

## **One Year In: An Update on the Lehigh Valley Breathes Project**

It's been about a year since we launched the public side of the Lehigh Valley Breathes project. We thought now might be a good time to look at where we are after a year.

The project design was for 40 PurpleAir monitors – 20 for each county – to be installed according to a plan worked out with our air quality expert at Carnegie Mellon. Those 40 monitors were to be located within eleven different categories of locations with at least two monitors in each category. Those categories were based on whether they were rural or urban and the monitor's proximity to traffic, highways, and warehouses. The details on the locations were explained in the October 2023 update.

Once all monitors are installed, it is necessary to collect data for at least a year so that they are exposed to all four seasons and as many disparate weather conditions as possible. Once a full year of data has been collected, the data needs to be analyzed to determine what conclusions can be drawn as to the effect of weather, location, traffic, wind, and other possible parameters on the level of PM 2.5. The project can then use that data analysis to identify reasonable policies and modifications that might be useful to mitigate the pollution.

### **So Where Are We?**

It's been a year full of learning experiences.

We first learned that the calibration process took much more time than was initially anticipated, which delayed installation since the monitors could not be used until they were calibrated to the DEP monitor in Freemansburg.

We had a great response from the public after our solicitation in August 2023 for locations for the monitors. Unfortunately, many of the locations did not fit into the project design primarily because most of the priority locations we needed were close to highways, warehouses, industrial parks, and other areas that are not optimal locations for private residences. Given the circumstances, it took us much longer than expected to find suitable locations for the monitors. As of now, we have 17 monitors installed in Lehigh County, with another two ready to be installed and 16 monitors installed in Northampton County, along with another two ready for installation. We have also been able to supplement our project monitors with four additional monitors owned by Lehigh University and three citizen-owned monitors that were calibrated to project standards and provide additional data.

As a result, we have some monitors that have been in place for more than a year (some were installed in July 2023), most have been in place for some portion of the year, and five are still waiting for installation. That means the start date for the one year of data acquisition will be this September, much later than expected. Data acquisition will run through August 2025, and the complete analysis will take place after that.

However, we have been able to do some preliminary analysis during the first year. In our January 2024, update we reported some variation in PM<sub>2.5</sub> levels based on geographic location. You can see the details in the January update. In the February update, we reported some very early findings on the differences between the monitor at Leaser Lake, our more rural location, and a number of other locations in both urban and rural sites.

Our May update included some preliminary data analysis on the effect of certain Meteorological variables on PM<sub>2.5</sub> pollution in the Lehigh Valley. These included windspeed and temperature inversions at the DEP monitors in the Valley. While the May update didn't use LVB data, it provided a direction for future analysis using specifically LVB data.

Based on identifying windspeed as an important variable in determining PM<sub>2.5</sub> pollution, Jeremy Mack, our data visualization guru, has added three icons to the Shiny App showing wind direction at each of the three locations. At the moment, it identifies either "no wind" or "wind direction" for each location. We expect to add wind speed to the icons in the future.

Even though this has taken significantly more time than we expected, we are still confident that the information we are gathering will deliver results that can be used to meet the project objectives of actionable PM<sub>2.5</sub> data.

## **A Look to the Future**

In the meantime, we are excited to share with you some initial work being done on using the data we're collecting for time series analysis.

Every two minutes, the database for the Lehigh Valley Breathes project grows larger. Currently, the project has collected roughly 6,428,000 readings of air quality at different locations throughout the Lehigh Valley. By the time you finish reading this update, the total will grow by almost 200 more readings. So, one question you may have is, what do we plan to do with all of that data?

