Unveiling Greenhouse Gas Giants: Tackling Leading Industry Emissions for a Cooler Future

By: Aliya Icasiano, Justin Bailey, Janeth Lara, Jimena Lara, Jianhao Zhu
BUS2 194A: Statistical Analysis | Dr. Etu
December 4, 2023

INTRODUCTION

Greenhouse gases (GHGs), like carbon dioxide, methane, and nitrous oxide, occur naturally in the Earth's atmosphere. They possess the capability to retain heat from the sun when present in excessive quantities, causing the Greenhouse Effect. Although crucial for regulating Earth's temperature, an overabundance of these gases contributes to global warming, marked by the gradual rise in the planet's average temperature. Human activities have intensified and exacerbated the natural greenhouse effect. According to the Environmental Protection Agency, the total warming effect from GHGs added by humans to the Earth's atmosphere increased by 45% from 1990 to 2019. This alarming rise highlights the urgency of addressing these excess emissions and implementing strategies to mitigate global warming effects like severe weather events and rising sea levels. It is essential to recognize that GHG emissions are linked to numerous supply chain sectors, including transportation, manufacturing, agriculture, construction, food and beverage, among others. These sectors collectively contribute to GHG release through energy usage, logistics, production methods, and land-use practices.

PROBLEM STATEMENT

Our Problem: What industries will see the greatest growth of greenhouse gas emissions in 2025? Addressing this issue is crucial because GHGs rank among the primary drivers of global warming. Recent years have marked some of the hottest on record, highlighting the urgency of identifying the primary GHG contributors. It is vital to minimize GHG emissions as failing to do so will lead to alarming consequences in the future.

OBJECTIVES(S)

The objective of our project is to use time series forecasting analysis to identify which industry is projected to experience the most significant increase in greenhouse gas emissions by 2025. We aim to create a forecasting model demonstrating the projected increase of GHGs in each industry.

METHOD

Data Collection

Our dataset was collected from climatewatchdata.org, encompassing GHG emissions across 195 countries, categorized by nine sectors: agriculture, buildings, electricity and heat, fugitive emissions, industry, land-use change and forestry, manufacturing and construction, transportation, and waste. The dataset spans from 1990 to 2019.

Descriptive statistics

```
               Electricity and heat transportation  manufacturing and construction  agriculture
               Min.   : 8.630       Min.   :4.940   Min.   :4.730                   Min.   :3.720
               1st Qu.: 9.955       1st Qu.:5.473   1st Qu.:5.045                   1st Qu.:3.900
               Median :12.250       Median :6.590   Median :5.290                   Median :4.750
               Mean   :12.255       Mean   :6.518   Mean   :5.517                   Mean   :4.741
               3rd Qu.:15.000       3rd Qu.:7.345   3rd Qu.:6.168                   3rd Qu.:5.633
               Max.   :15.880       Max.   :8.430   Max.   :6.450                   Max.   :5.820
 fugitive emissions   buildings        industry     land-use change and forestry     waste
 Min.   :2.740      Min.   :2.210   Min.   :1.450   Min.   :1.360                 Min.   :1.01
 1st Qu.:2.853      1st Qu.:2.362   1st Qu.:1.910   1st Qu.:1.470                 1st Qu.:1.15
 Median :2.935      Median :2.780   Median :2.005   Median :1.480                 Median :1.32
 Mean   :2.990      Mean   :2.684   Mean   :2.145   Mean   :1.521                 Mean   :1.29
 3rd Qu.:3.165      3rd Qu.:2.962   3rd Qu.:2.438   3rd Qu.:1.558                 3rd Qu.:1.43
 Max.   :3.400      Max.   :3.110   Max.   :3.060   Max.   :1.870                 Max.   :1.63
```

Figure 1: Descriptive Statistics of GHG Emissions in each Industry from 1990 to 2019. The highest mean is electricity and heat (12.26 billion tons of GHGs emitted).

