# Prediction of nitrous oxides emissions for remote sensed images and a prognosis of emissions data in agriculture and forestry

*A thesis submitted in partial fulfilment of the requirements for the*

*Degree of Master of Philosophy in Computer Science*

*By*

*Mohamed Ayoob*

*Supervised by*

*Dr Hafeez Ullah Amin*



*School of Computer Science*

*University of Nottingham Malaysia*

# Abstract

Land use activities such as agriculture and forestry have a significant impact on greenhouse gas emissions, but these impacts are often underestimated. This is because agriculture and forestry sectors cover a large land area, making on-ground surveys difficult to scale, expensive, and time-consuming. While there have been localized examinations of urban and industrial domains, it remains unclear how agriculture and forestry sectors would fare in a global examination using modern tools such as hyperspectral cameras.

Hyperspectral cameras can substantially bridge the gap between uncertainties in observations and offer little room for disparities in measured readings between countries, resulting in an unbiased survey. However, there are spatial (resolution) and temporal (date availability for long dates) gaps in hyperspectral satellite data. These gaps can be filled using machine learning algorithms.

The algorithm utilised paired data from the Sentinel 2 (RGB) images and the corresponding nitrous oxide reading from the Sentinel 5P satellite to obtain seasonal reading levels in 24 different locations spanning 8 different soil varieties. The data was analysed to obtain valuable insights on the variances between readings for a range of time between September 2018 to August 2022 (4 years). This data can be used to develop more effective policies to reduce emissions and mitigate climate change.

A Spatio Temporal Neural network was created that can ingress an RGB image and its associated date and obtain the associated nitrous oxide reading accurate within 12% (Mean Absolute Percentage Error) of the actual value. This model can be utilized as a prototype to fill in gaps in existing hyperspectral images with reasonable compromise in accuracy.

# Acknowledgements

# Table of Contents

# List of Figures

# Chapter 01: Introduction

## Preamble

*"Man did not weave the web of life, he is merely a strand in it. Whatever he does to the web, he does to himself"*

- Excerpts from a letter written by the Seattle Chief (accurately known as Seathl) to the President of Washington, 1855.

Humans have been existing on this planet for a small proportion of its existence. While we humans have a great flair for curiosity and unmatched problem-solving skills, we also possess a great capacity for ruin and destruction. During the past two centuries our efforts to achieve supremacy over each other politically, economically or militarily have set the wheels in motion towards a future that humankind would find neither pleasant nor conducive to conduct our life in the only home that we currently know of, Earth.

The excerpt above was part of a letter written by the Seattle Chief leader to the President of Washington, Pierce, in 1855. It was in response to Pierce's offer to buy the lands in Seattle. The Seattle Chief passionately describes the symbiotic relationship humans have had and should continue to have with nature.

## 1.1 Greenhouse Gases

In the sight of discussing about greenhouse gases, it is apt to have an explanation about atmospheres. The atmosphere can be defined as a gravitationally confined pool of elements that are volatile enough to exist in its gaseous state, above a celestial object at local temperatures (Pepin, 2006). Most celestial objects have atmospheres. As previously believed it is not just planetary bodies and stars (Pepin, 2006) that contain an atmosphere. Modern astronomy divulges that even extreme objects in the universe such as neutron stars (Ho and Heinke, 2009) and black holes (Jacobson and Nguyen, 2021) contain an atmosphere. Earth too contains an atmosphere that comprises of mainly Nitrogen ($N_2$), and Oxygen ($O_2$). Both of which compose nearly 99% of the Earth atmosphere.

Earth's atmosphere also consists of certain gases (less in proportions than $N_2$ and $O_2$), some of which can be considered greenhouse gases. Instances of such gases are carbon dioxide ($CO_2$), methane ($CH_4$), nitrogen dioxide ($NO_2$), ozone and sulphur dioxide ($SO_2$). These gases are both naturally occurring and can be emitted due to anthropogenic (synthetic) activities. Apart from these, there are Fluorinated gases such as hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), sulfur hexafluoride (SF6) and nitrogen trifluoride (NF3) which are emitted almost entirely due to anthropogenic activities such as air conditioning and refrigerating (Sovacool et al., 2021). Nitrogen and Oxygen, the most abundant gases in the atmosphere, are in fact small contributing factors to the greenhouse effect. $N_2$ and $O_2$ are however more potent as greenhouse gases in dry atmospheric conditions such as over the Antarctic continent, realistically though in non-ideal situations we can assume that $N_2$ and $O_2$ have only a fraction of the potency as other greenhouse gases. (Höpfner et al., 2012). For practical purposes, the most concerning gases are methane and nitrous oxide. It is estimated that they have nearly 25 times and

300 times the potency (as greenhouse gases) greater than carbon dioxide over 100 years ("AR4 Climate Change 2007," 2022).

A closely related artefact is Atmospheric Aerosols. Atmospheric Aerosols are fine particulate matter in the atmosphere emitted due to anthropogenic activities. Such gases and aerosols, if emitted in increasing proportions contribute disproportionately (to their compositions in the atmosphere) to the greenhouse effect and henceforth global warming (Sovacool et al., 2021) (Charlson et al., 1992) (Ramanathan et al., 2001).

## 1.2 Global warming and its adverse anthropocentric effects

Global warming is a causal event that arises due to greenhouse gas emissions. This is possible due to the greenhouse effect. The greenhouse effect is the phenomenon that recirculates longwave radiation within the atmosphere. The greenhouse effect as opposed to popular belief is a natural effect that helps to foster life. If the greenhouse effect were not to naturally occur, life on Earth would be difficult to imagine. (Mitchell, 1989). Global warming which arises due to the exacerbated effects of the greenhouse effect (due to increased emissions by anthropogenic activities) causes a plethora of adverse effects that leave, and will continue to leave, humanity reeling in its aftermath.

Global warming has positive correlations with several clinical impediments. Cancer has some of the highest fatality rates in most developed countries. Overall cancer is the cause for 26 % of the deaths globally. Cancer is particularly lethal in low-income countries, causing more than 50% of the deaths in such countries (Dagenais et al., 2020). A recent study shows that climate change, exacerbates the effects and symptoms of cancer, increases the likelihood of developing cancer and even disrupts the treatments to cancer patients. Further, the authors of the study point out various health benefits that would manifest as a result of mitigation of climate change (Nogueira et al., 2020).

Cardiovascular diseases are another impediment exacerbated by climate change, they manifest a variety of symptoms that affect people of all ages, genders, and occupations. According to a study done in America from 1999 to 2018, it has been estimated that one person dies of a cardiovascular disease, every 36 seconds ("Underlying Cause of Death, 1999-2020 Request," 2022). Furthermore, it has been estimated that 1of 4 deaths in the US is cardiovascular disease related ("Heart Disease and Stroke Statistics—2021 Update | Circulation," 2022). Studies have been conducted globally on the influence of climate change and rising air temperatures on people's health, particularly those with heart disease. Some of these researches found a direct and positive correlation of rising air temperature on cardiovascular health issues (Tsaloglidou et al., 2018), while others found this effect to be minor (Kakaei et al., 2021). Studies conducted in the middle east among blue-collar workers depict that high temperatures are in fact a strong contributing factor inducing cardiac arrests (Pradhan et al., 2019).

Data from the National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (NCHS) depicts that in 2019 more than 19% of US citizens consulted mental health professionals for treatments. 15% had taken prescribed medication for mental health anomalies. A comprehensive report by the United States Climate and Health Alliance on the psychological impacts of climate change outlines several psychological pressure-points due to climate change and an increase in natural disasters. Among them are trauma, shock, stress, anxiety, depression, grief, severe reactions (such as PTSD), and strains on social relationships. These pressure points further devolve into substance abuse, chronic depression, mental health emergencies, helplessness, resignation, loss of

personal identity and so on. Such individual traits often manifest as negative elements communally, leading to a lack of social cohesion, violence, and aggression ("Beyond Storms & Droughts," 2022). Other studies have noted proportional increments in domestic violence and assaults with increasing temperatures (Raj, 2022), and serious mental impairments that may even lead to a suicide epidemic (Berry et al., 2010).

To make matters bleak, the above was just an anthropocentric recollection of adverse effects. There are further meteorological effects such as the El Nino, La Nina, urban smog pollution (Xie et al., 2019); further ecological effects such as destruction of biota, marine pollution, such are the chain effects of climate change. In an almost clairvoyant fashion, the words of the Seattle Chief can be reiterated. Man is merely a strand in the web of life, indeed whatever he does to the web, he does to himself.

## 1.3 Sources and sinks of greenhouse gas emissions

Several anthropogenic activities contribute to greenhouse gas emissions. The Fifth Assessment Report (AR5) submitted by the Intergovernmental Panel on Climate Change (IPCC) in 2014, broadly categorises the sources of greenhouse gas emissions into classes such as Energy Systems, Transport, Buildings, Industry, Agriculture, Forestry and Other Land Use (AFOLU) and Human Settlements, Infrastructure, and Spatial Planning. Globally, the most concerning classes are Energy systems (electricity and heat production) (25%), transport (14%), buildings (6.4%) and agriculture, forestry and other land use (24%). ("AR5 Climate Change 2014," 2022)

An important concept, when discussing the sources of greenhouse gas are, sinks of greenhouse gases. Sinks are opposite to sources. Sources emit gases, while sinks absorb the greenhouse gases. Better management of resources with sound principles can even change the sources into sinks, for instance in the United States, since 1990, the agriculture sector has been acting as a sink rather than a source (Domke et al., 2020).

## 1.4 Research Problem

This research work aims to undertake an examination of literature on historical time-lapse of the most concerning greenhouse gas, nitrous oxide ("AR4 Climate Change 2007," 2022, p. 4) in the agricultural and forestry sectors. After which, a comparative analysis of a variety of soil varieties and agricultural practices would be considered and compared. An analysis of seasonal variations in emissions based on localized agricultural and forestry practices would be conducted. A comparison between the emission rates before COVID-19 lockdowns and during COVID-19 lockdowns would also be carried out. Work demonstrated by (Goldberg et al., 2020) showed significant differences in emissions in big cities. We would however be investigating the emissions in agriculture and forestry industries. The particular emphasis on agriculture and forestry is due to the following reasons:

1) impacts of land use (agriculture, forestry) are in fact underestimated (Mahowald et al., 2017)
2) agriculture and forestry sectors cover a large land area which makes any on ground surveys difficult to scale; on-ground surveys are also additionally expensive and time consuming; while there have been localized examinations of urban and industrial domains (Zheng et al., 2019) it remains unclear how agriculture and forestry sectors would fare in an examination using

modern tools (such as hyperspectral cameras) specialized to detect greenhouse gases in the atmosphere (Rolnick et al., 2019) in a global level

3) Remote sensing hyperspectral cameras substantially bridge the gap between uncertainties in observations and offer little room for disparities in measured readings between countries, resulting in an unbiased survey (Weiss et al., 2020). The "weekend effect" (Beirle et al., 2003) of trace gases in the domain of agriculture and forestry will also have to be investigated it remains unclear how the emissions reading would fare on a weekly basis.

The main purpose of this examination is to determine high priority locations (in assorted soil varieties) both spatially and temporally in agriculture and forestry domains that emit greenhouse gases using hyperspectral cameras. Our contribution consists of the following: (1) a dataset, on emissions of assorted soil varieties to conduct further research on remotely sensed images – and (2) an analysis of the dataset to determine GHG emissions on the domains of agriculture and forestry to quantify the observed data. As there has been increasing literature on utilizing deep learning algorithms for earth observation and remote sensing (Cheng et al., 2020), there is strong indication that the above dataset would be useful for researchers to conduct further study in the future (Persello et al., 2021).

Reduction of nitrous oxide emissions is a key element in mitigating the effects of global warming. Over 70% of nitrous oxide emissions come from agriculture and forestry (Ritchie et al., 2020). Insights from this examination would be a paramount benefit for many stakeholders and be an effort to minimize effects of climate change.

This examination also fills an important gap (Oertel et al., 2016) of determining emission rates on various soil varieties using a standardised tool and puts forward a methodical examination of nitrous oxide emissions from agriculture and forestry as an industry through the lens of an edaphic acumen.

## 1.5 Research Gap

Based on the above problem statement, the research gap can be elicited as follows. There is spatial (resolution) and temporal (date availability for long dates) gaps in hyperspectral satellite data, and such gaps can be filled using machine learning algorithms (Rolnick et al., 2022). We choose the domains of agriculture and forestry due to the 3 reasons mentioned in the problem statement.

Rather than stochastically determining locations to download images (to fill the gaps using machine learning algorithms) we decided to conduct an edaphic (pertaining to soil) examination and diversify our data collection methodology to cover several soil varieties. In addition to collecting data, this would also enable us to examine and set a cursory baseline for emission of various soil varieties due to various assorted agriculture and forestry practices. Such an examination additionally fills a gap of determining emission rates on various soil varieties using a standardised tool (Oertel et al., 2016).

## 1.6 European Space Agency (ESA) Copernicus Sentinel 2 and 5P

Several satellites have been launched to observe the Earth. Such satellites are called Earth observation satellites. Two main satellite families that observe the Earth using hyperspectral cameras are the LandSat 8 ("Landsat 8 | Landsat Science," 2021) and 9 ("Landsat 9 Spectral Specifications | Landsat Science," 2021), and Sentinel 2 and 5P. Sentinel 5P is the first Copernicus mission to be solely

dedicated to monitoring atmospheric composition of greenhouse gases. The Tropospheric Monitoring (TROPOMI) instrument onboard can estimate geolocated columns of various greenhouse gases including methane and nitrous oxide ("Products and Algorithms - Sentinel-5P Technical Guide - Sentinel Online - Sentinel Online," 2022). This makes this ideal for monitoring and tracking such greenhouse gases. The only drawback of using the Sentinel-5P data would be the comparably small magnitude of data available. The satellite was launched in October 2017 and has open data available to use from 10th July 2018.

## 1.7 Research Goal

This study's major goal is to use remote sensing to first identify and understand soil level variations in agriculture and forestry in the available timeframes of data (2018–2022). This would then be utilized in the next phase of the research, which utilizes the data to predict nitrous oxide emissions for a given high-res image thereby providing a viability assessment for filling the gaps in hyperspectral data. The exact goals and research objectives that go along with them are as follows:

**Objective 1**: To identify the differences in emissions rate of nitrous oxide in various soil levels from the Sentinel 5P data.

    a.   What are the differences in emissions of nitrous oxide in various soil levels?

    b.   How would they fare across 4 years of data?

**Objective 2**: To use the soil level data (along with the dates) with their corresponding images from the Sentinel-2 data to build and test a deep learning model.

    a.   Will deep learning model will be able to discern variances in the images to be able to predict the nitrous oxide emission rates for the area in the image?

    b.   How will they evaluate on intrinsic evaluation metrics?

## 1.8 Contributions

The first in-depth examination of monitoring nitrous oxide emissions and utilising machine learning and high-resolution satellite data to estimate the gas was made possible by this self-proposed study. In conclusion, the project's contributions included the following:

- An in-depth analysis of Nitrous oxide emissions data in and around agricultural and forestry areas (24 locations were surveyed) and the dataset for the analysis obtained. These locations spanned across the globe in an unbiased distribution of locations between the global north and the global south. A through data analysis will be conducted from the year 2018 to 2022 on various soil varieties.

- Utilizing the nitrous oxide emission dataset in conjunction with high-resolution satellite image data to obtain a fairly accurate machine learning model that is able to fill spatio-temporal gaps

at a workable accuracy or obtain the viability of utilizing a machine learning algorithm to test whether the aforementioned research gap mentioned (Rolnick et al., 2022) is viable in practise.

## 1.9 Project Scope

The project scope of this research would encompass the following aspects:

1. Data Collection: Acquire hyperspectral satellite imagery covering agricultural and forestry regions across a range of geographical locations and time periods.
2. Data Preprocessing: Process and prepare the hyperspectral imagery to ensure its quality and suitability for analysis.
3. Data Analysis: Analyse the collected data, see potential trends, based on seasons and soil varieties.
4. Machine Learning Model Development: Develop and train machine learning models capable of estimating greenhouse gas emissions from hyperspectral imagery. This involves developing neural networks.
5. Model Validation: Evaluate the performance of the developed machine learning models using ground-truth data or other reliable sources. This will assess the accuracy and reliability of the models in predicting greenhouse gas emissions.

Things which are out of scope for this project would be:

1. Spatial and Temporal Gap Filling: Utilize the trained machine learning models to fill in the spatial and temporal gaps present in hyperspectral satellite data. This will provide a more comprehensive and continuous representation of greenhouse gas emissions. This requires specialised access to data and high computing resources.
2. Policy Recommendations: Based on the findings, develop recommendations for policymakers on how to effectively reduce greenhouse gas emissions from agriculture and forestry sectors.

## 1.10   Thesis Structure

This thesis aims to investigate the evolution of air quality measurements using Sentinel 5P satellite data. Chapter 1 provides a general overview of the study, including the motivation, research questions, significance, and organization. Chapter 2 reviews previous research on the topic, discussing different Sentinel 5P data products, processing methods, and key findings. Chapter 3 describes the data collection and analysis process, examining trends in air quality from 2018 to 2022, analysing factors affecting air quality changes, and discussing implications for air quality management. Chapter 4 focuses on machine learning, describing the model architecture and evaluation, including the model type, training methods, and test set evaluation. Finally, Chapter 5 summarizes the key findings and provides recommendations for future research.

# Chapter 02: Literature Review

## 2.1 Overview

In Chapter 2, we look into the body of prior research on measurements of air quality made using information from the Sentinel 5P satellite. We will also look at some remote sensing studies. This chapter outlines the prevalent studies, methodology, and important discoveries in this area and serves as a crucial starting point for understanding the context in which our research is placed.

## 2.2 Remote sensing

Remote sensing is the capacity to record measurements of an unknown parameter from a distance. Described below is how remote sensing is relevant to our research and the types of cameras utilized.

### 2.2.1 Remote Sensing Primer

The capacity to discover details about a substance from a distance is known as remote sensing. This is accomplished by measuring the amount of light reflected from the planet using distant sensors, most frequently land observation satellites or aircraft. More specifically, distant satellite sensors assess the reflectance of electromagnetic (EM) radiation for a given place on earth, in addition to just detecting light.

We can roughly categorise remote sensing cameras into 2. 1) Optical cameras and 2) Synthetic Aperture cameras. Optical cameras read a specific portion of the electromagnetic spectrum for a given time and space. They are passive sensors. In other words, they gauge the environment's or the target's natural radiation or emission. Synthetic Aperture sensors sends radiofrequency (RF) radiation pulses in the direction of a target region. The Earth's surface is in contact with these radar frequencies. Depending on whether the characteristics are connected to land, water, or human activity, some of the signals are reflected to the instrument in patterns that convey extensive information about size, direction, composition, condition, and texture of the features encountered. In other words, active sensors produce energy and then analyse the signal that is reflected or dispersed.

Optical sensors are non-invasive means of measuring and Synthetics Aperture sensors are invasive means of measuring a parameter of a target.

### 2.2.2 Hyperspectral Remote Sensing

A sophisticated method called hyperspectral remote sensing employs specialised sensors to collect a wide variety of contiguous spectral bands from the electromagnetic spectrum. With the use of this cutting-edge technology, materials may be precisely identified based on their distinctive spectral characteristics. Hyperspectral remote sensing has a wide range of uses in forestry and agriculture, improving practises and resource management. This includes its function in crop classification, early pest and disease detection, crop water stress identification, accurate mapping of soil nutrient content,

and the facilitation of well-informed crop management decisions, such as optimising irrigation, fertilisation, and pesticide application.

Hyperspectral remote sensing has proved essential in the forestry industry for activities like full forest inventories, which require mapping different tree species, calculating biomass, and keeping track of the condition of the entire forest. It is also an essential tool for tracking trends in deforestation, seeing wildfires early on, identifying invasive species that harm ecosystems, and reliably differentiating between various tree species in forests.

Hyperspectral remote sensing is a key instrument for gathering crucial data on agriculture and forestry resources because of its adaptability and efficiency. Its increased use propels developments in both disciplines and aids in the development of more effective and sustainable resource management techniques.

## 2.3    Significance of measuring Nitrous oxides in agriculture

Monitoring nitrous oxide (NO2) emissions in agriculture and forestry holds significant environmental importance due to the potent greenhouse gas properties of NO2, which has a global warming potential (GWP) approximately 300 times that of carbon dioxide (CO2) over a 100-year timeframe. Agriculture stands as the largest anthropogenic contributor to NO2 emissions, accounting for around 4.1 teragrams of nitrogen per year, or roughly 60% of all anthropogenic NO2 emissions. Given the substantial impact of NO2 on climate change, effective monitoring and mitigation strategies are imperative (Audet et al., 2020).

One crucial facet of monitoring NO2 emissions lies in its potential to identify regions where emissions are notably high, thereby allowing for the precise targeting of mitigation efforts. By pinpointing these areas, strategies can be tailored to reduce emissions through improved fertilizer management, reduced tillage practices, and the implementation of cover crops. This proactive approach is vital for curbing the environmental impact of NO2 and mitigating its contribution to climate change.

Furthermore, as described in Section 1.6 (Research Gap) there is an existing need for soil level analysis of data as well. Furthermore, the monitoring of NO2 emissions serves as an essential tool for assessing the effectiveness of mitigation measures over time. By tracking changes in emissions following the implementation of specific strategies, researchers and policymakers can gauge the success of these initiatives and make necessary adjustments to optimize their impact.

