Forecasting NOx 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 NOx concentrations using Google Earth Engine APIs
2. Model the daily variations of the observed NOx 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 NOx concentrations

Technical Approach - What to expect ...

- This project implements a **modeling and forecasting system** for the heterogeneous, spatially-localized levels of atmospheric NOx.
- 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 NOx 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 CO_2 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

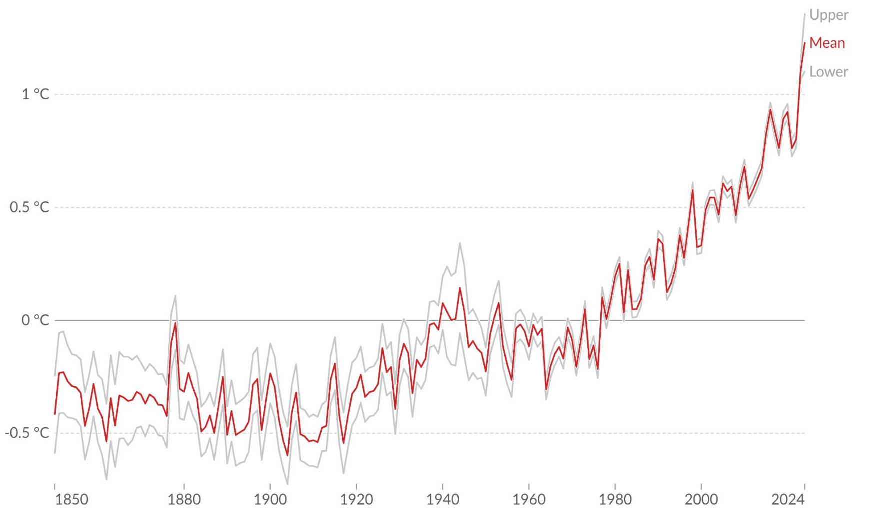
Listen to this article · 4:03 min Learn more

July 24, 2024

19

Average temperature anomaly, Global

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

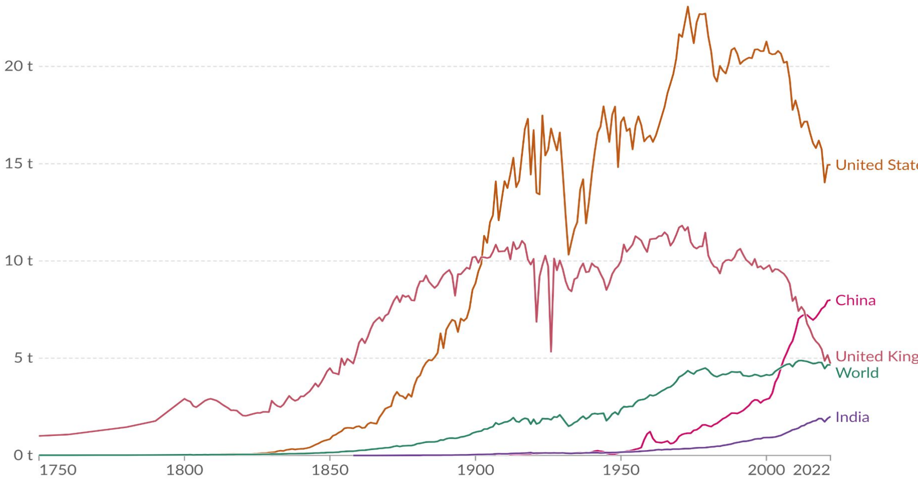


Data source: Met Office Hadley Centre (2024) OurWorldInData.org/co2-and-greenhouse-gas-emissions | CC BY

Note: The gray lines represent the upper and lower bounds of the 95% confidence intervals.

Per capita CO_2 emissions

Carbon dioxide (CO_2) emissions from fossil fuels and industry[1]. Land-use change is not included.



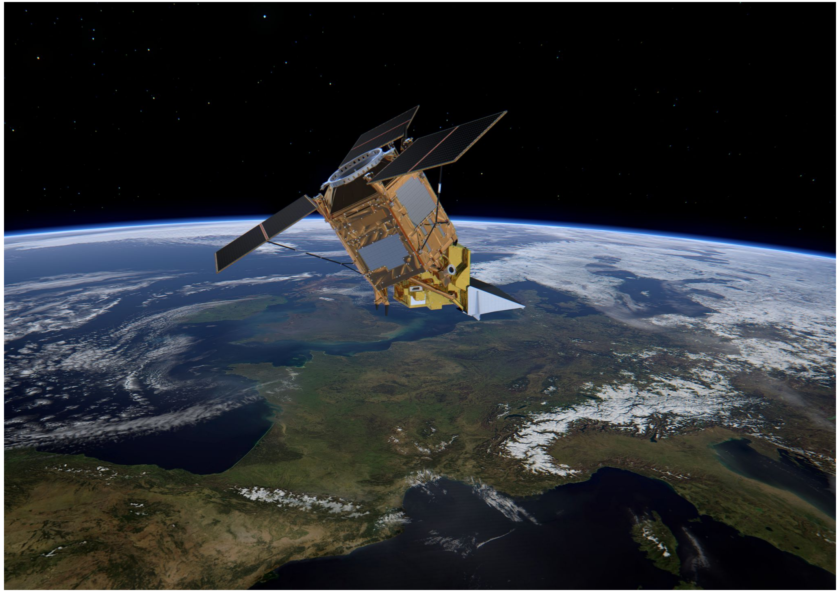
Data source: Global Carbon Budget (2023); Population based on various sources (2023)
OurWorldInData.org/co2-and-greenhouse-gas-emissions | CC BY

1. **Fossil emissions**: Fossil emissions measure the quantity of carbon dioxide (CO_2) emitted from the burning of fossil fuels, and directly from industrial processes such as cement and steel production. Fossil CO_2 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, CO_2, NO_2, SO_2, formaldehyde, and aerosol in the atmosphere.
- TROPOMI (version 02.03.0144) measures column-averaged tropospheric column NO_2 (mol/m^2) 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.