Data visualization

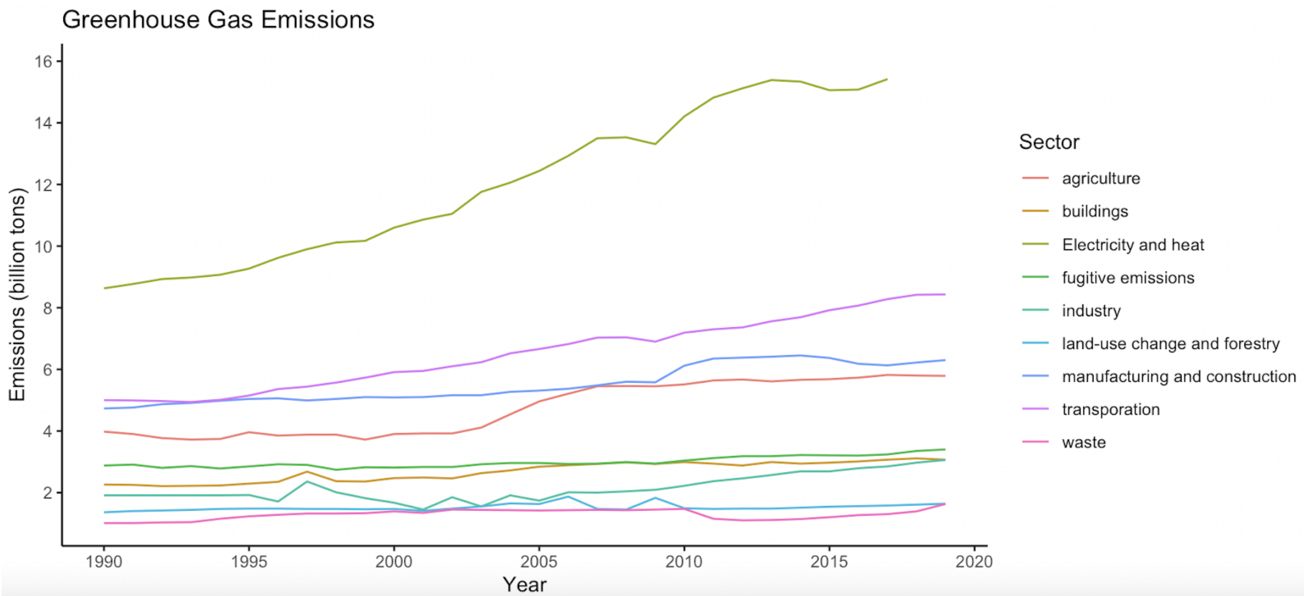

Figure 2: Line Graph of GHG Emissions in each Industry from 1990 to 2019.

Interpretations: The electricity and heat sector has seen an 83% increase in greenhouse gas emissions from 1990 to 2020, making it the top emitter in terms of GHGs. This industry has consistently been the largest contributor to greenhouse gas emissions since 1990.

Data analysis

This analysis involves using time series forecasting to study how greenhouse gas (GHG) emissions across various sectors relate to time. In this method, time (years) is the independent variable, while the GHG emissions for each industry is the dependent variable. The formula used for this analysis is Yt = b0 + b1(t), creating nine distinct equations where Yt represents sector-specific emissions in a given year. Here, b0 stands for the baseline emissions level unique to each sector, b1 signifies the estimated change in emissions for a sector with each year's passing, and t denotes the time in years. Assumptions include a significance level set at $\alpha = 0.05$ and the assumption of a linear relationship between time and industry-specific GHG emissions.

RESULTS

Final Model

Sector	Time Series Analysis Equation	P-Value	MAPE
Electricity and Heat	7.8219 + 0.2860t	0.00	2.80%
Transportation	4.5041 + 0.1299t	0.00	1.73%
Manufacturing and Construction	4.5381 + 0.0632t	0.00	3.27%
Agriculture	3.3281 + 0.0912t	0.00	5.81%
Fugitive Emissions	2.6892 + 0.0191t	0.00	2.27%
Buildings	2.1557 + 0.0341t	0.00	3.22%
Industry	1.5526 + 0.0382t	0.00	11.81%
Land-Use Change and Forestry	1.4259 + 0.0061t	0.01	4.09%
Waste	1.1733 + 0.0075t	0.03	10.70%

Figure 3: Model Summary.

