**DEPT. OF CEN**

**AMRITA SCHOOL OF ENGINEERING**

**III - B.TECH**

**CSE-AI**

**SEMISTER-5**

**21MAT301**

**Mathmatics for Intelligent System 5**

END SEMISTER PROJECT

**DATE:** 18-01-2023

# Predicting International Carbon Dioxide and Greenhouse Gas Emission Levels with Dynamic Mode Decomposition And LSTM

Submitted by:

TEAM : 11

DIVITH PHOGAT (CB.EN.U4AIE20013)

KARNATI SAI PRASHANTH (CB.EN.U4AIE20027)

SAI RISHITH REDDY (CB.EN.U4AIE20036)

S.V.S.DHANUSH (CB.EN.U4AIE20068)

T.S.GANESH KUMAR (CB.EN.U4AIE20073)

Under the supervision of

**Dr.Gopalakrishnan)**

## ABSTRACT

Dynamic Mode Decomposition (DMD) has been utilized to predict future CO2 and GHG emissions. The objective of this study was to determine if DMD can be utilized to accurately predict future emissions, and to compare the results of those DMD predictions using a data set of CO2 emission levels from countries and areas. For this project, a data-set containing annual CO2 emissions levels in 37 countries was used. The dataset was reconstructed using the CO2 DMD model, and it was used to predict GHG emission values and also CO2, while the 2015 and 2020 data were used for comparison. The results showed that DMD was able to provide accurate estimates of future CO2, GHG. -

# Contents

# Chapter 1

# Introduction

In this project we will use Dynamic Mode Decomposition (DMD) for two sections of the United Nations' International Greenhouse Gas (GHG) Inventory Data. The first part of the data set is a collection of international carbon dioxide (CO2) emissions across countries and other areas from 1990 to 2014. The second part of the data set is a collection of international GHG emissions across the same countries and areas from 1990 to 2014. The project goals are to determine whether DMD can provide accurate estimations of future CO2 and GHG emissions, and to compare the results from DMD on the CO2 and GHG parts of the data set. DMD could also potentially provide a linear regression of best fit to each part of the data set. One major question is if DMD will result in similar types of estimations for both CO2 and GHG emissions or if it could give seemingly contradictory estimations for CO2 and GHG emissions. An example of this would be if CO2 emissions are estimated to decrease but GHG emissions are estimated to increase.

# Chapter 2

# Methods

DMD is a mathematical tool that relies on measurements from different points in time like historical data, and the tool is generally considered to be equation-free [5]. This means DMD does not attempt to find a set of equations that directly calculate the entire data set [5]. Instead, DMD approximates a nonlinear set of data into a linear model. This linear model is a least-square fit system that is able to provide estimations for near-future states of the data set. In addition, DMD is useful because its process includes dimension reduction [3]. DMD can be utilized for reconstruction as well, such as recovering data with gaps in their measurements or signals [6]. If the United Nations data set is split into one part for the CO2 emissions and one part for the GHG emissions, the CO2 part should be able to go through DMD to provide a least-square fit system that estimates the amount of CO2 emissions per year in the near-future and the GHG part should be provided its own least-square fit system through DMD that estimates the near-future amount of GHG emissions per year. To utilize DMD, each part of the data set first needs to be split into two matrices. As an example, this section will assume DMD is being performed on the part of the data set that contains all of the information on CO2 emissions. The two matrices are

$$\begin{aligned} X &= [\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots \bar{x}_{m-1}] \\ \bar{X} &= [\bar{x}_2, \bar{x}_3, \bar{x}_4, \ldots \bar{x}_m] \end{aligned} \tag{2.1}$$

X is an array of vectors in which $\bar{x}_1$ is a column containing the CO2 emissions of every country or area in the year 1990 and $\bar{x}_{m-1}$ is is a column containing the CO2 emissions of every country or area in the year 2013. Similarly, $\bar{X}$ is an array of vectors in which$\bar{x}_2$ is a column containing the CO2 emissions of every country or area in the year 1991 and $\bar{x}_m$ is a column containing the CO2 emissions of every country or area in the year 2014.

