



Forecasting NO_x Atmospheric Concentrations for Pollution and Climate Change

By: Aaron Poulad, Elchonon Stein, and Ori Bach

[GitHub repo](#)



Our Team

- [Aaron Poulad](#) - post 1st- year student
- [Elchonon Stein](#) - post 1st-year student
- [Ori Bach](#) - post 1st-year student



Our Mentor

- [Ramesh Natarajan](#) - former Software Engineer, Google



High-level Project Goals and Deliverables

1. Collect, clean, and analyze Sentinel 5P satellite sensor data for atmospheric NO_x concentrations using Google Earth Engine APIs
2. Model the daily variations of the observed NO_x concentrations, perform a comparative analysis of model assumptions, and use the best model for forecasting.
3. Create a UI for selecting regions, obtaining satellite sensor data for these regions, and forecasting their NO_x concentrations



Technical Approach - What to expect ...

- This project implements a **modeling and forecasting system** for the heterogeneous, spatially-localized levels of atmospheric NO_x.
- To achieve this, we use the Sentinel 5P satellite sensor data available in Google Earth Engine to obtain **historical time series data of atmospheric column-averaged NO_x concentrations** over regions of interest.
- Using this data, we **develop state space forecasting models** to extract trends, seasonal effects and other important covariates from the data.
- Our models cover both **monthly-averaged data to identify long-term trends** as well as **daily-average data to identify short-term trends** and effects.
- Only results for daily-averaged data are presented here. Monthly-averaged TBD.

Importance of the Problem



Background

- **Monitor and forecast** atmospheric NOx concentrations, which are primarily caused by anthropogenic activity.
- Large emitters of NOx are due to combustion of fossil fuels for transportation, power generation, and industrial activity.
- Emissions can be mitigated by reducing activity, switching to alternative clean energy sources, and using emissions capture systems.
- NOx concentrations are important for two reasons:
 - **for environment and pollution studies** (e.g. smog), and health outcomes.
 - **as a proxy for the more hard-to-measure CO2 emissions**, the primary greenhouse gas responsible for long-term global warming.
- Goal is to provide reliable forecasts of NOx concentrations, leading to improved air quality management and reduced negative impacts on public health and the environment.

Looking From Space, Researchers Find Pollution Spiking Near E-Commerce Hubs

Research showed truck-related releases of nitrogen dioxide, which can cause asthma, concentrated around some 150,000 warehouses nationwide.



By Hiroko Tabuchi



Listen to this article · 4:03 min [Learn more](#)

July 24, 2024

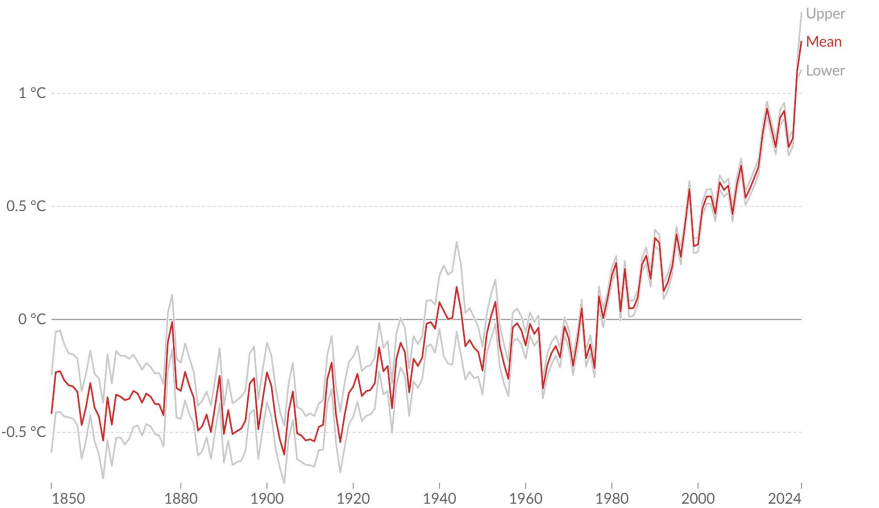


19



Average temperature anomaly, Global

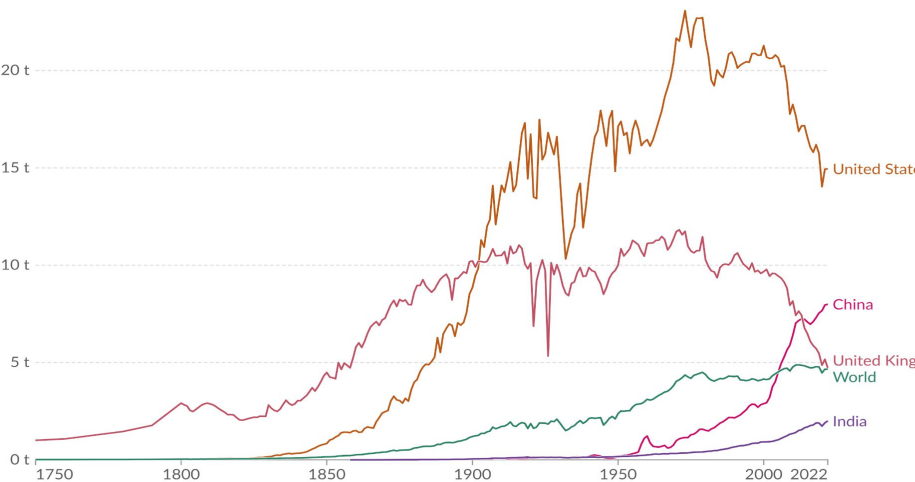
Global average land-sea temperature anomaly relative to the 1961-1990 average temperature.



Data source: Met Office Hadley Centre (2024) OurWorldInData.org/co2-and-greenhouse-gas-emissions | CC BY
Note: The gray lines represent the upper and lower bounds of the 95% confidence intervals.

Per capita CO₂ emissions

Carbon dioxide (CO₂) emissions from fossil fuels and industry¹. Land-use change is not included.



Data source: Global Carbon Budget (2023); Population based on various sources (2023)
OurWorldInData.org/co2-and-greenhouse-gas-emissions | CC BY

1. **Fossil emissions:** Fossil emissions measure the quantity of carbon dioxide (CO₂) emitted from the burning of fossil fuels, and directly from industrial processes such as cement and steel production. Fossil CO₂ includes emissions from coal, oil, gas, flaring, cement, steel, and other industrial processes. Fossil emissions do not include land use change, deforestation, soils, or vegetation.

