UNIVERSIDAD POLITÉCNICA DE CARTAGENA

Escuela Técnica Superior de Ingeniería de Telecomunicación

Calibration of Sensor Data using Machine Learning

PRÁCTICAS EXTERNAS

GRADO EN INGENIERÍA TELEMÁTICA

Author: Nakul Poudel

Student Number in SEC: SEC074BCT023

Programme in home university: Internship

Advisor: Rafael Asorey-Cacheda

Date: 23/07/4

ABSTRACT

Machine learning is a powerful technique that can be used for the calibration of sensor. Various environmental factor such as temperature, humidity, influences of other gases present in the atmosphere, dust etc have influence on the data collected by sensor. Their influence can be reduced by proper sensor calibration. Different machine learning algorithms like linear regression, polynomial regression and deep neural network along with various python libraries have been used in this project. Eventually, different models have been obtained which provides necessary adjustment to correct the error to an acceptable level.

TABLE OF CONTENT

List of Figures

CHAPTER 1: INTRODUCTION

Calibration is the process of configuring an instrument making adjustment or set of adjustment to make that instrument function as accurately, or error free, as possible. Sensors are used to collect the level of gases present in the atmosphere every hour. The data collected from the sensor are affected by various environmental factors. Machine learning algorithms like linear regression, polynomial regression and deep neural networks are used to create models for the calibration of sensor. The data corresponding to sensor from weather station collected every hour is assumed to be the standard value for calibration. Machine learning models are created using those sets of data and necessary adjustment are made to correct the data.

CHAPTER 2: METHODOLOGY

2.1 Data Set

Different sensors that belongs to alpha sense B4 family are used to measure the level of gases present in the atmosphere. The level of Nitrogen dioxide, Sulphur dioxide, Carbon monoxide, Ozone are measured with corresponding sensor along with the temperature and humidity at that time. The target value is obtained from the weather station. The dataset is organized in CSV format.

2.2 Model Overview

Machine learning algorithms like linear regression, polynomial regression and deep neural network are used to train the model.

Various python libraries like numpy, sklearn, pandas matplotlib etc are used for model training. Numpy is used for computation related to the arrays. Pandas is used to read CSV file. Plots are drawn using Matplotlib. The Sklearn library provides methods Linear Regression and Polynomial Features which are used to implement linear regression and polynomial regression. The Polynomial Features is used to create variable of provided degree as an argument.

Similarly, the deep neural network is created using the libraries keras. Four fully connected layers are defined with relu as an activation function. The relu activation function returns 0 if it receives any negative input, but for any positive value x, it returns that value back. So it can be written as:

$$f(x) = \max(0, x)$$

The model was trained with Adam as an optimizer and a mean square error as a loss function with a learning rate of 0.01.

2.3 Architecture of Neural Network

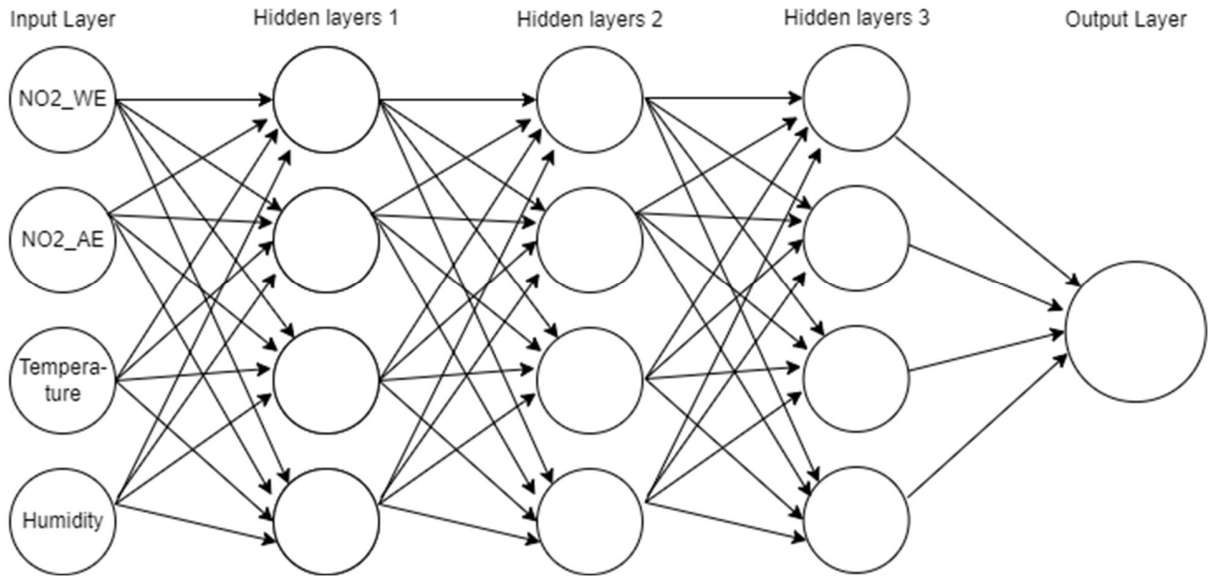

Figure 1: Architecture of Neural Network

A model for calibration has been created using the neural network. It consists of Input layer with four parameters i.e. data from sensor auxiliary pin (NO2_AE), data from sensor worker pin (NO2_WE), Temperature and Humidity of the atmosphere collected by the sensor. The three hidden layers with four hidden units are defined. These layers use Relu as an activation function which output the input directly if it is positive, otherwise, it will output zero. The model is trained for thirty epoch with Mean square error as loss function and Adam as optimizer.

CHAPTER 3: RESULT AND ANALYSIS

The models are first created first using linear regression, polynomial regression and deep neural network respectively. There are altogether 554 data which are split in the training and testing set in the ratio 9:1. The data so obtained is collected at different timing for multiple days.

3.1 Linear Regression

The curves obtained using linear regression model for different gases plotted on test dataset are shown below.