## 2.4    Convolutional Neural Networks (CNNs) in Remote Sensing

Convolutional Neural Networks (CNNs) are a powerful subset of deep learning models that are frequently used in tasks involving image processing and categorization. These networks are modelled after the complex operations of the human visual brain, which enables them to autonomously recognise and extract relevant characteristics from images, according to the task at hand.

Convolutional layers, which serve as the foundation of CNN performance, are located at their core. These layers are critical in extracting characteristics from photos. A convolutional layer is made up of a series of filters, which are effectively tiny matrices that are applied progressively to the input picture.

These filters move through the picture like sliding windows. This operation produces a feature map, which is an image representation that emphasises the characteristics judged relevant by the filter.

Furthermore, pooling layers are incorporated into the CNN architecture to minimise the size of the feature maps. This deliberate reduction serves several reasons, including reducing network computing complexity and preventing overfitting. Among the numerous pooling approaches, max pooling is the most prevalent. It works by choosing the largest value from the feature map's defined subregions and progressively compressing its size.

The fully connected layer in a CNN is like the layers in classic neural networks. This layer acts as a link between the convolutional layers' retrieved features and the network's output layer. The output layer's final task is to classify the input image into one of the specified classes.

## 2.5   Previous Work

Several attempts have been made to use machine learning to find deforestation in low- or medium-resolution satellite data in recent years. These studies have fundamentally altered how we identify significant alterations in forests, such as the detection of widespread deforestation in the Amazon Rainforest or the quantitative measurement of primary forest loss. Despite this, there is little research on machine learning with high-resolution satellite imagery and its associated date to predict it's associated greenhouse gas emission rate. Additionally, as far as we can tell from our study, no work has been done expressly for the task of predicting the emission value of a gas using an associated satellite image and the image date. In this section, we go over the earlier research that was done in agricultural emissions. Our goal is to assess the benefits and drawbacks of earlier studies to guide the methods and data that will be used later in this report.

## 2.5.1 Literature Review of selected agricultural emissions papers

*Table 1- Agricultural Emissions papers*

| Name | Year | Important findings | Locations | Seasonal | Frequency of measurement |
|---|---|---|---|---|---|
| Carbon dioxide spatial variability and dynamics for contrasting land uses in central Brazil agricultural frontier | 2022 | The findings revealed that the temporal variability of $CO_2$ flow is favourably connecta... | Brazil | Yes | 3 years |
| Estimation of methane emissions based on crop yield and remote sensing data in a paddy field | 2020 | According to the findings, $CH_4$ emissions are positively associated to rice output and | Jiangsu, China | Yes | 1 year |
| Estimating pasture quality using Landsat ETM+: application for the greenhouse gas inventory of New Zealand | 2011 | Suggests that we can estimate GHG emissions from satellties | New Zealand | Yes | 2 year |
| Quantifying methane emissions from rice paddies in Northeast China by integrating remote sensing with a mapping | 2011 | According to the modelled results, the plain's 1.44 million hectares of rice fields | Sanjiang Plain, China | N/A | N/A |
| Modeling impacts of carbon sequestration on net greenhouse gas emissions from agricultural soils in China | 2009 | Six Chinese agroecosystems were investigated. Model results showed that (1) when | China | Yes | N/A |
| Remote Sensing Time Series to Evaluate Direct Land Use Change of Recent Expanded Sugarcane Crop in Brazil | 2012 | The research area includes the South-central region of Brazil (states of Goiás | Brazil | Yes | N/A |
| Agricultural greenhouse gas emissions of an Indian village - Who's to blame: crops or livestock? | 2012 | Dairy buffalo and indigenous dairy cattle together contribute 60 % of the methane | India | N/A | N/A |
| Examining the relationships between land cover and greenhouse gas concentrations using remote-sensing data | 2012 | Forest and croplands can be viewed as $CH_4$ sources, while scrublands and grassland | East Asia | N/A | N/A |
| Agricultural opportunities to mitigate greenhouse gas emission | 2007 | Conservation practises that reduce soil erosion may simultaneously retain soil | N/A | N/A | N/A |

The table summarises many research studies on greenhouse gas emissions, with an emphasis on carbon dioxide (CO2) and methane (CH4), as well as the influence of land use changes and agricultural practises. Each research study provides distinct insights into the dynamics of greenhouse gas emissions and their interactions with diverse environmental conditions.

One of the research projects that stands out looks at the temporal and regional variability of CO2 flux and XCO2 concentrations. It was carried out in Brazil in 2022 (Rossi et al., 2022). While XCO2 is found to have a negative association with rainfall, CO2 flow is found to have a positive correlation. In addition, the study points out that areas that were once forested now support higher CO2 emissions. The effects of climate change are, however, tempered in areas with continuous forest cover since these areas serve as carbon sinks. The three-year duration of this study demonstrates how long-term it is.

A study conducted in Jiangsu, China ("Estimation of methane emissions based on crop yield and remote sensing data in a paddy field - Shi - 2020 - Greenhouse Gases: Science and Technology - Wiley Online Library," n.d.) in 2020 found that CH4 emissions from paddy fields are positively correlated with rice yield and biomass. This suggests that biomass plays a role in CH4 generation. The study also found that remote sensing can be used to estimate CH4 emissions from paddy fields, and that this method could be used to quantify regional greenhouse gas emissions. The study spanned one year.

A study conducted in New Zealand (Ausseil et al., 2011) in 2011 also explored the feasibility of estimating greenhouse gas emissions from satellites. The study found that it is possible to estimate emissions from satellites, but more research is needed to improve the accuracy of these estimates. The study spanned two years.

China's Sanjiang Plain research study (Zhang et al., 2011) from 2011 models methane emissions from rice fields. The findings raise concerns about the changes in CH4 emissions and show that this region is responsible for sizable emissions. There is no time frame for the investigation.

The 2009 Chinese study (Qiu et al., 2009) assesses the effects of carbon sequestration on the net emissions of greenhouse gases from agricultural soils. It emphasises the intricate connection between carbon sequestration rates, NO2 emissions, and CH4 emissions while concentrating on distinct agroecosystems. According to the research, various management strategies may boost carbon sequestration rates but may also raise N2O or CH4 emissions. The study advises reducing the use of synthetic fertiliser as a potential fix. There is no mention of the study's duration.

In 2012, a research (Adami et al., 2012) in Brazil evaluated how the increase of the sugarcane crop affected land use. It highlights how crucial these adjustments are for reducing greenhouse gas emissions and bringing about local climate cooling. The South-Central region of Brazil, a significant sugarcane growing zone, is the subject of the study. The study's duration is not made clear.

In a different East Asian study from 2012, (Guo et al., 2013) researchers look at the connections between land cover and greenhouse gas concentrations using remote sensing data. Scrublands and grasslands are classified as CH4 sinks, whereas forests and agricultural lands are CH4 sources.

Finally, a 2007 report (Johnson et al., 2007) examines ways that agriculture might reduce greenhouse gas emissions. To lower methane (CH4) and nitrous oxide (N2O) emissions from agriculture, it recommends conservation practises, effective nitrogen management, changes in animal nutrition, and manure management. There is no information about the study's duration.

Other than the above recent studies from France and Turkey focused extensively on nitrous oxide emissions via the Sentinel 5P satellite imagery. The Turkish studies depicted a significant correlation between urban population density and triphosphoric column density of $NO_2$ values (Kaplan et al.,

2019). The French study (Omrani et al., 2020) concluded with a seasonal variation of $NO_2$ values. During the winter and fall seasons (ie: cold seasons) a greater density of pollutants was observed.

In conclusion, the table displays a variety of research from different geographical areas, each of which offers insightful information about the dynamics of greenhouse gas emissions and their implications for mitigating climate change. The influence of land use changes, the contribution of biomass to CH4 emissions, and the possibility of remote sensing for emissions assessment are only a few of the issues covered in these works. These studies range in length, with some lasting several years to identify long-term trends and relationships.

## 2.5.2 Literature Review of hyperspectral remote sensing papers

*Table 2 - related papers to hyperspectral image processing*

| Name | Year | Intent | Models | Results | Datasets | Evaluation Metrics |
|---|---|---|---|---|---|---|
| Hyperspectral Super-Resolution with Spectral Unmixing Constraints (Lanaras et al., 2017) | 2017 | Hyperspectral Image Superresolution | Linear models with Linear | Got good results with 4 datasets | APEX and Pavia University CAVE and Harvard | ERGAS - 3.02 RMSE - 8.23 on Apex |
| Detection of Methane Plumes Using Airborne Midwave Infrared (3–5 μm) Hyperspectral Data (Scafutto and De Souza Filho, | 2018 | Detecting Methane (CH4) using hyperspe... | Here, a controlled field experime | The CH4 plume was detected, | SEBASS Data of MWIR and LWIR | N/A |
| SIPSA-Net: Shift-Invariant Pan Sharpening with Moving Object Alignment for Satellite Imagery (Lee et al., 2021) | 2021 | panchromatic to multispectral superreso... | Custom 2 stage alighment and | The superRes was successful | WorldView-3 satellite image dataset | SCC - 0.964 ERGAS - 0.858 PSNR – |
| An Approach To Super-Resolution Of Sentinel-2 Images Based On Generative Adversarial Networks (Zhang et al., 2020) | 2021 | SuperRes using GAN. | S2GAN | SuperRes is better than DSen, | SENTINEL-2 | RMSE - 33.1 SRE − 36.4 SAM – |
| A New Super Resolution Framework Based on Multi-Task Learning for Remote Sensing Images (Yan and Chang, 2021) | 2021 | A new multi task superRes method using blur | Used a recurrent feature capturing | Good results when compared | DIOR Dataset | PSNR and SSIM – numerous in the paper |
| A New Divergence Method to Quantify Methane Emissions Using Observations of Sentinel-5P TROPOMI (Liu et al., 2021) | 2021 | A new divergence method has been successful | Linear models with Linear | Fairly good results but | Sentinel 5P | N/A |
| First data set of H2O/HDO columns from the Tropospheric Monitoring Instrument (TROPOMI) (Schneider et al., | 2020 | Global measurements of atmospheric water | No models just stats | Global coverage but only for a few | Sentinel 5P | N/A |
| Weakly-supervised Semantic Segmentation in Cityscape via Hyperspectral Image (Huang et al., 2021) | 2021 | Used Hyperspectral images to semantic | Resnet 50 convolutional layers | Good results but not satellite | (a) Hyperspectral City V1.0 (b) Cityscape | mIoU - 69.4 Acc - 82.1 |
| An adaptive semantic dimensionality reduction approach for hyperspectral imagery classification ("An... classification") | 2018 | HSI classification is a challenge | PCA | | Indiana Pines and PAVIA | N/A |

| Title | Year | Method | Approach | Region | Data | Notes |
|---|---|---|---|---|---|---|
| A study of tropospheric NO2 variability over Pakistan using OMI data (ul-Haq et al., 2014) | 2014 | Top-down approach: they found hotspots | - | Pakistan | OMI data | N/A |
| Variability of nitrogen oxide emission fluxes and lifetimes estimated from Sentinel-5P TROPOMI observations | 2019 | | Linear models with Linear | Global | Sentinel 5P | N/A |

Hyperspectral Super-Resolution with Spectral Unmixing Constraints and A New Super Resolution Framework Based on Multi-Task Learning for Remote Sensing Images both propose new methods for super-resolving hyperspectral images. However, the first method uses spectral unmixing constraints, while the second method uses multi-task learning. Both methods are shown to outperform other super-resolution methods on their respective datasets.

Detection of Methane Plumes Using Airborne Midwave Infrared (3–5 µm) Hyperspectral Data and New Divergence Method to Quantify Methane Emissions Using Observations of Sentinel-5P TROPOMI both focus on the detection and quantification of methane emissions using hyperspectral data. However, the first study uses airborne data, while the second study uses satellite data. Both studies show that hyperspectral data can be used to detect and quantify methane emissions, but the second study shows that satellite data is more effective than airborne data for this purpose.

N/ASIPSA-Net: Shift-Invariant Pan Sharpening with Moving Object Alignment for Satellite Imagery and Weakly-supervised Semantic Segmentation in Cityscape via Hyperspectral Image both propose new methods for processing hyperspectral data for specific applications. The first method proposes a new pan sharpening method for satellite imagery that can handle moving objects. The second method proposes a new method for weakly-supervised semantic segmentation of cityscapes. Both methods are shown to achieve good results on their respective datasets.

First data set of H2O/HDO columns from the Tropospheric Monitoring Instrument (TROPOMI) and Variability of nitrogen oxide emission fluxes and lifetimes estimated from Sentinel-5P TROPOMI observations both present new findings from the Sentinel-5P TROPOMI instrument. The first study presents the first data set of H2O/HDO columns from TROPOMI. The second study presents a study on the variability of nitrogen oxide emission fluxes and lifetimes estimated from TROPOMI observations. Both studies provide new insights into the capabilities of the TROPOMI instrument and the potential of hyperspectral data for remote sensing applications.

## 2.5.3 Literature Review of deep learning and remote sensed papers

*Table 3 - assorted related work in deep learning and remote sensing*

| Paper | Year | Location | Dataset | Model | Findings | Evaluation metric |
|---|---|---|---|---|---|---|
| Satellite observations of atmospheric methane and their value for quantifying methane emissions (Jacob et al., 2016) | 2016 | USA | GOSAT, NOAA | - | Methane observations from GOSAT since | N/A |
| The added value of satellite observations of methane for understanding the contemporary methane budget ("The added value of satellite observations of methane for understanding the", 2021) | 2021 | Global | TROPOspheric Monitoring | - | Use data from the European TROPOspheric | N/A |
| Satellite-derived methane hotspot emission estimates using a fast data-driven method (Buchwitz et al., 2017) | 2016 | USA, Turkmenistan and Azerbaija | Landsat | Linear regression | We propose a new data-driven method | N/A |
| Automated Identification of Oil Field Features using CNNs (Dileep et al., 2020) | 2020 | UK | Landsat | CNNs | We propose a new method for | N/A |
| Spatio-temporal data on the air pollutant nitrogen dioxide derived from Sentinel satellite for France (Omrani et al., 2020) | 2020 | France | Sentinel-5P NO2 data | - | Developed a new spatio-temporal dataset of | N/A |
| Spatial Variation of NO2 and Its Impact Factors in China: An Application of Sentinel-5P Products (Zheng et al., 2019) | 2019 | China | Sentinel-5P NO2 data | - | Found that the monthly fluctuation of | R-Square – for GWR (0.643) -for OLS |
| Ground-based validation of the Copernicus Sentinel-5p TROPOMI NO2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks (Verhoelst et al., 2021) | 2021 | UK | Sentinel-5P NO2 data and ground- | - | Found that Sentinel-5P TROPOMI | N/A |
| Rotation Equivariant Deforestation Segmentation and Driver Classification (Mitton and Murray-Smith, 2021) | 2021 | - | Satellite images of deforestation | Rotation equivariant neural network | Developed a rotation equivariant neural | Accuracy - 63.0 mIoU - 72.3 |
| FIgLib & SmokeyNet: Dataset and Deep Learning Model for Real-Time Wildland Fire Smoke Detection ("Remote Sensing | Free Full-Text | FIgLib & SmokeyNet: Dataset and Deep Learning Model for", 2021) | 2021 | N/A | FIgLib dataset | Deep learning model | Proposed a new dataset and deep learning | A - 82.62 F1 – 82.83 P – 90.85 R - 89.89 |

| Paper | Year | Country | Data | Model | Contribution | Results |
|---|---|---|---|---|---|---|
| GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction (Fan et al., 2022) | 2021 | USA | Crop yield data, geospatial data, | GNN-RNN | Proposed a GNN-RNN approach for | Various assorted results |
| Data integration pipeline towards reliable monitoring of phytoplankton and early detection of harmful algal blooms (Guterres et al., 2021) | 2021 | N/A | Phytoplankton data, environmental data | - | Proposed a data integration pipeline towards | N/A |
| HyperionSolarNet: Solar Panel Detection from Aerial Images (Parhar et al., 2022) | 2021 | N/A | Aerial images of solar panels | Hyperion SolarNet | Proposed a Hyperion SolarNet model for | Mean Accuracy – 0.96, other |

Satellite observations of atmospheric methane and their value for quantifying methane emissions (2016)

The most recent methods for measuring methane emissions using satellite data are reviewed in this publication. It addresses the benefits and drawbacks of various satellite platforms and data products and emphasises the need for additional study to create more precise and dependable methane quantification techniques.

The added value of satellite observations of methane for understanding the contemporary methane budget (2021)

The utility of satellite data in comparison to data gathered by ground-based monitoring stations is evaluated in this research. It demonstrates that satellite data can offer ground-based data more thorough spatial coverage and higher temporal precision, which is crucial for comprehending the global methane budget.

Satellite-derived methane hotspot emission estimates using a fast data-driven method (2016)

The innovative data-driven approach for calculating methane hotspot emission estimates using satellite data is presented in this research. The approach is based on a linear regression model that was trained using data on known-source methane emissions. The technique has been demonstrated to be precise and effective, and it has the potential to be used to track methane emissions from a number of industries, including landfills, agriculture, and oil and gas extraction.

Automated Identification of Oil Field Features using CNNs (2020)

In this study, convolutional neural networks (CNNs) are used to introduce a novel technique for recognising characteristics in oil fields. A deep learning model called a CNN is ideally suited for image identification applications. It is demonstrated that the suggested strategy is effective and precise at locating oil field objects in Landsat pictures.

Spatio-temporal data on the air pollutant nitrogen dioxide derived from Sentinel satellite for France (2020)

The Sentinel-5P satellite obtained a brand-new spatio-temporal dataset of nitrogen dioxide (NO2) for France, which is presented in this study. The regional and temporal distribution of NO2 in France is examined using the dataset. According to the findings, urban regions and the winter have the highest

NO2 concentrations. The authors also list the key influences on NO2 column concentration as nighttime light intensity, the ratio of secondary and tertiary industries, and vehicle ownership.

Spatial Variation of NO2 and Its Impact Factors in China: An Application of Sentinel-5P Products (2019)

In this study, the regional and temporal distribution of NO2 in China is examined using Sentinel-5P data. The findings demonstrate that the tropospheric NO2 column concentration fluctuates monthly and clearly exhibits the seasonal pattern of "high in winter and low in summer," with the spatial distribution following a "high in East and low in west" trend. The authors also list the key influences on NO2 column concentration as nighttime light intensity, the ratio of secondary and tertiary industries, and vehicle ownership.

Ground-based validation of the Copernicus Sentinel-5p TROPOMI NO2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks (2021)

In this study, ground-based NO2 readings from the NDACC ZSL-DOAS, MAX-DOAS, and Pandonia worldwide networks are compared to Sentinel-5P measurements. The findings demonstrate some bias, especially at high NO2 concentrations, but overall Sentinel-5P NO2 values are compatible with ground-based data.

Rotation Equivariant Deforestation Segmentation and Driver Classification (2021)

In this paper, a new rotation equivariant neural network for driver classification and deforestation segmentation using satellite pictures is proposed. One kind of neural network that can learn from rotated data is a rotation equivariant neural network. The suggested method is demonstrated to be very precise and recall-efficient at segmenting deforestation areas and categorising the drivers of deforestation.

FIgLib & SmokeyNet: Dataset and Deep Learning Model for Real-Time Wildland Fire Smoke Detection (2021)

For the purpose of detecting smoke from wildfires in real time, this research suggests a new dataset and deep learning model. Over 100,000 photos of smoke from wildland fires and other objects have been labelled in the dataset, which goes by the name FIgLib. SmokeyNet, a deep learning model that is trained on FIgLib, has a high level of real-time accuracy while detecting smoke from wildland fires.

GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction (2021)

In order to use geospatial and temporal data for agricultural production prediction, a novel GNN-RNN technique is proposed in this study. A class of neural network known as a GNN-RNN is capable of learning from both sequential and graph-structured material. The suggested method has the potential to be used to direct agricultural decision-making because it has been demonstrated to be accurate and effective at predicting crop yields.

## 2.6   Chapter Summary

This chapter provides a thorough overview of current developments in hyperspectral remote sensing and remote sensing, which forms the basis for our own study. We learn more about the many approaches, datasets, and assessment criteria employed in the field by reading these papers.

Furthermore, it provides the appropriate data sources locations that have not had any surveys and the time periods in such locations with minimal survey and sets the groundwork for our own contributions to this developing topic in the next chapters.

All the papers listed above use satellite data to study various environmental problems. However, they use different satellite platforms and data products, and they focus on different applications. None focus on any singular objectives we outlined as the research goals above or work on the research objectives we have outlined.

# Chapter 03: Data Collection Methodology and Analysis

## 3.1    Overview

The important elements of the data collection process are covered in depth in Chapter 3 of our study project. In this chapter, we explain the techniques, resources, and equipment we used to gather the crucial data for our study. This chapter's importance rests in demonstrating the transparency and comprehension of the basis for our research.

To provide a complete overview of the background work needed for our investigation, Chapter 3 is essential. It provides readers with reassurance regarding the care and rigour used in gathering the data necessary for our investigation. The approaches and factors discussed in this chapter provide as a springboard for the chapters that follow, where we analyse, model, and derive insights from the data.

