Comparing the use of boosted regression trees to multiple regression models when determining the reef flattening effects on total richness and species responses in the Caribbean.

Eric Womack

ENVS 603: Environmental Research Methods, Virginia Commonwealth University

Spring 2020

Learning Objectives

- Create and interpret a boosted regression tree (BRT) model
- Create and interpret a multiple regression model
- Compare BRT and multiple regression model predictive power

Introduction

Coral reefs are the most biodiverse marine habitat per unit area (McIntyre, 2010). The physical architectural structure of a coral reef is complex and has a direct effect on the biodiversity in that habitat (Komyakova, Munday, & Jones, 2013). Over the past 40 years the Caribbean has seen a nonlinear decline in the architectural complexity of coral reefs due to bleaching events and subsequent breaking up of bleached reefs, converting a complex landscape into a flat homogenous one (Alvarez-Filip, Dulvy, Gill, Côté, & Watkinson, 2009). When trying to model how decreasing reef complexity is affecting species richness it becomes clear that reef complexity is complex and should be represented by multiple independent variables. One way of modeling this would be to use multiple regression which allows the researcher to predict the value of a variable based on the value of two or more independent variables. While multiple regression models are easy to generate and interpret, recent advances in machine learning have seen the increase in the adoption of BRTs which convey some advantages over generalized linear models. BRTs combine decision tree algorithms with boosting methods which improve the model accuracy. Comparing multiple regression models by the percent deviance explained and by the prediction error using a test dataset measured as root mean square error indicate that BRTs perform better than multiple regression models (Abeare, 2009).

In “Reef flattening effects on total richness and species responses in the Caribbean” Newman et al 2015 used BRTs to determine the relative influence that seven independent complexity variables had on the species richness of seven different taxa: anemone, annelid, arthropod, coral, fish, octocoral, sponge and total species richness in three countries in the Caribbean. Newman’s BRT models worked best for explaining the variability in total richness (82.3%), coral richness (80.6%), and fish species richness (77.3%). Newman and all used their BRTs to determine the relative influence that each of their seven complexity predictors had on the species richness of the taxa in question.

The ability to determine the relative influences independent predictors have on a model is a powerful tool which helps researchers assess which factors are having the biggest impact on richness still in the context of other factors. In this exercise we'll attempt to recreate the Newman et al. BRT model on fish richness data. We'll compare if our BRT model achieves the same relative influence percentages for the predictor variables. Next, we'll create a multiple regression model with the same data to compare the relative influence percentage values obtained through this method. Both models will be compared for predictive power using the guidelines established by Abeare 2009 described above. Finally, there is an assignment to recreate the analyses outlined in this lesson on total species richness.

Exercise Outline

1. PART 1: Overview of the models being used
    a. Boosted Regression Tree Models
    b. Multiple Regression Models
    c. Boosted Regression Tree Models vs Multiple Regression Models: Relative Influence
2. PART 2: Summary of relevant Newman et al. 2015 analyses
    a. Newman et al. Methods
        i. The Plots
        ii. The Dependent Variable: Fish Richness
        iii. The Independent Variables: Complexity Factors
        iv. The Model
    b. Newman et al. Results
3. PART 3: Creating and comparing our new models
    a. Creating a Boosted Regression Tree Model
    b. Creating a Multiple Regression Model
    c. Comparing the Predictive Power of our Models
4. PART 4: Assignment

PART 1

Boosted Regression Tree Models

The following description of BRTs was adapted from The Biodiversity and Climate Change Virtual Laboratory's guide to boosted regression trees:

Additional Useful Resources on BRTs:

"A working guide to boosted regression trees" (Elith, Leathwick, & Hastie, 2008)

StatQuest: Decision Trees

StatQuest: Regression Trees, Clearly Explained!!!

BRT models are a combination of two techniques: decision tree algorithms and boosting methods. BRTs repeatedly fit many decision trees to improve the accuracy of the model. BRTs take a random subset of all data for each new tree that is built. All random subsets have the same number of data points, and are selected from the complete dataset. Used data is placed back in the full dataset and can be selected in subsequent trees. BRTs use the boosting method in which the input data are weighted in subsequent trees. The weights are applied in such a way that data that was poorly modelled by previous trees has a higher probability of being selected in the new tree. This means that after the first tree is fitted the model will take into account the error in the prediction of that tree to fit the next tree, and so on. By taking into account the fit of previous trees that are built, the model continuously tries to improve its accuracy.

Boosted Regression Trees have two important parameters that need to be specified by the user:

1. Tree complexity (tc): this controls the number of splits in each tree. A tc value of 1 results in trees with only 1 split, and means that the model does not take into account interactions between environmental variables. A tc value of 2 results in two splits and so on.
2. Learning rate (lr): this determines the contribution of each tree to the growing model. As small value of lr results in many trees to be built.

These two parameters together determine the number of trees that is required for optimal prediction. The aim is to find the combination of parameters that results in the minimum error for predictions. As a rule of thumb, it is advised to use a combination of tree complexity and learning rate values that result in a model with at least 1000 trees. The optimal 'tc' and 'lr' values depend on the size of your dataset. For datasets with <500 occurrence points, it is best to model simple trees ('tc' = 2 or 3) with small enough learning rates to allow the model to grow at least 1000 trees.

