Indian Institute of Technology Bombay

Modeling & Identification of Gas sensor data for dynamic gas mixture

Course : CL 625 Process Modelling and System Identification

Author : Harsh Maheshwari

Guide : Prof. Sharad Bhartiya,

Chemical Engineering Department, IIT Bombay

Sunday 10^{th} November, 2019

Abstract

This paper discusses the use of different modelling techniques used for system identification. The system data is taken from the open source UCI Machine Learning Repository. Its is listed as Gas sensor array under dynamic gas mixtures data set.Different models from equation error family, output error family, Prediction Error methods and some non linear methods are discussed for the data. Best model is determined for simulation and prediction tasks. (Will be updated more in final submission)

1 Introduction

The Dynamic System used in this project is obtained from the UCI Machine Learning Repository and is titled as Gas sensor array under dynamic gas mixtures data set. This data set contains the acquired time series from 16 chemical sensors exposed to gas mixtures at varying concentration levels. In particular, it has two gas mixtures: Ethylene and Methane in air, and Ethylene and CO in air. Each measurement was constructed by the continuous acquisition of the 16-sensor array signals for a duration of about 12 hours without interruption.

1.1 The experimental setup and procedure as reported by the curators of dataset

The sensor array included 16 chemical sensors of 4 different types: TGS-2600, TGS-2602, TGS-2610, TGS-2620 (4 units of each type). The sensors were integrated with customized signal conditioning and control electronics. The operating voltage of the sensors, which controls the sensors operating temperature, was kept constant at 5 V for the whole duration of the experiments. The sensors conductivity were acquired continuously at a sampling frequency of 100 Hz. The sensor array was placed in a 60 ml measurement chamber, where the gas sample was injected at a constant flow of 300 ml/min.

Each measurement was constructed by the continuous acquisition of the 16-sensor array signals while concentration levels changed randomly. For each measurement (each gas mixture), the signals were acquired continuously for about 12 hours without interruption. The concentration transitions were set at random times (in the interval 80-120s) and to random concentration levels. The data set was constructed such that all possible transitions are present: increasing, decreasing, or setting to zero the concentration of one volatile while the concentration of the other volatile is kept constant (either at a fixed or at zero concentration level). At the beginning, ending, and approximately every 10,000 s.

The concentration ranges for Ethylene, Methane, and CO were selected such that the induced magnitudes of the sensor responses were similar. Moreover, for gas mixtures, lower concentration levels were favored. Therefore, the multivariate response of the sensors to the presented set of stimuli is challenging since none of the configurations (single gas or mixture presentation) can be easily identified from the magnitude of sensors responses. In particular Ethylene concentration ranges from 0-20 ppm; 0-600 ppm for CO; and 0-300 ppm for Methane. The data set contains columns given as Time (seconds), Methane conc (ppm), CO conc (ppm), sensor readings (16 channels) : TGS2602; TGS2602; TGS2600; TGS2600; TGS2610; TGS2610; TGS2620; TGS2620; TGS2602; TGS2602; TGS2600; TGS2600; TGS2610; TGS2610; TGS2620; TGS2620

2 Objective

This paper aims to develop and identify a model relationship that can predict the concentration of gases in air when given sensors readings at a particular time. For this different models from output error family and equation error family are used and then accuracy of these models is tested. Further it aims to build a model to predict the sensor readings when the concentration of gases are given. One might think why is the second model even needed? The answer lies in its business value one

can test if a sensor is working fine just by using the model. This can help industries check for faulty sensors very easily without much cost.
The two models can be shown as :

Model 1 : $Y = G * U + H * e$

where , y_1 = Conc. CO , y_2 = Conc. Ethylene, u_i = Sensor-i, $Y = [y_1, y_2]$, $U = [\{u_i\}]$

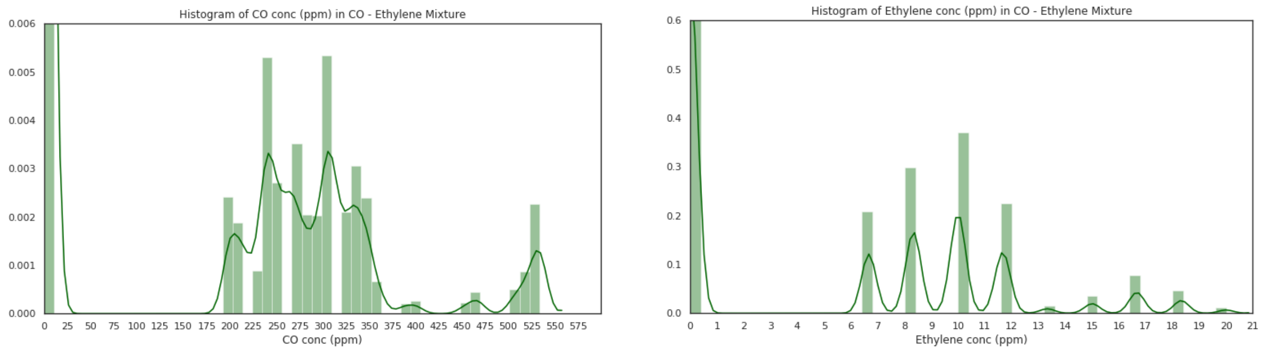
Model 2 : $Y = G * U + H * e$

where , u_1 = Conc. CO, u_2 = Conc. Ethylene, y_i = Sensor-i, $U = [u_1, u_2]$,$Y = [\{y_i\}]$

3 Exploratory Data Analysis

3.1 Input Output Analysis

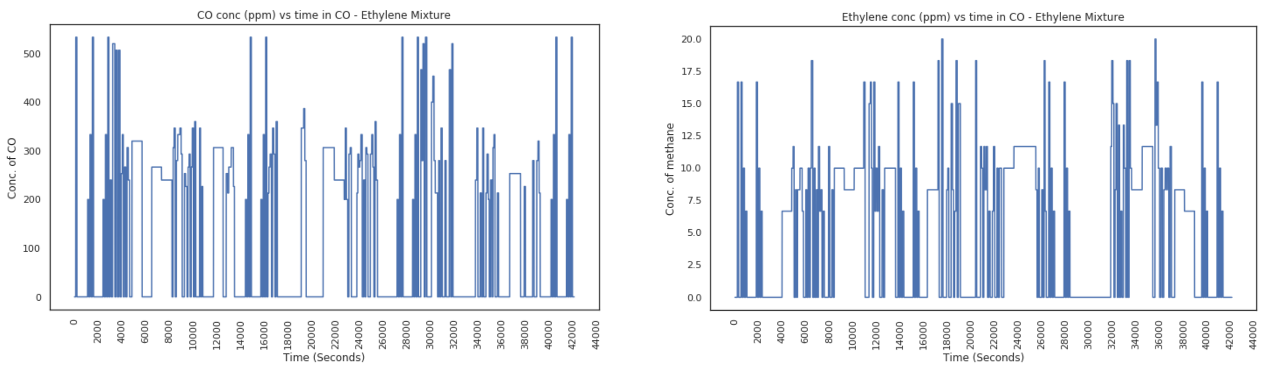
Following graphs were plotted while analysing the data set.



