impacts. In: Lembi, C.,Waaland, J.R. (Eds.), Algae and Human Affairs. Cambridge University Press, pp. 373– 401.

Steidinger, K. A., Vargo, G. A., Tester, P. A., & Tomas, C. R. 1998. Bloom dynamics and physiology of Gymnodinium breve with emphasis on the Gulf of Mexico. In D. M. Anderson, A. D. Cambella, & G.M. Hallegraeff (Eds.), Physiological Ecology of Harmful Algal Blooms (pp. 135−153). New York: Springer-Varleg.

Stumpf, R. P. 2001. Applications of satellite ocean color sensors for monitoring and predicting harmful algal blooms. Human and Ecological Risk Assessment, 7(5), 1363−1368.

Stumpf, R. P., Arnone, R. A., Gould, R.W.,Martinolich, P., Ransibrahmanakul, V., Tester, P. A., et al. 2000. SeaWiFS ocean color data for US Southeast coastal waters. Proceedings of the 6th International Conference on Remote Sensing forMarine and Coastal Environments, 1-3 May (pp. 25−27). Ann Arbor, MI, USA: Charleston-SC, Veridian ERIM Intl.

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

Strobl, C., T. Hothorn and A. Zeileis 2009. A New, Conditional Variable-Importance Measure for Random Forests Available in the party Package. The R Journal. Vol. 1/2. 14-17.

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

Tomlinson, M. C., T.T. Wynne and R. P. Stumpf. 2008. An evaluation of remote sensing techniques for enhanced detection of the toxic dinoflagellate, *Karenia brevis* Remote Sensing of the Environment 113: 598–609.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Ransibrahmanakul, V., Villareal, T.A., 2005. Detecting Karenia brevis blooms and algal resuspension in the western Gulf of Mexico with satellite ocean color imagery. Harmful Algae 4, 992–1006.

Wynne, T. T., Stumpf, R. P., Tomlinson, M. C., Warner, R. A., Tester, P. A., Dyble, J., et al. 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. International Journal of Remote Sensing, 29, 3665−3672.

Wood, S. N. 2006. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC.

Young, M and M. Christman. 2010. http://myfwc.com/research/redtide/archive/historical-database/evaluation/0

Table 1. Summary statistics of the potential predictor variables using in the modeling of HAB in the Gulf of Mexico. The χ^2 statistic is the test statistic from a logistic regression of presence- absence of HAB on the predictor variable.

Predictor	Mean (SEM)	Median	Min - Max	χ^2 statistic[1]
Chlorophyll Anomaly	40.67 (0.662)	1	1 – 249	563.62**
Chlorophyll	120.54 (0.588)	136	0 – 250	293.53**
Carder	55.80 (0.688)	1	1 – 211	43.85**
Morel	142.67 (0.646)	168	0 – 208	82.33**
CMbbp	137.99 (1.170)	172	0 – 249	25.67**
rrs670	76.53 (0.584)	80	1 – 210	7.40*
HAB Ensemble	13.99 (0.485)	1	0 – 249	106.73**
ssnLw490	126.68 (0.981)	133	0 – 249	268.07**
Easting	344204.9 (827.11)	342750	20450 – 497850	13.69*
Northing	2967297.1 (1535.41)	2975091	2709991 – 3352391	40.22**
Water Depth	13.87 (0.178)	10	0 – 255	6.70*
Sea Surface Temperature	24.63 (0.044)	24.53	11.78 – 31.74	47.30**
Week of Year	23.61 (0.156)	22	1 – 53	134.40**
Month	na	na	Jan-Dec.	554.77**

[1] ** = p< 0.0001; * = p < 0.01

Table 2. Significance table for the final GAM model (A. all area) for predicting HABs. The p-value of the test of significance is approximate.

Predictor Variable	Estimated Df	χ^2 statistic	p-value
as.factor(Month)	11	na	na
te(Chlorophyll Anomaly, Week Of Year)	7.058	31.49	<0.00001
Te(Chlorophyll a, Week Of Year)	14.363	68.46	<0.00001
te(Carder, Morel)	6.406	90.21	<0.00001
te(Easting, Northing)	13.096	157.71	<0.00001
te(Sea Surface Temperature, Water Depth)	7.065	48.62	<0.00001
te(RRS670,Week Of Year)	11.255	59.54	<0.00001
Te(HAB Ensemble, Week Of Year)	7.331	24.05	<0.00001
Te(Water Depth)	2.767	27.27	<0.00001

Table 3. Significance table for the final GAM model (B. Offshore >10m) restricted to only waters greater than 10 meters. The p-value of the test of significance is approximate.

Predictor Variable	**Estimated Df**	**χ^2 statistic**	**p-value**
as.factor(Month)	11	na	na
te(Chlorophyll Anomaly, Week Of Year)	5.086	40.619	<0.00001
Te(Chlorophyll a, Week Of Year)	10.720	54.537	<0.00001
te(Carder, Morel)	6.181	44.176	<0.00001
te(Easting, Northing)	5.508	34.372	<0.00001
te(Sea Surface Temperature, Water Depth)	5.275	20.594	<0.00001
te(RRS670,Week Of Year)	3.316	16.965	<0.00001
Te(HAB Ensemble, Week Of Year)	1.168	4.701	0.0167
Te(Water Depth)	0.000	0	0.8902

Table 4. Performance criteria for final models

	A. all area model	B. > 10 meter water depth model
N observations	9050	4294
Adjusted R^2	0.304	0.390
Deviance explained	36.6%	44.2%
ROC threshold	0.0827	0.0541
Area under ROC	0.9046	0.9295
Overall misclassification rate	17.4% (19.7%*)	20.89% (22.1%*)
False Positive rate	19.36% (19.8%*)	22.14% (23.1%*)
False negative rate	15.38% (19.3%*)	6.23% (10.4%*)

* cross-validation results

Table 5. Indices and index standard errors

	1. 10m_MCP75		2. 10m_MCP75 Aug_Dec		3. Inshore		4. Inshore Aug_Dec		5. All prediction areas		6. Grouper index		Threshold model
Model	B.offshore		B.offshore		A. All area		A. All area		A. All area		B.offshore		B.offshore
year	mean	se	mean	se	mean	se	mean	se	mean	se	mean	se	value
1998	0.016	0.000	0.018	0.001	0.007	0.006	0.017	0.006	0.006	0.011	0.012	0.008	0
1999	0.050	0.002	0.119	0.004	0.035	0.010	0.035	0.005	0.030	0.010	0.034	0.007	0
2000	0.038	0.001	0.099	0.003	0.025	0.008	0.038	0.005	0.024	0.009	0.026	0.006	0
2001	0.056	0.002	0.150	0.006	0.030	0.010	0.028	0.004	0.027	0.010	0.032	0.006	0
2002	0.061	0.002	0.127	0.004	0.044	0.010	0.079	0.008	0.038	0.011	0.044	0.007	0
2003	0.059	0.002	0.108	0.003	0.060	0.015	0.057	0.006	0.041	0.011	0.045	0.008	0
2004	0.049	0.001	0.117	0.005	0.031	0.009	0.034	0.004	0.028	0.010	0.033	0.006	0
2005	0.148	0.005	0.287	0.009	0.068	0.014	0.079	0.008	0.056	0.012	0.087	0.010	1
2006	0.058	0.002	0.144	0.004	0.037	0.011	0.055	0.006	0.032	0.011	0.037	0.006	0
2007	0.045	0.002	0.089	0.002	0.031	0.009	0.041	0.006	0.027	0.010	0.033	0.007	0
2008	0.057	0.002	0.109	0.004	0.034	0.011	0.023	0.005	0.030	0.010	0.043	0.010	0
2009	0.028	0.001	0.039	0.002	0.038	0.016	0.029	0.004	0.029	0.017	0.024	0.006	0
2010	0.058	0.002	0.115	0.004	0.043	0.014	0.043	0.008	0.032	0.002	0.047	0.011	0

Figure 1. Plot of all red tide water monitoring data (green points) for 1998-2010 and the spatial domain for satellite imagery. Blue represents cloud cover and shading represents satellite- derived chlorophyll.

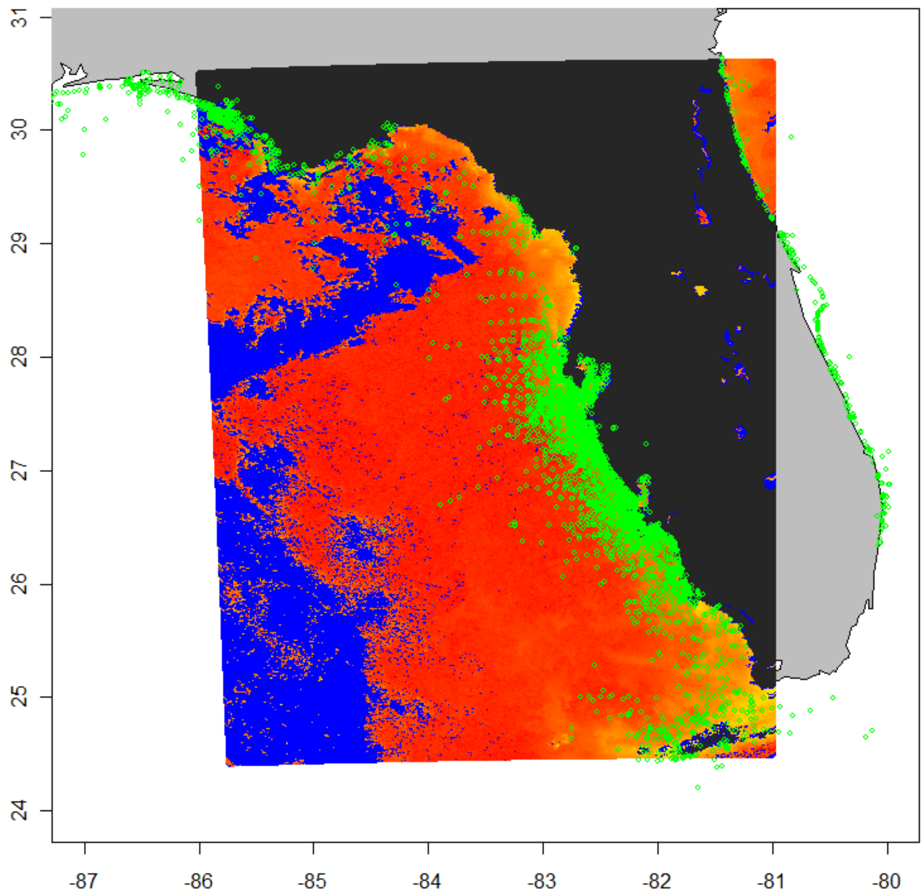


Figure 2. HAB data and Chlorophyll anomaly for 01/27/2000. Points are HAB stations for all of January 2000. Blue color is clouds.

