PARTICULATE MATTER 2.5 FORECAST

SUPREET DESHPANDE, NICK STERLING, HEQING SUN

CHELSEA THOMAS, CHESANEY WYSE

SEPT 13, 2019

TABLE OF CONTENTS

PARTICULATE MATTER 2.5 FORECAST

EXECUTIVE SUMMARY

The Environmental Protection Agency (EPA) sought out assistance in investigating the Air Quality at the Millbrook School station located in Wake County, North Carolina. Moreover, they would like to specifically focus on the average monthly Particulate Matter 2.5 ($PM_{2.5}$).

After trying out multiple Exponential Smoothing Models (ESMs) to forecast the last six months of air quality, we found the Holt-Winters Multiplicative model to be the best option. Using the Mean Absolute Percent Error (MAPE) as our accuracy measure, we found that our forecast model was off by only 20.9%. While other ESMs displayed similar accuracy rates, the Holt-Winters Multiplicative forecast model differentiated itself with its ability to capture seasonality. We specifically selected a multiplicative model after we observed that this model better explained the data, with the least error.

Overall in terms of trend, the air quality subtly improved from 2014 to the middle of 2017, and then slowly began to decline through the end of 2018.

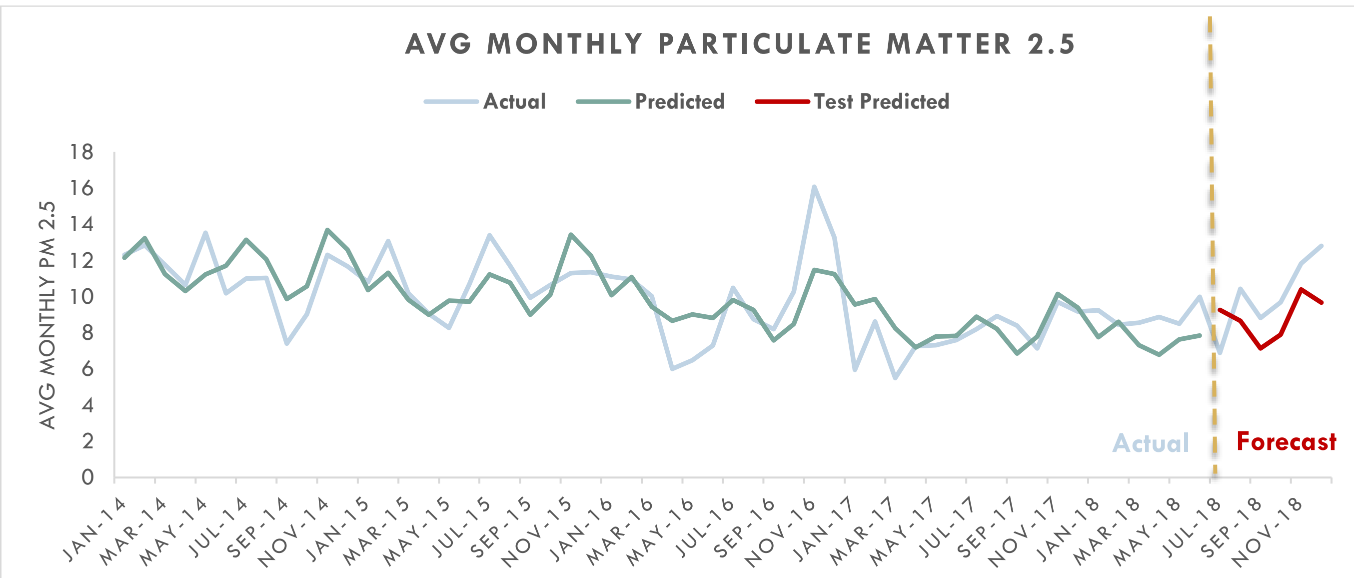

FIGURE 1: AVERAGE MONTHLY $PM_{2.5}$

RESULTS

TREND

The Seasonal and Trend using LOESS (STL) Decomposition identified an interesting trend in air quality. The solid red line in the middle section of Figure 2 displays this trend. From 2014 to mid-2017 we see a slight overall decrease in $PM_{2.5}$. Mid-2017 we notice a transition, as the trend shifts upward; increasing at a similar rate as it had previously decreased.