Data Sources and Data Pipelines

Sentinel 5p Satellite

- Sentinel-5P carries a the TROPOspheric Monitoring Instrument(Tropomi) spectrometer which can sense ozone, methane, CO₂, NO₂, SO₂, formaldehyde, and aerosol in the atmosphere.
- TROPOMI (version 02.03.01[44](#)) measures column-averaged tropospheric column NO₂ (mol/m²) at an unprecedented native resolution of 3.5 km x 5 km



API to Get Data

Input

API takes a date range and a metropolitan area and a boolean option for cloud masking.

Getting Satellite Image

Api calls data from GEE and receives an image collection. The API then limits the image to metropolitan location.

Spatial Mean

The API takes the spatial mean over the metropolitan area to create a daily average.

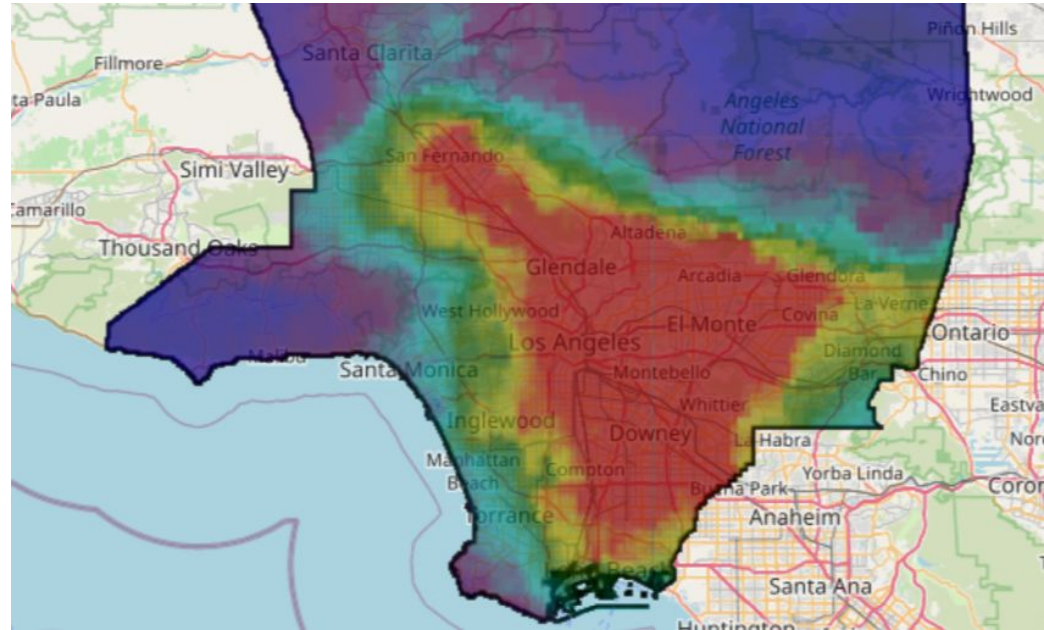
Data Frame for Time Series

The API converts this into a dataframe that can be graphed and modeled.

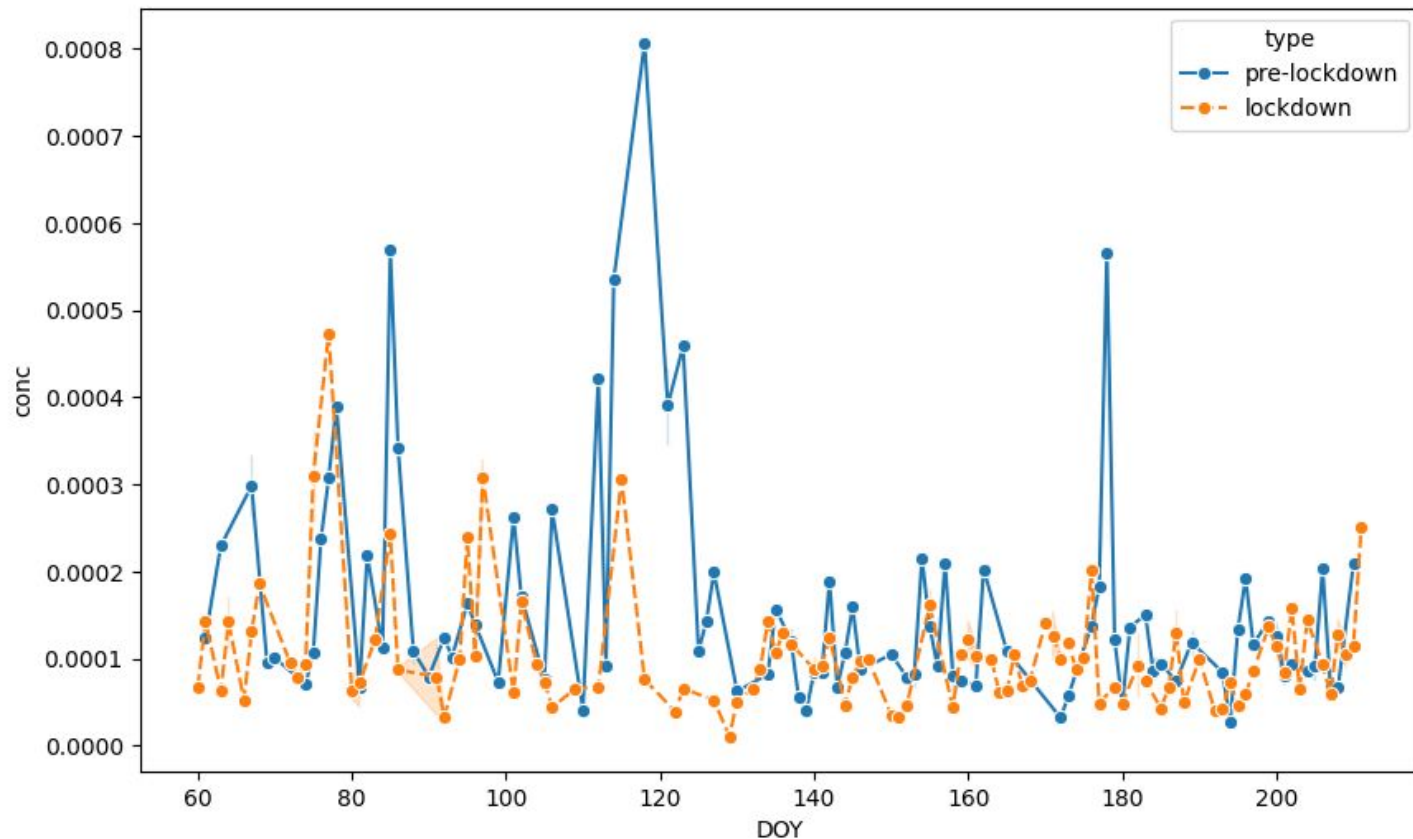
Google Earth Engine Satellite Data

- Used satellite data from google earth engine from the Sentinel-5P OFFL NO₂: Offline Nitrogen Dioxide dataset.

