Global Carbon Fluxes Associated with Livestock Feed and Emissions

Bhavya Vulchi - MH09851
Sindura Reddy Challa - OE37375
Yeswanth Sai Yerramasu - JG97937

Problem statement

- Given global annual carbon flux estimates related to livestock activities, this project aims to investigate the correlation between land use changes and carbon dynamics from 2000 to 2012. By mapping carbon flux data at a 0.05-degree resolution alongside land use changes, we look to uncover the relationship between human activities, land use alterations, and carbon emissions or sequestrations. Insights derived from this analysis will inform sustainable land management practices crucial for mitigating climate change impacts.

- Furthermore, recognizing the challenges of data collection in politically unstable or technically limited regions, this project will employ spatial interpolation and modeling techniques. By extrapolating from available global carbon flux data, we aim to estimate emissions in data-scarce areas, contributing to a more comprehensive understanding of global carbon dynamics and supporting international environmental oversight efforts.

Data Understanding

1. The dataset is taken from earthdata.nasa.gov (Carbon Monitoring System)

2. The data provides annual carbon flux estimates for global livestock activities, aggregated across all livestock types. These estimates are presented in two ways:
   - Average emissions per square meter (g C/m2/yr), calculated by dividing carbon fluxes by the total surface area of each grid cell. These values are not additive across grid cells.
   - Total annual carbon fluxes (Mg C/yr) summed over all livestock types for each grid cell. The total surface area of grid cells is provided.

3. The dataset consists of 56 comma-separated (.csv) files organized by year and divided into categories for solids and gases. Each category contains files for average carbon fluxes per unit area and annual carbon fluxes summed over all types. These files are grouped into four sets of 14 annual data files.

Data Understanding (contd.)

- Format of csv files: **lvst_solids_gCm2_YYYY**, **lvst_gases_gCm2_YYYY**,**lvst_solids_MgC_YYYY**,**lvst_gases_mgC_YYYY**, where YYYY range from 2000 to 2013 but for this project ,we used files in the format **lvst_gases_mgC_YYYY**

- Provides global estimates of livestock-related carbon emissions.
- Includes methane (CH_4) from digestion and manure, and carbon dioxide (CO_2) from respiration and manure decomposition.
- Data organized in grid cells for spatial analysis.
- Measurements in megagrams of carbon per year (Mg C/yr).

where CH_4 emissions from digestion and manure.
CO_2 emissions from respiration and manure.
Standard deviation provided for data reliability.

Tools and Technologies

Data Storage and Sharing:

- Local storage on device memory and cloud sharing via Box platform.
- Datasets could be accessible directly from Box using URLs within Databricks.

Computational Framework:

- Apache Spark and Hadoop installed locally for initial execution.
- Databricks utilized as the primary platform for project execution.

Data Handling and Analysis:

- Pyspark employed as the primary programming language.
- R used for parts of Exploratory Data Analysis (EDA) to efficiently plot geospatial data.

Tools and Technologies (contd.)

Data Workflow:

- Data downloaded from NASA's website. earthdata.nasa.gov
- Initial analysis performed on smaller data chunks to expedite error identification.
- Meaningful data extracted during cleaning phase.

Analytical Techniques:

- **Spatial Analysis**: Utilized latitude and longitude coordinates for spatial insights.
- **Trend Analysis**: Analyzed emissions with respect to corresponding years to identify trends over time.

Future Development:

- Plan to develop a predictive model to monitor high emission levels and forecast future variations both using data and graphs.

