The effects of air quality change on economic outcomes

Shreya Shanker, Samantha Virgil

Georgia Institute of Technology, Masters in Data Analytics

Abstract

Setting and meeting air quality standards is a global concern that must be addressed. Identifying possible trends between air quality and economic growth could help regions achieve their sustainable development goals. For this project, we investigated fifty-four cities spread across India to try to identify if there were any trends between changes in four emissions (CO, SO_2, NO_2, and $PM_{2.5}$) and changes in unemployment. After performing our analysis, we found no significant correlation between these factors, as the seasonality and trend observed in the emissions data did not align with the unemployment data.

1. Introduction

Pollution and air quality have been major global concerns in the past few years. Global air quality has been named as a driving factor for premature deaths around the world. [1] Identifying a trend between pollutant quantities in the air and economic growth could help regions set air quality standards in order to achieve economic goals. MacroX, a financial technology startup based in San Francisco, is currently researching a way to model the trade-offs that come with achieving different sustainable development goals (SDG). The following work outlined in this paper supports this overarching objective by identifying minor trends that exist in the data.

Our goal for this project is to analyze the relationship between the changes in air quality and the changes in unemployment in fifty-four cities throughout India. In particular, we aimed to conduct a time-series analysis of the changes in these values over time, using data from 2018 to 2022. As a sub-goal of this project, we sought to explore if the effects of the novel coronavirus pandemic introduced any specific trends to our data. This project will allow us to make determinations on the relationship between the environment and economic growth, along with exposing us to many opportunities to apply modeling principles and knowledge we have acquired over the course of our studies at Georgia Tech.

2. Data Sources

Our first step in beginning this project was to determine what data to use. There are a variety of emissions that have an effect on air quality, but we were able to narrow down our data to focus on four different emissions: carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide(SO_2), and particulate matter less than 2.5 microns in width ($PM_{2.5}$). We decided to move forward focusing on these four emissions because of an observational study done on global air quality in 2021. Since a sub-goal of this project was to analyze whether air quality and economic growth trended differently during the novel coronavirus pandemic, it was important that we chose emissions that were shown to have changed over the past few years. The 2021 study by Sokhi et. al showed major changes in CO, NO_2, and $PM_{2.5}$. [2] We decided to look at SO_2 as well despite it not appearing in this study because it is an emission that has been known to have reduced dramatically over the past few decades. [1] Because of this, we thought it would be interesting to explore the differences in trends with SO_2 and with the other three emissions.

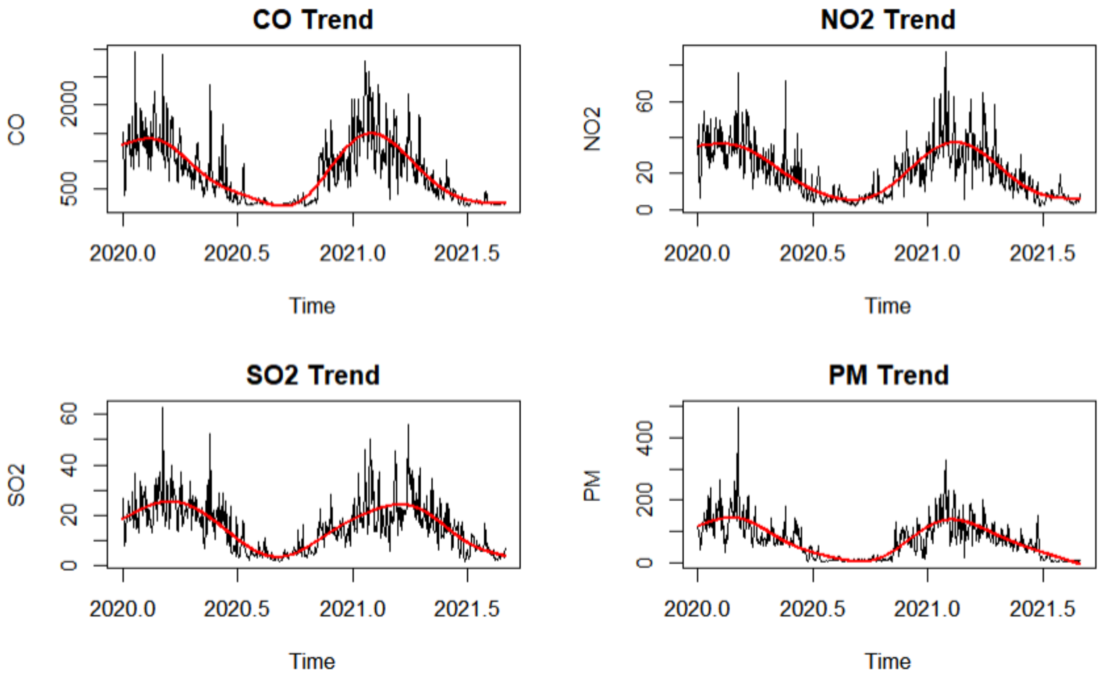
Our data for this project came from three separate sources. The first source was the Central Pollution Control Board (CPCB), an Indian organization that provides technical services to the Ministry of Environment and Forests. [3] As part of the work that CPCB performs, they collect air quality data for many cities across India. By accessing their data through an API, we were able to acquire data on NO_2, SO_2, CO, and $PM_{2.5}$ emissions from the fifty-four cities our project was focusing on. Each city had multiple stations reporting these emission levels, so we averaged the data when aggregating it to get a single observation point for each city on each date. We collected this data from January 2018 to August 2022.