NO_x Median for March 2021, Los Angeles region

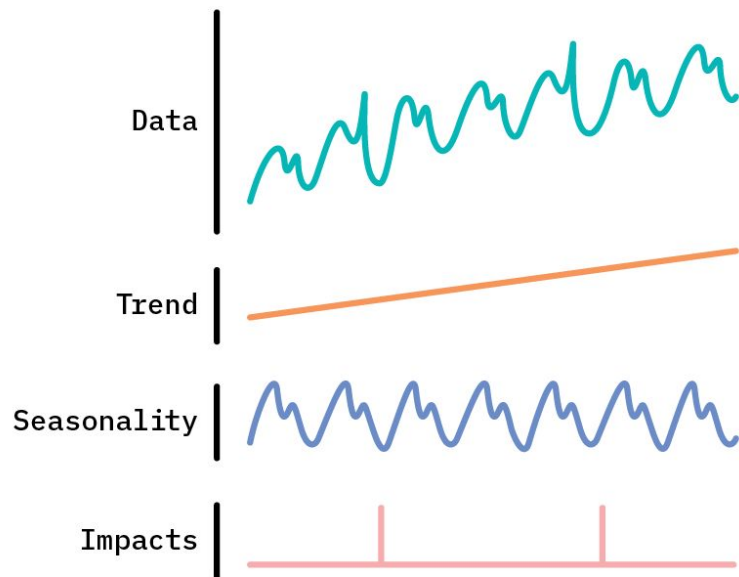


NOx
Emissions
Pre (2019)
and
During
(2020)
COVID
over NYC



Modeling and Forecasting approach

Structural Time Series (STS)



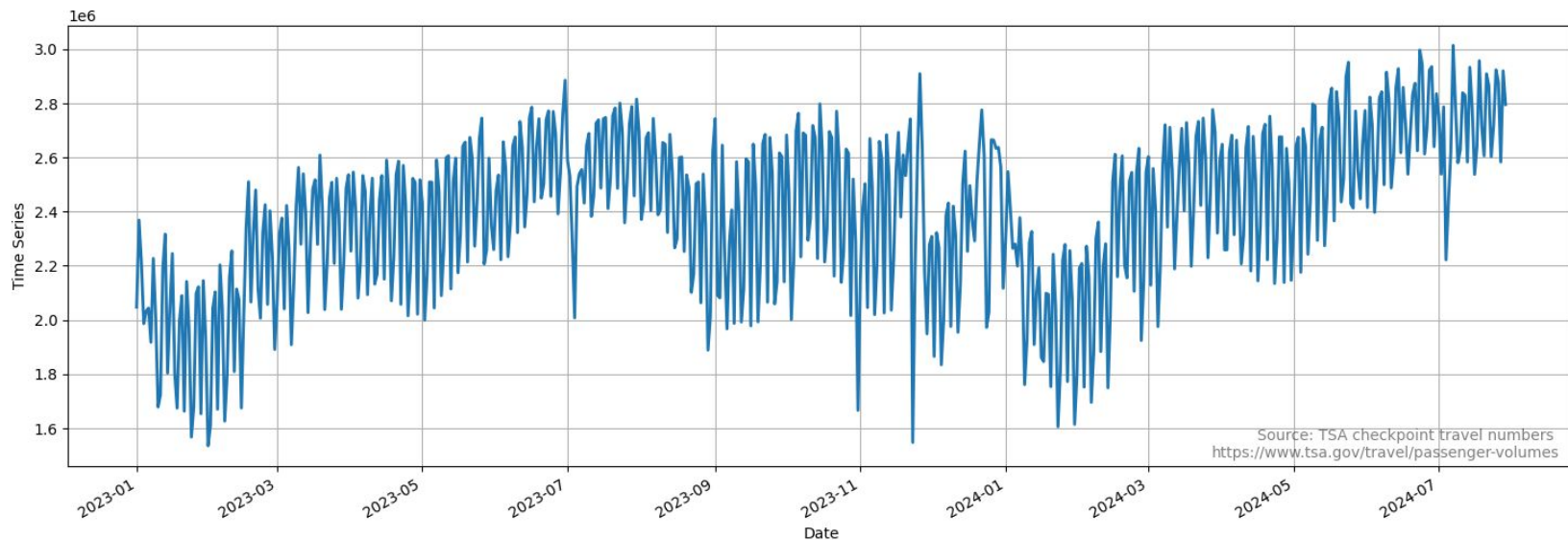
Decomposes observed data into a sum of latent components.

- These components capture various aspects of the data, including its trend, seasonality, and external covariates, e.g. sparse impacts due to holidays.
- Each component is modeled using Linear Gaussian State Space Models (LG-SSM), which have the capability to capture complex dynamics.
- Individual component models are combined into a comprehensive LG-SSM, which is then fitted to the observed data using Maximum Likelihood Estimation (MLE) to find the optimal parameters for the model.
- Backtesting is used to find the model with the best forecasting performance.

$$\text{Observation}(t) = \text{Trend}(t) + \text{Seasonality}(t) + \text{Impacts}(t) + \text{Noise}(t)$$

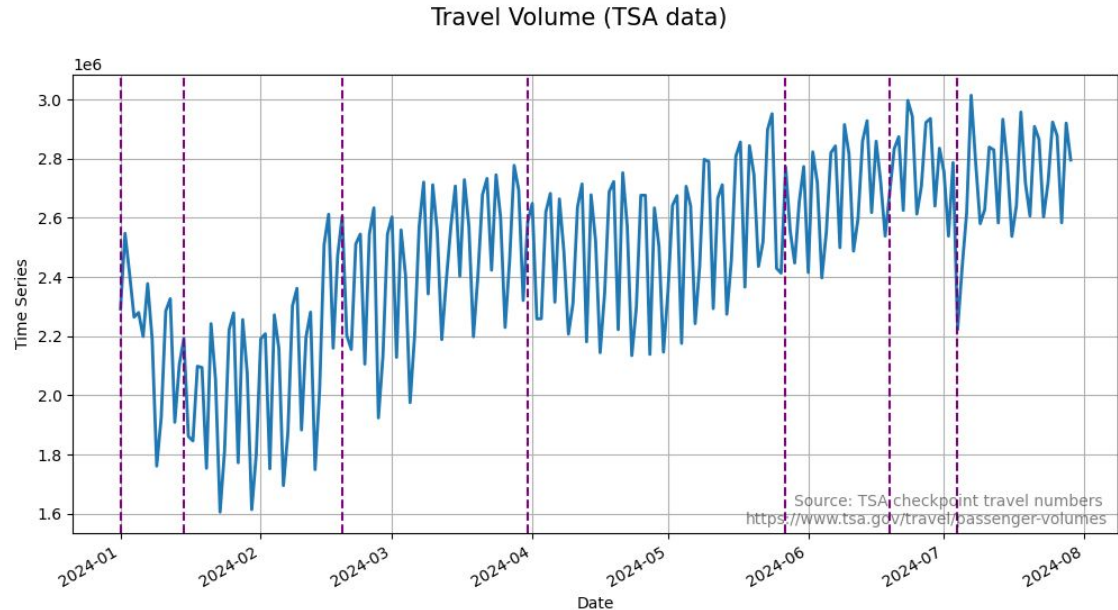
Example of STS - TSA Air Passenger Volumes

