Analysis of the UK’s greenhouse gas (GHG) emissions from 1990 – 2023 and future predictions

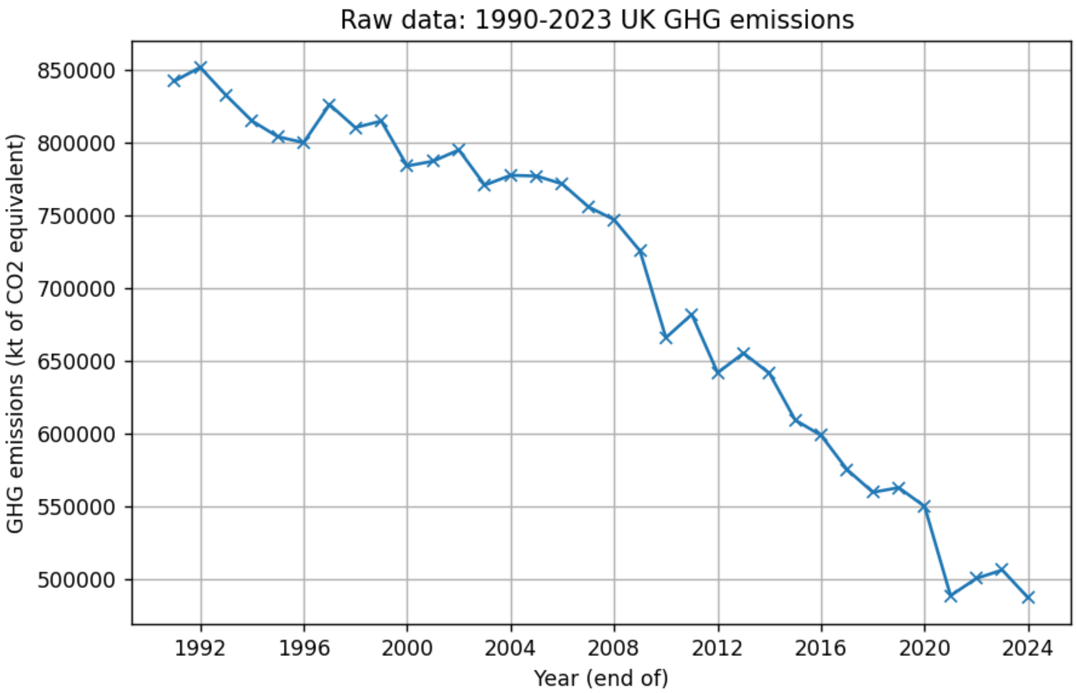
By Shohail Ismail

This report presents an analysis of the United Kingdom’s total greenhouse gas (GHG) emissions from 1990 to 2023 (provisional), gathered through a dataset from the Office For National Statistics (ONS). The dataset is an Excel file consisting of multiple sheets, each detailing the amount of GHG emissions, however for this task only the total yearly GHG values will be used, ignoring the impact of specific industries.

Data processing

The sheet being read (‘GHG total’) had 2 tables, and so only the first one, specifically just the ‘Year’ and ‘Total GHG emission’s (in kT CO_2 equivalent) columns were read. As the dataset was very small, visual inspection showed there to be no missing or invalid data, along with chronological ordering of the ‘Year’ column. Due to this, the only processing that was done was converting the data into appropriate formats for analysis; a notable point is that the ‘Year’ data was taken to mean end-of-year reporting for ease of analysis.

The data was then displayed through a graph to gauge visual trends and anomalies, the former of which there was a year-over-year downward trend and the latter of which there were none. This was part of the exploratory data analysis (EDA), which allowed for a better understanding of the data and the best model to interpret it.



Raw data displayed in graph

Data modelling

After careful consideration, a univariate ARIMA (autoregressive integrated moving average) model was chosen to capture the historical trend and autocorrelation in the given data. This was chosen over traditional machine learning (ML) models owing to the small size of the dataset, linear relationship in the data, and lack of a need for long-term forecasting or considering multivariate data. Additionally, while ML models are too complex, a simpler method such as regression could not be used owing to it being too simple and not being able to capture autocorrelations or subtle patterns in the data. Due to these reasons, the ARIMA model was chosen as the best fir.