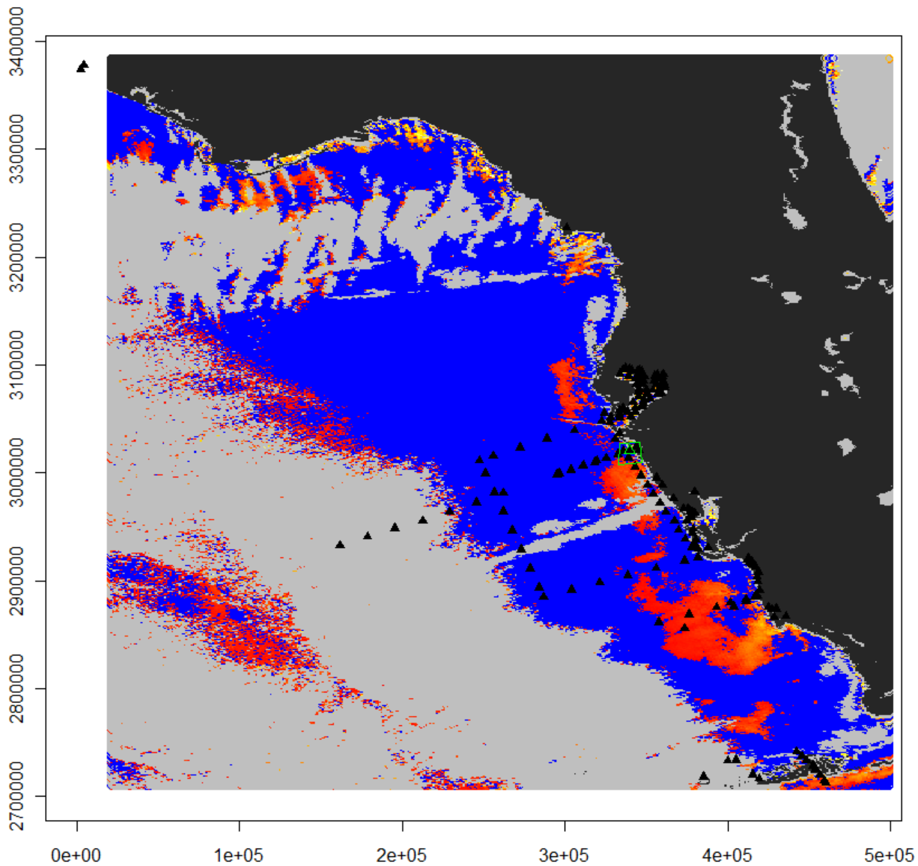


Figure 3. Inset of green polygon in Figure 2. Blue color is clouds. Units are meters. This demonstrates how a satellite measurement is assigned to a HAB data point. The satellite data is assigned to a set of grid nodes that define a cell. The SAT data for the lower left grid node is assigned for to the HAB value falling in the cell.

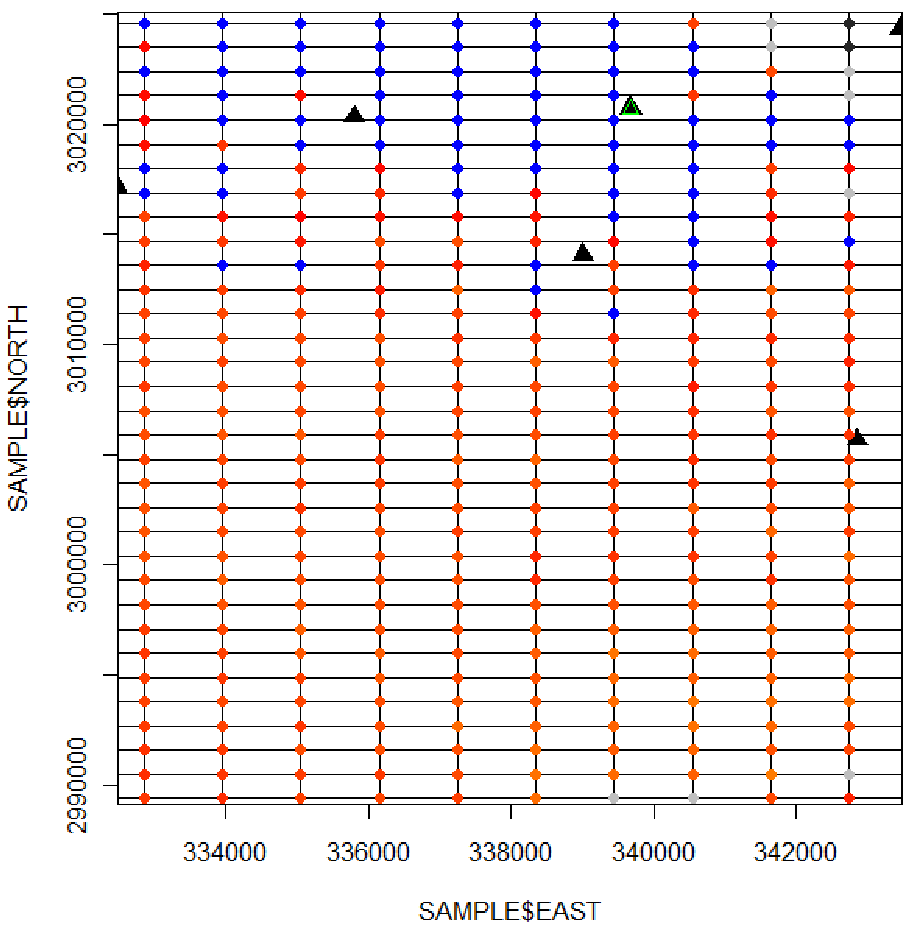


Figure 4. Spatial domains of the different index constructions.

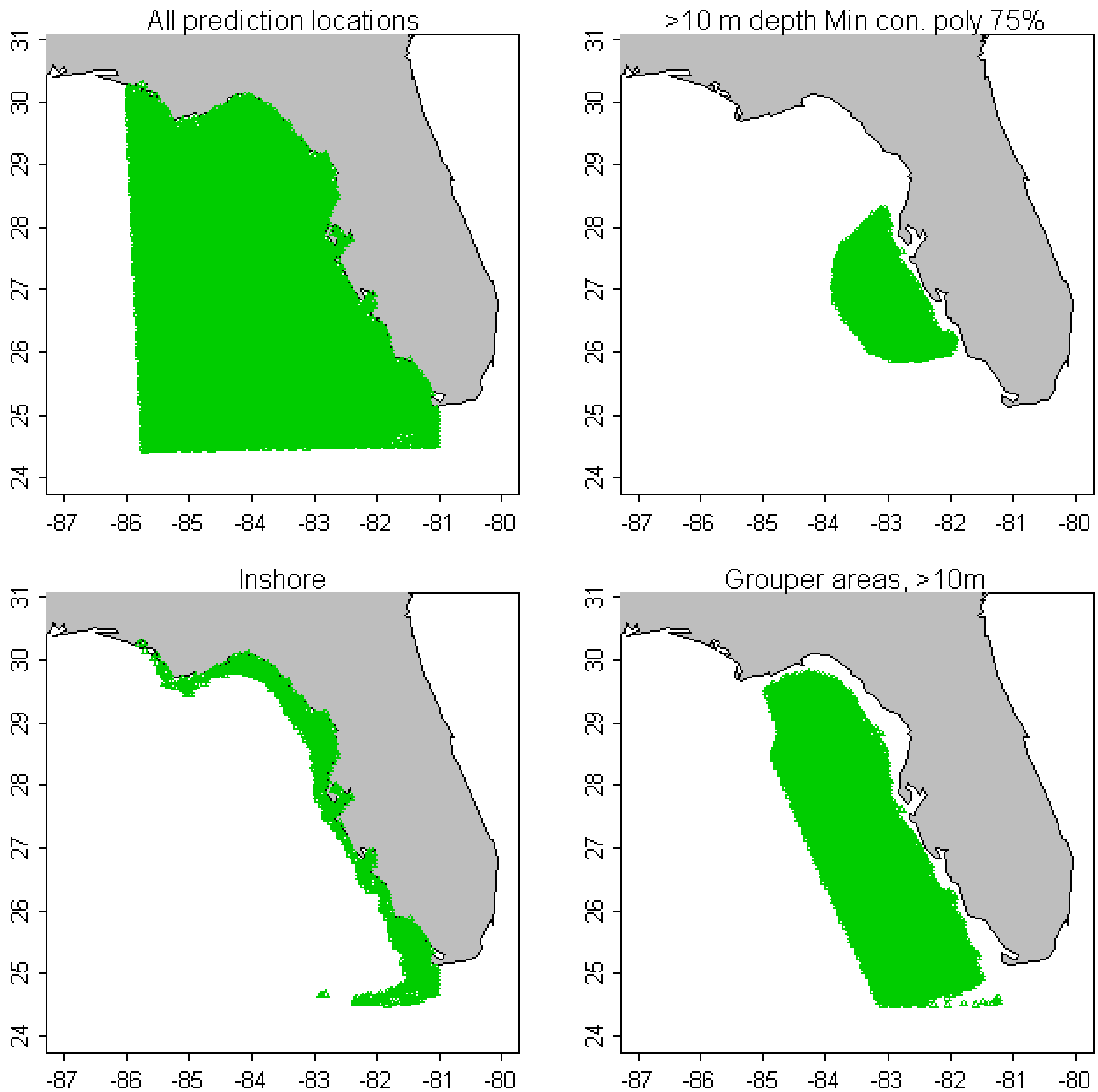
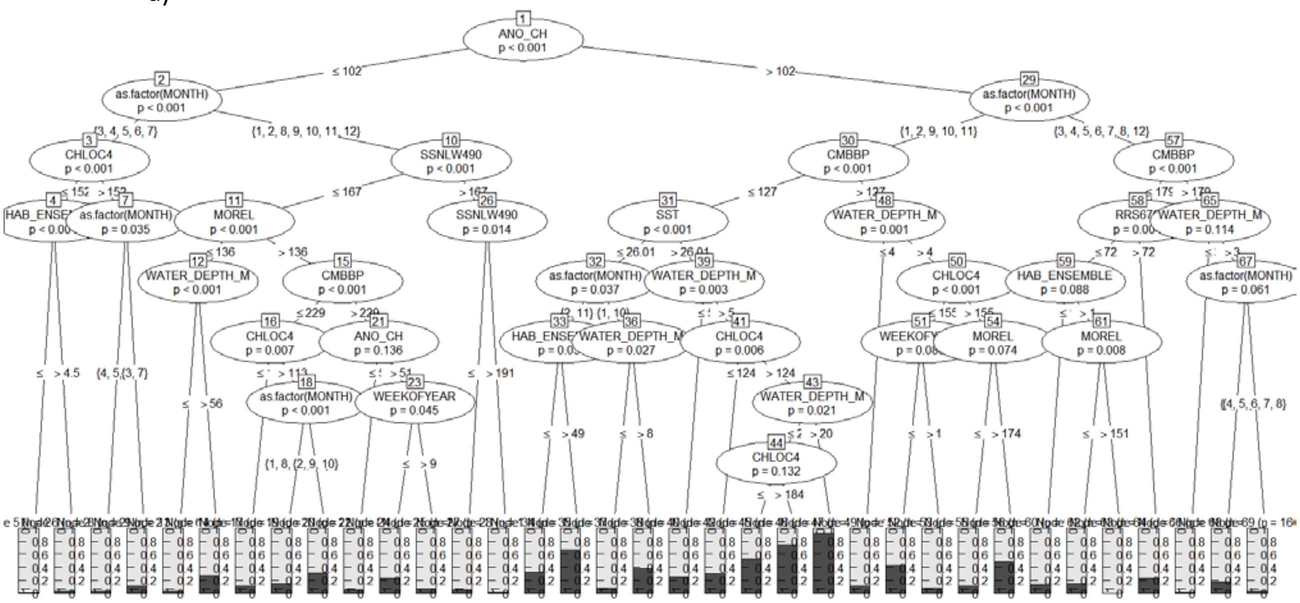


Figure 5. a) Classification tree for estimating the probability of a HAB presence. The p-value associated with the test of the best variable selected at each split is shown in each node with the name of the variable. b) bar chart showing variable importance based on a random forest.

a)



b)

