

1 **SUPPLEMENTAL MATERIAL**

2

3 In this supplemental material we specify the additional procedures and analyses. We also show
4 the figures and tables created to justify the use of Boosted Regression Trees (BRTs) for the
5 modeling process, and the evaluation of potential spatial autocorrelation in the data.

6

7 In our research we argued that BRTs are an adequate method to investigate the density
8 distribution of Brown shrimp in the southern Gulf of Mexico. Because historically, Generalized
9 Additive Models (GAMs) have been widely used in ecology and catch rate standardization
10 procedures we performed an initial model comparison of GAMs and BRTs.

11

12 **Choosing the final model**

13

14 During model fitting, to ensure that BRTs were an optimal method to model brown shrimp
15 density, we trained several candidate BRTs and GAMs to compare its performance (Table 1)
16 using the training data set. We first fitted two GAMs with two levels of interactions within the
17 predictor variables: no interaction term, and one interaction term (Model complexity, Table 1). In
18 GAMs it's not possible to fit a smooth function with more than two terms interacting (See Wood,
19 2006). Then, we repeated the same procedure for BRTs with one, two and five interaction terms.
20 BRTs have the capabilities of fitting complex variable interactions specifying the tree complexity
21 during the modeling process (Elith, 2008). As a result, we decided that the optimal model was a
22 BRT with a tree complexity of five (BRT-5, Table 1, Figure 1). The BRT-5 showed the highest
23 explained deviance (D^2) and adjusted R^2 ($1 - [SSE/SST]$) based on the testing set, and the lowest
24 residual deviance and Root Mean Square Error (RMSE).

25

26 **Table 1.** Comparison of five candidate models with different levels of interaction terms and its respective
27 statistics. RMSE = Root Mean Square Error, SSE = Sum of Square Error, SST = Sum of Square Total,
28 GAM = Generalized Additive Model, BRT = Boosted Regression Tree.

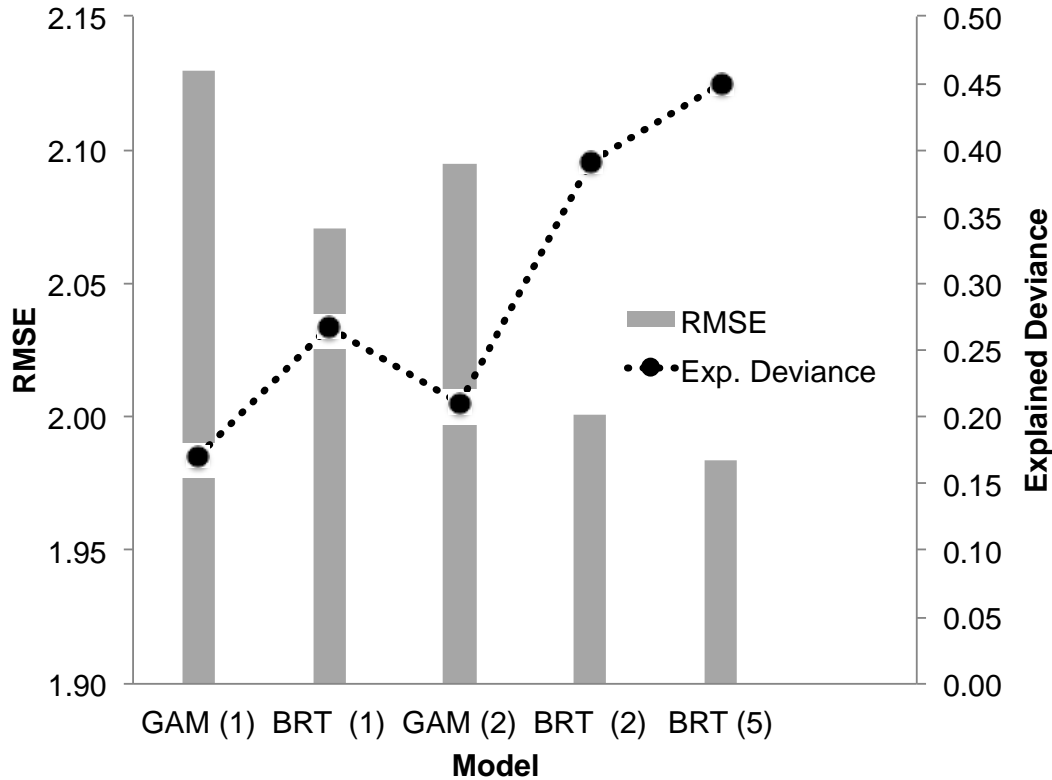
Model (complexity)	Model Complexity	Number of trees	Residual Deviance	Explained Deviance (D^2)	RMSE	$1 - (SSE/SST)$
GAM (1)	1	–	27481.06	0.17	2.13	0.18
BRT (1)	1	3050	24527.75	0.27	2.07	0.22
GAM (2)	2	–	26274.85	0.21	2.09	0.21
BRT (2)	2	2150	20313.46	0.39	2.00	0.27
BRT (5)	5	600	19381.39	0.45	1.98	0.29

