

Linköping University | Department of Computer and Information Science

Master's thesis, 30 ECTS | Statistics and Machine Learning

2021 | LIU-IDA/LITH-EX-A--2021/001--SE

Quantifying nitrogen oxides and ammonia via frequency modulation in gas sensors

Marcos Freitas Mourão dos Santos

Supervisor : Annika Tillander

Examiner : José M. Peña

External supervisor : Mike Andersson

Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtida ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innehåller rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida <http://www.ep.liu.se/>.

Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: <http://www.ep.liu.se/>.

Abstract

The use of Silicon Carbide Field Effect Transistor (SiC-FET) sensors in cyclic operation is a proven way to quantify different gases. Standard workflow involves extracting shape defining features such as average and slope of the sensor signal. This work's main goal is to verify if frequency modulation can be used to simultaneously quantify Nitric Oxide (NO), Nitrogen Dioxide (NO₂) and Ammonia (NH₃). Linear models were chosen, namely: Ordinary Least Squares (OLS), Principal Components Regression (PCR), Partial Least Squares Regression (PLSR) and Ridge regression. Results indicate that these models fail to predict concentrations completely for every gas. Analysis indicates that the features are not linear in terms of concentrations. This work is concluded by recommending a few other alternatives before discarding frequency cycling completely: non-parametric models of regression and different frequency regime, namely the use of triangular waves in future experiments.

Acknowledgments

Foremost, I would like to extend my deepest gratitude to my internal supervisor, Annika Tillander, for her insightful comments, feedback, and continuous support. I cannot stress enough her patience, availability, compassion, and constructive criticism enough.

I am also extremely grateful to my external supervisor, Mike Andersson, who offered me this project in the first place and whose support was indispensable throughout this thesis. Our formal and casual conversations, lab visits, and meetings surely made an already exciting topic even more enjoyable. Special thanks to Lida Khajavizadeh, who was co-responsible for lab experiments and data collection.

Moreover, I am deeply indebted to have José M. Peña as my examiner. I feel he greatly enhanced my work through his thorough review, appreciation, meaningful comments and valuable suggestions.

I also sincerely thank Samia Noreen Butt for her work as my opponent, helping my work to be as good as it can be through her feedback.

I would like to acknowledge the assistance of my colleagues Erik Rosendal and Mudith Chathuranga Silva, who helped shape this report via proof-reading and excellent suggestions. For that, I thank you.

Additionally, I would like to thank my friends Agustín Valencia, Bayu Brahmantio, José Mendez, and Ismail Khalil for their continuous support and meaningful discussions during our thesis work.

Finally, I would like to dedicate this work to my beloved parents, Katia and Marcio. Unfortunately, there is not enough space in this thesis to even begin describing how grateful I am to them.

To all of you, I offer my sincere appreciation, respect, and gratitude.

Sincerely,

Marcos Freitas Mourão dos Santos

Contents

Abstract	iii
Acknowledgments	iv
Contents	v
List of Figures	vii
List of Tables	ix
List of acronyms and abbreviations	x
1 Introduction	1
1.1 Motivation	1
1.2 Aim	2
1.3 Research questions	2
2 Data	3
2.1 Data acquisition	3
2.2 Raw data	5
2.3 Pre-processing	6
3 Theory	10
3.1 Notation	10
3.2 Ordinary Least Squares Regression	11
3.3 Principal Component Analysis	12
3.4 Principal Component Regression	14
3.5 Partial Least Squares Regression	15
3.6 Ridge Regression	17
3.7 Cross Validation	18
4 Methods	20
4.1 Ordinary Least Squares	21
4.2 Principal Components Regression	21
4.3 Partial Least Squares Regression	21
4.4 Ridge Regression	21

5 Results	22
5.1 Ordinary Least Squares	24
5.2 Principal Components Regression	25
5.3 Partial Least Squares Regression	28
5.4 Ridge Regression	31
6 Discussion	34
6.1 Results	34
6.2 Method	35
6.3 The work in a wider context	35
7 Conclusion	37
Bibliography	38
Appendix A Data acquisition time stamps	40
Appendix B Other data plots	42

List of Figures

2.1	Schema of the data acquisition process.	3
2.2	An example of raw sensor response	4
2.3	Feature measurements times per cycle. The width of the red line indicates the duration of one of the feature measurement windows as an example.	5
2.4	A visualization of the feature measurement process.	5
2.5	Feature naming convention.	7
2.6	Pre-processed data structure.	7
2.7	A visualization of the feature averaging process.	8
2.8	(a) Slope and (b) average features, both un-normalized and normalized.	9
5.1	Correlation matrix of features.	22
5.2	Actual vs. Predicted for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	24
5.3	PCA for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	25
5.4	Explained variance of PC for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	26
5.5	Cross-validation results for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	26
5.6	PCR for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	27
5.7	PLS scores for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	28
5.8	Explained variance of PLS components for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	29
5.9	Cross-validation results for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	29
5.10	PLSR for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	30
5.11	Cross-validation results for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	31
5.12	Coefficient shrinkage given λ (a) slopes and averages through exposures and (b) only averaged average features through mixtures. Each line corresponds to a coefficient/feature	32

5.13 Ridge regression predictions for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.	33
B.1 Averaged sensor average divided by predominant gas. Each line corresponds to an unique mixture.	43
B.2 Normalized sensor averaged per gas. Each line corresponds to an unique mixture. The levels are the concentrations of individual components of the mixture: (a) NO (b) NO ₂ , and (c) NH ₃	44

List of Tables

2.1	Data acquisition details	4
2.2	Raw data column details	6
2.3	Sample of raw data.	6
2.4	Sample of pre-processed data.	8
2.5	Sample of averaged features throughout all exposures data.	8
A.1	Data acquisition timestamps.	41

List of acronyms and abbreviations

AC Alternating Current. 2

CV Cross Validation. 19, 21, 35, 37

GBCO Gate Bias Cycled Operation. 2

Hz Hertz. 4, 6

mA milliamperes. 3

MSE Mean Squared Error. 18

NIPALS Nonlinear Iterative Partial Least Squares. 13, 14, 16

OLS Ordinary Least Squares. iii, v, vi, 11, 14, 16, 21, 24, 28, 34, 35

PC Principal Component. vii, 12, 13, 14, 15, 18, 21, 25, 26, 35

PCA Principal Components Analysis. vii, 12, 13, 14, 21, 25, 35

PCR Principal Components Regression. iii, v, vi, vii, 11, 14, 15, 21, 25, 26, 27, 28, 35

PLS Partial Least Squares. vii, 15, 16, 18, 21, 28, 29

PLSR Partial Least Squares Regression. iii, v, vi, vii, 11, 15, 16, 20, 21, 28, 29, 30, 35

ppb parts per billion. 20

ppm parts per million. 3, 6, 33

RMSE Root Mean Squared Error. 18, 21, 25, 28, 35

RSS Residual Sum of Squares. 11, 17

SAS Sensor and Actuator Systems. 3

SCR Selective Catalytic Reduction. 1, 2

SiC-FET Silicon Carbide Field Effect Transistor. iii, 2, 3, 20

TCO Temperature Cycled Operation. 2, 6, 20, 35



1 Introduction

1.1 Motivation

Nitric Oxide (NO) and Nitrogen Dioxide (NO₂), commonly referred together as NO_x, are hazardous gases to the environment and to humans. Its main sources are combustion processes in transportation, and industrial processes such as (but not limited to) auto mobiles, trucks, boats, industrial boilers, turbines, etc. (USEPA 2019).

NO_x exposure to humans can cause respiratory illnesses such bronchitis, emphysema and can worsen heart disease (Boningari and Smirniotis 2016). Environmentally, NO_x are deemed precursors of adverse phenomena such as smog, acid rain, and the depletion of ozone (O₃) (Bernabeo et al. 2019). It is of high interest, therefore, to reduce NO_x emissions.

One well studied and successful method of reducing emissions is Selective Catalytic Reduction (SCR), which consists in the reduction of NO_x by ammonia (NH₃) into nitrogen gas (N₂) and water (H₂O) (Forzatti 2001), both harmless components. The process is based in the following reactions (Forzatti 2001):

- 4 NH₃ + 4 NO + O₂ → 4 N₂ + 6 H₂O
- 2 NH₃ + NO + NO₂ → 2 N₂ + 3 H₂O
- 8 NH₃ + 6 NO₂ → 7 N₂ + 12 H₂O

One key element in these reactions, however, is the amount of ammonia dosed into the SCR systems. Ammonia itself is hazardous to humans, causing skin and respiratory irritation, among other illnesses (ASTDR 2004). More importantly, ammonia is one of the main sources of nitrogen pollution and it has direct negative impact on biodiversity via nitrogen deposition in soil and water (Guthrie et al. 2018). Hence it is also desired to keep ammonia emissions

to a minimum. Too much ammonia in the SCR catalyst will guarantee NO_x reduction at the expense of undesired ammonia emissions. Concurrently, too little ammonia will impede SCR to occur properly, beating the purpose of the catalyst and as a consequence, undesired NO_x emissions.

To monitor gasses concentrations, chemical sensors are deployed, one of which is the Silicon Carbide Field Effect Transistor (SiC-FET). The identification and quantification of gasses is normally achieved through multiple sensor in so called sensor arrays. Ideally each sensor in the array needs to have different responses to different compounds (Bastuck 2019). The deployment of multiple sensors, on the other hand, proves itself cumbersome due to the increased chances of failure, and decalibration of the system should one or multiple sensors be replaced (Bastuck 2019).

One solution to this problem is the cycled operation of one single sensor, referred as virtual multi-sensor (Bastuck 2019). By cycling the working point parameters of the sensor, different substances react differently in the sensor surface, which in turn produces different responses. Temperature Cycled Operation (TCO), Gate Bias Cycled Operation (GBCO), and the combination of the two have been proven to increase selectivity of SiC-FET sensors (Bastuck 2019).

TCO, in contrast with a constant temperature evaluation, produces unique transient sensor responses, i.e. each gas mixture yields a slightly different sensor output. This unique gas signature increases selectivity (Bur, Bastuck, Lloyd Spetz, et al. 2014). Additionally, the high temperatures reached in these cycles help in the cleansing of the sensor surface, preparing it for the new mixtures to come.

Frequency modulation tries to achieve the same goal: avoid steady state responses in exchange of unique signatures that could help identify/quantify the gasses at hand. It consists on operating the sensor in Alternating Current (AC). One then can regulate the frequency of this operation and create cycles of different frequencies, similar to what is done in TCO. This is equivalent to GBCO, but with more frequency changes and achieving overall higher frequencies.

1.2 Aim

The aim of this thesis is to investigate if frequency modulation can be used to simultaneously quantify the concentrations of NO_x and Ammonia in a particular gas mixture.

1.3 Research questions

1. Can frequency modulation be used to simultaneously quantify NO_x and Ammonia concentrations?
2. Does quality of fit vary over different prediction models?

2 Data

2.1 Data acquisition

The data was acquired at the Sensor and Actuator Systems (SAS) laboratory at Linköping University. The experiment — as shown on Figure 2.1 — consisted of exposing different gas combinations to two SiC-FET sensors under a particular frequency cycle and recording its response, measured in miliamperes (mA). This is then used to extract secondary features, namely average and slope values from certain regions of the frequency cycle.

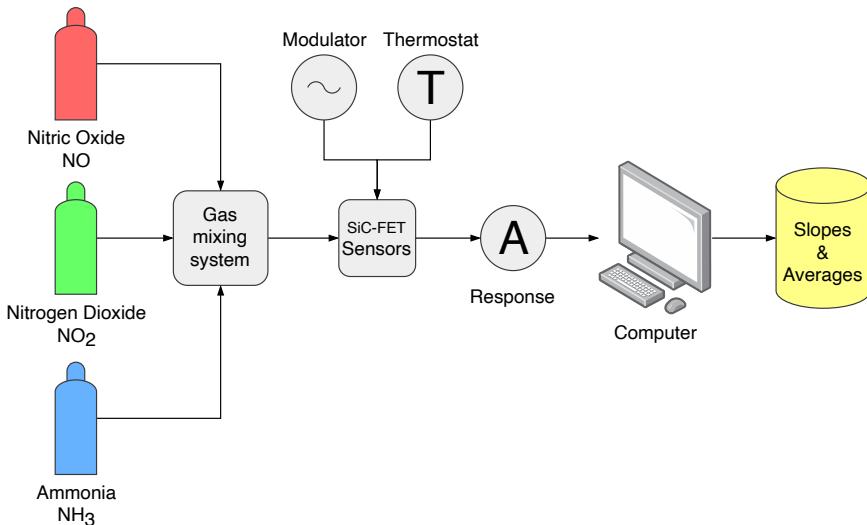


Figure 2.1: Schema of the data acquisition process.

In more detail, NO, NO₂ and NH₃ had five possible concentration values each: 5, 10, 20, 40, and 80 parts per million (ppm). The experiment was designed to encompass all possible combinations of these gases, amounting to 125 different gas mixtures. Each feature was sub-

mitted to the same frequency cycle four times. The cycle consists of 16 unique frequencies: 0.05, 0.1, 0.25, 0.5, 1, 2, 5, 10, 25, 50, 100, 200, 500, 1000, 2500 and 5000 Hertz (Hz). A typical raw sensor response for frequency modulation experiments is shown on Figure 2.2.

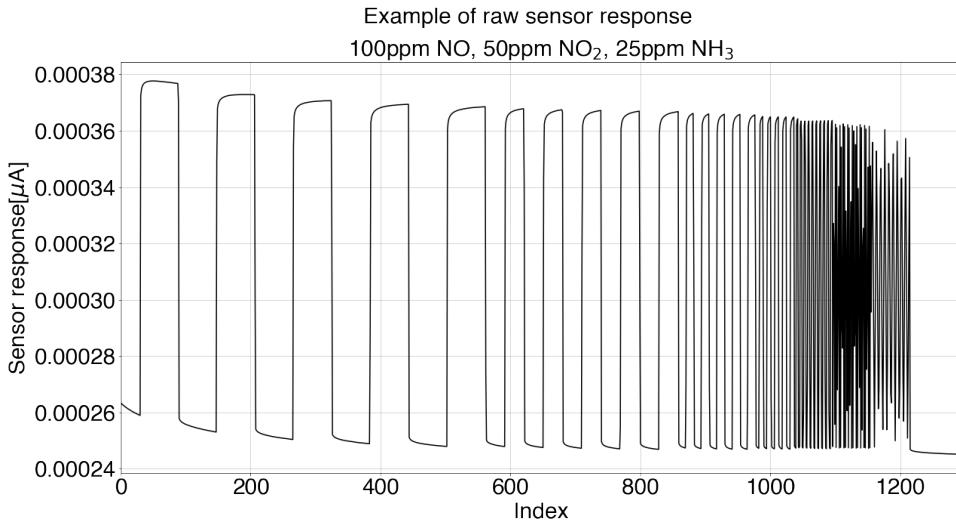


Figure 2.2: An example of raw sensor response

Throughout one cycle, several slope and average features were extracted. The sample rate for feature extraction was set at 4 Hz, i.e. in a cycle of 60 seconds, a total of $60\text{s} \times 4\frac{1}{\text{s}} = 240$ pairs of slopes and averages are recorded, which totals to 480 features per cycle. In other words, during one experiment – 4 cycles of 60 seconds – a total of $480 \times 4 = 1920$ features are extracted.