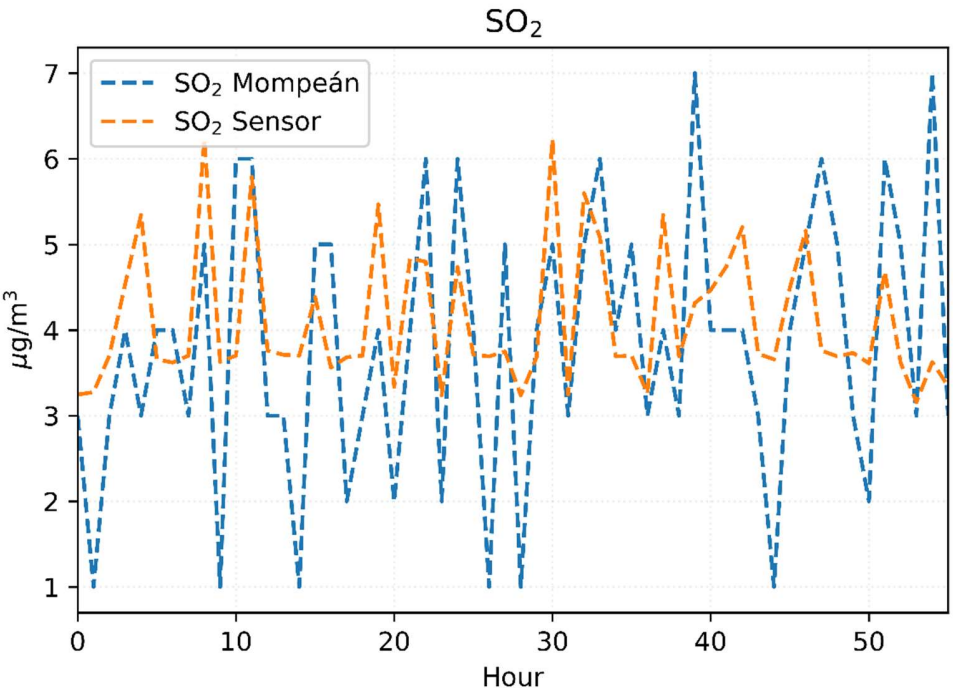

Figure 2: Plot between SO2 levels from sensor and weather station and hours

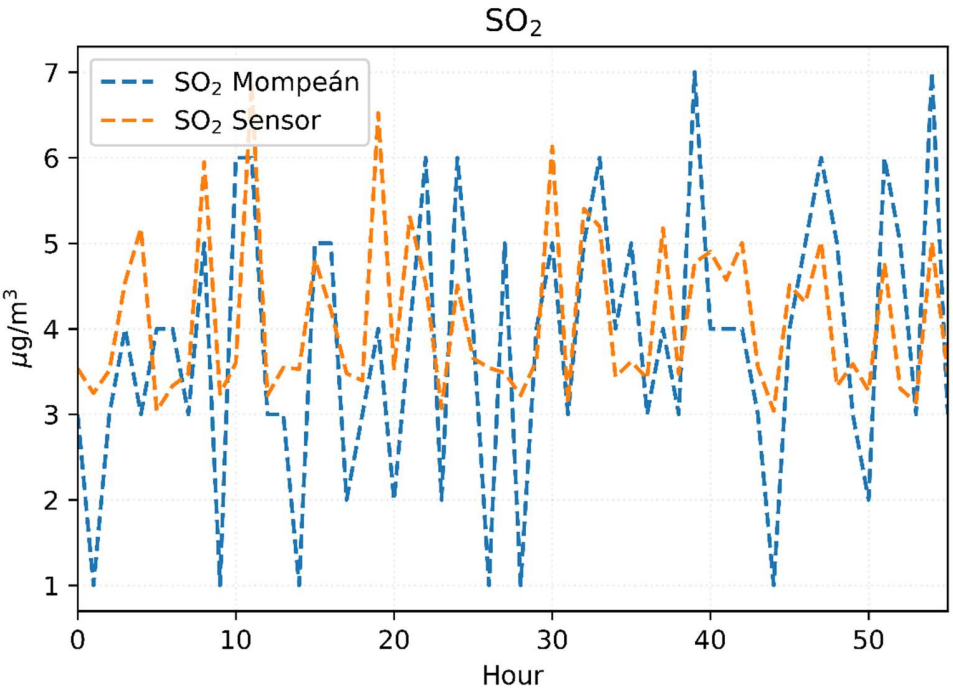

Figure 3: Plot between SO2 levels from sensor and weather station and hours considering temperature.

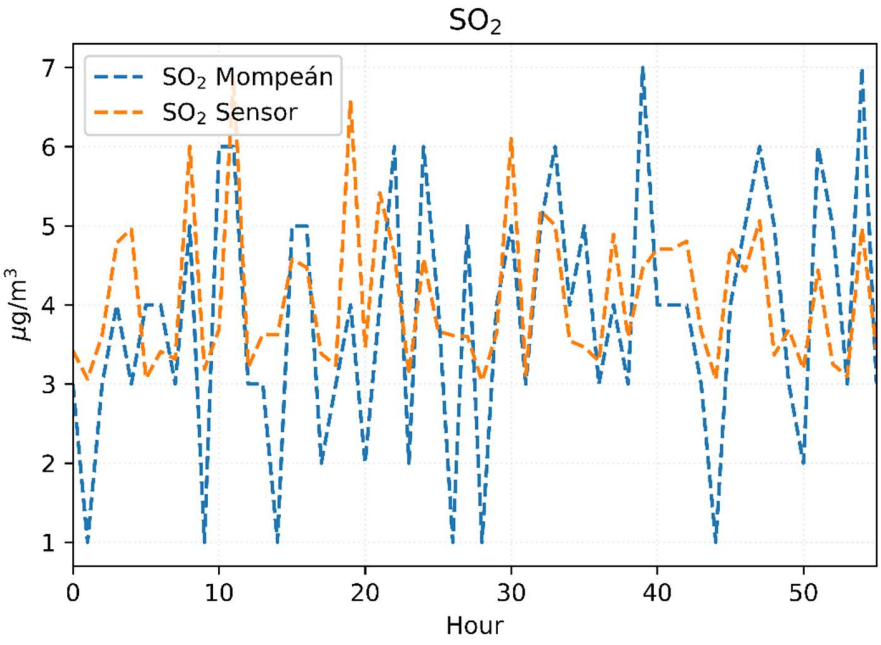

Figure 4: Plot between SO2 levels from sensor and weather station and hours considering temperature and humidity.

