CO_2 EMISSIONS AND ENERGY SOURCES

Ben Moss

Dominic Martorano

Daniel Kim

DSI 1211 February 2024

PROBLEM STATEMENT

- Communities across the United States are facing the dangers of climate change and air pollution. To address the issue, we are surveying air pollution and renewable energy datasets to create predictive models that can be utilized by policymakers, economists, and civil society organizations. The goal of this research is to better aid our understanding on how air pollution is affecting our society as a whole. The model will be capable of predicting air pollution levels as a function of changes in renewable energy and fossil fuel productions.

DATA CLEANING AND EXPLORATORY DATA ANALYSIS

- Import Data in from sources
- For all datasets - Set Date column as index
- In some cases (EIA datasets) utilize `melt` function to reformat the data
- Kaggle Datasets - Graph data to observe trends prior to modeling
- Commence Modeling

REGRESSION MODELING

- Most Successful Regression Models
  - Third Place: Linear Regression
  - Second Place: K-Nearest Neighbors
  - First Place: Bagging Tree Regressor

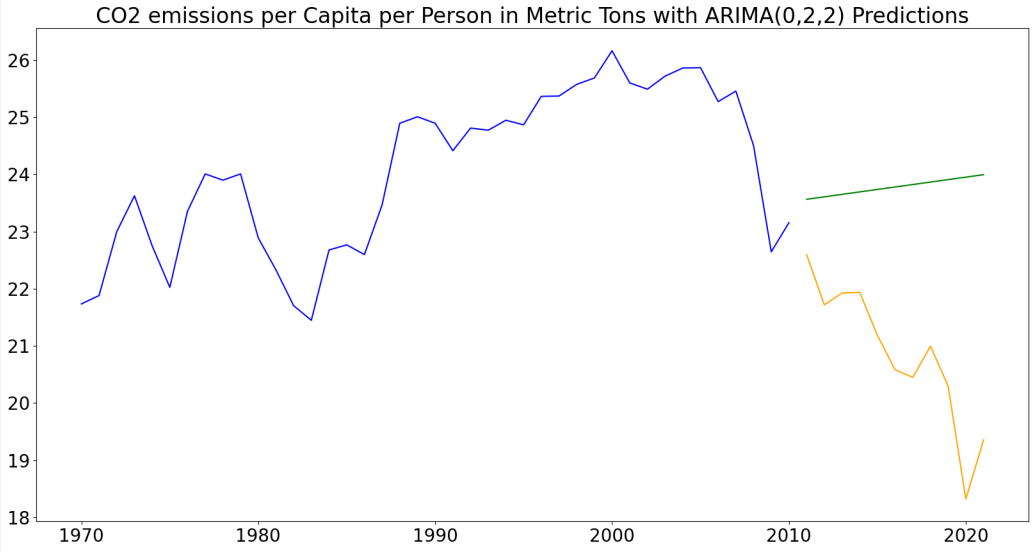
Regression Models	**R^2**		**RMSE**	
	Train	Test	Train	Test
Bagging Tree	0.9989	0.9953	0.5586	1.2230
KNN	1.0	0.99476	0.0	1.2920
Linear	0.9956	0.9941	1.1213	1.371