One way to visualize this process is shown in Figures 2.3. Note that the y-axis is in log-scale due to the different orders of magnitude of frequencies. Moreover, Figure 2.4 gives more insight into feature measurement and Table 2.1 summarizes the data acquisition details.

Table 2.1: Data acquisition details

Parameter	Value
Factors (gases)	3
Levels (concentrations)	5
Frequencies	16
Features per cycle	480
Number of cycles	4
Data points per mixture	1920
Number of mixtures	125
Features per experiment	240.000
Number of experiments	3
Total features	720.000

For specific timestamps and measurement durations, the reader is referred to Appendix A.

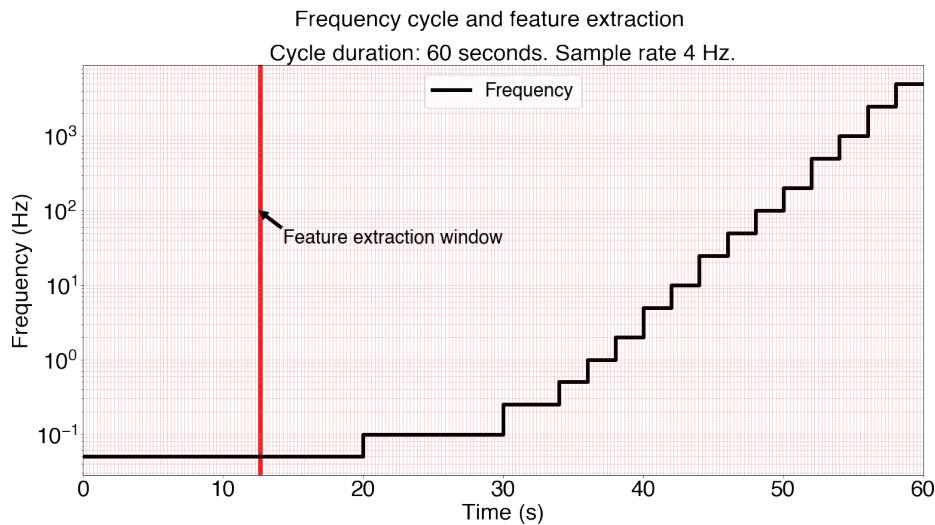


Figure 2.3: Feature measurements times per cycle. The width of the red line indicates the duration of one of the feature measurement windows as an example.

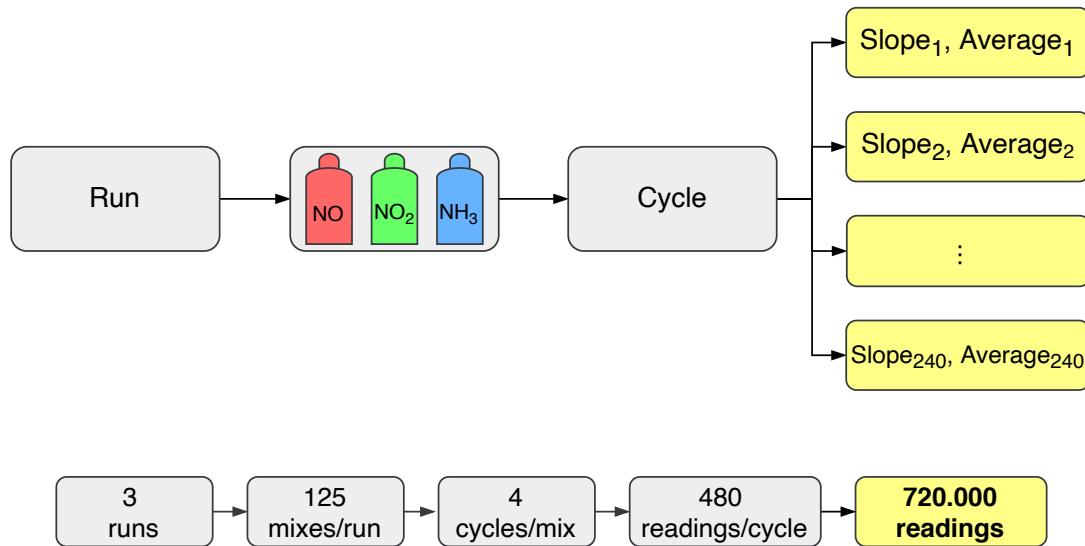


Figure 2.4: A visualization of the feature measurement process.

2.2 Raw data

The experiments were run between 26th and 29th March, 2021. The experiment data was exported as an excel file containing twelve columns, as specified in Table 2.2

Table 2.2: Raw data column details

Name	Description	Unit
Exposure nr	A particular mix of NO, NO ₂ and NH ₃ . Ranges from 1 to 375	-
Cycle nr	The cycle number. Ranges from 1 to 4.	-
Sample nr	Extracted feature index. Ranges from 1 to 240	-
NO	Nitric Oxide concentration	ppm
NO ₂	Nitrogen Dioxide concentration	ppm
NH ₃	Ammonia concentration	ppm
Freq	Frequency	Hz
Slope sensor 1	Slope	μA/s
Slope sensor 2	Slope	μA/s
Average sensor 1	Average	μA
Average sensor 2	Average	μA
Sensor temperature	Temperature	degrees Celsius (°C)

Table 2.3: Sample of raw data.

Index	Exposure nr	Cycle nr	Sample nr	NO [ppm]	NO ₂ [ppm]	NH ₃ [ppm]	Freq [Hz]	Slope sensor 1 [μA/s]	Slope sensor 2 [μA/s]	Average sensor 1 [μA/s]	Average sensor 2 [μA/s]	Sensor temperature [C]
0	1	1	1	10	5	20	0.05	-18.855169	-22.588416	32.926184	27.961554	274.994683
1	1	1	2	10	5	20	0.05	-28.289268	-28.185027	25.853867	20.915297	274.980487
2	1	1	3	10	5	20	0.05	-0.390916	-0.482129	25.756138	20.794765	274.985895
3	1	1	4	10	5	20	0.05	-0.234549	-0.156366	25.697501	20.755673	275.020372
4	1	1	5	10	5	20	0.05	-0.143336	-0.247580	25.661667	20.693778	275.014964
⋮												
100000	105	1	161	5	5	40	5.0	-38.366212	-48.495271	30.241896	24.821197	275.021724
100001	105	1	162	5	5	40	5.0	6.619507	8.521964	31.896773	26.951688	274.999415
100002	105	1	163	5	5	40	5.0	-1.941549	6.580416	31.411386	28.596792	275.011584
100003	105	1	164	5	5	40	5.0	27.401023	22.012900	38.261641	34.100017	275.009894
100004	105	1	165	5	5	40	5.0	-27.016623	-28.439121	31.507486	26.990236	275.014400
⋮												
359995	375	4	236	20	80	5	5000.0	-0.136821	-0.158538	34.129879	30.345597	275.002007
359996	375	4	237	20	80	5	5000.0	0.010859	0.010859	34.132593	30.348312	274.986797
359997	375	4	238	20	80	5	5000.0	-0.043435	0.030405	34.121734	30.355913	274.979811
359998	375	4	239	20	80	5	5000.0	-0.117275	-0.026061	34.092416	30.349398	274.984543
359999	375	4	240	20	80	5	5000.0	0.073840	0.039092	34.110876	30.359171	274.998063

2.3 Pre-processing

The features (slopes and averages) from the same target (a particular exposure) in the raw data file in Table 2.3 are spread across multiple rows, which is not suitable for analysis. Opposed to TCO, the experiments were conducted at constant temperature and therefore the temperature column is discarded. The data was, subsequently, modified to have the desired format: each row containing the predictors for one particular combination of gases. Additionally, the data from each sensor was split into two datasets.

The naming convention for the features is shown in Figure 2.5. First, the frequency in which the measurement was taken followed by the sensor number. After that, the feature name itself is followed by its index, i.e. where in the frequency cycle the measurement was made. This convention allows for easy identification of key information of the cycle and measurement.

The pre-processing results in the format shown in Figure 2.6. Recalling that there are 125 possible mixtures of gases, it is important to note that there are repeated exposures in the data set and those are treated as individual observations. Since each unique gas mixture was exposed 4 times during a cycle, and the experiment was repeated 3 times, this yields a total of $4 \times 3 \times 125 = 1500$ exposures. A snippet of the final data set is shown in Table 2.4.

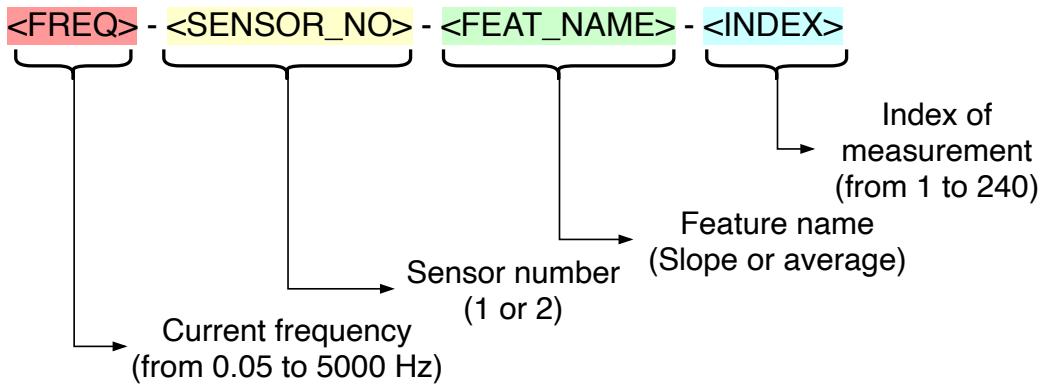


Figure 2.5: Feature naming convention.

Sensor 1

Gases			Features	
NO	NO ₂	NH ₃	Slopes	Averages

Sensor 2

Gases			Features	
NO	NO ₂	NH ₃	Slopes	Averages

1500 exposures

480 features

Figure 2.6: Pre-processed data structure.

In efforts to further analyze the data, the previous 1500 observations are averaged by unique mixtures, i.e. for each mixture, the features are averaged from its twelve exposures, yielding 125 observations. Figure 2.7 clarifies this further. Table 2.5 contains the averaged slopes and averages per unique mixture. Analysis will be run in this data set separately as means of comparison. The lower number of data points here gives opportunity to visualize the data in a plot; this is done in Figure 2.8.

2.3. Pre-processing

Table 2.4: Sample of pre-processed data.

Index	EXPOSURE	NO	NO2	NH3	0.05-1-slope-0	0.05-1-slope-1	...	5000.0-1-slope-239	0.05-1-avg-0	0.05-1-avg-1	0.05-1-avg-2	...	5000.0-1-avg-238	5000.0-1-avg-239
0	1.0	10.0	5.0	20.0	-18.855169	-28.289268	...	0.019546	32.926184	25.853867	25.756138	...	35.840135	35.845021
1	1.0	10.0	5.0	20.0	-28.979886	-9.251672	...	-0.056466	28.600050	26.287132	26.225237	...	35.884113	35.869996
2	1.0	10.0	5.0	20.0	-25.431240	-12.874158	...	-0.052122	29.512187	26.293647	26.238267	...	35.913432	35.900401
3	1.0	10.0	5.0	20.0	-30.126572	-8.196200	...	-0.156366	28.368758	26.319708	26.254555	...	35.939493	35.900401
4	2.0	20.0	40.0	40.0	-19.506695	-27.051368	...	-0.078183	33.180279	26.417437	26.303420	...	35.685397	35.665852
:														
700	176.0	40.0	20.0	40.0	-21.011721	-25.822155	...	-0.071668	31.458621	25.003082	24.902639	...	34.554999	34.537082
701	176.0	40.0	20.0	40.0	-27.505265	-10.847911	...	0.086870	27.660766	24.948788	24.918927	...	34.504506	34.526224
702	176.0	40.0	20.0	40.0	-27.516124	-10.750182	...	-0.097729	27.647193	24.959647	24.928700	...	34.531653	34.507221
703	176.0	40.0	20.0	40.0	-27.364102	-10.875058	...	0.086870	27.666195	24.947431	24.935215	...	34.537082	34.558800
704	177.0	80.0	40.0	40.0	-20.794546	-26.195696	...	0.041263	31.640505	25.091581	25.088324	...	34.695078	34.705393
:														
1495	374.0	80.0	80.0	40.0	-27.937445	-10.891346	...	-0.097729	27.166692	24.443855	24.392276	...	34.151596	34.127164
1496	375.0	20.0	80.0	5.0	-24.358394	-22.933723	...	-0.008687	30.315735	24.582305	24.530726	...	34.134765	34.132593
1497	375.0	20.0	80.0	5.0	-28.862612	-9.827186	...	-0.112931	26.916940	24.460144	24.410736	...	34.159740	34.131507
1498	375.0	20.0	80.0	5.0	-25.839531	-12.780772	...	-0.021718	27.671625	24.476432	24.430282	...	34.143452	34.138023
1499	375.0	20.0	80.0	5.0	-28.002598	-10.645937	...	0.073840	27.137373	24.475889	24.424853	...	34.092416	34.110876

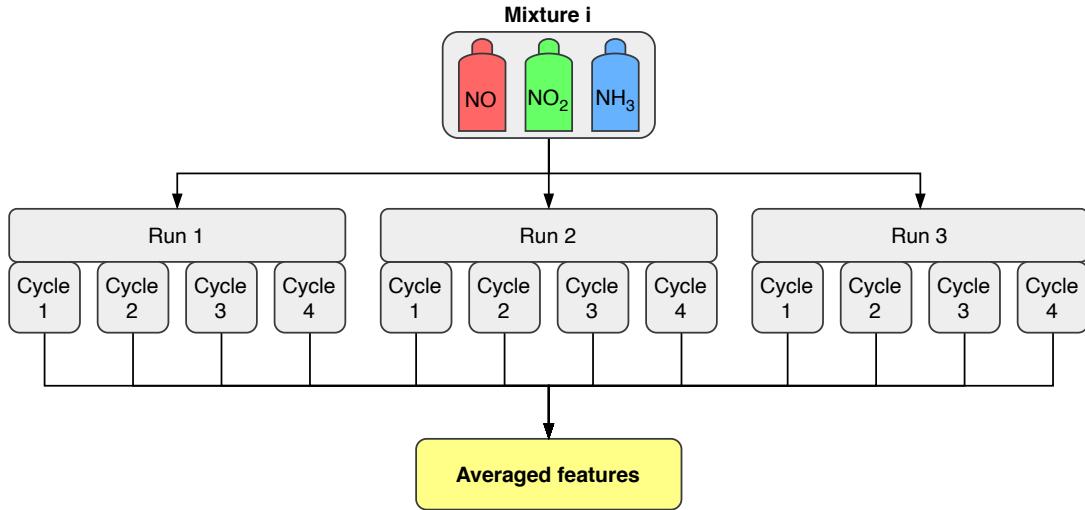


Figure 2.7: A visualization of the feature averaging process.

