

Marquette University

e-Publications@Marquette

Dissertations (2009 -)

Dissertations, Theses, and Professional
Projects

Analysis of Sensor Signals for Online Detection of Hydrocarbons in Liquids in the Presence of Interferents

Karthick Sothivelr
Marquette University

Follow this and additional works at: https://epublications.marquette.edu/dissertations_mu



Part of the [Signal Processing Commons](#)

Recommended Citation

Sothivelr, Karthick, "Analysis of Sensor Signals for Online Detection of Hydrocarbons in Liquids in the Presence of Interferents" (2018). *Dissertations (2009 -)*. 780.
https://epublications.marquette.edu/dissertations_mu/780

**ANALYSIS OF SENSOR SIGNALS FOR ONLINE DETECTION
OF HYDROCARBONS IN LIQUIDS IN THE
PRESENCE OF INTERFERENTS**

by

KARTHICK SOTHIVELR, B.S.E.E., M.S.E.E.

**A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in Electrical and Computer Engineering.**

Milwaukee, Wisconsin

May 2018

ABSTRACT
ANALYSIS OF SENSOR SIGNALS FOR ONLINE DETECTION
OF HYDROCARBONS IN LIQUIDS IN THE
PRESENCE OF INTERFERENTS

Karthick Sothivelr, B.S.E.E., M.S.E.E.

Marquette University, 2018

Current applicability of many chemical sensors is limited due to the lack of adequate selectivity to enable real-world applications. Often, the chemically sensitive element of the sensor is only partially selective to any specific target analyte, potentially giving rise to low probability of detection. Other challenges include the need to identify and quantify the target analytes in a mixture, especially in the presence of non-target interferents. In this dissertation, to enhance the selectivity of the sensor, analysis of sensor signals for detection and quantification of mixtures of hydrocarbon compounds in liquids in the presence of interferents using estimation theory and polymer-coated sensor devices is proposed. In particular, signal processing techniques are developed that can be employed for real-time detection and quantification of target analytes (specifically, petroleum hydrocarbons) in the presence of interferents using only a single polymer-coated shear horizontal surface acoustic wave (SH-SAW) sensor device. Estimation theory is used for signal processing because it enables near real-time data processing, minimal computational requirements, and minimal memory requirements for real-world implementations. The proposed techniques are based on bank of Kalman filters (BKFs) and/or exponentially weighted recursive least squares estimation (RLSE). The success of the approach depends on appropriate analytical modeling of the sensor response. A general n -analyte model is formulated that takes into account the responses due to the target analytes and non-target interferents that interact with the polymer coated sensor. The model, which assumes that sorption of one analyte does not prevent sorption of other analytes in the mixture, utilizes two sensor parameters, i.e. response time constant and sensitivity. Non-ideal cases, the non-step-like concentration versus time profile seen by the sensors, as well as concentration-dependent sensitivity are also considered.

The proposed techniques are tested using experimental sensor response data collected using polymer coated SH-SAW sensors with actual groundwater samples. The estimated analyte concentrations are compared to the results obtained independently using gas chromatography. Very good agreement (within about 10-15% accuracy) between the estimated and measured concentrations is found, even in the presence of non-target interferents. No complex training data set is required for the proposed technique.

ACKNOWLEDGEMENTS

Karthick Sothivelr, B.S.E.E., M.S.E.E.

I would like to express my sincere gratitude to my research advisor, Dr. Fabien Josse, for his guidance and encouragement throughout the completion of this work. The work in this dissertation would not have been possible without the countless hours of discussion on both technical aspects and writing style with Dr. Fabien Josse. My thanks are also due to Dr. Florian Bender for his time, guidance and helpful discussion on polymer coated sensor responses and for collecting some of the data analyzed in this work. I would also like to extend my sincere appreciation to Dr. Edwin Yaz and Dr. Susan Schneider for their time, guidance and helpful suggestions on estimation-theory-based techniques used throughout this work. My sincere gratitude also goes to my entire research committee: Dr. Fabien Josse, Dr. Florian Bender, Dr. Edwin Yaz, Dr. Susan Schneider, Dr. Antonio Ricco, and Dr. Henry Medeiros. Thank you for reading, reviewing, correcting and making constructive suggestions on how to improve this dissertation. I also would like to acknowledge my former and present fellow students in the Microsensor Research Laboratory and Department of Electrical and Computer Engineering.

Special thanks are due my parents, Sothivelr Muniandy and Manimala Muthu, my elder sister, Pungkulali Sothivelr, my younger brothers, Chandramohan Sothivelr and Vivisana Sothivelr, and the rest of my family for their support and encouragement in all my endeavors.

Finally, I would like to thank *Arut-Perun-Jothi Andavar* (the Divine Lord of the Supreme-Grace-Light) for continuously guiding me towards the attainment of blissful and deathless life.

*Arut-Perun-Jothi, Arut-Perun-Jothi
Thaniperung-Karunai, Arut-Perun-Jothi*

*Supreme-Grace-Light, Supreme-Grace-Light
Supreme Compassion, Supreme-Grace-Light*

May All Beings Receive the Supernal Grace and Live Blissfully!!!

May Prosperity, prosper at the Lotus Feet of the One who gives Freely!!!

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
LIST OF TABLES	v
LIST OF FIGURES	viii
1 INTRODUCTION	1
1.1 Chemical Sensors: General Background	1
1.2 Review of Signal Processing Techniques for Sensor Arrays	11
1.2.1 Signal Preprocessing	12
1.2.2 Feature Extraction and Dimensionality Reduction	14
1.2.3 Pattern Recognition and Classification.....	18
1.3 Problem Statement.....	20
1.4 Organization of the dissertation	27
2 ESTIMATION THEORY: A REVIEW	30
2.1 Introduction	30
2.2 Kalman Filter: A Review.....	33
2.3 Recursive Least Squares Estimation: A Review.....	41
2.3.1 Exponentially Weighted Recursive Least Squares Estimation.....	44
2.4 Bank of Kalman filters.....	48
3 MODELING THE SENSOR RESPONSE	55
3.1 Introduction	55
3.2 Model of the Single-Analyte Sensor Response.....	56
3.3 Model of the Multi-Analyte Sensor Response	65
3.4 Model of the Target Analytes in the Presence of Interferents	72
3.4.1 Example 1: Four-Analyte Model ($n = 4, p = 1$)	75
3.4.2 Example 2: Five-Analyte Model ($n = 5, p = 2$)	77
3.5 Model for the Non-Ideal Cases.....	78
3.5.1 Non-Step-Like Concentration versus Time Profile	79
3.5.2 Concentration-Dependent Sensitivity	82

4	ESTIMATION-THEORY-BASED SENSOR SIGNAL PROCESSING.....	88
4.1	Introduction	88
4.2	Online Detection and Quantification of Single Analyte in Solutions Containing any Arbitrary Single Analyte Using Bank of Kalman Filters.....	89
4.3	Online Detection and Quantification of Binary Mixtures in Solutions Containing any Arbitrary Binary Mixtures Using Bank of Kalman Filters	91
4.4	Online Identification and Quantification of Various Analytes in Multi- Analyte Mixtures Using Multi-Stage Exponentially Weighted RLSE	93
4.5	Near Real-Time Detection and Quantification of Multiple Target Analytes in a Chemical Sample Also Containing Various Interferents.....	98
5	COATED SH-SAW SENSOR DATA ACQUISITION.....	104
5.1	Introduction	104
5.2	Shear Horizontal Surface Acoustic Wave (SH-SAW) Devices.....	104
5.3	Measurement Setup and Data Acquisition	108
6	ESTIMATION RESULTS AND DISCUSSION.....	114
6.1	Introduction	114
6.2	Results of Detection and Quantification of Single Analytes Using Bank of Kalman Filters	115
6.3	Results of Detection and Quantification of Binary Mixtures Using Bank of Kalman Filters	126
6.4	Results of Identification and Quantification of Multi-Analyte Mixtures Using Multi-Stage Exponentially Weighted RLSE	133
6.5	Results of Detection and Quantification of Multi-Analyte in the Presence of Interferents	141
7	SUMMARY, CONCLUSIONS AND FUTURE WORK	166
7.1	Summary	166
7.2	Conclusions	173
7.3	Future Work	177
	REFERENCES.....	181

APPENDIX A: ADDITIONAL RESULTS	190
A.1 Additional Results for Identification and Quantification of Multi-Analyte Mixtures Using Multi-Stage Exponentially Weighted RLSE.....	190
APPENDIX B: MATLAB CODES.....	191
B.1 MATLAB Code for Detection and Quantification of Single Analytes Using Bank of Kalman Filters.....	191
B.2 MATLAB Code for Detection and Quantification of Binary Mixtures Using Bank of Kalman Filters.....	196
B.3 MATLAB Code for Identification and Quantification of Multi-Analyte Mixtures Using Multi-Stage Exponentially Weighted RLSE.....	201
B.4 MATLAB Code for Detection and Quantification of Multi-Analyte in the Presence of Interferents	206
B.4.1 Four-Analyte Model.....	206
B.4.2 Five-Analyte Model.....	233

LIST OF TABLES

Table 1.1: List of common steady-state features used for identification and quantification of chemical analytes [33].	16
Table 2.1: Applications of estimation theory [66].	32
Table 2.2: Summary of Kalman filter algorithm.	41
Table 2.3: Summary of exponentially weighted recursive least squares estimation (RLSE) algorithm.	47
Table 2.4: Summary of bank of Kalman filters algorithm.	53
Table 3.1: Measured Mean Sensitivities, $aKp - w$ (in Hz/ppm) and Response Time Constants, τ (in s), for sensors coated with 0.6 μm Poly(epichlorohydrin) (PECH) to Various BTEX Analytes and Common Interferents, Together with Their Standard Errors.	64
Table 3.2: Measured Mean Sensitivities, $aKp - w$ (in Hz/ppm) and Response Time Constants, τ (in s), for sensors coated with 0.8 μm Poly(isobutylene) (PIB) to Various BTEX Analytes and Common Interferents, Together with Their Standard Errors.	64
Table 4.1: Threshold values (approximate detection limits) of BTEX compounds and 1,2,4-trimethylbenzene (TMB) in water for SH-SAW sensors coated with either 0.6 μm PECH or 0.8 μm PIB.	98
Table 5.1: Basic properties of four polymers (PIB, PECH, PEA, PMMA) [83, 96].	110
Table 6.1: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.6μm PECH to 670 ppb benzene) collected for the first 1, 2, 3 and 5 minutes after the analyte has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.	118
Table 6.2: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.8μm PIB to 640 ppb toluene) collected for the first 4, 5, 6 and 9 minutes after the analyte has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.	121

Table 6.3: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.8 μ m PIB exposed to 370 ppb ethylbenzene) collected for the first 3, 4, 6 and 11 minutes after the analyte has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID..... 124

Table 6.4: Summary of identification and quantification results obtained using bank of Kalman filters and the measured response data of SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID. 125

Table 6.5: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.6 μ m PECH to 980 ppb benzene and 340 ppb toluene) collected for the first 3, 5, 6 and 9 minutes after the binary mixture sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID. 129

Table 6.6: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.8 μ m PIB to 510 ppb benzene and 260 ppb ethylbenzene) collected for the first 4, 6, 7 and 10 minutes after the binary mixture sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID. 131

Table 6.7: Summary of identification and quantification results obtained using bank of Kalman filters and the measured binary mixtures response data of SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID. 132

Table 6.8: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.8 μ m PIB to 590 ppb benzene, 420 ppb toluene, and 330 ppb ethylbenzene and xylenes) collected for the first 14, 16, 18 and 20 minutes after the multi-analyte sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID..... 136

Table 6.9: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.6 μ m PECH to 980 ppb benzene and 340 ppb toluene) collected for the first 3, 4, and 8 minutes after the binary mixture sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID. 138

Table 6.10: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.8 μ m PIB to 600 ppb 1,2,4-trimethylbenzene) collected for the first 5, 8, 10 and 14 minutes after the single analyte sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID. 139

Table 6.11: Summary of identification and quantification results obtained using multi-stage exponentially weighted RLSE and frequency transient data of SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB, compared to analyte concentrations in the mixture measured using GC-PID. In the table ‘B’ denotes benzene, ‘T’ denotes toluene, ‘EX’ denotes ethylbenzene and xylenes, and ‘TMB’ denotes 1,2,4-trimethylbenzene..... 141

Table 6.12: Estimated concentrations of BTEX compounds obtained using the measurement data of a LNAPL sample in groundwater (collected using a SH-SAW device coated with 0.6 μ m PECH) compared to concentrations measured using GC-PID..... 147

Table 6.13: Estimated concentrations of BTEX compounds obtained using the measurement data of a LNAPL sample in groundwater (collected using a SH-SAW device coated with 0.8 μ m PIB) compared to concentrations measured using GC-PID..... 151

Table 7.1: Summary of the proposed signal processing techniques for the investigated cases. 170

Table 7.2: Summary of the results obtained using the proposed signal processing techniques for the investigated cases..... 172

LIST OF FIGURES

Figure 1.1: Typical chemical sensor system [2].....	2
Figure 1.2: Schematic pathway for human exposure to hazardous compounds [1]..	6
Figure 1.3: Main signal processing components of a chemical sensor array.	12
Figure 1.4: Illustration of several baseline correction techniques [45]. In the figure y_{ss} represents the steady-state response and y_b represents the baseline response.	14
Figure 3.1: (a) Response of a SH-SAW sensor coated with 0.8 μm PIB, successively exposed to various samples of benzene in water (concentrations are indicated in the graph in parts per billion). (b) Response of a SH-SAW sensor coated with 0.8 μm PIB to various samples of toluene in water (concentrations in parts per billion).....	63
Figure 3.2: Response of a SH-SAW sensor coated with 1.0 μm PEA to toluene, ethylbenzene and their binary mixture. All concentrations are 10 ppm (binary mixture: 10 ppm toluene + 10 ppm ethylbenzene). Experimental data are modeled with single- and dual-exponential fits for single analytes and the binary mixture of analytes, respectively.	68
Figure 3.3: Concentration versus time profile of the measurement system used to collect the data analyzed in this dissertation.	82
Figure 3.4: Plot of equilibrium frequency shift, f^∞ versus maximum equilibrium ambient concentration, C_{eq} for concentration of ethylbenzene between 0 ppm to 70 ppm. Note that values for f^∞ were extracted from measured sensor response data for different C_{eq} collected using SH-SAW sensor coated with 0.6 μm PECH. Linear and nonlinear region are indicated in the figure. Also shown in the figure is the linear fit (red line) for the linear region of the plot where the slope of the line is the constant sensitivity, $aKp - w$	85
Figure 3.5: Exponential fit (red curve) to calibration data (blue dots) for ethylbenzene with concentration between 0 ppm to 70 ppm (large concentration range). The calibration data were collected using SH-SAW sensor coated with 0.6 μm PECH.....	87
Figure 4.1: Multi-stage exponentially weighted RLSE for online identification and quantification of various analytes in multiple analyte mixtures.	96
Figure 4.2: Estimation-theory-based sensor signal processing procedure that utilizes both sorption and desorption data for detection and quantification of BTEX analytes in the presence of interferences. “BT” and “EX” refers to the respective BTEX analytes.....	103

- Figure 5.1: Three-layer structure and coordinate system. The guided SH-SAW will propagate in the x_1 direction, x_2 is in the direction of the acoustic wave particle displacement, and x_3 is normal to the sensing surface [92]...... 107**
- Figure 5.2: The delay line configuration of a guided SH-SAW sensor device. The device is a two-port device and for simplicity, only one delay line is shown [83].... 108**
- Figure 5.3: Chemical structure of PMMA, PIB, PECH, and PEA [83]...... 110**
- Figure 5.4: Schematic of the measurement setup used for the data collection in the Microsensor Research Laboratory [97]...... 111**
- Figure 6.1: Measured response of a SH-SAW sensor coated with 0.6 μ m PECH to a single analyte sample of 670 ppb benzene (B). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentration is shown in the inset. 118**
- Figure 6.2: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to a single analyte sample of 640 ppb toluene (T). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentration is shown in the inset. 121**
- Figure 6.3: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to a single analyte sample of 370 ppb ethylbenzene (E). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentration is shown in the inset. 123**
- Figure 6.4: Measured response of a SH-SAW sensor coated with 0.6 μ m PECH to binary mixtures of 980 ppb benzene (B) and 340 ppb toluene (T). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentrations are shown in the inset. 128**
- Figure 6.5: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to a binary mixture of 510 ppb benzene (B) and 260 ppb ethylbenzene (E). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentrations are shown in the inset. 130**

Figure 6.6: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to multi-analyte mixtures of 590 ppb benzene (B), 420 ppb toluene (T), and 330 ppb ethylbenzene and xylenes (EX). Also shown (red line) is the sensor response estimated using the multi-stage exponentially weighted RLSE. The actual and estimated analyte concentrations are shown in the inset. 135

Figure 6.7: Measured response of a SH-SAW sensor coated with 0.6 μ m PECH to binary mixtures of 980 ppb benzene (B) and 340 ppb toluene (T). Also shown (red line) is the sensor response estimated using the multi-stage exponentially weighted RLSE. The actual and estimated analyte concentrations are shown in the inset. .. 137

Figure 6.8: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to 600 ppb 1,2,4-trimethylbenzene (TMB). Also shown (red line) is the sensor response estimated using the multi-stage exponentially weighted RLSE. The actual and estimated analyte concentrations are shown in the inset. 139

Figure 6.9: Measured response of a SH-SAW sensor coated with 0.6 μ m PECH to a LNAPL sample in groundwater containing 370 ppb benzene, 660 ppb toluene, and 330 ppb ethylbenzene/xylenes, and unknown concentration of interferents (top: sorption data, bottom: desorption data). Also shown (red dashed line) are the estimated sensor responses obtained using four-analyte model. 145

Figure 6.10: Measured response of a SH-SAW sensor coated with 0.6 μ m PECH to a LNAPL sample in groundwater containing 370 ppb benzene, 660 ppb toluene, and 330 ppb ethylbenzene/xylenes, and unknown concentration of interferents (top: sorption data, bottom: desorption data). Also shown (red dashed line) are the estimated sensor responses obtained using five-analyte model. 146

Figure 6.11: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to a LNAPL sample in groundwater containing 670 ppb benzene, 1340 ppb toluene, 510 ppb ethylbenzene/xylenes, and an unknown concentration of interferents (top: sorption data, bottom: desorption data). Also shown (red dashed line) are the estimated sensor responses obtained using four-analyte model. 149

Figure 6.12: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to a LNAPL sample in groundwater containing 670 ppb benzene, 1340 ppb toluene, 510 ppb ethylbenzene/xylenes, and an unknown concentration of interferents (top: sorption data, bottom: desorption data). Also shown (red dashed line) are the estimated sensor responses obtained using five-analyte model. 150

Figure 6.13: BTEX concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with 0.6 μ m PECH. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line. 156

Figure 6.14: BTEX concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line. 157

Figure 6.15: Benzene concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 15\%$ error line. 158

Figure 6.16: Toluene concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 15\%$ error line. 159

Figure 6.17: Ethylbenzene and xylenes (EX) concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line. 160

Figure 6.18: BTEX concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with 0.6 μ m PECH. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line. 161

Figure 6.19: BTEX concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line. 162

Figure 6.20: Benzene concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 15\%$ error line. 163

Figure 6.21: Toluene concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 15\%$ error line. 164

Figure 6.22: Ethylbenzene and xylenes concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line. 165

1 INTRODUCTION

1.1 Chemical Sensors: General Background

In recent years, there has been a growing need for chemical sensors for in-situ applications, which demand continuous real-time detection and monitoring of hazardous compounds in both gas and liquid environments (as opposed to sample collection on-site and transportation to an off-site laboratory for further analysis). Chemical sensors are devices or instruments that are used to determine the detectable presence, concentration (or quantity) of specific chemical substances (analytes) [1]. The working principle of a chemical sensor consists of converting a chemical stimulus due to the interaction between the target analytes and the sensor into an electrical signal (or optical signal in some cases). Further processing of this electrical (or optical) signal is performed to identify and quantify the analytes that induce the stimulus. A typical chemical sensor system consists of a recognition element (often a coating that is sensitive to the target analytes), a transduction element and a readout technology used to measure and convert the chemical perturbation into an electrical (or optical) signal as illustrated in Figure 1.1 [2].

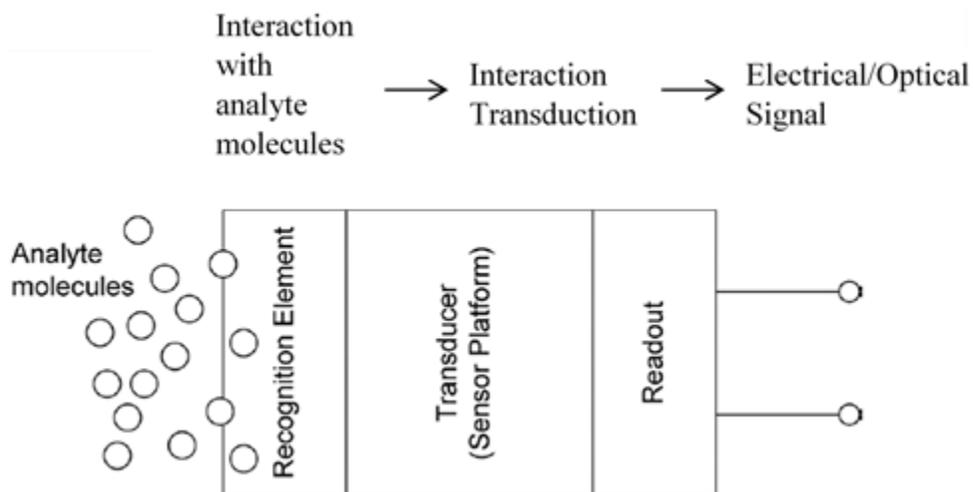


Figure 1.1: Typical chemical sensor system [2].

Chemical sensors are used in monitoring hazardous substances in many applications including manufacturing processes, environmental monitoring (both indoor and outdoor), and health monitoring. Of these applications, the most crucial application area for chemical sensors is in the detection of toxic chemicals in the environment. Toxic chemicals in the environment can be present in gas or liquid phase. Examples of toxic chemicals that need to be monitored in the gas phase include carbon monoxide, hydrogen cyanide, nitrites, arsine, carbon dioxide, volatile organic compounds (VOCs) and methane to name a few [3]. In the liquid phase, toxic chemicals that need to be monitored include VOCs, pesticides, pharmaceutical wastes etc. Often the presence of interfering species in the environment complicates the detection and quantification of the target hazardous compounds in both gas and aqueous environments. Interferents are the non-target analytes that are present along with the target analytes that can interfere with or disrupt the detection of target analytes.

The characteristics of the chemical sensor developed for a specific application are strongly dependent upon the requirements of the application. However, regardless of the application, the key characteristics that a chemical sensor must exhibit include:

- **Sensitivity:** Defined as the change in the measurement (output) signal per concentration unit change of the analyte [4, 5]. The slope of the calibration curve can be used for computing the sensitivity of the chemical sensor. The sensitivity of the sensor can be a constant, linear or nonlinear for the entire range of concentrations the sensor is exposed to.
- **Selectivity:** The ability of the chemical sensor to distinguish target analytes from non-target analytes [6] or to distinguish one target analyte from another. Strongly dependent upon the capability of the sensitive element of the sensor to recognize the size, shape, or dipolar properties of the target analytes [7].
- **Detection limit:** The smallest concentration of an analyte a chemical sensor can reliably detect. 'Reliably' here refers to an adequate signal-to-noise ratio [8].
- **Dynamic range (Span):** Concentration range that the chemical sensor can detect [9]. Concentrations outside this range may be unintelligible by the sensor.
- **Stability:** The ability of the chemical sensor to produce the same output value when measuring a fixed concentration over a period of time [9].
- **Repeatability:** Chemical sensor's ability to produce the same output for successive measurements of the same analyte concentration [9].
- **Reproducibility:** Ability to reproduce the same output responses for the same analyte concentration after some measurement condition has been altered [9].

- Time constant: The time taken for the sensor response to reach 63% of its final value when exposed to a step change in concentration [3].
- Linearity: Depends on the closeness of the calibration curve to an ideal straight line [9].

As indicated earlier, the primary task of chemical sensors is to detect and quantify the concentrations of target analytes in either gas or liquid phase. The detection and quantification of certain hazardous compounds has become of great importance for human health and ecosystems. In some cases, it is a legal obligation to monitor the level of certain hazardous compounds in the environment. These regulatory requirements are set by governmental agencies such as the Environmental Protection Agency, the Food and Drug Administration, the Occupational Safety and Health Administration, etc. The detection and quantification of hazardous compounds such as chemical warfare agents and VOCs like benzene, toluene, ethylbenzene and xylenes are of utmost importance due to the known hazard potentials of these compounds. The general pathways for human exposure to hazardous compounds are summarized in Figure 1.2 [1]. Based on Figure 1.2, it is vital to design a chemical sensor that can be placed at either the emission source of these hazardous compounds or in the media that transport these compounds so that the spreading of these hazardous compounds can be halted in the beginning stages (before human exposure).

Chemical sensors can be implemented using various sensing platforms such as optical devices, electrochemical devices, capacitive devices, resistive devices, micro-electromechanical systems (MEMS) and acoustic wave devices [10]. The overall

sensitivity of the chemical sensor is highly dependent on the chosen sensing platform. To be more precise, the overall device sensitivity is the product of the sensitivity of the chemically sensitive element to the target analytes and the sensitivity of the sensor platform to changes in the sensing element. Choice of sensing platform also plays a significant role in determining the suitable readout technique for a chemical sensor which in turn can greatly affect the overall complexity of the chemical sensor device.

Optical chemical sensors monitor the chemically induced changes in the optical parameters such as index of refraction, amount of absorbance, or intensity of photoluminescence in order to detect and quantify the target analytes [11]. Typically, a chemically sensitive layer will be placed on the optical device, which will be sensitive to the target analytes. The interaction between the target analytes and the chemically sensitive layer will result in the changes in a particular optical parameter, which can be related to the concentration of the target analytes [10, 11].

For electrochemical sensors, the sensing mechanism is based on monitoring the interaction between the target analytes and the electrodes of the device. A measurable signal (e.g. change in conductivity) will be induced due to the interaction between the target analytes and the electrode that can be related to the concentration of the target analytes [12]. Electrochemical sensors include voltammetric sensors, potentiometric sensors and amperometric sensors.

Capacitive and resistive chemical sensors have the advantage of the measurand signal already being an electrical signal (which eliminates the need for converting the measurand signal into electrical signal for further processing). In capacitive chemical sensors, the sensing principle is based on monitoring the changes in the capacitance of

the sensor device due to the interaction with the target analytes [13, 14]. The sensing principle of resistive chemical sensors (also known as chemiresistors) is based on monitoring the changes in the electrical resistance in response to interaction of the sensor with the target analytes [15].

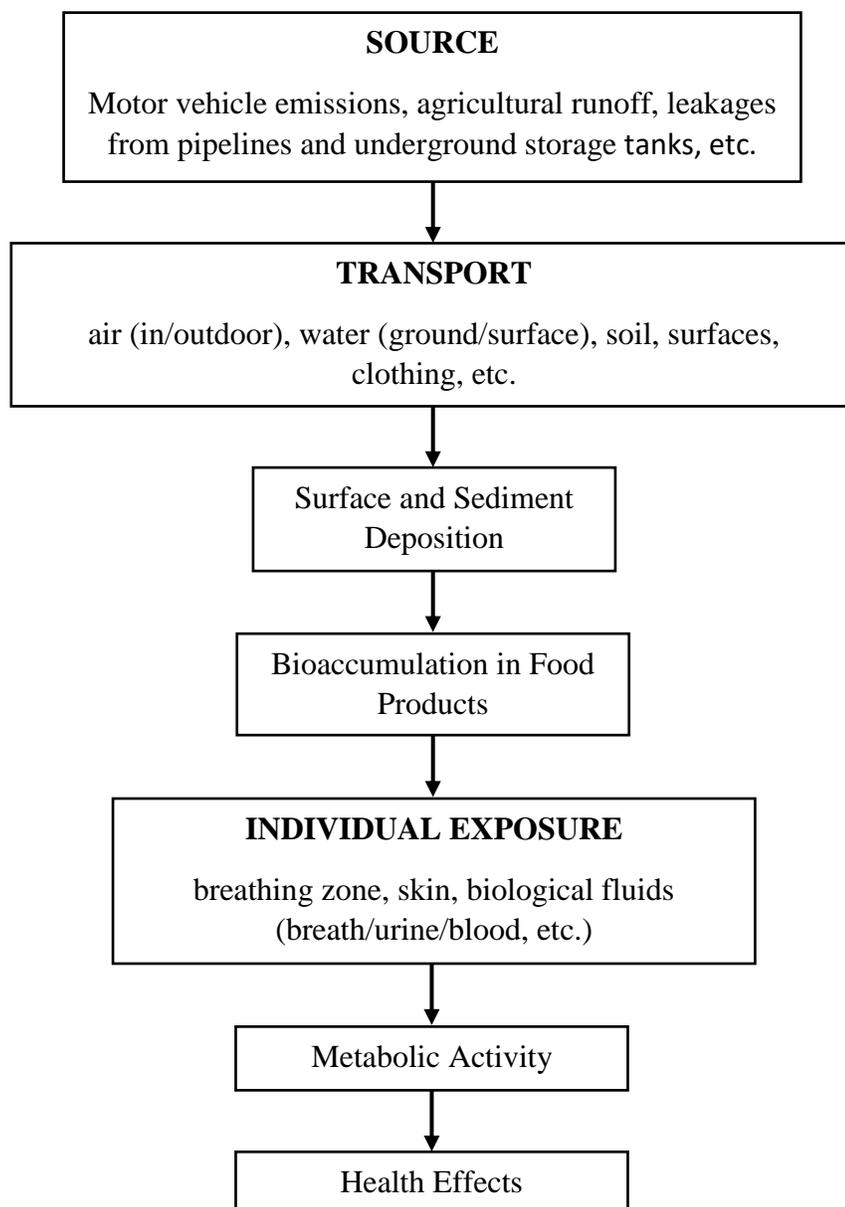


Figure 1.2: Schematic pathway for human exposure to hazardous compounds [1].

MEMS based chemical sensors belong to a relatively new class of chemical sensors which includes microbridges [16], microplates [17], and microcantilevers [18], etc. The mechanical or electrical properties of these micro-scaled devices are monitored, and any changes observed in these properties of the device can be related to the concentration of the target analytes [18]. One of the most promising MEMS based devices for chemical sensing is the microcantilever. A microcantilever is a diving-board-like structure usually only a few hundred microns in length and can be operated either in the dynamic mode or in the static mode. In the dynamic mode, the resonant frequency and quality factor of the device are monitored whereas in the static mode, the deflection of the microcantilever is monitored [19]. Microcantilevers can be used as a chemical sensor when a chemically sensitive layer is placed on the surface of a microcantilever. In dynamic mode, the analyte will interact with the polymer coating producing mass loading and stress effects, which will change the resonant frequency and quality factor of the microcantilever [19]. These changes can be related to the concentration of the analyte. On the other hand, in static mode, the differential surface stress from analyte sorption causes the microcantilever to bend and the magnitude of the bending of the microcantilever can be related to the concentration of analyte [19].

Acoustic wave chemical sensors include a wide variety of devices such as thickness shear mode (TSM) resonators [8, 20-22], surface acoustic wave (SAW) device [8, 24], shear-horizontal surface acoustic wave (SH-SAW) device [8, 26, 27], flexural plate wave (FPW) device [28] and acoustic plate mode (APM) device [29]. These devices can be used as chemical sensors for the detection of target analytes in gas and/or liquid phase. In general, acoustic wave devices utilize elastic waves at frequencies well above

the audible range propagating in piezoelectric crystals. Typically, acoustic wave devices are operated between the frequencies of 1 MHz to slightly above 1000 MHz [8]. The sensing mechanisms for acoustic wave devices are based on chemical interactions between the target analytes and the chemically sensitive coating placed on the sensor. This interaction will cause a perturbation in the propagation characteristics of the wave such as frequency and amplitude which can be related to the target analytes. TSM resonator and SAW devices are the most commonly used sensor platforms for chemical sensing applications [8, 20-24]. TSM resonators are capable of both gas and liquid phase detection of the target analytes. SAW devices, on the other hand, are only capable of gas phase detection. This is because when SAW devices are used for liquid phase sensing, the vertical displacement component of the wave will cause unacceptably high attenuation. Therefore, for liquid-phase detection, SH-SAW device is used rather than the standard SAW device [8, 26, 27]. For SH-SAW, the displacement component of the wave is parallel to the surface of the device and will not cause high attenuation of the wave in liquid phase.

The common issues for all chemical sensor technologies discussed earlier are lack of adequate sensitivity and selectivity. Both sensitivity and selectivity are highly dependent on the chemical interaction between the chemically sensitive layer (or polymer) and the target analytes. Therefore, the selection of the chemically sensitive element is critical to achieve the desired degree of sensitivity and selectivity for a particular application. To some extent, the issue with chemical sensitivity is solvable by selecting the appropriate sensor coating that can achieve the required degree of response for a given application. However, the issue with selectivity of a chemical sensor is very

difficult to rectify. Unlike biological sensors, chemical sensors are only partially selective i.e. the chemically sensitive element is not selectively sensitive to any specific analyte or group of analytes. For instance, if one wants to detect only methane in the air, it is difficult to develop a chemical sensor that only detects methane. The sensor is most likely to respond to ethane, propane and other similar molecules in the air making it difficult to determine which analyte is present and causing the response. Hence, the current applicability of many chemical sensors is limited due to the lack of sufficient selectivity to enable real-world applications. Lack of adequate selectivity of the chemical sensors further exacerbates the challenge in identifying and quantifying target analytes in a mixture of compounds. This is often caused by the chemical similarity within a group of target analytes, as well as the presence of non-target interferents to which the sensor is also responding.

One solution to enhance the selectivity of a chemical sensor is to use an array of partially selective chemical sensors instead of only a single sensor for a particular application [30-33]. All the sensors in the array respond to most if not all analytes in the sensing environment, but the pattern of responses from the sensors provides a unique fingerprint for each analyte or group of analytes [32]. The sensor array approach was initially proposed by Zaromb and Stetter who noted that while a single sensor could not discriminate a target analyte from possible interferents, many sensors would be able to detect multiple target analytes in the presence of interferents [30]. The sensor array approach to detect and quantify target analytes is often facilitated by dimensionality reduction and pattern-recognition techniques [34-39] including linear-discriminant analysis, principal-component analysis, and cluster analysis. Using a sensor array for

chemical sensing does, however, has potential drawbacks, the most critical drawback being the inability to obtain accurate results for multi-analyte mixtures. The sensor array approach often works well only for single analyte compounds or pre-defined mixtures. But as the complexity of the mixture increases, the approach will fail or produce inaccurate results for detection and quantification of target analytes. Other drawbacks of sensor arrays include increased signal- and data-processing time and potential misclassification. In general, sensor arrays require additional processing time compared to the processing time of only a single sensor. Unfortunately, this problem is further exacerbated with the increase in the number of sensors in the array, which will lead to increased data dimensionality and complexity. Furthermore, sensor arrays are also prone to misclassification errors which are particularly likely if the chemical diversity (“chemical orthogonality” [32]) and partial selectivity of the sensor coatings is insufficient. Moreover, the use of only one sensing parameter per sensor for classification, as is often the case, further intensifies the error due to misclassification [31]. Given all the drawbacks of the chemical sensors and chemical sensor arrays, there is a pressing need to develop a novel advanced signal processing approach that can overcome the existing challenges with the use of chemical sensors for in-situ applications. The goal is to develop a smart chemical sensor system by incorporating an advanced signal processing approach that enables the chemical sensor to operate effectively for in-situ applications. Since most of the existing approaches to develop a smart chemical sensor focus on the use of a sensor array, a brief review of signal processing techniques [40-42] with a sensor array will be discussed next.

1.2 Review of Signal Processing Techniques for Sensor Arrays

Signal processing of a chemical sensor array is a tedious process and demands significant computational capability to process the signal and data from each sensor in the array. Typically, in a sensor array, each of the sensors must be processed separately to eliminate noise, outliers and baseline fluctuations in the sensor response. Then, features (sensing parameters) are extracted from each of the individual sensor responses. The number of features used for classification dictates the dimensionality of the data collected from the sensor array for further processing. To avoid the issues that arise when analyzing data in high-dimensional spaces, often a dimensionality reduction technique is employed in a chemical sensor array by transforming the features to a new smaller dimensional feature space that retains the information most relevant for classification [43]. Finally, the sensor array data must be processed collectively using some form of pattern recognition technique to identify and quantify the target analytes causing the response. The main signal processing components of a chemical sensor array are depicted in Figure 1.3. In general, the main signal processing components of a chemical sensor array include signal preprocessing, feature extraction and dimensionality reduction algorithms, and pattern recognition and classification techniques [42]. In this section, a brief review of some of the common signal preprocessing techniques, feature extraction and dimensionality reduction techniques, and pattern recognition techniques employed in a chemical sensor array will be given.

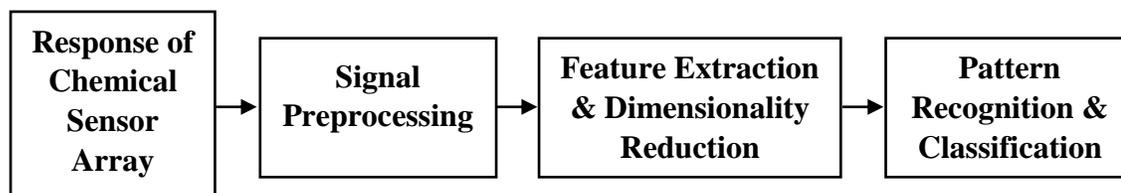


Figure 1.3: Main signal processing components of a chemical sensor array.

1.2.1 Signal Preprocessing

Chemical sensors in the array are susceptible to noise, outliers, and baseline fluctuations (or baseline drift) which could negatively influence the classification process. Therefore, the raw signal has to be preprocessed first to correct the data and remove these unwanted distortions.

Noise is basically any unwanted modifications that corrupt the true signal, whereas, an outlier is any observation point that is secluded from other observation points. Noise and outliers have significant impact on the signal-to-noise ratio (SNR) of a signal and thus, need to be filtered to improve the SNR. In order to eliminate the noise and outliers in the sensor response data, often digital filtering techniques are employed. Digital filters can be classified into two groups, i.e. infinite impulse response (IIR) filters and finite impulse response (FIR) filters [44]. For IIR filters, the output (filtered data) depends on the inputs and the previous outputs and for FIR filters, the output only depends on the inputs. Both types of digital filters have their own pros and cons. For instance, IIR filters are easier to implement but are susceptible to instability whereas FIR filters avoid instability with increased implementation complexity and longer propagation

delay [44]. Therefore, the choice of filter used to eliminate noise and outliers depends highly on the application and computational capability of the sensing system.

Baseline drift is another problem known to affect sensor signal and can greatly degrade the performance of a chemical sensor array, potentially resulting in misclassification errors. Baseline drift occurs due to fluctuations in environmental conditions such as changes in temperature, humidity, and pressure [45, 46]. Since it is difficult to control the environmental conditions where the sensor array is placed, often baseline correction techniques are employed to correct for the baseline drift. There are several baseline correction techniques that allow the estimation of the true baseline during the exposure to the analyte(s) such as linear extrapolation, linear interpolation, cubic interpolation and using estimation theory to estimate the baseline. The choice of the baseline correction technique to be employed depends on the duration of the sensor response. Linear extrapolation and linear interpolation are often used for sensors with short response time where it is implicitly assumed that the baseline slope remains constant during exposure. The linear extrapolation technique has the advantage of only requiring the data obtained before the sensor is exposed to the analyte(s) for estimating the baseline. On the other hand, linear interpolation requires data obtained both before the analyte(s) is added and after it has been flushed from the sensor for estimating the baseline [41], requiring additional time for the measurement to be completed before the estimation can be performed. However, linear interpolation is usually more accurate than linear extrapolation. For sensors with longer response time, linear extrapolation and linear interpolation could lead to a poor estimate of the baseline due to the possibility of baseline drift rate or direction change during a response [45]. In this case, cubic

interpolation function and a relatively new approach that utilizes estimation theory can be used. In most cases, as the complexity of the baseline correction techniques increases, it will yield a better result as depicted in Figure 1.4 [45]. It is imperative to note that the estimation theory approach for baseline drift correction can be implemented in real-time as the measurement is recorded, which could drastically shorten the overall time required to identify and quantify the analyte(s) from the sensor responses of the array [45, 46].

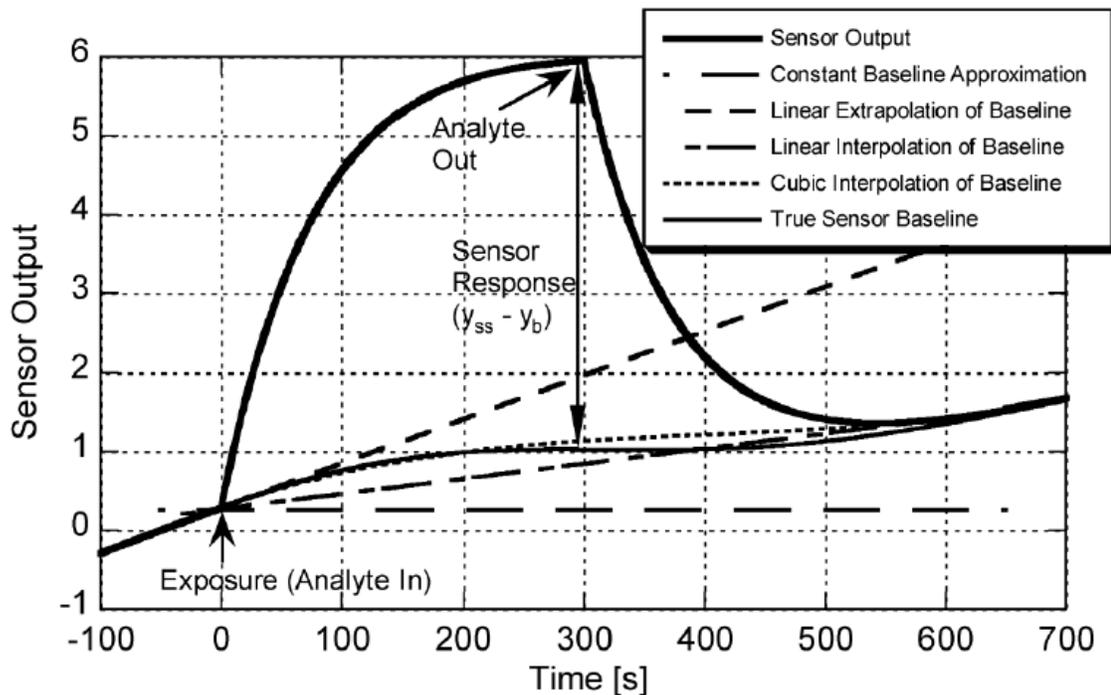


Figure 1.4: Illustration of several baseline correction techniques [45]. In the figure y_{ss} represents the steady-state response and y_b represents the baseline response.

1.2.2 Feature Extraction and Dimensionality Reduction

After preprocessing of the measured sensor responses from the array, the sensor responses are now ready for feature extraction (or sensor parameters extraction). The extraction of sensing parameters from each sensor in the array is the most crucial step

that will determine the successful identification and quantification of the target analytes. The sensor parameter most commonly used as the key feature to define the sensor response of each sensor in the array and to identify and quantify the target analytes is the steady-state feature [33]. Table 1.1 lists some common steady-state features and their formulae used for different types of sensor platforms [33]. There are several issues associated with the use of steady-state features to define a sensor response, the most critical issue being increased time-to-detection of the target analytes. If steady-state features are used, one must wait until the response reaches its steady-state before features can be extracted for further processing, i.e., before target analytes are identified and quantified. In some cases, the sensor could take a fairly long time before it reaches steady-state, especially in liquid phase detection where absorption can be slow (response time can be on the order of several minutes to hours). In these cases, if a dangerous level of a toxic chemical is present in the environment, the sensor would not be able to detect and quantify the toxic chemical rapidly, thus preventing the necessary remediation action to be carried out on time. In order to decrease the time-to-detection, the initial derivative of the sensor response is sometimes used instead of steady-state feature. By using initial derivative method, only the first few data points of the sensor response and an estimate of the initial derivative are required to quantify the target analytes [33]. However, the initial derivative method is prone to flow effects (i.e. how quickly the sensor is exposed to the sample) and higher noise (taking the derivative of the sensor response amplifies noise significantly).

Table 1.1: List of common steady-state features used for identification and quantification of chemical analytes [33].

Steady-State Feature	Formulae	Sensor Types
Difference	$x = y_{ss} - y_b$	Acoustic Wave Device Metal-Oxide Resistor
Relative	$x = y_{ss}/y_b$	Metal-Oxide Resistor Polymer Resistor
Fractional Change	$x = (y_{ss} - y_b)/y_b$	Metal-Oxide Resistor Polymer Resistor
Log Relative	$x = \ln(y_{ss}/y_b)$	Metal-Oxide Resistor

*** y_{ss} represents the steady-state response and y_b represents the baseline response.**

If only one feature or sensor parameter is available (or if the decision is made to ignore all but one parameter or feature) for analyte identification and quantification, misclassification errors are more likely to occur. For instance, if only steady-state feature is used for classification, there will be scenarios where two or more different analytes with different concentrations have the same steady-state value which will ultimately result in misclassification error. In this case, the sensor array will fail to identify the analyte most likely to have caused the sensor response. Hence, to improve the reliability of the classification of the sensor array, it is imperative to add another feature or sensing parameter for classification [47]. The addition of another feature will drastically improve the selectivity of the sensor array, thus permitting the construction of sensor arrays with fewer sensors. This will definitely contribute towards the reduction of the system cost, complexity and size. Another potential sensing parameter that could be utilized is the transient information of the sensor responses (assuming the transient information is

measurable). Often the sensor response to an analyte holds distinct transient information that can be used for analyte identification. If both transient and steady-state features are exploited, it can result in improved identification and increased recognition accuracy [46].

Once the features have been extracted from each of the individual sensor responses, a feature or observation vector is formed [33]. Typically, the feature vector, X is defined as

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix}, \quad (1.1)$$

where x_i is the i th feature from the sensor array and N is the total number of measured features. The dimension of the feature space is dependent on the number of features extracted from each of the individual sensor responses. Due to the issues that might arise when analyzing data in high-dimensional feature space, a dimensionality reduction technique is often employed in a chemical sensor array to reduce the dimensionality of the features by transforming the features to a new smaller dimensional feature space that retains the information most relevant for classification. Dealing with high-dimensional feature space can result in sub-optimal performance of the pattern recognition technique [43]. Thus, it is crucial to reduce the dimension of the higher-dimensional feature space. Usually, the dimension of the feature space will be reduced down to three- or even two-dimensional space for easier visualization. The general idea behind dimensionality

reduction techniques is to find a transformational matrix that can map an N -dimensional feature vector into smaller M -dimensional feature vector as shown in

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \\ \vdots \\ x'_M \end{bmatrix} = [A_{M \times N}] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix}, \quad (1.2)$$

where $[A_{M \times N}]$ is an M -by- N matrix with $M < N$. The transformation represented by $[A_{M \times N}]$ is usually optimized in some way to allow classification. Common dimensionality reduction techniques include principal component analysis (PCA) and linear discriminant analysis (LDA). In PCA, the transformation gives the best representation in the least-square sense of the feature vector, X , in the new M -dimensional feature space [34]. On the other hand, LDA tries to find the transformation that will maximize the separation between analyte classes in the new feature space [34].

1.2.3 Pattern Recognition and Classification

The final and important stage in chemical sensor array signal processing is pattern recognition and classification. Some common pattern recognition techniques include Bayesian decision rule, linear and nonlinear discriminant functions, nearest neighbor algorithm, neural networks, clustering algorithms, perceptron, etc. [33-39, 43]. The objective of pattern recognition and classification is to determine the possible analytes that are most likely to have caused the response, given the measured feature vector. The

pattern recognition process is divided into two steps, training and classification. The training step involves teaching the pattern recognition algorithm how the sensor array will respond to a known analyte or a group of analytes [38]. The classification step is where the information learned in training is used to determine the analyte (or group of analytes) that is most likely to have caused the given response [38]. The goal of pattern recognition is mathematically equivalent to finding the analyte(s) with the greatest probability given the measured feature vector,

$$a^* = \arg \max_i P(a_i|X), \quad (1.3)$$

where a^* represents the analyte or group of analytes identified as causing the response of the sensor array, a_i represents possible analytes and X represents the measured feature vector. Note that Bayesian probability theory can be used to find the conditional probability of each analyte in eq. (1.3). However, the use of Bayesian analysis requires the knowledge of the analyte (class) conditional probability density of the feature vector, $P(X|a_i)$, for each analyte that might be present [34]. In general, estimating $P(X|a_i)$ requires extensive training using the data that was collected from experiments for which the analyte information is known. There are also some pattern recognition techniques such as k nearest neighbor, linear discriminant functions, and neural networks that do not require direct calculation of the probability, $P(X|a_i)$, instead using a metric that is correlated to $P(X|a_i)$ [34]. The success of the pattern recognition algorithm in correctly identifying and quantifying the analyte(s) depends on the amount of training data used to train the algorithm. For instance, there is a high chance of misclassification error if a new

scenario that the algorithm has not been trained for were encountered. Generally, to design a sensor array that is robust to the wide variety of possible interferences requires a large number of sensors, which in turn, increases the dimensionality of the feature space and leads to the necessity of requiring more training data.

In summary, the signal processing involved with a chemical sensor array is complex and cumbersome. Even with these complex signal processing steps, there is no guarantee that the chemical sensor array will correctly detect and quantify the target analytes that might have caused the responses. This is especially true for applications which require the detection and quantification of the target analytes in the presence of interferences. Chemical sensor arrays will fail considerably as the complexity of the mixtures increases. Given all the drawbacks and the complex signal processing algorithms required with a chemical sensor array, there is a need to develop a novel signal processing approach that can enhance the ability of the chemical sensors for applications which require the detection and quantification of target analytes in the presence of interferences.

1.3 Problem Statement

The current applicability of many chemical sensors is limited due to the lack of adequate selectivity to enable real-world applications. In many chemical sensing applications, the chemically sensitive element of the sensor is only partially selective to any specific target analyte, potentially giving rise to low probability of detection and false positive or false negative results. Lack of adequate selectivity of the chemical sensors further exacerbates the challenge in identifying and quantifying the target analytes in a

mixture with potential presence of non-target interferents. Lack of selectivity is also often due to the chemical similarity within a group of target analytes.

One such real-world application which demands highly selective chemical sensors is in-situ monitoring of groundwater near underground storage tanks (USTs), pipelines, petrochemical processing facilities and military sites for accidental releases of petroleum hydrocarbons. The groundwater near these sites is always at risk of being contaminated with hazardous analytes and in some cases, it is a legal obligation to monitor such sites [1]. For example, the groundwater near USTs is subject to legal monitoring requirements by government agencies for accidental releases of fuel and oil [48]. The task of detecting and quantifying target contaminants in groundwater samples is challenging, not only because of the low concentrations that need to be detected, but also due to the presence of many chemically similar organic compounds as well as the presence of non-target interferents such as dissolved salts, aliphatic hydrocarbons, dissolved gases, particles and sediments, ethers, esters, ethanol, 1,2,4-trimethylbenzene, naphthalene, *n*-heptane, MTBE (methyl *tert*-butyl ether) etc. The volatile organic compounds (VOCs) benzene, toluene, ethylbenzene and xylenes (BTEX) are the target analytes of interest that can potentially contaminate groundwater because they are present in crude oil and its refined products and serve as good indicators of gasoline releases [49, 50]. The presence of BTEX compounds in groundwater is a major concern due their hazard potential and relatively high solubility in water [50-53]. The US Environmental Protection Agency (EPA) maximum contaminant levels of BTEX compounds in drinking water are in the low ppb ($\mu\text{g/L}$) to low ppm (mg/L) range [52]. Among the BTEX compounds, benzene is of particular importance due to its carcinogenicity and relatively higher water solubility [47,

53, 54]. Thus, it is imperative to monitor the presence of BTEX compounds in groundwater at critical sites regularly so that remediation actions can be carried out rapidly if leaks are detected. The chemical similarity of BTEX compounds as a group, and the frequent presence of other non-target interferents such as dissolved salts, aliphatic hydrocarbons, dissolved gases, particles and sediments, ethers, esters etc., further complicates the detection and quantification of BTEX analytes in groundwater.

Current methods for monitoring BTEX compounds in groundwater involves the use of either spectroscopy or gas chromatography [55-59]. Examples include infrared evanescent field spectroscopy [55-57] and Raman spectroscopy [58]. Although these methods are capable of accurately identifying and quantifying BTEX compounds, they are relatively impractical for use as a field-deployed system due to their total size and complexity in sample preparation procedures. The majority of the techniques that use spectroscopy and/or gas chromatography as the BTEX detection mechanism are only suitable for laboratory analysis. Hence, to utilize these techniques, groundwater samples have to be manually collected at the monitoring well and shipped to an ex-situ laboratory for analysis. Moreover, the existing methods are always at risk of losing vital information in the sample during sample collection, storage and transportation to an off-site laboratory for analysis. The entire process is expensive, labor-intensive and time-consuming, thus preventing frequent monitoring of groundwater near USTs, pipelines, petrochemical processing facilities and military sites. For instance, the groundwater near USTs is generally only monitored in 2-3 year intervals due to the high cost and labor-intensive process involved in sample collection at the monitoring well and shipping to an

ex-situ laboratory for analysis [60]. Due to such long monitoring intervals, leakages could go unnoticed for a long period of time.