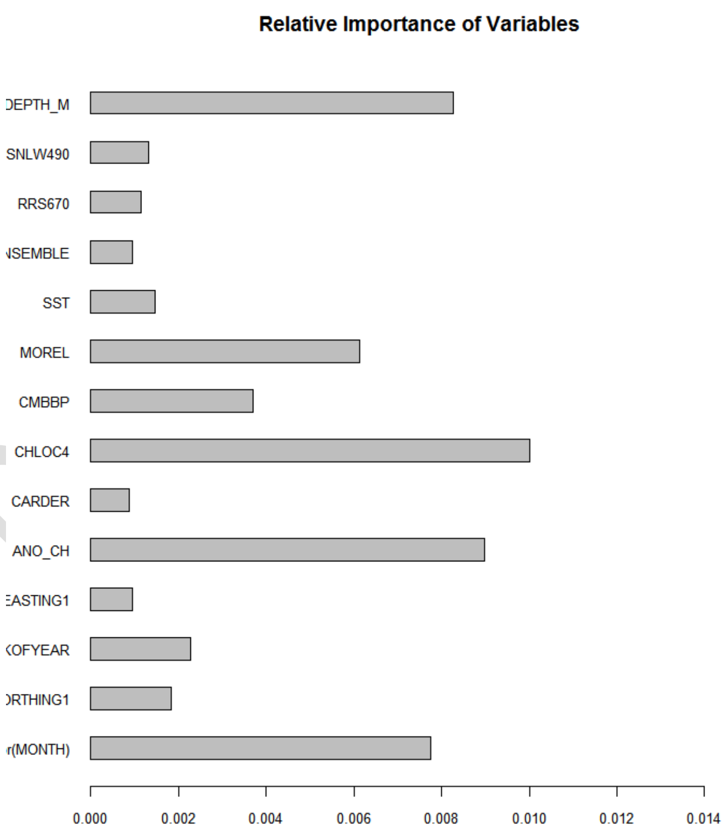


Figure 6. Plots of the smoothed components of best fitting GAM for predicting occurrence of a HAB.

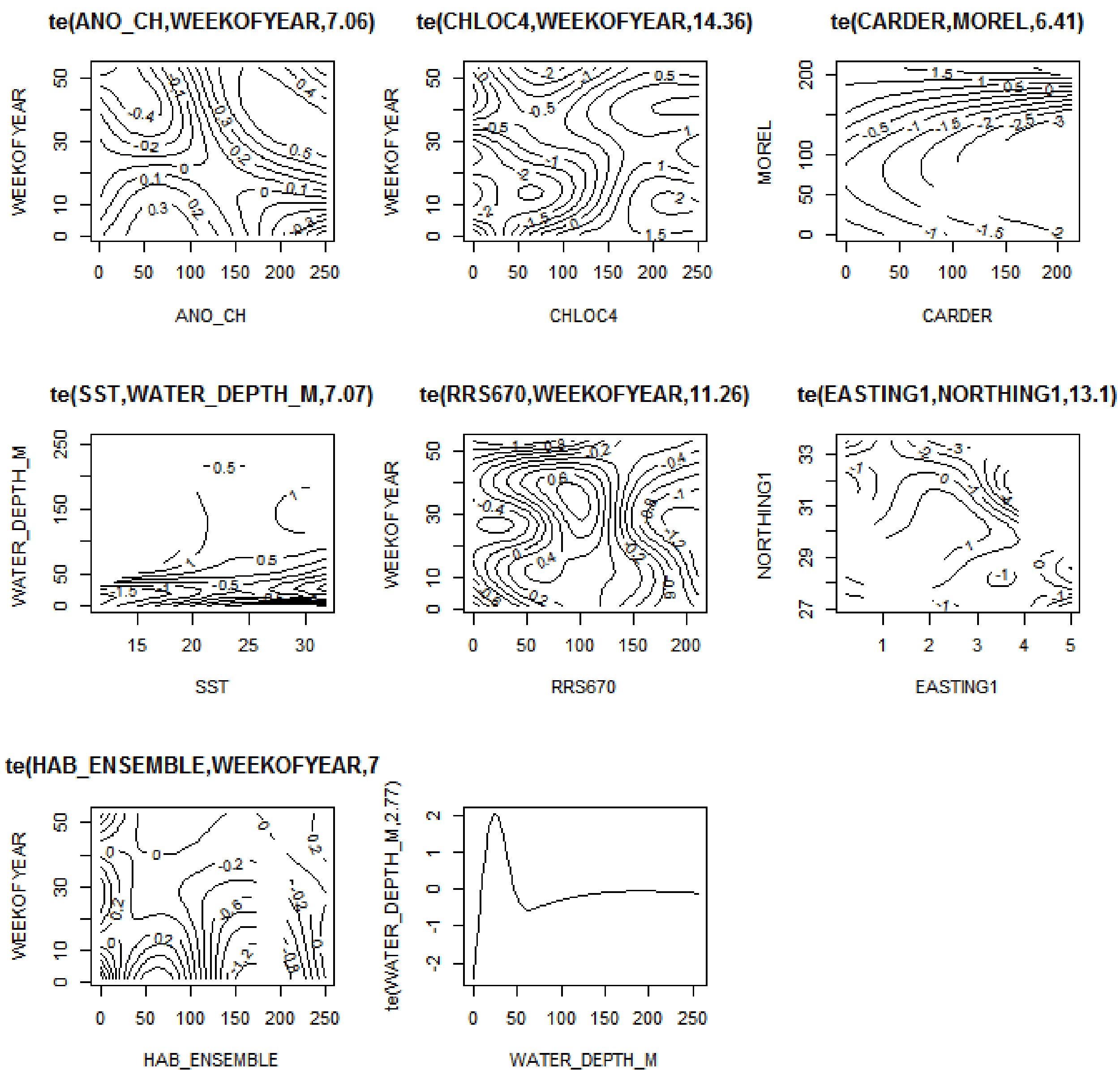


Figure 7. Monthly average predicted presence and average standard errors of prediction of algal concentrations greater than 100,000 (yellow) for October-December 2005 during a documented red tide bloom.

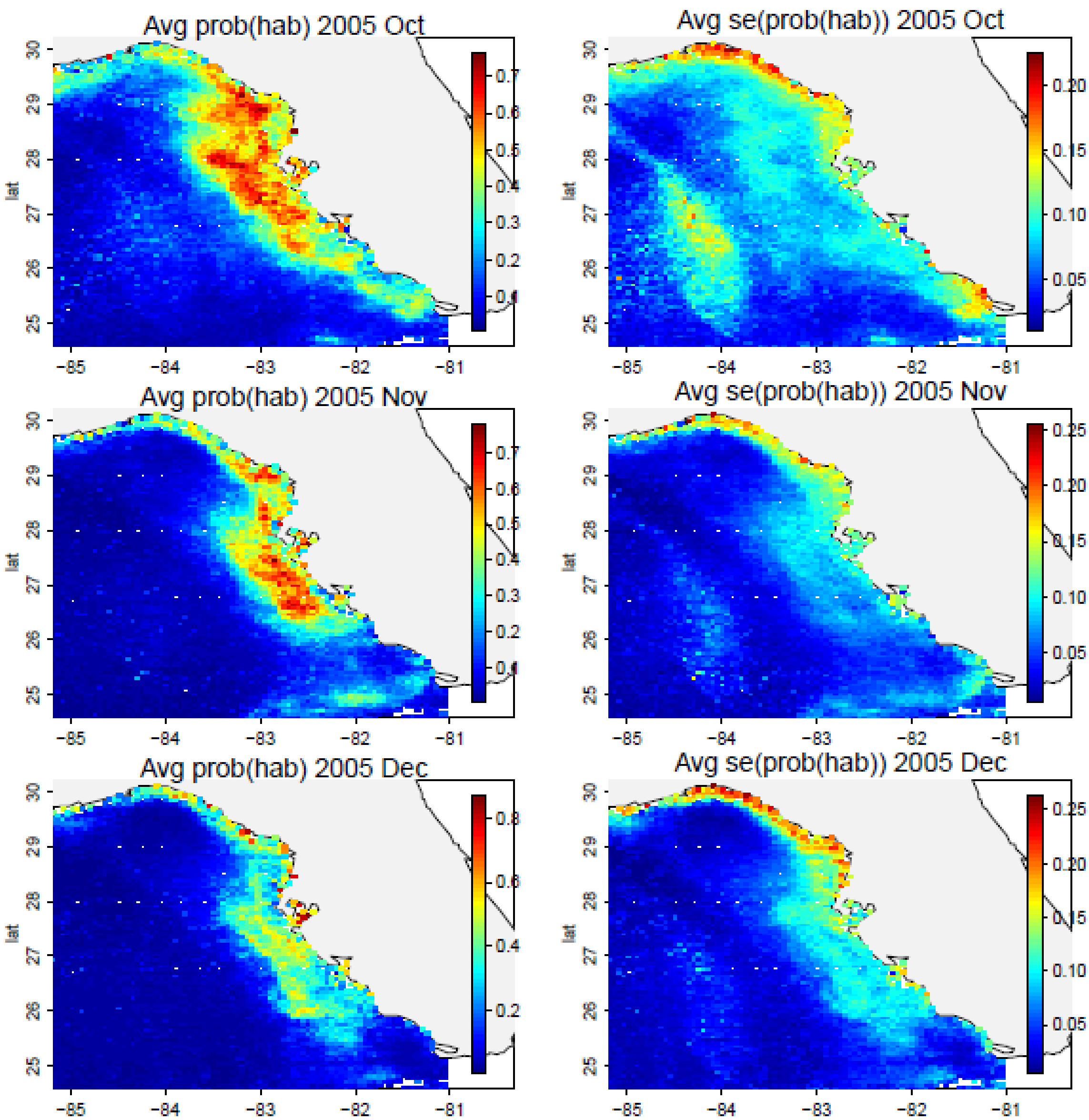


Figure 8. Monthly average predicted presence and average standard errors of prediction of algal concentrations greater than 100,000 (yellow) for 2010 during a time period of no documented red tide bloom.