Table 2.5: Sample of averaged features throughout all exposures data.

Index	UNIQUE MIXTURE	NO	NO2	NH3	0.05-1-slope-0	0.05-1-slope-1	...	5000.0-1-slope-239	0.05-1-avg-0	0.05-1-avg-1	0.05-1-avg-2	...	5000.0-1-avg-238	5000.0-1-avg-239
0	0	5.0	5.0	5.0	-26.386178	-14.165356	...	-0.041897	28.983749	25.442410	25.383750	...	35.162932	35.152458
1	1	5.0	5.0	10.0	-26.294332	-14.421532	...	0.004524	28.538652	24.933269	24.879247	...	34.622460	34.623591
2	2	5.0	5.0	20.0	-25.514491	-15.174588	...	-0.010135	29.038925	25.245278	25.181935	...	35.025637	35.023103
3	3	5.0	5.0	40.0	-25.906221	-14.565139	...	0.003801	28.698684	25.057399	24.980686	...	34.575699	34.576649
4	4	5.0	5.0	80.0	-26.731849	-13.795072	...	-0.022803	28.738748	25.289980	25.229714	...	34.860040	34.854340
:														
70	70	20.0	80.0	5.0	-26.743885	-13.981572	...	0.019546	28.142217	24.646824	24.596195	...	34.208650	34.213536
71	71	20.0	80.0	10.0	-25.832291	-14.650290	...	-0.056285	28.615026	24.952453	24.893228	...	34.511972	34.497900
72	72	20.0	80.0	20.0	-25.713207	-14.907192	...	-0.016469	28.432463	24.705665	24.649538	...	34.317554	34.313437
73	73	20.0	80.0	40.0	-26.080595	-14.410131	...	-0.066238	28.327675	24.725143	24.685825	...	34.213989	34.197429
74	74	20.0	80.0	80.0	-25.091813	-14.220736	...	0.004163	28.611836	25.056652	24.993128	...	34.592507	34.593548
:														
120	120	80.0	80.0	5.0	-27.073901	-13.562242	...	0.028142	28.548244	25.157684	25.103051	...	34.742313	34.749349
121	121	80.0	80.0	10.0	-26.329623	-14.337196	...	-0.012669	28.630183	25.045884	25.015773	...	34.678857	34.675690
122	122	80.0	80.0	20.0	-25.935631	-14.734990	...	-0.027690	28.420835	24.737087	24.687996	...	34.354338	34.347416
123	123	80.0	80.0	40.0	-25.821070	-14.855702	...	-0.032395	28.457189	24.743263	24.682929	...	34.327938	34.319839
124	124	80.0	80.0	80.0	-26.503363	-14.087625	...	-0.038820	28.615161	25.093255	25.046698	...	34.743829	34.734124

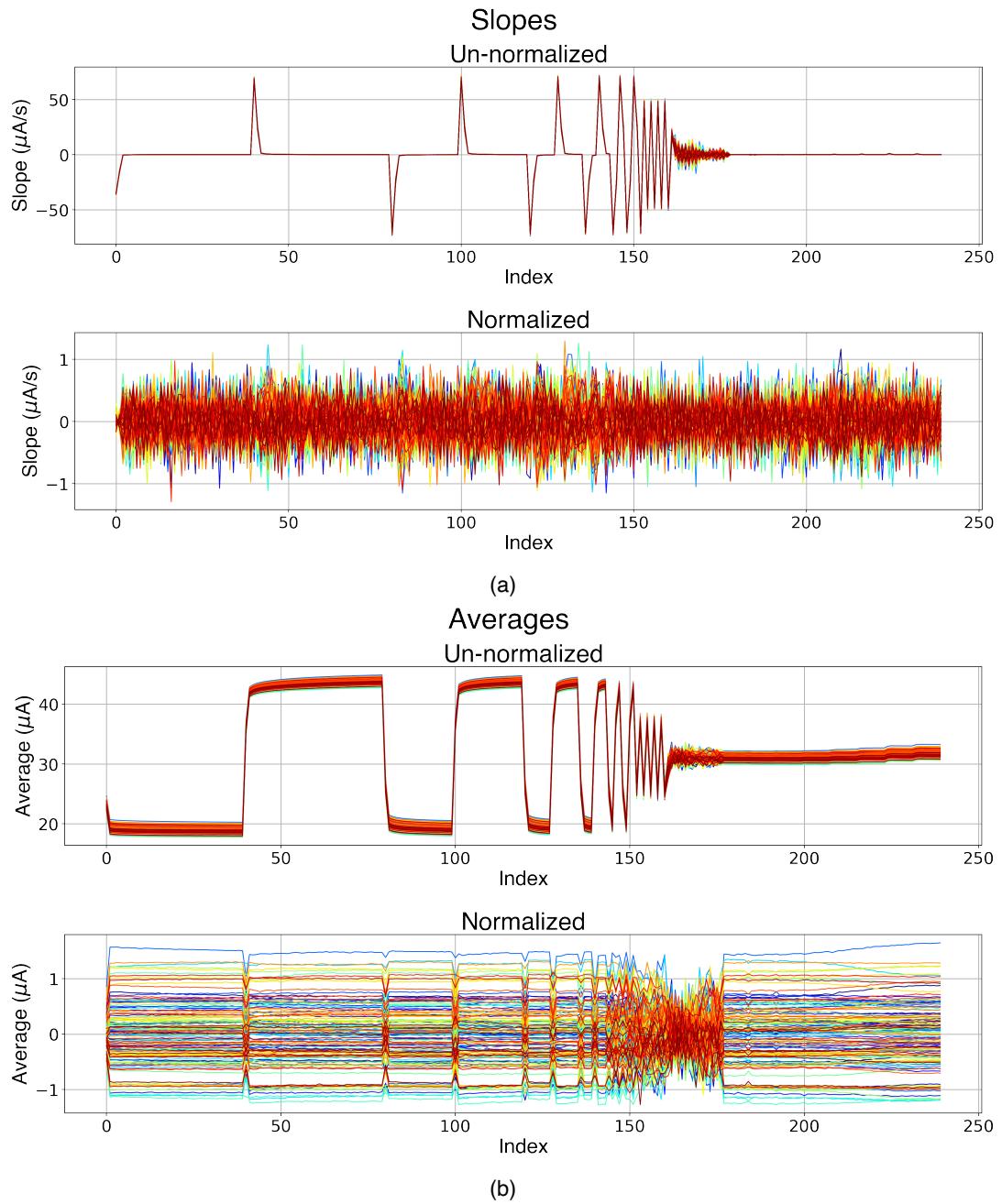


Figure 2.8: (a) Slope and (b) average features, both un-normalized and normalized.
In each plot, each line represents a unique gas mixture.



3 Theory

The quantification of gases based on the sensor response can be viewed as a multivariate multiple regression problem where the predictors, i.e. features derived from the sensor signal, are used to predict multiple responses, i.e. the concentrations of pertinent gases. This chapter exposes the theory behind some of these models.

The models here listed were chosen as a natural progression from a statisticians point of view: starting with simple models and progressively increasing complexity as insights from the data and the problem are gathered.

3.1 Notation

In favor of consistency and clarity, the notation used throughout this work is presented here. Bold capital letters, e.g. \mathbf{A} , are matrices while bold lower case letters are column vectors, e.g. \mathbf{a}_1 . Scalars, on the other hand, are denoted as standard lower case letters, e.g. a_{11} . Transposes and inverses are denoted, respectively with \cdot^T and \cdot^{-1} . This is valid unless explicitly noted. An example is shown below.

The data matrix \mathbf{X} :

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_p \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

The response matrix \mathbf{Y} :

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_m \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nm} \end{bmatrix}$$

3.2 Ordinary Least Squares Regression

A simple, first approach would be to tackle the problem with a Ordinary Least Squares (OLS) regression model. As Hastie et al. 2001 explains, each output in \mathbf{Y} has its own linear model. Now, given \mathbf{X} , a set of n observations and p features, the concatenation of all linear models can be written in matrix form as in Equation 3.1.

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad (3.1)$$

Where:

- \mathbf{B} : $[p+1 \times m]$ matrix of regression coefficients (with the $+1$ referring to the intercept term);
- \mathbf{E} : $[n \times m]$ matrix of random noise.

The Residual Sum of Squares (RSS), as the name suggests, is defined as the difference between real and predicted values, squared, which in matrix form is written as (Hastie et al. 2001):

$$\text{RSS}(\mathbf{B}) = \text{Tr}[(\mathbf{Y} - \mathbf{XB})^\top (\mathbf{Y} - \mathbf{XB})] \quad (3.2)$$

In turn, the objective is then to find the coefficients $\hat{\mathbf{B}}$ which minimizes the RSS, which is summarized by Equation 3.3 (Hastie et al. 2001):

$$\hat{\mathbf{B}}^{\text{OLS}} = \arg \min_{\mathbf{B}} \text{RSS}(\mathbf{B}) \quad (3.3)$$

Finally, solving for $\hat{\mathbf{B}}$ yields (Hastie et al. 2001):

$$\hat{\mathbf{B}}^{\text{OLS}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \quad (3.4)$$

For the problem at hand, in addition to the high number of features, it is often the case that sensor data points are acquired in quick succession, which in turn leads to highly correlated features (Bastuck 2019), which can result in high variance in a least squares model (Hastie et al. 2001). It is natural, therefore, to progress towards methods that incorporate dimensionality reduction such as Principal Components Regression (PCR) and Partial Least Squares Regression (PLSR) or shrinkage such as Ridge Regression.

3.3 Principal Component Analysis

One way to define Principal Components Analysis (PCA) is to view it as a orthogonal projection of the data into a principal space of lower dimension such that the variance of this projection is maximized (Bishop 2006).

Just as before, consider the collection of n observations \mathbf{X} with covariance matrix Σ . Additionally, consider a matrix $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_p]$ where $\mathbf{p}_j, j = 1, 2, \dots, p$ is a vector of weights of the linear combination (Johnson and Wichern 2013):

$$\mathbf{t}_i = \mathbf{p}_i^\top \mathbf{X} \quad i = 1, 2, \dots, p \quad (3.5)$$

The variance and covariance of these new variables t_i can be written as follows:

$$\text{Var}(t_i) = \mathbf{p}_i^\top \Sigma \mathbf{p}_i \quad i = 1, 2, \dots, p \quad (3.6)$$

$$\text{Cov}(t_i, t_k) = \mathbf{p}_i^\top \Sigma \mathbf{p}_k \quad i, k = 1, 2, \dots, p \quad (3.7)$$

The first Principal Component (PC) is then the linear combination with maximum variance, i.e. the linear combinations that maximizes $\text{Var}(t_1)$, with the constraint that the coefficient vector \mathbf{p}_1 has unit length. In summary, the first PC is computed as (Johnson and Wichern 2013):

$$\begin{aligned} PC_1 &= t_1 = \mathbf{p}_1^\top \mathbf{X} \\ &\text{that maximizes } \text{Var}(t_1 = \mathbf{p}_1^\top \mathbf{X}) \\ &\text{subject to } \mathbf{p}_1^\top \mathbf{p}_1 = 1 \end{aligned} \quad (3.8)$$

The second PC, similarly to the first, is the linear combination with maximum variance, but with an added extra constraint: this new linear combination must be orthogonal to the previous one, i.e. they must be linearly independent:

$$\begin{aligned} t_2 &= \mathbf{p}_2^\top \mathbf{X} \\ &\text{that maximizes } \text{Var}(t_2 = \mathbf{p}_2^\top \mathbf{X}) \\ &\text{subject to } \mathbf{p}_2^\top \mathbf{p}_2 = 1 \\ &\text{and } \text{Cov}(t_1, t_2) = 0 \end{aligned} \quad (3.9)$$

The k-th PC is then:

$$\begin{aligned}
 t_k &= p_k^\top X \\
 \text{that maximizes } \text{Var}(t_k) &= p_k^\top X \\
 \text{subject to } p_k^\top p_k &= 1 \\
 \text{and } \text{Cov}(t_i, t_k) &= 0 \quad \text{for } k < i
 \end{aligned} \tag{3.10}$$

In summary, the objective of PCA is find a matrix P such that the linear transformation

$$T = X P^\top \tag{3.11}$$

yields new variables that are uncorrelated and arranged in decreasing order of variance.

It can be shown that these desired linear combinations can be written in terms of the eigenvalues (ϕ) and eigenvectors (e) of Σ , the covariance matrix of X (Johnson and Wichern 2013). The elements of eigenvectors are called loadings, while the new features T are . In short, for the k-th PC:

$$\begin{aligned}
 t_k &= e_k^\top X_k \\
 \text{Var}(t_k) &= e_k^\top \Sigma e_k = \phi_k \\
 \text{Cov}(t_k, t_j) &= e_k^\top \Sigma e_j = 0 \quad \text{for } k \neq j
 \end{aligned} \tag{3.12}$$

There are several ways of computing PCs. Many of which involving finding aforementioned eigenvalues and eigenvectors. These calculations can be computationally expensive, depending on the desired number of extracted PCs (Bishop 2006). One option is the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm, also called Power Method. It has two clear advantages: "it can handle missing data and computes the components sequentially" (Dunn 2021).

