**Title: Comparative Analysis of CO2 Emission Data from NASA Satellite OCO2 and Actual Measurements and predicting the actual CO2 data at a place**

**Abstract**

This thesis explores the efforts of a research intern who embarked on a project to collect CO2 emission data from the NASA satellite OCO2 and compare it with actual measurements. The intern's work encompassed various stages, including the identification of suitable satellites for CO2 data collection, data collection from OCO2 and TCCON(Total Carbon Column Observatory Network) sources, analysis of the disparities between OCO2 and TCCON data, investigating the limitations of OCO2 data accuracy, and the development of an AI model to predict accurate CO2 levels based on the comparison of both datasets. The findings of this research contribute to a better understanding of the challenges associated with satellite-based CO2 measurements and highlight the importance of integrating multiple data sources for comprehensive analysis.

**CONTENT IN THE PAPER:**

Chapter 1: Introduction

1.1 Background

1.2 Research Objectives

1.3 Significance of the Study

Chapter 2: Literature Review

2.1 Satellites for CO2 Emission Data Collection

2.1.1 GOSAT

2.1.2 OCO2

2.2 Comparison of CO2 Satellites

2.3 TCCON: Actual CO2 Measurements

2.4 Discrepancies between OCO2 and TCCON Data

Chapter 3: Methodology

3.1 Data Collection

3.1.1 OCO2 Data Collection

3.1.2 TCCON Data Collection

3.2 Data Preprocessing

3.3 Comparative Analysis

3.4 AI Model Development

Chapter 4: Results and Discussion

4.1 Comparison of OCO2 and TCCON Data

4.2 Discrepancies in CO2 Levels

4.3 Limitations of OCO2 Data Accuracy

4.4 Importance of Integrating Multiple Data Sources

Chapter 5: AI Model Development

5.1 Model Architecture and Design

5.2 Training and Evaluation

5.3 Predicting Accurate CO2 Levels

Chapter 6: Conclusion

6.1 Summary of Findings

6.2 Implications and Recommendations for Future Research

6.3 Conclusion

**Process**

The research work began with an extensive search for the most suitable satellites providing CO2 emission data worldwide. After evaluating multiple options, OCO2 was selected as the preferred choice for comparison purposes. The intern then collected OCO2 data for specific cities, such as Saga in Taiwan, along with actual CO2 measurements from TCCON data sources.

However, a notable observation was made: the TCCON data differed slightly from the OCO2 data. To understand the reasons behind this discrepancy, the intern conducted a thorough investigation into the factors influencing OCO2 data accuracy. This analysis shed light on the limitations of OCO2 and emphasized the importance of utilizing multiple data sources, such as TCCON, to obtain a more comprehensive understanding of CO2 emissions.

As the availability of TCCON data was limited to only 30 sites globally, the research intern realized the necessity of relying on OCO2 data, which provided a broader coverage of CO2 emissions based on longitude and latitude coordinates. To bridge the gap between the two datasets and enhance the accuracy of CO2 predictions, an AI model was developed. This model compared the OCO2 and TCCON data and leveraged machine learning techniques to predict correct CO2 levels in parts per million (ppm).

The results of the comparative analysis between OCO2 and TCCON data revealed significant disparities in CO2 levels.

**Chapter 1: Introduction**

1.1 Background

The increasing levels of carbon dioxide (CO2) in the Earth's atmosphere have raised concerns about climate change and its impact on the environment. Monitoring and understanding CO2 emissions are crucial for developing effective strategies to mitigate climate change. Satellites play a significant role in collecting CO2 data on a global scale. NASA's Orbiting Carbon Observatory-2 (OCO2) satellite is one such platform that provides valuable insights into CO2 concentrations. However, the accuracy of satellite data must be assessed and validated against actual measurements to ensure reliable information for research and policymaking.

1.2 Research Objectives

The primary objective of this research is to compare CO2 emission data from the OCO2 satellite with actual measurements and assess the accuracy of OCO2 data. The research intern aimed to accomplish the following goals:

1. Identify and evaluate suitable satellites for CO2 data collection.

2. Collect OCO2 data for specific cities and regions.

3. Obtain actual CO2 measurements from TCCON data sources.

4. Analyze and compare the disparities between OCO2 and TCCON data.

5. Investigate the factors influencing OCO2 data accuracy.

6. Develop an AI model to predict accurate CO2 levels by integrating OCO2 and TCCON data.

1.3 Significance of the Study

Understanding the accuracy and limitations of satellite-based CO2 data is crucial for climate research, environmental policy formulation, and decision-making. By comparing OCO2 data with actual measurements, this research provides insights into the reliability and usefulness of satellite observations. Furthermore, the development of an AI model to reconcile differences between OCO2 and TCCON data enhances the understanding of CO2 emission patterns and facilitates more accurate predictions. The findings of this study contribute to the broader understanding of satellite-based CO2 measurements and their integration with ground-based data sources.