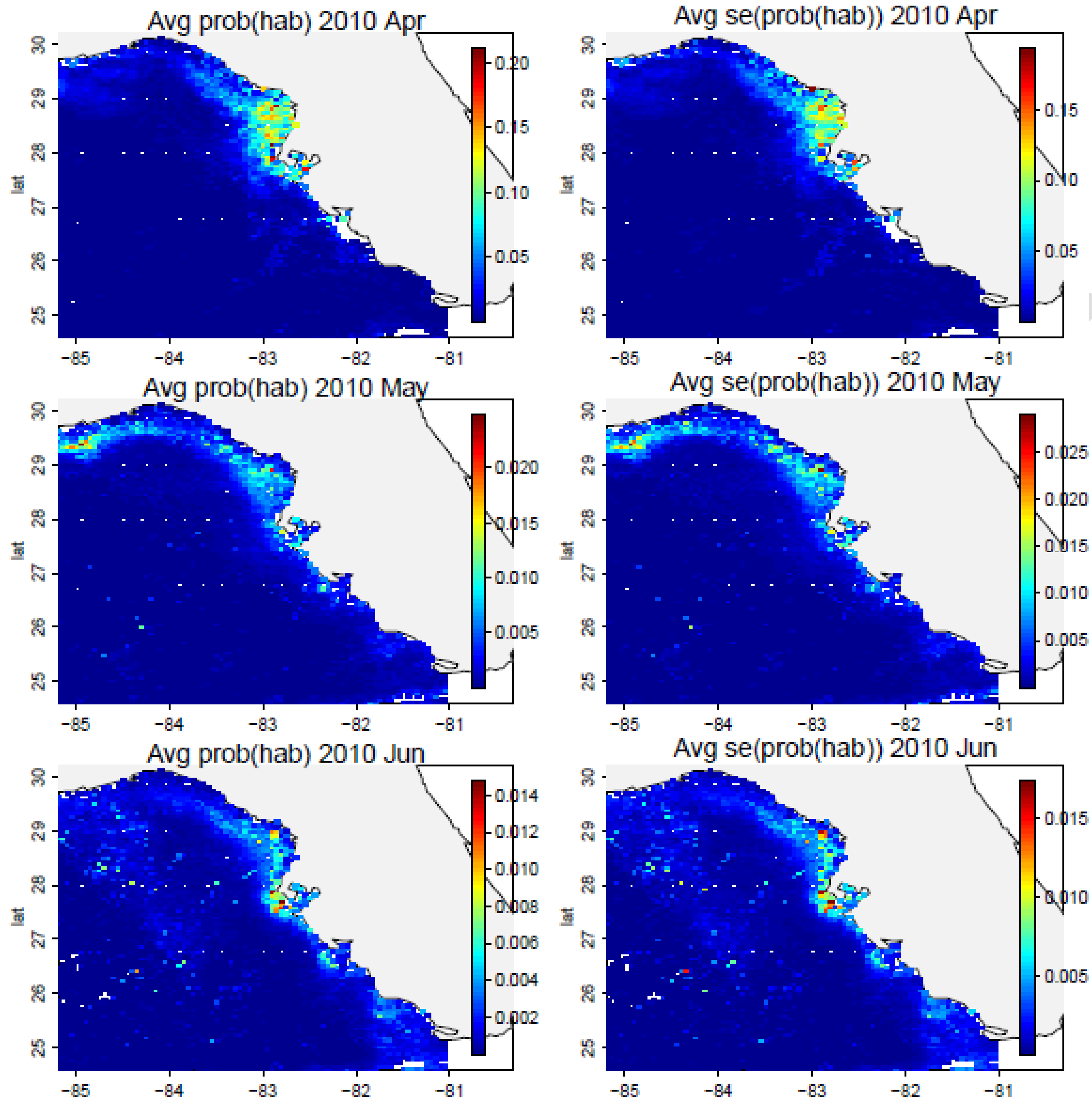


Figure 9. Indices and index standard errors.

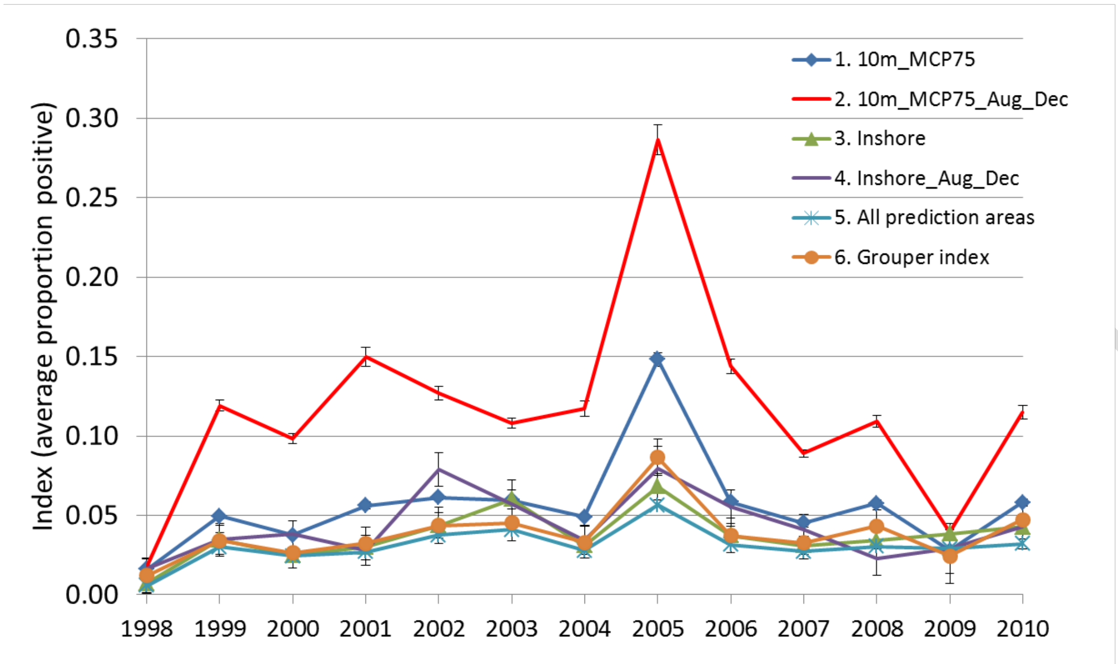
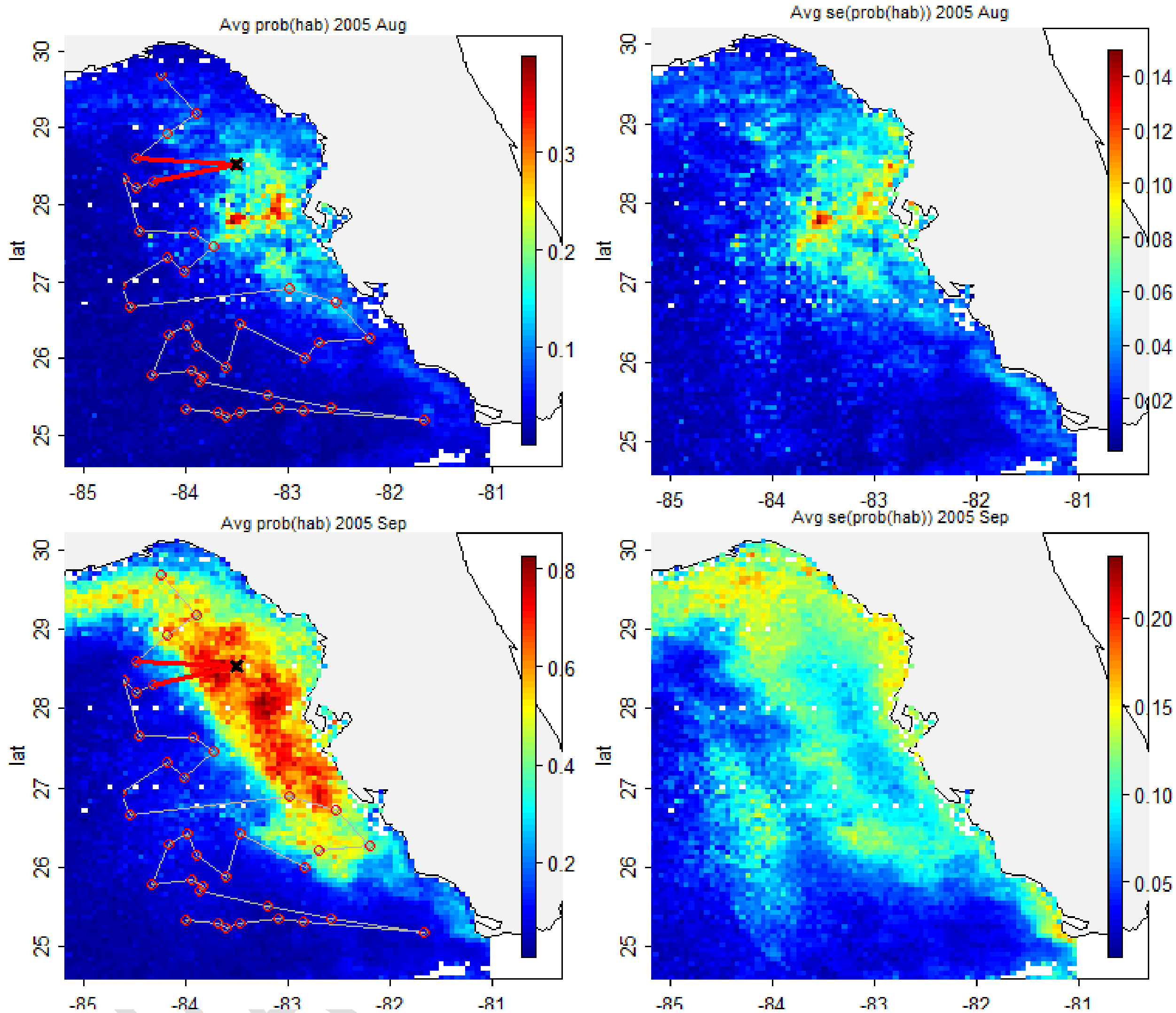
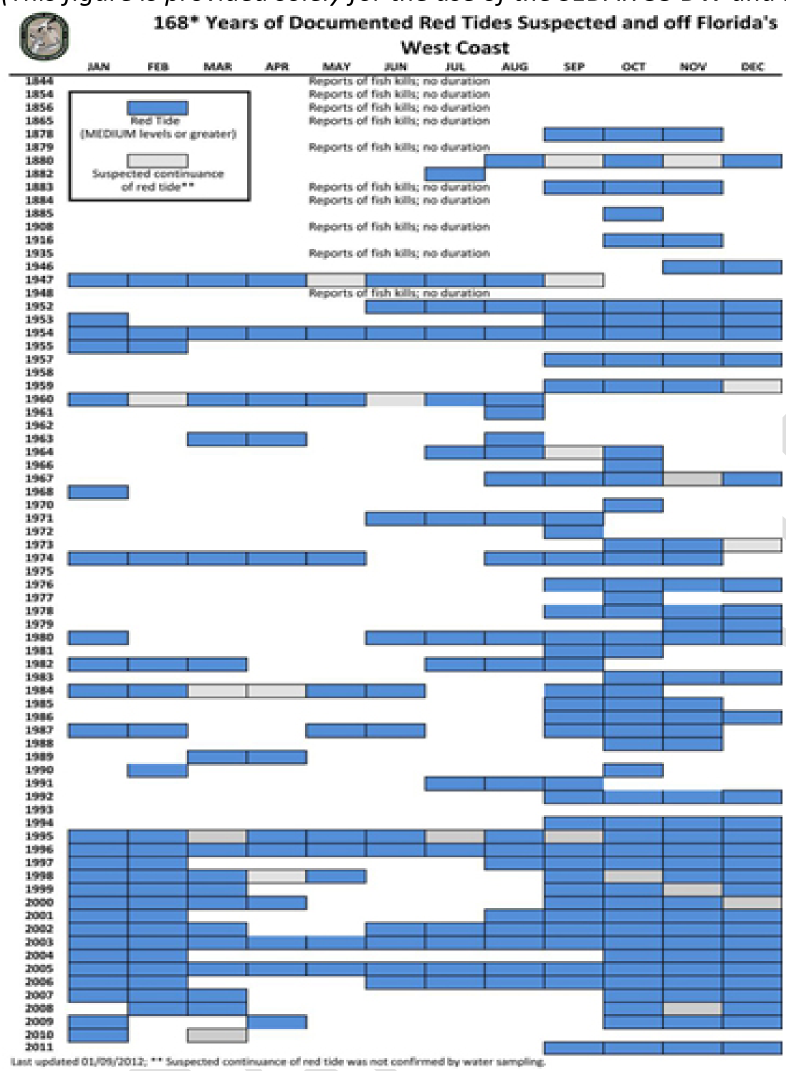


Figure 10. NMFS bottom longline cruise track for August 2005 overlaid on predicted red tide. Cruise track in bold red and station noted with ‘x’ is referenced to the August 24, 2005 station log, quoted above.

