# Brown shrimp (*Farfantepenaeus aztecus*) density distribution in the Northern Gulf of Mexico: an approach using boosted regression trees

**JOSE T. MONTERO,**[1]* **TANYA A. CHESNEY,**[2] **JENNIFER R. BAUER,**[2,3,4] **JOHN T. FROESCHKE**[5] **AND JIM GRAHAM**[6]

[1]*Center of Applied Ecology and Sustainability, Facultad de Ciencias Biologicas, Pontificia Universidad Católica de Chile, Avda. Libertador Bernardo O'Higgins 340, Santiago, Chile*

[2]*College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, 104 CEOAS Administration Building, Corvallis, OR 97331-5503, U.S.A.*

[3]*U.S. Department of Energy, National Energy Technology Laboratory, 1450 Queen Avenue SW, Albany, OR 97321*

[4]*AECOM, 1450 Queen Avenue SW, Albany, OR 97321*

[5]*Gulf of Mexico Fishery Management Council, 2203 North Lois Ave. Suite 1100, Tampa, FL 33607, U.S.A.*

[6]*Environmental Science & Management, Humboldt State University, Arcata, CA 95521, U.S.A.*

## ABSTRACT

The estuarine-dependent brown shrimp, *Farfantepenaeus aztecus*, is a significant commercial fishery and important species in the Gulf of Mexico (GOM) ecosystem as well as being a key component in energy transfer between benthic and pelagic food web systems. Because of the economical and ecological importance of brown shrimp, we developed a spatial population model to identify places of high shrimp density under a set of spatial, environmental and temporal variables in the Northern Gulf of Mexico (NGOM). We used fisheries-independent data collected by the Southeast Area Monitoring and Assessment Program (SEAMAP) from 1992 to 2007 (summer and fall seasons). The relationship between the predictor variables and shrimp density was modeled using Boosted Regression Trees (BRT). Within the environmental variables included in the model, bottom type and depth of the water column were the most important predictors of shrimp density in the NGOM. Spatial predictions performed using the trained BRT model for summer and fall seasons showed a spatial segregation of shrimp density. During the summer, higher densities were predicted near the Texas and Louisiana coast and during the fall, higher densities were predicted further offshore. The model performed well and allowed successful prediction of brown shrimp hot spots in the NGOM. Model results allow fisheries managers to evaluate the potential impact from fisheries on the resource and to develop future fisheries management strategies, understand the biology of brown shrimp as well as assess the potential impacts of oil spills or climate change.

**Key words**: boosted regression trees, brown shrimp, Gulf of Mexico, spatial prediction

## INTRODUCTION

The estuarine-dependent brown shrimp, *Farfantepenaeus aztecus*, is a significant commercial fishery and important species in the Gulf of Mexico (GOM) ecosystem. In 2011, the brown shrimp fishery yielded 117.8 million pounds in commercial landings in Alabama, Florida, Louisiana, Mississippi and Texas worth approximately $200.3 million (NOAA NMFS Office of Science and Technology, 2013). In addition to their economic role, brown shrimp are a key component in energy transfer between benthic and pelagic food web systems (Patillo *et al.*, 1997; Daewel *et al.*, 2011). Throughout their lifecycle, they feed on phytoplankton, zooplankton, detritus, benthic organisms, and other organic materials and species (Pérez-Farfante, 1969; Sheridan and Ray, 1981; Louisiana Office of Fisheries, 1992; Patillo *et al.*, 1997). Additionally, throughout their lifecycle, brown shrimp are prey to fish and large crustaceans, many of which are commercially important species (Pérez-Farfante, 1969; Sheridan and Ray, 1981; Louisiana Office of Fisheries, 1992; Patillo *et al.*, 1997). As a result of their economic and ecological value, decades of population dynamics research and management efforts in the GOM have occurred (Haas *et al.*, 2001a); however, the interannual variability of brown shrimp has made it difficult to predict their abundance.

Brown shrimp exhibit an annual life cycle, although individuals have been observed to live up to 2 yr (Pérez-Farfante, 1969). Like other estuarine-dependent species, adult brown shrimp spawn offshore over the shelf, eggs and larvae are transported towards the coast, post-larvae and juveniles grow within the estuaries and late-stage juveniles, or subadults migrate offshore (Etzold and Christmas, 1977; Benfield and Downer, 2001; Haas *et al.*, 2001b). Brown shrimp spawn year-round, but peak spawning is believed to occur from September through to November (Lassuy, 1983) and April through to May (St. Amant *et al.*, 1966; Lassuy, 1983). Peak recruitment into estuaries occurs in February through to March (Li and Clarke, 2005) and August through to September (Saoud and Davis, 2003). As brown shrimp have an annual life cycle, their population size is dependent on the recruitment in the same year (Caillouet *et al.*, 1980). Therefore, shrimp recruitment and abundance are significantly influenced by annual environmental conditions and survival rates as opposed to the success of previous cohorts or historical catch (Caillouet *et al.*, 1980; Minello *et al.*, 1989; Haas *et al.*, 2001a; Li and Clarke, 2005).

In the GOM, research has been conducted to model the relationship between environmental conditions and the spatial distribution and abundance of shrimp for decades. The majority of these studies have been limited spatially and temporally by focusing on small-scale regions, earlier life stages in estuaries and a limited set of environmental conditions. Environmental parameters that have been incorporated into correlative studies include water temperature (Haas *et al.*, 2001a; Clark *et al.*, 2004; Li and Clarke, 2005; Diop *et al.*, 2007), salinity (Haas *et al.*, 2001a; Clark *et al.*, 2004; Diop *et al.*, 2007), dissolved oxygen (Craig *et al.*, 2005; Craig, 2012), depth (Craig *et al.*, 2005), water clarity, precipitation, water level (Haas *et al.*, 2001a), bottom type (Clark *et al.*, 2004) and wetland loss (Diop *et al.*, 2007). However, we believe there is still a gap of knowledge about the environmental and spatial factors that affect the distribution and abundance of brown shrimp at a larger spatial scale in the GOM. Additionally, only a few studies have considered using an empirical model to describe the relationships of brown shrimp on the GOM using an extensive environmental dataset combined with the trawl survey data.

The goal of this study was to use an empirical model that allows us to identify the main spatial and environmental variables that affect the spatial and temporal distribution of observed density hot spots of brown shrimp in the Northern Gulf of Mexico (NGOM). We estimated the relative density of shrimp using catch per unit effort (CPUE) and a model of the relationship of shrimp density with environmental conditions using boosted regression trees (BRT; Elith *et al.*, 2008; Froeschke *et al.*, 2010). The main objective was to identify an empirical model that detects under which environmental condition the brown shrimp population will be denser in the NGOM.