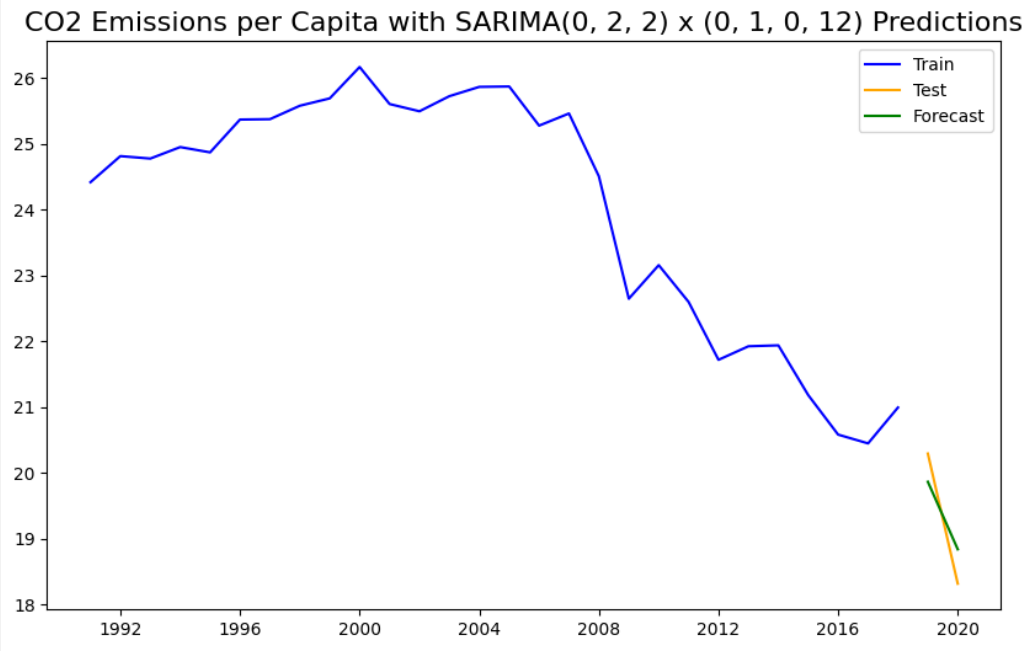
TIME SERIES MODELING

- ARIMA
- SARIMA (Small Sample Size)