## 3.2    Data Sources

Images from the Sentinel 2 and the Sentinel 5P satellites were used for our research. The RGB composite image from the Sentinel 2 were used as inputs images and the mean value of that $NO_2$ gas in that particular image area (obtained from the Sentinel 5P) was used as regression targets in the model. (Explained subsequently). The images were obtained from 24 diverse soil conditions, spanning a period of 4 years (September 2018 to September 2022)

## 3.3    Description of the data collection tools

Agricultural soils are a strong contributing factor in GHG emissions. They produce nitrous oxide when nitric fertilizers are applied. The United States Department of Agriculture (USDA) classified soil types around the world in an extensive survey (given below). The soils are classified ("Keys to Soil Taxonomy | NRCS Soils," n.d.) into 12 distinct categories. In addition to the 12 mentioned there are 3 more classes (shifting sand, ice glacier, and rocky land) depicted in a Global Soil Region map below ("Global Soil Regions Map | NRCS Soils," n.d.).  The codes for the data extraction tool from the Sentinel 5P provided in Data Extraction Tool.5P in the Code and Data section)

The codes above are designed for processing and analyzing satellite data from the Sentinel-5P mission. Below is a summary of its functionality.

- Library Installation and Imports: The code begins by installing the geemap library and importing necessary modules such as tifffile, matplotlib, numpy, geemap, math, ee, and os. These libraries are used for image processing, mathematical calculations, and interacting with the Google Earth Engine API.
- Helper Functions: It defines helper functions for formatting dates, mosaicking daily images, and calculating bounding boxes based on latitude and longitude.

- Main Analysis Function: A primary function get_gas_value_for_location is detailed, which retrieves greenhouse gas readings for a specified location within a given time frame. It utilizes parameters like latitude, longitude, start and end dates, dataset, gas band, and others to perform the analysis.
- Visuals: The code includes commented-out sections for visualizing data using tifffile and matplotlib, as well as for measuring execution time of code blocks.
- Date Configuration: It automates the creation of a date range for analysis and prepares a list of formatted dates.
- Location Data and File Handling: The notebook contains lists of geographical locations with associated soil types, commands for downloading and unpacking datasets, and integration with Sentinel-2 data.
- Data Processing Loop: A loop processes the data for each location over the specified date range, capturing greenhouse gas values and handling exceptions.
- Data Integration and Output: The code integrates data from Sentinel-2, mounts Google Drive for storage, and writes the processed data to a CSV file. It also includes code for zipping and downloading the dataset.

The codes for the data extraction tool from the Sentinel 2 provided in Data Extraction Tool.2 in the Code and Data section)

The codes above are designed for processing and analyzing satellite data from the Sentinel-2 mission. Below is a summary of its functionality.

- Geospatial Library Utilization: The code employs the geemap Python library, which is a popular tool for geospatial data analysis, particularly with the Google Earth Engine.
- Authentication and Initialization: It includes a procedure for authenticating and initializing the Earth Engine API, which is essential for accessing and processing the vast satellite imagery data available on the platform.
- Mathematical Functions for Geodesy: The notebook defines mathematical functions to convert degrees to radians and vice versa, and to calculate the Earth's radius at a given latitude based on the WGS-84 ellipsoid model.
- Image Processing Workflow: The core functionality of the code involves defining a bounding box for a region of interest, masking clouds from satellite images, calculating the percentage of black pixels (potentially indicating missing data), and retrieving images for specific locations and time frames.
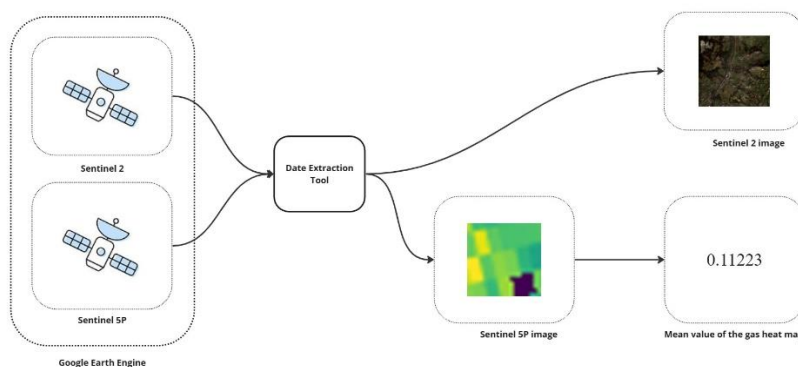


*Figure 1 - Data flow diagram*
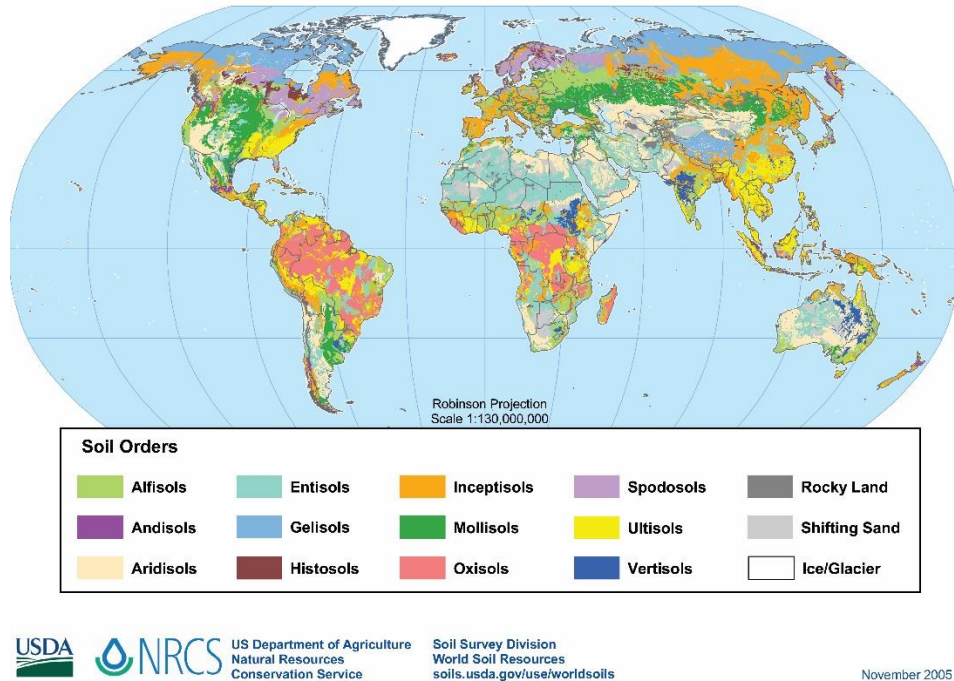
# Global Soil Regions



*Figure 2- Global Soil Region USDA*

Our analysis entails an agricultural and forestry vantage point. We selected a minimum of 3 locations from each soil order. Since agricultural lands and forestry consists of large land areas, we would get an average of at readings from a swath of land area surrounding a given point.

The equations to calculate the square shaped surface area bounding box coordinates of readings to be considered at a chosen point is described and depicted below. First and foremost, for a given latitude and longitude, the degree coordinates would have to be converted to radians.

The following equation describes the above process. This conversion ensures that we work with angular measurements in radians, which are commonly used in trigonometric calculations.

$$\text{(latitude or longitude)}\text{in radians} = \frac{\text{(latitude or longitude) in degrees}}{180} \times \pi$$

Contrary to popular belief the Earth is not a perfect sphere. WGS-84 (World Geodetic System of 1984) considers the Earth as a spheroid or an ellipsoid of revolution. Spheroids are ellipsoids of which 2 of 3 semi-axes are equal. As such WGS-84 defines the 2 semi diameters (and by extension radius) of the Earth as follows; the equatorial radius and the polar radius. They are given the following values.

$$\text{equatorial radius of Earth(WSG84A)} = 6378137.0m$$

$$\text{polar radius of Earth(WSG84B)} = 6356752.3142m$$

The equations to find the radius and the parallel radius at a given latitude is given below.

$$\text{A} = \text{WSG84 A}^2 \times \cos(\text{latitude in radians})$$

27

$$B = \text{WSG84 B}^2 \times \sin(\text{latitude in radians})$$

$$C = \text{WSG84 A} \times \cos(\text{latitude in radians})$$

$$D = \text{WSG84 B} \times \sin(\text{latitudeinradians})$$

$$\text{radius at a given latitude} = \sqrt{\frac{(A^2 + B^2)}{(C^2 + D^2)}}$$

$$\text{parallel radius at a given latitude} = \text{radius at a given latitude} \times \cos(\text{latitude in radians})$$

The coordinates of bounding square of a given point can be calculated as below. Half-length of a side of the square is denoted by H.

$$\text{latitude } min = \text{latitude in radians} - \frac{H}{\text{radius at a given latitude}}$$

$$\text{latitude } max = \text{latitude in radias} + \frac{H}{\text{radius at a given latitude}}$$

$$\text{longitude } min = \text{longitude in radians} - \frac{H}{\text{parallel radius at a given latitude}}$$

$$\text{longitude } max = \text{longitude in radians} + \frac{H}{\text{parallel radius at a given latitude}}$$

All these bounding square coordinates are to be converted to degrees with the following equation.

$$\text{(latitude or longitude) in degree} = \frac{\text{(latitude or longitude) in radians}}{\pi} \times 180$$

With the above coordinates we can create a bounding box for a given latitude and longitude, of length (2H) from the centroid to the sides. We take the mean values of all the readings within in the bounding box to determine a net average value per day. In our experiments we use a bounding box that is a square of side 10 km. This would be able to estimate the average emissions within the area enclosed by the bounding box.

## 3.4   Singular results obtained from the tools

An example of the results obtainable from the above software tools to monitor Nitrogen Dioxide are depicted below in the examples.

- For a localized region of agricultural land and forests (and basically covers majority of the countries in the world)
- Global coverage

### 3.4.1 Global Coverage

Often it is crucial to have a holistic view before conducting further scrutiny. For such purposes, we have a tool to view the average value of emissions all over the coverage of the Sentinel 5P satellite. Figure 2 shows the average values of $NO_2$ obtained for a period between 7th June 2020 and 30th June 2020.

It is observed that Japan, China, South Korea, India, Iran, Lebanon, Egypt, Russia, Germany, Italy and the USA are significant emitters of $NO_2$ gas during the time frame. This enables us to atomize parts of areas relevant to forestry and agriculture better.

Below is an example of the results obtained. This heatmap uses a color scale that transitions from 'black' to 'blue', then 'purple', 'cyan', 'green', 'yellow', and finally 'red'. The lowest values (0.0) in the heatmap will be represented by 'black', while the highest values (0.0002) will be represented by 'red', with intermediate values being represented by the other colors in the palette above.
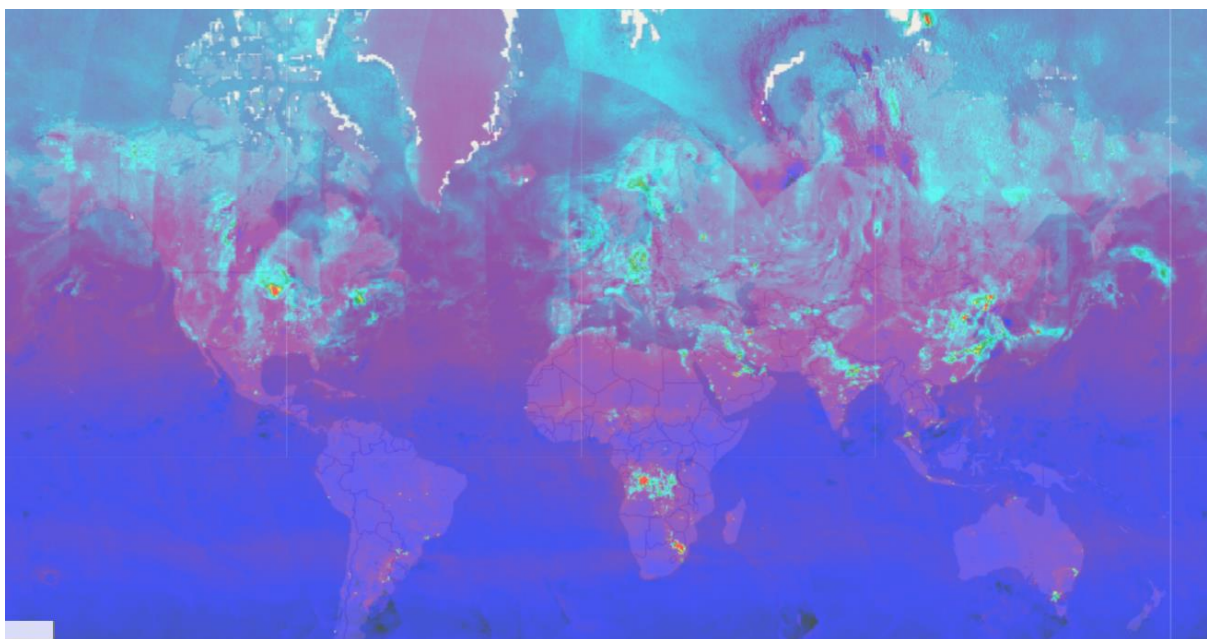


*Figure 3- Global average emissions of NO2 between 7th June 2020 to 30th June 2020*

### 3.4.2 Localized Coverage

Let's consider one region for example. The North Island of New Zealand houses nearly 65% of the New Zealand population. The particular area that was surveyed is projected to increase it's agricultural productivity by nearly 100% in 2040s ("Integrated National Policy Approaches to Climate-Smart Agriculture. Insights from Brazil, Ethiopia, and New Zealand," 2014). This area is also home to the Whirinaki Forest Park and the Te Urewera homeland of the Tuhoe Maori people. The soil variety in this region is of Andisol type. The specific longitude and latitude we surveyed was [-38.44145537791065, 176.80194705056425] respectively. The time frame of survey is between September 2018 to August 2019

The tool has a feature to annotate the specific area being surveyed as displayed below in Figure 3. The line plot of this region is also obtainable as shown in Figure 4.
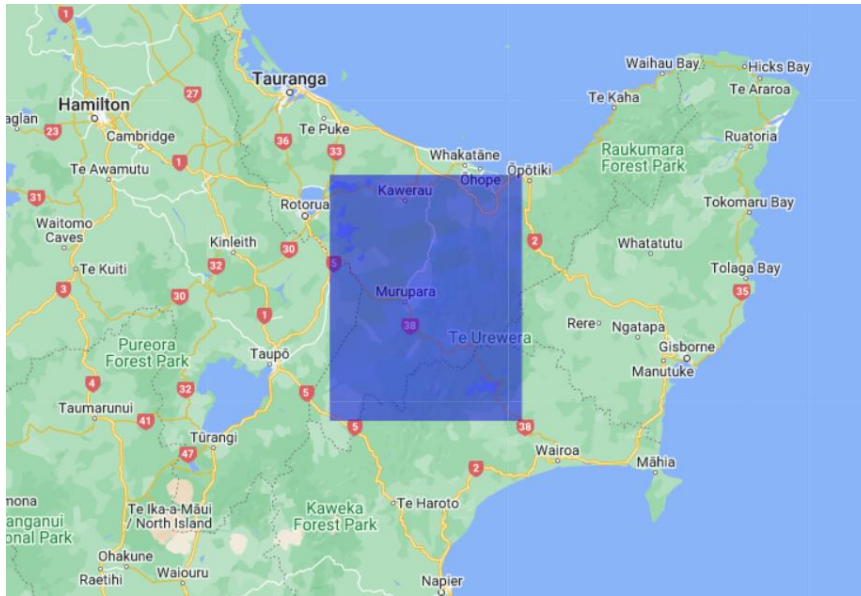
*Figure 4- Annotated Area of survey*
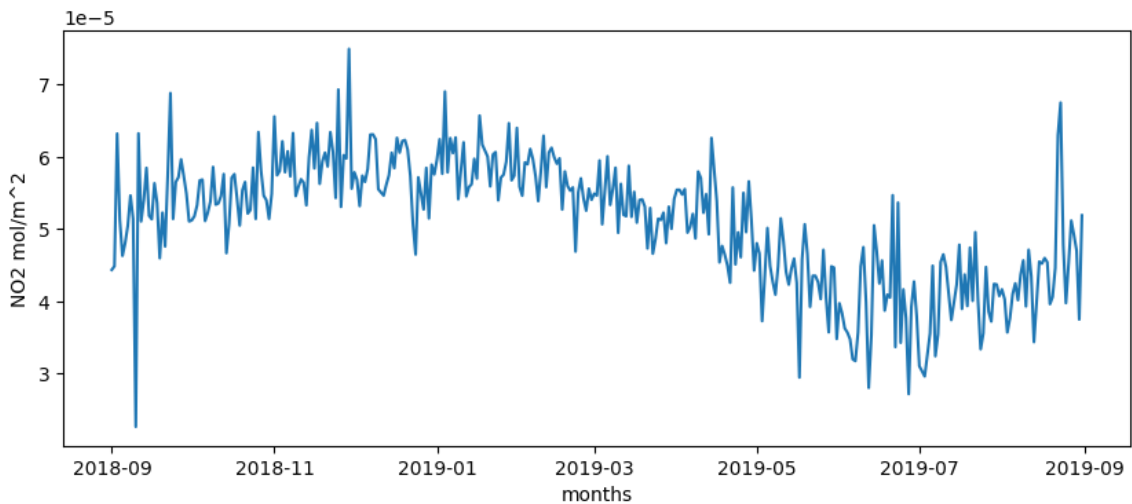


*Figure 5- Daily emission value for a year in the above area*

## 3.5 Description of the locations surveyed

This dataset comprises historical data collected from a total of 24 distinct geographical locations, representing a diverse array of eight different soil orders. It is important to note that the data collection process was conducted impartially, encompassing regions spanning from the global northern hemisphere to the global southern hemisphere. The data acquisition itself was facilitated through the utilization of the Google Earth Engine, a sophisticated geospatial analysis platform. The temporal scope of this data compilation spans from September 2018 to September 2022. The locations of survey is displayed below.
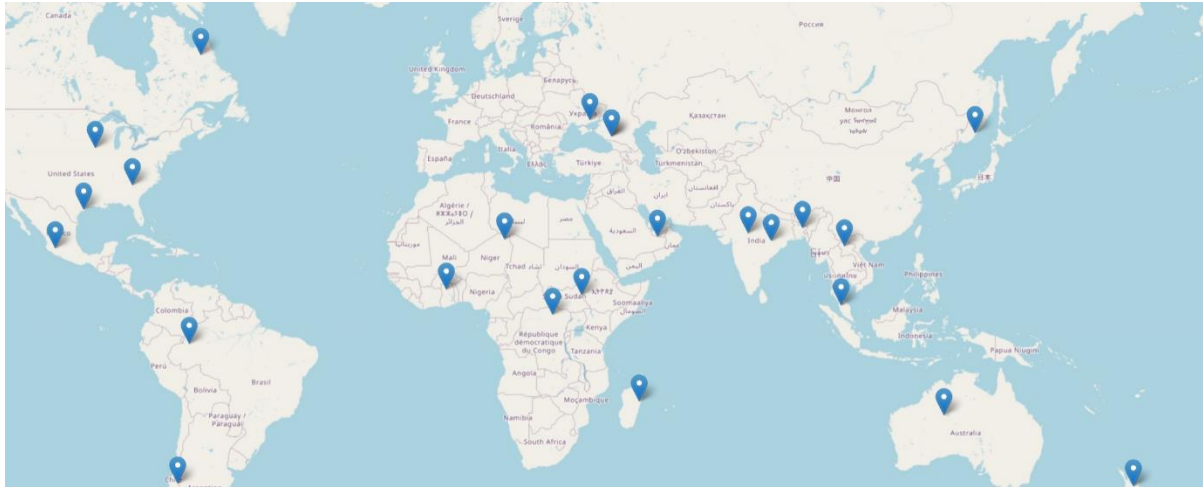
*Figure 6- Locations of the data collection sites*

*Table 4- Soil order and crops grown in the surveyed areas*

| Serial No. | Location | Latitude | Longitude | Soil Order | Crops or Forests |
|---|---|---|---|---|---|
| 1 | Karachay-Cherkessia, Russia | 44.02549592 | 41.27779137 | Mollisol | Wheat, corn, sunflowers |
| 2 | Melitopol, Ukraine | 47.04296361 | 35.75185611 | Mollisol | Wheat, corn, sunflowers |
| 3 | Iowa, America | 41.90193949 | -92.75029809 | Mollisol | Corn, soybeans, hay |
| 4 | Tver, Russia | 44.02549592 | 41.27779137 | Alfisol | Wheat, rye, potatoes |
| 5 | Odisha, India | 21.4111331 | 82.71054963 | Alfisol | Rice, jute, wheat |
| 6 | North Gonja, Ghana | 9.388339935 | -1.526934298 | Alfisol | Cocoa, plantains, cassava |
| 7 | North Island, NewZealand | -38.44145538 | 176.8019471 | Andisol | Beech forests, podocarp forests |
| 8 | Jalisco, Mexico | 19.79403165 | -103.1927506 | Andisol | Corn, beans, coffee |
| 9 | Santa Barbara, Chile | -37.68590304 | -71.41446515 | Andisol | Vineyards, olive groves |
| 10 | Madama, Niger | 22.01110157 | 13.46676007 | Entisol | Millet, sorghum, peanuts |
| 11 | Shubaytah, SaudiArabia | 22.61869806 | 53.24287929 | Entisol | Date palms, wheat |

| | | | | | |
|---|---|---|---|---|---|
| 12 | Gibson Desert North, Australia | -22.54613578 | 127.6495604 | Entisol | Shrubs, grasses |
| 13 | Primorsky Krai, Russia | 44.57940166 | 135.6788987 | Inceptisol | Soybeans, corn, wheat |
| 14 | Muzaffarpur, Bangladesh | 24.60487059 | 90.89032422 | Inceptisol | Rice, wheat, jute |
| 15 | Quebec, Canada | 57.27976996 | -65.46318238 | Inceptisol | Forests, crops |
| 16 | Jutai, Brazil | -4.549269446 | -68.29210095 | Oxisol | Soybeans, corn, coffee |
| 17 | Bambesa, Congo | 3.016830986 | 26.05157797 | Oxisol | Coffee, cassava, bananas |
| 18 | Moramanga, Madagascar | -19.13854286 | 48.51022224 | Oxisol | Shrubs, grasses |
| 19 | Georgia, America | 34.24755459 | -83.11264726 | Ultisol | Corn, soybeans, hay |
| 20 | Perak, Malaysia | 5.30571406 | 100.8891812 | Ultisol | Rubber, oil palm |
| 21 | Oudomxay, Laos | 20.17307836 | 101.761504 | Ultisol | Rice, corn, coffee |
| 22 | Madhya Pradesh, India | 23.36609533 | 76.86099548 | Vertisol | Millet, sorghum, peanuts |
| 23 | Aboro, Ethiopia | 8.082542473 | 33.49662823 | Vertisol | Enset, coffee |
| 24 | Houston, Texas | 28.9770527 | -95.74814063 | Vertisol | Corn, sugarcane |

Table 4 offers a summary of the geographic coordinates, soil order, and types of crops or forests that can be found in many regions across the world. To ensure a diverse dataset for study, these places were chosen to reflect agricultural lands in various geographic areas and climatic conditions. Data on changes and trends in crop agriculture and forest cover were gathered from September 2018 to September 2022.