## METHODS

### Data

Monitoring brown shrimp abundance and density requires a proxy of measurement with minimum variability in fishing effort as a result of gear modifications and changes in management efforts. This study utilizes fishery-independent trawl survey and oceanographic data collected by the Southeast Area Monitoring and Assessment Program (SEAMAP) Gulf of Mexico to model and predict the density of brown shrimp across the NGOM because of the datasets extensive spatial and temporal coverage and consistent sampling procedures (Ocean Studies Board, N.R.C., 2000). SEAMAP has conducted fishery-independent trawl surveys targeting shrimp and groundfish since 1983 across Florida, Alabama, Mississippi, Louisiana and Texas between 88°W and 97°W longitude (Hart, 2011) (Fig. 1). From 1992 to 2007, surveys occurring in the summer (June–August) and fall (October–December) used similar gears, protocols and random stratified sampling design based on depth and statistical area (Craig *et al.*, 2005). Trawl surveys performed by the National Marine Fisheries Service Center, Alabama, Mississippi, and Louisiana were made perpendicular to the depth strata from 5 to 60 fathoms for a maximum of 55 min using 40 ft nets and parallel to the depth strata for 10 min using 20 ft nets by Texas (Gulf States Marine Fisheries Commission, 1983–2010). At the start of each trawl, bottom measurements of salinity and dissolved oxygen (DO), sea surface temperature (SST) and depth were collected and upon completion, all species collected were identified and counted, including brown shrimp.

### Density of brown shrimp

Assuming that fishing effort in the SEAMAP dataset was standardized as a result of the consistent sampling protocols (Craig *et al.*, 2005), the brown shrimp density was calculated in our study area (Fig. 1) based on CPUE, which is used extensively as an index for the relative abundance of fisheries research (Haas *et al.*, 2001a). To calculate the density of brown shrimp, data were binned into a 0.1 by 0.1-degree grid, with each cell representing an area of approximately 121 km$^2$,

**Figure 1.** A figure of our study area in the Northern Gulf of Mexico (NGOM) and the Southeast Area Monitoring and Assessment Program (SEAMAP) statistical regions where Brown shrimp trawls were made. The grey line represents the 200 m isobath contour.
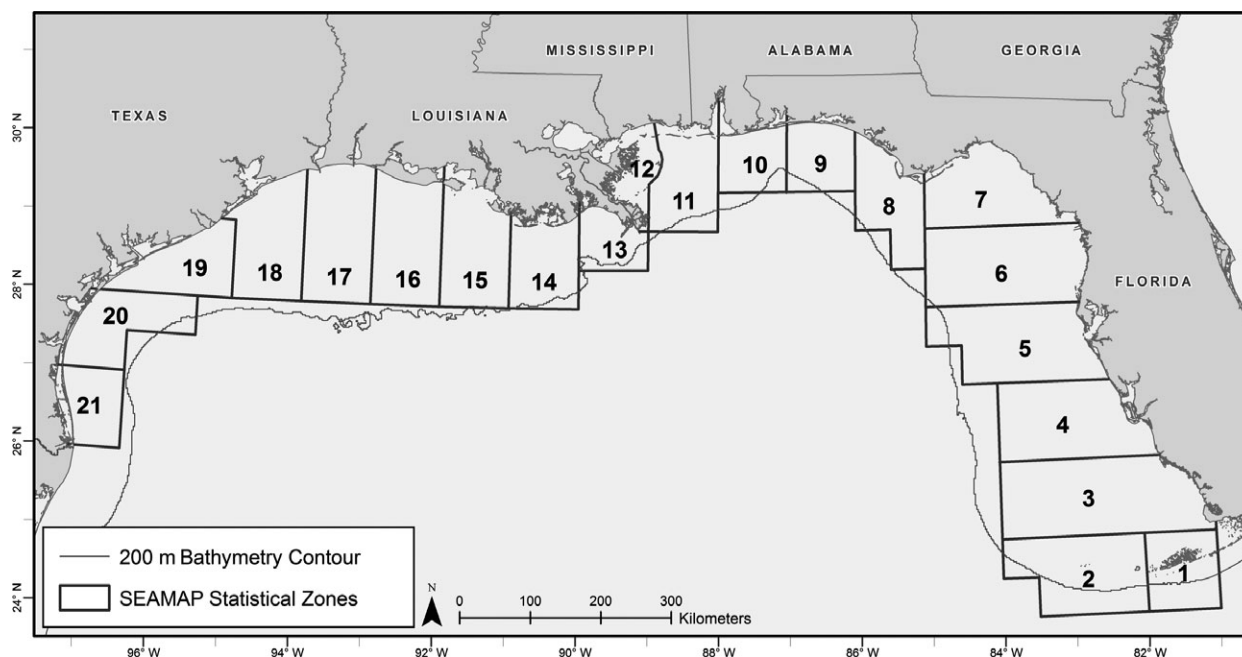


across the entire study area. Then, CPUE was calculated for each grid cell per summer and fall months using the equation:

$$CPUE_t = C_t/E_t$$

where $C_t$ represents the catch at time $t$ and $E_t$ the fishing effort, by taking the total number of brown shrimp for all trawls in each grid cell for each summer and fall months ($C_t$) divided by the total number of hours for all trawls within the same grid cell per month ($E_t$). To reduce skewness of the distribution for calculating brown shrimp density, density was log-transformed as Log(density+1) (Legendre and Legendre, 1998; Lauria *et al.*, 2011). The density of brown shrimp was then used to produce a map of the overall annual mean density of brown shrimp between 1988 and 2007 throughout the NGOM and as the response variable for our modeling approach using SEAMAP data from 1992 to 2007. For modeling purposes, we left out the years 1983–1991 and 2008–2011 because of changes in the sampling techniques. The data left out (1983–1991 and 2008–2011) of the modeling process was later used to generate brown shrimp log–CPUE prediction maps.

*Predictor variables*

For modeling brown shrimp density, SST, bottom DO, bottom salinity, depth, the percentage of mud on the bottom (Mud %), longitude, latitude and season (summer & fall) were used as environmental and temporal predictor variables (Table 1). SST, bottom DO, bottom salinity and depth were measured during each SEAMAP survey. Mud % was determined from over 2 30 000 bottom-type samples within the GOM, provided by the Institute of Arctic and Alpine Research (INSTAAR), that identified the proportion of mud, clay, silt and sand in each sample (Jenkins, 2010; Graham *et al.*, 2012). The environmental predictor variables were averaged for all measurements taken in each grid cell for the summer and fall months of each year.

*Boosted regression trees*

Booted regression trees (BRTs) are a combination of the statistical techniques boosting and regression trees. Boosting is a machine learning method, where the results of several competing models are merged. Through boosting, a stage-wise procedure fits tree models iteratively to a subset of the data that is being modeled, where the subsets are randomly selected without replacement. This procedure is known as stochastic gradient boosting and introduces a stochasticity element that improves model accuracy and reduces overfitting (Elith *et al.*, 2008).

Like other model-averaging methods, BRTs differ fundamentally from the more conventional regression-based techniques such as generalized additive models

**Table 1.** Predictor variables used in the analyses.

| Variable (Unit) | Description | Mean | Range |
|---|---|---|---|
| Latitude (degrees North) | Latitude of the observed CPUE | 28.5 | 24.6–30.2 |
| Longitude (degrees West) | Longitude of the observed CPUE | 92.8 | 97.3–81.5 |
| Mud (%) | Mud percentage provided by INSTAAR | 64.2 | 0–100 |
| Depth (m) | Depth of the water column where sample occurred | −42.6 | 0–400 |
| Salinity (ppt) | Salinity measured during sampling | 34.4 | 0–40 |
| SST (°C) | Sea surface temperatures measured during sampling | 26.7 | 15–32 |
| DO (mg $O_2$ $L^{-1}$) | Dissolved oxygen concentration measured during sampling | 5.1 | 0–15 |
| Season (summer & fall) | Seasons derived from months when sample occurred | NA | NA |

CPUE, catch per unit effort; SST, sea surface temperature; DO, dissolved oxygen; NA, not applicable.