(a) Histogram of CO Conc (ppm) in CO - Ethylene Mixture

(b) Histogram of Ethylene Conc (ppm) in CO - Ethylene Mixture

Figure 1: Histograms of Concentration



(a) CO Conc (ppm) vs time in CO - Ethylene Mixture

(b) Ethylene Conc (ppm) vs time in CO - Ethylene Mixture

Figure 2: Conc. vs Time

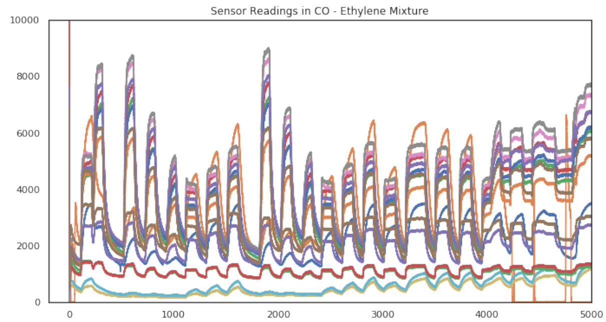
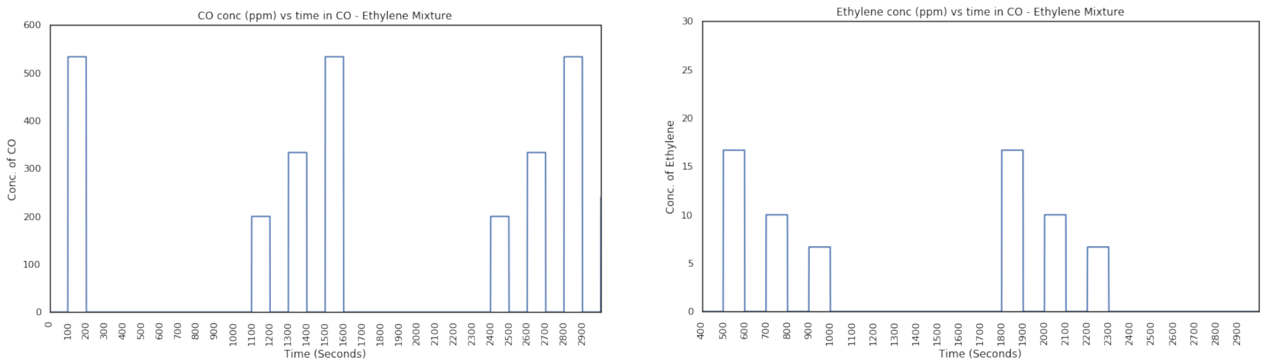


Figure 3: Sensors Readings in CO vs Time

Figure 1: Histograms of concentration for different gases, Figure 2: Concentration vs time plots and Figure 3 Sensor readings vs time plots. It can be seen from Figure 2 that concentration of CO and Ethylene do not change in a linear fashion. Figure 3 shows that different sensor readings generally follow a patterns with respect to time. To get a better understanding of Sensor variations with time see Appendix Figure 5 and Figure 6.

3.2 Sampling Time

The sampling time in the original data is that of 0.01 sec i.e 1 millisecond. This is a really low sampling time i.e high sampling frequency



(a) Zoomed CO Conc (ppm) in CO - Ethylene Mixture

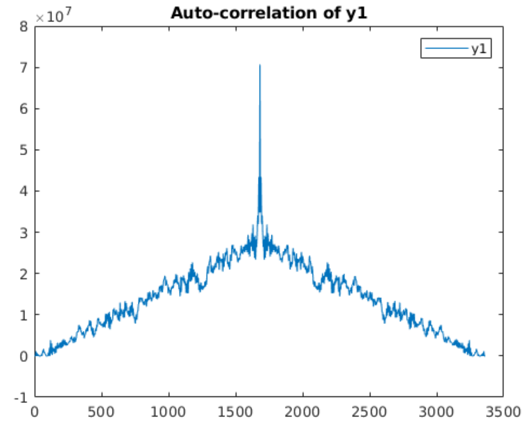
(b) Zoomed Ethylene Conc (ppm) in CO - Ethylene Mixture

Figure 4: Histograms of Concentration

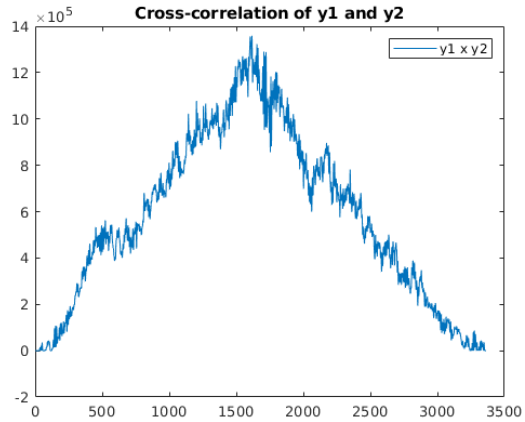
From Figure 4 we can see that the time period in the most dense regions in Figure 2 of the the data is about 1400 sec and hence we should take minimum sampling time of 700 sec. (Shannon sampling theorem). Now if we increase the sampling time from 0.01 sec to 700 sec the number of data points will reduce drastically.
For $T_s = 70$ sec we get $10 * T_s$ in one $T_p/2$. Hence take a sampling time much bigger than 0.01 sec but much lower than 700 sec say, 25 sec.
After sampling the data at sampling time 25 sec there are 1680 timestamps remaining in the new data set and from this point on this data will be used for all analysis

3.3 Correlation in the data

Auto-correlations in the output concentrations: from Figure 5 can see that their is a general increase in auto correlation for both the outputs and Figure 6 tells us that this auto correlation is runs around 0 when deviation from mean is used for plotting. Similar features can be observed in Figure 7 for cross correlations. No Delay is observed as Auto correlations exist on t = 1680