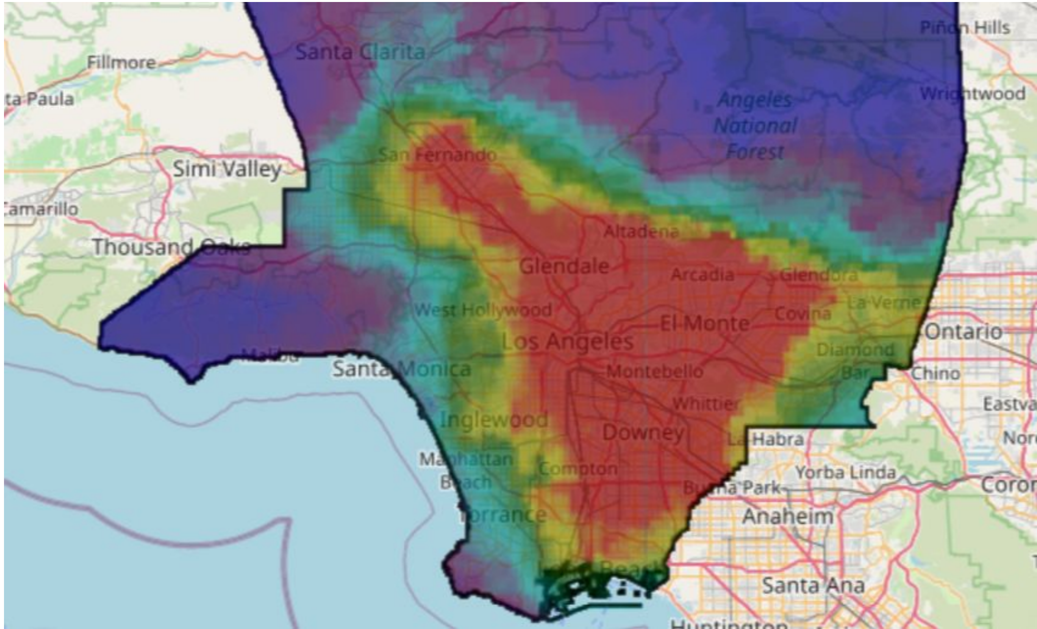
Data Frame for Time Series

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

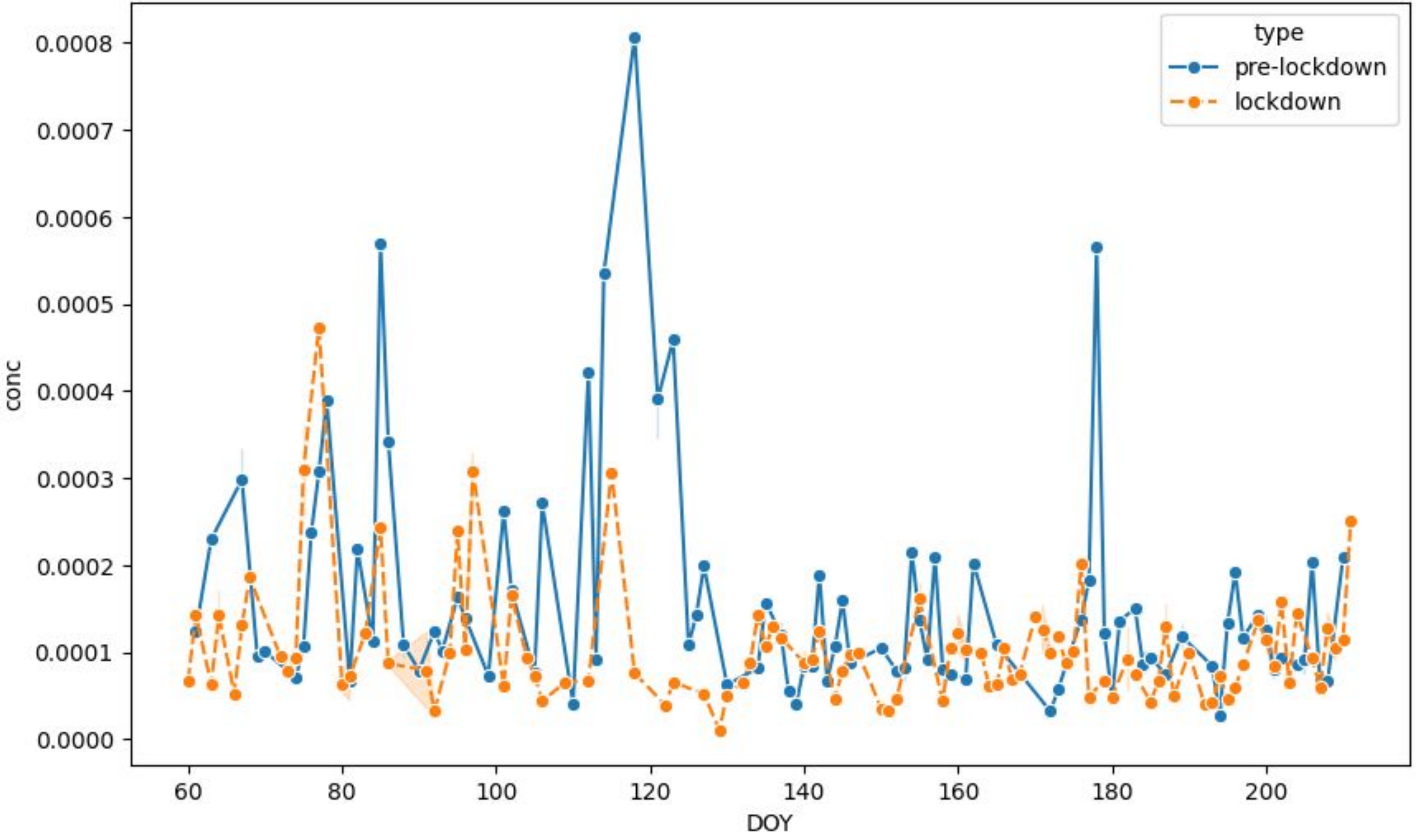
Google Earth Engine Satellite Data

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

NOx Median for March 2021, Los Angeles region

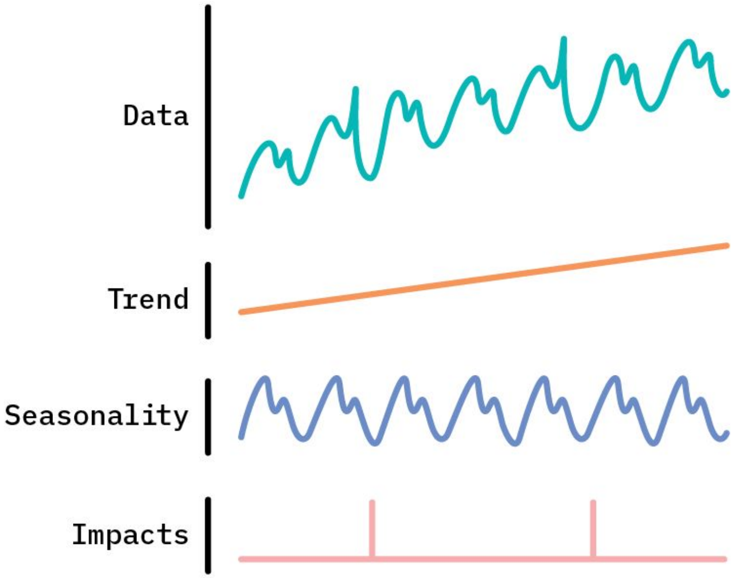


NOx Emissions Pre (2019) and During (2020) COVID over NYC



Modeling and Forecasting approach

Structural Time Series (STS)