(GAMs) (Hastie and Tibshirani, 1990). GAMs fit a single most parsimonious model that best describes the relationship between a response variable and a set of predictors, whereas BRTs fit a large number of simple models, which predictions are later combined to give a more robust estimate of the response (Elith *et al.*, 2008). BRTs can be fit with a variety of predictor variables, are immune to extreme outliers and are flexible when fitting interactions between variables (Friedman and Meulman, 2003).

Only recently, BRTs have been applied to answer ecological questions (Friedman, 2001; Leathwick *et al.*, 2006; Elith *et al.*, 2008; Froeschke *et al.*, 2010; Froeschke and Froeschke, 2011; Martínez-Rincón *et al.*, 2012). Because of the ability of BRTs to fit a model with interactions and automatically select important variables, as well as its robustness to outliers and missing data, BRT models are growing in popularity (Froeschke and Drymon, 2013). BRTs can fit complex non-linear relationships and in many cases have a superior predictive performance than Generalized Linear Models (GLMs) and GAMs, which are commonly used to develop standardized abundance indices (Lo *et al.*, 1992). Generally speaking, BRTs have been shown to be more robust than GAMs and GLMs and can be implemented to answer the same type of ecological questions.

BRT model fitting requires the specification of three parameters: (i) learning rate, which controls the rate at which model complexity increases, (ii) bag fraction and (iii) tree complexity which represents the number of splits in each tree and controls the size of the trees, as well as the complexity of interaction within predictor variables. A tree complexity value of one corresponds to an additive model with no interaction between the predictor variables where each tree has a single node or decision way. In contrast, a value of two or higher indicates that more than one node is used in each tree, which represents a model with two or more interaction ways (Elith *et al.*, 2008).

For our analyses, the brown shrimp data from 1992 to 2007 were randomly divided into two datasets, a training set (70% of the data) and testing set (30% of the data), using a stratified random sampling based on seasons (Summer and Fall) to reduce the bias. The training set was then used to fit the BRT model whereas the test set was used to evaluate model performance. We fitted a BRT model using R (R Development Core Team, 2012) version 2.15.2 with the 'gbm' and 'dismo' packages (Ridgeway, 2010; Hijmans *et al.*, 2011). The 'gbm' library implements the formulae developed by Friedman (2001) to estimate the relative influence of the predictor variable over the response. These values are based on the number of times a variable is selected for splitting, weighted by the squared improvements to the model based on each split, and averaged over all trees (Friedman and Meulman, 2003). Adding its sum to 100, the relative contribution of each variable was scaled so that a higher number corresponds to a stronger influence on the response.

*Relationship between response and predictor variables*

The relationship between brown shrimp density and each predictor variable was explored by plotting the partial dependence plots using the *gbm.plot* function in R from the *gbm* package. These plots show the effect of a variable over shrimp density after accounting for the overall effect of all other variables in the model (Friedman, 2001). The predicted values for each variable of the partial dependence plots are calculated by holding the values of all other variables at their mean. The partial dependence plots also indicate how shrimp density partially depends on each predictor variable.

*Model fitting and evaluation*

Before fitting a final BRT, we went through several steps to find the best candidate model (see Supporting Information). We first fitted a BRT model without interactions among variables (tree complexity of 1), and recorded the performance parameters: residual

deviance, mean residual deviance, and pseudo R$^2$ (or D$^2$), calculated as D$^2$ = 1-(mean residual deviance/ mean total deviance). The pseudo R$^2$ was used as a measure of model fitting and performance for the trained model. The same procedure was performed for a model with tree complexity 2 and 5, to evaluate the performance of a model with multiple interactions. We repeated the same procedure using Generalized Additive Models (GAMs). Although a model comparison was not the goal of our research, we wanted to verify that the BRT approach performs better than the widely used GAMs because BRTs have only recently been used in the ecological and fisheries literature (see Supporting Information for model comparison statistics). We also calculated R$^2$ for the testing dataset contrasting the predicted and observed values for all models (See Supporting Information). We finally selected as the best model a BRT with a tree complexity of 5.

Additional to the general model statistics and goodness of fit tests, we built omnidirectional semi-variograms to investigate spatial autocorrelation and the ability of the BRT and GAM models. We first plotted the semi-variogram of the train set CPUE and then compared it with the residuals semi-variogram for the different candidate models. This analysis allowed us to identify how the model handles spatial autocorrelation if present (see Supporting Information).

We used several statistics using the testing data to look at the model performance of the final trained BRT model in terms of accuracy and bias using the testing data set. To look at model accuracy, we calculated the proportion of variation explained in the outcome of the testing data comparing the observations (test data set, log-CPUE) versus predictions as 1-(SSR/SST), as well as the Root Mean Square Error (RMSE). We also performed a Spearman's rank correlation coefficient ($r_s$) and calculated its respective P-value. We used the Spearman's $r_s$ because it does not assume linearity. In the Spearman's correlation test, if $P < 0.05$ and $r_s > 0.1$, the test for model accuracy is considered passed (Lauria et al., 2011). To assess model bias in the predictions, we used Wilcoxon's signed-rank test to compare observed versus predicted values deviations. The Wilcoxon signed-rank test compares the median observed densities of brown shrimp with median predicted densities. The null hypothesis for the sign test is that the observations and predictions are unbiased, therefore, the model was considered passed if $P > 0.05$.

## Mapping model predictions

The trained model was used to predict the mean density of brown shrimp for summer and fall seasons with SEAMAP survey data excluded from our initial modeling process (1983–1991 and 2008–2011) which contains measurements of all the predictor variables. Maps of the mean observed model prediction output were created in ArcMap 10.1 (Esri, 2012). Additionally, we performed a spatial interpolation over the CPUE predicted surface within a 200-m isobath by fitting a Thin Spline Regression Model to fill the gaps where data were not available to predict shrimp density distribution. This method has been previously used by Martínez-Rincón et al. (2012) and Montero et al. (2016) to fill in gaps in the predictions when there is missing data.

## RESULTS

Based on the SEAMAP dataset, two spatial trends were observed in the overall mean density of brown shrimp throughout the NGOM between 1992 and 2007 (Fig. 2). Overall, the higher densities were located in the western region of the Northern GOM between the 20–60 m isobaths whereas, lower densities of brown shrimp were observed in the GOM region east of Mississippi, between 82 and 88°W.