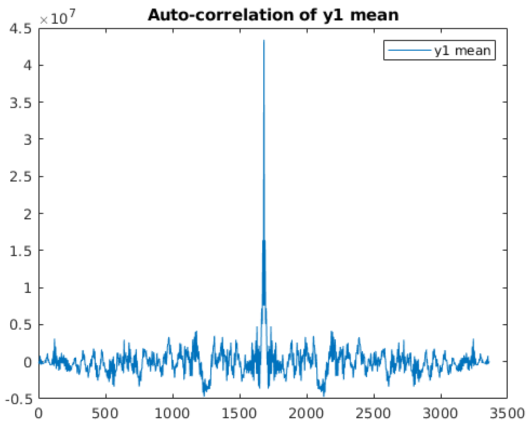


(a) Auto Correlation in y1

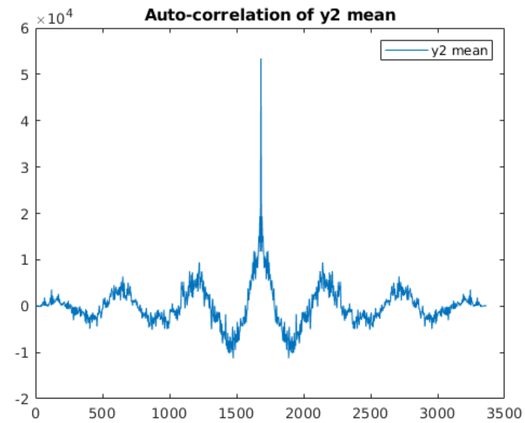


(b) Auto Correlation in y2

Figure 5: Auto Correlation of the output concentration

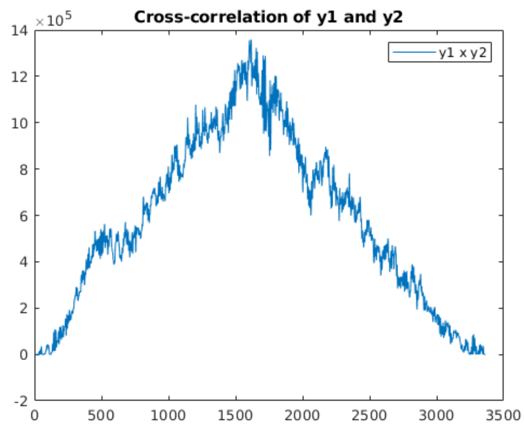


(a) Auto Correlation in y1-mean_Of_y1

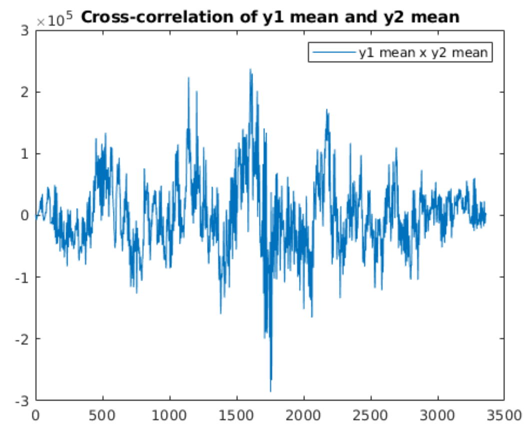


(b) Auto Correlation in y2-mean_Of_y2

Figure 6: Auto Correlation of the output concentration in deviation from mean



(a) Cross Correlation in y1 x y2

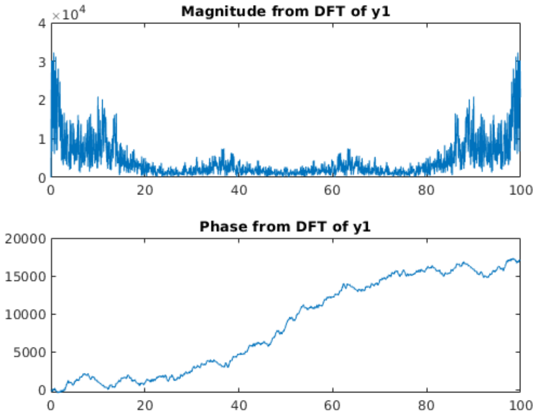


(b) Auto Correlation in y1-mean_Of_y1 and y2-mean_Of_y2

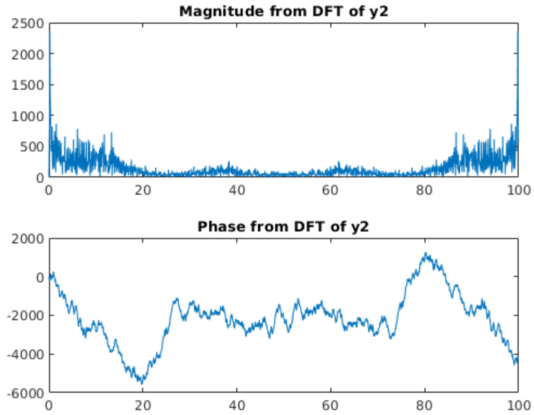
Figure 7: Cross Correlation of the output concentration in deviation from mean

3.4 Discrete Fourier Transformation

The Discrete Fourier Transformation of y1 and y2 has been plotted in Figure 8 and we observe that their is just a spike on 0 and then it tends to drop down.



(a) DFT of y1

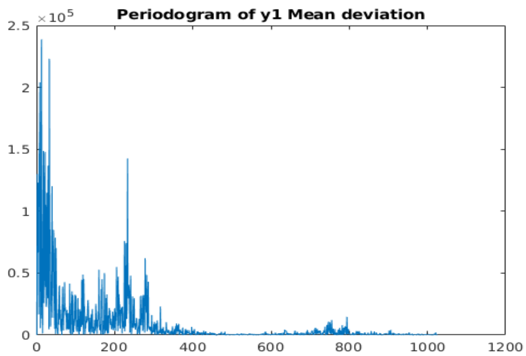


(b) DFT of y2

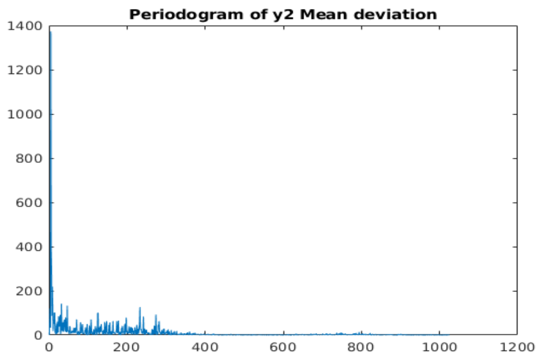
Figure 8: Discrete Fourier Transformation

3.5 Power Spectral Density Analysis

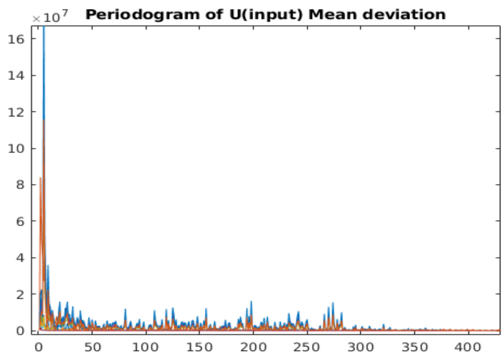
From the Figure 9 we can see that the power is majorly at lower frequencies in both input and output variables.