The Mollisol soil order, which is prominently displayed in the table, is represented by regions like Iowa in the United States and Karachay-Cherkessia in Russia. Wheat, corn, and sunflowers are the main crops in these areas. These agricultural staples are essential to local economies and the production of food on a worldwide scale.

The chart also features the Alfisol soil order, which is found in places like Tver, Russia, and Odisha, India. Farmers grow crops like rice, jute, wheat, and potatoes in these regions. certain crops are necessary for the food security and economic viability of certain areas.

Moving on to Andisol, the table emphasises places like Jalisco, Mexico, and North Island, New Zealand. Here, varied landscapes, such as vineyards and olive orchards in Mexico and beech and podocarp forests in New Zealand, demonstrate how adaptable andisols are for maintaining different kinds of vegetation.

Primorsky Krai in Russia, characterized by Inceptisol soil, predominantly features soybeans, corn, and wheat cultivation. This region, situated at approximately 44.58° latitude and 135.68° longitude, showcases the versatility of Inceptisols in supporting various agricultural activities.

Moving to Muzaffarpur, Bangladesh, with its Inceptisol soil, rice, wheat, and jute are the primary crops cultivated. This area, located at approximately 24.60° latitude and 90.89° longitude, is known for its significant contribution to rice production in the country.

Quebec, Canada, falls within the Inceptisol category, and it exhibits a blend of forests and croplands. Positioned around 57.28° latitude and -65.46° longitude, Quebec's landscape highlights the coexistence of natural woodlands and agricultural practices. Jutai, Brazil, is situated in an Oxisol region, where soybeans, corn, and coffee are the primary crops. The coordinates at approximately -4.55° latitude and -68.29° longitude place this area within the Amazon rainforest, where agricultural activities are interspersed with dense vegetation.Bambesa in the Congo, also with Oxisol soil, relies on coffee, cassava, and bananas for its agricultural output. Located at around 3.02° latitude and 26.05° longitude, this region plays a vital role in the production of coffee beans in Central Africa.

Moramanga, Madagascar, characterized by Oxisol soil, features shrubs and grasses within its landscape. Positioned at approximately -19.14° latitude and 48.51° longitude, this region's terrain is marked by its natural vegetation and grassy expanses. Georgia in the United States, with its Ultisol soil, engages in corn, soybean, and hay cultivation. Located at around 34.25° latitude and -83.11° longitude, this American state is renowned for its agricultural activities.

Perak, Malaysia, also categorized under Ultisol soil, specializes in rubber and oil palm cultivation. Positioned at approximately 5.31° latitude and 100.89° longitude, Perak contributes significantly to Malaysia's rubber and palm oil industries.

Oudomxay, Laos, situated in the Ultisol soil category, focuses on rice, corn, and coffee production. Located at approximately 20.17° latitude and 101.76° longitude, this region in Southeast Asia plays a vital role in the country's crop cultivation.

Madhya Pradesh, India, marked by Vertisol soil, showcases a landscape where millet, sorghum, and peanuts are the primary crops. Positioned at around 23.37° latitude and 76.86° longitude, this Indian state demonstrates the suitability of Vertisols for diverse crop cultivation.

Aboro, Ethiopia, situated in an area characterized by Vertisol soil, is known for the cultivation of enset and coffee. Positioned at approximately 8.08° latitude and 33.50° longitude, this region showcases the significance of Vertisols in supporting the growth of enset, a drought-tolerant crop, and coffee, one of Ethiopia's major agricultural exports. The ability of Vertisols to retain moisture and nutrients contributes to the successful cultivation of these crops in this East African region.

Houston, Texas, falls within the Vertisol soil category, where corn and sugarcane are the primary crops grown. With coordinates at around 28.98° latitude and -95.75° longitude, Houston exemplifies the agricultural diversity of Vertisols in the United States. Corn and sugarcane are essential components of the region's agricultural landscape, with corn serving as a staple crop and sugarcane supporting various industries, including sugar production and biofuel development.

## 3.6   Description of the collected data

The initial phase of data retrieval involved obtaining data from the Sentinel-2 satellite, a satellite renowned for its capability to perform multispectral Earth observations. However, a key consideration in this process was the discrepancy in the orbital periods of the Sentinel-2 and Sentinel-5P satellites. Consequently, an approach was adopted, wherein data points were exclusively selected for analysis during time intervals when both satellites orbited over the same geographic locations. This stringent selection criterion was implemented to ensure the alignment of data acquired from both satellite sources, thereby enhancing the dataset's accuracy and reliability.

Subsequently, employing the analytical tools previously developed and demonstrated, the dataset underwent an averaging process. Specifically, the gas concentration values, expressed in terms of molar concentration per square meter (mol/m²), were averaged for each respective region using data acquired from the Sentinel-5P satellite. The region around each latitude and longitude is a square of 10km each. This gives it an effective surface area of 100 $km^2$. This data processing step aimed to derive a representative and consolidated value that encapsulated the gas concentrations within the specified regions.

The following depiction demonstrates one example of the collected data and the relationship between the data. The images are 3 channel images and of size 200 x 200.
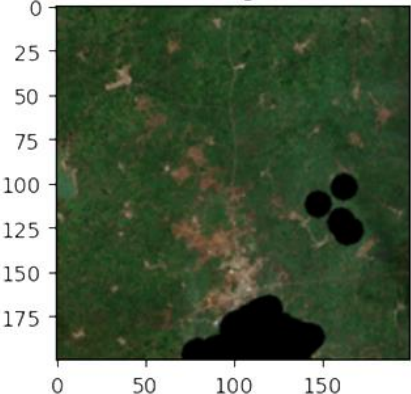
*Table 5- Example of one unit of data*

| Image | Soil Order | Date of image | Average NO emission in nano moles per sq m |
|---|---|---|---|
|  | Molllisol | (2019,3,1) | 81.1224 |

Table 6 - Dataset number of images per soild order

| Soil Type | No of Images |
|-----------|-------------:|
| Vertisol | 496 |
| Mollisol | 416 |
| Ultisol | 268 |
| Andisol | 500 |
| Alfisol | 383 |
| Oxisol | 185 |
| Entisol | 896 |
| Inceptisol | 362 |

This is an unbalanced dataset on a soil order level. However since the task is regression the unbalanced nature of the dataset in the soil order level would not impact the prediction capabilities of the task. The dataset was split into 80% and 20% of training and testing data respectively.

## 3.7   Results based on soil types

Nitrous oxide (NO2) emissions from various soil orders have been thoroughly examined, and the results have provided important new insights into the global environmental dynamics of many different nations. This study explores the emissions data of eight different soil orders over the course of the last four years, from September 2018 to September 2022. These orders are Entisol, Inceptisol, Mollisol, Alfisol, Ultisol, Oxisol, Andisol, and Vertisol. The rigorously recorded and studied annual data provides significant insight into the dynamics of NO2 emissions within these soils, illuminating the environmental difficulties these areas face. This introduction lays the groundwork for a thorough examination of the conclusions and ramifications of this multi-year investigation.
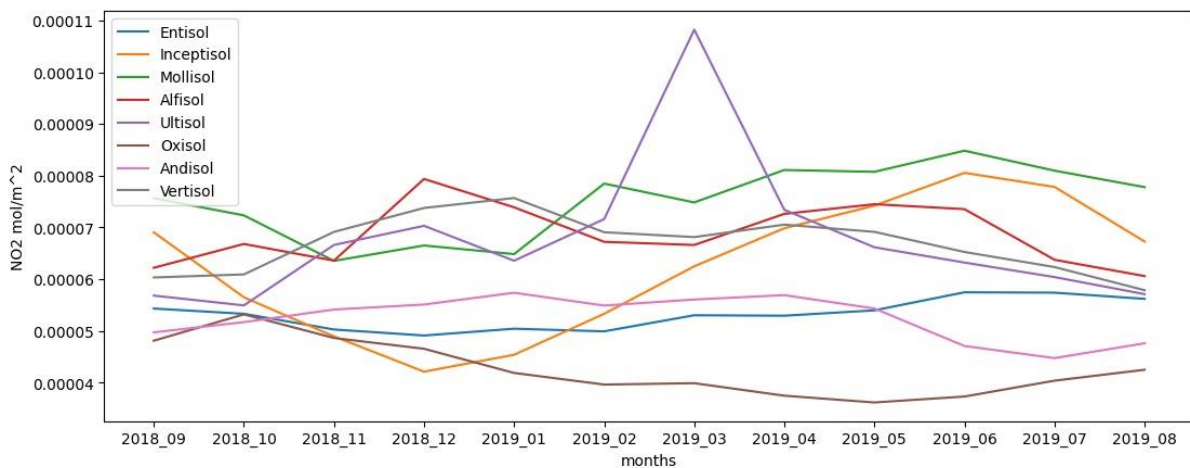
### 3.7.1   2018-2019



*Figure 7- Soil level emissions between 2018-2019*

35

Table 7- Statistics of NO2 emission for the year 2018-2019

| Soil order | Mean | Median | Std Deviation | Min | 25th percentile | 50th percentile | 75th percentile | Max | Range |
|---|---|---|---|---|---|---|---|---|---|
| Entisol | 0.000053 | 0.000053 | 0.000003 | 0.000049 | 0.00005 | 0.000053 | 0.000055 | 0.000057 | 0.000008 |
| Inceptisol | 0.000062 | 0.000065 | 0.000013 | 0.000042 | 0.000052 | 0.000065 | 0.000071 | 0.000081 | 0.000038 |
| Mollisol | 0.000075 | 0.000077 | 0.000007 | 0.000064 | 0.000071 | 0.000077 | 0.000081 | 0.000085 | 0.000021 |
| Alfisol | 0.000069 | 0.000067 | 0.000006 | 0.000061 | 0.000064 | 0.000067 | 0.000074 | 0.000079 | 0.000019 |
| Ultisol | 0.000068 | 0.000065 | 0.000014 | 0.000055 | 0.00006 | 0.000065 | 0.000071 | 0.000108 | 0.000053 |
| Oxisol | 0.000043 | 0.000041 | 0.000005 | 0.000036 | 0.000039 | 0.000041 | 0.000047 | 0.000053 | 0.000017 |
| Andisol | 0.000052 | 0.000054 | 0.000004 | 0.000045 | 0.000049 | 0.000054 | 0.000055 | 0.000057 | 0.000013 |
| Vertisol | 0.000067 | 0.000069 | 0.000006 | 0.000058 | 0.000062 | 0.000069 | 0.000069 | 0.000076 | 0.000018 |

In the year 2018-2019, as depicted in (Table 7- Statistics of NO2 emission for the year 2018-2019) the NO2 emissions within different soil orders displayed various trends. Entisols remained stable, with consistent NO2 levels throughout the year. Mollisols showed a slight increase in NO2 emissions, suggesting a gradual environmental change affecting these soils. Inceptisols followed a pattern of declining NO2 emissions towards the end of the year, followed by a subsequent increase, ultimately returning to levels similar to the previous year. Alfisols exhibited a modest decrease in NO2 emissions, indicating a possible reduction in NO2 sources or improved environmental conditions. Ultisols, in contrast, experienced a rapid increase in NO2 emissions at the beginning of the year, only to return to levels comparable to the previous year. Oxisols showed a decrease in NO2 emissions compared to the previous year, suggesting a potential positive environmental development. Andisols displayed a slight decrease in NO2 emissions, indicating some improvement in air quality within these soils. Vertisols exhibited a moderate decrease in NO2 emissions, potentially reflecting environmental changes influencing this soil type during the specified period.

The analysis of NO2 emissions from different soil orders over the year's period from September 2018 to September 2019 indicates important trends. Mollisols consistently showed the greatest mean emissions among the examined soil orders at 0.000075, while Oxisols consistently showed the lowest mean emissions at 0.000043. The results show that, on average, with some slight fluctuations, NO2 emissions were steady during this time. The emissions of Entisols, Inceptisols, Alfisols, Andisols, and Vertisols all fell within a small range, whereas Ultisols showed more noticeable variability, peaking at 0.000108. This data indicates that soil types have a substantial impact on NO2 emissions, with Mollisols consistently making the largest contribution of the emissions throughout this year.
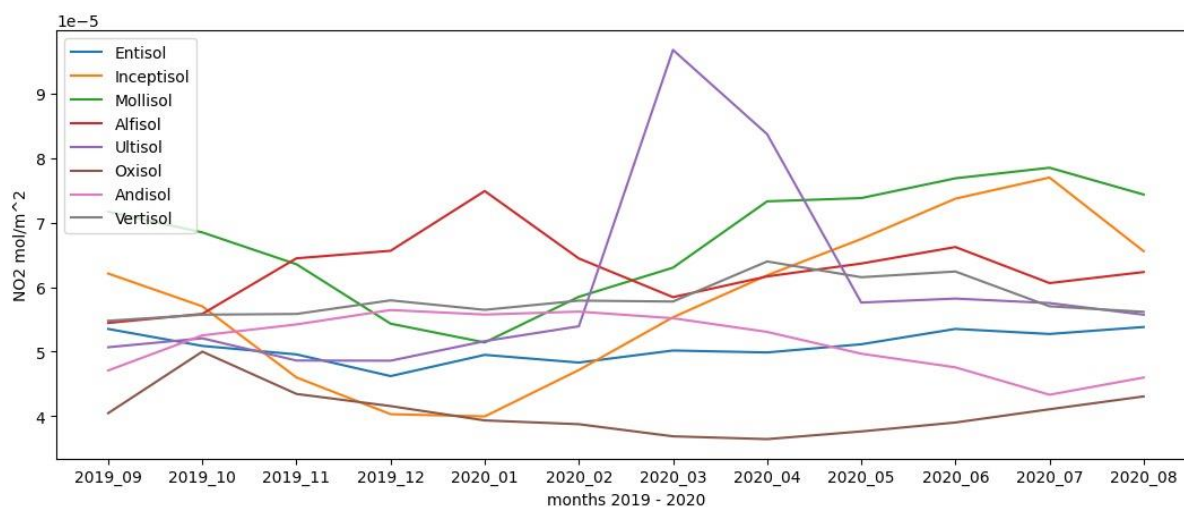
## 3.7.2  2019-2020



*Figure 8- Soil level emissions between 2019-2020*

*Table 8- Statistics of NO2 emission for the year 2019-2020*

| Soil order | Mean | Median | Std Deviation | Min | 25th percentile | 50th percentile | 75th percentile | Max | Range |
|---|---|---|---|---|---|---|---|---|---|
| Entisol | 0.000051 | 0.000051 | 0.000002 | 0.000046 | 0.00005 | 0.000051 | 0.000053 | 0.000054 | 0.000008 |
| Inceptisol | 0.000058 | 0.000059 | 0.000012 | 0.00004 | 0.000047 | 0.000059 | 0.000066 | 0.000077 | 0.000037 |
| Mollisol | 0.000067 | 0.00007 | 0.000009 | 0.000051 | 0.000062 | 0.00007 | 0.000074 | 0.000079 | 0.000027 |
| Alfisol | 0.000063 | 0.000063 | 0.000005 | 0.000054 | 0.00006 | 0.000063 | 0.000065 | 0.000075 | 0.00002 |
| Ultisol | 0.00006 | 0.000055 | 0.000015 | 0.000049 | 0.000051 | 0.000055 | 0.000058 | 0.000097 | 0.000048 |
| Oxisol | 0.000041 | 0.00004 | 0.000004 | 0.000036 | 0.000038 | 0.00004 | 0.000042 | 0.00005 | 0.000014 |
| Andisol | 0.000051 | 0.000053 | 0.000005 | 0.000043 | 0.000047 | 0.000053 | 0.000055 | 0.000056 | 0.000013 |
| Vertisol | 0.000058 | 0.000057 | 0.000003 | 0.000055 | 0.000056 | 0.000057 | 0.000059 | 0.000064 | 0.000009 |

In the year 2019-2020, as depicted in the (Table 8- Statistics of NO2 emission for the year 2019-2020) the NO2 emissions within different soil orders showed distinct patterns. Entisols remained relatively stable in terms of NO2 emissions, indicating consistent levels over this period. Mollisols, on the other hand, experienced a notable increase in NO2 emissions, suggesting a potential environmental shift in these soils. Inceptisols exhibited a fluctuating pattern, with a decrease towards the end of the year followed by a slight increase in NO2 emissions. Alfisols continued their modest upward trend in NO2 emissions. Ultisols displayed a slight increase as well, indicating a gradual rise in NO2 levels. Oxisols, in contrast, saw a decrease in NO2 emissions compared to the previous year. Andisols remained stable, with no significant changes. Vertisols, similar to Mollisols, exhibited a further increase in NO2 emissions, suggesting dynamic environmental factors influencing these soil types during this

timeframe. Notably we can notice a greater than 10% drop in emissions during this period. This likely correlates with the COVID-19 lockdown which resulted in an economic slowdown.

The analysis of NO2 emissions from various soil orders for the period of September 2019 to September 2020 reveals some interesting patterns. Mollisols consistently showed the greatest mean emissions among the soil orders under study, at 0.000067, while Oxisols consistently showed the lowest mean emissions, at 0.000041. Notably, NO2 emissions showed a general upward trend throughout all soil categories, with certain months indicating significant increases, especially in Ultisols and Mollisols. The widest range of emissions was shown by ultisols, which peaked in March 2020 at 0.000097.
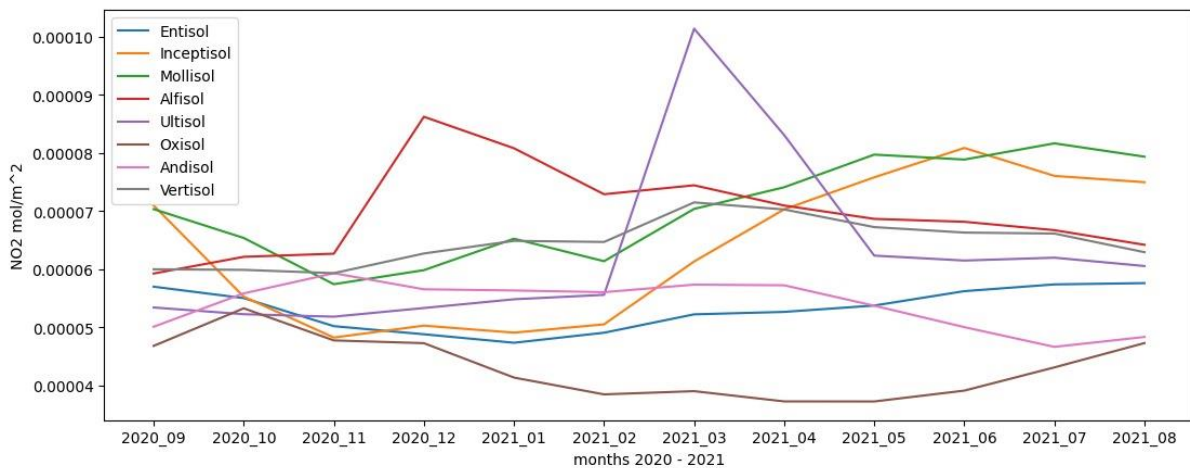
### 3.7.3   2020-2021



*Figure 9- Soil level emissions between 2020-2021*

*Table 9- Statistics of NO2 emission for the year 2020-2021*

| Soil order | Mean | Median | Std Deviation | Min | 25th percentile | 50th percentile | 75th percentile | Max | Range |
|---|---|---|---|---|---|---|---|---|---|
| Entisol | 0.000053 | 0.000053 | 0.000004 | 0.000047 | 0.00005 | 0.000053 | 0.000056 | 0.000058 | 0.00001 |
| Inceptisol | 0.000064 | 0.000066 | 0.000012 | 0.000048 | 0.00005 | 0.000066 | 0.000075 | 0.000081 | 0.000033 |
| Mollisol | 0.00007 | 0.00007 | 0.000008 | 0.000057 | 0.000064 | 0.00007 | 0.000079 | 0.000082 | 0.000024 |
| Alfisol | 0.00007 | 0.000068 | 0.000008 | 0.000059 | 0.000064 | 0.000068 | 0.000073 | 0.000086 | 0.000027 |
| Ultisol | 0.000063 | 0.000058 | 0.000015 | 0.000052 | 0.000053 | 0.000058 | 0.000062 | 0.000101 | 0.00005 |
| Oxisol | 0.000043 | 0.000042 | 0.000005 | 0.000037 | 0.000039 | 0.000042 | 0.000047 | 0.000053 | 0.000016 |
| Andisol | 0.000054 | 0.000056 | 0.000004 | 0.000047 | 0.00005 | 0.000056 | 0.000057 | 0.000059 | 0.000013 |
| Vertisol | 0.000065 | 0.000065 | 0.000004 | 0.000059 | 0.000062 | 0.000065 | 0.000067 | 0.000071 | 0.000012 |

In the year 2020-2021, as depicted in the (Table 9- Statistics of NO2 emission for the year 2020-2021) the NO2 emissions within different soil orders exhibited diverse patterns. Entisols remained stable,

showing consistent NO2 levels throughout the year. Mollisols displayed an increase in NO2 emissions, suggesting a potential environmental shift affecting these soils. Inceptisols followed a pattern of declining NO2 emissions towards the end of the year, followed by a subsequent increase, ultimately returning to levels similar to the previous year. Alfisols exhibited a modest decrease in NO2 emissions, indicating a possible reduction in NO2 sources or improved environmental conditions. Ultisols, in contrast, experienced a rapid increase in NO2 emissions at the beginning of the year, with an overall increase compared to the previous year, potentially indicating changing environmental factors. Oxisols showed a decrease in NO2 emissions initially, returning to levels similar to the previous year, suggesting some degree of stabilization. Andisols displayed a slight decrease in NO2 emissions, followed by a return to the previous year's levels, indicating stability in air quality within these soils. Vertisols exhibited a moderate increase in NO2 emissions, reflecting potential environmental changes influencing this soil type during the specified time period.