An alternative approach for on-site BTEX monitoring would be to use a sensor array constructed with appropriately chosen partially selective chemical sensors [61, 62]. All sensors in the array respond to most if not all analytes, but the pattern of responses provides a unique fingerprint for each single target species or target mixture. Chemical sensor arrays based on various sensing platform technologies such as optical fibers [63], chemiresistors [61, 62] and SAW (surface acoustic wave) sensors [64, 65] have been investigated for the detection and quantification of BTEX compounds in water and soil. Identification and quantification based on arrays of partially selective sensor responses is often facilitated by pattern-recognition techniques as reviewed above. However, as indicated earlier, there are various drawbacks of using a sensor array. The drawbacks include increased signal- and data-processing time and potential misclassification. Data dimensionality and complexity, and hence processing resources and time, increase with the number of sensors in the array. Misclassification errors are particularly likely if the chemical diversity (“chemical orthogonality” [32]) and partial selectivity of the sensor coatings is insufficient. This problem is intensified if only one sensing parameter per sensor is used for classification, as is often the case [32]. Moreover, chemical sensor arrays often work well only for single compounds or pre-defined mixtures. A chemical sensor array alone is not a suitable candidate for applications which require the detection and quantification of the target analytes in the presence of interferents. Therefore, a need clearly exists to develop a chemical sensor system that is accurate, fast and inexpensive for in-situ and long-term monitoring of groundwater. The sensor system should be

capable of identifying and quantifying the hazardous compounds (i.e. BTEX compounds) present in the groundwater near the USTs, pipelines, petrochemical processing facilities and military sites in an automated way. Given the need for high accuracy in the estimation of target analyte concentrations under field conditions, the system needs a capable chemical sensor and an advanced signal processing unit. In-situ chemical sensors capable of direct liquid phase sensing are currently under development and SH-SAW devices coated with certain types of polymers are showing promise in detecting BTEX compounds in trace amounts in the presence of interferents commonly found in groundwater.

The work described in this dissertation focuses on the development of a capable signal processing unit that can enhance the ability of an array of polymer coated SH-SAW sensors or even a single polymer coated SH-SAW sensor to detect and quantify BTEX compounds in the presence of interferents in near real-time. The development of the advanced signal processing technique will be based on estimation theory. Utilizing estimation theory for this purpose offers several advantages which include the ability to process data in near real-time and low memory requirements for implementation using microcontrollers for in-situ applications. Specifically, the investigated novel signal processing approach utilizes estimation-theory-based technique comprised of both exponentially weighted recursive least-squares estimation (RLSE) and bank of Kalman filters (BKFs). With the awareness that the use of advanced signal processing techniques is only possible when the response of the sensor (SH-SAW in this work) to the target analytes in the presence of common interferents is accurately modeled, a n -analyte model that take into account the responses due to the target analytes and non-target interferents

for the selected polymer coated sensor is developed and investigated. The sensor response model is developed based on empirical results of individual target analytes, multi-analyte mixtures of target analytes and also based on the investigations of the common interferents found in the groundwater. The model utilizes two sensor parameters, i.e. the equilibrium frequency shift and the response time (for individual analyte), the latter being specific for each combination of coated device and analyte. Once the model is formulated, it will then be transformed into state-space form, so that exponentially weighted RLSE and BKF's can be used to estimate BTEX concentrations in the presence of interferents. Moreover, the use of desorption data to provide more information about the analyte-specific interactions with the polymer film will also be investigated. Estimation algorithms tailored for the application of detecting and quantifying BTEX compounds in the presence of interferents will be proposed and tested using the experimental data obtained using polymer coated SH-SAW sensors. All the estimation results obtained will be recorded, processed and analyzed.

Furthermore, the necessary modifications that need to be made in the sensor response model when non-ideal conditions occur are also addressed. The two non-ideal cases addressed are non-step-like concentration versus time profile (i.e. the transition from clean water to the sample in the flow cell containing the sensor is not sufficiently fast) and concentration-dependent sensitivity of the sensor. The ideal concentration versus time profile is the step-like concentration versus time profile where it is assumed that the sensor is exposed rapidly to the analyte containing sample (i.e. instantaneous transition from the clean water to the sample in the flow cell containing the sensor). In the real-world scenario, the concentration versus time profile seen by the sensor depends

on a number of measurement system parameters such as the average flow speed of the sample, the total length of the tube separating the point of sample introduction and the sensor device and diffusion coefficient of the soluble substance in the sample. These measurement system parameters determine whether the concentration versus time profile seen by the sensor is step-like or non-step-like. Hence, the modifications that need to be made in the sensor response model for the non-step-like concentration versus time profile is discussed. As for the sensitivity of the sensor, constant sensitivity is preferred for easier analysis. However, in most cases, for a larger concentration of the analyte (or sample), sensitivity will vary based on the concentration of the analyte which will result in non-constant value for the sensitivity. Thus, the possible modifications that can be made in the sensor response model when the sensor exhibits concentration-dependent sensitivity are also discussed.

Moreover, online sensor signal processing techniques based on either exponentially weighted RLSE or BKF's for the detection and quantification of multi-analyte mixtures (including single analyte responses) of target analytes in the absence of the interferences are also discussed. Specifically, a sensor signal processing technique based on BKF's is proposed for the detection and quantification of single and binary mixtures of analytes whereas a sensor signal processing technique based on multi-stage exponentially weighted RLSE is proposed for the detection and quantification of multi-analyte mixtures (including single analyte responses). These sensor signal processing techniques will be tested using experimental data obtained using polymer coated SH-SAW sensors. All estimation results obtained will be recorded, processed and analyzed.

1.4 Organization of the dissertation

This dissertation is presented in seven chapters. Chapter 1 gives an introduction to the chemical sensors and discusses the issues currently limiting in-situ applications of many chemical sensors. Significant emphasis is placed on chemical sensor arrays and issues related to their implementation. A brief review of signal processing techniques for chemical sensor arrays is given to illustrate the complexity in the successful implementation of sensor arrays for applications requiring detection of target analytes in the presence of interferents. Chapter 1 also introduces the main objective of this dissertation, namely, to develop a capable signal processing unit that can enhance the ability of an array of polymer coated SH-SAW sensors or even a single polymer coated SH-SAW sensor to detect and quantify BTEX compounds in the presence of interferents in near real-time. The importance for monitoring BTEX compounds in groundwater is also addressed in Chapter 1. Chapter 2 gives a review of all the estimation theory-based algorithms utilized in this dissertation. Specifically, the derivation and implementation of recursive least squares estimation, Kalman filter and bank of Kalman filters are discussed. Chapter 3 presents the model of the sensor responses to the target analytes. The use of estimation theory-based signal processing techniques is only possible if the response of the sensor to the analytes can be modeled analytically. These sensor response models were developed based on empirical results. Since the single analyte sensor response model serves as the basis for the multi-analyte sensor response model, the former is reviewed first. Then, by making necessary assumptions based on empirical observations, the single analyte model is extended to multi-analyte sensor responses.

Finally, by taking into account the sensor responses to the common non-target interferences found in the groundwater, the generic model was used for the detection and quantification of target analytes in the presence of possible interferences. Specifically, two different example models were used for the detection and quantification of BTEX compounds in the presence of interferences, i.e. four-analyte model and five-analyte model. Also discussed in Chapter 3 are the model for the non-ideal cases i.e. non-step-like concentration versus time profile and concentration-dependent sensitivity of the sensor. In chapter 4, the sensor signal processing techniques employed for the detection and quantification of multi-analyte mixtures of BTEX compounds with or without the presence of interferences (including single analyte responses of BTEX compounds) are presented. The justifications behind the proposed signal processing algorithm are also given. In Chapter 5, the specifics of SH-SAW sensors that were used to collect the data analyzed in this dissertation and the process of data acquisition using the SH-SAW sensor are discussed. In Chapter 6, the estimation results for the detection and quantification of multi-analyte mixtures of BTEX compounds with or without the presence of the interferences (including single analyte responses of BTEX compounds) using the formulated signal processing techniques are presented. These estimation results will be compared to the results obtained independently using gas chromatograph-photoionization detector (GC-PID) and gas chromatograph-mass spectrometry (GC-MS). All the proposed signal processing techniques will be tested using measured SH-SAW sensor responses (time-dependent frequency shift transient) to multi-analyte mixtures of BTEX compounds with or without the presence of the interferences (including single analyte responses of BTEX compounds). The sensors data were collected by the research team at

the Microsensors Research Laboratory, Marquette University. The results obtained from some of these tests are discussed in detail to highlight the effectiveness of the proposed sensor signal processing techniques. Finally, Chapter 7 provides a summary of the work performed in this dissertation, and also gives some suggestions regarding possible extensions of this work for future research in the field.

2 ESTIMATION THEORY: A REVIEW

2.1 Introduction

In this dissertation, estimation-theory-based techniques are used extensively for sensor signal processing to facilitate the detection and quantification of the target analytes in the presence of measurement noise and interferences. Estimation-theory-based techniques are utilized because they offer various advantages, including near real-time data processing, minimal computational requirements, and minimal memory requirements for real-world implementations. In general, estimation theory is a branch of statistics and signal processing that deals with estimating the values of unknown parameters or unmeasured states based on the measurement data [66]. These estimates are obtained using an estimator that attempts to approximate the unknown parameters or the unmeasured states using measured outputs. There are various forms of estimator and estimation techniques that are commonly used in different applications, including Kalman filter (KF) and its various derivatives, maximum likelihood estimators, Bayes estimators, Cramer-Rao bound, Wiener filter, Particle filter, least-squares filtering, Markov Chain Monte Carlo (MCMC) etc. Some of these estimators can be used to solve three common problems in state estimation, namely:

- Smoothing: Estimating the past values of states using the available measurements.
- Filtering: Estimating the present value of states using the available measurements.
- Prediction: Estimating the future value of states using the available measurements.

In fact, most of these modern estimation-theory-based techniques can be found at the heart of many electronic signal processing systems designed to extract information [66]. Typical application areas and example applications in areas utilizing estimation theory are listed in Table 2.1 [66].

All the signal processing techniques proposed in this dissertation are based on Bank of Kalman filters (BKFs) and/or exponentially weighted recursive least squares estimation (RLSE). Both BKFs and exponentially weighted RLSE are related to the Kalman filter (KF). As such, KF will be reviewed first in this chapter. A very detailed discussion on KF will be given including discussions on possible extension (or modification) to KF for nonlinear systems. A detailed review on RLSE technique will then be given starting with a brief introduction into RLSE, followed by a derivation of RLSE and exponentially weighted RLSE. Finally, a review on the derivation and implementation of BKFs is given.

Table 2.1: Applications of estimation theory [66].

Area of application	Example application
Control Systems	Estimation of the position of a powerboat for correcting navigation in the presence of sensor and environmental noise.
Communications	Estimation of the carrier frequency of a signal for demodulation to the baseband in the presence of degradation noise.
Seismology	Estimation of the underground distance of an oil deposit in the presence of noisy sound reflections.
Biomedical	Estimation of the heart rate of a fetus in the presence of environmental noise.
Image Processing	Estimation of the position and orientation of an object from a camera image in the presence of illumination changes and background noise.
Radar Communications	Estimation of the delay in the received pulse echo in the presence of noise.
Speech Signal Processing	Estimation of the parameters of the speech model in the presence of speech variability and environmental noise.
Sensor Signal Processing	Estimation of the baseline drifts in the sensor response in the presence of noise.

2.2 Kalman Filter: A Review

In this section, the derivation of the KF and possible modifications that can be made to KF for estimating the states of a nonlinear system are reviewed [67-74]. This detailed discussion will provide a better understanding of the least squares based filter and BKFs described in the later sections.

Kalman filter is named after R.E. Kalman, one of the primary developers of the theory behind it. In 1960, R.E. Kalman first used Kalman Filter to obtain a recursive solution to the discrete-data linear filtering problem [67, 68]. Today KFs are being used in many areas of applications, particularly in the area of assisted or autonomous navigation.

In general, KF is an online state estimation technique that is based on a set of mathematical equations providing an efficient recursive means to estimate the states of a system in a way that minimizes the mean of the squared error [68-71]. KF supports the estimation of the present states of a system given the available measurement data and with appropriate modifications can also be used to estimate the past and even future values of the states of a system [69]. Using KF, good state estimates can be obtained even when the precise nature of the modeled system is unknown [68].

For the derivation of KF, consider a general linear stochastic discrete-time system with internal states, x_k , outputs, y_k , inputs, u_k , and time-varying system matrices A_k , B_k , C_k and D_k as following (note that the system matrices can also be time-invariant),

$$x_{k+1} = A_k x_k + B_k u_k + F_k v_k , \quad (2.1a)$$

$$y_k = C_k x_k + D_k u_k + G_k w_k, \quad (2.1b)$$

where, v_k is the process or state noise with zero mean and covariance, V_k , and w_k represent the measurement noise with zero mean and covariance, W_k . It is assumed that both process and measurement noises are normally-distributed white noises (uncorrelated in time) and uncorrelated with each other. This assumption is made in line with most systems in real-world having white measurement and process noises. With the assumption that the system in eq. (2.1) meets the detectability criteria (i.e. if all unstable modes of the system are observable [75]), then it is possible to estimate the unknown states, x_k of the system by using only the available measurement data, y_k [68-71].

In order to estimate the new states of the system, \hat{x}_{k+1} , an estimator is formed using the information that will be available at any time, k . Typically, at any time, k , one will have access to three sources of information which include the present value of the state estimate, \hat{x}_k , the present value of the input, u_k and the present value of the measurement, y_k . Using \hat{x}_k , u_k , and y_k , an estimator of the form given in eq. (2.2) can be obtained as [69].

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + K_k (y_k - \hat{y}_k), \quad (2.2)$$

where \hat{y}_k is the estimate of the measurement,

$$\hat{y}_k = C_k \hat{x}_k + D_k u_k. \quad (2.3)$$

In eq. (2.2), K_k is called the estimator gain matrix (also commonly referred to as Kalman gain) and the quantity $(y_k - \hat{y}_k)$ is called the correction/innovation term. By substituting eq. (2.3) into eq. (2.2), eq. (2.2) can be rewritten as,

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + K_k (y_k - [C_k \hat{x}_k + D_k u_k]). \quad (2.4)$$

Note that the convergence of the state's estimator, eq. (2.4) to the actual states of system are highly dependent on the gain, K_k . Therefore, to obtain the best possible estimate for the states, optimal gain, K_k need to be found. The optimal gain, K_k can be found by minimizing the error covariance. For that, the estimation error, e_{k+1} and the error covariance, $P_{k+1} = E\{(e_{k+1})(e_{k+1})^T\}$ have to be defined first. Note that $(\cdot)^T$ denotes matrix transpose. The estimation error can be defined as the difference between the true state of the system and the state estimate,

$$e_{k+1} = x_{k+1} - \hat{x}_{k+1}. \quad (2.5)$$

Substitution of eqs. (2.1) and (2.4) into eq. (2.5) yields,

$$e_{k+1} = (A_k - K_k C_k) e_k + F_k v_k - K_k G_k w_k. \quad (2.6)$$

From eq. (2.6), the estimation error mean can be computed as,

$$E[e_{k+1}] = (A_k - K_k C_k)E[e_k] + F_k E[v_k] - K_k G_k E[w_k]. \quad (2.7)$$

From eq. (2.7), if $E[v_k] = 0$, $E[w_k] = 0$ and $E[e_k] = 0$, then $E[e_{k+1}] = 0$. This implies that if the process noise, v_k and measurement noise, w_k is zero-mean for all k , and if the initial estimate of x is set equal to the expected value of x , i.e. $\hat{x}_0 = E[x]$, then the expected value of \hat{x}_{k+1} will be equal to x_{k+1} for all k . Thus, the assumed estimator form of eq. (2.2) is an unbiased estimator [69]. Next, the expression for error covariance, P_{k+1} can be found using eq. (2.6),

$$P_{k+1} = E\{(e_{k+1})(e_{k+1})^T\}$$

$$P_{k+1} = A_k P_k A_k^T - A_k P_k C_k^T K_k^T - K_k C_k P_k A_k^T + K_k C_k P_k C_k^T K_k^T + F_k V_k F_k^T + K_k G_k W_k G_k^T K_k^T. \quad (2.8)$$

As indicated earlier, to determine the Kalman gain, K_k , the optimality criterion that was chosen to be minimized is the error covariance, P_{k+1} . Note that the error covariance matrix, P_{k+1} is a diagonal and symmetric matrix. Therefore, minimizing the error covariance matrix is equivalent to minimizing the trace of P_{k+1} (i.e. $Tr\{P_{k+1}\}$) [68]. Thus, the Kalman gain, K_k , can be found by taking the partial derivative of $Tr\{P_{k+1}\}$ with respect to K_k and solving it for K_k by setting the resulting equation to zero. Taking the partial derivative of $Tr\{P_{k+1}\}$ with respect to K_k yields

$$\frac{\delta Tr\{P_{k+1}\}}{\delta K_k} = -2A_k P_k C_k^T + 2K_k (C_k P_k C_k^T + G_k W_k G_k^T). \quad (2.9)$$

Setting eq. (2.9) to zero and solving it for K_k yields

$$K_k = A_k P_k C_k^T (C_k P_k C_k^T + G_k W_k G_k^T)^{-1}. \quad (2.10)$$

Therefore, the Kalman gain of eq. (2.10) is the value of the gain that would result in minimum error covariance at any given time k . Using eq. (2.10), alternate form of the error covariance of eq. (2.8) can be obtained,

$$P_{k+1} = A_k P_k A_k^T + F_k V_k F_k^T - A_k P_k C_k^T (C_k P_k C_k^T + G_k W_k G_k^T)^{-1} C_k P_k A_k^T. \quad (2.11)$$

Equations (2.4), (2.10) and (2.11) form the KF estimator which leads to a recursive algorithm for updating the state estimate based on the measurement data. Only the current estimate of the states and the latest measurement value is required to update the state estimate. Therefore, by using Kalman filter, the estimation can be performed in real-time under strict memory requirements as it is not required to store all the measurement values.

Note that if the system is a nonlinear stochastic discrete-time system, then some modifications need to be made, so that the estimation of the states of a nonlinear system can be performed using the KF equations developed earlier. This modification is done in the form of Taylor series expansion about the current state-estimate and neglecting the higher order terms (i.e. terms higher than first order). Since this modification is just an extension to the original KF, it is commonly referred to as extended Kalman filter (EKF).

Consider a general non-linear stochastic discrete-time system with internal states x_k , outputs y_k , and inputs u_k , given by,

$$x_{k+1} = f(x_k, u_k, v_k), \quad (2.12a)$$

$$y_k = h(x_k, u_k, w_k). \quad (2.12b)$$

In eq. (2.12), $f(\cdot)$ and $h(\cdot)$ are the nonlinear function of the state and output of the nonlinear system. As mentioned earlier, to perform the estimation for the nonlinear system of eq. (2.12) using the KF equations developed earlier, the nonlinear system has to be linearized by performing a Taylor series expansion about the current state estimate and by neglecting the higher-order terms. This will lead to the following approximation,

$$x_{k+1} \cong f(\hat{x}_k, u_k, \bar{v}) + A_k e_k + F_k v_k, \quad (2.13a)$$

$$y_k \cong h(\hat{x}_k, u_k, \bar{w}) + C_k e_k + G_k w_k, \quad (2.13b)$$

where e_k represents the error term (i.e. $e_k = x_k - \hat{x}_k$), \hat{x}_k is used to represent the state estimate, \bar{v} and \bar{w} represent the expected (mean) value of process and measurement noise, respectively, and matrices A_k , C_k , F_k , and G_k are defined as,

$$A_k = \left(\frac{\partial f}{\partial x} \right)_{\substack{x=\hat{x}_k \\ u=u_k \\ v_k=\bar{v}}},$$

$$C_k = \left(\frac{\partial h}{\partial x} \right)_{\substack{x=\hat{x}_k \\ u=u_k \\ w_k=\bar{w}}}$$

$$F_k = \left(\frac{\partial f}{\partial v} \right)_{\substack{x=\hat{x}_k \\ u=u_k \\ v_k=\bar{v}}}$$

$$G_k = \left(\frac{\partial h}{\partial w} \right)_{\substack{x=\hat{x}_k \\ u=u_k \\ w_k=\bar{w}}}$$

Note that the partial derivatives are actually time-varying Jacobian matrices and are evaluated at the current state estimate, known input value and mean of noise. For nonlinear systems, these Jacobian matrices will serve as the system matrices and can be used to perform the estimation in a similar fashion as KF by using eqs. (2.4), (2.10) and (2.11). However, for nonlinear systems, some modifications need to be made to eq. (2.4),

$$\hat{x}_{k+1} = f(\hat{x}_k, u_k, \bar{v}) + K_k[y_k - h(\hat{x}_k, u_k, \bar{w})]. \quad (2.14)$$

Thus, for nonlinear systems, eqs. (2.10), (2.11) and (2.14) can be used for estimating and updating the states based on the measurement data. This recursive algorithm for the nonlinear system is often referred to as EKF algorithm. It is important to note that EKF is not an optimal filter because Gaussianity of the probability distributions will not be preserved under a nonlinear transformation [76]. However, the EKF does give useful

estimates of the states and will demonstrate convergence for certain conditions. The convergence of EKF is dependent on the initial value of the error covariance and the value of process and measurement noise [76].

The implementation of the KF is rather straightforward using the equations developed earlier. KF estimates the states of a system using a form of feedback control which estimates and updates the states of a system using the feedback obtained in the form of output measurement [68]. Basically, the new state estimate will be updated or corrected using the new measurement value with a weighted average, where more weight is assigned to estimates with higher certainty.

Before applying the KF algorithm, it is crucial to convert the system into state-space model and determine the system matrices A_k , B_k , C_k and D_k for the linear system, and for the nonlinear system, one has to determine the general expression of the Jacobian matrices. Next, it is important to determine the measurement noise covariance, W_k and process noise covariance, V_k . In actual implementation of KF, the measurement noise covariance, W_k is usually measured prior to the operation of the filter. Typically, W_k is determined using some off-line sample measurements. On the other hand, determination of the process noise covariance, V_k , is generally more difficult as the states are not observed directly. Therefore, often, V_k is set by tuning the parameter to its optimum value. The general KF algorithm is outlined in Table 2.2.

Table 2.2: Summary of Kalman filter algorithm.

Step 1: Initialize the state estimate and the error covariance as follows:

$$\hat{x}_0 = E[x]$$

$$P_0 = E[(x - \hat{x}_0)(x - \hat{x}_0)^T]$$

(Note: If no knowledge of x is available before measurements are taken, then set $P_0 = \infty I$)

Step 2: Calculate Kalman gain:

$$K_k = A_k P_k C_k^T (C_k P_k C_k^T + G_k W_k G_k^T)^{-1}$$

Step 3: Update the state estimate:

[Linear system]

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + K_k (y_k - [C_k \hat{x}_k + D_k u_k])$$

[Nonlinear system]

$$\hat{x}_{k+1} = f(\hat{x}_k, u_k, \bar{v}) + K_k [y_k - h(\hat{x}_k, u_k, \bar{w})]$$

Step 4: Update the error covariance:

$$P_{k+1} = A_k P_k A_k^T + F_k V_k F_k^T - A_k P_k C_k^T (C_k P_k C_k^T + G_k W_k G_k^T)^{-1} C_k P_k A_k^T$$

Step 5: Repeat Step 2 – 4, as long as new output data are collected.

2.3 Recursive Least Squares Estimation: A Review

In this section, RLSE is reviewed [69, 72, 77, 78]. RLSE belongs to the class of adaptive filters, which recursively estimate the coefficients or the unknown parameters by minimizing the weighted least squares cost (or objective) function related to the known signal (i.e. unknown parameters are estimated by minimizing the error between the actual

measured data and the estimated measured data). It is commonly accepted that RLSE was first developed by Gauss in the 1800's but was ignored for almost a century and half before it was redeveloped by Plackett in 1950 [77]. It was only after the advent of Kalman filter in 1960 that RLSE gained popularity. RLSE is an attractive adaptive filtering technique because it overcomes some practical limitations of other adaptive techniques such as least mean squares (LMS) by providing a faster rate of convergence and a performance insensitive to variations in the eigenvalue spread of the correlation matrix of the known signal [77]. These advantages come with the penalty of increased computational complexity compared to adaptive techniques such as LMS.

In reality, RLSE is a special case of KF (that emphasizes the notion of a state and state-space model) [72]. In fact, in many applications, RLSE has long been utilized as a viable alternative to the KF, especially in applications that require the estimation of constant unknown parameters. The striking difference between RLSE and KF is that, RLSE does not place the importance on the notion of state and state-space model, whereas KF as an estimator is well known to be optimal under the requirement of complete prior knowledge of the state-space model and its parameters [77]. Moreover, the implementation of RLSE is computationally cheaper compared to the implementation of the KF [77].

The derivation of RLSE equations is very similar to the derivation of KF equations. To be more precise, there is a one-to-one correspondence between RLSE and KF. For the derivation of RLSE, the following linear system is considered:

$$y_k = H_k \theta_k + w_k. \quad (2.15)$$

In eq. (2.15), y_k represents the measurement vector, θ_k represents a vector of unknown parameters, H_k represents the vector or matrix of known signals or observations and w_k represents the measurement noise term. The measurement noise is assumed to be white noise with zero mean and covariance, W_k . The easiest way to derive the RLSE equations is through the utilization of the KF equations derived earlier. For that purpose, the linear system of eq. (2.15) can be rewritten in the state-space form as,

$$\theta_{k+1} = I\theta_k, \quad (2.16a)$$

$$y_k = H_k\theta_k + w_k. \quad (2.16b)$$

By direct comparison between eq. (2.1) and eq. (2.16), it is obvious that the state-space model of eq. (2.16) is obtained through the following assignment of the internal states and system matrices of eq. (2.1),

$$x_k = \theta_k, \quad (2.17)$$

$$A_k = I, \quad (2.18)$$

$$B_k = [0], \quad (2.19)$$

$$F_k = [0], \quad (2.20)$$

$$C_k = H_k, \quad (2.21)$$

$$G_k = [1]. \quad (2.22)$$

Therefore, by substituting eqs. (2.17) through (2.22) into eqs. (2.4), (2.10) and (2.11), RLSE update equations are obtained,

$$\hat{\theta}_{k+1} = \hat{\theta}_k + K_k(y_k - H_k\hat{\theta}_k), \quad (2.23)$$

$$K_k = P_k H_k^T (H_k P_k H_k^T + W_k)^{-1}, \quad (2.24)$$

$$P_{k+1} = P_k - P_k H_k^T (H_k P_k H_k^T + W_k)^{-1} H_k P_k. \quad (2.25)$$

Equations (2.23), (2.24) and (2.25) form the recursive least squares estimator. The implementation of RLSE to estimate the unknown parameters using the measurement data is very similar to the implementation of KF. In fact, RLSE can be implemented in real-time with lower computational costs.

2.3.1 Exponentially Weighted Recursive Least Squares Estimation

Typically, measurement data will be received sequentially over time and the information contained in the new set of data is crucial to improve the estimate of the

unknown parameters, $\hat{\theta}_k$. However, as the time, k increases, the error covariance, P_k and the gain, K_k will start decreasing to a very small value [70, 78]. This means that the corrections made to estimate the unknown parameters over time get small, independent of the newly collected measurement data. This event is sometimes referred to as the RLSE estimator 'going to sleep'. Therefore, to prevent the RLSE estimator from failing to respond adequately to the new data, exponential data weighting is often employed [70, 78].

The objective of exponential data weighting is to prevent the error covariance, P_k and the gain, K_k from getting too small as new data are processed using the RLSE estimator. This will ensure more credibility is given to the recent data. Data weighting can be performed by setting the measurement noise covariance equal to

$$W_k = W\lambda^{2(k+1)}, \quad (2.26)$$

for λ (forgetting factor) between 0 and 1 ($0 < \lambda \leq 1$; $\lambda=0$ means no memory and $\lambda=1$ indicates infinite memory) and constant W . Since $0 < \lambda \leq 1$, as time k increases, the measurement noise covariance decreases, giving more credibility to the recent data [70]. Substituting eq. (2.26) into eq. (2.25), the new error covariance is found to be equal to

$$P_{k+1} = P_k - P_k H_k^T (H_k P_k H_k^T + W\lambda^{2(k+1)})^{-1} H_k P_k. \quad (2.27)$$

Multiplying both sides of eq. (2.27) with $\lambda^{-2(k+1)}$ yields,

$$P_{k+1} = \frac{1}{\lambda^2} [P_k \lambda^{-2k} - P_k \lambda^{-2k} H_k^T (H_k P_k \lambda^{-2k} H_k^T + W \lambda^2)^{-1} H_k P_k \lambda^{-2k}]. \quad (2.28)$$

By defining the weighted error covariance as

$$P_k^\lambda = P_k \lambda^{-2k}, \quad (2.29)$$

eq. (2.28) can be rewritten as

$$P_{k+1}^\lambda = \frac{1}{\lambda^2} [P_k^\lambda - P_k^\lambda H_k^T (H_k P_k^\lambda H_k^T + W \lambda^2)^{-1} H_k P_k^\lambda]. \quad (2.30)$$

Note that the initial condition for the weighted error covariance, P_k^λ is the same as the initial condition for error covariance, P_k (i.e. $P_0^\lambda = P_0$). Substitution of eq. (2.26) into the gain, K_k eq. (2.24) yields

$$K_k = P_k H_k^T (H_k P_k H_k^T + W \lambda^{2(k+1)})^{-1}, \quad (2.31)$$

which can be written in terms of weighted error covariance as

$$K_k = P_k^\lambda H_k^T (H_k P_k^\lambda H_k^T + W \lambda^2)^{-1}. \quad (2.32)$$

In summary, exponentially weighted RLSE estimator can be implemented using eqs. (2.23), (2.30) and (2.32). Exponentially weighted RLSE algorithm is outlined in Table 2.3, which can be used to estimate the unknown parameters, $\hat{\theta}_k$ recursively.

Table 2.3: Summary of exponentially weighted recursive least squares estimation (RLSE) algorithm.

Step 1: Initialize the algorithm by setting:

$$\hat{\theta}_0 = E[\theta]$$

$$P_0^\lambda = E \left[(\theta - \hat{\theta}_0)(\theta - \hat{\theta}_0)^T \right]$$

$$0 < \lambda \leq 1$$

(Note: If no knowledge of θ is available before measurements are taken, then set $P_0 = \infty I$; $\lambda=0$ means no memory and $\lambda=1$ indicates infinite memory)

Step 2: When y_k and H_k becomes available, update the gain, K_k , unknown parameters, $\hat{\theta}_k$ and the weighted covariance matrix, P_k^λ :

$$K_k = P_k^\lambda H_k^T (H_k P_k^\lambda H_k^T + W \lambda^2)^{-1}.$$

$$\hat{\theta}_{k+1} = \hat{\theta}_k + K_k (y_k - H_k \hat{\theta}_k)$$

$$P_{k+1}^\lambda = \frac{1}{\lambda^2} \left[P_k^\lambda - P_k^\lambda H_k^T (H_k P_k^\lambda H_k^T + W \lambda^2)^{-1} H_k P_k^\lambda \right]$$

Step 3: Repeat Step 2, as long as the new measurement data are collected or until the unknown parameters converge to the actual value.

2.4 Bank of Kalman filters

Kalman filters require an exact knowledge of the system parameters (system model) and noise statistics for the accurate estimation of the states of a linear dynamical system [79]. However, in many engineering applications, it is difficult to construct a precise system model due to the underlying processes being too complex for scientific analysis [71]. Often, for systems with parameters that do not vary with time, it is possible to construct a model for the system using test input-output data off-line or on-line from the measurement data itself [71]. However, some systems have parameters which vary slowly in some random manner. Therefore, it is often preferred to come up with an adaptive filtering scheme that can estimate and adapt to variations in the system parameters on-line.

Typically for simultaneously estimating the unknown system parameters and the states on-line, it is common practice in many adaptive estimation schemes to treat the unknown system parameters as states too. The implication of treating the unknown parameters as states is that it often makes the adaptive estimation scheme to become too complex and highly nonlinear to solve directly without any simplifying assumptions. By making the simplifying assumption that the unknown parameters belong to a discrete set, a parallel processing technique which consists of multiple KFs (bank of KFs) can be used to estimate and adapt to variations in the system parameters on-line [71, 79]. This simplifying assumption ensures the adaptive estimation scheme does not become too complex and highly nonlinear. In most cases, the simplifying assumption allows the

system to stay linear. In general, the bank of KFs (BKFs) approach for parallel processing can be formulated as follows [79]:

- It is assumed that the unknown parameters vector, θ is discrete or suitably quantized to a finite number of grid points $\{\theta_1, \dots, \theta_N\}$, with known or assumed ‘*a priori*’ probability for each θ_j .
- The conditional mean estimator includes a parallel bank of N Kalman filters where the j^{th} filter is a standard Kalman filter designed on the assumption that $\theta = \theta_j$ and yielding conditional state estimates $\hat{x}_{k+1|k, \theta_j}$.
- The filter bank is driven by the noisy signal measurements, y_k .
- The conditional mean state estimate, $\hat{x}_{k+1|k}$ is given by a weighted sum of the conditional state estimate of the Kalman filters, $\hat{x}_{k+1|k, \theta_j}$.
- The weighted coefficient of the state of the j^{th} Kalman filter is the ‘*a posteriori*’ probability of θ_j , which can be updated recursively using the noisy signal measurements and the state of the j^{th} Kalman filter.

The recursion equation to update the ‘*a posteriori*’ probability of θ_j is crucial to the implementation of BKFs. In order to derive the recursion equation, first, a general linear stochastic discrete-time system (similar to eq. (2.1)) expressed in terms of the unknown parameter vector, θ that is assumed to belong to the discrete set $\{\theta_1, \dots, \theta_N\}$ is considered:

$$x_{k+1, \theta_j} = A_{k, \theta_j} x_{k, \theta_j} + B_k u_k + F_k v_k, \quad (2.33a)$$

$$y_k = C_{k, \theta_j} x_{k, \theta_j} + D_k u_k + G_k w_k, \quad (2.33b)$$

where $j = 1, 2, \dots, N$. Then, the conditional probability of θ_j assuming a sequence of measurements, $Y_k = \{y_0, y_1, \dots, y_k\}$ is equal to

$$\begin{aligned} p(\theta_j|Y_k) &= \frac{p(Y_k, \theta_j)}{p(Y_k)} \\ &= \frac{p(y_k, Y_{k-1}, \theta_j)}{p(y_k, Y_{k-1})}, \end{aligned} \tag{2.34}$$

where $p(\theta_j|Y_k)$ ('*a posteriori*' probability) is shorthand for $p(\theta = \theta_j|Y_k)$ and lower-case p is used interchangeably to denote a probability or probability density. Now, through the application of Bayes' rule, eq. (2.34) can be manipulated into the following relationships

$$\begin{aligned} p(\theta_j|Y_k) &= \frac{p(y_k, \theta_j|Y_{k-1})p(Y_{k-1})}{p(y_k|Y_{k-1})p(Y_{k-1})} \\ &= \frac{p(y_k|Y_{k-1}, \theta_j)p(\theta_j|Y_{k-1})}{\sum_{j=1}^N p(y_k|Y_{k-1}, \theta_j)p(\theta_j|Y_{k-1})}. \end{aligned} \tag{2.35}$$

The denominator of eq. (2.35) is just a normalizing constant. In order to determine the recursive equation to update the '*a posteriori*' probability, the computation of

$p(y_k|Y_{k-1}, \theta_j)$ is crucial [71, 79-81]. For Gaussian signal models, $p(y_k|Y_{k-1}, \theta_j)$ will be Gaussian with mean \hat{y}_{k,θ_j} and covariance S_{k,θ_j} which is equal to

$$\begin{aligned} S_{k,\theta_j} &= E \left[(y_k - \hat{y}_{k,\theta_j})(y_k - \hat{y}_{k,\theta_j})^T \right] \\ &= C_{k,\theta_j} P_{k,\theta_j} C_{k,\theta_j}^T + G_k W_k G_k^T. \end{aligned} \quad (2.36)$$

Hence, for a γ -vector y_k , the Gaussian probability density, $p(y_k|Y_{k-1}, \theta_j)$ [71, 79-81] will be equal to

$$\begin{aligned} p(y_k|Y_{k-1}, \theta_j) &= \left(\frac{1}{2\pi} \right)^{\gamma/2} |S_{k,\theta_j}|^{-0.5} \times \exp \left\{ -0.5 \left[y_k - \hat{y}_{k,\theta_j} \right]^T \left[S_{k,\theta_j} \right]^{-1} \left[y_k \right. \right. \\ &\quad \left. \left. - \hat{y}_{k,\theta_j} \right] \right\}, \end{aligned} \quad (2.37)$$

and clearly $p(\theta_j|Y_k)$ can be calculated recursively from

$$\begin{aligned} p(\theta_j|Y_k) &= \frac{1}{c} |S_{k,\theta_j}|^{-0.5} \times \exp \left\{ -0.5 \left[y_k - \hat{y}_{k,\theta_j} \right]^T \left[S_{k,\theta_j} \right]^{-1} \left[y_k \right. \right. \\ &\quad \left. \left. - \hat{y}_{k,\theta_j} \right] \right\} p(\theta_j|Y_{k-1}), \end{aligned} \quad (2.38)$$

where c is a normalizing constant independent of θ_j , chosen to ensure that $\sum_{j=1}^N p(\theta_j|Y_k) = 1$. Based on eq. (2.38) it is evident that the ‘*a posteriori*’ probabilities of all the filters in the bank are computed by evaluating how well the measurement, y_k compares to each KFs estimate of the measurement, \hat{y}_{k,θ_j} . This means that the KF which produces the smallest difference between the actual measurement, y_k and the estimate of the measurement, \hat{y}_{k,θ_j} will be weighted more heavily (larger ‘*a posteriori*’ probability) compared to the other KFs in the bank. Finally, to obtain parallel processing state estimation and its error covariance, the following equations can be used [71, 79],

$$\hat{x}_{k+1|k} = \sum_{j=1}^N p(\theta_j|Y_k) \hat{x}_{k+1|k, \theta_j}, \quad (2.39)$$

$$P_{k+1} = \sum_{j=1}^N p(\theta_j|Y_k) \left(P_{k+1, \theta_j} + [\hat{x}_{k+1|k, \theta_j}] [\hat{x}_{k+1|k, \theta_j}]^T \right) - [\hat{x}_{k+1|k}] [\hat{x}_{k+1|k}]^T. \quad (2.40)$$

For the case of time-varying unknown parameters, various modifications to the BKFs scheme are possible. Two common modifications include the use of exponential data weighting and reinitializing the states and ‘*a priori*’ probabilities of the KFs in the bank with the frequency of reset related to the rate of time variation of the unknown parameters [71, 79]. Note that the exponential data weighting can be implemented on the KFs in bank and/or on the ‘*a posteriori*’ probability, $p(\theta_j|Y_k)$ update equation. Table 2.4

summarizes the BKFs algorithm that can be used to obtain the best estimate for the unknown parameters vector, θ and the states, $\hat{x}_{k+1|k}$ recursively.

Table 2.4: Summary of bank of Kalman filters algorithm.

Step 1: Suitably quantize the unknown parameters vector, θ into a finite number of grid points $\{\theta_1, \dots, \theta_N\}$. (Note: this step will determine the number of parallel filters in the bank)

Step 2: Initialize the state estimate, \hat{x}_{0, θ_j} , the ‘*a priori*’ probabilities, $p(\theta_j|Y_0)$ and the error covariance, P_{0, θ_j} for all the filters in the bank.

Step 3: Calculate the Kalman gain, K_{k, θ_j} for all the filters in the bank:

$$K_{k, \theta_j} = A_{k, \theta_j} P_{k, \theta_j} C_{k, \theta_j}^T \left(C_{k, \theta_j} P_{k, \theta_j} C_{k, \theta_j}^T + G_k W_k G_k^T \right)^{-1}$$

Step 4: Update the state estimate, \hat{x}_{k, θ_j} and error covariance, P_{k, θ_j} of all the filters in the bank:

$$\hat{x}_{k+1, \theta_j} = A_{k, \theta_j} \hat{x}_{k, \theta_j} + B_k u_k + K_{k, \theta_j} (y_k - \hat{y}_{k, \theta_j})$$

$$P_{k+1, \theta_j} = A_{k, \theta_j} P_{k, \theta_j} A_{k, \theta_j}^T + F_k V_k F_k^T - A_{k, \theta_j} P_{k, \theta_j} C_{k, \theta_j}^T \left(C_{k, \theta_j} P_{k, \theta_j} C_{k, \theta_j}^T + G_k W_k G_k^T \right)^{-1} C_{k, \theta_j} P_{k, \theta_j} A_{k, \theta_j}^T$$

where,

$$\hat{y}_{k, \theta_j} = C_{k, \theta_j} \hat{x}_{k, \theta_j} + D_k u_k$$

Step 5: Compute the ‘*a posteriori*’ probabilities:

$$p(\theta_j|Y_k) = \frac{1}{c} |S_{k, \theta_j}|^{-0.5} \times \exp \left\{ -0.5 \left[y_k - \hat{y}_{k, \theta_j} \right]^T \left[S_{k, \theta_j} \right]^{-1} \left[y_k - \hat{y}_{k, \theta_j} \right] \right\} p(\theta_j|Y_{k-1})$$

where,

$$S_{k, \theta_j} = C_{k, \theta_j} P_{k, \theta_j} C_{k, \theta_j}^T + G_k W_k G_k^T$$

$$\hat{y}_{k, \theta_j} = C_{k, \theta_j} \hat{x}_{k, \theta_j} + D_k u_k$$

Step 6: Calculate the conditional mean state estimate and its covariance:

$$\hat{x}_{k+1|k} = \sum_{j=1}^N p(\theta_j | Y_k) \hat{x}_{k+1|k, \theta_j}$$

$$P_{k+1} = \sum_{j=1}^N p(\theta_j | Y_k) \left(P_{k+1, \theta_j} + [\hat{x}_{k+1|k, \theta_j}] [\hat{x}_{k+1|k, \theta_j}]^T \right) - [\hat{x}_{k+1|k}] [\hat{x}_{k+1|k}]^T$$

Step 7: Repeat Step 3 – Step 6 for all the available measurement data, y_k . (Note: For the time varying unknown parameters, the states and ‘*a priori*’ probabilities might need to be reinitialized before repeating Step 3 – 6)

Step 8: Based on the $p(\theta_j | Y_k)$, the best case for the unknown parameters vector, θ can be determined.

3 MODELING THE SENSOR RESPONSE

3.1 Introduction

In this chapter, models of the polymer coated shear-horizontal surface acoustic wave (SH-SAW) sensor responses to single analyte and multiple analytes are presented. SH-SAW sensors are used because of their ability for direct liquid phase detection of organic compounds. Note that, in this dissertation, three different polymer coatings were chosen as the chemically sensitive layer for the analytes of interest: poly(epichlorohydrin) (PECH), poly(isobutylene) (PIB) and poly(ethyl acrylate) (PEA). More details on SH-SAW devices, the chosen polymer coatings and experimental setup used to collect the sensor data analyzed in this work are given in Chapter 5. The use of estimation-theory-based techniques (reviewed in Chapter 2) for the detection and quantification of BTEX compounds in the presence of interferents requires an accurate analytical model that describes the response of the polymer coated SH-SAW sensor to the BTEX containing samples. The sensor response models presented here were developed based on empirical results. First, the model of sensor responses to single analytes are discussed. The single analyte sensor response model serves as the basis for the multi-analyte sensor response model. The general model of the multi-analyte sensor responses was then developed and modified based on empirical results obtained for the sensor responses to common non-target interferents found in contaminated groundwater. Specifically, most of the common groups of non-target interferents that are known to cause a response of the selected polymer coated SH-SAW sensors were studied in detail. Based on studies performed on these commonly known non-target interferents, a general

model for a system containing n analytes consisting of both analytes of interest and interferents was formulated. Finally, discussions on possible modifications to the sensor response models for non-ideal cases are given. Specifically, possible modifications for the case of non-step-like concentration versus time profile and concentration-dependent sensitivity of the sensor are discussed.

3.2 Model of the Single-Analyte Sensor Response

The model for the polymer coated SH-SAW sensor response to single-analyte samples was formulated based on the experimental data for various SH-SAW sensor responses to the individual BTEX compounds, as well as to some non-target interferents in the groundwater that are known to interact with the selected polymer coatings. Typical polymer coated SH-SAW sensor responses to single-analyte samples are shown in Figure 3.1. Based on such measured responses for BTEX compounds and for non-target interferents in the groundwater, several assumptions were made in order to model the single-analyte sensor response. These assumptions are listed below:

- 1) It is assumed that the coated SH-SAW sensor is exposed to a step change in the analyte concentration for the transition from clean (filtered) water to the sample (for sorption) or vice versa (for desorption). In other words, the sensor sees an instantaneous transition in the concentration (i.e. step-like concentration versus time profile) when the sample containing the analyte is introduced to the sensor or when the sample is switched with the clean water. This implies that the

concentration presented to the sensor, $C_{amb}(t)$, is almost equal to the maximum equilibrium ambient concentration, C_{eq} of the sample containing the analyte.

Mathematically, it can be represented using the following equation,

$$C_{amb}(t) = C_{eq}u_s(t), \quad (3.1)$$

where $u_s(t)$ is the unit step function. Note that achieving a step-like concentration versus time profile will depend on the measurement system set-up and parameters used for the experiments. For the measurement system used to collect the data analyzed in this dissertation, this assumption is valid. More details will be given later in section 3.5.

- 2) It is assumed that only physisorption occurs since the sorption process is reversible. Note that sensor reversibility is the ability of the sensor to return to its original baseline condition after exposure to an analyte [82]. As can be seen from the sample measurement data of Figure 3.1, the polymer coated SH-SAW sensor responses to the analytes of interest are reversible and have rapid sorption and desorption response time. Such responses indicate that only physisorption occurs. In physisorption, only weaker intermolecular forces (van der Waals force) occur between the coating and the analyte [82, 83].
- 3) It is assumed that the investigated concentration range for the analytes obeys the linear sorption isotherm. Linear sorption isotherm represents ideal sorption

behavior and is classified as type I sorption where the analyte concentration in the coating will be proportional to the ambient concentration of the analyte [83, 84].

Therefore, only low concentration range (analyte concentrations in the 0 – 10 ppm (parts per million) range) was investigated.

- 4) It is assumed that the rate of change in analyte concentration in the coating is proportional to the difference in the concentration of the analyte in the polymer coating at time t , $C(t)$ and the equilibrium concentration in the polymer coating. Therefore, when the polymer coated SH-SAW sensor is exposed to the aqueous solution containing the analyte, the sensor will respond rapidly at first and then slowly as the analyte concentration in the polymer coating approaches equilibrium (where the amount of analyte molecules entering and leaving the polymer coating will be equal). The extent of analyte sorption and the distribution of analyte between the polymer coating and the aqueous solution is characterized by the equilibrium constant which is also known as polymer-water partition coefficient, K_{p-w} [2, 82, 83]. Note that, at equilibrium, the concentration of the analyte in the polymer coating will be equal to the product of the maximum equilibrium ambient concentration, C_{eq} and the polymer-water partition coefficient, K_{p-w} . The constant K_{p-w} determines the overall strength of the interactions between the polymer coating and analyte. As a result, larger K_{p-w} , indicates stronger analyte sorption.
- 5) It is assumed that the rate of frequency change is proportional to the rate of change of the analyte concentration in the coating.

6) It is assumed that the absolute magnitude of the equilibrium frequency shift is the same for both analyte sorption response (measurements made immediately after exposing the sensor to the analyte containing sample) and analyte desorption response (measurements made by exposing the sensor to clean (or filtered) water after the analyte sorption process reaches equilibrium). As can be seen from the sample measurement data shown in Figure 3.1, the behavior of both analyte sorption response and analyte desorption response are identical. For analyte desorption it is also observed that the sensor signal changes rapidly at first and then more slowly as it reaches equilibrium. However, for the investigated sensor and coatings, analyte sorption will produce a negative frequency shift and desorption a positive frequency shift with the same absolute magnitude of equilibrium frequency shift. This means that the sensor responses are reversible. However, sorption and desorption responses might not necessarily have the same response-time constants. Empirically, it is observed that the desorption time constants are slightly different from the sorption time constants, but the difference is usually within the experimental error margins.

Based on these assumptions and the measured response in Figure 3.1, the single analyte sensor response (both analyte sorption and desorption responses) can be effectively modeled by an exponential rise to a steady-state governed by a single time constant, τ ,

$$\Delta f(t) = a K_{p-w} C_{eq} \left[1 - e^{-t/\tau} \right] u_s(t), \quad (3.2)$$

where $\Delta f(t)$ is the frequency shift as a function of time, a represents the sensitivity of the sensor platform (which also depends on coating thickness), and K_{p-w} is the polymer/water partition coefficient for a given analyte/coating pair. Note that the term aK_{p-w} represents the overall sensitivity of the sensor. Alternatively, eq. (3.2) can be written in terms of $C(t)$ as

$$\Delta f(t) = aC(t), \quad (3.3a)$$

where $C(t)$, the concentration of analyte in the coating at time t , is the solution to the following first-order differential equation:

$$\dot{C}(t) = -\frac{1}{\tau} C(t) + \frac{K_{p-w}}{\tau} C_{amb}(t). \quad (3.3b)$$

Equations (3.2) or (3.3) can be used to represent the single analyte responses (BTEX compounds or non-target interferents in the groundwater that are known to interact with the selected polymer coatings).

By fitting the single-analyte sensor response to eq. (3.2) or eq. (3.3), two sensor parameters could be extracted, namely, response time constant, τ and sensitivity, aK_{p-w} (Hz of frequency shift per ppm-by-mass of analyte concentration) for each coating/analyte combination. Experimental results show that both τ and aK_{p-w} are independent of analyte concentration in the range of interest. In particular, the

characteristic value of τ can be used to identify the analyte(s) and the characteristic value of aK_{p-w} can be used to quantify the analyte(s) (i.e. to determine its concentration) for a given sensor coating. Note that at equilibrium ($t = \infty$), eq. (3.2) can be represented as

$$\Delta f(\infty) = f_{\infty} = aK_{p-w}C_{eq}, \quad (3.4)$$

where f_{∞} denotes the equilibrium frequency shift. Thus, by using the experimentally determined sensitivity, aK_{p-w} for a given analyte/coating pair, the maximum equilibrium ambient concentration, C_{eq} can be computed by dividing the equilibrium frequency shift, f_{∞} by the sensitivity, aK_{p-w} ,

$$C_{eq} = \frac{f_{\infty}}{aK_{p-w}}. \quad (3.5)$$

Note that eq. (3.5) can be used to calculate the concentration of the analyte if the sensitivity of the analyte for a given coating is known. Therefore, the determination of average response time constant, τ and average sensitivity, aK_{p-w} of an analyte for a given coating is sufficient for successful the detection and quantification of the analyte. As such, several (more than 10) single-analyte measurements were made to determine the average values of the response time constants and sensitivities of the analytes of interest (here BTEX) and some interferents that are commonly found in contaminated groundwater. The measured mean sensitivities and response time constants for various

coatings are listed in Tables 3.1 and 3.2. For the investigated sensor coatings, chemical isomers ethylbenzene and the mixtures of three xylenes (*m*, *o*, *p*-xylenes) are found to have nearly identical values for the response times and sensitivities. Therefore, no attempt was made to distinguish the response time constants and sensitivities between them.

The single-analyte responses of the common interferences found in the contaminated groundwater that were analyzed experimentally include ethanol, 1,2,4-trimethylbenzene, naphthalene, *n*-heptane, and MTBE (methyl *tert*-butyl ether). Based on the single-analyte experiments for the selected polymer coatings, no significant response to ethanol was found up to concentrations of 100 ppm. For MTBE, a very low sensitivity was found for the selected coatings (~ 1 Hz/ppm) and since this compound is usually present at low concentrations in groundwater, its contribution to the sensor response is insignificant and can be ignored. High sensitivity and long response time were found for 1,2,4-trimethylbenzene and naphthalene for both PECH and PIB sensor coatings. In addition, PIB also exhibits high sensitivity and long response time for *n*-heptane. Therefore, the response to these compounds cannot be ignored and must be included in the model. Their mean sensitivities and response time constants are also listed in Tables 3.1 and 3.2 for different polymer coatings.