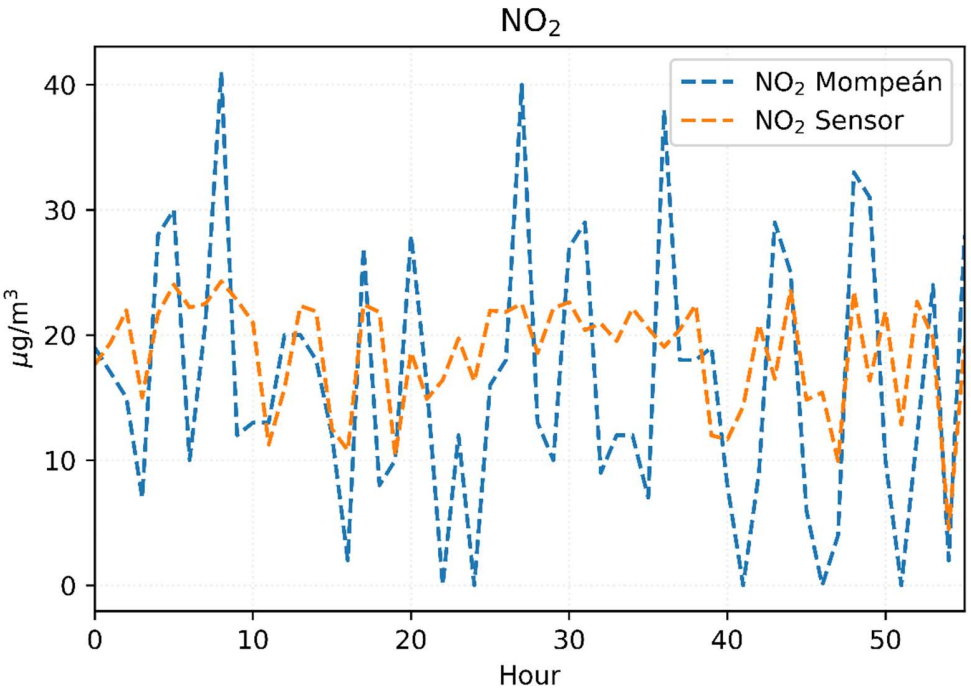

Figure 5: Plot between NO2 levels from sensor and weather station and hours.

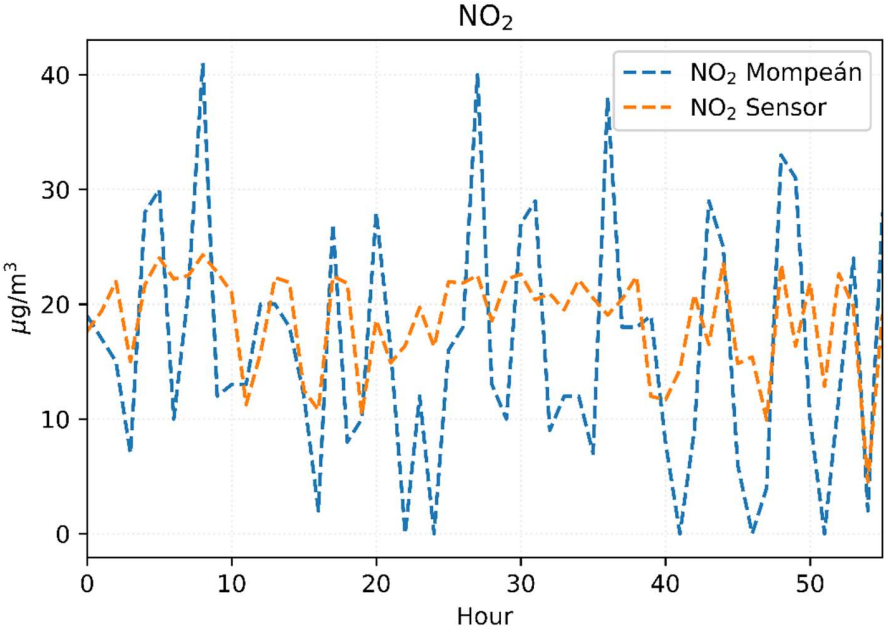

Figure 6: Plot between SO2 levels from sensor and weather station and hours considering temperature.

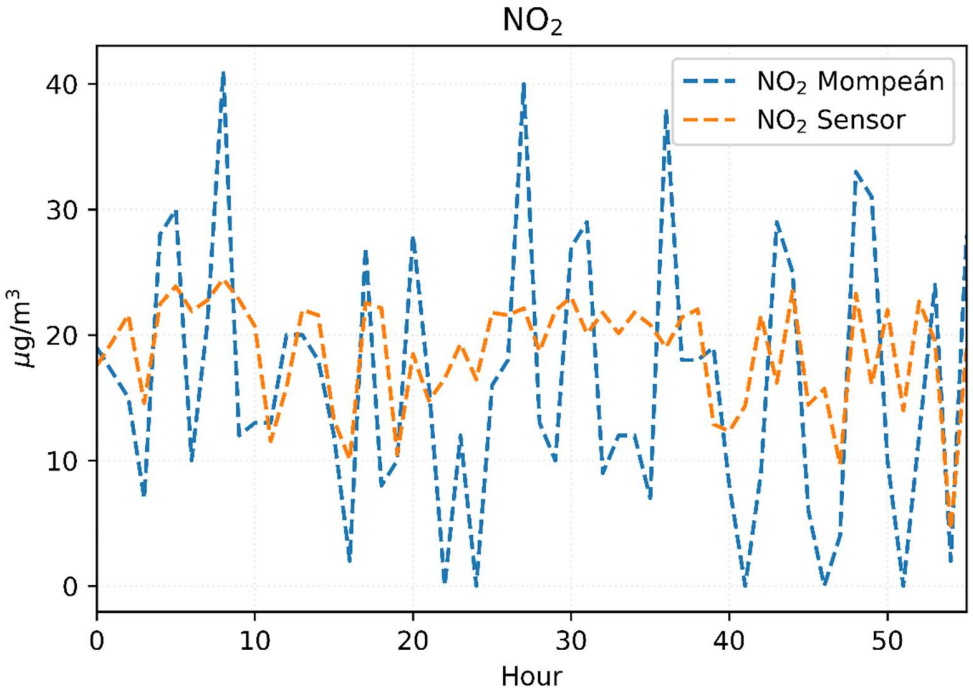

Figure 7: Plot between NO2 levels from sensor and weather station and hours considering temperature and humidity.