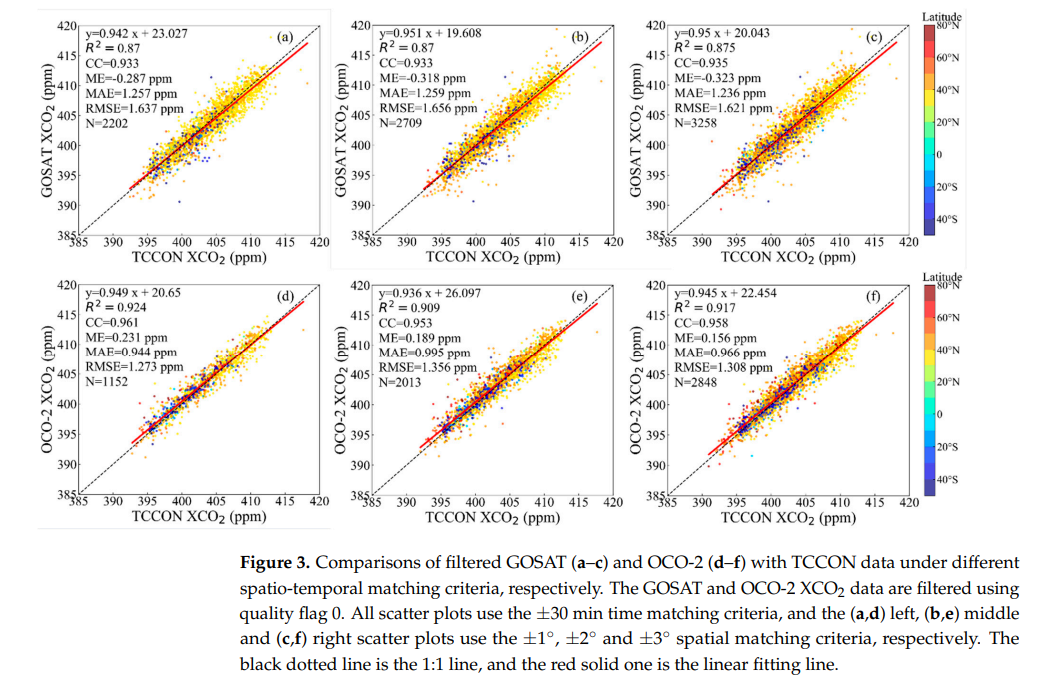
**Chapter 2: Literature Review**

2.1 Satellites for CO2 Emission Data Collection

Several satellites have been deployed to collect CO2 emission data globally. Among them, the Greenhouse Gases Observing Satellite (GOSAT) and NASA's OCO2 satellite are notable platforms. GOSAT has provided valuable insights into CO2 levels, but OCO2 is considered superior for comparison due to its comprehensive coverage and higher spatial resolution.