The second source was satellite data, sourced from the Ozone Monitoring Instrument. This data consisted of daily observations of different emissions in various cities measured by satellite. We were provided data on CO, NO_2, and SO_2 values from this satellite that we then proceeded to use in our analysis. This data existed has been collected for a long time, but for our project, we focused on data collected by the satellite between January 2018 and July 2022.

Our third and final source was data from OpenWeatherMap, an excellent API that provides many data measures related to weather and air quality index. We sourced data from OpenWeatherMap through their API from November 2020 onwards. This data contained measurements of CO, NO_2, SO_2, and $PM_{2.5}$ for the fifty-four cities we were looking at. The OpenWeatherMap dataset was the most complete dataset - however, the set only started in November 2020. Therefore, to conduct a more thorough analysis, the data would have to be backfilled to match the date ranges for the satellite and government data we were provided.

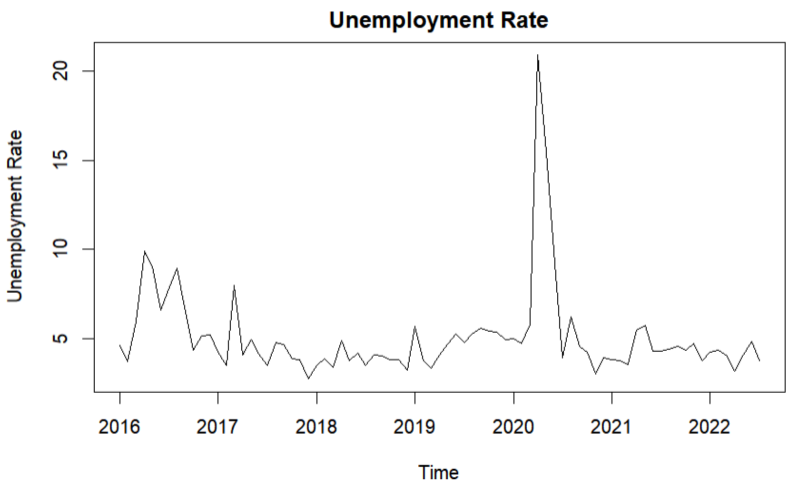
3. Exploratory Data Analysis

We began our exploratory data analysis with the data from OpenWeatherMap. In this analysis, we first examined the time series and autocorrelation function (ACF) plots for each pollutant for each city. The time series plots show us the movement of the pollutants in time and the ACF plots show us the coefficients of correlation between each time series and its lagged values. ACF plots can help us determine if a time series contains trend and/or seasonality components. The ACF plots for many of the pollutants across all cities indicated trend and some may have seasonality components. We modeled the trend for each pollutant for each city using splines regression. The time series plots for the city of Pune are very similar to the pollutant plots of many of the other cities in India. The air pollutant trends in Pune are modeled below.



We can see a clear drop in all emissions once the COVID pandemic began affecting the country and another drop after the first quarter of 2021.

Next, we tackled the unemployment data for each city. We analyzed the time series plots as well as the ACF plots to see if any trend or seasonality was present. The ACF plots showed no trend in the unemployment rate for all cities, but suggested seasonality present for some. Below is a plot of the unemployment rate in Pune from 2016 to 2022. Many cities had a large spike in the unemployment rate in 2020 similar to Pune.