Travel Volume (TSA data)

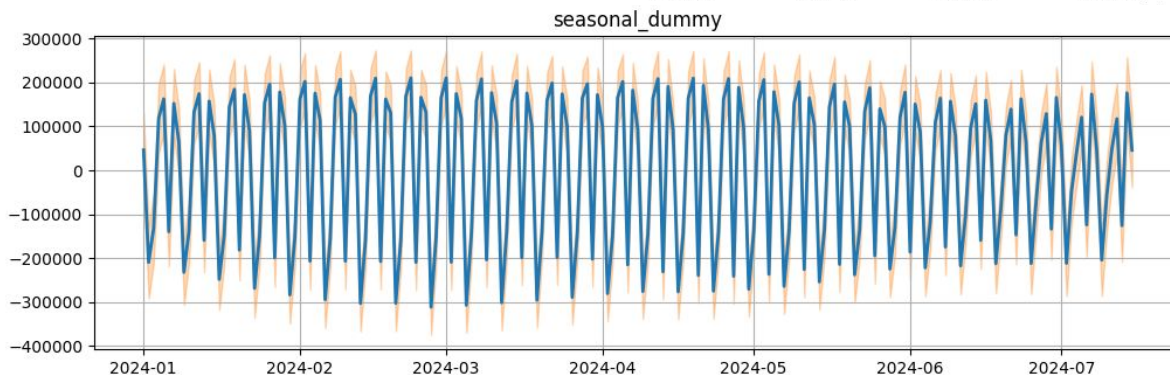
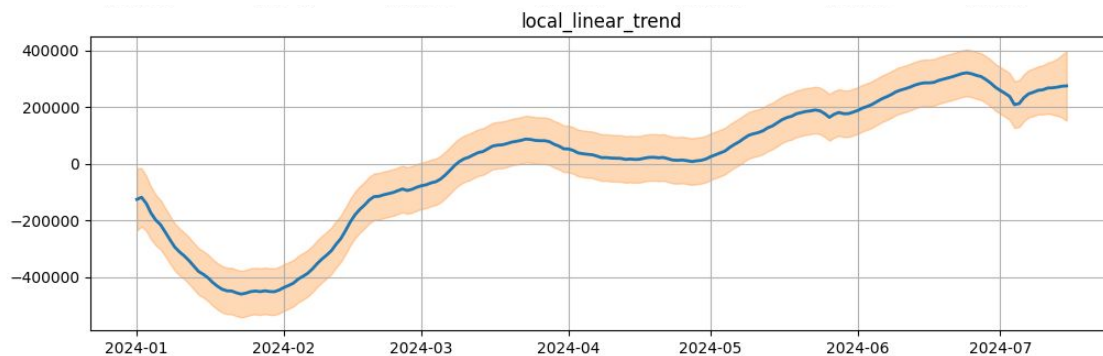


Example of STS - TSA Air Passenger Volumes YTD

- For our purpose of short-term forecasting, zooming in on just this year is helpful
- A trend as the year goes on is visible, as well as a **major trough on July 4** and a subsequent **record high on July 7**

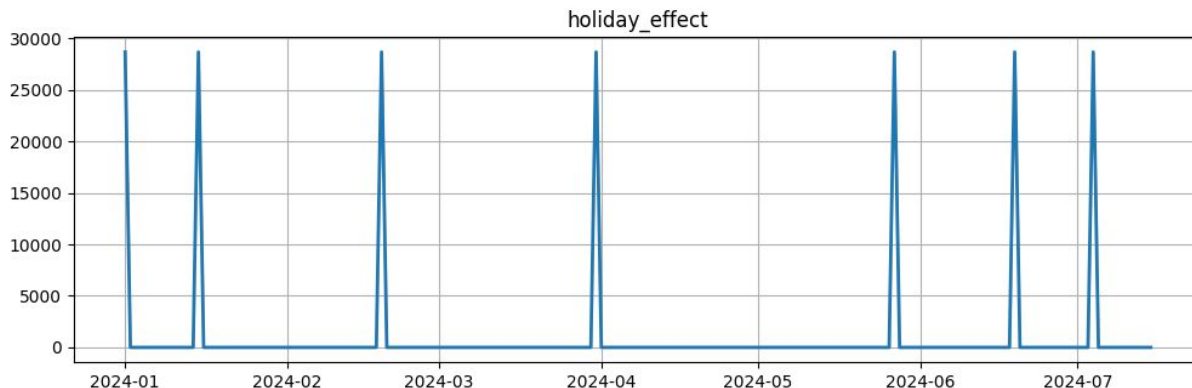
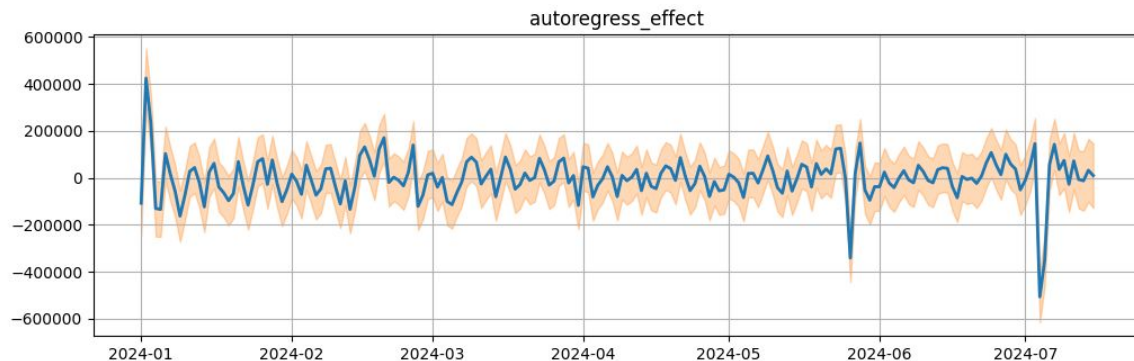