Analyzing NO2 emissions from various soil types between September 2020 and September 2021 reveals interesting details. The highest mean emissions were continuously found in mollisols, at 0.000070, indicating a considerable contribution to NO2 emissions. Oxisols, on the other hand, consistently showed the lowest mean emissions at 0.000043. The data indicates an overall increased trend in NO2 emissions for all soil types, with some months, especially in March 2021, displaying notable jumps, especially in Mollisols and Ultisols. The largest range of emissions, with a maximum of 0.000101, was shown by ultisols.
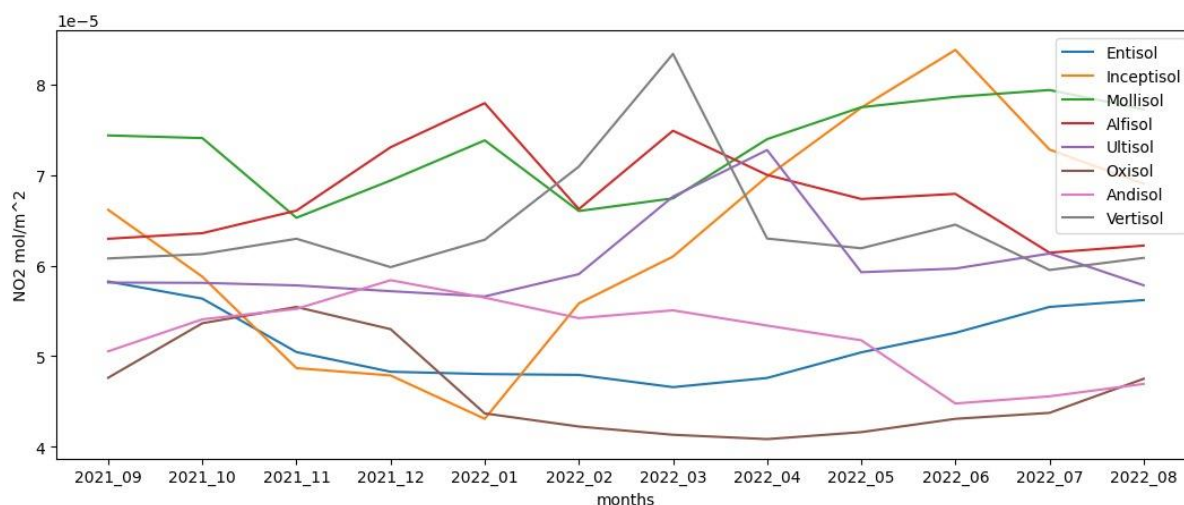
### 3.7.4  2021-2022



*Figure 10- Soil level emissions between 2021-2022*

*Table 10- Statistics of NO2 emission for the year 2021-2022*

| Soil order | Mean | Median | Std Deviation | Min | 25th percentile | 50th percentile | 75th percentile | Max | Range |
|---|---|---|---|---|---|---|---|---|---|
| Entisol | 0.000052 | 0.00005 | 0.000004 | 0.000047 | 0.000048 | 0.00005 | 0.000056 | 0.000058 | 0.000012 |
| Inceptisol | 0.000063 | 0.000064 | 0.000013 | 0.000043 | 0.000054 | 0.000064 | 0.000071 | 0.000084 | 0.000041 |
| Mollisol | 0.000073 | 0.000074 | 0.000005 | 0.000065 | 0.000069 | 0.000074 | 0.000077 | 0.000079 | 0.000014 |
| Alfisol | 0.000068 | 0.000067 | 0.000005 | 0.000061 | 0.000063 | 0.000067 | 0.000071 | 0.000078 | 0.000017 |
| Ultisol | 0.00006 | 0.000059 | 0.000005 | 0.000057 | 0.000058 | 0.000059 | 0.000049 | 0.000055 | 0.000015 |
| Oxisol | 0.000046 | 0.000044 | 0.000005 | 0.000041 | 0.000042 | 0.000044 | 0.000049 | 0.000055 | 0.000015 |
| Andisol | 0.000052 | 0.000054 | 0.000004 | 0.000045 | 0.00005 | 0.000054 | 0.000055 | 0.000058 | 0.000014 |
| Vertisol | 0.000064 | 0.000062 | 0.000007 | 0.00006 | 0.000061 | 0.000062 | 0.000063 | 0.000083 | 0.000024 |

In the year 2021-2022, as depicted in (Table 10- Statistics of NO2 emission for the year 2021-2022), the NO2 emissions within different soil orders displayed distinctive trends. Entisols showed a slight decrease in NO2 emissions. Mollisols, on the other hand, exhibited an increase in NO2 emissions, indicating a potential shift in environmental conditions affecting these soil types. Inceptisols followed a pattern of declining NO2 emissions towards the end of the year, followed by a rapid increase slightly above the previous year's levels, suggesting some fluctuations in air quality. Alfisols remained stable, with consistent NO2 levels. Ultisols showed an increase at the beginning of the year, leveling off to the previous year's levels, indicating fluctuations in NO2 emissions. Oxisols displayed a decrease in NO2 emissions initially, returning to levels similar to the previous year, suggesting some degree of stabilization. Andisols, however, showed a significant drop in NO2 emissions compared to the previous year, indicating potential improvements in air quality within these soils. Vertisols exhibited a rapid

increase in NO2 emissions at the beginning of the year, followed by a return to the previous year's levels, suggesting dynamic environmental changes influencing this soil type during the specified time period.

Analyzing NO2 emissions from different soil orders from September 2021 to September 2022 reveals interesting trends. The fact that mollisols continuously had the greatest mean emissions, at 0.000073, shows how much they contribute to NO2 emissions. Oxisols, on the other hand, consistently had the lowest mean emissions, at 0.000046. According to the data, NO2 emissions in all soil categories follow a predictable seasonal pattern with wintertime peaks (December and January) and summers lows (July and August). Emissions considerably rose, especially in March 2022, primarily as a result of Mollisols and Vertisol. The widest range of emissions was seen in ultisols, which had a maximum emission of 0.000084.
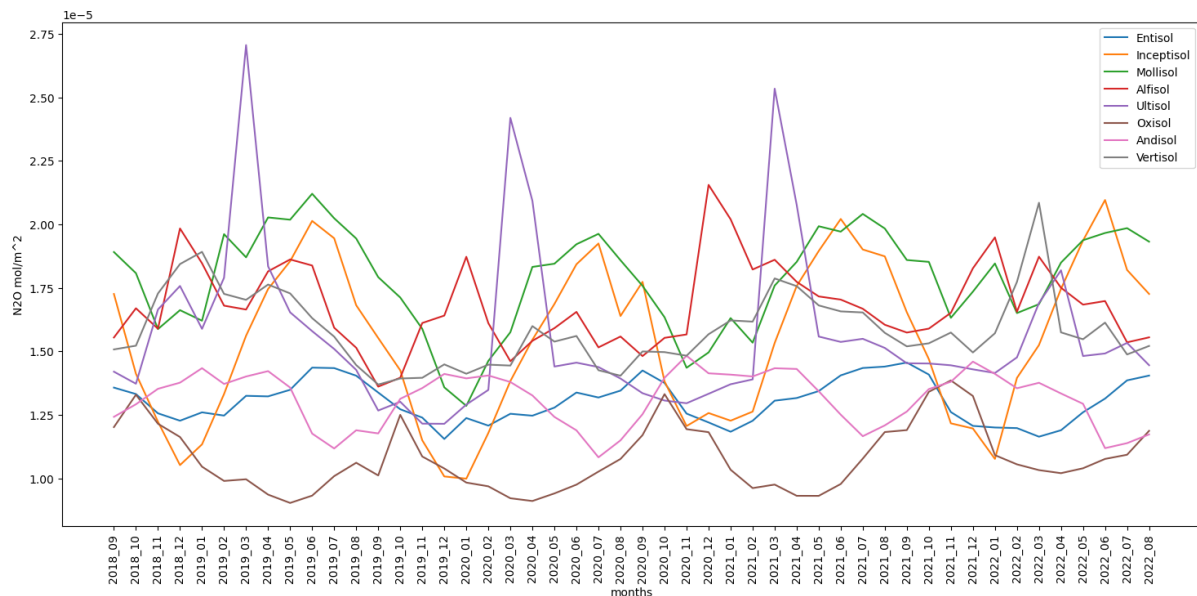
### 3.7.5 2018-2022



*Figure 11- Soil level emissions between 2018-2022*

*Table 11- Statistics of NO2 emission for the year 2018-2022*

| Soil order | Mean | Median | Std Deviation | Min | 25th percentile | 50th percentile | 75th percentile | Max | Range |
|---|---|---|---|---|---|---|---|---|---|
| Entisol | 0.000013 | 0.000013 | 8.40E-07 | 0.000012 | 0.000012 | 0.000013 | 0.000014 | 0.000015 | 0.000003 |
| Inceptisol | 0.000015 | 0.000015 | 3.10E-06 | 0.00001 | 0.000013 | 0.000015 | 0.000018 | 0.000021 | 0.000011 |
| Mollisol | 0.000018 | 0.000018 | 1.97E-06 | 0.000013 | 0.000016 | 0.000018 | 0.000019 | 0.000021 | 0.000008 |
| Alfisol | 0.000017 | 0.000017 | 1.65E-06 | 0.000014 | 0.000016 | 0.000017 | 0.000018 | 0.000022 | 0.000008 |
| Ultisol | 0.000016 | 0.000015 | 3.22E-06 | 0.000012 | 0.000014 | 0.000015 | 0.000016 | 0.000027 | 0.000015 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Oxisol | 0.000011 | 0.00001 | 1.29E-06 | 0.000009 | 0.00001 | 0.00001 | 0.000012 | 0.000014 | 0.000005 |
| Andisol | 0.000013 | 0.000014 | 1.07E-06 | 0.000011 | 0.000012 | 0.000014 | 0.000014 | 0.000015 | 0.000004 |
| Vertisol | 0.000016 | 0.000016 | 1.46E-06 | 0.000014 | 0.000015 | 0.000016 | 0.000017 | 0.000021 | 0.000007 |

Over the course of the four-year period from 2018 to 2022, (as depicted in Table 11- Statistics of NO2 emission for the year 2018-2022) the NO2 emissions within different soil orders exhibited a range of patterns and trends. In 2018-2019, Entisols remained stable, while Mollisols showed a slight increase. Inceptisols followed a fluctuating pattern, and Alfisols exhibited a modest decrease. Ultisols experienced a rapid increase at the beginning of the year, and Oxisols showed a decrease. Andisols displayed a slight decrease, and Vertisols exhibited a moderate decrease.

Moving to 2019-2020, Entisols remained stable, but Mollisols exhibited a notable increase, while Inceptisols followed a fluctuating pattern. Alfisols continued their modest upward trend, and Ultisols displayed a slight increase. Oxisols saw a decrease, Andisols remained stable, and Vertisols exhibited a further increase.

In 2020-2021, Entisols stayed stable, Mollisols displayed an increase, Inceptisols followed a fluctuating pattern, and Alfisols exhibited a modest decrease. Ultisols had a rapid increase at the beginning of the year, while Oxisols showed a decrease initially. Andisols displayed a slight decrease, and Vertisols exhibited a moderate increase. Finally, in 2021-2022, Entisols showed a slight decrease, Mollisols exhibited an increase, Inceptisols followed a fluctuating pattern, and Alfisols remained stable. Ultisols had an increase at the beginning of the year, while Oxisols showed a decrease initially. Andisols had a significant drop, and Vertisols exhibited a rapid increase.

## 3.8   Results based on a year-level

This introduction lays the groundwork for the year-level analysis we undertook with the data. We analysed the variations on a year level.
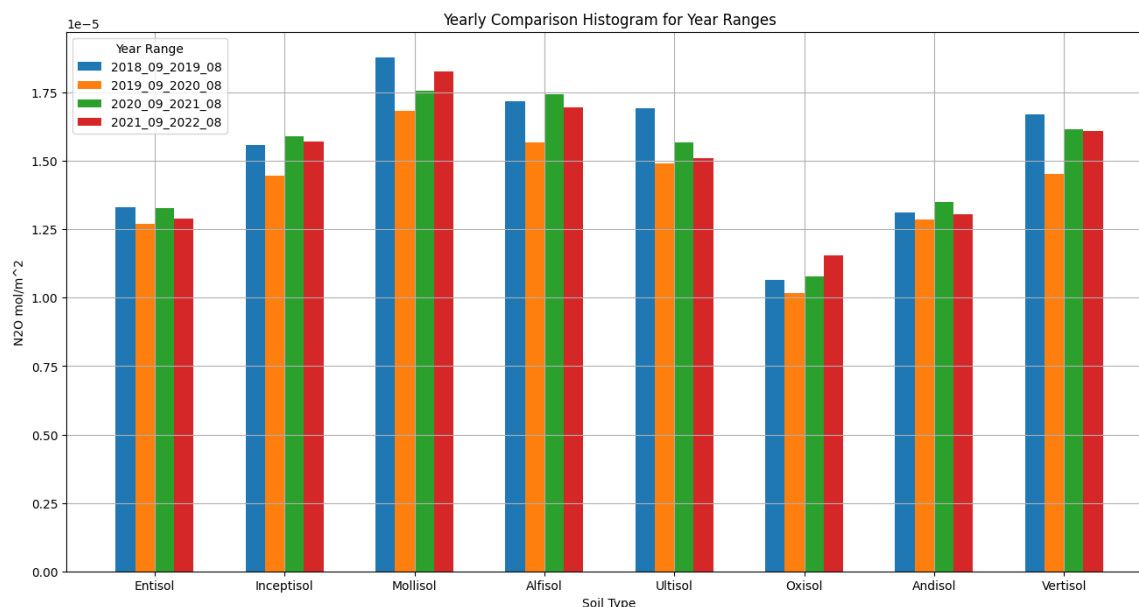
*Figure 12 - Year level analysis*

Since our data ranges from September 2018 to August 2022, we annotated a year to the period between September and August.

Cumulatively we can see mollisols and alfisols (ie the best and second best soil for agriculture and forestry) emit the highest. An interesting observation on the orange bars (denoting the COVID lockdown year) has a noticeable dip in all the soil orders.

## 3.9    Results based on equatorial distance

This analysis is based on the surveyed locations and its proximity to the equator. This serves the purpose of discerning the differences between emissions of the same soil variety in different geographical locations. To discern the equatorial locations and non-equatorial locations we selected locations with latitude in the range of -15.5 and 23.5 as equatorial locations and the locations outside the range as non-equatorial locations. Mollisols and Inceptisols where not surveyed at an equatorial location as most mollisol soils are prevalent in the non-equatorial latitudes.
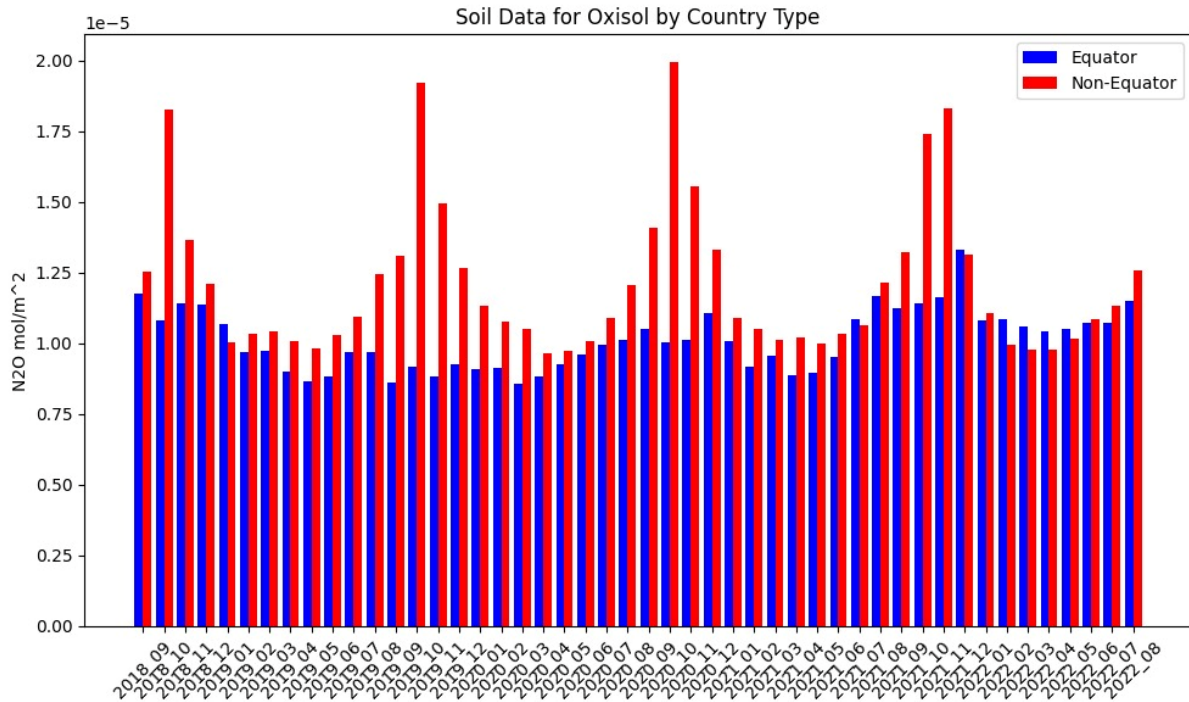
*Figure 13 - Oxisol equatorial analysis*

Oxisols exhibit a periodic pattern. They peak during the Septembers in the higher latitudes and crest during the March-April window. They don't exhibit periodic peaking patterns near the equator. They don't exhibit any significant reduction in emissions separately (ie- equator and non-equator) though cumulatively they exhibit some significance as explained in the Section 3.7.