3.1. Approach to Missing Data

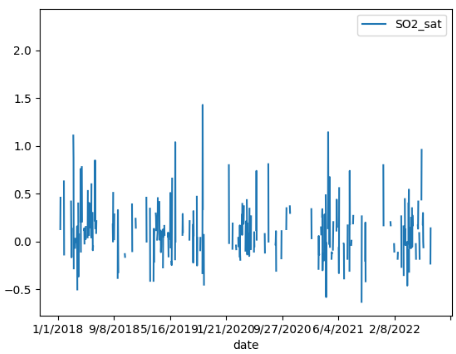
A major challenge of this project was to handle significant amounts of missing data. For both our satellite data and our government data, we had many individual dates where one or more emissions were missing values. Because of the volume of missing data, removal of missing observations did not make sense. Therefore, in order to make our analysis more thorough, we needed to impute the data that was missing with estimates.

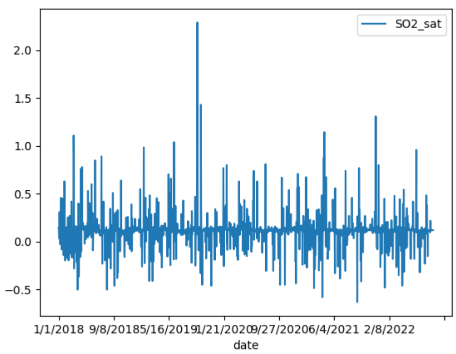
3.1.1. Satellite and Government Data

When imputation is performed, it is important to estimate missing values based on as many true values as possible. However, the data we were looking at was made up of cities across India, and emissions data varied widely between different regions. Thus, instead of using all the data we had to calculate and impute values, we chose to cluster similar cities together, so that we could get more accurate results. We used the Köppen climate classification system in order to cluster the cities into groups with similar climates, as we thought emissions data would be similar between these cities. [4] We sourced our information on what Köppen classification each city belonged to from the website WeatherBase. With this information, we were able to group the fifty-four cities from our data source into six climate groups. When moving forward with imputation, we imputed values for cities in each climate group at a time, instead of using all the data as inputs for the imputation calculation.

In order to impute the missing values for our satellite and government data, we used the imputers provided by Python's scikit-learn library. [5] This library offers three imputers: simple, KNN, and iterative. We knew right away that the simple imputer was not the right choice for the job, as it replaces the missing values with a simple mean or mode of the data, which would not correctly reflect the seasonality of the data. Therefore, we tried comparing the KNN Imputer and the Iterative Imputer. We found that the Iterative Imputer worked much better than the KNN Imputer in preserving the seasonality of the data when filling in the values. This is likely because the Iterative Imputer takes a multivariate approach to filling in the missing data, while the KNN Imputer only uses data from nearby rows to fill in missing observations.

After using the Iterative Imputer, we once again performed some basic analysis to confirm that the seasonality and general trends of our data was preserved. For an example, we have included the graphs of raw satellite data and imputed satellite data for SO_2 emissions over time in Chennai.





3.1.2. OpenWeatherData

Now that we had imputed data from our government and satellite sources, we moved on to the OpenWeatherData source. Imputing data for this source was a little more difficult. Unlike with our government and satellite data, every observation we had from our OpenWeatherData source was complete. However, we only had observations from November 2020 to July 2022, while we had government and satellite data dating back to January 2018. Therefore, instead of using an imputer to backfill this data, we instead aimed to create a regression model that would take the government and satellite data in as predictor variables, and output the OpenWeatherData value as a response variable.

Before creating any models, we started by cleaning and organizing the dataset. We grouped the data based on the Köppen climate classification system we had used earlier, then aggregated the data by taking the mean of each climate group. This allowed us to back-fill the data for similar regions, instead of using all the data to back-fill each city individually. We then dropped any columns of data values that did not exist in the OpenWeatherData set, such as the ammonia values in the government data and the black carbon values in the satellite data. Finally, we scaled the nitrogen dioxide satellite data to make sure it matched the scale of the government and OpenWeatherData.

While we knew that this was a regression problem, there were many regression methods we could have used to create the best model for backfilling the OpenWeatherData values prior to November 2020. We decided to test and compare five different methods: linear regression, LASSO regression, recursive feature elimination, random forest regression, and support vector regression (SVR). In order to analyze these methods, we collected all OpenWeatherData from November 2020 to July 2022. We then split this data into a training and testing set, with 80% of the data going to the training set and 20% of the data going to the testing set.

