

MULTI-CALIBRATION BASED DRIFT COMPENSATION
FOR CHEMICAL SENSOR ARRAYS

A Thesis

by

SUDIP PAUL

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee, Ricardo Gutierrez-Osuna
Committee Members, Radu Stoleru
Pao-Tai Lin

Head of Department, Scott Schaefer

May 2022

Major Subject: Computer Engineering

Copyright 2022 Sudip Paul

ABSTRACT

Long-term application of chemical sensor arrays for continuous monitoring is challenging as a result of sensor aging and drift. A number of techniques have been proposed to compensate for drift, but the issue remains a challenge in this domain. In addition, most drift correction approaches require periodic recalibration of the sensors, which may not be feasible for sensors deeply embedded and deployed for uninterrupted continuous monitoring. In this thesis, we propose a multi-calibration ensemble approach for compensating sensor drift. Our method characterizes drift in the sensor measurements by using past sensor measurements for which ground-truth is available and treating them as “pseudo-calibration” samples along with the recording time of those measurements. Then, we build a regression model that learns to predict the concentration of target analytes given (1) the current sensor measurements and (2) a history of these prior pseudo-calibration samples. The approach is agnostic to the particular regression method used. For this purpose, we evaluate the efficacy of the approach using three different regression techniques, partial least squares, extreme gradient boosting, and neural networks, and compare it against two baselines: regression models that do not use the history of prior pseudo-calibration samples, and a state-of-the-art drift correction autoencoder (DCAE) technique. We evaluated these systems on two experimental datasets from a bioprocess control application, and also characterize their performance as a function of array cross-selectivity and amount of drift in simulation.

Our proposed approach outperforms the calibration-free model and DCAE in the first experimental dataset with errors reduced by as much as 50% in some cases. The correlation between the prediction and the ground truth also improves significantly compared to the comparison methods. On the second dataset, the proposed approach shows improvement in most of the cases compared to the calibration-free model. However, in comparison to the DCAE, only the neural network model shows significant improvement in some cases. In our analysis of the simulated datasets, we have found that the proposed approach shows significant robustness to the presence of drift compared to the other methods. All the three regression techniques using the proposed technique produced a drift-free performance as the amount of drift is increased in the data. As for the analysis with varying cross-selectivity in the simulated sensors, the proposed approach shows significantly lower error compared to the comparison methods. These findings indicate that the proposed technique can generate robust predictions with low error variance and can enhance the reliability of chemical sensing arrays for continuous and long-term applications.

ACKNOWLEDGEMENTS

I thank my advisor Dr. Ricardo Gutierrez-Osuna for all his insights and support throughout my graduate studies. I admire him, and his work ethic and his commitment towards his work.

I would also like to thank Dr. Radu Stoleru and Dr. Pao-Tai Lin for serving as my thesis committee and for their interest in my research and their feedback. I would also like to thank my colleagues Anurag Das and Dennis Rodrigo da Cunha Silva for all their help and support throughout my project and the thesis. It would have been a difficult journey without their help.

Finally, I would like to thank my family and my friends for their never-ending support and for uplifting me during trying times.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis (or) dissertation committee consisting of Dr. Ricardo Gutierrez-Osuna [advisor] and Dr. Radu Stoleru of the Department of Computer Science & Engineering and Dr. Pao-Tai Lin of the Department of Electrical & Computer Engineering.

The data analyzed for Chapter V was provided by Applied Biosensors Inc.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

Graduate study was supported by a graduate research assistantship at Texas A&M University.

This work was funded by a SECO supplement to SBIR Phase II award #1831249 from the National Science Foundation (United States). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Science Foundation (United States).

NOMENCLATURE

ANN	Artificial Neural Networks
CC	Component Correction
DAELM	Domain Adaption Extreme Learning Machine
DCAE	Drift Correction Auto-Encoder
MOX	Metal Oxide
MPC	Multiple Pseudo-Calibration
PLS	Partial Least Squares
QMB	Quartz Microbalance
SEMI	Standardization Error based Model Improvement
XGBoost	Extreme Gradient Boosting

TABLE OF CONTENTS

	Page
ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
CONTRIBUTORS AND FUNDING SOURCES.....	v
NOMENCLATURE.....	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES.....	ix
LIST OF TABLES	xi
CHAPTER I INTRODUCTION	1
1.1 Specific Aims of the Research	4
1.2 Summary of Findings.....	4
1.3 Outline.....	5
CHAPTER II BACKGROUND.....	6
2.1 Chemical Sensing: Drift in Sensor Array.....	6
2.2 Related Work in Drift Correction.....	7
2.2.1 Drift Compensation based on Modeling Sensor Drift.....	8
2.2.2 Drift Correction using Data Standardization.....	10
2.2.3 Drift Compensation using Non-Parametric and Non-Linear Approaches ...	12
2.2.4 Drift Correction using Domain Adaptation.....	14
CHAPTER III METHODS	17
3.1 On-site calibration for drift compensation	17
3.1.1 Problem Statement	17
3.1.2 Proposed on-site calibration technique.....	18
3.2 Regression Models Used for Evaluating the Proposed Approach	22
3.2.1 Partial Least Squares (PLS).....	22
3.2.2 Extreme Gradient Boosting (XGBoost)	22
3.2.3 Multi-Layer Perceptron (MLP)	23

3.3	Evaluation Strategy	24
3.3.1	Proposed approach vs baseline approach	24
3.3.2	Comparison against a state-of-the-art technique	25
CHAPTER IV EXPERIMENTAL SETUP.....		27
4.1	Setups for the Experimental Datasets.....	27
4.1.1	First Experimental Dataset	30
4.1.2	Second Experimental Dataset.....	33
4.2	Simulated Dataset.....	35
CHAPTER V RESULTS		37
5.1	Results on the First Experimental Data:.....	38
5.1.1	Proposed Approach vs Baseline Approach	38
5.1.2	Proposed Approach vs DCAE.....	41
5.2	Results on the Second Experimental Data	43
5.2.1	Proposed approach vs baseline approach	43
5.2.2	Proposed Approach vs DCAE.....	46
5.3	Results on the Simulated Data	48
CHAPTER VI DISCUSSION AND CONCLUSION		56
6.1	Summary of Findings	56
6.2	Limitations and Future Work	60
REFERENCES.....		62

LIST OF FIGURES

	Page
Figure 1 A bioreactor system with a single probe of sensor arrays for continuous monitoring of mixture analytes.....	18
Figure 2 Block diagram of the DCAE model (Yan and Zhang 2016)	26
Figure 3 Experimental Setup.....	28
Figure 4 Raw Sensor Data Collected in the First Experiment.	30
Figure 5 Cross-sensitivity profile of the sensor array used for the first experimental dataset. Correlation coefficients between the target variables and each of the sensors are used as a measure of sensitivity.	31
Figure 6 Sensor reading vs Ground truth concentrations showing presence of drift in sensor measurements collected for the first dataset. The color bar to the right shows the recording time in hours.	32
Figure 7 Cross-sensitivity profile of the sensor array used for the second experimental dataset. Correlation coefficients between the target variables and each of the sensors are used as a measure of sensitivity.	33
Figure 8 Sensor reading vs Ground truth concentrations showing presence of drift in sensor measurements collected for the second dataset. The color bar to the right shows the recording time in hours.	34
Figure 9 Sensor reading vs Ground truth concentrations showing presence of drift in the simulated sensor measurements. The color bar to the right shows the recording time in hours.	36
Figure 10 Performance comparison (using normalized RMSE) between the baseline approach and the proposed (MPC) approach on the first experimental dataset.	39
Figure 11 Correlation between prediction and ground truth concentrations for all analytes for the baseline model (top) and the MPC model (bottom) using MLPs as the underlying regression technique. (* $p < 0.05$; ** $p < 0.001$).....	41
Figure 12 Performance comparison (using normalized RMSE) between proposed (MPC) approach and DCAE.....	42

Figure 13 Correlation between prediction and ground truth concentrations for all the analytes; tested on the first experimental dataset.	43
Figure 14 Performance comparison (using normalized RMSE) between the baseline approach and the proposed (MPC) approach.....	44
Figure 15 Correlation between prediction and ground truth concentrations for all the analytes; tested on the second experimental dataset.....	45
Figure 16 Performance comparison (using normalized RMSE) between proposed (MPC) approach and DCAE.....	46
Figure 17 Correlation between prediction and ground truth concentrations for all the analytes.	47
Figure 18 Visualizing the condition number for sensitivity matrices. (a) Sensitivity matrix with high cross-sensitivity and high condition number, (b) Sensitivity matrix with no cross sensitivity and very low condition number...	49
Figure 19 Normalized RMSE obtained at different complexity (different condition numbers) of the regression task.	51
Figure 20 Normalized RMSE obtained at different amount of drift added to the data. The amount of drift is shown as a percentage of the range of amplitude of the original signal.	53
Figure 21 Normalized RMSE obtained at different amount of training data used. The size of the training data is represented as a percentage of the total number of data samples used during training the three MPC models in the default setting.....	55

LIST OF TABLES

	Page
Table 1 Model Parameters Considered for XGBoost.....	23
Table 2 Target Analytes and Sensors.....	29
Table 3 p-values for the pairwise 2 sample t-test between the baseline approach and the MPC approach	40

CHAPTER I

INTRODUCTION

Reliable and continuous monitoring of biomarkers and bioprocess variables finds application in healthcare and related fields such as drug manufacturing and pharmaceutical industries (Guiseppi-Elie, Brahim, Slaughter, & Ward, 2005; Kumar et al., 2001; O'Mara, Farrell, Bones, & Twomey, 2018). As such, sensor systems that provide simultaneous and continuous measurements find high commercial demand (Nguyen, Tathireddy, & Magda, 2018). A plethora of sensing modalities offers diverse functionality in detecting and monitoring different biomarkers. However, regardless of the technology or sensing materials used, these sensing modalities have several limitations. These include accumulation of temporal drift over time, cross-sensitivity of the sensor arrays, thermal noise in the sensors, drift and aging effects being the most common phenomena (Sasaki, Josowicz, Janata, & Glezer, 2006). Drift can occur in sensor measurements due to environmental fluctuations and can accumulate over time (Di Carlo & Falasconi, 2012; Ma, Luo, Qin, Wang, & Niu, 2018), while aging or poisoning effect on the sensors often happens due to degradation and sensitivity changes of the sensing materials over long-term usage (Anik, Guilley, Danger, & Karimi, 2020; Wenzel, Mensah-Brown, Josse, & Yaz, 2010). These can cause degradation of sensing performance, resulting in unreliable measurements and inaccurate identification of target analytes.

To overcome the issue of drift in chemical sensor arrays, several techniques have been proposed. The most effective drift compensation technique is periodic recalibration

of the sensor arrays using a stable reference solution (Ricardo Gutierrez-Osuna, 2002; Haugen, Tomic, & Kvaal, 2000). Several data-driven drift compensation techniques also exist in the literature, using various signal processing techniques (Marco & Gutierrez-Galvez, 2012; Wenzel et al., 2010), dimension reduction methods (Perera, Papamichail, Bârsan, Weimar, & Marco, 2006), and machine learning approaches (Marco & Gutierrez-Galvez, 2012; Verma, Asmita, & Shukla, 2015). However, most of these methods do not work well on long-term application cases and require recalibration/retraining of the model using intermittent data collection (Leon-Medina, Pineda-Muñoz, & Burgos, 2020). Sensor recalibration using a reference solution or retraining of the prediction models are not suitable for applications where uninterrupted monitoring for long periods is required. For example, in an application where the sensors are deeply embedded into the system (e.g., a bioreactor), it is impossible to recalibrate the sensors by introducing them to an external reference solution or perform controlled sample collection for retraining without interrupting the ongoing process. As a result, the drift compensation technique must work with the available information within the system.

In this work, I propose an on-site calibration scheme based on the assumption that the user can take out samples from the bioreactor during an experiment and provide ground truth concentrations for that sample using an offline analyzer. The proposed system can store all the sensor measurements and when the concentrations of a past sample are available, it can generate a calibration point using the stored measurements and the provided concentrations. The proposed approach uses such calibration points to generate the current input to the prediction model. In my approach, the input is constructed using

the difference between the current sensor measurement and the past sensor measurement at the calibration point. In addition, I also feed in the ground truth concentrations provided by the user and the time elapsed between the calibration point and the current sample. Finally, this input is fed to a previously trained regression model for predicting the concentrations of the current sample. In this work, I show that this approach improves accuracy regardless of the prediction model used. The system also employs an ensembling step, where it can combine any number of past calibration measurements provided by the user. This system has two major advantages: it does not require the sensor array to be interrupted at any moment in time, and it can utilize an arbitrary number of past measurements as calibration updates. The proposed approach accepts such past sensor measurements as calibration points and following a mathematical model, learns the temporal behavior of the drift in the measurements. This, in turn, allows the predictive model to generate drift-free predictions without requiring explicit recalibration of the sensors.

To evaluate the performance of the proposed approach, I propose to implement our approach on top of three regression techniques: partial least squares (PLS), extreme gradient boosting (XGBoost), and multi-layer perceptron (MLP). For each model, along with the proposed approach, I have implemented a baseline approach where we use only the sensor measurements as input. Finally, I have also implemented a state-of-the-art drift correction method, drift correction autoencoder (DCAE) (Yan & Zhang, 2016b), as the comparison method. I have tested the different models on two experimental datasets

collected using different hydrogel-based magneto-resistive sensor arrays developed by Applied Biosensors, Inc (Nguyen et al., 2018).

In addition, I have generated a simulated dataset to mimic the experimental setup and evaluate the performance of the proposed approach in scenarios involving different amounts of drift and different levels of cross-sensitivity of the sensor array. In the simulated dataset, I vary the complexity of the prediction problem by changing the amount of cross-sensitivity in the sensor array.

1.1 Specific Aims of the Research

The specific aims of this work are summarized as follows:

Aim – 1: Develop an on-site multi-calibration drift compensation technique for analyte concentration prediction in chemical sensor arrays.

Aim – 2: Evaluate the performance of the developed technique on experimental and simulated datasets using different regression models and compare against a state-of-the-art drift correction method.

1.2 Summary of Findings

I have tested our approach on top of three regression techniques and compared the performance with respect to a baseline approach and a state-of-the-art drift correction technique (DCAE). The results from the experiments can be summarized as:

- ❖ The proposed approach, when implemented with PLS and MLP, performs better than the baseline approach on all three datasets. In the case of XGBoost, the

proposed approach performs better than the baseline approach on two of three datasets.

- ❖ Compared to the DCAE, the proposed approach performs better in two of the three datasets. For the third dataset, the performance of the two techniques are comparable.
- ❖ In the experiment with increasing amounts of complexity in the data, the three regression models combined with the proposed approach perform better than both the baseline approach and the DCAE.

1.3 Outline

The rest of the manuscript is organized as follows. Chapter 2 discusses prior work on chemical sensor calibration, and different drift correction methods for chemical sensor arrays. Chapter 3 describes the proposed multi-calibration drift correction method and the evaluation techniques employed in this work. In Chapter 4, we discuss the details of the experimental setup and then in Chapter 5 we discuss the results obtained through our experiments and explain the implications of the results. Finally, Chapter 6 includes a discussion of the work, present ideas for future work and concludes the thesis.

CHAPTER II

BACKGROUND

In this chapter, I first provide a brief background of the chemical sensing technology, the limitations of the present technology and describe the effect of drift on sensing systems that are deployed for continuous monitoring. Then, I provide a review of prior work on drift correction techniques available in the literature.

2.1 Chemical Sensing: Drift in Sensor Array

Chemical sensors measure and detect chemical qualities in an analyte (a chemical substance being observed) and convert the sensed chemical data into electronic data. Chemical sensors are used in various applications, such as medical, automotive, nanotechnology and home detection systems (i.e. carbon monoxide detectors). As such, sensor systems that provide simultaneous and continuous measurements find high commercial demand (Nguyen et al., 2018). A plethora of sensing modalities offers diverse functionality in detecting and monitoring different chemical targets. However, reliable deployment of such sensor systems for long-term monitoring is limited by the presence of temporal drift and the limited temporal validity of the calibration models (Rudnitskaya, 2018). In addition to accumulation of temporal drift over time, cross-sensitivity of the sensor arrays, thermal noise in the sensors, and aging are the most concerning issues (Sasaki et al., 2006). Drift can occur in sensor measurements due to environmental fluctuations and can accumulate over time (Di Carlo & Falasconi, 2012; Ma et al., 2018). This can cause degradation of sensing performance, resulting in unreliable measurements

and inaccurate identification of target analytes. The most effective drift compensation technique is the periodic recalibration of the sensor arrays using a stable reference solution (Ricardo Gutierrez-Osuna, 2002; Haugen et al., 2000). However, frequent recalibration of multisensor systems is often excessively costly and time consuming due to the large number of necessary reference sample and their limited availability.