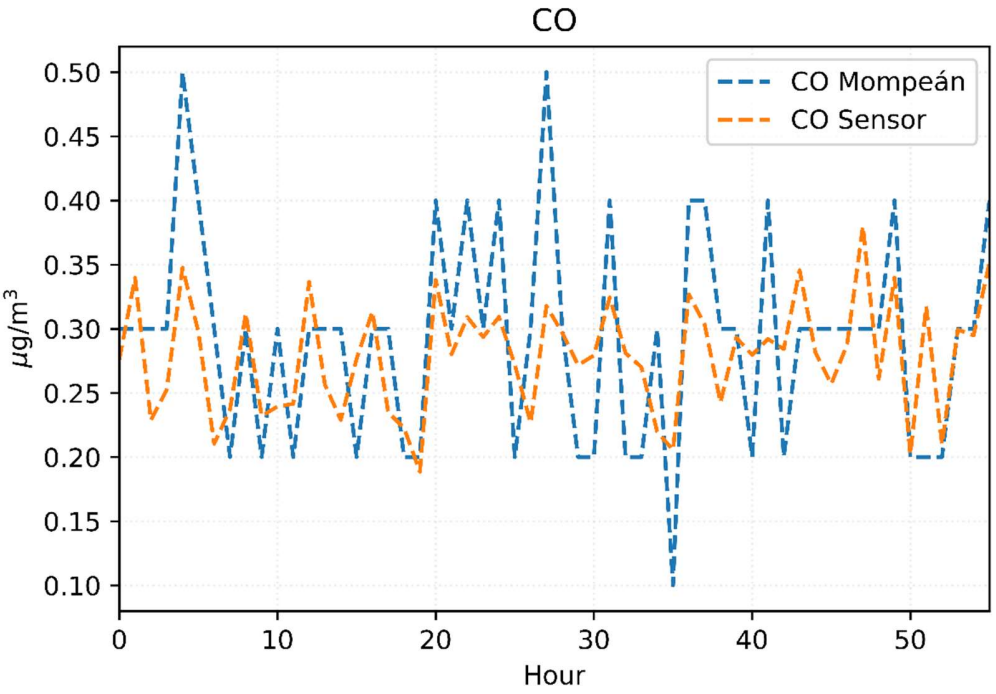

Figure 8: Plot between CO levels from sensor and weather station and hours.

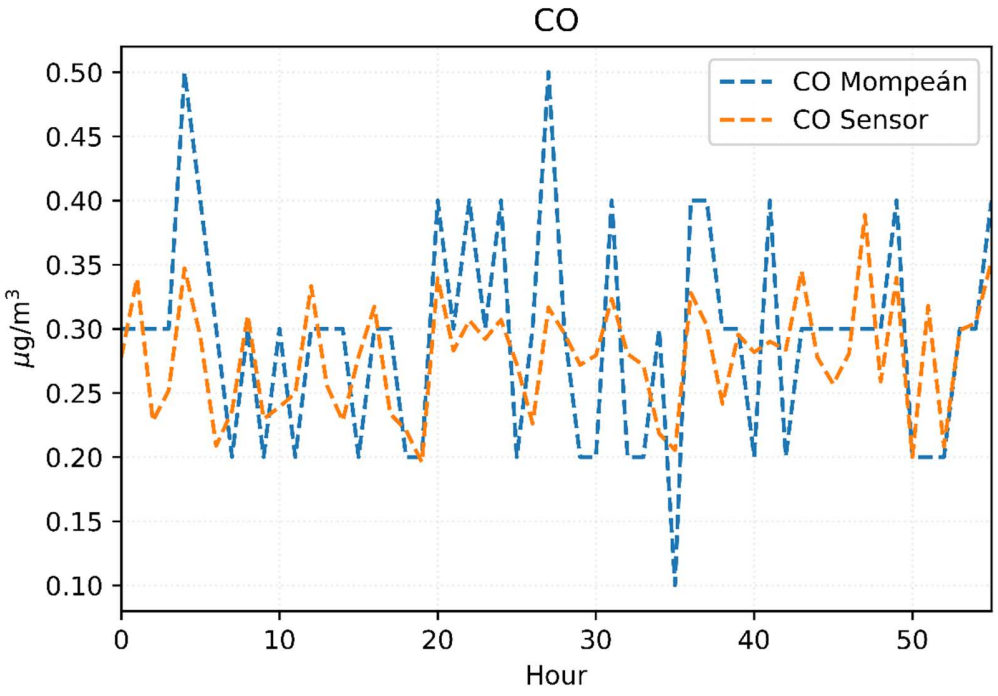

Figure 9: Plot between CO levels from sensor and weather station and hours considering temperature.

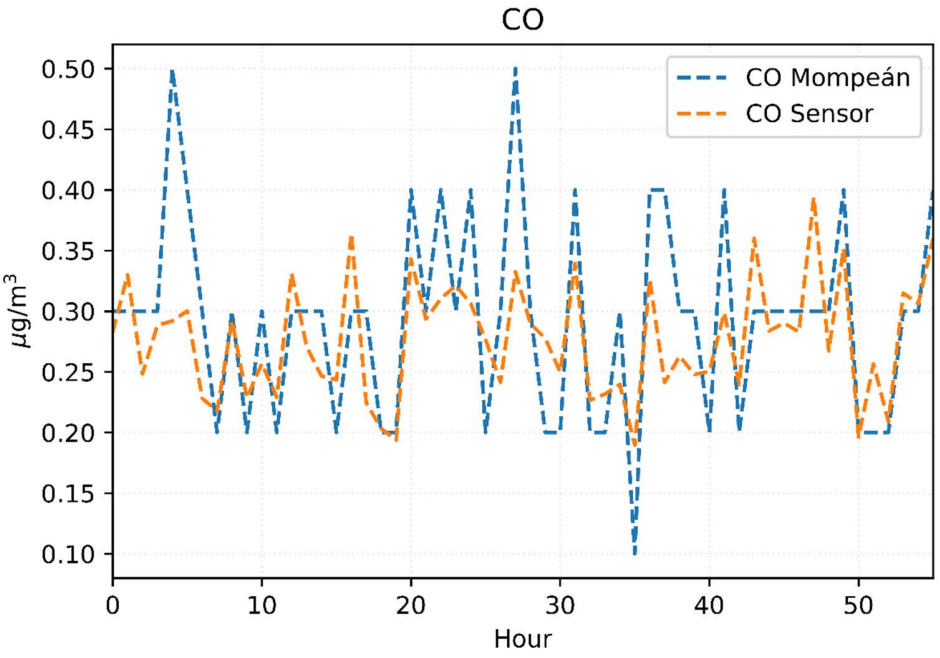

Figure 10: Plot between CO levels from sensor and weather station and hours considering temperature and humidity.