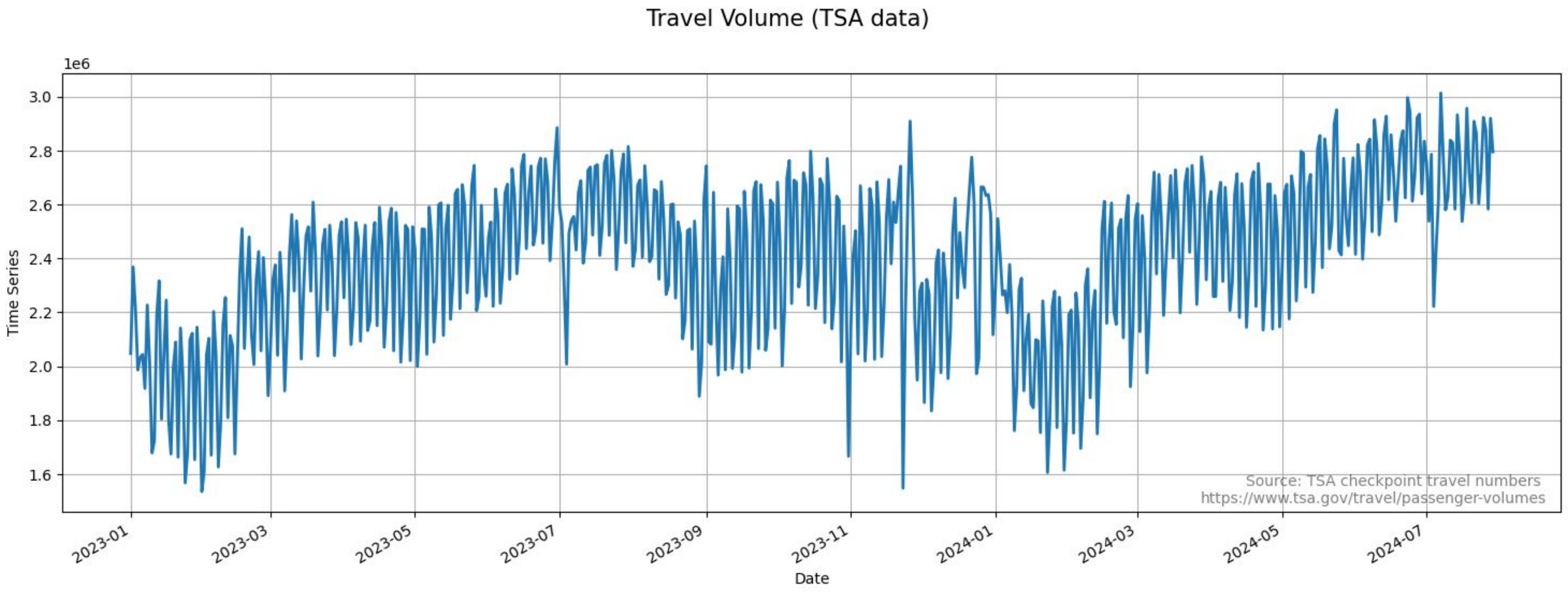


Decomposes observed data into a sum of latent components.

- 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.

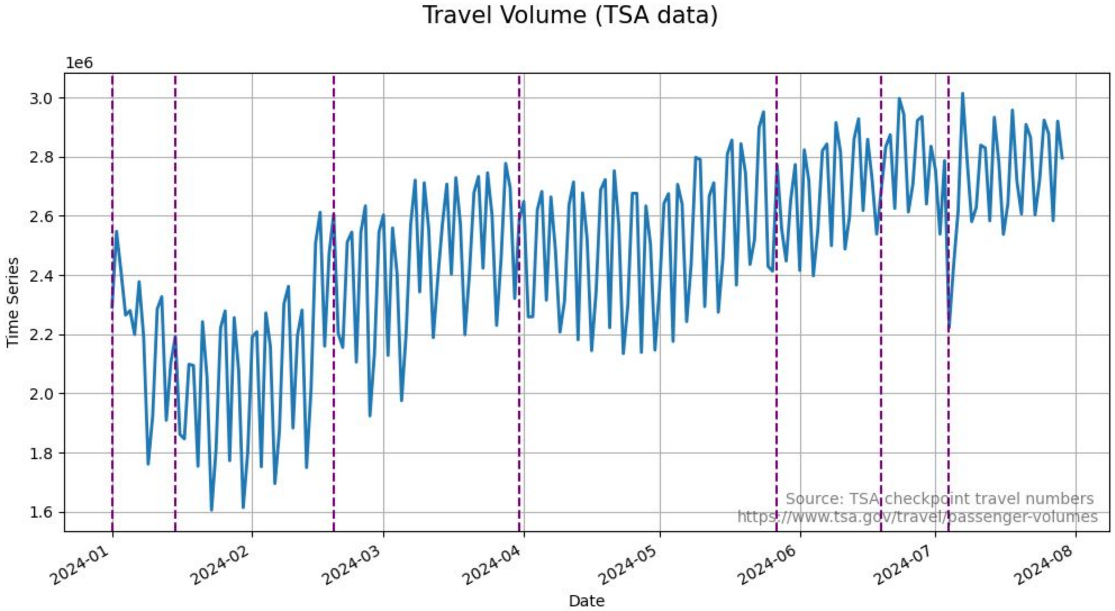