Then, for each climate and emission combination, we tested each of the five models by building a model on the training set, then testing the model on the testing set. We did this for each combination in case there were situations where one model performed better for a given climate than another model. In order to compare the models, we collected the scores of the models (provided by the Python scikit-learn library) and the mean absolute errors of the models. These scores and errors can be seen in Appendix A and Appendix B for each climate and emission combination.

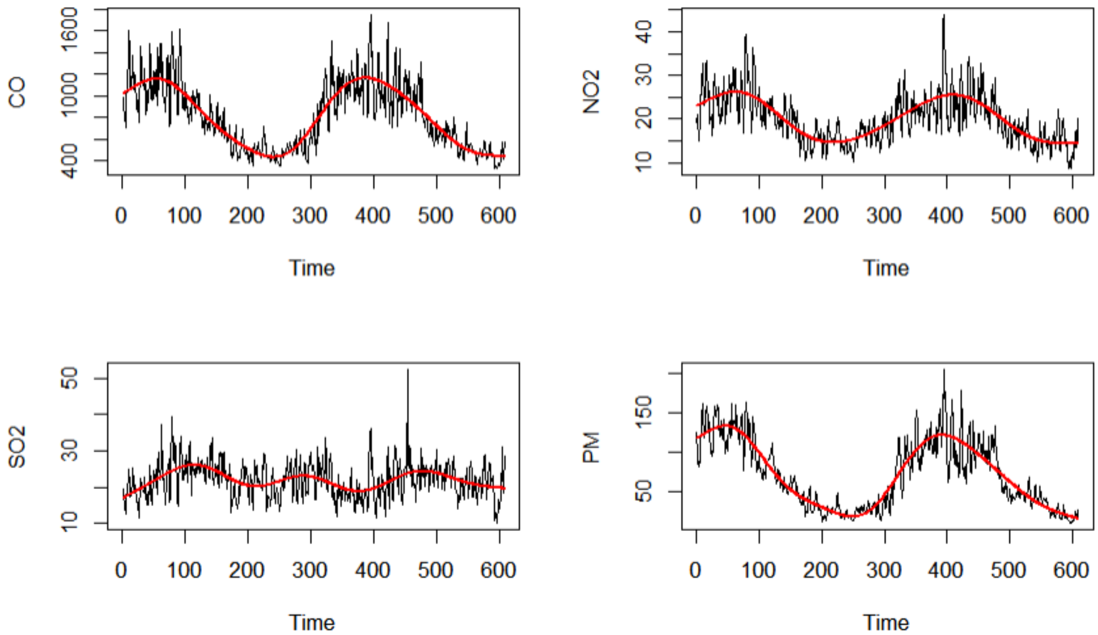
When comparing the scores and errors, the random forest regression model significantly outperformed all other models. Therefore, we decided to further analyze the random forest model by passing different combinations of variables to Python's Random Forest Regressor. For each combination of climate and emission, we tested different parameters for the random forest model, varying the numbers of trees (between 50, 100, 200, and 500) and maximum depths of said trees (between 2, 5, 7, and 9). We then stored the best performing parameters for each climate and emission comparison, as can be seen in Appendix C.

Once we had this information, we went ahead and back-filled the data for each climate and emission combination. We did this by building the random forest model using the entire OpenWeatherData dataset from November 2020 to July 2022, with the ideal parameters found earlier. We then used that model to predict the values for January 2018 to November 2020.

Despite our thorough testing, some of our back-filled data did not match the same seasonality trends that we expected to see based on the November 2020 to July 2022 data. We attributed these issues to the fact that some climates had more imputed data than other climates, and therefore the regression models would not be as accurate for those climates as they were for others.

3.2. Data Aggregation

Once we had imputed the data, we felt that we could get more readable, interpretable results from our analysis by performing it on each of the six climate groups, instead of performing it on each city as an isolated data value. Therefore, we decided to aggregate our data before moving forward. Before aggregation, we compared the emissions plots for all cities within each climate to confirm all cities had similar trends. We then aggregated the data by averaging emissions data for each climate group, and did a second exploratory data analysis to ensure the emissions still followed the same trends and seasonal patterns. An example of the emissions plots for one of our Savannah climate is below. This climate includes Pune, the city mentioned above in our initial analysis.