29

30

31

32

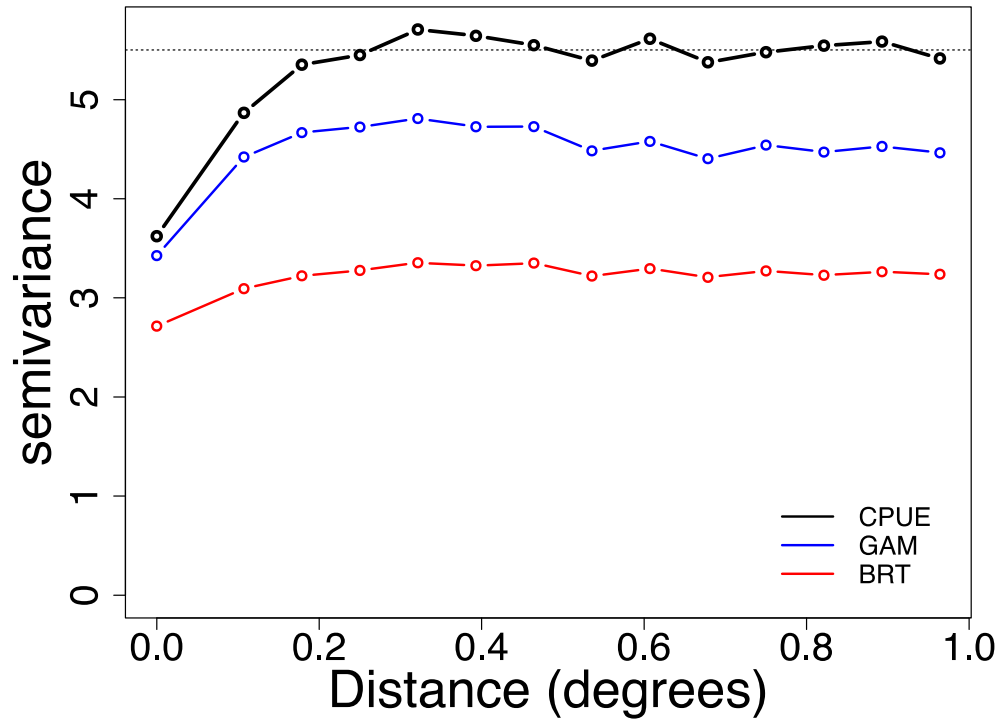


52 **Figure 1.** Comparison of the Root Mean Square Error (RMSE) (bars), and explained deviance (dotted
 53 line) of the five trained models using the test data set.

54
 55 **Spatial autocorrelation**

56
 57 Before deciding whether or not to use a mixed or autoregressive model to include the effect of
 58 spatial autocorrelation, we evaluated the potential presence of a considerable spatial
 59 autocorrelation that could affect the model performance and outputs. We first plotted an
 60 omnidirectional semivariogram of the response variable (log-CPUE, Figure 2), and then plotted
 61 the semivariogram for the model residuals of the BRT-5 and GAM-2, which were the best
 62 candidate models (Table 1). Even though none of the fitted model's structure is equipped to deal
 63 with auto covariance, the level of autocorrelation in the residuals decreases relative to the native
 64 levels (Figure 2). The highest decrease in the level of autocorrelation is observed with the BRT-5
 65 model. Consequently, some of the proportion of the observed autocorrelation resulted from
 66 interaction effects. This indicates the importance of the inclusion of higher interaction levels in
 67 the model. Additionally, The semivariogram for nominal log (CPUE), levels up quickly which
 68 suggest that the autocorrelation effect is not as strong to significantly affect the model output.

69
 70
 71
 72



93 **Figure 2.** Omnidirectional semivariogram of the model residuals for the best candidate BRT and GAM,
 94 compared with the nominal log(CPUE) (black line). Dashed dotted line represents the variogram sill = 5.5
 95 equal to the sample variance; x-axis = lag measured as Euclidean distance between latitude and longitude
 96 coordinates.