### Boosted regression trees

The final BRT model was constructed with a learning rate of 0.05, 600 trees and tree complexity of 5 to allow multiple interactions and a larger number of splits (Table 2 and Table S2). The error was assumed to fit a Gaussian distribution. Figure 3 summarizes the relative contribution of each predictor variable in the model. The most important predictor variable for brown shrimp density was Mud % which contributed 20% of the overall model response showing higher predicted densities of brown shrimp corresponding with regions that had a higher percentage of mud. The second most important predictor variable was depth (Bathymetry) with a relative contribution of 19.3% with a greater density of brown shrimp predicted at depths between 20 and 100 m. Longitude, bottom salinity, SST, and bottom DO, and Latitude has less influence on the model, with contributions of 12.5, 12.1, 11.1 and 7.3%, respectively, with a higher density of brown shrimp predicted on waters of higher longitude, with bottom salinity between 10 and 20 ppt, SST between 23 and 30°C, and bottom DO from 2 to 7 ppm, and Latitude between 28 and 30°N. Seasonality did not significantly affect the brown shrimp density. Spatially, the model shows a decrease of brown shrimp density. In contrast, seasons (summer and fall) had a minimal contribution to the overall model (0.6%, Fig. 3). The partial dependence plots (Fig. 4) depict the relationship between brown

**Figure 2.** Mean spatial distribution of observed mean brown shrimp density for summer and fall between the years 1992-2007. The black line represents the 200 m isobath contour. Overall, higher densities were located in the western region of the Northern GOM between 20–60 m depth. The filled contours were made using natural breaks classification (Jenks) in ArcMap 10.1.
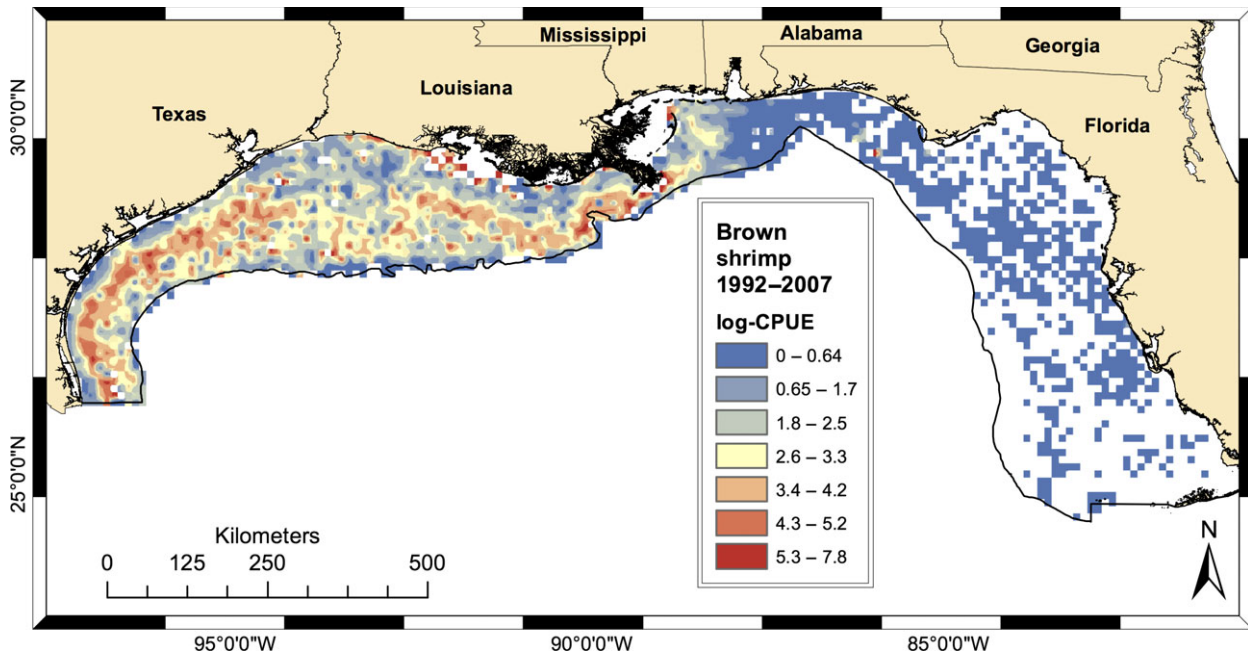


**Table 2.** Summary statistics of the candidate Boosted Regression Trees (BRT) models for brown shrimp.

| Model | Model complexity | Number of trees | Residual deviance | Explained deviance (D2) | RMSE | 1-(SSE/ SST) |
|---|---|---|---|---|---|---|
| BRT (1) | 1 | 3050 | 24 527.75 | 0.27 | 2.07 | 0.22 |
| BRT (2) | 2 | 2150 | 20 313.46 | 0.39 | 2 | 0.27 |
| BRT* (5) | 5 | 600 | 19 381.39 | 0.45 | 1.98 | 0.29 |

RMSE, Root Mean Square error, 1-(SSE/SST), Adjusted R2.
*NOTE, The model BRT-5 was the final model selected for training and prediction.

shrimp density and each predictor variable explained by the BRT model.

*Model performance*

Model evaluation, based on the Spearman's rank correlation test for our BRT model, showed a significant positive correlation between observed and predicted values: P-value < 0.001 and $r_s$ = 0.54, suggesting a fair accuracy in the model predictions. Also, Wilcoxon's signed rank test was considered successful, with a P-value >> 0.05. In contrast, when we looked at the proportion of variation explained in the outcome of the testing data, this only explains ~30% of the variation, which suggests a low precision in model predictions (Table 2). These results imply that the BRT model

was fairly successful in describing the density trends of brown shrimp and the potential factors affecting its distribution. However, the model has a weakness in predicting with precision the overall mean density of brown shrimp in the environment.

*Spatial distribution of brown shrimp density*

The predicted brown shrimp mean densities trend showed variation between the two seasons (Fig. 5a,b). Greater densities were predicted during the summer season compared with the fall season. In addition, the spatial distribution of brown shrimp varied between seasons. During the summer, higher density values were predicted closer to inshore waters in the western region of the NGOM near the Texas and Louisiana

**Figure 3.** Graphical representation of the relative contribution of each predictor variable for brown shrimp density by the BRT model. Y-axis shows the predictor variables and X-axis depicts the relative contribution of each predictor variable on a scale of 0-100%. The most important predictor variable for brown shrimp density was Mud %, contributing 20% of the overall model response.
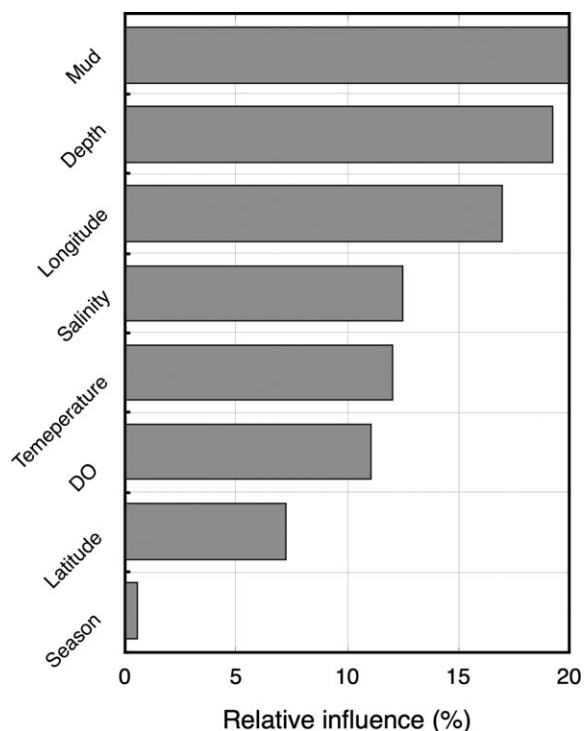


coast, between 26–29°N and 94.5–97°W. In contrast, during the fall, higher brown shrimp densities were predicted further offshore, with greater densities near Louisiana between 28–23°N and 89–92°W. Very low densities of brown shrimp were predicted for the region of the NGOM east of Mississippi in both the summer and fall.