Data preparation

- The data set **lvst_gases_mgC_YYYY** contains the following columns:

Attribute	Description	Units
longitude	Center longitude of 0.05°x0.05° grid cell	Decimal degrees
latitude	Center latitude of 0.05°x0.05° grid cell	Decimal degrees
lvstk_km2	Area assigned to livestock per grid cell	km^2
efCH4_MgC	Enteric fermentation CH4-C emissions	Mg C yr-1
efCH4_SD_MgC	Standard deviation of enteric fermentation CH4-C emissions	Mg C yr-1
mmCH4_MgC	Manure management CH4-C emissions	Mg C yr-1
mmCH4_SD_MgC	Standard deviation of manure management CH4-C emissions	Mg C yr-1
totCH4_MgC	Total CH4-C emissions (manure + fermentation)	Mg C yr-1
totCH4_SD_MgC	Standard deviation of total CH4-C emissions	Mg C yr-1
rspCO2_MgC	Livestock respiration CO2-C emissions	Mg C yr-1
rspCO2_SD_MgC	Standard deviation of respiration CO2-C emissions	Mg C yr-1
mmCO2_MgC	Manure management CO2-C emissions	Mg C yr-1
mmCO2_SD_MgC	Standard deviation of manure management CO2-C emissions	Mg C yr-1
totCO2_MgC	Total CO2-C emissions (manure management + respiration)	Mg C yr-1
totCO2_SD_MgC	Standard deviation of total CO2-C emissions	Mg C yr-1

Data preparation (contd.)

During the data cleaning process, we focused on refining the dataset to ensure its quality and relevance for analysis.

- 1. **Importing Data**: Initially, we imported the dataset into our environment, selecting only the necessary columns essential for our analysis. These columns include latitude, longitude, and the total methane (CH4) emissions in megagrams of carbon (MgC).

- 2. **Filtering Zero Emission Values**: Recognizing that zero values for methane emissions may indicate missing or irrelevant data, we proceeded to remove all rows where the total methane emissions (totCH4_MgC) value equaled zero. This step ensures that our analysis focuses solely on regions with measurable methane emissions, enhancing the accuracy of our findings and interpretations.

By implementing these cleaning procedures, we refined the dataset to contain only the pertinent information required for our analysis, thereby facilitating more robust and insightful investigations into global methane emissions associated with livestock activities.

Exploratory Data Analysis

- The following results were observed while working with one year's data

▸ (2) Spark Jobs

▸ non_zero_totCH4_MgC: pyspark.sql.dataframe.

non_zero_totCH4_MgC
1027563

▸ (2) Spark Jobs

▸ zero_columns: pyspark.sql.dataframe.DataFrame = [longitude: lon

longitude	latitude	totCH4_MgC
0	0	5111215

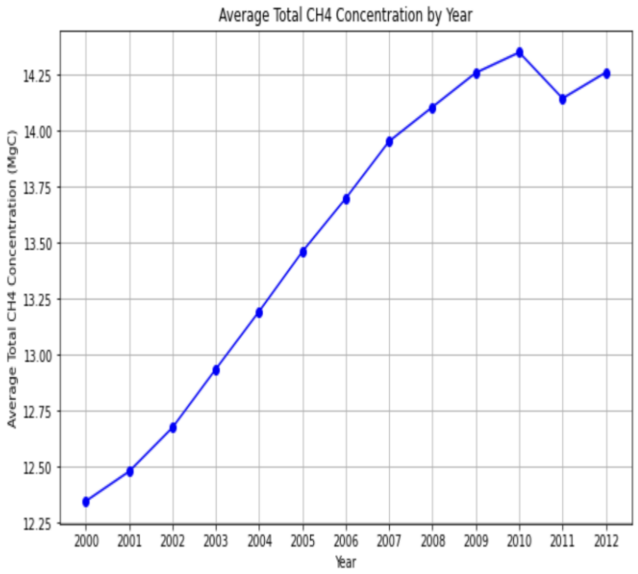
```
Minimum value: 0.0
Maximum value: 2618.74609375
```