Multiple Regression Models

The following description of multiple regression models was adapted from [The Biodiversity and Climate Change Virtual Laboratory's guide to generalized linear models]:

Additional Useful Resources on Multiple Regression Models:

[StatQuest: Linear Models Pt.1 - Linear Regression]

[StatQuest: Linear Models Pt.1.5 - Multiple Regression]

[StatQuest: Multiple Regression in R]

Multiple regression models are a form of generalized linear models (GLM) and are an extension of 'simple' linear regression models, which predict the response variable as a function of multiple predictor variables. Linear regression models work on a few assumptions, such as the assumption that we can use a straight line to describe the relationship between the response and the predictor variables. This implies that a constant change in a predictor leads to a constant change in the response variable. This assumption is often violated in ecological data, and therefore these models are extended into GLMs to be able to deal with non-normal distributed data. GLMs find the equation that best predicts the occurrence of a species for the values of the environmental variables. The model has three important components:

1. The probability distribution of the response variable.
2. The linear predictor (LP): a combination of all predictor variables, which represents an overall score for the environmental suitability.
3. The link function: this describes how the mean of the response depends on the linear predictor.

The estimation of the values of the variable coefficients is obtained by maximum likelihood estimation (MLE), which maximizes the "agreement" of the predicted species occurrences with the observed data. In other words, MLE finds the values of the coefficients that result in a model under which you would be most likely to get the observed results.

Boosted Regression Tree Models vs Multiple Regression Models: Relative Influence

Ostensibly BRT models and multiple regression models are attempting similar things: explain how multiple predictor variables influence a response variable. Recall how Newman et al used their BRT models to establish the relative influence of different complexity variables on the species richness of different taxa. These relative influence numbers represent the contribution of each variable and are scaled so that the sum adds to 100, with higher numbers indicating a stronger influence on the response. It is also possible to determine relative influence from multiple regression models, as long as all regressors are uncorrelated: Each regressor's contribution is just the R^2 from univariate regression, and all univariate R^2-values add up to the full model R^2.

PART 2

Newman et al. Methods

The plots

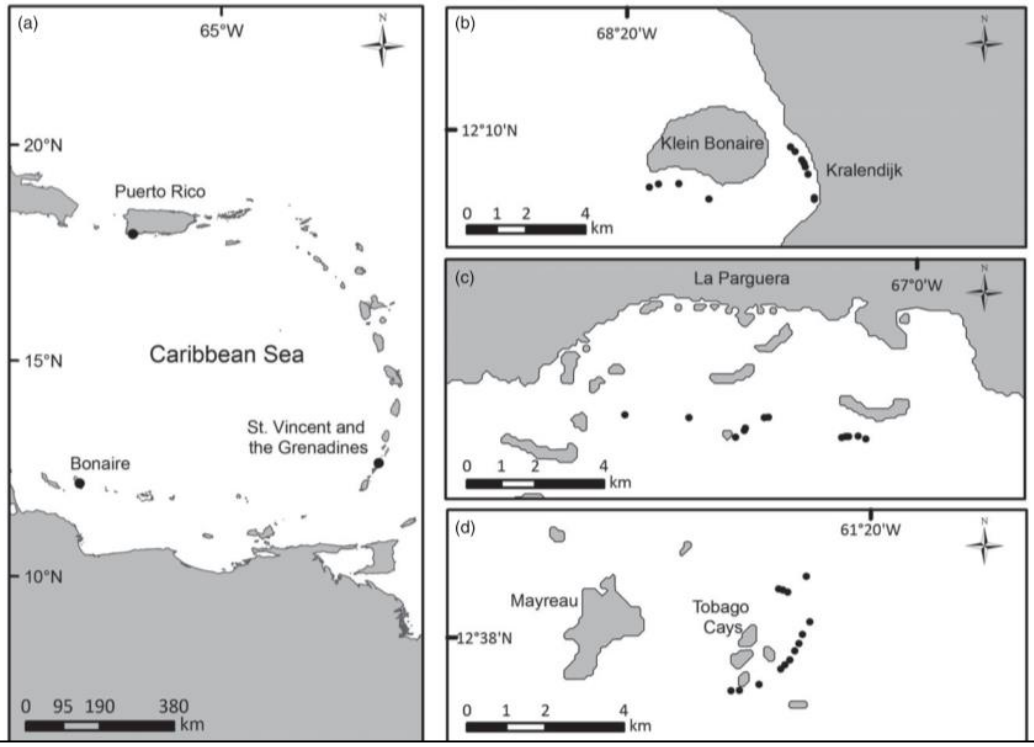


Fig. 1 Survey locations: (a) in eastern Caribbean, (b) west coast of Bonaire, (c) La Parguera, south-west Puerto Rico, and (d) Tobago Cays in St. Vincent and the Grenadines (Newman, et al., 2015).

Adapted from Newman et al. 2015:

The presence of reef macrofauna was recorded on reefs with different levels of topographic complexity in three marine reserves in the Caribbean: Bonaire National Marine Park (BON), La Parguera Natural Reserve in south-west Puerto Rico (PR) and the Tobago Cays Marine Park in St. Vincent and the Grenadines (SVG; Fig. 1). Species presence was quantified in twenty 25-m2 plots at each of five levels of reefscape complexity in each country (100 plots total per country). The maximum distance between surveys was 5.5km in Bonaire, 6.9 km in PR and 3.6 km in SVG. Plots were haphazardly situated in areas of uniform complexity at least 10 m from a boundary between different complexity levels or from other plots, on coral fore-reefs at depths of 5–15 m (mean 10.15 ± 0.14 SE n = 300).