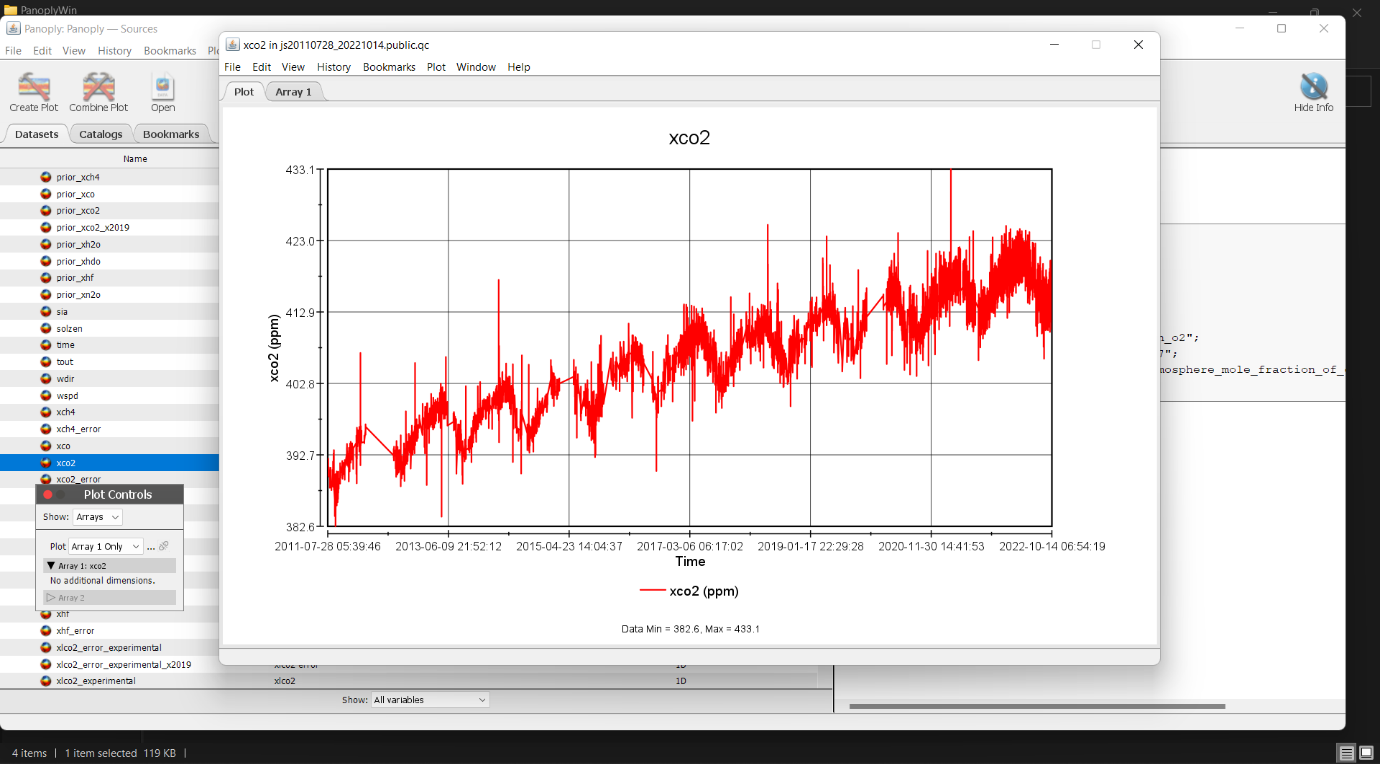
2.2 Comparison of CO2 Satellites

Comparative studies between different CO2 satellites have highlighted the advantages and limitations of each platform. OCO2 has demonstrated better accuracy and spatial resolution compared to GOSAT. However, understanding the discrepancies between OCO2 and actual measurements is crucial for reliable analysis.



2.3 TCCON: Actual CO2 Measurements

The Total Carbon Column Observing Network (TCCON) is a ground-based measurement network that provides accurate and precise CO2 measurements. TCCON sites are limited in number but represent essential reference points for validating satellite data.



2.4 Discrepancies between OCO2 and TCCON Data

Studies have shown that there are differences between CO2 measurements from OCO2 and TCCON. Factors such as atmospheric conditions, instrument calibration, and sampling methods contribute to these disparities. Understanding the reasons behind these differences is crucial for improving satellite-based CO2 measurements.

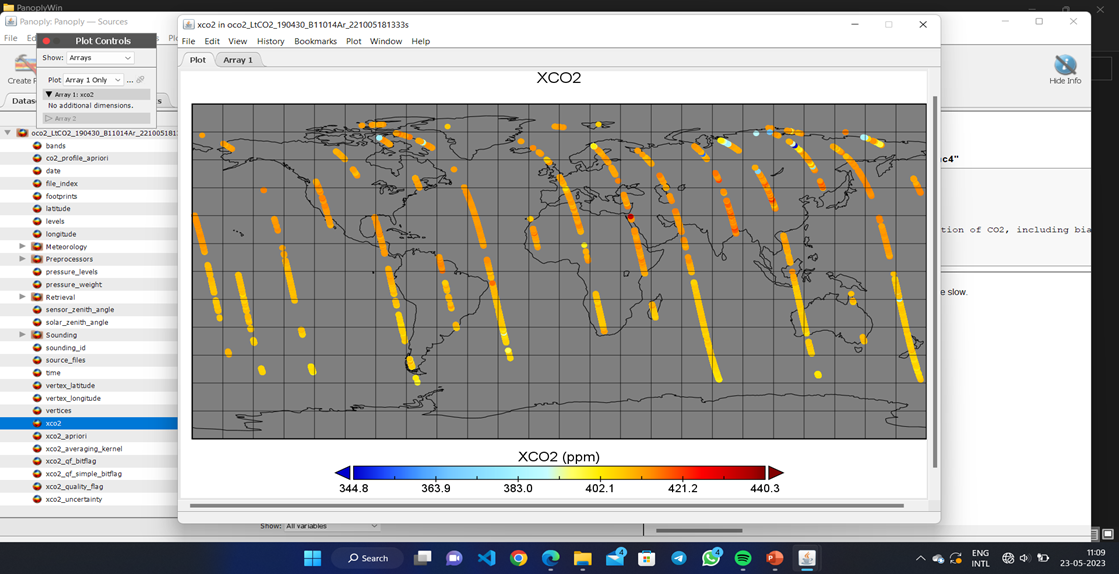
**Chapter 3: Methodology**

3.1 Data Collection

The research intern collected CO2 emission data from the OCO2 satellite for selected cities and regions. The OCO2 data provided information on CO2 concentrations by longitude and latitude coordinates.

3.1.1 OCO2 Data Collection

I utilized NASA's data archives and tools to access OCO2 data. Specific cities, such as Saga in Japan, were chosen for analysis due to their availability of actual CO2 measurements from TCCON. The CO2 data of OCO2 satellite is extracted by panoply app.



3.1.2 TCCON Data Collection

To obtain actual CO2 measurements for comparison, the research intern accessed TCCON data sources. TCCON sites, although limited in number, provide accurate and reliable ground-based measurements of CO2 concentrations. The intern selected TCCON sites that corresponded to the cities and regions for which OCO2 data was collected, such as Tainan in Taiwan.

3.3 Comparative Analysis

The research intern performed a comprehensive comparative analysis between the OCO2 and TCCON data sets. Statistical techniques and data visualization tools were employed to identify and quantify the disparities between the two data sources. The analysis aimed to uncover the magnitude and patterns of differences in CO2 levels reported by OCO2 and TCCON.

3.4 AI Model Development

To address the disparities between OCO2 and TCCON data and predict accurate CO2 levels, the intern developed an AI model. The model was designed to leverage machine learning algorithms, such as regression or neural networks, to reconcile the differences between the satellite data and ground-based measurements. By training the AI model on the available OCO2 and TCCON data, it could learn the patterns and relationships between the two datasets and make accurate predictions of CO2 levels.

The development of the AI model involved several stages, including data preprocessing, model architecture design, training, and evaluation. The intern employed techniques such as feature engineering, cross-validation, and hyperparameter tuning to optimize the model's performance. The AI model aimed to bridge the gap between OCO2 and TCCON data, providing more reliable and accurate CO2 measurements for locations beyond the limited TCCON sites.