Interpretation: The time series variables are significant; all P-values < 0.05. Additionally, the mean absolute percentage error (MAPE) is low for all industries and therefore, our forecasting accuracy is good.

Sector	2025 GHG Prediction
Electricity and Heat	18.12
Transportation	9.18
Manufacturing and Construction	6.81
Agriculture	6.61
Fugitive Emissions	3.39
Buildings	3.38
Industry	2.93
Land-Use Change and Forestry	1.65
Waste	1.44

Figure 4: GHG Prediction Values

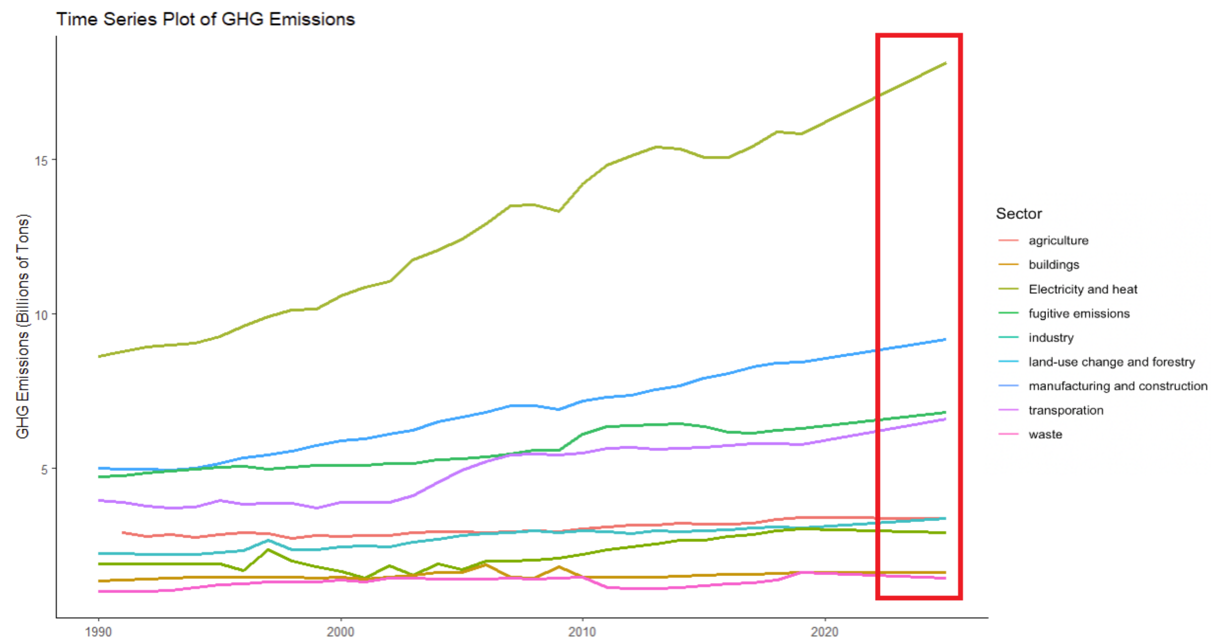

Figure 5: Time Series Plot of GHG Emissions.

Interpretations: The projected GHG output for 2025 is highlighted by the red box on the graph, displaying a pattern of linear increase. By 2025, the electricity and heat industry is expected to continue being the leading emitter of GHGs, showcasing the most rapid growth rate among all sectors.

Model assumption checks

A linear relationship between the independent variable and dependent variables.

- We checked for linear correlation using R and found that every sector has anywhere between a moderate and positive linear correlation with time.