## DISCUSSION

The results of this study suggest a significant relationship of brown shrimp density with the environmental, spatial and temporal variables used for the BRT model. The model suggested that the bottom-type composition (Mud %) and depth of the water column as the strongest predictors of brown shrimp density in the NGOM. The predicted spatial distribution and density of brown shrimp were consistent with the spatial pattern and brown shrimp densities observed in the NGOM based on the SEAMAP dataset corroborating the performance of the BRT model to accurately predict the spatial distribution of brown shrimp density using the selected predictor variables.

Our results suggest that the variation in brown shrimp density in the NGOM is mostly as a result of Mud %. This is consistent with previous research conducted in estuaries in the GOM, indicating that brown shrimp prefer soft bottom substrates because they are rich in food material (Williams, 1955, 1959; Van Lopik *et al.*, 1979; Turner and Brody, 1983). Furthermore, research suggests that although juvenile and adult brown shrimp are found on sand, silt, clay and shell bottoms, a greater abundance of brown shrimp are located in areas with mud bottoms (Pérez-Farfante, 1969; Etzold and Christmas, 1977; Lassuy, 1983). In regard to the spatial distribution of brown shrimp densities, the most distinct trend observed in our results was the presence of a distinct 'line' separating areas of a relative high density and areas of low-density present in both observed and predicted brown shrimp densities around the 88°W meridian. This invisible spatial divider of brown shrimp can be attributed to distinct regional differences in bottom type, with regions in the NGOM west of the 88°W meridian characterized predominantly by muddy clay-silts and muddy sands, whereas regions east of the 88°W meridian are predominantly sand, gravel and shell (Wilkinson *et al.*, 2009).

Depth also plays a critical role in determining brown shrimp density, with our results indicating that brown shrimp are most abundant in depths between 20 and 100 m. This is consistent with previous research in which brown shrimp were distributed broadly on the GOM shelf in areas up to 110 m, with highest densities occurring between 20 and 40 m from June to July (Zein-Eldin and Renaud, 1986; Craig *et al.*, 2005). The spatial distribution of the observed and predicted mean density of brown shrimp, despite seasonal differences, supported these results. The slight seasonal variations observed between the summer and fall months predicted densities might be the result of different stages of their life cycle, which have been strongly correlated to depth (Pérez-Farfante, 1969; Renfro and Brusher, 1982). Increased densities of brown shrimp during the summer months closer to the Texas and Louisiana shoreline might be attributed to increased emigration of late juveniles from the shallow estuaries, which are prolific along the Texas and Louisiana coast, into the GOM during June and July (Zein-Eldin and Renaud, 1986). Higher densities observed offshore in the fall corresponds with brown shrimp's peak spawning season from September to November in waters between 27 and 110 m (Renfro and Brusher, 1982), which overlaps with SEAMAP's

**Figure 4.** Partial dependence plots used to explore the effect of each predictor variable, mud %, depth, longitude, bottom salinity, SST, bottom DO, latitude and season (summer & fall) over shrimp density used in the BRT model. Y-axes show the centered scale of the data distribution. The X-axes represent the range of the eight predictor variables and their relative contribution. The shaded region represents the confidence interval.
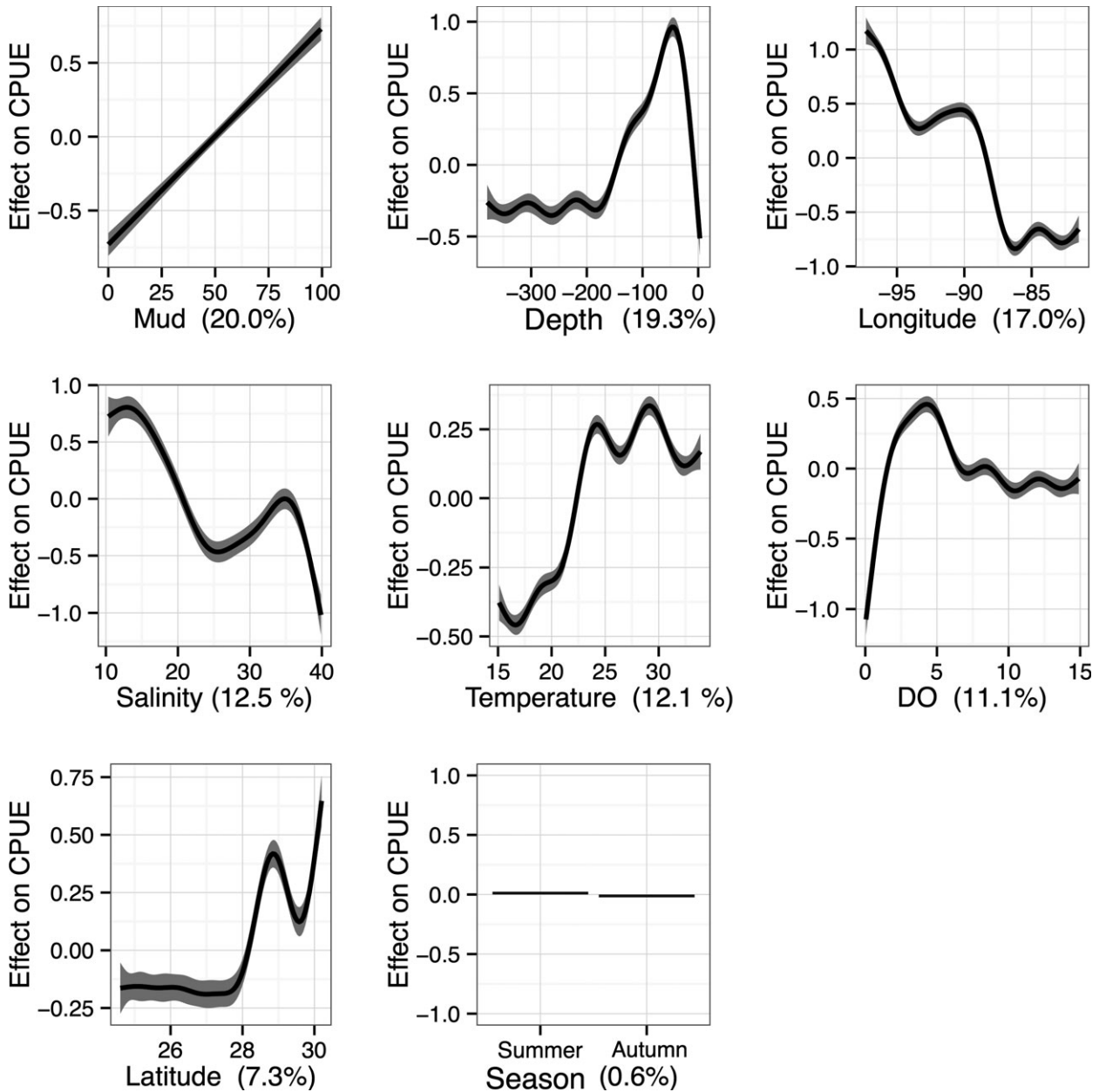


fall sampling season. Furthermore, this aggregation between 27 and 100 m depths for spawning could explain the slightly higher densities observed off Eastern Louisiana and Mississippi as the slope of the continental shelf is steeper off Eastern Louisiana and Mississippi and a more gradual slope off Western Louisiana and Texas.

