Modeling Particulate Matter 2.5 as a Time Series

Ethan Biegeleisen

Manhattan College
MATG 635 - Probabilistic Methods
Spring 2020

What is the Project Goal?

- Use a time series to model Particulate Matter 2.5 (PM 2.5) levels from the Environmental Protection Agency's (EPA) Air Quality Index (AQI) Data Values Report [4]
- First step: Model Manhattan's daily PM 2.5 levels as a time series [4]
- Second step: Approximate Manhattan's PM 2.5 levels with the Holt-Winters method
- Third step: Predict future PM 2.5 levels in Manhattan
- Fourth step: Compare predicted and observed PM 2.5 levels in Manhattan for 2019

What is Particulate Matter 2.5?

- An air pollutant often coming from vehicle exhausts and burning fuel [3]
- Irritates places like eyes and throat [3]
- Causes coughing, sneezing, and shortness of breath [3]
- Worsens medical conditions like heart disease and asthma [3]

How is the Data Structured?

- The EPA's website has PM 2.5 levels from 1999 onward, but Manhattan's PM 2.5 levels are only recorded daily from 2013 onward
- Data sets are searched for by year and location (by general city area or by county)
- AQI category is determined by the daily pollutant level [1]
- AQI category is Good if the level is 0-50, Moderate if the level is 51-100, Unhealthy for Sensitive Groups if the level is 101-150, and Unhealthy if the level is 151-200 [1]
- Pollutant level can go above 200 into more severe AQI categories, but that is less relevant for this data [1]

What Does the Data Look Like?

- Note: Pollutant levels are not measured at the same site every day

Date	PM2.5 AQI	AQI Category	Site Name	Site ID	Source
1/1/2018	51	Moderate	Intermediate School 143	36-061-01	AQS
1/2/2018	56	Moderate	Intermediate School 143	36-061-01	AQS
1/3/2018	61	Moderate	Intermediate School 143	36-061-01	AQS
1/4/2018	48	Good	PS 19	36-061-01	AQS
1/5/2018	48	Good	Intermediate School 143	36-061-01	AQS
1/6/2018	46	Good	Intermediate School 143	36-061-01	AQS
1/7/2018	52	Moderate	Intermediate School 143	36-061-01	AQS
1/8/2018	67	Moderate	Intermediate School 143	36-061-01	AQS
1/9/2018	61	Moderate	Intermediate School 143	36-061-01	AQS
1/10/2018	55	Moderate	PS 19	36-061-01	AQS
1/11/2018	74	Moderate	IS 45	36-061-00	AQS
1/12/2018	41	Good	PS 19	36-061-01	AQS
1/13/2018	44	Good	Intermediate School 143	36-061-01	AQS
1/14/2018	43	Good	Intermediate School 143	36-061-01	AQS
1/15/2018	58	Moderate	PS 19	36-061-01	AQS

What is Exponential Smoothing? [2]

- Exponential model for x at time t: $x_t = \mu_t + \omega_t$
- μ_t is the non-stationary mean at time t
- ω_t are independent random deviations with mean 0 and standard deviation σ
- Approximate a_t, the exponentially weighted moving average at time t:

$$a_t = \alpha x_t + (1-\alpha)a_{t-1}, \qquad 0 < \alpha < 1 \tag{1}$$

- α is the smoothing parameter
- a_t gets closer to x_t as α approaches 1
- Holt-Winters method applies this approach to seasonally adjusted mean, change in this mean between periods, and seasonal effect

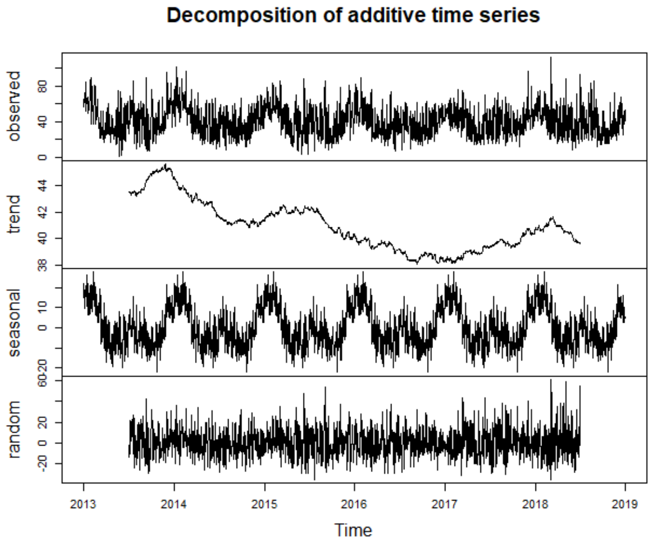
What Forecasting Model is Used? [2]

- The forecasting model applied from the Holt-Winters method estimates x_{n+k} over period p as:

$$\hat{x}_{n+k|n} = a_n + kb_n + s_{n+k-p}, \qquad k \leq p \tag{2}$$

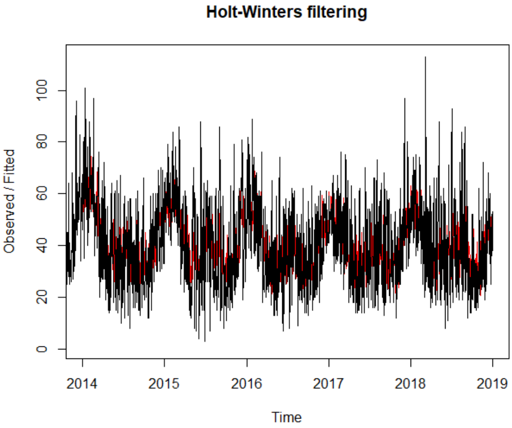
- $a_n + kb_n$ is the expected seasonally adjusted mean at time $n + k$
- s_{n+k-p} is the exponentially weighted estimate of the seasonal effect at time $n = k - p$
- RStudio's predict function forecasts based on the Holt-Winters fit

