# Environmental Influences on Albacore Tuna (*Thunnus alalunga*) Distribution in the Coastal and Open Oceans of the Northeast Pacific: Preliminary Results from Boosted Regression Trees Models [1]

Yi Xu[2], Steven L.H. Teo[2] and John Holmes[3]

[2]NOAA Fisheries
Southwest Fisheries Science Center
8604 La Jolla Shores Drive
La Jolla, CA 92037, USA

[3]Fisheries and Oceans Canada
Pacific Biological Station
3190 Hammond Bay Road
Nanaimo, B.C., V9T 6N7, Canada

Email: Yi.Xu@noaa.gov

[1]Working document submitted to the ISC Albacore Working Group Meeting, 19-25 March 2013, Shanghai Ocean University, Shanghai, China. Document not to be cited without author's permission.

## ABSTRACT

A boosted regression tree (BRT) model is used to study the distribution of albacore tuna (*Thunnus alalunga*) in the northeastern Pacific Ocean, based on logbook data from the US and Canadian troll and pole-and-line fisheries. The model domain covered the Northeast Pacific Ocean and was divided into two sub-regions to study the coastal ocean and the open ocean processes. The logbooks from US and Canada vessels provided time, location, catch and effort over two decades from 1992 to 2011. Satellite data including sea surface temperature, sea surface height (SSH) anomaly, meridional and zonal geostrophic currents and chlorophyll-a (chl-a) concentration were used as environmental predictors for the BRT model. We used data from 1998-2008 as the training dataset and 2009 as an independent testing dataset. The preliminary results showed that the open ocean and coastal ocean oceanographic dynamics affected albacore tuna distribution differently. In the open ocean, meridional geostrophic currents, SSH anomaly and zonal geostrophic currents were important influences on albacore CPUE (catch-per-unit-effort) changes. In the coastal ocean, chl-a concentration was the leading factor, followed by SSH anomaly. The predicted albacore CPUE showed a near 1:1 relationship to both training and testing data. In the future, if these relationships are found to be robust, these types of analyses may be integrated into population dynamic models to help improve fisheries management in the face of environmental changes.


## INTRODUCTION

Albacore tuna (*Thunnus alalunga*) is a highly migratory species found in the Atlantic, Indian and Pacific Oceans. The north Pacific albacore troll and pole-and-line fisheries (surface fisheries) are the most important commercial fisheries for highly migratory species for both the U.S. and Canada in the north eastern Pacific Ocean (NEPO). In 2009, the combined albacore catch by U.S. and Canadian vessels in this fishery was more than 18,000 metric tons (ISC, 2012). The albacore that are caught by troll and pole-and-line vessels are widely distributed in the NEPO, primarily from 30 to 45°N in offshore waters and a broader latitudinal band from 30 to 54°N in North American coastal waters (Figure 1). The catch-per-unit-effort (CPUE) is highly variable in both the coastal and open oceans, with strong seasonal and interannual variability (Figure 2).

Previous studies have linked the distribution north Pacific albacore with environmental features, with water temperature and food availability being the most influential factors. Albacore abundance is highest in waters 16-19°C (Johnson, 1962; Laurs and Lynn, 1977) and aggregations are found in the eastern sector of the North Pacific Transition Zone concurrent with frontal structure during the early part of the fishing season (May-June) (Laurs and Lynn, 1991). By mid-July, some albacore migrate into coastal waters and appear to concentrate in the vicinity of oceanic fronts related to upwelling (Laurs and Lynn, 1977; Laurs, et al., 1984). Recently, the availability of a sufficiently long time series of high resolution satellite remote sensing data have allowed analysis of the relationship between albacore distribution and the detailed structure of oceanographic features such as sea surface temperature (SST), sea surface height (SSH) and chlorophyll-a. For example, Polovina et al. (2001) found that the highest albacore catch rates in 1998 were associated with the the chlorophyll front in the north Pacific Transition Zone. Hot spots for albacore were detected in the northwestern Pacific based on multi-sensor satellite data from 1998 to 2003 occurred in areas with warmer SST (19.78°C), relatively high chlorophyll concentrations (0.31 mg·m$^{-3}$), and high eddy kinetic energy and geostrophic currents (Zainuddin et al., 2006; Zainuddin et al., 2008).

The objectives of this study are: 1) to develop a quantitative predictive model of the influence of environmental variables (SST, chlorophyll-a, currents, etc.) on albacore distribution; 2) to identify key environmental factor(s) in the coastal ocean and open ocean in NEPO; and 3) to evaluate the model performance by hindcasting and comparing with available fisheries data. It is important to note that this study is in its early phases and the results presented here are highly preliminary. A series of boosted regression trees (BRT) were used to model the influence of environmental variables on albacore distribution in the coastal and the open ocean. The results from this study can improve our understanding of the effects of environmental variability on albacore distribution and may potentially be integrated into population dynamic models to help improve fisheries management in the face of environmental changes.

## DATA AND METHODS

### Data

***Fisheries data***

A database of catch and effort was assembled from the logbook records of US and Canadian albacore troll and pole-and-line vessels, which were extracted from databases maintained by US National Marine Fisheries Service and Canadian Department of Fisheries and Oceans. Currently, we use US data from 1992 to 2011 and Canadian data from 2004 to 2011. The assembled database has fishing date (year, month, day), fishing location (latitude and longitude in degrees and minutes), and total catch (the number of albacore caught plus discarded) for individual vessels. The unit of effort was vessel-days. Fishing locations of all years for US and Canada are shown in Figure 1.

***Environmental data***

Gridded sea surface temperature (SST), sea surface height (SSH) anomaly, meridional and zonal geostrophic currents (GCM, GCZ), and chlorophyll-a concentration (chl-a) were obtained from the NOAA Ocean Watch data repository (http://las.pfeg.noaa.gov/oceanWatch/oceanwatch.php). Our model domain is from 20° to 60°N and 180° to 120°W. The SST data were from AVHRR Pathfinder Version 5.0 (1992 to 2005, 8-day) and MODIS on Aqua (2006 to 2011, 8-day); SSH anomaly, GCM and GCZ were from AVISO (Archiving, Validation and Interpretation of Satellite Oceanographic data, 1992 to 2009, weekly) program; and chl-a were from SeaWiFS (Sea-viewing Wide Field of view Sensor, 1997 to 2010, 8-day).

We also developed an algorithm to deal with missing environmental data. If the data in a pixel were not available due to cloud cover, the missing value was replaced by the average value of pixels in the surrounding square area. The size of the square area used depended on the spatial resolution of the data (SST-Pathfinder: 54x54 $km^2$, SST-MODIS: 56x56 $km^2$, and chl-a: 65x65 $km^2$). By doing so, the percentage of missing data to total data was reduced from >15% to 3.90% for SST-Pathfinder, 3.76% for SST-MODIS, and 6% for chl-a. There were no missing data in the SSH anomaly, GCM and GCZ datasets.