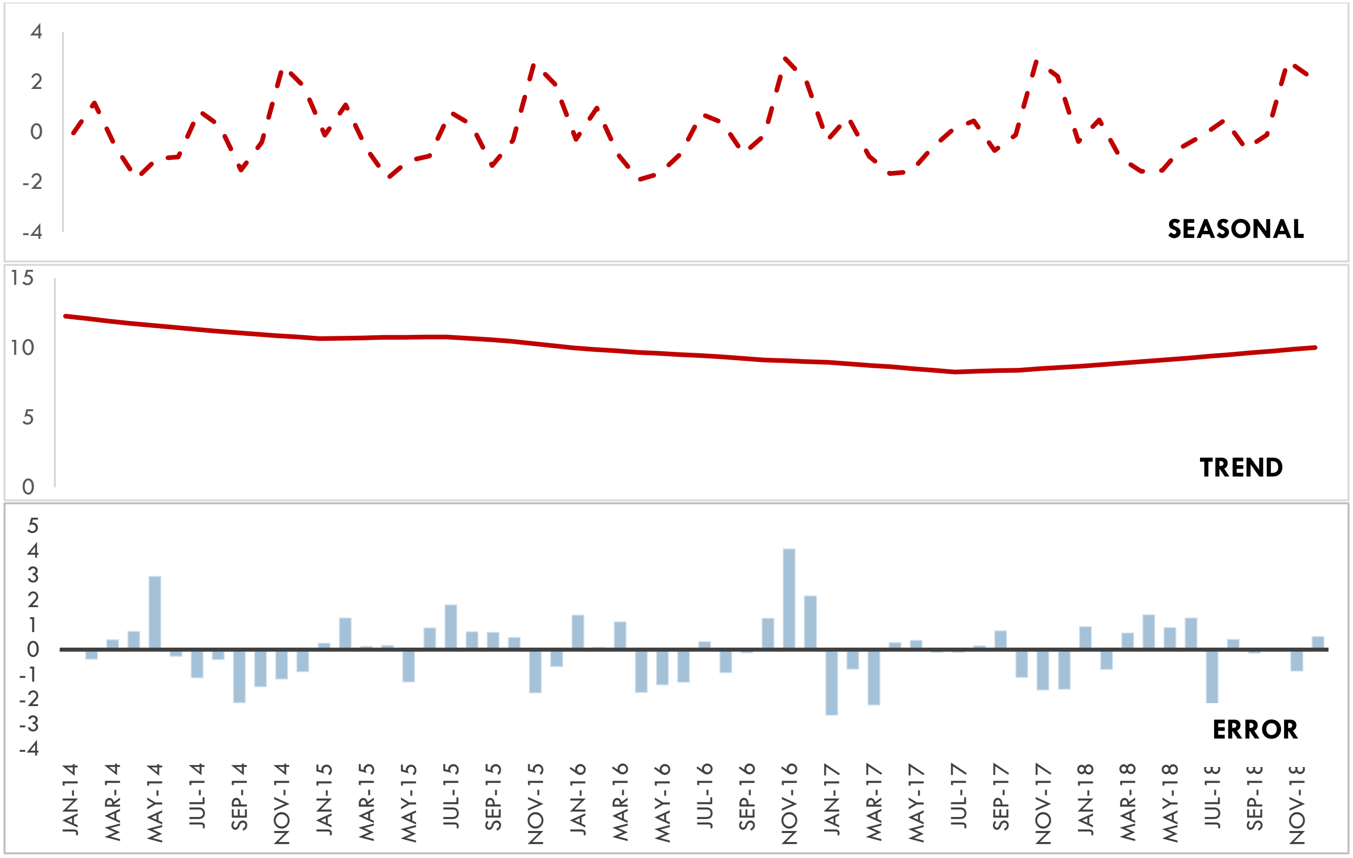

FIGURE 2: STL DECOMPOSITION

MODEL SELECTION

With our two main criteria points being pattern representation and overall model accuracy, the multiplicative Holt-Winters model proved to be the best fit. As seen in the upper section of Figure 2, there is a clear seasonality piece. Trend-only models fail to capture this informative pattern. The multiplicative Holt-Winters model displayed the least error out of the models that do contain this seasonality piece.

RECOMMENDATIONS

After testing various models, we determined that the multiplicative Holt-Winters model best represents the data. We recommend the use of this model in forecasting average monthly $PM_{2.5}$.

ADVANTAGES

We believe it is essential to consider both the patterns of the data and model accuracy. In this instance, our final model captures the trend and seasonality components, while maintaining an acceptable accuracy level. The STL decomposition uncovers the necessary components of the model, and accuracy measurements such as the MAPE and MAE Mean Absolute Error) help confirm the model selection.