CONCLUSIONS AND RECOMMENDATIONS

Based on the analysis, it is evident that by 2025, the primary contributors to greenhouse gas emissions will be electricity and heat, transportation, and manufacturing and construction. This projection suggests a consistent linear increase in emissions within these sectors. To address this trend, prioritizing the development and adoption of renewable energy sources is a crucial recommendation. The government should consider implementing robust policies and strategies geared towards a gradual transition away from fossil fuels in energy production. Furthermore, setting precise and measurable targets aligned with global climate objectives becomes imperative to effectively curb greenhouse gas emissions and mitigate their impact on the environment.

R CODE

```
# GHG Time Series Forecasting
library(readxl)
library(ggplot2)

# Load the data
df <- read_excel(file.choose())

####### correlation analysis ###############################################
cor_elec <- cor.test(df$Year, df$Electricity_and_heat, method = "pearson")
cor_transp <- cor.test(df$Year, df$transportation, method = "pearson")
cor_manu <- cor.test(df$Year, df$manufacturing_and_construction, method = "pearson")
cor_agriculture <- cor.test(df$Year, df$agriculture, method = "pearson")
cor_fugitive <- cor.test(df$Year, df$fugitive_emissions, method = "pearson")
cor_buildings <- cor.test(df$Year, df$buildings, method = "pearson")
cor_industry <- cor.test(df$Year, df$industry, method = "pearson")
cor_land <- cor.test(df$Year, df$land_use_change_and_forestry, method = "pearson")
cor_waste <- cor.test(df$Year, df$waste, method = "pearson")

cor_elec # R = 0.99
cor_transp # 0.99
cor_manu # 0.94
cor_agriculture # 0.93
cor_fugitive # 0.89
cor_buildings # 0.95
cor_industry # 0.77
cor_land # 0.47
cor_waste # 0.41

# Every sector has anywhere between a moderate and a positive linear correlation
# with time
###################################################################

####### Time series plot ###################################################################################

ggplot(df, aes(x = Year)) +
  theme_gray() +
  geom_line(aes(y = Electricity_and_heat, color = "Electricity and Heat"), size = 0.8) +
  geom_line(aes(y = transportation, color = "Transportation"), size = 0.8) +
  geom_line(aes(y = manufacturing_and_construction, color = "Manufacturing and Construction"), size = 0.8) +
  geom_line(aes(y = agriculture, color = "Agriculture"), size = 0.8) +
  geom_line(aes(y = fugitive_emissions, color = "Fugitive Emissions"), size = 0.8) +
  geom_line(aes(y = buildings, color = "Buildings"), size = 0.8) +
  geom_line(aes(y = industry, color = "Industry"), size = 0.8) +
  geom_line(aes(y = as.numeric(land_use_change_and_forestry), color = "Land Use Change and Forestry"), size = 0.8) +
  geom_line(aes(y = waste, color = "Waste"), size = 0.8) +
  xlab("Year") +
  ylab("GHG Emissions (Billions of Tons)") +
  ggtitle("Time Series Plot of GHG Emissions") +
  scale_color_manual(values = c("Electricity and Heat" = "blue",
                                "Transportation" = "green",
                                "Manufacturing and Construction" = "red",
                                "Agriculture" = "lightblue",
                                "Fugitive Emissions" = "purple",
                                "Buildings" = "yellow",
                                "Industry" = "pink",
                                "Land Use Change and Forestry" = "chocolate",
                                "Waste" = "magenta"),
                     breaks = c("Electricity and Heat",
                                "Transportation",
                                "Manufacturing and Construction",
                                "Agriculture",
                                "Fugitive Emissions",
                                "Buildings",
                                "Industry",
                                "Land Use Change and Forestry",
                                "Waste")) +
  labs(color = "Sector")  # Setting the legend title