### Model

We used boosted regression trees (BRT) to model the distribution of albacore. Boosted regression trees are a tree-based method, which can easily fit complex nonlinear relationships with multiple predictors. BRTs are able to deal with missing values, handle categorical and continuous predictor variables simultaneously, and are relatively insensitive to outliers and transformation. The BRT was performed using the “gbm” package (Version2.0-8, http://cran.r-project.org/web/packages/gbm/gbm.pdf) in R (version 2.15.1, R Developmental Core Team, 2012). Elith et al. (2008) provide a detailed working guide for BRT. In our model, we chose the Gaussian family and set tree complexity = 1, learning rate = 0.05, and bag fraction = 0.5. We use year, month, longitude, latitude, SST, SSH anomaly, chl-a, GCM and GCZ, as predictor variables, and log (CPUE+1) as the response variable.

**Simulations, predictions and model performance evaluation**

We divided the assembled data into two subsets: the coastal ocean (120°W-130°W) and the open ocean (130°W-180°). Two BRT models were constructed independently from the subsets. For each subset, we fit the model to the data and calculated the mean total deviance and mean residual deviance. The BRT model also generated the relative importance of each predictor, given the total variance equals 100%. A 10-fold cross-validation was performed to evaluate the robustness of the model. Cross-validation deviance and standard errors were calculated for each model.

We constructed each BRT model with data from 1998 to 2008, which was the training dataset. The 2009 data were used to validate the estimated model. The coastal and open ocean log (CPUE+1) were predicted by the coastal and open ocean BRT models, respectively. We evaluated the model performance by comparing the predicted and observed CPUE data - for both training and testing data. We plotted the residual Q-Q plots to examine whether the predictions follow normal distributions. In the last section, we fitted three regression lines between predictions and observations: linear regression, geometric mean regression and orthogonal regression. Statistical analyses of three regressions are performed to evaluate the uncertainty of observed and predicted CPUE.

## RESULTS

**Environmental variables at albacore fishing locations**

The frequency histograms of SST, chl-a concentration (in log scale), GCM, GCZ, SSH anomaly, at albacore fishing locations are shown in Figure 3. All variables followed normal or near normal distributions. The mean, standard deviation, minimum and maximum values of the environmental variables for the coastal and open ocean are summarized in Table 1. Although most of the SSTs were approximately 16-19°C sea surface temperature, the coastal ocean SSTs were slightly warmer (0.7°C) than the open ocean. Similarly, coastal chl-a concentration was higher and had larger variation than the open ocean. The mean GCZ in the open ocean was $0.03 m \cdot s^{-1}$ eastward and the mean GCM was negligible. In the coastal ocean, mean geostrophic currents flowed to the south toward California at the speed about $0.05 m \cdot s^{-1}$. Coastal SSH anomaly was 1cm higher and had a smaller variation at the locations where albacore were caught.

**BRT model results**

Table 2 summarizes the mean total deviance, mean residual deviance, 10-fold cross-validation deviance mean and standard errors from the BRT models. The relative influence of all predictors is listed in Table 3. For the open ocean, BRT can explain 34.2% of total variance. The most important factors are year, longitude and latitude, which accounted for more than 60% of the total deviance. Among the environmental variables, GCM and SSH anomaly are relatively important compared to log (chl-a) and SST. For the coastal ocean, the model can explain 19% of the total variance, with year, longitude and latitude still accounting for 55% of the deviance. Log (chl-a), SSH anomaly and month are relatively important factors compared to GCZ, GCM and SST. Interannual variability was larger than seasonal variability in both the coastal ocean and the open ocean. The monthly variance (9.28%) of the coastal ocean is larger than the open ocean (2.81%).

Figure 4 shows the relationships between albacore abundance (in log scale) and the predictor variables in the BRT model. In the open ocean, the CPUE is positively related to the latitude with the peak at 48-52°N, near the Haida Gwaii islands (This area can be considered as coastal region.). In the horizontal direction, there are several peaks at 170°W, 160°W, 145°W and the CPUE generally decreases as reaching the eastern Pacific coast (130°W). The expected CPUE is high when GCM is within -0.15 to +0.15 $m \cdot s^{-1}$ and low when GCM is increasing, no matter the direction. There appears to be a negative relationship between CPUE and SSH anomaly, where high CPUE is associated with strong negative SSH anomaly. In the coastal ocean, CPUE has a bimodal longitudinal distribution with peaks at 133°W and 123°Wwhile there is a peak between 43-48°N with respect to latitude. The peak CPUE in the coastal ocean occurs at relatively high chl-a concentration with a peak at 20.1$mg \cdot m^{-3}$ and a minimum at 7.4 $mg \cdot m^{-3}$ but gradually increasing with low chl-a values. Unlike the open ocean, coastal SSH anomaly is positively related to CPUE with the highest value over 0.2 meters and the lowest value less than -0.1 meters.

**Model predictions to training and testing data**

Figure 5 shows the residuals normal Q-Q plots between observed and predicted log (CPUE+1) for the open ocean and coastal ocean. The predictions to both the training dataset from 1998 to 2008 (Fig 5A and 5B) and to the independent testing dataset in 2009 are shown (Fig. 5C and 5D). All four Q-Q plots showed relatively straight lines between residuals and theoretical quantiles. The linear regressions, geometric mean regression and orthogonal regression between observed and predicted log (CPUE+1) for the open and coastal ocean models are shown in Figure 6. There are positive relationships between observations and predictions in all simulations.  The standard linear regression lines have the smallest slopes (ranging from 1.03-1.36) among three regressions; the geometric mean regression slopes are larger (1.87-3.11) and the orthogonal regression slopes are the largest (2.66-6.90). For both coastal and open ocean BRT models, there is a significant linear relation between predictions and observations in the training data (Table 4). The slopes are 1.10 for the open ocean and 1.03 for the coastal ocean (Figure 6A and 6B). For the testing data, the slopes are statistically significant (Table 4, Figure 6C and 6D). The standard errors for geometric mean regression and orthogonal regression are generally larger than linear regression. The correlation coefficients for open ocean and coastal ocean to the training data are 0.59 (n=11721) and 0.44(n=77719), and to testing data are 0.41(n=127) and 0.50 (n=1428) respectively.