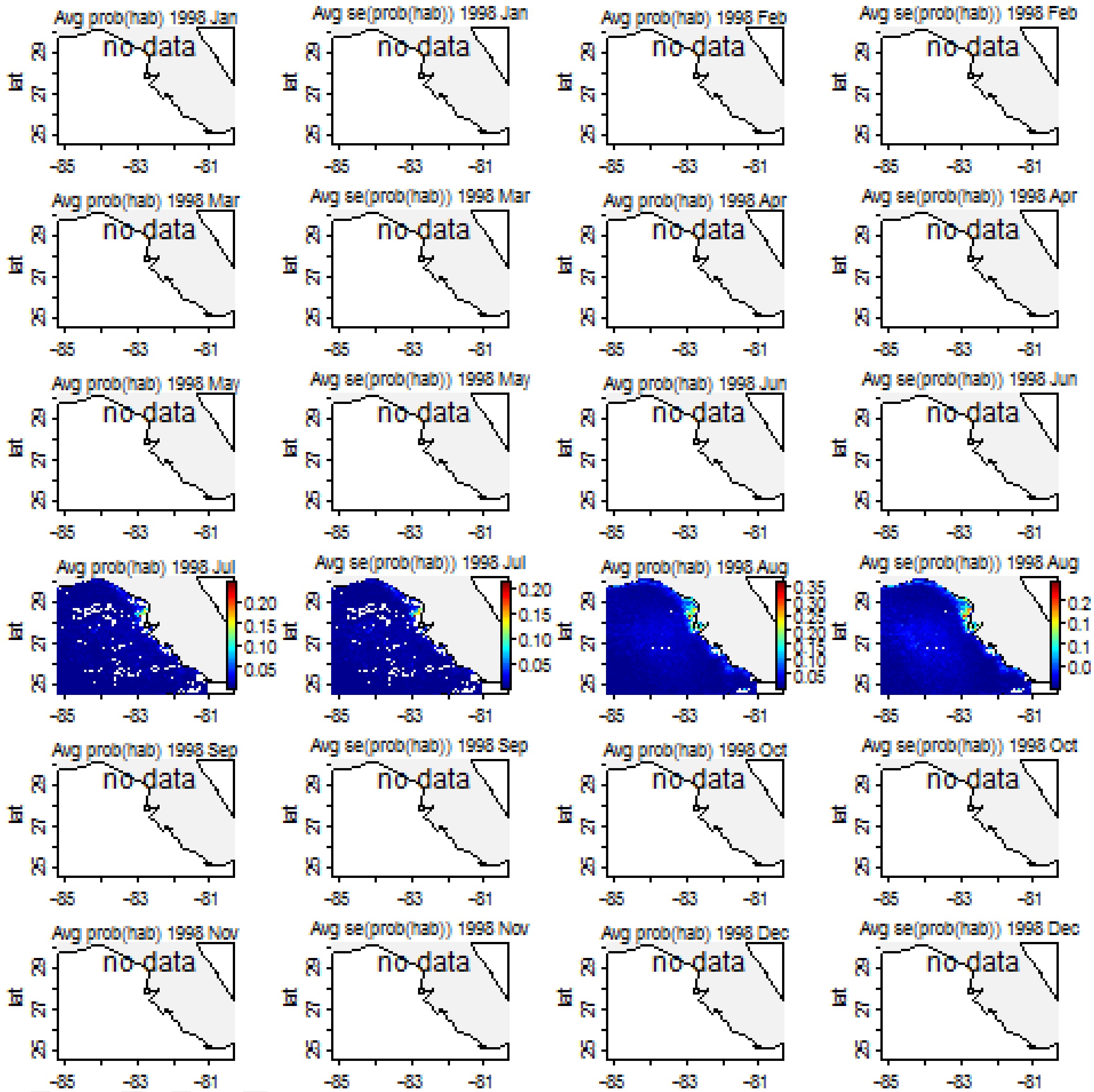


Appendices

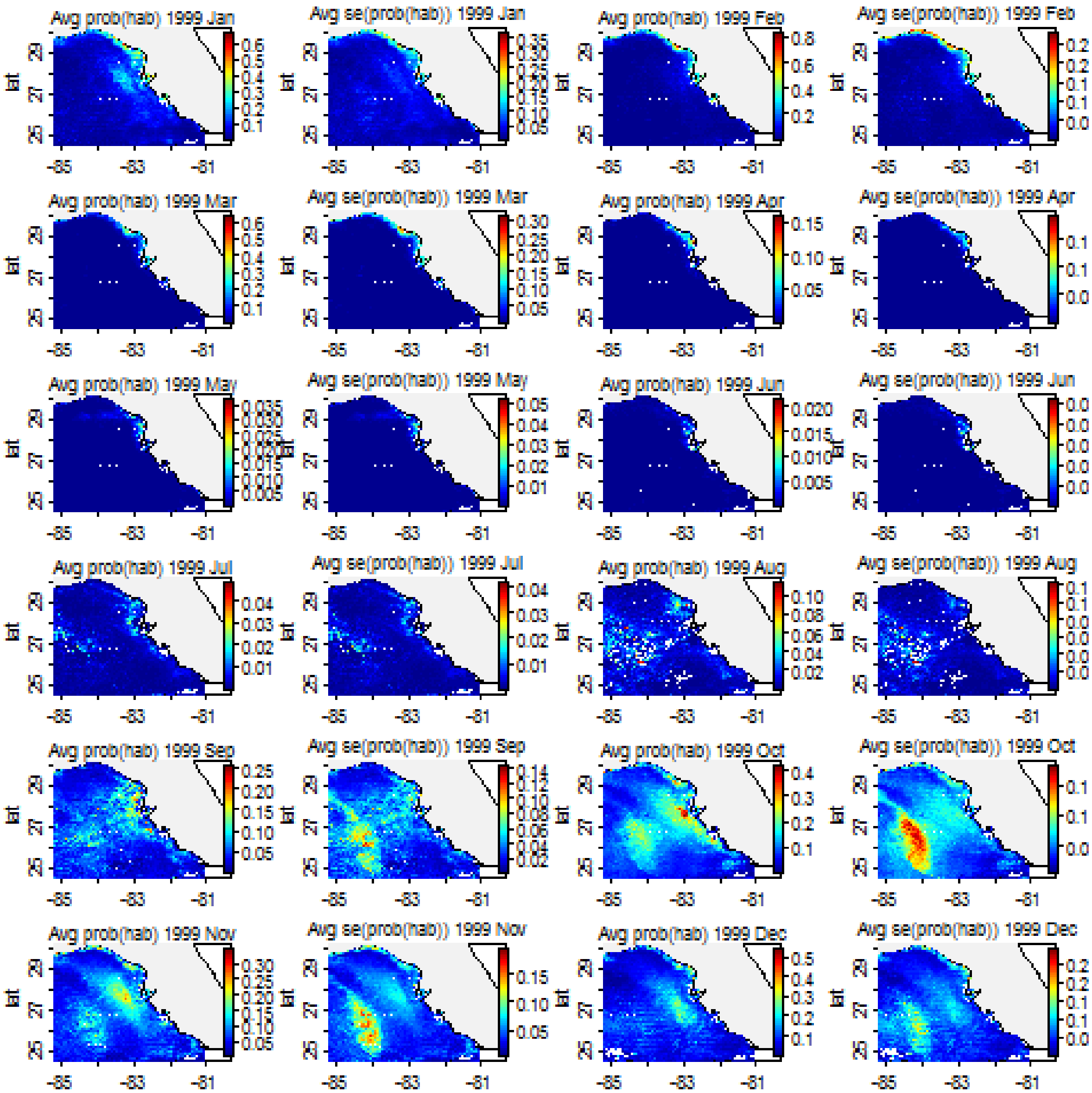
Appendix Figure 1. Distribution of recorded harmful algal bloom events off the west coast of Florida.
From: http://myfwc.com/research/redtide/archive/historical-database/timeline-red-tides-fl-w-coast/
(This figure is provided solely for the use of the SEDAR 33 DW and may not be part of a final manuscript)



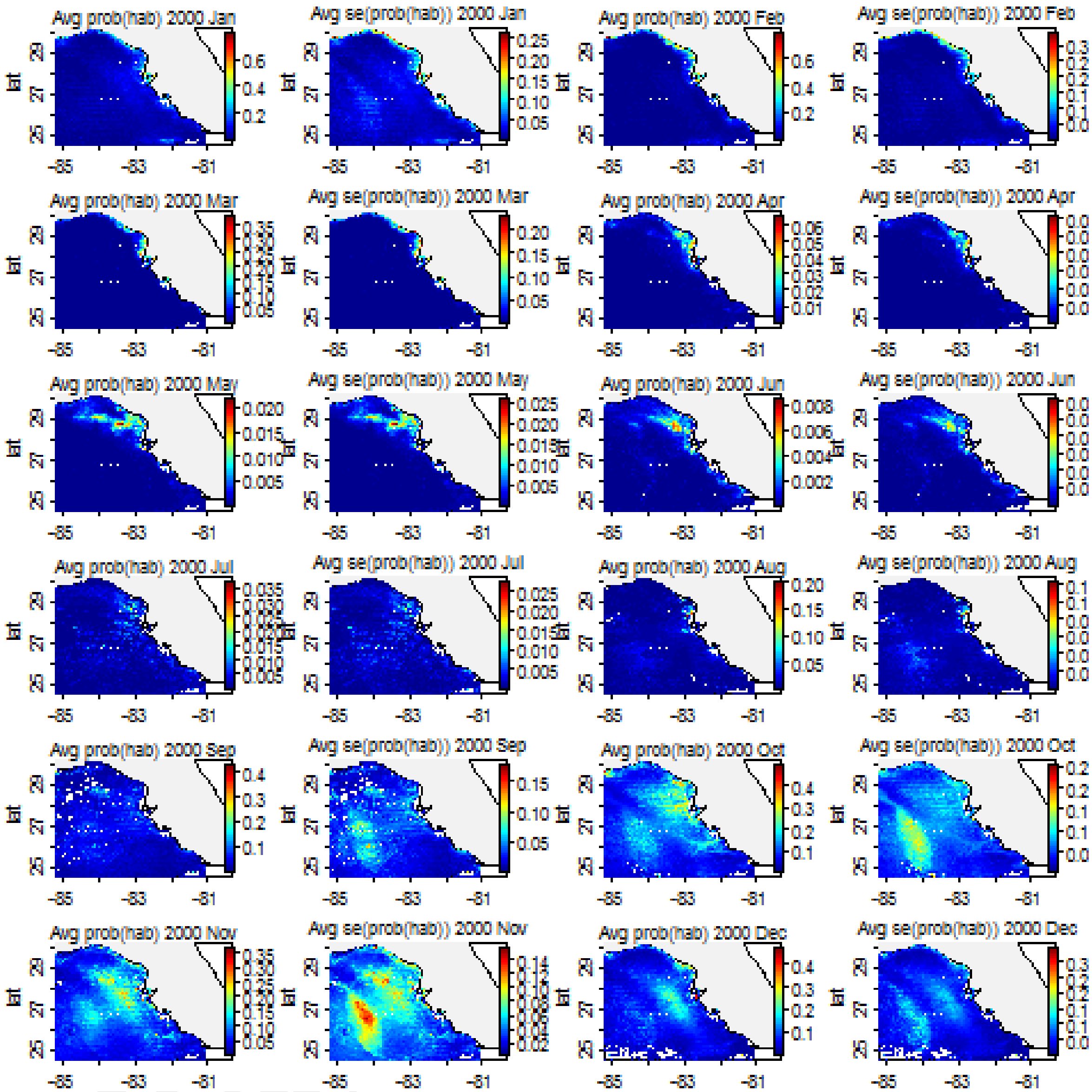
Appendix Figures 2a-k. Monthly predictions of prob(red tide).
a. Monthly plots for 1998, no sat data means incomplete sat data.