The Dependent Variable: Fish Richness

Adapted from Newman et al. 2015:

Plots were delineated with tape measures in a 'T' shape, after first recording larger, more mobile fish species each plot half was then carefully searched for fish (by S Newman and C Dryden). Unknown species were photographed for later identification. Survey time was greater in more complex plots due to greater surface area and the necessity to thoroughly search for cryptic species, with total plot survey times varying between 10 and 20 min.

The Independent Variables: Complexity Factors

Adapted from Newman et al. 2015:

Factor	Name in Data	Description
Reef Complexity	Reef.complexity	Reefscape complexity was visually estimated on a scale of 1 (flat, little relief) to 5 (highly complex with high vertical relief and overhangs) (Appendix A)
Number of Large Corals	No.tall.corals	Number of live corals > 50 cm in height
Number of Corals	No.corals	Number of live corals > 4 cm in diameter
Maximum Octocoral Height	Octocoral.max.height	Measure of soft coral height
Maximum Sponge Height	Sponge.max.height	Measure of sponge height
Slope	Slope.angle	Slope angle was visually estimated in degrees from the horizontal plane at each plot edge perpendicular to the reef slope and averaged
Country	Country	Country where the plot was located

The Model

Adapted from Newman et al. 2015:

Relationships between species richness of different taxonomic groups and reef complexity and structural components, and the relative importance of each complexity variable, were examined using boosted regression trees. All models were fitted to allow interactions using a tree complexity (tc) of 5, a bag fraction of 0.6 and a learning rate (lr) of 0.003 or 0.001 to minimize predictive deviance and maximize performance. Predictor variables that increased variance and reduced model performance were dropped using the 'gbm.step' function from Elith, Leathwick & Hastie (2008). Ten-fold cross-validation (CV) was used to identify the optimum number of trees (1000–2650 for taxonomic group models).

Newman et al. Results

For our purposes we're just looking at the results for fish richness. Newman et al. found that their boosted regression tree model for fish richness explained 77.3% of the variation. Newman et al. lists four most important predictor variables ranked by percentage relative influence on fish richness as: country (31.6%), reef complexity (24.9%), number of large corals (14.0%), and sponge max height (8.9%) and generated fitted function plots (Fig 2). Each graph represents the fitted function of fish richness response due to the x -axis (predictor) variable, while keeping all other predictor variables average. From these plots Newman et al. concluded: Reef complexity and country had the greatest relative influence on fish species richness, with fish species richness lowest in PR. Fish species richness was highest in Bonaire and SVG, at high reef complexity levels, where there were more than 15 large corals per plot and with sponges over 75 cm tall. Fish species richness declined below reef complexity level three, and confidence intervals indicate a greater variability in the number of fish species at lower levels of complexity. Now let's create our own boosted regression tree model and see how our results compare.

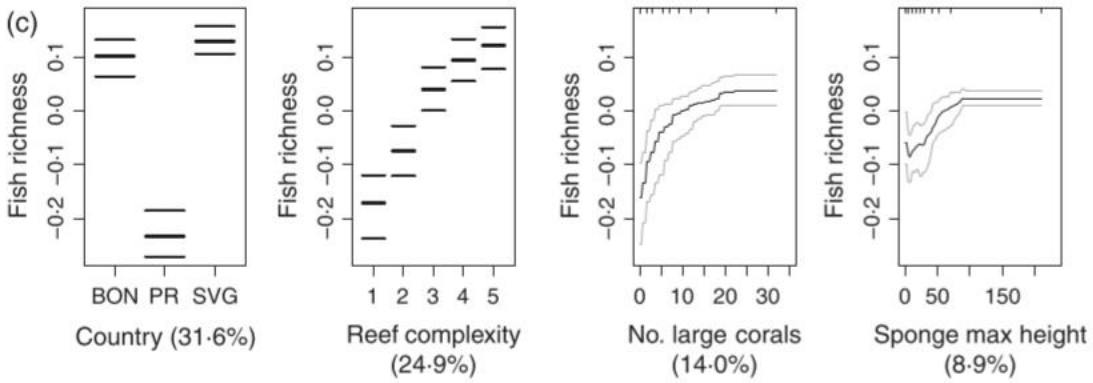


Fig. 2 Functions fitted for the four most important predictor variables (ranked by percentage relative influence left to right) (Newman, et al., 2015).

PART 3

Important Note: PART 3 of this exercise utilizes the accompanying Rmd file: “PART3.Rmd”

Lines highlighted in blue correspond to chunks to be run in the Rmd file.