(a) PSD of y1 Mean Deviation



(b) PSD of y2 Mean Deviation

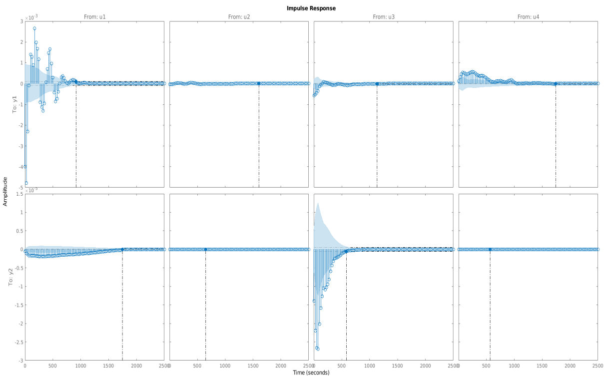


(c) PSD of (Input) U Mean Deviation

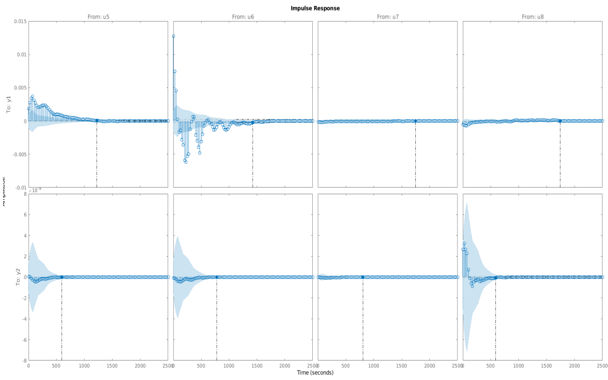
Figure 9: Periodograms

3.6 Impulse Response

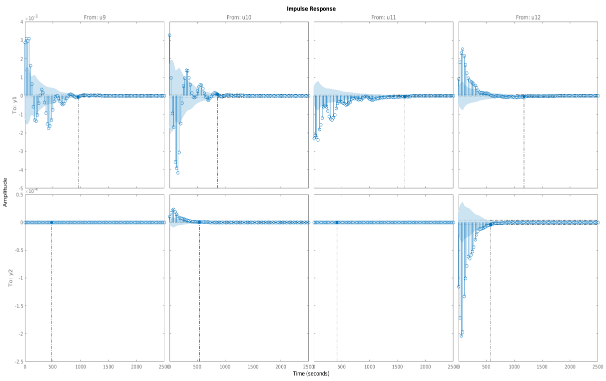
Before moving further in our analysis we will divide our data in test and train sets. First 1400 points are taken as train remaining are considered for test.
Now Impulse response and step response coefficients as a solution to the least squares problem is found using Matlab and the below impulse response and step responses are plotted.



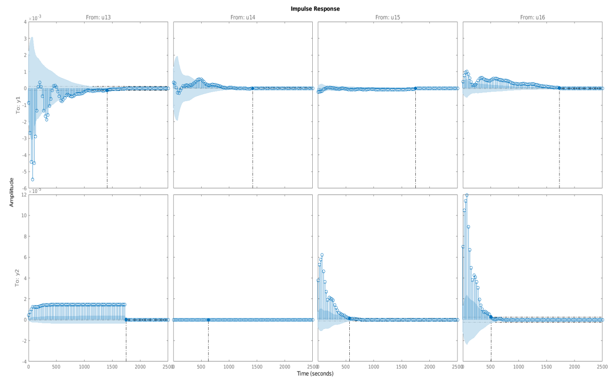
(a) Impulse estimates u1 to u4



(b) Impulse estimates u5 to u8



(c) Impulse estimates u9 to u12



(d) Impulse estimates u13 to u16

Figure 10: Impulse responses

Now , white noise is injected in the output of the process. Use a noise to signal ratio of 1% and 10%. Using these data sets the impulse response coefficients are calculated again.
The final noisy data is plotted with the original data and is shown in figure below

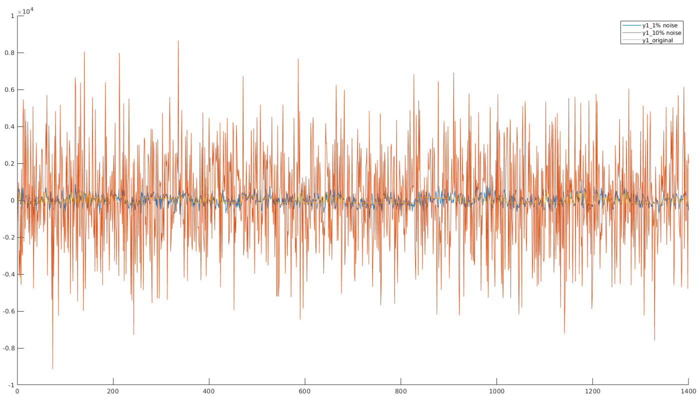
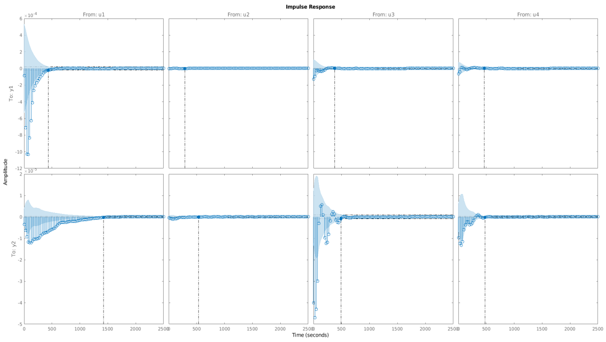
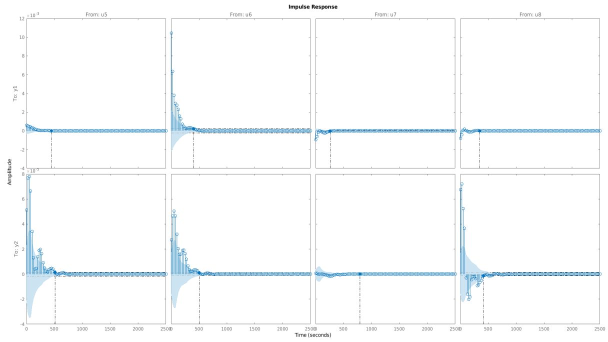


Figure 11: One percent and Ten percent Noise data

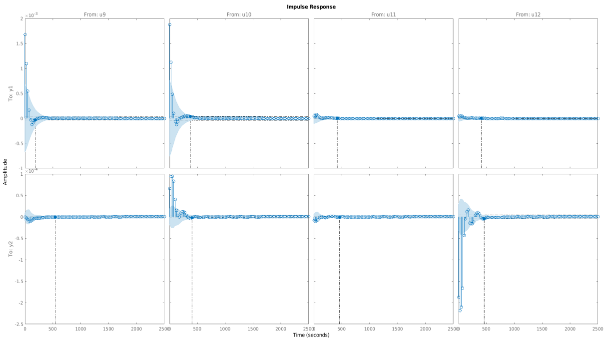
The impulse response coefficients are calculated again for the new data sets with 1% and 10% noise and are plotted below.



