Predicting Hourly Bike Rentals

Abdullah Farouk
University of British Columbia

April 08 2019

Outline

Introduction

- Bike sharing systems are a new mode of transportation.
- Highly flexible. Rent for short periods of time. Return when done.
- No need to store or maintain them.
- Estimated that $\sim$ 500 systems and 50,000 bicycles worldwide.
- Data on usage can be used to
  - Monitor traffic.
  - Identify congested areas in city.

Goal

Predict Capital Bike Share system's hourly rental counts in Washington D.C.

Exploring Data

Data Description

- Data set contains hourly ride/rental counts of Capital Bikeshare systems' users in Washington D.C. from 2011 to 2012.
- Has 16 columns (temp, humidity, month and time of day etc.).
- Mix of categorical and continuous variables.

Visualizing Bike Rides over Time

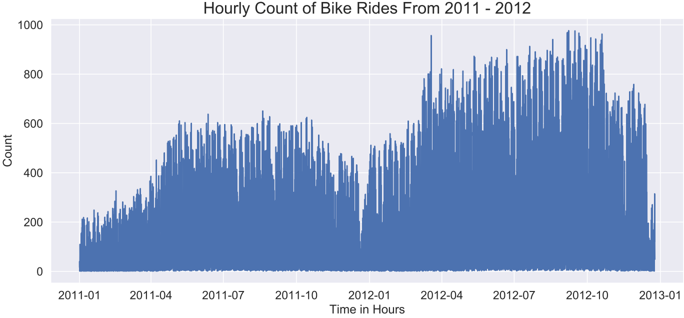

Figure 1: Hourly count of bike rides in Washington D.C. from 2011 - 2012

Features of Ride Count Time Series

- Seasonality - Periodic patterns that repeat over time
- Trend - Non-periodic patterns observed over time

Seasonality

- Hourly counts measured over two years.
- Long time series with high frequency.
- Multiple seasonalities observed in the data.