TS Models	AIC	RMSE
ARIMA	89.124	3.330
SARIMA	36.0	0.4773

TIME SERIES PLOTS





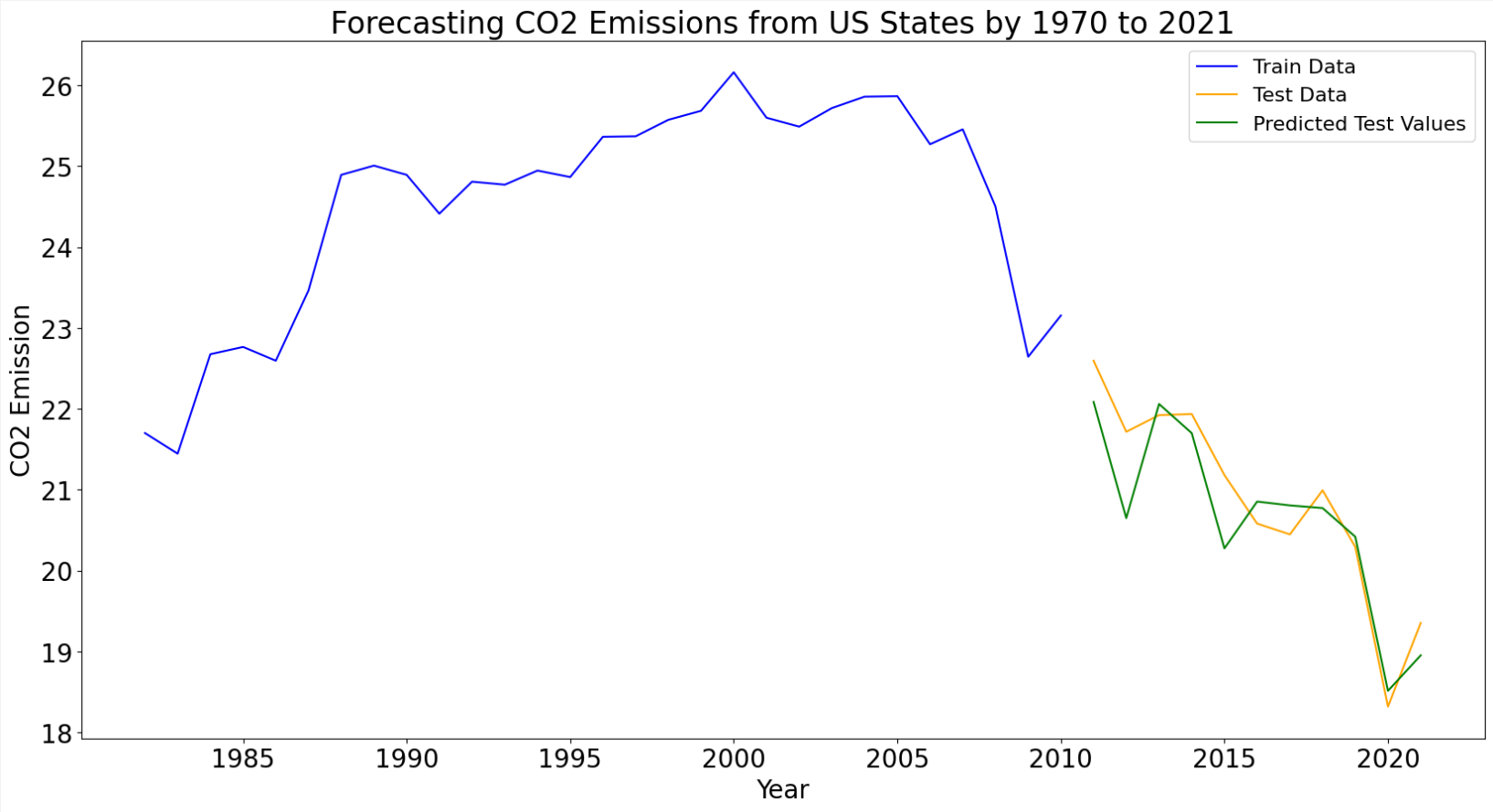
LINEAR TIME SERIES MODEL

- Year column as index
- Convert Year to DateTime Index
- CO2 Value as the target variable
- All energy values as features
- Test size as 20 percent and Shuffle to False

MODEL SUMMARY

- AIC Score: -27.74 (Negative AIC indicates less information loss than a positive AIC and therefore a better model.)
- R-squared: 0.997
- RMSE: 0.5009
- R2 score of test data: 0.8212