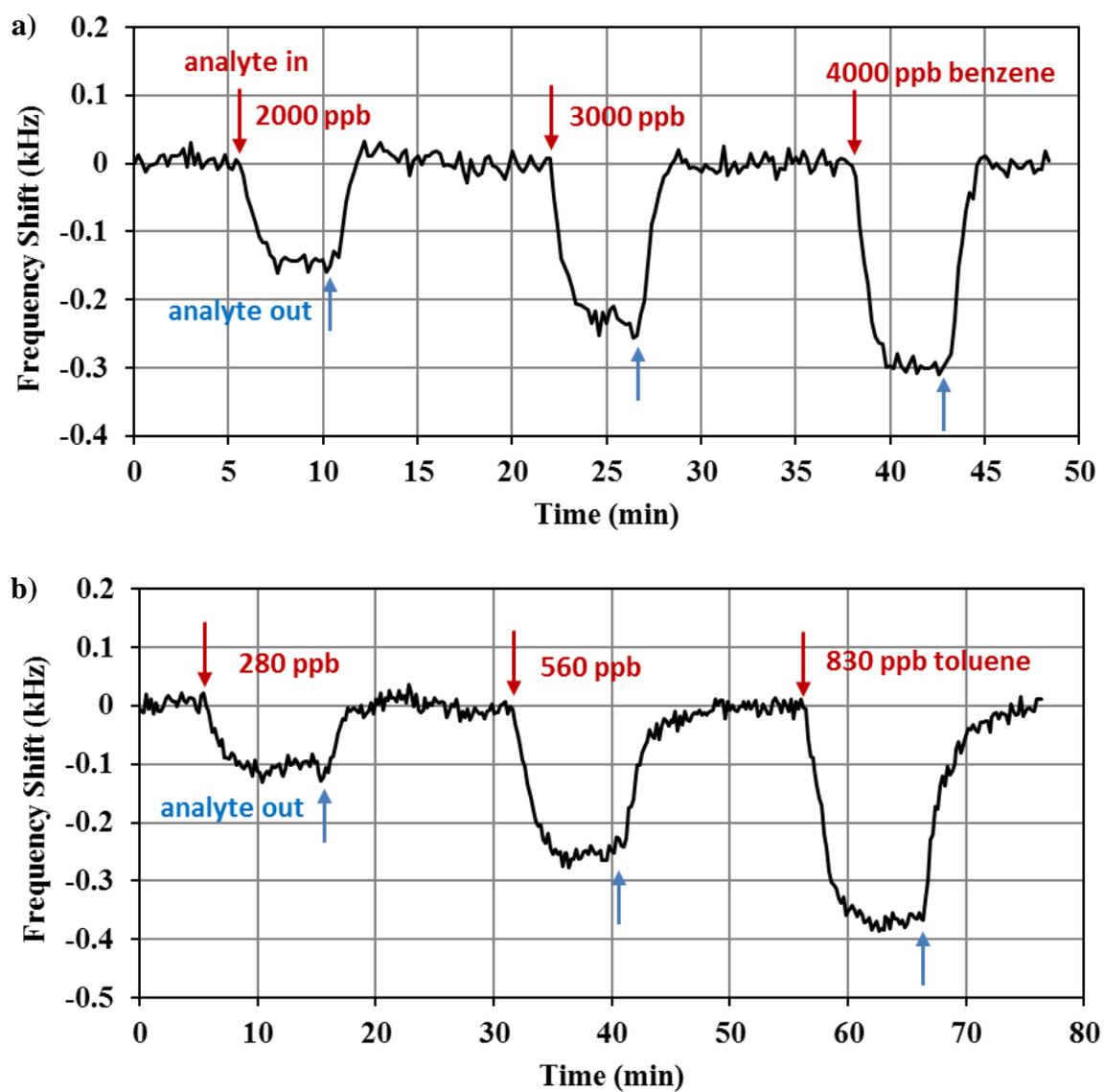


Figure 3.1: (a) Response of a SH-SAW sensor coated with 0.8 μm PIB, successively exposed to various samples of benzene in water (concentrations are indicated in the graph in parts per billion). (b) Response of a SH-SAW sensor coated with 0.8 μm PIB to various samples of toluene in water (concentrations in parts per billion).

Table 3.1: Measured Mean Sensitivities, aK_{p-w} (in Hz/ppm) and Response Time Constants, τ (in s), for sensors coated with 0.6 μm Poly(epichlorohydrin) (PECH) to Various BTEX Analytes and Common Interferents, Together with Their Standard Errors.

	Mean Sensitivities, aK_{p-w} (in Hz/ppm)		Mean Response Times, τ (in s)	
	Sorption	Desorption	Sorption	Desorption
Benzene	109 (± 9)	110 (± 11)	27 (± 8)	27 (± 9)
Toluene	435 (± 25)	422 (± 4)	78 (± 3)	74 (± 5)
Ethylbenzene	1450 (± 240)	1410 (± 240)	175 (± 13)	171 (± 13)
1,2,4-trimethylbenzene	3540 (± 420)	3390 (± 410)	428 (± 22)	461 (± 27)
Naphthalene	1562 (± 30)	1605 (± 27)	495 (± 87)	665 (± 24)
<i>n</i> -heptane	≈ 0	≈ 0	not applicable	not applicable
Ethanol	≈ 0	≈ 0	not applicable	not applicable
MTBE	\ll benzene	\ll benzene	$<$ benzene	$<$ benzene

Table 3.2: Measured Mean Sensitivities, aK_{p-w} (in Hz/ppm) and Response Time Constants, τ (in s), for sensors coated with 0.8 μm Poly(isobutylene) (PIB) to Various BTEX Analytes and Common Interferents, Together with Their Standard Errors.

	Mean Sensitivities, aK_{p-w} (in Hz/ppm)		Mean Response Times, τ (in s)	
	Sorption	Desorption	Sorption	Desorption
Benzene	78 (± 7)	78 (± 12)	36 (± 7)	31 (± 3)
Toluene	403 (± 39)	408 (± 85)	88 (± 7)	88 (± 9)
Ethylbenzene	1160 (± 57)	1100 (± 85)	230 (± 12)	215 (± 11)
1,2,4-trimethylbenzene	3640 (± 230)	3440 (± 165)	610 (± 18)	667 (± 31)
Naphthalene	621	650	250	254
<i>n</i> -heptane	\gg ethylbenzene	5932	\gg ethylbenzene	9177
Ethanol	≈ 0	≈ 0	not applicable	not applicable
MTBE	\ll benzene	\ll benzene	$<$ benzene	$<$ benzene

3.3 Model of the Multi-Analyte Sensor Response

The single-analyte sensor response model developed in the previous section can readily be extended to multi-analyte sensor responses. This extension is made based on the experimental data for various polymer coated SH-SAW sensor responses to multi-analyte mixtures, like the one depicted in Figure 3.2. Figure 3.2 shows the responses of a polymer coated SH-SAW sensor to a binary mixture of BTEX analytes (specifically, a mixture of toluene and ethylbenzene) and to the individual analytes that comprise the mixture. Based on Figure 3.2, it is evident that the measured total frequency shift for the binary mixture at any time, t is the sum of the frequency shifts of the individual analytes (i.e. $\Delta f(t) = \Delta f_1(t) + \Delta f_2(t)$). Based on such measured responses for multi-analyte mixtures, several assumptions were made in order to formulate the general analytical model for the sensor response to a mixture of n analytes. These assumptions are listed below:

- 1) It is assumed that all the assumptions made previously for the single-analyte sensor responses are still valid for the sensor responses of the multi-analyte mixtures.
- 2) It is assumed that for analyte concentrations in the range of 0 to 50 ppm (depending on the analyte), both sorption and desorption of the multi-analyte mixture by the polymer coating obey Henry's law [47, 50, 64]. Based on this assumption, it can be inferred that for a dilute mixture of multiple soluble species,

the sorption of any given species into the polymer does not affect the sorption of the other species in any way. Similar to the single analyte response, free partitioning of analytes between polymer and aqueous phase is assumed, including the implication that the sorption process is fully reversible at room temperature (i.e. only physisorption occurs). All this implies that the concentration of the mixture in the coating at any time t is the sum of the concentrations of each individual i th analyte, $C_i(t)$, as it would be measured in a single-analyte response, i.e. $C_{mixture}(t) = C_1(t) + C_2(t) + \dots + C_n(t)$.

- 3) Likewise, it is assumed that the equilibrium frequency shifts are also mutually independent, i.e. the frequency shift due to the mixture at any time t is the sum of the frequency shifts due to each analyte in the mixture at that time.
- 4) It is assumed that the values of the sensor parameters, response time constant, τ and sensitivity, aK_{p-w} of each analyte in the mixture is the same as those obtained from single analyte experiment for a given polymer coating.

Considering these assumptions, the sensor response to a mixture of n analytes can be modeled as the sum of the individual responses of each analyte in the mixture and is given by,

$$\Delta f(t) = \sum_{i=1}^n aK_{p-w,i} C_{eq,i} \left[1 - e^{-t/\tau_i} \right] u_s(t).$$

(3.6)

Alternatively, eq. (3.6) can be written in terms of $C_i(t)$ as

$$\Delta f(t) = \sum_{i=1}^n a C_i(t),$$

(3.7a)

where $C_i(t)$, the concentration of each analyte in the coating at time t , is the solution to the following first-order differential equation:

$$\dot{C}_i(t) = -\frac{1}{\tau_i} C_i(t) + \frac{K_{p-w,i}}{\tau_i} C_{amb,i}(t).$$

(3.7b)

All variables are as previously defined, with subscript $i = 1, 2, \dots, n$ referring to each analyte in the mixture. Equations (3.6) or (3.7) represent the general analytical model for the sensor response to any number of analytes in the sample, provided that each analyte and possible interferences in the sample have been separately characterized for the appropriately selected coating and in the concentrations range of interest.

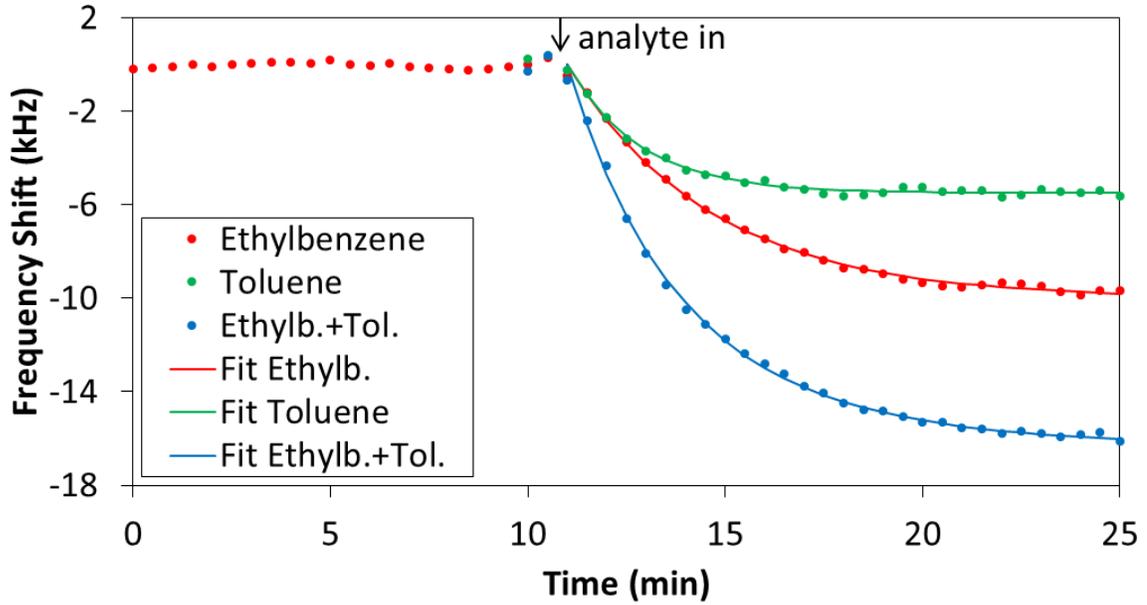


Figure 3.2: Response of a SH-SAW sensor coated with 1.0 μm PEA to toluene, ethylbenzene and their binary mixture. All concentrations are 10 ppm (binary mixture: 10 ppm toluene + 10 ppm ethylbenzene). Experimental data are modeled with single- and dual-exponential fits for single analytes and the binary mixture of analytes, respectively.

In order to utilize the estimation-theory-based techniques presented in Chapter 2 to extract recognition and quantification, the general multi-analyte sensor response model of eq. (3.7) was normalized and discretized. Equation (3.7) was normalized by dividing with $K_{p-w,i}C_{eq,i}$. By defining new variables as

$$m_i(t) = \frac{C_i(t)}{K_{p-w,i} C_{eq,i}}, \quad (3.8a)$$

$$u_s(t) = \frac{C_{amb,i}(t)}{C_{eq,i}}, \quad (3.8b)$$

and

$$f_{\infty,i} = aK_{p-w,i}C_{eq,i}, \quad (3.8c)$$

the following normalized differential equations and output equation are obtained as,

$$\dot{m}_i(t) = -\frac{1}{\tau_i}m_i(t) + \frac{1}{\tau_i}u_s(t), \quad (3.9a)$$

and

$$\Delta f(t) = \sum_{i=1}^n f_{\infty,i}m_i(t), \quad (3.9b)$$

where, for analyte i , $m_i(t)$ represents the normalized concentration absorbed/desorbed at time t , $f_{\infty,i}$ is the equilibrium frequency shift, and $u_s(t)$ represents the unit step concentration versus time profile observed by the sensor during the rapid transition from

clean water to the sample (for $t < 0$, $C_{amb,i}(t) = 0$; for $t > 0$, $C_{amb,i}(t) = C_{eq,i} u_s(t)$) or vice versa for desorption.

Since the frequency shift data are collected at discrete-time instants, $t = kT_s$, where T_s is the sampling period and k is a non-negative integer, the normalized model of eq. (3.9) was converted into a discrete-time model using Euler's first-order forward method,

$$\dot{m}_i(t) = \frac{m_{i,k+1} - m_{i,k}}{T_s}, \quad (3.10)$$

which yields the following discrete-time equations,

$$m_{i,k+1} = (1 - S_i)m_{i,k} + S_i u_{s,k}, \quad (3.11a)$$

$$\Delta f_k = \sum_{i=1}^n f_{\infty,i} m_{i,k} + w_k. \quad (3.11b)$$

In eq. (3.11), S_i is defined as

$$S_i = \frac{T_s}{\tau_i}, \quad (3.12)$$

and is commonly referred to as the sorption/desorption rate constant. Also note that in eq. (3.11), the term w_k is added to represent the measurement noise with variance σ_w^2 , which is likely to be present during data collection. It is assumed that the measurement noise is white noise (uncorrelated in time).

From eq. (3.11), the state-space form of the general multi-analyte sensor response model can be obtained by assigning state variables to the normalized concentrations, $m_{i,k}$ absorbed/desorbed at time instant k ,

$$\begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+1}^{(2)} \\ \vdots \\ x_{k+1}^{(n)} \end{bmatrix} = \begin{bmatrix} 1 - S_1 & 0 & 0 & 0 \\ 0 & 1 - S_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 - S_n \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \\ \vdots \\ x_k^{(n)} \end{bmatrix} + \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix} u_{s,k} = Ax_k + Bu_{s,k}, \quad (3.13a)$$

$$y_k = \Delta f_k = \sum_{i=1}^n f_{\infty,i} x_k^{(i)} + w_k = Cx_k + w_k, \quad (3.13b)$$

where,

$$x_k = [x_k^{(1)} \quad x_k^{(2)} \quad \dots \quad x_k^{(n)}]^T,$$

$$A = \begin{bmatrix} 1 - S_1 & 0 & 0 & 0 \\ 0 & 1 - S_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 - S_n \end{bmatrix},$$

$$B = [S_1 \quad S_2 \quad \cdots \quad S_n]^T,$$

$$C = [f_{\infty,1} \quad f_{\infty,2} \quad \cdots \quad f_{\infty,n}].$$

In eq. (3.13), $x_k^{(i)}$ represents the normalized concentration of absorbed/desorbed analyte at time instant k , and A , B and C represent the system matrices. Note that for near real-time n -analyte quantification, the unknown parameters that need to be estimated in the model defined by eq. (3.13) are the equilibrium frequency shifts (i.e., $f_{\infty,i}$, $i = 1, 2, \dots, n$). By using the estimated equilibrium frequency shift of each analyte in the mixture and known sensitivity, aK_{p-w} value for a given analyte/coating pair, the equilibrium ambient concentration for each analyte can be determined using eq. (3.5). It is recalled here that, because of K_{p-w} , the actual concentration of the analyte in the coating is different from the equilibrium ambient concentration for each analyte.

3.4 Model of the Target Analytes in the Presence of Interferents

In general, the sample under the test (or liquid environment of interest) may contain various chemical compounds other than the target chemical compounds. The presence of these non-target interferents often complicates the detection and quantification of the target analytes from the sensor response data. Typically, through the selection of appropriate sorbent polymer coatings, the sensor would not respond to most of these non-target interferents. Such polymer coatings were selected to collect the sensor

response data. However, the chosen polymer coatings do respond to some non-target interferents. Therefore, to enable the detection and quantification of the target analytes in the presence of interferents from the polymer coated SH-SAW sensor response data, the sensor response due to target analytes and the detectable interferents have to be taken into account in the sensor response model. Based on the sensor response to the detectable interferents, several different representations of the target analytes in the presence of interferents are possible using the general discrete-time sensor response model of eq. (3.11). All possible representations of the sensor response due to the target analytes and interferents in the sensor response model are given below:

Case 1: If all the detectable interferents have distinct sensor parameters (i.e. response time constants and sensitivities), the sensor response due to each of these interferents must be explicitly represented using individual exponential term (or single-analyte term) in the sensor response model. Therefore, in this case, eq. (3.11) can be used to represent the response due to the p detectable interferents using p individual exponential terms and q target analytes using q exponential terms where $n = q + p$. Note that n represents the total number of exponential terms in the sensor response model.

Case 2: If the all the detectable interferents have very low sensitivity to the chosen polymer coatings, the response due to the interferents can be ignored in the sensor response model. In this case, eq. (3.11) can be used to only represent the response due to the target analytes.

Case 3: If the response due to all the detectable interferents have nearly similar time constant and sensitivity, the combined response due to the interferents can be represented using only one exponential term. In this case, eq. (3.11) can be used to represent the response due to all detectable interferents using only one exponential term (i.e. $p = 1$) and q target analytes using q exponential terms where $n = q + 1$.

Case 4: If the response due to the different groups of interferents have different time constant and sensitivity, the combined response due to each group of interferents with nearly similar sensor parameters can be represented using one exponential term. In this case, eq. (3.11) can be used to represent the response due each group of detectable interferents with nearly similar time constant and sensitivity using only one exponential term (i.e. $p = p_1 + p_2 + p_3 + \dots$) and q target analytes using q exponential terms where $n = q + p$.

The general discrete-time sensor response model of eq. (3.11) will be used for the detection and quantification of BTEX compounds in the presence of interferents. For the chosen polymer coatings, based on the experimental observations of the sensor responses to the common interferents found in the contaminated groundwater, it has been found that the non-target interferents will either have slower response time constants or lower sensitivities than the target analytes (see Table 3.1 and Table 3.2). This indicates that the chosen polymer coatings have significantly larger partition coefficients for the target

analytes than some non-target interferents (thus, low sensitivities), or the target analytes have low solubility in water and are present at very small concentrations (thus leading to slow response times). Therefore, based on the experimental data on the sensor responses to the common interferents found in the contaminated groundwater, two specific example models, i.e. four-analyte model and five-analyte model, for the detection and quantification of BTEX compounds in the presence of interferents will be discussed.

3.4.1 Example 1: Four-Analyte Model ($n = 4$, $p = 1$)

For the four-analyte model, the total number of single-analyte terms, n is four ($n = 4$) and all the detectable interferents in the mixture with a response time constant longer or sensitivity lower than that of all BTEX analytes are modeled using a single exponential term, i.e. $p = 1$. Note that, as indicated earlier, ethylbenzene and the three xylenes are chemical isomers that have similar characteristic response time constants (for the selected coatings) which permits representation of the response to all of them by a single exponential term. This means that, for the four-analyte model, the subscript $i = 1, 2, 3$ represents the response due to the target analytes benzene, toluene, and the combination of the four C8 isomers, respectively, and the subscript $i = 4$ represents the combined response due to all the interferents in the mixture. The investigations performed on contaminated groundwater samples using polymer coated SH-SAW sensors indicates that the response to 1,2,4-trimethylbenzene is more pronounced than the response to other interferents; 1,2,4-trimethylbenzene is often the most common constituent of petroleum C9 aromatic compounds. Therefore, the time constant for $i = 4$

can be set close to that of the characteristic time constant of 1,2,4-trimethylbenzene. The four-analyte discrete-time model is as follows:

$$m_{i,k+1} = (1 - S_i)m_{i,k} + S_i u_{s,k} \quad \text{where } i = 1, 2, 3, 4 \quad (3.14a)$$

$$\Delta f_k = \sum_{i=1}^4 f_{\infty,i} m_{i,k} + w_k \quad (3.14b)$$

and the corresponding state-space form of the four-analyte model is as follows,

$$\begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+1}^{(2)} \\ x_{k+1}^{(3)} \\ x_{k+1}^{(4)} \end{bmatrix} = \begin{bmatrix} 1 - S_1 & 0 & 0 & 0 \\ 0 & 1 - S_2 & 0 & 0 \\ 0 & 0 & 1 - S_3 & 0 \\ 0 & 0 & 0 & 1 - S_4 \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \\ x_k^{(3)} \\ x_k^{(4)} \end{bmatrix} + \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{bmatrix} U_k \quad (3.15a)$$

$$y_k = \Delta f_k = \sum_{i=1}^4 f_{\infty,i} x_k^{(i)} + w_k \cdot \quad (3.15b)$$

Note that for four-analyte state-space model of eq. (3.15), the unknown parameters that need to be estimated for near real-time detection and quantification of target analytes are $f_{\infty,i}$, $i = 1, 2, 3, 4$.

3.4.2 Example 2: Five-Analyte Model ($n = 5, p = 2$)

For the five-analyte model, two exponential terms, i.e. $p = 2$ are used to represent the response due to all the detectable interferences in the mixture with a response time constant longer or sensitivity lower than that of target analytes. As a result, for the five-analyte model, the total number of single-analyte terms, n is five ($n = 5$). Here, the subscript $i = 1, 2, 3$ represents the response due to the target analytes benzene, toluene, and the combination of the four C8 isomers, respectively, and the subscript $i = 4, 5$ represents the combined response due to all the interferences in the mixture. In this model, the response to the interferent 1,2,4-trimethylbenzene, which is often dominant among the responses of the interferences, is treated individually as an analyte ($i = 4$) whereas the responses due to all the other interferences in the mixture are modeled collectively as a single-analyte ($i = 5$) with its characteristic time constant set higher than that of 1,2,4-trimethylbenzene; this has empirical reasons but is also consistent with the fact that besides 1,2,4-trimethylbenzene and similar C9 aromatic compounds, the second largest group of interferences identified in the contaminated groundwater samples are C10 aromatics which are expected to show even longer response times. The five-analyte discrete-time model is as follows:

$$m_{i,k+1} = (1 - S_i)m_{i,k} + S_i u_{s,k} \quad \text{where } i = 1, 2, 3, 4, 5 \quad (3.16a)$$

$$\Delta f_k = \sum_{i=1}^5 f_{\infty,i} m_{i,k} + w_k$$

(3.16b)

and the corresponding state-space form of the five-analyte model is as follows,

$$\begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+1}^{(2)} \\ x_{k+1}^{(3)} \\ x_{k+1}^{(4)} \\ x_{k+1}^{(5)} \end{bmatrix} = \begin{bmatrix} 1 - S_1 & 0 & 0 & 0 & 0 \\ 0 & 1 - S_2 & 0 & 0 & 0 \\ 0 & 0 & 1 - S_3 & 0 & 0 \\ 0 & 0 & 0 & 1 - S_4 & 0 \\ 0 & 0 & 0 & 0 & 1 - S_5 \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \\ x_k^{(3)} \\ x_k^{(4)} \\ x_k^{(5)} \end{bmatrix} + \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{bmatrix} U_k \quad (3.17a)$$

$$y_k = \Delta f_k = \sum_{i=1}^5 f_{\infty,i} x_k^{(i)} + w_k \cdot$$

(3.17b)

Note that for five-analyte state-space model of eq. (3.17), the unknown parameters that need to be estimated for near real-time detection and quantification of target analytes are $f_{\infty,i}$, $i = 1, 2, 3, 4, 5$.

3.5 Model for the Non-Ideal Cases

The coated SH-SAW sensor response models discussed in the earlier sections were developed for the ideal cases where it is assumed that the sensor is exposed rapidly to the analyte containing sample (i.e. step-like concentration versus time profile: instantaneous transition from the clean water to the sample or vice versa for desorption)

and constant sensitivity (i.e. linear relationship between the frequency shift at equilibrium, f_{∞} and the maximum equilibrium ambient concentration, C_{eq}). However, these ideal assumptions are often not met due to the limitations of the measurement system and sensor platform used to collect the sensor response data. If such non-ideal cases are encountered, the sensor response model must be modified in order to accurately represent the measured data. In this section, the possible modifications to the sensor response models for the case of non-step-like concentration versus time profile and concentration-dependent sensitivity of the sensor are discussed.

3.5.1 Non-Step-Like Concentration versus Time Profile

In many chemical sensing applications, the sample collected from the environment (with concentration, C_{eq}) is transferred rapidly to a cell containing the sensor(s). Typically, in such cases, it is assumed that the concentration presented to the sensor(s), $C_{amb}(t)$, is almost equal to C_{eq} ,

$$\frac{C_{amb}(t)}{C_{eq}} \approx 1.$$

(3.18)

Hence, if the flow of the sample to the sensor is sufficiently fast, the concentration versus time profile observed by the sensor during the transition from clean water to the sample or vice versa for desorption, $u(t)$ can be assumed to be equal to unit step function (for $t < 0$, $C_{amb}(t) = 0$; for $t > 0$, $C_{amb}(t) = C_{eq}$). However, in practice, non-step-like

concentration versus time profile is often observed when the sample is introduced to the sensor. This is due to the dispersion which occurs when the sample is introduced into a tube containing steadily flowing clean water which will transport the sample to the sensor [85]. Dispersion is caused by the combined action of convection parallel to the axis and molecular diffusion in the radial direction, which in turn, depends on the measurement system parameters such as the average flow speed, v , the total length of the tube separating point of sample introduction and the sensor device, l , and diffusion coefficient of the soluble substance, D [85]. Therefore, the concentration versus time profile seen by the sensor depends on these measurement system parameters and can be accurately described using a modified error function [85, 86],

$$u(t) = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{l - vt}{2\sqrt{Dt}} \right) \right]. \quad (3.19)$$

Equation (3.19) is the generalized expression for the concentration versus time profile and also can be used to determine whether the concentration versus time profile observed by the sensor can be assumed to be unit step function, $u_s(t)$ or not. For measurement systems with sufficiently high v , short l , and very small D , eq. (3.19) reduces to unit step function (i.e. step-like concentration versus time profile). If this is not the case, the value of the concentration versus time profile observed by the sensor at time, t , has to be determined using eq. (3.19). Note that, for non-step-like concentration versus time profile cases, one can still use the sensor response model developed in earlier sections but with

$u_s(t)$ replaced with $u(t)$ given by eq. (3.19). Therefore, eq. (3.19) is a simple adaptation that can be employed for measurement systems with non-step-like concentration versus time profile.

For the measurement system with sampling period, $T_s = 12s$, used to collect the data analyzed in the present work, the parameters are as follows,

$$l = 185.7 \text{ mm} ,$$

$$v = 6.19 \text{ mm/s} ,$$

$$D = 0.8 \times 10^{-9} \text{ m}^2 \text{ s}^{-1} \text{ to } 1.03 \times 10^{-9} \text{ m}^2 \text{ s}^{-1} \text{ (for BTEX compounds).}$$

These parameters will yield concentration versus time profile as shown in Figure 3.3.

Based on Figure 3.3, it is evident that the concentration versus time profile, for the measurement system used to collect the data analyzed in this dissertation, can be assumed to be a unit step function, $u(t) = u_s(t)$ after neglecting several initial data points immediately after transition from clean water to sample or from sample to clean water, due to a time delay of $t_D = l/v$.

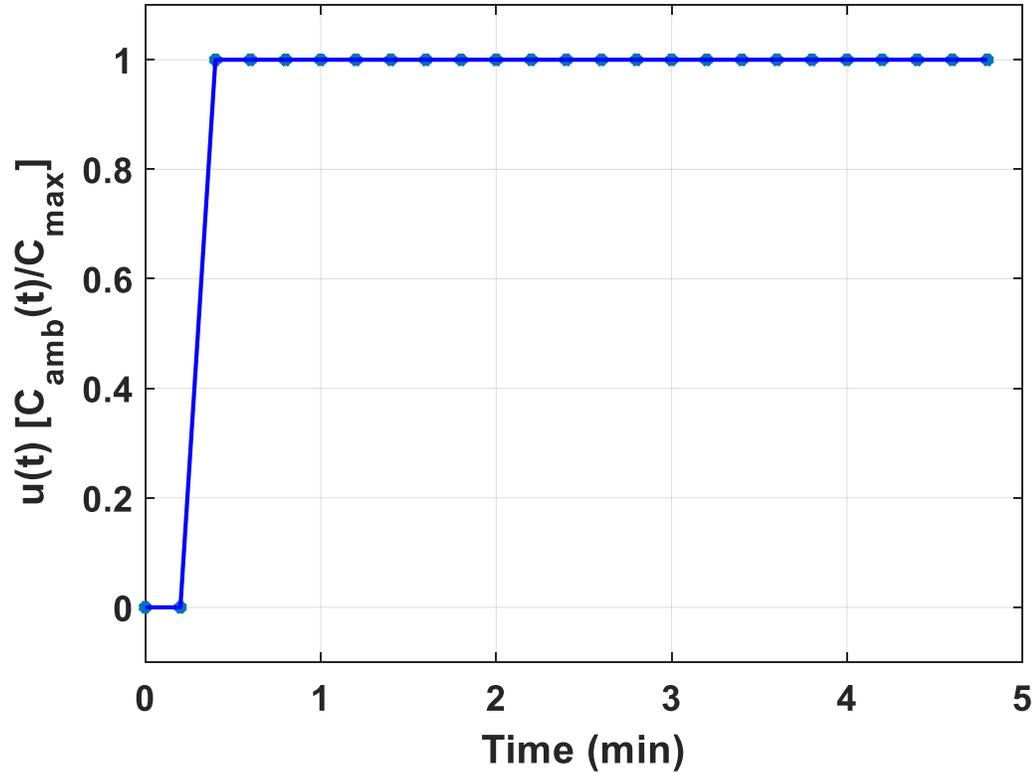


Figure 3.3: Concentration versus time profile of the measurement system used to collect the data analyzed in this dissertation.

3.5.2 Concentration-Dependent Sensitivity

As mentioned earlier, the characteristic value of sensitivity, aK_{p-w} is needed to accurately determine the concentration of the analyte(s) from a given coated SH-SAW sensor response. The slope of the calibration curve (plot of equilibrium frequency shift, f_{∞} versus equilibrium ambient concentration, C_{eq}) can be used to compute the sensitivity of the coated SH-SAW sensor. Ideally, it is desired for the relationship between f_{∞} and C_{eq} to be linear, so that the sensitivity, aK_{p-w} determined from the slope of the linear line is a constant,

$$aK_{p-w} = \frac{f_{\infty}}{C_{eq}}. \quad (3.20)$$

If the sensitivity, aK_{p-w} is constant, then eq. (3.5) can be used to compute the concentration of the analyte using the measured or estimated value for f_{∞} . However, linear relationship between f_{∞} and C_{eq} is only observed for small concentration range of BTEX compounds. For a larger concentration range, the relationship between f_{∞} and C_{eq} is usually nonlinear. As an example, Figure 3.4 shows the plot of equilibrium frequency shift, f_{∞} versus maximum equilibrium ambient concentration, C_{eq} for concentrations of ethylbenzene between 0 ppm to 70 ppm. As can be seen from Figure 3.4, linear relationship between f_{∞} and C_{eq} is observed for ethylbenzene concentrations between 0 ppm to 40 ppm and nonlinear relationship between f_{∞} and C_{eq} is observed for ethylbenzene concentrations between 0 ppm to 70 ppm. The nonlinear relationship between f_{∞} and C_{eq} for a larger concentration range will cause the sensitivity to vary based on the concentration of the analyte (i.e. non-constant sensitivity). In such cases, it is not possible to extract the concentration for the analyte using eq. (3.5). Therefore, a more general approach that could be utilized for a larger concentration range with non-constant sensitivity is desired.

One simple approach that could be employed is to express the nonlinear relationship between f_{∞} and C_{eq} (for a large concentration range) using a mathematical function, $F(.)$ that adequately captures the nonlinearity between f_{∞} and C_{eq} ,

$$f_{\infty} = F(C_{eq}). \quad (3.21)$$

Based on the relationship between f_{∞} and C_{eq} , it is preferable to utilize analytic functions such as exponential function or n^{th} order polynomial function to express the nonlinear relationship between f_{∞} and C_{eq} . Once the function of eq. (3.21) is determined using the calibration data for the required large concentration range of the analyte(s), the inverse of the function of eq. (3.21) can then be used to extract the concentration of the analyte(s) from the sensor response using the measured or estimated value for f_{∞} ,

$$C_{eq} = F^{-1}(f_{\infty}). \quad (3.22)$$

Note that eq. (3.22) is a generalized equation that can be used to extract the concentration of analyte(s) even from a highly nonlinear calibration curve.

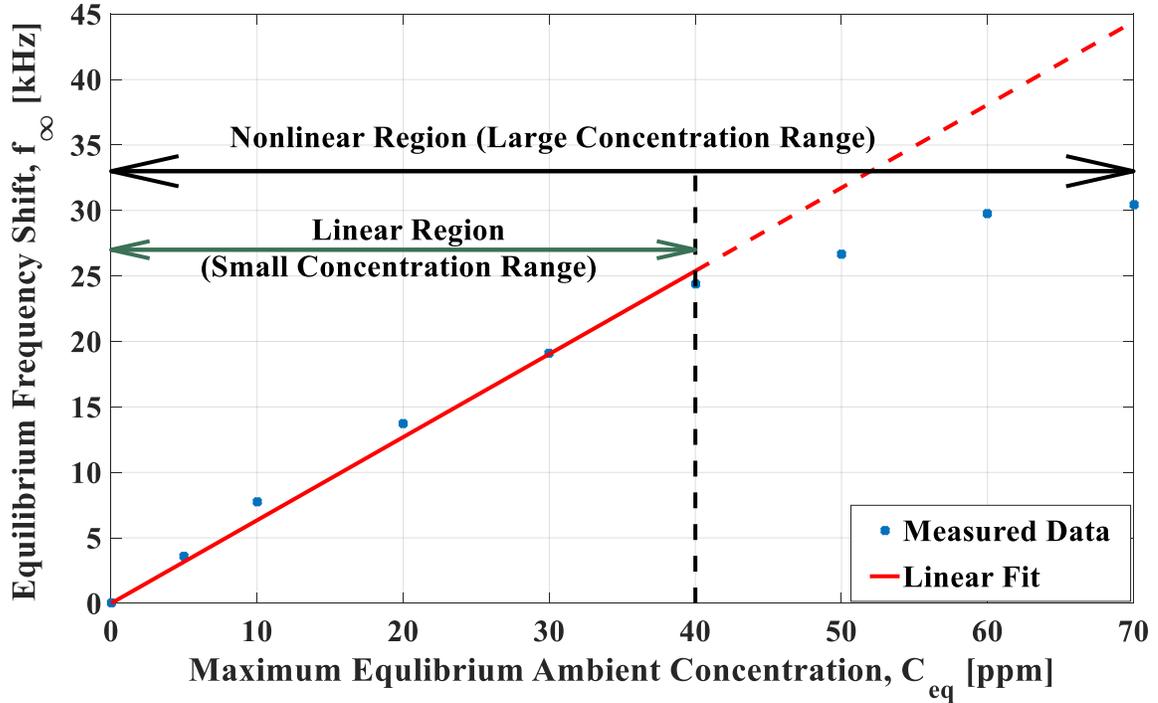


Figure 3.4: Plot of equilibrium frequency shift, f_{∞} versus maximum equilibrium ambient concentration, C_{eq} for concentration of ethylbenzene between 0 ppm to 70 ppm. Note that values for f_{∞} were extracted from measured sensor response data for different C_{eq} collected using SH-SAW sensor coated with $0.6 \mu\text{m}$ PECH. Linear and nonlinear region are indicated in the figure. Also shown in the figure is the linear fit (red line) for the linear region of the plot where the slope of the line is the constant sensitivity, aK_{p-w} .

For example, the calibration data of Figure 3.4, can be adequately well-fitted using exponential function of the form,

$$f_{\infty} = F(C_{eq}) = \alpha(1 - e^{-\beta C_{eq}}), \quad (3.23)$$

where α and β are constant parameters that can be determined by fitting the calibration data of Figure 3.4 with eq. (3.23). The exponential fit along with the calibration data is shown in Figure 3.5. In this case, once the equilibrium frequency shift, f_{∞} was measured

or estimated from the sensor response, the concentration of the analyte can be determined using the following equation,

$$C_{eq} = \frac{-\ln\left(1 - f_{\infty}/\alpha\right)}{\beta}. \quad (3.24)$$

In summary, for the large concentration range of the analyte(s) where the calibration data is usually nonlinear, the nonlinear relationship between f_{∞} and C_{eq} can be expressed using an analytic function. The function can then be used to directly extract the concentration of the analyte(s) from the sensor response using the measured or estimated value of f_{∞} . Note that this approach is highly dependent on the chosen mathematical function to express the nonlinearity between f_{∞} and C_{eq} . If a suitable function is properly selected, the approach is capable of yielding highly accurate results for the extracted concentration of the analyte(s). This approach can also be adopted to deal with multi-analyte responses as long as the analytes in the sample are chemically inactive to each other (i.e. the analytes do not interact with each other).

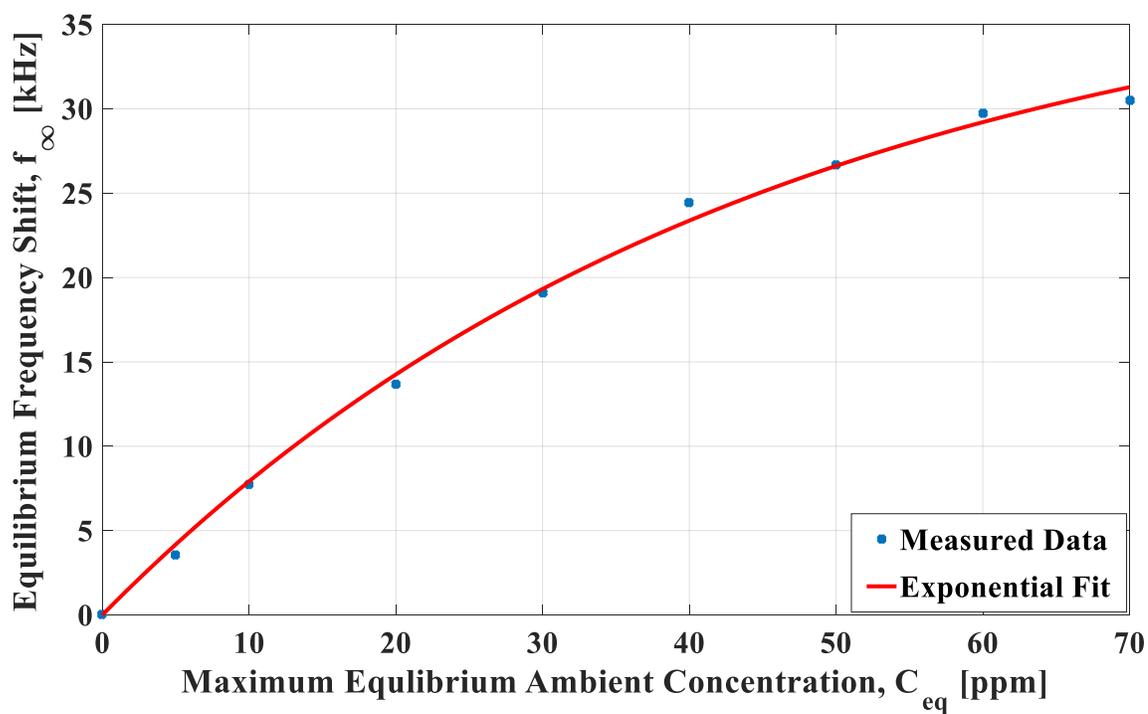


Figure 3.5: Exponential fit (red curve) to calibration data (blue dots) for ethylbenzene with concentration between 0 ppm to 70 ppm (large concentration range). The calibration data were collected using SH-SAW sensor coated with 0.6 μm PECH.

4 ESTIMATION-THEORY-BASED SENSOR SIGNAL PROCESSING

4.1 Introduction

In this chapter, the sensor signal processing techniques used for the detection and quantification of chemical samples for four different cases are discussed. It is noted that these techniques could be utilized for the detection and quantification of any chemical analytes, provided that the characteristic response time constants and sensitivities of the analytes in the sample are known for the selected coating. In this dissertation, the sensor signal processing techniques will be demonstrated for the detection and quantification of BTEX compounds. The four different cases discussed are as following:

- Case 1: Detection and quantification of single analyte from sensor response to solution containing any arbitrary single analyte.
- Case 2: Detection and quantification of binary mixtures from sensor response to solution containing any arbitrary binary mixtures.
- Case 3: Identification and quantification of various analytes in a sample mixture containing n different analytes.
- Case 4: Detection and quantification of various analyte targets in the presence of interferences in a chemical sample.

The signal processing techniques for all those cases are based on estimation theory because it offers various advantages, including near real-time data processing and

minimal computational and memory requirements for real-world implementations. Specifically, as indicated in Chapter 2, all signal processing techniques proposed in this dissertation are based on Bank of Kalman filters (BKFs) and/or exponentially weighted recursive least squares estimation (RLSE). Also, it is recalled here that the estimation-theory-based signal processing techniques are developed based on the coated SH-SAW sensor response models and their corresponding state-space models discussed in Chapter 3.

The sensor signal processing techniques proposed for Case 1 and 2 are based on BKFs. For Case 3, sensor signal processing based on multi-stage exponentially weighted RLSE is proposed, which will be implemented in series (multiple stages) in order to obtain a more accurate identification and quantification result. Finally, for Case 4, a two-step sensor signal processing technique based on both exponentially weighted RLSE and BKFs is proposed. In the two-step signal processing, an initial estimation of the unknown parameters is performed using exponentially weighted RLSE, which is then used as input for the second step using BKFs to refine the estimation results. More details on the implementation of the proposed techniques will be discussed in the subsequent sections. All the proposed techniques will be tested using experimental data obtained from chemical samples which contains BTEX compounds.

4.2 Online Detection and Quantification of Single Analyte in Solutions Containing any Arbitrary Single Analyte Using Bank of Kalman Filters

In this section, a novel approach based on BKFs for the online detection and quantification of a single analyte from an unknown single analyte sensor response is

discussed. The goal here is to identify and quantify the analyte which is most likely to have caused the measured sensor response using SH-SAW devices. The proposed approach will be demonstrated for detection and quantification of single analytes, in the present case, benzene, toluene, and chemical isomers ethylbenzene and xylenes.

The single analyte sensor response model has already been discussed in Chapter 3 and its state-space model can be obtained directly from eq. (3.13) with the substitution of $n = 1$,

$$x_{k+1} = (1 - S) x_k + S u_{s,k} = A x_k + B u_{s,k}, \quad (4.1a)$$

$$y_k = \Delta f_k = f_\infty x_k + w_k = C x_k + w_k, \quad (4.1b)$$

where, $S = \frac{T_s}{\tau}$ is the sorption/desorption rate constant of the analyte. For the implementation of BKF to identify and quantify the analyte in real-time, the parameters, S which is related to the response time constant, τ of the analyte, and the equilibrium frequency shift, f_∞ which is related to the sensitivity, aK_{p-w} of the analyte will be treated as the unknown parameters that need to be identified based on the measured sensor response. Note that for the investigated coatings, the parameter S , like the response time constant of the analyte, will be unique for the analytes of interest, i.e. benzene, toluene, and chemical isomers ethylbenzene and xylenes (refer to Tables 3.1 and 3.2 from the previous chapter). Hence, S can be used to identify the analyte that is most likely to have caused the sensor response. Subsequently, f_∞ can be used to quantify the analyte using eq. (3.5) based on the sensitivity, aK_{p-w} of the identified analyte. Note that aK_{p-w} is

also unique for the analytes of interest and the selected coating (refer to Tables 3.1 and 3.2 from the previous chapter). The implementation of BKF requires these unknown parameters to be suitably quantized into a finite set. Based on the knowledge of possible analytes that can cause the sensor response, S can be quantized to a finite number of grid points, $\{S_{benzene}, S_{toluene}, S_{ethylbenzene \& xylenes}\}$ and based on the possible equilibrium concentration range for the analyte, f_{∞} can be quantized to a finite number of grid points, $\{f_{\infty; 1}, f_{\infty; 2}, \dots, f_{\infty; N}\}$. Therefore, the unknown parameters vector, θ , which is formed based on all possible combinations of the discrete points of the unknown parameters S and f_{∞} , will belong to a finite number of grid points $\{\theta_1, \dots, \theta_N\}$. For each θ_j , the corresponding system matrices A_{θ_j} , B_{θ_j} and C_{θ_j} of eq. (4.1) can be computed and using the BKF algorithm as outlined in Table 2.4, the analyte most likely to have caused the sensor response can be identified and quantified in real-time. Note that this technique can be used as an alternative to a sensor array for single analyte identification and quantification. In fact, as opposed to a sensor array, the proposed approach has the advantages of not requiring a complex training data set and enabling real-time detection and quantification of the analyte based on the sensor response from only a single sensor device.

4.3 Online Detection and Quantification of Binary Mixtures in Solutions Containing any Arbitrary Binary Mixtures Using Bank of Kalman Filters

In this section, a novel approach based on BKFs for the online detection and quantification of binary mixtures from an unknown binary mixture sensor response is discussed. The objective here is, based on the measured coated SH-SAW sensor response

to binary mixtures samples, to identify and quantify the analytes which are most likely to have caused the measured response. The proposed approach is demonstrated here for various binary mixture responses of BTEX compounds. The state-space model for binary mixtures can be obtained directly from eq. (3.13) with the substitution of $n = 2$,

$$\begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+1}^{(2)} \end{bmatrix} = \begin{bmatrix} 1 - S_1 & 0 \\ 0 & 1 - S_2 \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} + \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} u_{s,k} = Ax_k + Bu_{s,k}, \quad (4.2a)$$

$$y_k = \Delta f_k = [f_{\infty,1} \quad f_{\infty,2}] \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} + w_k = Cx_k + w_k. \quad (4.2b)$$

In using BKF to identify and quantify the binary mixture of analytes in real-time, the parameters S_1 , S_2 , $f_{\infty,1}$, and $f_{\infty,2}$ can be treated as the unknown parameters that need to be determined based on the measured binary mixture response. All these unknown parameters have to be quantized to a finite number of grid points. From the knowledge of possible binary mixtures of analytes to have caused the sensor response, $\begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$ can be quantized to a finite number of grid points

$\left\{ \begin{bmatrix} S_{benzene} \\ S_{toluene} \end{bmatrix}, \begin{bmatrix} S_{benzene} \\ S_{ethylbenzene \& xylenes} \end{bmatrix}, \begin{bmatrix} S_{toluene} \\ S_{ethylbenzene \& xylenes} \end{bmatrix} \right\}$. The unknown parameters,

$\begin{bmatrix} f_{\infty,1} \\ f_{\infty,2} \end{bmatrix}$ can be quantized to a finite number of grid points based on the expected

concentration range of the identified analytes. Therefore, the unknown parameters vector, θ , which is formed based on all possible combinations of the discrete points of all the unknown parameters, will belong to a finite number of grid points $\{\theta_1, \dots, \theta_N\}$. Since θ

belongs to a discrete set, the state-space model of eq. (4.2) and the BKF algorithm as outlined in Table 2.4 can be utilized to identify and quantify the binary mixture of analytes that is most likely to have caused the sensor response in real-time. Note that, based on each θ_j , the corresponding system matrices A_{θ_j} , B_{θ_j} and C_{θ_j} of eq. (4.2) can be computed. The proposed approach is a viable alternative for binary mixture identification and quantification using a sensor array and offers similar advantages as mentioned in the previous section. It is also important to note that, direct extension of this technique for the detection and quantification of multi-analyte mixtures (i.e. mixtures of analytes with more than two analytes) is rather ineffective due to the exponential rise in the implementation complexity of the technique. Note that as the number of analytes required to be detected in the sample increases, the number of Kalman filters in the bank (needed for accurately identifying and quantifying the analytes) increases as well, thus, increasing the implementation complexity of the technique based on BKFs for multi-analyte detection and quantification. Alternative approaches for detection and quantification of multiple analytes in a mixture are discussed in the subsequent sections.

4.4 Online Identification and Quantification of Various Analytes in Multi-Analyte Mixtures Using Multi-Stage Exponentially Weighted RLSE

In this section, a sensor signal processing technique for online identification and quantification of various analytes in multi-analyte mixtures (including single analyte samples) using only the measured sensor response from a single polymer coated SH-SAW sensor device is discussed. The objective is to develop a technique that can be used for the identification and quantification of n analytes in a mixture that are most likely to

have caused the measured sensor response, provided the characteristic response time constants and sensitivities of all possible analytes in the mixture are known. As indicated in Chapter 3, this is achieved by characterizing all possible analytes to which the selected coated sensor responds.

In order to achieve the aforementioned objective, a sensor signal processing technique based on multi-stage exponentially weighted RLSE that utilizes two sensing parameters (i.e. response time constant and sensitivity) is proposed. The initial stages of exponentially weighted RLSE will be used to eliminate analytes that are erroneously identified as present in the mixture (i.e. analytes that are not present or analytes with concentrations close to zero) and the final stage of exponentially weighted RLSE with the corresponding sensor response model representing the analytes present in the mixture will be used to obtain a more accurate quantification result of the analytes. In this technique, exponentially weighted RLSE is used to estimate the equilibrium frequency shifts of the various analytes in the mixture, $f_{\infty,i}$ that collectively produced the measured total frequency response. Based on the estimated $f_{\infty,i}$, the concentrations of the analytes assumed to be present in the mixture can be determined using eq. (3.5) and sensitivities, $aK_{p-w,i}$ of the analytes. In the proposed approach, initial estimation of $f_{\infty,i}$ is performed using exponentially weighted RLSE and the n -analyte sensor response model. If erroneous analyte detections are identified from the initial estimation step, the estimation process is repeated using the appropriate sensor response model (i.e. sensor response model for analytes less than n) and exponentially weighted RLSE to obtain a more accurate analyte quantification result. This is because the estimated concentrations associated with erroneously detected analytes can greatly influence the quantification

result obtained for the analytes that are present in the mixture. Therefore, by repeating the estimation using exponentially weighted RLSE with the corresponding sensor response model for only the analytes identified to be in the mixture, a more accurate analytes quantification result can be obtained. Note that erroneous analyte detections can be identified by comparing the estimated concentrations of the analytes to the detection limit of the chosen sensor polymer-coated sensor platform for that analyte. If the estimated concentration of a particular analyte is negative or lower than a threshold value that was selected to be close to the detection limit for that analyte, the concentration of the analyte can be assumed to be zero.

A block diagram depicting the proposed sensor signal processing technique is illustrated in Figure 4.1. In summary, for the proposed approach, the initial $(M - 1)$ exponentially weighted RLSE stages along with appropriate sensor response models is used to identify erroneously detected analytes in the mixture and finally, using the M^{th} exponentially weighted RLSE and the appropriate sensor response model, the concentrations of the analytes that are present in the mixture are determined based on the estimated equilibrium frequency shifts of the analytes. Note that for as long as erroneous analyte detections are identified, the estimation process will be repeated using exponentially weighted RLSE and the appropriate sensor response model to obtain a more accurate analyte quantification result.

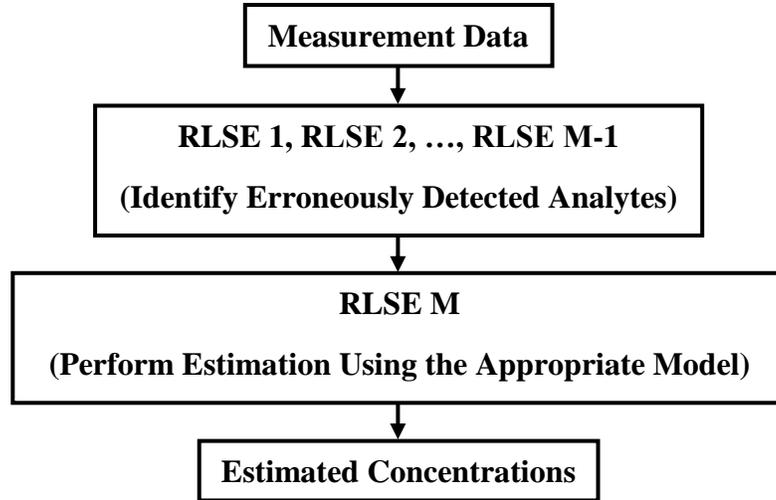


Figure 4.1: Multi-stage exponentially weighted RLSE for online identification and quantification of various analytes in multiple analyte mixtures.

The implementation of exponentially weighted RLSE to estimate the unknown parameters $f_{\infty,i}$ is demonstrated using the state-space form of the general multi-analyte sensor response model of eq. (3.13). For the implementation of exponentially weighted RLSE to estimate the unknown parameters $f_{\infty,i}$, eq. (3.13b) can be rearranged into the following form,

$$y_k = \Delta f_k = [x_k^{(1)} \quad \dots \quad x_k^{(n)}] \begin{bmatrix} f_{\infty,1} \\ \vdots \\ f_{\infty,n} \end{bmatrix} + w_k = H_k \theta + w_k, \quad (4.3)$$

where $H_k = [x_k^{(1)} \quad \dots \quad x_k^{(n)}]^T$ is a vector of known signals (regressor) which can be determined at every time step, k using eq. (3.13a), and $\theta = [f_{\infty,1} \quad \dots \quad f_{\infty,n}]^T$ is a vector of unknown parameters that need to be estimated. Note that the form of eq. (4.3) is similar to that of eq. (2.15) in Chapter 2. Therefore, the exponentially weighted RLSE

algorithm as outlined in Table 2.3 can be used to estimate the unknown parameters, $f_{\infty,i}$ recursively.