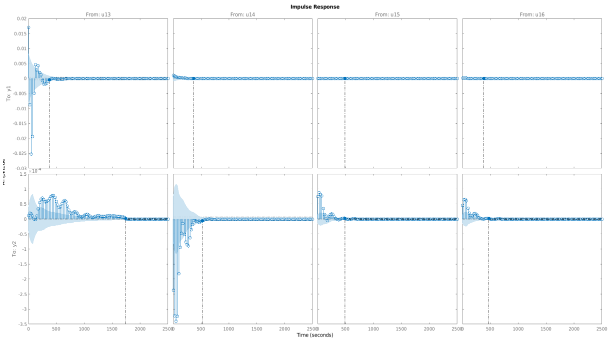
(a) Impulse estimates u1 to u4



(b) Impulse estimates u5 to u8

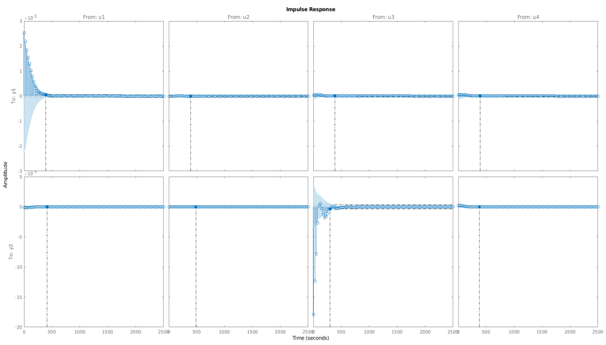


(c) Impulse estimates u9 to u12

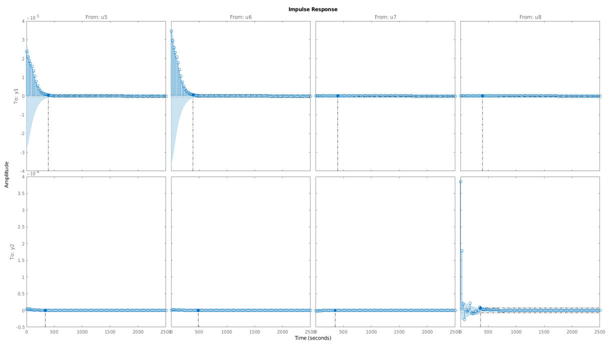


(d) Impulse estimates u13 to u16

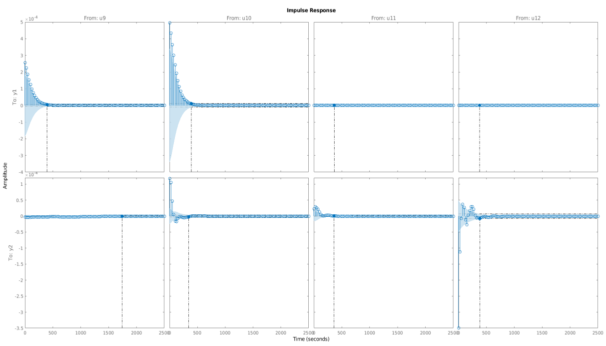
Figure 12: Impulse responses for One percent noise to signal ratio



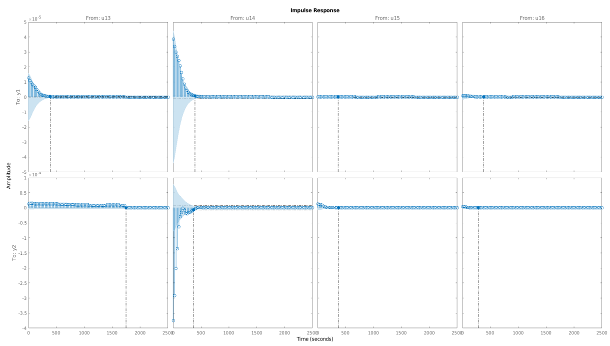
(a) Impulse estimates u1 to u4



(b) Impulse estimates u5 to u8



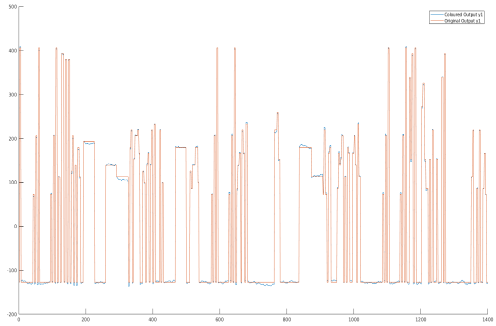
(c) Impulse estimates u9 to u12



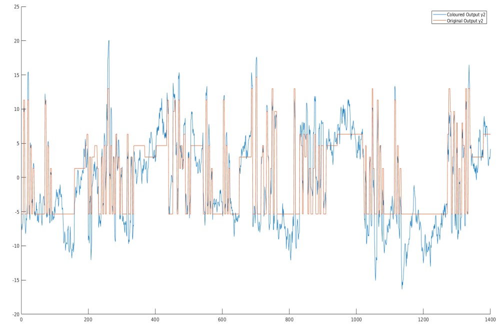
(d) Impulse estimates u13 to u16

Figure 13: Impulse responses for Ten percent noise to signal ratio

Now coloured noise is injected in the output of the process modeled as $v(t) = \frac{1}{(1-0.95q^{-1})}e(t)$



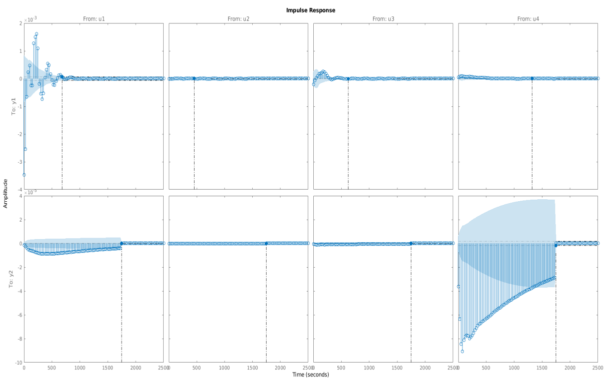
(a) Coloured Noise Y1 vs Original Y1



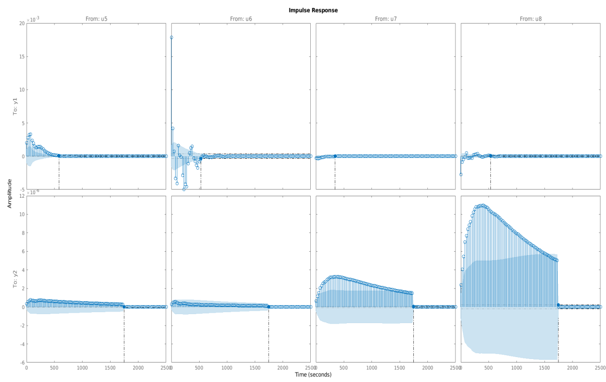
(b) Coloured Noise Y2 vs Original Y2

Figure 14: Colour Noise Data

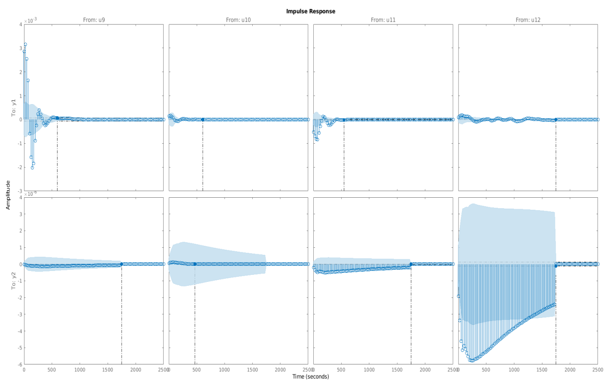
With this noise impulse response and step response is calculated and is given in the figure below.