Essentially, X is a matrix containing the CO2 emissions for every country or area each year from 1990 to 2013 and $\bar{X}$ is a matrix containing the CO2 emissions for every country or area each year from 1991 to 2014. The next step is to find the rank of X and let the rank be equal to r, such that r is equal to the number of linearly independent columns in X. Singular Value Decomposition (SVD) is performed next with rank truncation, which creates a reduced square matrix, Xr. The number of rows and columns Xr contains is equal to the rank of X such that

$$X_r = U_r \Sigma_r V_r^* \tag{2.2}$$

U is an array of the eigenvectors of $XX^*$, V is an array of the eigenvectors of $X^*X$, and $\Sigma$ is a rectangular diagonal matrix whose diagonal entries are called $\sigma$. $U_r$ is a matrix containing the first r columns of U, $V_r$ is a matrix containing the first r columns of V , and $\Sigma$ r is a square matrix containing the first r rows and the first r columns of $\Sigma$. $V_r$ is the conjugate transpose of $V_r$. The square matrix $A_r$ is calculated afterwards:

$$\begin{aligned} A_r &= U_r^* \bar{X} X^+ U_r \\ A_r &= U_r^* \bar{X} V_r \Sigma_r^{-1} \end{aligned} \tag{2.3}$$

$X^+$ is the pseudo-inverse of X. A non square matrix like X does not have a standard inverse, so the pseudo-inverse is considered the closest possible solution to finding the inverse of a non square matrix. The next step is to find the eigenvectors and eigenvalues of Ar such that W is an array in which each column is an eigenvector of Ar and $[\lambda 1, \lambda 2, ..., \lambda_r]$ is an array in with each entry is an eigenvalue of Ar. After this, the following is computed:

$$\Phi = \bar{X} V_r \Sigma_r^{-1} W \tag{2.4}$$

$\Phi$ is a matrix whose columns are called the DMD modes. The next is to calculate the following equation.

$$x_k = \sum_{j=1}^{r} \phi_j \lambda_j^{k-1} b_j$$

The countries and areas included in the United Nations data set that contain 1990-2014 data on CO2 emissions and GHG emissions are the following: Australia (1), Austria (2), Belarus (3), Belgium (4), Canada (5), Croatia (6), Cyprus (7), the Czech Republic (8), Denmark (9), Estonia (10), the European Union (11), Finland (12), France (13), Germany (14), Greece (15), Hungary (16), Iceland (17), Ireland (18), Italy (19), Japan (20), Latvia (21), Lithuania (22), Malta (23), the Netherlands (24), New Zealand (25), Norway (26), Poland (27), Portugal (28), Romania (29), the Russian Federation (30), Slovakia (31), Slovenia (32), Sweden (33), Switzerland (34), Turkey (35), Ukraine (36), and the United States (37). The number next to each location corresponds to where the location can be found in the figures for the Results section. Additional countries and areas in the data set are Bulgaria, Liechtenstein, Luxembourg, Monaco, Spain, and the United Kingdom. However, Monaco's CO2 data covers 1990-2013 instead of 1990-2014 so it needs to be excluded from the data set in order to accurately utilize DMD. Bulgaria, Liechtenstein, Luxembourg, Spain, and the United Kingdom also need to be excluded because they include CO2 data, but not the corresponding GHG data. To prepare the data set for DMD, a Numerical Python (NumPy) array is created from the CO2 emission values for each country or area from 2014 to 1990. The array is then flipped so the values go from the year 1990 to the year 2014. Once each country or area has its own NumPy array, the above matrix X is created using the CO2 values for each country or area from 1990 to 2013 and the above matrix $X^{-}$ is created using the CO2 values for each country or area from 1991 to 2014. This process is repeated for the GHG emission values for each country or area except the resulting matrices are called Y and $Y^{-}$ instead. Once X, $X^{-}$, Y, and $Y^{-}$ are created, the data is ready for DMD. Table 1 displays a small section of X from equation(1) with labels added for each location and year.

# Chapter 3

# Analysis and Results

## 3.1 DMD vs Dataset

We were able to perform DMD successfully on the CO2 emissions data set and the GHG emissions data set. The numbers on the x-axis of each figure in this section match the numbers given in the previous section to each country or area in the data set. For example, Australia corresponds to 1 on the x-axis, Austria corresponds to 2 on the x-axis, and the United States corresponds to 37 on the x-axis.

Below figure 3.1 is the Eigenvalue Spectrum for CO2 and figure 3.2 for GHG data which helps to find the stable and unstable modes, if the eigenvalues are on the line or inside the unit circle they are considered stable modes and in other cases, it is considered unstable modes.