Several data-driven drift compensation techniques exist in the literature that has applied different signal processing techniques (Marco & Gutierrez-Galvez, 2012; Wenzel et al., 2010), dimension reduction methods (Perera et al., 2006), and machine learning approaches (Marco & Gutierrez-Galvez, 2012; Verma et al., 2015). However, most of these methods do not work well on long-term applications and require recalibration/retraining of the model using intermittent data collection (Leon-Medina et al., 2020). As a result, data-driven drift compensation techniques requiring small number of standard samples or no standard samples at all or techniques leveraging calibration transfer or update can be advantageous.

2.2 Related Work in Drift Correction

Data driven techniques have been used for sensor drift compensation for several decades. However, the nature and the complexity of the drift varies quite a bit with different sources of drift. As a result, several different strategies have been used for drift compensation.

2.2.1 *Drift Compensation based on Modeling Sensor Drift*

Drift compensation using modeling methods presume that sensor drift can be separately modeled from the analytical signal and such a model can be used for the correction of drift from the sensor array response when deployed. One such approach is the component correction (CC) which assumes that sensors in an array have similar behavior with respect to drift and that this drift has the same direction for all measured samples and reference gas (Artursson et al., 2000; R Gutierrez-Osuna, 2000). Therefore, drift correction can be done by identifying and modeling drift direction in the reference samples and subtracting it from the new data. Several different techniques have been used to implement this concept. Artursson et al. (Artursson et al., 2000) proposed a PCA application to drift modeling based on the following concept. Given that sensor responses in the reference samples have significant drift, the first components in a PCA model, calculated using only measurements in reference sample, will describe the direction of the drift. Therefore, the loading vector of the PCA model is attributed to the noise and used to calculate projection of drift in the new measurements. Drift correction is thus performed by subtracting this calculated drift component from the new samples. It is reported in the article that the proposed drift correction technique improves recognition and classification performance when applied on a MOX sensor array system. Ziyatdinov et al. (Ziyatdinov et al., 2010) proposed a more generalized approach considering sensors behaving differently in different samples. Their technique uses Common Principal Component Analysis (CPCA) to calculate the loading vector p so that it expresses common covariance for all classes (gases) instead of variance observed in the reference gas. When applied to

a sensor array containing 17 polymeric sensors, CPCA is reported to be performing better than both PCA and uncorrected data. Grover et al. (Grover & Lall, 2020) proposed an adaptive Kalman filter based drift correction model on top of the PLS based regression analysis which shows that removing baseline drift plays a crucial role in improving predictive performances.

Drift correction using PLS and Canonical Correlation Analysis (CCA) has been described in (R Gutierrez-Osuna, 2000). This technique calculated projections of both the reference gas and the samples that are maximally correlated and use the projection of the sample to model the drift component in the samples. PLS and CCA were used to find the projection matrices in the first step of the algorithm and Ordinary least squares (OLS) regression is used to subtract the drift component from the test samples. This technique was applied to the measurements made by an array of 10 MOX sensors. Results indicate that correct classification rate of 95% was maintained for up to 10 consecutive measurement sessions when at least 5 days of measurements were used for calibration. This was significant improvement in comparison to uncorrected data, for which classification rate varied between 70 and 80% in the same settings. It was also reported that success of the drift correction depended on the size of the calibration data set and on the period of time elapsed between last calibration measurement and the test sample.

A major limitation with drift modeling methods is that a relatively long series of measurements is required to produce a drift correction model. Measurements made during several weeks are typically used. Furthermore, it can be expected that the drift correction model would become invalid after a while as the sensing continues and would need an

update, which logically requires collecting further reference measurements. Another issue is that most of these approaches rely on the assumption that sensor drift has a regular direction either in the actual signal or the projected subspace and is limited by the fact that long-term drift do not always follow a consistent trend (Liu, Hu, Ye, Cheng, & Li, 2015).

2.2.2 Drift Correction using Data Standardization

Data standardization techniques aim to correct new samples by eliminating new drift components calculated using a relationship between two experimental conditions. These corrected samples are then used for concentration prediction in place of the raw samples. Reduced set of standards measured at both conditions, also called standardization or transfer data set, is used for such correction. Two approaches are possible for data standardization: correction of the concentrations predicted at new conditions and correction of signals measured in the new conditions.

Signal standardization is by far the most widely used approach among calibration update methods. It has been applied to drift removal and calibration transfer for both electronic noses and electronic tongues. Signal standardization techniques try to capture a relationship between sensor responses in the initial (at the time of calibration) and new conditions (in standardization/transfer sample set) for correction of the data measured in unknown samples in new conditions. Calibration transfer from one electronic nose to the other using robust regression has been proposed used for drift correction in several previous works (Deshmukh, Bandyopadhyay, Bhattacharyya, Pandey, & Jana, 2015; Shaham, Carmel, & Harel, 2005). Calibration transfer between two electronic noses

employing different types of sensors, one with QMB sensors and another with conducting polymer sensors, has been described in (Shaham et al., 2005). Performance in data standardization of Multiple Linear Regression (MLR), Partial Least Squares (PLS2), Principal Component regression (PCR), ANNs and a method introduced in this work called Tessellation-based linear interpolation (TLT) was compared. The TLT consists of two stages: tessellation and prediction, where Tessellation is done in such a way that all vertices of all simplexes are calibration set X points. Prediction of class membership of a new sample is done by first locating simplex enclosing vector of sensor responses and calculating barycentric coordinates of this sample relative to the vertices of simplex that encloses it. However, the results indicate that among all data standardization methods, ANN proves to be the most effective for the studied data. Unfortunately, no comparison with uncorrected data was shown. One other drawback of such technique is that relatively large data sets or even entire calibration sets are necessary for an efficient calibration transfer.

Application of the Robust Weighted Least Squares (RWLS) to data standardization was described in (Zhang et al., 2011). RWLS is a robust regression algorithm with a property of being less sensitive or more “robust” in the presence of outliers in the data (Heiberger & Becker, 1992). Calibration transfer was done from one master to five slave electronic noses, all equipped with 3 MOX sensors. Transfer data comprising 5 samples of a reference gas were selected by the Kennard-Stone algorithm. Data standardization by RWLS allowed to achieve lower concentration prediction errors compared to uncorrected

data. It was also observed that efficiency of data standardization varied between instruments.

Methods used for data standardization differ in the way the relationship between two sets of sensor signals is calculated: single wave standardization (SWS) calculates relationship between each signal individually, piecewise direct standardization (PDS) between groups of signals and direct standardization (DS) between all signals. SWS and DS have been used for data standardization in (Panchuk et al., 2016) along with LS regression and MLR for the data mapping. While both standardization methods were shown to be effective in drift removal, SWS performed better with lower errors compared to both DS and uncorrected data. The same two techniques along with MLR and Tikhonov regularization were used for the calibration transfer between two identical arrays of potentiometric sensors (Khaydukova, Panchuk, Kirsanov, & Legin, 2017). In this study, however, DS with Tikhonov regularization performed better, producing lowest errors for all tested models. Results reported in these two studies indicate that performance of the data standardization methods is dependent on the data and probably on the composition of standardization data sets as well.

2.2.3 Drift Compensation using Non-Parametric and Non-Linear Approaches

Non-linear approaches like artificial neural networks (ANN), specially deep neural networks (DNN) are becoming more and more popular for drift compensation because of their ability to approximate complex behavior of long-term drift (Jiao, Hu, Xu, & Wang, 2020; Zhao et al., 2019). Verma et al. (Verma et al., 2015) reported that classification performance of an ensemble of classifiers is improved by the introduction of a non-smooth

convex regularized loss to compensate for the sensor drift. This study employs a weighted support vector machine (SVM) classifier ensemble and show that the use of the regularization technique improves classification performance by compensating for the time dependent drift component. SVR is also used for continuous gas concentration estimation in a gas mixture in real time (De Vito et al., 2007). This proposed method implements a time delay SVR to compensate for slow transients in the sensor array and the results indicate that the TD-SVR slightly outperforms a time delay Neural Network (TDNN).

A multiple classifier approach integrating Artificial Neural Networks (ANN) and K-Nearest Neighbors (KNN) is proposed in (Adhikari & Saha, 2014), which is reported to improve the performance of classification while mitigating the sensor drift. Abbatangelo et al. (Abbatangelo, Nunez-Carmona, Sberveglieri, Comini, & Sberveglieri, 2020) has proposed a hybrid KNN-ANN approach to counteract sensor drift in chemical sensing. Their approach uses k-nearest neighbors based on Euclidean distance of the test data and the training data in the feature space and use the mean values of the k neighbors as input to the ANN. More recently, deep neural network architectures have been used for drift compensation by utilizing temporal properties of the input space. Zhao et al. (Zhao et al., 2019) combined Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) units with a custom loss function based on cross-entropy loss. The ensemble technique used in this method is based on combining the advantages of SVM for small samples and the advantages of LSTMs to deal with time-varying drift; it is reported that the method performs well with a classification accuracy around 99%.

A Gated Recurrent Unit (GRU) neural network model is proposed for predicting gas concentration in (Wang, Hu, Burgués, Macro, & Liu, 2020). A single GRU layer is used on raw data from low-cost MOX sensors and the proposed model achieves promising result in dealing with the drift and cross-sensitivity issues of the sensors. An attention based GRU model is proposed in (Chaudhuri, Wu, Zhang, Liu, & Li, 2020) for drift compensation. This study focuses on predicting sensor drift, rather than correcting for it, by leveraging the GRU's ability to capture time-dependencies to model temporal information of drift while utilizing an attention mechanism to reduce information loss.

2.2.4 Drift Correction using Domain Adaptation

While non-parametric approaches have been shown to perform better than most classical approaches, their efficacy is sometimes limited by the domain knowledge and the amount of data available to train such models. In recent times, several domain adaptation and transfer learning approaches have emerged that focus on eliminating the effects of cross-domain sensitivity to make the prediction models more robust (Rudnitskaya, 2018). The prediction models employed in these techniques range from classical regression techniques to state-of-the-art deep neural network architectures.

One way to achieve domain adaptation, as reported in literature, is based on the expansion of calibration model by joining initial calibration data set and measurement made in new conditions in the subset of standardization or transfer samples, and recalculating the calibration model (Rudnitskaya, 2018). In this way, new sources of variance are incorporated in the updated calibration model, which allows to decrease prediction errors for the samples measured in new conditions. Application of three

methods of calibration model expansion, namely weighting, Tikhonov regularization and Joint-Y PLS has been reported in (Rudnitskaya, Costa, & Delgadillo, 2017) and compared against data standardization methods, slope and bias correction of predicted values and DS with PLS2 regression and ANN for data mapping. Model expansion methods are shown to perform better achieving lowest root mean square error of prediction (RMSEP) with smaller number of transfer samples compared to the data standardization techniques. Model expansion by a variant of ridge regression called the transfer sample-based coupled task learning (TCTL) has been reported for an electronic nose in (Yan & Zhang, 2016a). Two tasks were addressed: calibration transfer using data set described in (Yan & Zhang, 2015) and calibration update using long-term drift data set described in (Vergara et al., 2012). For both tasks and data sets, TCTL obtained better results compared to the uncorrected data and performed similarly to the combination of variable standardization with Standardization Error based Model Improvement (SEMI) (Yan & Zhang, 2015).

Zhang et al. (2017) proposed an unsupervised subspace projection approach for drift reduction in. The key idea is to use mean distribution discrepancy (MDD) to remove the drift in the latent subspace. The proposed domain regularized component analysis (DRCA) model combines multi-class SVM with RBF kernel and is shown to outperform several classical ML techniques in terms of gas identification. Another calibration model expansion technique is reported in (Zhang & Zhang, 2014) which achieves domain adaptation by inclusion of the transfer samples in the calibration model. Two approaches based on extreme learning machines or back-propagation neural networks were employed. The first one, called the source domain adaption extreme learning machine (DAELM-S),

uses transfer samples for regularization or update of the calibration model while the second one, target DAELM or DAELM-T, works similarly to a semi-supervised adaptive neural network. Both algorithms were shown to be more successful in drift reduction compared to CC by PCA and SVM classification models with correct classification rate close or above 90%. One important thing to note is that relatively large number of transfer samples (20 to 30 for DAELM-S and 40 to 50 for DAELM-T) were necessary for these algorithms to successfully function.

Yan et al. (2016b) proposed an autoencoder based transfer learning approach for adaptive domain adaptation in order to correct for drift and instrumental variation. Their method incorporates domain features in addition to original features from both source and target domains and introduces a correction layer for dealing with complex drift. The performance of the DCAE was evaluated using data sets from three different domains with different tasks. DCAE is shown to perform better than most drift correction approaches and more robustly specially when training and test data were collected several months apart. While transfer learning approaches show promising results in drift compensation, they require retraining in the target domain making the approach less effective in a real-life deployment. Also, the transfer samples used in this method requires the source and target domains to have similar patterns limiting the availability of transfer samples and in turn adversely affecting the performance of the model.

CHAPTER III

METHODS

The objective of this study is to develop a data-driven drift compensation technique for chemical sensor arrays that can be executed on-site without sensor recalibration. For this purpose, I propose a drift compensation approach that takes in user-provided ground truth for one or multiple past measurements (pseudo-calibration samples) and utilizes a mathematical drift model on top of any regression technique. In this chapter, I will first discuss the context of the drift compensation problem at hand, and then explain the mathematical model used in this approach. Next, I will describe the different regression models used and the baseline systems.

3.1 On-site calibration for drift compensation

3.1.1 Problem Statement

Chemical sensor arrays deployed for continuous monitoring tend to suffer from drift accumulation over the course of the application period (Sasaki et al., 2006). The problem is especially dire when the sensors are active for a very long period at a stretch. Traditionally, periodic sensor recalibration is performed to compensate for the accumulated drift from the sensors (Haugen et al., 2000). However, in most bioreactors, it is not viable to recalibrate sensors during an ongoing process. To illustrate the situation, Figure 1 shows a bioreactor setup where a single probe is submerged into the analyte mixture for monitoring. The probe containing the sensor array is attached to the bioreactor and has outgoing connections to the digital electronics circuitry for recording the sensor

measurements. In this setup, the sensor array needs to be completely taken out of the bioreactor so that a sensor recalibration can be performed. As a result, the continuous monitoring will have to be halted. An additional problem is the possibility of contamination in the bioreactor during the sensor replacement procedure. For these reasons, there is a need for an on-site sensor calibration technique which will not interrupt the ongoing bioreaction process.

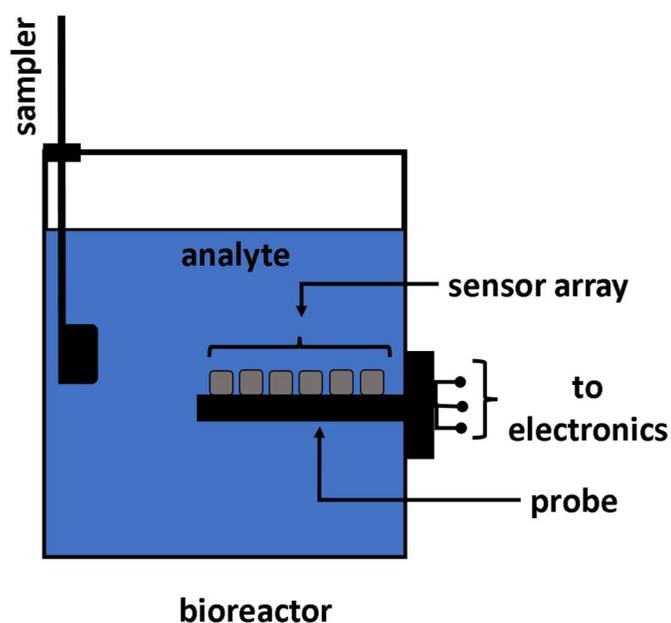


Figure 1 A bioreactor system with a single probe of sensor arrays for continuous monitoring of mixture analytes.