In this dissertation, the effectiveness of the proposed approach is tested using the measured responses of polymer coated SH-SAW sensors to various multi-analyte mixtures of benzene, toluene, ethylbenzene and xylenes, and 1,2,4-trimethylbenzene (TMB). Basically, the approach will be demonstrated for identification and quantification of single, binary, ternary or quaternary mixtures of BTEX compounds and TMB that are most likely to have caused a given measured sensor response. The threshold values close to the detection limits of BTEX compounds and TMB that were used for the implementation of this approach are listed in Table 4.1. The proposed approach can be used as a viable alternative to chemical sensor arrays for multi-analyte detection and quantification, offering the advantages of not requiring a complex training data set and allowing near real-time detection and quantification of the analytes. In direct comparison to the sensor signal processing techniques based on BKF's discussed in the preceding sections, the approach based on multi-stage exponentially weighted RLSE discussed in this section is far superior. This is because the technique discussed in this section is capable of detecting and quantifying n analytes in a mixture, provided the characteristic response time constants and sensitivities of the analytes are known. However, as mentioned earlier, this cannot be achieved using the technique based on BKF's due to the exponential rise in the implementation complexity of the technique for the multi-analyte detection and quantification.

Table 4.1: Threshold values (approximate detection limits) of BTEX compounds and 1,2,4-trimethylbenzene (TMB) in water for SH-SAW sensors coated with either 0.6 μm PECH or 0.8 μm PIB.

Analyte	Detection Limit (ppb)
Benzene	100
Toluene	50
Ethylbenzene & Xylenes	30
1,2,4-trimethylbenzene	10

4.5 Near Real-Time Detection and Quantification of Multiple Target Analytes in a Chemical Sample Also Containing Various Interferents

In this section, a novel near real-time signal processing technique for the detection and quantification of target analytes (i.e. BTEX compounds in this dissertation) in the presence of interferents is discussed. The objective of the proposed signal processing technique is to facilitate the detection and accurate quantification of the target analytes in the presence of noise and interferents. In Chapter 3, two examples of state-space models (i.e. four-analyte state-space model, eq. (3.15) and five-analyte state-space model, eq. (3.17)) for the detection and quantification of BTEX compounds in water in the presence of interferents were presented. Also, as indicated in Chapter 3, in those state-space models, the unknown parameters that need to be estimated for near real-time detection and quantification of target analytes are the equilibrium frequency shifts, $f_{\infty,i}$ of the analytes. This is because, by estimating $f_{\infty,i}$, the corresponding concentration of analytes can be determined using eq. (3.5) based on the sensitivity, $aK_{p-w,i}$ of the analytes. This means that, by estimating $f_{\infty,i}$ in near real-time, the target analytes can also be identified and quantified in near real-time. In order to estimate the unknown parameters in near real-time, suitable estimation-theory-based techniques are selected for sensor signal

processing based on the formulated state-space models. Specifically, based on our investigations, an estimation-theory-based technique comprised of exponentially weighted RLSE and bank of KFs (two-step processing) was selected as the most suitable signal processing approach for accurate multi-analyte detection and quantification [87].

In the two-step processing, an initial estimation of the unknown parameters is performed using exponentially weighted RLSE, which is then used as input for the second step using BKFs to refine the estimation results. For the implementation of the two-step processing, the unknown parameters vector (i.e. $\theta^4 = [f_{\infty,1} \ f_{\infty,2} \ f_{\infty,3} \ f_{\infty,4}]^T$ for the four-analyte model and $\theta^5 = [f_{\infty,1} \ f_{\infty,2} \ f_{\infty,3} \ f_{\infty,4} \ f_{\infty,5}]^T$ for the five-analyte model) is first formed. Then, for the application of the exponentially weighted RLSE algorithm to estimate the unknown parameters vector, as outlined in Table 2.3, eq. (3.15b) for the four-analyte model and eq. (3.17b) for the five-analyte model are rearranged into the form of eq. (2.15). For the four-analyte model, the rearrangement will result in,

$$y_k = \Delta f_k = \begin{bmatrix} x_k^{(1)} & x_k^{(2)} & x_k^{(3)} & x_k^{(4)} \end{bmatrix} \begin{bmatrix} f_{\infty,1} \\ f_{\infty,2} \\ f_{\infty,3} \\ f_{\infty,4} \end{bmatrix} + w_k = H_k^4 \theta^4 + w_k, \quad (4.4)$$

where $H_k^4 = [x_k^{(1)} \ x_k^{(2)} \ x_k^{(3)} \ x_k^{(4)}]^T$ is a vector of known signals (regressor) which can be determined at every time step, k using eq. (3.15a). For the five-analyte model, the rearrangement will result in the same form as eq. (4.4) but with H_k^4 and θ^4 replaced with H_k^5 and θ^5 respectively, where $H_k^5 = [x_k^{(1)} \ x_k^{(2)} \ x_k^{(3)} \ x_k^{(4)} \ x_k^{(5)}]^T$ is a vector of known signals (regressor) which can be determined at every time step, k using eq.

(3.17a). Note that the form of eq. (4.4) is similar to that of eq. (2.15) in Chapter 2. Thus, the exponentially weighted RLSE algorithm as outlined in Table 2.3 can be used to estimate the unknown parameters vector, θ recursively.

By using the exponentially weighted RLSE algorithm, an initial estimate for the unknown parameters vector, θ^n (where $n = 4$ for the four-analyte model and $n = 5$ for the five-analyte model) can be obtained. Next, based on the results from exponentially weighted RLSE, the unknown parameters vector, θ^n can be suitably quantized into a finite number of grid points $\{\theta_1^n, \dots, \theta_N^n\}$ with assumed *a priori* probability for each θ_j^n (where subscript $j = 1, \dots, N$ (total number of filters)). Note that θ^n are quantized into finite number of grid points based on the estimated equilibrium frequency shifts, $f_{\infty,i}$ obtained from exponentially weighted RLSE. The values for $f_{\infty,i}$ are increased and decreased by 20% (or by a user defined value) to define limits for a range of possible equilibrium frequency shifts for the analytes. Within these limits, by uniformly quantizing the equilibrium frequency shifts, a finite number of grid points $\{\theta_1^n, \dots, \theta_N^n\}$ are obtained. Since θ^n belongs to a discrete set, the state-space model of eq. (3.15) for the four-analyte model and eq. (3.17) for the five-analyte model, and the BKF algorithm as outlined in Table 2.4 can be used to obtain the best estimate for the unknown parameters vector, θ^n . Note that for each θ_j^n , the corresponding system matrices $A_{\theta_j^n}$, $B_{\theta_j^n}$ and $C_{\theta_j^n}$ can be computed. By following all the steps in the BKF algorithm as outlined in Table 2.4, the best (i.e., most likely) case for the unknown parameters vector, θ_j^n can be determined by finding the case that produces the highest *a posteriori* probability, $p(\theta_j^n | Y_k)$.

In an effort to obtain more reliable and accurate estimation results using the two-step processing for the detection and quantification of the target analytes in the presence of interferents, both sorption and desorption data including the error range in the measured mean response times as listed in Tables 3.1 and 3.2 are utilized. The sensor signal processing procedure used to obtain the final estimated concentrations of BTEX compounds in the presence of interferents using the data collected from only a single polymer-coated SH-SAW device is summarized in Figure 4.2. By incorporating desorption transients (which are often more sensitive to energies of desorption of analytical targets) in the signal processing procedure, more information about the analyte/polymer interactions can be obtained. The proper use of this additional information can potentially result in a more accurate estimated concentration for BTEX compounds compared to only using sorption transients alone. Moreover, utilizing both sorption and desorption data in the signal processing can reduce the error in the estimation due to measurement noise, making this procedure highly tolerant of such noise. In summary, the utilization of both sorption and desorption data can improve the accuracy of the extracted concentrations and the reliability of chemical speciation.

As can be seen from Figure 4.2, the sorption data are used to estimate benzene and toluene (BT) concentrations whereas the desorption data are used to obtain an estimate for the combined concentrations of ethylbenzene and xylenes (EX). This is because based on the initial investigations using the proposed signal processing technique, it was observed consistently that using the sorption data always results in more accurate estimates of BT concentrations whereas using the desorption data always results in more accurate estimates for the combined concentrations of EX. The observed scenario

might be due to the analyte's general interaction with (and diffusion through) the polymer coatings. During analyte sorption, compounds with shorter response times (BT) are absorbed into a relatively clean coating, similar to the case of single analyte detection that was used to determine individual response time constants and sensitivities. The remaining compounds (EX, interferents) are then absorbed into a coating that already contains BT, thus introducing a slight error in sorption time constant. The reverse will be true for analyte desorption. Therefore, in the proposed signal processing procedure, the sorption data are used to estimate BT concentrations whereas the desorption data are used to obtain an estimate for the combined concentrations of EX. For both sorption and desorption data, the algorithm obtains the estimates using the two-step processing as depicted in Figure 4.2. Note that this technique is independent of the initial values of the unknown parameters, $f_{\infty,i}$. Therefore, any concentration range of BTEX compounds can be analyzed and the results can be obtained in near real-time. Note that this approach can, in general, be used for the detection and quantification of multi-analyte mixtures in the presence of interferents if the sensor parameters such as characteristic response time constants and sensitivities associated with the target analytes are known.

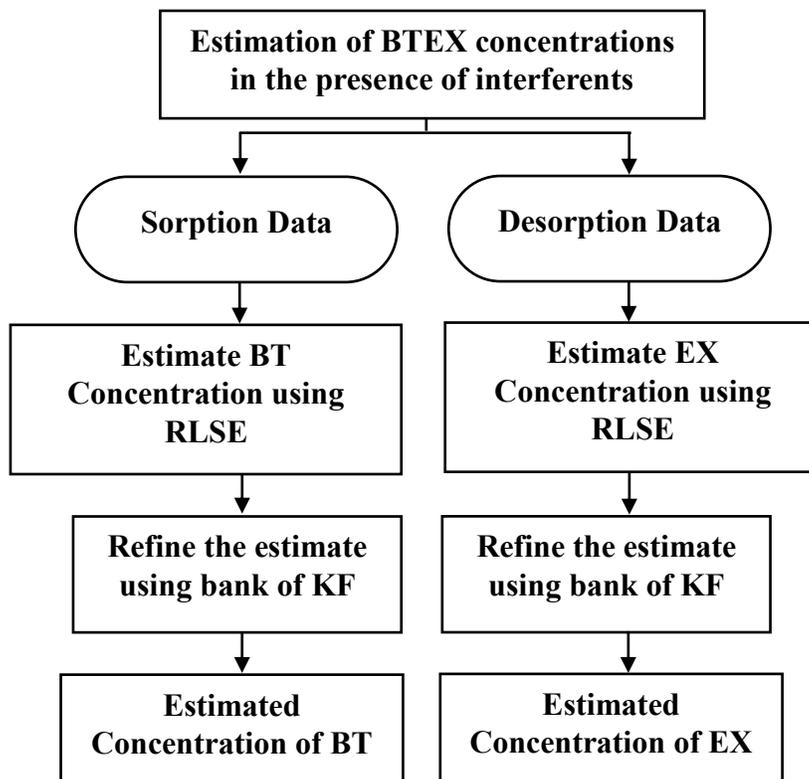


Figure 4.2: Estimation-theory-based sensor signal processing procedure that utilizes both sorption and desorption data for detection and quantification of BTEX analytes in the presence of interferents. “BT” and “EX” refers to the respective BTEX analytes.

5 COATED SH-SAW SENSOR DATA ACQUISITION

5.1 Introduction

The signal processing techniques discussed in Chapter 4 are tested and validated using experimental sensor response data collected using a polymer-coated shear horizontal surface acoustic wave (SH-SAW) sensor platform. In this chapter, the specifics of this SH-SAW device are discussed. The physics of the SH-SAW devices are briefly discussed. The SH-SAW sensor platform has been chosen for the sensing experiments because it has been shown in [47, 50, 64, 65] that it has the potential of being used as an in-situ chemical sensor for direct aqueous phase detection of BTEX compounds. Moreover, the information on the types of the polymer films used in the present work to detect the target analytes is also given. Note that these polymer films were coated for the sensor platform to enhance the sensitivity and to provide partial selectivity for the target analytes. Furthermore, the details of the measurement setup and data acquisition are also discussed in this chapter. It is noted that all data acquisitions for the selected coated sensors were done by various research team members of the Microsensor Research Laboratory, Marquette University.

5.2 Shear Horizontal Surface Acoustic Wave (SH-SAW) Devices

The shear horizontal surface acoustic wave (SH-SAW) devices are a specific type of sensor platform that belong to the family of acoustic wave sensors that can be used for

direct liquid-phase sensing applications, where the device is in direct contact with the chemical solutions of interest. SH-SAW devices are fabricated on rotated Y-cut LiTaO₃, a piezoelectric substrate cut and orientation, which supports shear horizontal surface acoustic wave (SH-SAW) [8, 88], thus making the device suitable for liquid-phase sensing applications.

There exist only selected few acoustic wave devices that can operate efficiently in aqueous environments, and they include the thickness shear mode (TSM), shear horizontal acoustic plate mode (SH-APM), flexural plate wave (FPW) and SH-SAW devices. This is because, the generation of compressional waves in the liquid (that are associated with ‘ordinary’ SAW or Rayleigh SAW) is absent with those devices. Therefore, these devices cannot radiate compressional waves in the liquid, hence resulting in a small amount of acoustic energy being dissipated into the liquid. Among the different types of acoustics wave sensors for liquid-phase sensing, SH-SAW devices are often preferred because SH-SAW devices have high operational frequency and surface waves are more sensitive to surface perturbations. Other interesting features of SH-SAW devices include high quality factor, compactness, robustness and low fabrication cost. All these attractive features of SH-SAW devices will enable the development of a small, portable, cost-effective sensor system for field applications.

SH-SAW and surface acoustic wave (SAW) sensing platforms are structurally similar. However, despite their structural similarity, SH-SAW devices often propagate slightly deeper (1 ~ 5 wavelengths) within the substrate [83, 89]. As a result, SH-SAW’s sensitivity to surface perturbations are reduced. In order to increase SH-SAW devices sensitivity to surface perturbations, a thin guiding layer is often deposited on the device

surface. The purpose of the thin guiding layer is to trap the energy near the surface of the device so that the device sensitivity to surface perturbations can be increased significantly and enable the device to be operated efficiently in liquid. The amount of acoustic energy trapped in the guiding layer depends on the thickness of the layer. Empirical results show that acoustic energy trapped increases with the thickness up to an optimum value.

Therefore, coated SH-SAW sensors are typically modeled as a multilayered structure [90]. The SH-SAW sensor structure used in this dissertation is commonly referred to as the three-layer structure as shown in Figure 5.1. The three-layer structure consists of the piezoelectric crystal substrate (LiTaO_3) with input and output interdigital transducers (IDTs) arranged in a delay line configuration, a viscoelastic polymer layer of finite thickness, h and a liquid layer for transport of chemical samples. The purpose and characteristics of each layer is as following:

- | | |
|--------------------------|--|
| Piezoelectric substrate: | <ul style="list-style-type: none"> – Convert the electrical signal into a mechanical signal (strain) (i.e. the acoustic wave) and vice versa. – Serve as a support for the entire device. |
| Polymer layer: | <ul style="list-style-type: none"> – Have lower shear wave velocity than the substrate (precondition for the confinement of the SH-SAW to the surface). – Can also serves as both a wave-guiding layer and a chemically sensitive layer [2, 91, 92]. |

- Liquid layer:
- Assumed to be a Newtonian fluid (because the solutions being tested are dilute aqueous solutions).
 - Used for transport of chemical species.

Note that since the polymer layer is of thickness, h , the polymer layer is considered as a finite layer while the substrate and the liquid layer are considered as semi-infinite layers [91].

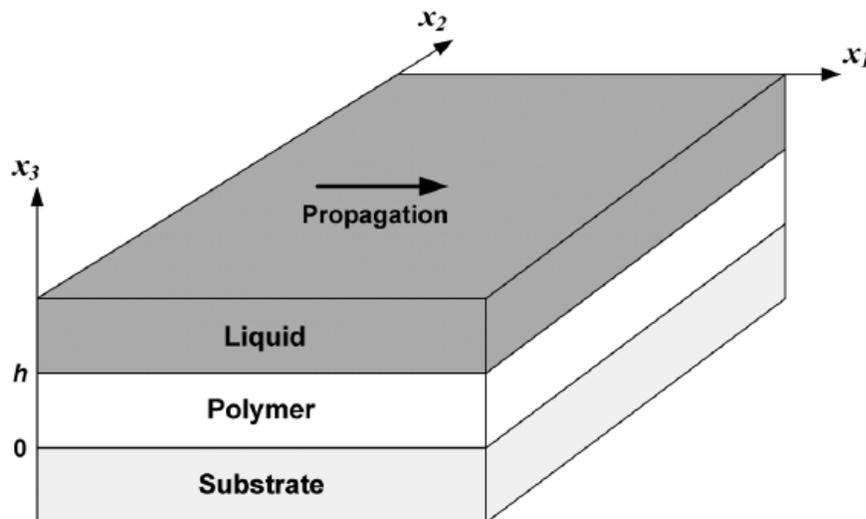


Figure 5.1: Three-layer structure and coordinate system. The guided SH-SAW will propagate in the x_1 direction, x_2 is in the direction of the acoustic wave particle displacement, and x_3 is normal to the sensing surface [92].

The basic delay line configuration of a guided SH-SAW sensor device used in the experiments is shown in Figure 5.2. In Figure 5.2, only one delay line is shown for simplicity but for actual data acquisition a dual delay line configuration is used. In a dual delay line configuration, one line serves as a sensing line and the other as a reference line. This design enables for the common environmental interactions (such as temperature and

pressure) producing responses from both lines to be eliminated by subtraction (i.e. differential measurement) [26, 92]. In order to eliminate acoustoelectric interactions with the load, a thin metal layer is used between the two IDTs (input and output IDTs) to create an electric short. As a result, only sensing caused by mechanical loading (i.e. changes in the mechanical properties of the polymer coating) is monitored.

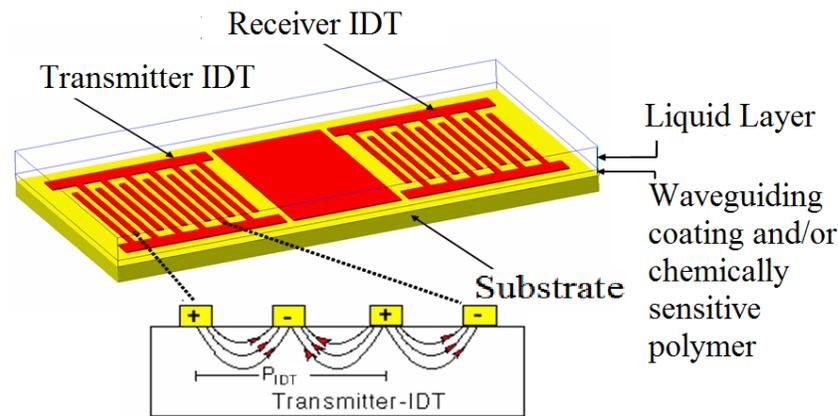


Figure 5.2: The delay line configuration of a guided SH-SAW sensor device. The device is a two-port device and for simplicity, only one delay line is shown [83].

5.3 Measurement Setup and Data Acquisition

As indicated earlier, the sensor response data analyzed in this work were collected using the 36° YX-LiTaO₃ guided SH-SAW device as the sensing platform [47, 64, 65]. The SH-SAW devices were fabricated with 10/80 nm-thick Ti/Au interdigital transducers (IDTs) using a multielectrode design that produces an operating frequency for the third harmonic SH-SAW of 103 MHz for polymer-coated devices [47, 64, 65]. As dual-delay-line configuration was used where both delay lines include a metalized path between the IDTs. The sensing and reference line were coated with different types of polymer

coatings. The sensing line was coated with sorbent polymer coatings that respond to the analytes of interest. They include poly(ethyl acrylate) (PEA), poly(epichlorohydrin) (PECH), and poly(isobutylene) (PIB), all purchased from Sigma-Aldrich, St. Louis, MO. These polymers were deposited on the sensing line from solution in toluene (PEA) or chloroform (PECH, PIB) by spin coating and baking for 15 min at 55°C. This will result in thicknesses of 1.0 μm for PEA, 0.6 μm for PECH and 0.8 μm for PIB. Note that the baking step is crucial to ensure repeatability of the sensor responses. On the other hand, the reference line was coated with poly(methyl methacrylate) (PMMA) (purchased from Scientific Polymer Products, Ontario, NY) and baked for 120 min at 180 °C, resulting in a glassy, nonsorbent coating, which will be chemically insensitive (i.e. does not absorb appreciable amounts of analyte) at the concentration ranges of interest. The chemical structure and basic properties of the four polymers utilized for data acquisition is shown in Figure 5.3 and Table 5.1, respectively. Note that the basic properties of these polymers as listed in Table 5.1, served as a guide for the selection of these coatings for the detection of BTEX compounds. In particular, the glass transition temperature, T_g of the chosen polymer coatings influence the rate of sensor responses to analytes of interest. For most applications, it is desirable that the chosen polymer coatings have a T_g that is below room temperature and/or the operating temperature of the sensor in order to promote rapid sorption/desorption process [2, 93-95].

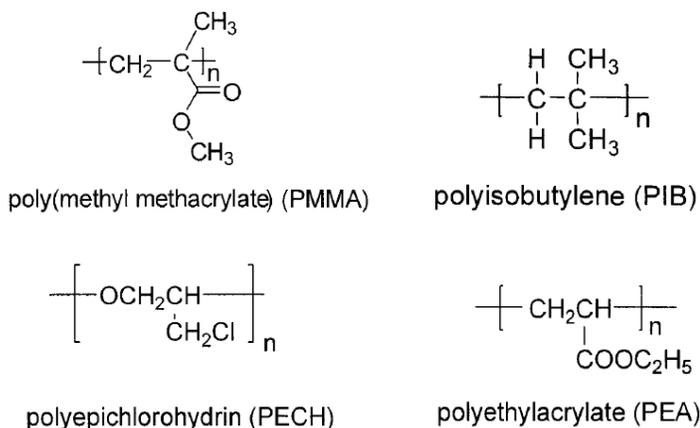


Figure 5.3: Chemical structure of PMMA, PIB, PECH, and PEA [83].

Table 5.1: Basic properties of four polymers (PIB, PECH, PEA, PMMA) [83, 96].

	PMMA	PIB	PECH	PEA
Density (g/cm³) at 25°C	1.19	0.92	1.36	1.12
Glass transition temperature, <i>T_g</i> (°C)	105	-73	-15 ~ -22	-21
Repeat unit	C ₅ H ₈ O ₂	C ₄ H ₈	C ₃ H ₅ ClO	C ₅ H ₈ O ₂
Uses	thermoplastic	elastomer	elastomer	elastomer
Monomer	methyl methacrylate	isobutylene	epichlorohydrin	ethyl acrylate

The target analytes, i.e. BTEX analytes used in the experiments were purchased from Sigma-Aldrich with purities of at least 98.5%. The BTEX analytes were diluted to various concentrations using either deionized (DI) water or groundwater. The groundwater and light non-aqueous phase liquid (LNAPL) samples used in the experiments were collected from actual groundwater monitoring wells in California. Commonly found interferents in groundwater were also tested in the experiments, including *n*-heptane, 1,2,4-trimethylbenzene, naphthalene, MTBE (methyl *tert*-butyl

ether), and ethanol. The interferents were all purchased from Sigma-Aldrich and had purities of $\geq 98\%$, except ethanol which was denatured and had $\geq 90\%$ purity. The concentrations of these samples are reported in parts per million (ppm) or parts per billion (ppb) by weight.

The measurement setup used to collect the sensor response data consists of a network analyzer (Agilent E5061B, Santa Clara, CA), a switch/control system (Agilent 34980A) to switch between the two SH-SAW delay lines on each device, a PC based HPVVE (Hewlett-Packard Visual Engineering Environment) program for data acquisition (insertion/device loss, frequency, phase and temperature), a measurement chamber and a liquid delivery system. The measurement chamber contains a flow cell with the coated guided SH-SAW device, chemical samples, reference sample, and 3-way valve. The liquid delivery system is comprised of a peristaltic pump (IDEX Ismatec Reglo Digital MS, Oak Harbor, WA) and a waste container. The schematic of the measurement setup used for the data collection in the Microsensor Research Laboratory is shown in Figure 5.4.

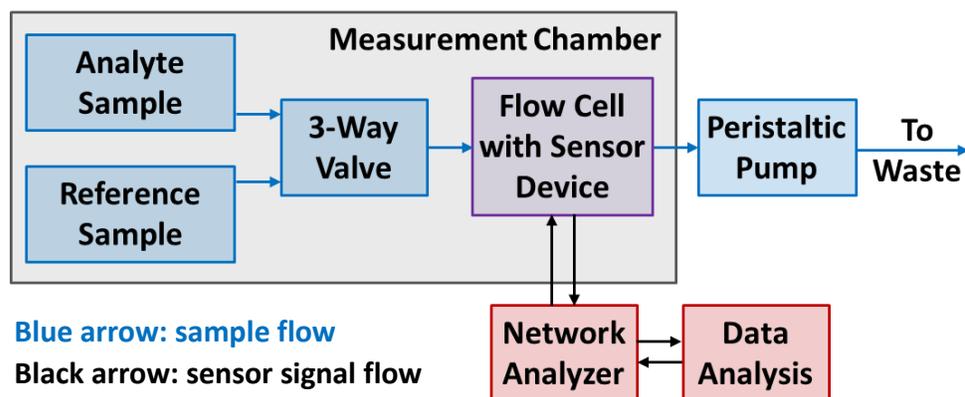


Figure 5.4: Schematic of the measurement setup used for the data collection in the Microsensor Research Laboratory [97].

In order to provide an independent measurement of BTEX concentrations for the aqueous LNAPL solutions, a portable GC-PID (gas chromatograph-photoionization detector) system (Defiant Technologies FROG-4000, Albuquerque, NM) was used. Note that the BTEX concentrations determined using GC-PID are subject to an average error of $\pm 7\%$ [98]. In some cases, analyte concentrations were further confirmed using GC-MS (gas chromatography–mass spectroscopy). Note that the LNAPL samples tested were prepared by placing the LNAPL above DI water in a separatory funnel for 3 days to create a saturated aqueous solution, which was further diluted with either DI water or groundwater for the respective experiments to yield concentrations of 1 ppm or less for each BTEX compound. PTFE tubing, PTFE valves, and PTFE sealed glass vials were used throughout the experiments to minimize the loss of volatile analytes, and for the same reason the headspace in the sample vials was kept negligible. As indicated above, actual sample concentrations as seen by the sensor were determined by the subsequent GC-PID measurements. The experiments were performed by placing the SH-SAW sensor inside a flow cell that was designed in the Microsensor Research Laboratory, Marquette University [82, 83]. A peristaltic pump was used to pump the solutions through the flow cell and to minimize the hydrodynamic forces from the flowing fluid, the solutions were pumped at a constant sample flow rate of 7 $\mu\text{L/s}$.

It should be noted that prior to the introduction of the samples containing the analyte(s) into the flow cell, a reference solution (DI water or groundwater) was drawn through the cell to obtain a stable baseline output signal. For groundwater samples, filtration was first performed to remove sediments and other physical interferents in the sample, before pumping these samples into the flow cell. After the sensor signal reached

the equilibrium response to the analytes, the reference solution was again pumped through the system to flush the flow cell. This latter step causes the analyte(s) to desorb from the polymer coating of the sensor. The process was repeated for different analyte samples and concentrations. Note that all the measurements were conducted at an approximately constant temperature of 22 ± 0.1 °C. The experiments were conducted using samples containing multi-analyte mixtures of BTEX compounds, trimethylbenzene (TMB), and (diluted) LNAPL samples. The tested LNAPL samples contain BTEX compounds as well as chemical and physical interferents.

All the experimental sensor response data collected using the sensor system described above exhibit some degree of baseline drift during the response and sometimes outlier points are also recorded probably due to the presence of bubbles or minute changes in the local concentration in the cell. Therefore, pre-processing had to be done first to correct the data for baseline drift and to eliminate any outlier points in the data before further signal processing. There are various baseline drift and outlier correction techniques that can be utilized for pre-processing. In this work, linear interpolation was used for baseline drift compensation. Online techniques for baseline and outlier corrections are discussed in [45, 46]. These online pre-processing techniques can be used in a field deployable smart sensor system.

6 ESTIMATION RESULTS AND DISCUSSION

6.1 Introduction

In this chapter, the estimation results obtained using the proposed sensor signal processing techniques for the various cases discussed in Chapter 4 are presented. The estimation results are obtained by testing and validating the sensor signal processing techniques using experimental sensor response data collected using polymer coated SH-SAW sensors with actual groundwater samples. The process of data acquisition has already been discussed in Chapter 5. It is also pertinent to note that the signal processing techniques discussed in Chapter 4 are applicable for the detection and quantification of any analytes, provided that the characteristic response time constants and sensitivities of the analytes are known. As mentioned earlier, in this dissertation, the proposed signal processing techniques will be demonstrated for the detection and quantification of multi-analyte mixtures of BTEX compounds with or without the presence of interferents (including single analyte responses of BTEX compounds).

The signal processing techniques will be used to estimate the unknown parameters needed for the detection and quantification of the analytes. The analyte(s) can be quantified using eq. (3.5) based on the estimated equilibrium frequency shift, f_{∞} and the sensitivity, aK_{p-w} of the corresponding analyte (listed in Table 3.1 and 3.2 for SH-SAW sensors coated with PECH and PIB, respectively). In order to evaluate the performance of the proposed techniques, the estimated analyte concentration(s) will be compared to the results obtained independently using gas chromatograph-photoionization

detector (GC-PID) and gas chromatograph-mass spectrometry (GC-MS). It is noted that analyte concentrations determined using GC-PID are subject to an average error of $\pm 7\%$ [93].

All proposed sensor signal processing techniques are tested extensively using multiple measured SH-SAW sensor responses (time-dependent frequency shift transients) to multi-analyte mixtures of BTEX compounds with or without the presence of the interferences. In order to highlight the effectiveness of the proposed techniques, only a few selected representative results will be shown and discussed here. More results are presented in the Appendix section. However, whenever possible, the results obtained from these tests are summarized either graphically or by using a table.

6.2 Results of Detection and Quantification of Single Analytes Using Bank of Kalman Filters

In this section, the results of detection and quantification of unknown single BTEX compounds using a Bank of Kalman filters (BKFs) are presented. The technique used to perform the detection and quantification of single analytes has already been discussed in detail in Chapter 4. The objective here is to accurately identify and quantify the analyte which is most likely to have caused the measured polymer coated SH-SAW sensor response. For the implementation of the proposed approach, the unknown parameters, sorption/desorption rate constant of the analyte, S and equilibrium frequency shift, f_{∞} were quantized to a finite number of grid points. In this work, the parameter S was quantized as $\{S_{benzene}, S_{toluene}, S_{ethylbenzene \& \ xylenes}\}$ and the parameter f_{∞} was quantized to a finite number of grid points, $\{f_{\infty; 1}, f_{\infty; 2}, \dots, f_{\infty; N}\}$ by assuming that the

concentration of the analyte will be below 5 ppm. Thus, the unknown parameters vector, θ which is formed based on all possible combinations of the discrete points of the unknown parameters S and f_{∞} , will belong to a finite number of grid points. By utilizing the BKF algorithm as outlined in Table 2.4, the most probable case for the unknown parameters vector, θ is determined based on the highest *a posteriori* probability, $p(\theta_j | Y_k)$. Based on the selected probable case for θ , the values for unknown parameters, sorption/desorption rate constant, S and equilibrium frequency shift, f_{∞} are determined. Note that based on the selected S , the analyte most likely to have caused the response is identified and based on the selected f_{∞} , the identified analyte concentration is determined using eq. (3.5). The results are compared to the actual analyte concentration measured independently using GC-PID (and GC-MS).

The proposed approach was tested using several time-dependent frequency shift transients measured with polymer-coated SH-SAW sensors. The single analyte identification and quantification results obtained from these tests are summarized in Table 6.4 at the end of this section. Three representative results are presented first in order to highlight the attractive features of the proposed technique. First, the results obtained using the response of a SH-SAW sensor coated with 0.6 μ m PECH to 670 ppb benzene are discussed. Figure 6.1 and Table 6.1 show the results obtained using the aforementioned data and the proposed signal processing technique. Figure 6.1 shows the measurement data (in blue) and the estimated sensor response curve (in red) along with the identification and quantification result obtained using all the measurement data points. As can be seen from Figure 6.1, the estimated sensor response curve is a good fit to the measured data points, indicating correct identification and accurate quantification

of the analyte. Given the data of Figure 6.1, the proposed technique has correctly identified the analyte as benzene and estimated its concentration as 660 ppb which is very close ($\pm 1.5\%$ difference) to the benzene concentration measured using GC-PID. Table 6.1 shows the identification and quantification results obtained using just the measurement data collected for the first 1, 2, 3 and 5 minutes after the analyte has been introduced to the sensor. Based on the results shown in Table 6.1, it is evident that in this case the proposed technique is able to accurately identify the analyte most likely to have caused the measured response using just the data collected for the first 1 minute of analyte introduction to the sensor. In addition, a good estimated analyte concentration with less than $\pm 10\%$ difference to the analyte concentration measured using GC-PID can be obtained using just the measurement data collected for the first 2 minutes. Therefore, the results obtained imply that an extremely rapid analyte identification and quantification is possible using the proposed technique which utilizes BKF. This result certainly highlights the advantage of the proposed technique compared to using the conventional technique based on sensor arrays for analyte identification and quantification.

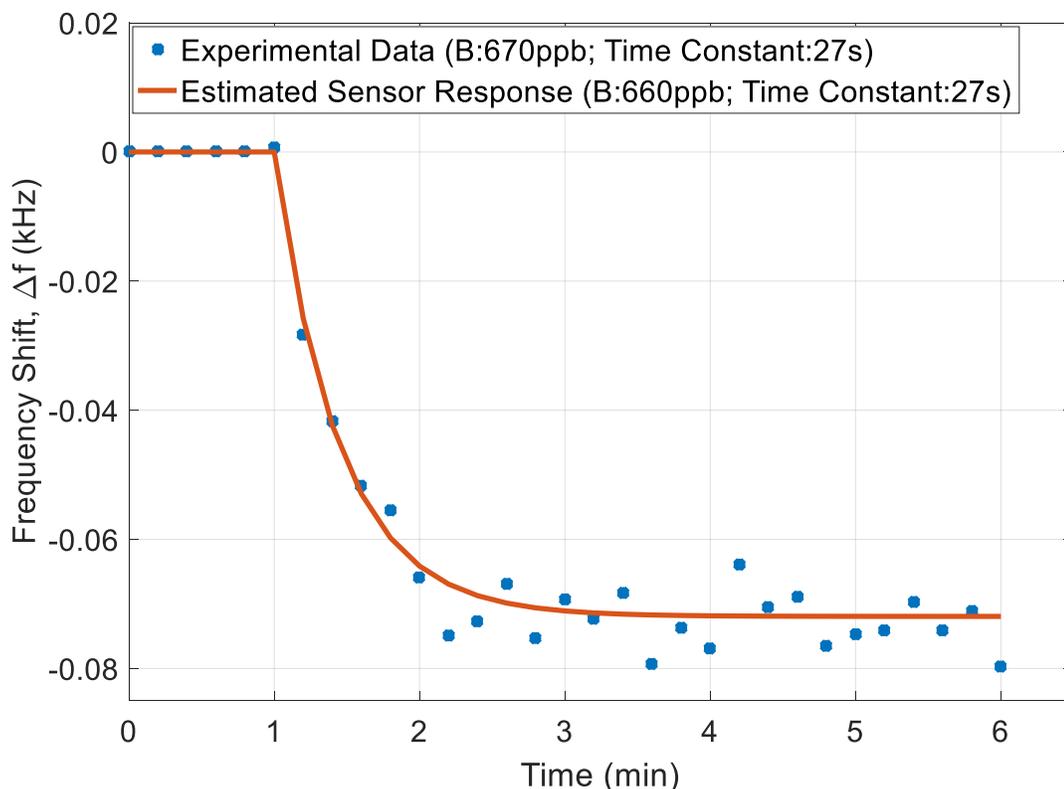


Figure 6.1: Measured response of a SH-SAW sensor coated with 0.6 μ m PECH to a single analyte sample of 670 ppb benzene (B). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentration is shown in the inset.

Table 6.1: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.6 μ m PECH to 670 ppb benzene) collected for the first 1, 2, 3 and 5 minutes after the analyte has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Identification and Quantification Result	Identification Results	Estimated Concentration in ppb (% difference to actual analyte concentration)
After 1 minute	Benzene	570 (15 %)
After 2 minutes	Benzene	630 (6 %)
After 3 minutes	Benzene	650 (3 %)
After 5 minutes	Benzene	660 (1.5 %)

Figure 6.2 and Table 6.2 show the results of another test of the proposed technique, in this case using the response of a SH-SAW sensor coated with 0.8 μ m PIB to 640 ppb toluene. As can be seen from Figure 6.2, the estimated sensor response curve (in red) shows good agreement with the measured data points. This means that the selected best case for the unknown parameters vector, θ should contain parameter values that are close to the actual values. Also shown in Figure 6.2, is the identification and quantification result obtained using the proposed approach. Again, the proposed approach accurately identified and quantified the detected analyte as toluene with a concentration of 640 ppb (0 % difference to the toluene concentration measured using GC-PID). Table 6.2 shows the results obtained using measured data collected for the first 4, 5, 6 and 9 minutes after the analyte has been introduced to the sensor. Based on the results, the data collected for the first 5 minutes are required to accurately identify and quantify the analyte (with less than ± 10 % difference to the toluene concentration measured using GC-PID). This is slightly longer than the time required for the identification and quantification of benzene in the previous case which is about 2 minutes (see Table 6.1). Several factors might contribute to the overall time to detection and quantification. These include the noise of the measured data and the total response length of the measured analyte, i.e. the time required for the analyte to cause equilibrium frequency shift response for the selected coating. Data with high noise level might increase the time required for accurately identifying and quantifying the analyte using the proposed approach. In cases with extremely poor signal-to-noise ratio (SNR), misidentification is also possible. The total time required for the analyte sensor response to reach equilibrium frequency shift may also affect the time required for accurate detection and

quantification. Note that for times much shorter than the response time, the onset of an exponential response is approximately linear, which precludes identification of the analyte at this point. Therefore, analytes with longer response time will result in slightly more time for identification and quantification using the proposed technique compared to analytes with shorter response time. Toluene (tested in the present case) has a longer response time compared to benzene (tested in the previous case). Therefore, toluene is expected to take slightly longer for accurate identification and quantification than benzene. Moreover, the measured data of Figure 6.2 (for the present case) has slightly higher noise compared to the measured data of Figure 6.1 (the previous case). The aforementioned factors might have influenced the time required for accurate analyte identification and quantification in the present case. Nonetheless, accurately identifying and quantifying toluene in just 5 minutes with less than $\pm 10\%$ error using the proposed technique is still extremely fast compared to using the conventional technique of a sensor array.

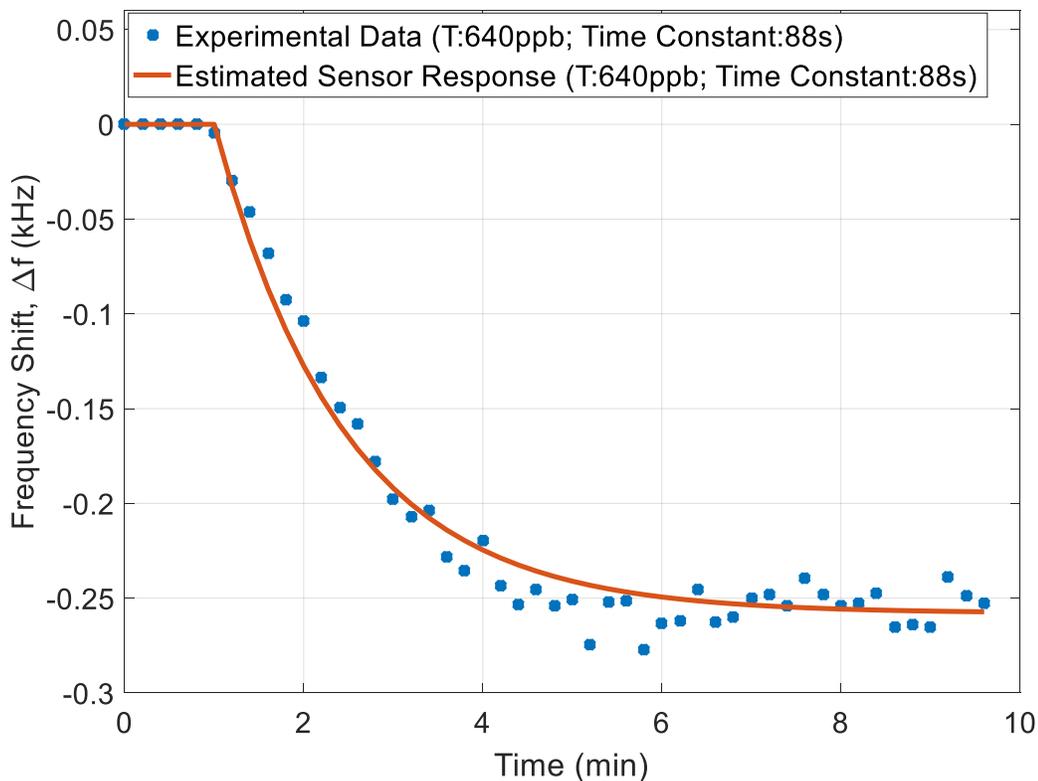


Figure 6.2: Measured response of a SH-SAW sensor coated with $0.8\mu\text{m}$ PIB to a single analyte sample of 640 ppb toluene (T). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentration is shown in the inset.

Table 6.2: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with $0.8\mu\text{m}$ PIB to 640 ppb toluene) collected for the first 4, 5, 6 and 9 minutes after the analyte has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Identification and Quantification Result	Identification Results	Estimated Concentration in ppb (% difference to actual analyte concentration)
After 4 minutes	Ethylbenzene	360 (N/A: Error in Analyte Identification)
After 5 minutes	Toluene	650 (2 %)
After 6 minutes	Toluene	660 (3 %)
After 9 minutes	Toluene	640 (0 %)

Finally, for the third example, the proposed approach was tested using the response data of a SH-SAW sensor coated with 0.8 μ m PIB to a 370 ppb ethylbenzene sample. Figure 6.3 and Table 6.3 show the results. By using the proposed approach and all the measurement data points, it has been identified that the measured response is most probably caused by ethylbenzene at the concentration of 370 ppb (0 % difference to the ethylbenzene concentration measured using GC-PID). Once again, the proposed approach is capable of accurately identifying and quantifying the analyte most likely to have caused the measured response. Table 6.3 shows the identification and quantification results obtained using only the measurement data collected for the first 3, 4, 6 and 11 minutes after the analyte has come in contact with the sensor. Based on the results, in order to accurately identify and quantify the analyte (with less than ± 10 % difference to the analyte concentration measured using GC-PID), the measured data collected for the first 3 minutes of analyte introduction to the sensor are sufficient. It is noted that the minimum time required to obtain accurate analyte identification and quantification for the third case is slightly longer than that of the first case but faster than that of the second case. Since the response time to ethylbenzene is slightly longer than for both benzene (first case) and toluene (second case), the reason for this might be due to the noise level in the data for the second case being higher than the third case. This indicates that the noise level in the measured data may play a significant role in determining time to accurate identification and quantification of the analyte. Nonetheless, in all three representative results presented here, by using the proposed technique based on BKFs, the analyte which is most likely to have caused the measured response was identified and quantified well before the sensor response reaches steady-state, i.e., equilibrium

frequency shift. The ability to rapidly detect and quantify the analyte is necessary in most real-world applications, particularly where detection and quantification of hazardous analytes is required.

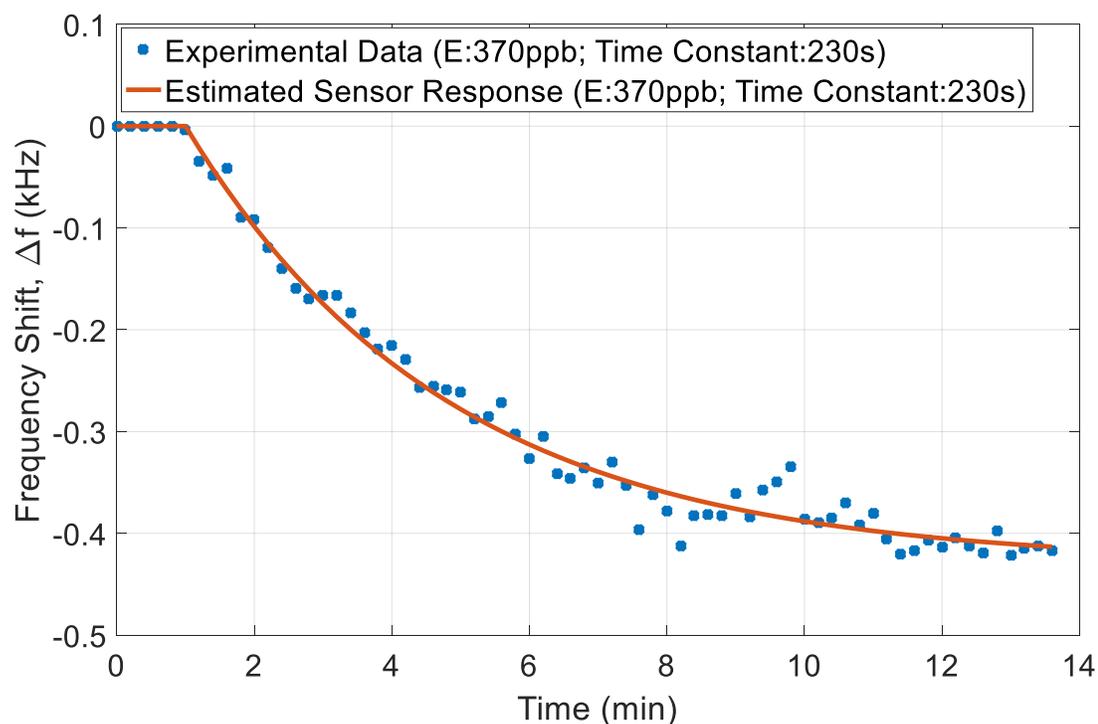


Figure 6.3: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to a single analyte sample of 370 ppb ethylbenzene (E). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentration is shown in the inset.

Table 6.3: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.8 μ m PIB exposed to 370 ppb ethylbenzene) collected for the first 3, 4, 6 and 11 minutes after the analyte has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Identification and Quantification Result	Identification Results	Estimated Concentration in ppb (% difference to actual analyte concentration)
After 3 minutes	Ethylbenzene	340 (8 %)
After 4 minutes	Ethylbenzene	350 (5 %)
After 6 minutes	Ethylbenzene	360 (3 %)
After 11 minutes	Ethylbenzene	370 (0 %)

In addition to the three representative results presented above, several additional experiments with different concentrations of benzene, toluene and ethylbenzene were conducted to test the proposed signal processing technique. The results are summarized in Table 6.4, which includes estimates obtained using the measured data of SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. In all the cases shown in Table 6.4, the proposed technique was able to accurately identify and quantify the analyte in real-time. Overall, the percentage error between the estimated and the measured analyte concentration using GC-PID is less than ± 10 %. For single analyte sensor responses, the results discussed in this section clearly demonstrate the potential of the proposed signal processing technique based on BKF's to rapidly identify and quantify the analyte which is most likely to have caused the response. These results highlight the advantages of the proposed technique as a viable alternative to using the conventional technique of sensor arrays for single analyte identification and quantification. Specifically, the proposed technique has the advantages of not requiring a complex training data set and offering

real-time detection and quantification of the analyte based on the sensor response of only a single sensor device.

Table 6.4: Summary of identification and quantification results obtained using bank of Kalman filters and the measured response data of SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Data	Identification Results	Estimated Concentration (ppb) (% difference to actual analyte concentration)
1: Benzene (1930 ppb)	Benzene	1910 (1 %)
2: Toluene (360 ppb)	Toluene	360 (0 %)
3: Ethylbenzene (1400 ppb)	Ethylbenzene	1390 (1 %)
4: Benzene (670 ppb)	Benzene	660 (1.5 %)
5: Toluene (880 ppb)	Toluene	870 (1 %)
6: Ethylbenzene (380 ppb)	Ethylbenzene	380 (0 %)
7: Benzene (1920 ppb)	Benzene	1850 (4 %)
8: Toluene (930 ppb)	Toluene	920 (1 %)
9: Toluene (640 ppb)	Toluene	640 (0 %)
10: Ethylbenzene (370 ppb)	Ethylbenzene	370 (0 %)

6.3 Results of Detection and Quantification of Binary Mixtures Using Bank of Kalman Filters

In this section, the novel approach based on BKF's will be applied to the rapid detection and quantification of binary mixtures of analytes from polymer-coated SH-SAW sensor responses to unknown binary mixtures of BTEX compounds. For a detailed discussion on the proposed technique, refer to Chapter 4.

For the implementation of the proposed technique, the unknown parameters: the sorption/desorption rate constant of each analyte (i.e. S_1 and S_2) and the corresponding equilibrium frequency shift of each analyte (i.e. $f_{\infty,1}$, and $f_{\infty,2}$) were quantized to a finite number of grid points. The parameters S_1 and S_2 were quantized based on the knowledge of possible binary mixtures of analytes to have caused the response. In this work, $\begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$ were quantized as $\left\{ \begin{bmatrix} S_{benzene} \\ S_{toluene} \end{bmatrix}, \begin{bmatrix} S_{benzene} \\ S_{ethylbenzene \& xylenes} \end{bmatrix}, \begin{bmatrix} S_{toluene} \\ S_{ethylbenzene \& xylenes} \end{bmatrix} \right\}$. In order to quantize the parameters $f_{\infty,1}$, and $f_{\infty,2}$, it is assumed that the concentration of the analytes will be below 5 ppm, as it is often the case for actual groundwater contamination from gasoline spills. Based on this assumption, $f_{\infty,1}$ and $f_{\infty,2}$ were quantized into a finite number of grid points. Thus, the unknown parameters vector, θ which is formed based on all possible combinations of the discrete points of all the unknown parameters, will belong to a finite number of grid points. This enables the application of the BKF algorithm (presented in Table 2.4) in order to determine the best case for the unknown parameters vector, θ based on the highest *a posteriori* probability, $p(\theta_j|Y_k)$. Based on the selection of the best case, the analytes most likely to have caused the response were identified using the selected sorption/desorption rate constant of the analytes (i.e. S_1 and

S_2) and quantified using the selected equilibrium frequency shifts (i.e. $f_{\infty,1}$ and $f_{\infty,2}$) and eq. (3.5). Similar to the previous section, the quantification results obtained will be compared to the concentration of the analyte measured independently using GC-PID (and GC-MS).

The proposed approach was tested using several measured polymer-coated SH-SAW sensor responses to binary mixture samples. However, only two representative results are presented in detail to highlight the potential of the proposed technique to accurately identify and quantify the analytes most likely to have caused the measured binary mixture response. At the end of this section, the results obtained from some of these tests are summarized in Table 6.7.

First, the results obtained using the measured SH-SAW sensor coated with $0.6\mu\text{m}$ PECH to a binary mixture of 980 ppb benzene and 340 ppb toluene are discussed. The results are shown in Figure 6.4 and Table 6.5. Figure 6.4 shows the measurement data (in blue) and the estimated sensor response curve (in red) along with the identification and quantification result obtained using all the measurement data points. As can be seen from the inset of Figure 6.4, the proposed technique managed to identify and quantify the analytes most likely to have caused the measured response accurately with less than $\pm 10\%$ difference to the analyte concentrations measured using GC-PID. Table 6.5 shows the identification and quantification results obtained using just the measurement data collected for the first 3, 5, 6 and 9 minutes after the analyte has been introduced to the sensor. It is evident that the proposed technique can accurately identify and quantify the analytes using just the data collected for the first 3 minutes. This means that by utilizing

the proposed technique, binary mixtures of analytes can be identified and quantified well before the measured response reaches equilibrium frequency shift.