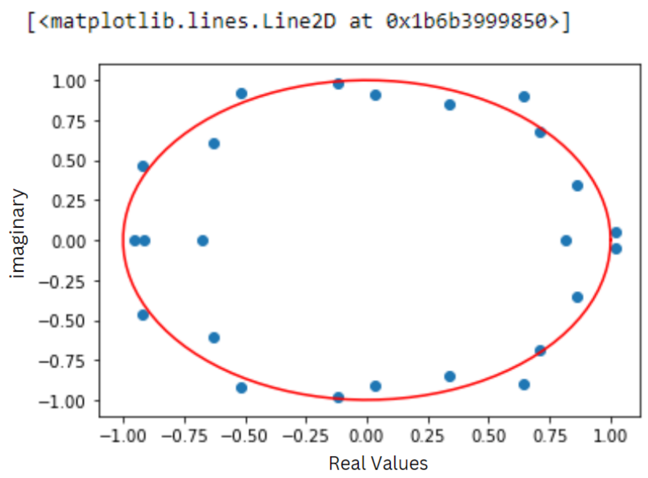


Figure 3.1: Eigenvalue Spectrum for CO2

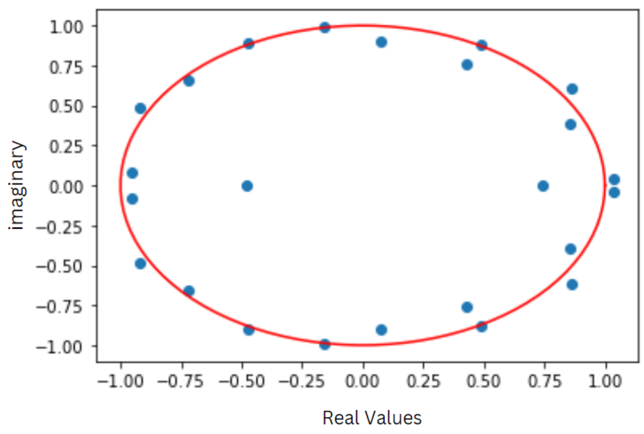


Figure 3.2: Eigenvalue Spectrum for GHG

It should be noted that all of the DMD models for both CO2 emissions and GHG emissions resulted in entries with complex values. When plotting these data points, Python only plots the real number in the entry and ignores the imaginary number. As a simplified example, if one of the entries is 50 + 3i, then Python only plots the number 50. However, the loss of the imaginary values in the plots does not seem to significantly impact the plots because the real values in each entry are many times larger than the imaginary values in each entry.

Figure 3.3 shows that the DMD model gave the same results for CO2 estimates in 2010. The CO2 emissions estimate was close in accuracy to the large majority of countries or areas included in the data set.

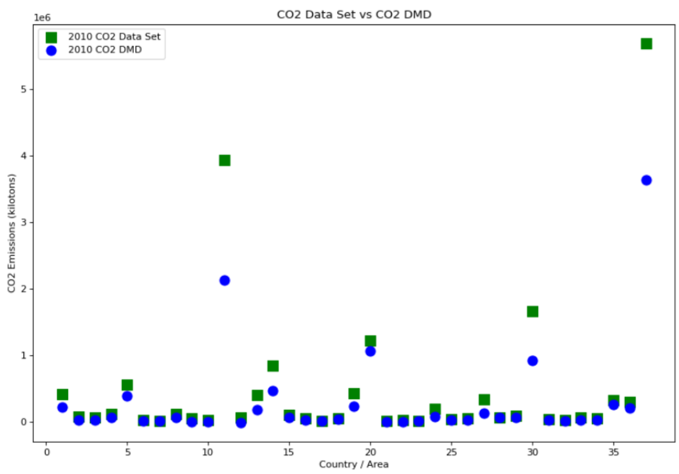


Figure 3.3: Comparison between data set and DMD for CO2 in year 2010

The three major exceptions were the European Union, the Russian Federation, and the United States which all had drastically higher records of CO2 emissions than the DMD model expected. However, the DMD model still seemed to mostly capture the general trend of CO2 emissions in each country or area. The DMD model for GHG estimates in 2010 (figure 3.4) generally displayed similar types of results to the CO2 estimates with one major exception.

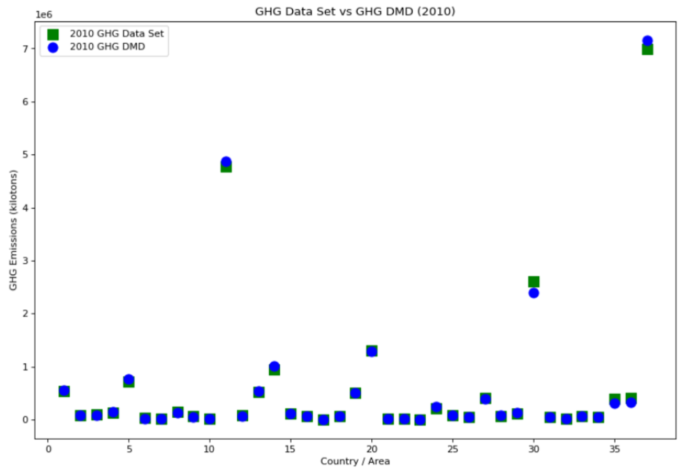


Figure 3.4: Comparison between the data set and DMD for GHG in the year 2010

However, the DMD model for GHG were much more accurate at predicting 2010 GHG emissions from the European Union, the Russian Federation, and the United States than the DMD model for CO2 was at predicting the 2010 CO2 emissions from the same three locations. Overall, the DMD model for GHG emissions appear to be highly accurate.