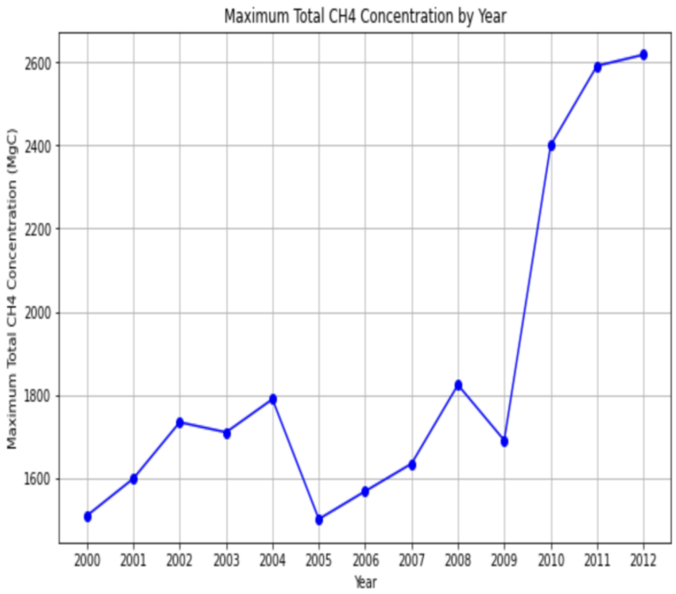
Exploratory Data Analysis (contd.)

- Calculated the following values for every year and stored the results in a list. The data in the list was then parsed into a csv file in DBFS for further analysis.

```
Year: 2012
Number of rows with zero total CH4: 5111215
Number of rows with non-zero total CH4: 1027563
Maximum total CH4 value: 2618.74609375
Average total CH4 value: 14.26200687847394
Latitude corresponding to max value: 40.36185836791992
Longitude corresponding to max value: -86.57499694824219
==================================================
Results:
(2000, 34.811859130859375, -77.7750015258789, 1508.7110595703125, 12.347304100479397)
(2001, 34.811859130859375, -77.92499542236328, 1597.8809814453125, 12.479837051489236)
(2002, 34.561859130859375, -78.7750015258789, 1733.541015625, 12.675741673489803)
(2003, 34.811859130859375, -78.875, 1709.383056640625, 12.937029309061653)
(2004, 34.61185836791992, -78.7750015258789, 1789.0989990234375, 13.194954877207048)
(2005, 34.76185989379883, -78.7249984741211, 1500.3609619140625, 13.462445854783773)
(2006, 34.76185989379883, -78.7249984741211, 1567.4300537109375, 13.700029112761252)
(2007, 34.76185989379883, -78.7249984741211, 1633.7220458984375, 13.954655166427)
(2008, 34.561859130859375, -78.7750015258789, 1823.5699462890625, 14.107779113902934)
(2009, 34.76185989379883, -78.7249984741211, 1690.06298828125, 14.260533156026694)
(2010, 40.36185836791992, -86.375, 2400.509033203125, 14.3513505770897)
(2011, 40.2118606567382, -86.4749984741211, 2591.85400390625, 14.146149881729231)
```