###############################################################################################

# Developing linear trend equations

# electricity and heat sector
electricity_model <- lm(Electricity_and_heat ~ Period, data = df)
summary(electricity_model)

# transportation sector
transportation_model <- lm(transportation ~ Period, data = df)
summary(transportation_model)

# Manufacturing sector
manufacturing_model <- lm(manufacturing_and_construction ~ Period, data = df)
summary(manufacturing_model)

# Agriculture sector
agriculture_model <- lm(agriculture ~ Period, data = df)
summary(agriculture_model)

# Fugitive emissions sector
fugitive_emissions_model <- lm(fugitive_emissions[-1] ~ Period[-1], data = df) #taking out the first period because it's blank in the data
summary(fugitive_emissions_model)

# Buildings sector
buildings_model <- lm(buildings ~ Period, data = df)
summary(buildings_model)
```

```
# Industry sector
industry_model <- lm(industry ~ Period, data = df)
summary(industry_model)

# land_use change and forestry sector
land_use_forestry_model <- lm(`land_use_change_and_forestry` ~ Period, data = df)
summary(land_use_forestry_model)

# Waste sector
waste_model <- lm(waste ~ Period, data = df)
summary(waste_model)


######################### Calculate MAPE for each sector ##############################
calc_MAPE <- function(actual, predicted) {
  MAPE <- mean(abs((actual - predicted) / actual)) * 100
  return(MAPE)
}

# Putting each calculation into a variable
MAPE_electricity <- calc_MAPE(df$Electricity_and_heat, df$electricity_predicted)
MAPE_transportation <- calc_MAPE(df$transportation, df$transportation_predicted)
MAPE_manufacturing <- calc_MAPE(df$manufacturing_and_construction, df$manufacturing_predicted)
MAPE_agriculture <- calc_MAPE(df$agriculture, df$agriculture_predicted)
###### because first value is blank, calculating manually
MAPE_fugitive_emissions <- mean(abs((as.numeric(df$fugitive_emissions[-1]) - as.numeric(df$fugitive_emissions_predicted[-1])) /
 as.numeric(df$fugitive_emissions[-1]))) * 100
MAPE_buildings <- calc_MAPE(df$buildings, df$buildings_predicted)
MAPE_industry <- calc_MAPE(df$industry, df$industry_predicted)
MAPE_land_use_forestry <- calc_MAPE(df$`land_use_change_and_forestry`, df$land_use_forestry_predicted)
MAPE_waste <- calc_MAPE(df$waste, df$waste_predicted)

# Record each MAPE
MAPE_electricity # Result:2.80%
MAPE_transportation # Result: 1.73%
MAPE_manufacturing # Result: 3.27%
MAPE_agriculture # Result: 5.81%
MAPE_fugitive_emissions # Result: 2.27%
MAPE_buildings # Result: 3.22%
MAPE_industry # Result: 11.81%
MAPE_land_use_forestry # Result: 4.09%
MAPE_waste # Result: 10.70%
############################################################################


######## Forecast for each sector
electric_forecast_2025 <- predict(electricity_model, newdata = data.frame(Period = 36)) # period 1 = starting date of 1990 [adding 35 years to it]
transportation_forecast_2025 <- predict(transportation_model, newdata = data.frame(Period = 36))
manufacturing_forecast_2025 <- predict(manufacturing_model, newdata = data.frame(Period = 36))
agriculture_forecast_2025 <- predict(agriculture_model, newdata = data.frame(Period = 36))
fugitive_emissions_forecast_2025 <- predict(fugitive_emissions_model, newdata = data.frame(Period = 1:36)) # because first row is void
buildings_forecast_2025 <- predict(buildings_model, newdata = data.frame(Period = 36))
industry_forecast_2025 <- predict(industry_model, newdata = data.frame(Period = 36))
land_use_forestry_forecast_2025 <- predict(land_use_forestry_model, newdata = data.frame(Period = 36))
waste_forecast_2025 <- predict(waste_model, newdata = data.frame(Period = 36))

electric_forecast_2025 # 18.12 (billion tons of GHG gasses)
transportation_forecast_2025 # 9.18
manufacturing_forecast_2025 # 6.81
agriculture_forecast_2025 # 6.61
fugitive_emissions_forecast_2025 #3.39
buildings_forecast_2025 # 3.38
industry_forecast_2025 # 2.93
land_use_forestry_forecast_2025 # 1.65
waste_forecast_2025 # 1.44

##################################################################################################
```