We can see only the trend in SO_2 differs a bit from Pune's, but represents all of the cities within the climate group well. This also reflects our hypothesis from the 2021 study that SO_2 would trend differently than the other emissions. Having seen this evidence that the aggregated data followed the same trends as the individual data, we decided that the aggregated data would be best to use moving forward.

4. Methods

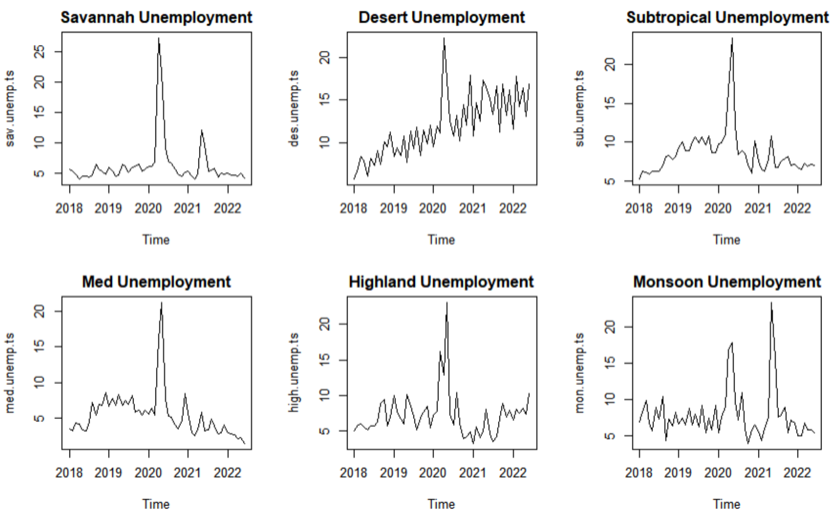
With our data ready to analyze, we moved on to address the major driving question of our project. In order to assess the relationship between the unemployment rate of the cities we were observing and the emission trends in those cities, we utilized many time series analysis methods. For each climate, we created time series and ACF plots for each air pollutant. We also created these plots for the unemployment rates for each climate, then compared the plots to see if any trend and/or seasonality was present. We confirmed that trends and seasonality were present in all of the emissions data, and aside from the DESERT climate, no trend or seasonality was present in the unemployment data.

Moving on, we used a non-parametric ANOVA model to simultaneously estimate the trend and seasonality components of each air pollutant. We then subtracted out the trend and seasonality components so we could properly analyze the emissions against the unemployment data. Following this, we combined the time series information for all of the de-trended and de-seasonalized emissions and the unemployment data. With the combined multivariate time series, we modeled the data using a Vector Autoregressive (VAR) model. A VAR model is used to linearly model the relationships between multiple time series as they change over time. The number of time lags can be adjusted within the model. This model was chosen due to its flexibility, multivariate functionality, and statistical testing methods. The specific statistical test we were interested in is called the Wald Test. This test assesses a feature called 'Granger causality', which is used to determine if one time series within a multivariate time series is useful in forecasting another time series. We used the VAR select function in R to choose the best number of lags based on Akaike information criterion (AIC).

5. Analysis and Results

The results from the VAR models for each climate were underwhelming, as all of our models had little to no significant coefficients. We tried restricting the VAR models to get better outcomes, resulting in slightly more significant results. However, we still did not see significant enough results to draw a clear conclusion. The Wald Test on each VAR model came back negative, meaning none of the emissions for any of the climates were significant in estimating the unemployment rate. Our time series plots along with these results lead us to believe that emissions and unemployment are not significantly related.

Analyzing the unemployment rate time series plots for each climate group, we saw the smallest spike in 2020 for the cities within the Tropical Monsoon climate. However, later on we see the largest spike in 2021 for the cities within the Tropical Monsoon climate. It seems all cities were drastically affected by the novel coronavirus pandemic. Another data value of note was that the cities within the desert climate had a constantly increasing rate of unemployment. Even though there was a large spike in unemployment around the onset of the coronavirus pandemic, it had the least impact on the already increasing unemployment rate in those Desert climate cities. The time series plots of the unemployment rates for all climate groups are modeled below.



For the purpose of comparing trends, we created matrices that mapped unemployment trends against emission trends for each quarter of data we analyzed. Looking at these matrices, it's difficult to detect any true trends between emissions changes and unemployment changes. Changes for many climates seem almost random. These matrices can be found in Appendix D.