Figure 3.5 displays predicted CO2 emissions from 2015 to 2020 using the DMD model.

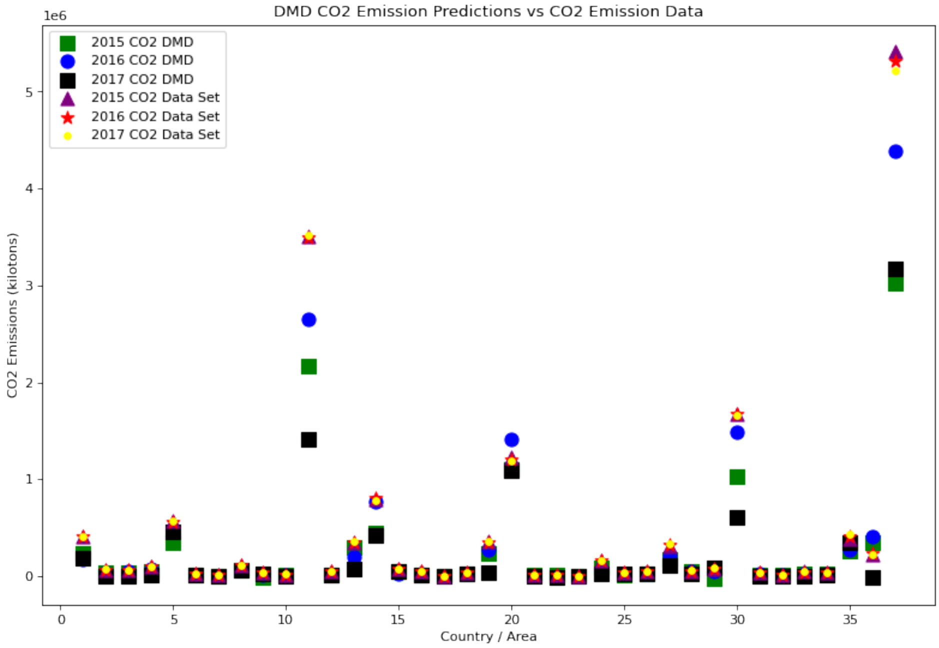


Figure 3.5: Plot for DMD predicted values from 2015 - 2020 for CO2

2015 is the first year included in the figure because the original data set ended at 2014. Most of the locations are predicted to have roughly the same level of CO2 emissions or slight increases in emissions from 2015 to 2020. The outliers with higher annual CO2 emissions like the European Union, Japan, the Russian Federation, and the United States show much more variance in the year-to-year predictions than the locations with lower annual CO2 emissions. By 2019 and 2020, the DMD model also starts predicting some locations to have a negative annual CO2 emissions value. Figure 3.6 displays predicted GHG emissions from 2015 to 2020 using the DMD model,