To optimise the model's parameters, the 'pmdarima.auto_arima' function was used to automatically select the best-fitting model by evaluating various configurations and minimising the Akaike Information Criterion.

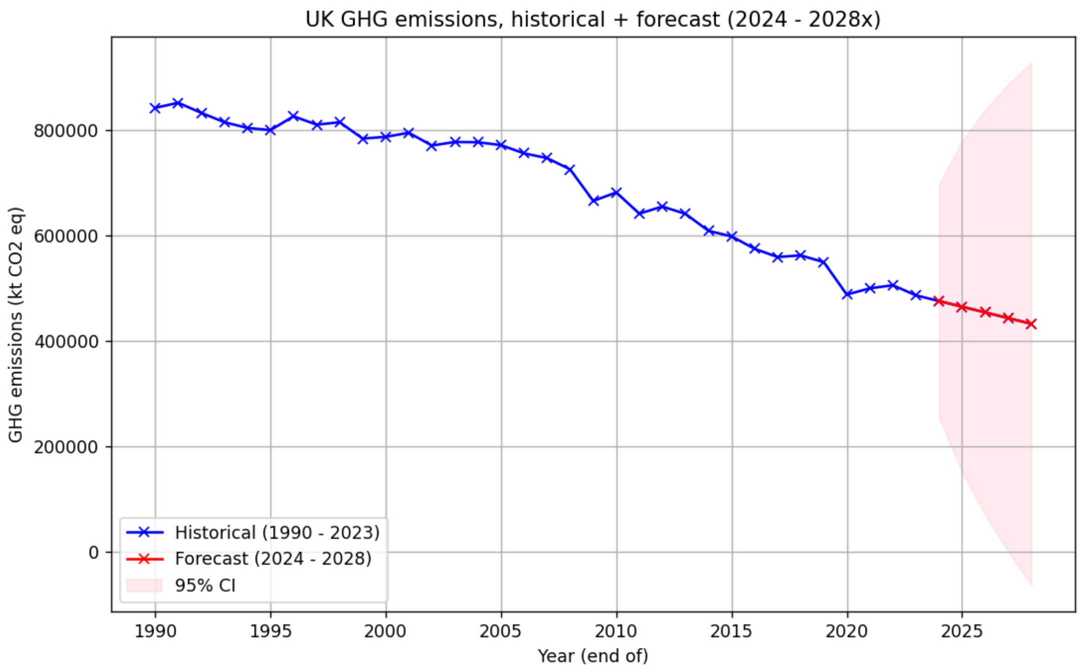
The main assumption made by using this model is that the historical trend will continue into the future without any major fluctuations (either due to policy changes or technological breakthroughs). An assumption was also made regarding the data, specifically that it is stationary when fed into the ARIMA model, which was not the case.

To rectify this, an Augmented Dicky-Fuller (ADF) test was performed on the data to gauge its stationarity , after which first-order differencing was applied to turn the data from a quadratic trend to a linear trend. This was then fed into 'auto_arima' with a linear trend component included during parameter estimation.