ADDITIONAL CONSIDERATIONS

Further exploration may provide a better understanding of the fluctuation of air quality at the Millbrook School Station. Comparing conditions during peaks and troughs of $PM_{2.5}$ may reveal insights as to why there is a consistent fluctuation. Additionally, as solutions to the increase in $PM_{2.5}$ are explored, it will be useful to monitor the overall trend. This will help in determining the effectiveness of solutions.

METHODOLOGY

DATA DESCRIPTION AND CLEANING

Daily averages of $PM_{2.5}$ concentrations were recorded at the Millbrook School Station over the five years from January 2014 through December 2018. Missing values were inserted for the 353 days without observations. Monthly averages of $PM_{2.5}$ concentrations were calculated from the daily averages, ignoring missing values. The data was split into a training set, consisting of observations from January 2014 through June 2018, and a test set, containing the six months recorded after June 2018.

DECOMPOSITION

The training set was employed to construct a Seasonal and Trend using LOESS (STL) decomposition. The STL method was selected over the classical decomposition because a changing trend was observed. Plots visualizing the trend/cycle, and seasonally adjusted components were constructed using the STL decomposition.

EXPONENTIAL SMOOTHING MODEL

We constructed five exponential smoothing models and calculated the MAPE of the forecast produced by each on the test set. The Holt-Winters multiplicative model was selected based on MAPE values and relative explanatory value.

ANALYSIS

STL DECOMPOSITION

With the aggregated monthly training data set, we created an STL decomposition. The output, demonstrated in Figure 2, provided an in-depth look at the seasonal, trend, and error components of the data.

Figure 3 overlays actual $PM_{2.5}$ values with the trend/cycle component of the data. We observed that the trend gradually decreased over time until mid-2017, and then began increasing since then.

Additionally, Figure 4 overlays actual $PM_{2.5}$ values with the seasonal component of the data. We observed that the $PM_{2.5}$ values spike during November and take a dip during April.