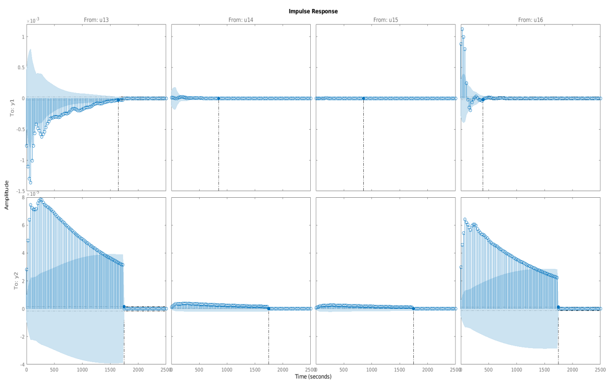
(a) Impulse estimates u1 to u4



(b) Impulse estimates u5 to u8



(c) Impulse estimates u9 to u12

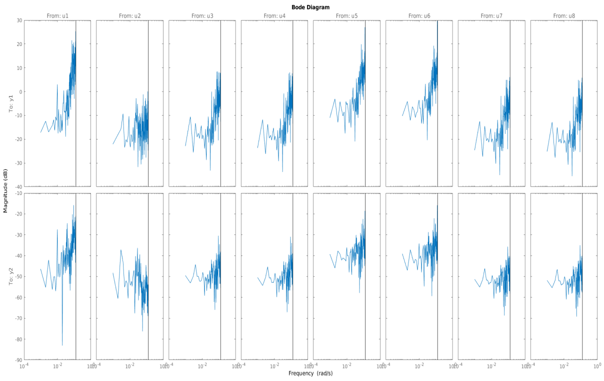


(d) Impulse estimates u13 to u16

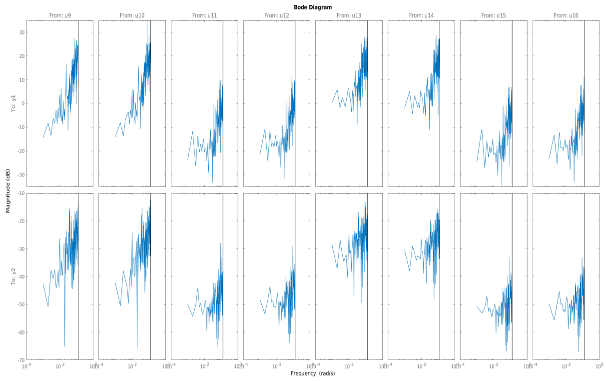
Figure 15: Impulse responses for Ten percent noise to signal ratio

3.7 ETFE

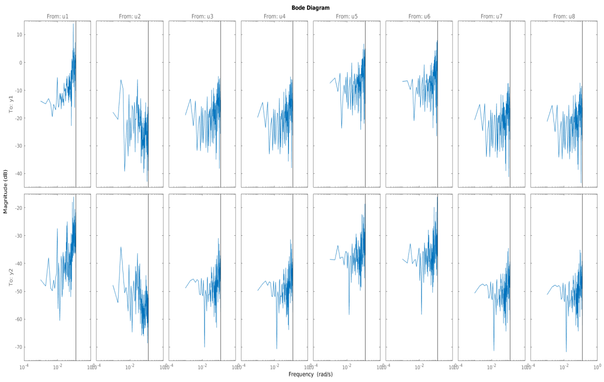
Using the data obtained for the white noise with snr 1% a ETFE for the system is plotted in the form of a bode plot. And then same is done for the Coloured noise.



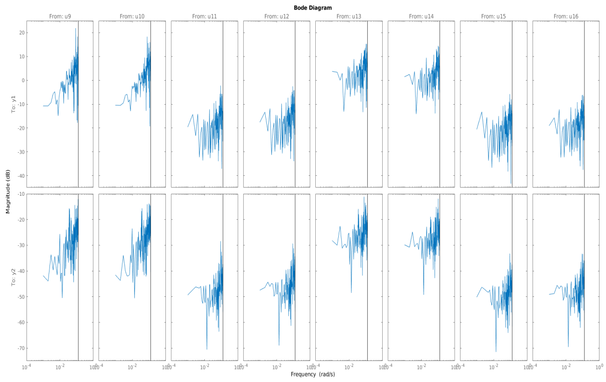
(a) ETFE estimates u1 to u8 1% noise



(b) ETFE estimates u9 to u16 1% noise



(c) ETFE estimates u1 to u8 coloured noise



(d) ETFE estimates u9 to u16 coloured noise

Figure 16: ETFE for 1% noise and coloured noise

4 Results

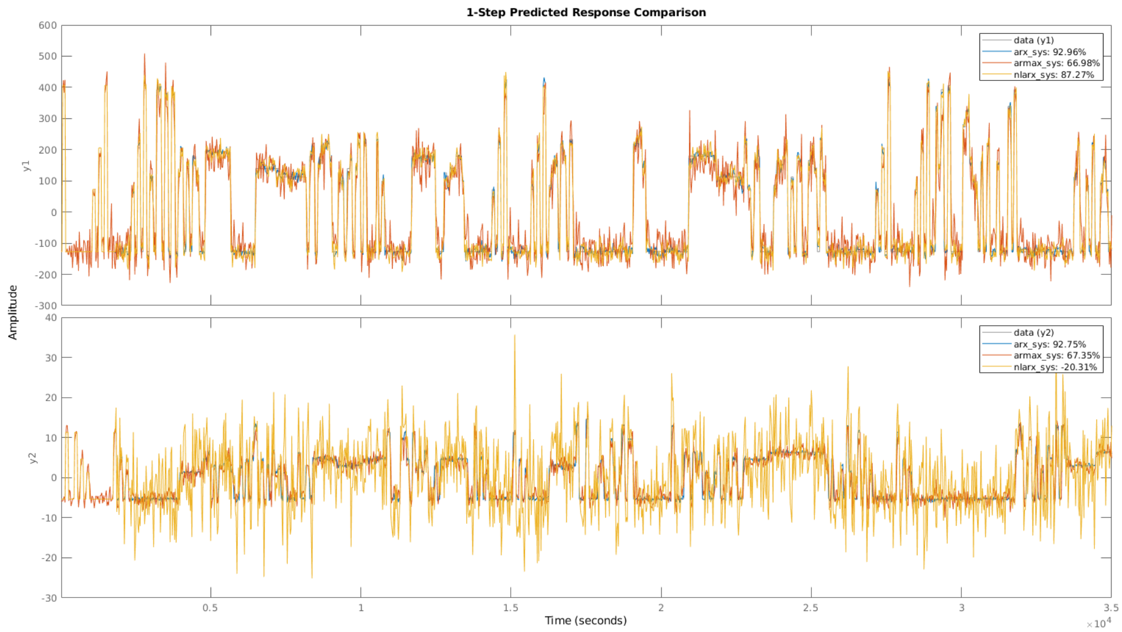
Now we model the Output from different models with one step ahead prediction. Comparison between best ARX best ARMAX and best N4sid models is shown below