Using Dash in Python, we created a dynamic visualization to look at the trends for each climate. This visualization scaled the emissions data and placed line graphs of the emissions over bar graphs of the unemployment data. An example screenshot can be shown as follows:

This visualization helped us confirm our that the unemployment trends did not align with emissions trends.

6. Conclusion

Considering all of the results from our analyses, we've determined that there's no significant relationship between emissions and unemployment rate across our city groupings in India. The results of the Granger causality tests along with the trend and seasonality plots were the largest indicators that these variables don't quite move with each other in time. When analyzing the VAR model, we only considered the unemployment rate based on emissions and not emissions based on the unemployment rate.

If we were to do further work on this project, we could test the VAR model going in the opposite direction. We could also explore different methods for data imputation, as the random forest regression model was not ideal for all climates. A different combination of air pollutants could be used for analysis as there are many more than the ones analyzed in this project. There could also be other factors affecting the unemployment rate and air pollutants such as political and sociocultural factors that were not taken into account in this project.

Appendix A. Model Scores

Humid Subtropical	CO	NO2	SO2	PM2.5
lasso	0.665035191	0.676384027	0.344221962	0.763258731
linreg	0.686490007	0.699340696	0.352568642	0.789616488
rf	0.958087386	0.957237232	0.910035983	0.975910653
rfe	0.620825282	0.687520645	0.28202032	0.324499748
svr	0.038719059	0.628262594	0.254910153	0.520473087

Mediterran	CO	NO2	SO2	PM2.5
lasso	0.580782192	0.570528217	0.556622239	0.724547404
linreg	0.586675071	0.577154048	0.561175636	0.728655661
rf	0.946859453	0.943888294	0.945424634	0.965769294
rfe	0.565871321	0.552248186	0.516937464	0.710108196
svr	0.154668916	0.565743728	0.575863396	0.580834125

Mid-Latitude Steppe and Desert	CO	NO2	SO2	PM2.5
lasso	0.557891691	0.590328618	0.429959454	0.646743103
linreg	0.562344505	0.59350797	0.437760225	0.648791968
rf	0.938872144	0.946082462	0.912493112	0.953457043
rfe	0.536681667	0.586603487	0.417422716	0.568634287
svr	0.027462935	0.506464569	0.35869293	0.446572465

Oceanic Subtropical Highland	CO	NO2	SO2	PM2.5
lasso	0.346171827	0.166840482	0.303352107	0.494962309
linreg	0.349397711	0.167114895	0.308063781	0.500218239
rf	0.924686247	0.921896692	0.917667955	0.940774912
rfe	0.009447162	0.121912659	0.283274404	0.473328746
svr	0.156802428	0.227025974	0.322913596	0.417573694

Tropical Monsoon	CO	NO2	SO2	PM2.5
lasso	0.437234941	0.305864924	0.368372712	0.485950466
linreg	0.445762775	0.307971178	0.370718376	0.489495844
rf	0.927663759	0.907480566	0.918745658	0.938176915
rfe	0.27609623	0.129135174	0.205704593	0.083191663
svr	0.059848644	0.231385123	0.351329349	0.285494568

Tropical Savanna	CO	NO2	SO2	PM2.5
lasso	0.683546844	0.551796788	0.028043989	0.850119613
linreg	0.690957117	0.552988969	0.037666696	0.856185713
rf	0.964274043	0.945100189	0.874918642	0.981420679
rfe	0.655719958	0.460014985	0.002511178	0.829912087
svr	0.0706331	0.499139327	0.012732044	0.554633269