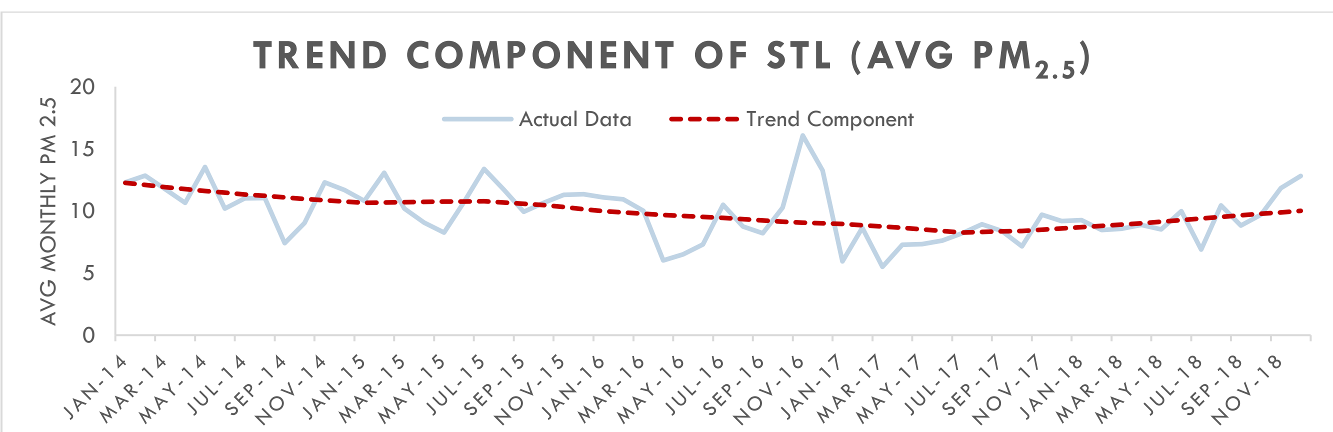

FIGURE 3: TREND COMPONENTS

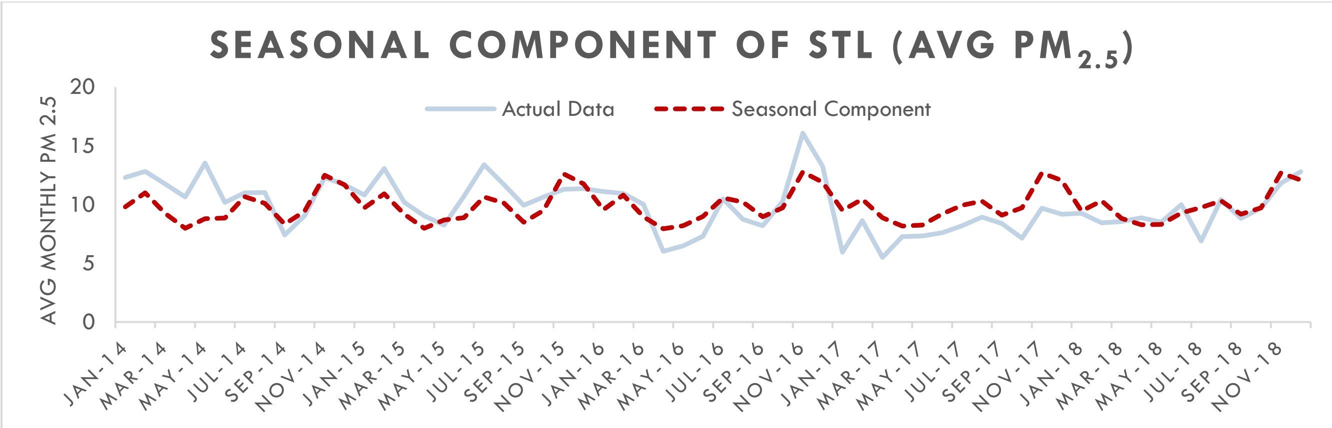

FIGURE 4: SEASONAL COMPONENTS

ESM EVALUATION

Using a 54-month training data set, we built five exponential smoothing models and evaluated each model by forecasting the last six months of data.

The five models we explored were very close in terms of the accuracy measurements. Although the Simple, Linear, and Damped Trend ESMs gave smaller a MAPE than the Holt-Winters Models, their predictions are all straight lines. Therefore, we chose the Holt-Winters Multiplicative Model to capture the trend and seasonal components seen in the STL decomposition. Considering over-predictions might be problematic in calculating the MAPE, we also looked at the MAE for confirmation. Table 1 displays MAPE and MAE values for all five ESM models we evaluated.

TABLE 1: MAPE AND MAE FOR ESMs

Model	MAPE (%)	MAE
Simple	18.10	1.89
Linear/Holt	20.75	2.20
Damped Trend	19.07	2.01
Holt-Winters Additive	22.85	2.25
Holt-Winters Multiplicative	20.90	2.03

CONCLUSION

For the decomposition of the air quality data, we decided to use the STL method. The versatility and robustness of the STL method allowed us to view how the seasonal component changed over time. Upon exploring the patterns of the data, we selected the Holt-Winters exponential smoothing model to reflect both trend and seasonality. When comparing to the additive model, we observed that a multiplicative model better explained the data, with the least error.

As seen in Figure 5 below, the Holt-Winters model closely represents the actual data. The forecast carries on the trend of the data well.

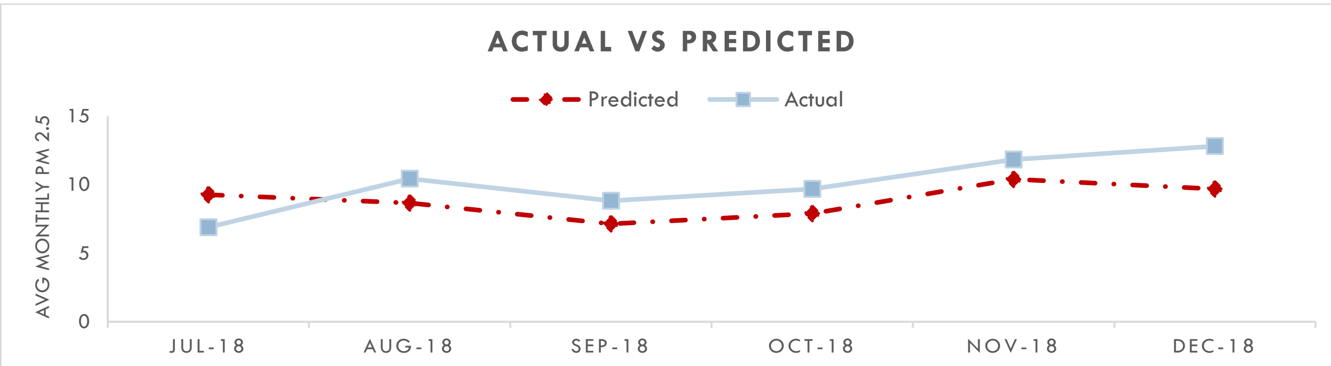

FIGURE 5: HOLT WINTER'S ACTUAL VS PREDICTED

Moving forward, we recommend that the EPA take a closer look into what factors caused the air quality to improve from 2014 to the middle of 2017, as well as what caused the air quality to deteriorate after 2017.