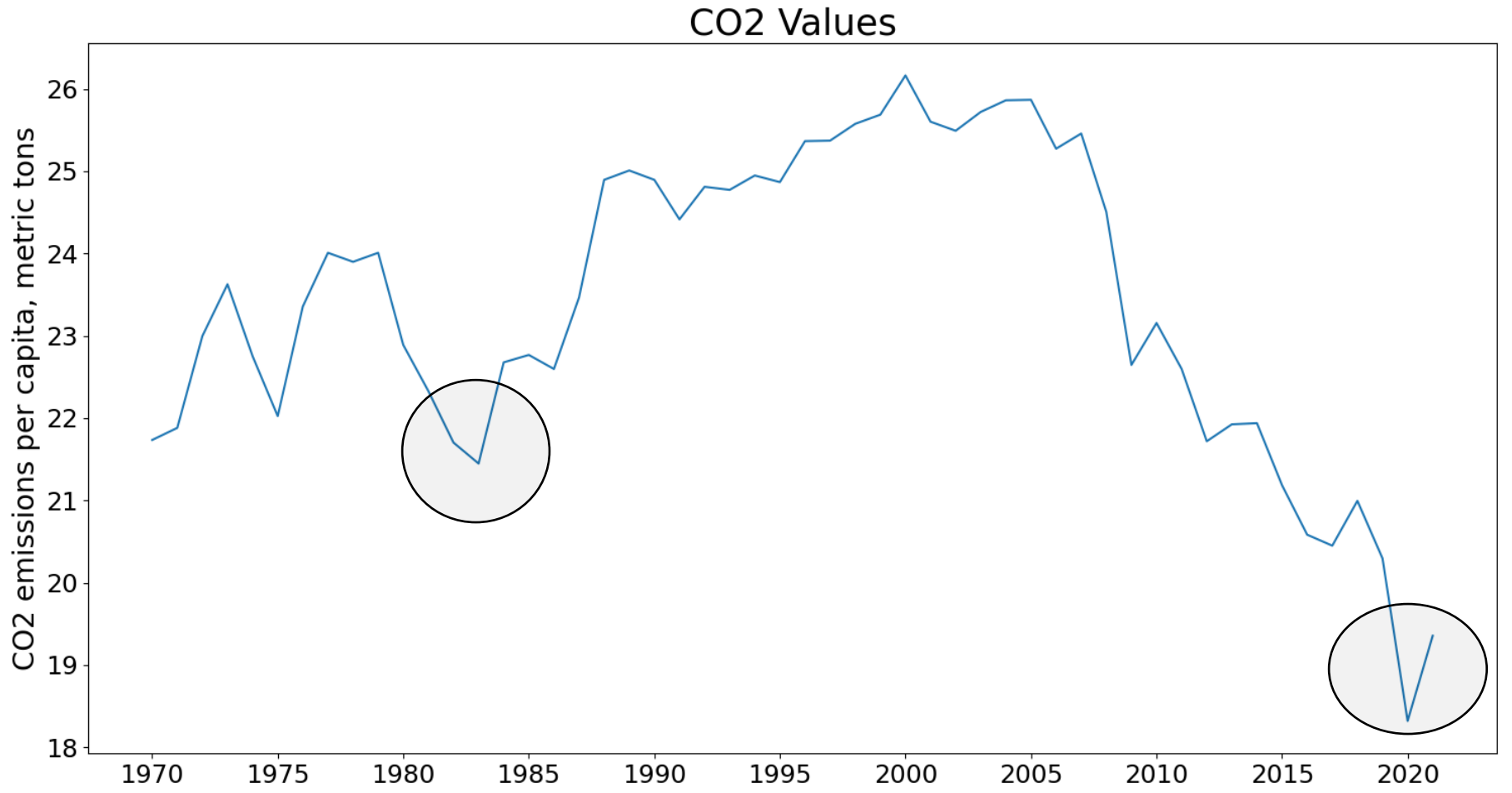
LINEAR TIME SERIES GRAPH



OBSERVING TRENDS IN CO_2 EMISSIONS

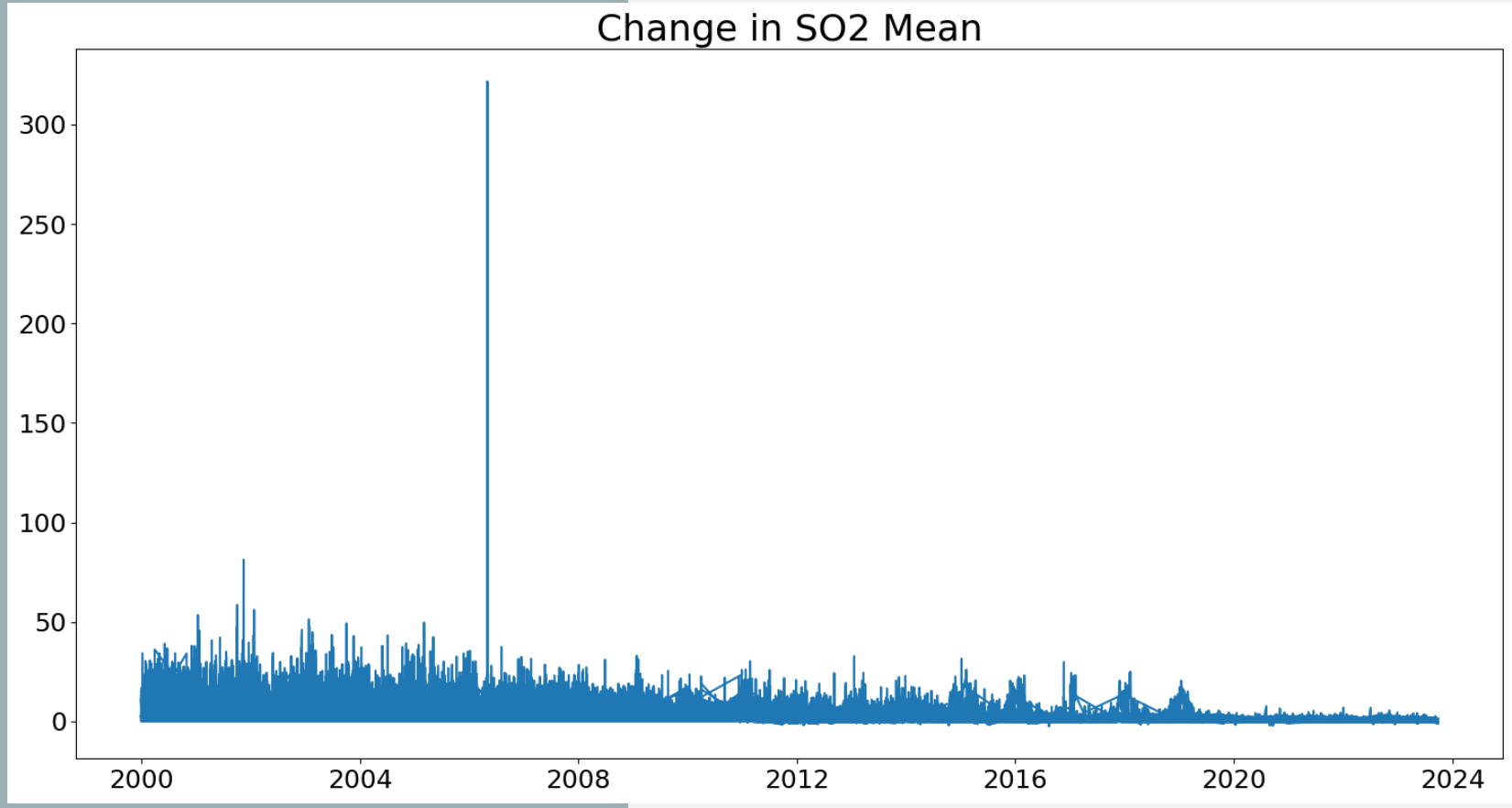
Drop in emissions in the early 80s

Decrease in 2020 followed by a slight rebound



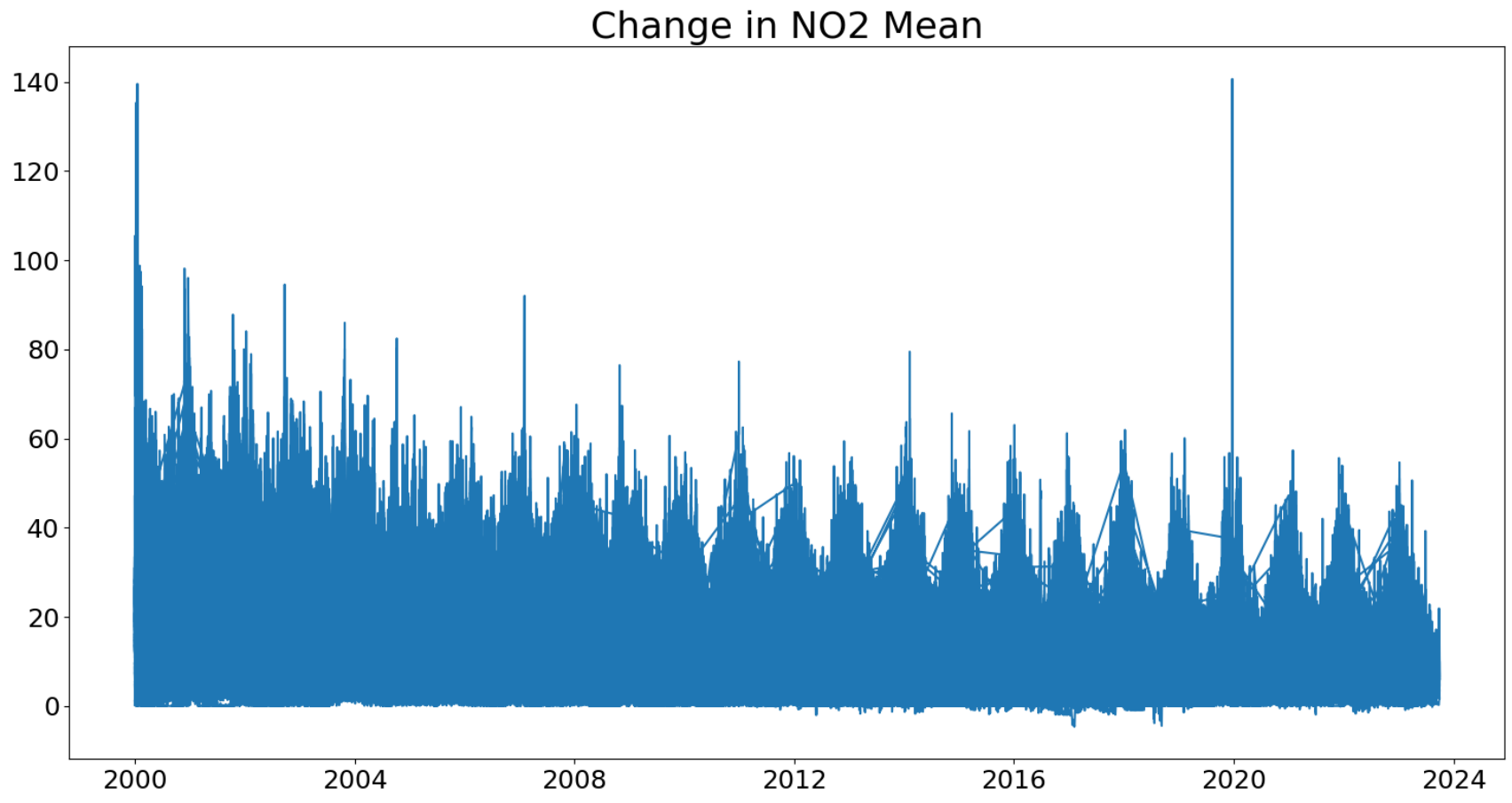
OBSERVING TRENDS IN SO_2 EMISSIONS

Sharp increase in SO_2 emission in 2006



OBSERVING TRENDS IN NO_2 EMISSIONS

Sharp increase in NO_2 emissions in 2020 as a result of COVID-19 pandemic



MULTIVARIATE TIME SERIES MODEL

- CO_2 equation of multivariate model demonstrates anticipated rebound in CO_2 emissions after COVID-19 pandemic
- Slight rebound in emissions in post-COVID era forecasted to be followed by continued downward trend in CO_2 emissions
- Energy consumption harder to predict, but general downward trends in fossil fuels and upward trend in renewables forecasted
- VARIMA AIC: 86.9

CONCLUSIONS AND RECOMMENDATIONS

- Regression, time series, and multivariate time series models are highly effective at solving many kinds of problems
- Today, we demonstrated how these models may successfully predict pollution index values from energy sources.
- These techniques are highly appealing for individuals and organizations interested in learning how our environment is affected by anthropological activity.

DATA SOURCES

1. **https://www.kaggle.com/datasets/sogun3/uspollution**
2. **https://www.kaggle.com/datasets/anshtanwar/global-data-on-sustainable-energy**
3. **https://www.eia.gov/state/seds/seds-data-complete.php**
4. **https://www.eia.gov/electricity/data/state/**