The NIPALS algorithm to compute the first k-th PCs, PC_i , $i = 1, 2, \dots, k$ us displayed below as Algorith 1 (Dunn 2021) (Ng 2013) (Wright 2017). Since it computes the loadings and scores sequentially, it is possible to stop it as early as desired. The "truncated" loadings and scores that project X into the principal subspace of k PCs is defined in Equation 3.13 :

$$T_{|k} = X P_{|k}^\top \tag{3.13}$$

Algorithm 1: Nonlinear Iterative Partial Least Squares (NIPALS) for PCA

Result: Matrices of loadings $\mathbf{P}_{|k}$ and scores $\mathbf{T}_{|k}$ of the k-th first Principal Components

```

1 Initialize  $\mathbf{T}_{|k}$  and  $\mathbf{P}_{|k}$ 
2  $i = 1$ 
3  $\mathbf{X}_1 := \mathbf{X}$ 
4 while  $i < k$  do
5   repeat
6     Choose  $\mathbf{t}_i$  as any column of  $\mathbf{X}_i$ 
7     Compute loadings  $\mathbf{p}_i = (\mathbf{t}_i^\top \mathbf{t}_i)^{-1} \mathbf{t}_i^\top \mathbf{X}_i$ 
8     Scale  $\mathbf{p}_i = \frac{\mathbf{p}_i}{\sqrt{\mathbf{p}_i^\top \mathbf{p}_i}}$ 
9     Compute scores  $\mathbf{t}_i = (\mathbf{p}_i^\top \mathbf{p}_i)^{-1} \mathbf{p}_i^\top \mathbf{X}_i$ 
10    until  $\mathbf{t}_i$  converges
11    Append  $\mathbf{t}_i$  to  $\mathbf{T}_{|k}$ 
12    Append  $\mathbf{p}_i$  to  $\mathbf{P}_{|k}$ 
13    Deflate:  $\mathbf{X}_{i+1} = \mathbf{X}_i - \mathbf{t}_i \mathbf{p}_i^\top$ 
14     $i += 1$ 
15 end
16 return  $\mathbf{T}_{|k}, \mathbf{P}_{|k}$ 
```

In words, the main idea of the algorithm can be summarized as choosing an arbitrary column of \mathbf{X} as the scores vector \mathbf{t}_i , shown in line 6. Next, the computation of the i-th loadings vector \mathbf{p}_i by regressing every column of \mathbf{X} via OLS onto the scores \mathbf{t}_i . \mathbf{p}_i is then scaled to have unit length in Line 8, which in turn is used to compute the i-th scores vector \mathbf{t}_i by regressing every column of \mathbf{X} via OLS onto the loadings \mathbf{p}_i , shown in Line 9. This procedure is repeated until change in \mathbf{t}_i between iterations is small enough. Once convergence is achieved, scores \mathbf{t}_i and loadings \mathbf{p}_i are stored as the i-th column of matrices \mathbf{T} and \mathbf{P} of Equation 3.11, respectively. Finally, the variability explained by \mathbf{t}_i and \mathbf{p}_i from \mathbf{X} is subtracted in a procedure called deflation.

3.4 Principal Component Regression

With the inner workings of PCA explained in the previous section, PCR can be simply reduced to a Least Squares regression on the first k-th PCs, i.e. performing linear regression on $\mathbf{T}_{|k}$ instead of \mathbf{X} :

$$\mathbf{Y} = \mathbf{T}_{|k} \mathbf{B} + \mathbf{E} \quad (3.14)$$

And the regression coefficients are found analogously to Equation 3.4:

$$\hat{\mathbf{B}}^{\text{PCR}} = (\mathbf{T}_{|k}^\top \mathbf{T}_{|k})^{-1} \mathbf{T}_{|k}^\top \mathbf{Y} \quad (3.15)$$

Although useful, PCR has a potential flaw: while the new found projection of \mathbf{X} is guaranteed to best explain the variance of predictors, this cannot be said about the responses \mathbf{Y} (Gareth et al. 2013). PLSR, on the other hand, solves this issue by supervising the identification of PCs (Gareth et al. 2013).

3.5 Partial Least Squares Regression

PLSR, much like PCR, also aims to reduce dimensionality via linear combinations of the inputs. This technique, however, also takes into account the response variables \mathbf{Y} . One key advantage of PLSR is that it seeks axes with most variance (like PCR) and high correlation with response variables (Hastie et al. 2001).

The main idea can be described as finding linear combinations for the data matrix \mathbf{X} and response matrix \mathbf{Y} as follows (Ng 2013), similarly to what was done in Section 3.3. Here, the matrices \mathbf{W} and \mathbf{U} are score matrices, i.e. the transformed PLS variables, and \mathbf{L} and \mathbf{Q} are loading matrices, i.e. the weights of this transformation (projection), similar to Equation 3.11.

$$\mathbf{W} = \mathbf{XL}^\top \quad (3.16)$$

Where:

- \mathbf{W} : $[n \times k]$ matrix of PLS scores referring to the data \mathbf{X} ;
- \mathbf{L} : $[k \times p]$ matrix PLS coefficients referring to the data \mathbf{X} .

$$\mathbf{U} = \mathbf{YQ}^\top \quad (3.17)$$

Where:

- \mathbf{U} : $[n \times k]$ matrix of PLS scores referring to the response \mathbf{Y} ;
- \mathbf{Q} : $[k \times m]$ matrix PLS coefficients referring to the response \mathbf{Y} .

Instead of simply running NIPALS on \mathbf{X} and \mathbf{Y} separately, PLSR uses information from \mathbf{Y} to decompose \mathbf{X} and vice-versa (Ng 2013). Algorithm 2 is an adaptation of Algorithm 1 to incorporate this intended behavior.

Algorithm 2: NIPALS for Partial Least Squares Regression (PLSR)

Result: Matrices of loadings $\mathbf{L}_{|k}$, $\mathbf{Q}_{|k}$ and scores $\mathbf{W}_{|k}$, $\mathbf{U}_{|k}$ of the k -th first Partial Least Squares directions

```

1 Initialize  $\mathbf{L}_{|k}$ ,  $\mathbf{Q}_{|k}$  and  $\mathbf{W}_{|k}$ ,  $\mathbf{U}_{|k}$ 
2  $i = 1$ 
3  $\mathbf{X}_1 := \mathbf{X}$ 
4  $\mathbf{Y}_1 := \mathbf{Y}$ 
5 while  $i < k$  do
6   repeat
7     Choose  $\mathbf{u}_i$  as any column of  $\mathbf{Y}_i$ 
8     Compute loadings of  $\mathbf{X}_i$  based on score of  $\mathbf{Y}_i$ :  $\ell_i = (\mathbf{u}_i^\top \mathbf{u}_i)^{-1} \mathbf{u}_i^\top \mathbf{X}_i$ 
9     Scale  $\ell_i = \frac{\ell_i}{\sqrt{\ell_i^\top \ell_i}}$ 
10    Compute score of  $\mathbf{X}_i$ :  $\mathbf{w}_i = (\ell_i^\top \ell_i)^{-1} \ell_i^\top \mathbf{X}_i$ 
11    Compute loadings of  $\mathbf{Y}_i$  based on score of  $\mathbf{X}_i$ :  $\mathbf{q}_i = (\mathbf{w}_i^\top \mathbf{w}_i)^{-1} \mathbf{w}_i^\top \mathbf{Y}_i$ 
12    Scale  $\mathbf{q}_i = \frac{\mathbf{q}_i}{\sqrt{\mathbf{q}_i^\top \mathbf{q}_i}}$ 
13    Compute score of  $\mathbf{Y}_i$ :  $\mathbf{u}_i = (\mathbf{q}_i^\top \mathbf{q}_i)^{-1} \mathbf{q}_i^\top \mathbf{Y}_i$ 
14  until  $\mathbf{u}_i$  converges
15  Append  $\mathbf{w}_i$  to  $\mathbf{W}_{|k}$ 
16  Append  $\ell_i$  to  $\mathbf{L}_{|k}$ 
17  Append  $\mathbf{u}_i$  to  $\mathbf{U}_{|k}$ 
18  Append  $\mathbf{q}_i$  to  $\mathbf{Q}_{|k}$ 
19  Deflate  $\mathbf{X}_i$ :  $\mathbf{X}_{i+1} = \mathbf{X}_i - \mathbf{w}_i \ell_i^\top$ 
20  Deflate  $\mathbf{Y}_i$ :  $\mathbf{Y}_{i+1} = \mathbf{Y}_i - \mathbf{u}_i \mathbf{q}_i^\top$ 
21   $i += 1$ 
22 end
23 return  $\mathbf{W}_{|k}$ ,  $\mathbf{L}_{|k}$ ,  $\mathbf{U}_{|k}$ ,  $\mathbf{Q}_{|k}$ 

```

As with Algorithm 1, Algorithm 2 can be summarized as choosing a column of \mathbf{Y}_i as the initial response score vector \mathbf{u}_i . After that, the i -th loadings vector \mathbf{w}_i of \mathbf{X} is computed in Line 8 by regressing every column of \mathbf{X} via OLS onto scores vector of \mathbf{Y} , \mathbf{u}_i . Similarly to before, the data loadings vector \mathbf{w}_i is scaled to have unit length, which in turn is used to compute the i -th data scores vector \mathbf{w}_i by regressing every column of \mathbf{X}_i via OLS onto the column ℓ_i in Line 10. Now, the i -th response loadings vector \mathbf{q}_i of \mathbf{Y}_i by regressing every column of \mathbf{Y} via OLS onto scores vector of \mathbf{X} , \mathbf{w}_i , shown in Line 11; This loadings vector is also scaled to have unit length.

Following in Line 13, the i -th response scores vector \mathbf{u}_i is computed by regressing every column of \mathbf{Y}_i via OLS onto the column \mathbf{q}_i . This procedure is repeated until change in \mathbf{u}_i between iterations is small enough. In that case the results \mathbf{w}_i and ℓ_i are stored as the i -th

column of matrices \mathbf{W} and \mathbf{L} of Equation 3.16 and u_i and q_i are stored as the i -th column of matrices \mathbf{U} and \mathbf{Q} of Equation 3.17. Finally, the variability explained by w_i, ℓ_i and u_i, q_i from \mathbf{X}_i and \mathbf{Y}_i , respectively, are removed.

After finding the k partial least squares directions from Algorithm 2 above, the problem, as in Section 3.4, reduces to performing Least Squares Regression using the newfound transformations.

$$\mathbf{Y} = \mathbf{W}_{|k}\mathbf{B} + \mathbf{E} \quad (3.18)$$

Which in turn, analogously to Equations 3.4 and 3.15, yields the coefficients:

$$\hat{\mathbf{B}}^{\text{PLSR}} = (\mathbf{W}_{|k}^\top \mathbf{W}_{|k})^{-1} \mathbf{W}_{|k}^\top \mathbf{Y} \quad (3.19)$$

3.6 Ridge Regression

Ridge regression is also a viable alternative to reduce the problem of highly correlated features (Hastie et al. 2001). Instead of fitting a least squares model on a subset of predictors or a transformation of them, Ridge allows the use of all features with a continuous shrinkage of its coefficients, which results in less variance (Hastie et al. 2001).

For the multi-output case, there are two options: use the same penalization parameter λ for all outputs $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m]^\top$ or apply different parameters $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_m]^\top$. In this work, the latter is preferred over the former, as it allows a more fine tuned control of the regression models for each studied gas.

Analogous to Section 3.2, the goal is to minimize the RSS, but now with the penalization term taken into account. Equation 3.20 below shows this objective function in matrix form.

$$\text{RSS}^{\text{Ridge}}(\mathbf{B}, \boldsymbol{\lambda}) = \text{Tr}[(\mathbf{Y} - \mathbf{XB})^\top (\mathbf{Y} - \mathbf{XB})] + \text{Tr}[\mathbf{B}^\top \mathbf{B} + \boldsymbol{\lambda} \mathbf{I}] \quad (3.20)$$

Where \mathbf{I} is the $[m \times m]$ identity matrix.

$$\hat{\mathbf{B}}^{\text{Ridge}} = \arg \min_{\mathbf{B}} \text{RSS}^{\text{Ridge}}(\mathbf{B}) \quad (3.21)$$

The coefficients that minimize the RSS is shown in Equation 3.22 below.

$$\hat{\mathbf{B}}^{\text{Ridge}} = (\mathbf{X}^\top \mathbf{X} + \boldsymbol{\lambda} \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{Y} \quad (3.22)$$

It is important to note that the parameters β_0 are calculated separately as the response mean \bar{Y} (Hastie et al. 2001), i.e.:

$$\hat{\beta}_0^{\text{Ridge}} = \begin{bmatrix} \hat{\beta}_{0,1} \\ \hat{\beta}_{0,2} \\ \vdots \\ \hat{\beta}_{0,m} \end{bmatrix} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_m \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n y_{i,1} \\ \frac{1}{n} \sum_{i=1}^n y_{i,2} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n y_{i,m} \end{bmatrix} \quad (3.23)$$

The choice of hyper-parameters $\lambda \geq 0$ controls how much shrinkage is applied to the coefficients: larger λ implies more penalization to complex models. Although the coefficients are shrunk towards zero, they never reach zero, which makes Ridge regularization unsuitable for feature selection (Hastie et al. 2001).

Finally, predictions in a ridge regression setting are computed as:

$$\hat{Y}^{\text{Ridge}} = \hat{\beta}_0^{\text{Ridge}} + \hat{B}^{\text{Ridge}} X \quad (3.24)$$

3.7 Cross Validation

There are several choices to make for the aforementioned models: How many PCs/PLS components to use? How much penalization to impose in Ridge regression?

A first answer to this would be to split the data into training and validation sets. After fitting models to the former, the latter is used to measure the prediction error via some scoring function. In that sense, it is important to distinguish test error rate from training error rate. The first, also called generalization error, is the score of the fit on an independent, previously unseen test sample. The second, on the other hand, is the average score over the training sample (Hastie et al. 2001).

Scoring functions measure how much the data deviates from the fit and can be used as a qualitative tool for model selection and comparison. Once this is done, the choice of the model that yields minimum error is trivial. Two examples of wildly used score functions are Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). For the multi-output case of m responses and n observations, they are defined respectively as :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^m (y_{ij} - \hat{y}_{ij}) \right)^2 \quad (3.25)$$

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (3.26)$$

This approach, however, is sensitive to the choice of these sets. Additionally, reserving part of the data just for validation might be detrimental for the model fitting process, specially if the number of observations is low (Gareth et al. 2013).

One tool that can help alleviating these problems is Cross Validation (CV). More specifically, F -fold CV: it involves equally dividing the training data into F sets. For each subset, the desired model is trained using $F - 1$ folds and the prediction error is computed on the remaining fold (Hastie et al. 2001). As for the final evaluation, it is performed in the held-out test set.



4 Methods

This work's main question is: given sensor responses, how one can quantify the gasses that produced them? Multivariate regression techniques have been shown to be successful, namely Partial Least Squares Regression (PLSR) has been used in chemometrics extensively and it has been proven to be good at this task (Bastuck 2019) (Wold et al. 2001).

As examples, Bur, Bastuck, Puglisi, et al. (2015) uses TCO of SiC-FET sensors alongside PLSR to quantify naphthalene sufficiently enough to monitor its concentration for indoor air quality monitoring. Additionally, Bastuck et al. (2016) shows that, also using TCO, PLSR can quantify ethanol and naphthalene mixtures down to the parts per billion (ppb) level.