Observation(t) = Trend(t) + Seasonality(t) + Impacts(t) + Noise(t)

Example of STS - TSA Air Passenger Volumes

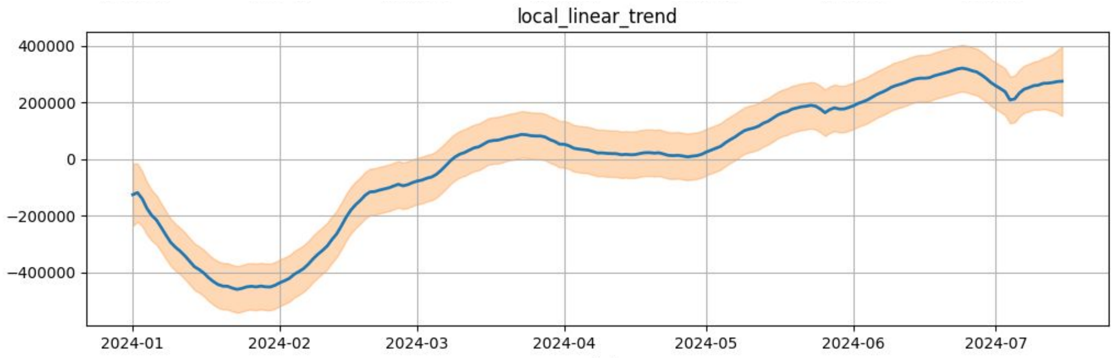


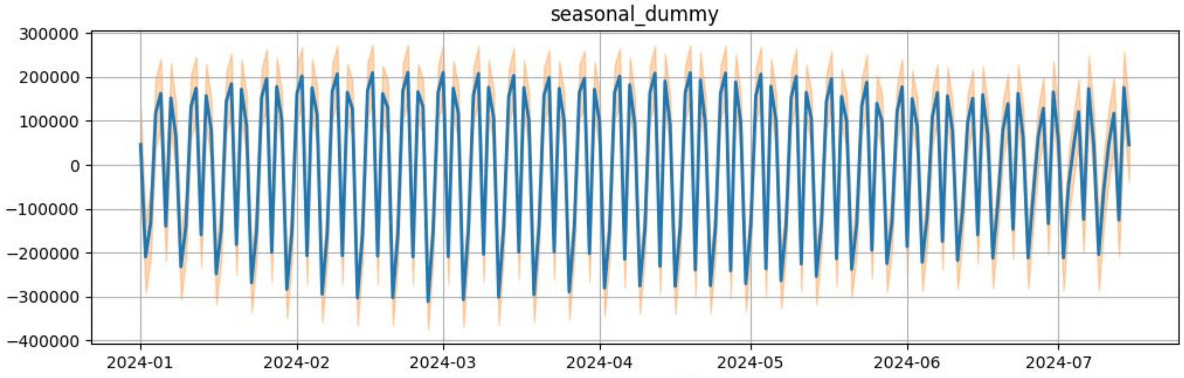
Example of STS - TSA Air Passenger Volumes YTD