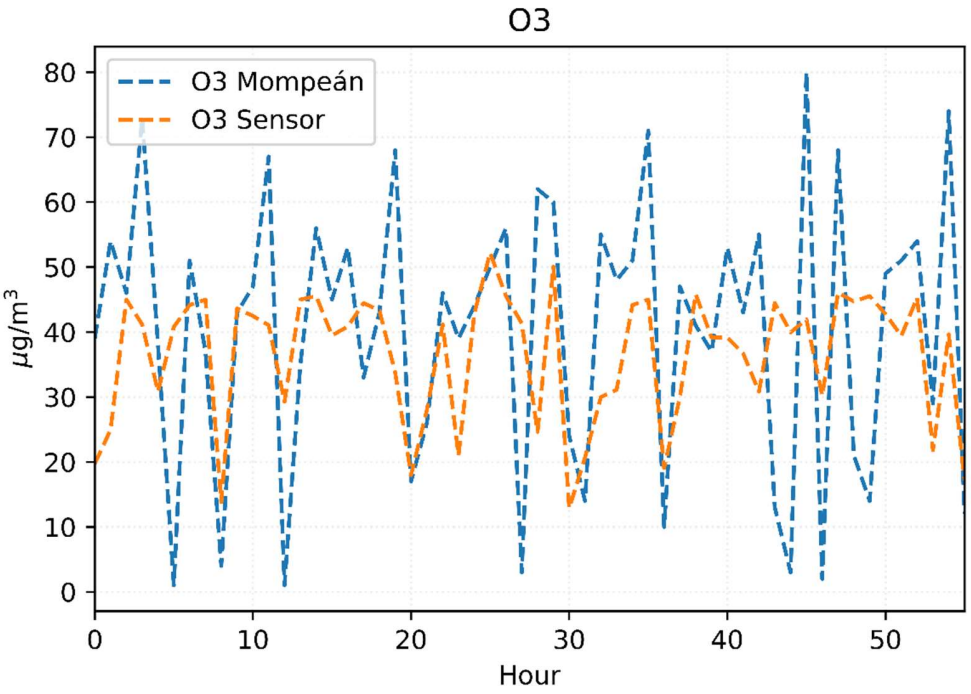

Figure 11: Plot between O3 levels from sensor and weather station and hours.

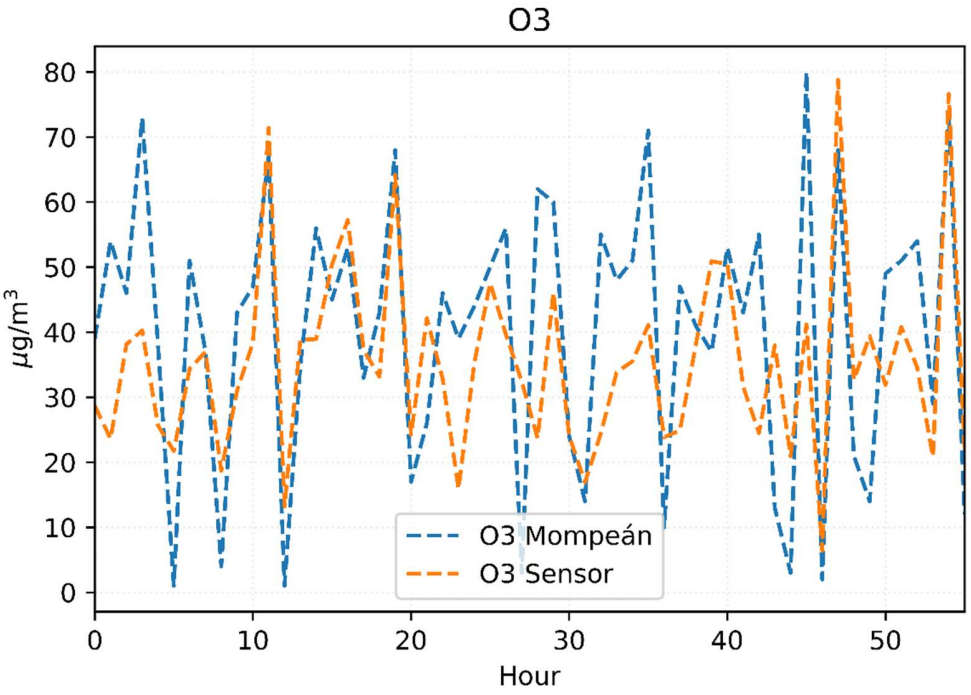

Figure 12: Plot between O3 levels from sensor and weather station and hours considering temperature.

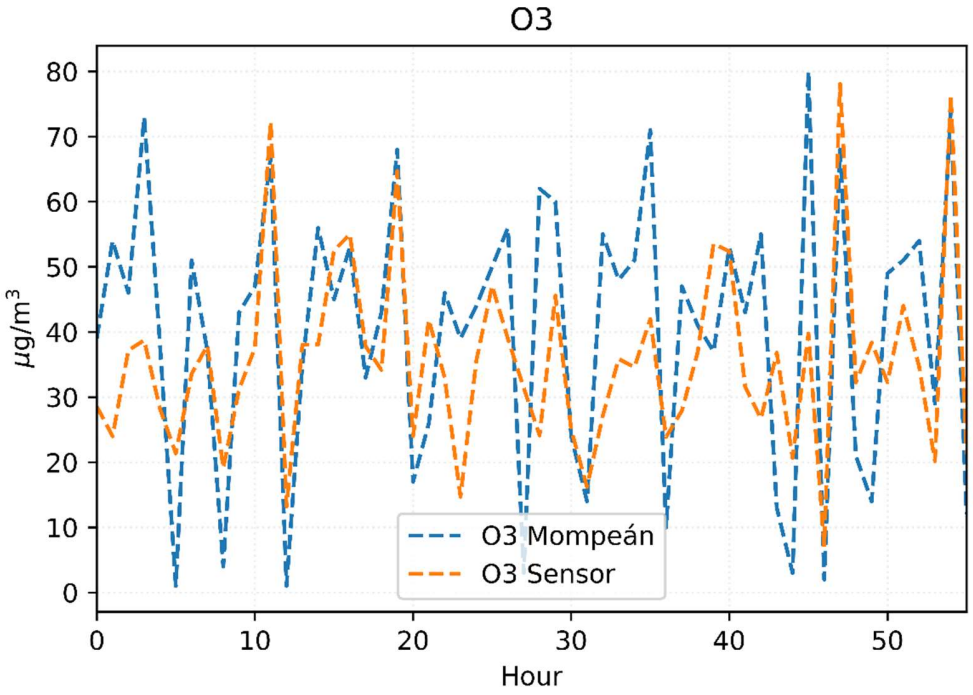

Figure 13: Plot between O3 levels from sensor and weather station and hours considering temperature and humidity.

Figures above are the plot of gases level from weather station and calibrated sensor with hours. The curve with blue dashed line represents weather station and orange dashed line represent sensor calibrated values.

The dependency of gases level in the atmosphere collected by sensor with temperature and humidity can be seen from the above curves. Considering temperature and humidity makes the model to predict value nearer to the target value. The curve drawn with calibrated data from sensor align more with the value from weather station. Thus, there is the dependency of gases level in atmosphere with temperature and humidity. It is seen more clearly from the plot of 03. There is large gap between sensor calibrated curve and weather station in first plot of O3 but it improves and the curves aligns more when considering temperature in second plot. In the third plot, there is very less changes thus, O3 sensor has more dependency with temperature and slightly on humidity.