All code was done in Python, namely in Jupyter notebooks both for it's simplicity and for its easiness in code and data exploration. In general, the use of Python's library Scikit-Learn and its pipeline class alongside linear models made analysis straightforward. Additionally, Scikit-Learn's GridSearchCV allowed for a faster evaluation of different hyper parameters such as number of components and shrinkage factors.

Throughout all methods, training, test and validation sets were split using Python's `train-test_split` function. 80% of observations were assigned to the training set and the remaining 20% to the validation set. The test set is embedded in the training set, as k-fold cross-validation is used. A fixed random seed of 42 was set to ensure reproducibility of results.

Before any model implementation, however, the data needed to be pre-processed as described in Section 2.3. This was done by "transposing" the row-based, per sample measurements into column-based, per exposure features. This reduce the number of row significantly: from 360.000 to 1500. The features were named according to Figure 2.5.

The evaluation and comparison of different models was done via the actual vs. predicted plot. In this plot, it is possible to qualitatively see how good the predictions are. Additionally, fit metrics such as R^2 and RMSE are shown as a quantitative means of comparison.

This methodology was carried out using data from both sensors: 1 and 2. Moreover, two variations of the data from these sensors were used: first using 1500 exposures and all 480 features. The second variation was done by averaging all exposures into unique mixtures as described in Figure 2.7 and only using average features. The reasoning behind this will become apparent in the following chapters.

For further details regarding coding and implementation, the reader is referred to this work's [repository](#) where all notebooks can be found.

4.1 Ordinary Least Squares

Here treated as a baseline, OLS is fit using all 480 features and evaluated using the validation set.

4.2 Principal Components Regression

First, a PCA with two PCs was performed. With that, cumulative variance and score plots were made in an attempt of better visualizing and understanding the data.

Following that, a linear regression on the PCs was made. The number of components was set to be between 1 and 200, and the ideal number of components was chosen via cross-validation with RMSE as the scoring function. Just as before, the ideal model was evaluated in a held-out validation set and a actual vs predicted plot was constructed.

4.3 Partial Least Squares Regression

Here, a similar procedure to Section 4.2 was conducted. Initially, two PLS components were extracted and some informative plots were made: cumulative explained variance and score plots in an attempt to better visualize the data. Here, once more, the grid of the number of PLS components was set between 1 and 200. The regression model was trained with the ideal number of components given by CV and later evaluated in the held-out validation set.

4.4 Ridge Regression

For the shrinkage factor ϕ , a logarithmic grid of 1000 values of ϕ , ranging from 10^{-10} to 10^4 was set. This deliberate choice of unusually high values of ϕ are further explained in the following sections. Regardless of that, CV was used to find the best fit and that was evaluated in the held-out validation set.

5 Results

This chapter is dedicated to show the analysis' results. In favor of clarity and organization, this chapter will be divided into sections, each corresponding to a different model. The plots presented in this section were made using *ad-hoc* plotting functions.

Initially the regression analysis was done in the pre-processed data presented in Table 2.4, i.e. each observation corresponds to a gas exposure. It is important to remind the reader that in this data, each unique gas mixture was exposed (i.e. an exposure) twelve times: four frequency cycles through three experiment repetitions, yielding 1500 observations. Subsequently, the same analysis was conducted, but this time using the only average features of the mixture averages, shown in Table 2.5.

Before beginning analysis, an assessment of correlation between features is first conducted and shown in Figure 5.1. From the correlation matrix, it is possible to see that slope features are not correlated at all with one another, while average features, on the other hand, are mostly perfectly positively correlated.

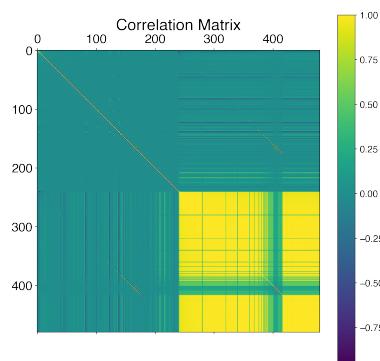


Figure 5.1: Correlation matrix of features.

In Figure 5.1, the first, mainly green, quadrant correspond to slope features, while the fourth, mainly yellow, quadrant, averages.

5.1 Ordinary Least Squares

As explained in Chapter 4, OLS is treated here as a baseline. The actual vs. predicted plot in Figure 5.2a shows the predictions for both data sets.

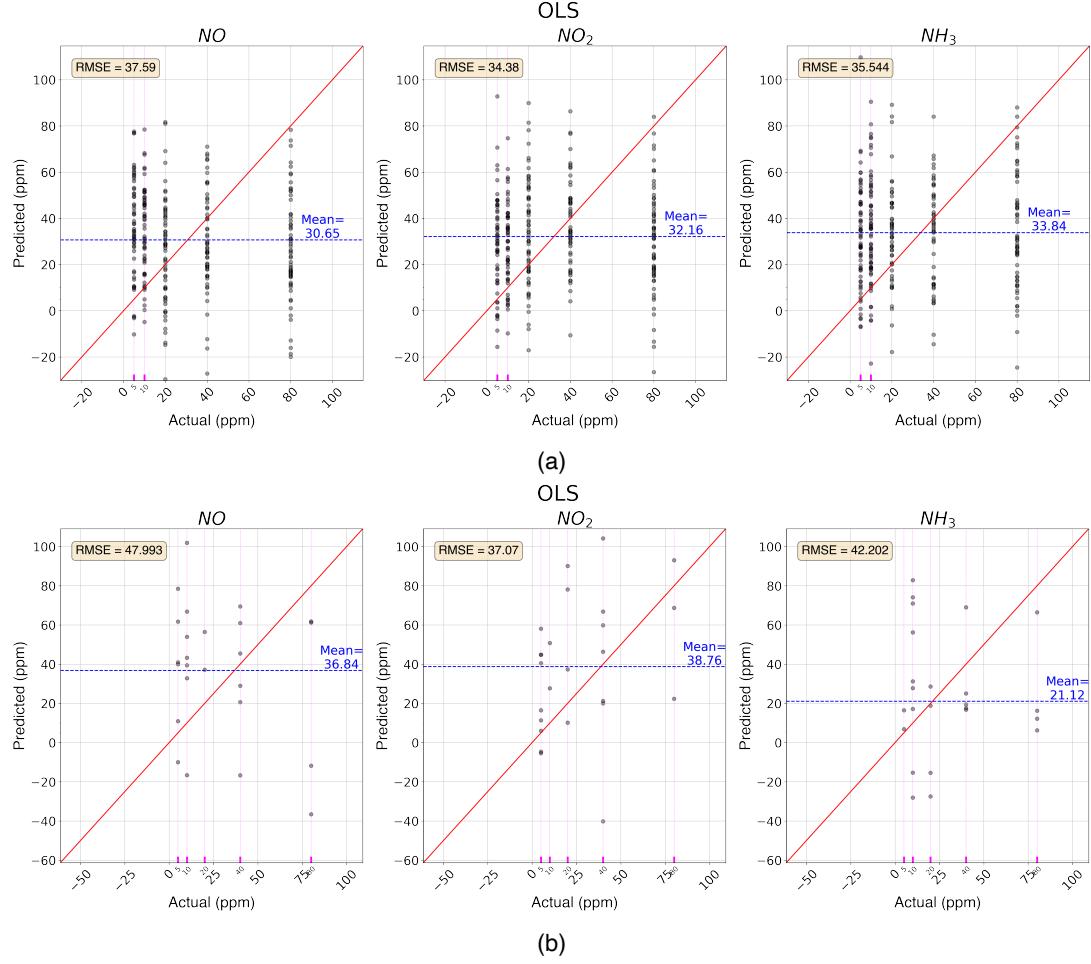


Figure 5.2: Actual vs. Predicted for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

Each subplot in Figure 5.2 corresponds to predictions of NO_x and ammonia concentrations, respectively. The red diagonal line is the identity line. The blue line is the mean of predicted concentrations. The text box contains more information regarding the model fit.

5.2 Principal Components Regression

Following the methodology of Chapter 4, a PCA is conducted with two components in an attempt to visualize the data in a lower dimensional space in Figure 5.3. It is not possible to see any separation between the levels (i.e. concentrations) in both attempts.

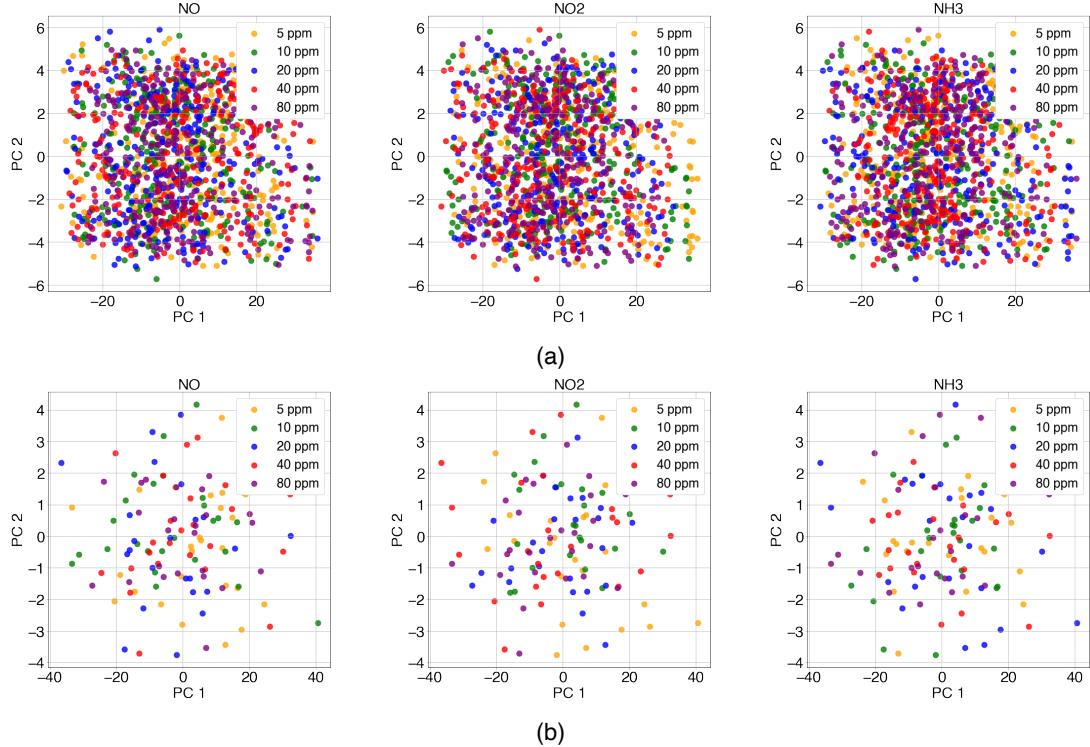


Figure 5.3: PCA for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

Furthermore, an explained variance plot is shown in Figure 5.4. For the first case, the first two PCs explain approximately 40% of the total variance, reaching 80% around 100 components. On the other hand, the second case achieves 90% of explained variability with two components.

After this exploration of PCA, the analysis proceeds to fit a PCR model to the data. The choice of number of PCs was made via cross-validation using RMSE as the loss function, as can be seen in Figure 5.5. Choosing only one components yields the minimum loss for both cases, around 27.

After choosing the number of components, the regression was fit to the training data and used to predict unseen validation data. The results are shown in Figure 5.6. Once again, predicted concentrations are evenly spread around the mean .

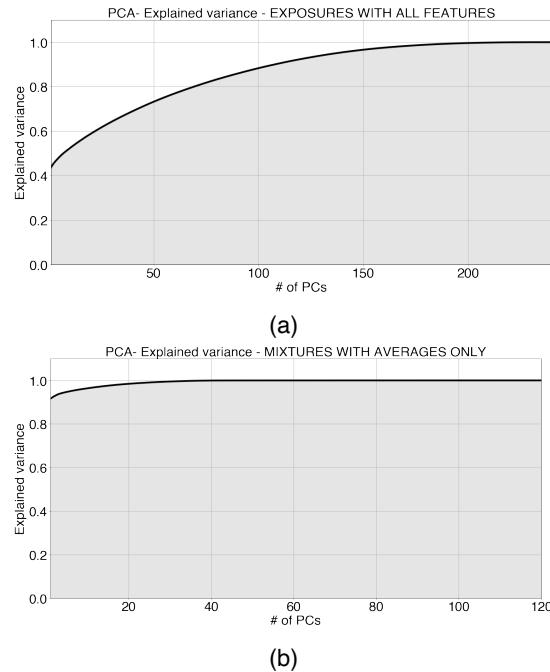


Figure 5.4: Explained variance of PC for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

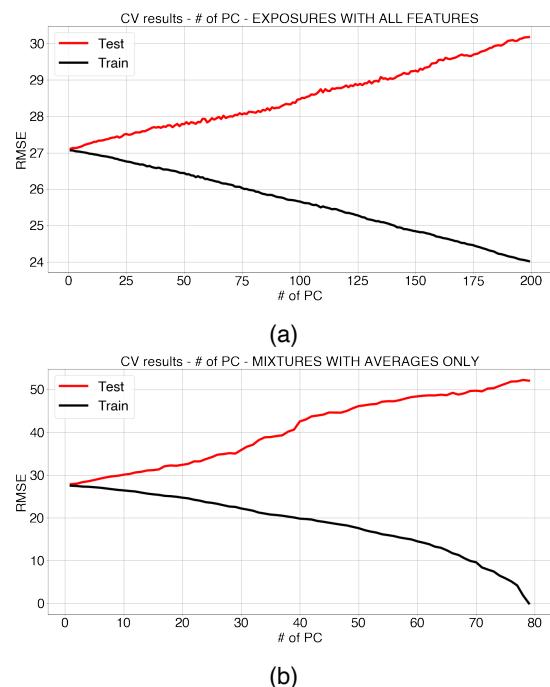


Figure 5.5: Cross-validation results for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