Daily Seasonality

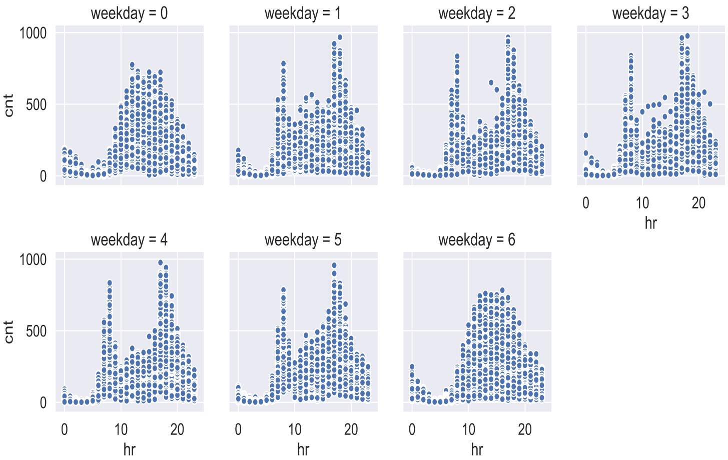

Figure 2: Hourly variation in ride count on each day of the week

Weekly Seasonality

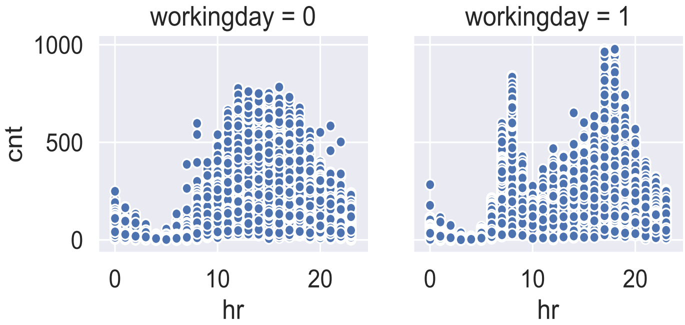

Figure 3: Differences between weekday and weekend ride count distribution

Yearly Seasonality

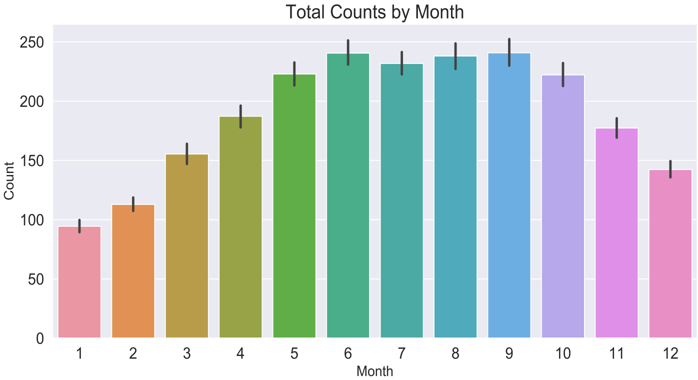

Figure 4: Total count of bike rides for each month of the year

Trend

- General increase in counts from 2011 to 2012
- Aggregating counts (e.g. reducing frequency to daily or monthly) makes this more apparent.

Monthly Bike Ride Count

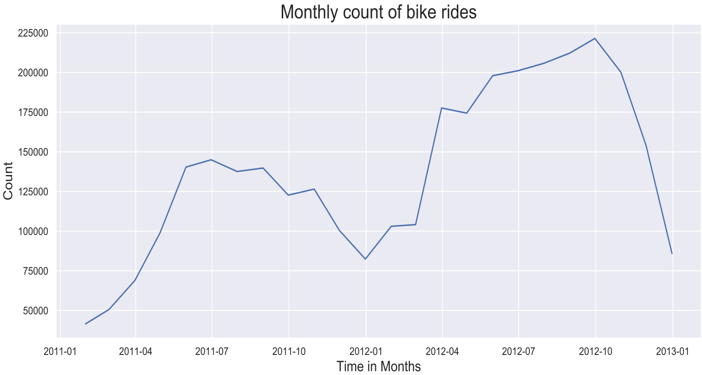

Figure 5: Sum of counts for each month in 2011 and 2012

Feature Exploration

Identifying useful features to include as predictors

Weather Situation

Categorical variable that encodes prevailing weather conditions.

It has 4 levels. They are:

- 1: Clear, Few clouds, Partly cloudy, Partly cloudy.
- 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist.
- 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds.
- 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog.

Effect of Weather Situation on Ride Counts

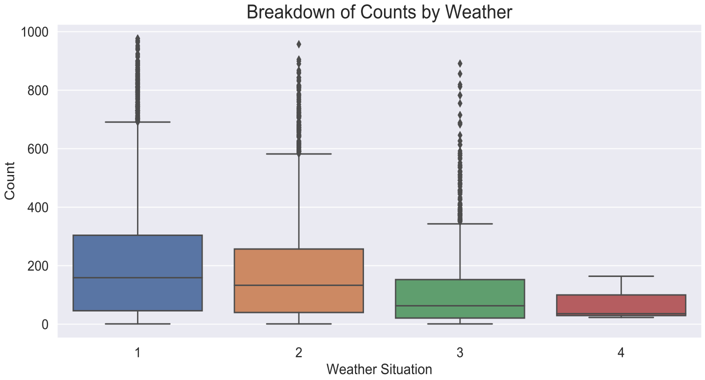

Figure 6: Distribution of bike ride counts under each weather condition

Potential Interactions of Interest

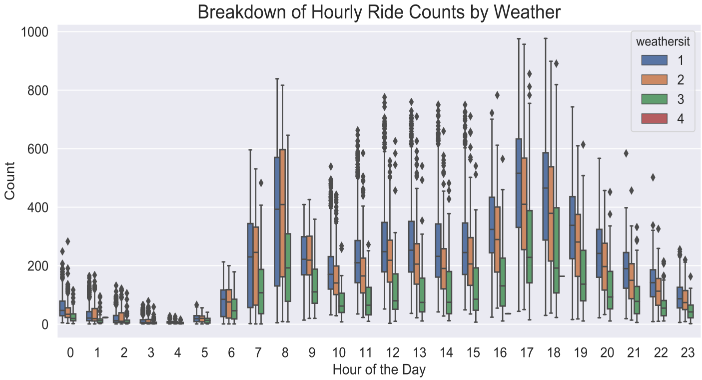

Figure 7: Ride counts under different weather conditions every hour of the day

Environmental Conditions (EC)

- Data set contains information on real temperature, humidity and wind speed, on an hourly basis.
- Interested in determining if there is a linear/non-linear relationship between them and ride counts.
- They are continuous variables. Easy to plot and calculate correlations with.