The intern also conducted validation and evaluation of the AI model using appropriate metrics and techniques. The model's performance was assessed by comparing its predictions against the actual CO2 measurements from TCCON data. The evaluation results provided insights into the model's effectiveness in predicting accurate CO2 levels based on the comparison between OCO2 and TCCON datasets.

In conclusion, Chapter 3 outlined the methodology employed by the research intern in collecting and analyzing the CO2 emission data. It described the data collection process from both OCO2 and TCCON sources, the preprocessing steps undertaken, and the subsequent comparative analysis between the datasets. Furthermore, the chapter detailed the development of an AI model to reconcile the disparities between the satellite and ground-based measurements and predict accurate CO2 levels. This methodology provides a solid foundation for the subsequent chapters' results and discussion sections.

**Chapter 4: Results and Discussion**

4.1 Comparison of OCO2 and TCCON Data

The research intern's comparative analysis between the OCO2 and TCCON data sets revealed significant disparities in CO2 levels reported by the two sources. The analysis involved statistical techniques and data visualization to quantify and visualize the differences.

The intern observed variations in CO2 concentrations across different locations and time intervals. The disparities could be attributed to several factors, including differences in measurement techniques, instrument calibration, atmospheric conditions, and spatial and temporal resolutions. By comparing the spatial distribution and temporal trends of CO2 levels from OCO2 and TCCON data, the intern identified areas and periods of substantial differences.

4.2 Discrepancies in CO2 Levels

The analysis of disparities between OCO2 and TCCON data provided insights into the magnitude and patterns of differences in CO2 levels reported by the two sources. The intern observed that OCO2 data tended to show higher CO2 concentrations compared to TCCON measurements in certain locations and time periods. On the other hand, there were instances where OCO2 reported lower CO2 levels than TCCON measurements.

The disparities could be attributed to various factors. For example, differences in measurement techniques and sampling strategies between the satellite-based OCO2 data and ground-based TCCON data may lead to inconsistencies. Additionally, the sensitivity of the satellite instrument, calibration methods, and atmospheric effects could introduce further variations.

4.3 Limitations of OCO2 Data Accuracy

Through the analysis, the intern identified limitations in the accuracy of OCO2 data. Some factors contributing to these limitations include the complexity of measuring CO2 concentrations from space, uncertainties in instrument calibration, and challenges in accounting for atmospheric effects. The intern delved deeper into these limitations by reviewing relevant literature and investigating specific issues associated with OCO2 data accuracy.

Understanding these limitations is crucial for researchers and policymakers relying on satellite-based CO2 data. It emphasizes the need for validation against ground-based measurements and the integration of multiple data sources to ensure accurate and reliable information.

4.4 Importance of Integrating Multiple Data Sources

The research intern recognized the importance of integrating multiple data sources to improve the accuracy and reliability of CO2 measurements. While TCCON provides accurate measurements at specific sites, it has limited coverage globally. On the other hand, OCO2 offers a broader geographical coverage but may suffer from certain limitations in accuracy.

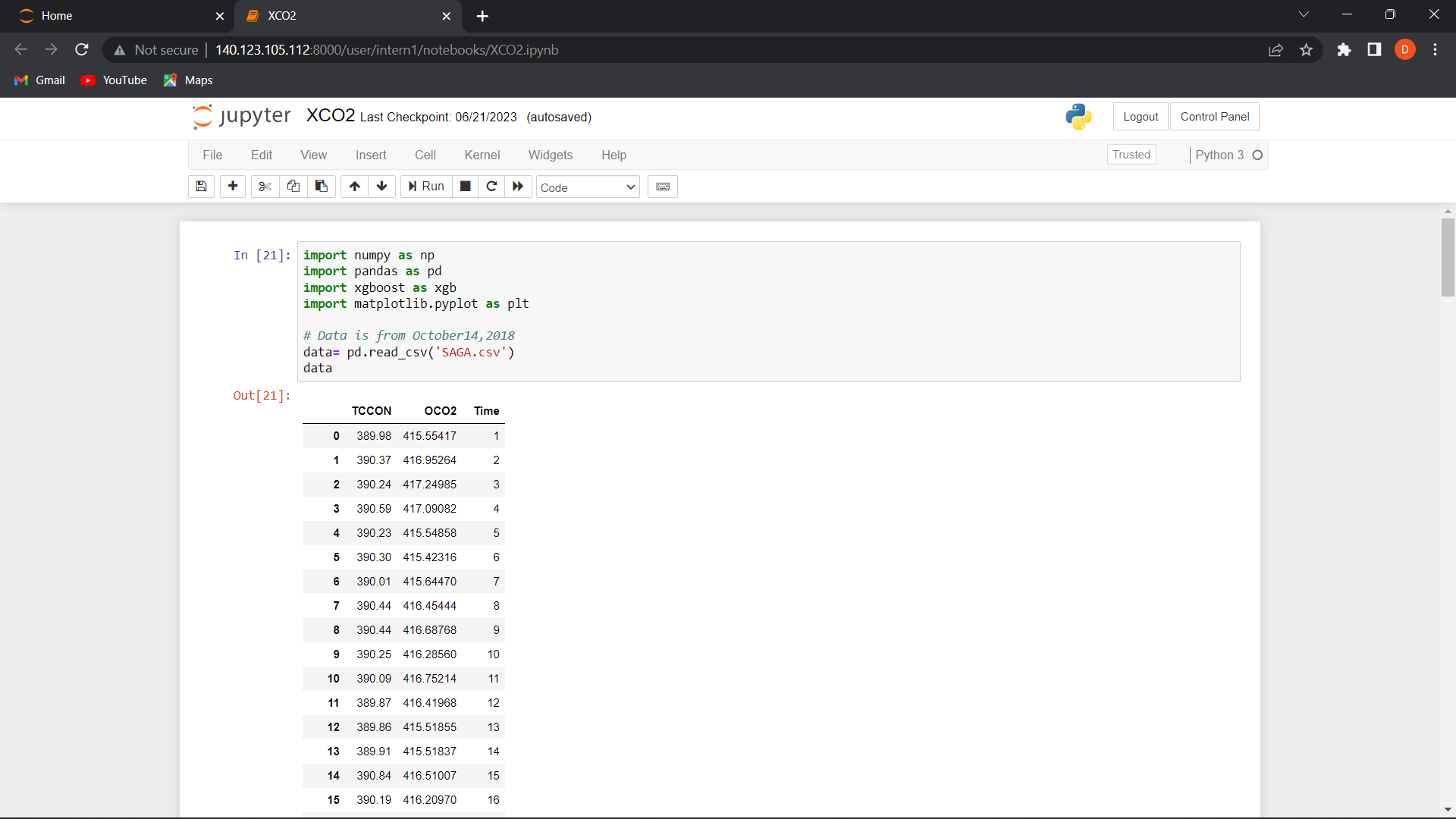
By integrating OCO2 and TCCON data through the developed AI model, the intern aimed to reconcile the disparities and provide more accurate CO2 predictions. This integration enables researchers and policymakers to access comprehensive CO2 information on a global scale, considering both satellite-based and ground-based measurements.