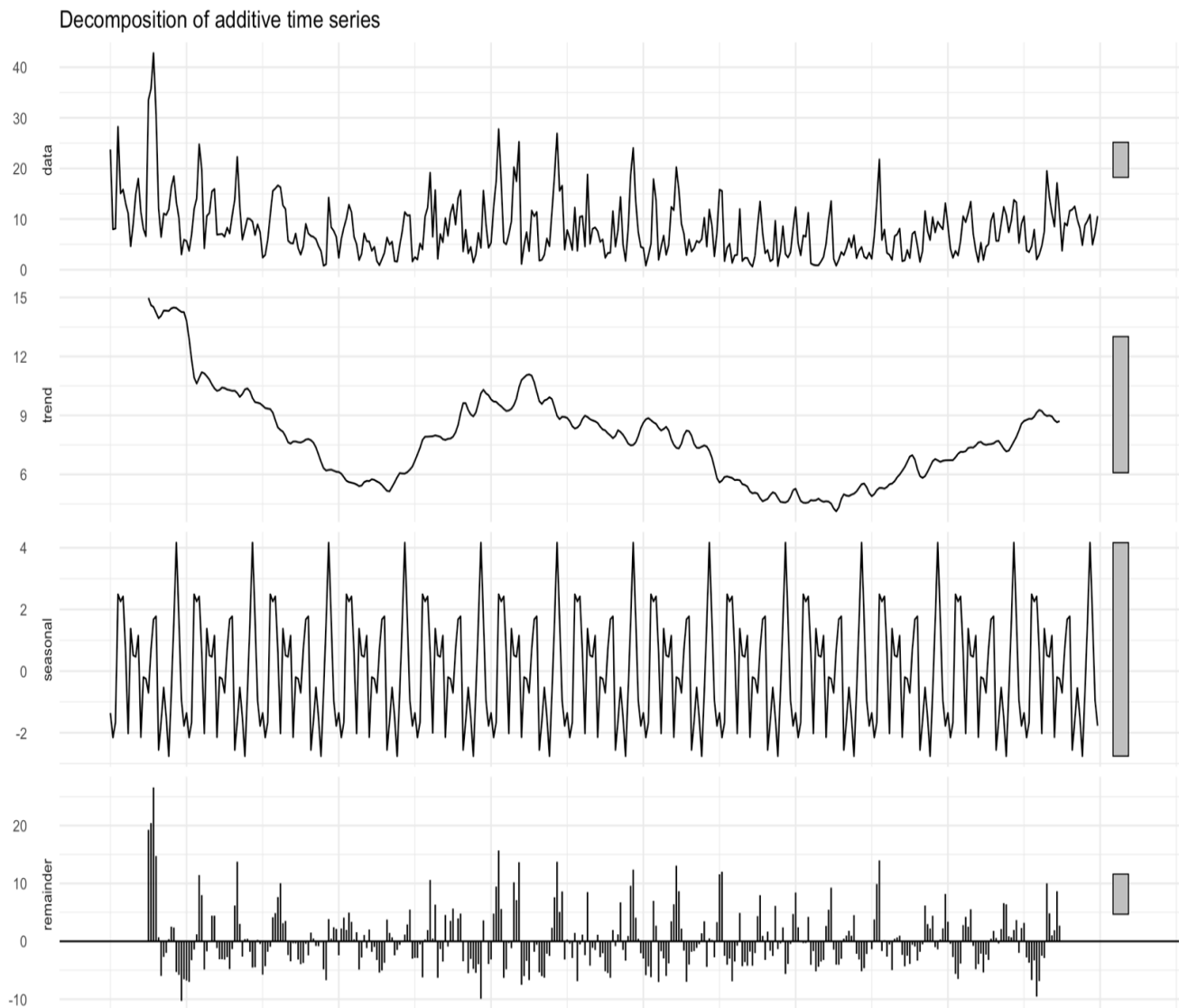
A few things have already been discussed in previous updates. For example, how do measurements of air quality differ between locations? Does truck traffic influence air quality? What about wind? How much impact does the weather have on air quality? Another interesting question might be, with such a large dataset, can we forecast, much like the weather, air quality throughout the Lehigh Valley? This latter question is the focus of this update.

## **Time series analysis**

First, let's cover what a time series is. A time series is a sequence of data points recorded in chronological order. These data points are often measured at consistent intervals, such as daily, monthly, or yearly. As noted above, our time series data is recorded at two-minute intervals.

Time series data can be used to conduct a time series analysis, which simply looks at how a variable, in our case, air quality, changes over time. This type of analysis looks to identify patterns in the data over certain time scales. Think about temperature. We experience daily patterns in temperature, where the hottest part of the day is around 2 PM. There are seasonal patterns in temperature - hot in the summer and cold in the winter. There's also a long-term trend in temperature where on average, temperatures have been getting warmer over time.