Appendix B. Model Errors

Humid Subtropical	CO	NO2	SO2	PM2.5
rf	239.4170332	4.822064584	3.141660094	26.84891319
linreg	249.238461	4.893995929	3.02790951	30.41355377
lasso	259.952893	5.138734225	3.082184321	32.51899979
rfe	276.2690414	5.215012799	3.184822864	50.02038978
svr	414.4304831	5.483474803	3.171309188	41.7979463
Mediterran	CO	NO2	SO2	PM2.5
lasso	157.5620871	3.537777391	2.278066806	23.14036539
linreg	158.3284759	3.531892715	2.262819029	22.89081649
rf	160.3351292	3.555545799	2.067215832	20.91821938
rfe	163.0389248	3.435986926	2.320092989	23.05609863
svr	223.6201352	3.608475898	2.227720029	27.85677028
Mid-Latitude Steppe and Desert	CO	NO2	SO2	PM2.5
rf	189.976246	4.258112527	2.619425321	19.28013647
lasso	200.0680823	4.608787471	2.564159117	21.77792082
linreg	200.5380478	4.627172743	2.567913828	21.70758573
rfe	205.6861595	4.65504759	2.565713447	24.10267489
svr	303.1120578	4.639131588	2.606996573	26.14506893
Oceanic Subtropical Highland	CO	NO2	SO2	PM2.5
rf	44.49169367	1.296235326	0.346116932	7.224553974
linreg	49.26404582	1.676340976	0.402518725	8.802060655
lasso	49.54913739	1.677367561	0.402842468	8.94947245
svr	58.38542293	1.56182815	0.365735565	9.182577246
rfe	63.41997369	1.609061201	0.371166862	8.516110246
Tropical Monsoon	CO	NO2	SO2	PM2.5
rf	96.42452297	2.884924024	1.840808628	13.23499173
linreg	110.7672532	3.012537045	2.063587115	15.53112556
lasso	111.4901703	3.025808171	2.051540091	15.63605877
rfe	115.6055004	3.06727661	2.201388227	20.60377993
svr	138.5335697	2.976529782	2.039504943	17.10368448
Tropical Savanna	CO	NO2	SO2	PM2.5
rf	120.1802886	2.905131151	3.53501264	12.42590269
rfe	141.4090618	3.644260107	4.015717028	14.17621101
linreg	144.7870757	3.233583535	4.046403618	13.80891045
lasso	146.7020775	3.217824176	4.03164009	13.86471945
svr	259.1213586	3.264278386	3.984574409	24.03781571

Appendix C. Random Forest Optimal Parameters

	CO	NO2	SO2	PM2.5
Tropical Savanna	max_depth: 9, n_estimators: 200	max_depth: 9, n_estimators: 200	max_depth: 7, n_estimators: 500	max_depth: 9, n_estimators: 100
Mid-Latitude Steppe and Desert	max_depth: 5, n_estimators: 200	max_depth: 7, n_estimators: 100	max_depth: 2, n_estimators: 100	max_depth: 9, n_estimators: 200
Humid Subtropical	max_depth: 9, n_estimators: 200	max_depth: 7, n_estimators: 100	max_depth: 5, n_estimators: 200	max_depth: 7, n_estimators: 500
Mediterran	max_depth: 5, n_estimators: 200	max_depth: 5, n_estimators: 500	max_depth: 5, n_estimators: 200	max_depth: 5, n_estimators: 200
Tropical Monsoon	max_depth: 7, n_estimators: 50	max_depth: 7, n_estimators: 200	max_depth: 7, n_estimators: 200	max_depth: 5, n_estimators: 500
Oceanic Subtropical Highland	max_depth: 7, n_estimators: 100	max_depth: 5, n_estimators: 50	max_depth: 9, n_estimators: 50	max_depth: 7, n_estimators: 100
Marine West Coast	max_depth: 5, n_estimators: 500	max_depth: 5, n_estimators: 500	max_depth: 5, n_estimators: 200	max_depth: 7, n_estimators: 200

Appendix D. Quarterly Matrices

Legend:

TS - Tropical Savannah	MLSD - Mid-Latitude Steppe and Desert
HS - Humid Subtropical	M - Mediterran
TM - Tropical Monsoon	OSH - Oceanic Subtropical Highland

Q2 2018	Emissions +	Emissions -
Unemployment +	MLSD, HS	
Unemployment -		TS, M, TM, OSH

Q3 2018	Emissions +	Emissions -
Unemployment +	TS, OSH	MLSD, HS, M, TM
Unemployment -		

Q4 2018	Emissions +	Emissions -
Unemployment +	TS, MLSD, HS, M	OSH
Unemployment -	TM	

Q1 2019	Emissions +	Emissions -
Unemployment +	OSH	TS, HS
Unemployment -	TM	MLSD, M

Q2 2019	Emissions +	Emissions -
Unemployment +		TS, MLSD, HS, M, TM, OSH
Unemployment -		

Q3 2019	Emissions +	Emissions -
Unemployment +	HS	TS
Unemployment -		MLSD, M, TM, OSH

Q4 2019	Emissions +	Emissions -
Unemployment +	TS, MLSD, OSH, TM	
Unemployment -	HS, M	