- 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**



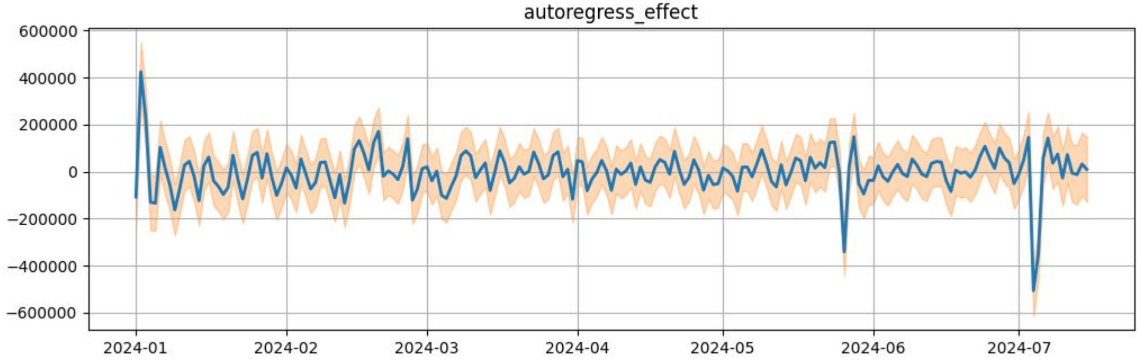
Example of STS - Trend (2024 YTD) with Seasonality (weekly)

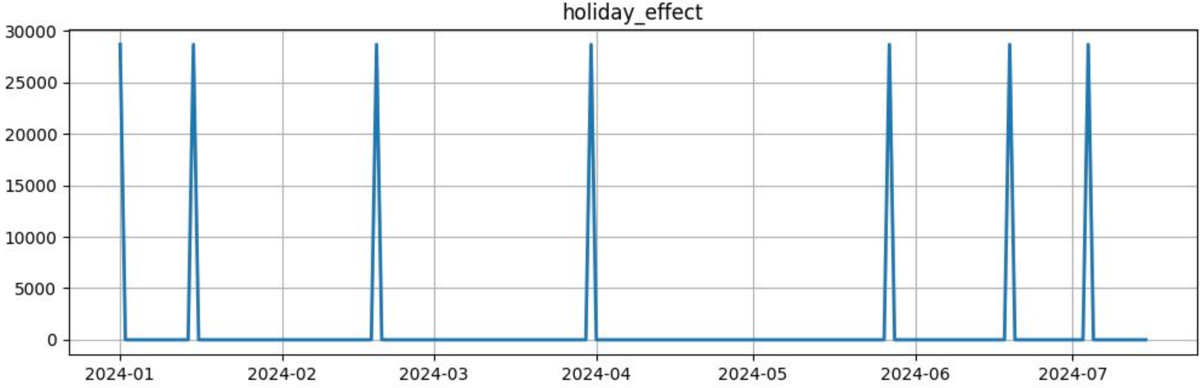




- The **trend** for the year is clearly visible with a **small margin of error**
- The model also does a **great job of accounting for the weekly** cycle of spikes and troughs