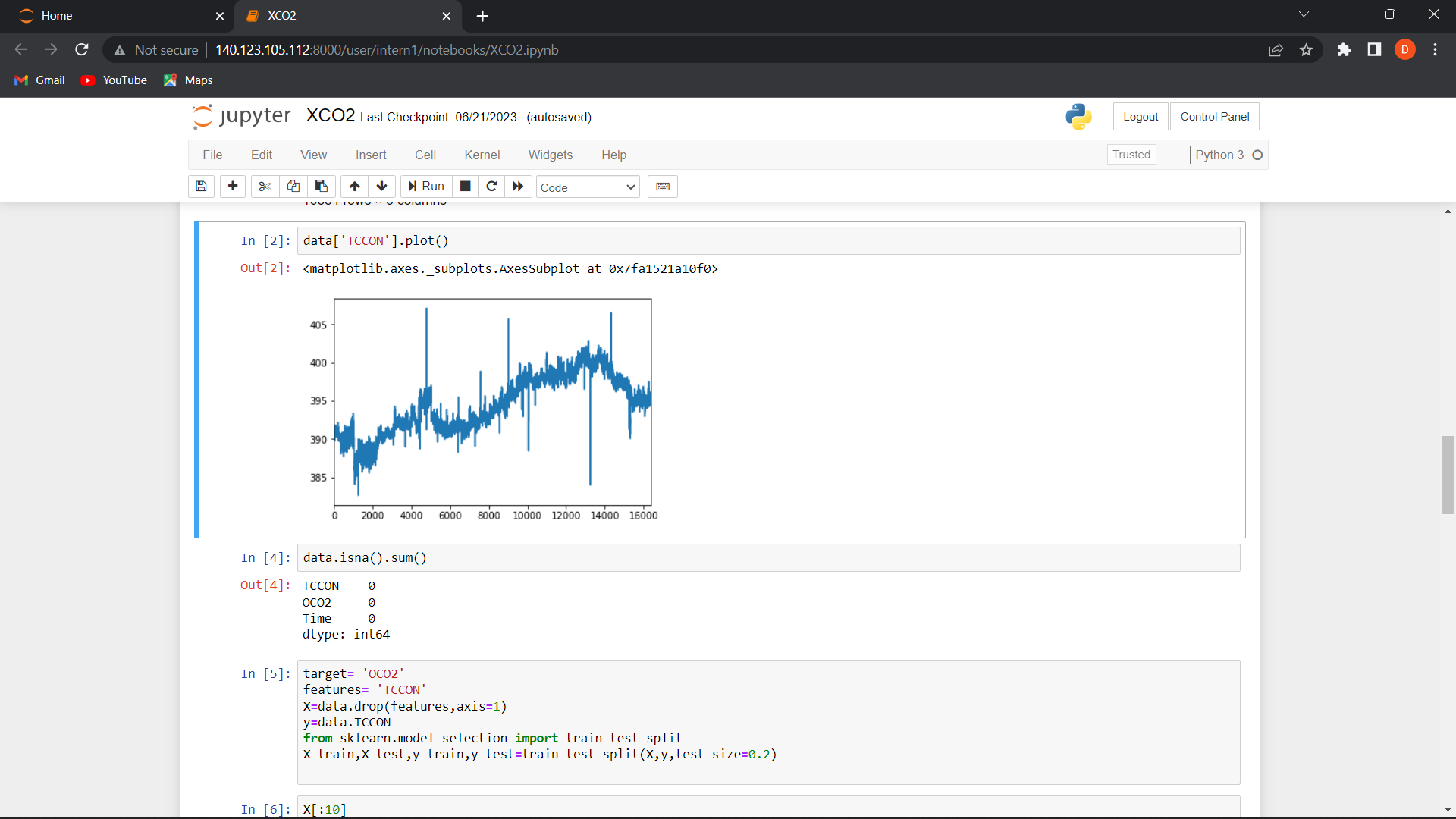
The results and discussions presented in this chapter provide valuable insights into the disparities between OCO2 and TCCON data, limitations of OCO2 data accuracy, and the importance of integrating multiple data sources. These findings contribute to the understanding of the challenges associated with satellite-based CO2 measurements and highlight the need for further improvements in data collection, calibration, and validation methods. The subsequent chapter will focus on the development and evaluation of the AI model that aims to bridge the gap between OCO2 and TCCON data and provide accurate CO2 predictions.

