# CAMBIUM DATA SET AND PROHET MODEL ESTIMATION

## Cambium Data Set :

<https://data.nrel.gov/submissions/183>

<https://data.nrel.gov/submissions/230>

## Organize the cambium data

1. If we consider CO2 emission we have both pre combustion and direct combustion
2. Sum up the pre combustion and direct combustion for all 8 cases ( Mid-case , Low Renewable Energy Costs, High Renewable Energy Costs, Low Natural Gas Prices, High Natural Gas Prices, High Demand Growth, 95% Decarbonization by 2050, 100% Decarbonization by 2035 )

i.e similar to that to combustion + pre combustion = combined in the given example

A table with numbers and a yellow line

Description automatically generated

1. Once data is combined – next step is to segregate all the 8 cases data, because the prophet model is trained for each use case for all the 3 emissions (CO2,N2O,CH4)

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. The final step to tune the data to feed to the Prophet Model is to calculate the Harmonic mean of emissions for each month

A math equation with black text

Description automatically generated with medium confidence

To calculate the Harmonic mean –

Ther’s a python code to calculate the harmonic mean In folder we code – please find the file “HarmonicMean.ipynb”

Note : Make sure the input\_dir and output\_dir paths are correctly placed

A screenshot of a computer

Description automatically generated

After Calculating the harmonic mean the data looks like below

A screenshot of a computer

Description automatically generated

A screenshot of a spreadsheet

Description automatically generated

## Prophet Model

Input to the Prophet Model is the Harmonic Mean of ease scenario of all 3 emissions, once the model is trained with the years 2025, 2030, 2035, 2040, 2045, 2050 . The Model estimates the data from all the years between 2025 to 2050 based on the code configurations

(Note: Make sure the paths of input\_dir and output\_file\_path are as per required)

A screenshot of a computer

Description automatically generated

Output :

The Model is trained and estimated for all the years between 2025 to 2050

All the forecasted results are stored in an excel named “forecasted\_results.xlsx”

Also the code captures all the performance matrix like MAPE, MSE, R2 value etc..

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Visualization

To visualize the data there are two python codes

1. To generate graphs which shows emission data of California, Georgia, Washington, New York for each state is “graphs\_allyears.ipynb”

(Note : Make sure the filepaths are accurate)

A screenshot of a computer

Description automatically generated

Output :

A graph of a graph showing the amount of emission

Description automatically generated with medium confidence 

1. To populate graphs for single year for example 2025 (Jan to Dec) which shows emission data of California, Georgia, Washington, New York for each state is “graphs\_singleyear.ipynb”

(Note : Make sure the filepaths are accurate)

A screenshot of a computer

Description automatically generated

A graph with different colored lines

Description automatically generated