1. ARX:
Fit to estimation data: [92.96;92.75]% (prediction focus)
MSE: 132.3

2. ARMAX:
Fit to estimation data: [66.98;67.35]% (prediction focus)
MSE: 2911

3. N4sid:
Fit to estimation data: [87.27;-20.31]% (prediction focus)
MSE: 478.8

We found that an ARX model is best suited for this process and we can predict the system with accuracy of more than 90%. Them closest is the ARMAX Model with accuracy of around 67% for



(a)

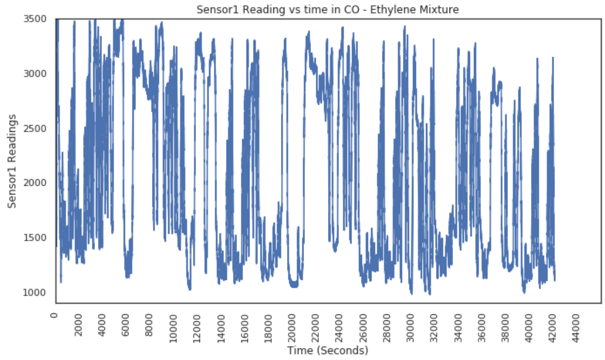
Figure 17: Comparison of different models

both the output concentrations. N4sid Models the system output 1 pretty correctly but it fails to model output 2, hence is not a good option

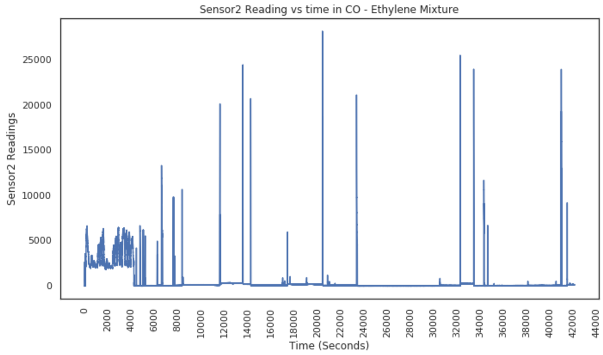
5 References

1. Fonollosa et al. 'Reservoir Computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring'
2. https://in.mathworks.com/products/sysid.html [Sys Id matlab Toolbox]

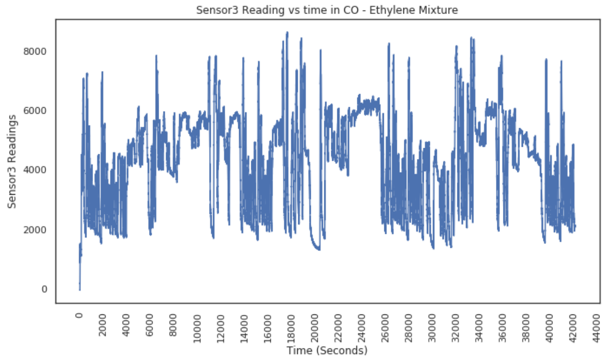
6 Appendix



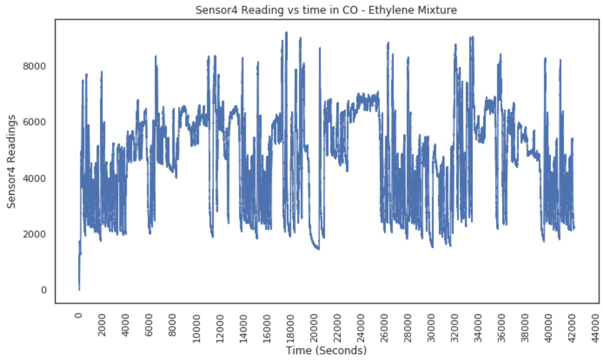
(a) Sensor1 Reading vs time in CO - Ethylene Mixture



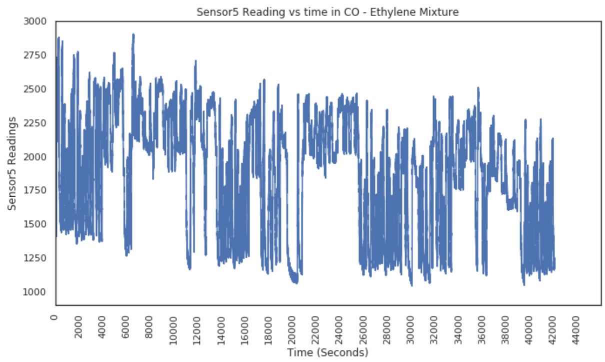
(b) Sensor2 Reading vs time in CO - Ethylene Mixture



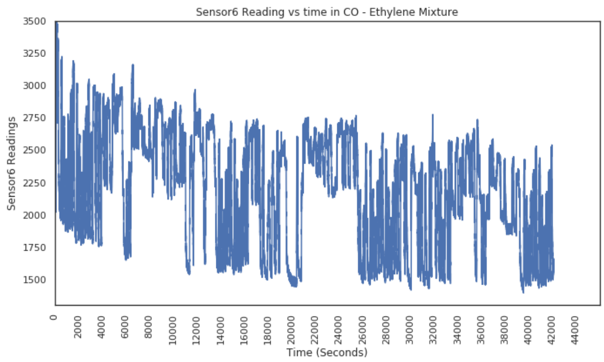
(c) Sensor3 Reading vs time in CO - Ethylene Mixture



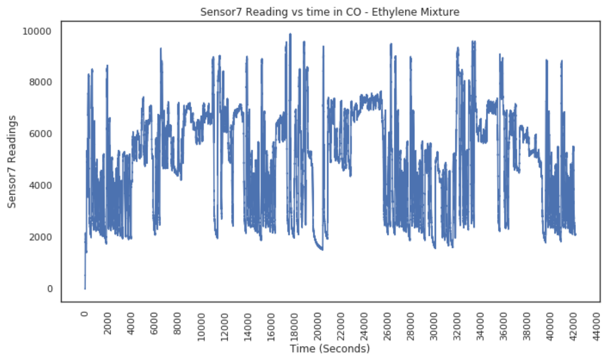
(d) Sensor4 Reading vs time in CO - Ethylene Mixture



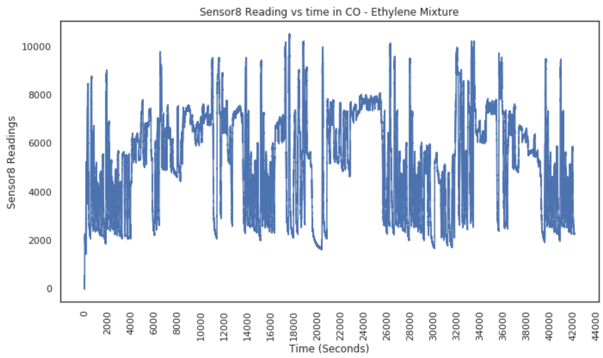
(e) Sensor5 Reading vs time in CO - Ethylene Mixture