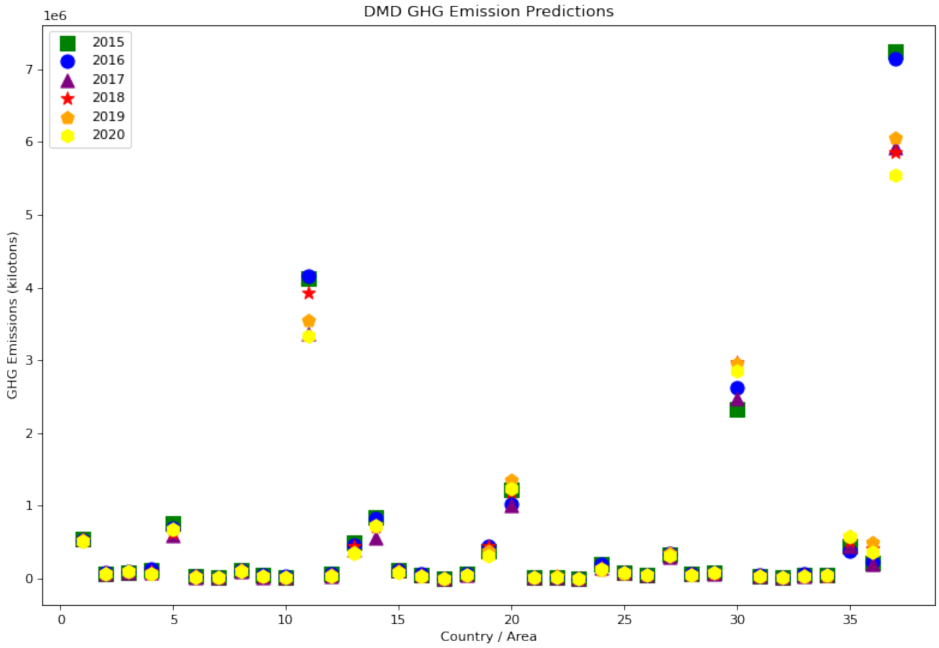


Figure 3.6: Plot for DMD predicted values from 2015 - 2020 for GHG

For the GHG prediction, it has less variation in the outlier locations than CO2 prediction. Most of the locations are still predicted to have roughly the same emission levels or slightly higher emission levels from 2015 to 2020. In addition, none of the DMD estimates in GHG emission seem to predict a negative GHG emission value from any of the countries or areas in 2019 or 2020.

Figure 3.7: Comparing predicted 2015,2016,2017 CO2 emissions from the DMD model for CO2 emissions with reported 2015,2016,2017 CO2 emission values. Data set from.

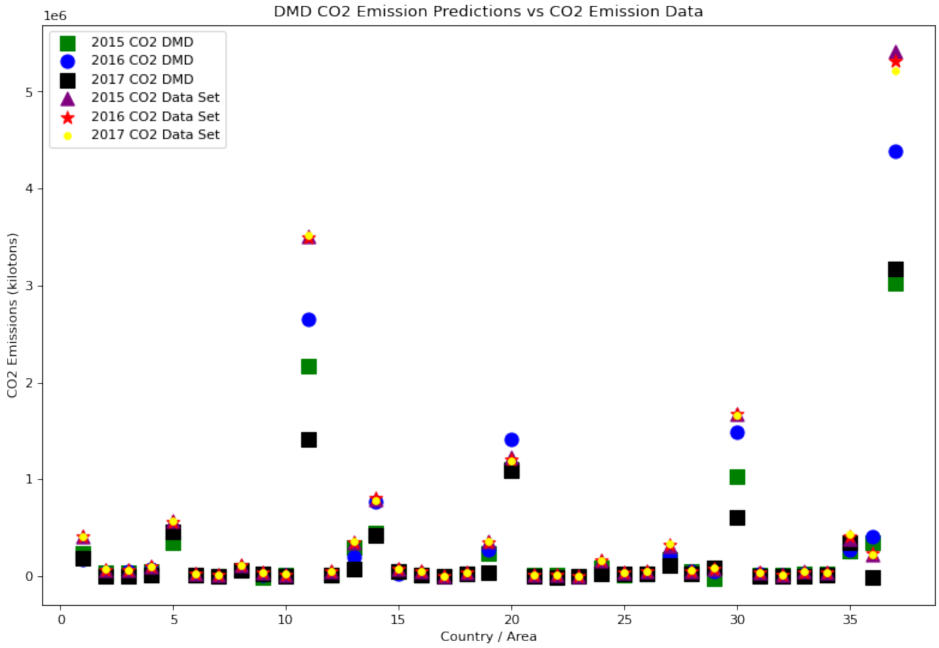


Figure 3.7

Figure 3.8: Comparing predicted 2015,2016 and 2017 GHG emissions from the DMD model for GHG emissions with reported 2015,2016,2017 GHG emission values.