Example of STS - Trend (2024 YTD) with Seasonality (weekly)



- The **trend** for the year is clearly visible with a **small margin of error**
- The model also does a **great job of accounting for the weekly cycle of spikes and troughs**

Example of STS - Holiday Effects on TSA Air Passenger Volumes in 2024



- Interestingly, overall in 2024 so far, holidays had a positive effect on travel volumes, making the July 4 trough even more pronounced in the autoregressive effect



Advantages of using STS models to forecast NOx concentrations

- **Forecasting of complex data sets:** STS models can accurately predict future NOx concentrations, incorporating localized trends, seasonality, and other effects.
- **Daily forecasting capabilities:** STS models can reliably forecast NOx concentrations for weeks into the immediate future, as NOx concentrations are measured daily.
- **Nowcasting capabilities:** STS models can fill the 10-day gap in NOx concentration measurements, providing real-time estimates and enabling timely decision-making.

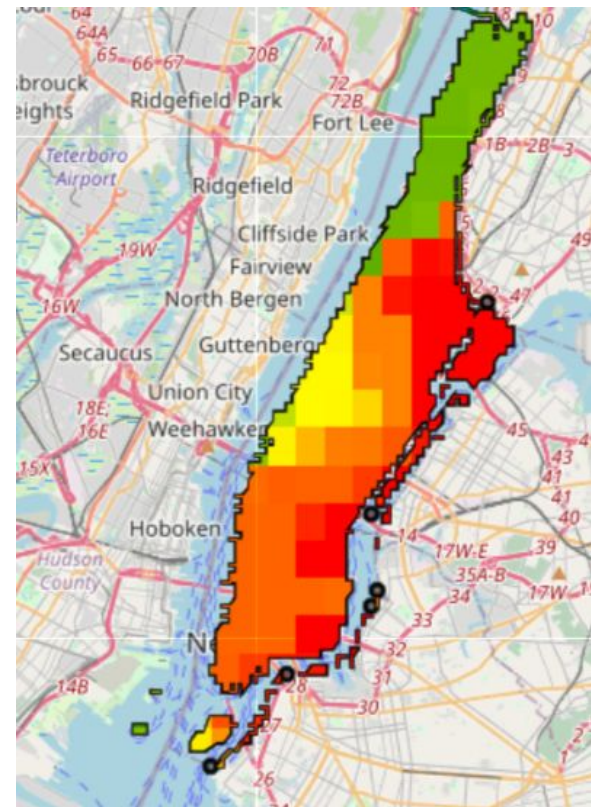
We used the sts-jax library and fixed a bug in the covariate-modeling program, which we plan on PR

Design of the UI for forecasting application

Goals for the Interactive Dashboard

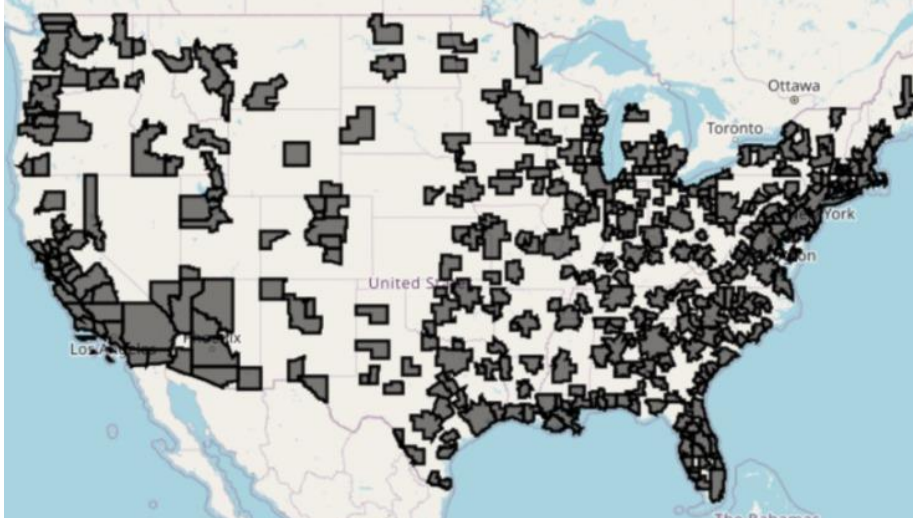
1. Quickly receive historical NOx data for requested metropolitan areas
2. Plot historical data in a variety of ways to accommodate analysis
3. Effortlessly train a model based on historical data, and view forecasts

NOx Median for March 2021, NYC

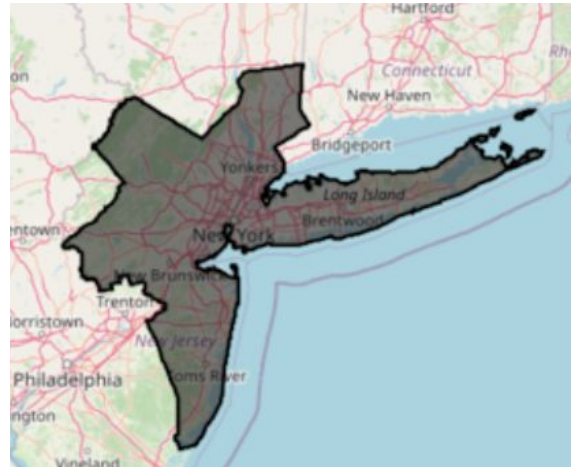


Receiving Metropolitan Statistical Areas (MSAs)

- US Office of Management and Budget defines MSAs
 - Not solely based on legal administrative divisions
- Created a script to download all MSAs in the US and PR
 - Provides name, geometry, and other data for each MSA
- After selecting an MSA, the application uses its geometry to collect its NOx data