## DISCUSSION

**Environmental influence on albacore tuna**

The open ocean and the coastal ocean have quite different oceanographic dynamics. Albacore tuna, as a highly migratory species may be able to take advantage of the oceanographic heterogeneity to seek optimal conditions, forage prey or migrate to favored habitat. In the open ocean, SSH anomaly is an important indicator of albacore habitat. Negative SSH anomaly is associated with cyclonic eddies and divergent upper ocean waters. Under these conditions, cool and nutrient-rich waters may be upwelled from deeper ocean as compensation to the divergent surface layer. These upwelling eddies are areas of increased productivity in the open ocean and may attract various prey sources for albacore tuna. In contrast coastal SSH anomaly has a weak positive relationship with albacore CPUE, which suggests that albacore tends not to be within the strongest upwelling region, but are at a distance offshore from the coast. The SSH anomaly is high offshore and becomes lower towards the coast in the coastal upwelling region. There was evidence that highest abundance of phytoplankton is very close to shore (15-40 km), whereas zooplankton peak is not coincident with phytoplankton but is further offshore (~100km, Chelton, et al., 1982). One can therefore expect that upper trophic level species like albacore may have offshore hotspots, where prey sources are feeding on the zooplankton. These hot spots may also be coincident with positive SSH anomaly. In the open ocean, high CPUE is related with a small to moderate GCM (absolute values are $<0.15 m \cdot s^{-1}$). This is likely due to albacore migrating primarily from west to east during the early part of the fishing season, following the north Pacific current. In the coastal ocean, the GCM is less important than GCZ, which shows a weak relationship with CPUE. Small to moderate GCZ with absolute values less than $0.15 m \cdot s^{-1}$ correspond to areas with high CPUE. This suggests that albacore prefer weak north-south flow, which is characteristic of the California Current ecosystem. Our model shows that chl-a concentration is an important factor in the coastal ocean, but the relationship to CPUE is nonlinear. The CPUE is high at low chl-a levels ($0.14 mg \cdot m^{-3}$) and generally decreases with higher chl-a concentrations. However, at very high chl-a concentrations ($>20 mg \cdot m^{-3}$), CPUE apparently increases to relative high levels, but relatively few observations were made at these chl-a levels. Surprisingly, our preliminary results do not show a significant relationship between CPUE and SST for both the coastal and the open ocean.

**Prediction to training and testing data**

Considering that these are preliminary models, the predicted CPUE correspond relatively well to the observed CPUE for both open and coastal ocean models. Predicted CPUEs to both training and testing data shows positive relationships with observed CPUEs. The correlation coefficients between observed and predicted CPUEs are greater than 0.4 for all models. The residuals between predicted and observed CPUE follow the normal distribution $N(\mu,\sigma)$ but not standard normal distribution $N(0,1)$ (Y-intercepts are near but not equal to zero, and slopes are not 1:1.). Based on these results, there may be some scaling and a small deviation between the model predictions and observations, which can be improved in the future.

Three linear regressions help us to study the uncertainty in both the predicted and observed CPUE. The standard linear assumes that all the variation in the data is related to the dependent variable- the observed CPUE, and variation in the independent variable is minimal or negligible relative to the variation in the dependent variable. The geometric mean regression minimizes the sum of the product of the horizontal and vertical deviation, and the orthogonal regression minimizes the distance perpendicular to the regression line. Both methods put some weights of uncertainty to the observations, therefore, the results showed that slopes are greater than standard linear regression. These uncertainties were evaluated quantitatively.

This paper is a first step toward a comprehensive assessment of environmental effects on albacore distribution in the northeastern Pacific Ocean. An immediate next step would be identifying more environmental variables such as frontal probability index, North Pacific Currents positions, Pacific Decadal Ossilation, El Nino and Southern Oscillation, and North Pacific Gyre Oscillation Index. By introducing those predictors in BRT, we may find better indicators and improve our understanding of the North Pacific ecosystem and its impact on albacore distribution. Another direction is comparing BRT to traditional statistical models and machine learning approaches, such as GLM, GAM, random forest or maximum entropy. Using the same training and testing datasets, we can construct those models and make predictions. We can study the results and compare the similarities and differences among different modeling techniques. Such a comparison between different modeling approaches may help us tease apart actual albacore-environment relationships and inherent model effects.

## ACKNOWLEDGEMENTS

The authors acknowledge all US and Canadian fishermen who reported the albacore catch data in logbooks. We also thank Dr. Jane Elith (The University of Melbourne, Australia) to help us understand the BRT model. We are grateful to Dr. Nancy Lo (NOAA-Southwest Fisheries Science Center, USA) for her suggestions and comments on statistical analyses.

## REFERENCES

Chelton, D., Bernal, P., McGowan, J. 1982. Large-scale interannual physical and biological interaction in the California Current. Journal of Marine Research 40: 1095–1125.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology.77:802-813. doi:10.1111/j.1365-2656.2008.01390.x.

ISC. 2012. Report of the Albacore Working Group Workshop. 14 July 2012, Sapporo, Japan.

John T. Froeschke and Bridgette F. Froeschke, 2011. Spatio-temporal predictive model based on environmental factors for juvenile spotted seatrout in Texas estuaries using boosted regression trees. Fisheries Research 111:131-138. doi:10.1016/j.fishres.2011.07.008

Johnson, J.H. 1962. Sea temperatures and the availability of albacore off the coasts of Oregon and Washington. Trans. Am. Fish. Soc. 91: 269-274.

Laurs, R. M. and R.J. Lynn. 1977. Seasonal migration of North Pacfic albacore, *Thunnus alalunga*, into North American coastal waters: Distribution, relative abundance, and association with transition zone waters. Fisheries Bulletin, U.S. 75: 795.

Laurs, R.M. and Lynn, R.J. 1991. North Pacific albacore ecology and oceanography. In J.A. Wetherall (Ed.), Biology, oceanography and fisheries and the North Pacific Transition Zone and Subarctic Frontal Zone. NOAA Technical Report NMFS 105, pp. 69-87.

Laurs, R.M., Fielder, P.C. and Montgomery, D.R. 1984. Albacore tuna catch distributions relative to environmental features observed from satellites. Deep-Sea Research 31: 1085-1099.

Polovina, J.J., Howell, E., Kobayashi, D.R. and Seki, M.P., 2001. The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources. Progress in oceanography 49: 469-483.

Zainuddin, M., Kiyofugi, H., Saitoh, K. and Saitoh, S. 2006. Using multi-sensor satellite remote sensing and catch data to detect ocean hot spots for albacore (*Thunnus alalunga*) in the northwestern North Pacific. Deep-Sea Research II 53: 419-431.

Zainuddin, M., Saitoh, K. and Saitoh, S. 2008. Albacore (*Thunnus alalunga*) fishing ground in relation to oceanographic conditions in the western North Pacific Ocean using remotely sensed satellite data. Fisheries Oceanography 17:61-73.