Creating a Boosted Regression Tree Model

1. Load the required packages and import the data
   ```
   library(dismo)
   data <- read.csv("data.csv")
   ```

2. Create the model
   ```
   fish.step<- gbm.step(data=data, gbm.x = 10:16, gbm.y = 1, family = "poisson",
   tree.complexity = 5, learning.rate = 0.003, bag.fraction = 0.6)
   ```
   a. Let’s break this down, we’re attempting to recreate the model generated by Newman et al. these individual parameters are taken from the Model section of PART 1. Here’s what they mean:
      i. “gbm.x=” Defines the range of the independent variables you want to use. In our data set: data.csv the complexity variables occupied the last seven columns. We want to use all of our complexity variables to make this model so we select 10:16.
      ii. “gbm.y=” Defines the dependent variable. In our data.csv fish richness occupies the first column so we select “1”.
      iii. “tree.complexity=” (tc) Defines the number of nodes in a tree
      iv. “learning.rate=” (lr) Learning rate is used to shrink the contribution of each tree as it is added to the model. Decreasing the learning rate will increase the number of trees required.
      v. “bag.fraction=” Bag fraction selects the proportion of data to be selected at each step

3. Look at the output
   a. Fig 3 shows a graphical representation of how the algorithm decided how many trees to use to generate the model. The solid black curve is the mean, and the dotted curves about 1 standard error, for the measure of predictive deviance. The goal of the model is to maximize the predictive deviance by minimizing the holdout deviance using the least number of trees possible. The red line shows the minimum of the mean and the green line shows the number of trees at which that occurs in this case 1350 trees. So far this is consistent with Newman et al. who found the optimum number of trees to be between 1000 and 2650.
      i. IMPORTANT NOTE: Your numbers will not be identical, BRTs are stochastic and are therefore slightly different each time they’re run.

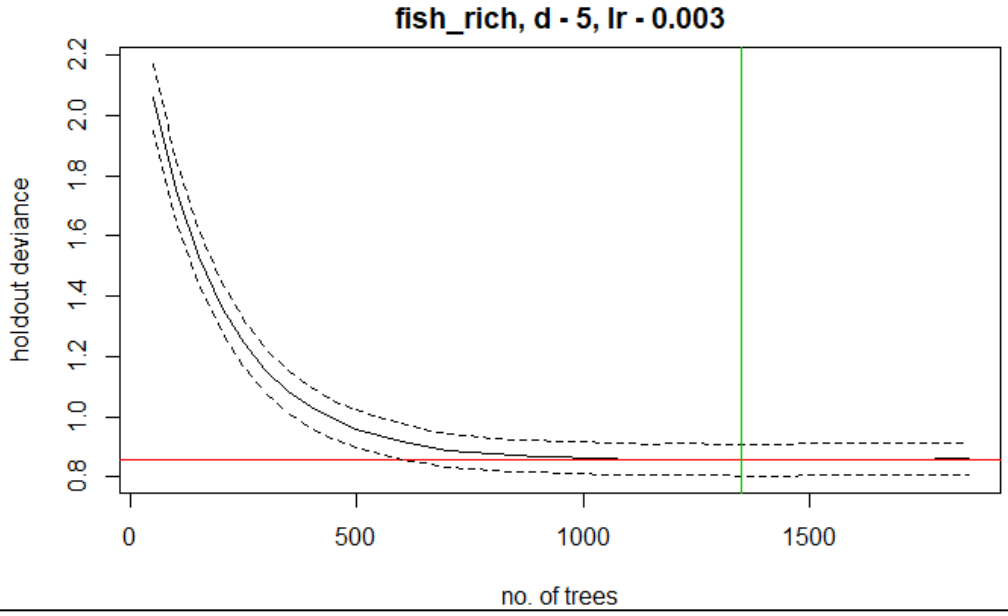


Fig. 3 Graphical representation of boosted regression tree model fitting

b. Also outputted was a brief summary of our model which contains the information necessary to calculate the deviance explained.
   i. $Deviance\ Explained = \left(1 - \left(\frac{mean\ residual\ deviance}{mean\ total\ deviance}\right)\right) * 100$
   ii. $Deviance\ Explained = \left(1 - \left(\frac{0.528}{2.469}\right)\right) * 100$
   iii. $Deviance\ Explained = 78.6\%$
   iv. Recall that Newman et al. had a deviance explained of 77.3% for fish richness, so far our models seem comparable.

4. Assessing the performance of our model
   a. The gmb.step function's ultimate goal is to develop a model in order to get predictions on new data it does this through cross validation.
   b. Cross validation works by:
      i. Splitting the dataset into separate training and test subsets
      ii. hold out a set at a time and train the model on the remaining set
      iii. Test model on hold out set
      iv. Repeat this process for each subset of the dataset
   c. The output from number 3 above also provided us with some cross validation statistics that are useful in assessing the performance of our model:
      i. training data correlation = 0.885
      ii. cv correlation = 0.787
   d. These numbers tell us how well the model is fitting our data

Additional Useful Resource on Cross Validation:

Towards data science: Cross Validation Explained: Evaluating estimator performance.

i. 0.885 is a good score while 0.787 is acceptable (Dedman, Officer, Brophy, Clarke, & Reid, 2017)

e. The relatively minor decrease from training to CV indicates that overfitting was not a serious issue and predictions were more likely to be correct. These scores indicate that we can have confidence in the model predictions.

5. Examine Residuals Plot

```
plot(fish.step$fitted, fish.step$residuals)
abline(0, 0)
```

a. Recall that this plot shows if residuals have nonlinear patterns. Ideally residuals should be spread around a horizontal line without obvious patterns.

b. The residuals are relatively pattern free this plot gets a pass.