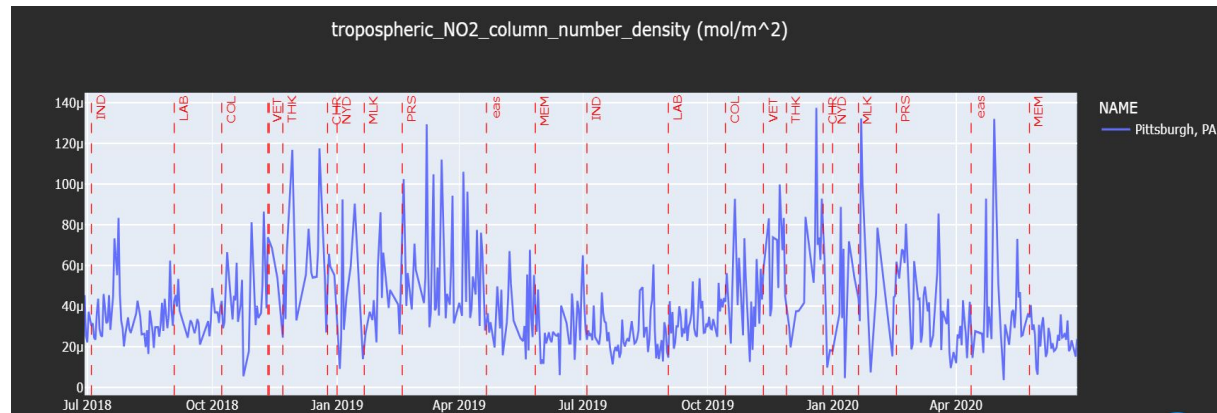


New York-Newark-Jersey City MSA



1. *Chlorophyll a* (Chl *a*)
 2. *Chlorophyll b* (Chl *b*)
 3. *Chlorophyll c* (Chl *c*)
 4. *Chlorophyll d* (Chl *d*)
 5. *Chlorophyll e* (Chl *e*)
 6. *Chlorophyll f* (Chl *f*)
 7. *Chlorophyll g* (Chl *g*)
 8. *Chlorophyll h* (Chl *h*)
 9. *Chlorophyll i* (Chl *i*)
 10. *Chlorophyll j* (Chl *j*)
 11. *Chlorophyll k* (Chl *k*)
 12. *Chlorophyll l* (Chl *l*)
 13. *Chlorophyll m* (Chl *m*)
 14. *Chlorophyll n* (Chl *n*)
 15. *Chlorophyll o* (Chl *o*)
 16. *Chlorophyll p* (Chl *p*)
 17. *Chlorophyll q* (Chl *q*)
 18. *Chlorophyll r* (Chl *r*)
 19. *Chlorophyll s* (Chl *s*)
 20. *Chlorophyll t* (Chl *t*)
 21. *Chlorophyll u* (Chl *u*)
 22. *Chlorophyll v* (Chl *v*)
 23. *Chlorophyll w* (Chl *w*)
 24. *Chlorophyll x* (Chl *x*)
 25. *Chlorophyll y* (Chl *y*)
 26. *Chlorophyll z* (Chl *z*)
 27. *Chlorophyll aa* (Chl *aa*)
 28. *Chlorophyll ab* (Chl *ab*)
 29. *Chlorophyll ac* (Chl *ac*)
 30. *Chlorophyll ad* (Chl *ad*)
 31. *Chlorophyll ae* (Chl *ae*)
 32. *Chlorophyll af* (Chl *af*)
 33. *Chlorophyll ag* (Chl *ag*)
 34. *Chlorophyll ah* (Chl *ah*)
 35. *Chlorophyll ai* (Chl *ai*)
 36. *Chlorophyll aj* (Chl *aj*)
 37. *Chlorophyll ak* (Chl *ak*)
 38. *Chlorophyll al* (Chl *al*)
 39. *Chlorophyll am* (Chl *am*)
 40. *Chlorophyll an* (Chl *an*)
 41. *Chlorophyll ao* (Chl *ao*)
 42. *Chlorophyll ap* (Chl *ap*)
 43. *Chlorophyll aq* (Chl *aq*)
 44. *Chlorophyll ar* (Chl *ar*)
 45. *Chlorophyll as* (Chl *as*)
 46. *Chlorophyll at* (Chl *at*)
 47. *Chlorophyll au* (Chl *au*)
 48. *Chlorophyll av* (Chl *av*)
 49. *Chlorophyll aw* (Chl *aw*)
 50. *Chlorophyll ax* (Chl *ax*)
 51. *Chlorophyll ay* (Chl *ay*)
 52. *Chlorophyll az* (Chl *az*)
 53. *Chlorophyll ba* (Chl *ba*)
 54. *Chlorophyll bb* (Chl *bb*)
 55. *Chlorophyll bc* (Chl *bc*)
 56. *Chlorophyll bd* (Chl *bd*)
 57. *Chlorophyll be* (Chl *be*)
 58. *Chlorophyll bf* (Chl *bf*)
 59. *Chlorophyll bg* (Chl *bg*)
 60. *Chlorophyll bh* (Chl *bh*)
 61. *Chlorophyll bi* (Chl *bi*)
 62. *Chlorophyll bj* (Chl *bj*)
 63. *Chlorophyll bk* (Chl *bk*)
 64. *Chlorophyll bl* (Chl *bl*)
 65. *Chlorophyll bm* (Chl *bm*)
 66. *Chlorophyll bn* (Chl *bn*)
 67. *Chlorophyll bo* (Chl *bo*)
 68. *Chlorophyll bp* (Chl *bp*)
 69. *Chlorophyll bq* (Chl *bq*)
 70. *Chlorophyll br* (Chl *br*)
 71. *Chlorophyll bs* (Chl *bs*)
 72. *Chlorophyll bt* (Chl *bt*)
 73. *Chlorophyll bu* (Chl *bu*)
 74. *Chlorophyll bv* (Chl *bv*)
 75. *Chlorophyll bw* (Chl *bw*)
 76. *Chlorophyll bx* (Chl *bx*)
 77. *Chlorophyll by* (Chl *by*)
 78. *Chlorophyll bz* (Chl *bz*)
 79. *Chlorophyll ca* (Chl *ca*)
 80. *Chlorophyll cb* (Chl *cb*)
 81. *Chlorophyll cc* (Chl *cc*)
 82. *Chlorophyll cd* (Chl *cd*)
 83. *Chlorophyll ce* (Chl *ce*)
 84. *Chlorophyll cf* (Chl *cf*)
 85. *Chlorophyll cg* (Chl *cg*)
 86. *Chlorophyll ch* (Chl *ch*)
 87. *Chlorophyll ci* (Chl *ci*)
 88. *Chlorophyll cj* (Chl *cj*)
 89. *Chlorophyll ck* (Chl *ck*)
 90. *Chlorophyll cl* (Chl *cl*)
 91. *Chlorophyll cm* (Chl *cm*)
 92. *Chlorophyll cn* (Chl *cn*)
 93. *Chlorophyll co* (Chl *co*)
 94. *Chlorophyll cp* (Chl *cp*)
 95. *Chlorophyll cq* (Chl *cq*)
 96. *Chlorophyll cr* (Chl *cr*)
 97. *Chlorophyll cs* (Chl *cs*)
 98. *Chlorophyll ct* (Chl *ct*)
 99. *Chlorophyll cu* (Chl *cu*)
 100. *Chlorophyll cv* (Chl *cv*)
 101. *Chlorophyll cw* (Chl *cw*)
 102. *Chlorophyll cx* (Chl *cx*)
 103. *Chlorophyll cy* (Chl *cy*)
 104. *Chlorophyll cz* (Chl *cz*)
 105. *Chlorophyll da* (Chl *da*)
 106. *Chlorophyll db* (Chl *db*)
 107. *Chlorophyll dc* (Chl *dc*)
 108. *Chlorophyll dd* (Chl *dd*)
 109. *Chlorophyll de* (Chl *de*)
 110. *Chlorophyll df* (Chl *df*)
 111. *Chlorophyll dg* (Chl *dg*)
 112. *Chlorophyll dh* (Chl *dh*)
 113. *Chlorophyll di* (Chl *di*)
 114. *Chlorophyll dj* (Chl *dj*)
 115. *Chlorophyll dk* (Chl *dk*)
 116. *Chlorophyll dl* (Chl *dl*)
 117. *Chlorophyll dm* (Chl *dm*)
 118. *Chlorophyll dn* (Chl *dn*)
 119. *Chlorophyll do* (Chl *do*)
 120. *Chlorophyll dp* (Chl *dp*)
 121. *Chlorophyll dq* (Chl *dq*)
 122. *Chlorophyll dr* (Chl *dr*)
 123. *Chlorophyll ds* (Chl *ds*)
 124. *Chlorophyll dt* (Chl *dt*)
 125. *Chlorophyll du* (Chl *du*)
 126. *Chlorophyll dv* (Chl *dv*)
 127. *Chlorophyll dw* (Chl *dw*)
 128. *Chlorophyll dx* (Chl *dx*)
 129. *Chlorophyll dy* (Chl *dy*)
 130. *Chlorophyll dz* (Chl *dz*)
 131. *Chlorophyll ea* (Chl *ea*)
 132. *Chlorophyll eb* (Chl *eb*)
 133. *Chlorophyll ec* (Chl *ec*)
 134. *Chlorophyll ed* (Chl *ed*)
 135. *Chlorophyll ee* (Chl *ee*)
 136. *Chlorophyll ef* (Chl *ef*)
 1