## LIST OF TABLES



## LIST OF FIGURES

Table 1. Mean, standard deviation and range of environmental variables at the locations where albacore were caught for the open ocean and coastal ocean during multiple years.

| | Variable | Mean | Standard deviation | Range |
|---|---|---|---|---|
| Open Ocean | SST | 16.7 | 2.97 | [7.57,28.4] |
| | chl-a | 0.34 | 0.40 | [0.018,6.816] |
| | Meridional Geostrophic Currents | 0.0006 | 0.029 | [-0.325,0.275] |
| | Zonal Geostrophic Currents | 0.028 | 0.040 | [-0.373,0.466] |
| | SSH anomaly | 0.021 | 0.059 | [-0.273,0.468] |
| Coastal Ocean | SST | 17.4 | 3.26 | [7.05,29.0] |
| | chl-a | 0.68 | 1.25 | [0.024,58.79] |
| | Meridional Geostrophic Currents | -0.047 | 0.031 | [-0.250,0.152] |
| | Zonal Geostrophic Currents | 0.0098 | 0.030 | [-0.256,0.318] |
| | SSH anomaly | 0.0295 | 0.045 | [-0.184,0.277] |

Table 2. The mean total deviance, mean residual deviance, cross-validation deviance (mean and standard errors) modeled by BRT.

| Model | Mean total deviance | Mean residual deviance | Cross-Validation Deviance Mean (Standard Errors) |
|---|---|---|---|
| Open Ocean | 2.892 | 1.904 | 2.051(0.033) |
| Coastal Ocean | 1.334 | 1.081 | 1.092(0.008) |

Table 3. Relative importance of variables modeled by BRT

| Model | Rank | Predictors | Relative Influences (%) |
|---|---|---|---|
| Open Ocean | 1 | Year | 21.72 |
| | 2 | Longitude | 19.68 |
| | 3 | Latitude | 19.17 |
| | 4 | Meridional Geostrophic Currents | 11.02 |
| | 5 | SSH anomaly | 9.86 |
| | 6 | Zonal Geostrophic Currents | 8.35 |
| | 7 | Log(chl-a) | 4.87 |
| | 8 | Month | 2.81 |
| | 9 | SST | 2.52 |
| Coastal Ocean | 1 | Year | 22.37 |
| | 2 | Longitude | 18.76 |
| | 3 | Log(chl-a) | 14.10 |
| | 4 | Latitude | 13.98 |
| | 5 | SSH anomaly | 9.75 |
| | 6 | Month | 9.28 |
| | 7 | Zonal Geostrophic Currents | 6.36 |
| | 8 | Meridional Geostrophic Currents | 4.21 |
| | 9 | SST | 1.19 |

Table 4. Linear regression, geometric mean regression, and orthogonal regression results between observed and predicted CPUEs to training set (1998-2008) and testing set (2009) for the coastal and the open ocean.

| | | Predictions | Results | | Standard Error | t-value | Probability(>\|t\|) | |
|---|---|---|---|---|---|---|---|---|
| Linear Regression | Open Ocean | Training data (1998-2008) | Intercept | -0.40311 | 0.05791 | -6.961 | 3.56e-12 | *** |
| | | | Slope | 1.10038 | 0.01402 | 78.487 | <2e-16 | *** |
| | | Testing data (2009) | Intercept | -0.3955 | 0.9404 | -0.421 | 0.675 | |
| | | | Slope | 1.2803 | 0.2531 | 5.058 | 1.47e-06 | *** |
| | Coastal Ocean | Training data (1998-2008) | Intercept | -0.152200 | 0.033333 | -4.566 | 4.98e-06 | *** |
| | | | Slope | 1.034597 | 0.007511 | 137.75 | <2e-16 | *** |
| | | Testing data (2009) | Intercept | -0.72043 | 0.24380 | -2.955 | 0.00318 | ** |
| | | | Slope | 1.36738 | 0.06109 | 22.385 | <2e-16 | *** |
| Geometric Mean Regression | Open Ocean | Training data (1998-2008) | Intercept | -3.523144 | 0.0192095 | -183.406 | <2e-16 | *** |
| | | | Slope | 1.874647 | 2.247175 | 0.834 | 0.769 | |
| | | Testing data (2009) | Intercept | -5.889360 | 0.0008665 | -679.708 | <2e-16 | *** |
| | | | Slope | 2.335463 | 44.30226 | 0.05272 | 1 | |
| | Coastal Ocean | Training data (1998-2008) | Intercept | -7.117932 | 13.57366 | -0.52439 | 0.600934 | |
| | | | Slope | 3.106039 | -0.121241 | -25.6188 | 3.11e-02 | ** |
| | | Testing data (2009) | Intercept | -5.939825 | 0.428064 | -13.8760 | <2e-16 | *** |
| | | | Slope | 2.681562 | 0.779820 | 3.43869 | 0.229 | |
| Orthogonal Regression | Open Ocean | Training data (1998-2008) | Intercept | -6.690942 | 0.001074 | -622.969 | <2e-16 | *** |
| | | | Slope | 2.660765 | 6.339117 | 0.41974 | 0.983 | |
| | | Testing data (2009) | Intercept | -15.55178 | 0.000444 | -3496.20 | <2e-16 | *** |
| | | | Slope | 4.526357 | 157.7373 | 0.028696 | 1 | |
| | Coastal Ocean | Training data (1998-2008) | Intercept | -21.08448 | 3.541569 | -5.95343 | 2.474e-08 | *** |
| | | | Slope | 6.899159 | -0.11391 | -6.05678 | 1.315e-02 | ** |
| | | Testing data (2009) | Intercept | -14.10922 | 0.342857 | -41.1519 | <2e-16 | *** |
| | | | Slope | 4.738524 | 2.44366 | 1.93911 | 0.394 | |