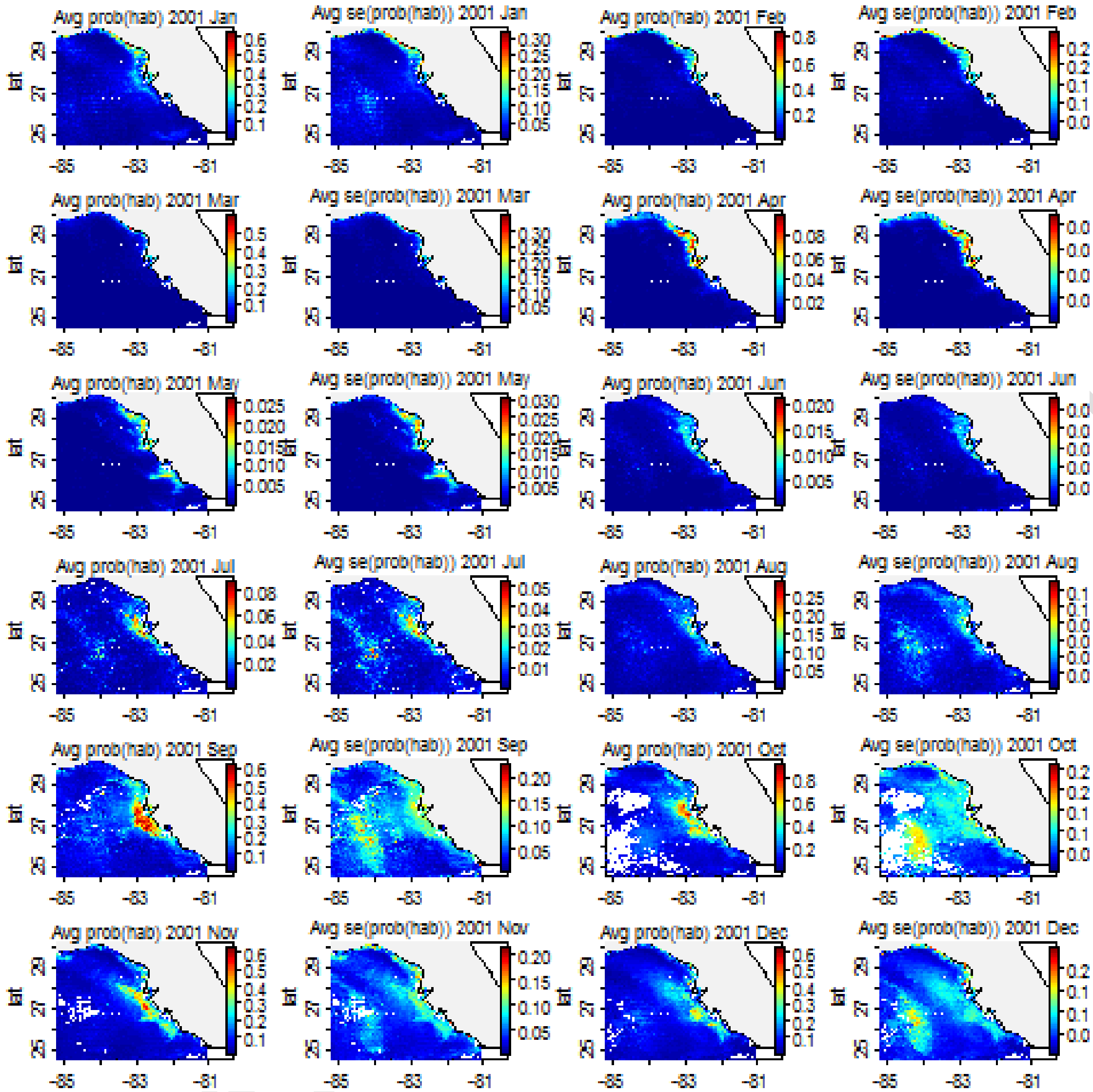
b. Monthly plots for 1999



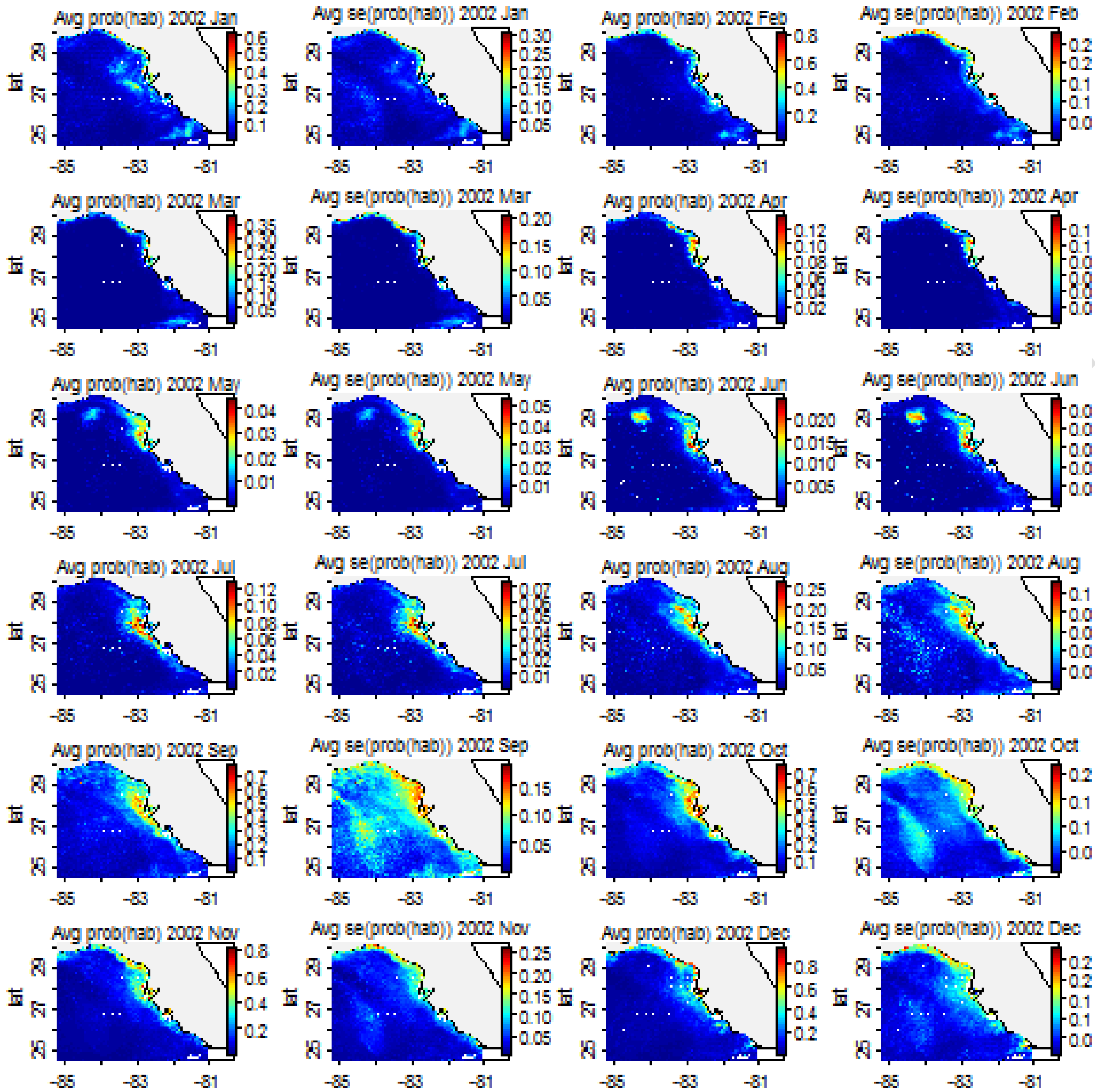
c. Monthly plots for 2000



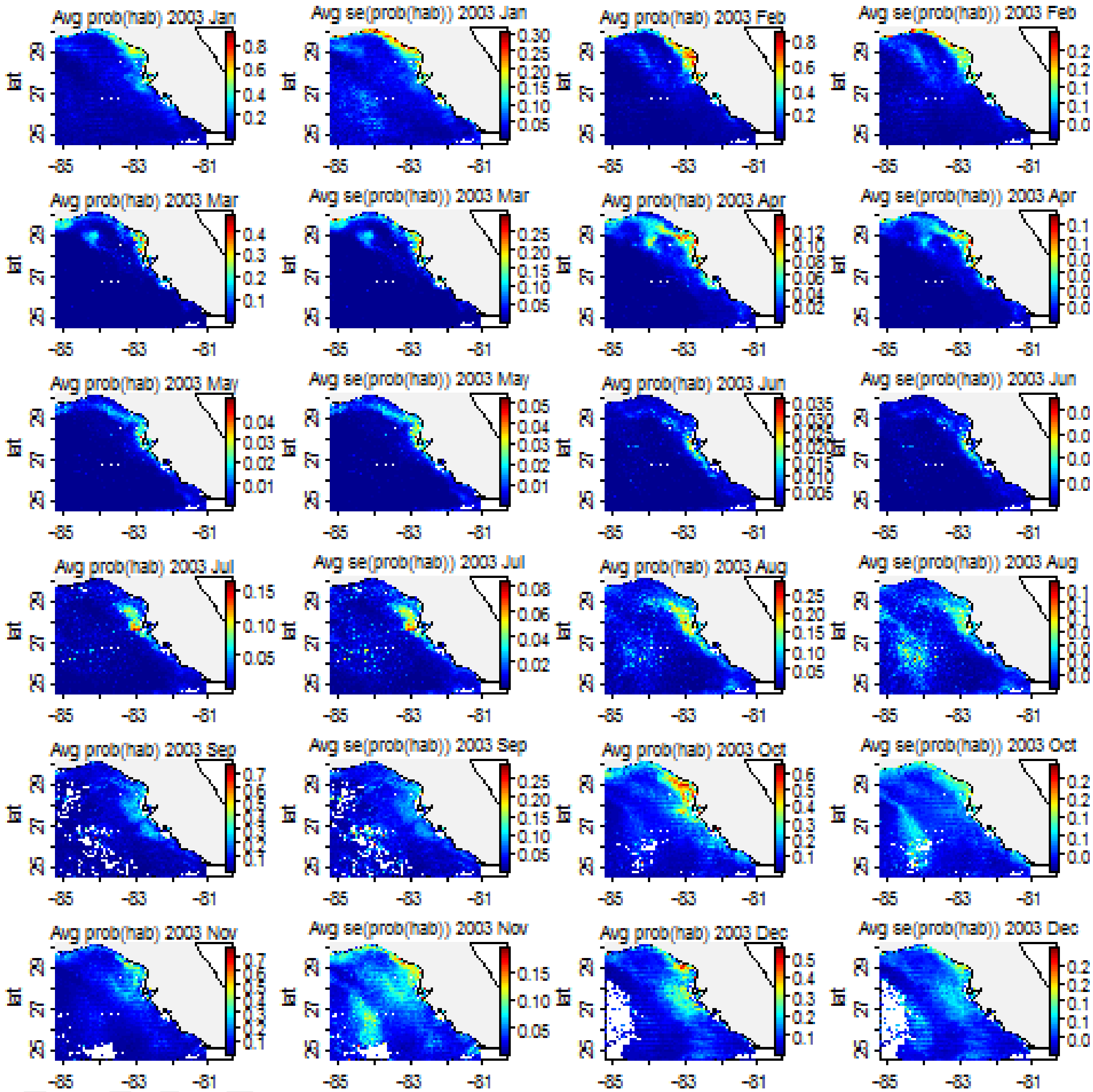
d. Monthly plots for 2001



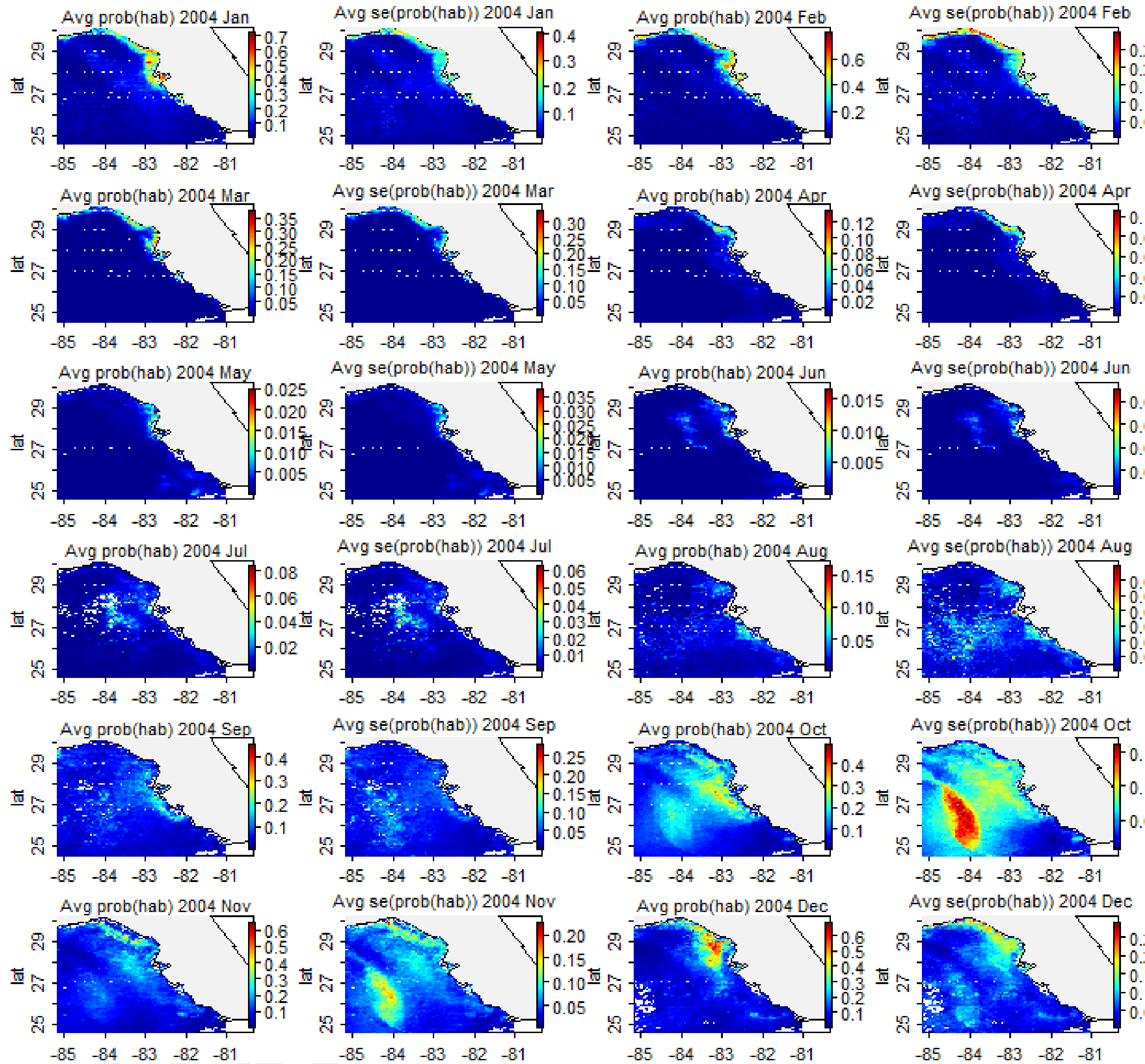