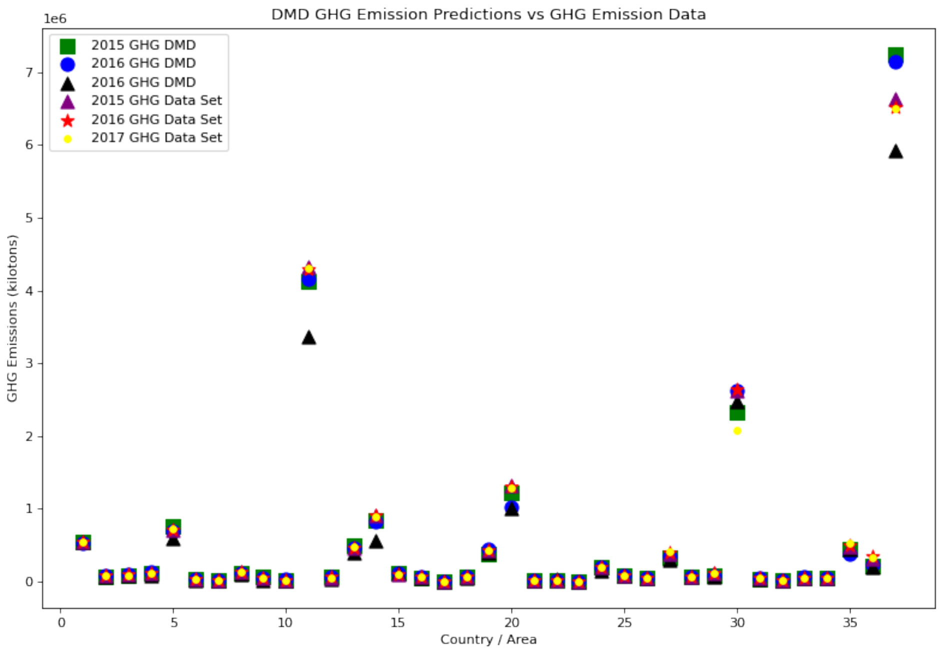


Figure 3.8

Figure 3.9: Comparing predicted 2018, 2019, and 2020 CO2 emissions from the DMD model for CO2 emissions with reported 2018, 2019, and 2020 CO2 emission values. Data

set from.

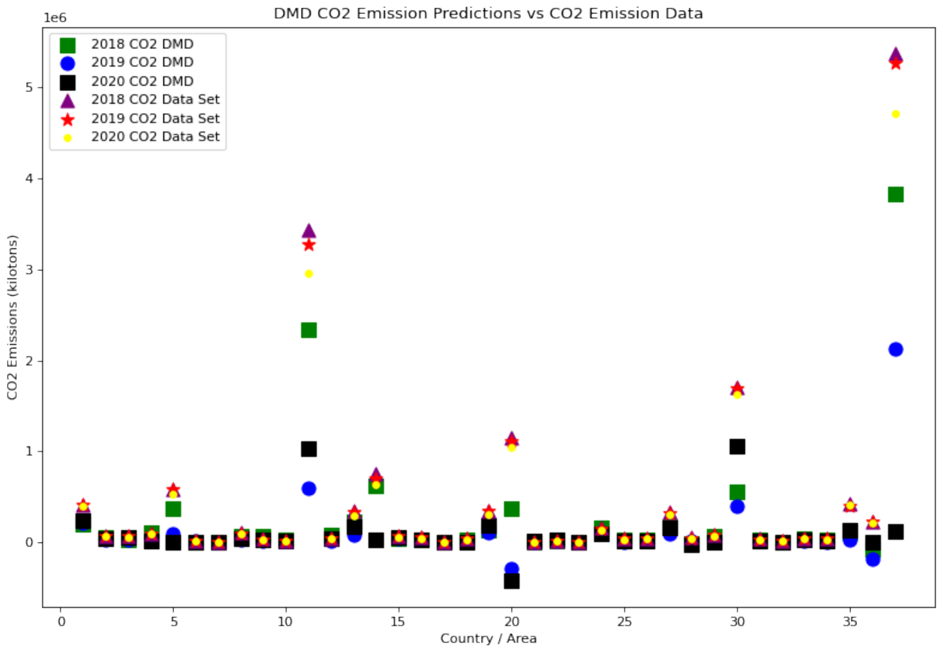


Figure 3.9

Figure 3.10: Comparing predicted 2018, 2019, and 2020 GHG emissions from the DMD model for GHG emissions with reported 2018, 2019, and 2020 GHG emission values.