*** Statistically significant values, probability is less than 0.001.

** Statistically significant values, probability is less than 0.01.

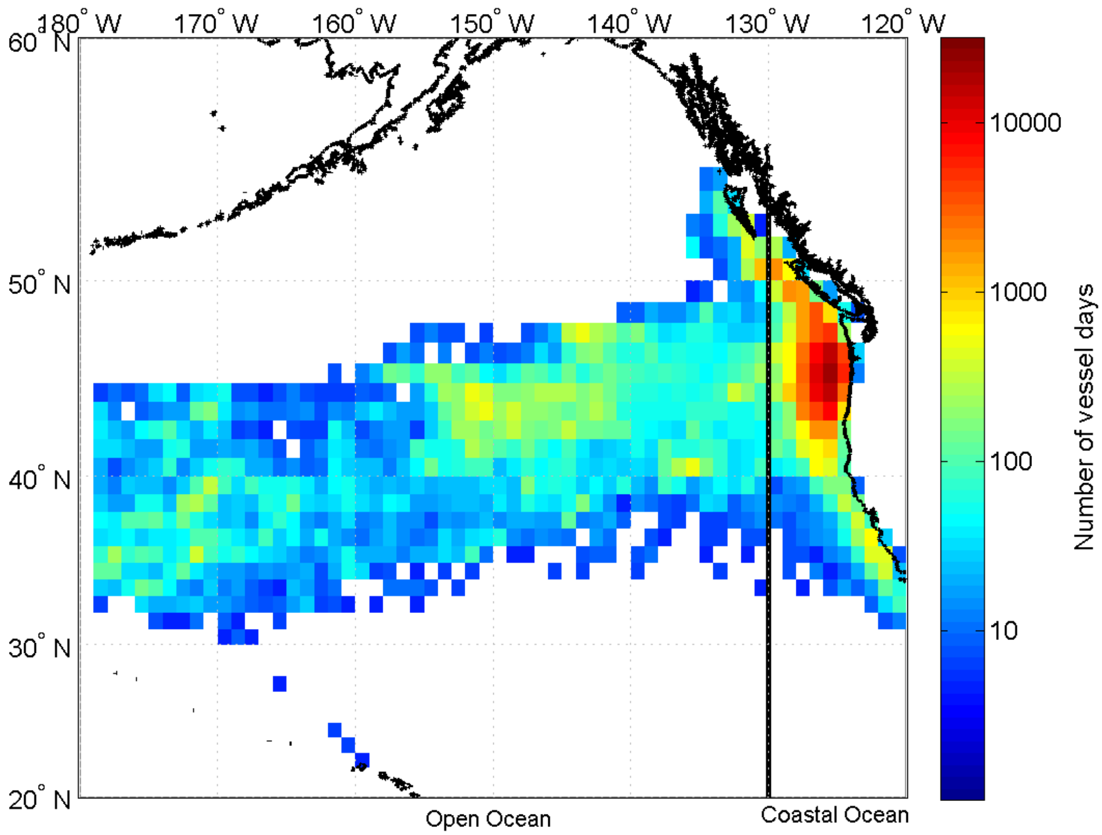


Figure 1. The numbers of record of all catch data from US (1992-2011) and Canada (2004-2011) logbooks. We divided into two data subsets. The open ocean is defined west of 130°W and the coastal ocean is defined east of 130°W.   Strata with <=3 vessels are not shown due to data confidentiality.

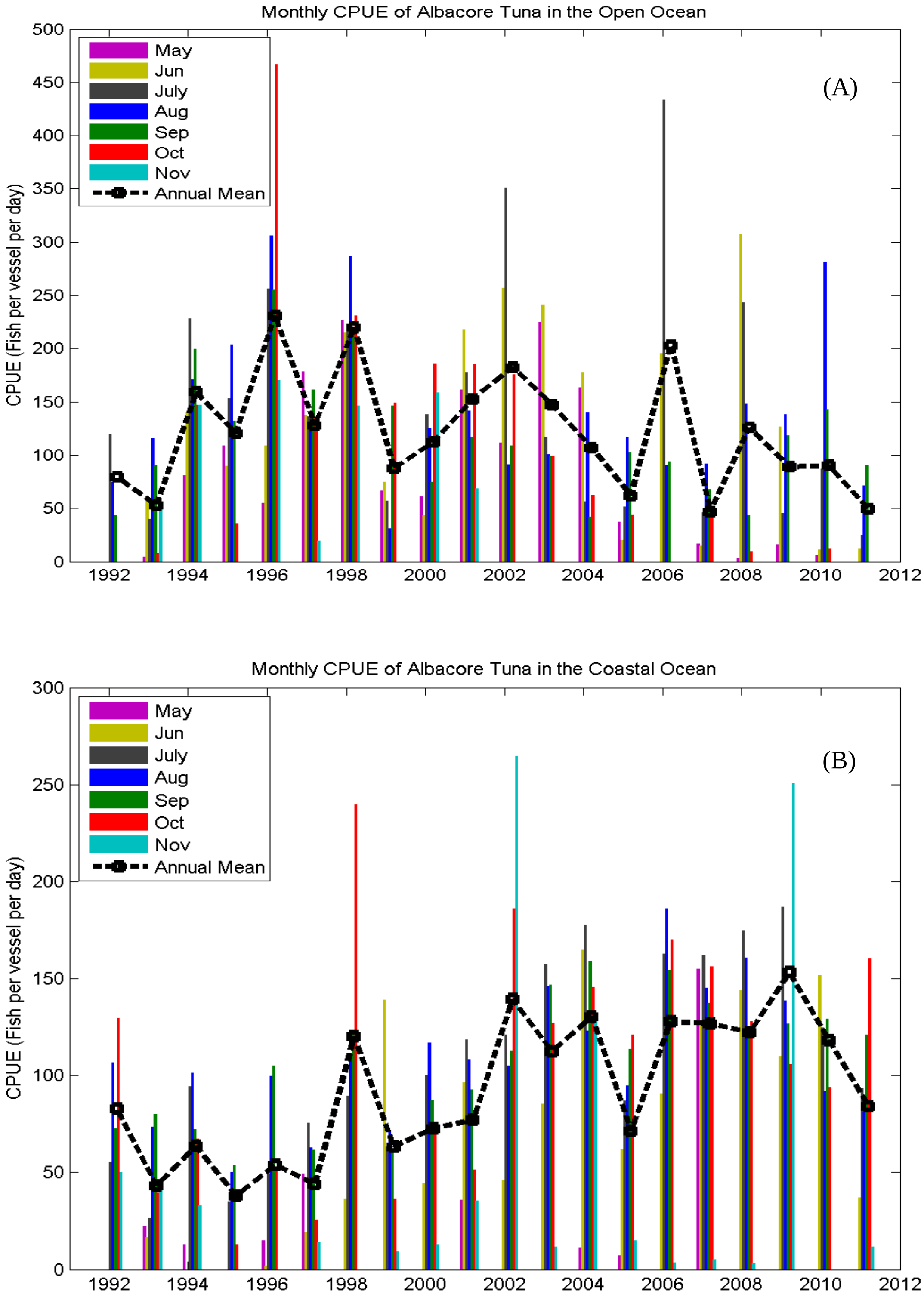


Figure 2. Monthly CPUE of albacore troll fisheries time series for (A) the open ocean and (B) the coastal ocean from 1992 to 2011.