3.1.2 Proposed on-site calibration technique

The proposed method takes care of the calibration issue based on the following assumptions:

- I. Given an available past sample (the sensor measurements and the ground truth concentrations at that point in time), the current sample can be drift compensated

by modeling the target concentration as a function of the current sample and the past sample. We call this past sample as a calibration point. We will explain the mathematical modeling in the section below.

- II. The user can provide the concentrations at that calibration point by extracting a small sample from the bioreactor and obtaining the concentrations by running it through an offline analyzer. In Figure 1, we show such a setup with a sampler dipped inside the bioreactor. The assumption is that the user can extract a sample from the bioreactor without disrupting the ongoing reaction.

Based on the above assumptions, first I explain the mathematical model used in this work. Let $c(t)$ be the concentrations of the analytes in the bioreactor at time t and $x(t)$ be the corresponding sensor response. The relationship between the concentrations and the sensor response can be written as:

$$x(t) = A * c(t). \quad (1)$$

Here, A is a sensitivity matrix that maps concentrations into sensor responses. In this equation, x is a vector of dimension $n \times 1$, c is a vector of dimension $m \times 1$, A is a matrix of dimension $n \times m$, n is the number of sensors in the sensor array, and m is the number of analytes/target variables in the bioreactor. Ideally, we can use Equation (1) to solve for the concentrations as a simple regression task. However, in reality, the sensor measurements have additional noise and drift components along with the actual sensor response. In an initial step, we assume a linear drift model:

$$s(t) = x(t) + \delta(t). \quad (2)$$

Here $s(t)$ is the sensor measurement at time t and $\delta(t)$ is the time dependent drift. In this linear model, we further assume drift to be additive and linearly dependent on time:

$$\delta(t) = a * t. \quad (3)$$

Here a is a constant term. Now, we can replace $x(t)$ in (1) using the relationship in (2) and write,

$$s(t) - \delta(t) = A * c(t). \quad (4)$$

In this equation, the sensor drift $\delta(t)$ is unknown, so we cannot solve for the concentrations. To address this issue, we propose to use a previous calibration point, as mentioned in the assumptions earlier. Now, assuming we have a calibration point at time t' , the sensor measurement and the drift component for that sample can be written as,

$$s(t') - \delta(t') = A * c(t'). \quad (5)$$

Now, subtracting the above two equations we get the difference between the current sample and the calibration sample as,

$$[s(t) - s(t')] - [\delta(t) - \delta(t')] = A * [c(t) - c(t')]. \quad (6)$$

We can replace $\delta(t)$ in this relationship using Equation (3) and rearrange the equation to arrive at:

$$c(t) = A^+ [[s(t) - s(t')] - a * (t - t') + A * c(t')] \quad (7)$$

Equation (7) indicates that the concentration of the analytes at the current time t can be modeled as a function of the sensor measurements at current and a past time as well as concentrations from that past time. But most importantly, the drift components in the equation are now replaced with $(t - t')$, a measurement of elapsed time from a past

calibration point. This provides us with a way to estimate target concentrations using parameters already known. Now, for a more general case, we can rewrite the above relationship as below:

$$c(t) = \psi(s(t) - s(t'), (t - t'), c(t')). \quad (8)$$

Here, $\psi()$ represents a generalized relationship between the target and the proposed inputs. An advantageous feature of this approach is that the generalized function can be approximated using different regressor models depending on the complexity of the problem at hand.

The second part of the proposed approach is an ensemble of predictions utilizing multiple calibration points, if available, to improve the performance of the final prediction. In the subsection above, I have shown that an available past calibration point can be used to model the drift behavior of the sensor measurements and thus might help compensate for drift while predicting the target concentrations. I extend this idea further considering if the user provides multiple past calibration points over the course of the experiment. The idea here is to include multiple available calibration points to make the predictions more robust. To this end, I propose an ensemble technique that takes in predictions generated using multiple calibration points and produces the final prediction as an average, as shown below.

$$c(t) = \frac{1}{N_c} \sum_{i=1}^{N_c} \hat{c}_i(t) \quad (9)$$

Here, $c(t)$ is the predicted concentration using the i^{th} calibration sample and N_c is the number of calibration points available. The advantage of this approach is that,

assuming that the individual predictions $\hat{c}_i(t)$ are i.i.d., the variance of the ensemble estimate $\hat{c}(t)$ is reduced by a factor of N_c . In what follows, I refer to the proposed model as MPC, to indicate that it exploits Multiple Pseudo-Calibration samples.

3.2 Regression Models Used for Evaluating the Proposed Approach

In this section, I describe the different regression techniques used to validate the performance of the proposed approach as shown in (8). The generic function $\psi()$ can be implemented using any regression model. In this work, I have considered three regression techniques, namely, partial least squares, extreme gradient boosting, and multilayer perceptron. Below each of these models are explained further.

3.2.1 Partial Least Squares (PLS)

The first regression technique I consider is partial least square (PLS). PLS finds a linear regression model between the target and the inputs by projecting both variables to a latent space such that the covariance between the variables can be explained the most. In this study, I have implemented the classical PLS variant with the NIPALS algorithm (Wold, 1975).

3.2.2 Extreme Gradient Boosting (XGBoost)

The second method considered is the state-of-the-art regression technique known as Extreme Gradient Boosting (Chen & Guestrin, 2016). XGBoost utilizes an ensemble of decision trees using the gradient boosting technique and a regularized model formulation to reduce over-fitting. The method iteratively includes decision trees to reduce the residual error from the previous iteration of a simpler model i.e. a model with a smaller number of trees. In this study, I have used the XGBoost library implemented in Python. In order to

optimize the model parameters, I have implemented the Grid Search algorithm on the parameters mentioned in Table 1. I have performed this optimization process each time a new model is trained. I have chosen the final set of hyper-parameters for testing based on the performance on a validation set, obtained by randomly splitting the training dataset using a 75%-25% split.

Table 1 Model Parameters Considered for XGBoost

Parameter	Values Considered
Number of estimators	10, 25, 50, 75
Maximum Depth	1, 3, 5, 7
Learning Rate	0.001, 0.01, 0.1
L ₂ Regularization Parameter	0.1, 1, 2

3.2.3 *Multi-Layer Perceptron (MLP)*

The third and final model considered is a multilayer perceptron. The MLP is a fully connected feedforward neural network model with three hidden layers. The number of neurons used in each layer are 40, 20, and 10 respectively. The network hyperparameters were optimized separately. Namely, networks with 2, 3, and 4 layers with number of neurons varied in the range of 10 to 60 neurons per layer were evaluated. The final architecture was selected based on the model with lowest NRMSE. I have used rectified linear units (ReLU) as the activation function in each of the hidden layers. This activation function is linear for input values greater than or equal to zero but returns zero when the

input is negative, thus providing a non-linear activation at the output of a layer, as shown in (10).

$$y = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

The output layer activation function is linear as I am using the model for a regression problem. I used Huber loss (Huber, 1992) as the loss function to train this model, which is defined by,

$$L_h = \begin{cases} \frac{1}{2}(c_{gt} - \hat{c})^2, & \text{if } |c_{gt} - \hat{c}| \leq \delta \\ \delta \left(|c_{gt} - \hat{c}| - \frac{1}{2}\delta \right), & \text{otherwise} \end{cases} \quad (11)$$

Here, c_{gt} and \hat{c} are the ground truth and predicted concentrations respectively and δ is a small value denoting the threshold for the function. Finally, the MLP model is optimized using the ADAM optimizer (Kingma & Ba, 2014) with a batch of size 16. I train this model for a maximum iteration number of 1000 with an early stopping criterion set to be the minimum validation loss on a validation set.

3.3 Evaluation Strategy

In this thesis, I evaluate the approach in two steps, which will be explained in the following subsections. For the evaluation steps, I have used both experimental and simulated datasets, details of which will be explained in the following chapter.

3.3.1 Proposed approach vs baseline approach

To evaluate if the proposed approach (MPC) can successfully compensate for drift, I have implemented a baseline approach using the same regression techniques as the

proposed approach. The baseline approach considers only the sensor measurements at time t as input and tries to predict the concentrations of the analytes, as shown in (12).

$$c(t) = \sigma(s(t)) \quad (12)$$

Here, $\sigma()$ is a generalized function which we implement using the regression techniques listed above.

3.3.2 Comparison against a state-of-the-art technique

I have considered a domain adaptation based transfer learning technique proposed in (Yan & Zhang, 2016b) as a comparison method. This method is based on transfer learning, and uses a drift correction autoencoder (DCAE) to compensate for sensor drift and instrumental variations. The underlying idea is to map the target domain (drifted data) into the source domain (drift-free data) so that a model trained on the source domain can also be used on the target domain. In this method, the original features from the sensors and as well as a vector of domain features are fed into an autoencoder (AE) and the hidden representation in the last layer of the AE is treated as drift-compensated data for the final predictor model. Domain features considered in this method are device identity and sample acquisition time. However, the domain features are not fed to the AE directly. Rather, an additional feedforward network, called the correction layer, is used on the domain features to enhance the DCAE's ability to correct complex time-dependent drift. To perform transfer learning, transfer sample pairs i.e., samples that are similar in characteristics, from both domains are used to train the model and learn a mapping between the domains. An overview of the DCAE architecture is shown in Figure 2.

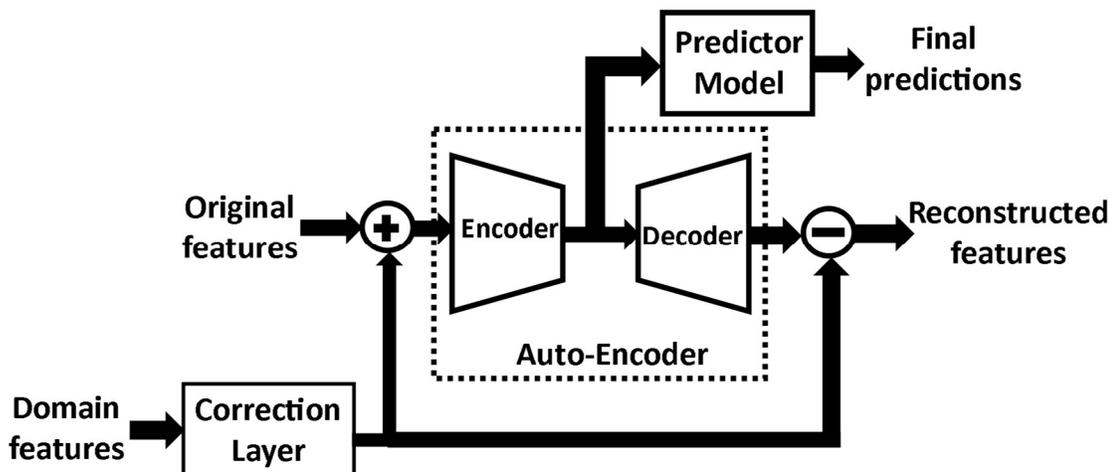


Figure 2 Block diagram of the DCAE model (Yan & Zhang, 2016b)

In this study, I have implemented the DCAE as reported by the authors with three hidden layers with the number of neurons being 30, 20, and 30 respectively. I have also included the correction layer for drift correction having a single layer of 12 neurons. Finally, for the prediction of the target concentrations, I have used the linear ridge regression model. This method requires transfer samples for domain adaptation and as such I have used three transfer samples from the test data. The source domain transfer samples are selected from the training data by matching the concentration profiles of the target transfer samples and the source domain samples. For the target domain transfer samples, first, I randomly select three samples from the initial 10 samples of the test data. For each of the samples, I calculate the L2-norm of the normalized concentrations of the target analytes. I then search the training space for samples that has similar L2-norm as the target transfer samples.

CHAPTER IV

EXPERIMENTAL SETUP

This chapter presents the details of the experimental setups used to collect the data used in this study. The experiments are conducted twice with a different set of sensors each time having different chemical compositions and thus different sensitivity characteristics. I have also created a simulated dataset to mimic the setup of the actual experiments in order to further analyze the proposed technique in different scenarios. The details of the simulated dataset are also described in this chapter.

In the following subsections, I first discuss the details of the experimental setup used and then explain the specific details of each of the experimental datasets. Finally, I provide the details of the simulated dataset.

4.1 Setups for the Experimental Datasets

In this study, I have used data collected by a hydrogel-based magneto-resistive sensor array system developed by Applied Biosensors, Inc (Nguyen et al., 2018). The system collects real-time data using a sensor array consisting of 5 different polyacrylamide-based hydrogel formulations. These sensors are designed to specifically respond to glucose and lactate concentrations in the solution as well as the pH and osmolality levels of the solution. Each of these hydrogels are integrated with a magnet, as shown in Figure 3(b). In addition, a reference sensor (bare magnet) is also included to monitor environmental noise and temperature variations. The reusable reader, as shown

in Figure 3, consists of 6 magnetometers each of which aligns with one of the individual magnets sitting on top of the hydrogels. The reusable reader also consists of 6 thermistors that are placed next to magnetometers. These thermistors keep track of the changes in temperature levels throughout the experiment. The reader is capable of continuous recording (every second) from all the magnetometers and thermistors. In these experiments, 4 such readers or probes are deployed inside the test vessel for simultaneous measurements. Independent ground truth data for osmolality, glucose, lactate, pH and temperature is also collected from offline analyzers.

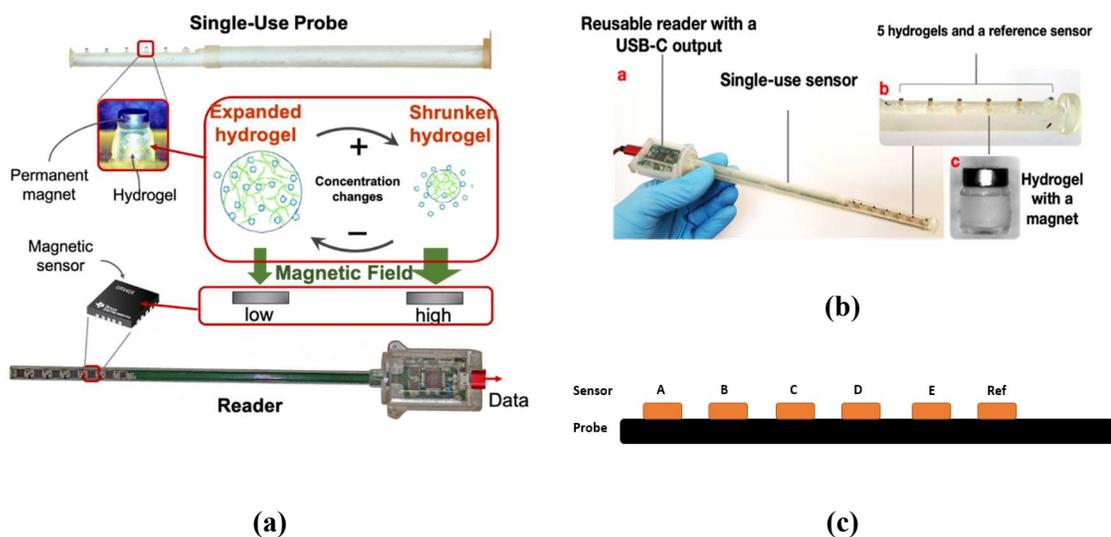


Figure 3 Experimental Setup.

The concentrations used in the experiment for each of the analytes and the corresponding hydrogel sensor module are included in Table 2. The sensor modules are named sensor A to sensor E, as shown in Figure 3(c), and I will use this notation throughout this thesis.

Table 2 Target Analytes and Sensors

Analytes	Concentration Range	Target Sensor
Glucose	0 – 50 mM	Sensor A, sensor C
Lactate	0 – 50 mM	Sensor B, sensor C
Osmolality	200 – 500 mOsm/Kg	Sensor D
pH	6.3 – 7.7	Sensor E

For the experiments, sensor data was collected for approximately 280 hours with a sampling rate of 1 Hz. During this time, the concentrations of the analytes was changed every 1-3 hours; thus, creating around 160 observations in total. Figure 4 includes sample data from one of the probes used in the first experiment. Here, the individual observations can be seen as different levels. The plot also shows the presence of noise in the measurements, specially in sensor E1. For the final datasets, I have removed some observations due to the lack of reliable ground-truth concentrations, so the final datasets had a total of 81 observations in the first experiment and 72 observations in the second. After data collection, as a preprocessing step, I applied a median filter to smooth out the data and remove noise. This filtered data is then used to calculate the average sensor response during the steady-state portion of an observation by taking the mean of the measurements over that period.