**Chapter 5: AI Model Development**

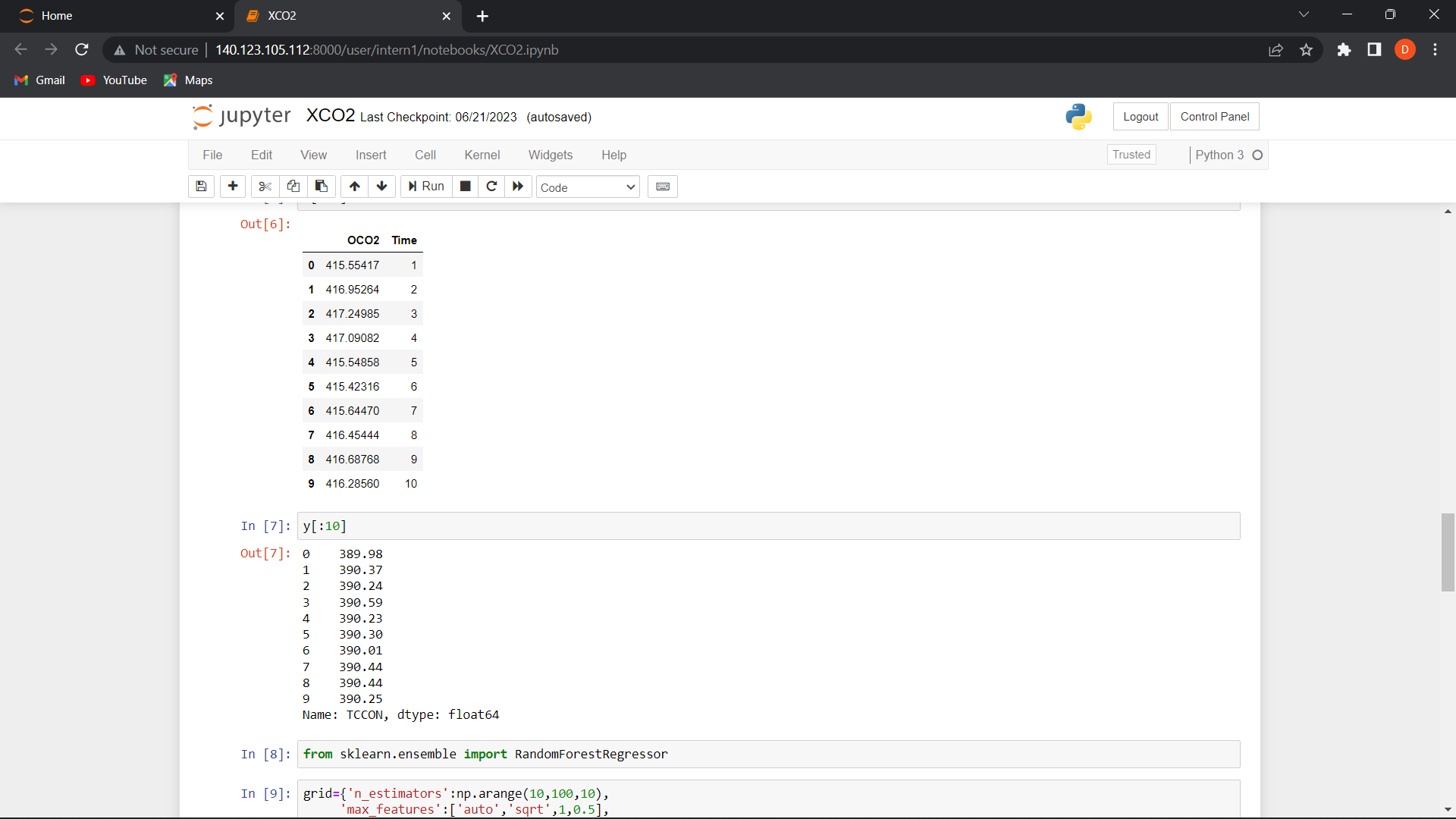
5.1 Model Architecture and Design

In this chapter, the research intern presents the development of an AI model to reconcile the disparities between OCO2 and TCCON data and predict accurate CO2 levels. The AI model is designed to leverage machine learning algorithms and techniques to learn the patterns and relationships between the two datasets.