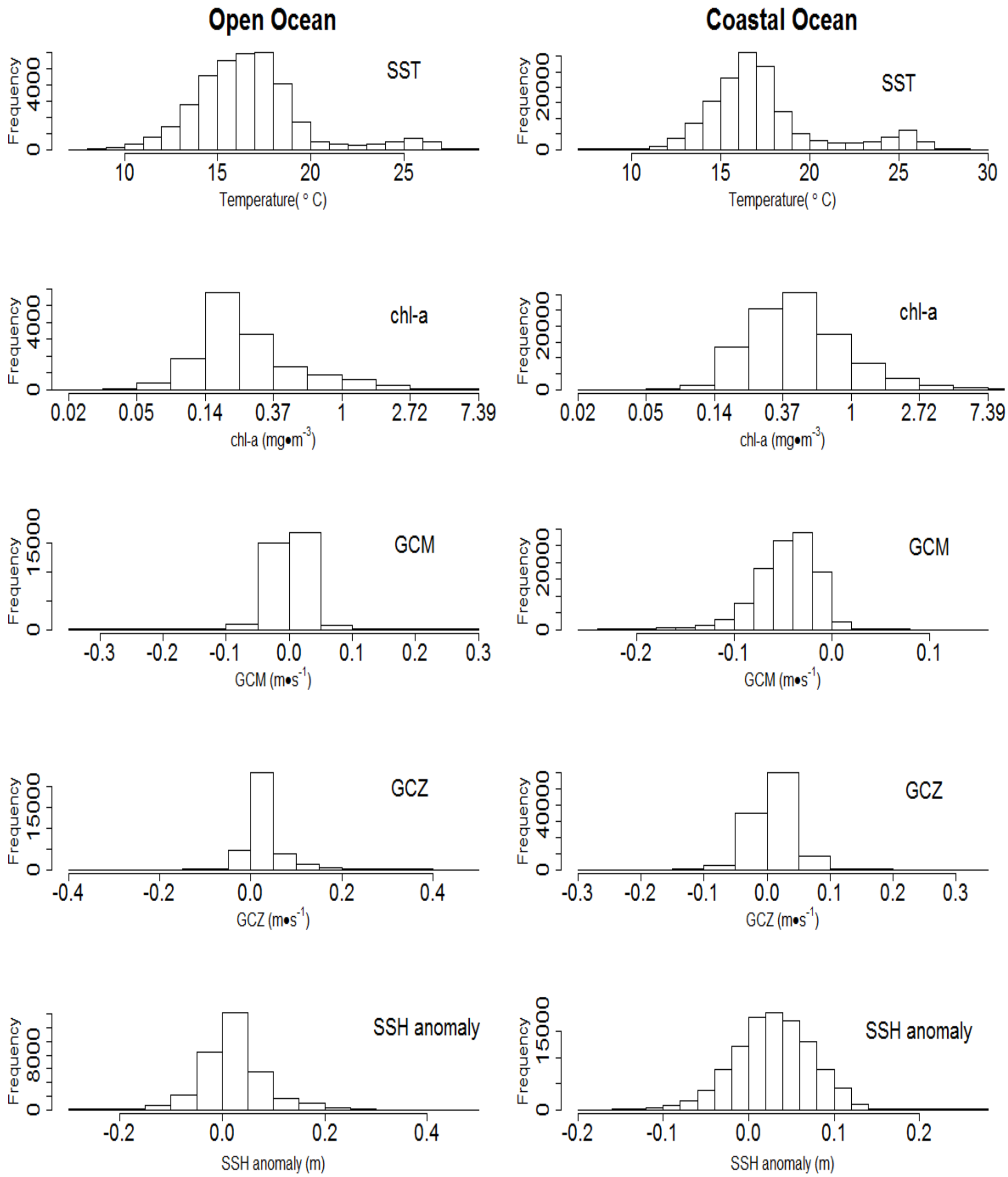


Figure 3. The histograms of SST, chl-a concentration (in log scale), GCM, GCZ, and SSH anomaly at albacore fishing locations.