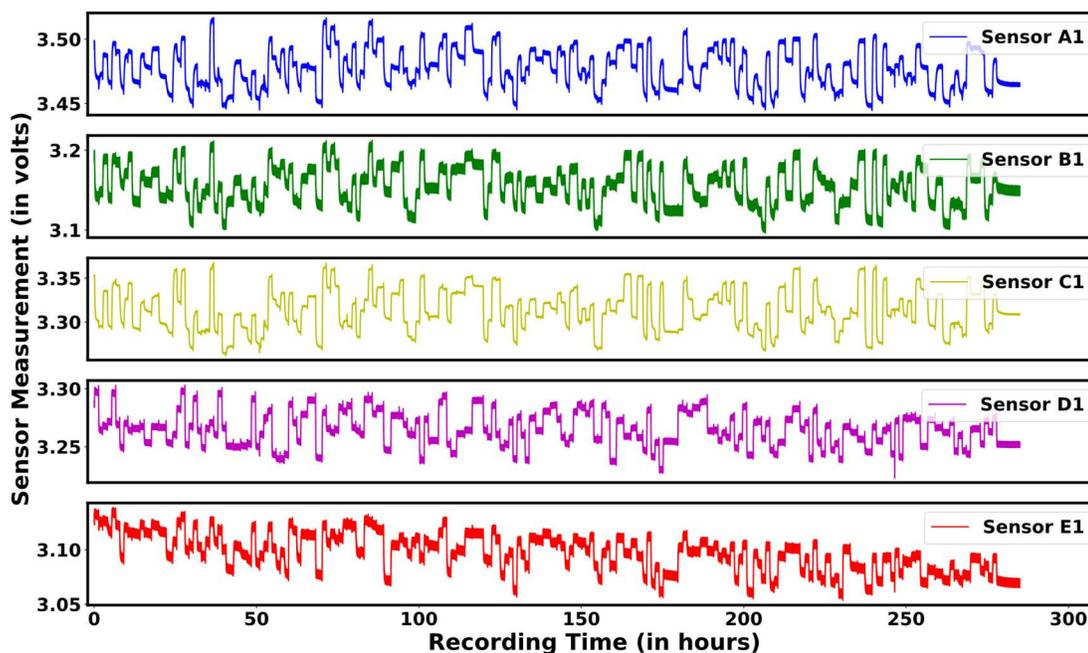


Figure 4 Raw Sensor Data Collected in the First Experiment.

Using this experimental setup, two separate datasets were collected. In each of the datasets, everything except for the hydrogel formulations were changed. In the following subsection, I discuss the details of each of the datasets in terms of drift and cross-sensitivity characteristics.

4.1.1 First Experimental Dataset

The hydrogel sensors developed for this experiment have cross-sensitivity to non-target analytes and process variables, that is, the hydrogel formulation responds to another analyte or process variable to some degree. In the first experimental dataset, the hydrogel sensors used have strong cross-sensitivity, as shown in Figure 5. In this figure, the correlation coefficients between each sensor and each analyte is shown as bar plots. Referring to Table 2, it is seen that the sensors do not always follow the expected

sensitivity profile. For example, sensor A1 has similar correlation to Glucose and Lactate and very high correlation to pH. Similar behavior can be seen for sensor C1. Sensor B1 has high correlation to Lactate as expected, but it has high correlation to pH as well. Similar behavior can be seen for sensor E1, where the sensor is highly reactive to a second analyte. Sensor D1, however, has very low cross-sensitivity with strong correlation to the target variable.

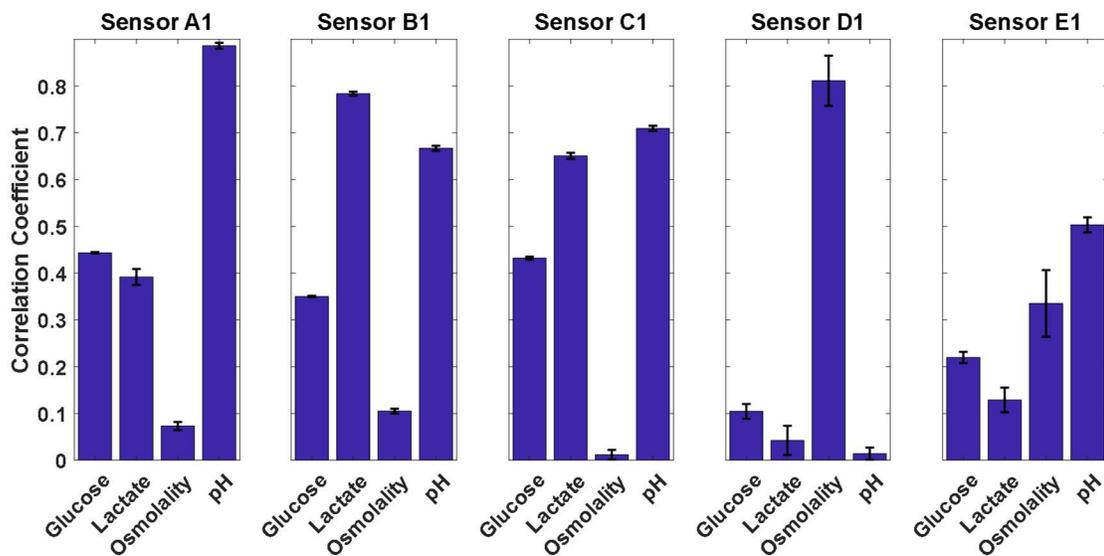


Figure 5 Cross-sensitivity profile of the sensor array used for the first experimental dataset. Correlation coefficients between the target variables and each of the sensors are used as a measure of sensitivity.

Next, I analyze the presence of drift in the dataset. In this experiment, the hydrogel and magnetic sensor combination shows the presence of time-dependent drift. To visualize the presence of drift in the sensor measurements, I have included a scatter plot of the osmolality sensor readings (sensor D1) vs the ground truth osmolality values in Figure 6. We have chosen sensor D1 because it is sensitive to only changes in the osmolality, as

shown in Figure 5, and thus provides a better understanding of the time-dependent behavior of the sensor.

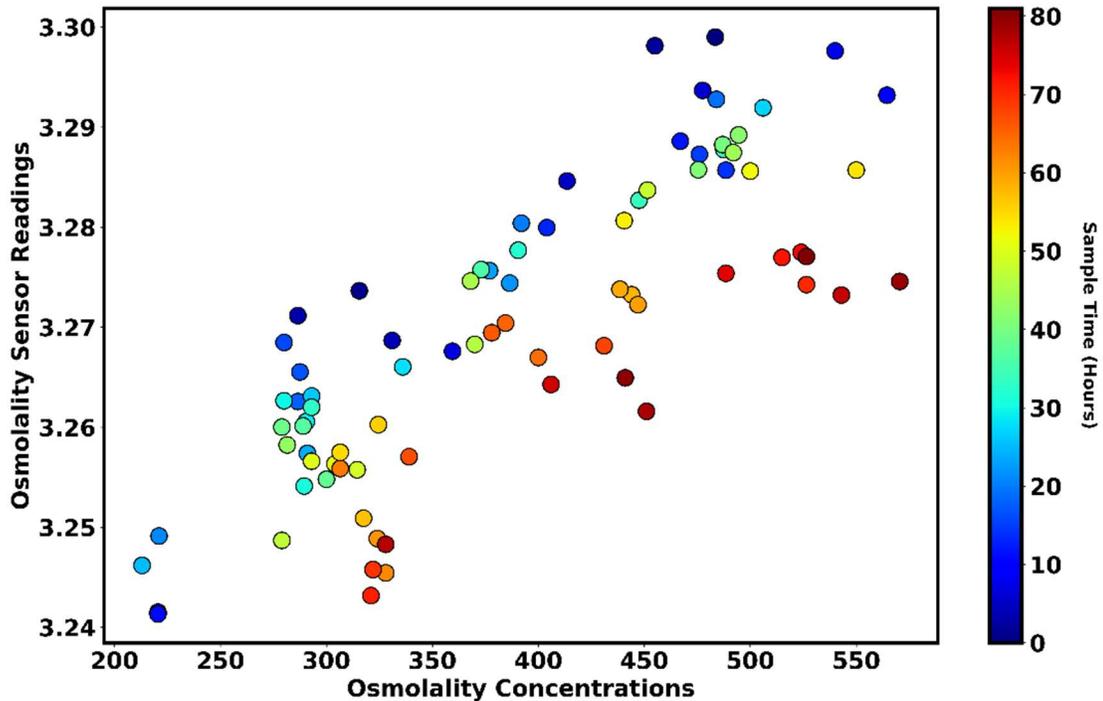


Figure 6 Sensor reading vs Ground truth concentrations showing presence of drift in sensor measurements collected for the first dataset. The color bar to the right shows the recording time in hours.

In Figure 6, each of the samples is color coded to indicate the recording time of that particular sample. As can be seen from the plot, for the same osmolality value the sensor values drift quite a bit as the sample recording goes on. For example, when osmolality values are around 500, the sensor readings change from around 3.30 volts to about 3.27 volts between samples collected in the first 50 hours to the samples collected after 250 hours of the beginning of the experiment. Similar behavior can be observed throughout the whole experiment. For the same osmolality values, samples that are collected after 200 hours have a drifted sensor reading of around 0.3 volts from the

samples collected in the first 20 hours. This indicates that the sensor measurements have embedded drift component affecting the measurement quality over time.

4.1.2 Second Experimental Dataset

The second experimental dataset uses different hydrogel formulations for the sensor array, and in this case the sensors showed less cross-sensitivity, as shown in Figure 5. However, the correlations for each of the sensors are lower than that of the first experimental dataset. For example, sensor A2 has a correlation around 0.25 with glucose and around 0.1 for the other targets. Similarly, sensors D2 and E2 has higher correlations to their respective target variables and low correlations for all others but the correlation values are down to 0.4 and 0.49 respectively. This indicates that while the sensors are slightly less cross-sensitive, the relationship between the sensor measurements and the ground truth is more non-linear in nature.

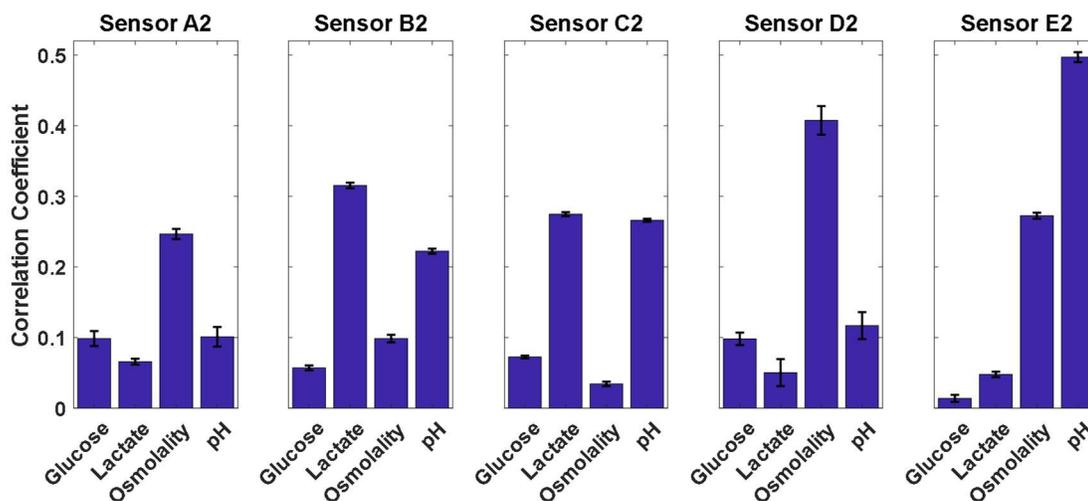


Figure 7 Cross-sensitivity profile of the sensor array used for the second experimental dataset. Correlation coefficients between the target variables and each of the sensors are used as a measure of sensitivity.

Finally, Figure 8 shows the presence of drift in this dataset. To maintain consistency, I have used sensor D2 and the osmolality ground truth values for this plot as well. The scatter plot shows clear presence of drift in this dataset as well. When the osmolality value is around 400, the samples collected in the 160-hour mark have sensor measurements below 3.2 volts. Whereas, for the same osmolality value, samples collected in the first 20 hours have sensor measurements around 3.22 volts, showing a difference of about 0.3 volts in the measurements.

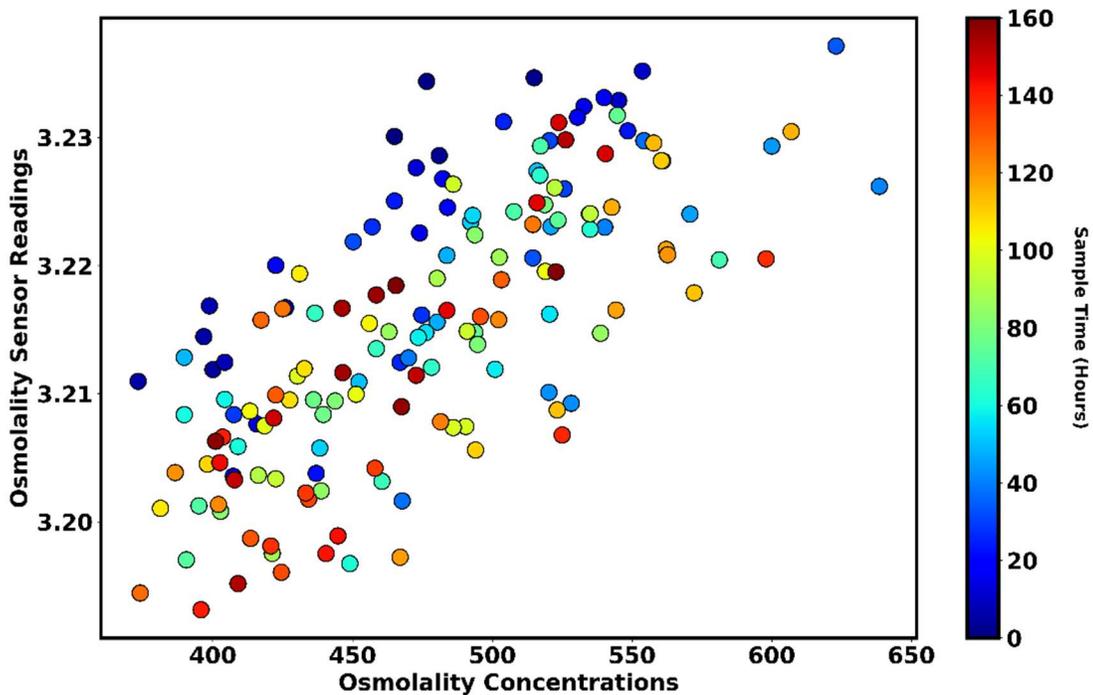


Figure 8 Sensor reading vs Ground truth concentrations showing presence of drift in sensor measurements collected for the second dataset. The color bar to the right shows the recording time in hours.