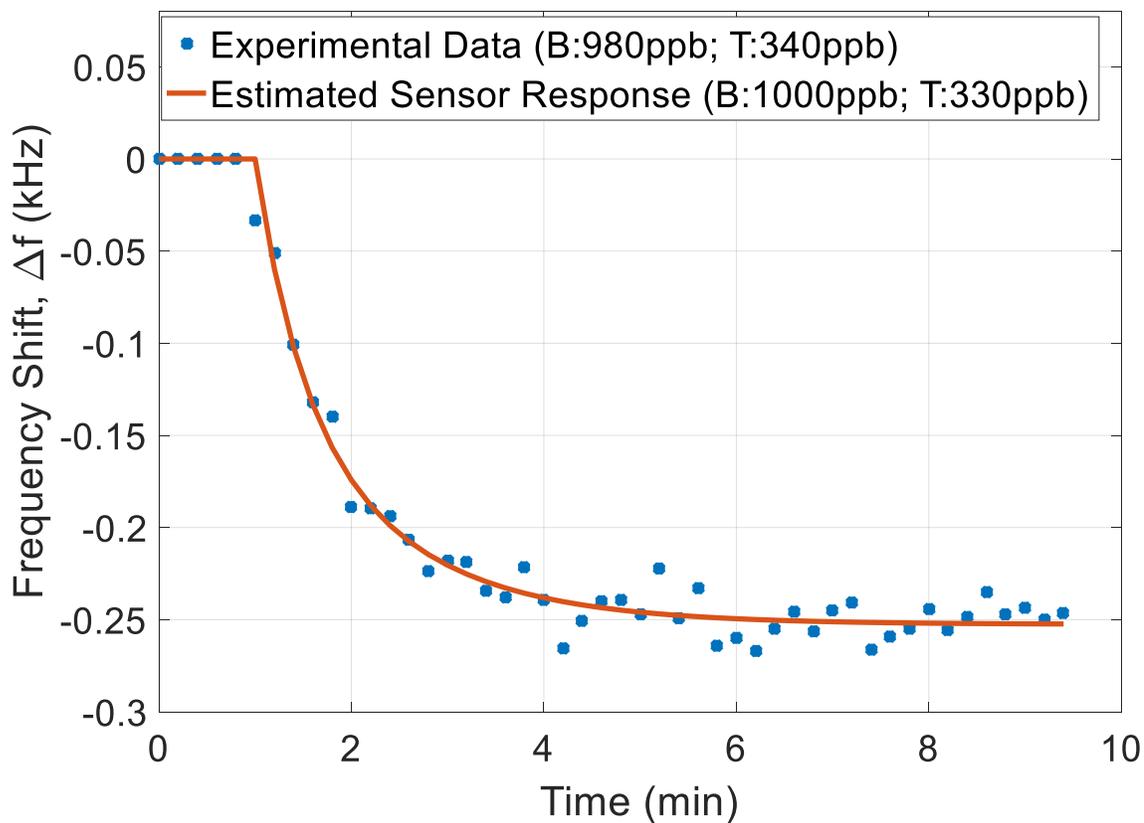


Figure 6.4: Measured response of a SH-SAW sensor coated with 0.6 μm PECH to binary mixtures of 980 ppb benzene (B) and 340 ppb toluene (T). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentrations are shown in the inset.

Table 6.5: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.6 μ m PECH to 980 ppb benzene and 340 ppb toluene) collected for the first 3, 5, 6 and 9 minutes after the binary mixture sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Identification and Quantification Result	Identification Results	Estimated Concentration (ppb) (% difference to actual analyte concentration)
After 3 minutes	Benzene	890 (9 %)
	Toluene	370 (9 %)
After 5 minutes	Benzene	1020 (4 %)
	Toluene	320 (6 %)
After 6 minutes	Benzene	940 (4 %)
	Toluene	350 (3 %)
After 9 minutes	Benzene	1000 (2 %)
	Toluene	330 (3 %)

For the second example, the proposed technique was tested using the sensor response data of SH-SAW sensor coated with 0.8 μ m PIB to binary mixtures of 510 ppb benzene and 260 ppb ethylbenzene. The results are shown in Figure 6.5 and Table 6.6. From the results, it can be concluded that the proposed technique is capable of accurately identifying and quantifying the analytes most likely to have caused the measured response as being benzene and ethylbenzene with less than ± 10 % difference to the analytes concentration measured using GC-PID, and it is able to do so using only the measured data collected for the first 7 minutes. This is slightly longer than the time required for the identification and quantification of a binary mixture of benzene and toluene in the previous example, which was about 3 minutes (see Table 6.5). The difference in the minimum time required to identify and quantify the analytes accurately

might be due to the total response time of the binary mixture of benzene and ethylbenzene being longer than for the binary mixture of benzene and toluene. Nevertheless, accurately identifying and quantifying the analytes causing the response in only 7 minutes for a sample instead of a total response time (equilibrium response) of about 15 minutes is still an important improvement.

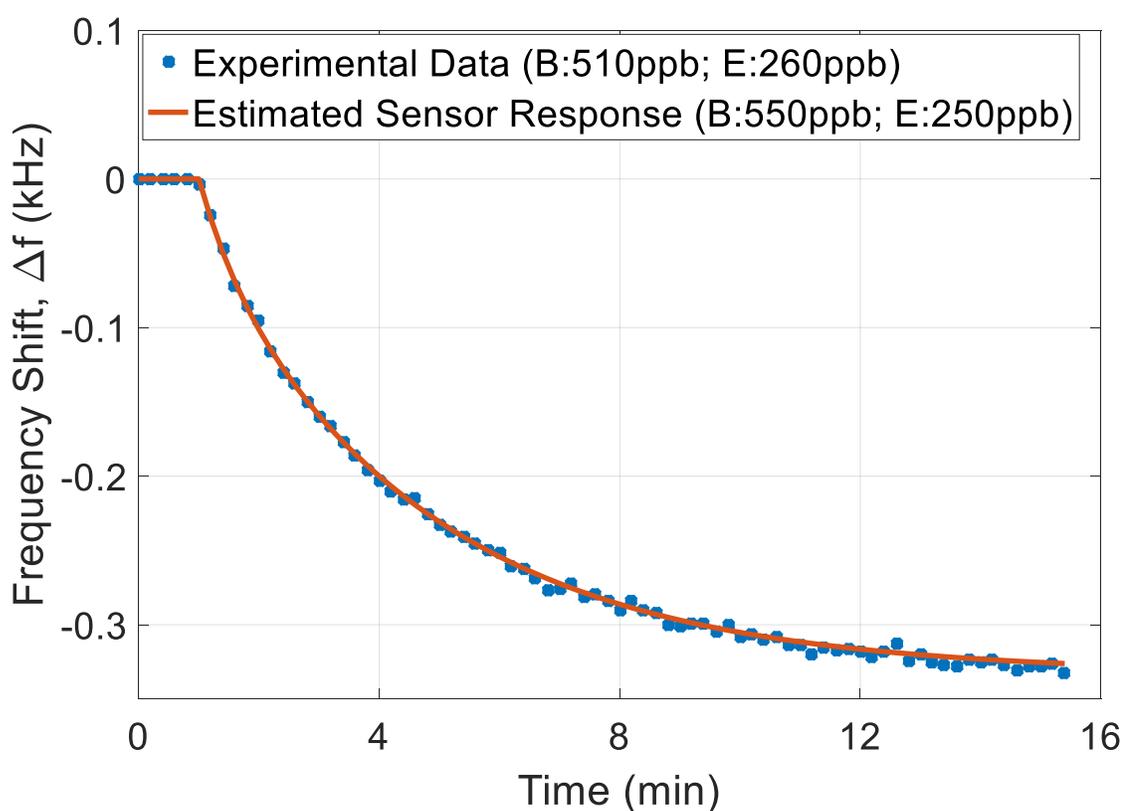


Figure 6.5: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to a binary mixture of 510 ppb benzene (B) and 260 ppb ethylbenzene (E). Also shown (red line) is the estimated sensor response obtained using the most probable unknown parameters determined using bank of Kalman filters. The actual and estimated analyte concentrations are shown in the inset.

Table 6.6: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.8 μ m PIB to 510 ppb benzene and 260 ppb ethylbenzene) collected for the first 4, 6, 7 and 10 minutes after the binary mixture sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Identification and Quantification Result	Identification Results	Estimated Concentration (ppb) (% difference to actual analyte concentration)
After 4 minutes	Toluene	50 (N/A)
	Ethylbenzene	250 (N/A)
After 6 minutes	Toluene	350 (N/A)
	Ethylbenzene	150 (N/A)
After 7 minutes	Benzene	550 (8 %)
	Ethylbenzene	250 (4 %)
After 10 minutes	Benzene	550 (8 %)
	Ethylbenzene	250 (4 %)

Results for all the tested binary mixtures of BTEX compounds are summarized in Table 6.7, which includes estimates obtained using the measured data of SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. For all the examples shown in Table 6.7, the proposed technique was able to accurately identify and quantify the two analytes in the tested samples rapidly. Specifically, for most of the tested samples the analytes could be identified and quantified well before the sensor response reached equilibrium frequency shift. It is also pertinent to note that using the proposed technique, high quantification accuracy i.e. less than ± 15 % difference to the analyte concentrations measured using GC-PID can be achieved. It is noted that this accuracy can still be further improved by appropriately selecting or finding a coating with higher sensitivity to the analytes of interest or with higher liquid-phase polymer-analyte partition coefficients.

In summary, the results shown and discussed in this section clearly demonstrate the potential of the proposed technique, which is based on BKF, to rapidly identify and quantify the two analytes in a sample based on the measured response from only a single polymer coated SH-SAW sensor. By utilizing the proposed technique, a more tedious sensor array signal processing protocol could be avoided. Moreover, using the proposed technique, the analytes could be identified and quantified in real-time, a feat that cannot be achieved using a sensor array and only equilibrium frequency shift as the sole sensing parameter.

Table 6.7: Summary of identification and quantification results obtained using bank of Kalman filters and the measured binary mixtures response data of SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Binary Mixture Data	Identified Analytes and their Estimated Concentrations in ppb (% difference with actual analyte concentration)
1: Benzene (1000 ppb) Toluene (500 ppb)	Benzene: 940 (6 %) Toluene: 510 (2 %)
2: Benzene (1060 ppb) Ethylbenzene (1410 ppb)	Benzene: 970 (9 %) Ethylbenzene: 1360 (4%)
3: Benzene (620 ppb) Ethylbenzene (1260 ppb)	Benzene: 590 (5 %) Ethylbenzene: 1200 (5 %)
4: Benzene (2260 ppb) Toluene (740 ppb)	Benzene: 2010 (11 %) Toluene: 790 (7 %)
5: Benzene (500 ppb) Toluene (1000 ppb)	Benzene: 490 (2 %) Toluene: 1000 (0 %)
6: Benzene (1000 ppb) Ethylbenzene (800 ppb)	Benzene: 920 (8 %) Ethylbenzene: 800 (0 %)

6.4 Results of Identification and Quantification of Multi-Analyte Mixtures Using Multi-Stage Exponentially Weighted RLSE

In this section, the results of detection and quantification of multi-analyte mixtures using a sensor signal processing technique based on multi-stage exponentially weighted RLSE are presented. Note that this technique could be used for the detection and quantification of n analytes in a mixture that are most likely to have caused the measured sensor response, provided the characteristic response time constants and sensitivities of the analytes-coating pairs are known. However, in this dissertation, the proposed technique was demonstrated for the detection and quantification of multi-analyte mixtures of BTEX compounds and TMB using the time-transient frequency shift response of a single polymer-coated SH-SAW sensor. The measured sensor responses were either obtained from SH-SAW sensors coated with 0.6 μm PECH or 0.8 μm PIB. For reference, analyte concentrations were also independently measured using GC-PID. A detailed discussion on the implementation of the proposed technique can be found in Chapter 4 (specifically section 4.4). It is noted that the proposed technique is insensitive to the initial values of the unknown parameters (i.e. the equilibrium frequency shifts, $f_{\infty,i}$). Thus, for the implementation of the proposed technique, the unknown parameters, $f_{\infty,i}$ were set to zero for all the tested data. Once the unknown parameters, $f_{\infty,i}$ are estimated, the corresponding concentrations associated with each of these unknown parameters are calculated using eq. (3.5). In this technique, the unknown parameters, $f_{\infty,i}$ are no longer being discretized into a finite number of grid points as in the previous techniques.

In order to validate the proposed technique, multiple tests were performed on detection and quantification of multi-analyte samples, but also including tests using single-analyte samples. Selected results from these tests are summarized at the end of this section. The results obtained from three different tests are discussed first in order to highlight the effectiveness of the proposed technique.

The first sample result is shown in Figure 6.6. This result was obtained using the sensor response data of a SH-SAW sensor coated with 0.8 μ m PIB to multi-analyte mixtures of 590 ppb benzene (B), 420 ppb toluene (T), and 330 ppb ethylbenzene and xylenes (EX). Note that ethylbenzene and the three xylenes are grouped together because they are chemical isomers which have similar response time constants and sensitivities for the investigated coatings. Figure 6.6 shows the measurement data (in blue) and the estimated sensor response curve (in red) along with the identification and quantification results. The measured and estimated frequency shift response in Figure 6.6 are in good agreement, resulting in accurate identification of the analytes in the sample with the estimated analyte concentrations within ± 10 % of reference analyte concentrations measured using GC-PID. Table 6.8 shows the identification and quantification results obtained using the measured data collected for the first 14, 16, 18 and 20 minutes after the multi-analyte sample has come in contact with the sensor. As can be seen from Table 6.8, for this multi-analyte sample, at least the data collected for the first 14 minutes are needed to accurately identify and quantify the analytes with less than ± 20 % difference to the analyte concentrations measured using the GC-PID. Using just the measured data collected for less than the first 14 minutes will either result in misidentification or inaccurate quantification of the analytes. In this case, a longer time is needed to

accurately identify and quantify the analytes present in the mixture. This indicates that, as the number of analytes in the mixture increases, the minimum time to accurately detect and quantify the analytes in the sample might increase.

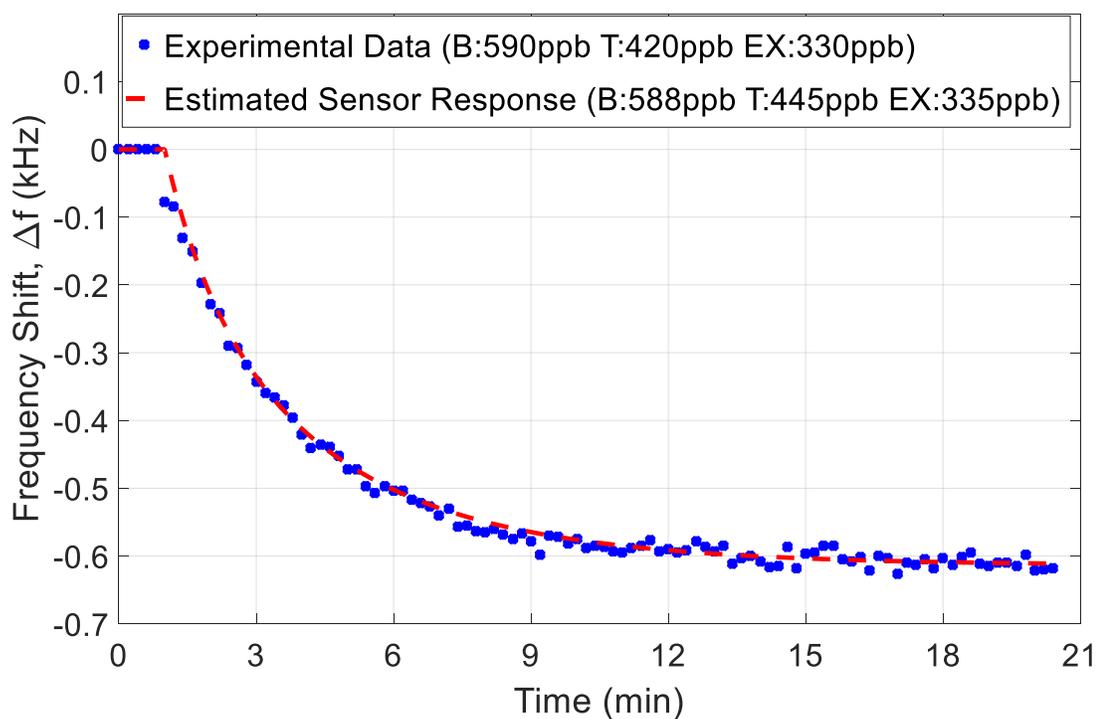


Figure 6.6: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to multi-analyte mixtures of 590 ppb benzene (B), 420 ppb toluene (T), and 330 ppb ethylbenzene and xylenes (EX). Also shown (red line) is the sensor response estimated using the multi-stage exponentially weighted RLSE. The actual and estimated analyte concentrations are shown in the inset.

Table 6.8: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.8 μ m PIB to 590 ppb benzene, 420 ppb toluene, and 330 ppb ethylbenzene and xylenes) collected for the first 14, 16, 18 and 20 minutes after the multi-analyte sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Identification and Quantification Result	Identified Analytes and their Estimated Concentrations in ppb (% difference to actual analyte concentration)
After 14 minutes	Benzene: 714 (20 %) Toluene: 392 (7 %) Ethylbenzene & Xylenes: 347 (5 %)
After 16 minutes	Benzene: 527 (11 %) Toluene: 469 (12 %) Ethylbenzene & Xylenes: 330 (0 %)
After 18 minutes	Benzene: 534 (10 %) Toluene: 466 (11 %) Ethylbenzene & Xylenes: 330 (0 %)
After 20 minutes	Benzene: 588 (1 %) Toluene: 445 (6 %) Ethylbenzene & Xylenes: 335 (2 %)

Next, the results obtained using the measured data of a 0.6 μ m PECH coated SH-SAW sensor to a binary mixture sample containing 980 ppb benzene and 340 ppb toluene are presented. Figure 6.7 and Table 6.9 show the results. Once again, using the proposed technique, the two analytes in the sample were identified and quantified accurately. For the second sample, estimated analyte concentrations are within ± 10 % difference to the analyte concentrations measured using the GC-PID. The accurate identification and quantification result obtained is also consistent with the excellent agreement between the measured data points and the estimated sensor response. From the results shown in Table 6.9, a minimum of the data collected for the first 3 minutes are sufficient to detect and

quantify the two analytes in the sample with less than $\pm 20\%$ difference to the analyte concentrations measured using the GC-PID. This implies that, using the proposed technique, rapid identification and quantification of the analytes is possible. In comparison to the results of the first sample discussed earlier, the minimum time required to accurately detect and quantify the analytes for the second sample is shorter. This result is consistent with the inference made earlier that the number of analytes in the mixture dictates the minimum time required to accurately detect and quantify the analytes in the sample. Moreover, the total response time of a binary mixture sample containing benzene and toluene is shorter than the total response time of the multi-analyte mixture sample containing BTEX compounds discussed earlier.

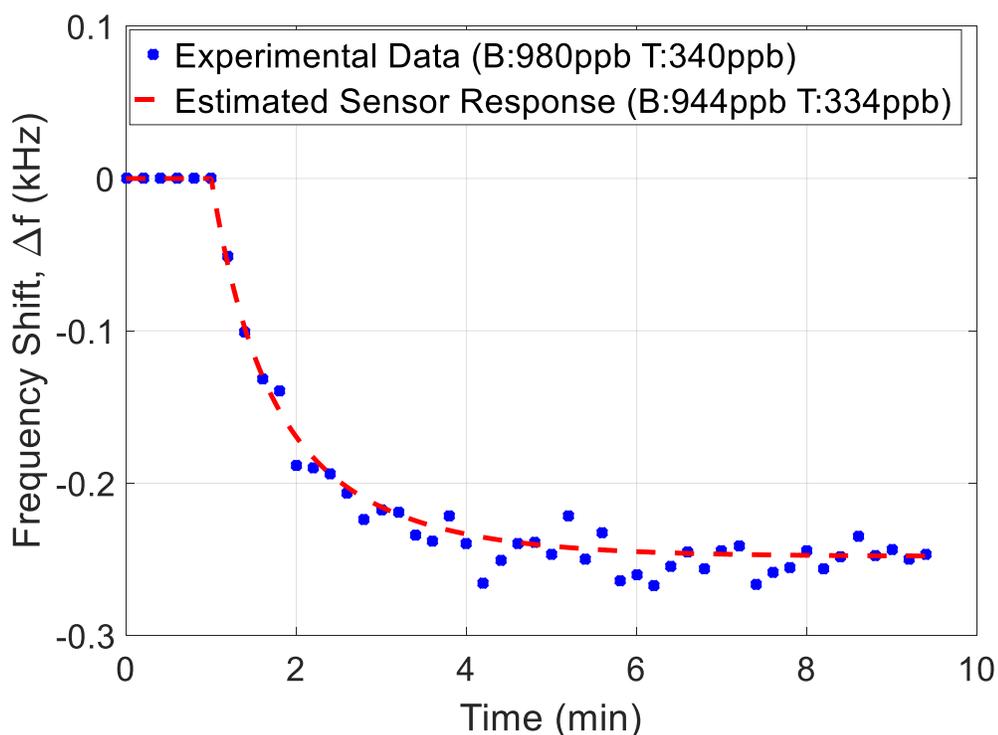


Figure 6.7: Measured response of a SH-SAW sensor coated with 0.6 μ m PECH to binary mixtures of 980 ppb benzene (B) and 340 ppb toluene (T). Also shown (red line) is the sensor response estimated using the multi-stage exponentially weighted RLSE. The actual and estimated analyte concentrations are shown in the inset.

Table 6.9: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.6 μ m PECH to 980 ppb benzene and 340 ppb toluene) collected for the first 3, 4, and 8 minutes after the binary mixture sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Identification and Quantification Result	Identified Analytes and their Estimated Concentrations in ppb (% difference to actual analyte concentration)
After 3 minutes	Benzene: 1044 (7 %) Toluene: 271 (20 %)
After 4 minutes	Benzene: 994 (1 %) Toluene: 307 (10 %)
After 8 minutes	Benzene: 943 (4 %) Toluene: 334 (2 %)

For the third example, the proposed approach is tested using the sensor response data of a SH-SAW sensor coated with 0.8 μ m PIB to 600 ppb 1,2,4-trimethylbenzene (TMB). The results for the third case are shown in Figure 6.8 and Table 6.10. Based on these results, it can be concluded that the proposed technique is able to accurately identify and quantify the analyte in the sample as TMB with less than ± 15 % difference to the TMB concentration measured using the GC-PID. In fact, the analyte is accurately identified and quantified (with ± 20 % difference to the TMB concentration measured using GC-PID) using only the measured data collected for the first 5 minutes, which is just a fraction of the total response time of TMB. In this case, valuable time could be saved (approximately 20 minutes of response time of TMB) by utilizing the proposed technique.

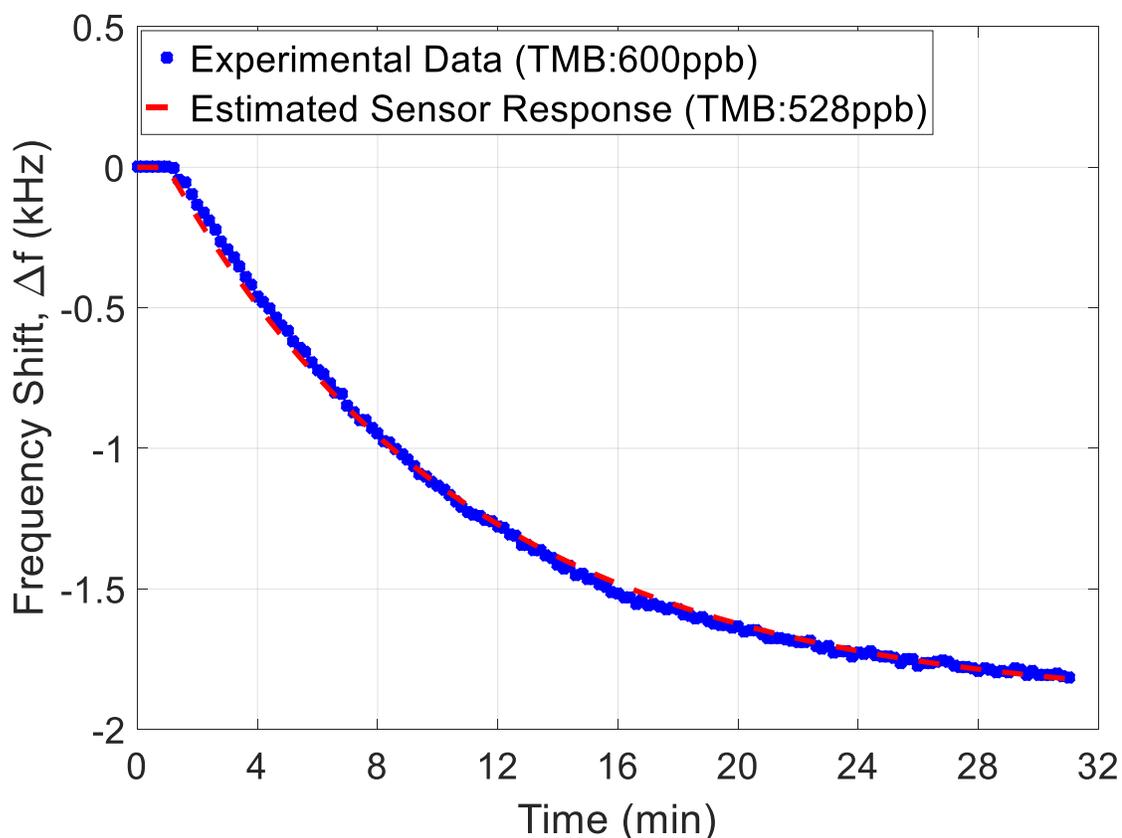


Figure 6.8: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to 600 ppb 1,2,4-trimethylbenzene (TMB). Also shown (red line) is the sensor response estimated using the multi-stage exponentially weighted RLSE. The actual and estimated analyte concentrations are shown in the inset.

Table 6.10: Identification and quantification results obtained using the measured response data (of a SH-SAW sensor coated with 0.8 μ m PIB to 600 ppb 1,2,4-trimethylbenzene) collected for the first 5, 8, 10 and 14 minutes after the single analyte sample has been introduced to the sensor. Also shown in the table are the percentage differences between the estimated concentration and the concentration of the analyte determined using GC-PID.

Identification and Quantification Result	Identified Analytes and their Estimated Concentrations in ppb (% difference with actual analyte concentration)
After 5 minutes	1,2,4-trimethylbenzene: 484 (19 %)
After 8 minutes	1,2,4-trimethylbenzene: 509 (15 %)
After 10 minutes	1,2,4-trimethylbenzene: 518 (14 %)
After 14 minutes	1,2,4-trimethylbenzene: 527 (12 %)

As mentioned earlier, in addition to these three sample results, multiple tests were performed on the proposed technique for the detection and quantification of multi-analyte samples (including tests using single analyte samples). Some of the results obtained from these tests are summarized in Table 6.11, which includes estimates obtained using the measured data of SH-SAW sensors coated with either 0.6 μm PECH or 0.8 μm PIB. These results demonstrate the ability of the proposed technique to accurately identify and quantify the analytes (with less than $\pm 20\%$ difference to the analytes concentration measured using GC-PID) in the mixture using just the measured response from a single polymer coated SH-SAW device. It is pertinent to note that using the proposed technique, the detection and quantification of the analytes can be performed in real-time as the measurements are recorded. The results clearly demonstrate the ability of the multi-stage exponentially weighted RLSE to identify and quantify multi-analyte mixtures using only the measured sensor response from a single polymer coated SH-SAW sensor in real-time. Clearly, this technique can be used as a viable alternative to chemical sensor arrays for multi-analyte detection and quantification, offering the advantages of not requiring a complex training data set, reduced cost of implementation, and rapid detection and quantification of the analytes.

Table 6.11: Summary of identification and quantification results obtained using multi-stage exponentially weighted RLSE and frequency transient data of SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB, compared to analyte concentrations in the mixture measured using GC-PID. In the table ‘B’ denotes benzene, ‘T’ denotes toluene, ‘EX’ denotes ethylbenzene and xylenes, and ‘TMB’ denotes 1,2,4-trimethylbenzene.

Data	Actual Concentrations (in ppb)				Estimated Concentrations (in ppb)			
	B	T	EX	TMB	B	T	EX	TMB
1	1920	0	0	0	1880	0	0	0
2	0	360	0	0	0	360	0	0
3	0	0	370	0	0	0	370	0
4	620	0	1260	0	660	0	1240	0
5	870	520	70	0	950	510	67	0
6	390	810	70	0	450	720	82	0
7	170	450	360	0	140	470	350	0
8	0	0	0	500	0	0	0	430
9	0	0	0	1000	0	0	0	1030
10	0	640	0	0	0	630	0	0

6.5 Results of Detection and Quantification of Multi-Analyte in the Presence of Interferents

In this section, the results of the detection and quantification of multi-analyte samples in the presence of interferents using the formulated signal processing approach based on exponentially weighted RLSE and BKF's are discussed. The proposed sensor signal processing procedure was introduced and discussed in detail in Chapter 4 (section 4.5). In general, the proposed technique can be used for the detection and quantification of any number of target analytes in the presence of interferents, provided that the characteristic response time constants and sensitivities of the target analytes and the

dominant interferents for the selected coatings are known. In this dissertation, the proposed technique was tested for the detection and quantification of BTEX compounds in the presence of interferents commonly found in the contaminated groundwater using only the time-transient frequency shift response from a single polymer-coated SH-SAW sensor. The results reported here are obtained using the formulated signal processing approach and either four-analyte model or five-analyte model. For the four-analyte model, the response due to all the detectable interferents in the sample is modeled using a single exponential term, whereas for the five-analyte model, the response due to all the detectable interferents in the sample is modeled using a dual-exponential term. Since the sensor response to the interferents is often dominated by 1,2,4-trimethylbenzene, the parameters of the single exponential term designated for the interferents in the four-analyte model were set close to those of 1,2,4-trimethylbenzene. In the five-analyte model, one of the exponential terms is used to represent the response due to 1,2,4-trimethylbenzene and the other is used to represent the contributions of the other less dominant interferents by setting the parameters of the exponential term slightly higher than those of 1,2,4-trimethylbenzene. Note that more details on these example models can be found in Chapter 3.

The proposed approach was tested extensively using measured SH-SAW sensor responses (frequency shifts) to BTEX compounds in actual groundwater (which contains various non-target interferents such as dissolved salts, aliphatic hydrocarbons, dissolved gases, particles and sediments, ethers, esters, ethanol, 1,2,4-trimethylbenzene, naphthalene, *n*-heptane and MTBE (methyl *tert*-butyl ether)). The tested groundwater sample were collected from groundwater monitoring wells in California. The measured

sensor responses to contaminated groundwater were either obtained from SH-SAW sensors coated with 0.6 μm PECH or 0.8 μm PIB. Since the proposed technique is independent of the initial values of the unknown parameters (i.e. the equilibrium frequency shifts, $f_{\infty,i}$), the initial values of all the unknown parameters (i.e. four unknown parameters in the four-analyte model and five unknown parameters in the five-analyte model) were set to zero for all the tested data. Once the unknown parameters are estimated, the corresponding concentrations associated with each of these unknown parameters are determined using eq. (3.5). The estimated concentrations are then compared to the BTEX concentrations in the samples measured independently using a GC-PID (and GC-MS). These results are summarized graphically in Figures 6.13-6.22 at the end of this section. A pair of representative estimation results for each coating (i.e. PECH and PIB) are presented first to highlight the effectiveness of the proposed technique.

Figures 6.9 and 6.10 show results for sensor response data collected using a SH-SAW sensor coated with 0.6 μm PECH to a LNAPL sample containing 370 ppb of benzene, 660 ppb of toluene, and 330 ppb of ethylbenzene and xylenes in the presence of interferences (such as dissolved salts, aliphatic hydrocarbons, dissolved gases, particles and sediments, ethers, esters, ethanol, 1,2,4-trimethylbenzene, naphthalene, *n*-heptane and MTBE (methyl *tert*-butyl ether) commonly found in groundwater. Figure 6.9 shows the results obtained using the four-analyte model and Figure 6.10 shows the results obtained using the five-analyte model. Table 6.12 summarizes the results obtained using the two different models. Both Figures 6.9 and 6.10 show very good agreement between the measured data and the estimated sensor response obtained using the proposed signal

processing technique. This implies that the estimated equilibrium frequency shifts, $f_{\infty,i}$ (especially, the $f_{\infty,i}$ associated with the target analytes) should be close to the actual values of the equilibrium frequency shifts of the analytes. As explained in section 4.5, sorption data was used to estimate the concentrations of benzene and toluene whereas the desorption data was used to estimate the combined concentrations of chemical isomers ethylbenzene and xylenes. The results in Table 6.12 indicate very good agreement between the concentrations measured using GC-PID and the concentrations estimated using the proposed approach. For the four-analyte model all the estimated concentrations are well within 20% of the GC-PID measurements and for the five-analyte model the estimated concentrations are well within 15% of the GC-PID measurements. Note that, since the proposed technique utilizes both sorption and desorption data, the target analytes can only be quantified accurately soon (i.e. within several minutes) after the desorption data was collected. Thus, the proposed technique is capable of near-real time detection and quantification of target analytes in the presence of interferents.

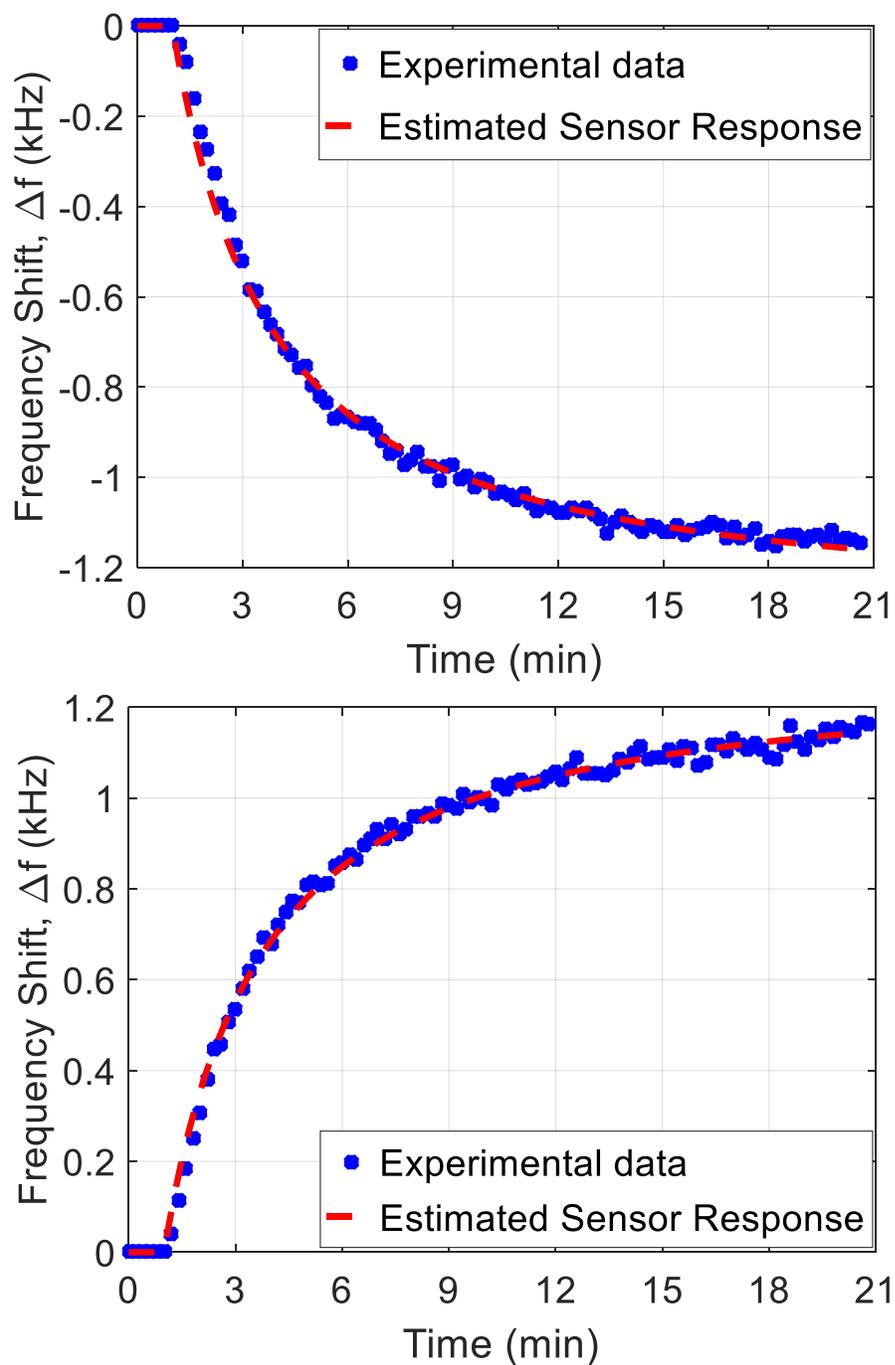


Figure 6.9: Measured response of a SH-SAW sensor coated with 0.6 μm PECH to a LNAPL sample in groundwater containing 370 ppb benzene, 660 ppb toluene, and 330 ppb ethylbenzene/xylenes, and unknown concentration of interferents (top: sorption data, bottom: desorption data). Also shown (red dashed line) are the estimated sensor responses obtained using four-analyte model.

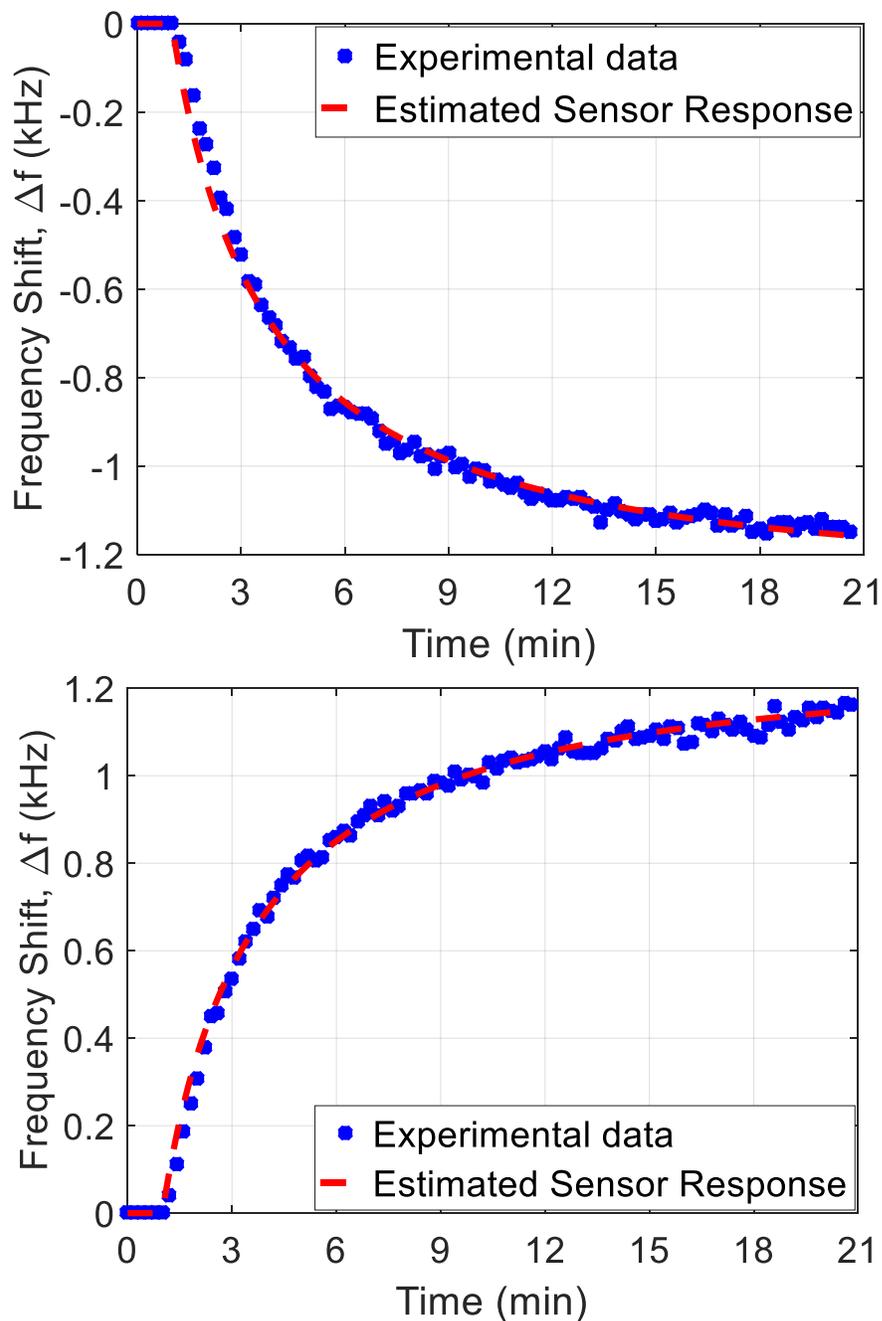


Figure 6.10: Measured response of a SH-SAW sensor coated with 0.6 μ m PECH to a LNAPL sample in groundwater containing 370 ppb benzene, 660 ppb toluene, and 330 ppb ethylbenzene/xylenes, and unknown concentration of interferences (top: sorption data, bottom: desorption data). Also shown (red dashed line) are the estimated sensor responses obtained using five-analyte model.

Table 6.12: Estimated concentrations of BTEX compounds obtained using the measurement data of a LNAPL sample in groundwater (collected using a SH-SAW device coated with 0.6 μ m PECH) compared to concentrations measured using GC-PID.

Target Analyte	Analyte Concentrations from GC-PID (ppb)	Estimated Concentrations (ppb) [% difference with actual concentration]	
		Four-Analyte Model	Five-Analyte Model
Benzene	370	309 [16%]	317 [14%]
Toluene	660	531 [20%]	649 [2%]
Ethylbenzene & Xylenes	330	370 [12%]	296 [10%]

The estimation results obtained using the response of a SH-SAW sensor coated with 0.8 μ m PIB to a LNAPL sample containing 670 ppb of benzene, 1340 ppb of toluene, and 510 ppb of ethylbenzene and xylenes in the presence of interferents (such as dissolved salts, aliphatic hydrocarbons, dissolved gases, particles and sediments, ethers, esters, ethanol, 1,2,4-trimethylbenzene, naphthalene, *n*-heptane and MTBE (methyl *tert*-butyl ether)) are shown in Figure 6.11 (for four-analyte model) and Figure 6.12 (for five-analyte model). Table 6.13 summarizes the results obtained using both models. Based on Figures 6.11 and 6.12, it is observed that the estimated response curves are in close agreement with the measured data points, indicating the estimated equilibrium frequency shifts, $f_{\infty,i}$ obtained using either four-analyte or five-analyte model are indeed close to the actual values. As can be seen from Table 6.13, the estimated concentrations obtained using either four-analyte or five-analyte model are in good agreement with the concentrations determined using GC-PID. In this case, all the estimated concentrations for both four-analyte and five-analyte models are within 15% of the GC-PID measurements. However, the estimated concentrations obtained using the five-analyte

model are slightly more accurate than the quantification results obtained using the four-analyte model. It is important to note that for the sample results shown, if the desorption data were not used in the proposed signal processing procedure, the estimation error for ethylbenzene and xylenes would have been much higher (about 30% for both four-analyte and five-analyte model). As for the time to quantification, as mentioned earlier, the target analytes can be quantified accurately within several minutes after the desorption data was collected. The accurate results obtained for the two sample data discussed here clearly illustrate the effectiveness of the proposed technique for the detection and quantification of the target analytes in the presence of interferents.

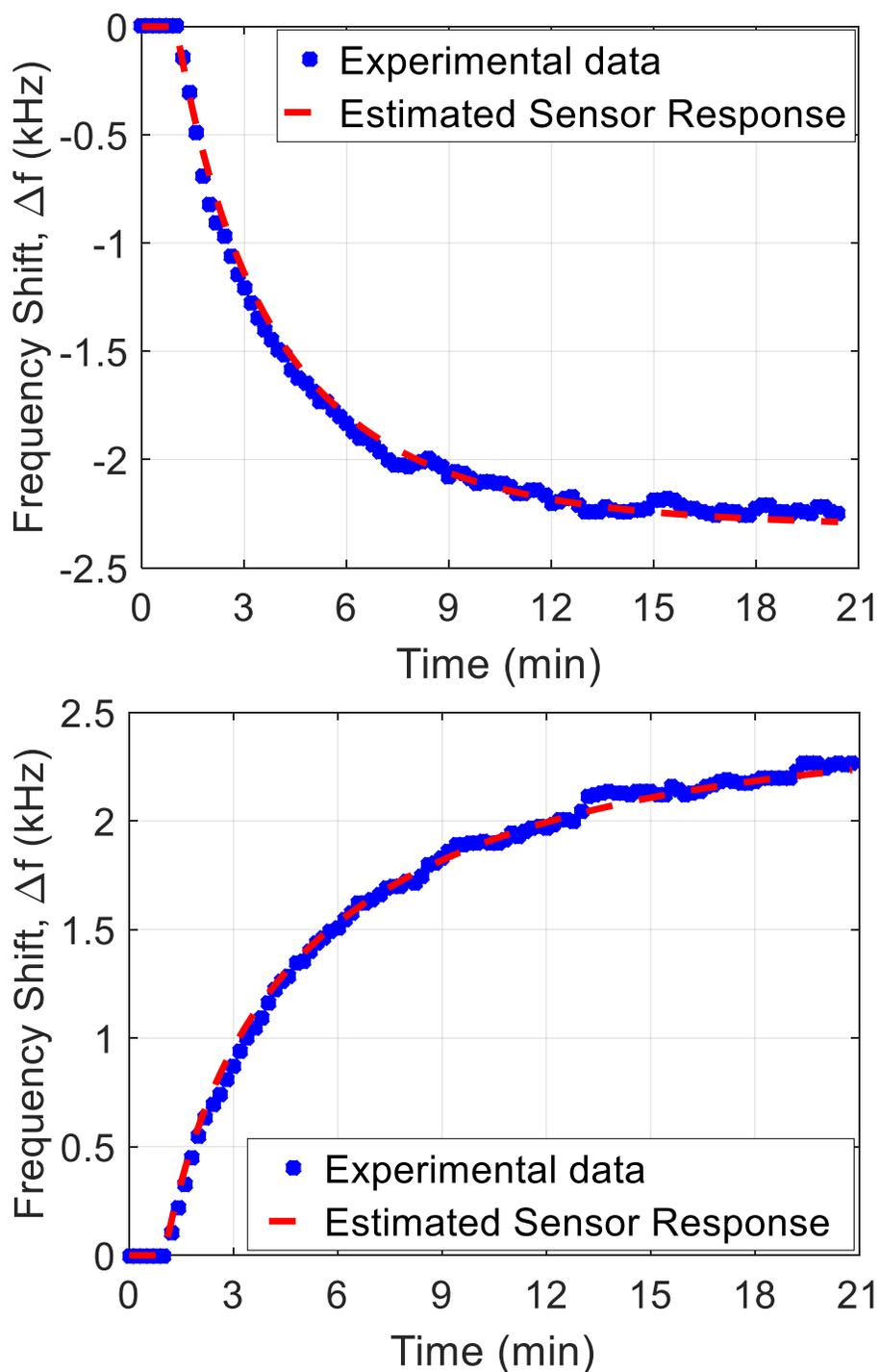


Figure 6.11: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to a LNAPL sample in groundwater containing 670 ppb benzene, 1340 ppb toluene, 510 ppb ethylbenzene/xylenes, and an unknown concentration of interferences (top: sorption data, bottom: desorption data). Also shown (red dashed line) are the estimated sensor responses obtained using four-analyte model.

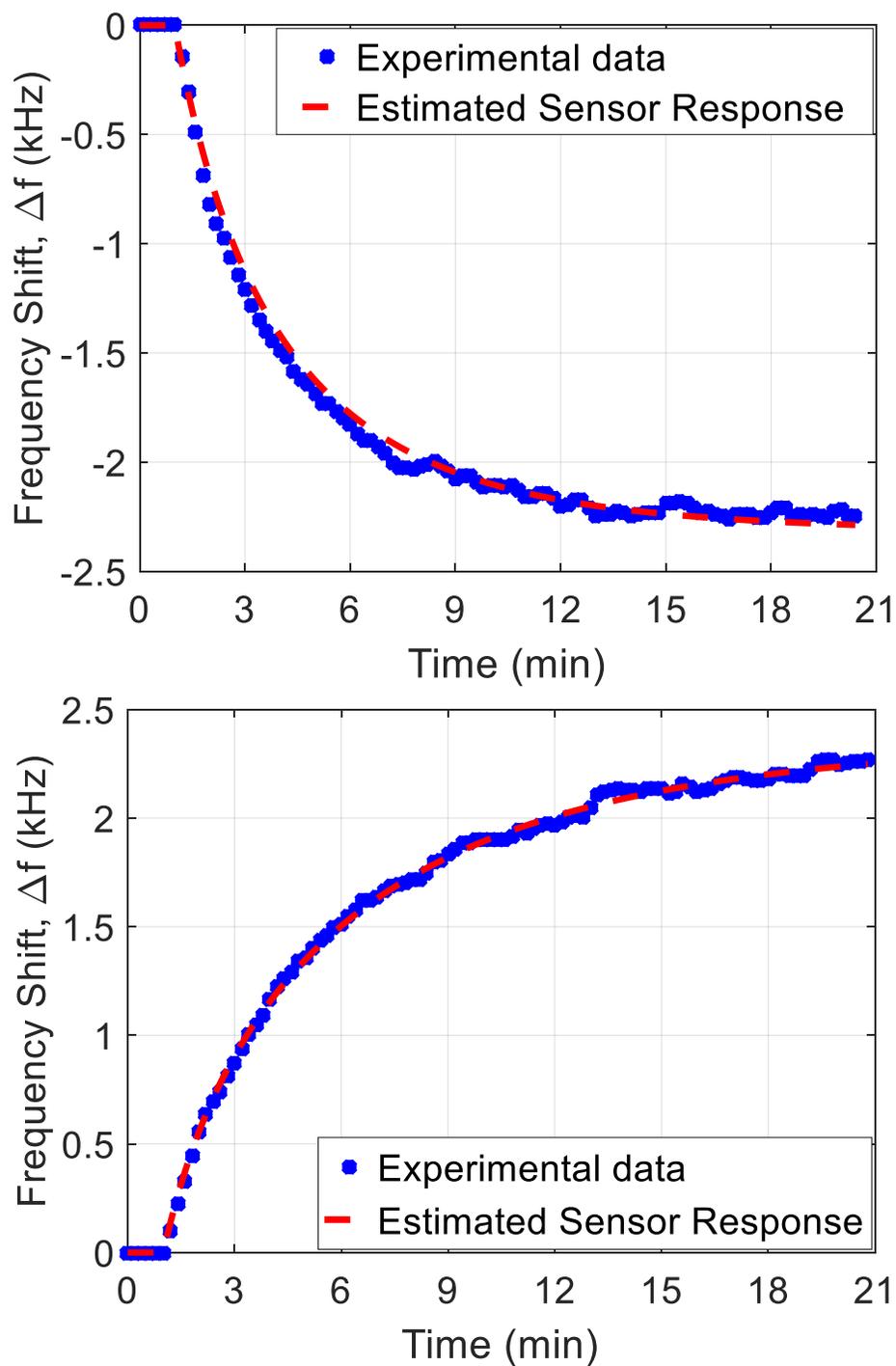


Figure 6.12: Measured response of a SH-SAW sensor coated with 0.8 μ m PIB to a LNAPL sample in groundwater containing 670 ppb benzene, 1340 ppb toluene, 510 ppb ethylbenzene/xylenes, and an unknown concentration of interferents (top: sorption data, bottom: desorption data). Also shown (red dashed line) are the estimated sensor responses obtained using five-analyte model.

Table 6.13: Estimated concentrations of BTEX compounds obtained using the measurement data of a LNAPL sample in groundwater (collected using a SH-SAW device coated with 0.8 μ m PIB) compared to concentrations measured using GC-PID.

Target Analyte	Analyte Concentrations from GC-PID (ppb)	Estimated Concentrations (ppb) [% difference with actual concentration]	
		Four-Analyte Model	Five-Analyte Model
Benzene	670	644 [4%]	688 [3%]
Toluene	1340	1491 [11%]	1398 [4%]
Ethylbenzene & Xylenes	510	584 [15%]	570 [12%]

About 100 measured sensor data with LNAPL/groundwater samples containing various BTEX concentrations ranging from low ppb to low ppm levels were tested using the proposed technique. The results from these tests are summarized in Figures 6.13, 6.14, 6.15, 6.16 and 6.17 for the four-analyte model and in Figures 6.18, 6.19, 6.20, 6.21 and 6.22 for the five-analyte model by plotting the estimated concentrations versus the concentrations measured by GC-PID. The concentrations measured by GC-PID provide an independent reference for comparison of the results obtained using the proposed technique. However, it is also noted that the GC-PID data are subject to an average of $\pm 7\%$ error [93]. Also shown in Figures 6.13-6.22 are the relative percentage errors between the estimated concentrations and concentrations determined using GC-PID, the $\pm 20\%$ error line (or in some cases, $\pm 15\%$ error line) and the ideal line with the slope of one which represents the ideal case when the estimated concentrations of the analytes are equal to the concentrations determined using GC-PID. Note that the closer a point in those figures to the ideal line, the more accurate the concentration estimate of the analytes. Based on Figures 6.13-6.22, it can be seen that most of the estimated

concentrations of BTEX compounds lie in close proximity to the ideal line. This indicates that the estimated concentrations are in very good agreement with the concentrations determined using GC-PID for the tested samples. Specifically, for the four-analyte model the estimated concentrations (utilizing both sorption and desorption) are within $\pm 9\%$ for benzene (as shown in Figure 6.15), $\pm 10\%$ for toluene (as shown in Figure 6.16), and $\pm 14\%$ for ethylbenzene and xylenes (as shown in Figure 6.17) and for the five-analyte model the estimated concentrations are within $\pm 7\%$ for benzene (as shown in Figure 6.20), $\pm 8\%$ for toluene (as shown in Figure 6.21), and $\pm 11\%$ for ethylbenzene and xylenes (as shown in Figure 6.22). Given the measurement error of the GC-PID instrument [93], this implies the estimated concentrations are in excellent agreement with the concentrations determined using GC-PID. Further analysis into the results obtained from these tests reveals that, for the five-analyte model, about 90% and about 95% of the tested data produces estimates which are within $\pm 15\%$ and $\pm 20\%$, respectively, of the concentrations determined using GC-PID.