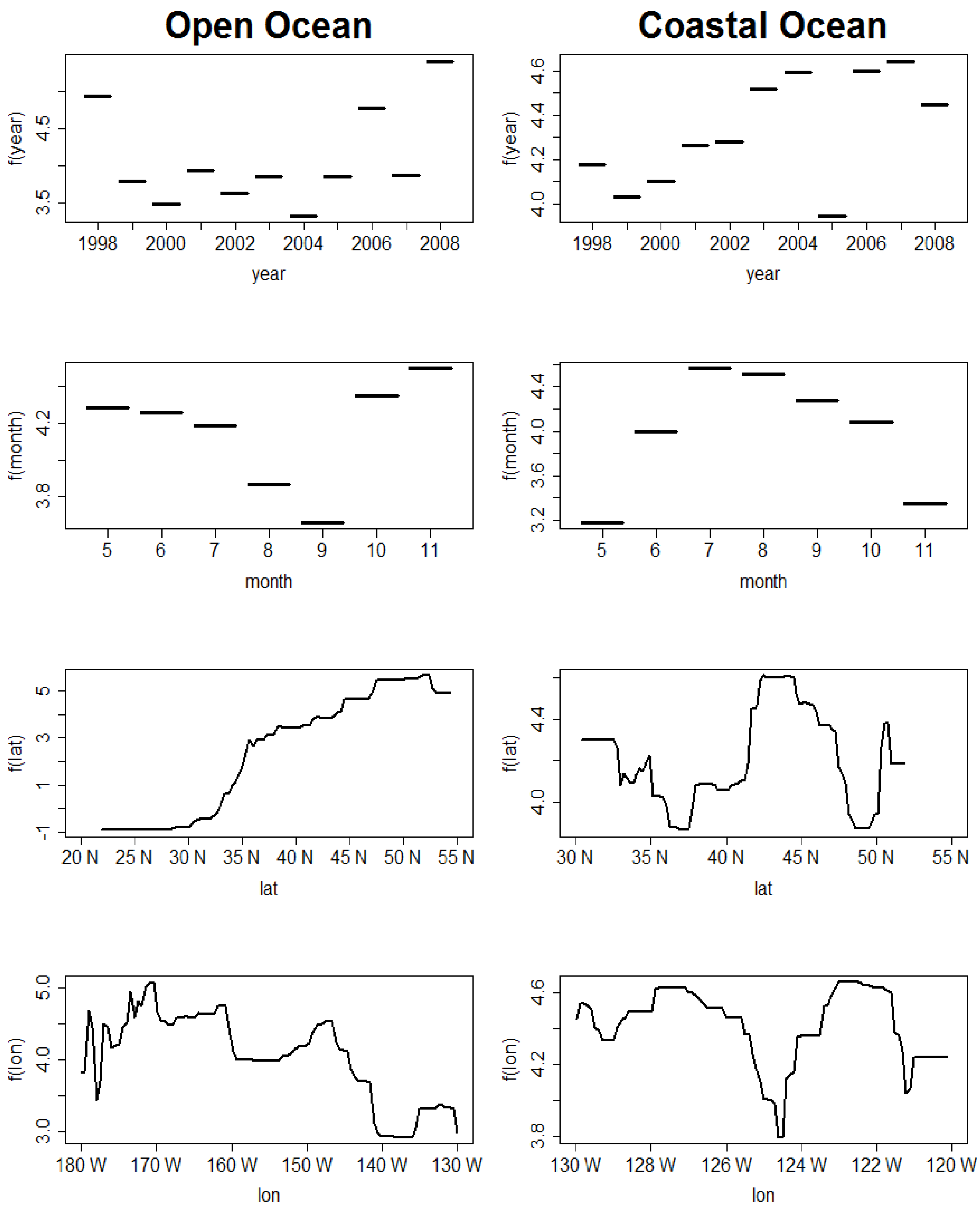


Figure 4. Fitted functions of all predictors (year, month, latitude, longitude, SST, chl-a, GCM, GCZ and SSH anomaly) for the open ocean and coastal ocean by BRT model.

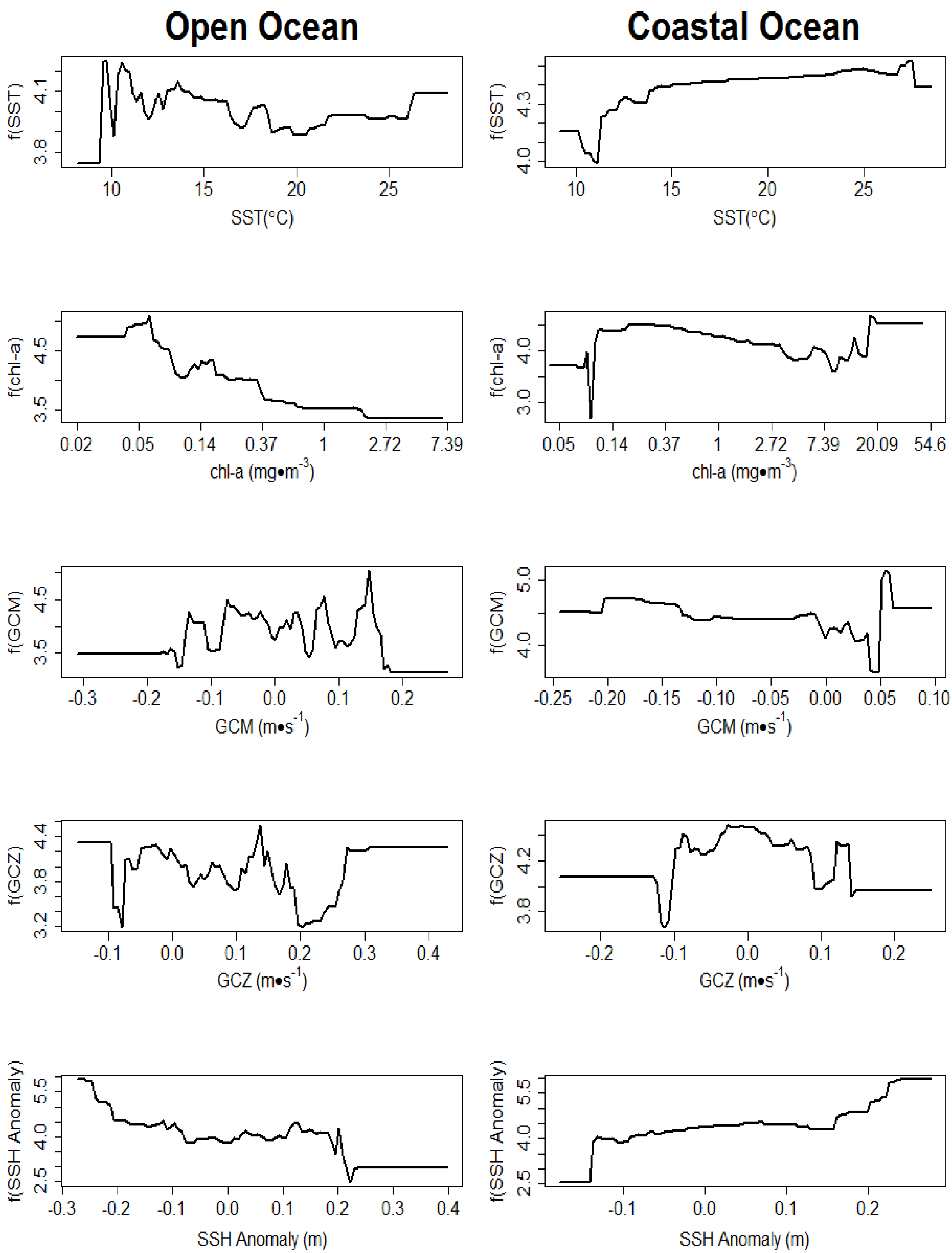


Figure 4. Continued.