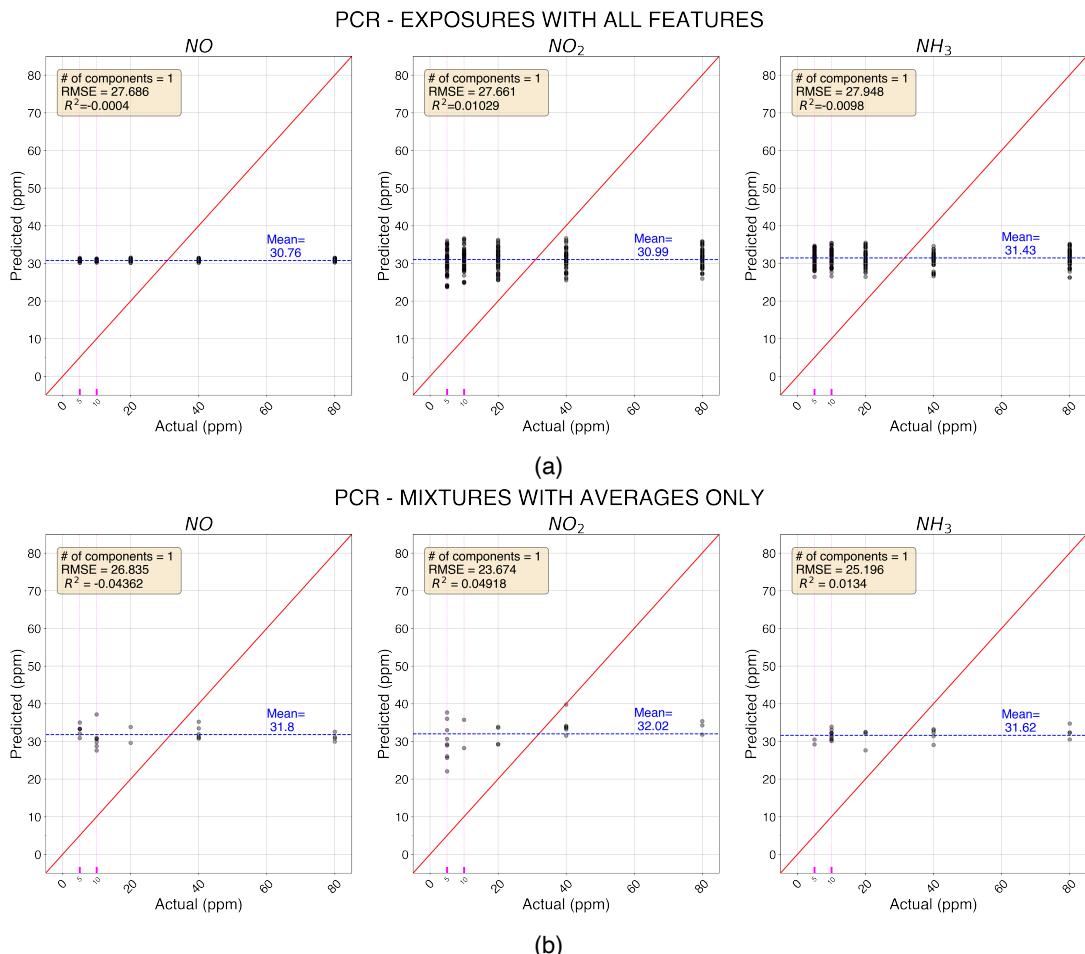


Figure 5.6: PCR for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

5.3 Partial Least Squares Regression

Following the proposed model progression, the analysis proceeds to fit the PLSR model. A similar pipeline to Section 5.2 was used. First, in Figure 5.7, the choice of only two PLS components allowed visualization of data in a two dimensional plot. Moreover, total explained variance is shown in Figure 5.8. Once again, cross-validation using RMSE yields a single component as the best choice with a RMSE of approximately 27, which is then used to fit and predict gas concentrations in Figure 5.10.

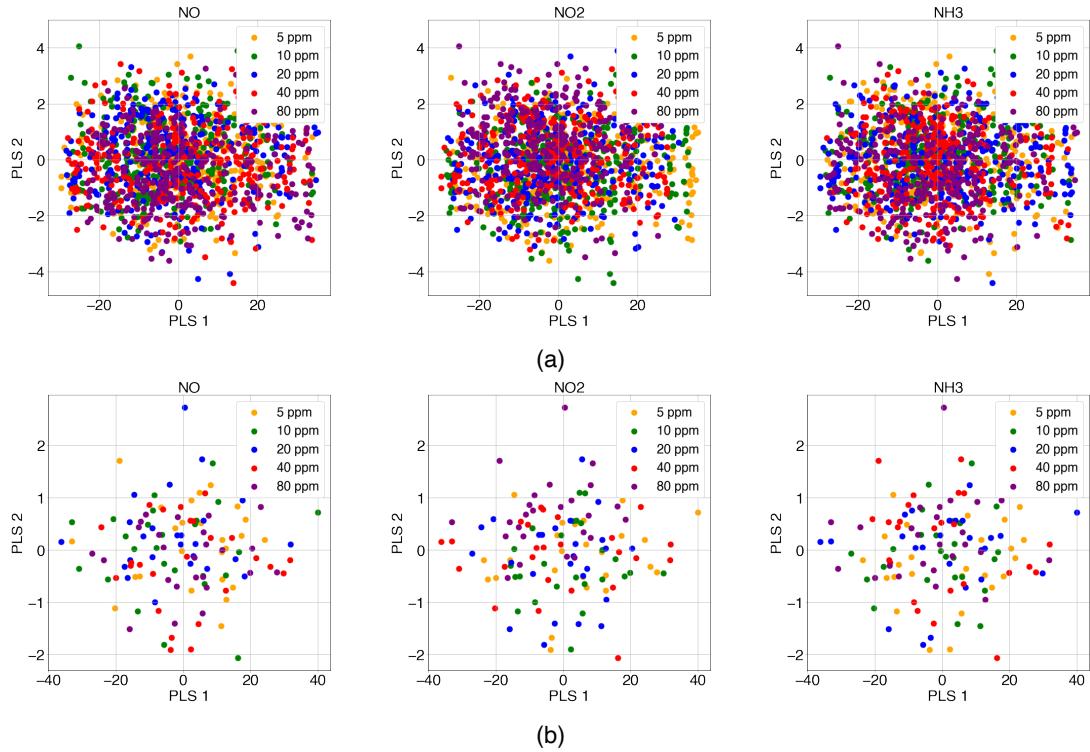


Figure 5.7: PLS scores for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

The results here are similar to PCR, with significant smaller variance in predictions in comparison to OLS and equally spread around the mean (blue lines).

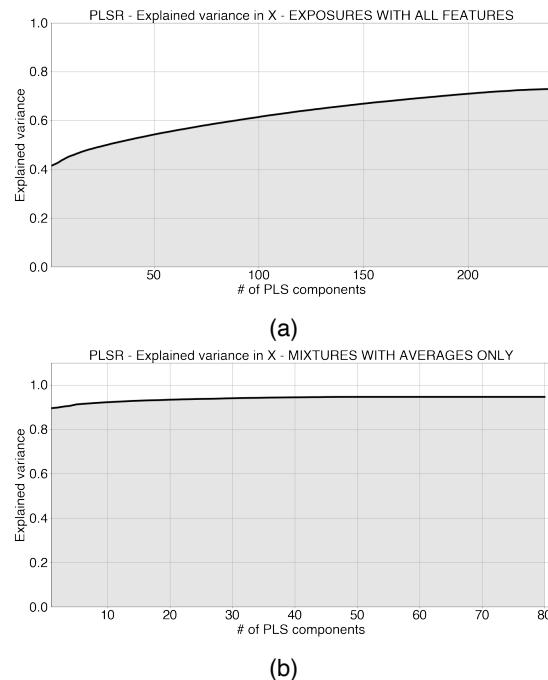


Figure 5.8: Explained variance of PLS components for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

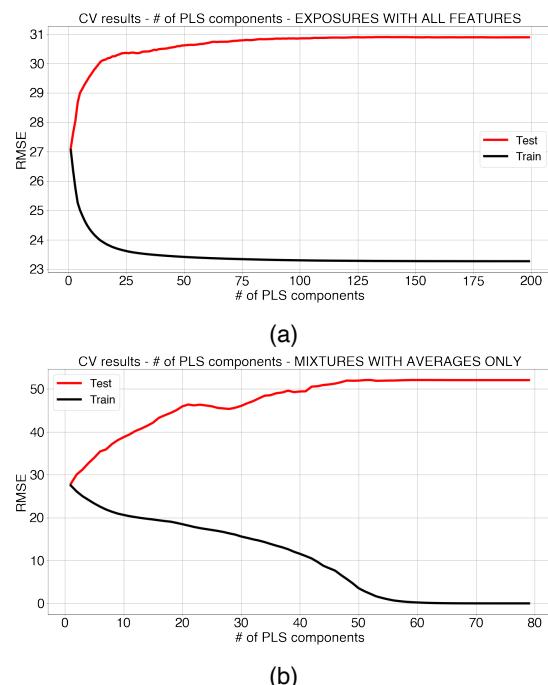


Figure 5.9: Cross-validation results for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

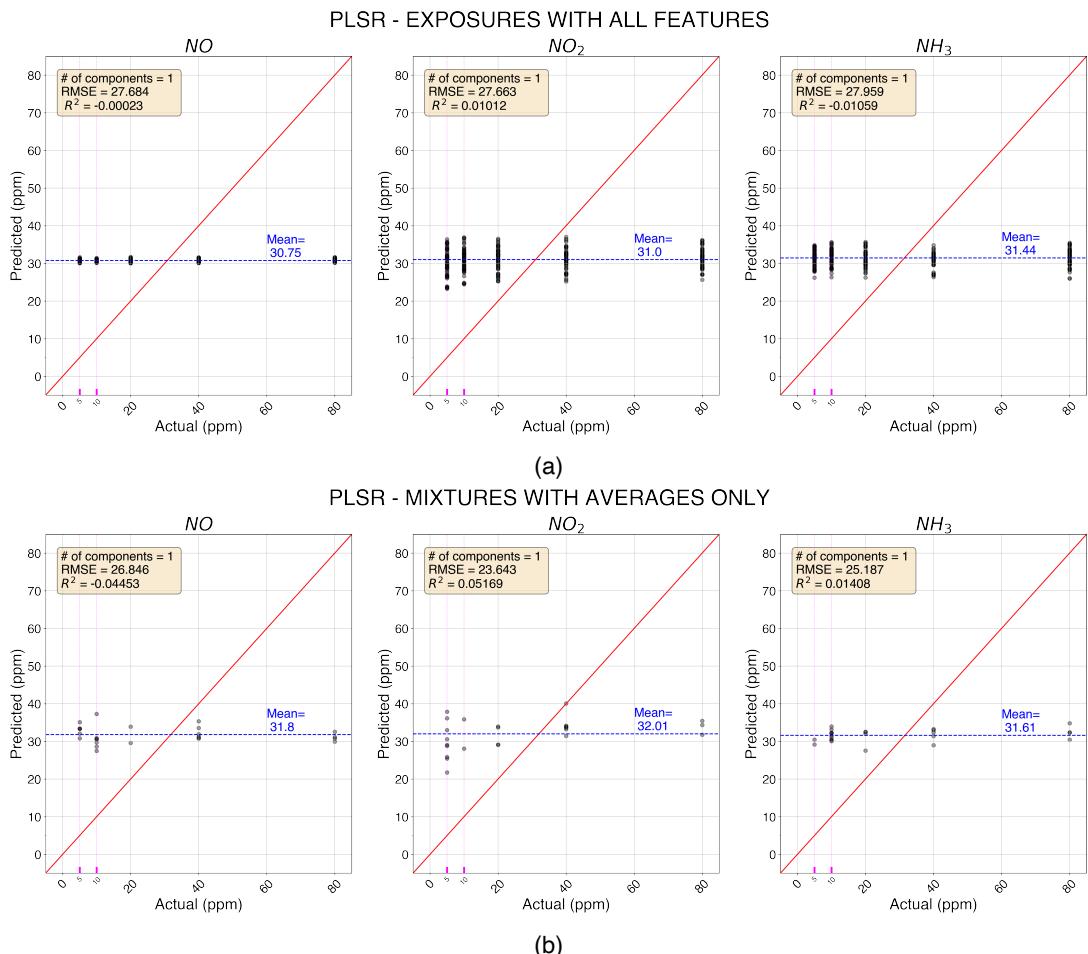


Figure 5.10: PLSR for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

5.4 Ridge Regression

For Ridge regression, the regularization term λ was chosen via cross-validation, as shown in Figure 5.11. Additionally, the shrinkage of coefficients can be seen in Figure 5.12. As expected, the coefficients shrink asymptotically towards zero.

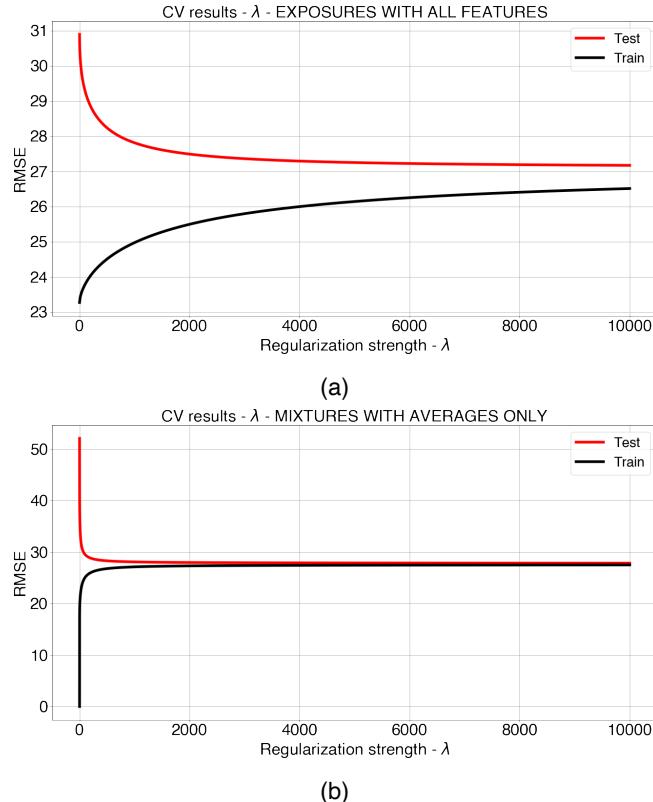


Figure 5.11: Cross-validation results for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.