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

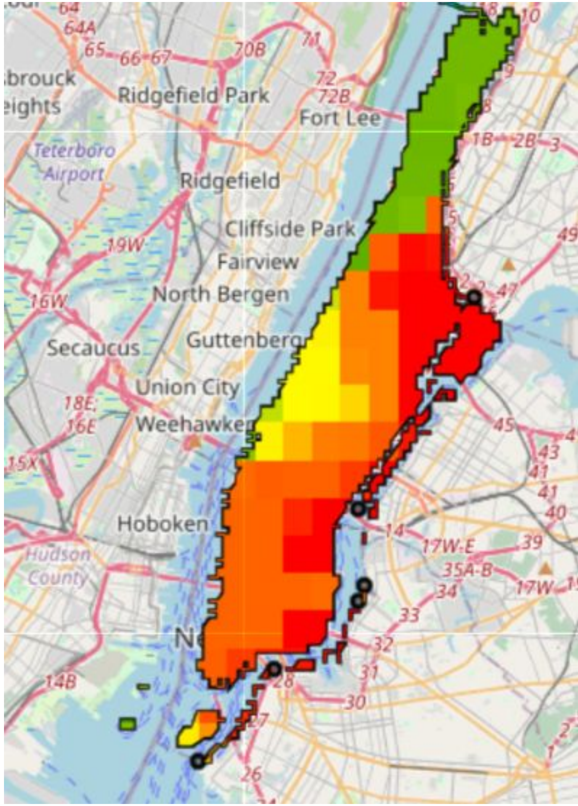
Goals for the Interactive Dashboard

1. Quickly receive historical NOx data for requested metropolitan areas

2. Plot historical data in a variety of ways to accommodate analysis

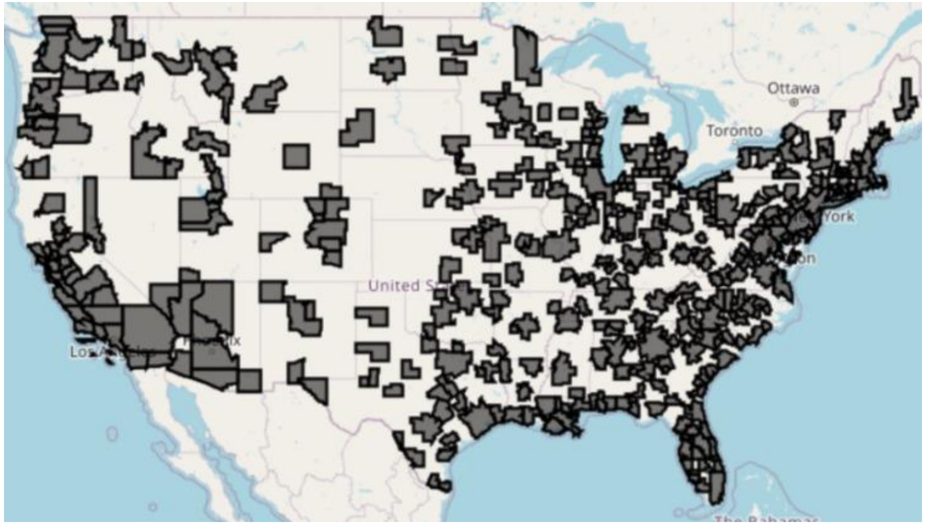
3. Effortlessly train a model based on historical data, and view forecasts

NOx Median for March 2021, NYC

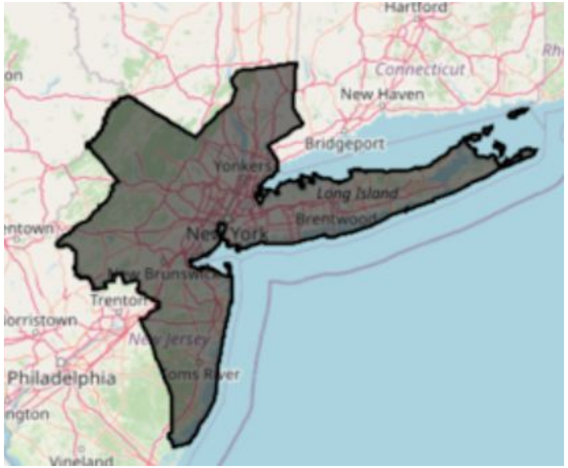


Receiving Metropolitan Statistical Areas (MSAs)