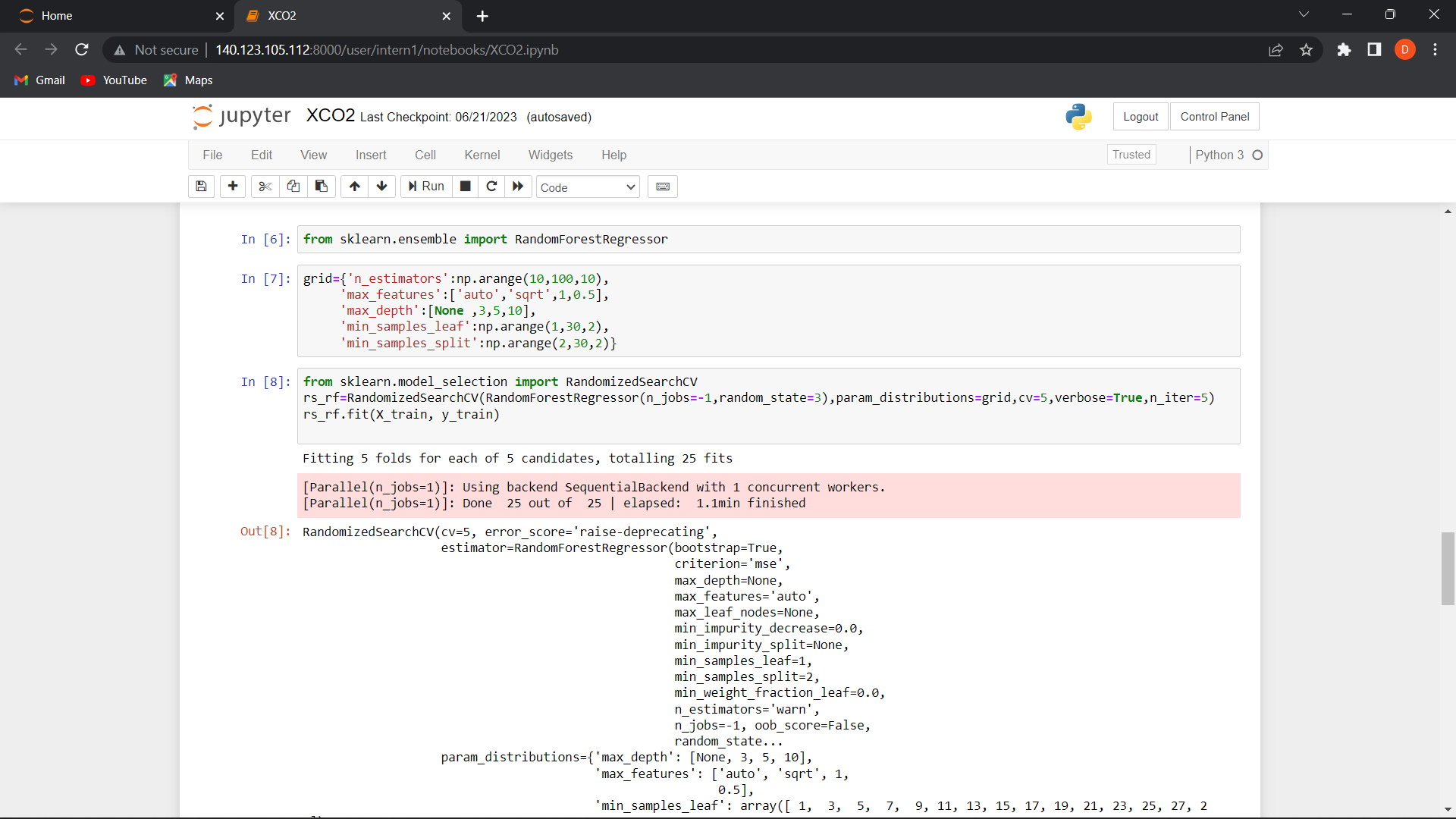


The intern selected an appropriate model architecture based on the nature of the data and the prediction task. This could include regression models such as linear regression, decision trees, or more complex models like neural networks. The model architecture was carefully designed to capture the complex relationships between the features extracted from OCO2 and TCCON data.