Ultisols below (Figure 13) exhibit some periodic pattern over the 4 years. They peak around the February and March in equatorial countries and crest around December. Non equatorial countries exhibit some degree of periodic pattern where they crest around September, but not significant peaks are observed. Both country types don't exhibit any significant changes due to the COVID-19 lockdown. Equatorial region tend to show significantly less emissions than non- equatorial regions.
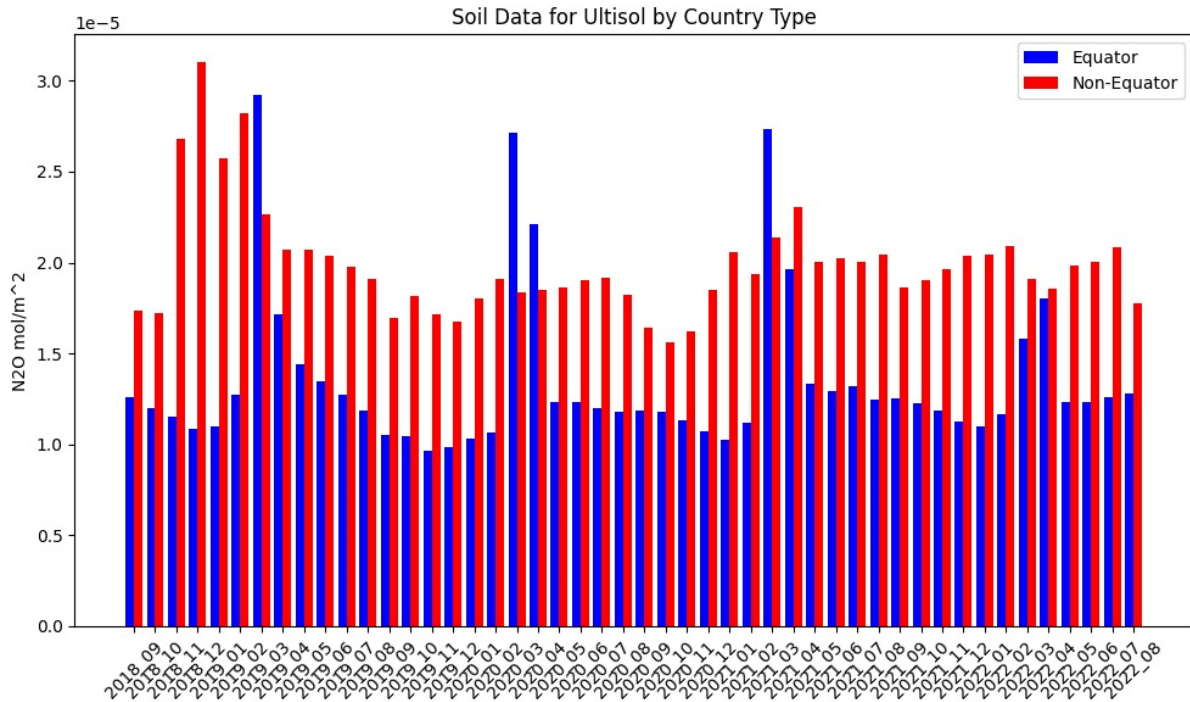
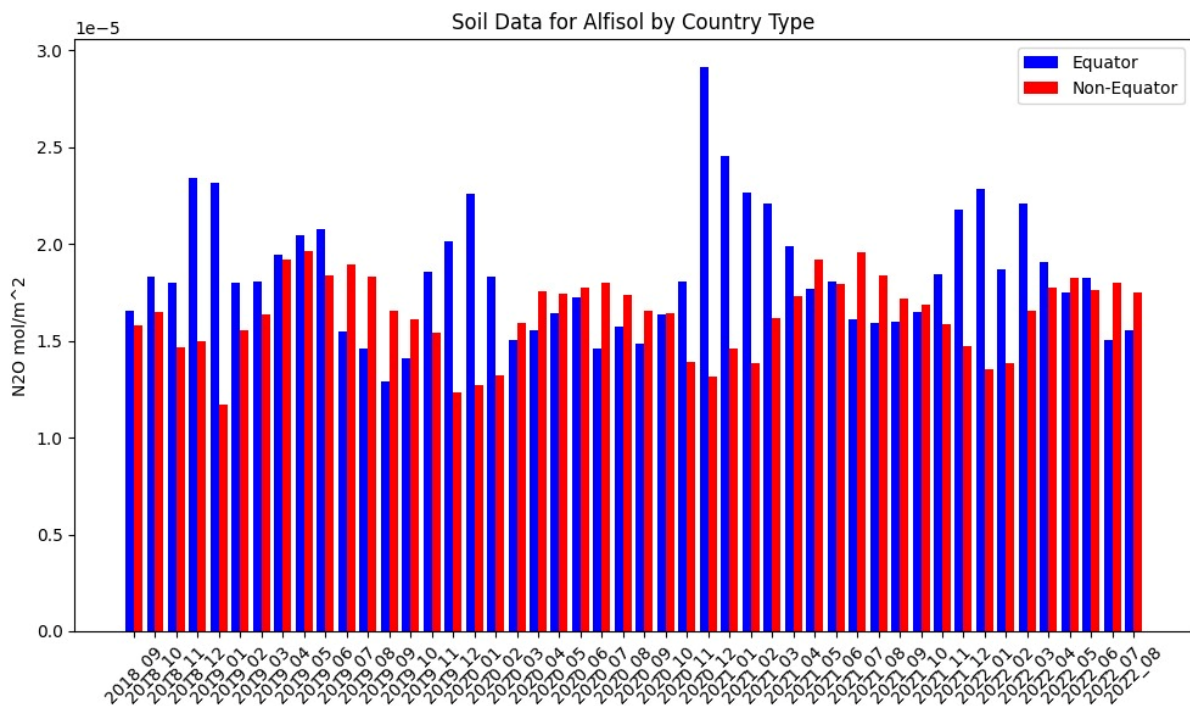*Figure 14 - Ultisol equatorial analysis*



*Figure 15 - Alfisol equatorial analysis*

Alfisols (Figure 14), Entisols (Figure 15) and Andisols (Figure 16) in non-equatorial regions show a periodic pattern, peaking in March-April window and cresting in November and vice versa.
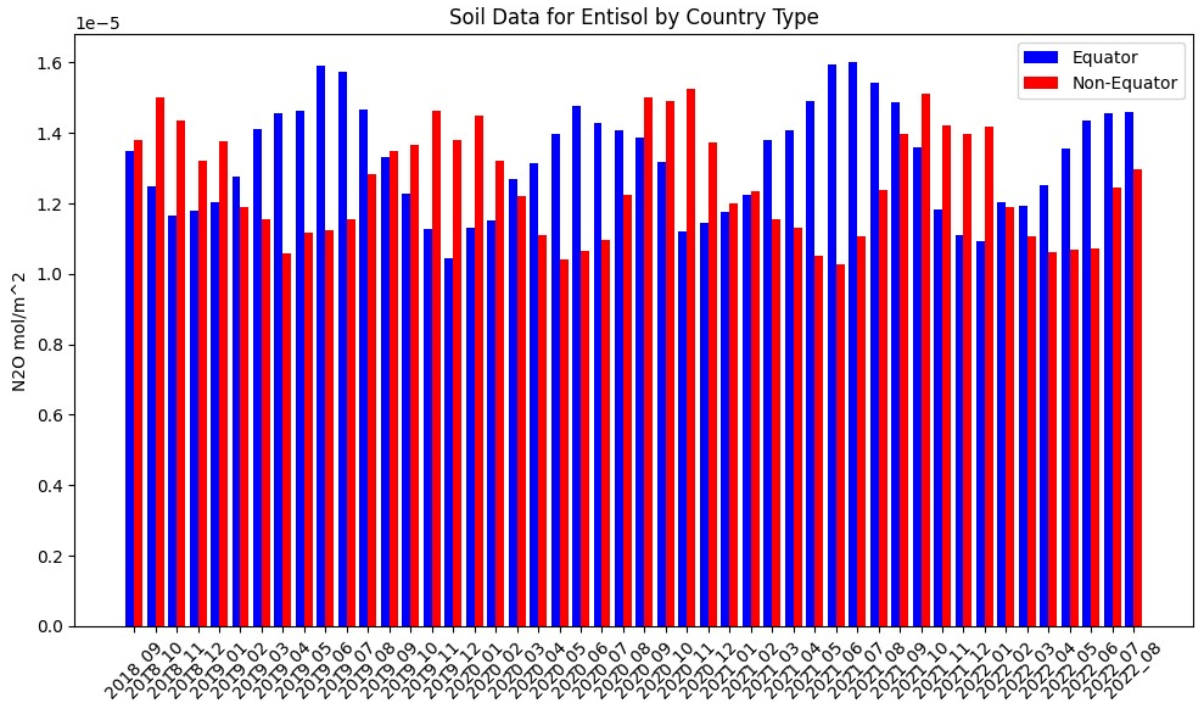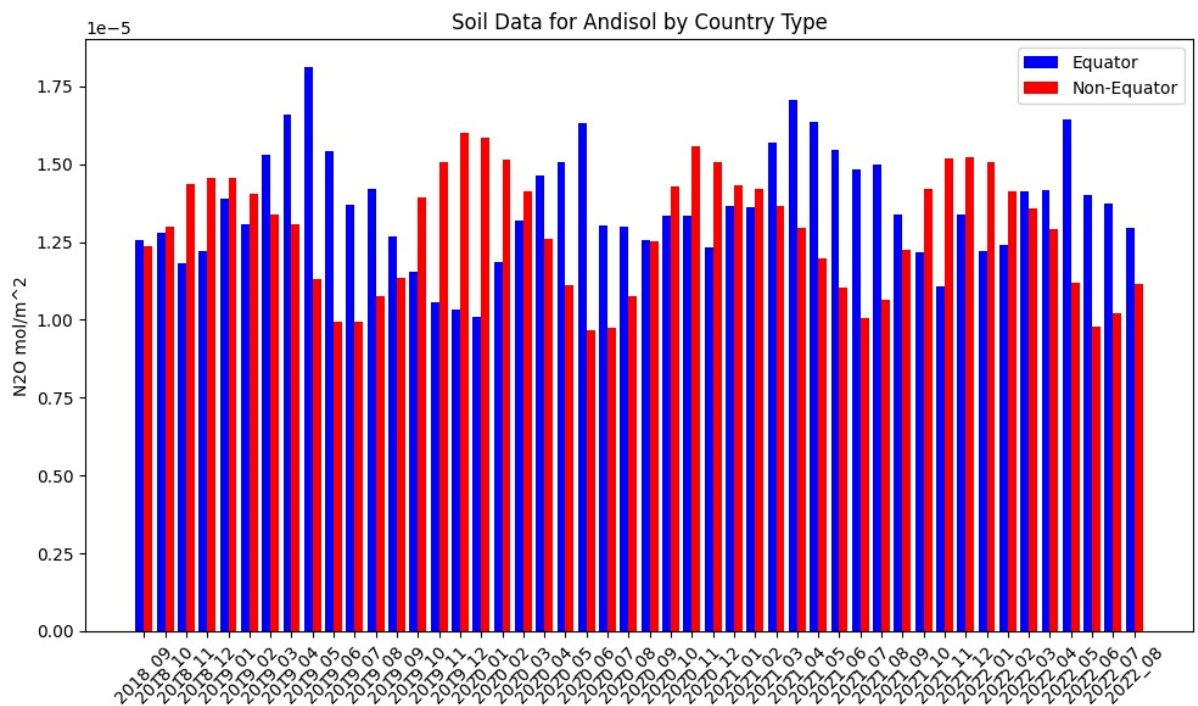
45

*Figure 16 - Entisol equatorial analysis*



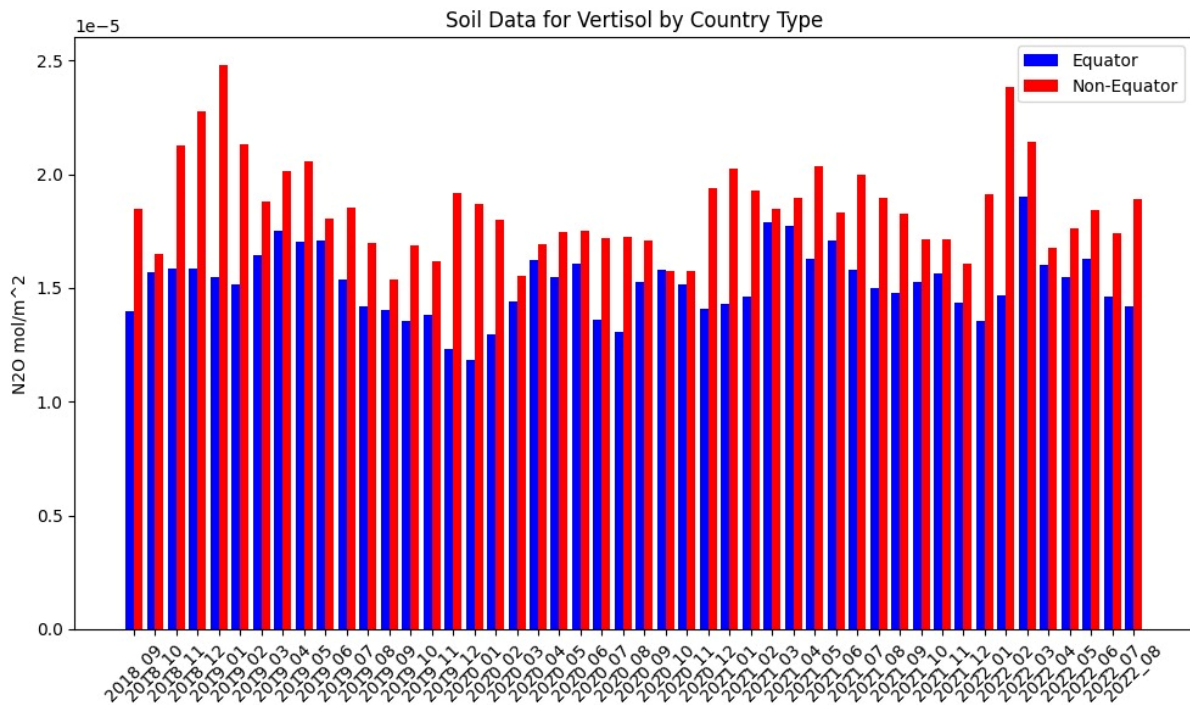*Figure 17 - Andisol equatorial analysis*

*Figure 18 - Vertisol geographical analysis*

Vertisols show significantly higher readings for the non-equatorial regions as compared to equatorial ones.

## 3.10  Chapter Summary

The data on nitrous oxide (NO2) emissions from different soil orders during a four-year period from September 2018 to September 2022 indicates significant trends and oscillations. Mollisols stood out for constantly having the greatest average emissions, demonstrating their major contribution to NO2 emissions over the course of this time. Conversely, the mean emissions of oxisols were consistently the lowest. The data also shows a cyclical seasonal trend, with wintertime emissions being higher and summertime emissions being lower. The largest range of emissions was shown by ultisols, which peaked in April 2022. Overall, these results highlight the significance of seasonal variations and soil types in NO2 emissions, emphasising the need for targeted mitigation strategies and sustainable soil management practises, especially in soil orders like Mollisols, Alfisols and Ultisols, which demonstrated both high mean emissions and significant variability.

In summary, different soil types exhibit various emission patterns, with some showing periodic peaks in emissions during specific months. The impact of the COVID-19 lockdown on emissions is not significant in most soil types, and in the locations the data was collected, the equatorial regions generally exhibited lower emissions compared to non-equatorial regions.

Across these years, it's apparent that different soil orders responded differently to NO2 emissions, with some showing stability, others fluctuating, and some indicating potential environmental shifts.

Data Availability: (Data link in Data.1 in the Code and Data section)

# Chapter 04: Machine Learning Model and Results

## 4.1 Chapter Overview

The focal point of our study, analysis, and prediction efforts is Chapter 4. It demonstrates the specifics of the computational methods utilised to get important conclusions from the methodically acquired data. Readers will have a thorough understanding of the models, their functionality, and the insights they offer by the end of this chapter, laying the groundwork for the conclusions and suggestions covered in the following chapter. It further discusses the results of the machine learning models and the evaluation metrics. The associated codes and data are provided in the links in the "Code and Data" section.

## 4.2 Data Preprocessing

The preprocessing steps for the inputs with their respective shapes are crucial for preparing the data for analysis or machine learning tasks. Let's break down the preprocessing steps for each of these tensors:

1. **Images (images.shape: (3506, 200, 200, 3)):**

   - **Resizing:** The images have been resized to a shape of (200, 200, 3). The original images came in different dimensions and resizing them to a consistent shape is essential for compatibility and efficient processing. The (200, 200) dimensions represent the width and height of the images, and the '3' indicates that these are color images with three color channels (typically Red, Green, and Blue).

   - **Data Type Conversion:** The image data are also converted to a specific data type, such as float32, depending on the requirements of the PyTorch framework.

2. **NO2 Readings (no2_reading.shape: (3506, 1)):**

   - **Feature Selection:** In this case, only a single feature, NO2 readings, is retained for analysis or modeling. The shape (3506, 1) indicates that there are 3506 data samples, each with one NO2 reading.

   - **Scale:** The mol/m.sq is converted to nano mol/m.sq for better convergence as extremely small numbers take time for the back prop.

3. **Dates Readings (dates_reading.shape: (3506, 3)):**

   - **Feature Engineering:** The dates_readings tensor has a shape of (3506, 3), it contains three date-related features. Date-related features could include year, month, and day, for example. These features are extracted from the original date/time information for each data sample.

# 4.3   Model Architecture

We conducted our experiments the dataset, comparing the performance of two different models. The first model was custom-built from the ground up, with its architecture designed from scratch. The second model, on the other hand, was based on ResNet, a well-known convolutional neural network architecture used in transfer learning. A diagrammatic description of the two models are provided below in Figure 19 - Model architecture diagram.
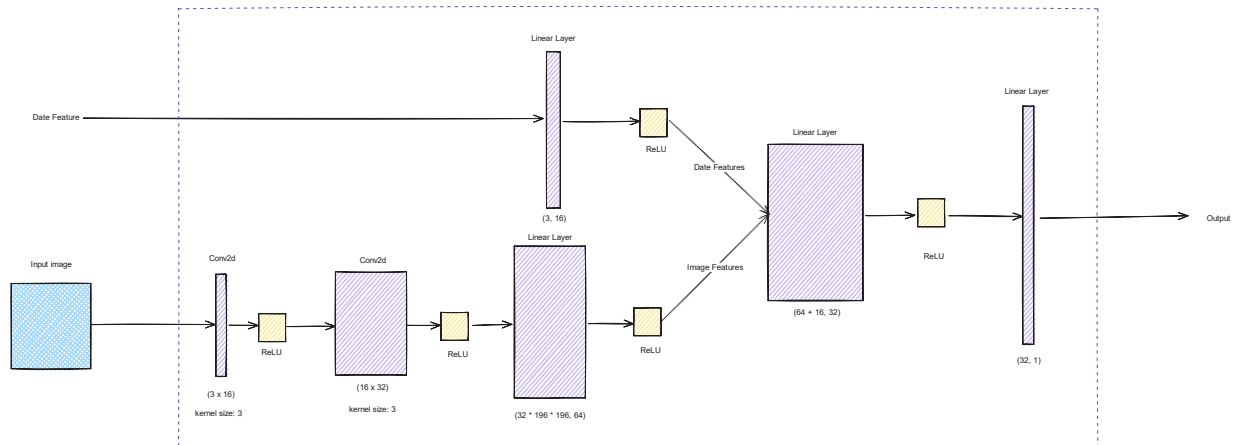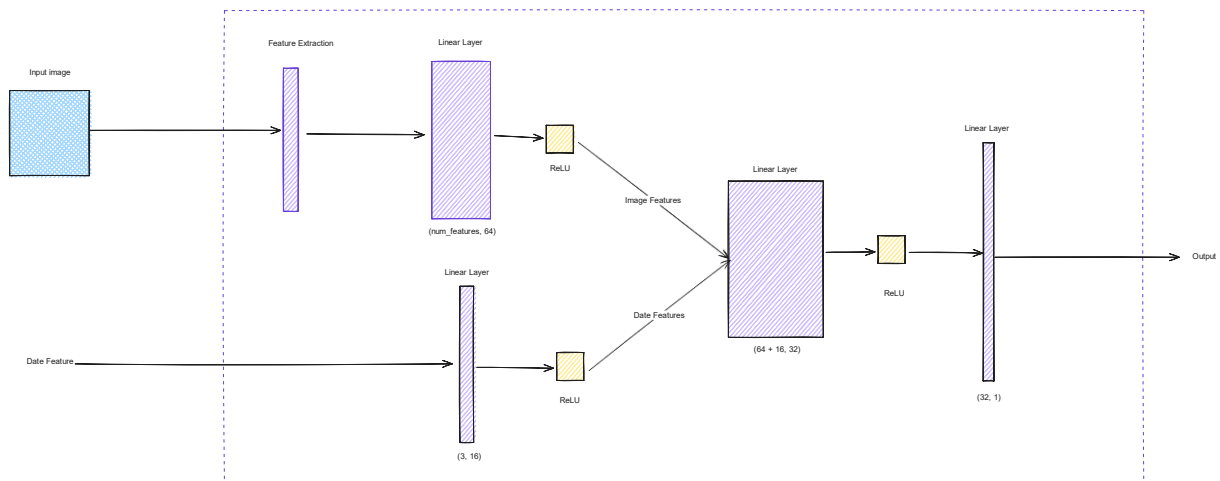


*Figure 19 - Model architecture diagram*

### 4.3.1  Custom Model

This model combines two branches: one for processing satellite image data and another for processing date-related features.

**Model Architecture:**

1. **Image Branch:**

   - The image branch is designed to process image data.

   - **image conv1** is the first convolutional layer with 3 input channels (assuming RGB images) and 16 output channels. It uses a 3x3 kernel for feature extraction.

   - **image relu1** is the ReLU activation function applied after the first convolution.

   - **image conv2** is the second convolutional layer, which takes the output of the first convolutional layer and further extracts features. It has 16 input channels and 32 output channels with 3x3 kernels.

   - **image relu2** is the ReLU activation following the second convolution.

   - **image fc** is a fully connected (linear) layer that takes the flattened output from the convolutional layers and maps it to a 64-dimensional feature space.

   - **Image relu3** is another ReLU activation function applied to the output of the fully connected layer.

2. **Date Branch:**

   - The date branch is responsible for processing date-related features.

   - **date fc** is a fully connected layer that takes 3 date-related features (e.g., year, month, day) and maps them to a 16-dimensional feature space.

   - **date relu** is the ReLU activation applied after the date feature layer.

3. **Combined Layers:**

   - **fc combined1** is a fully connected layer that takes the concatenated output from both the image and date branches. It combines the information from the image and date features and maps it to a 32-dimensional feature space.

   - **fc combined2** is the final fully connected layer that maps the 32-dimensional feature vector to a single output, which is the regression output.

   - **fc relu** is a ReLU activation applied after the combined layers.

In the **forward** method, the model performs the following steps. The image data is passed through the image branch, going through convolutional and fully connected layers, and then ReLU activations. The date-related features are processed through the date branch, which consists of a single fully connected layer and a ReLU activation. The outputs of the image and date branches are concatenated along the feature dimension (dimension 1). The concatenated features are passed through the combined fully

connected layers. The final output is obtained, which represents the NO emissions in (nano mol per sq m) regression prediction.