(f) Sensor6 Reading vs time in CO - Ethylene Mixture

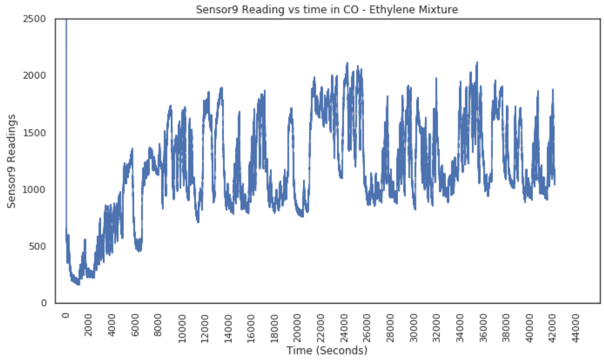


(g) Sensor7 Reading vs time in CO - Ethylene Mixture

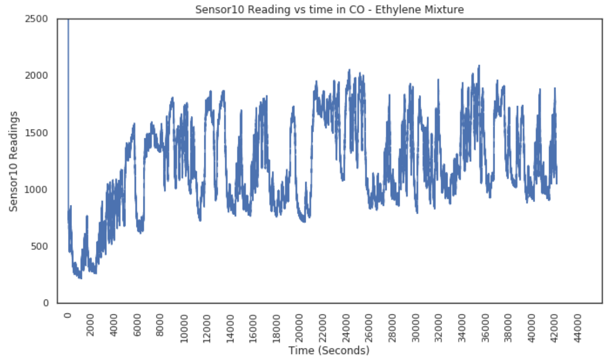


(h) Sensor8 Reading vs time in CO - Ethylene Mixture

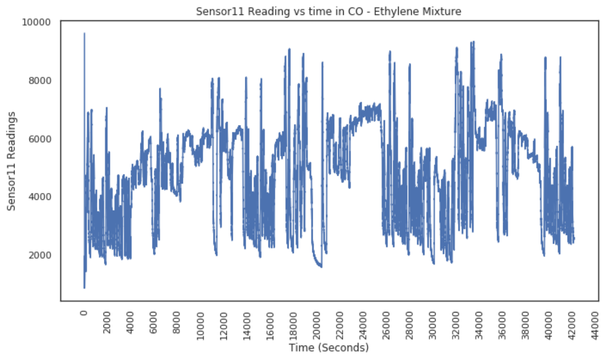
Figure 18: Sensor(1-8) vs Time



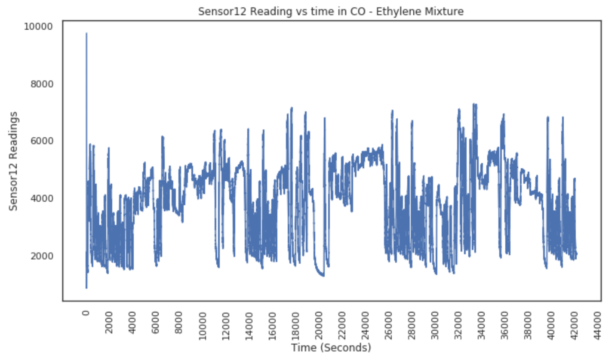
(a) Sensor9 Reading vs time in CO - Ethylene Mixture



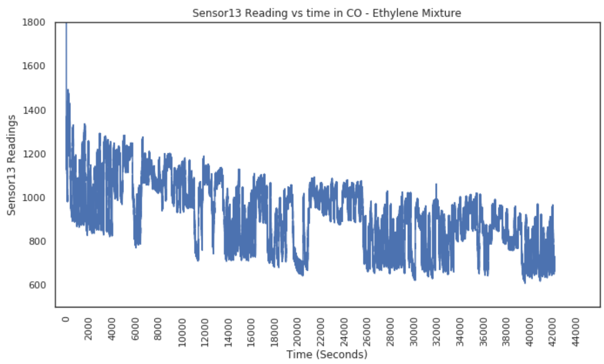
(b) Sensor10 Reading vs time in CO - Ethylene Mixture



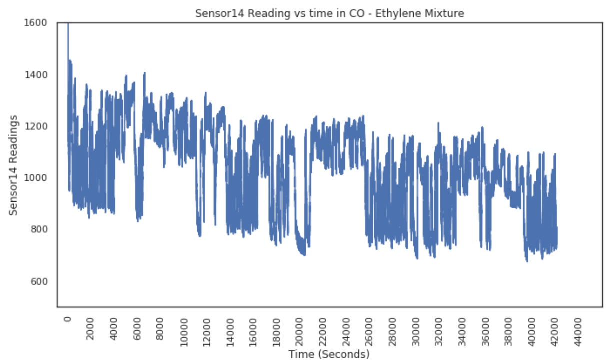
(c) Sensor11 Reading vs time in CO - Ethylene Mixture



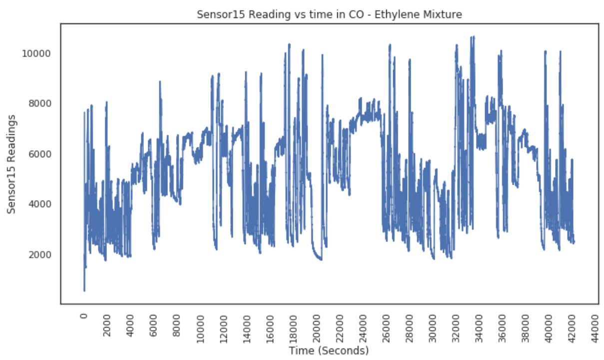
(d) Sensor12 Reading vs time in CO - Ethylene Mixture



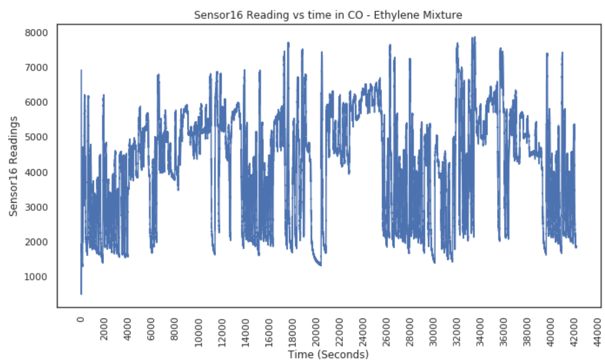
(e) Sensor13 Reading vs time in CO - Ethylene Mixture



(f) Sensor14 Reading vs time in CO - Ethylene Mixture



(g) Sensor15 Reading vs time in CO - Ethylene Mixture



(h) Sensor16 Reading vs time in CO - Ethylene Mixture

Figure 19: Sensor(9-16) vs Time