Decomposition of 2013-2018 Data

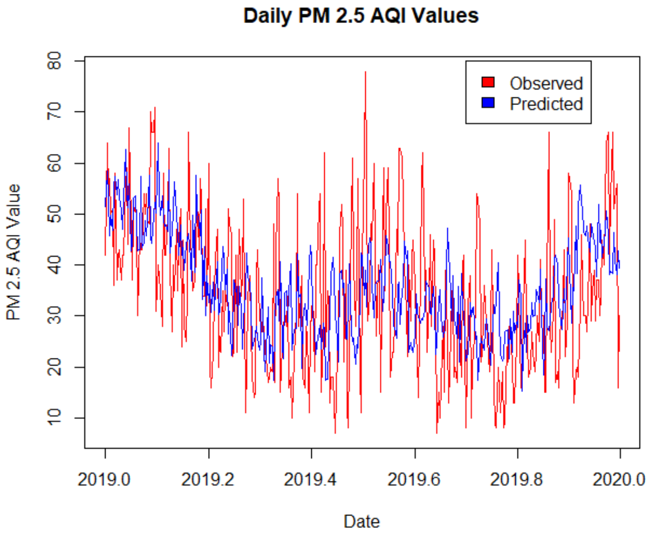


Observed Data (Black) vs Holt-Winters Fit (Red)

- Note: In order to run the function with less than 8 full years of data, the entry from February 29, 2016 was excluded

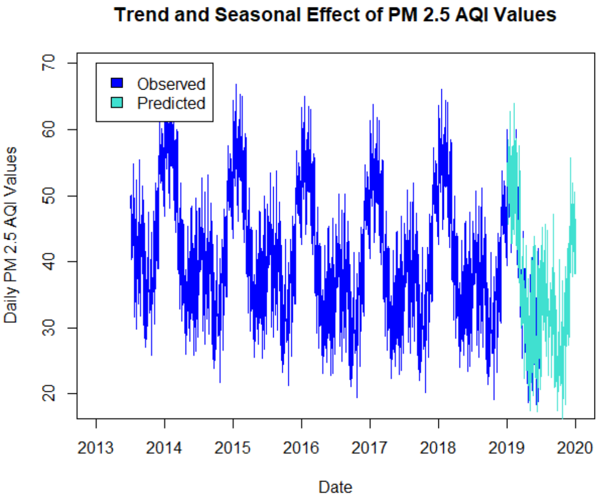


Observed vs Predicted 2019 PM 2.5 Levels



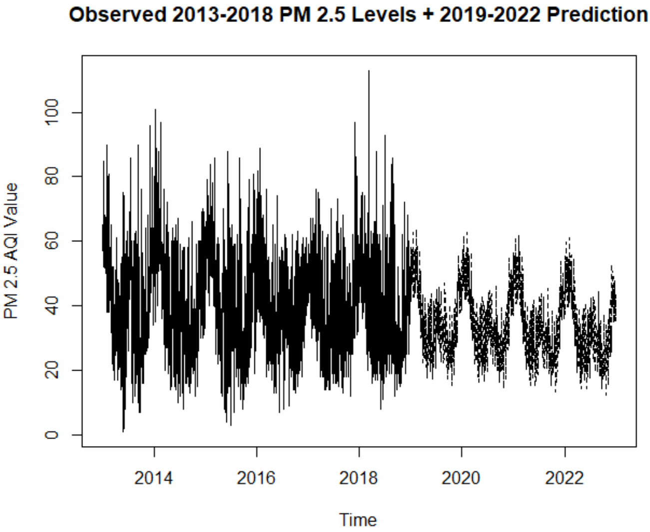
Filtering Out Noise from the 2013-2019 Data

- The blue plot is the trend + the seasonal effect of the observed data



Long-Term Prediction After 2018

- Note: The 2013-2018 plot is the original data with the noise included



References

Thanks for listening! Any questions?

References:

[1] AirNow (June 2019). Air Quality Index (AQI) Basics. Retrived from: `https://www.airnow.gov/index.cfm?action=aqibasics.aqi`

[2] Cowpertwait, P. S. P., Metcalfe, A. V. (2009), *Introductory Time Series with R*. Springer Science + Business Media.

[3] Department of Health (February 2018). Fine Particles (PM 2.5) Questions and Answers. Retrieved from: `https://www.health.ny.gov/environmental/indoors/air/pmq_a.htm`

[4] Environmental Protection Agency (2020). Air Quality Index Daily Values Report. Retrieved from: `https://www.epa.gov/outdoor-air-quality-data/air-quality-index-daily-values-report`

Extra: Holt-Winters Method in More Detail [2]

- Builds on exponential smoothing
- Searches for level (seasonally adjusted mean), slope (change in level between periods), and seasonal effect
- Key equations over period p:

$$a_t = \alpha(x_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{3}$$

$$b_t = \beta(a_t - a_{t-1}) = (1-\beta)b_{t-1} \tag{4}$$

$$s_t = \gamma(x_t - a_t) + (1-\gamma)s_{t-p} \tag{5}$$

- a_t, b_t, and s_t are the estimated level, slope, and seasonal effects at time t
- α, β, and γ are the smoothing parameters
- RStudio's HoltWinters function approximates α, β, and γ