This model is designed for regression tasks where the goal is to predict nitrous emissions based on input satellite image data and image date. It combines information from both branches to make predictions. Given in the Appendix A is the computation graph. (Code in Code.1 in the Code and Data section)

## 4.3.2  ResNet Model

This model leverages transfer learning by using a pre-trained ResNet-18 (He et al., 2015) architecture to extract features from input images. It combines these image features with date information and uses fully connected layers to make a regression prediction. The choice of using transfer learning can be beneficial when dealing with image data, as it allows the model to leverage the knowledge learned from a large dataset (ImageNet) and adapt it to the specific regression task at hand. Given in the Appendix B is the computation graph. (Code in Code.2 in the Code and Data section)

*Table 12 - model details*

| Parameter/Hyperparameter | Model 1 (Custom CNN) | Model 2 (Transfer Learning with ResNet) |
|---|---|---|
| **Image Branch** | | |
| Input Channels | 3 (assuming RGB images) | 3 (assuming RGB images) |
| Kernel Size (Conv Layers 1 & 2) | 3x3 | Defined by pre-trained ResNet model |
| Number of Filters (Conv Layer 1) | 16 | Defined by pre-trained ResNet model |
| Number of Filters (Conv Layer 2) | 32 | Defined by pre-trained ResNet model |
| Activation Function (Conv Layers) | ReLU | Defined by pre-trained ResNet model |
| Fully Connected Layer Size | 32 * 196 * 196 (calculated based on input size) | Defined by pre-trained ResNet model |
| Activation Function (FC Layer) | ReLU | Defined by pre-trained ResNet model |
| **Date Branch** | | |
| Input Features | 3 | 3 |
| Hidden Layer Size | 16 | 16 |
| Activation Function (FC Layer) | ReLU | ReLU |
| **Combined Layers** | | |
| Input Features (Combined) | 64 (image) + 16 (date) | 64 (image) + 16 (date) |
| Hidden Layer 1 Size | 32 | 32 |
| Hidden Layer 2 Size (Output) | 1 | 1 |
| Activation Function (FC Layers) | ReLU | ReLU |

51

| | | |
|---|---|---|
| **Learning Rate** | 0.0001 | 0.0001 |
| **Batch Size** | 16 | 16 |
| **Optimizer** | RMSprop | RMSprop |
| **Epochs** | 200 | 200 |

## 4.4   Model Training

This model appears to be a custom neural network architecture designed for regression tasks. It's important to note that the model is explicitly moved to the computing device (either CPU or GPU) specified earlier. This step ensures that all subsequent computations involving the model will be performed on the chosen hardware. For these experiments all the codes were run on Google Colab instances. Most of the training and prototyping was completed using the free T4 GPU instance. However, the final training was done on a Colab Pro with A100 GPU.

Two loss functions are defined for training: Mean Squared Error (MSE) and L1 Loss. These loss functions are fundamental in training regression models, as they quantify the difference between predicted outputs and actual target readings. MSE emphasizes larger errors by taking squared differences, while L1 loss treats all errors uniformly by considering absolute differences.

The optimizer is RMSprop optimizer, which is set up to manage the model's parameter updates during training. It employs a small learning rate of 0.0001, which controls the step size in the optimization process. The training loop iterates for a predefined number of epochs (200 epochs in this instance). Within each epoch, the code processes the training data in batches)

## 4.5   Loss functions

For the model where we have two inputs, satellite images (**x1**) and image date (**x2**), and one output, **y**, we defined a custom loss function. We define the loss function mathematically as follows:

Let $y_{\mathrm{pred}}$ be the predicted output the model, which is a function of the x1 and x2

i.e. $y_{\mathrm{pred}} = f(x_1, x_2)$

Where:

- N is the total number of training samples.
- $y_{\mathrm{pred,i}}$ is the predicted output for the i-th sample
- $y_{\mathrm{true,i}}$ is the predicted output for the i-th sample

The loss function can be defined as the error between the predicted and true outputs. A common choice is to use Mean Squared Error (MSE), which is the squared difference between $y_{true}$ and $y_{\mathrm{pred}}$ averaged over the dataset.

$$\text{Loss (MSE)} = \frac{1}{N} \sum_{i=1}^{N} \left( y_{\text{true},i} - y_{\text{pred},i} \right)^2$$

Mean Absolute Error loss function measures the average of the absolute differences between the predicted and true values.

$$\text{Loss (MAE) or L1} = \frac{1}{N} \sum_{i=1}^{N} \left| y_{\text{true},i} - y_{\text{pred},i} \right|$$

We use a combined loss that balances the contributions of both MSE and L1 losses for both outputs. Here, α and β are hyperparameters that control the weight of each loss term in the combined loss.

$$\text{Combined Loss} = \alpha \cdot \text{MSE Loss}\left(Y_{\text{true}}, Y_{\text{pred}}\right) + \beta \cdot \text{L1 Loss}\left(Y_{\text{true}}, Y_{\text{pred}}\right)$$

We gave both α and β the value of 1 to provide equal weighting in terms of loss back propagation.

## 4.6    Evaluation of results

Regression model evaluation metrics provide insight into how effectively a regression model predicts continuous values (numeric values) in comparison to the actual target values. Some standard metrics for regression model evaluation are:

1. Mean Squared Error (MSE): MSE calculates the average squared deviations between anticipated and observed values. It is determined as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left( y_{\text{true}}^{(i)} - y_{\text{pred}}^{(i)} \right)^2$$

2. Mean Absolute Error (MAE): The average absolute difference (MAE) between the predicted and actual values is measured. It is determined as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_{\text{true}}^{(i)} - y_{\text{pred}}^{(i)} \right|$$

3. R-squared (R²): R-squared measures the percentage of the target variable's variance that the model accounts for. A better fit is indicated by higher values, which range from 0 to 1. It is determined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left( y_{\text{true}}^{(i)} - y_{\text{pred}}^{(i)} \right)^2}{\sum_{i=1}^{n} \left( y_{\text{true}}^{(i)} - \overline{y_{\text{true}}} \right)^2}$$

4. Mean Absolute Percentage Error (MAPE): The average percentage difference between expected and actual values is measured by MAPE. It is determined as:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\left| y_{\text{true}}^{(i)} - y_{\text{pred}}^{(i)} \right|}{\left| y_{\text{true}}^{(i)} \right|} \right) \times 100\%$$

*Table 13- results*

| Metric | Custom Model | | ResNet Model | |
|---|---|---|---|---|
| | Training | Test | Training | Test |
| Loss | 2.6245 | 20.8887 | 8.3281 | 12.0502 |
| MSE Loss | 2.3645 | 20.2746 | 8.0767 | 11.5781 |
| MAE Loss | 0.2599 | 0.6141 | 0.2514 | 0.4721 |
| R² Score | 0.9439 | 0.1499 | 0.8089 | 0.5151 |
| MAPE | 0.0724 | 0.1596 | 0.0644 | 0.1197 |

The regression model's evaluation provides important information about how well it performs on both the training and test datasets. These measurements give a thorough knowledge of how accurately the model represents the fundamental patterns and connections in the data.

## 4.6.1  Loss

Loss represents the error or discrepancy between the model's predictions and the actual target values.

In the "Custom Model," the training loss is 2.6245, which means that, on average, the model's predictions during training are close to the actual values. In the test phase, the loss increases to 20.8887, indicating that the model's performance is less accurate on unseen data.

For the "ResNet Model," the training loss is 8.3281, while the test loss is 12.0502. These numbers suggest that the ResNet Model has a slightly higher loss during training but performs better on the test data. It can be observed in Figure 18 that the Resnet converges faster than the Custom CNN.
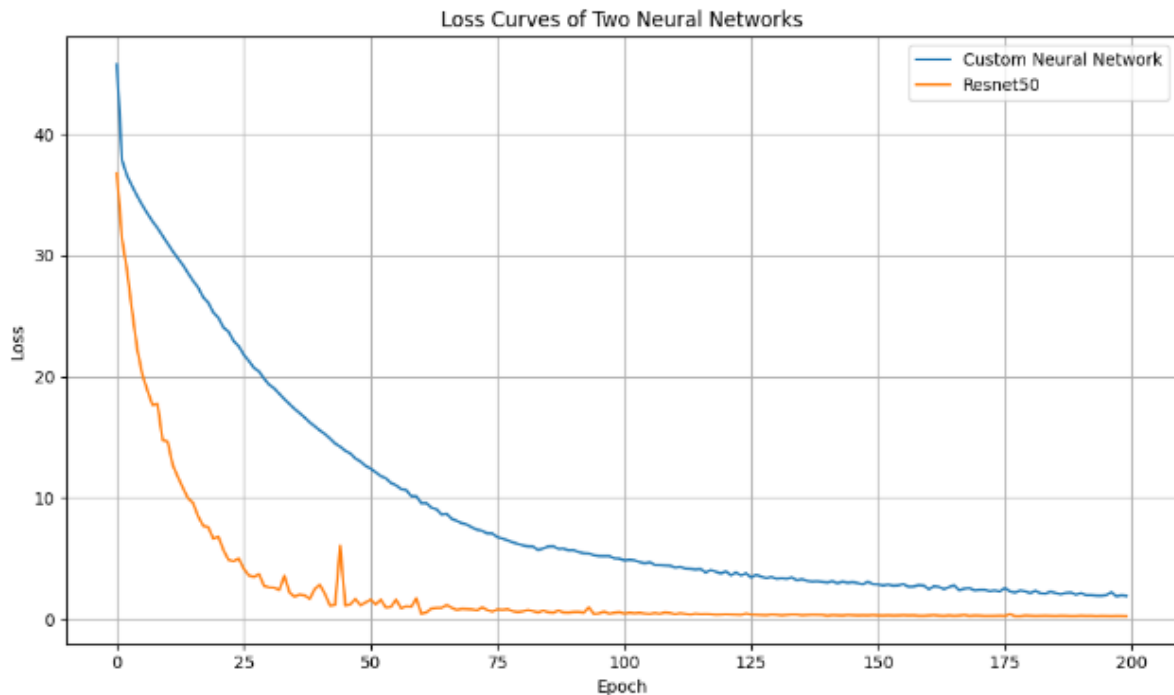


*Figure 20 - Loss curves of the two models*

54

### 4.6.2  MSE Loss (Mean Squared Error):

MSE loss measures the average of the squared differences between the predicted and actual values. It provides a measure of how far the predictions are from the actual values.

In the "Custom Model," both the training and test MSE losses are higher compared to the "ResNet Model." This implies that the "ResNet Model" has lower squared errors, indicating better performance in terms of minimizing prediction errors.

### 4.6.3  MAE Loss (Mean Absolute Error):

MAE loss measures the average of the absolute differences between the predicted and actual values. It provides a robust measure of the average prediction error.

The "Custom Model" has higher MAE losses for both training and test data compared to the "ResNet Model." This indicates that the "ResNet Model" has smaller absolute prediction errors.

### 4.6.4  R Squared Score (Coefficient of Determination):

The $R^2$ score measures the proportion of the variance in the dependent variable (target) that is predictable from the independent variables (model predictions).

In the "Custom Model," the training $R^2$ score is 0.9439, suggesting that 94.39% of the variance in the target variable can be explained by the model. However, in the test phase, the $R^2$ score drops to 0.1499, indicating that only 14.99% of the variance is explained by the model on unseen data.

For the "ResNet Model," the $R^2$ score during training is 0.8089, and it remains relatively high in the test phase at 0.5151. This suggests that the "ResNet Model" performs better in explaining the variance in the target variable in both training and testing.

### 4.6.5  MAPE (Mean Absolute Percentage Error):

MAPE measures the average percentage difference between the predicted and actual values. It provides insight into the relative prediction error.

The "Custom Model" has a MAPE of 0.0724 for training and 0.1596 for the test. In contrast, the "ResNet Model" has a lower MAPE in both training (0.0644) and test (0.1197), indicating that it produces predictions that are closer to the actual values in terms of percentage error.

## 4.7  Chapter Summary

In summary, the model performs well on the training dataset, exhibiting low loss values, a high R2 score, and a low MAPE. The "ResNet Model" generally performs better than the "Custom Model" across these metrics. It achieves lower loss values, better $R^2$ scores, and smaller absolute and percentage prediction errors, suggesting its superior performance in the regression task. Given the low MAPE of 11% on the test set, the model can perform with a good margin error on the task of filling spatio-temporal gaps in NO2 gas emission estimation for various soil varieties in diverse conditions.

# Chapter 05: Conclusions

The two primary research objectives of this project was to investigate to what extent could we be able to identify the differences in emissions rate of nitrous oxides in various soil orders from the Sentinel 5P data, and to what extent could we use the soil level data (along with the dates) with their corresponding images from the Sentinel-2 data to build and test a deep learning model that is capable of predicting emissions rates. The first research objective was answered in detail in the Chapter 3 of the thesis. The second research objective was answered in detail in the Chapter 4 of this thesis.

To achieve this, we developed a data extraction tool using the Google Earth Engine API's to download data locally. Then we determined locations for the survey. We utilized the nitrous oxide readings data into various statistical techniques and analysis, visualized the dataset to gain additional insights. We then utilized the RGB images and created a pipeline for data ingress and egress for deep learning tasks.

Predicting the presence of nitrous gas in plain view RGB images using satellite imagery is difficult because the images do not have enough variation to estimate the amount of gas in a particular image. However, we managed to get predictions within accurate to within 12% of the actual value (as tested on our test dataset). This shows that the model captures nuances in pictures that is not visible to humans.

## 5.1    Implications

The first assessment of the utilisation of satellite imaging and image date for the purpose of recognising nitrous oxide gas has been aided by this research on a soil level. The findings show the viability of predicting the presence of nitrous gases, a task that was previously not possible with satellite imagery but could have significant effects if improved to address some of the practical difficulties of using remote sensing data to monitor air quality projects. This project's first objective was to determine whether and how fluctuations in nitrous oxide gas emissions on a soil level would be possible. The study's finding that soil type plays a significant effect in emissions was significant.

This dissertation's secondary goal was to assess whether high-resolution satellite imagery and the image's related date were appropriate for calculating emissions via a deep learning neural network. We found out that our model's predictions were within 12% of the actual value. This would help in a variety of ways such as building predictive models and ecosystems.

The prevalence of these 2 research artifacts would help with planning, monitoring, and managing agricultural resources as well as educating decision-makers on crucial environmental issues including programmes to mitigate climate change will benefit from the knowledge from this study.

## 5.2    Future Work and Limitations

There are numerous potential directions for further work addressing some limitations in our research.

First and foremost, improving data diversity and quality should be the next goal. The dataset's capacity for generalisation can be increased by adding more high-quality samples. Additionally, if there are data

imbalances, determining their causes and putting data balancing procedures into practise may improve performance. In this case we don't have data imbalances (for the context that is regression), but it is possible that in future when authors collect data it is likely that the collected data may contain imbalances during the collection process.

Second, significant consideration must be given to model complexity and hyperparameter adjustment. It is important to assess the model's architecture and complexity to make sure it is appropriate for the task at hand. To avoid overfitting, advanced regularisation techniques can be used, and hyperparameters should be consistently adjusted to improve model performance.

Third, using transfer learning is a worthwhile direction to pursue. For the objective, pre-trained models or characteristics from models trained on larger and more varied datasets can be adjusted and fine-tuned. This could speed up model training and possibly produce better outcomes. There are numerous potential directions for further work.

Finally, the satellite photos are resized to 200 x 200 pixels, but the images themselves are gaining access to higher-compute power resources might greatly improve the model's ability to grasp subtleties and features in the data. The limits of the current model can be solved, and its performance on the test dataset can be enhanced, by pursuing these new lines of inquiry.

# Code and Data

Code1:

AgNES-Network/SentinelRegression_CNN.ipynb at main · Ayoob7/AgNES-Network (github.com)

Code2:

AgNES-Network/SentinelRegression_ResNet.ipynb at main · Ayoob7/AgNES-Network (github.com)

Data1:

https://drive.google.com/drive/folders/1cd9Y3c4Y8Quto0_efS_M88iKgy6FoSdE?usp=sharing

Data Extraction Tool.5P:

AgNES-Network/Sentinel5P_DataExtraction.ipynb at main · Ayoob7/AgNES-Network (github.com)

Data Extraction Tool.2:

AgNES-Network/Sentinel2_DataExtraction.ipynb at main · Ayoob7/AgNES-Network (github.com)
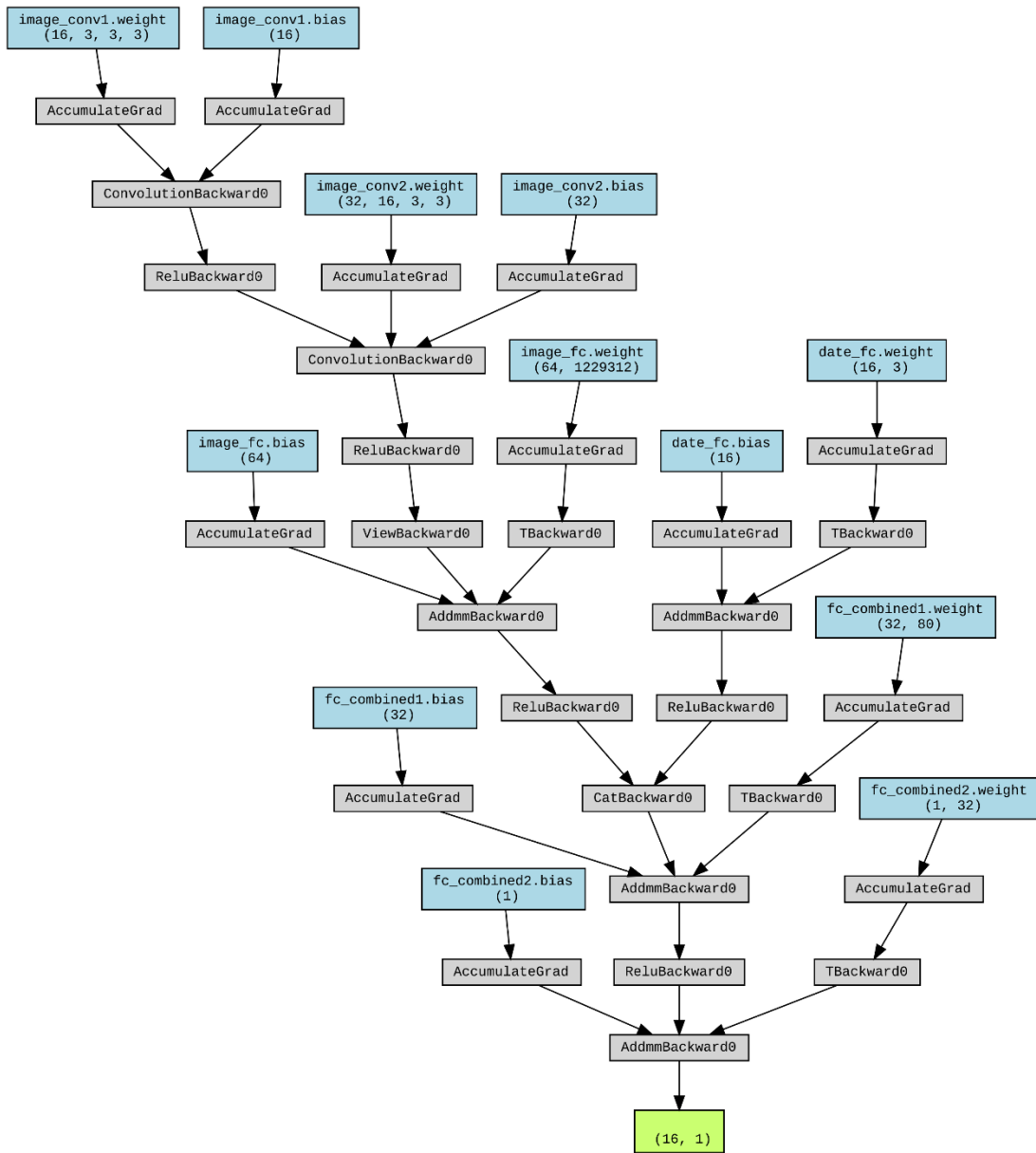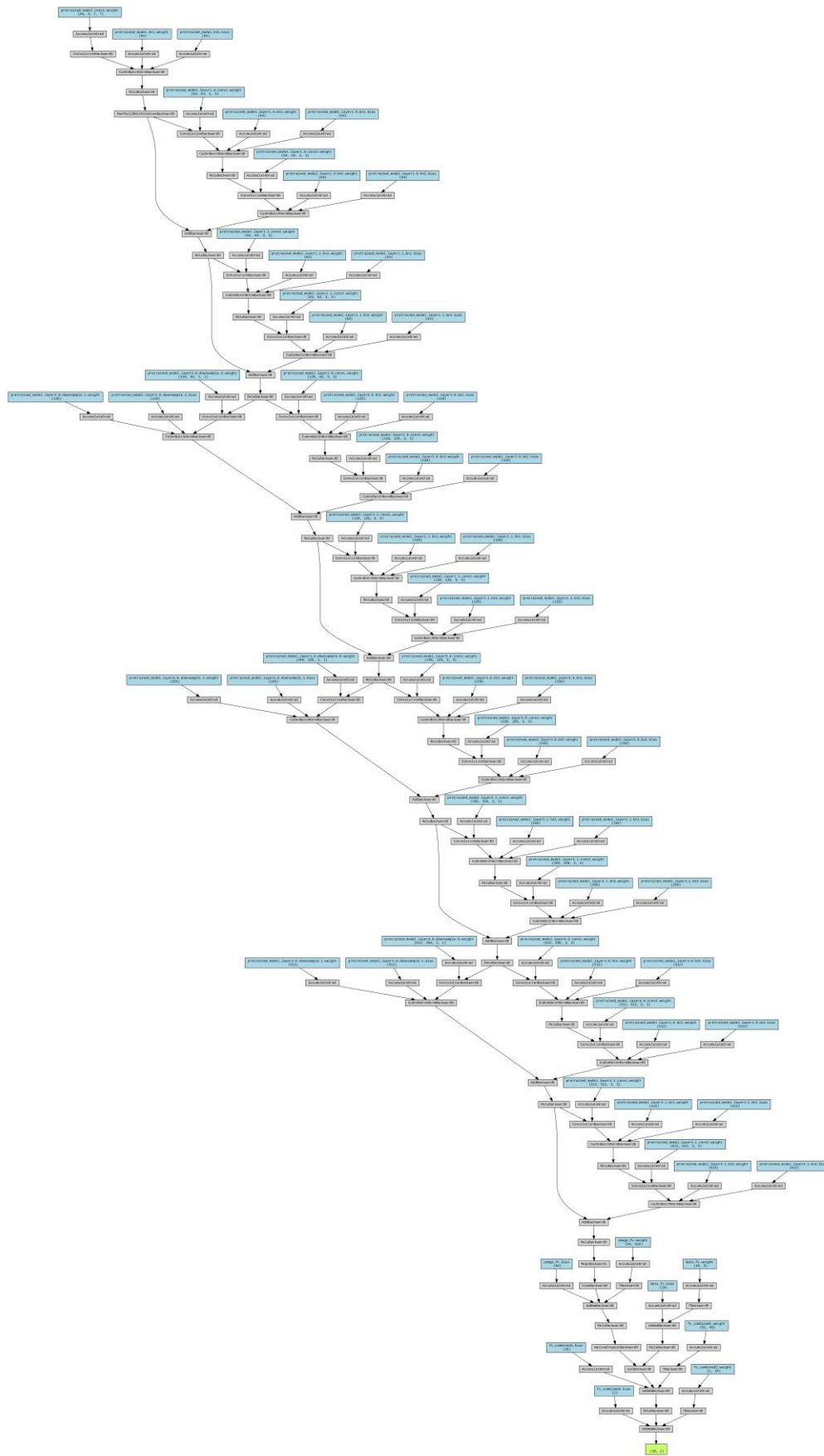
# Appendix



*Figure 21 - Appendix A*

*Figure 22 - Appendix B*

# References

Adami, M., Rudorff, B.F.T., Freitas, R.M., Aguiar, D.A., Sugawara, L.M., Mello, M.P., 2012. Remote Sensing Time Series to Evaluate Direct Land Use Change of Recent Expanded Sugarcane Crop in Brazil. Sustainability 4, 574–585. https://doi.org/10.3390/su4040574

An adaptive semantic dimensionality reduction approach for hyperspectral imagery classification | IEEE Conference Publication | IEEE Xplore [WWW Document], n.d. URL https://ieeexplore.ieee.org/document/8364504 (accessed 10.12.23).