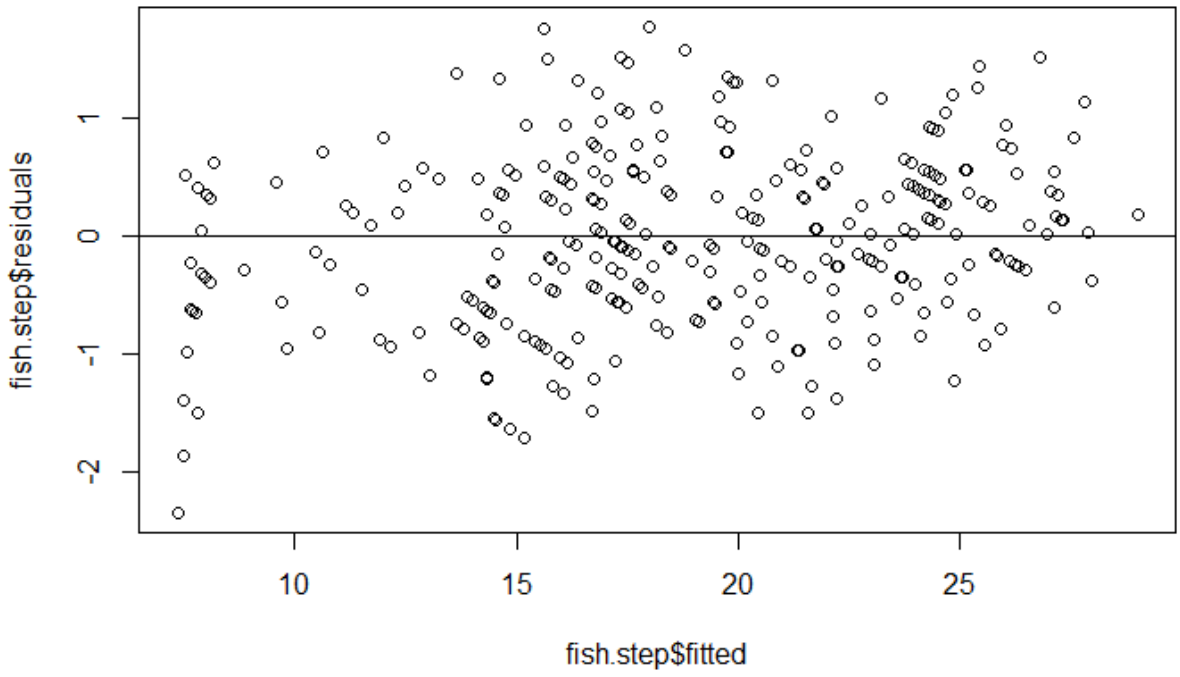


Fig. 4 Residuals vs Fitted for our boosted regression tree model

6. Display relative influence

```
summary(fish.step)
```

a. Let's compare the relative influence of the predictors in our model to the Newman et al. model. While the relative influence percentages are slightly different the order of the most influential predictors is the same for the top four predictors. Still comparable so far so let's keep going.

Predictor	Newman et al. BRT Relative Influence (%) on Fish Richness	Our BRT Relative Influence (%) on Fish Richness
Country	31.6	33.9
7Reef Complexity	24.9	27.5
Number Large Corals	14.0	10.9
Sponge Max Height	8.9	8.0
Number of Corals	Unknown	7.5
Octocoral Max Height	Unknown	6.8
Slope Angle	Unknown	5.4

7. Plotting the fitted functions
   gbm.plot(fish.step,n.plots=4,common.scale = TRUE, write.title = FALSE, plot.layout = c(2,4))
   a. Here we plot the first four predictors by relative influence to generate a plot similar to the Newman et al. partial dependency plot seen in Fig 2. For comparison the Newman et al. plots will be added on the top row while our new plots will be on the bottom row.