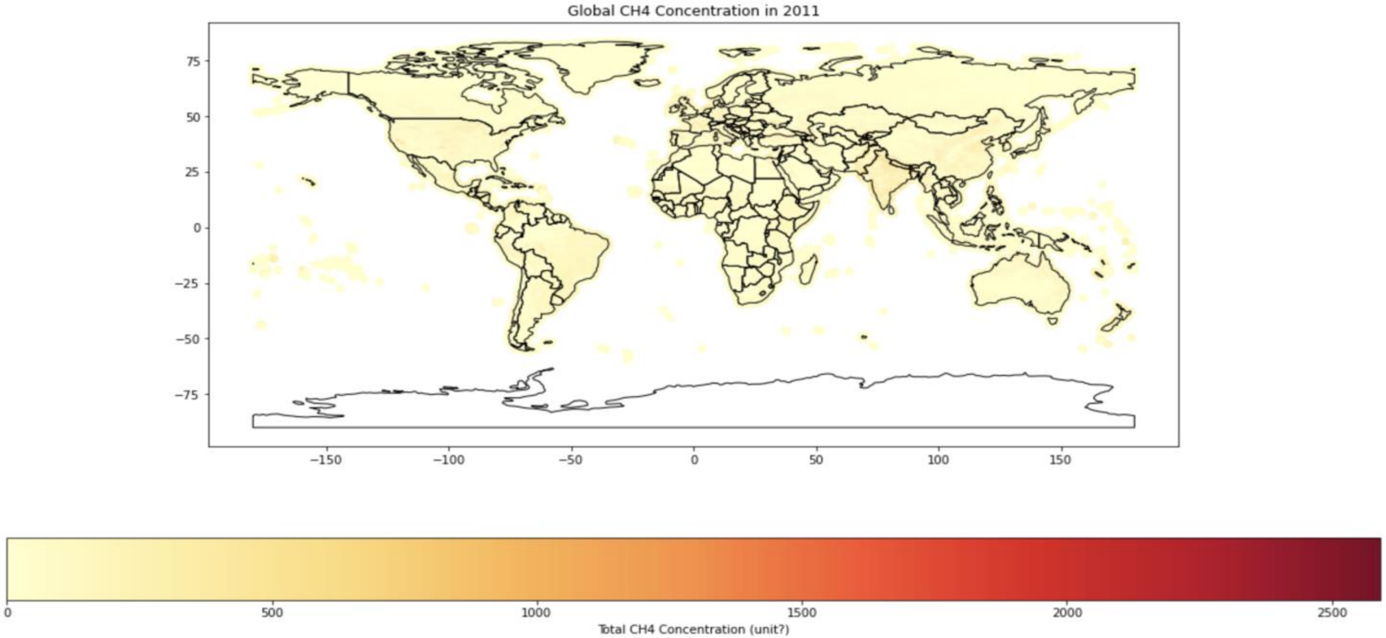
EDA(Contd.)





The above graphs show average and total CH4 concentration over the years.

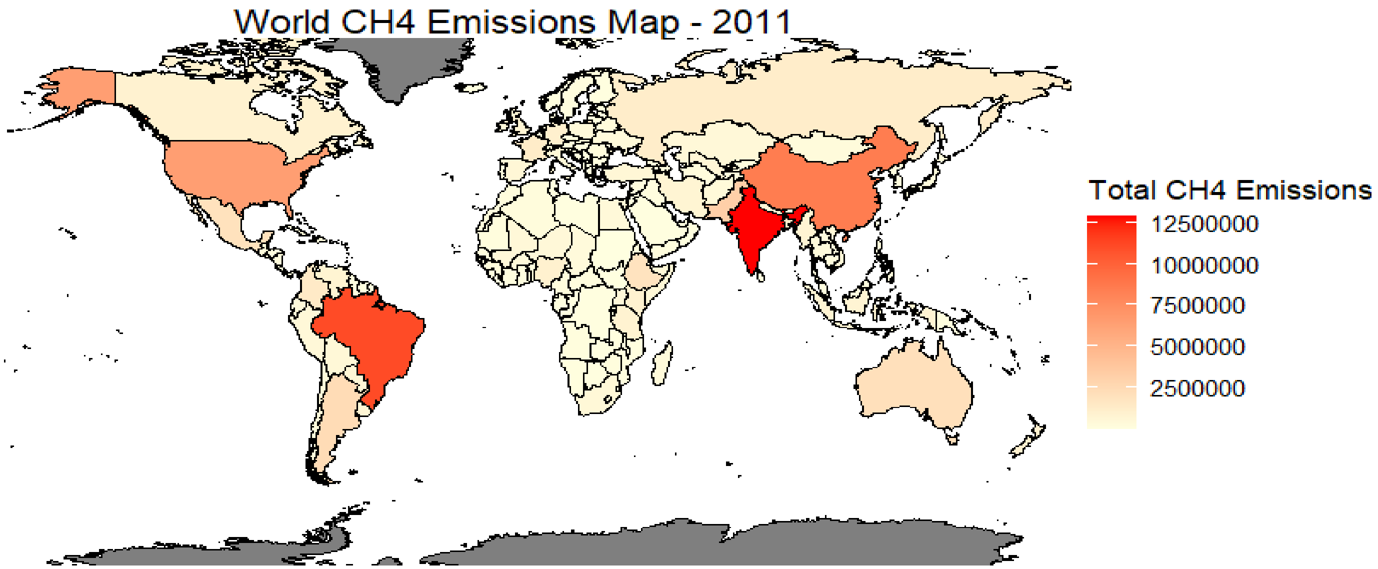
EDA (Contd.)