4.2 Simulated Dataset

To further analyze the performance of the proposed method, I generated a simulated dataset mimicking the properties of the experimental dataset. For this dataset, sensor data is generated by simulating the sensitivity of different sensors to different analytes using (13). Here, S is the sensor measurement, C is the concentration matrix, and A is the sensitivity matrix. The subscripts used in the equation indicate the dimensions of each of the matrices with N being the number of observations, p is the number of sensors in each device, and q is the number of analytes. In this experiment, I have considered $N=100$, $p=4$ and $q=4$. In order to include variations among the probes within the same dataset, random offset b is added to the sensor measurements. An additive drift term D is also added to the equation to simulate time-varying drift. All the matrices in the equation are taken from a uniform distribution.

$$S_{N \times p} = C_{N \times q} * A_{q \times p} + b_{N \times p} + D_{N \times p} \quad (13)$$

The sensor readings obtained using (13) contains simulated data for one experimental probe/device. In this experiment, I have simulated 4 such probes or devices with same set of parameters but different amount of drift as shown in (14). Here, $d_{N \times m}$ is the common drift term across the probes and the additive term is probe specific with η being random white noise and β is a scaling parameter which is varied across the probes. The common drift term d is simulated to be a linear time-dependent term with α being the slope of the drift term, as shown in (15). In this analysis, the amount of drift in the data is measured using the α parameter. To mimic the experimental conditions, I have measured

a similar parameter for the experimental datasets by fitting a linear model in the data and estimating the slope of the linear model.

$$D_{N \times m} = d_{N \times m} + \beta * \eta_{N \times m} \quad (14)$$

$$d = \alpha * t \quad (15)$$

To show the presence of drift in the sensors, Figure 9 includes a scatter plot between the concentration of analyte 1 and readings from sensor 1. In this plot, each sample is color coded to indicate the recording time of that sample. As can be seen, for the same concentration level, the sensor readings diverge quite a bit as time goes on. For example, at concentration level 0.8 the sensor readings are at 1.0 when the recording time is below 20 and increases to above 2.5 when recording time is above 80. This indicates the presence of significant amount of drift in the sensor readings.

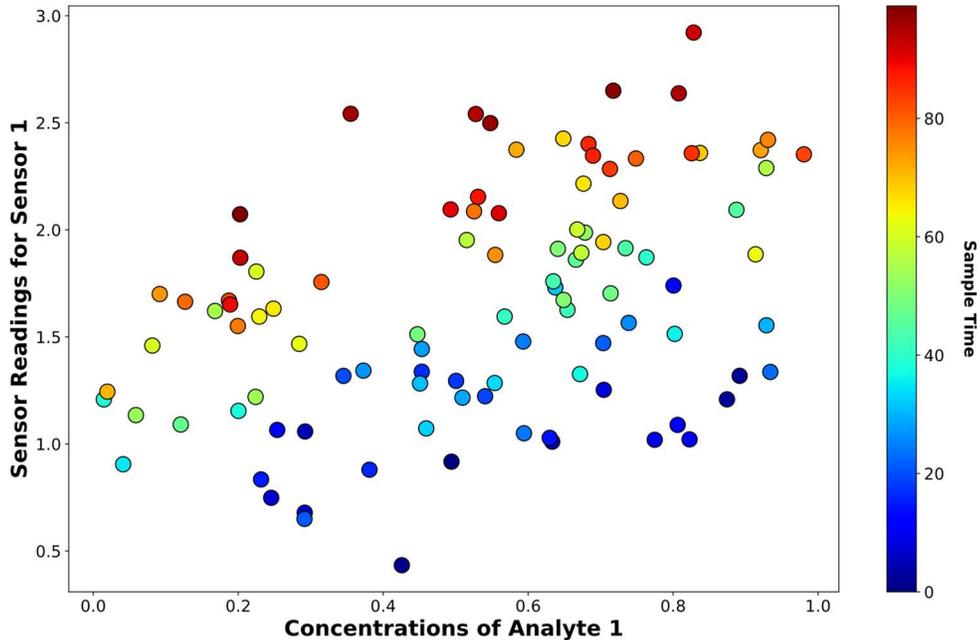


Figure 9 Sensor reading vs Ground truth concentrations showing presence of drift in the simulated sensor measurements. The color bar to the right shows the recording time in hours.

CHAPTER V

RESULTS

In this chapter, I provide the analysis of the results obtained in our experiments on the different datasets. For each of the datasets, I first analyze the performance of the proposed calibration technique against the baseline approach using the three regression techniques considered. Then, I compare the performance of the proposed approach against the DCAE method. To evaluate the performance of the various models, I have used a leave-one-probe-out cross-validation technique. Namely, for each dataset, I use 3 of the 4 probes for training and the remaining one for testing, and repeat the process 4 times. In order to evaluate the effect of drift compensation, I have further divided each of the probe's data into two segments with a 75%-25% split. During training, I use the first 75% of the data from the training probes and testing was performed on the last 25% of the data from the test probe.

In this analysis, I have used two different metrics for comparing the different techniques. First, I have considered normalized root mean squared error (NRMSE) as a performance metric as shown in equation (16). Here, \hat{c}_i and c_i are the predicted and ground truth concentration of the i^{th} sample respectively, and N is the total number of samples.

$$NRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{(\hat{c}_i - c_i)^2}{c_i^2}} \quad (16)$$

The second performance metric is the Pearson correlation coefficient between the predicted concentrations and the ground truth concentrations, calculated using equation (17). Here, μ_c and $\mu_{\hat{c}}$ are the average ground truth and average prediction respectively, and σ_c and $\sigma_{\hat{c}}$ are the standard deviations of the ground truth and prediction respectively.

$$r_{c,\hat{c}} = \frac{E[(c - \mu_c)(\hat{c} - \mu_{\hat{c}})]}{\sigma_c \sigma_{\hat{c}}} \quad (17)$$

5.1 Results on the First Experimental Data:

5.1.1 Proposed Approach vs Baseline Approach

First, I will compare the performance of our proposed approach against the baseline approach using each of the models. In what follows, I denote the proposed approach as the MPC approach. To evaluate the statistical significance of the results for each of the models, I have also performed pairwise two sample t-tests between the two approaches for each of the models. The p-values obtained in the significance tests are summarized in Table 3.

Figure 10 shows the average NRMSE for the two approaches (baseline and MPC) using PLS, XGBoost and MLP, and the 4 process variables: glucose, lactate, osmolality, and pH. As seen in the plot, MPC performs better than the baseline in most of the cases. For glucose, the baseline achieves average NRMSE of 49.8%, 49.3%, and 48.2% for PLS, XGB, and MLP, respectively. In contrast, MPC achieves average NRMSE of 46.1%, 44.8%, and 38.5%. Using two-sample t-test, we found significant difference in the all the

cases ($p < 0.01$). For lactate, the baseline achieves average NRMSE of 94.3%, 94.2%, and 59.2% for PLS, XGB, and MLP respectively. When MPC is used, the errors decrease to 55.6% for PLS, 74.4% for XGB, and 41.2% for MLP. Two sample t-test finds significant differences for all cases ($p < 0.01$). In the case of osmolality, the average NRMSE decreases significantly for PLS from 25.6% for the baseline to around 13.8% for MPC ($p < 0.001$). For XGB and MLP, the errors decrease significantly from 23.1% and 23.2% to 17.1% and 9.8% respectively (XGB: $p = 0.001$; MLP: $p < 0.001$). Similar results are observed for pH as well with significant improvement for all the three MPC models.

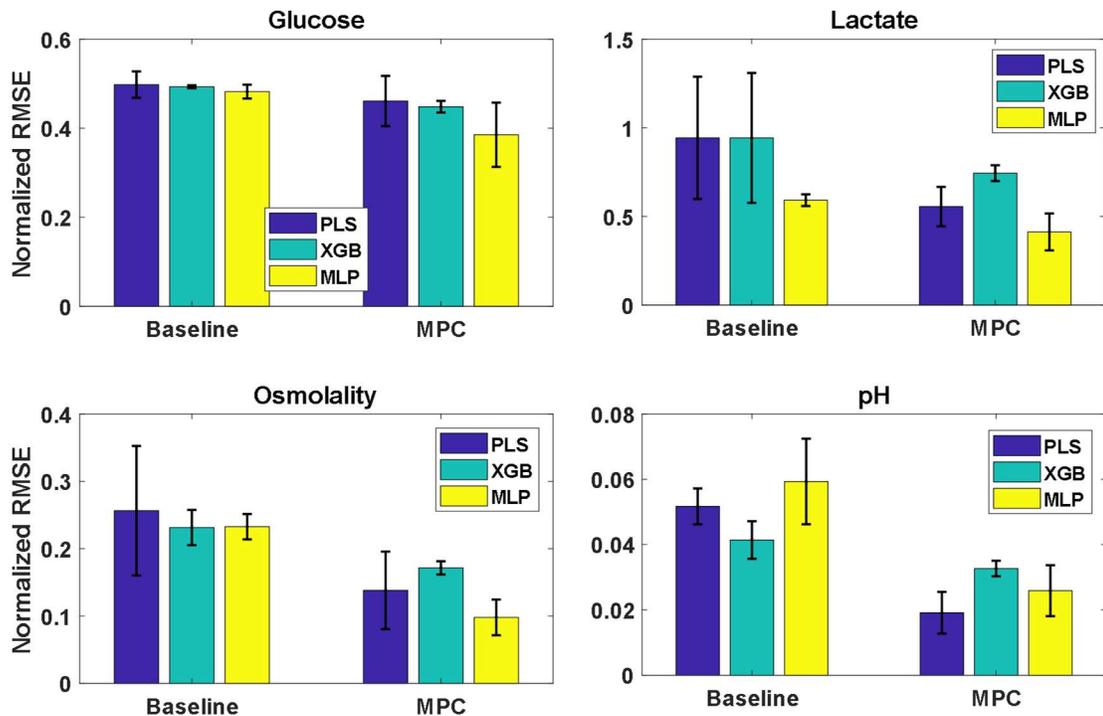


Figure 10 Performance comparison (using normalized RMSE) between the baseline approach and the proposed (MPC) approach on the first experimental dataset.

Similar results are observed in the case of pH as well with significant improvement for all the three models when the MPC approach is applied. The p-values for these cases are listed in Table 3 below.

Table 3 p-values for the pairwise 2 sample t-test between the baseline approach and the MPC approach

Models	Glucose	Lactate	Osmolality	pH
PLS	0.07	0.009**	< 0.001***	< 0.001***
XGB	0.03	0.06	0.001**	0.002**
MLP	0.07	0.005**	< 0.001***	< 0.001***

In the results shown above, it is observed that while the MPC approach improves the prediction errors in most cases, it is not significantly better than the baseline approach in a few cases. Next, we examine the correlation between predictions and ground truth for both approaches. Illustrative results are shown in Figure 11 for the baseline model and the MPC approach, in both cases using MLP as the underlying regression technique. Here, we used predictions on the test data and ground truths for all the probes. Each MLP model was ran 10 times, but the figure only shows one of the ten runs. These scatterplots indicate that, for the baseline model, there is little to no correlation between predictions and ground truth concentrations for any of the analytes except for osmolality with correlation is 0.35 with a p-value less than 0.05. In comparison, MPC has much higher correlation in all the four cases. The p-values from a significance test also indicate that the positive correlations are significant for MPC ($p < 0.001$) for all analytes except for glucose, where $p < 0.05$.

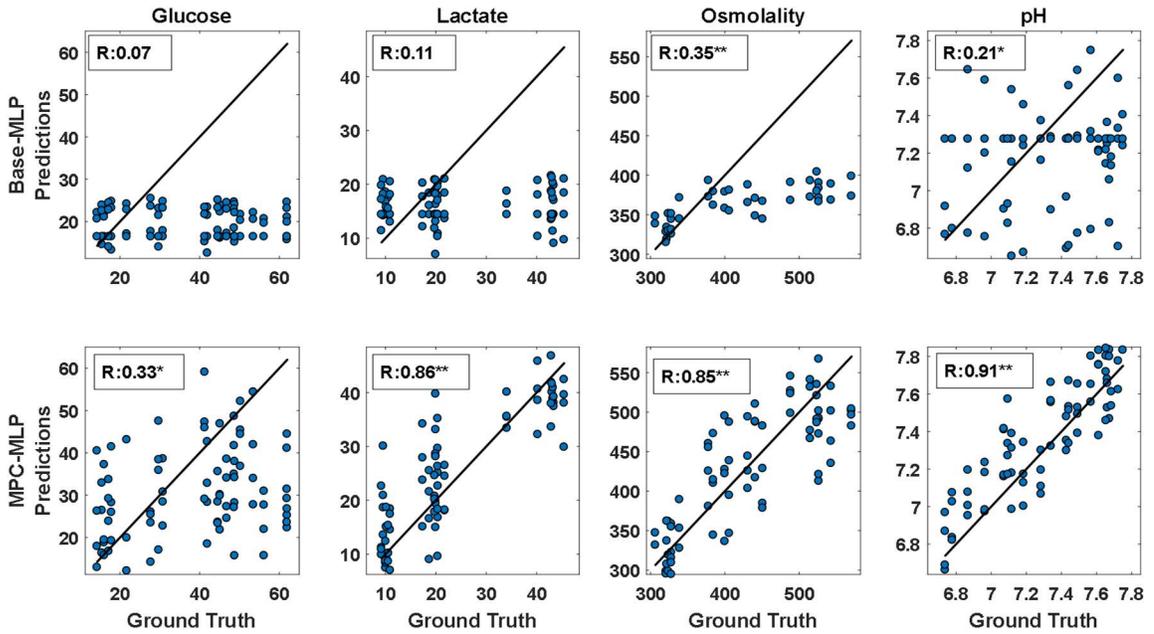


Figure 11 Correlation between prediction and ground truth concentrations for all analytes for the baseline model (top) and the MPC model (bottom) using MLPs as the underlying regression technique. (* $p < 0.05$; ** $p < 0.001$).

5.1.2 Proposed Approach vs DCAE

Next, I compare the performance of the MPC approach against the state-of-the-art drift correction approach, DCAE. As before, I first analyze the NRMSE for the two techniques and then compare the correlations obtained.

Figure 12 shows bar plots of the average and standard deviations of the NRMSE. In this plot, I have used all the three models using the MPC approach, denoted as MPC-PLS, MPC-XGB, and MPC-MLP. From the plots, it is obvious that the three models using the MPC approach performs better than DCAE. The p-values obtained from the pairwise t-test also indicate that the differences are significant in all the cases (p-value = 0.01 for glucose and p-value < 0.001 for others).

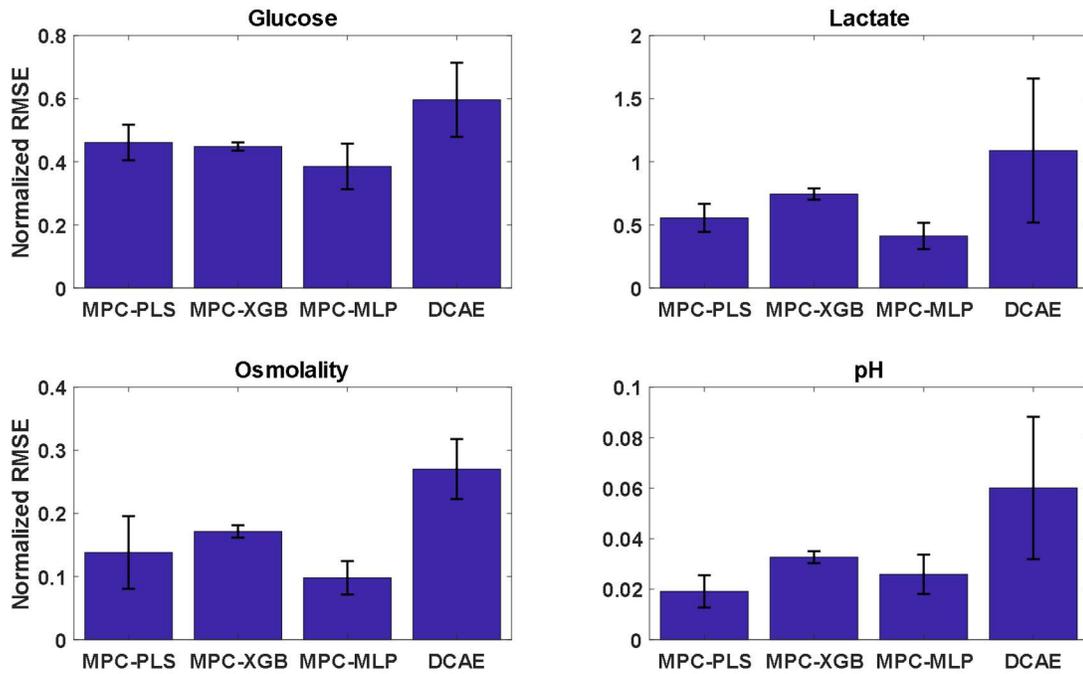


Figure 12 Performance comparison (using normalized RMSE) between proposed (MPC) approach and DCAE.