Q1 2020	Emissions +	Emissions -
Unemployment +	TS, OSH	HS
Unemployment -	TM	MLSD, M

Q2 2020	Emissions +	Emissions -
Unemployment +		TS, HS
Unemployment -		MLSD, M, OSH, TM

Q3 2020	Emissions +	Emissions -
Unemployment +	MLSD, M	
Unemployment -	HS, OSH, TM	TS

Q4 2020	Emissions +	Emissions -
Unemployment +	TS, MLSD, HS, M, TM, OSH	
Unemployment -		

Q1 2021	Emissions +	Emissions -
Unemployment +	OSH	TM, MLSD
Unemployment -		TS, OSH, M

Q2 2021	Emissions +	Emissions -
Unemployment +		TS, TM
Unemployment -		MLSD, HS, M, OSH

Q3 2021	Emissions +	Emissions -
Unemployment +	TS	HS, M, TM, OSH
Unemployment -	MLSD	

Q4 2021	Emissions +	Emissions -
Unemployment +	TS, M, TM	
Unemployment -	MLSD, HS, OSH	

Q1 2022	Emissions +	Emissions -
Unemployment +	HS	MLSD, TM, OSH
Unemployment -		TS, M

Q2 2022	Emissions +	Emissions -
Unemployment +		MLSD, HS
Unemployment -	OSH	TS, M, TM

Q3 2022	Emissions +	Emissions -
Unemployment +		TS, HS, M
Unemployment -		MLSD, OSH, TM

References

[1] D. Fowler, J. A. Pyle, M. A. Sutton, M. L. Williams, Global air quality, past present and future: an introduction, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 378 (2020) 20190323. URL: `https://doi.org/10.1098/rsta.2019.0323`. doi:10.1098/rsta.2019.0323.

[2] R. S. Sokhi, V. Singh, X. Querol, S. Finardi, A. C. Targino, M. de Fatima Andrade, R. Pavlovic, R. M. Garland, J. Massagué, S. Kong, A. Baklanov, L. Ren, O. Tarasova, G. Carmichael, V.-H. Peuch, V. Anand, G. Arbilla, K. Badali, G. Beig, L. C. Belalcazar, A. Bolignano, P. Brimblecombe, P. Camacho, A. Casallas, J.-P. Charland, J. Choi, E. Chourdakis, I. Coll, M. Collins, J. Cyrys, C. M. da Silva, A. D. D. Giosa, A. D. Leo, C. Ferro, M. Gavidia-Calderon, A. Gayen, A. Ginzburg, F. Godefroy, Y. A. Gonzalez, M. Guevara-Luna, S. M. Haque, H. Havenga, D. Herod, U. Hõrrak, T. Hussein, S. Ibarra, M. Jaimes, M. Kaasik, R. Khaiwal, J. Kim, A. Kousa, J. Kukkonen, M. Kulmala, J. Kuula, N. L. Violette, G. Lanzani, X. Liu, S. MacDougall, P. M. Manseau, G. Marchegiani, B. McDonald, S. V. Mishra, L. T. Molina, D. Mooibroek, S. Mor, N. Moussiopoulos, F. Murena, J. V. Niemi, S. Noe, T. Nogueira, M. Norman, J. L. Pérez-Camaño, T. Petäjä, S. Piketh, A. Rathod, K. Reid, A. Retama, O. Rivera, N. Y. Rojas, J. P. Rojas-Quincho, R. S. José, O. Sánchez, R. J. Seguel, S. Sillanpää, Y. Su, N. Tapper, A. Terrazas, H. Timonen, D. Toscano, G. Tsegas, G. J. Velders, C. Vlachokostas, E. von Schneidemesser, R. VPM, R. Yadav, R. Zalakeviciute, M. Zavala, A global observational analysis to understand changes in air quality during exceptionally low anthropogenic emission conditions, Environment International 157 (2021) 106818. URL: `https://doi.org/10.1016/j.envint.2021.106818`. doi:10.1016/j.envint.2021.106818.

[3] N. P. of India, Website of central pollution control board, 2012. URL: `https://www.india.gov.in/official-website-central-pollution-control-board`.

[4] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, E. F. Wood, Present and future köppen-geiger climate classification maps at 1-km resolution, Scientific Data 5 (2018). URL: `https://doi.org/10.1038/sdata.2018.214`. doi:10.1038/sdata.2018.214.

[5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: machine learning in python, Journal of machine learning research 12 (2011) 2825–2830.