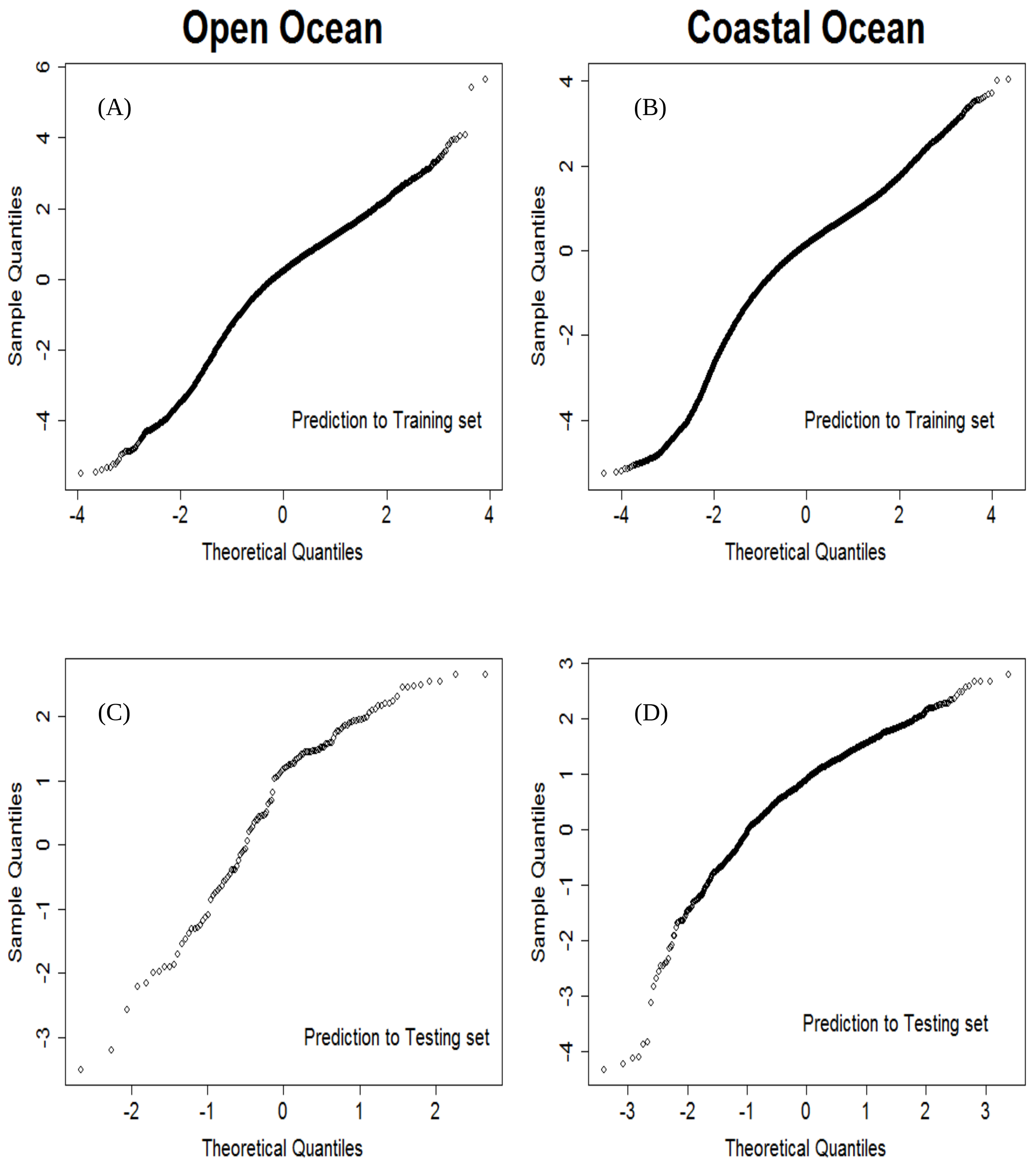


Figure 5. Residuals normal Q-Q plots to training set (1998-2008) and testing set (2009) for the coastal and the open ocean.

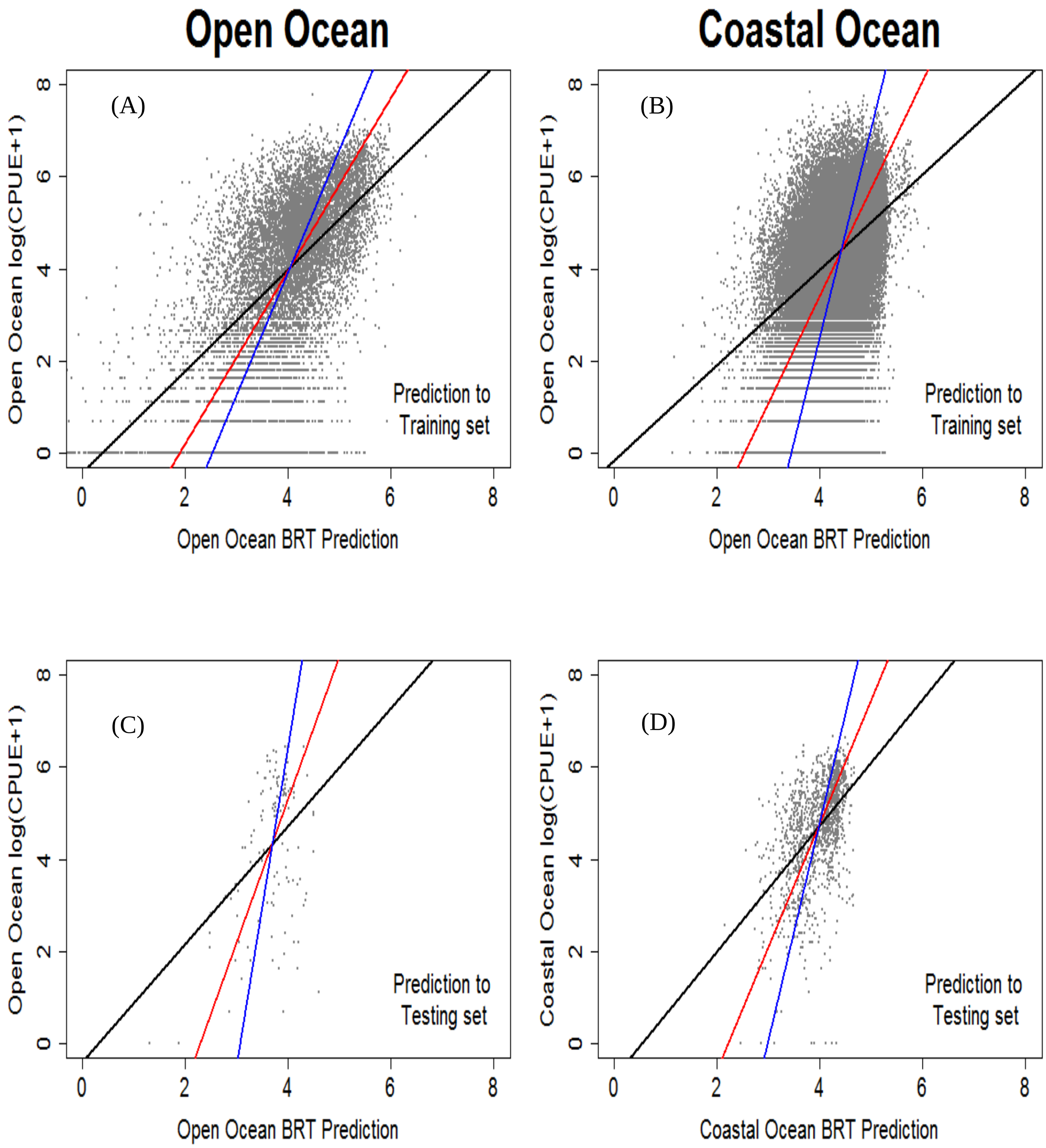


Figure 6. Linear regression(black), geometric mean regression (red), and orthogonal regression (blue) between data and BRT model prediction to training set (1998-2008) and testing set (2009) for the coastal and the open ocean.