Figure 13 shows the scatter plots for all the target variables. For simplicity, only the MPC-MLP model is included in this plot. As can be seen in the plots, the correlations improve significantly between the DCAE and the MPC-MLP methods. In case of glucose, the correlation similar with slight improvement from 31% to 33% with p-value less than 0.05 for both cases. For the other target variables, it improves from 79% to 86% for lactate, from 49% to 85% for osmolality, and 31% to 91% for pH. These results indicate that the MPC approach performs better than a state-of-the-art drift compensation technique even when a simple regression model, for example PLS, is used.

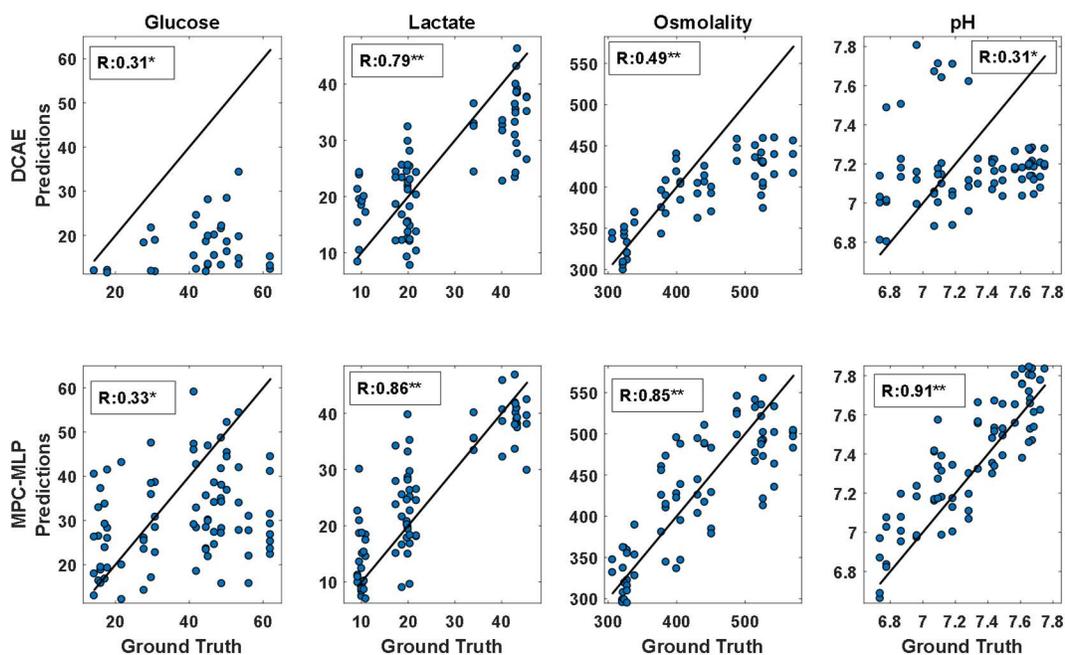


Figure 13 Correlation between prediction and ground truth concentrations for all the analytes; tested on the first experimental dataset.

5.2 Results on the Second Experimental Data

5.2.1 Proposed approach vs baseline approach

For the second experimental dataset, I have performed similar analysis as before and the following figures show the obtained results. First, Figure 14 shows the bar plots of the errors for all the models using the baseline and the MPC approaches. In this figure, the errors improve for the MPC approach in most cases. For glucose, there is a slight improvement in case of PLS and XGB, but for MLP the improvement is much more pronounced. The average NRMSEs for the baseline approach are 45.02 for PLS, 42.44 for XGB, and 46.85 for MLP respectively. Whereas for the MPC approach, the average NRMSEs are 42.41 for PLS, 41.25 for XGB, and 36.69 for MLP respectively. The p-values from a pairwise t-test indicate that the improvement is significant in the case of PLS and

MLP (p-value = 0.04 and 0.03 respectively) but for XGB there is no significant improvement (p-value = 0.08).

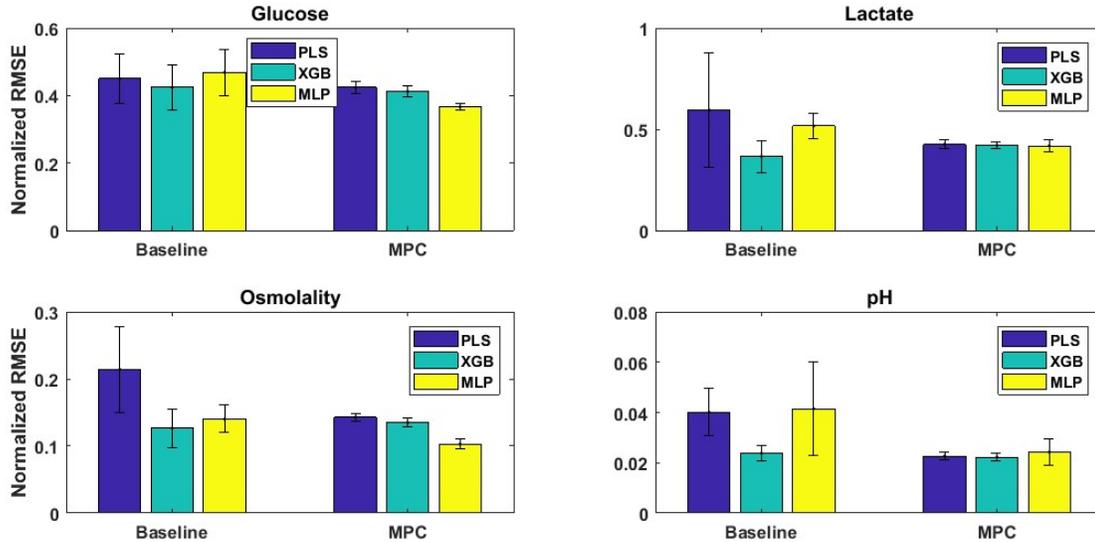


Figure 14 Performance comparison (using normalized RMSE) between the baseline approach and the proposed (MPC) approach.

For lactate, the MPC approach seems to perform much better than the baseline approach, with the average NRMSE decreasing from more than 59.51% to about 42.6% for PLS, 36.6% to 42% for XGB, and 51.6% to 41.8% for MLP. The p-values indicate marginally significant improvement at the 5% confidence level for PLS (p-value = 0.05 and 0.05 respectively). For MLP, there is significant improvement with a p-value of 0.03.

In the case of osmolality, the average NRMSE decreases from 14.03% to 10.2% for MLP (p-value < 0.001). For XGB, the errors remain the similar 12% to about 13% and for PLS the errors decrease from about 21% to about 14% with a marginally significant improvement (p-value = 0.04). Similar results can be observed for pH as well. In this case, MLP and PLS show significant improvement from around 2.5% of NRMSE to around

1.5% with p-values of 0.001 and 0.003 respectively for MLP and PLS. However, for XGB the improvement is not significant with a p-value of 0.06.

To further analyze these results, the correlation coefficients are calculated for each of the cases. In Figure 15, the scatter plot of the prediction vs the ground truth is shown for all the three models. From the plots, it is seen that the correlations do not improve much when the MPC approach is applied. This indicates that the baseline approach performs as well as the MPC approach indicating that the data may not contain as much drift to affect the performance of the baseline approach. The similarities in the results also indicate that the regression task for this dataset may not be as complicated as the previous dataset given that the sensors have less cross-selectivity, as shown in the previous chapter. In the next subsection we discuss more on the effects of the problem complexity on the performance of different models using a simulated dataset.

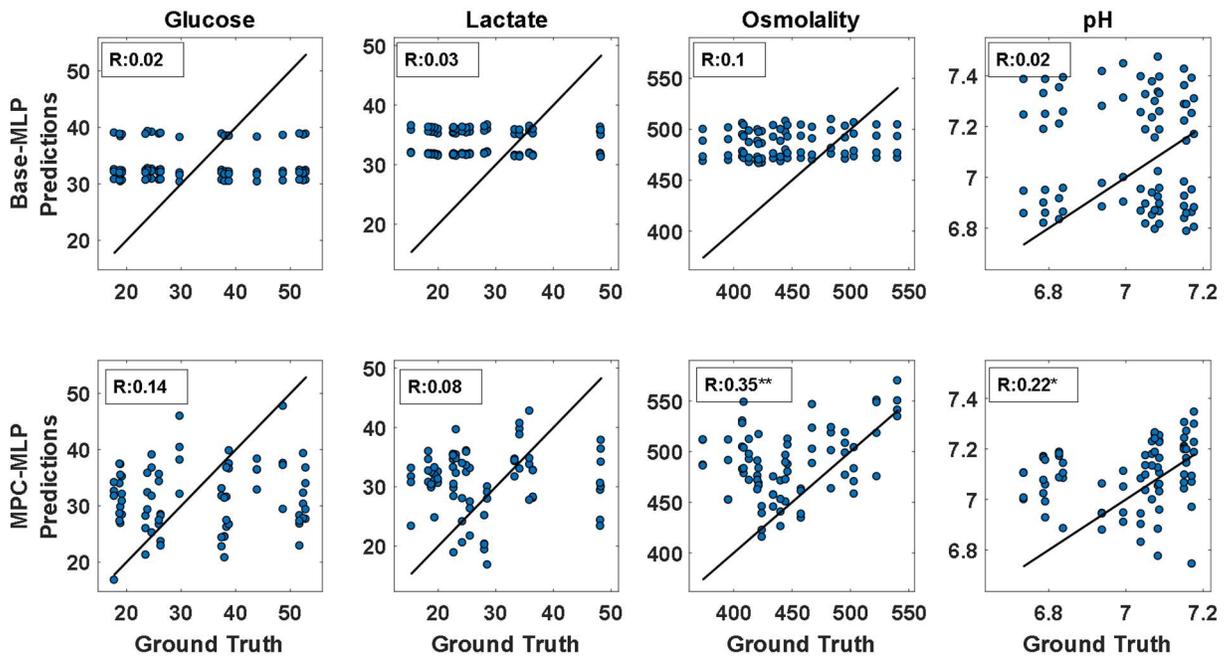


Figure 15 Correlation between prediction and ground truth concentrations for all the analytes; tested on the second experimental dataset.

5.2.2 Proposed Approach vs DCAE

In this subsection, I compare the performance of the MPC approach against the DCAE method on the second dataset. Figure 16 shows the bar plots of the NRMSE obtained using the three MPC models and the DCAE. From the plot, it is seen that the errors obtained by the MPC-MLP model is lower than that for the DCAE in case of glucose (36.6% for MPC-MLP and 41.5% for DCAE). For the PLS and XGB models there is little to no improvement with average NRMSE being 41.45% for PLS and 41.2% for XGB. The p-values from the t-test show that there is no significant difference in the case of PLS and XGB (p-value = 0.07 and 0.09 respectively). In the case of MLP the improvement is significant with a p-value of 0.03.

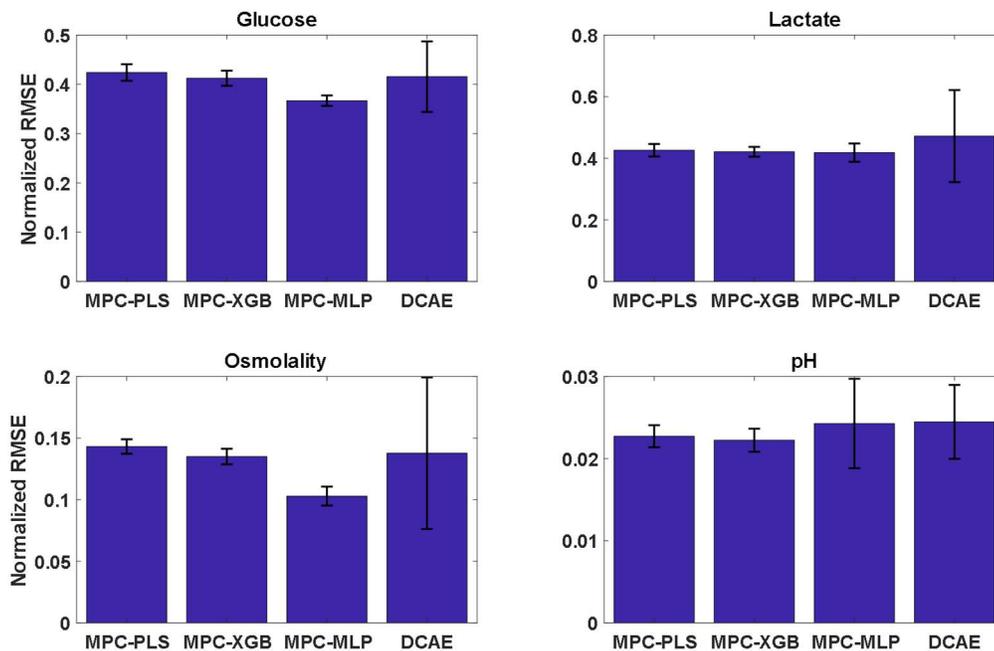


Figure 16 Performance comparison (using normalized RMSE) between proposed (MPC) approach and DCAE.

For lactate, the errors improve significantly for PLS, XGB, and MLP from about 47.2% for DCAE to about 41.8% for MLP and around 42.1% for XGB and 42.6% for PLS. However, the improvements are found not to be significant in any of the cases. In the case of osmolality, using the MLP significantly improves the errors from about 13.7% for DCAE to about 10.2% with a p-value 0.04. PLS and XGB, however, do not show any significant difference with average NRMSEs of 14.3% and 13.5% respectively (p-values of 0.08 and 0.07 respectively). In the case of pH, DCAE seems to perform better than PLS and XGB and almost similar to MLP. However, p-values from the t-test indicate that there are no significant differences between the two approaches (p-values = 0.6, 0.4, and 0.8 respectively for PLS, XGB, and MLP).

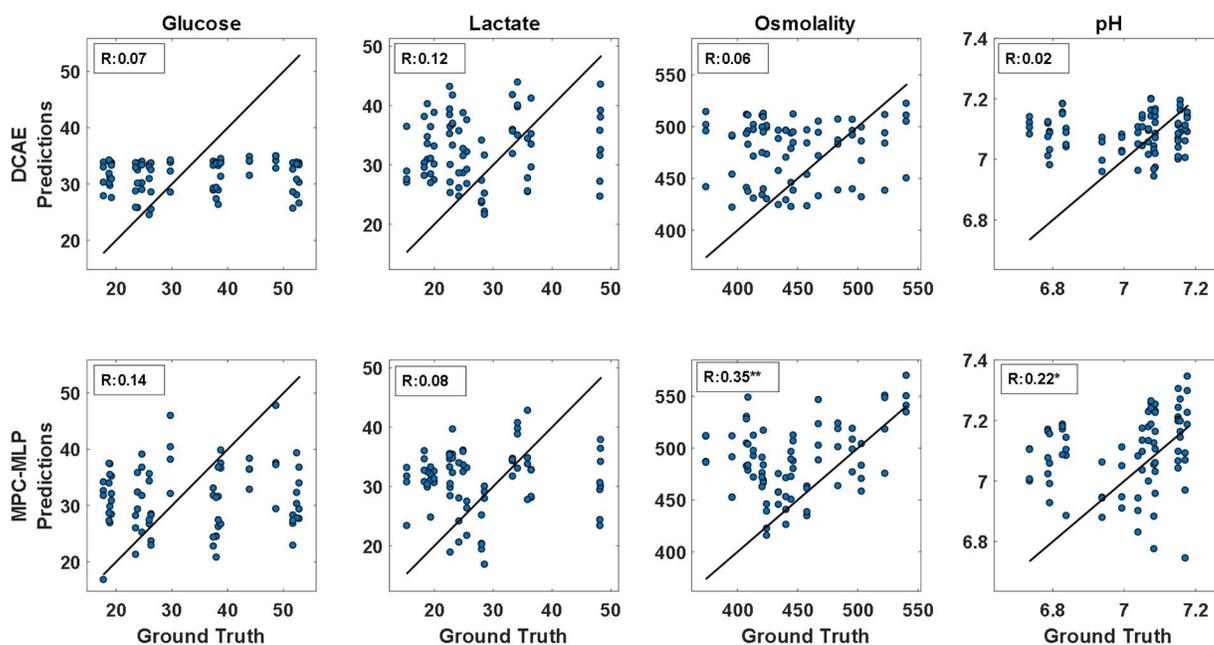


Figure 17 Correlation between prediction and ground truth concentrations for all the analytes.