The estimation results obtained show the ability of the proposed approach to detect and quantify target analytes in the presence of non-target interferents using only a single polymer coated sensor. Specifically, for the detection and quantification of BTEX compounds in the presence of interferents, the results obtained using the five-analyte model are slightly more accurate than the results obtained using the four-analyte model. This may be related to the representation of all the detectable interferents using these models. For the four-analyte model, all the detectable interferents in the sample were represented using only one exponential term whereas for the five-analyte model, all the detectable interferents in the sample were represented using two exponential terms. If

multiple interferents are present at significant concentrations, their representations using a single exponential term will introduce an error in the estimation of the analyte concentrations. This means that the target analytes can only be detected and quantified with high accuracy if the responses due to all the detectable interferents are modelled separately and accurately. Ideally, to obtain accurate estimation results, the model should accurately describe the sensor responses due to each analyte (which includes both the target analytes and non-target analytes) that interact with the selected polymer coatings. Thus, it is imperative to determine the interferents that interact with the chosen polymer coatings and the resulting sensor parameters associated with these interferents. Based on experiments conducted on several interferents commonly found in the contaminated groundwater using the selected polymer coatings, it has been identified that 1) the polymer coatings respond to only some interferents such 1,2,4-trimethylbenzene, naphthalene, *n*-heptane and MTBE, 2) the response to these interferents will either have larger response time constants or lower sensitivities than the target analytes and 3) the sensor response to 1,2,4-trimethylbenzene is more pronounced than the response to other detectable interferents. Based on the investigations, two example models were formulated i.e. for-analyte model and five-analyte model. In the four-analyte model, the response parameters of 1,2,4-trimethylbenzene were used as a representative of the response due to the interferents whereas, in the five-analyte model, the contribution of 1,2,4-trimethylbenzene to the sensor response is treated individually as one analyte, and the contributions of the other less dominant interferents in the groundwater are represented by a single common term in the model like a separate analyte. Therefore, the representation of interferents in the five-analyte model is slightly more accurate than the

four-analyte model. Hence, the results obtained using five-analyte model is slightly more accurate than the results obtained using four-analyte model. This clearly indicates that for a sample which contains multiple dominant interferents, a more accurate result is obtainable if all the dominant interferents are modeled separately using one exponential term. It should be noted that inaccurate representation of the response due to the detectable interferents in the sensor response model can greatly increase the error in the detection and quantification of the target analytes.

The utilization of both sorption and desorption data for signal processing provide more information about the analyte-specific interactions with the polymer film. This enables a more accurate detection and quantification of target analytes with improved accuracy, high tolerance to measurement noise and improved selectivity. The motivation to selectively include desorption data in the signal processing technique is based on the realization that the desorption transients, which are often more sensitive to energies of desorption of analytical targets, could provide additional information about analyte/polymer interactions.

Moreover, the results obtained indicate that among the BTEX compounds, the estimation errors for ethylbenzene and xylenes are the largest (in percent). This may be due to the low concentration range (low ppb range) tested for these compounds as well as the sensor detection limit for these compounds. In the low ppb range, the signal noise limits the accuracy of the estimated concentrations. Another factor contributing to these inaccuracies may be the simplifying assumption made in modeling the combined response of interferents in the mixture (i.e. as a single analyte using one exponential term in the four-analyte model and as two analytes using two exponential terms in the five-

analyte model). It is also noted that the estimation error for ethylbenzene and xylenes is not as critical as that for benzene, because the latter has a greater hazard potential and lower maximum contaminant level for drinking water. For benzene, an excellent estimation error within $\pm 10\%$ was found for both sensor coatings investigated.

By utilizing the proposed approach, BTEX compounds can be quantified in near real-time with high accuracy within approximately two minutes after data collection (including both sorption and desorption responses). For applications which only require high accuracy for the estimated concentrations of benzene, even faster processing time can be achieved by just processing the sorption response data. In this case, benzene can be quantified with high accuracy in real-time even before the sorption response reaches steady-state [46, 84]. Furthermore, the accurate results obtained clearly highlight the potential of the proposed approach to detect and quantify BTEX compounds in aqueous phase using only the data collected from a single polymer-coated SH-SAW device. If needed, in some applications, redundancy and improved detection limits could be achieved using the proposed approach with a small sensor array (consisting of 2 to 3 devices) coated with appropriately selected coatings.

It is important to point out here that the signal-to-noise ratio (SNR) of the measured frequency response also plays a significant role in the ability of the proposed technique to accurately detect and quantify the target analytes in the presence of interferences, especially for analyte concentrations close to the detection limit of the sensor. The presence of high noise in the measured frequency response data can greatly decrease the accuracy of the results obtained using the proposed technique. This is

especially true if the concentrations of the target analytes are in the low ppb range.

Therefore, it is necessary to ensure that the measured response has a very high SNR.

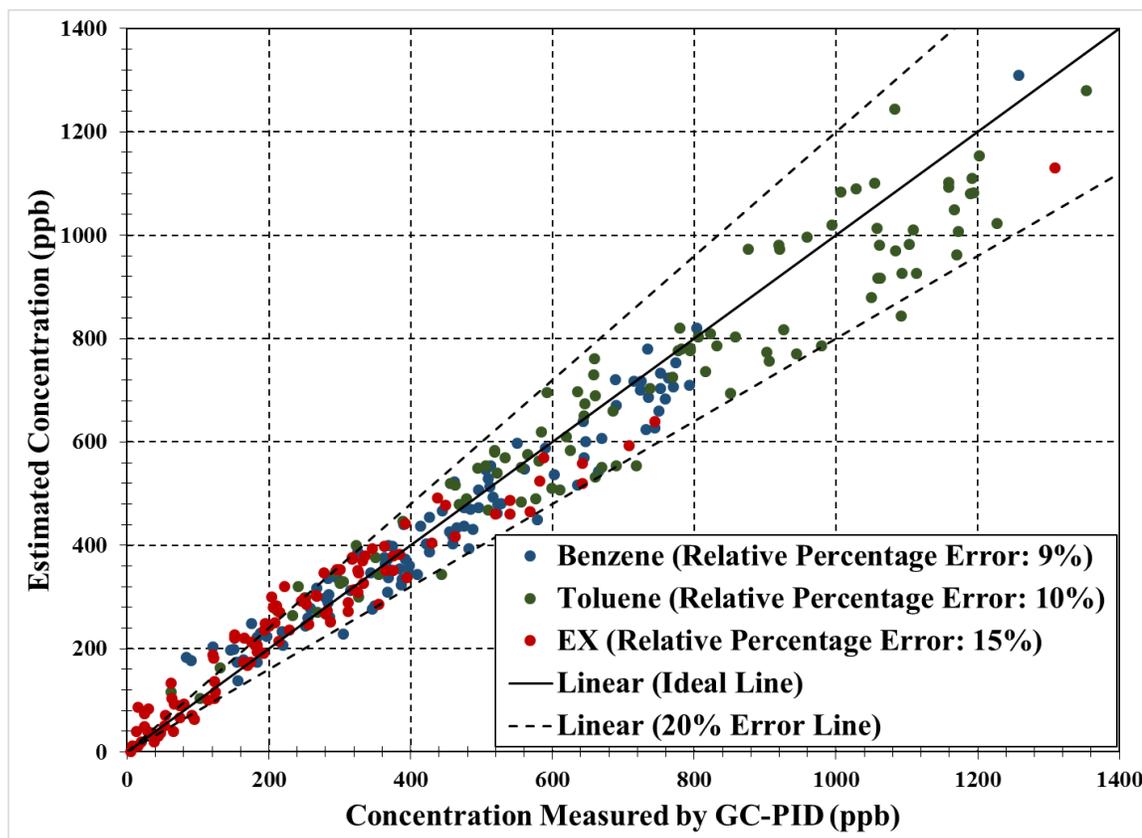


Figure 6.13: BTEX concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with $0.6\mu\text{m}$ PECH. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line.

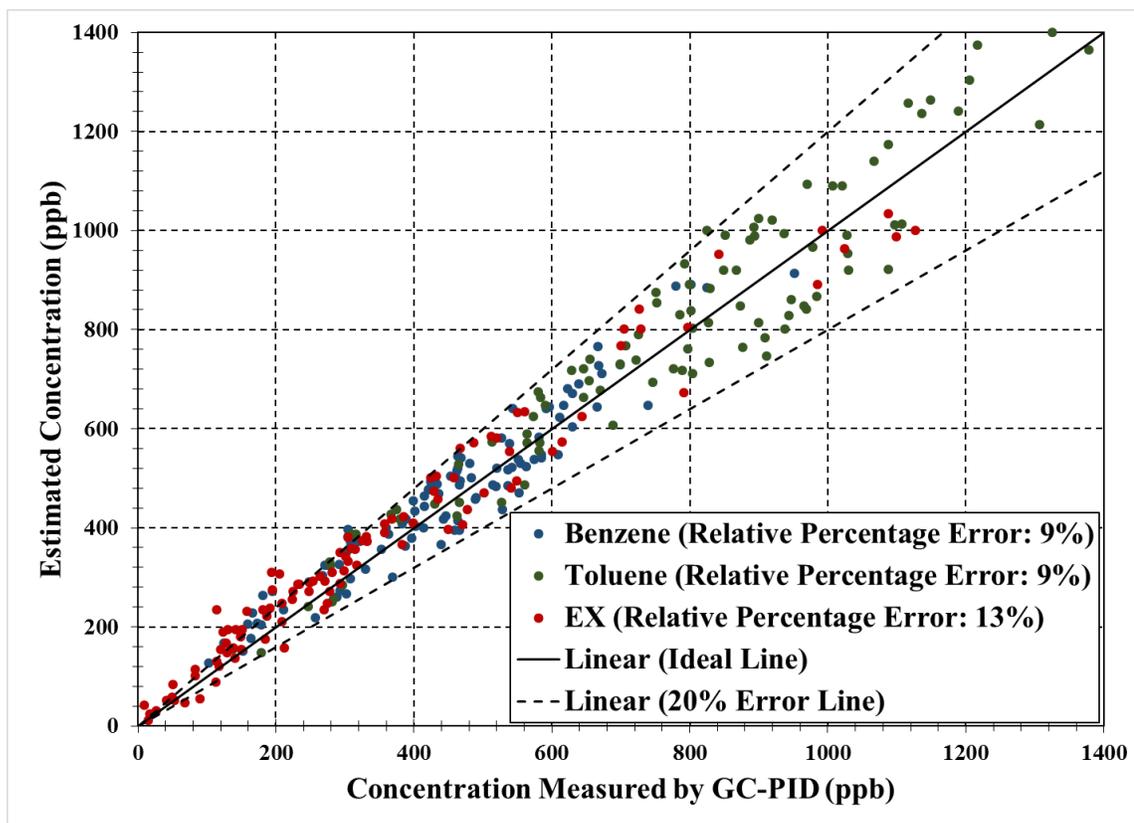


Figure 6.14: BTEX concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line.

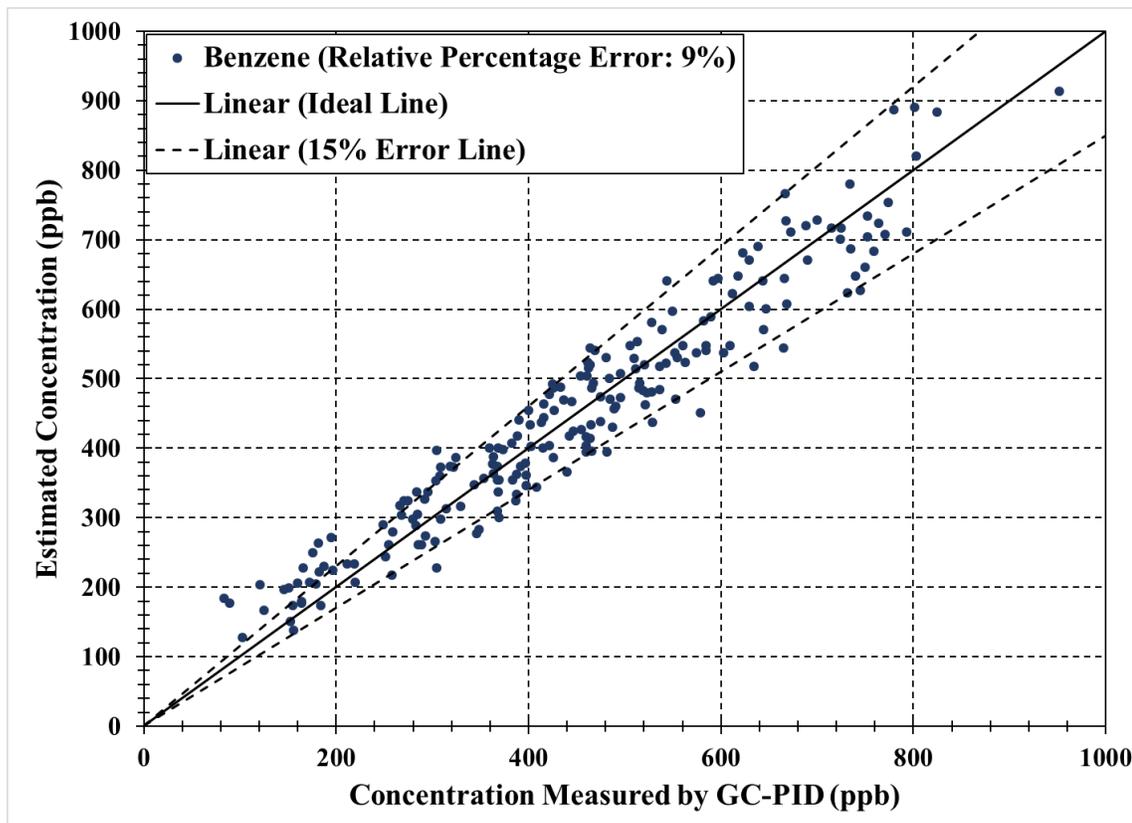


Figure 6.15: Benzene concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 15\%$ error line.

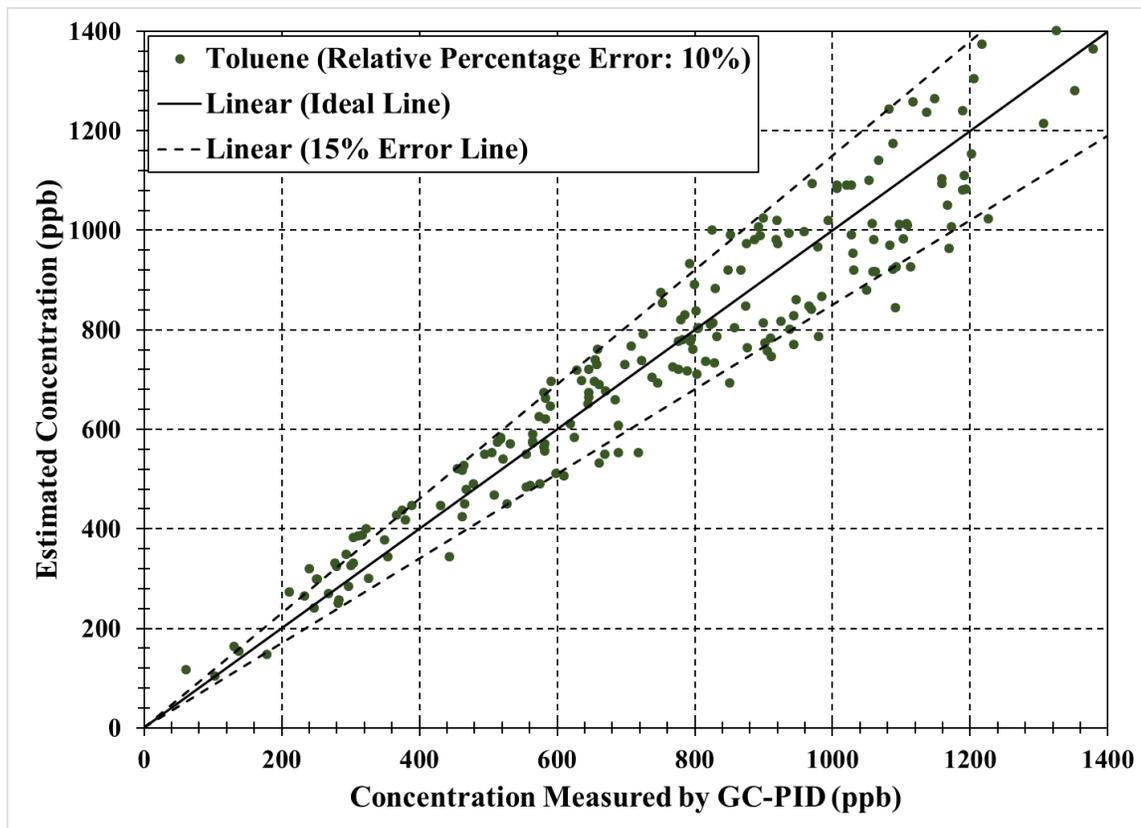


Figure 6.16: Toluene concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 15\%$ error line.

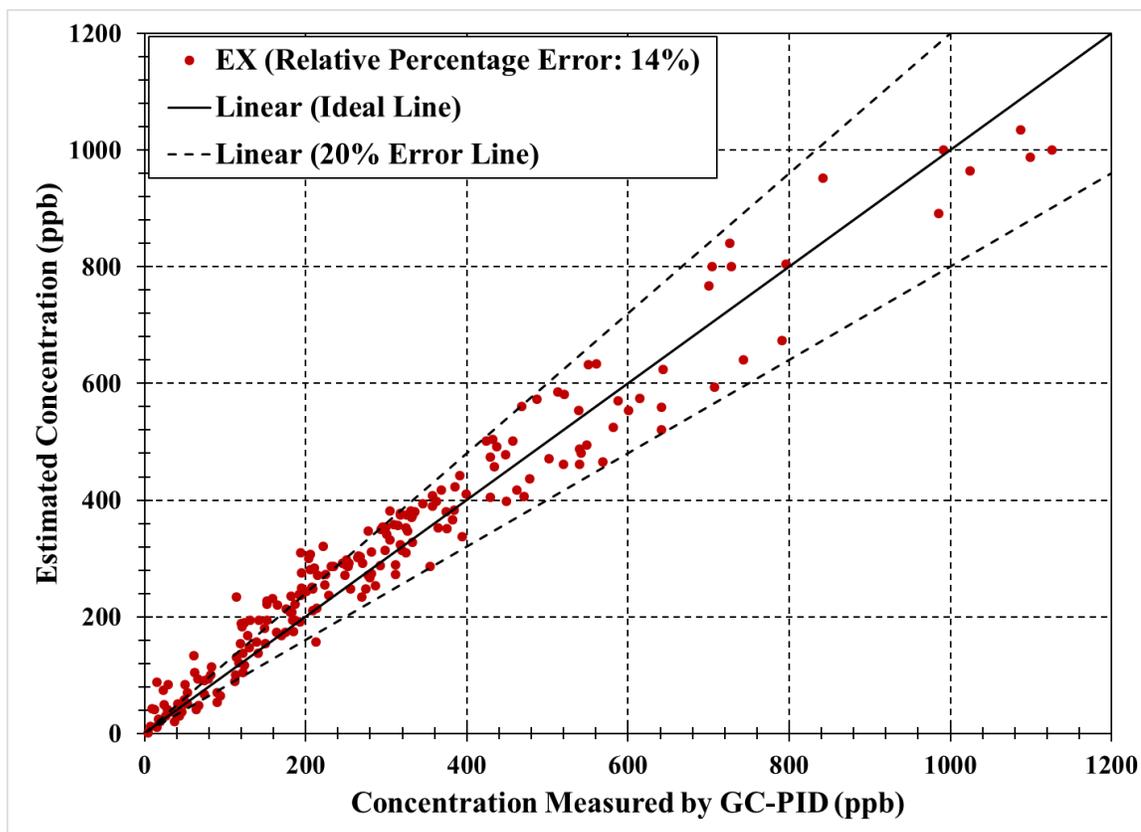


Figure 6.17: Ethylbenzene and xylenes (EX) concentrations estimated using the proposed signal processing technique and four-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line.

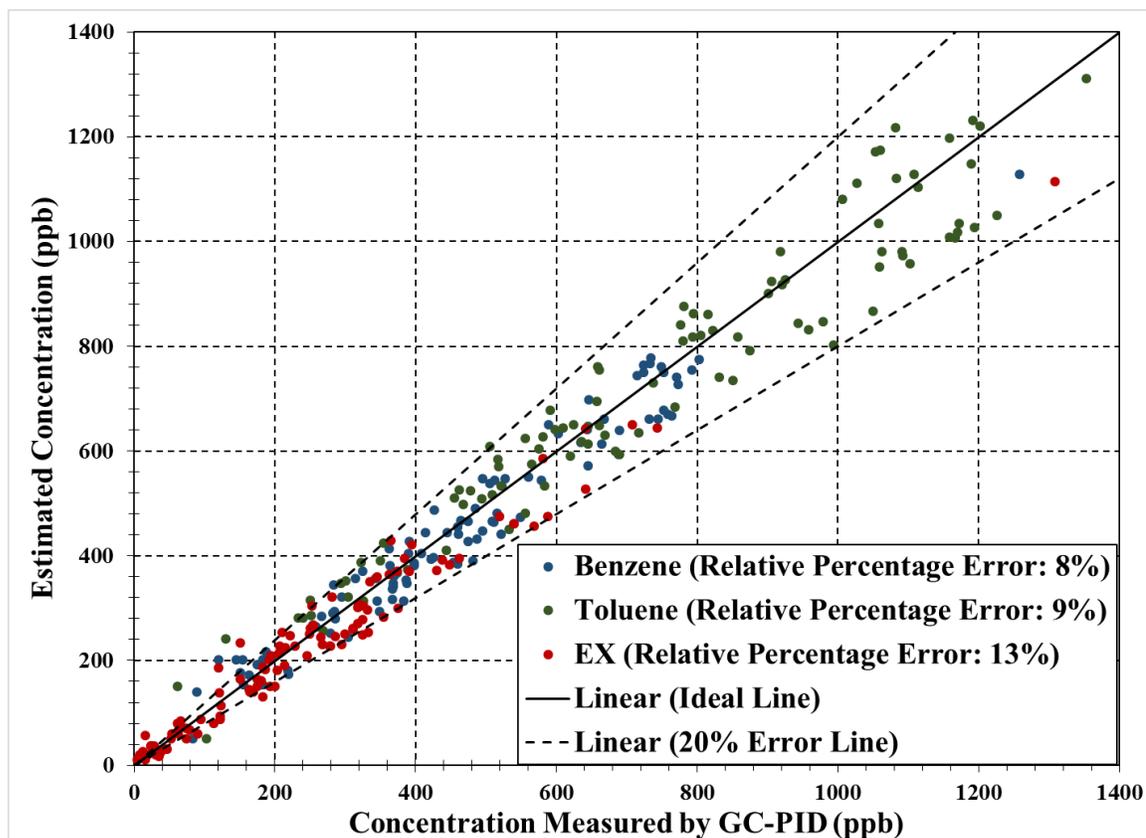


Figure 6.18: BTEX concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with 0.6 μ m PECH. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line.

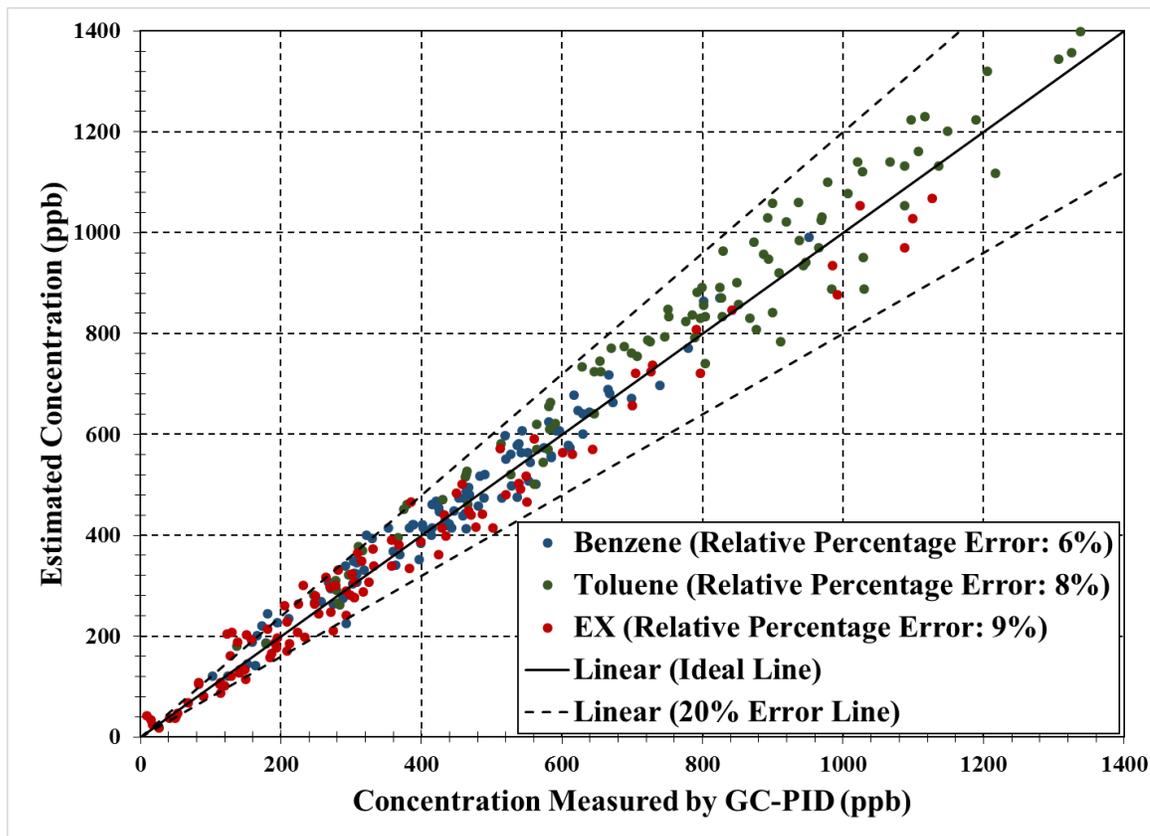


Figure 6.19: BTEX concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line.

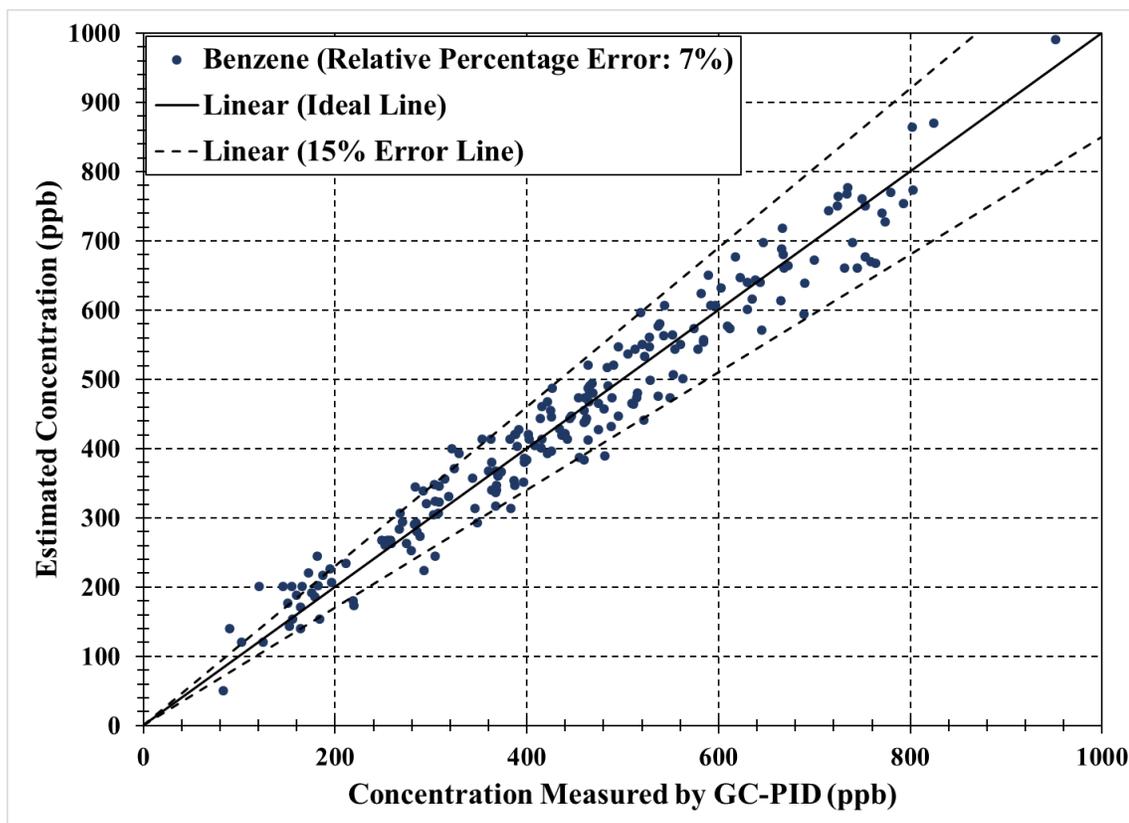


Figure 6.20: Benzene concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 15\%$ error line.

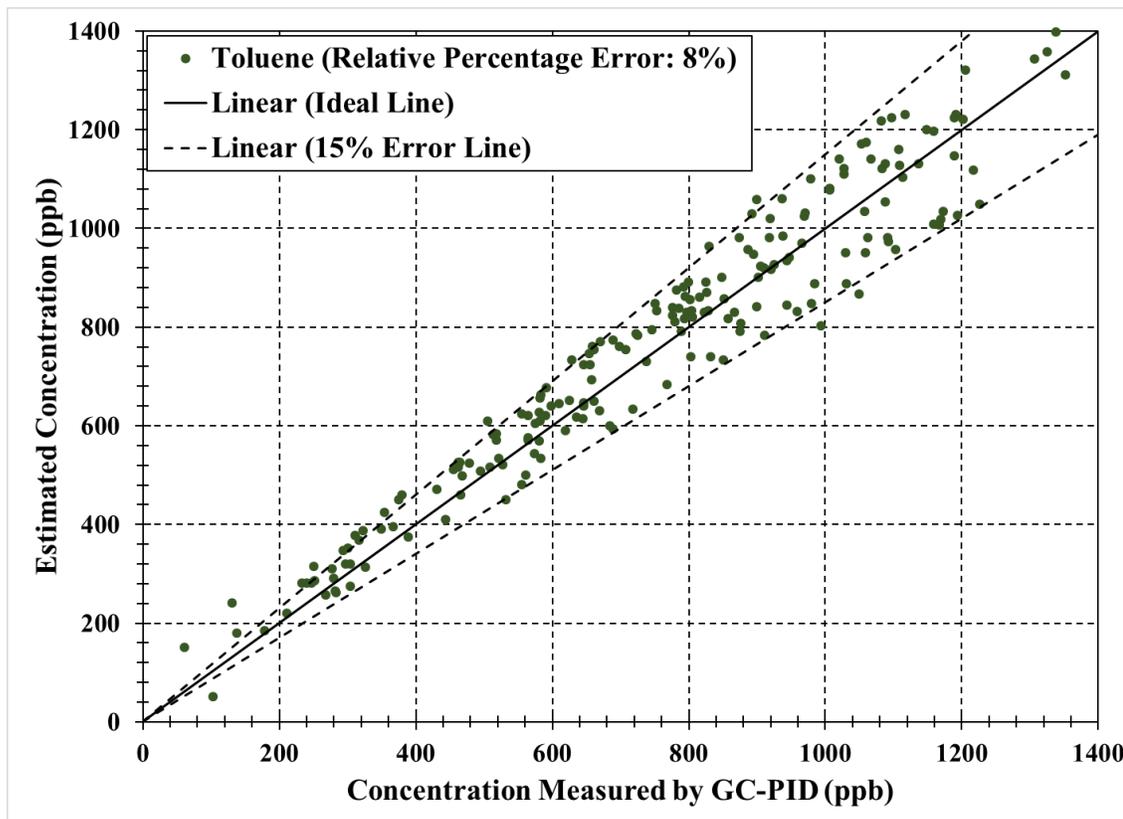


Figure 6.21: Toluene concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 15\%$ error line.

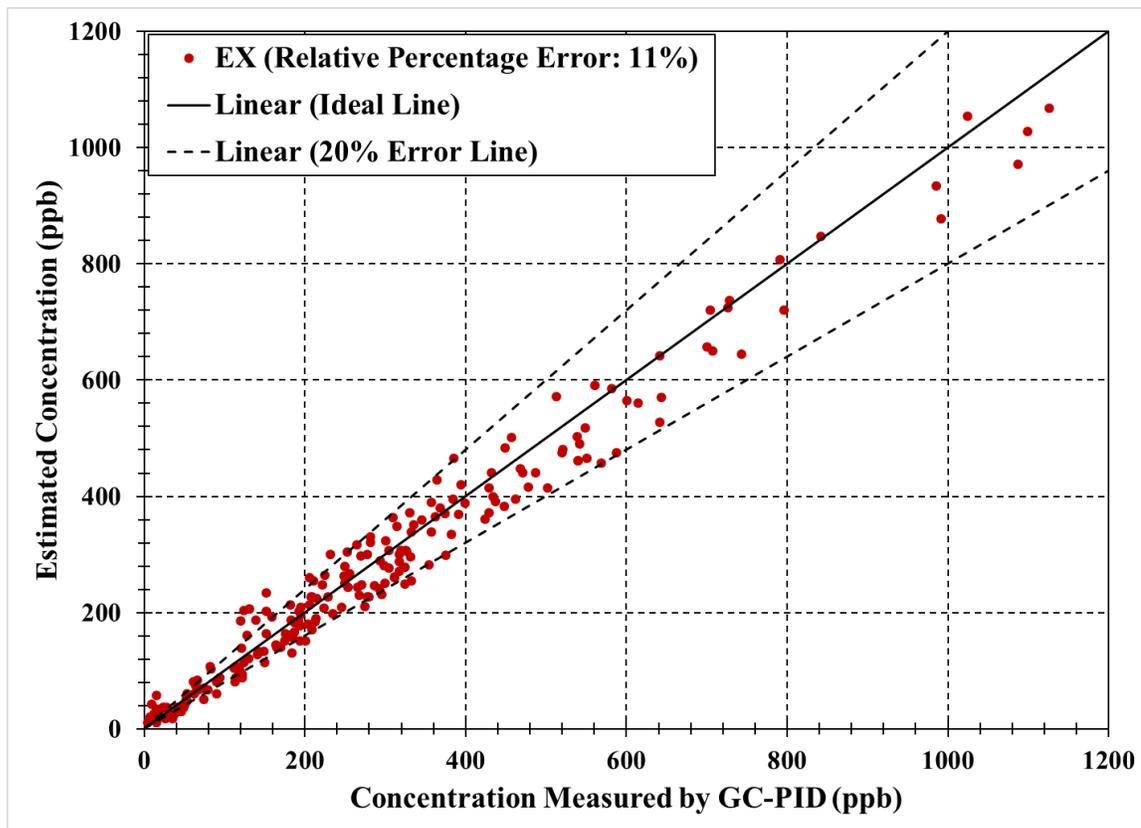


Figure 6.22: Ethylbenzene and xylenes concentrations estimated using the proposed signal processing technique and five-analyte model. The measured data were obtained from SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB. The legends show the average relative percentage error between the estimated and the GC-PID measured concentrations. The diagonal line represents the ideal case (estimated concentration = measured concentration). Also shown in the figure is the $\pm 20\%$ error line.

7 SUMMARY, CONCLUSIONS AND FUTURE WORK

7.1 Summary

The main objective of this work was to develop a signal processing technique that can enhance the ability of an array of polymer coated SH-SAW sensors or even a single polymer coated SH-SAW sensor to detect and quantify target analytes (i.e. benzene, toluene, ethylbenzene and xylenes (BTEX) compounds) in the presence of interferents in near real-time. Apart from the main objective, several other signal processing techniques for different cases leading towards the main objective were also investigated. All the cases investigated in this work are listed below:

- Case 1: Investigation of a signal processing technique for the detection and quantification of single analyte from sensor response to solution containing any arbitrary single analyte.
- Case 2: Investigation of a signal processing technique for the detection and quantification of binary mixtures from sensor response to solution containing any arbitrary binary mixture.
- Case 3: Investigation of a signal processing technique for the identification and quantification of multi-analyte mixtures using only the measured sensor response

from a single polymer coated sensor, given the characteristic response time constants and sensitivities of the analytes are known.

Case 4: Investigation of a signal processing technique for the detection and quantification of multiple target analytes in a chemical also containing various interferences using only the measured sensor response from a single polymer coated sensor.

Note that the investigations of Case 1 to Case 3 serve as building blocks towards the development of the signal processing technique for realizing the main objective, i.e. to detect and quantify target analytes in the presence of interferences in near real-time based only on the measured response of a single polymer coated SH-SAW sensor device (Case 4). In many applications, the signal processing techniques proposed in this work can be used as an alternative to that of a sensor array. In fact, in some applications, as opposed to the use of sensor arrays which are complex and have various drawbacks, the proposed technique will enable the development of a smart sensor system that can detect and quantify the analytes of interest in near real-time without the need of a complex training data set.

All the signal processing techniques proposed in this dissertation are based on estimation theory. Estimation-theory-based techniques were utilized because they offer various advantages, including near real-time data processing, minimal computational requirements, and minimal memory requirements for real-world implementations. In particular, the signal processing techniques proposed in this dissertation were based on

Bank of Kalman filters (BKFs) and/or exponentially weighted recursive least squares estimation (RLSE). Therefore, a detailed review of the theory and implementation of exponentially weighted RLSE and BKFs was given in Chapter 2.

The use of estimation-theory-based techniques for the detection and quantification of target analytes in the presence (or absence) of interferents requires an accurate analytical model that describes the response of the SH-SAW sensor to the samples containing the target analytes. Therefore, sensor response models were developed for different cases based on the empirical data collected for the investigated polymer coatings. These models were discussed in detail in Chapter 3. All the formulated sensor response models utilize two sensor parameters, i.e. the equilibrium frequency shift and the response time constant (for individual analyte), the latter being specific for each combination of coated device and analyte. First, the model of sensor responses to the single analytes was discussed. The single analyte model was formulated based on suitable assumptions (these assumptions were listed in Chapter 3, section 3.2) and serves as the basis for the multi-analyte sensor response model. The general model of the multi-analyte sensor responses was then developed and modified based on the empirical results obtained for the sensor responses to the common non-target interferents found in the groundwater. A detailed study on the common groups of non-target interferents that are known to interact with the selected polymer coated SH-SAW sensors was performed in order to formulate accurate models for the detection of BTEX compounds in the presence of non-target interferents. Based on those studies, several assumptions were made to formulate a general model for detection of samples containing n analytes consisting of both analytes of interest and interferents. The formulated general model of the multi-

analyte sensor responses was then normalized, discretized and transformed into state-space form, so that exponentially weighted RLSE and BKF's can be applied to identify and quantify the analytes of interest. Furthermore, the necessary modifications that need to be made in the sensor response model when non-ideal cases occur were also addressed in this dissertation. Specifically, two non-ideal cases were discussed: non-step-like concentration versus time profile (i.e. the transition from clean water to the sample in the flow cell containing the sensor is not sufficiently fast) and concentration-dependent sensitivity of the sensor. The concentration versus time profile seen by the sensor depends on a number of measurement system parameters such as the average flow speed of the sample, the total length of the tube separating point of sample introduction and the sensor device, and diffusion coefficient of the soluble substance in the sample. Thus, the modifications that need to be made in the sensor response model for the non-step-like concentration versus time profile were discussed. For most sensors, for large analyte concentrations, the sensitivity of the sensor depends on the concentration range of the analyte (or sample). Therefore, the modifications that can be made in the sensor response model when the sensor exhibits concentration-dependent sensitivity were discussed.

Moreover, the details and implementations of the proposed sensor signal processing techniques for all the investigated cases were discussed. The proposed signal processing techniques for all the four cases discussed above, along with the advantages and disadvantages of the proposed techniques, are summarized in Table 7.1.

Table 7.1: Summary of the proposed signal processing techniques for the investigated cases.

Descriptions	Proposed Signal Processing Technique
<p>Case 1: Detection and quantification of single analyte from single analyte sensor response</p>	<ul style="list-style-type: none"> • Bank of Kalman filters <ul style="list-style-type: none"> – Advantages: <ol style="list-style-type: none"> 1) Online detection and quantification of the analyte. 2) Information about the analyte causing the response is not required. 3) Alternative to sensor array. – Disadvantage: Requires considerable computational power.
<p>Case 2: Detection and quantification of binary mixtures of analytes from binary mixture sensor response</p>	<ul style="list-style-type: none"> • Bank of Kalman filters <ul style="list-style-type: none"> – Advantages: <ol style="list-style-type: none"> 1) Online detection and quantification of the analytes. 2) Information about the analytes causing the response is not required. 3) Alternative to sensor array. – Disadvantage: Requires considerable computational power.
<p>Case 3: Identification and quantification of multi-analyte mixtures</p>	<ul style="list-style-type: none"> • Multi-stage exponentially weighted RLSE <ul style="list-style-type: none"> – Advantages: <ol style="list-style-type: none"> 1) Online identification and quantification of the analytes. 2) Alternative to sensor array. 3) Can be used for analysis of up to n analytes in the mixture, provided the characteristic response time constants and sensitivities of the analytes are known.
<p>Case 4: Detection and quantification of target analytes in the presence of interferents</p>	<ul style="list-style-type: none"> • Two-step processing: Exponentially weighted RLSE and bank of Kalman filters <ul style="list-style-type: none"> – Advantages: <ol style="list-style-type: none"> 1) Near real-time detection and quantification of the target analytes. 2) Alternative to sensor array 3) Does not require information on approximate initial concentration range of the BTEX analytes.

In order to show the validity of the proposed signal processing techniques for all the studied cases, these techniques were tested using experimental sensor response data collected in the Microsensor Research Laboratory using polymer coated SH-SAW sensors. Thus, before discussing the results obtained using the proposed signal processing techniques, the fundamentals of SH-SAW sensors, the polymer coatings chosen for this work and the details of measurement setup and data acquisition were reviewed first. In particular, the data analyzed in this dissertation were collected using three-layer (piezoelectric substrate, polymer layer and liquid layer) SH-SAW structure in a delay line configuration. The chosen polymer coatings for the detection of the analytes of interest were poly(ethyl acrylate) (PEA), poly(epichlorohydrin) (PECH), and poly(isobutylene) (PIB). All the experimental data collected using the sensor system exhibit baseline drift during the response and sometimes outlier points were also recorded. Therefore, pre-processing was performed first in order to correct the measured data for baseline drift and to eliminate outlier points before testing the data using the proposed signal processing techniques.

Finally, the results for the experimental verification of the proposed sensor signal processing techniques for all the four different cases were presented and discussed. In order to evaluate the performance of the proposed signal processing techniques, the estimated analyte concentrations were compared to the results obtained independently using GC-PID. The results obtained from these tests for all the four different cases are summarized in Table 7.2.

Table 7.2: Summary of the results obtained using the proposed signal processing techniques for the investigated cases.

Proposed Techniques	Summary of Results
<p>Case 1: Online Detection and Quantification of Single Analyte Using Bank of Kalman Filters</p>	<ul style="list-style-type: none"> - The proposed technique was able to accurately identify and quantify the analyte in real-time. - The percentage difference between the estimated and the measured analyte concentration is less than $\pm 10\%$. - Minimum time-to-detection: well before the response reaches steady-state.
<p>Case 2: Online Detection and Quantification of Binary Mixtures Using Bank of Kalman Filters</p>	<ul style="list-style-type: none"> - The proposed technique was able to accurately identify and quantify the two analytes in the samples rapidly. - The percentage difference between the estimated and the measured analyte concentrations is within $\pm 15\%$. - Minimum time-to-detection: well before the response reaches steady-state (slightly longer than Case 1).
<p>Case 3: Online Identification and Quantification of Multi-Analyte Mixtures Using Multi-Stage Exponentially Weighted RLSE</p>	<ul style="list-style-type: none"> - The proposed technique was able to accurately identify and quantify the analytes in the samples rapidly. - The technique can be used for the identification and quantification of n analytes (up to 5 analytes tested) in a sample (given the characteristic response time constants and sensitivities of the analytes are known). - The percentage difference between the estimated and the measured analyte concentrations is within $\pm 20\%$. - Minimum time-to-detection: well before the response reaches steady-state (slightly longer than Case 1 and 2 but computationally less expensive than Case 1 and 2).

<p>Case 4: Near Real-Time Detection and Quantification of Multi-Analyte Mixtures in the Presence of Interferents Using Exponentially Weighted RLSE and Bank of Kalman Filters</p>	<ul style="list-style-type: none"> - The proposed technique was able to accurately detect and quantify the target analytes in the presence of interferents in near real-time. - Four-analyte model: The estimated concentrations for target analytes i.e. benzene, toluene, and ethylbenzene-plus-xylenes fall within $\pm 9\%$, $\pm 10\%$, and $\pm 15\%$, respectively, of the measured concentration. - Five-analyte model: The estimated concentrations for benzene, toluene, and ethylbenzene-plus-xylenes fall within $\pm 8\%$, $\pm 9\%$, and $\pm 13\%$, respectively, of the measured concentration. - Time-to-detection: Approximately two minutes after data collection (including both sorption and desorption responses).
--	--

7.2 Conclusions

One of the major contributions of this work is in the development of analytical models that describe the sensor response of the polymer coated SH-SAW sensors to the samples containing the analytes of interest. Specifically, the models of the coated SH-SAW sensor responses to single analytes, multiple analytes and multiple target analytes in the presence of non-target interferents were formulated. These models were developed based on suitable assumptions (as listed in Chapter 3) and empirical data collected for the investigated polymer coatings. Two sensor parameters were utilized in these models, i.e. the sensitivity (equilibrium frequency shift) and the response time constant which is unique to each coated device and analyte. The formulated general model of the multi-analyte sensor response can be used to describe the SH-SAW sensor response to any

number of analytes in the sample, provided that each analyte and possible interferences in the sample (in the concentration range of interest) has been separately characterized for the selected polymer coating. In order to use the formulated general model of the multi-analyte sensor response in combination with the estimation-theory-based techniques, the general model was normalized, discretized and transformed into state-space form.

Another main contribution of this work is in the development of novel sensor signal processing techniques for several different cases based on estimation theory. In particular, BKFs and/or exponentially weighted RLSE was used for sensor signal processing. The proposed techniques were tested and validated using extensive measured data. Based on the results obtained from these tests, the following conclusions can be made:

- 1) In general, all the proposed sensor signal processing techniques (for all the cases investigated) utilized two sensing parameters, i.e. sensitivity and response time constant. The utilization of two sensing parameters contributes significantly towards the accurate detection and quantification results obtained using these techniques.
- 2) All the proposed signal processing techniques have the ability to enhance the selectivity of a single polymer coated sensor to enable real-world applications. Typically, a sensor array is used to improve the selectivity of a chemical sensor for a particular application. However, there are various drawbacks associated with using a sensor array including increased signal- and data- processing time and potential misclassification for complex mixtures (i.e. mixtures with more than two analytes).

Thus, the proposed sensor signal processing techniques can be used as a viable alternative to a sensor array. In comparison to a sensor array, the proposed techniques have the advantage of not requiring a complex training data set and enabling real-time (or near real-time) detection and quantification of the analytes of interest (even in the presence of interferences) based on the sensor response from only a single sensor device.

- 3) The ability to process the sensor data in real-time (or in near real-time) using the proposed signal processing techniques will enable rapid detection of analytes of interest. The sensor response to some analytes may take a relatively long time to reach steady-state. In such cases, the use of conventional signal processing techniques would require waiting until the sensor response reaches steady-state before the analytes of interest could be identified and quantified. However, through the utilization of the proposed sensor signal processing techniques, the analytes of interest (even in the presence of interferences) could be identified and quantified rapidly. Thus, it can be concluded that by utilizing the proposed techniques, time to detection and quantification of analytes of interest even in the presence of interferences could be reduced significantly (in some cases well before the sensor response reaches steady-state). Note that the ability to identify and quantify hazardous analytes in the environment is critical. By using the proposed sensor signal processing techniques and appropriate sensor platform, these hazardous compounds in the environment could be detected and quantified rapidly, so that mitigation plans and remediation actions could be carried out earlier and efficiently. Additionally, rapid detection and

quantification of analytes of interest will also shorten the sensor exposure time which in turn will improve the accuracy, repeatability and longevity of the polymer coatings used with the sensor.

- 4) All the proposed sensor signal processing techniques can be implemented using a microcontroller. Therefore, these techniques can be employed for the development of a small, portable, and cost-effective smart sensor system for in-situ applications such as groundwater monitoring, the monitoring of the plume in a sub-surface marine oil spill, and spill clean-ups.
- 5) The proposed sensor signal processing techniques were demonstrated for the detection and quantification of BTEX compounds (in the presence or absence of interferences). However, these techniques could also be used for the detection and quantification of other chemical analytes, provided the analytes have been characterized appropriately for the selected polymer coatings.
- 6) For Case 4 (near real-time detection and quantification of multi-analyte mixtures in the presence of interferences using exponentially weighted RLSE and BKFs), the technique is only capable of near real-time detection and quantification of analytes of interest because the proposed technique is based on a two-step processing. In the two-step processing, initial detection and quantification is performed using exponentially weighted RLSE and the results obtained from the first step is used as an input to the second step using BKFs. The first step can be performed in real-time as the data are being collected, however, the second step can only be performed after the first step is

completed. Nonetheless, the results obtained indicate that the proposed technique is capable of accurately detecting and quantifying the target analytes in approximately two minutes after the data collection.

- 7) For Case 4, it can be concluded that the results obtained using the five-analyte model are slightly more accurate than the results obtained using the four-analyte model, as expected. This is due to the difference in the representation of the detectable interferents between these two models. Hence, it can be inferred that the target analytes can only be detected and quantified with high accuracy if the response due to all the detectable interferents is modelled accurately.

It is pertinent to note that the sensor signal processing techniques presented in this dissertation are independent of the sensing platform used to detect the analytes of interest (i.e. not specific to the SH-SAW sensor platform). This means that the proposed signal processing techniques could also be used together with the response of other sensor platforms such as MEMS-based sensors (e.g. microcantilevers), optical chemical sensors and other types of acoustic wave-based sensors, provided the sensor responses to these platforms can be modeled analytically.

7.3 Future Work

The work presented in this dissertation could be expanded upon and further improved. Further improvements in the sensor response models, sensitivity and

selectivity of the sensor and the sensor signal processing techniques may be possible. In this section a few possible future research proposals are listed:

1) In this dissertation, the models of the polymer coated SH-SAW sensor responses to single analyte and multiple analytes were presented. These models were formulated by making several assumptions. As a possible extension of this work, some of these assumptions could be relaxed in order to obtain a more general model that could better represent the sensor responses in some cases. Some possible investigations that could be performed to generalize the sensor response models are listed below:

- The formulated sensor response models are for type I sorption where the analyte concentration in the coating are proportional to the ambient concentration of the analyte [83, 84]. However, in general, this assumption is only valid for a low concentration range of the analytes. Thus, as a possible extension of the sensor response model, one could investigate and develop a more general sensor response model that can account for larger concentration ranges of the analyte. In order to develop such a model, different sorption isotherms need to be considered. For instance, the BET sorption isotherm (named after the founders of the theory Brunauer, Emmett and Teller) can be considered to model the physisorption at higher concentrations of the analytes [99, 100]. Therefore, a study could be performed to generalize further the existing sensor response model for a larger concentration range of the analytes.

- The sensor response times to analytes of interest are dependent upon the sample flow rate. Choosing the optimum sample flow rate, especially for liquid phase detection, is crucial for minimizing the hydrodynamic coupling and sensor noise. Often, the flow rate will be adjusted until a reasonable trade-off is found between the sample flow-rate and the resulting sensor noise. Any changes in the sample flow rate would affect the response time of the sensor. Therefore, one could also study the possibility of developing a sensor response model that explicitly depends upon the sample flow rate. Through the development of such a model, the response time of a sensor for different flow rates could be predicted.
- 2) The accuracy of the results obtained using the proposed sensor signal processing techniques are dependent upon the accuracy of the analytical model that describes the response of the sensor to the analytes of interest. Any mismatch between the analytical model and the actual physical processes causing the response would trigger inaccuracy in the detection and quantification of the analytes. For instance, a mismatch between the response time constant of a particular analyte used in the sensor response model and the actual physical processes causing the response could lead to undesirable results. Therefore, a study on the effect of model mismatch could be performed in order to develop a model mismatch compensator. Through the development of such compensator, the signal processing technique can be made more robust against mismatch in the sensor response model. For example, this would help the sensor system to compensate for slight changes in flow rate, temperature, etc. that will affect response times.

- 3) Common issues for any chemical sensor technologies are the lack of adequate sensitivity and selectivity. Several advanced sensor signal processing techniques based on estimation theory were introduced to enhance the selectivity of a chemical sensor. As for improving the sensitivity of a chemical sensor, the selection of the chemically sensitive element is critical. This is because the chemically sensitive layer plays a significant role in dictating the desired degree of sensitivity to the analytes of interest. Therefore, an investigation on finding and synthesizing more sensitive polymer layers to the analytes of interest could be carried out. Through the development of highly sensitive polymer layers, a very low concentration of analytes of interest could be detected.
- 4) All the sensor signal processing techniques presented in this dissertation were based on estimation theory, in particular, exponentially weighted RLSE and/or BKF. Besides exponentially weighted RLSE and BKF, there are several other estimation-theory-based techniques that could also be utilized for sensor signal processing. Therefore, as a future work in this area, one could investigate the feasibility of using other estimation-theory-based techniques such as Monte Carlo approaches, least mean squares filter and Wiener filter for sensor signal processing. Additionally, one could also investigate a completely different approach for sensor signal processing that is not based on estimation theory. For instance, one could investigate the use of machine learning techniques such as deep learning, sparse dictionary learning, clustering, neural networks, and Bayesian networks for sensor signal processing.