3.2 Polynomial Regression

Polynomial regression models with degree two are created. The parameters are temperature and sensor data from worker and auxiliary pins. The curves obtained using Polynomial regression model for different gases plotted on test dataset are shown below.

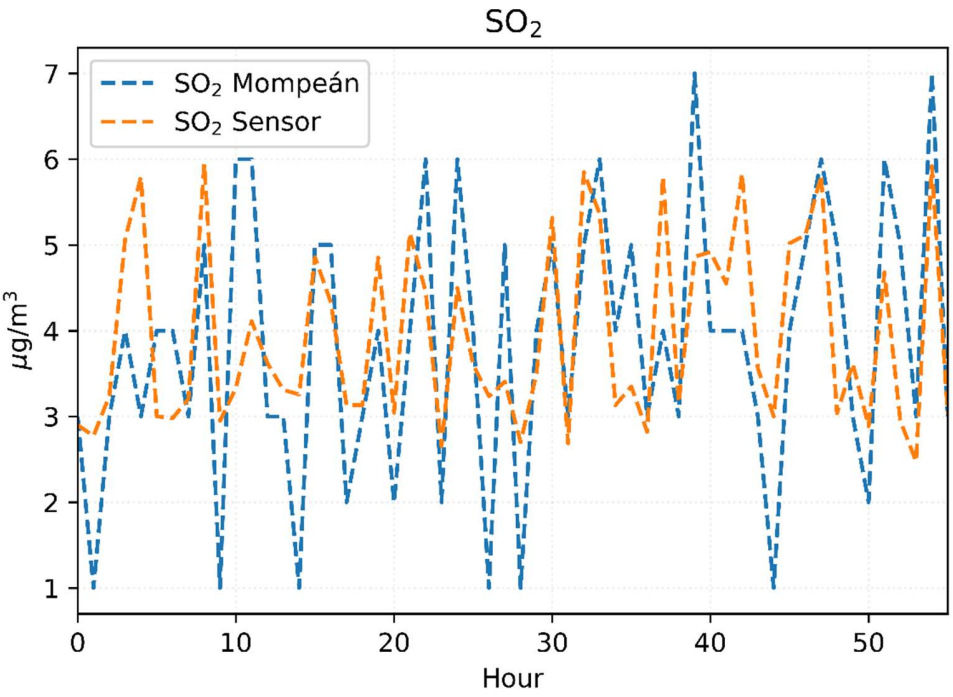

Figure 14: Polynomial plot between SO2 levels from sensor and weather station and hours.

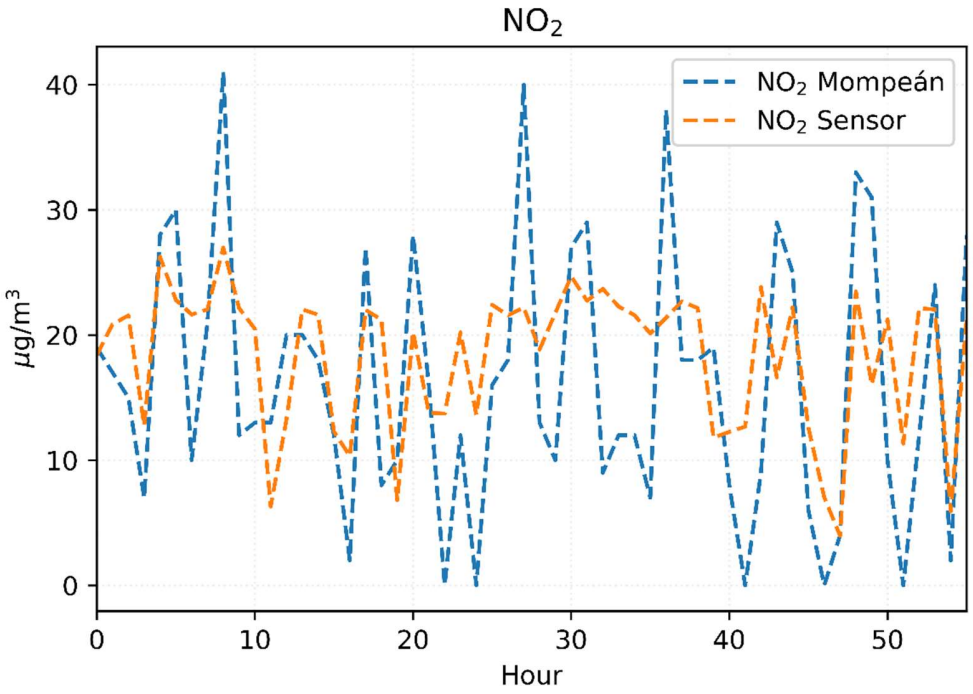

Figure 15: Polynomial plot between NO2 levels from sensor and weather station and hours.

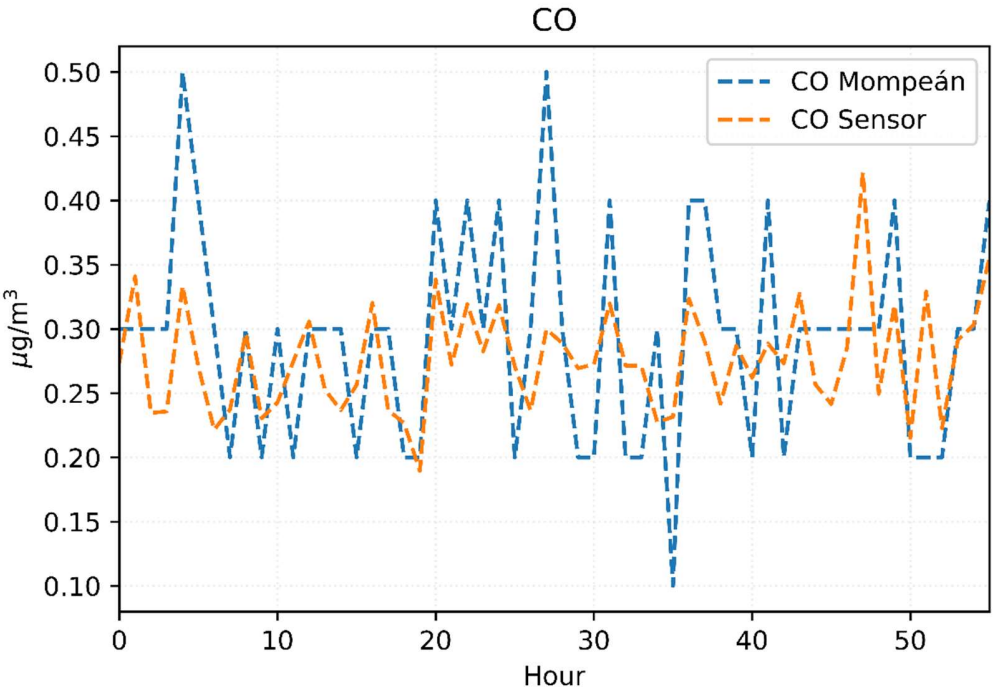

Figure 16: Polynomial plot between CO levels from sensor and weather station and hours.