e. Monthly plots for 2002



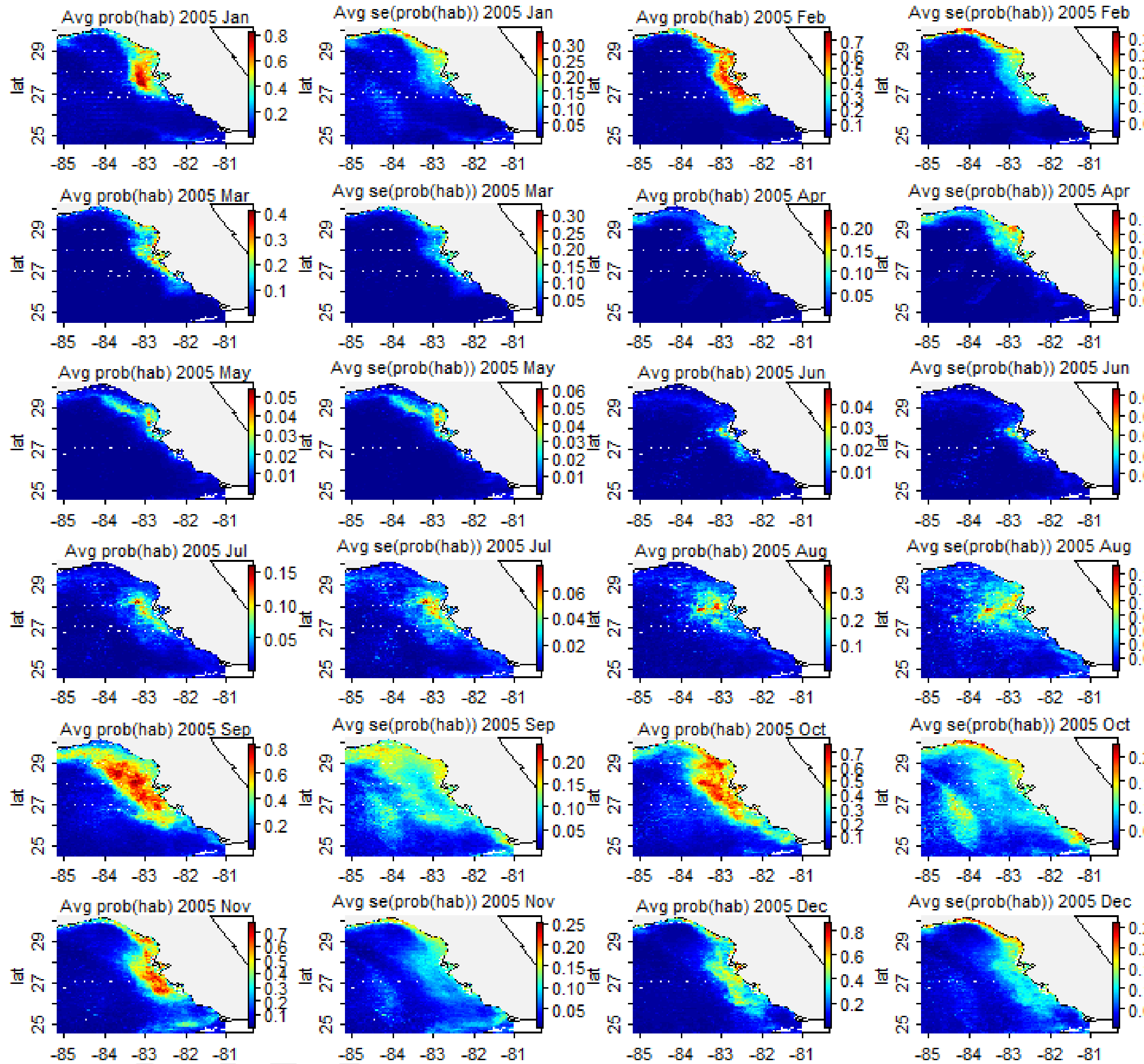
f. Monthly plots for 2003



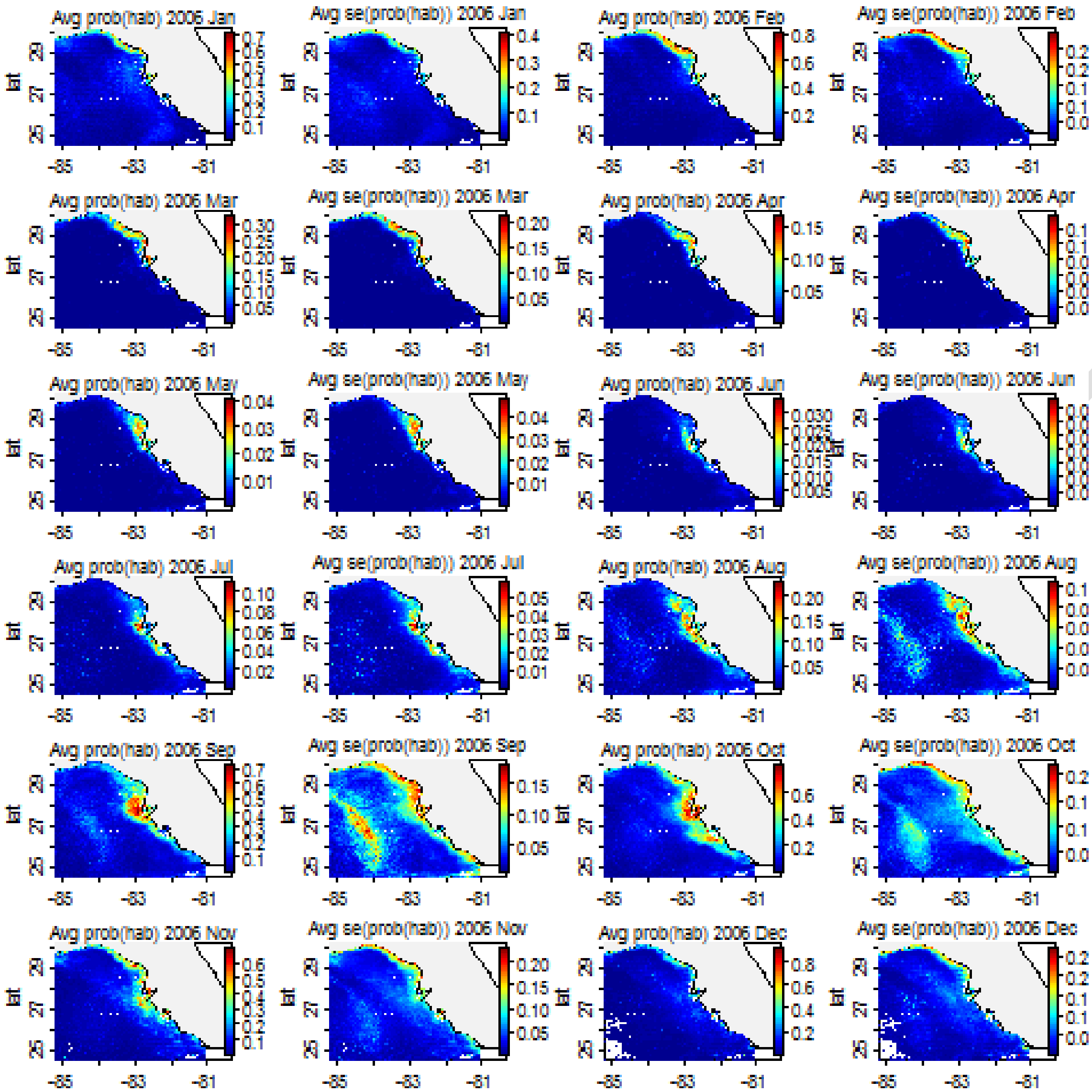
g. Monthly plots for 2004



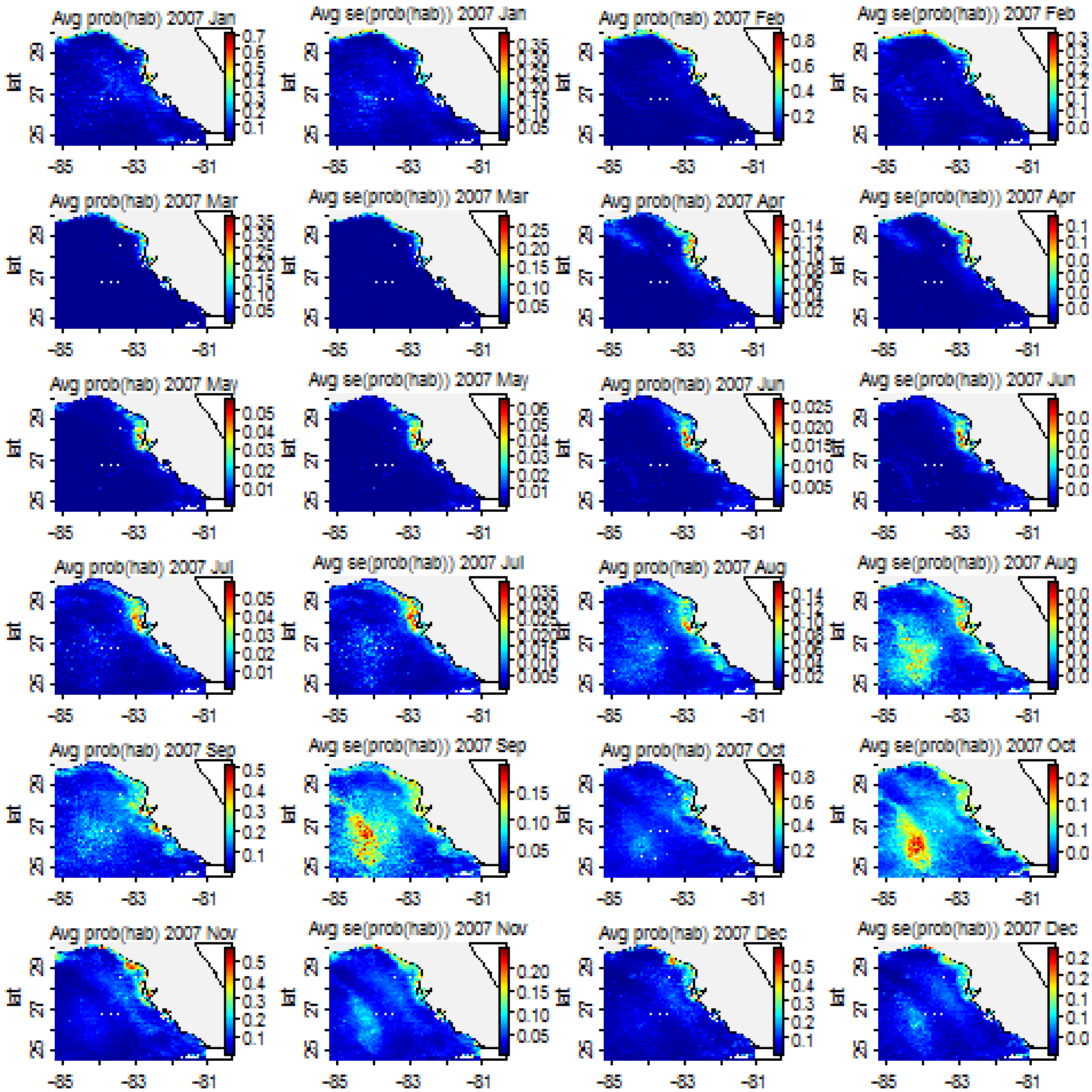