- Date range selector
- Multiple graphing options
- Labeled holiday indicators



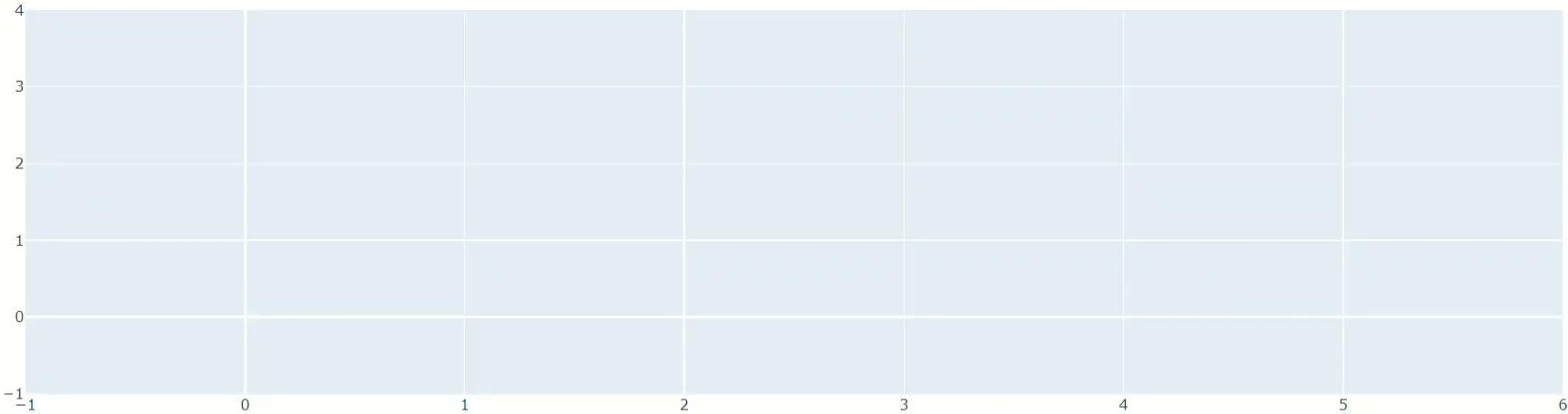
Select a City

06/28/2018 → 06/30/2024

Historical Data

Forecasts

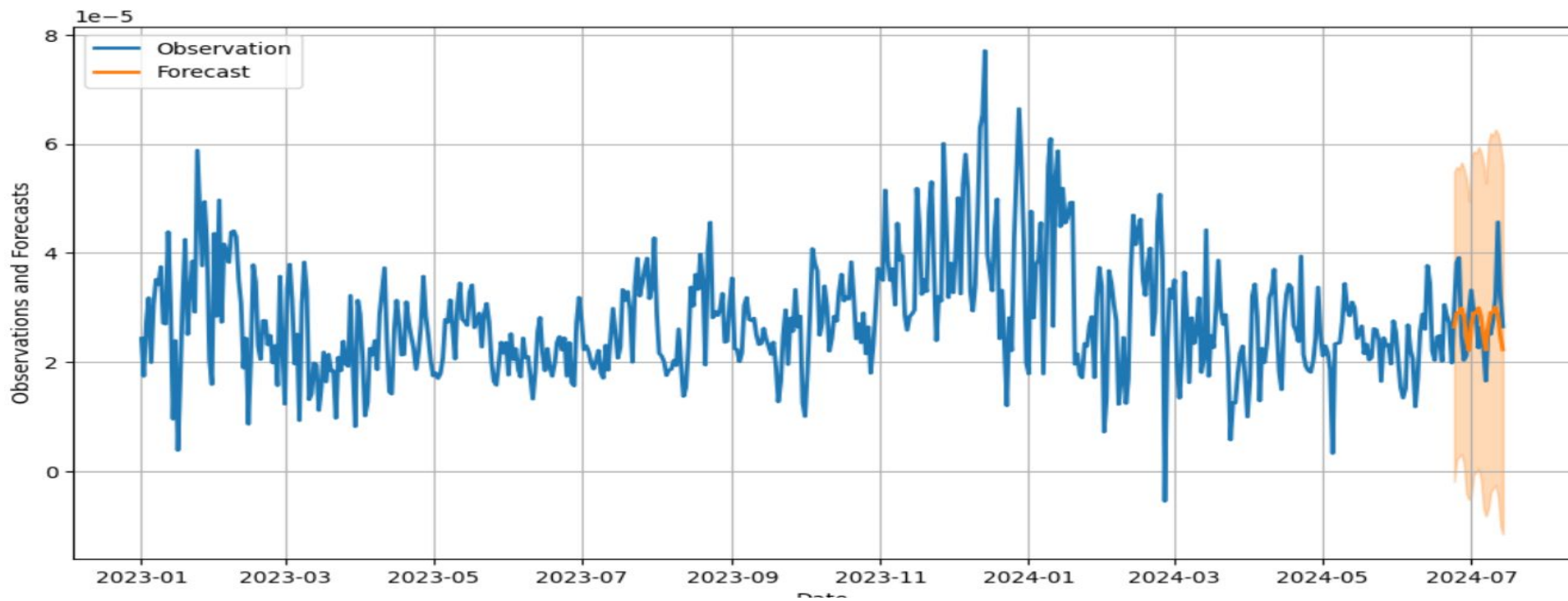
☒ Daily ☐ Monthly Average ☐ Yearly

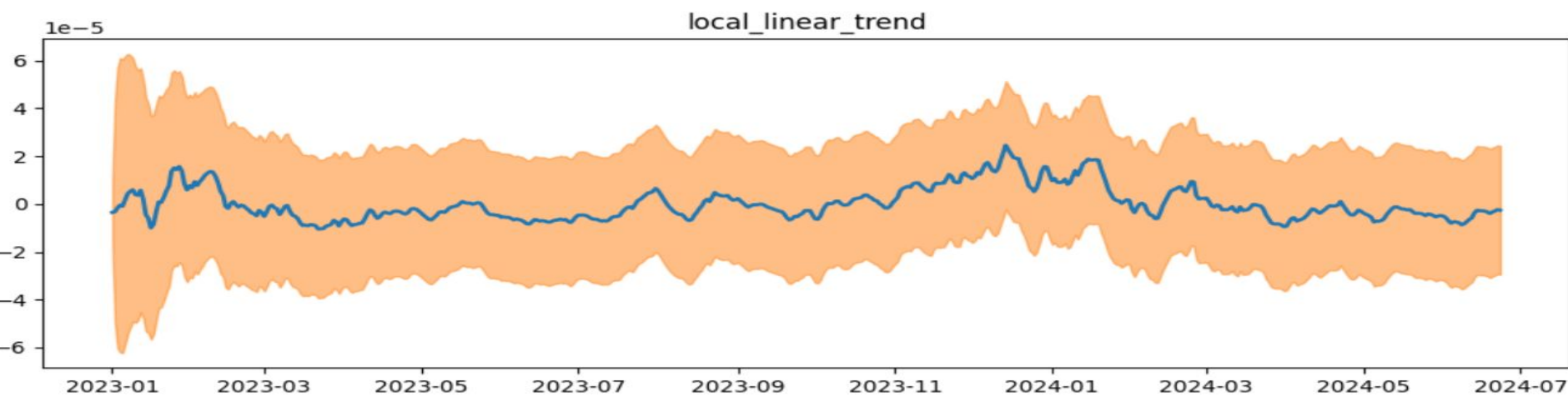
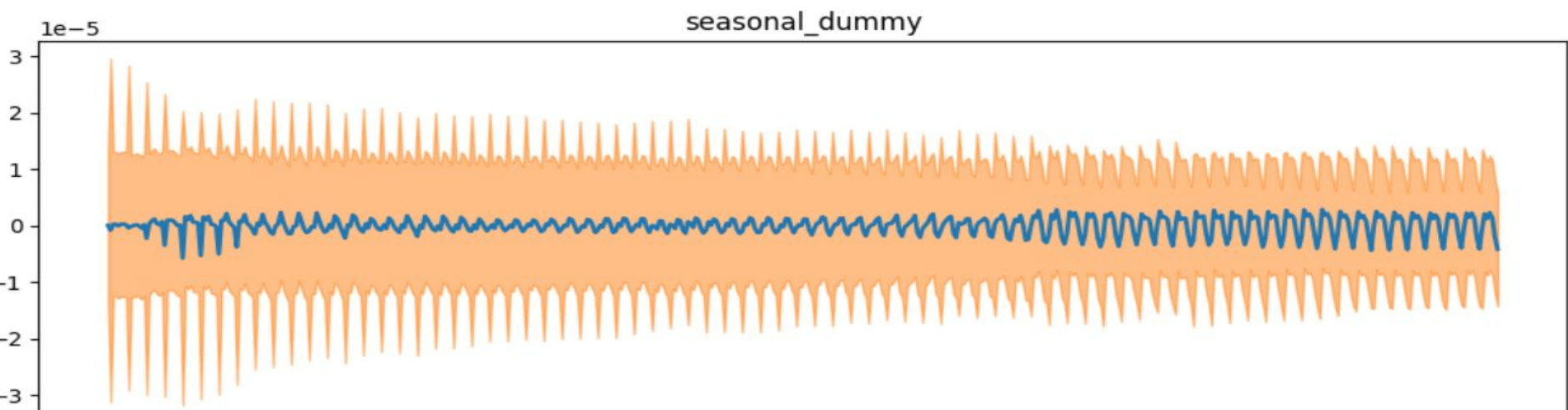


Forecast Models and Accuracy.

Results with Phoenix, AZ

Model was fitted with data from *Phoenix-Mesa-Chandler*, AZ region, using data from 2023-01-01 to 2024-06-30





Out-of-Sample (OOS) Forecast Accuracy

Utilizing the Phoenix, AZ data shown previously, forecast is for 2 weeks ahead

Model	mape	smape	rmse (mol/m^2)
Trend only	0.16195	0.15806	6.437 e-6
Trend + seasonal	0.17732	0.17297	5.681 e-6
Trend + seasonal + holiday	TBD	TBD	TBD
Trend + seasonal + holiday + AR	0.26935	0.23353	7.767 e-6

$$\frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad \frac{100}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2} \quad \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}}$$



Future Work - Improvements to the Model

- Factoring in **covariates**
 - **Transportation** data as a covariate for the models
 - **Weather** data as a covariate for the models
 - **Population** data, which we got, as a covariate for the models
 - More **precise modeling of holidays**
 - Weighing each holiday differently
 - Accounting for variations in DOW for each holiday
- Use the **monthly models** for forecasting long-term variations in the NOx concentrations
 - Current models are all short term, daily models
- Models for **other countries** and regions
 - (e.g. EU, China and India)