Finally, I have analyzed the correlation between the prediction and the ground truth for the two techniques. Figure 17 shows the scatter plots for each of the target variables obtained using DCAE and the MPC-MLP model. The correlations improve for the MPC model in all the four target variables. However, except for osmolality and pH where the correlations are about 35% and 22% respectively, the correlations are still low and insignificant.

5.3 Results on the Simulated Data

In this section, I further analyze the performance of the proposed approach and the two comparison techniques using the simulated dataset explained in the previous chapter. For the first analysis, the robustness of the proposed approach is tested by varying the complexity of the problem, defined using the condition number of the sensitivity matrix. The condition number of a $n \times n$ matrix A is defined by,

$$\text{cond}(A) = \frac{\max(\lambda)}{\min(\lambda)}. \quad (18)$$

where $\max(\lambda)$ and $\min(\lambda)$ indicate the largest and the smallest eigen values of the matrix A respectively. In this analysis, we use the condition number of the sensitivity matrix of the simulated dataset to measure the complexity of the problem. The intuition behind is that if the condition number of a sensitivity matrix is low then the sensors have nearly ideal relationship with the target variables with little to no cross-sensitivity. On the other hand, a sensitivity matrix with high condition number indicates more cross-sensitivity and thus a more complicated task of separating the effects of the analytes from the sensor measurements. This intuition is shown in Figure 18. Here, the color black indicates high

correlation, and a blank space indicates zero correlation. The color gray indicates a non-zero correlation between the sensor and the target variable. In Figure 18(b), a matrix with some degree of cross-sensitivity is shown. This sensitivity matrix would have a high condition number whereas the matrix shown in Figure 18(a) has very low condition number as there is no cross-sensitivity in the sensors.

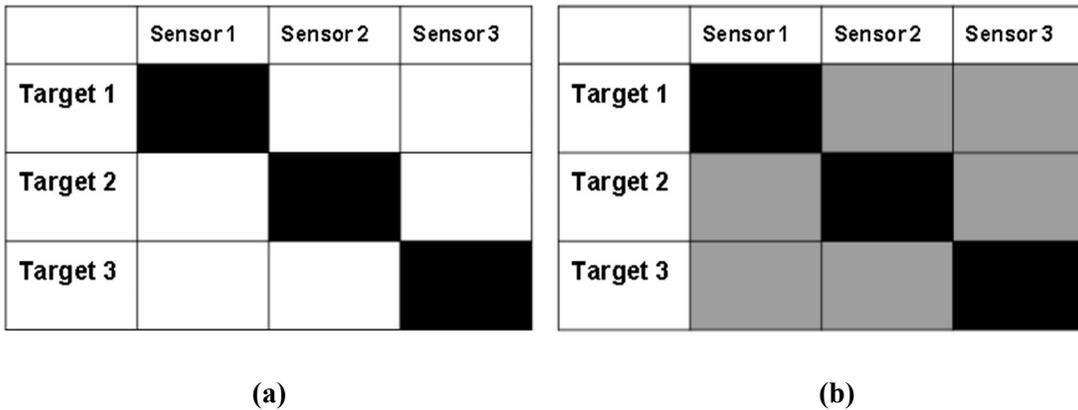


Figure 18 Visualizing the condition number for sensitivity matrices. (a) Sensitivity matrix with high cross-sensitivity and high condition number, (b) Sensitivity matrix with no cross sensitivity and very low condition number.

To simulate sensor arrays with different cross-selectivity profiles, we generated sensitivity matrices A with different condition numbers, as follows. Starting from a randomly generated matrix A , we perform singular value decomposition $A = U \Sigma V^T$. In a second step, we replace the diagonal elements in Σ (i.e., the eigenvalues) with a set of values linearly interpolated between the desired condition number \mathcal{N} (largest eigenvalue) and 1 (smallest eigenvalue). This yields a new eigenvalue matrix $\hat{\Sigma}$, which by construction has condition number \mathcal{N} . In a third step, we reverse the singular value decomposition to obtain sensitivity matrix $\hat{A} = U \hat{\Sigma} V^T$. In a last step, we compute the final sensitivity

matrix as the outer product $A^* = \hat{A} \times \hat{A}^T$. For $\mathcal{N} = 1$, this yields a matrix A^* whose columns (i.e., sensors) are orthonormal (i.e., each sensor responds to only one analyte), as shown in Figure 18(a). As \mathcal{N} increases, so does the degree of cross-sensitivity.

Following this procedure, we generated different datasets with condition numbers ranging from $\mathcal{N} = 1$ to 30. For each \mathcal{N} , we generated 10 different sensitivity matrices A^* , each starting from a random matrix A , as described above. Then, for each sensitivity matrix A^* , we generated 100 simulated measurements with different concentration of the target variables. The amount of drift in each of the datasets is kept constant at 0.5% of its dynamic range every hour. Figure 19 shows the NRMSE for the various models, averaged across the 10 simulations and the four simulated analytes. The DCAE curve in each of the plots is identical and is included to facilitate visual comparison. Here, as the condition number increases, the complexity of the problem increases and as such the prediction errors should increase. However, for a robust model the error should not be increasing as much with the increasing condition number.

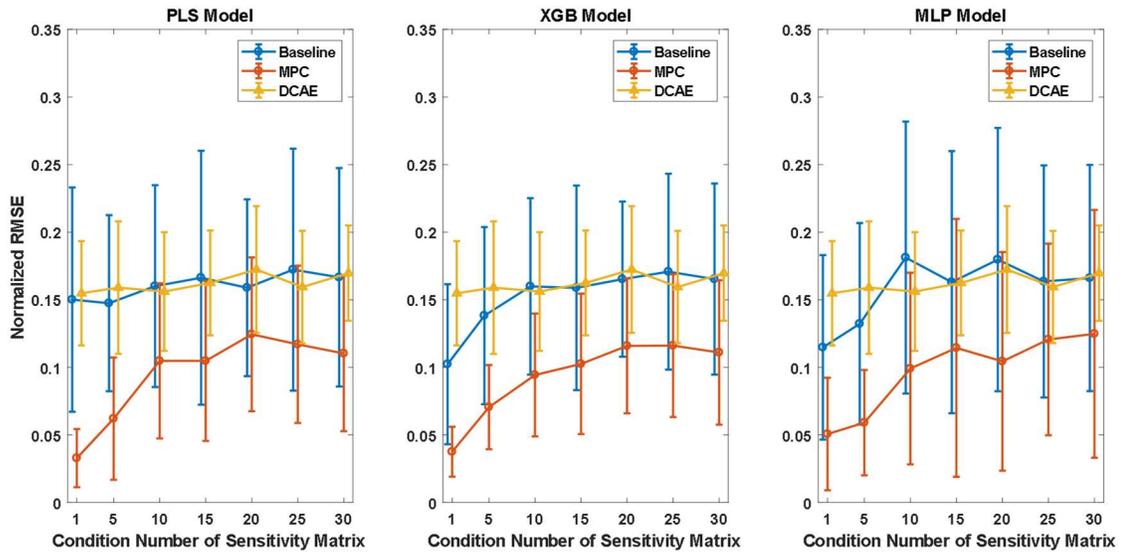


Figure 19 Normalized RMSE obtained at different complexity (different condition numbers) of the regression task.

In the case of PLS, the baseline approach achieves higher NRMSE compared to the MPC approach right from the beginning. At low condition numbers, the MPC-PLS model performs well with very low NRMSE of 3.2% and with increasing complexity the NRMSE increases steadily to about 11%. In comparison, the baseline PLS approach shows high NRMSE of 15.01% even at very low condition numbers. The DCAE approach shows similar error pattern as the baseline approach with high NRMSE across the board. As for the XGB and MLP models, both the baseline and MPC approaches show similar trend in error with the MPC approach performing significantly better than the other two approaches. For XGB, the MPC approach achieves NRMSE of 3.7% at condition number 1 and the error increases to 11.1% at the highest condition number. For the baseline approach, error increases from 10.2% to 16.9%. As for the MLP model, the NRMSE for

MPC approach ranges from 5% to 12.4% whereas in case of the baseline approach the NRMSE increases from 11.4% to 16.6%.

In a second analysis, I varied the amount of drift in the data to investigate how the proposed approach fares with increasing amounts of drift in the data. For this analysis, I changed the coefficient of the drift term in equation (15) as a percentage of the amplitude range of the signal without drift, as shown in (19).

$$\alpha = r (\max(x_i) - \min(x_i)). \quad (19)$$

Here, x_i denotes the i -th sensor in the array, and r represents the amount of drift as a percentage. For example, for $r = 1\%$, the sensor drifts by 1% of its dynamic range every hour. For this analysis, we used r in the range 0-2%, and sensitivity matrices with a condition number of 10. Figure 20 shows the normalized RMSE obtained for different datasets with increasing amounts of drift. For this plot, I used 10 different datasets with varying amounts of drift and the normalized RMSE is calculated by averaging the four target variables. The plot shows the NRMSE for all the models using the baseline and MPC approaches along with DCAE. As in Figure 19, the DCAE curve is replicated in each of the plots to facilitate comparison.

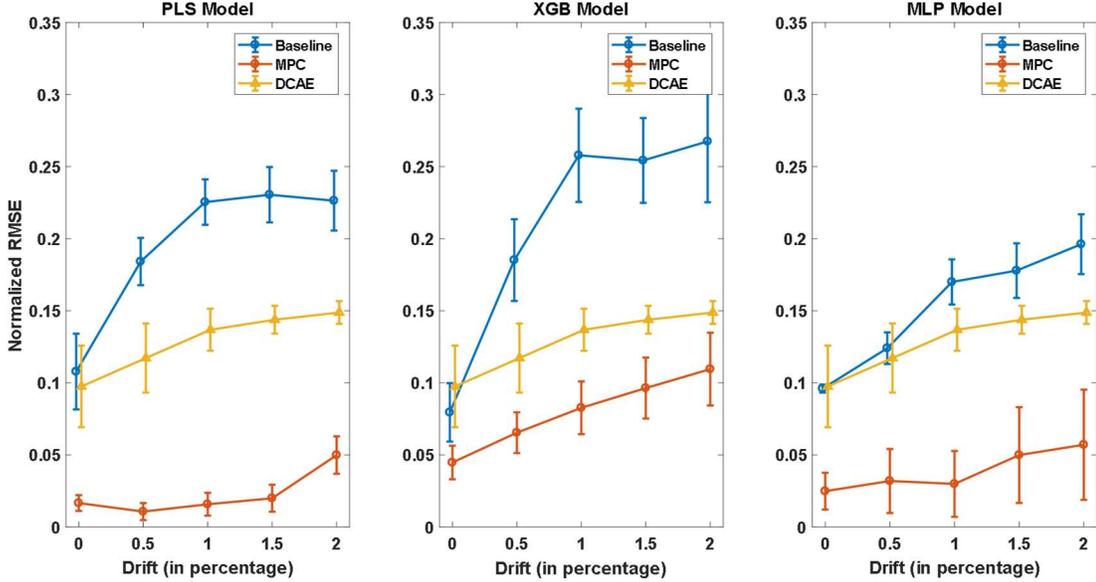


Figure 20 Normalized RMSE obtained at different amount of drift added to the data. The amount of drift is shown as a percentage of the range of amplitude of the original signal.

As can be seen from the plot, the NRMSE for MPC-PLS is 1.6% at 0% drift, remains fairly steady until 1.5% drift (NRMSE of 1.9%) and finally increases to 4.9% as the amount of drift is increased to 2%. In comparison, both the baseline PLS and the DCAE approaches show increasing trend in error as the drift increases. The baseline PLS achieves significantly higher NRMSE of 10.7% at the beginning with the error increasing to 22.6% at 2% drift. The DCAE shows slightly better performance than the baseline with NRMSE of 9.7% at the beginning and increasing up to 14.9 at highest amount of drift. In case of the XGB model, the MPC approach does show some increasing trend in error with NRMSE increasing from 4.4% to 10.5%. However, in comparison to the baseline approach, which achieves NRMSE ranging from 7.9% to 26.7%, MPC performs significantly better. The errors are lower than the DCAE model as well. Similar

performance is seen in the case of MLP as well. In this case, the errors for the MPC approach are steadier across increasing drift (increasing from 2.4% to 5.7%) compared to the baseline approach (increasing from 9.7% to 19.6%).

Finally, we analyzed how the number of training data samples affect the performance of the MPC approach. For this analysis, we used the three MPC models on 10 simulated datasets with condition number $\mathcal{N} = 5$ and the amount of drift set to be 0.5%. During training in the default MPC setting, for the t -th sample, we used all the previous samples from 0 to $t-1$ as the pseudo-calibration points. This way, the total number of training samples increase from n (the size of the training dataset) to $n*(n-1)/2$. In this analysis, we chose to use varying amount of training data ranging from 10% of the default setting to 100% of the default setting. Equation (20) shows how the number of training samples is set for each data sample. Here, TR_t is the number of training samples created for the t -th data sample, k is a constant that is varied between 0.1 and 1.0 with a step of 0.1. For each data sample, then we randomly chose TR_t past samples as pseudo-calibration points. To make sure that each data sample has at least one training sample included, we used the *ceil()* function.

$$TR_t = \text{ceil}(k * (t - 1)). \quad (20)$$

Figure 21 shows the normalized RMSE vs different amounts of training data used in training. Each point on the plot indicates the average NRMSE obtained across the 10 datasets and the error bars indicate the standard deviations. In addition, the NRMSEs obtained for the corresponding baseline models and the DCAE is also included in the plots.

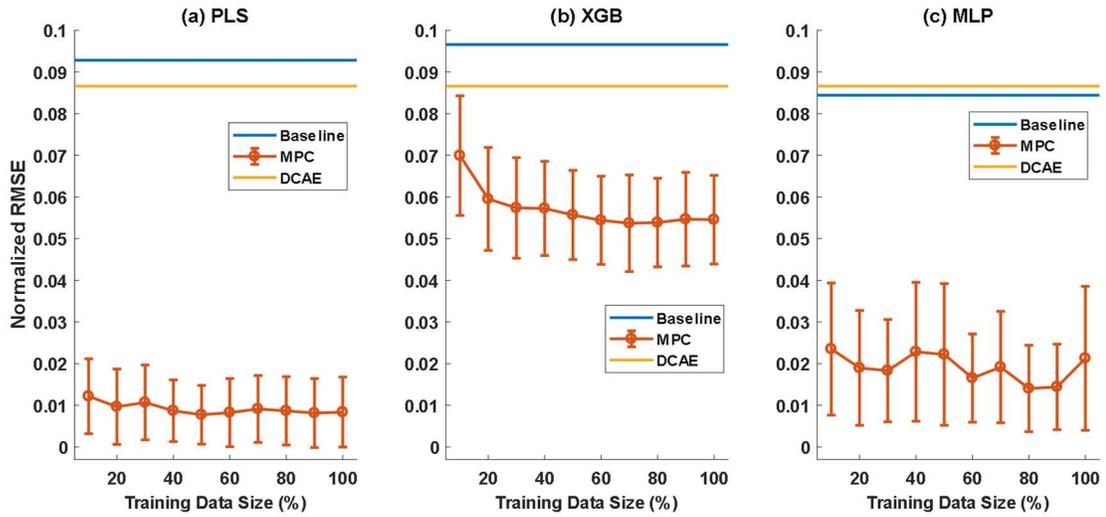


Figure 21 Normalized RMSE obtained at different amount of training data used. The size of the training data is represented as a percentage of the total number of data samples used during training the three MPC models in the default setting.

As can be seen from the plot, there is a slightly downward trend in the average NRMSE as the amount of training data is increased. When 10% of the training data is used the error obtained is 1.22% whereas for 100% of the training data being used, the error decreases to 0.84%. The NRMSEs for the baseline approach and the DCAE are 9.28% and 8.66% respectively. As for the XGB model, the downward trend is more significant with the NRMSE decreasing from 6.99% when 10% of the training data is used to 5.46% at 100%. Similar pattern can be seen for the MLP as well, with the NRMSE being 2.35% at 10% of the training data and decreasing to 1.4% at 90%. Comparing to the other two approaches, the difference in error for the MPC models remain fairly similar across different settings.