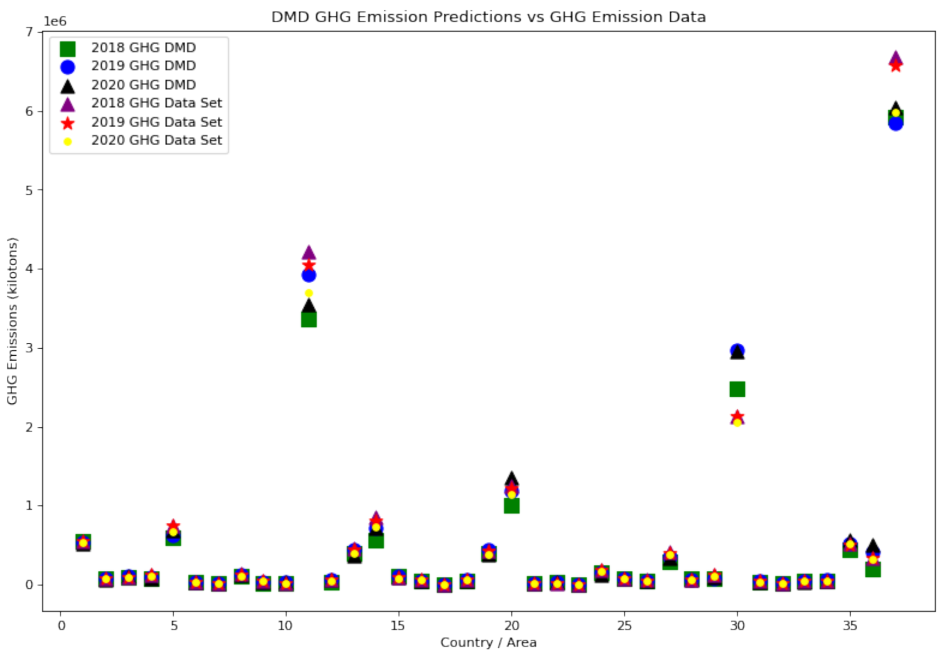


Figure 3.10

Figure 3.11: Comparing the DMD model for Australia's CO2 emission levels with Australia's documented CO2 emission levels

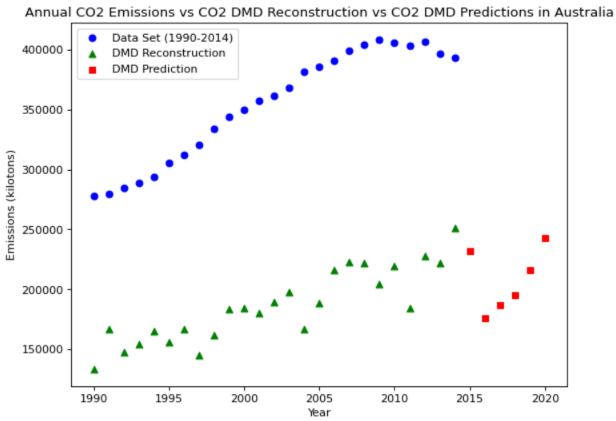


Figure 3.11

Figure 3.12: Comparing the DMD model for Australia's GHG emission levels with Australia's documented GHG emission levels

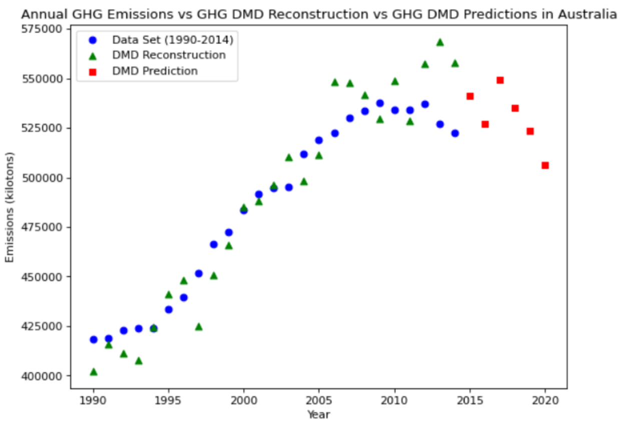


Figure 3.12

The RMSE value for CO2 prediction using DMD model is 0.547 and for GHG is 0.133, less the RMSE value less the deviation of predicted values from original values.

## 3.2 Comparison with LSTM

We have also tried to predict our data using LSTM for both CO2 and GHG data in the year 2019 for all the countries, It is a variety of recurrent neural networks (RNNs) that are capable of learning long-term dependencies, especially in sequence prediction problems.

It has feedback connections, i.e., it is capable of processing the entire sequence of data, apart from single data points such as images.

Below figure 3.13 is the comparison between the LSTM predicted values for CO2 emissions in the year 2019 and the actual values.