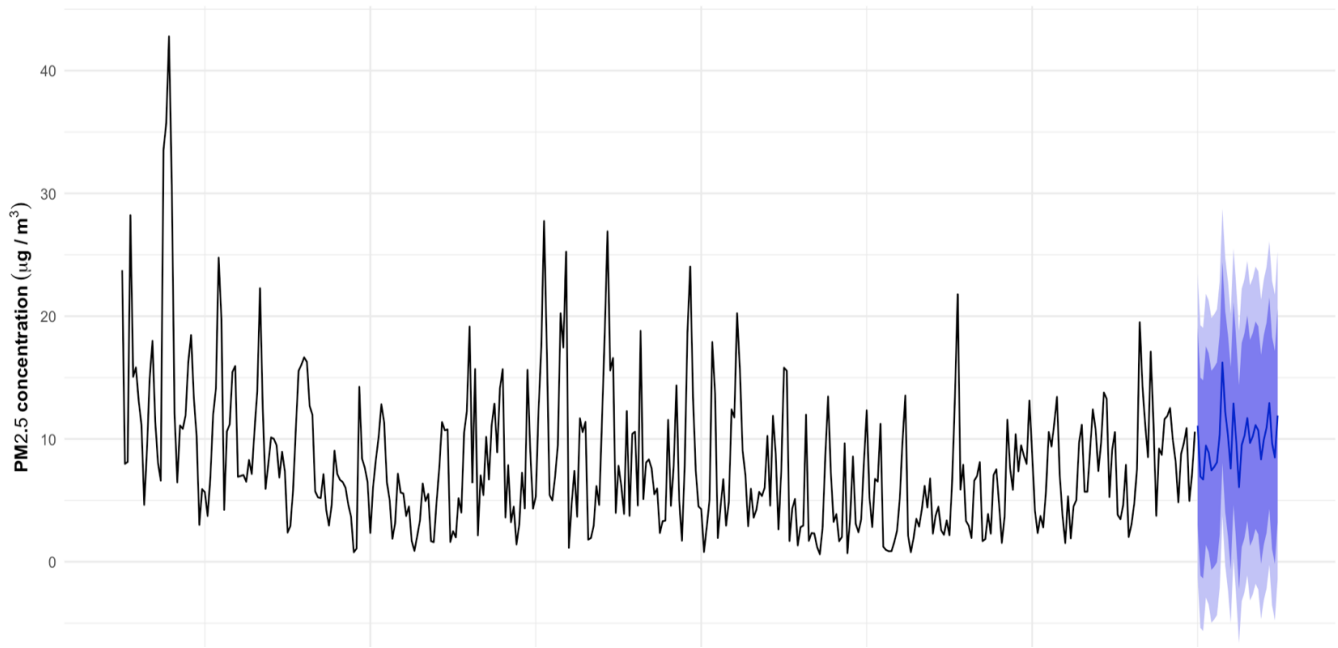
To do this, a time series is broken down (“decomposed”) into pieces – trends, seasonal variation, and random variation. Here’s an example from the sensor located in north Bethlehem:



## Forecasting

After uncovering patterns in the data, a time series analysis can be used for forecasting or predicting the future based on what's happened in the past. What might that look like? Here's an example of forecasting for the next 30 days for the sensor located in north Bethlehem:

Forecasted PM2.5 data for next 30 days in North Bethlehem



In this figure, the black line is the time series data – here we’re using daily averaged PM2.5 values, the blue line at the end is the forecasted (“predicted”) values, and the light and dark blue areas around the blue line represent the error or uncertainty around the prediction. For example, the darker blue area represents the 80% confidence interval, meaning there’s an 80% chance that the true value will be within that area. One way to increase the accuracy of the forecast or decrease the blue area around the blue line is to increase the length of the time series. So, as time goes on and the Lehigh Valley Breathes database grows larger and larger, the ability to analyze and predict air quality conditions increases as well.

We hope you’ve found this annual update and the time series analysis interesting. We’ll be back with another update at the end of September. In the meantime, breathe easy.