REFERENCES

- [1] Expanding the vision of sensor materials. Washington, D.C.: National Academy Press, 1995.
- [2] A.K.Mensah-Brown. "Analysis of the Detection of Organophosphate Pesticides in Aqueous Solutions Using Polymer-Coated SH-SAW Devices," Ph.D. Dissertation, Marquette University, Milwaukee, WI, U.S.A, 2010.
- [3] P. Carson and C. Mumford, Hazardous Chemicals Handbook, 2nd ed. Woburn, MA: Butterworth-Heinemann, 2002.
- [4] L. M. Dorozhkin and I. A. Rozanov, "Acoustic Wave Chemical Sensors for Gases," Analytical Chemistry, vol. 56, no. 5, pp. 399-416, 2001.
- [5] P. Grundler, Chemical Sensors: An Introduction for Scientists and Engineers. Leipzig: Springer, 2007.
- [6] J. Fraden, Handbook of Modern Sensors: Physics, Designs and Applications. New York: Springer-Verlag Inc, 2004.
- [7] National Research Council, Expanding the Vision of Sensor Materials. Washington D.C.: National Academy Press, 1995, pp. 3.
- [8] D. S. Ballantine, R. M. White, S. J. Martin, A. J. Ricco, G. C. Frye, E. T. Zellers and H. Wohltjen, Acoustic Wave Sensors: Theory, Design and Physico-Chemical Applications San Diego: Academic Press, 1997.
- [9] K. Kalantar-zadeh and B. Fry, Nanotechnology enabled sensors. Boston, MA: Springer, 2008.
- [10] A. Hulanicki, S. Glab and F. Ingman, "Chemical Sensors Definitions and Classification," Pure & Appl. Chem., Vol. 63, No.9, pp. 1247-1250, 1991.
- [11] A. Lobnik, M. Turel, and S. K. Urek, "Optical Chemical Sensors: Design and Applications," in Advances in Chemical Sensors, W. Wen, Ed. ISBN: 978-953-307-792-5, InTech, DOI: 10.5772/31534. [Online] Available: <https://www.intechopen.com/books/advances-in-chemical-sensors/optical-chemical-sensors-design-and-applications>. [Accessed: 13- Nov- 2017].
- [12] E. Bakker and Y. Qin, "Electrochemical Sensors", Analytical Chemistry, vol. 78, no. 12, pp. 3965-3984, 2006.
- [13] L. Baxter, Capacitive sensors: Design and Applications. New York, NY: The Institute of Electrical and Electronics Engineers, Inc., 1997.

- [14] A. Mamishev, K. Sundara-Rajan, Fumin Yang, Yanqing Du and M. Zahn, "Interdigital sensors and transducers", *Proceedings of the IEEE*, vol. 92, no. 5, pp. 808-845, 2004.
- [15] F. Banica, *Chemical sensors and biosensors: fundamentals and applications*. Chichester, West Sussex, United Kingdom: Wiley, 2012.
- [16] T. Stievater, W. Rabinovich, N. Papanicolaou, R. Bass and J. Boos, "Measured limits of detection based on thermal-mechanical frequency noise in micromechanical sensors", *Applied Physics Letters*, vol. 90, no. 5, p. 051114, 2007.
- [17] A. Nannini, D. Paci, F. Pieri and P. Toscano, "A CMOS-compatible bulk technology for the fabrication of magnetically actuated microbalances for chemical sensing", *Sensors and Actuators B: Chemical*, vol. 118, no. 1-2, pp. 343-348, 2006.
- [18] N. Lavrik, M. J. Sepaniak and P. G. Datskos, "Cantilever transducers as a platform for chemical and biological sensors," *Review of Scientific Instruments*. Vol. 75, no.7, pp. 2229-2253, July 2004.
- [19] M. Wenzel, "Polymer-Coated and Polymer-Based Microcantilever Chemical Sensors: Analysis and Sensor Signal Processing", Ph.D, Marquette University, 2009.
- [20] D. Ivanov, "Chemical Sensitivity of the Thickness-Shear-Mode Quartz-Resonator Nanobalance", *Journal of The Electrochemical Society*, vol. 143, no. 9, p. 2835, 1996.
- [21] S. Rösler, R. Lucklum, R. Borngräber, J. Hartmann and P. Hauptmann, "Sensor system for the detection of organic pollutants in water by thickness shear mode resonators", *Sensors and Actuators B: Chemical*, vol. 48, no. 1-3, pp. 415-424, 1998.
- [22] R. Patel, R. Zhou, K. Zinszer, F. Josse and R. Cernosek, "Real-Time Detection of Organic Compounds in Liquid Environments Using Polymer-Coated Thickness Shear Mode Quartz Resonators", *Analytical Chemistry*, vol. 72, no. 20, pp. 4888-4898, 2000.
- [23] H. Wohltjen and R. Dessy, "Surface Acoustic Wave Probes for Chemical Analysis. I and II," *Analytical Chemistry*, vol. 51, no. 9, pp. 1458-1470, 1979.
- [24] S. Martin, G. Frye and S. Senturia, "Dynamics and Response of Polymer-Coated Surface Acoustic Wave Devices: Effect of Viscoelastic Properties and Film Resonance", *Analytical Chemistry*, vol. 66, no. 14, pp. 2201-2219, 1994.

- [25] F. Bender, R. Cernosek and F. Josse, "Love-wave biosensors using cross-linked polymer waveguides on LiTaO₃ substrates", *Electronics Letters*, vol. 36, no. 19, p. 1672, 2000.
- [26] F. Josse, F. Bender and R. Cernosek, "Guided Shear Horizontal Surface Acoustic Wave Sensors for Chemical and Biochemical Detection in Liquids", *Analytical Chemistry*, vol. 73, no. 24, pp. 5937-5944, 2001.
- [27] F. Josse, "Acoustic wave liquid-phase-based microsensors", *Sensors and Actuators A: Physical*, vol. 44, no. 3, pp. 199-208, 1994.
- [28] J. Grate, S. Wenzel and R. White, "Flexural plate wave devices for chemical analysis", *Analytical Chemistry*, vol. 63, no. 15, pp. 1552-1561, 1991.
- [29] G. McHale, M. Newton and F. Martin, "Theoretical mass sensitivity of Love wave and layer guided acoustic plate mode sensors", *Journal of Applied Physics*, vol. 91, no. 12, p. 9701, 2002.
- [30] S. Zaromb and J. Stetter, "Theoretical basis for identification and measurement of air contaminants using an array of sensors having partly overlapping selectivities", *Sensors and Actuators*, vol. 6, no. 4, pp. 225-243, 1984.
- [31] K. Albert, N. Lewis, C. Schauer, G. Sotzing, S. Stitzel, T. Vaid and D. Walt, "Cross-Reactive Chemical Sensor Arrays", *Chemical Reviews*, vol. 100, no. 7, pp. 2595-2626, 2000.
- [32] A. Ricco, R. Crooks and G. Osbourn, "Surface Acoustic Wave Chemical Sensor Arrays: New Chemically Sensitive Interfaces Combined with Novel Cluster Analysis To Detect Volatile Organic Compounds and Mixtures", *Accounts of Chemical Research*, vol. 31, no. 5, pp. 289-296, 1998.
- [33] J. Gardner and P. Bartlett, *Electronic noses: Principles and Applications*. New York: Oxford University Press, 1999.
- [34] R. Duda, P. Hart and D. Stork, *Pattern classification*. New York: Wiley, 2001.
- [35] A. Webb and K. Copsey, *Statistical pattern recognition*. Hoboken: Wiley, 2011.
- [36] T. Cover and P. Hart, "Nearest neighbor pattern classification", *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, 1967.
- [37] L. Devroye, "On the Inequality of Cover and Hart in Nearest Neighbor Discrimination", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. -3, no. 1, pp. 75-78, 1981.

- [38] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press, 1995.
- [39] J. Gardner, "Detection of vapours and odours from a multisensor array using pattern recognition Part 1. Principal component and cluster analysis", *Sensors and Actuators B: Chemical*, vol. 4, no. 1-2, pp. 109-115, 1991.
- [40] A. Jeremic and A. Nehorai, "Landmine detection and localization using chemical sensor array processing", *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1295-1305, 2000.
- [41] T. Braun, F. Huber, M. Ghatkesar, N. Backmann, H. Lang, C. Gerber and M. Hegner, "Processing of kinetic microarray signals", *Sensors and Actuators B: Chemical*, vol. 128, no. 1, pp. 75-82, 2007.
- [42] S. Marco and A. Gutierrez-Galvez, "Signal and Data Processing for Machine Olfaction and Chemical Sensing: A Review", *IEEE Sensors Journal*, vol. 12, no. 11, pp. 3189-3214, 2012.
- [43] C. W. Therrien, *Decision, Estimation and Classification: An Introduction to Pattern Recognition and Related Topics*. New York: John Wiley & Sons, Inc., 1989.
- [44] A.V. Oppenheim and R.W. Schaffer, *Discrete-Time Signal Processing*. 3rd Ed., New Jersey: Pearson, 2010.
- [45] M. Wenzel, A. Mensah-Brown, F. Josse and E. Yaz, "Online Drift Compensation for Chemical Sensors Using Estimation Theory", *IEEE Sensors Journal*, vol. 11, no. 1, pp. 225-232, 2011.
- [46] K. Sothivelr, F. Bender, F. Josse, E. Yaz, A. Ricco and R. Mohler, "Online Chemical Sensor Signal Processing Using Estimation Theory: Quantification of Binary Mixtures of Organic Compounds in the Presence of Linear Baseline Drift and Outliers", *IEEE Sensors Journal*, vol. 16, no. 3, pp. 750-761, 2016.
- [47] F. Bender, F. Josse, A. J. Ricco, "Influence of Ambient Parameters on the Response of Polymer-Coated SH-Surface Acoustic Wave Sensors to Aromatic Analytes in Liquid-Phase Detection," In *Joint Conference of the IEEE IFCS and EFTF Proceedings*, pp 422-427, 2011.
- [48] US Environmental Protection Agency: *Underground Storage Tanks (USTs)* [Online]. Available: <https://www.epa.gov/ust>.
- [49] "What is BTEX and why is it important?", Aeroqual, 2017. [Online]. Available: <http://www.aeroqual.com/what-is-btex>. [Accessed: 22-Jun-2017].

- [50] F. Bender, R. Mohler, A. J. Ricco, F. Josse, "Quantification of Benzene in Groundwater Using SH-Surface Acoustic Wave Sensors," IMCS Proc., pp 473–476, 2012.
- [51] M. Mehlman, G. Henstreet, J. Thorpe and N. Weaver, Renal Effects of Petroleum Hydrocarbons. Princeton: Princeton Scientific Publ. Co., 1984.
- [52] US Environmental Protection Agency: National Primary Drinking Water Regulations, [Online]. Available: <https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations>.
- [53] D. R. Lide, Aqueous Solubility and Henry's Law Constants of Organic Compounds. In CRC Handbook of Chemistry and Physics, 82nd ed. Boca Raton, FL: CRC Press, pp 8-86–8-112, 2001–2002.
- [54] Y. Jones, Zhonghui Li, M. Johnson, F. Josse and J. Hossenlopp, "ATR-FTIR spectroscopic analysis of sorption of aqueous analytes into polymer coatings used with guided SH-SAW sensors", IEEE Sensors Journal, vol. 5, no. 6, pp. 1175-1184, 2005.
- [55] M. Karlowatz, M. Kraft and B. Mizaikoff, "Simultaneous Quantitative Determination of Benzene, Toluene, and Xylenes in Water Using Mid-Infrared Evanescent Field Spectroscopy", Analytical Chemistry, vol. 76, no. 9, pp. 2643-2648, 2004.
- [56] T. Schädle, B. Pejčić, M. Myers and B. Mizaikoff, "Fingerprinting Oils in Water via Their Dissolved VOC Pattern Using Mid-Infrared Sensors", Analytical Chemistry, vol. 86, no. 19, pp. 9512-9517, 2014.
- [57] K. Lima, I. Raimundo and M. Pimentel, "Improving the detection limits of near infrared spectroscopy in the determination of aromatic hydrocarbons in water employing a silicone sensing phase", Sensors and Actuators B: Chemical, vol. 125, no. 1, pp. 229-233, 2007.
- [58] B. Wittkamp and D. Tilotta, "Determination of BTEX Compounds in Water by Solid-Phase Microextraction and Raman Spectroscopy", Analytical Chemistry, vol. 67, no. 3, pp. 600-605, 1995.
- [59] M. Kamal and P. Klein, "Estimation of BTEX in groundwater by using gas chromatography–mass spectrometry", Saudi Journal of Biological Sciences, vol. 17, no. 3, pp. 205-208, 2010.
- [60] US Environmental Protection Agency: Semiannual Report of UST Performance Measures, End Of Fiscal Year 2016, [Online]. Available: <https://www.epa.gov/sites/production/files/2016-11/documents/ca-16-34.pdf>.

- [61] J. Cooper, B. Raguse, E. Chow, L. Hubble, K. Müller and L. Wieczorek, "Gold Nanoparticle Chemiresistor Sensor Array that Differentiates between Hydrocarbon Fuels Dissolved in Artificial Seawater", *Analytical Chemistry*, vol. 82, no. 9, pp. 3788-3795, 2010.
- [62] C. Ho and R. Hughes, "In-Situ Chemiresistor Sensor Package for Real-Time Detection of Volatile Organic Compounds in Soil and Groundwater", *Sensors*, vol. 2, no. 1, pp. 23-34, 2002.
- [63] L. Silva, A. Panteleitchouk, A. Freitas, T. Rocha-Santos and A. Duarte, "Microscale optical fibre sensor for BTEX monitoring in landfill leachate", *Analytical Methods*, vol. 1, no. 2, p. 100, 2009.
- [64] F. Bender, R. Mohler, A. Ricco and F. Josse, "Identification and Quantification of Aqueous Aromatic Hydrocarbons Using SH-Surface Acoustic Wave Sensors", *Analytical Chemistry*, vol. 86, no. 3, pp. 1794-1799, 2014.
- [65] F. Bender, F. Josse, R. E. Mohler, A. J. Ricco, "Design of SH-Surface Acoustic Wave Sensors for Detection of ppb Concentrations of BTEX in Water," Joint UFFC. EFTF and PFM Symposium, pp 628-631, 2013.
- [66] S. Kay, *Fundamentals of statistical signal processing*. Englewood Cliffs, N.J.: Prentice-Hall PTR, 1993.
- [67] R. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering*, vol. 82, no. 1, p. 35, 1960.
- [68] G. Welch and G. Bishop, "An Introduction to the Kalman Filter", Cs.unc.edu, 2006. [Online]. Available: http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf. [Accessed: 20-Nov- 2017].
- [69] D. Simon, *Optimal state estimation*. Hoboken (N.J.): Wiley-Interscience, 2006.
- [70] F. Lewis, L. Xie, D. Popa and F. Lewis, *Optimal and robust estimation*, 2nd ed. Boca Raton: CRC Press, 2008.
- [71] B. Anderson, *Optimal Filtering*. Dover Publications, 2012.
- [72] S. Haykin, *Adaptive filter theory*, 4th ed. Upper Saddle River, New Jersey: Prentice Hall, 2002.
- [73] R. Brown and P. Hwang, *Introduction to random signals and applied Kalman filtering*, 2nd ed. New York: John Wiley and Sons, Inc, 1992.

- [74] J. Candy, Bayesian signal processing. Hoboken, N.J.: Wiley, 2009.
- [75] E. Costa, J. do Val and M. Fragosa, "On a Detectability Concept of Discrete-Time Infinite Markov Jump Linear Systems", *Stochastic Analysis and Applications*, vol. 23, no. 1, pp. 1-14, 2005.
- [76] K. Reif, S. Gunther, E. Yaz and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter", *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 714-728, 1999.
- [77] T. Li, "On Exponentially Weighted Recursive Least Squares for Estimating Time-Varying Parameters and its Application to Computer Workload Forecasting", *Journal of Statistical Theory and Practice*, vol. 2, no. 3, pp. 339-354, 2008.
- [78] W. Brogan, *Modern control theory*. Englewood Cliffs: Prentice-Hall, 1991.
- [79] P. Tam and J. Moore, "Adaptive estimation using parallel processing techniques", *Computers & Electrical Engineering*, vol. 2, no. 2-3, pp. 203-214, 1975.
- [80] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations", *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439-448, 1972.
- [81] G. Terejanu, P. Singla, T. Singh and P. Scott, "Adaptive Gaussian Sum Filter for Nonlinear Bayesian Estimation", *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2151-2156, 2011.
- [82] Y. Jones, "Spectroscopic and Computational Investigation of Polymer Coatings and Analyte Systems for use with Guided Shear Horizontal Surface Acoustic Wave (SH-SAW) Sensors for Liquid Phase Detection", Ph.D Dissertation, Marquette University, Milwaukee, WI, 2005.
- [83] Z. Li, " Guided shear-horizontal surface acoustic wave (SH-SAW) chemical sensors for detection of organic contaminants in aqueous environments", Ph.D Dissertation, Marquette University, Milwaukee, WI, 2005.
- [84] C. Rogers, "Permeation of Gases and Vapours in Polymers", in *Polymer Permeability*, J. Comyn, Ed. New York: Elsevier, 1985, pp. 11-73.
- [85] G. Taylor, "Dispersion of Soluble Matter in Solvent Flowing Slowly through a Tube", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 219, no. 1137, pp. 186-203, 1953.

- [86] F. Bender, C. Kim, T. Mlsna and J. Vetelino, "Characterization of a WO₃ thin film chlorine sensor", *Sensors and Actuators B: Chemical*, vol. 77, no. 1-2, pp. 281-286, 2001.
- [87] K. Sothivelr, F. Bender, F. Josse, E. Yaz and A. Ricco, "Sensor-based estimation of BTEX concentrations in water samples using recursive least squares and Kalman filter techniques", in *IEEE Sensors*, Orlando, 2016.
- [88] B. Drafts, "Acoustic wave technology sensors", *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 4, pp. 795-802, 2001.
- [89] C. McMullan, H. Mehta, E. Gizeli, H. Mehta and C. Lowe, "Modelling of the mass sensitivity of the Love wave device in the presence of a viscous liquid", *Journal of Physics D: Applied Physics*, vol. 33, no. 23, pp. 3053-3059, 2000.
- [90] T. Zhou, "Theoretical Modeling of Acoustic Waves in Layered Structure Chemical Sensors and Biosensors," M.S. Thesis, Marquette University, Milwaukee, WI, 1992.
- [91] Z. Li, Y. Jones, J. Hossenlopp, R. Cernosek and F. Josse, "Analysis of Liquid-Phase Chemical Detection Using Guided Shear Horizontal-Surface Acoustic Wave Sensors", *Analytical Chemistry*, vol. 77, no. 14, pp. 4595-4603, 2005.
- [92] A. Mensah-Brown, D. Mlambo, F. Josse and S. Schneider, "Analysis of the Detection of Organophosphate Pesticides in Aqueous Solutions Using Hydrogen-Bond Acidic Coating on SH-SAW Devices", *IEEE Sensors Journal*, vol. 12, no. 5, pp. 893-903, 2012.
- [93] J. Grate and D. Nelson, "Sorptive polymeric materials and photopatterned films for gas phase chemical microsensors", *Proceedings of the IEEE*, vol. 91, no. 6, pp. 881-889, 2003.
- [94] R. McGill, M. Abraham and J. Grate, "Choosing Polymer-Coatings for Chemical Sensors", *Chemtech*, vol. 24, no. 9, pp. 27-37, 1994.
- [95] J. Grate, S. Kaganove, S. Patrash, R. Craig and M. Bliss, "Hybrid Organic/Inorganic Copolymers with Strongly Hydrogen-Bond Acidic Properties for Acoustic Wave and Optical Sensors", *Chemistry of Materials*, vol. 9, no. 5, pp. 1201-1207, 1997.
- [96] J. Mark, *Polymer data handbook*. Oxford: Oxford University Press, 2009.
- [97] P. Adhikari, L. Alderson, F. Bender, A. Ricco and F. Josse, "Investigation of Polymer-Plasticizer Blends as SH-SAW Sensor Coatings for Detection of

Benzene in Water with High Sensitivity and Long-Term Stability", *ACS Sensors*, vol. 2, no. 1, pp. 157-164, 2016.

- [98] "A Micro-GC Based Chemical Analysis System", Defiant Technologies, Inc., 2018. [Online]. Available: <http://www.defiant-tech.com/pdfs/Pittcon%202014%20A%20Micro-GC%20Based%20Chemical%20Analysis%20System.pdf>. [Accessed: 08- Jan-2018].
- [99] E. MacCash, *Surface chemistry*. Oxford [u.a.]: Oxford Univ. Press, 2007.
- [100] G. C. Frye, S. J. Martin and A. J. Ricco, "Monitoring Diffusion in Real Time in Thin Polymer Films Using SAW Devices", *Sensors and Materials*, vol. 1, no. 6, pp. 335-357, 1989.

APPENDIX A: ADDITIONAL RESULTS

A.1 Additional Results for Identification and Quantification of Multi-Analyte Mixtures Using Multi-Stage Exponentially Weighted RLSE

Table A. 1: Additional identification and quantification results obtained using multi-stage exponentially weighted RLSE and frequency transient data of SH-SAW sensors coated with either 0.6 μ m PECH or 0.8 μ m PIB, compared to analyte concentrations in the mixture measured using GC-PID. In the table ‘B’ denotes benzene, ‘T’ denotes toluene, ‘EX’ denotes ethylbenzene and xylenes, and ‘TMB’ denotes 1,2,4-trimethylbenzene.

Data	Actual Concentrations (in ppb)				Estimated Concentrations (in ppb)			
	B	T	EX	TMB	B	T	EX	TMB
1	980	340	0	0	944	334	0	0
2	0	0	1400	0	0	0	1440	0
3	1930	0	0	0	1945	0	0	0
4	1950	0	1670	0	2113	0	1636	0
5	1060	0	1410	0	878	0	1415	0
6	2260	740	0	0	1997	794	0	0
7	4200	0	0	0	4086	0	0	0
8	1990	0	0	0	1944	0	0	0
9	0	930	0	0	0	923	0	0
10	850	370	750	0	743	396	771	0
11	593	420	330	0	588	445	335	0
12	157	1014	350	0	222	967	353	0
13	0	0	0	1000	0	0	0	918
14	0	0	0	2000	0	0	0	1929
15	0	0	0	2000	0	0	0	2011
16	0	0	0	600	0	0	0	528

APPENDIX B: MATLAB CODES

The MATLAB codes used to implement the proposed sensor signal processing techniques in this dissertation are contained in this Appendix.

B.1 MATLAB Code for Detection and Quantification of Single Analytes Using Bank of Kalman Filters

```
%%
% Author: Karthick Sothivelr
%%
% Description: Code for Single Analyte Identification and Quantification
% Using Bank of Kalman Filters

%% Sensor Details
% SH-SAW Coated with 0.6 um PECH

%% Cleaning
clear all
close all
clc
tic;

%% Open and read the measurement file
FID = fopen('single_binary_test.ini','r'); % Type in the file name
data = textscan(FID,'%f %f');
fclose(FID);

%% USER INPUT
% Type in the 'End Point' of data
data_end = [65];

% Actual Concentration from FROG (GC-PID)
CA = 0; CB = 0.360; CC = 0; CD = 0; start=0+10; % 1st

%% Data Preprocessing (Selection of Analyte-In Point)
T = 12; % Sampling period T=12s
yy = data{2}((start+1:data_end(1)));
vk_raw=0:length(yy)-1;

% Plot of Baseline Corrected Data
figure,
h=plot(vk_raw*(T/60),yy,'*');
xlabel('Time (min)','FontSize',14); ylabel('Frequency Shift, \Deltaf (kHz)','FontSize',14)
set(h,'LineWidth',3)
```

```

grid on

%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=27; % Benzene
tauB=78; % Toluene
tauC=175; % Ethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; % Sorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; % Steady-State Sensitivity

%% Analyte Identification and Quantification Using Bank of Kalman Filters

% Generating Unknown Parameters Vector Theta
tau_rate = [T/tauA T/tauB T/tauC]; % T/tau

C_range = 0:0.01:5; % Concentration Range in ppm

alphaA = -C_range*ssA;
alphaB = -C_range*ssB;
alphaC = -C_range*ssC;

theta = zeros(2,20);
j=1;
for i=1:length(alphaA)
    theta(:,j) = [tau_rate(1);alphaA(i)];
    j=j+1;
end
for i=1:length(alphaB)
    theta(:,j) = [tau_rate(2);alphaB(i)];
    j=j+1;
end
for i=1:length(alphaC)
    theta(:,j) = [tau_rate(3);alphaC(i)];
    j=j+1;
end

ne = length(theta); % total number of filters
kmax = length(yy); y = yy;

%% Intialize the weights (probability) of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight (probability)
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat=zeros([ne,kmax]); inno=yhat;
x = zeros([1,kmax+1,ne]);

```

```

P=zeros(1,1,kmax+1,ne);
xest=zeros(1,kmax+1); Pest=zeros(1,1,kmax+1);

Pt = diag(100*ones(1,1));

for i=1:ne
    P(:, :, 1,i) = P(:, :, 1,i) + Pt;
end

%% Process and Measurement Noise
V = 0;
W = 5e-4;
Psum = zeros(1,1);

%% System Matrices G and F
G = 1; % Matrix G (1 by 1 Matrix)
F = 1;

%% Bank of Kalman Filters Scheme
for i=1:kmax
    for j=1:ne
        % System Matrices
        A = (1-theta(1,j));
        B = theta(1,j);
        C = theta(2,j);

        % Estimate of y: yhat
        h = C*x(:,i,j);
        yhat(j,i)=h;

        % Finding the innovation
        inno(j,i) = y(i)-h;

        % Estimator
        K = (A*P(:, :, i,j)*C')/(C*P(:, :, i,j)*C' + G*W*G'); % Kalman gain
        x(:,i+1,j) = A*x(:,i,j) + B + K*inno(j,i); % State Update
        P(:, :, i+1,j) = (A-K*C)*P(:, :, i,j)*(A-K*C)' + K*G*W*G'*K' + F*V*F'; % Error
        Covariance Update

        % Weight Update Equations
        S = C*P(:, :, i+1,j)*C' + W;

        % Weight before normalization
        cw(j,i) = ((abs(S))^-0.5)*(exp((( -0.5)*(inno(j,i))*(inno(j,i)))/(S))))*w(j,i);
    end
    % Normalized Weight

```

```

c = sum(cw(:,i));
w(:,i+1) = cw(:,i)./c;

% Estimate of state and error covariance
xx = x(:,i+1,:);
xx = permute(xx,[3 1 2]);
xest(:,i+1) = w(:,i+1)*xx(:,:,1);

for j=1:ne
    Psum = w(j,i+1)*(P(:,:,i+1,j)+(x(:,i+1,j)*(x(:,i+1,j)))) + Psum;
end
Pest(:,:,i+1) = Psum - xest(:,i+1)*xest(:,i+1)';
Psum = zeros(1,1);
end

%% Plot and Analysis

% Choosing the Maximum A posteriori Probability MAP
[Y, I]=max(w(:,kmax+1));

% Selecting the unknown parameters corresponding to the MAP
theta_hat = theta(:,I);

% Display Identification and Quantification Results in Command Window
tau_est = 12/theta_hat(1);
if tau_est==tauA
    Con_est = -theta_hat(2)/ssA;
    fprintf('\nIdentification Result: Analyte A \n')
    fprintf('Time Constant (s):')
    disp(tau_est)
end
if tau_est==tauB
    Con_est = -theta_hat(2)/ssB;
    fprintf('\nIdentification Result: Analyte B \n')
    fprintf('Time Constant (s):')
    disp(tau_est)
end
if tau_est==tauC
    Con_est = -theta_hat(2)/ssC;
    fprintf('\nIdentification Result: Analyte C \n')
    fprintf('Time Constant (s):')
    disp(tau_est)
end

fprintf('The estimated concentration (in ppb) is \n')
disp(Con_est)

```

```

% Plot of Results
vk=0:kmax-1;
y_est = (theta_hat(2))*(1-exp(-theta_hat(1).*vk)); % Estimated Output

yb = zeros(1,5);
y=[yb y']; y_est=[yb y_est];
vk=((0:length(y)-1)*(T/60)); vt=(0:0.1:length(y)-1)*(T/60);

% Plot of y and estimate of y
figure,
h=plot(vk, y, '*', vk, y_est, '-');
xlabel('Time (min)')
ylabel('Frequency Shift, \Delta f (kHz)')
legend(['Experimental Data (Time Constant: ' num2str(tauA) 's)'], ['Estimated Sensor
Response (Estimated Time Constant: ' num2str(tau_est) 's)'])
set(h,'LineWidth',3)
grid on

t1=0:kmax; t=0:kmax-1;
% A posteriori plot
figure,
plot(t1*(T/60),w(1:ne,:))
title('Plot of A Posteriori Probabilities')
ylabel('A Posteriori Probability')
xlabel('Time (min)')
grid on

% Innovation
figure,
h=plot(t, inno);
%set(h,'LineWidth',2)
title('Innovation Propagation Over Time (y-y_h_a_t)')
xlabel('Number of Iterations, k')
ylabel('y-y_h_a_t')
grid on

toc;

```

B.2 MATLAB Code for Detection and Quantification of Binary Mixtures Using Bank of Kalman Filters

```

%%
% Author: Karthick Sothivelr
%%
% Description: Code for Binary Mixtures Identification and Quantification
% Using Bank of Kalman Filters

%% Sensor Details
% SH-SAW Coated with 0.8 um PIB

%% Cleaning
clear all
close all
clc
tic;

%% Open and read the measurement file
FID = fopen('binary_data.ini','r'); % Type in the file name
data = textscan(FID,'%f %f');
fclose(FID);

%% USER INPUT
% Type in the 'End Point' of data
%data_end = [76 127];
data_end = [127];

% Actual Concentration from FROG (GC-PID)
%CA = 0.850; CB = 0; CC = 0.260; CD = 0; start=0; % 1st
CA = 1.170; CB = 0.470; CC = 0; CD = 0; start=76; % 2nd

%% Data Preprocessing (Selection of Analyte-In Point)
T = 12; % Sampling period T=12s

yy = data{2}((start+1:data_end(1)));

% Plot of Baseline Corrected Data
vk_raw=0:length(yy)-1;
figure,
h=plot(vk_raw*(T/60),yy,'*');
xlabel('Time (min)','FontSize',14); ylabel('Frequency Shift, \Deltaf (kHz)','FontSize',14)
set(h,'LineWidth',3)
grid on

%% Time Constant of Analytes (Coating: 0.8um PIB)

```

```

T = 12; % Sampling Period
tauA = 36;
tauB = 88;
tauC = 230;

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; % Absorption Rate
ssA=0.078; ssB=0.403; ssC=1.160; % Steady-State Sensitivity

%% Binary Identification and Quantification Using Bank of Kalman Filters

% Generating Unknown Parameters Vector Theta
tau_rate = zeros(2,3);

tau_rate(:,1) = [Sa;Sb]; tau_rate(:,2) = [Sa;Sc]; tau_rate(:,3) = [Sb;Sc];

C_rangeA = 0.01:0.01:1.50;
C_rangeB = 0.01:0.01:1.50;
C_rangeC = 0.01:0.01:1.50;

alphaA = -C_rangeA*ssA;
alphaB = -C_rangeB*ssB;
alphaC = -C_rangeC*ssC;

theta = zeros(4,2000);
j=1;
for i=1:length(alphaA)
    for k=1:length(alphaB)
        theta(:,j) = [tau_rate(1,1); tau_rate(2,1); alphaA(i); alphaB(k)];
        j=j+1;
    end
end
for i=1:length(alphaA)
    for k=1:length(alphaC)
        theta(:,j) = [tau_rate(1,2); tau_rate(2,2); alphaA(i); alphaC(k)];
        j=j+1;
    end
end
for i=1:length(alphaB)
    for k=1:length(alphaC)
        theta(:,j) = [tau_rate(1,3); tau_rate(2,3); alphaB(i); alphaC(k)];
        j=j+1;
    end
end

ne = length(theta); % total number of filters
kmax = length(yy); y = yy;

```

```

%% Initialize the weights (probability) of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight (probability)
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat=zeros([ne,kmax]); inno=yhat; %thetha_hat=zeros(2,kmax);
x = zeros([2,kmax+1,ne]);
P=zeros(2,2,kmax+1,ne);
xest=zeros(2,kmax+1); Pest=zeros(2,2,kmax+1);

Pt = diag(100*[1 1]);

for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = 0;
W = 1e-5;
Psum = zeros(2,2);

% System Matrices G and F
G = 1;
F = [1;1];

%% Bank of Kalman Filters Scheme
for i=1:kmax
    for j=1:ne
        % System Matrices
        A = [(1-thetha(1,j)) 0; 0 (1-thetha(2,j))];
        B = [thetha(1,j); thetha(2,j)];
        C = [thetha(3,j) thetha(4,j)];
        %h = C*x(:,i,j);
        h = C(1,1)*(1-exp(-B(1,1).*(i-1))) + C(1,2)*(1-exp(-B(2,1).*(i-1)));
        yhat(j,i)=h;

        % Finding the innovation
        inno(j,i) = y(i)-h;

        % Estimator
        K = (A*P(:,:,i,j)*C')/(C*P(:,:,i,j)*C' + G*W*G'); % Kalman Gain
        x(:,i+1,j) = A*x(:,i,j) + B + K*inno(j,i); % State Update
        P(:,:,i+1,j) = (A-K*C)*P(:,:,i,j)*(A-K*C)' + K*G*W*G'*K' + F*V*F'; % Error
        Covariance Update
    end
end

```

```

% Weight Update Equations
S = C*P(:,i+1,j)*C' + W;

% Weight before normalization
cw(j,i) = ((abs(S))^-0.5)*(exp((-0.5)*(inno(j,i))*(inno(j,i))/(S)))*w(j,i);
end
% Normalized Weights
c = sum(cw(:,i));
w(:,i+1) = cw(:,i)./c;

% Estimate of state and error covariance
xx = x(:,i+1,:);
xx = permute(xx,[3 1 2]);
xest(:,i+1) = w(:,i+1)'*xx(:,:,1);

for j=1:ne
    Psum = w(j,i+1)*(P(:,i+1,j)+(x(:,i+1,j)*(x(:,i+1,j)))') + Psum;
end

Pest(:,i+1) = Psum - xest(:,i+1)*xest(:,i+1)';
Psum = zeros(2,2);
end

%% Plot and Analysis

% Choosing the Maximum A posteriori Probability (MAP)
[Y, I]=max(w(:,kmax+1));

% Selecting the unknown parameters corresponding to the MAP
theta_hat = theta(:,I);

% Display Identification and Quantification Results in Command Window
tau_est1 = 12/theta_hat(1,1);
if tau_est1==tauA
    Con_est1 = -theta_hat(3,1)/ssA;
    fprintf('\nIdentification Result: Analyte A \n')
    fprintf('Time Constant (s):')
    disp(tau_est1)
end
if tau_est1==tauB
    Con_est1 = -theta_hat(3,1)/ssB;
    fprintf('\nIdentification Result: Analyte B \n')
    fprintf('Time Constant (s):')
    disp(tau_est1)
end
if tau_est1==tauC

```

```

Con_est1 = -thetha_hat(3,1)/ssC;
fprintf('\nIdentification Result: Analyte C \n')
fprintf('Time Constant (s):')
disp(tau_est1)
end

tau_est2 = 12/thetha_hat(2,1);
if tau_est2==tauA
    Con_est2 = -thetha_hat(4,1)/ssA;
    fprintf('\nIdentification Result: Analyte A \n')
    fprintf('Time Constant (s):')
    disp(tau_est2)
end
if tau_est2==tauB
    Con_est2 = -thetha_hat(4,1)/ssB;
    fprintf('\nIdentification Result: Analyte B \n')
    fprintf('Time Constant (s):')
    disp(tau_est2)
end
if tau_est2==tauC
    Con_est2 = -thetha_hat(4,1)/ssC;
    fprintf('\nIdentification Result: Analyte C \n')
    fprintf('Time Constant (s):')
    disp(tau_est2)
end

fprintf('\nThe estimated concentration of Analyte 1 (in ppb) is \n')
disp(Con_est1)
fprintf('\nThe estimated concentration of Analyte 2 (in ppb) is \n')
disp(Con_est2)

% Plot of Results
vk=0:kmax-1;
y_est = (thetha_hat(3,1))*(1-exp(-thetha_hat(1,1).*vk)) + (thetha_hat(4,1))*(1-exp(-
thetha_hat(2,1).*vk)); % Estimated Output

yb = zeros(1,5);
y=[yb y']; y_est=[yb y_est];
vk=((0:length(y)-1)*(T/60)); vt=(0:0.1:length(y)-1)*(T/60);

figure,
h=plot(vk, y, '* ', vk, y_est, '-');
xlabel('Time (min)')
ylabel('Frequency Shift, \Delta f (kHz)')

```

```

legend(['Experimental Data (Time Constants: [' num2str(tauA) 's] [' num2str(tauC) 's]'],
['Estimated Sensor Response (Estimated Time Constants: [' num2str(tau_est1) 's] ['
num2str(tau_est2) 's]')]
set(h,'LineWidth',3)
grid on

toc;

```

B.3 MATLAB Code for Identification and Quantification of Multi-Analyte Mixtures Using Multi-Stage Exponentially Weighted RLSE

```

%%
% Author: Karthick Sothivelr
%%
% Description: Code for Identification and Quantification of Multi-Analyte
% Mixtures Using Multi-Stage Exponentially Weighted Recursive Least Squares
% Estimation (EW-RLSE)

%% Sensor Details
% SH-SAW Coated with 0.8 um PIB

%% Cleaning
clear all
close all
clc
tic;

%% Open and read the measurement file
FID = fopen('pib_ternary_data.ini','r'); % Type in the file name
data = textscan(FID,'%f %f');
fclose(FID);

%% USER INPUT
% Type in the 'End Point' of data
data_end = [106];

% Actual Concentration from FROG (GC-PID)
CA = 0.593; CB = 0.420; CC = 0.330; CD = 0; start=0+8; % 1st

%% Data Preprocessing (Selection of Analyte-In Point)
T = 12; % Sampling period T=12s

yy = data{2}((start+1:data_end(1)));

```

```

% Plot of Baseline Corrected Data
vk_raw=0:length(yy)-1;
figure,
h=plot(vk_raw*(T/60),yy,'*');
xlabel('Time (min)','FontSize',14); ylabel('Frequency Shift, \Deltaf (kHz)','FontSize',14)
set(h,'LineWidth',3)
grid on

%% Time Constant of Analytes (Coating: 0.8um PIB)
tauA=36; % Benzene
tauB=88; % Toluene
tauC=230; % Ethylbenzene
tauD=610; % Trimethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; Sd=T/tauD; % Absorption Rate
ssA=0.078; ssB=0.403; ssC=1.160; ssD=3.640; % Steady-State Sensitivity

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=yy(1:end); yb = zeros(1,5); b=y; %y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(1,4);

Q=10000;
lambdha=0.99; % Weighting factor
W = 2.8; % Measurement Noise Covariance

A = 1;
B = 1;
C = 1;
D = 1;

for j=1:5
    P=Q*eye(4); % Error covariance

    theta=zeros(4,kmax+1); % Initialize the unknown parameters
    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0;

    % Call the function EW_RLSE
    [theta,innovation] =
    EW_RLSE(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd,A,B,C,D,kmax);

    a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
    a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);

```

```

a(1,:) = [a1 a2 a3 a4];

% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);

% Estimated Concentration Compared with Detection Limit
if a1>=0 || Con_A<0.100
    A=0;
end
if a2>=0 || Con_B<0.050
    B=0;
end
if a3>=0 || Con_C<0.030
    C=0;
end
if a4>=0 || Con_D<0.015
    D=0;
end
end

%% Analysis
disp([a1 a2 a3 a4])

fprintf('The estimated concentration of Analyte A (in ppm) is \n')
disp(Con_A)

fprintf('The estimated concentration of Analyte B (in ppm) is \n')
disp(Con_B)

fprintf('The estimated concentration of Analyte C (in ppm) is \n')
disp(Con_C)

fprintf('The estimated concentration of Analyte D (in ppm) is \n')
disp(Con_D)

Per_CA = ((Con_A-CA)/CA)*100;
Per_CB = ((Con_B-CB)/CB)*100;
Per_CC = ((Con_C-CC)/CC)*100;
Per_CD = ((Con_D-CD)/CD)*100;

fprintf('\nPercentage Error of Concentration\n')
disp (Per_CA)
disp (Per_CB)

```

```

disp (Per_CC)
disp (Per_CD)

%% Plot

vk=0:kmax; vvk=0:kmax-1;
vt=(0:0.1:kmax-1)';
% Estimated Frequency Shift
vest = a1*(1-exp(-Sa.*vt)) + a2*(1-exp(-Sb.*vt)) + a3*(1-exp(-Sc.*vt)) + a4*(1-exp(-
Sd.*vt));

% Converting time step number to minutes and add baselines
y = [yb y']; %yhat = [yb';yhat'];
vest=[((0:0.1:4.9)*0);vest];
vk=((0:length(y)-1)*(T/60)); vt=(0:0.1:length(y)-1)*(T/60);

figure,
% Plot of Frequency Shift vs Time step
h=plot(vk,y, '*b', vt, vest, '-r');
%title ('Frequency Shift vs Time','FontSize',24)
%LineWidth = [3];
xlabel('Time (min)','FontSize',14); ylabel('Frequency Shift, \Deltaf (kHz)','FontSize',14)
h_legend = legend('Experimental Data (B:590ppb T:420ppb EX:330ppb)', 'Estimated
Sensor Response (B:588ppb T:445ppb EX:335ppb)');
set(h_legend,'FontSize',15);
set(h,'LineWidth',3)
grid on

%% Time Evolution of Estimated Concentrations
vk=0:kmax;

CA_Actual = ones(1,length(vk))*CA*1000;
CB_Actual = ones(1,length(vk))*CB*1000;
CC_Actual = ones(1,length(vk))*CC*1000;
CD_Actual = ones(1,length(vk))*CD*1000;

CA_1 = ones(1,length(vk))*CA*1000*0.75;
CB_1 = ones(1,length(vk))*CB*1000*0.75;
CC_1 = ones(1,length(vk))*CC*1000*0.75;
CD_1 = ones(1,length(vk))*CD*1000*0.75;

CA_2 = ones(1,length(vk))*CA*1000*1.25;
CB_2 = ones(1,length(vk))*CB*1000*1.25;
CC_2 = ones(1,length(vk))*CC*1000*1.25;
CD_2 = ones(1,length(vk))*CD*1000*1.25;

```

```
a1_time = 1000*(abs(theta(1,:))/ssA); a2_time = 1000*(abs(theta(2,:))/ssB);
a3_time = 1000*(abs(theta(3,:))/ssC); a4_time = 1000*(abs(theta(4,:))/ssD);
```

```
figure,
h=plot(vk*(T/60),CA_Actual,'-b',vk*(T/60),a1_time,'-r',vk*(T/60),CA_1,'-
k',vk*(T/60),CA_2,'-k');
xlabel('Time (min)','FontSize',14); ylabel('Concentration (ppb)','FontSize',14)
h_legend = legend('Actual Concentration', 'Estimated Concentration','25% Difference
Line');
set(h_legend,'FontSize',15);
set(h,'LineWidth',3)
```

```
figure,
h=plot(vk*(T/60),CB_Actual,'-b',vk*(T/60),a2_time,'-r',vk*(T/60),CB_1,'-
k',vk*(T/60),CB_2,'-k');
xlabel('Time (min)','FontSize',14); ylabel('Concentration (ppb)','FontSize',14)
h_legend = legend('Actual Concentration', 'Estimated Concentration','25% Difference
Line');
set(h_legend,'FontSize',15);
set(h,'LineWidth',3)
```

```
figure,
h=plot(vk*(T/60),CC_Actual,'-b',vk*(T/60),a3_time,'-r',vk*(T/60),CC_1,'-
k',vk*(T/60),CC_2,'-k');
xlabel('Time (min)','FontSize',14); ylabel('Concentration (ppb)','FontSize',14)
h_legend = legend('Actual Concentration', 'Estimated Concentration','25% Difference
Line');
set(h_legend,'FontSize',15);
set(h,'LineWidth',3)
```

```
figure,
h=plot(vk*(T/60),CD_Actual,'-b',vk*(T/60),a4_time,'-r',vk*(T/60),CD_1,'-
k',vk*(T/60),CD_2,'-k');
xlabel('Time (min)','FontSize',14); ylabel('Concentration (ppb)','FontSize',14)
h_legend = legend('Actual Concentration', 'Estimated Concentration','25% Difference
Line');
set(h_legend,'FontSize',15);
set(h,'LineWidth',3)
```

```
toc;
```

```
function [theta,innovation] =
EW_RLSE(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd,A,B,C,D,kmax)
% Exponentially Weighted RLSE Scheme
for i=1:kmax
H=[mA mB mC mD]';
innovation=y(i)-theta(:,i)*H;
```

```

%EW-RLSE identifier
K = (P*H)/(H'*P*H+W*lambda^2);
theta(:,i+1) = theta(:,i) + K*[innovation];
P = (1/(lambda^2))*(P - P*(H*H')*P/(H'*P*H+(W*lambda^2)));

mA = ((1-Sa)*mA + Sa)*A;
mB = ((1-Sb)*mB + Sb)*B;
mC = ((1-Sc)*mC + Sc)*C;
mD = ((1-Sd)*mD + Sd)*D;
end

```

B.4 MATLAB Code for Detection and Quantification of Multi-Analyte in the Presence of Interferents

B.4.1 Four-Analyte Model

a) Sorption Data

```

%%
% Author: Karthick Sothivelr
%%
% Description: Code for Detection and Quantification of Target Analytes in
% the Presence of Interferents using EW-RLSE and Bank of Kalman Filters
% 4-Analyte Model (Sorption)
%% Sensor Details
% SH-SAW Coated with 0.6 um PECH

%% Cleaning
clear all
close all
clc
tic;

%% Open and read the measurement file
FID = fopen('pech_sorp_data.ini','r'); % Type in the file name
data = textscan(FID,'%f %f');
fclose(FID);

%% USER INPUT
% Type in the 'End Point' of data
data_end = [105];

% Actual Concentration from FROG (GC-PID)

```

```

CA = 0.325; CB = 0.465; CC = 0.540; start=3; q=10000;

%% Data Preprocessing (Selection of Analyte-In Point)
nn = 3; % Number of points to check to determine the best curve fit
T = 12; % Sampling period T=12s

tau = zeros(1,length(data_end));
Fe = zeros(1,length(data_end));
Eps = zeros(1,nn);

for i=1:length(data_end)
    % Section to find the best curve fit
    for j=1:nn
        if i==1
            y = data{2}((1+(j-1)):data_end(i));
        end
        if i>1
            y = data{2}((data_end(i-1)+1+(j-1)):data_end(i));
        end
        kmax = length(y);
        try
            [h_fit, G] = fit((0:kmax-1)',y,'fe*(1-exp(-S*x))','StartPoint',[1 1]);
        catch err
            j = j-1;
            break
        end
        Eps(j) = G.sse; % Sum of error squared
    end

    % Section to generate the best curve
    [Y,I] = min(Eps);
    if i==1
        y = data{2}((1+(I-1)):data_end(i));
    end
    if i>1
        y = data{2}((data_end(i-1)+1+(I-1)):data_end(i));
    end
    kmax = length(y);
    vt=(0:0.1:kmax-1)'; vk=(0:kmax-1)';
end

% Plot of Baseline Corrected Data
figure,
plot(y(start:end),'*')

%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5; % Benzene

```

```

tauB=77.6; % Toluene
tauC=175; % Ethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=2; ssE=6;% Steady-State Sensitivity

tauD = [700]; % Trimethylbenzene
Sd = T./tauD;

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=y(start:end); yb = zeros(1,5); b=y; % y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Sd),4); eps = zeros(length(Sd),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 0.1; % Measurement Noise Covariance

for j=1:length(Sd)

    P=Q*eye(4); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0;

    theta=zeros(4,kmax+1); % Initialize the unknown parameters

    % Call the function EW_RLSE_4A
    [theta,innovation] =
EW_RLSE_4A(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd(j),kmax);

    %% Analysis
    a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
    a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);
    a(j,:) = [a1 a2 a3 a4];

    vvk=0:kmax-1;
    vt=(0:0.1:kmax)';

    % Estimated Frequency Shift
    verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
exp(-Sd(j).*vvk)); verr=verr';

    ytest = y(1:length(vvk));
    eps(j,1) = immse(ytest,verr);

```

```

    if a1>0 || a2>0 || a3>0 || a4>0
        eps(j,1)=100;
    end

end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);

clearvars Sd
Sd = T/tauD(Ind);

%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);

%% Bank of Kalman Filters Section

%% Set the concentration ranges for each analyte
% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;

% Generate the unknown parameters vector
% Frequency shift range
alpha1=-B*ssA;
alpha2=-Toluene*ssB;
alpha3=-EX*ssC;
alpha4=-TMB*ssD;

ne = length(B)*length(Toluene)*length(EX)*length(TMB); % Number of filters
C = zeros(ne,4);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l)];
            end
        end
    end
end

```

```

        ind = ind+1;
    end
end
end
end

%% Initialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([4,kmax+1,ne]);
P = zeros(4,4,kmax+1,ne);

Pt = diag([1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;

%% Bank of Kalman Filters Scheme
for i=1:kmax
    % Setting the first measurement to zero
    if i==1
        y(i) = y(i)*0;
    end

    for j=1:ne
        [x(:,i+1,j),P(:,:,i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter4A(x(:,i,j),P(:,:,i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,i);
    end
    % Normalized Weights
    c = sum(cw(:,i));
    w(:,i+1) = cw(:,i)./c;
end

%% Analysis

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);

```

```

Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);

```

```

% Estimated Concentration

```

```

Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);

```

```

Con = [Con_A; Con_B; Con_C];

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

%% Section for Lower Limit of Time Constants

```

```

%% Time Constant of Analytes (Coating: 0.6um PECH)

```

```

tauA=26.5-8; % Benzene
tauB=77.6-3; % Toluene
tauC=175-13; % Ethylbenzene

```

```

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; % Absorption Rate

```

```

ssA=0.109; ssB=0.435; ssC=1.450; ssD=2; ssE=6;% Steady-State Sensitivity

```

```

tauD = [700]; % Trimethylbenzene

```

```

Sd = T./tauD;

```

```

%% Exponentially Weighted Recursive Least Square Estimation Section

```

```

%% Select Data to be analyzed

```

```

y=y(1:end); yb = zeros(1,5); b=y; %y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Sd),4); eps = zeros(length(Sd),1);
Q=q;

```

```

lambdha=0.99; % Weighting factor

```

```

W = 0.1; % Measurement Noise Covariance

```

```

for j=1:length(Sd)

```

```

    P=Q*eye(4); % Error covariance

```

```

    % Initial normalized concentration

```

```

    mA=0; mB=0; mC=0; mD=0;

```

```

    theta=zeros(4,kmax+1); % Initialize the unknown parameters

```

```

[theta,innovation] =
EW_RLSE_4A(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd(j),kmax);

%% Analysis
a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);

a(j,:) = [a1 a2 a3 a4];

vvk=0:kmax-1;
vt=(0:0.1:kmax)';

% Estimated Frequency Shift
verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
exp(-Sd(j).*vvk)); verr=verr';

ytest = y(1:length(vvk));
eps(j,1) = immse(ytest,verr);

if a1>0 || a2>0 || a3>0 || a4>0
    eps(j,1)=100;
end

end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);

clearvars Sd
Sd = T/tauD(Ind);

%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);

%% Bank of Kalman Filters Section
%% Set the concentration ranges for each analyte
% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;

```

```

% Generate the unknown parameters vector
% Frequency shift range
alpha1=-B*ssA;
alpha2=-Toluene*ssB;
alpha3=-EX*ssC;
alpha4=-TMB*ssD;

ne = length(B)*length(Toluene)*length(EX)*length(TMB); % Number of filters
C = zeros(ne,4);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l)];
                ind = ind+1;
            end
        end
    end
end

%% Initialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([4,kmax+1,ne]);
P = zeros(4,4,kmax+1,ne);
xest = zeros(4,kmax+1); Pest = zeros(4,4,kmax+1);

Pt = diag([1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;

%% Bank of Kalman Filters Scheme
for i=1:kmax
    % Setting the first measurement to zero
    if i==1

```

```

    y(i) = y(i)*0;
end

for j=1:ne
    [x(:,i+1,j),P(:, :,i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter4A(x(:,i,j),P(:, :,i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,i);
end
% Normalized Weights
c = sum(cw(:,i));
w(:,i+1) = cw(:,i)./c;
end

%% Analysis

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);

% Estimated Concentration
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);

Con = [Con_A; Con_B; Con_C]+Con;

%% Section for Upper Limit of Time Constants
%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5+8; % Benzene
tauB=77.6+3; % Toluene
tauC=175+13; % Ethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=2; ssE=6;% Steady-State Sensitivity

tauD = [700]; % Trimethylbenzene
Sd = T./tauD;

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=y(1:end); yb = zeros(1,5); b=y; % y(1)=0;
kmax = length(y); % Length of the measurement data points

```

```

a = zeros(length(Sd),4); eps = zeros(length(Sd),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 0.1; % Measurement Noise Covariance

for j=1:length(Sd)

    P=Q*eye(4); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0;

    theta=zeros(4,kmax+1); % Initialize the unknown parameters

    [theta,innovation] =
    EW_RLSE_4A(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd(j),kmax);

    %% Analysis
    a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
    a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);

    a(j,:) = [a1 a2 a3 a4];

    vvk=0:kmax-1;
    vt=(0:0.1:kmax)';

    % Estimated Frequency Shift
    verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
    exp(-Sd(j).*vvk)); verr=verr';

    ytest = y(1:length(vvk));
    eps(j,1) = immse(ytest,verr);

    if a1>0 || a2>0 || a3>0 || a4>0
        eps(j,1)=100;
    end

end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);

clearvars Sd