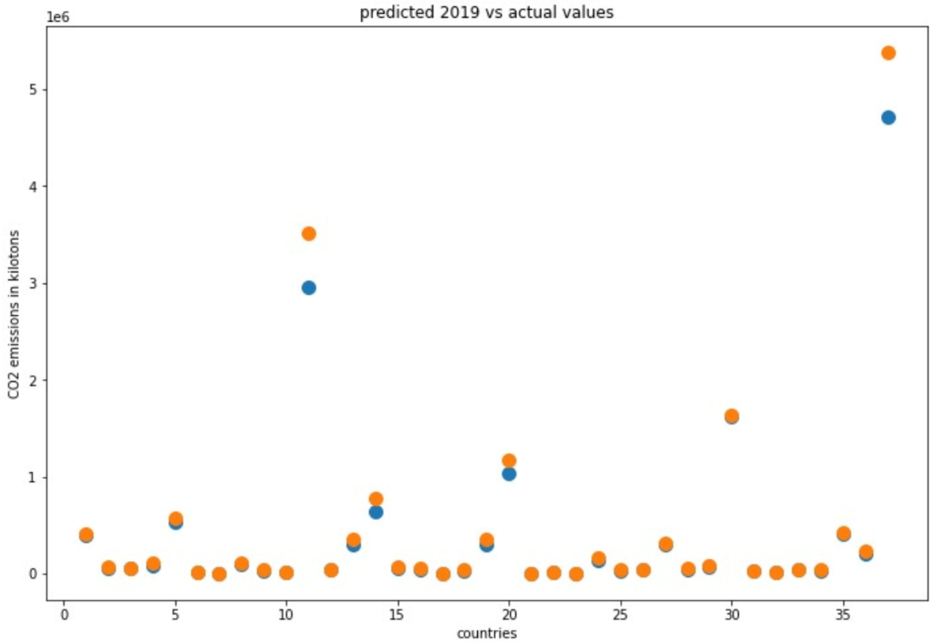


Figure 3.13

figure 3.14 is the comparison between the LSTM predicted values for CO2 emissions in the year 2019 and the actual values.

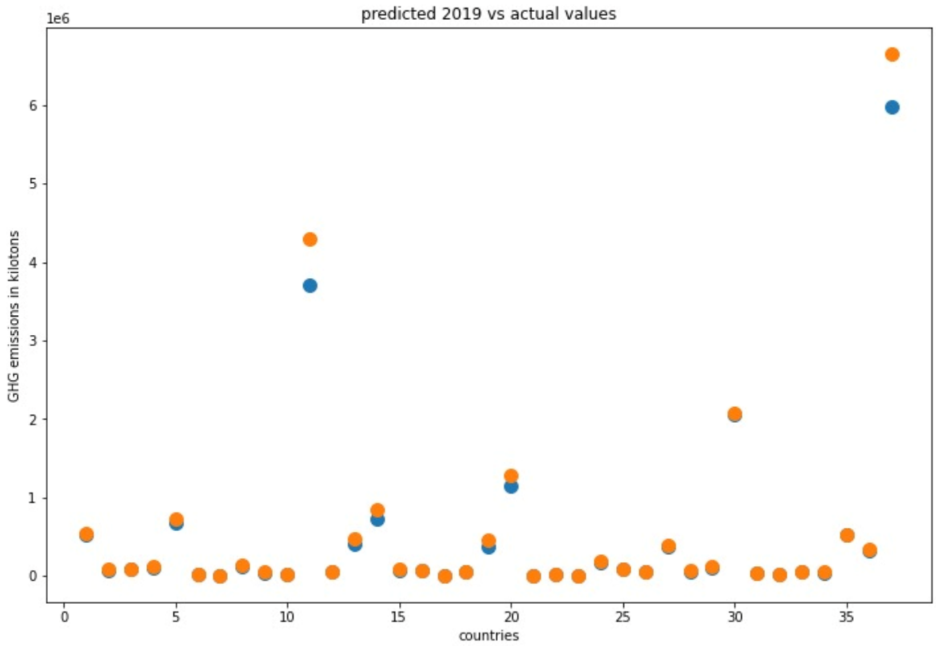


Figure 3.14

The RMSE value for CO2 prediction using DMD model is 0.00823 and for GHG is 0.0072, less the RMSE value less the deviation of predicted values from original values.

# Chapter 4

# Conclusion

The CO2 and GHG DMD models show similar trends in their results. The GHG DMD model is consistently more accurate than the CO2 DMD model. Perhaps differences in the CO2 and GHG data sets. DMD showed mixed results for predicting CO2 emission levels but powerful results for predicting GHG emission levels. Even the DMD trend results for CO2 and GHG emission levels appeared consistent overall but the LSTM prediction was more accurate than the DMD model, but when compared to LSTM, DMD is more efficient.

## 4.1 References

Mohan, N., Soman, K.P., Sachin Kumar, S. (2018). A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. Applied Energy, 232, 229-244. Retrieved from: https://doi.org/10.1016/j.apenergy.2018.09.190.

Proctor, J.L., Eckho, P.A. (2015). Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. International Health, 7(2), 139-145. doi: 10.1093/inthealth/ihv009.

Clainche, S.L., Lorente, L.S., Vega, J.M. (2018). Wind predictions upstream wind turbines from a LiDAR database. Energies, 11(4), 543. Retrieved from: https://www.mdpi.com/1996-1073/11/3/543

Henson, R. (2014). The thinking person's guide to climate change. Boston: American Meteorological Society.

**Thank You**