Effect of Environmental Conditions

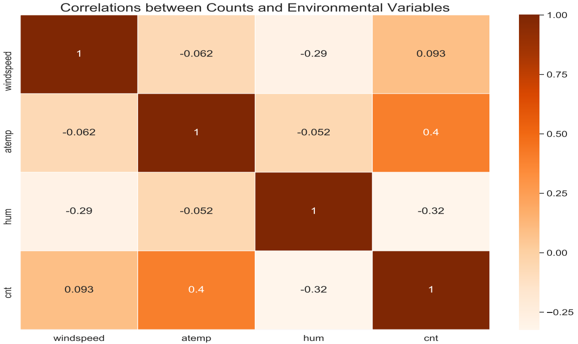

Figure 8: Strength of linear relationship between EC and ride counts

Findings

- Ride counts significantly higher under weather conditions 1 & 2.
- Weak linear relationship between EC and ride counts
- Accounting for multiple seasonalities and increasing trend are more important for predictive accuracy.

Time Series Models

How can we accurately predict rentals per hour?

Features of Ride Count Time Series

- Restricted range - counts are > 0.
- Linear trend - Counts are increasing over time.
- Multiple seasonalities due to high frequency and length of time series.
- Non-constant variance - Amplitude of peaks go up over time (i.e. counts in summer of 2012 $>$ counts in summer of 2011).
- Use log transformation to stabilize variance.
- Interpret-able & allows modelling of series dynamics additively.

Decomposable Models

Class of models that break down time series into predictable components.

$$Y_t \; = \; \text{Trend}_t \; + \; \text{Seasonal}_t + \text{Remainder}_t \tag{1}$$

- Very intuitive.
- Highly flexible (in my opinion).

Types of Decomposition Models

1. **Classical Decomposition Models**:-
   1.1 STLF - STL Decomposition + ETS Forecasts.

   1.2 Linear Regression with ARIMA errors (Not included in presentation)

2. **General Additive Models (GAM)**:-
   2.1 Facebook's General Additive Model (FBG)

   2.2 XGBoost (XGB)

How Does STLF Work?

- Uses seasonal & trend decomposition with LOESS (STL) algorithm to decompose time series as:

$$Y_t = S_t + A_t \tag{2}$$

- Uses last periodic observation for each seasonal component S_t to predict future seasonal values (seasonal naive method)

- Uses exponential smoothing (ETS) to forecast seasonally adjusted component A_t (trend + remainder).

How Does Facebook's GAM Work?

- FBG is a decomposable model that can be expressed as:

$$Y_t = g_t + s_t + h_t + X_t + \epsilon_t \tag{3}$$

- Can model trend (g_t) accounting for unexpected level changes (e.g. using piece-wise linear regression).
- Can include multiple Fourier terms (sine & cosine pairs) to model each type of seasonality (s_t) precisely.
- Uses normal prior to account for one-time effects of holidays (h_t).
- Easy to add multiple regressors due to model's linear formulation (e.g. temperature).

How Does XGBoost Work?

- Boosting is an ensemble learning method.
- Final predictions are usually averaged predictions from multiple weak learners (usually non-deep decision trees).
- Trees are built sequentially.
- Each subsequent tree aims to reduce errors from the previous tree (i.e. the next tree in the sequence is fit onto the residuals from the previous iteration).
- Gradient boosting (GB) fits weak learners to the gradient of the loss function (LF). Allows it to work on *all* commonly used LFs.
- XGB is an efficient implementation of GB.