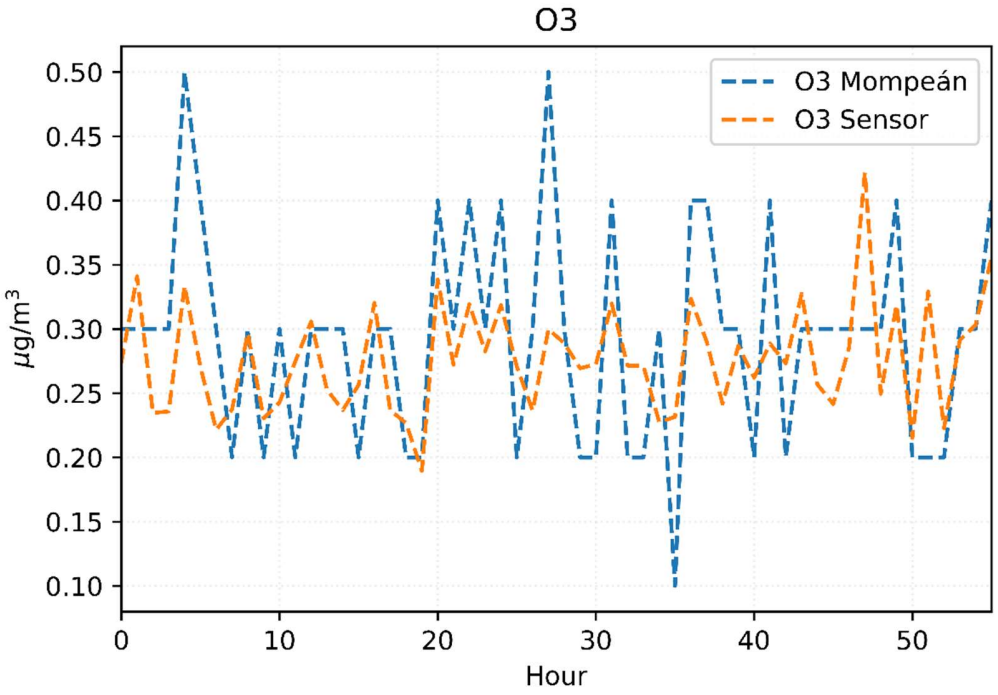

Figure 17: Polynomial plot between O3 levels from sensor and weather station and hours.

Using polynomial regression of degree two, the curve fits more closely to the actual values. Thus, a more improved model is obtained.

Table 1: Regression Score Table

Sensor	Polynomial regression score	Linear regression score
NO2	0.28550	0.20165
SO2	0.3280	0.2366
CO	0.4122	0.3819
O3	0.5845	0.5115

The table above consists of the regression scores of different sensors obtained using both polynomial and linear regression. The regression score for a particular gas sensor using polynomial regression is always higher than linear regression.

3.3 Deep Neural Network

A deep neural network with three hidden layers is created. The input layer consists of data of sensor auxiliary and worker pin, Temperature and Humidity. The plots obtained using deep neural network are shown below.

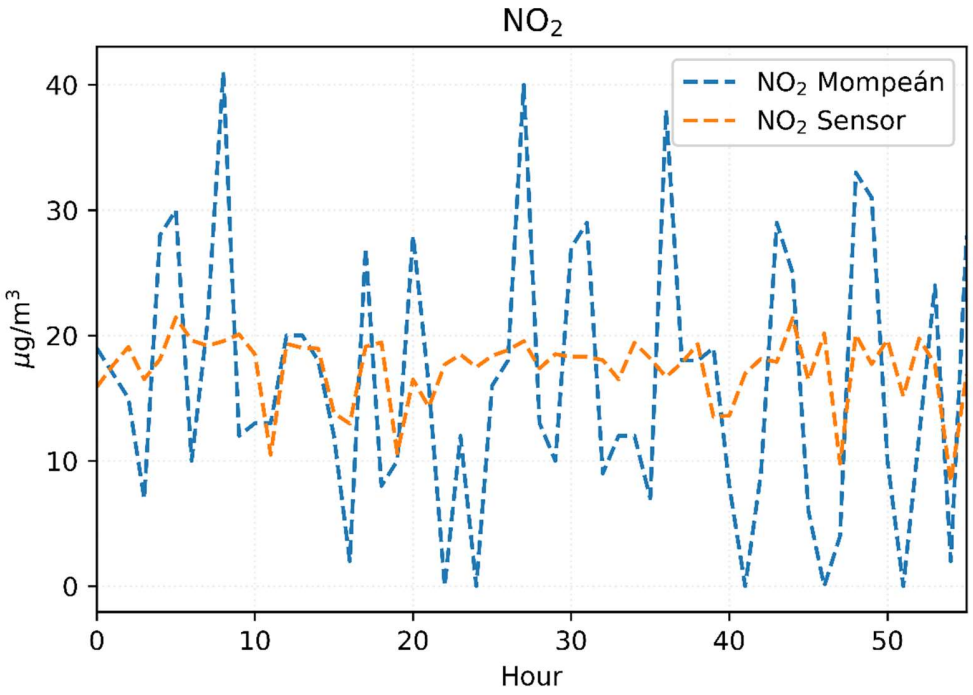

Figure 18: Plot between NO2 levels from sensor and weather station and hours.