- 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



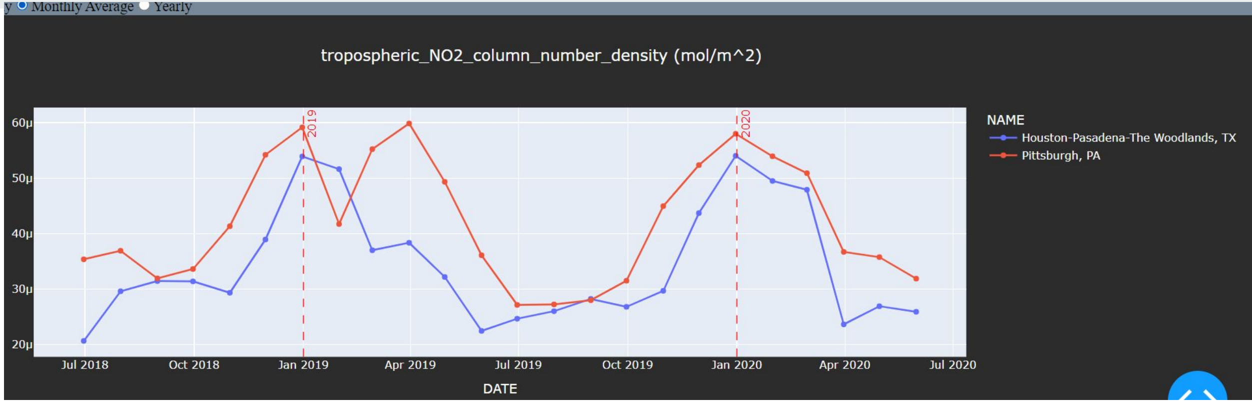
New York-Newark-Jersey City MSA

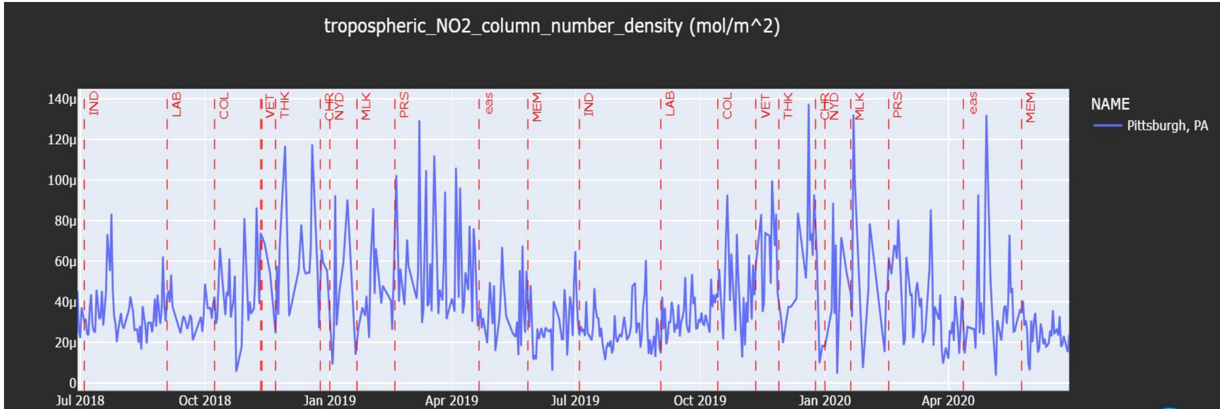


Features in UI

- Search bar

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



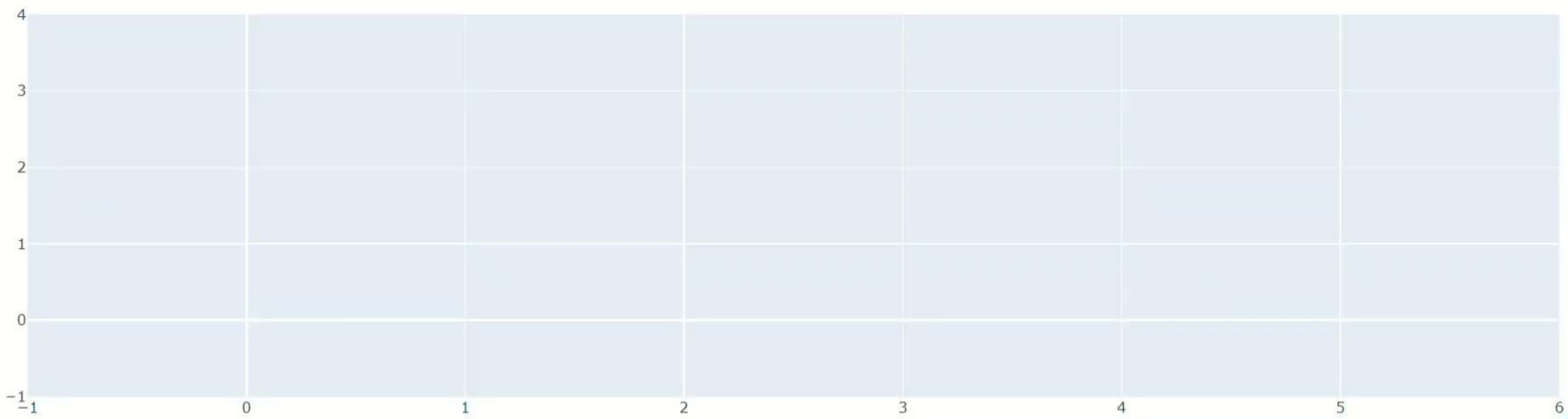


Select a City

06/28/2018 → 06/30/2024

Historical Data | Forecasts

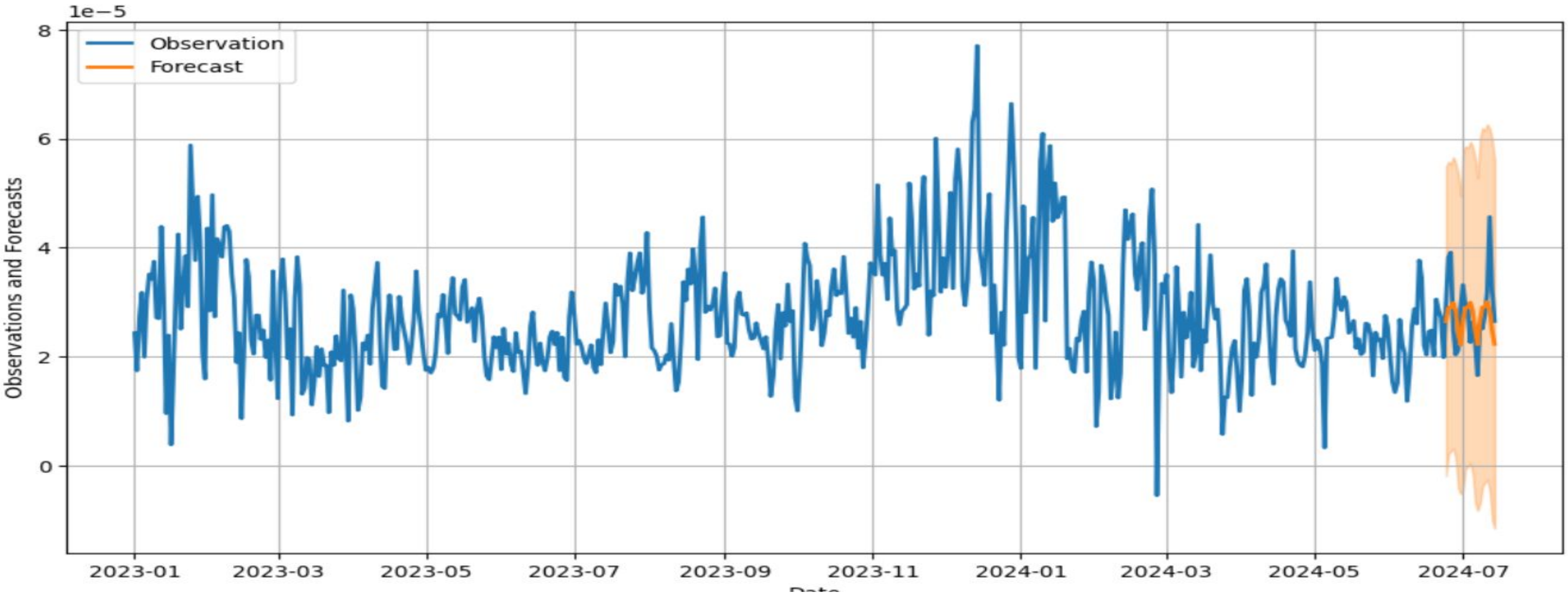
Daily Monthly Average Yearly

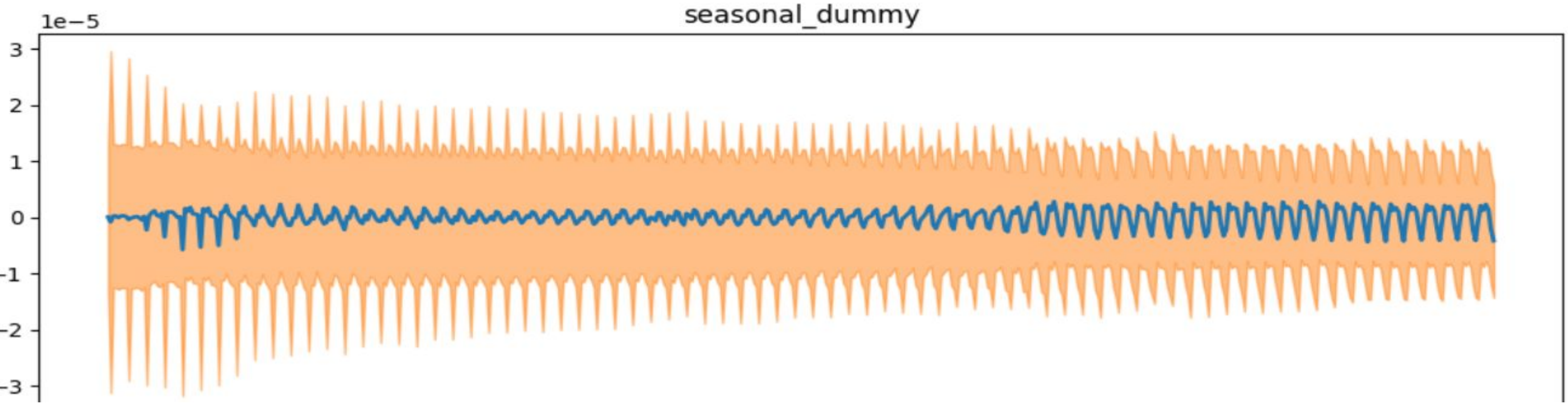


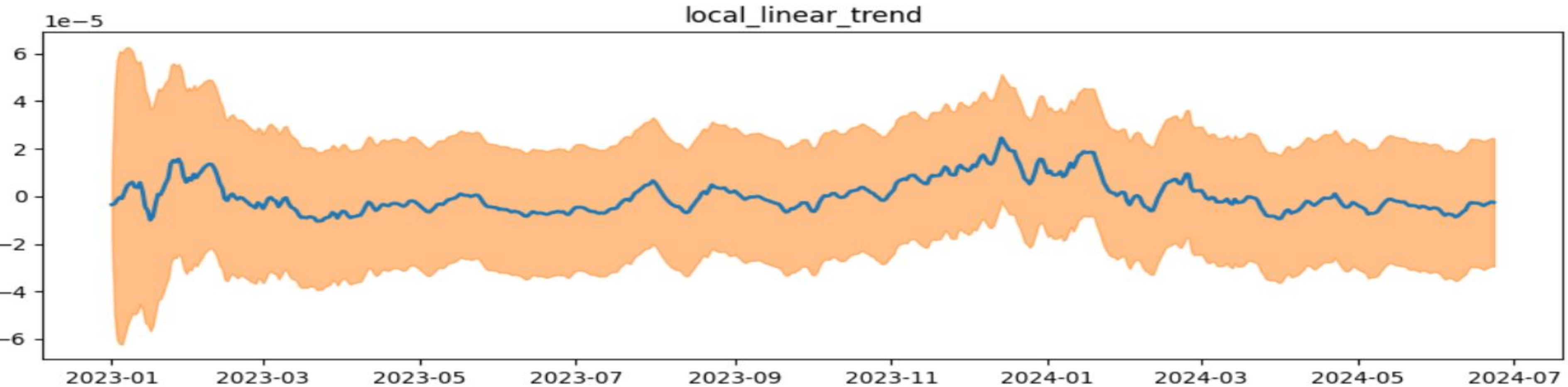
Forecast Models and Accuracy.

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




seasonal_dummy
1e−5
3
2
1
0
-1
-2
-3


local_linear_trend
1e−5
6
4
2
0
−2
−4
−6
2023-01
2023-03
2023-05
2023-07
2023-09
2023-11
2024-01
2024-03
2024-05
2024-07

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

$$\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

$$\frac{100}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

$$\sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2}{N}}$$

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)