Results

Comparing predictive accuracy of different models

Train - Test Split

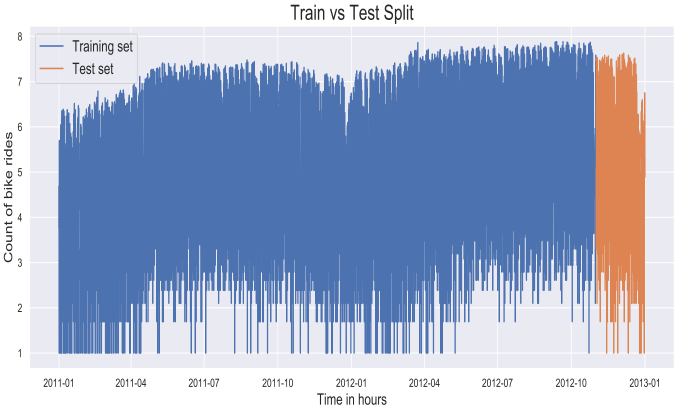

Figure 9: Training models to predict counts in the last 2 months of 2012

STLF Results

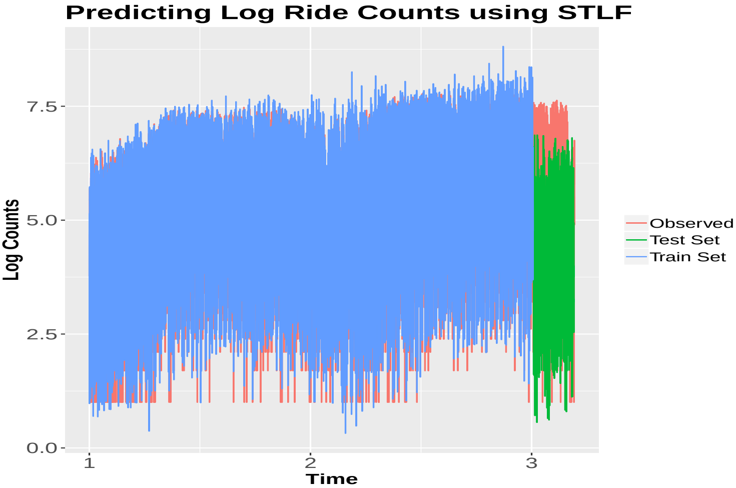

Figure 10: In-sample and out-of-sample predictions from STLF model

STLF Results - Close Up

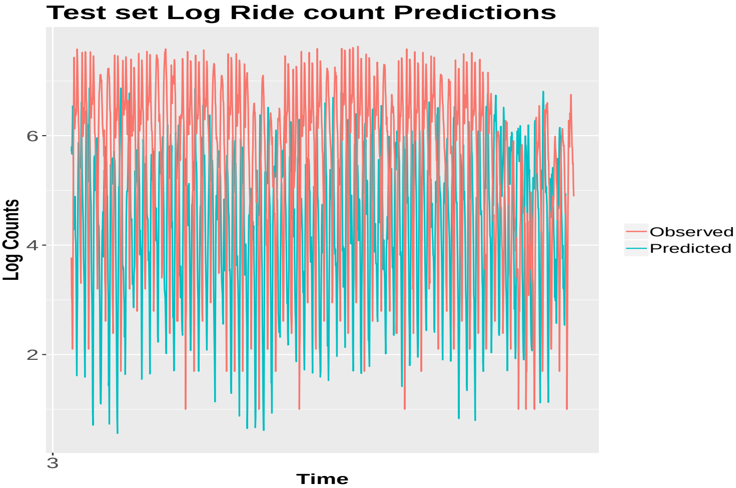

Figure 11: STLF seems to consistently under predict counts

What did STLF learn?

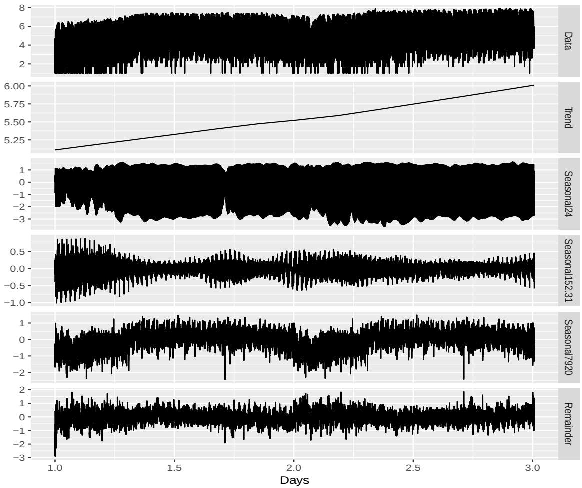

Figure 12: STL applied iteratively to filter out multiple seasonalities in train set