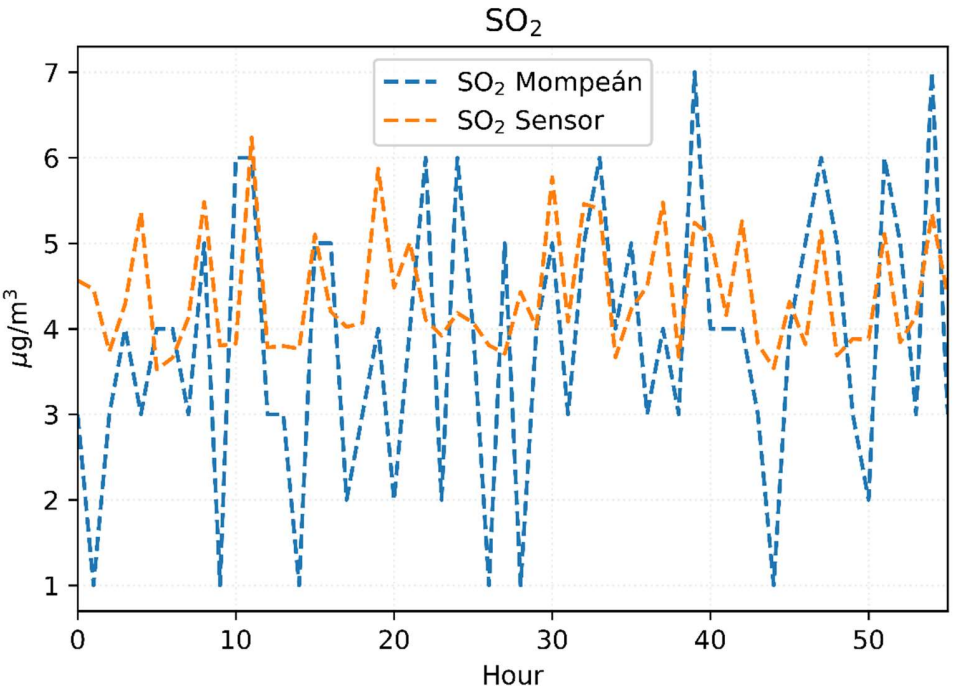

Figure 19: Plot between SO2 levels from sensor and weather station and hours.

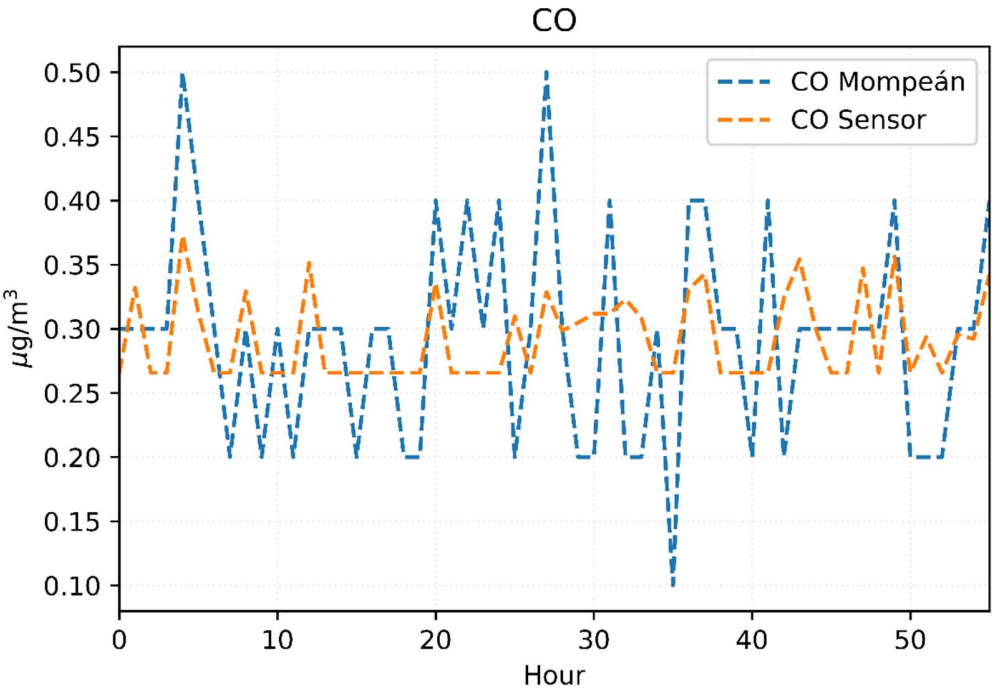

Figure 20: Plot between CO levels from sensor and weather station and hours.

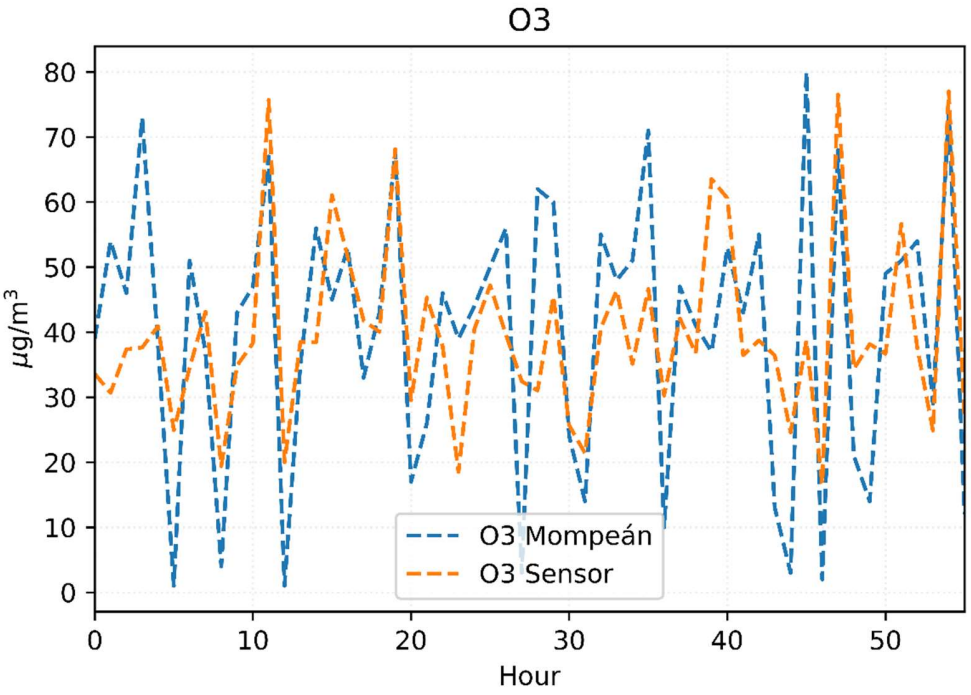

Figure 21: Plot between O3 levels from sensor and weather station and hours.

CHAPTER 4: CONCLUSION

The main aim of the project was to calibrate the data from the sensor and to make necessary adjustment to correct it. The system was implemented in Python programming language along with various other python libraries. The different graphs obtained using various algorithms shows the calibration of data of sensor. The model is tested on the data that was separated for testing and it showed satisfactory result.

CHAPTER 5: LIMITATIONS

There are some limitations of this project. The data obtained from the sensor is not consistent. There is huge variation in temperature, humidity range for a similar target output. One reason for that may be due to the sensor being used for collection of data from different manufacturer. As the deep learning model works well with large amount of data. The data set for this project was also limited. These limitation cause model not to work well on unseen data. The adjustment parameters obtained to correct data was not perfect.