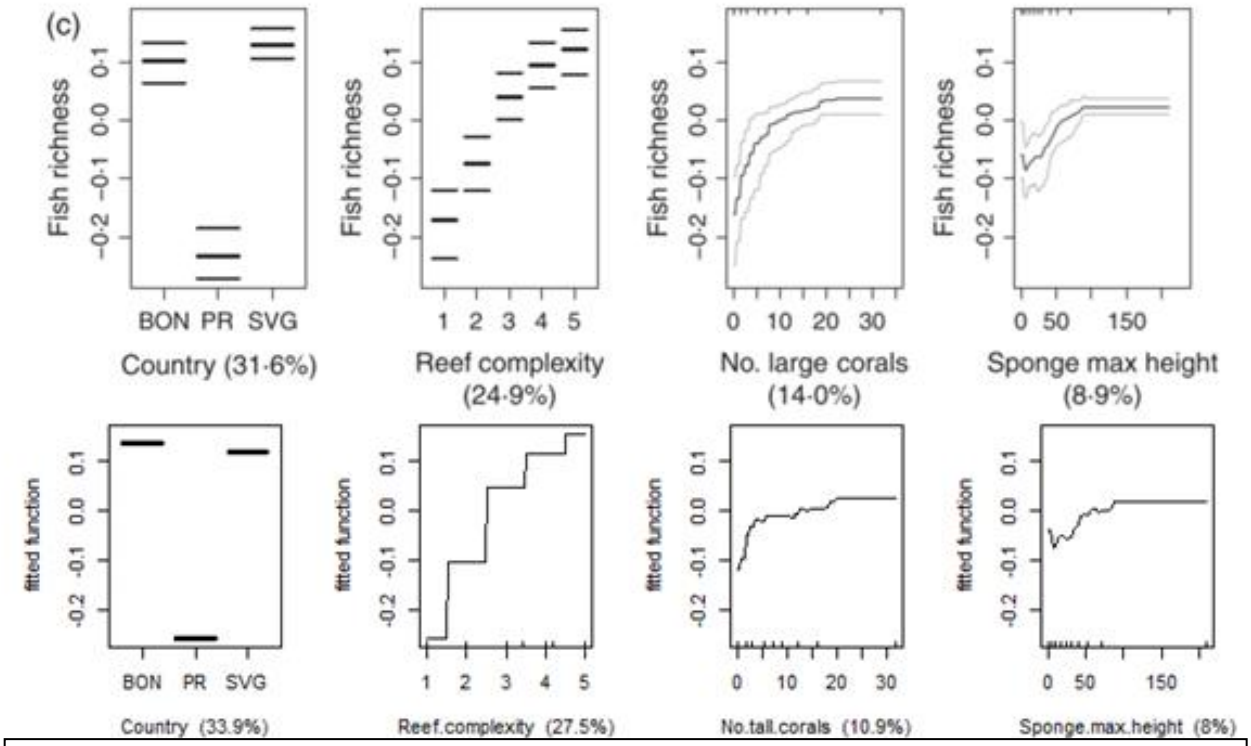


Fig. 5 Functions fitted for the four most important predictor variables (ranked by percentage relative influence left to right). **TOP ROW**: Newman et al. plots **BOTTOM ROW**: Our plots

b. Recall that each graph represents the fitted function of fish richness response due to the x -axis (predictor) variable, while keeping all other predictor variables average.
c. Let's compare our results with Newman et al. going line by line to see if we draw the same conclusions from our model
    i. "Reef complexity and country had the greatest relative influence on fish species richness"
        1. Yes, our results are similar though not identical
            a. Newman found that combined reef complexity and country We found that combined reef complexity and country had a relative influence of 61.4%
        2. "With fish species richness lowest in PR"
            a. Yes, our results are similar
    ii. "Fish species richness was highest in Bonaire and SVG"
        1. Yes, our results are similar but not identical
            a. While Bonaire and SVG both have higher richness than PR Newman has SVG slightly higher than Bonaire and we have Bonaire slightly higher than SVG
    iii. "At high reef complexity levels, where there were more than 15 large corals per plot"
        1. Yes, our results are similar
    iv. "And with sponges over 75 cm tall"
        1. Yes, our results are similar
    v. "Fish species richness declined below reef complexity level three"
        1. Yes, our results are similar

Creating a Multiple Regression Model

1. Load the required packages
   ```
   library(relaimpo)
   library(car)
   ```

2. Create the multiple regression model
   ```
   fish.glm<- glm(fish_rich ~ Reef.complexity + No.tall.corals + No.corals +
   Sponge.max.height + Octocoral.max.height + Slope.angle + Country, data = data)
   ```
   a. This creates the GLM "fish" using the dependent variable "fish_rich" as a function of our multiple independent variables which are the structural components described above.

3. Examine diagnostic plots
   ```
   par(mfrow=c(2,2))
   plot(fish.glm)
   ```
   a. Let's check if the model works well for these data