```

```

Sd = T/tauD(Ind);

%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);

%% Bank of Kalman Filters Section
%% Set the concentration ranges for each analyte

% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;

% Generate the unknown parameters vector
% Frequency shift range
alpha1=-B*ssA;
alpha2=-Toluene*ssB;
alpha3=-EX*ssC;
alpha4=-TMB*ssD;

ne = length(B)*length(Toluene)*length(EX)*length(TMB); % Number of filters
C = zeros(ne,4);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l)];
                ind = ind+1;
            end
        end
    end
end

%% Intialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]); % weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([4,kmax+1,ne]);

```

```

P = zeros(4,4,kmax+1,ne);
xest = zeros(4,kmax+1); Pest = zeros(4,4,kmax+1);

Pt = diag([1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;
Psum = zeros(4,4);

%% Bank of Kalman Filters Scheme
for i=1:kmax
    % Setting the first measurement to zero
    if i==1
        y(i) = y(i)*0;
    end

    for j=1:ne
        [x(:,i+1,j),P(:, :, i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter4A(x(:,i,j),P(:, :, i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,i);
    end
    % Normalized Weights
    c = sum(cw(:,i));
    w(:,i+1) = cw(:,i)./c;
end

%% Plot and Analysis

vk=0:kmax; vvk=0:kmax-1;
vt=(0:0.1:kmax-1)';

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);

% Estimated Concentration
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);

```

```

Con_D = abs(Ed)/(ssD);

Con = [Con_A; Con_B; Con_C]+Con;

% Estimated Frequency Shift
vest = Ea*(1-exp(-Sa.*vt)) + Eb*(1-exp(-Sb.*vt)) + Ec*(1-exp(-Sc.*vt)) + Ed*(1-exp(-
Sd.*vt));

% Converting time step number to minutes and add baselines
yp = [yb y']; % yhat = [yb';yhat'];
vest=[((0:0.1:4.9)*0);vest];
vk=((0:length(yp)-1))*(T/60); vt=(0:0.1:length(yp)-1)*(T/60);

% Plot of Frequency Shift vs Time
figure,
h=plot(vk,yp, '*b', vt, vest, '--r');
xlabel('Time (min)',FontSize,14); ylabel('Frequency Shift, \Deltaf (kHz)',FontSize,14)
h_legend = legend('Experimental data', 'Estimated Sensor Response');
set(h_legend,FontSize,15);
set(h,'LineWidth',3)
grid on

%%
gg = inno';
eps = mean(abs(gg));
[X, Ix]=min(eps);

% Average Concentrations
Con = Con/3;

% Display Results
fprintf('\n~~~~~\n')
fprintf('\n Average Results \n')

fprintf('The estimated concentration of Analyte A (in ppm) is \n')
disp(Con(1))

fprintf('The estimated concentration of Analyte B (in ppm) is \n')
disp(Con(2))

fprintf('The estimated concentration of Analyte C (in ppm) is \n')
disp(Con(3))

Per_CA = ((Con(1)-CA)/CA)*100;
Per_CB = ((Con(2)-CB)/CB)*100;
Per_CC = ((Con(3)-CC)/CC)*100;

```

```

fprintf('\nPercentage Error of Concentration\n')
disp (Per_CA)
disp (Per_CB)
disp (Per_CC)

toc;

function [theta,innovation] =
EW_RLSE_4A(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd,kmax)

for i=1:kmax
    H=[mA mB mC mD]';
    innovation=y(i)-theta(:,i)*H;

    %EW-RLSE identifier
    K = (P*H)/(H'*P*H+W*lambdha^2);
    theta(:,i+1) = theta(:,i) + K*[innovation];
    P = (1/(lambdha^2))*(P - P*(H'*H)*P/(H'*P*H+(W*lambdha^2)));

    mA = ((1-Sa)*mA + Sa);
    mB = ((1-Sb)*mB + Sb);
    mC = ((1-Sc)*mC + Sc);
    mD = ((1-Sd)*mD + Sd);
end

function [x,P,h,cw,inno] = kalman_filter4A(x,P,y meas,C,W,V,w,Sa,Sb,Sc,Sd,i)
% KF Kalman Filter for linear dynamic systems
% Returns state estimate x, Error Covariance P, yhat(or h)
% Inputs:  x: "a priori" state estimate
%          P: "a priori" estimated state covariance
%          y meas: current measurement
%          V: process noise covariance
%          W: measurement noise covariance
% Output:  x: "a posteriori" state estimate
%          P: "a posteriori" state covariance

G = [1]; % Matrix G (1 by 1 Matrix)
F = eye(4);
U = 1; % Step Input
A = diag([1-Sa, 1-Sb, 1-Sc, 1-Sd]);
B = [Sa; Sb; Sc; Sd];

% Finding the innovation
h = C(1,1)*(1-exp(-Sa.*(i-1))) + C(1,2)*(1-exp(-Sb.*(i-1))) + C(1,3)*(1-exp(-Sc.*(i-1)))
+ C(1,4)*(1-exp(-Sd.*(i-1)));

```

```

inno = ymeas-h;

% Estimator
K = (A*P*C')/(C*P*C' + G*W*G');
x = A*x + B*U + K*inno;
P = (A-K*C)*P*(A-K*C)' + K*G*W*G'*K' + F*V*F';

% Weight Update Equations
S = C*P*C' + W;

% Weight before normalization
cw = ((abs(S))(-0.5))*(exp((-0.5)*(inno)*(inno))/(S))*w;

```

b) Desorption Data

```

%%
% Author: Karthick Sothivelr
%%
% Description: Code for Detection and Quantification of Target Analytes in
% the Presence of Interferents using EW-RLSE and Bank of Kalman Filters
% 4-Analyte Model (Desorption)
%% Sensor Details
% SH-SAW Coated with 0.6 um PECH

%% Cleaning
clear all
close all
clc
tic;

%% Open and read the measurement file
FID = fopen('pech_desorp_data.ini','r'); % Type in the file name
data = textscan(FID,'%f %f');
fclose(FID);

%% USER INPUT
% Type in the 'End Point' of data
data_end = [101];

% Actual Concentration from FROG (GC-PID)
CA = 0.325; CB = 0.565; CC = 0.540; start=0+1; q=10000;

%% Data Preprocessing (Selection of Analyte-In Point)
T = 12; % Sampling period T=12s

```

```

y0 = data{2}((start+1:data_end(1))); kmax = length(y0); y0 = y0-y0(1);
y1 = data{2}((start+2:data_end(1))); kmax1 = length(y1); y1 = y1-y1(1);
y2 = data{2}((start+3:data_end(1))); kmax2 = length(y2); y2 = y2-y2(1);
y3 = data{2}((start+4:data_end(1))); kmax3 = length(y3); y3 = y3-y3(1);
y4 = data{2}((start+5:data_end(1))); kmax4 = length(y4); y4 = y4-y4(1);
y5 = data{2}((start+6:data_end(1))); kmax5 = length(y5); y5 = y5-y5(1);
y6 = data{2}((start+7:data_end(1))); kmax6 = length(y6); y6 = y6-y6(1);
y7 = data{2}((start+8:data_end(1))); kmax7 = length(y7); y7 = y7-y7(1);

y=y0;

% Plot of Baseline Corrected Data
vk=0:length(y)-1;
figure,
h=plot(vk*(T/60),y,'*');
xlabel('Time (min)', 'FontSize',14); ylabel('Frequency Shift, \Deltaf (kHz)', 'FontSize',14)
set(h, 'LineWidth',3)
grid on

%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5; % Benzene
tauB=77.6; % Toluene
tauC=175; % Ethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=2; ssE=6; % Steady-State Sensitivity

tauD = [700]; % Trimethylbenzene
Sd = T./tauD;

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=y(1:end); yb = zeros(1,5); b=y; %y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Sd),4); eps = zeros(length(Sd),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 1; % Measurement Noise Covariance

for j=1:length(Sd)

    P=Q*eye(4); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0;

```

```

theta=zeros(4,kmax+1); % Initialize the unknown parameters

[theta,innovation] =
EW_RLSE_4A(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd(j),kmax);

%% Analysis
a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);

a(j,:) = [a1 a2 a3 a4];

vvk=0:kmax-1;
vt=(0:0.1:kmax)';

% Estimated Frequency Shift
verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
exp(-Sd(j).*vvk)); verr=verr';

ytest = y(1:length(vvk));
eps(j,1) = immse(ytest,verr);

if a1<0 || a2<0 || a3<0 || a4<0
    eps(j,1)=100;
end
end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);

clearvars Sd
Sd = T/tauD(Ind);

%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);

%% Bank of Kalman Filter Section
%% Set the concentration ranges for each analyte
% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;

```

```

EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;

% Generate the unknown parameters vector
% Frequency shift range
alpha1=B*ssA;
alpha2=Toluene*ssB;
alpha3=EX*ssC;
alpha4=TMB*ssD;

ne = length(B)*length(Toluene)*length(EX)*length(TMB); % Number of filters
C = zeros(ne,4);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l)];
                ind = ind+1;
            end
        end
    end
end

%% Intialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([4,kmax+1,ne]);
P = zeros(4,4,kmax+1,ne);
xest = zeros(4,kmax+1); Pest = zeros(4,4,kmax+1);

Pt = diag([1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;

%% Bank of Kalman Filters Scheme
for i=1:kmax

```

```

% Setting the first measurement to zero
if i==1
    y(i) = y(i)*0;
end

for j=1:ne
    [x(:,i+1,j),P(:,:,i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter4A(x(:,i,j),P(:,:,i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,i);
end
% Normalized Weights
c = sum(cw(:,i));
w(:,i+1) = cw(:,i)./c;
end

%% Analysis

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);

% Estimated Concentration
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);

Con = [Con_A; Con_B; Con_C];

%% Section for Lower Limit of Time Constants
%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5-8; % Benzene
tauB=77.6-3; % Toluene
tauC=175-13; % Ethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=2; ssE=6;% Steady-State Sensitivity

tauD = [700]; % Trimethylbenzene
Sd = T./tauD;

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed

```

```

y=y(1:end); yb = zeros(1,5); b=y; % y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Sd),4); eps = zeros(length(Sd),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 1; % Measurement Noise Covariance

for j=1:length(Sd)

    P=Q*eye(4); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0;

    theta=zeros(4,kmax+1); % Initialize the unknown parameters

    [theta,innovation] =
    EW_RLSE_4A(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd(j),kmax);

    %% Analysis
    a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
    a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);

    a(j,:) = [a1 a2 a3 a4];

    vvk=0:kmax-1;
    vt=(0:0.1:kmax)';

    % Estimated Frequency Shift
    verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
    exp(-Sd(j).*vvk)); verr=verr';

    ytest = y(1:length(vvk));
    eps(j,1) = immse(ytest,verr);

    if a1<0 || a2<0 || a3<0 || a4<0
        eps(j,1)=100;
    end

end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);

```

```

clearvars Sd
Sd = T/tauD(Ind);

%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);

%% Bank of Kalman Filters Section
%% Set the concentration ranges for each analyte

% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;

% Frequency shift range
alpha1=B*ssA;
alpha2=Toluene*ssB;
alpha3=EX*ssC;
alpha4=TMB*ssD;

ne = length(B)*length(Toluene)*length(EX)*length(TMB); % Number of filters
C = zeros(ne,4);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l)];
                ind = ind+1;
            end
        end
    end
end

%% Intialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]); % weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;

```

```

x = zeros([4,kmax+1,ne]);
P = zeros(4,4,kmax+1,ne);
xest = zeros(4,kmax+1); Pest = zeros(4,4,kmax+1);

Pt = diag([1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;

%% Bank of Kalman Filters Scheme
for i=1:kmax
    % Setting the first measurement to zero
    if i==1
        y(i) = y(i)*0;
    end

    for j=1:ne
        [x(:,i+1,j),P(:, :, i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter4A(x(:,i,j),P(:, :, i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,i);
    end
    % Normalized Weights
    c = sum(cw(:,i));
    w(:,i+1) = cw(:,i)./c;
end

%% Analysis

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);

% Estimated Concentration
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);

Con = [Con_A; Con_B; Con_C]+Con;

```

```

%% Section for Upper Limit of Time Constants
%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5+8; % Benzene
tauB=77.6+3; % Toluene
tauC=175+13; % Ethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=2; ssE=6;% Steady-State Sensitivity

tauD = [700]; % Trimethylbenzene
Sd = T./tauD;

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=y(1:end); yb = zeros(1,5); b=y; %y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Sd),4); eps = zeros(length(Sd),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 1; % Measurement Noise Covariance

for j=1:length(Sd)

    P=Q*eye(4); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0;

    theta=zeros(4,kmax+1); % Initialize the unknown parameters

    [theta,innovation] =
    EW_RLSE_4A(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd(j),kmax);

    %% Analysis
    a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
    a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);

    a(j,:) = [a1 a2 a3 a4];

    vvk=0:kmax-1;
    vt=(0:0.1:kmax)';

    % Estimated Frequency Shift

```

```

    verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
    exp(-Sd(j).*vvk)); verr=verr';

```

```

    ytest = y(1:length(vvk));
    eps(j,1) = immse(ytest,verr);

```

```

    if a1<0 || a2<0 || a3<0 || a4<0
        eps(j,1)=100;
    end

```

```

end

```

```

[ms_error, Ind] = min(eps);

```

```

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);

```

```

clearvars Sd
Sd = T/tauD(Ind);

```

```

%% Estimated Concentration

```

```

Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);

```

```

%% Bank of Kalman Filters Section

```

```

%% Set the concentration ranges for each analyte

```

```

% Concentration Range in ppm

```

```

B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;

```

```

% Frequency shift range

```

```

alpha1=B*ssA;
alpha2=Toluene*ssB;
alpha3=EX*ssC;
alpha4=TMB*ssD;

```

```

ne = length(B)*length(Toluene)*length(EX)*length(TMB); % Number of filters
C = zeros(ne,4);

```

```

ind=1;
for i=1:length(B)

```

```

for j=1:length(Toluene)
  for k=1:length(EX)
    for l=1:length(TMB)
      C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l)];
      ind = ind+1;
    end
  end
end
end

%% Initialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([4,kmax+1,ne]);
P = zeros(4,4,kmax+1,ne);
xest = zeros(4,kmax+1); Pest = zeros(4,4,kmax+1);

Pt = diag([1 1 1 1]);
for i=1:ne
  P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;
Psum = zeros(4,4);

%% Bank of Kalman Filters Scheme
for i=1:kmax
  % Setting the first measurement to zero
  if i==1
    y(i) = y(i)*0;
  end

  for j=1:ne
    [x(:,i+1,j),P(:,:,i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
    kalman_filter4A(x(:,i,j),P(:,:,i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,i);
  end
  % Normalized Weights
  c = sum(cw(:,i));
  w(:,i+1) = cw(:,i)./c;
end

```

%% Plot and Analysis

```
vk=0:kmax; vvk=0:kmax-1;
vt=(0:0.1:kmax-1)';

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');
```

```
Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);
```

% Estimated Concentration

```
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);
```

```
Con = [Con_A; Con_B; Con_C]+Con;
```

% Estimated Frequency Shift

```
vest = Ea*(1-exp(-Sa.*vt)) + Eb*(1-exp(-Sb.*vt)) + Ec*(1-exp(-Sc.*vt)) + Ed*(1-exp(-Sd.*vt));
```

% Converting time step number to minutes and add baselines

```
yp = [yb y']; %yhat = [yb';yhat'];
vest=[((0:0.1:4.9)*0);vest];
vk=((0:length(yp)-1))*(T/60); vt=(0:0.1:length(yp)-1)*(T/60);
```

% Plot of Frequency Shift vs Time

```
figure,
h=plot(vk,yp, '*b', vt, vest, '--r');
xlabel('Time (min)','FontSize',14); ylabel('Frequency Shift, \Deltaf (kHz)','FontSize',14)
h_legend = legend('Experimental data', 'Estimated Sensor Response');
set(h_legend,'FontSize',15);
set(h,'LineWidth',3)
grid on
```

%% Average Concentrations

```
gg = inno';
eps = mean(abs(gg));
[X, Ix]=min(eps);
```

```
Con = Con/3;
```

```

fprintf('\n~~~~~\n')
fprintf('\n Average Results \n')

fprintf('The estimated concentration of Analyte A (in ppm) is \n')
disp(Con(1))

fprintf('The estimated concentration of Analyte B (in ppm) is \n')
disp(Con(2))

fprintf('The estimated concentration of Analyte C (in ppm) is \n')
disp(Con(3))

Per_CA = ((Con(1)-CA)/CA)*100;
Per_CB = ((Con(2)-CB)/CB)*100;
Per_CC = ((Con(3)-CC)/CC)*100;

fprintf('\nPercentage Error of Concentration\n')
disp (Per_CA)
disp (Per_CB)
disp (Per_CC)

toc;

function [theta,innovation] =
EW_RLSE_4A(y,theta,P,W,lambdha,mA,mB,mC,mD,Sa,Sb,Sc,Sd,kmax)

for i=1:kmax
    H=[mA mB mC mD]';
    innovation=y(i)-theta(:,i)*H;

    %EW-RLSE identifier
    K = (P*H)/(H'*P*H+W*lambdha^2);
    theta(:,i+1) = theta(:,i) + K*[innovation];
    P = (1/(lambdha^2))*(P - P*(H*H')*P/(H'*P*H+(W*lambdha^2)));

    mA = ((1-Sa)*mA + Sa);
    mB = ((1-Sb)*mB + Sb);
    mC = ((1-Sc)*mC + Sc);
    mD = ((1-Sd)*mD + Sd);
end

function [x,P,h,cw,inno] = kalman_filter4A(x,P,y meas,C,W,V,w,Sa,Sb,Sc,Sd,i)
% KF Kalman Filter for linear dynamic systems
% Returns state estimate x, Error Covariance P, yhat(or h)
% Inputs: x: "a priori" state estimate
%         P: "a priori" estimated state covariance

```

```

%      ymeas: current measurement
%      V: process noise covariance
%      W: measurement noise covariance
% Output:  x: "a posteriori" state estimate
%      P: "a posteriori" state covariance

G = [1]; % Matrix G (1 by 1 Matrix)
F = eye(4);
U = 1; % Step Input
A = diag([1-Sa, 1-Sb, 1-Sc, 1-Sd]);
B = [Sa; Sb; Sc; Sd];

% Finding the innovation
h = C(1,1)*(1-exp(-Sa.*(i-1))) + C(1,2)*(1-exp(-Sb.*(i-1))) + C(1,3)*(1-exp(-Sc.*(i-1)))
+ C(1,4)*(1-exp(-Sd.*(i-1)));
inno = ymeas-h;

% Estimator
K = (A*P*C')/(C*P*C' + G*W*G');
x = A*x + B*U + K*inno;
P = (A-K*C)*P*(A-K*C)' + K*G*W*G'*K' + F*V*F';

% Weight Update Equations
S = C*P*C' + W;

% Weight before normalization
cw = ((abs(S))^-0.5)*(exp((-0.5)*(inno)*(inno))/(S))*w;

```

B.4.2 Five-Analyte Model

a) Sorption Data

```

%%
% Author: Karthick Sothivelr
%%
% Description: Code for Detection and Quantification of Target Analytes in
% the Presence of Interferents using EW-RLSE and Bank of Kalman Filters
% 5-Analyte Model (Sorption)
%% Sensor Details
% SH-SAW Coated with 0.6 um PECH

%% Cleaning
clear all

```

```

close all
clc
tic;

%% Open and read the measurement file
FID = fopen('pech_sorp_data.ini','r'); % Type in the file name
data = textscan(FID,'%f %f');
fclose(FID);

%% USER INPUT
% Type in the 'End Point' of data
data_end = [105];

% Actual Concentration from FROG (GC-PID)
CA = 0.325; CB = 0.465; CC = 0.540; start=3; q=10000;

%% Data Preprocessing (Selection of Analyte-In Point)
nn = 3; % Number of points to check to determine the best curve fit
T = 12; % Sampling period T=12s

tau = zeros(1,length(data_end));
Fe = zeros(1,length(data_end));
Eps = zeros(1,nn);

for i=1:length(data_end)
    % Section to find the best curve fit
    for j=1:nn
        if i==1
            y = data{2}((1+(j-1)):data_end(i));
        end
        if i>1
            y = data{2}((data_end(i-1)+1+(j-1)):data_end(i));
        end
        kmax = length(y);
        try
            [h_fit, G] = fit((0:kmax-1)',y,'fe*(1-exp(-S*x))','StartPoint',[1 1]);
        catch err
            j = j-1;
            break
        end
        Eps(j) = G.sse; % Sum of error squared
    end

    % Section to generate the best curve
    [Y,I] = min(Eps);
    if i==1

```

```

        y = data{2}((1+(I-1)):data_end(i));
    end
    if i>1
        y = data{2}((data_end(i-1)+1+(I-1)):data_end(i));
    end
    kmax = length(y);
    vt=(0:0.1:kmax-1)'; vk=(0:kmax-1)';
end
% Plot of Baseline Corrected Data
figure,
plot(y(start:end), '*')

%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5; % Benzene
tauB=77.6; % Toluene
tauC=175; % Ethylbenzene
tauD=460; % Trimethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; Sd=T/tauD; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=1.85; ssE=6; % Steady-State Sensitivity

tauE = [1000]; % Trimethylbenzene
Se = T./tauE;

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=y(start:end); yb = zeros(1,5); b=y; % y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Se),5); eps = zeros(length(Se),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 15; % Measurement Noise Covariance

for j=1:length(Se)

    P=Q*eye(5); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0; mE=0;

    theta=zeros(5,kmax+1); % Initialize the unknown parameters

    [theta,innovation] =
    EW_RLSE_5A(y,theta,P,W,lambdha,mA,mB,mC,mD,mE,Sa,Sb,Sc,Sd,Se(j),kmax);

```

```

%% Analysis
a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);
a5 = theta(5,kmax+1);

a(j,:) = [a1 a2 a3 a4 a5];

vvk=0:kmax-1;
% Estimated Frequency Shift
verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
exp(-Sd.*vvk)) + a5*(1-exp(-Se(j).*vvk));

ytest = y(1:length(vvk))';
eps(j,1) = immse(ytest,verr);

if a1>0 || a2>0 || a3>0 || a4>0 || a5>0
    eps(j,1)=100;
end

end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);
a5 = a(Ind,5);

clearvars Se
Se = T/tauE(Ind);

%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);
Con_E = abs(a5)/(ssE);

%% Bank of Kalman Filters Section
%% Set the concentration ranges for each analyte
% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;
fifth_a = 0.8*Con_E:0.02:1.2*Con_E;

```

```

% Frequency shift range
alpha1=-B*ssA;
alpha2=-Toluene*ssB;
alpha3=-EX*ssC;
alpha4=-TMB*ssD;
alpha5=-fifth_a*ssE;

% Generate the unknown parameters vector
ne = length(B)*length(Toluene)*length(EX)*length(TMB)*length(fifth_a); % Number of
filters
C = zeros(ne,5);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                for m=1:length(fifth_a)
                    C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l) alpha5(1,m)];
                    ind = ind+1;
                end
            end
        end
    end
end

%% Intialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([5,kmax+1,ne]);
P = zeros(5,5,kmax+1,ne);
xest = zeros(5,kmax+1); Pest = zeros(5,5,kmax+1);

Pt = diag([1 1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;
Psum = zeros(5,5);

```

```

%% Kalman Filter Scheme
for i=1:kmax
    % Setting the first measurement to zero
    if i==1
        y(i) = y(i)*0;
    end

    for j=1:ne
        [x(:,i+1,j),P(:, :,i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter(x(:,i,j),P(:, :,i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,Se,i);
    end
    % Normalized Weights
    c = sum(cw(:,i));
    w(:,i+1) = cw(:,i)./c;
end

%% Analysis

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);
Ee = C(I,5);

% Estimated Concentration
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);
Con_E = abs(Ee)/(ssE);

Con = [Con_A; Con_B; Con_C];

%% Section for Lower Limit of Time Constants
%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5-8; % Benzene
tauB=77.6-3; % Toluene
tauC=175-13; % Ethylbenzene
tauD=460; % Trimethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; Sd=T/tauD; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=1.85; ssE=6; % Steady-State Sensitivity

```

```

tauE = [1000]; % Trimethylbenzene
Se = T./tauE;

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=y(1:end); yb = zeros(1,5); b=y; %y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Se),5); eps = zeros(length(Se),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 15; % Measurement Noise Covariance

for j=1:length(Se)

    P=Q*eye(5); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0; mE=0;

    theta=zeros(5,kmax+1); % Initialize the unknown parameters

    [theta,innovation] =
    EW_RLSE_5A(y,theta,P,W,lambdha,mA,mB,mC,mD,mE,Sa,Sb,Sc,Sd,Se(j),kmax);

    %% Analysis
    a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
    a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);
    a5 = theta(5,kmax+1);

    a(j,:) = [a1 a2 a3 a4 a5];

    vvk=0:kmax-1;
    % Estimated Frequency Shift
    verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
    exp(-Sd.*vvk)) + a5*(1-exp(-Se(j).*vvk));

    ytest = y(1:length(vvk))';
    eps(j,1) = immse(ytest,verr);

    if a1>0 || a2>0 || a3>0 || a4>0 || a5>0
        eps(j,1)=100;
    end
end
end

```

```

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);
a5 = a(Ind,5);

clearvars Se
Se = T/tauE(Ind);
%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);
Con_E = abs(a5)/(ssE);

%% Bank of Kalman Filters Section
%% Set the concentration ranges for each analyte
% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;
fifth_a = 0.8*Con_E:0.02:1.2*Con_E;

% Frequency shift range
alpha1=-B*ssA;
alpha2=-Toluene*ssB;
alpha3=-EX*ssC;
alpha4=-TMB*ssD;
alpha5=-fifth_a*ssE;

% Generate the unknown parameters vector
ne = length(B)*length(Toluene)*length(EX)*length(TMB)*length(fifth_a); % Number of
filters
C = zeros(ne,5);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                for m=1:length(fifth_a)
                    C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l) alpha5(1,m)];
                    ind = ind+1;
                end
            end
        end
    end
end

```

```

    end
  end
end

%% Initialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([5,kmax+1,ne]);
P = zeros(5,5,kmax+1,ne);
xest = zeros(5,kmax+1); Pest = zeros(5,5,kmax+1);

Pt = diag([1 1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;

%% Kalman Filter Scheme
for i=1:kmax
    % Setting the first measurement to zero
    if i==1
        y(i) = y(i)*0;
    end

    for j=1:ne
        [x(:,i+1,j),P(:,:,i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter(x(:,i,j),P(:,:,i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,Se,i);
    end
    % Normalized Weights
    c = sum(cw(:,i));
    w(:,i+1) = cw(:,i)./c;
end

%% Analysis

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);

```

```
Ec = C(I,3);
Ed = C(I,4);
Ee = C(I,5);
```

```
% Estimated Concentration
```

```
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);
Con_E = abs(Ee)/(ssE);
```

```
Con = [Con_A; Con_B; Con_C]+Con;
```

```
%% Section for Upper Limit of Time Constants
```

```
%% Time Constant of Analytes (Coating: 0.6um PECH)
```

```
tauA=26.5+8; % Benzene
tauB=77.6+3; % Toluene
tauC=175+13; % Ethylbenzene
tauD=460; % Trimethylbenzene
```

```
Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; Sd=T/tauD; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=1.85; ssE=6; % Steady-State Sensitivity
```

```
tauE = [1000]; % Trimethylbenzene
Se = T./tauE;
```

```
%% Exponentially Weighted Recursive Least Square Estimation Section
```

```
%% Select Data to be analyzed
```

```
y=y(1:end); yb = zeros(1,5); b=y; % y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Se),5); eps = zeros(length(Se),1);
Q=q;
```

```
lambdha=0.99; % Weighting factor
W = 15; % Measurement Noise Covariance
```

```
for j=1:length(Se)
```

```
    P=Q*eye(5); % Error covariance
```

```
% Initial normalized concentration
```

```
    mA=0; mB=0; mC=0; mD=0; mE=0;
```

```
    theta=zeros(5,kmax+1); % Initialize the unknown parameters
```

```

[theta,innovation] =
EW_RLSE_5A(y,theta,P,W,lambdha,mA,mB,mC,mD,mE,Sa,Sb,Sc,Sd,Se(j),kmax);

%% Analysis
a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);
a5 = theta(5,kmax+1);

a(j,:) = [a1 a2 a3 a4 a5];

vvk=0:kmax-1;
vt=(0:0.1:kmax)';
% Estimated Frequency Shift
verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
exp(-Sd.*vvk)) + a5*(1-exp(-Se(j).*vvk));

ytest = y(1:length(vvk))';
eps(j,1) = immse(ytest,verr);

if a1>0 || a2>0 || a3>0 || a4>0 || a5>0
    eps(j,1)=100;
end

end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);
a5 = a(Ind,5);

clearvars Se
Se = T/tauE(Ind);
%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);
Con_E = abs(a5)/(ssE);

%% Bank of Kalman Filters Section
%% Set the concentration ranges for each analyte
% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;

```

```

TMB = 0.8*Con_D:0.02:1.2*Con_D;
fifth_a = 0.8*Con_E:0.02:1.2*Con_E;

% Frequency shift range
alpha1=-B*ssA;
alpha2=-Toluene*ssB;
alpha3=-EX*ssC;
alpha4=-TMB*ssD;
alpha5=-fifth_a*ssE;

% Generate the unknown parameters vector
ne = length(B)*length(Toluene)*length(EX)*length(TMB)*length(fifth_a); % Number of
filters
C = zeros(ne,5);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                for m=1:length(fifth_a)
                    C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l) alpha5(1,m)];
                    ind = ind+1;
                end
            end
        end
    end
end

%% Intialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([5,kmax+1,ne]);
P = zeros(5,5,kmax+1,ne);
xest = zeros(5,kmax+1); Pest = zeros(5,5,kmax+1);

Pt = diag([1 1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5 1e-5]);

```

```

W = 1e-5;

%% Bank of Kalman Filters Scheme
for i=1:kmax
    % Setting the first measurement to zero
    if i==1
        y(i) = y(i)*0;
    end

    for j=1:ne
        [x(:,i+1,j),P(:,:,i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter(x(:,i,j),P(:,:,i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,Se,i);
    end
    % Normalized Weights
    c = sum(cw(:,i));
    w(:,i+1) = cw(:,i)./c;
end

%% Plot and Analysis

vk=0:kmax; vvk=0:kmax-1;
vt=(0:0.1:kmax-1)';

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);
Ee = C(I,5);

% Estimated Concentration
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);
Con_E = abs(Ee)/(ssE);

Con = [Con_A; Con_B; Con_C]+Con;

% Estimated Frequency Shift
vest = Ea*(1-exp(-Sa.*vt)) + Eb*(1-exp(-Sb.*vt)) + Ec*(1-exp(-Sc.*vt)) + Ed*(1-exp(-
Sd.*vt)) + Ee*(1-exp(-Se.*vt));

% Converting time step number to minutes and add baselines

```

```

yp = [yb y']; % yhat = [yb';yhat'];
vest=[((0:0.1:4.9)*0);vest];
vk=((0:length(yp)-1))*(T/60); vt=(0:0.1:length(yp)-1)*(T/60);

% Plot of Frequency Shift vs Time step
figure,
h=plot(vk,yp, '*b', vt, vest, '--r');
xlabel('Time (min)',FontSize,14); ylabel('Frequency Shift, \Deltaf (kHz)',FontSize,14)
h_legend = legend('Experimental data', 'Estimated Sensor Response');
set(h_legend,FontSize,15);
set(h,LineWidth,3)
grid on

%% Average Concentrations
gg = inno';
eps = mean(abs(gg));
[X, Ix]=min(eps);

Con = Con/3;

fprintf('\n~~~~~\n')
fprintf('\n Average Results \n')

fprintf('The estimated concentration of Analyte A (in ppm) is \n')
disp(Con(1))

fprintf('The estimated concentration of Analyte B (in ppm) is \n')
disp(Con(2))

fprintf('The estimated concentration of Analyte C (in ppm) is \n')
disp(Con(3))

Per_CA = ((Con(1)-CA)/CA)*100;
Per_CB = ((Con(2)-CB)/CB)*100;
Per_CC = ((Con(3)-CC)/CC)*100;

fprintf('\nPercentage Error of Concentration\n')
disp (Per_CA)
disp (Per_CB)
disp (Per_CC)

toc;

function [theta,innovation] =
EW_RLSE_5A(y,theta,P,W,lambdha,mA,mB,mC,mD,mE,Sa,Sb,Sc,Sd,Se,kmax)

```

```

for i=1:kmax
    H=[mA mB mC mD mE]';
    innovation=y(i)-theta(:,i)*H;

    %EW-RLSE identifier
    K = (P*H)/(H'*P*H+W*lambda^2);
    theta(:,i+1) = theta(:,i) + K*[innovation];
    P = (1/(lambda^2))*(P - P*(H*H')*P/(H'*P*H+(W*lambda^2)));

    mA = ((1-Sa)*mA + Sa);
    mB = ((1-Sb)*mB + Sb);
    mC = ((1-Sc)*mC + Sc);
    mD = ((1-Sd)*mD + Sd);
    mE = ((1-Se)*mE + Se);
end

function [x,P,h,cw,inno] = kalman_filter(x,P,y meas,C,W,V,w,Sa,Sb,Sc,Sd,Se,i)
% KF Kalman Filter for linear dynamic systems
% Returns state estimate x, Error Covariance P, yhat(or h)
% Inputs:  x: "a priori" state estimate
%          P: "a priori" estimated state covariance
%          y meas: current measurement
%          V: process noise covariance
%          W: measurement noise covariance
% Output:  x: "a posteriori" state estimate
%          P: "a posteriori" state covariance

G = [1]; % Matrix G (1 by 1 Matrix)
F = eye(5);
U = 1; % Step Input
A = diag([1-Sa, 1-Sb, 1-Sc, 1-Sd, 1-Se]);
B = [Sa; Sb; Sc; Sd; Se];
%h = C*x;

% Finding the innovation
h = C(1,1)*(1-exp(-Sa.*(i-1))) + C(1,2)*(1-exp(-Sb.*(i-1))) + C(1,3)*(1-exp(-Sc.*(i-1)))
+ C(1,4)*(1-exp(-Sd.*(i-1))) + C(1,5)*(1-exp(-Se.*(i-1)));
inno = y meas-h;

% Estimator
K = (A*P*C')/(C*P*C' + G*W*G');
x = A*x + B*U + K*inno;
P = (A-K*C)*P*(A-K*C)' + K*G*W*G'*K' + F*V*F';

% Weight Update Equations
S = C*P*C' + W;

```

```
% Weight before normalization
cw = ((abs(S))(-0.5))*(exp((-0.5)*(inno)*(inno))/(S))*w;
```

b) Desorption Data

```
%%
% Author: Karthick Sothivelr
%%
% Description: Code for Detection and Quantification of Target Analytes in
% the Presence of Interferents using EW-RLSE and Bank of Kalman Filters
% 5-Analyte Model (Desorption)
%% Sensor Details
% SH-SAW Coated with 0.6 um PECH

%% Cleaning
clear all
close all
clc
tic;

%% Open and read the measurement file
FID = fopen('pech_desorp_data.ini','r'); % Type in the file name
data = textscan(FID,'%f %f');
fclose(FID);

%% USER INPUT
% Type in the 'End Point' of data
data_end = [101];

% Actual Concentration from FROG (GC-PID)
CA = 0.325; CB = 0.565; CC = 0.540; start=0+1; q=10000;

%% Data Preprocessing (Selection of Analyte-In Point)
T = 12; % Sampling period T=12s

y0 = data{2}((start+1:data_end(1))); kmax = length(y0); y0 = y0-y0(1);
y1 = data{2}((start+2:data_end(1))); kmax1 = length(y1); y1 = y1-y1(1);
y2 = data{2}((start+3:data_end(1))); kmax2 = length(y2); y2 = y2-y2(1);
y3 = data{2}((start+4:data_end(1))); kmax3 = length(y3); y3 = y3-y3(1);
y4 = data{2}((start+5:data_end(1))); kmax4 = length(y4); y4 = y4-y4(1);
y5 = data{2}((start+6:data_end(1))); kmax5 = length(y5); y5 = y5-y5(1);
y6 = data{2}((start+7:data_end(1))); kmax6 = length(y6); y6 = y6-y6(1);
y7 = data{2}((start+8:data_end(1))); kmax7 = length(y7); y7 = y7-y7(1);
```

```

y=y0;

% Baseline Corrected Data
vk=0:length(y)-1;
figure,
h=plot(vk*(T/60),y,'*');
xlabel('Time (min)',FontSize,14); ylabel('Frequency Shift, \Deltaf (kHz)',FontSize,14)
set(h,'LineWidth',3)
grid on

%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5; % Benzene
tauB=77.6; % Toluene
tauC=175; % Ethylbenzene
tauD=460; % Trimethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; Sd=T/tauD; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=1.85; ssE=6; % Steady-State Sensitivity

tauE = [1000]; % Trimethylbenzene
Se = T./tauE;

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=y(1:end); yb = zeros(1,5); b=y; % y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Se),5); eps = zeros(length(Se),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 1; % Measurement Noise Covariance

for j=1:length(Se)

    P=Q*eye(5); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0; mE=0;

    theta=zeros(5,kmax+1); % Initialize the unknown parameters

    [theta,innovation] =
    EW_RLSE_5A(y,theta,P,W,lambdha,mA,mB,mC,mD,mE,Sa,Sb,Sc,Sd,Se(j),kmax);

    %% Analysis

```

```

a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);
a5 = theta(5,kmax+1);

a(j,:) = [a1 a2 a3 a4 a5];

vvk=0:kmax-1;
% Estimated Frequency Shift
verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
exp(-Sd.*vvk)) + a5*(1-exp(-Se(j).*vvk));

ytest = y(1:length(vvk))';
eps(j,1) = immse(ytest,verr);

if a1<0 || a2<0 || a3<0 || a4<0 || a5<0
    eps(j,1)=100;
end

end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);
a5 = a(Ind,5);

clearvars Se
Se = T/tauE(Ind);
%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);
Con_E = abs(a5)/(ssE);

%% Bank of Kalman Filters Section
%% Set the concentration ranges for each analyte

% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;
fifth_a = 0.8*Con_E:0.02:1.2*Con_E;

% Frequency shift range

```

```

alpha1=B*ssA;
alpha2=Toluene*ssB;
alpha3=EX*ssC;
alpha4=TMB*ssD;
alpha5=fifth_a*ssE;

% Generate the unknown parameters vector
ne = length(B)*length(Toluene)*length(EX)*length(TMB)*length(fifth_a); % Number of
filters
C = zeros(ne,5);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                for m=1:length(fifth_a)
                    C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l) alpha5(1,m)];
                    ind = ind+1;
                end
            end
        end
    end
end

%% Intialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([5,kmax+1,ne]);
P = zeros(5,5,kmax+1,ne);
xest = zeros(5,kmax+1); Pest = zeros(5,5,kmax+1);

Pt = diag([1 1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;

%% Bank of Kalman Filters Scheme
for i=1:kmax

```

```

% Setting the first measurement to zero
if i==1
    y(i) = y(i)*0;
end

for j=1:ne
    [x(:,i+1,j),P(:,:,i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter(x(:,i,j),P(:,:,i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,Se,i);
end
% Normalized weights
c = sum(cw(:,i));
w(:,i+1) = cw(:,i)./c;
end

%% Analysis

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);
Ee = C(I,5);

% Estimated Concentration
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);
Con_E = abs(Ee)/(ssE);

Con = [Con_A; Con_B; Con_C];

%% Section for Lower Limit of Time Constants
%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5-8; % Benzene
tauB=77.6-3; % Toluene
tauC=175-13; % Ethylbenzene
tauD=460; % Trimethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; Sd=T/tauD; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=1.85; ssE=6; % Steady-State Sensitivity

tauE = [1000]; % Trimethylbenzene
Se = T./tauE;

```

```

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=y(1:end); yb = zeros(1,5); b=y; %y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Se),5); eps = zeros(length(Se),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 1; % Measurement Noise Covariance

for j=1:length(Se)

    P=Q*eye(5); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0; mE=0;

    theta=zeros(5,kmax+1); % Initialize the unknown parameters

    [theta,innovation] =
    EW_RLSE_5A(y,theta,P,W,lambdha,mA,mB,mC,mD,mE,Sa,Sb,Sc,Sd,Se(j),kmax);

    %% Analysis
    a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
    a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);
    a5 = theta(5,kmax+1);

    a(j,:) = [a1 a2 a3 a4 a5];

    vvk=0:kmax-1;
    % Estimated Frequency Shift
    verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
    exp(-Sd.*vvk)) + a5*(1-exp(-Se(j).*vvk));

    ytest = y(1:length(vvk))';
    eps(j,1) = immse(ytest,verr);

    if a1<0 || a2<0 || a3<0 || a4<0 || a5<0
        eps(j,1)=100;
    end
end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);

```

```

a3 = a(Ind,3); a4 = a(Ind,4);
a5 = a(Ind,5);

clearvars Se
Se = T/tauE(Ind);

%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);
Con_E = abs(a5)/(ssE);

%% Bank of Kalman Filters Section
%% Set the concentration ranges for each analyte
% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;
fifth_a = 0.8*Con_E:0.02:1.2*Con_E;

% Frequency shift range
alpha1=B*ssA;
alpha2=Toluene*ssB;
alpha3=EX*ssC;
alpha4=TMB*ssD;
alpha5=fifth_a*ssE;

ne = length(B)*length(Toluene)*length(EX)*length(TMB)*length(fifth_a); % Number of
filters
C = zeros(ne,5);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                for m=1:length(fifth_a)
                    C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l) alpha5(1,m)];
                    ind = ind+1;
                end
            end
        end
    end
end
end
end
end
end

```

```

%% Initialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([5,kmax+1,ne]);
P = zeros(5,5,kmax+1,ne);
xest = zeros(5,kmax+1); Pest = zeros(5,5,kmax+1);

Pt = diag([1 1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;

%% Bank of Kalman Filters Scheme
for i=1:kmax
    % Setting the first measurement to zero
    if i==1
        y(i) = y(i)*0;
    end

    for j=1:ne
        [x(:,i+1,j),P(:,:,i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter(x(:,i,j),P(:,:,i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,Se,i);
    end
    % Normalized Weights
    c = sum(cw(:,i));
    w(:,i+1) = cw(:,i)/c;
end

%% Analysis

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);
Ee = C(I,5);

```

```

% Estimated Concentration
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);
Con_E = abs(Ee)/(ssE);

Con = [Con_A; Con_B; Con_C]+Con;

%% Section for Upper Limit of Time Constants
%% Time Constant of Analytes (Coating: 0.6um PECH)
tauA=26.5+8; % Benzene
tauB=77.6+3; % Toluene
tauC=175+13; % Ethylbenzene
tauD=460; % Trimethylbenzene

Sa=T/tauA; Sb=T/tauB; Sc=T/tauC; Sd=T/tauD; % Absorption Rate
ssA=0.109; ssB=0.435; ssC=1.450; ssD=1.85; ssE=6; % Steady-State Sensitivity

tauE = [1000]; % Trimethylbenzene
Se = T./tauE;

%% Exponentially Weighted Recursive Least Square Estimation Section
%% Select Data to be analyzed
y=y(1:end); yb = zeros(1,5); b=y; % y(1)=0;
kmax = length(y); % Length of the measurement data points
a = zeros(length(Se),5); eps = zeros(length(Se),1);
Q=q;

lambdha=0.99; % Weighting factor
W = 1; % Measurement Noise Covariance

for j=1:length(Se)

    P=Q*eye(5); % Error covariance

    % Initial normalized concentration
    mA=0; mB=0; mC=0; mD=0; mE=0;

    theta=zeros(5,kmax+1); % Initialize the unknown parameters

    [theta,innovation] =
    EW_RLSE_5A(y,theta,P,W,lambdha,mA,mB,mC,mD,mE,Sa,Sb,Sc,Sd,Se(j),kmax);

    %% Analysis

```

```

a1 = theta(1,kmax+1); a2 = theta(2,kmax+1);
a3 = theta(3,kmax+1); a4 = theta(4,kmax+1);
a5 = theta(5,kmax+1);

a(j,:) = [a1 a2 a3 a4 a5];

vvk=0:kmax-1;
% Estimated Frequency Shift
verr = a1*(1-exp(-Sa.*vvk)) + a2*(1-exp(-Sb.*vvk)) + a3*(1-exp(-Sc.*vvk)) + a4*(1-
exp(-Sd.*vvk)) + a5*(1-exp(-Se(j).*vvk));

ytest = y(1:length(vvk))';
eps(j,1) = immse(ytest,verr);

if a1<0 || a2<0 || a3<0 || a4<0 || a5<0
    eps(j,1)=100;
end

end

[ms_error, Ind] = min(eps);

a1 = a(Ind,1); a2 = a(Ind,2);
a3 = a(Ind,3); a4 = a(Ind,4);
a5 = a(Ind,5);

clearvars Se
Se = T/tauE(Ind);

%% Estimated Concentration
Con_A = abs(a1)/(ssA);
Con_B = abs(a2)/(ssB);
Con_C = abs(a3)/(ssC);
Con_D = abs(a4)/(ssD);
Con_E = abs(a5)/(ssE);

%% Bank of Kalman Filters Section
%% Set the concentration ranges for each analyte
% Concentration Range in ppm
B = 0.8*Con_A:0.02:1.2*Con_A;
Toluene = 0.8*Con_B:0.02:1.2*Con_B;
EX = 0.8*Con_C:0.02:1.2*Con_C;
TMB = 0.8*Con_D:0.02:1.2*Con_D;
fifth_a = 0.8*Con_E:0.02:1.2*Con_E;

% Frequency shift range

```

```

alpha1=B*ssA;
alpha2=Toluene*ssB;
alpha3=EX*ssC;
alpha4=TMB*ssD;
alpha5=fifth_a*ssE;

% Generate the unknown parameters vector
ne = length(B)*length(Toluene)*length(EX)*length(TMB)*length(fifth_a); % Number of
filters
C = zeros(ne,5);

ind=1;
for i=1:length(B)
    for j=1:length(Toluene)
        for k=1:length(EX)
            for l=1:length(TMB)
                for m=1:length(fifth_a)
                    C(ind,:) = [alpha1(1,i) alpha2(1,j) alpha3(1,k) alpha4(1,l) alpha5(1,m)];
                    ind = ind+1;
                end
            end
        end
    end
end

%% Intialize the weights of each filter
w = zeros([ne,kmax+1]); cw=zeros([ne,kmax]);% weight
w(:,1) = w(:,1) + 1/ne;

%% Initialization of State and Error Covariance
yhat = zeros([ne,kmax]); inno=yhat;
x = zeros([5,kmax+1,ne]);
P = zeros(5,5,kmax+1,ne);
xest = zeros(5,kmax+1); Pest = zeros(5,5,kmax+1);

Pt = diag([1 1 1 1 1]);
for i=1:ne
    P(:,:,1,i) = P(:,:,1,i) + Pt;
end

%% Process and Measurement Noise
V = diag([1e-5 1e-5 1e-5 1e-5 1e-5]);
W = 1e-5;
Psum = zeros(5,5);

%% Bank of Kalman Filters Scheme

```

```

for i=1:kmax
    % Setting the first measurement to zero
    if i==1
        y(i) = y(i)*0;
    end

    for j=1:ne
        [x(:,i+1,j),P(:, :, i+1,j),yhat(j,i),cw(j,i),inno(j,i)] =
kalman_filter(x(:,i,j),P(:, :, i,j),y(i),C(j,:),W,V,w(j,i),Sa,Sb,Sc,Sd,Se,i);
    end
    c = sum(cw(:,i));
    w(:,i+1) = cw(:,i)./c;

    % Estimate of state and error covariance
    xx = x(:,i+1,:);
    xx = permute(xx,[3 1 2]);
    xest(:,i+1) = w(:,i+1)*xx(:, :, 1);

    for j=1:ne
        Psum = w(j,i+1)*(P(:, :, i+1,j)+(x(:,i+1,j)*(x(:,i+1,j)))') + Psum;
    end
    Pest(:, :, i+1) = Psum - xest(:,i+1)*xest(:,i+1)';
    Psum = zeros(5,5);

end

%% Plot and Analysis

vk=0:kmax; vvk=0:kmax-1;
vt=(0:0.1:kmax-1)';

[Y, I]=max(w(:,kmax+1));
[Bs, Is]=sort(w(:,kmax+1),'descend');

Ea = C(I,1);
Eb = C(I,2);
Ec = C(I,3);
Ed = C(I,4);
Ee = C(I,5);

% Estimated Concentration
Con_A = abs(Ea)/(ssA);
Con_B = abs(Eb)/(ssB);
Con_C = abs(Ec)/(ssC);
Con_D = abs(Ed)/(ssD);
Con_E = abs(Ee)/(ssE);

```

```

Con = [Con_A; Con_B; Con_C]+Con;

% Estimated Frequency Shift
vest = Ea*(1-exp(-Sa.*vt)) + Eb*(1-exp(-Sb.*vt)) + Ec*(1-exp(-Sc.*vt)) + Ed*(1-exp(-
Sd.*vt)) + Ee*(1-exp(-Se.*vt));

% Converting time step number to minutes and add baselines
yp = [yb y']; %yhat = [yb';yhat'];
vest=[((0:0.1:4.9)*0);vest];
vk=((0:length(yp)-1))*(T/60); vt=(0:0.1:length(yp)-1)*(T/60);

% Plot of Frequency Shift vs Time
figure,
h=plot(vk,yp, '*b', vt, vest, '--r');
xlabel('Time (min)', 'FontSize',14); ylabel('Frequency Shift, \Deltaf (kHz)', 'FontSize',14)
h_legend = legend('Experimental data', 'Estimated Sensor Response');
set(h_legend, 'FontSize',15);
set(h, 'LineWidth',3)
grid on

%% Average Concentrations
gg = inno';
eps = mean(abs(gg));
[X, Ix]=min(eps);

Con = Con/3;

% Display in Command Window
fprintf('\n~~~~~\n')
fprintf('\n Average Results \n')

fprintf('The estimated concentration of Analyte A (in ppm) is \n')
disp(Con(1))

fprintf('The estimated concentration of Analyte B (in ppm) is \n')
disp(Con(2))

fprintf('The estimated concentration of Analyte C (in ppm) is \n')
disp(Con(3))

Per_CA = ((Con(1)-CA)/CA)*100;
Per_CB = ((Con(2)-CB)/CB)*100;
Per_CC = ((Con(3)-CC)/CC)*100;

fprintf('\nPercentage Error of Concentration\n')

```

```
disp (Per_CA)
disp (Per_CB)
disp (Per_CC)
```

```
toc;
```

```
function [theta,innovation] =
EW_RLSE_5A(y,theta,P,W,lambdha,mA,mB,mC,mD,mE,Sa,Sb,Sc,Sd,Se,kmax)
```

```
for i=1:kmax
```

```
    H=[mA mB mC mD mE]';
    innovation=y(i)-theta(:,i)'*H;
```

```
    %EW-RLSE identifier
```

```
    K = (P*H)/(H'*P*H+W*lambdha^2);
    theta(:,i+1) = theta(:,i) + K*[innovation];
    P = (1/(lambdha^2))*(P - P*(H*H')*P/(H'*P*H+(W*lambdha^2)));
```

```
    mA = ((1-Sa)*mA + Sa);
    mB = ((1-Sb)*mB + Sb);
    mC = ((1-Sc)*mC + Sc);
    mD = ((1-Sd)*mD + Sd);
    mE = ((1-Se)*mE + Se);
```

```
end
```

```
function [x,P,h,cw,inno] = kalman_filter(x,P,y meas,C,W,V,w,Sa,Sb,Sc,Sd,Se,i)
```

```
% KF Kalman Filter for linear dynamic systems
% Returns state estimate x, Error Covariance P, yhat(or h)
% Inputs:  x: "a priori" state estimate
%          P: "a priori" estimated state covariance
%          ymeas: current measurement
%          V: process noise covariance
%          W: measurement noise covariance
% Output:  x: "a posteriori" state estimate
%          P: "a posteriori" state covariance
```

```
G = [1]; % Matrix G (1 by 1 Matrix)
```

```
F = eye(5);
```

```
U = 1; % Step Input
```

```
A = diag([1-Sa, 1-Sb, 1-Sc, 1-Sd, 1-Se]);
```

```
B = [Sa; Sb; Sc; Sd; Se];
```

```
%h = C*x;
```

```
% Finding the innovation
```

```
h = C(1,1)*(1-exp(-Sa.*(i-1))) + C(1,2)*(1-exp(-Sb.*(i-1))) + C(1,3)*(1-exp(-Sc.*(i-1)))
+ C(1,4)*(1-exp(-Sd.*(i-1))) + C(1,5)*(1-exp(-Se.*(i-1)));
```

```
inno = ymeas-h;
```

```
% Estimator
```

```
K = (A*P*C')/(C*P*C' + G*W*G');
```

```
x = A*x + B*U + K*inno;
```

```
P = (A-K*C)*P*(A-K*C)' + K*G*W*G'*K' + F*V*F';
```

```
% Weight Update Equations
```

```
S = C*P*C' + W;
```

```
% Weight before normalization
```

```
cw = ((abs(S))-0.5)*(exp(((-0.5)*(inno)*(inno))/(S))))*w;
```