Based on our results, bottom salinity and SST contribute to brown shrimp density in the NGOM. Salinity and SST have been known to affect brown shrimp physiology, migration and abundance (Zein-Eldin and Aldrich, 1965; Lassuy, 1983; Haas *et al.*, 2001a; Saoud and Davis, 2003; Diop *et al.*, 2007; Piazza *et al.*, 2010; Rozas and Minello, 2011). Previous research found

**Figure 5.** Spatial distribution of predicted mean brown shrimp density (log-CPUE) using the trained BRT-5 model. (a) Shows the predictions made for Summer season and (b) for Autumn (fall). *NOTE: These maps were made averaging the predicted CPUE by latitude and longitude for all the years and months included in the study to identify the major trends in the prediction of brown shrimp density. The filled contours were made using natural breaks classification (Jenks) in ArcMap 10.1.
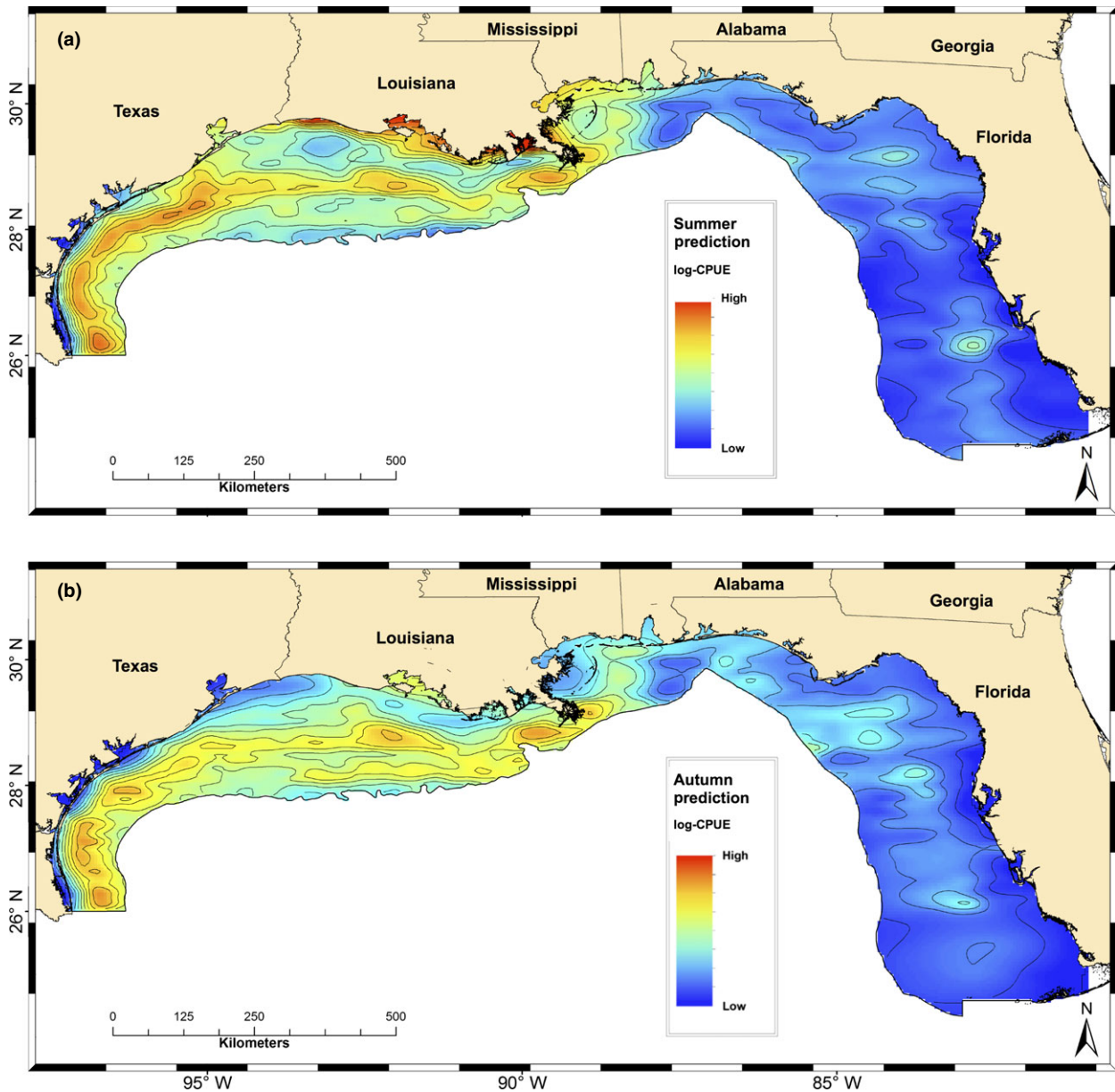


that there is a positive correlation between temperature and salinity for brown shrimp, indicating that high river discharge and climate variability play a significant role in the success of juvenile and adult brown shrimp (Haas *et al.*, 2001a; Saoud and Davis, 2003; Diop *et al.*, 2007; Piazza *et al.*, 2010). The variation in observed densities of brown shrimp in salinity ranges between 10 and 20 and 30 and 40 ppt in our results could be an indication of regional differences in freshwater input. Areas of high brown shrimp densities with bottom salinities between 10 and 20 ppt were observed closer to estuaries and major river outputs along Eastern Texas and Western Louisiana, whereas regions of high density with bottom salinities between 30 and 40 ppt occurred further offshore.

Brown shrimp are sensitive to low oxygen conditions. We observed greater densities of brown shrimp in locations where the bottom DO was between 2

and 7 ppm. Brown shrimp have been found to detect and avoid hypoxic waters where dissolved oxygen concentrations range less than or equal to 2.0 ppm (Renaud, 1986; Zimmerman, 2003; Craig, 2012). Low DO avoidance affects brown shrimp emigration and leads to an alteration in trophic interactions, density and distribution patterns (Zimmerman, 2003; Craig, 2012). An increase in hypoxic conditions on the Louisiana shelf has led to shifts in brown shrimp distribution in the NGOM and has been found to affect brown shrimp catchability (Craig, 2012). This is consistent with our findings in which brown shrimp densities were greatest near the Texas shelf suggesting migration to Texas waters during summer months when hypoxic conditions are frequent in the Louisiana shelf (Renaud, 1986; Zimmerman, 2003; Craig, 2012).

*Uncertainty*

While this study was partially successful in modeling the density of brown shrimp and matches expected outcomes, there were sources of uncertainty, especially in the data used (Szuwalski and Punt, 2012) and the model. The sampling techniques of fisheries-independent data can result in both temporal and spatial biases in the data. Although SEAMAP performs surveys over a short-time period, mostly from June to July, species distributions change over time both vertically and horizontally (Ocean Studies Board, 2000). Additionally, the total trawl catch from the SEAMAP dataset containing no catch located near trawls with large catches, indicating that brown shrimp may be clustered, which would reduce the overall model performance. Each of the environmental predictors provided by SEAMAP and INSTAAR would also have a spatial and measurement error. Averaging the datasets together and modeling with relatively large cell sizes would reduce some of this.