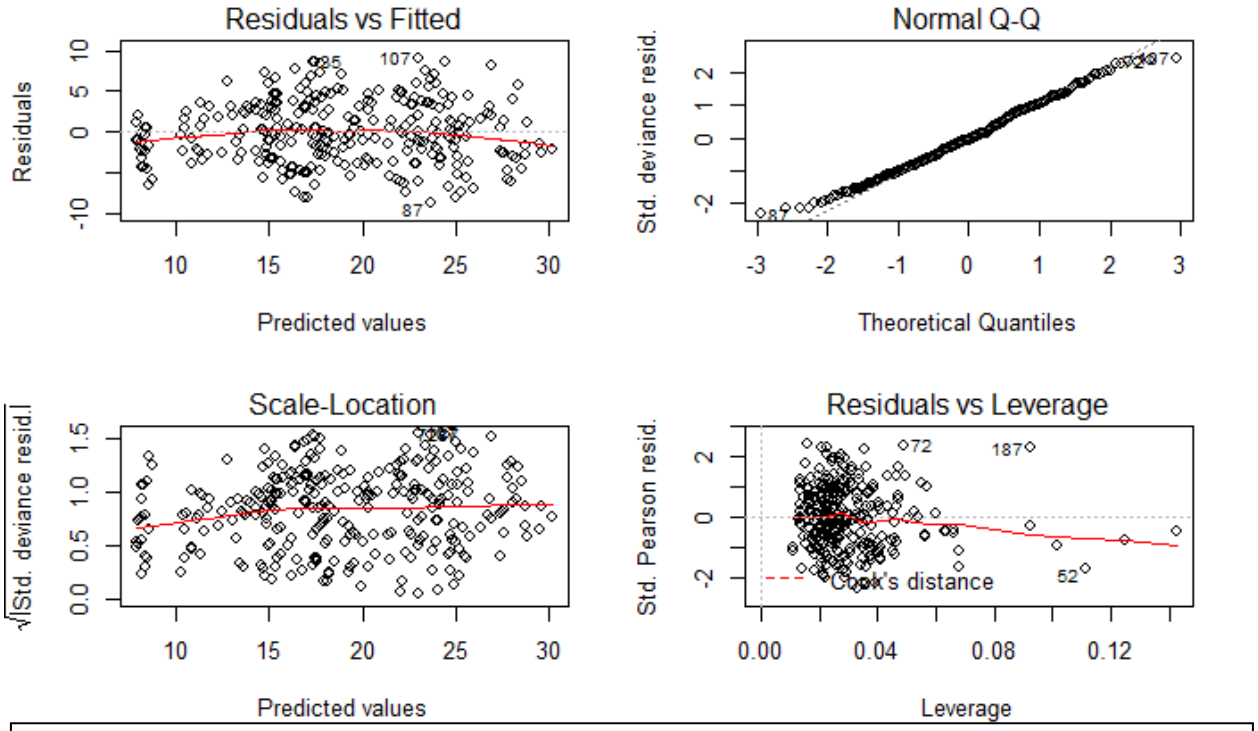


Fig. 6 Diagnostic plots for our multiple regression model

b. Fig 6: Residuals vs Fitted
    i. Recall that this plot shows if residuals have nonlinear patterns. Ideally residuals should be spread around a horizontal line without obvious patterns.
    ii. There is a slight curve to our line but nothing dramatic and the residuals are relatively pattern free, this plot gets a pass.
c. Fig 6: Normal Q-Q
    i. This plot shows if residuals are normally distributed. Residuals should follow a straight line.
    ii. There is some deviation at each end of the line but nothing too dramatic. This plot gets a pass.
d. Fig 6: Scale Location
    i. This plot shows if residuals are spread equally along the ranges of predictors. Similar to the residuals vs fitted plot the residuals should be dispersed around a horizontal line without obvious patterns.
    ii. A slight curve but nothing too dramatic and no clear pattern to the residuals. This plot gets a pass.
e. Fig 6: Residuals vs Leverage
    i. This plot looks for outliers which might be skewing our data. These will be represented by points falling outside of Cook's distance which is represented by the dashed line.
    ii. No points are falling outside of Cook's distance in this plot so it gets a pass.

4. Check for multicollinearity
   vif(fish.glm)
   a. Multicollinearity refers to a situation in which two or more explanatory variables in a multiple regression model are highly linearly related. This can be a problem because it undermines the statistical significance of an independent variable.
   b. Recall also that for the relative influence % we hope to generate from our multiple regression model depends on the explanatory variables being independent of each other.
   c. One way to check for multicollinearity in our model is by calculating the variance inflation factor (VIF) of our predictors which is done by regressing it against every other predictor in our model.
   d. Predictors with a VIF > 5 indicate high correlation and may need to be removed from the model. All of our VIFs are < 5 so we'll continue with the predictors we have.

5. Calculate the deviance explained
   summary(fish.glm)
   a. This summary provides us with the information we need to calculate the deviance explained by this model
      i. $Deviance\ Explained = \left(1 - \left(\frac{residual\ deviance}{total\ deviance}\right)\right) * 100$
      ii. $Deviance\ Explained = \left(1 - \left(\frac{4107}{12853}\right)\right) * 100$
      iii. $Deviance\ Explained = 68.0\%$
      iv. Recall that Newman et al. had a deviance explained of 77.3% and our BRT model had a deviance explained of 78.6%

6. Calculate relative influence of the predictors
   calc.relimp(fish.glm, type = "lmg", rela = TRUE)

Predictor	Newman et al. BRT Relative Influence (%) on Fish Richness	Our BRT Relative Influence (%) on Fish Richness	Our GLM Relative Influence (%) on Fish Richness
Country	31.6	33.9	33.6
Reef Complexity	24.9	27.5	26.7
Number Large Corals	14.0	10.9	15.9
Sponge Max Height	8.9	8.0	4.5
Number of Corals	Unknown	7.5	6.8
Octocoral Max Height	Unknown	6.8	3.5
Slope Angle	Unknown	5.4	9.0

   a. Comparing the percent relative influence of the predictors in our models above shows that our multiple regression model agrees with our BRT model about the top three predictors but deviate for the remaining four.
   b. Which model should we use?

Comparing the Predictive Power of our Models

In “Comparisons of boosted regression tree, GLM and GAM performance in the standardization of yellowfin tuna catch-rate data from the Gulf of Mexico lonline [sic] fishery” Abeare 2009 compared the predictive power of generalized linear models and BRT models using the deviance explained by the model and the root mean squared error

1. Deviance explained by the model: Recall that we’ve already calculated the deviance explained:

Fish Richness Model	Deviance Explained (%)
Newman et al. BRT	77.3
Our BRT	78.6
Our GLM	68.0

   a. This shows that both of the BRT models were able to explain ~10% more of the variation in our data than our multiple regression model.

2. Root mean square error

```
RMSE <- function(error) { sqrt(mean(error^2)) }
RMSE(fish.glm$residuals)
RMSE(fish.step$residuals)
```

   a. Root mean square error is the measure of the differences between values predicted by a model and the values observed.

Fish Richness Model	RMSE
Our BRT	0.75
Our GLM	3.70

   b. RMSE is an absolute measure of fit and is reported in the same unit as the response variable (fish richness) therefore these numbers should be directly comparable. Smaller RMSE values imply a better fit.

3. Conclusion
   a. According to these metrics our BRT model fit the data better while simultaneously explaining more of the deviation making it the superior model.

PART 4

Assignment

To answer the questions below use the PART3.Rmd file and PART 3 walkthrough in this document as a template to generate your own multiple regression and BRT models using total species richness (the measure of all species richness described by the field “total_rich” in data.csv) as the dependent variable.