h. Monthly plots for 2005



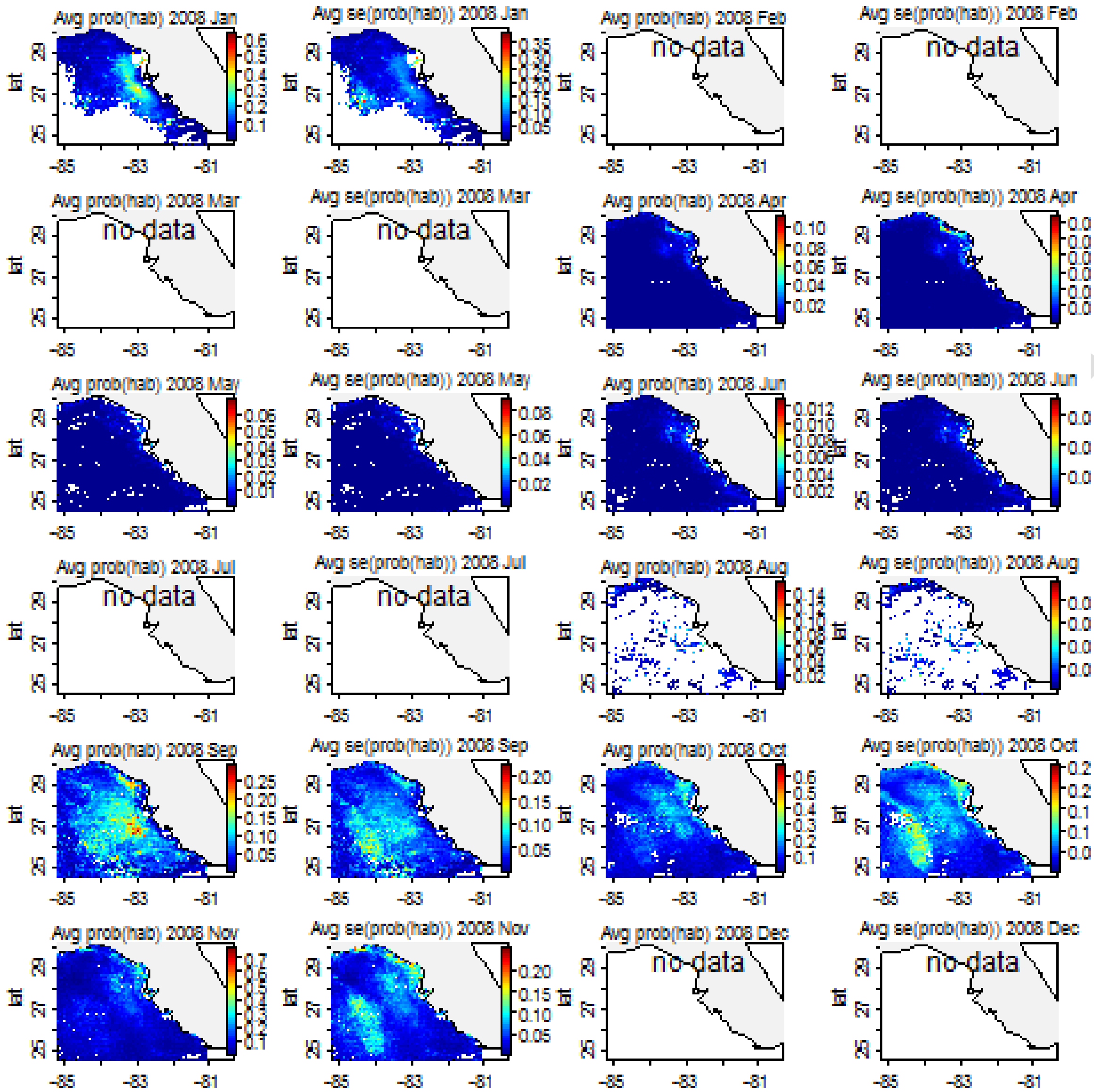
i. Monthly plots for 2006



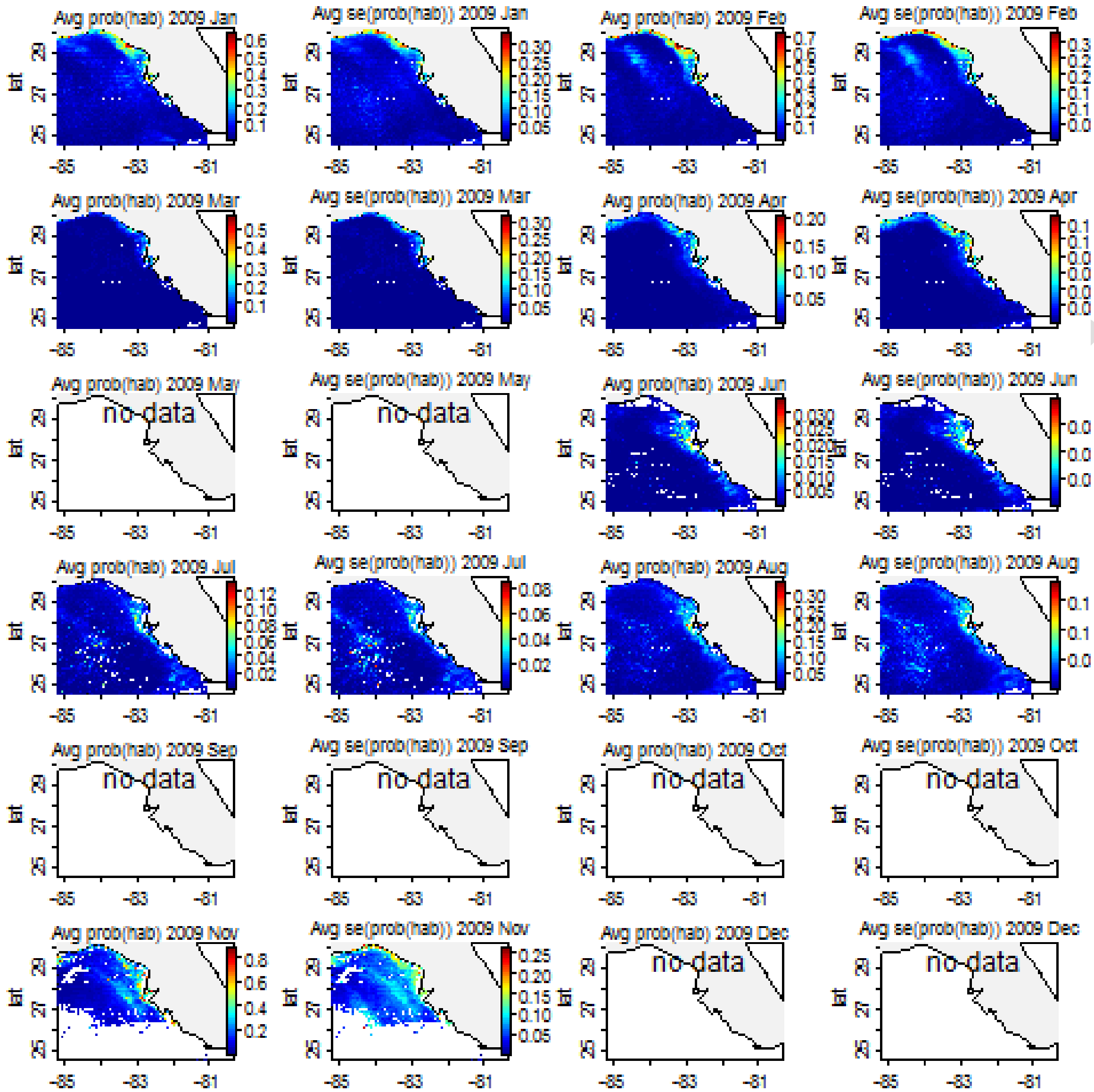
k. Monthly plots for 2007



l. Monthly plots for 2008



J. Monthly plots for 2009



K. Monthly plots for 2010