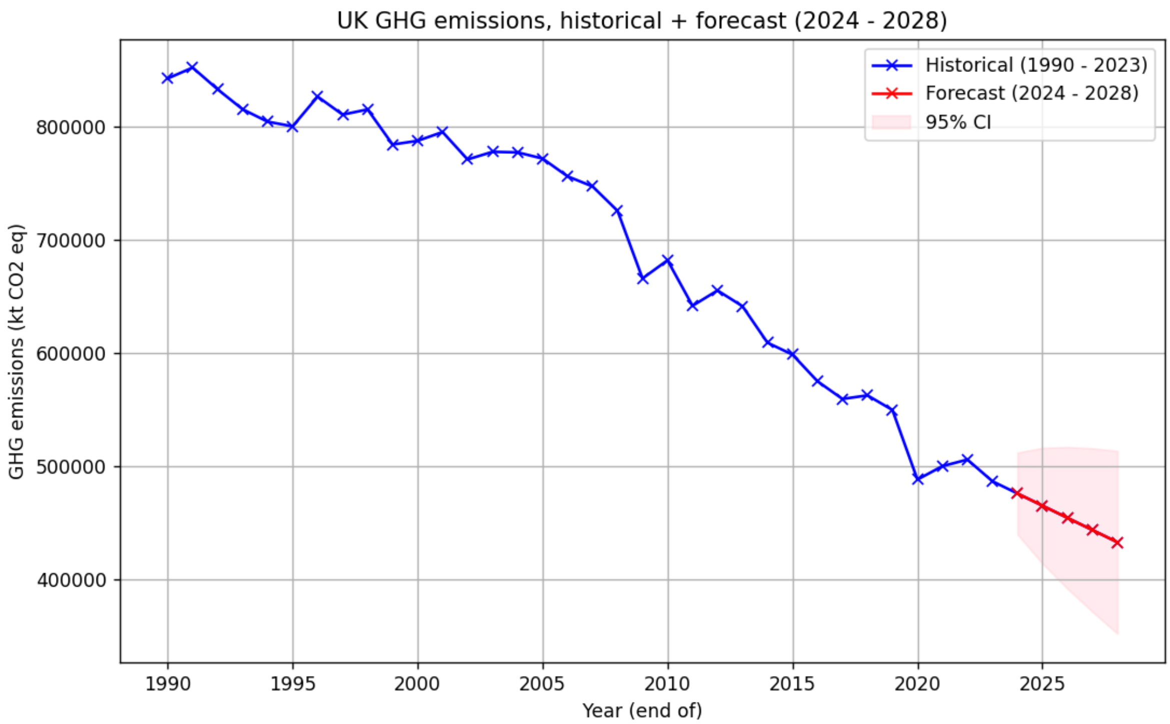
```
########## ADF Test ###########
ADF stat: 0.7903
p-value: 0.9915
Critical values:
   1%: -3.6535
   5%: -2.9572
   10%: -2.6176
Non-stationary (p-value >= 0.05)

Applying first-order differencing...
Checking stationarity of the differenced series:

########## ADF Test ###########
ADF stat: -7.6517
p-value: 0.0000
Critical values:
   1%: -3.6535
   5%: -2.9572
   10%: -2.6176
Stationary: (p-value < 0.05)
Performing stepwise search to minimize aic
 ARIMA(2,1,2)(0,0,0)[0] intercept   : AIC=757.455, Time=0.30 sec
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=750.149, Time=0.03 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=752.830, Time=0.04 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=751.869, Time=0.04 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=756.828, Time=0.02 sec
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=753.394, Time=0.16 sec

Best model:  ARIMA(0,1,0)(0,0,0)[0] intercept
Total fit time: 0.611 seconds
```

Results of ADF test before and after applying differencing

It should be noted that 'auto_arima' does have a function to convert non-stationary data into stationary data before parameter selection, however after using this feature the confidence intervals were large and inaccurate, and so manual differencing was done instead.



Forecasted data with confidence intervals when differencing is handled by 'auto_arima'



Forecasted data with confidence intervals when differencing is handled manually

Results analysis

The UK's annual GHG emissions have shown a general downward trend since 1990, reflecting the move away from coal and into cleaner energy, improved energy efficiency, and environmental regulation. Plotting the forecasted data suggests a reasonable fit with historical data, with no strong patterns left such as clustering. The projection indicates continued gradual decline in total GHG emissions, although the margin of uncertainty increases with each year.

Although emissions are likely to continue declining under current conditions, achieving goals such as net zero by 2050 may require further measures to maintain/accelerate reductions, such as considering specific industries to curb CO_2 emissions of.

Model limitations

One limitation of this model is its reliance on a small dataset, which restricts the quality of predictions and contributes to challenges during the parameter-fitting stage, as reflected through displayed warnings. Furthermore, there is no elaboration into the specific industries and types of GHG being produced, which would lead to more actionable insights and help shape policy changes to lead to continued decrease in GHG emission. This would most likely involve switching to a machine learning model to handle the complexity of the data.

In conclusion, while the model is sufficient for a rough estimate of total GHG emissions in the UK for the next 5 years, it would benefit from a larger dataset (i.e., more granular data such as monthly observations) and the inclusion of the influence of external factors such as economic growth, technological advancements, and global events.