CHAPTER VI

DISCUSSION AND CONCLUSION

In this thesis, I have proposed an on-site calibration technique for chemical sensor arrays to eliminate time-dependent drift from the sensor measurements. The proposed technique applies a mathematical model to characterize the drift components in the sensor measurements with the help of user-provided past calibration points. This approach allows for an arbitrary number of calibration points to be used through an ensemble of predictions generated by the calibration points fed to the system. In our experiments, we have found that the proposed technique can be implemented on top of any regression techniques and yields robust performances in predicting target concentrations.

In this chapter, I will first summarize the key findings of this thesis. Finally, I will discuss some of the limitations of the proposed approach and the experiments conducted followed by possible future directions of this work.

6.1 Summary of Findings

In this thesis, a multi-calibration drift correction approach is proposed, and it is tested on two separate real-life experimental datasets as well as simulated datasets. The details of the experiments conducted, and the results obtained are described in detail in the previous chapters. From the experimental results, it is fairly clear that the proposed approach, in most cases, can handle drifted data better than the baseline approach and the state-of-the-art DCAE technique. Regardless, there are some key points in the results that need to be highlighted.

In the experiments, the task is to predict the concentrations of the different analytes/process variables from continuous measurements of an array of hydrogel sensors. The results across the two datasets indicate that the proposed approach outperform both the comparison methods in most cases. On the first dataset, the proposed approach does significantly better (lower NRMSE) than both the baseline approach and the DCAE in all four target variables. These results are true irrespective of the prediction models used. To further investigate these results, correlation between the predicted and the ground truth is used as a second metric. It can be seen in the results, that the proposed approach does improve the correlation in all cases and generates predictions that are significantly correlated to the ground truth values. However, comparing the performance across different targets, it is evident that both the NRMSE and the correlations are worse in the case of glucose and lactate. This outcome might explain the effect of cross-sensitivity in the sensor array as shown in the sensitivity plots in Figure 5.

In the case of the second dataset, while we can see that the proposed approach improves the errors on all the target variables, the difference between the proposed approach and the baseline approach is much closer, with non-significant differences in some cases. In case of DCAE, the proposed approach performs better when the MLP model is used except in the case of pH. The other two models fail to show significant improvement. In terms of correlation, the proposed approach improves the correlations in all the cases compared to the comparison techniques. However, the correlations are still fairly low even with the proposed approach. This result can be explained using the sensitivity profile of the second dataset, as the sensor measurements show less correlation

to the target concentrations. As a result, the performance does not only suffer from sensor cross-sensitivity but also the lack of linear relationship between the measurements and the targets.

As shown in Chapter IV, the two datasets were collected from similar experimental setups and the amount of drift in the datasets are similar. However, the sensors used in the experiments are different and so are the sensitivity matrices of the two datasets. This is also obvious from the correlation between the sensor readings and the ground truth concentrations values shown in Figure 5 and Figure 7. This implies that the task of predicting the analyte concentrations would be much harder in the second experimental data than in the first one. The results obtained on these datasets also show that the errors on the second dataset are higher than those on the first dataset, irrespective of the prediction model using either of the approaches. In terms of comparing the baseline approach to MPC, we see that the difference between the two approaches is more pronounced in the first dataset than the second one. Further improvement can be achieved on the second dataset by using a non-linear drift model to explain the highly non-linear drift relationship in the data.

These results are further corroborated through experiments on synthetic data that simulated different degrees of sensor cross-selectivity and drift. In the first experiment using simulated data, I synthesized the sensitivity matrix of each of the datasets to have increasing amount of cross-sensitivity to the target variables, thus increasing the complexity of the prediction problem among the datasets. The results obtained in this experiment, included in Figure 19, show that the MPC approach outperforms both

comparison methods, with much lower NRMSE across different condition numbers. The significant part of this result is that even when the simplest regression model is used (*i.e.* PLS) the MPC approach achieves significantly lower NRMSE compared to even a complex model like the DCAE. Although the MPC approach does show lower NRMSE, as the condition number is increased, the average NRMSE increases in all cases. This indicates that the MPC approach does suffer from the increasing complexity of the problem at hand. This result, however, reinforces the findings found in the experimental datasets. The DCAE model shows more stable performance but that might be because it performs poorly when the model complexity is lower.

In the second analysis with varying amount of drift in the simulated data, we see that the proposed approach remains robust with increasing drift compared to the baseline approach. The errors with the baseline approach increase markedly, as this approach does not employ any drift correction. As for the DCAE, the errors are comparatively more stable than those of baseline, a result which is to be expected. However, compared to the proposed approach, the errors in the DCAE are significantly higher. These results prove that the proposed approach does perform drift compensation effectively and perform significantly better than the state-of-the-art DCAE model. Specially, the performance of the PLS model, when the MPC approach is used, shows the efficacy of the approach. Using the pseudo calibration points in the proposed approach allows the model to learn drift patterns on the data and in turn it makes the regression task much simpler, even allowing the PLS to perform better than the more complex models at times.

Finally, I analyzed the effect of training sample size on the performance of the MPC approach using the MPC-PLS model, as shown in Figure 21. For the MPC-XGB model there is a significant improvement in error when the amount of training data is increased. For the other two models, the improvements in error with the increase in training volume is comparatively lower but there is a definite trend which shows that the models improve with the increase in training data being used. This indicates that the training scheme used in the proposed approach does contribute to the performance observed for the regression models used. However, the difference in performance between the MPC approach and the two comparison methods indicate that the MPC models outperform both of those even with the lowest training data volume used. This indicates that while there is improvement due to training data volume, the MPC approach's better performance owes mostly to the pseudo-calibration based drift correction technique used.

6.2 Limitations and Future Work

In this thesis, I have proposed a drift correction technique which shows promising results in terms of handling drift in chemical sensor arrays deployed for long term monitoring. The proposed approach employs a mathematical model to characterize the drift component in the data. However, the drift model considered in this approach takes into account a linear drift model. This assumption may not always hold true, and in the case of a more complex drift scenario, the approach may not perform as expected. The datasets used in this thesis show that the proposed approach can indeed handle drift better

than the comparison techniques. However, further testing is required on more complex real-world datasets or simulated datasets with non-linear drift model considered.

The results obtained in this thesis indicate that the proposed approach can be implemented on top of any regression technique. In this work, I have shown the performance of the approach using three regression techniques, namely, PLS, XGB and MLP. The promising results obtained with a simple regression model like PLS indicate that the proposed approach is indeed viable for continuous monitoring systems for long-term applications. However, more state-of-art regression technique like XGB did not always achieve the expected results. Given the characteristics of XGB coupled with the results obtained through PLS, might indicate that the XGB model can be further optimized to obtain even better performance. As for the case of the deep MLP network utilized in this work, further hyper parameter tuning can be performed to optimize the performance of the MLP model as well. Since, our approach also resolves the issue of lack of training data, the MLP network might be able to improve the results more if the proper set of optimized parameters are employed.

In addition to testing the approach on two separate experimental datasets, I have also employed a simulated dataset to test the efficacy of the approach in terms of complexity of the regression task. However, the drift model used in the synthetic datasets is a linear additive model. In real-world data the drift component may not be linear or additive. Further analysis using non-linear drift models might be more useful to explain the performance of the proposed approach.

REFERENCES

- Abbatangelo, M., Nunez-Carmona, E., Sberveglieri, V., Comini, E., & Sberveglieri, G. (2020). k-NN and k-NN-ANN Combined Classifier to Assess MOX Gas Sensors Performances Affected by Drift Caused by Early Life Aging. *Chemosensors*, 8(1), 6.
- Adhikari, S., & Saha, S. (2014). *Multiple classifier combination technique for sensor drift compensation using ANN & KNN*. Paper presented at the 2014 IEEE International Advance Computing Conference (IACC).
- Anik, M. T. H., Guilley, S., Danger, J.-L., & Karimi, N. (2020). *On the effect of aging on digital sensors*. Paper presented at the 2020 33rd International Conference on VLSI Design and 2020 19th International Conference on Embedded Systems (VLSID).
- Artursson, T., Eklöv, T., Lundström, I., Mårtensson, P., Sjöström, M., & Holmberg, M. (2000). Drift correction for gas sensors using multivariate methods. *Journal of chemometrics*, 14(5-6), 711-723.
- Chaudhuri, T., Wu, M., Zhang, Y., Liu, P., & Li, X. (2020). An Attention based deep sequential GRU model for sensor drift compensation. *IEEE Sensors Journal*.
- Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system*. Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.
- De Vito, S., Castaldo, A., Loffredo, F., Massera, E., Polichetti, T., Nasti, I., . . . Di Francia, G. (2007). Gas concentration estimation in ternary mixtures with room temperature operating sensor array using tapped delay architectures. *Sensors and Actuators B: Chemical*, 124(2), 309-316.
- Deshmukh, S., Bandyopadhyay, R., Bhattacharyya, N., Pandey, R., & Jana, A. (2015). Application of electronic nose for industrial odors and gaseous emissions measurement and monitoring—an overview. *Talanta*, 144, 329-340.

- Di Carlo, S., & Falasconi, M. (2012). *Drift correction methods for gas chemical sensors in artificial olfaction systems: techniques and challenges*: INTECH Open Access Publisher.
- Grover, A., & Lall, B. (2020). A novel method for removing baseline drifts in multivariate chemical sensor. *IEEE Transactions on Instrumentation and Measurement*, 69(9), 7306-7316.
- Guiseppi-Elie, A., Brahim, S., Slaughter, G., & Ward, K. R. (2005). Design of a subcutaneous implantable biochip for monitoring of glucose and lactate. *IEEE Sensors Journal*, 5(3), 345-355.
- Gutierrez-Osuna, R. (2000). Drift reduction for metal-oxide sensor arrays using canonical correlation regression and partial least squares. *Electronic Noses and Olfaction*, 147-152.
- Gutierrez-Osuna, R. (2002). Pattern analysis for machine olfaction: A review. *IEEE Sensors Journal*, 2(3), 189-202.
- Haugen, J.-E., Tomic, O., & Kvaal, K. (2000). A calibration method for handling the temporal drift of solid state gas-sensors. *Analytica chimica acta*, 407(1-2), 23-39.
- Heiberger, R. M., & Becker, R. A. (1992). Design of an S function for robust regression using iteratively reweighted least squares. *Journal of Computational and Graphical Statistics*, 1(3), 181-196.
- Huber, P. J. (1992). Robust estimation of a location parameter *Breakthroughs in statistics* (pp. 492-518): Springer.
- Jiao, Z., Hu, P., Xu, H., & Wang, Q. (2020). Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications. *ACS Chemical Health & Safety*, 27(6), 316-334.
- Khaydukova, M., Panchuk, V., Kirsanov, D., & Legin, A. (2017). Multivariate calibration transfer between two potentiometric multisensor systems. *Electroanalysis*, 29(9), 2161-2166.

- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kumar, M., Thakur, M., Senthuran, A., Senthuran, V., Karanth, N., Hatti-Kaul, R., & Mattiasson, B. (2001). An automated flow injection analysis system for on-line monitoring of glucose and L-lactate during lactic acid fermentation in a recycle bioreactor. *World journal of Microbiology and Biotechnology*, *17*(1), 23-29.
- Leon-Medina, J. X., Pineda-Muñoz, W. A., & Burgos, D. A. T. (2020). Joint distribution adaptation for drift correction in electronic nose type sensor arrays. *IEEE Access*, *8*, 134413-134421.
- Liu, Q., Hu, X., Ye, M., Cheng, X., & Li, F. (2015). Gas recognition under sensor drift by using deep learning. *International Journal of Intelligent Systems*, *30*(8), 907-922.
- Ma, Z., Luo, G., Qin, K., Wang, N., & Niu, W. (2018). Online sensor drift compensation for E-nose systems using domain adaptation and extreme learning machine. *Sensors*, *18*(3), 742.
- Marco, S., & Gutierrez-Galvez, A. (2012). Signal and data processing for machine olfaction and chemical sensing: A review. *IEEE Sensors Journal*, *12*(11), 3189-3214.
- Nguyen, T., Tathireddy, P., & Magda, J. J. (2018). Continuous hydrogel-based glucose sensors with reduced pH interference and contact-free signal transduction. *IEEE Sensors Journal*, *19*(6), 2330-2337.
- O'Mara, P., Farrell, A., Bones, J., & Twomey, K. (2018). Staying alive! Sensors used for monitoring cell health in bioreactors. *Talanta*, *176*, 130-139.
- Panchuk, V., Lvova, L., Kirsanov, D., Gonçalves, C. G., Di Natale, C., Paolesse, R., & Legin, A. (2016). Extending electronic tongue calibration lifetime through mathematical drift correction: Case study of microcystin toxicity analysis in waters. *Sensors and Actuators B: Chemical*, *237*, 962-968.

- Perera, A., Papamichail, N., Bârsan, N., Weimar, U., & Marco, S. (2006). On-line novelty detection by recursive dynamic principal component analysis and gas sensor arrays under drift conditions. *IEEE Sensors Journal*, 6(3), 770-783.
- Rudnitskaya, A. (2018). Calibration update and drift correction for electronic noses and tongues. *Frontiers in chemistry*, 6, 433.
- Rudnitskaya, A., Costa, A. M. S., & Delgadillo, I. (2017). Calibration update strategies for an array of potentiometric chemical sensors. *Sensors and Actuators B: Chemical*, 238, 1181-1189.
- Sasaki, I., Josowicz, M., Janata, J., & Glezer, A. (2006). Enhanced performance of a filter-sensor system. *Analyst*, 131(6), 751-756.
- Shaham, O., Carmel, L., & Harel, D. (2005). On mappings between electronic noses. *Sensors and Actuators B: Chemical*, 106(1), 76-82.
- Vergara, A., Vembu, S., Ayhan, T., Ryan, M. A., Homer, M. L., & Huerta, R. (2012). Chemical gas sensor drift compensation using classifier ensembles. *Sensors and Actuators B: Chemical*, 166, 320-329.
- Verma, M., Asmita, S., & Shukla, K. (2015). A regularized ensemble of classifiers for sensor drift compensation. *IEEE Sensors Journal*, 16(5), 1310-1318.
- Wang, S., Hu, Y., Burgués, J., Macro, S., & Liu, S.-C. (2020). *Prediction of gas concentration using gated recurrent neural networks*. Paper presented at the 2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS).
- Wenzel, M. J., Mensah-Brown, A., Josse, F., & Yaz, E. E. (2010). Online drift compensation for chemical sensors using estimation theory. *IEEE Sensors Journal*, 11(1), 225-232.
- Wold, H. (1975). Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. *Journal of Applied Probability*, 12(S1), 117-142.

- Yan, K., & Zhang, D. (2015). Improving the transfer ability of prediction models for electronic noses. *Sensors and Actuators B: Chemical*, 220, 115-124.
- Yan, K., & Zhang, D. (2016a). Calibration transfer and drift compensation of e-noses via coupled task learning. *Sensors and Actuators B: Chemical*, 225, 288-297.
- Yan, K., & Zhang, D. (2016b). Correcting instrumental variation and time-varying drift: A transfer learning approach with autoencoders. *IEEE Transactions on Instrumentation and Measurement*, 65(9), 2012-2022.
- Zhang, L., Liu, Y., He, Z., Liu, J., Deng, P., & Zhou, X. (2017). Anti-drift in E-nose: A subspace projection approach with drift reduction. *Sensors and Actuators B: Chemical*, 253, 407-417.
- Zhang, L., Tian, F., Kadri, C., Xiao, B., Li, H., Pan, L., & Zhou, H. (2011). On-line sensor calibration transfer among electronic nose instruments for monitoring volatile organic chemicals in indoor air quality. *Sensors and Actuators B: Chemical*, 160(1), 899-909.
- Zhang, L., & Zhang, D. (2014). Domain adaptation extreme learning machines for drift compensation in E-nose systems. *IEEE Transactions on Instrumentation and Measurement*, 64(7), 1790-1801.
- Zhao, X., Li, P., Xiao, K., Meng, X., Han, L., & Yu, C. (2019). Sensor drift compensation based on the improved LSTM and SVM Multi-Class ensemble learning models. *Sensors*, 19(18), 3844.
- Ziyatdinov, A., Marco, S., Chaudry, A., Persaud, K., Caminal, P., & Perera, A. (2010). Drift compensation of gas sensor array data by common principal component analysis. *Sensors and Actuators B: Chemical*, 146(2), 460-465.