STLF - No inference possible

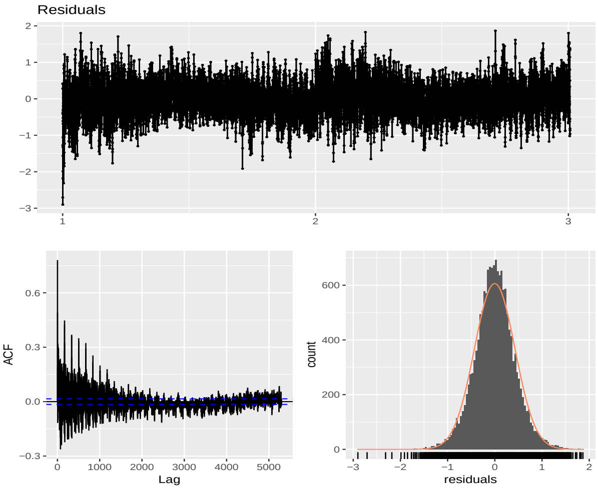

Figure 13: Decomposition does not capture all the correlation in the data

Facebook's GAM - Base Model + Holiday Effects

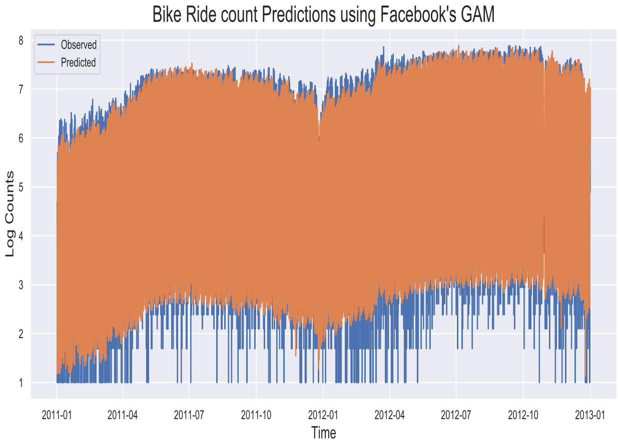

Figure 14: Facebook's base GAM produces accurate predictions

Facebook's GAM - Base Model (Close Up)

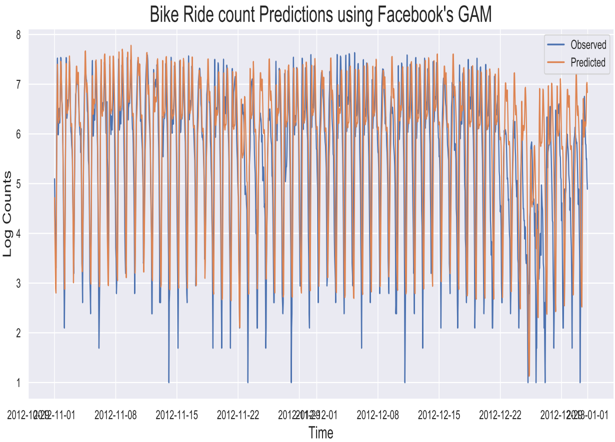

Figure 15: Captures large holiday spikes relatively well

Facebook's GAM - Adding EC Variables

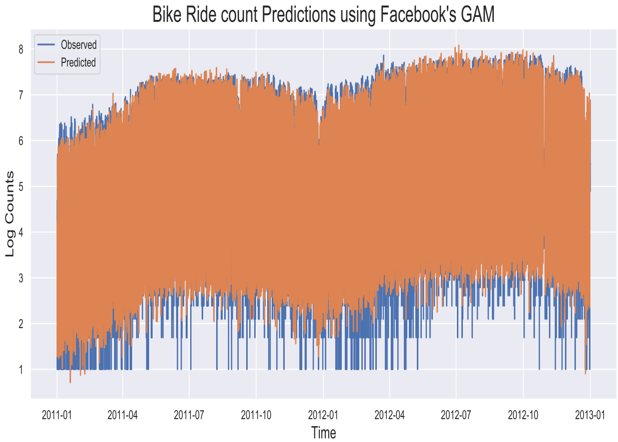

Figure 16: Adding EC regressors significantly improves *low count* predictions

What did Facebook's GAM Learn?