Global CH4 Concentration in 2011
75
50
25
0
−25
−50
−75
−150
−100
−50
0
50
100
150
0
500
1000
1500
2000
2500
Total CH4 Concentration (unit?)

EDA(Contd.)

- We then modified the CSV, to label the country name with corresponding latitude and longitude values and then calculated the total Methane Emission by every country.
- We then generated maps using R for better map images for use in ML modelling.



Modelling

- We then Built a Liner regression model to predict the average emission of CH4 for every year using the CSV file created during EDA.

▸ (7) Spark Jobs

▾ (1) MLflow run

Logged 1 run to an experiment in MLflow. Learn more

► Loading the widget is taking longer than expected. We suggest the following...

► Loading the widget is taking longer than expected. We suggest the following...

```
Coefficients:  [0.18027794374500436]
Intercept: -348.10833847917746
```

1

▸ (3) Spark Jobs

▸ new_data: pyspark.sql.dataframe.DataFrame

▸ predictions: pyspark.sql.dataframe.DataFrame

```
+----------------+
|      prediction|
+----------------+
|14.7911622795163|
+----------------+
```

Evaluation

- We evaluate our model(s) by observing the actual value for the year 2013.
- Every approach in EDA has been performed with the assumption that we not have access to 2013 data
- When we generate Key Values/ Maps predicting the Methane emissions using values from the previous years, we directly compare them to actual values by generating 2013 values
- The actual and predicted average value for 2013:

```
+------------------+
|Average_totCH4_MgC|
+------------------+
|14.319763825145744|
+------------------+
```

Next Works: Predicting Geographic Map Images

- Complexity and data Requirements:

Understanding and predicting changes in geographic maps through ML involves analyzing complex spatial patterns and environmental influences.

Traditional ML models are less suitable; computer vision techniques are recommended for handling raw image data.

- Potential Methodologies:

Convolutional Neural Networks (CNNs): Effective for processing spatial data and identifying temporal patterns in map images.

Next Works(Contd.)

- **RNNs with CNN Features**: Combining Recurrent Neural Networks (e.g., LSTM, GRU) with CNNs can enhance predictions by analyzing time series image data.

- **Generative Models (GANs)**: Utilizing Generative Adversarial Networks to create new map images from learned patterns, though requiring extensive tuning.

By leveraging these advanced techniques, future projects can potentially generate accurate predictions of geographic maps for years not covered by existing data

Conclusion

- This big data project, utilizing the NASA dataset on carbon flux emissions, provided crucial insights into the global patterns of methane (CH4) emissions. By focusing on the correlation between specific latitudes and longitudes and total CH4 emissions, we have highlighted significant variations in methane release across different regions and years.

- Although this analysis is confined to data up until 2013 and solely focuses on methane emissions, the methodologies applied here can be expanded to include a wider array of greenhouse gases such as CO2 from various sources.

Conclusion

- The fluctuating political landscape worldwide underscores the necessity for governments to adapt and evolve their environmental policies effectively

- By understanding the geographic and temporal patterns of these emissions, each country, and potentially each state within, can develop tailored approaches that consider both environmental and socio-economic factors, aiming for a more sustainable and politically feasible path forward.

- By leveraging big data and analytics, stakeholders can better understand environmental challenges and craft strategies that address the urgent need for sustainability in our global response to climate change.

References

- Wolf, J., G. Asrar, and T.O. West. 2017. CMS: Global Carbon Fluxes Associated with Livestock Feed and Emissions, 2000-2013. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1329
- https://daac.ornl.gov/CMS/guides/CMS_Global_Livestock_CH4_CO2.html