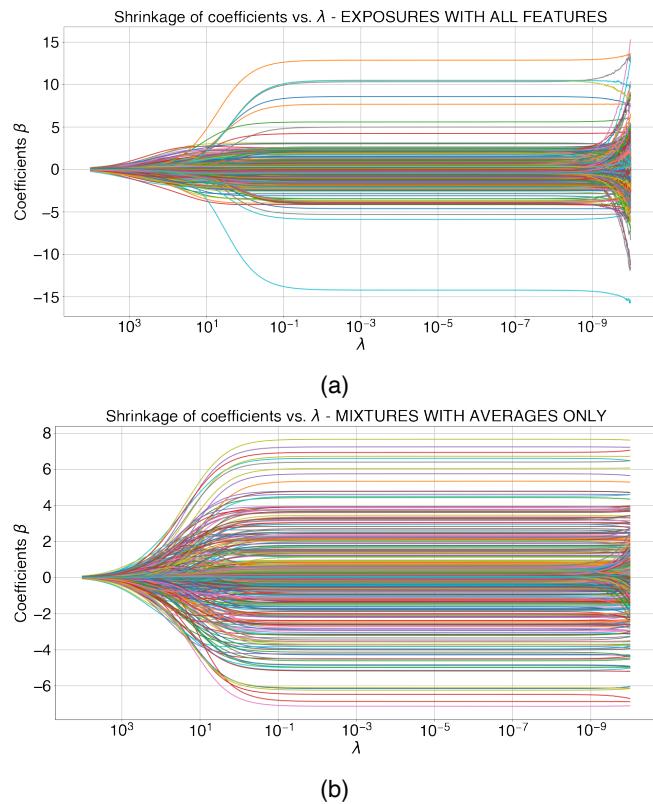


Figure 5.12: Coefficient shrinkage given λ (a) slopes and averages through exposures and (b) only averaged average features through mixtures. Each line corresponds to a coefficient/feature

Finally, after the choice of $\lambda = 10000$, the actual vs. predicted plot is presented if Figure 5.13. Results here, one more time, seem to be centered around the concentration mean of approximately 31 ppm.

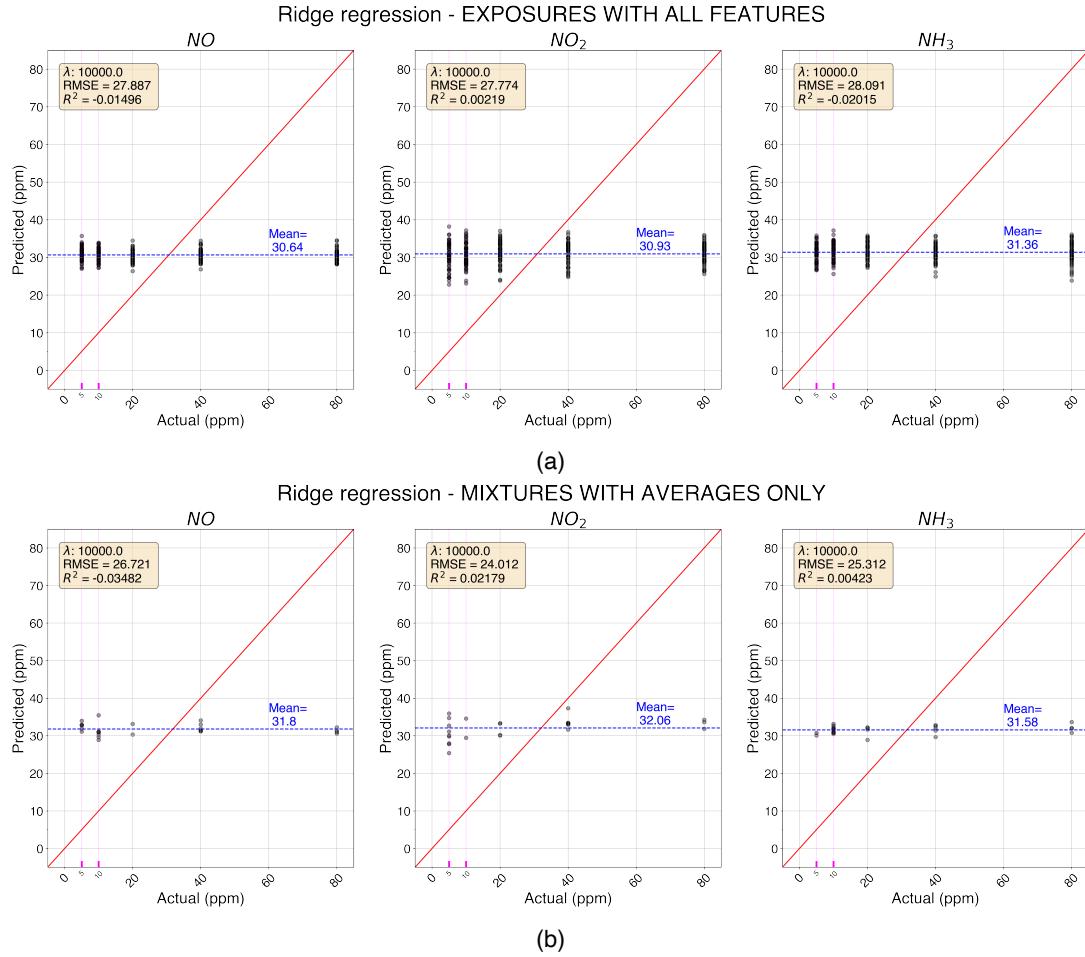


Figure 5.13: Ridge regression predictions for (a) slopes and averages through exposures and (b) only averaged average features through mixtures.



6 Discussion

This chapter is dedicated to explain the results obtained in the previous sections and relating them to statistical theory. The main objective here is to explain why the results are not satisfactory for gas concentration predictions.

6.1 Results

From the results shown in Chapter 5, it is clear that all models fail in predicting gas concentrations. As a first visual assessment, it can be seen from Figures 2.8a and 2.8b that there seems to be no clear order of response variables, i.e. simultaneous gas concentrations.

Slope features, for example, are approximately zero throughout the cycle, with the exception of some particular measurements as seen in Figure 2.8a. Even then, visual inspection indicates that this feature is not informative of gas concentrations. For this reason, the second part of the analysis was done without these features.

Average features, on the other hand, seem to have some separation and indicate that gas concentrations might be explained by it. Nonetheless, it is not possible to order in any clear way. For example, mixtures with high concentrations of NO have average features that vary wildly, and it does not seem to follow any particular linear order of ammonia or NO_x concentration levels. Further attempts at ordering the data can be found in Appendix B.

Ordinary Least Squares was expected to perform poorly, due to the high correlation of average features evidenced in Figure 5.1. Indeed, the results fail to predict gas concentrations at a reasonable level as seen on Figure 5.2a. Looking at the actual vs. predicted plot, it can be seen that the predictions are centered at the mean of concentrations (31 ppm) and have high variance, indicated by the wide range of prediction in all concentration levels.

PCA with two components confirmed previous suspicions: there is no clear separation of gas concentrations for any of the gasses. Although Figure 5.4 shows that 80% of the variance can be explained by approximately 80 components, cross-validation indicate that only one PC yields minimal error before it starts to over fit the training data. Predictions in Figure 5.6 are poor, but have significantly less variance around the concentration mean than OLS. This is expected from this method, as the extraction of PCs is tightly related to the explained variability of the predictors, selecting linear combinations ordered by "importance" to the result.

PLSR, a method that has been shown to work in this type of problem, also performed poorly. Once again, there is not clear separation of concentration levels as shown in Figure 5.7, and CV shows that only one component, again, yields minimal RMSE. Prediction results in Figure 5.10 is very similar to predictions from PCR: centered around the mean with lower variance than OLS.

The final proposed model, Ridge regression, also fails completely in prediction, but brings meaningful insights for analysis. The CV plot for the shrinkage factor λ shows a curious behavior: with low values of λ , regression seems to fit training data well, with a relatively low RMSE of approximately 23. However, this is not the case for validation set. From the plot, extremely high values of regularization yields lowest RMSE in the test set, around 27. From Equations 3.22, 3.23 and 3.24, it can be seen that for high λ , the regression converges to a model that only predicts the mean (in this case, 31 ppm). A virtually "infinite" regularization would achieve best prediction in this case.

6.2 Method

In hindsight, the results indicate that the selection of linear models was ill-advised. Although PLSR seems to work well for problems using TCO, this is not the case for frequency modulation with NO_x and ammonia. For future work, non-parametric models are recommended. Perhaps their high flexibility and lack of assumptions about data could be of aid in achieving better prediction metrics.

Additionally, the frequency cycle itself could be changed. Most notably, instead of a square wave signal, a triangular wave could be more desired, as it would imply in less "stable" sections of the sensor response, possibly yielding more meaningful features, specially slopes.

6.3 The work in a wider context

From a statisticians point of view, it is always a misfortune when analysis' results are not "good" in the sense of high accuracy or low prediction error. These arbitrarily "bad" results, however, bring more insight to the problem, and knowing what does not work might be as valuable as knowing what works.

The quantification of NO_x and Ammonia, as shown in Chapter 1, is of paramount importance in the current world, where combustion processes are still commonplace. Although

the advent of ever-improving electric vehicles is a silver lining regarding gas emissions and combustion processes, some industrial processes cannot avoid it.

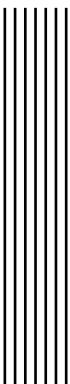


7 Conclusion

Given previous discussions and analysis, the answer to the research question "Can frequency modulation be used to simultaneously quantify NO_x and Ammonia concentrations?" seems to be: "perhaps not". Although poor predictions, the methods used here are far from exhausting the several other, possibly more flexible, models in the statisticians toolbox. Moreover, experiments using different frequency modulations (e.g. triangular waves instead of square) and/or different sensor calibrations and/or different temperatures could be further investigated in order to answer this question conclusively.

As for the second question "Does quality of fit vary over different prediction models?", the correct answer would be "no". All models failed to predict gas concentrations in every concentration and CV results indicate that an underfitted, predict-the-mean model seems to work best in every situation, indicating that the models were not suitable for the regression analysis, as indicated by the R² tending to zero in most models.

The author finds comfort in perhaps pointing future work towards better methods and possibly better quantification of these gases in hopes of addressing the problem more efficiently than current practices. In this sense, this thesis work is considered successful.



Bibliography

- ASTDR (2004). "Sheet for ammonia published by the Agency for Toxic Substance and Disease Registry (ASTDR)." In: 2672, pp. 1–18. URL: <https://www.atsdr.cdc.gov/MHMI/mmg126.pdf> <https://www.atsdr.cdc.gov/mmg/mmg.asp?id=7&tid=2#bookmark02>.
- Bastuck, M. (Jan. 2019). "Improving the performance of gas sensor systems with advanced data evaluation, operation, and calibration methods." PhD thesis, p. 267.
- Bastuck, M. et al. (2016). "Exploring the selectivity of WO₃ with iridium catalyst in an ethanol/naphthalene mixture using multivariate statistics." In: *Thin Solid Films* 618. IX International Workshop on Semiconductor Gas Sensors – SGS'2015, pp. 263–270. ISSN: 0040-6090. DOI: <https://doi.org/10.1016/j.tsf.2016.08.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0040609016304242>.
- Bernabeo et al. (2019). "Health and Environmental Impacts of Nox: An Ultra-Low Level of Nox (Oxides of Nitrogen) Achievable with A New Technology." In: *Global Journal of Engineering Sciences* 2.3, pp. 2–7. DOI: 10.33552/gjes.2019.02.000540.
- Bishop, Christopher M (2006). *Pattern recognition and machine learning*. Springer.
- Boningari, T. and P. Smirniotis (2016). "Impact of nitrogen oxides on the environment and human health: Mn-based materials for the NOx abatement." In: *Current Opinion in Chemical Engineering* 13.x, pp. 133–141. ISSN: 22113398. DOI: 10.1016/j.coche.2016.09.004. URL: <http://dx.doi.org/10.1016/j.coche.2016.09.004>.
- Bur, C., M. Bastuck, A. Lloyd Spetz, et al. (2014). "Selectivity enhancement of SiC-FET gas sensors by combining temperature and gate bias cycled operation using multivariate statistics." In: *Sensors and Actuators B: Chemical* 193, pp. 931–940. ISSN: 0925-4005. DOI: <https://doi.org/10.1016/j.snb.2013.12.030>. URL: <https://www.sciencedirect.com/science/article/pii/S0925400513015037>.
- Bur, C., M. Bastuck, D. Puglisi, et al. (2015). "Discrimination and quantification of volatile organic compounds in the ppb-range with gas sensitive SiC-FETs using multivariate statis-

- tics." In: *Sensors and Actuators B: Chemical* 214, pp. 225–233. ISSN: 0925-4005. DOI: <https://doi.org/10.1016/j.snb.2015.03.016>. URL: <https://www.sciencedirect.com/science/article/pii/S0925400515003391>.
- Dunn, K. (2021). *Process Improvement Using Data*. McMaster University. ISBN: 9781292037578. URL: <https://learnche.org/pid/>.
- Forzatti, P. (2001). "Present status and perspectives in de-NOx SCR catalysis." In: *Applied Catalysis A: General* 222.1. Celebration Issue, pp. 221–236. ISSN: 0926-860X. DOI: [https://doi.org/10.1016/S0926-860X\(01\)00832-8](https://doi.org/10.1016/S0926-860X(01)00832-8). URL: <https://www.sciencedirect.com/science/article/pii/S0926860X01008328>.
- Gareth, J. et al. (2013). *An introduction to statistical learning*. Vol. 112. Springer.
- Guthrie, S. et al. (2018). *Impact of ammonia emissions from agriculture on biodiversity: An evidence synthesis*. Santa Monica, CA: RAND Corporation. DOI: 10.7249/RR2695.
- Hastie, T., R. Tibshirani, and J. Friedman (2001). *The elements of statistical learning*. Vol. 1. 10. Springer series in statistics New York.
- Johnson, R.A. and D.W. Wichern (2013). *Applied Multivariate Statistical Analysis: Pearson New International Edition*. Pearson Education Limited. ISBN: 9781292037578. URL: <https://books.google.se/books?id=xCipBwAAQBAJ>.
- Ng, Kee Siong (2013). "A simple explanation of partial least squares." In: *The Australian National University, Canberra*.
- USEPA (2019). *Nitrogen Oxides Control Regulations*. <https://www3.epa.gov/region1/airquality/nox.html>. Accessed 2021-02-09.
- Wold, S., M. Sjöström, and L. Eriksson (2001). "PLS-regression: a basic tool of chemometrics." In: *Chemometrics and Intelligent Laboratory Systems* 58.2. PLS Methods, pp. 109–130. ISSN: 0169-7439. DOI: [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1). URL: <https://www.sciencedirect.com/science/article/pii/S0169743901001551>.
- Wright, K. (2017). *The NIPALS algorithm*. https://cran.r-project.org/web/packages/nipals/vignettes/nipals_algorithm.html. Accessed: 2021-03-12.



A

Data acquisition time stamps

Table A.1: Data acquisition timestamps.