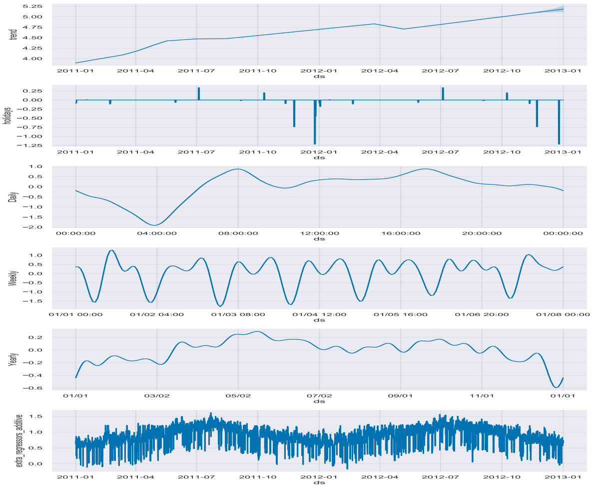

Figure 17: Model is able to accurately learn daily and annual seasonalities

XGBoost Predictions

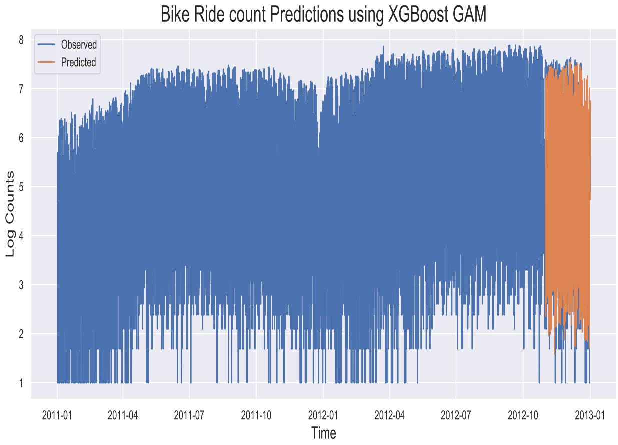

Figure 18: Model is unable to predict low counts as well as FBG

What did XGBoost learn?

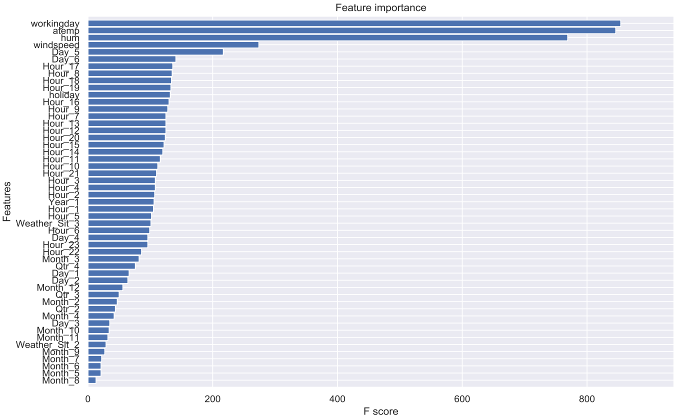

Figure 19: Time (hour, month, year, day type), EC and weather condition variables were the most useful in maximizing accuracy gains

Prediction Accuracy

Model	Training Set			Test Set		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
STLF	0.28	0.21	0.049	2.64	2.19	0.45
FBG - B	0.71	0.52	0.17	0.74	0.53	0.13
FBG - TB	0.48	0.37	0.12	0.55	0.36	0.087
FBG - TBA	0.46	0.34	0.11	0.51	0.33	0.084
XGB - 1000	0.35	0.24	0.061	0.51	0.37	0.087

Table 1: FBG - B: Facebook GAM - Baseline model,FBG - TB: Facebook GAM - Tuned Baseline model, FBG - TBA: Facebook GAM - Tuned Baseline + Add. regressors, XBG - 1000 - XGBoost Model with 1000 boosted trees

Concluding Remarks

- STL is very sensitive to choice of seasonal smoothing parameter.
- It uses seasonal differencing to estimate seasonal effects. Not very appropriate for high freq. data.
- FBG uses priors (numeric values) as a form of regularization. Difficult to specify without sensible priors.
- Accuracy metrics computed are point estimates. Can generate distributions for each using grid search + CV.
- XGB performs as well as FBG - TBA and takes less time to fit.
- More in-depth model details are available in my GitHub Repository