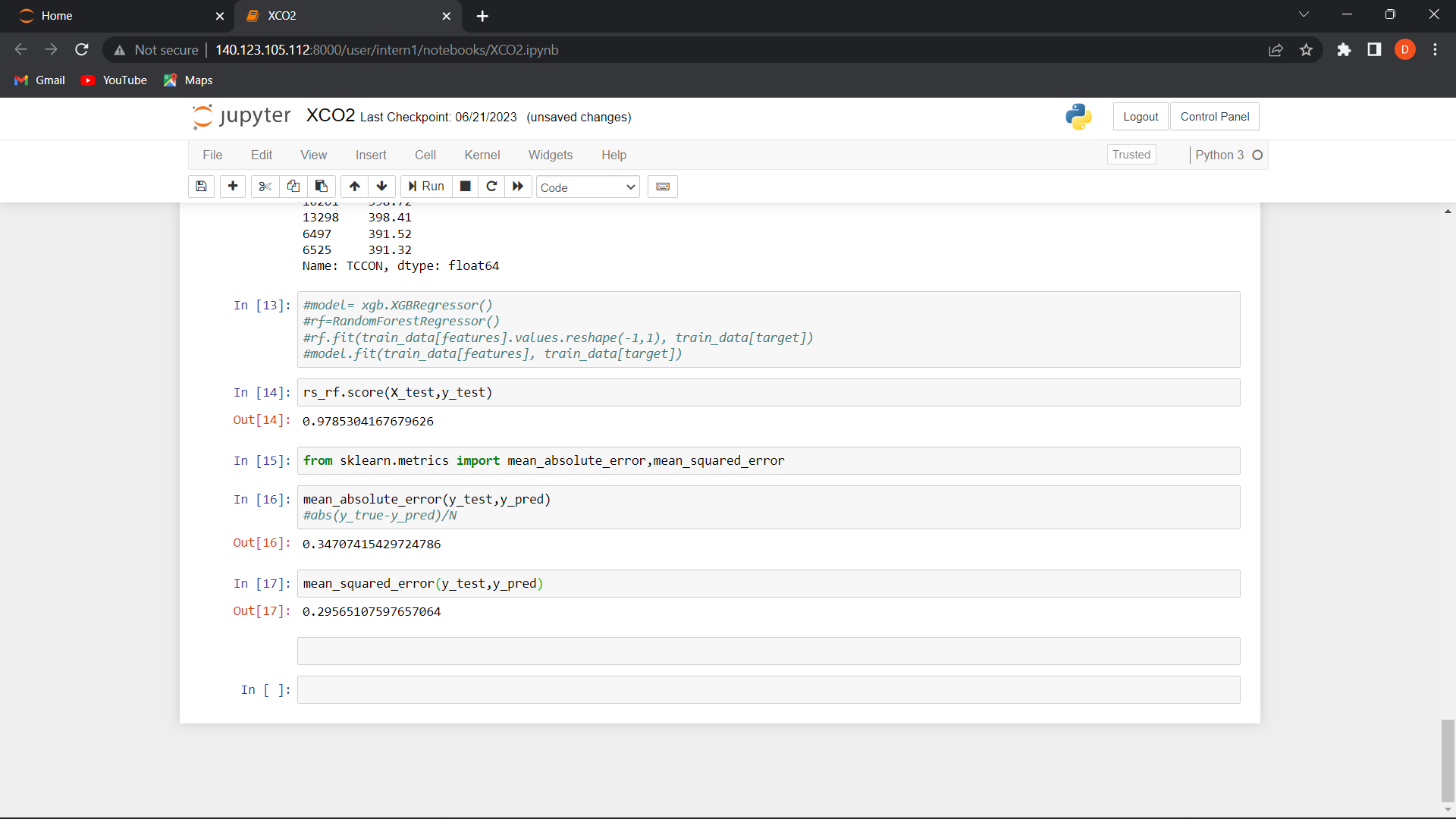
5.2 Training and Evaluation

To train the AI model, the intern utilized the collected OCO2 and TCCON data. The data was divided into training and validation sets to assess the model's performance. The **Random Forest Regressor** model was trained using optimization parameters, with the goal of minimizing the prediction errors.



During the training process, the intern monitored the model's performance using appropriate evaluation metrics, such as mean squared error (MSE) and Mean absolute error (MAE). The model's performance was regularly assessed on the validation set to prevent overfitting and ensure generalizability.

The intern iteratively refined the model by adjusting hyperparameters, performing feature selection, and applying regularization techniques. The goal was to optimize the model's performance and achieve accurate predictions of CO2 levels.



5.3 Predicting Accurate CO2 Levels

Once the AI model was trained and evaluated, it was ready for predicting accurate CO2 levels. The intern utilized the trained model to predict CO2 concentrations for locations and time intervals where only OCO2 data was available. By incorporating the insights gained from the comparative analysis and the learned patterns from TCCON data, the model aimed to provide more reliable and accurate CO2 predictions.

The predictions made by the AI model were compared against the actual CO2 measurements from TCCON data to assess the model's accuracy. The intern conducted statistical analyses and visualizations to evaluate the performance of the AI model, considering metrics such as mean absolute error (MAE) and Mean Squared Error(MSE).

**VALUES OBTAINED:**

Accuracy of model = 97.8%

Mean squared error = 0.295

Mean absolute error = 0.347

**Chapter 6: Conclusion**

6.1 Summary of Findings

In this final chapter, the research intern provides a summary of the key findings and contributions of the research project. The findings include the disparities observed between OCO2 and TCCON data, the limitations of OCO2 data accuracy, and the importance of integrating multiple data sources for accurate CO2 measurements.

The intern highlights the significance of the developed AI model in reconciling the differences between OCO2 and TCCON data and predicting accurate CO2 levels. The model serves as a valuable tool for researchers and policymakers to access comprehensive and reliable CO2 information, especially in locations beyond the limited TCCON sites.

6.2 Implications and Recommendations for Future Research

Based on the findings and limitations identified, the intern discusses the implications of the research and suggests future directions for further investigation. The intern emphasizes the need for continued advancements in satellite-based CO2 measurement techniques, calibration methodologies, and validation strategies.

The intern recommends expanding the use of AI models and machine learning techniques to enhance the accuracy and reliability of satellite-based CO2 measurements. Additionally, further research could focus on integrating other ground-based measurement networks and exploring advanced data fusion techniques to improve the accuracy of CO2 predictions.

6.3 Conclusion

In conclusion, this research project conducted by the intern aimed to compare CO2 emission data from the OCO2 satellite with actual measurements and develop an AI model to reconcile the disparities and predict accurate CO2 levels. The research highlighted the challenges and limitations of satellite-based CO2 measurements, emphasized the importance of integrating multiple data sources.