The statistics of the trained models and test set suggested a fair predictive power of spatial distribution of brown shrimp density in the GOM. Even although the goal of this study was to predict, as close as possible, the values observed in the environment, our model output informs under which environmental condition brown shrimp will be more abundant in the GOM, regardless of the mismatch between predicted and real values. We think the model is not fully capable of predicting brown shrimp density in the GOM. However, it gave us an insight of where, when and under which environmental and spatial condition is more probable to find higher densities of brown shrimp during fishing operations.

## CONCLUSION

The results of our study provided broader-scale analyses and spatial predictions of brown shrimp based on environmental and temporal parameters which have significant implications for fisheries management. The capabilities of predicting the spatial distribution of brown shrimp density based on environmental predictors, many of which are readily available through previous research or remotely sensed data, can help fisheries managers evaluate potential impacts to the fishery from sources including alternative fisheries management strategies, oil spills, extreme weather events, or climate change.

## REFERENCES

Benfield, M.C. and Downer, R.G. (2001) Spatial and temporal variability in the nearshore distributions of postlarval Farfantepenaeus aztecus along Galveston Island, Texas. *Estuar. Coast. Shelf Sci.* **52:**445–456.

Caillouet, C.W., Patella, F.J. and Jackson, W.B. (1980) Trends toward decreasing size of brown shrimp, *Penaeus aztecus*, and white shrimp, *Penaeus setiferus*, in reported annual catches from Texas and Louisiana. *Fish. Bull.* **77:**985–989.

Clark, R.D., Christensen, J.D., Monaco, M.E., Caldwell, P.A., Matthews, G.A. and Minello, T.J. (2004) A habitat-use model to determine essential fish habitat for juvenile brown shrimp (Farfantepenaeus aztecus) in Galveston Bay, Texas. *Fish. Bull.* **102:**264–277.

Craig, J.K. (2012) Aggregation on the edge: effects of hypoxia avoidance on the spatial distribution of brown shrimp and

demersal fishes in the Northern Gulf of Mexico. *Mar. Ecol. Prog. Ser.*, **445:**75–95.

Craig, J.K., Crowder, L.B. and Henwood, T.A. (2005) Spatial distribution of brown shrimp (Farfantepenaeus aztecus) on the northwestern Gulf of Mexico shelf: effects of abundance and hypoxia. *Can. J. Fish Aquat. Sci.* **62:**1295–1308.

Daewel, U., Schrum, C. and Temming, A. (2011) Towards a more complete understanding of the life cycle of brown shrimp (Crangon crangon): modelling passive larvae and juvenile transport in combination with physically forced vertical juvenile migration. *Fish Oceanogr.* **20:**479–496.

Diop, H., Keithly, W.R., Kazmierczak, R.F. and Shaw, R.F. (2007) Predicting the abundance of white shrimp (Litopenaeus setiferus) from environmental parameters and previous life stages. *Fish. Res.* **86:**31–41.

Elith, J., Leathwick, J.R. and Hastie, T. (2008) A working guide to boosted regression trees. *J. Anim. Ecol.* **77:**802–813.

Esri (2012) ArcGIS 10.1. Redlands, CA: Environmental Systems Research Institute.

Etzold, D.J. and Christmas, J.Y. (1977) A Comprehensive Summary of the Shrimp Fishery of the Gulf of Mexico United States: A Regional Management Plan. Gulf Coast Research Laboratory Technical Report Series **2:**20.

Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29:**1189–1232.

Friedman, J.H. and Meulman, J.J. (2003) Multiple additive regression trees with application in epidemiology. *Stat. Med.* **22:**1365–1381.

Froeschke, J. and Drymon, M. (2013) Atlantic sharpnose shark: standardized index of relative abundance using boosted regression trees and generalized linear models. In: SEDAR34-WP-12. North Charleston, SC: SEDAR, 31 pp.

Froeschke, J.T. and Froeschke, B.F. (2011) Spatio-temporal predicitive model based on environmental factors for juvenilen spotted seatrout in Texas estuaries using boosted regression trees. *Fish. Res.* **111:**131–138.

Froeschke, J., Stunz, G.W. and Wildhaber, M.L. (2010) Environmental influences on the occurrence of coastal sharks in estuarine waters. *Mar. Ecol. Prog. Ser.* **407:**279–292.

Graham, J., Rose, K., Bauer, J. *et al.* (2012) Integration of spatial data to support risk and impact assessments for deep and ultra-deepwater hydrocarbon activities in the Gulf of Mexico. In: ICES Document NETL-TRS-4-2012; EPAct Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV; 36 pp.

Gulf States Marine Fisheries Commission (1983-2010) SEAMAP Environmental and Biological Atlas of the Gulf of Mexico. Gulf States Marine Fisheries Commission, Ocean Springs, MS.

Haas, H.L., Lamon, E.C., Rose, K.A. and Shaw, R.F. (2001a) Environmental and biological factors associated with the stage-specific abundance of brown shrimp (Penaeus aztecus) in Louisiana: applying a new combination of statistical techniques to long-term monitoring data. *Can. J. Fish Aquat. Sci.* **58:**2258–2270.

Haas, H.L., Shaw, R.F., Rose, K.A., Benfield, M.C. and Keithly, W.R. (2001b) Regression analysis of the relationships among life-stage abundances of brown shrimp (*Penaeus aztecus*) and envrionmental variables in southern Louisiana, USA. *Proc. Gulf. Caribb. Fish. Inst.* **52:**231–241.

Hart, R.A. (2011) Stock Assessment of Brown Shrimp (Farfantepenaeus aztecus) in the U.S. Gulf of Mexico.

Galveston, TX: NOAA Fisheries Southeast Fisheries Science Center.

Hastie, T. and Tibshirani, R. (1990) Generalized Additive Models. Volume 43 of Monographs on Statistics and Applied Probability: Chapman & Hall/CRC.

Hijmans, R.J., Phillips, S., Leathwick, J. and Elith, J. (2011) dismo: Species Distribution Modeling. R Package Version 0.6-3. University of California-Davis, Davis, CA.

Jenkins, C.J. (2010) Seafloor Substrates. Boulder, CO: INSTAAR, University of Colorado.

Lassuy, D.R. (1983) Species profiles: life histories and environmental requirements of coastal fishes and invertebrates (Gulf of Mexico)–brown shrimp. In: U.S. Fish and Wildlife Service Biological Report FWS/OBS 82(11.1). 15pp.

Lauria, V., Vaz, S., Martin, C.S., Mackinson, S. and Carpentier, A. (2011) What influences European plaice (Pleuronectes platessa) distribution in the eastern English Channel? Using habitat modelling and GIS to predict habitat utilization. *ICES J. Mar. Sci.* **68:**1500–1510.

Leathwick, J.R., Elith, J., Francis, M.P., Hastie, T. and Taylor, P. (2006) Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Mar. Ecol. Prog. Ser.* **321:**267–281.

Legendre, P. and Legendre, L. (1998) Numerical Ecology. Amsterdam: Elseveir.

Li, J. and Clarke, A.J. (2005) Sea surface temperature and the brown shrimp (Farfantepenaeus aztecus) population on the Alabama, Mississippi, Louisiana and Texas continental shelves. *Estuar. Coast. Shelf Sci.* **64:**261–266.