Index	Frequency (Hz)	Frequency duration (s)	Measurement start time (mm:ss)	Measurement end time (mm:ss)	Index	Frequency (Hz)	Frequency duration (s)	Measurement start time (mm:ss)	Measurement end time (mm:ss)
0	0.05	20	00:00:00	00:00:25	120	0.25	4	00:30:00	00:30:25
1			00:00:25	00:00:50	121			00:30:25	00:30:50
2			00:00:50	00:00:75	122			00:30:50	00:30:75
3			00:00:75	00:01:00	123			00:30:75	00:31:00
4			00:01:00	00:01:25	124			00:31:00	00:31:25
5			00:01:25	00:01:50	125			00:31:25	00:31:50
6			00:01:50	00:01:75	126			00:31:50	00:31:75
7			00:01:75	00:02:00	127			00:31:75	00:32:00
8			00:02:00	00:02:25	128			00:32:00	00:32:25
9			00:02:25	00:02:50	129			00:32:25	00:32:50
10			00:02:50	00:02:75	130			00:32:50	00:32:75
11			00:02:75	00:03:00	131			00:32:75	00:33:00
12			00:03:00	00:03:25	132			00:33:00	00:33:25
13			00:03:25	00:03:50	133			00:33:25	00:33:50
14			00:03:50	00:03:75	134			00:33:50	00:33:75
15			00:03:75	00:04:00	135			00:33:75	00:34:00
16	0.1	10	00:04:00	00:04:25	136	0.50	2	00:34:00	00:34:25
17			00:04:25	00:04:50	137			00:34:25	00:34:50
18			00:04:50	00:04:75	138			00:34:50	00:34:75
19			00:04:75	00:05:00	139			00:34:75	00:35:00
20			00:05:00	00:05:25	140			00:35:00	00:35:25
21			00:05:25	00:05:50	141			00:35:25	00:35:50
22			00:05:50	00:05:75	142			00:35:50	00:35:75
23			00:05:75	00:06:00	143			00:35:75	00:36:00
24			00:06:00	00:06:25	144	1	2	00:36:00	00:36:25
25			00:06:25	00:06:50	145			00:36:25	00:36:50
26			00:06:50	00:06:75	146			00:36:50	00:36:75
27			00:06:75	00:07:00	147			00:36:75	00:37:00
28			00:07:00	00:07:25	148			00:37:00	00:37:25
29			00:07:25	00:07:50	149			00:37:25	00:37:50
30			00:07:50	00:07:75	150			00:37:50	00:37:75
31			00:07:75	00:08:00	151			00:37:75	00:38:00
32			00:08:00	00:08:25	152	2	2	00:38:00	00:38:25
33			00:08:25	00:08:50	153			00:38:25	00:38:50
34			00:08:50	00:08:75	154			00:38:50	00:38:75
35			00:08:75	00:09:00	155	5	2	00:38:75	00:39:00
36			00:09:00	00:09:25	156			00:39:00	00:39:25
37			00:09:25	00:09:50	157			00:39:25	00:39:50
38			00:09:50	00:09:75	158			00:39:50	00:39:75
39			00:09:75	00:10:00	159			00:39:75	00:40:00
40	0.05	20	00:10:00	00:10:25	160	10	2	00:40:00	00:40:25
41			00:10:25	00:10:50	161			00:40:25	00:40:50
42			00:10:50	00:10:75	162			00:40:50	00:40:75
43			00:10:75	00:11:00	163	25	2	00:40:75	00:41:00
44			00:11:00	00:11:25	164			00:41:00	00:41:25
45			00:11:25	00:11:50	165			00:41:25	00:41:50
46			00:11:50	00:11:75	166			00:41:50	00:41:75
47			00:11:75	00:12:00	167			00:41:75	00:42:00
48			00:12:00	00:12:25	168	50	2	00:42:00	00:42:25
49			00:12:25	00:12:50	169			00:42:25	00:42:50
50			00:12:50	00:12:75	170			00:42:50	00:42:75
51			00:12:75	00:13:00	171	100	2	00:42:75	00:43:00
52			00:13:00	00:13:25	172			00:43:00	00:43:25
53			00:13:25	00:13:50	173			00:43:25	00:43:50
54			00:13:50	00:13:75	174			00:43:50	00:43:75
55			00:13:75	00:14:00	175			00:43:75	00:44:00
56	0.1	10	00:14:00	00:14:25	176	25	2	00:44:00	00:44:25
57			00:14:25	00:14:50	177			00:44:25	00:44:50
58			00:14:50	00:14:75	178			00:44:50	00:44:75
59			00:14:75	00:15:00	179	200	2	00:44:75	00:45:00
60			00:15:00	00:15:25	180			00:45:00	00:45:25
61			00:15:25	00:15:50	181			00:45:25	00:45:50
62			00:15:50	00:15:75	182			00:45:50	00:45:75
63			00:15:75	00:16:00	183			00:45:75	00:46:00
64			00:16:00	00:16:25	184	1000	2	00:46:00	00:46:25
65			00:16:25	00:16:50	185			00:46:25	00:46:50
66			00:16:50	00:16:75	186			00:46:50	00:46:75
67			00:16:75	00:17:00	187	2500	2	00:46:75	00:47:00
68			00:17:00	00:17:25	188			00:47:00	00:47:25
69			00:17:25	00:17:50	189			00:47:25	00:47:50
70			00:17:50	00:17:75	190			00:47:50	00:47:75
71			00:17:75	00:18:00	191			00:47:75	00:48:00
72	0.1	10	00:18:00	00:18:25	192	500	2	00:48:00	00:48:25
73			00:18:25	00:18:50	193			00:48:25	00:48:50
74			00:18:50	00:18:75	194			00:48:50	00:48:75
75			00:18:75	00:19:00	195	2000	2	00:48:75	00:49:00
76			00:19:00	00:19:25	196			00:49:00	00:49:25
77			00:19:25	00:19:50	197			00:49:25	00:49:50
78			00:19:50	00:19:75	198			00:49:50	00:49:75
79			00:19:75	00:20:00	199			00:49:75	00:50:00
80			00:20:00	00:20:25	200	200	2	00:50:00	00:50:25
81			00:20:25	00:20:50	201			00:50:25	00:50:50
82			00:20:50	00:20:75	202			00:50:50	00:50:75
83			00:20:75	00:21:00	203	5000	2	00:50:75	00:51:00
84			00:21:00	00:21:25	204			00:51:00	00:51:25
85			00:21:25	00:21:50	205			00:51:25	00:51:50
86			00:21:50	00:21:75	206			00:51:50	00:51:75
87			00:21:75	00:22:00	207			00:51:75	00:52:00
88	0.1	10	00:22:00	00:22:25	208	1000	2	00:52:00	00:52:25
89			00:22:25	00:22:50	209			00:52:25	00:52:50
90			00:22:50	00:22:75	210			00:52:50	00:52:75
91			00:22:75	00:23:00	211	2500	2	00:52:75	00:53:00
92			00:23:00	00:23:25	212			00:53:00	00:53:25
93			00:23:25	00:23:50	213			00:53:25	00:53:50
94			00:23:50	00:23:75	214			00:53:50	00:53:75
95			00:23:75	00:24:00	215			00:53:75	00:54:00
96			00:24:00	00:24:25	216	5000	2	00:54:00	00:54:25
97			00:24:25	00:24:50	217			00:54:25	00:54:50
98			00:24:50	00:24:75	218			00:54:50	00:54:75
99			00:24:75	00:25:00</					



B

Other data plots

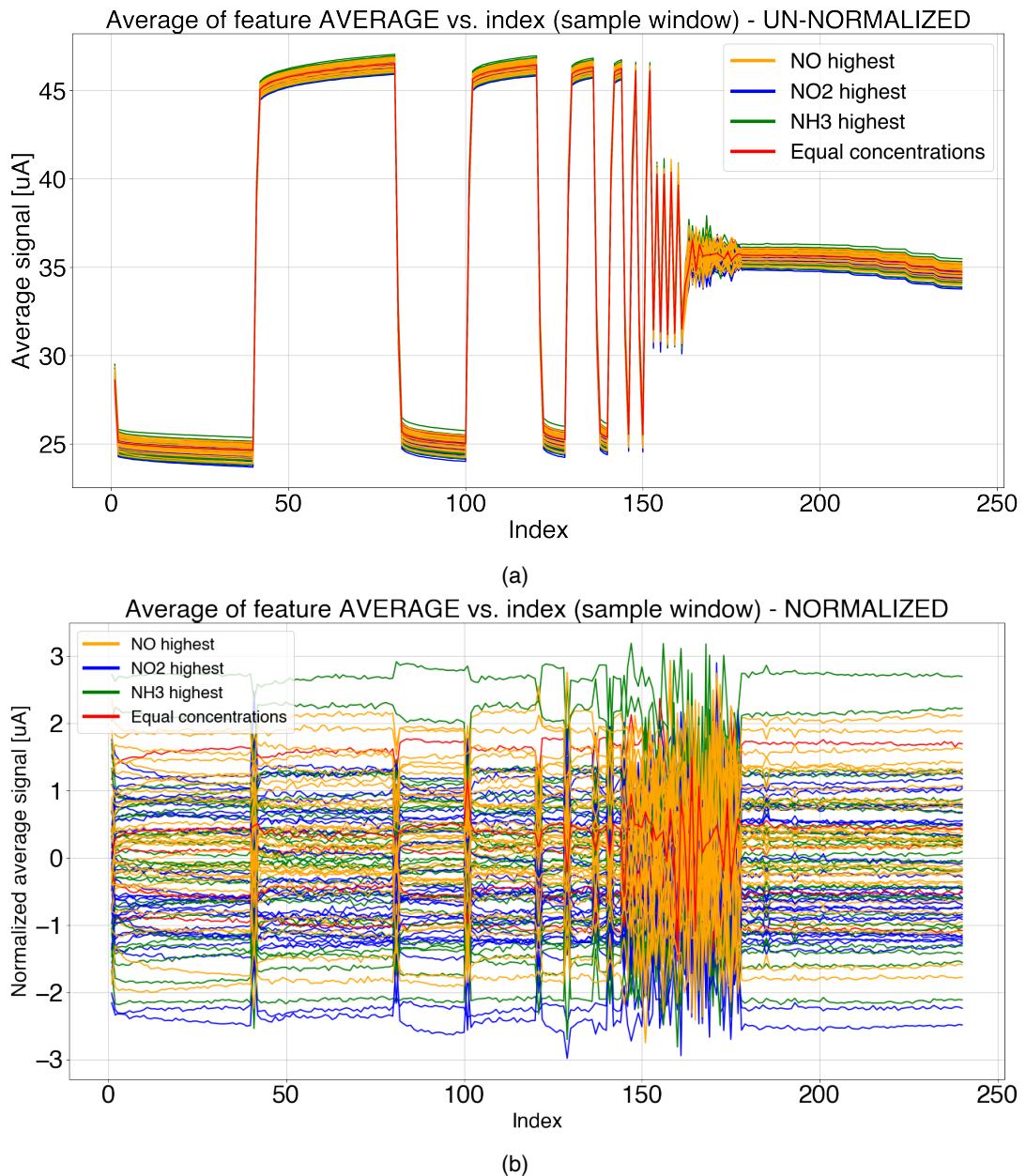


Figure B.1: Averaged sensor average divided by predominant gas. Each line corresponds to an unique mixture.

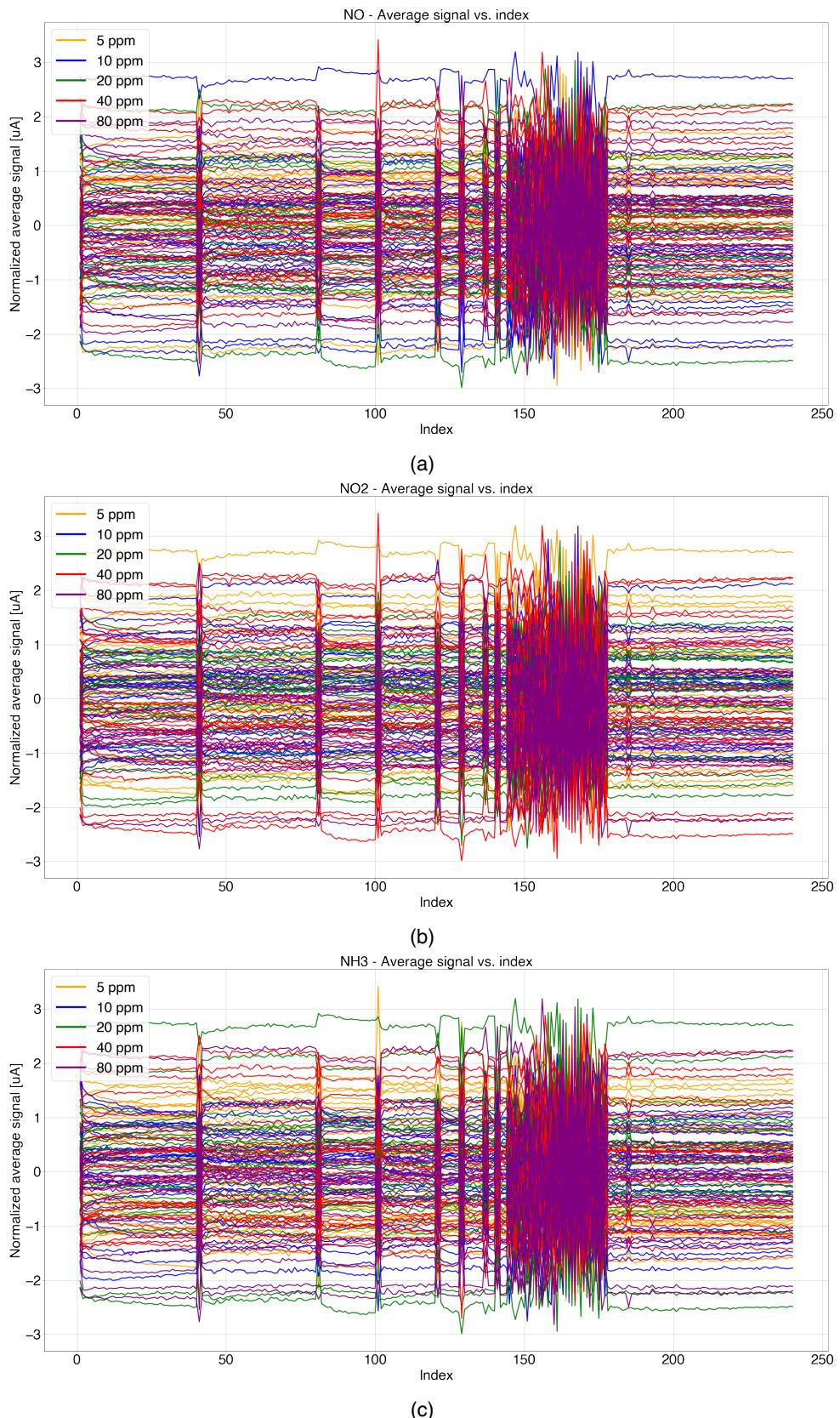


Figure B.2: Normalized sensor averaged per gas. Each line corresponds to an unique mixture. The levels are the concentrations of individual components of the mixture: (a) NO (b) NO₂, and (c) NH₃