AR4 Climate Change 2007: Synthesis Report — IPCC, 2022. URL https://www.ipcc.ch/report/ar4/syr/ (accessed 1.19.22).

AR5 Climate Change 2014: Mitigation of Climate Change — IPCC, 2022. URL https://www.ipcc.ch/report/ar5/wg3/ (accessed 1.19.22).

Audet, J., Bastviken, D., Bundschuh, M., Buffam, I., Feckler, A., Klemedtsson, L., Laudon, H., Löfgren, S., Natchimuthu, S., Öquist, M., Peacock, M., Wallin, M.B., 2020. Forest streams are important sources for nitrous oxide emissions. Glob Chang Biol 26, 629–641. https://doi.org/10.1111/gcb.14812

Ausseil, A.-G., Dymond, J., Dynes, R., Shepherd, J., Vantier, B., Sutherland, A., 2011. Estimating pasture quality using Landsat ETM+: application for the greenhouse gas inventory of New Zealand.

Beirle, S., Platt, U., Wenig, M., Wagner, T., 2003. Weekly cycle of $NO_2$ by GOME measurements: a signature of anthropogenic sources. Atmospheric Chemistry and Physics 3, 2225–2232. https://doi.org/10.5194/acp-3-2225-2003

Berry, H.L., Bowen, K., Kjellstrom, T., 2010. Climate change and mental health: a causal pathways framework. Int J Public Health 55, 123–132. https://doi.org/10.1007/s00038-009-0112-0

Beyond Storms & Droughts: The Psychological Impacts of Climate Change, 2022. . US Climate and Health Alliance. URL https://usclimateandhealthalliance.org/post_resource/beyond-storms-droughts-the-psychological-impacts-of-climate-change/ (accessed 1.18.22).

Buchwitz, M., Schneising, O., Reuter, M., Heymann, J., Krautwurst, S., Bovensmann, H., Burrows, J.P., Boesch, H., Parker, R.J., Somkuti, P., Detmers, R.G., Hasekamp, O.P., Aben, I., Butz, A., Frankenberg, C., Turner, A.J., 2017. Satellite-derived methane hotspot emission estimates using a fast data-driven method. Atmospheric Chemistry and Physics 17, 5751–5774. https://doi.org/10.5194/acp-17-5751-2017

Charlson, R.J., Schwartz, S.E., Hales, J.M., Cess, R.D., J. A. Coakley, J., Hansen, J.E., Hofmann, D.J., 1992. Climate Forcing by Anthropogenic Aerosols. Science. https://doi.org/10.1126/science.255.5043.423

Dagenais, G.R., Leong, D.P., Rangarajan, S., Lanas, F., Lopez-Jaramillo, P., Gupta, R., Diaz, R., Avezum, A., Oliveira, G.B.F., Wielgosz, A., Parambath, S.R., Mony, P., Alhabib, K.F., Temizhan, A., Ismail, N., Chifamba, J., Yeates, K., Khatib, R., Rahman, O., Zatonska, K., Kazmi, K., Wei, L., Zhu, J., Rosengren, A., Vijayakumar, K., Kaur, M., Mohan, V., Yusufali, A., Kelishadi, R., Teo, K.K., Joseph, P., Yusuf, S., 2020. Variations in common diseases, hospital admissions, and deaths in middle-aged adults in 21 countries from five continents (PURE): a prospective cohort study. The Lancet 395, 785–794. https://doi.org/10.1016/S0140-6736(19)32007-0

Dileep, S., Zimmerle, D., Beveridge, R., Vaughn, T., 2020. Automated Identification of Oil Field Features using CNNs, in: Climate Change AI. Presented at the NeurIPS 2020 Workshop on Tackling Climate Change with Machine Learning, Climate Change AI.

Domke, G.M., Walters, B.F., Nowak, D.J., Smith, J., Ogle, S.M., Coulston, J.W., Wirth, T.C., 2020. Greenhouse gas emissions and removals from forest land, woodlands, and urban trees in the United States, 1990-2018. https://doi.org/10.2737/FS-RU-227

Estimation of methane emissions based on crop yield and remote sensing data in a paddy field - Shi - 2020 - Greenhouse Gases: Science and Technology - Wiley Online Library [WWW Document], n.d. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ghg.1946 (accessed 9.15.23).

Fan, J., Bai, J., Li, Z., Ortiz-Bobea, A., Gomes, C.P., 2022. A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction. https://doi.org/10.48550/arXiv.2111.08900

Global Soil Regions Map | NRCS Soils [WWW Document], n.d. URL https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/worldsoils/?cid=nrcs142p2_054013 (accessed 8.8.22).

Goldberg, D.L., Anenberg, S.C., Griffin, D., McLinden, C.A., Lu, Z., Streets, D.G., 2020. Disentangling the Impact of the COVID-19 Lockdowns on Urban NO2 From Natural Variability. Geophysical Research Letters 47, e2020GL089269. https://doi.org/10.1029/2020GL089269

Guo, M., Wang, X., Li, J., Wang, H., Tani, H., 2013. Examining the relationships between land cover and greenhouse gas concentrations using remote-sensing data in East Asia. International Journal of Remote Sensing 34, 4281–4303. https://doi.org/10.1080/01431161.2013.775535

Guterres, B., Khalid, S., Pias, M., Botelho, S., 2021. A data integration pipeline towards reliable monitoring of phytoplankton and early detection of harmful algal blooms, in: Climate Change AI. Presented at the NeurIPS 2021 Workshop on Tackling Climate Change with Machine Learning, Climate Change AI.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. https://doi.org/10.48550/arXiv.1512.03385

Heart Disease and Stroke Statistics—2021 Update | Circulation [WWW Document], 2022. URL https://www.ahajournals.org/doi/10.1161/CIR.0000000000000950 (accessed 4.4.22).

Ho, W.C.G., Heinke, C.O., 2009. A neutron star with a carbon atmosphere in the Cassiopeia A supernova remnant. Nature 462, 71–73. https://doi.org/10.1038/nature08525

Höpfner, M., Milz, M., Buehler, S., Orphal, J., Stiller, G., 2012. The natural greenhouse effect of atmospheric oxygen (O2) and nitrogen (N2). Geophysical Research Letters 39. https://doi.org/10.1029/2012GL051409

Huang, Y., You, S., Fu, Y., Shen, Q., 2021. Weakly-supervised Semantic Segmentation in Cityscape via Hyperspectral Image. https://doi.org/10.48550/arXiv.2012.10122

Integrated National Policy Approaches to Climate-Smart Agriculture. Insights from Brazil, Ethiopia, and New Zealand [WWW Document], 2014. URL https://ccafs.cgiar.org/resources/publications/integrated-national-policy-approaches-climate-smart-agriculture (accessed 8.10.22).

Jacob, D.J., Turner, A.J., Maasakkers, J.D., Sheng, J., Sun, K., Liu, X., Chance, K., Aben, I., McKeever, J., Frankenberg, C., 2016. Satellite observations of atmospheric methane and their value for quantifying methane emissions. Atmospheric Chemistry and Physics 16, 14371–14396. https://doi.org/10.5194/acp-16-14371-2016

Jacobson, T., Nguyen, P., 2021. Scrambling and the black hole atmosphere. arXiv:2111.00669 [gr-qc, physics:hep-th].

Johnson, J.M.-F., Franzluebbers, A.J., Weyers, S.L., Reicosky, D.C., 2007. Agricultural opportunities to mitigate greenhouse gas emissions. Environmental Pollution 150, 107–124. https://doi.org/10.1016/j.envpol.2007.06.030

Kakaei, S., Zakerimoghadam, M., Rahmanian, M., Abbasi Dolatabadi, Z., 2021. The Impact of Climate Change on Heart Failure: A Narrative Review Study. Shiraz E-Med J 22. https://doi.org/10.5812/semj.107895

Kaplan, G., Avdan, Z.Y., Avdan, U., 2019. Spaceborne Nitrogen Dioxide Observations from the Sentinel-5P TROPOMI over Turkey. Proceedings 18, 4. https://doi.org/10.3390/ECRS-3-06181

Keys to Soil Taxonomy | NRCS Soils [WWW Document], n.d. URL https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/class/taxonomy/?cid=nrcs142p2_053580 (accessed 8.8.22).

Lanaras, C., Baltsavias, E., Schindler, K., 2017. Hyperspectral Super-Resolution with Spectral Unmixing Constraints. Remote Sensing 9, 1196. https://doi.org/10.3390/rs9111196

Landsat 8 | Landsat Science, 2021. URL https://landsat.gsfc.nasa.gov/satellites/landsat-8/ (accessed 1.24.22).

Landsat 9 Spectral Specifications | Landsat Science, 2021. URL https://landsat.gsfc.nasa.gov/satellites/landsat-9/landsat-9-instruments/landsat-9-spectral-specifications/ (accessed 1.24.22).

Lange, K., Richter, A., Burrows, J.P., 2022. Variability of nitrogen oxide emission fluxes and lifetimes estimated from Sentinel-5P TROPOMI observations. Atmospheric Chemistry and Physics 22, 2745–2767. https://doi.org/10.5194/acp-22-2745-2022

Lee, J., Seo, S., Kim, M., 2021. SIPSA-Net: Shift-Invariant Pan Sharpening with Moving Object Alignment for Satellite Imagery. Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, pp. 10161–10169. https://doi.org/10.1109/CVPR46437.2021.01003

Liu, M., van der A, R., van Weele, M., Eskes, H., Lu, X., Veefkind, P., de Laat, J., Kong, H., Wang, J., Sun, J., Ding, J., Zhao, Y., Weng, H., 2021. A New Divergence Method to Quantify Methane Emissions Using Observations of Sentinel-5P TROPOMI. Geophysical Research Letters 48, e2021GL094151. https://doi.org/10.1029/2021GL094151

Mahowald, N.M., Ward, D.S., Doney, S.C., Hess, P.G., Randerson, J.T., 2017. Are the impacts of land use on warming underestimated in climate policy? Environ. Res. Lett. 12, 094016. https://doi.org/10.1088/1748-9326/aa836d

Mitchell, J.F.B., 1989. The "Greenhouse" effect and climate change. Reviews of Geophysics 27, 115–139. https://doi.org/10.1029/RG027i001p00115

Mitton, J., Murray-Smith, R., 2021. Rotation Equivariant Deforestation Segmentation and Driver Classification. https://doi.org/10.48550/arXiv.2110.13097

Nogueira, L.M., Yabroff, K.R., Bernstein, A., 2020. Climate change and cancer. CA: A Cancer Journal for Clinicians 70, 239–244. https://doi.org/10.3322/caac.21610

Oertel, C., Matschullat, J., Zurba, K., Zimmermann, F., Erasmi, S., 2016. Greenhouse gas emissions from soils—A review. Geochemistry 76, 327–352. https://doi.org/10.1016/j.chemer.2016.04.002

Omrani, H., Omrani, B., Parmentier, B., Helbich, M., 2020. Spatio-temporal data on the air pollutant nitrogen dioxide derived from Sentinel satellite for France. Data Brief 28, 105089. https://doi.org/10.1016/j.dib.2019.105089

Parhar, P., Sawasaki, R., Todeschini, A., Reed, C., Vahabi, H., Nusaputra, N., Vergara, F., 2022. HyperionSolarNet: Solar Panel Detection from Aerial Images. https://doi.org/10.48550/arXiv.2201.02107

Pepin, R.O., 2006. Atmospheres on the terrestrial planets: Clues to origin and evolution. Earth and Planetary Science Letters 252, 1–14. https://doi.org/10.1016/j.epsl.2006.09.014

Persello, C., Wegner, J.D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M., Camps-Valls, G., 2021. Deep Learning and Earth Observation to Support the Sustainable Development Goals. arXiv:2112.11367 [cs].

Pradhan, B., Kjellstrom, T., Atar, D., Sharma, P., Kayastha, B., Bhandari, G., Pradhan, P.K., 2019. Heat Stress Impacts on Cardiac Mortality in Nepali Migrant Workers in Qatar. Cardiology 143, 37–48. https://doi.org/10.1159/000500853

Products and Algorithms - Sentinel-5P Technical Guide - Sentinel Online - Sentinel Online [WWW Document], 2022. URL https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/products-algorithms (accessed 1.24.22).

Qiu, J., Li, C., Wang, L., Tang, H., Li, H., Van Ranst, E., 2009. Modeling impacts of carbon sequestration on net greenhouse gas emissions from agricultural soils in China. Global Biogeochemical Cycles 23. https://doi.org/10.1029/2008GB003180

Raj, A.R., Ajai, 2022. Feeling Hot Can Fuel Rage [WWW Document]. Scientific American. https://doi.org/10.1038/scientificamericanmind0114-16

Ramanathan, V., Crutzen, P.J., Kiehl, J.T., Rosenfeld, D., 2001. Aerosols, Climate, and the Hydrological Cycle. Science. https://doi.org/10.1126/science.1064034

Remote Sensing | Free Full-Text | FIgLib &amp; SmokeyNet: Dataset and Deep Learning Model for Real-Time Wildland Fire Smoke Detection [WWW Document], n.d. URL https://www.mdpi.com/2072-4292/14/4/1007 (accessed 11.16.23).

Ritchie, H., Roser, M., Rosado, P., 2020. $CO_2$ and Greenhouse Gas Emissions. Our World in Data.

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Kording, K.P., Gomes, C., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J., Bengio, Y., 2019. Tackling Climate Change with Machine Learning. arXiv:1906.05433 [cs, stat].

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A.S., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Kording, K.P., Gomes, C.P., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J., Bengio, Y., 2022. Tackling Climate Change with Machine Learning. ACM Comput. Surv. 55, 42:1-42:96. https://doi.org/10.1145/3485128

Rossi, F.S., de Araújo Santos, G.A., de Souza Maria, L., Lourençoni, T., Pelissari, T.D., Della-Silva, J.L., Oliveira Júnior, J.W., Silva, A. de A. e, Lima, M., Teodoro, P.E., Teodoro, L.P.R., de Oliveira-Júnior, J.F., La Scala Jr, N., Silva Junior, C.A. da, 2022. Carbon dioxide spatial variability and dynamics for contrasting land uses in central Brazil agricultural frontier from remote sensing data. Journal of South American Earth Sciences 116, 103809. https://doi.org/10.1016/j.jsames.2022.103809

Scafutto, R.D.P.M., De Souza Filho, C.R., 2018. Detection of Methane Plumes Using Airborne Midwave Infrared (3–5 μm) Hyperspectral Data. Remote Sensing 10, 1237. https://doi.org/10.3390/rs10081237

Schneider, A., Borsdorff, T., aan de Brugh, J., Aemisegger, F., Feist, D.G., Kivi, R., Hase, F., Schneider, M., Landgraf, J., 2020. First data set of $H_2O$/HDO columns from the Tropospheric Monitoring Instrument (TROPOMI). Atmospheric Measurement Techniques 13, 85–100. https://doi.org/10.5194/amt-13-85-2020

Sovacool, B.K., Griffiths, S., Kim, J., Bazilian, M., 2021. Climate change and industrial F-gases: A critical and systematic review of developments, sociotechnical systems and policy options for reducing synthetic greenhouse gas emissions. Renewable and Sustainable Energy Reviews 141, 110759. https://doi.org/10.1016/j.rser.2021.110759

The added value of satellite observations of methane forunderstanding the contemporary methane budget - PubMed [WWW Document], n.d. URL https://pubmed.ncbi.nlm.nih.gov/34565220/ (accessed 11.16.23).

Tsaloglidou, A., Koukourikos, K., Papathanasiou, I., Koutsochristou, D., Dalagozi, P., Kourkouta, L., 2018. The Effects of Global Warming on Cardio-Respiratory System, in: Kallel, A., Ksibi, M., Ben Dhia, H., Khélifi, N. (Eds.), Recent Advances in Environmental Science from the Euro-Mediterranean and Surrounding Regions, Advances in Science, Technology & Innovation. Springer International Publishing, Cham, pp. 1949–1951. https://doi.org/10.1007/978-3-319-70548-4_565

ul–Haq, Zia., Tariq, Salman., Ali, Muhammad., Mahmood, Khalid., Batool, S.Adila., Rana, A.Daud., 2014. A study of tropospheric NO2 variability over Pakistan using OMI data. Atmospheric Pollution Research 5, 709–720. https://doi.org/10.5094/APR.2014.080

Underlying Cause of Death, 1999-2020 Request [WWW Document], 2022. URL https://wonder.cdc.gov/ucd-icd10.html (accessed 4.4.22).

Verhoelst, T., Compernolle, S., Pinardi, G., Lambert, J.-C., Eskes, H.J., Eichmann, K.-U., Fjæraa, A.M., Granville, J., Niemeijer, S., Cede, A., Tiefengraber, M., Hendrick, F., Pazmiño, A., Bais, A., Bazureau, A., Boersma, K.F., Bognar, K., Dehn, A., Donner, S., Elokhov, A., Gebetsberger, M., Goutail, F., Grutter de la Mora, M., Gruzdev, A., Gratsea, M., Hansen, G.H., Irie, H., Jepsen, N., Kanaya, Y., Karagkiozidis, D., Kivi, R., Kreher, K., Levelt, P.F., Liu, C., Müller, M., Navarro Comas, M., Piters, A.J.M., Pommereau, J.-P., Portafaix, T., Prados-Roman, C., Puentedura, O., Querel, R., Remmers, J., Richter, A., Rimmer, J., Rivera Cárdenas, C., Saavedra de Miguel, L., Sinyakov,

V.P., Stremme, W., Strong, K., Van Roozendael, M., Veefkind, J.P., Wagner, T., Wittrock, F., Yela González, M., Zehner, C., 2021. Ground-based validation of the Copernicus Sentinel-5P TROPOMI $NO_2$ measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks. Atmospheric Measurement Techniques 14, 481–510. https://doi.org/10.5194/amt-14-481-2021

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sensing of Environment 236, 111402. https://doi.org/10.1016/j.rse.2019.111402

Xie, R., Wei, D., Han, F., Lu, Y., Fang, J., Liu, Y., Wang, J., 2019. The effect of traffic density on smog pollution: Evidence from Chinese cities. Technological Forecasting and Social Change 144, 421–427. https://doi.org/10.1016/j.techfore.2018.04.023

Yan, L., Chang, K., 2021. A New Super Resolution Framework Based on Multi-Task Learning for Remote Sensing Images. Sensors 21, 1743. https://doi.org/10.3390/s21051743

Zhang, K., Sumbul, G., Demir, B., 2020. An Approach to Super-Resolution of Sentinel-2 Images Based on Generative Adversarial Networks, in: 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS). pp. 69–72. https://doi.org/10.1109/M2GARSS47143.2020.9105165

Zhang, Y., Wang, Y.Y., Su, S.L., Li, C.S., 2011. Quantifying methane emissions from rice paddies in Northeast China by integrating remote sensing mapping with a biogeochemical model. Biogeosciences 8, 1225–1235. https://doi.org/10.5194/bg-8-1225-2011

Zheng, Z., Yang, Z., Wu, Z., Marinello, F., 2019. Spatial Variation of NO2 and Its Impact Factors in China: An Application of Sentinel-5P Products. Remote Sensing 11, 1939. https://doi.org/10.3390/rs11161939