Lo, N.C.H., Jacobson, L.D. and Squire, J.L. (1992) Indices of relative abundance from fish spotter data based on delta-lognormal models. *Can. J. Fish. Aquat. Sci.* **49:**2515–2526.

Louisiana Office of Fisheries (1992) A Fisheries Management Plan for Louisiana's Penaeid Shrimp Fishery. Baton Rouge, LA: Louisiana Department of Wildlife and Fisheries Office of Fisheries.

Martínez-Rincón, R.O., Ortega-García, S. and Vaca-Rodríguez, J.G. (2012) Comparative performance of generalized additive models and boosted regression trees for statistical modeling of incidental catch of wahoo (Acanthocybium solandri) in the Mexican tuna purse-seine fishery. *Ecol. Model.* **233:**20–25.

Minello, T.J., Zimmerman, R.J. and Martinez, E.X. (1989) Mortality of young brown shrimp Penaeus aztecus in estuarine nurseries. *Trans. Am. Fish. Soc.* **118:**693–708.

Montero, J.T., Martinez-Rincon, R.O., Heppell, S.S., Hall, M. and Ewal, M. (2016) Characterizing enviromental and spatial variables associated with the incidental catch of olive ridley (*lepidochelis olivacea*) in the Eastern Tropical Pacific puse-seine fishery. *Fish. Oceanogr.* **25:**1–14.

NOAA NMFS Office of Science and Technology (2013) Annual Commercial Landing Statistics. NOAA NMFS Office of Science and Technology: Silver Spring, MD.

Ocean Studies Board, N.R.C. (2000) "General issues in the collection, MANAGEMENT, and use of fisheries data." In: Improving the Collection, Management, and Use of Marine Fisheries Data. Washington: The National Academies Press, pp. 59–130.

Patillo, M.E., Czapla, T.E., Nelson, D.M. and Monaco, M.E. (1997) Distribution and abundance of fishes and invertebrates in Gulf of Mexico estuaries. Volume II: species life history summaries. In: ELMR Report No. 11. Silver

Spring: NOAA/NOS Strategic Environmental Assessments Division, pp. 377.

Pérez-Farfante, I. (1969) Western Atlantic shrimps of the genus *Penaeus*. *Fish. Bull.* **67:**461–591.

Piazza, B.P., La Peyre, M.K. and Keim, B.D. (2010) Relating large-scale climate variability to local species abundance: ENSO forcing and shrimp in Breton Sound, Louisiana, USA. *Climate Res.* (Open Access for articles 4 years old and older) **42:**195.

R Development Core Team (2012) R: A Language and Environment for Statistical Computing, 2.15.2 edn. Vienna: R Foundation for Statistical Computing.

Renaud, M.L. (1986) Detecting and avoiding oxygen deficient sea-water by brown shrimp, *Penaeus aztecus* (Ives), and white shrimp *Penaeus setiferus* (Linnaeus). *J. Exp. Mar. Biol. Ecol.* **98:**283–292.

Renfro, W.C. and Brusher, H.A. (1982) Seasonal Abundance, Size Distribution, and Spawning of Three Shrimps (Penaeus aztecus, P. setiferus, and P. duorarum) in the Northwestern Gulf of Mexico, 1961-1962, NOAA NMFS Southeast Fisheries Center, Galveston, TX.

Ridgeway, G. (2010) GBM: Generalized Boosted Regression Models. R package version 1.6-3.1. http://CRAN.R-project.org/package=gbm

Rozas, L.P. and Minello, T.J. (2011) Variation in penaeid shrimp growth rates along an estuarine salinity gradient: implications for managing river diversions. *J. Exp. Mar. Biol. Ecol.* **397:**196–207.

Saoud, I.P. and Davis, D.A. (2003) Salinity tolerance of brown shrimp Farfantepenaeus aztecus as it relates to postlarval and juvenile survival, distribution, and growth in estuaries. *Estuaries* **26:**970–974.

Sheridan, P.F. and Ray, S.M. (1981) Report of the Workshop on the Ecological Interactions between Shrimp and Bottomfishes, April, 1980. Galveston, TX: National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southeast Fisheries Center, Galveston Laboratory, pp. 132.

St. Amant, L.S., Lindner, M.J., Allen, G.W., Ingle, R.M., Demoran, W.J. and Leary, T.R. (1966) The shrimp fishery of the Gulf of Mexico. In: Gulf States Marine Fisheries Commission Information Series No. 3. Gulf States Marine Fisheries Commission, New Orleans, LA. 10 pp.

Szuwalski, C. and Punt, A.E. (2012) Identifying research priorities for management under uncertainty: The estimation ability of the stock assessment method used for eastern Bering Sea snow crab (Chionoecetes opilio). *Fish. Res.* **134:**82–94.

Turner, R.E. and Brody, M.S. (1983) Habitat suitability index models: northern Gulf of Mexico brown shrimp and white shrimp. In: U.S. Department of the Interior Fish and Wildlife Service FWS/OBS-82/10.54. U.S. Department of Interior Fish and Wildlife Service, Washington, D.C.

Van Lopik, J.R., Drummond, K.H. and Condrey, R.E. (1979) Draft Environmental Impact Statement and Fishery Management Plan for the Shrimp Fishery of the Gulf of Mexico, United States. Tampa: Gulf of Mexico Fishery Management Council.

Wilkinson, T.A.C., Wiken, E.B., Creel, J.B. *et al.* (2009) Marine Ecoregions of North America. Montreal: Commission for Environmental Cooperation.

Williams, A.B. (1955) A contribution to the life histories of commercial shrimps (Penaeidae) in North Carolina. *Bull. Mar. Sci. Gulf Caribb.* **5:**116–146.

Williams, A.B. (1959) Spotted and brown shrimp postlarvae (Penaeus) in North Carolina. *Bull. Mar. Sci.* **9:**281–290.

Zein-Eldin, Z.P. and Aldrich, D.V. (1965) Growth and survival of postlarval penaeus aztecus under controlled conditions of temperature and salinity. *Biol. Bull.* **129:**199–216.

Zein-Eldin, Z.P. and Renaud, M.L. (1986) Inshore envrionmental-effects on brown shrimp, *Penaeus-aztecus*, and white shrimp, *Penaeus-setiferus*, populations in coastal waters, particularly of Texas. *Mar. Fish. Rev.*, **48:**9–19.

Zimmerman, R.J. (2003) Basis for Concern about the Hypoxic Zone in Coastal Louisiana. Galveston, TX: NOAA, National Marine Fisheries Service, Southeast Fisheries Science Center, Galveston Laboratory, Habitat Connections.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Table S1** Comparison of five candidate models with different levels of interaction terms and its respective statistics. RMSE, Root Mean Square Error; SSE, Sum of Square Error; SST, Sum of Square Total; GAM, Generalized Additive Model, BRT, Boosted Regression Tree.

**Figure S1** Comparison of the Root Mean Square Error (RMSE) (bars), and explained deviance (dotted line) of the five trained models using the test data set.

**Figure S2** Omnidirectional semivariogram of the model residuals for the best candidate BRT and GAM, compared with the nominal log(CPUE) (black line).