1. How did the relative influence of your predictors compare to those reported by Newman et al.?

Predictor	Newman et al. BRT Relative Influence (%) on Total Richness	Your BRT Relative Influence (%) on Total Richness	Your GLM Relative Influence (%) on Total Richness
Number Large Corals	21.5		
Sponge Max Height	18.6		
Reef Complexity	17.3		
Octocoral Max Height	14.9		
Number of Corals	Unknown		
Country	Unknown		
Slope Angle	Unknown		

2. How much of the deviance was explained by your models compared to Newman et al.?

Total Richness Model	Deviance Explained (%)
Newman et al. BRT	82.3
Your BRT	
Your GLM	

3. What were the RMSE values for your models?

Total Richness Model	RMSE
Your BRT	
Your GLM	

4. Based on these results which model would you choose to describe these data?

Works Cited

Abeare, S. (2009). Comparisons of boosted regression tree, GLM and GAM performance in the standardization of yellowfin tuna catch-rate data from the Gulf of Mexico lonline [sic] fishery. *Louisiana State University and Agricultural and Mechanical College*. Retrieved from https://digitalcommons.lsu.edu/gradschool_theses/2880

Alvarez-Filip, L., Dulvy, N. K., Gill, J. A., Côté, I. M., & Watkinson, A. R. (2009). Flattening of Caribbean coral reefs: region-wide declines in architectural complexity. *The Royal Society Publishing*. Retrieved from https://doi.org/10.1098/rspb.2009.0339

Dedman, S., Officer, R., Brophy, D., Clarke, M., & Reid, D. G. (2017). Advanced Spatial Modeling to Inform Management of Data-Poor Juvenile and Adult Female Rays. *Marine and Freshwater Research Centre*.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 802-813.

Komyakova, V., Munday, P. L., & Jones, G. P. (2013). Relative Importance of Coral Cover, Habitat Complexity and Diversity in Determining the Structure of Reef Fish Communities. *PLoS One*, e83178. doi:10.1371/journal.pone.0083178

McIntyre, A. D. (2010). *Life in the World's Oceans: Diversity, Distribution, and Abundance.* Oxford: Blackwell Publishing Ltd. doi:10.1002/9781444325508

Newman, S. P., Meesters, E. H., Dryden, C. S., Williams, S. M., Sanchez, C., Mumby, P. S., & Polunin, N. V. (2015). Reef flattening effects on total richness and species responses in the Caribbean. *Journal of Animal Ecology*, 1678-1689. doi:10.1111/1365-2656.12429

APPENDIX A

Examples of visually assessed levels of complexity (Newman, et al., 2015)

Grade 1:

No or very low vertical relief

Grade 2:

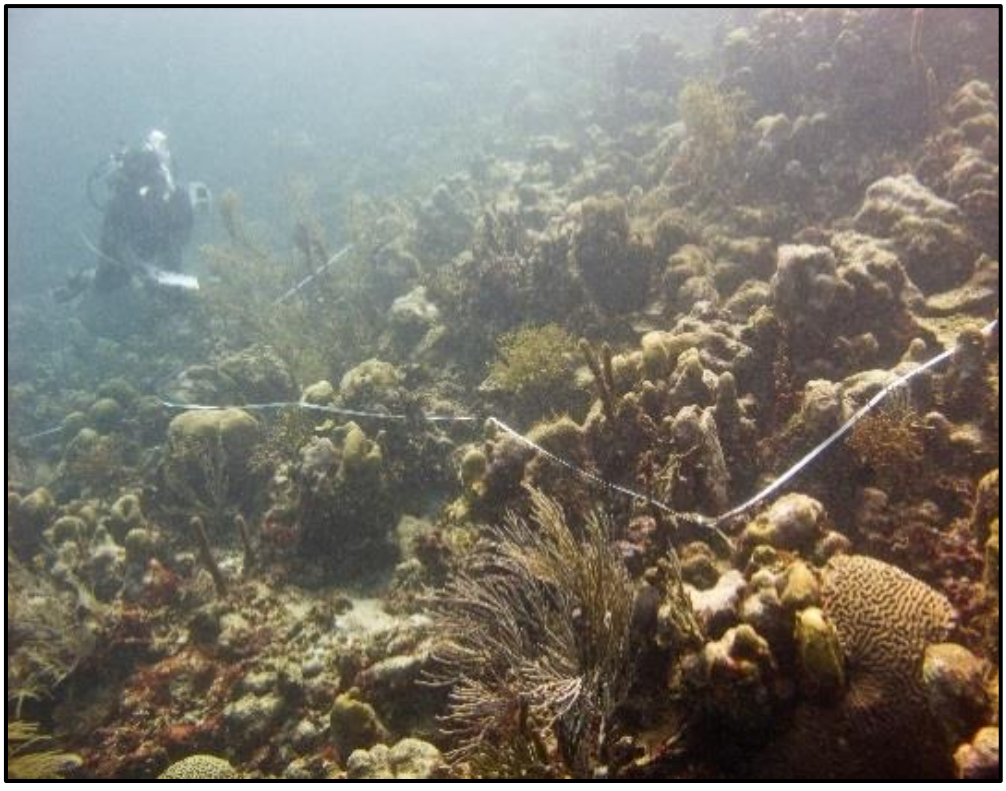
Low relief

Grade 3:

Moderate complexity

Grade 4:

Very complex with numerous fissures



Grade 5:

Exceptionally complex with numerous ledges or overhangs