

AUTOMATIC DETECTION OF SENSOR CALIBRATION ERRORS IN MINING
INDUSTRY

By

Rambabu Pothina, M.S., B.E.

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Engineering

University of Alaska Fairbanks

December 2017

© Rambabu Pothina

APPROVED:

Rajive Ganguli, Committee Chair

Tathagata Ghosh, Committee Member

Orion Lawlor, Committee Member

Ronald Barry, Committee Member

Margaret Darrow, Chair

*Department of Mining and Geological
Engineering*

Douglas Goering, Dean

College of Engineering and Mines

Michael Castellini, *Dean of the Graduate School*

Abstract

Sensor errors cost the mining industry millions of dollars in losses each year. Unlike gross errors, “calibration errors” are subtle, develop over time, and are difficult to identify. Economic losses start accumulating even when errors are small. Therefore, the aim of this research was to develop methods to identify calibration errors well before they become obvious. The goal in this research was to detect errors at a bias as low as 2% in magnitude. The innovative strategy developed relied on relationships between a variety of sensors to detect when a given sensor started to stray. Sensors in a carbon stripping circuit at a gold processing facility (Pogo Mine) in Alaska were chosen for the study.

The results from the initial application of classical statistical methods like correlation, aggregation and principal component analysis (PCA), and the signal processing methods (FFT), to find bias ($\pm 10\%$) in “feed” sensor data from a semi-autogenous (SAG) grinding mill operation (Fort Knox mine, Alaska) were not promising due to the non-linear and non-stationary nature of the process characteristics. Therefore, those techniques were replaced with some innovative data mining techniques when the focus shifted to Pogo Mine, where the task was to detect calibration errors in strip vessel temperature sensors in the carbon stripping circuit. The new techniques used data from two strip vessel temperature sensors (S1 and S2), four heat exchanger related temperature sensors (H1 through H4), barren flow sensor (BARNFL) and a glycol flow sensor (GLYFL). These eight sensors were deemed to be part of the same process. To detect when the calibration of one of the strip vessel temperature sensors, S1, started to stray, tests were designed to detect changes in relationship between the eight temperature sensors. Data was filtered (“threshold”) based on process characteristics prior to being used in tests. The tests combined basic concepts such as moving windows of time, ratios (ratio of one sensor data to data from a set of

sensors), tracking of maximum values, etc. Error was triggered when certain rules were violated. A 2% error was randomly introduced into one of the two strip vessel temperature data streams to simulate calibration errors. Some tests were less effective than others at detecting the simulated errors. The tests that used GLYFL and BARNFL were not very effective. On the other hand, the tests that used total “Heat” of all the heat exchanger sensors were very effective. When the tests were administered together (“Combined test”), they have a high success rate (95%) in terms of True alarms, i.e., tests detecting bias after it is introduced. In those True alarms, for 75% of the cases, the introduction of the error was detected within 39.5 days. A -2% random error was detected with a similar success rate.

Dedicated to

My wife Lakshmi Prasuna, and my Chito, a feline grace, for their unconditional love, support, and understanding. And, to the memories of my beloved grandfather Sreeramulu Kolla, the smartest and most industrious person I have ever known, whom I miss dearly.

Table of Contents

Page

Title Page.....	i
Abstract.....	iii
Dedication.....	v
Table of Contents.....	vi
List of Figures.....	xiv
List of Tables.....	xx
Acknowledgements.....	xxiii
Chapter 1: Introduction.....	1
1.1 Background.....	1
1.1.1 Organization of the Dissertation.....	2
1.1.2 Objective of the Work.....	3
1.1.3 Problem Statement.....	4
1.1.4 Scope of the Research.....	4
1.1.5 Challenges.....	5
1.2 Literature Review.....	6
1.2.1 Functionality of Sensors.....	6
1.2.2 Accuracy of a Sensor and Role of Calibration.....	7
1.2.3 Classification of Sensor Faults.....	8

1.2.4	Sensor Faults and Impact on Operating Costs.....	9
1.2.5	Fault Detection: Statistical Methods.....	10
1.2.6	Fault Detection: Signal Processing Methods.....	11
1.3	Introduction to Gold Mining, Fort Knox and Pogo	13
1.3.1	Carbon Stripping Circuit	14
1.4	Methods and Materials.....	17
1.5	Overview of Work Done and Conclusions	17
1.6	Future Work	20
1.7	Acknowledgements.....	20
1.8	References.....	21
Chapter 2:	Statistical Methods in the Detection of Industrial Sensor Calibration Errors.....	25
2.1	Abstract.....	25
2.2	Introduction.....	26
2.3	Literature Review.....	27
2.3.1	Estimation Based Methods	27
2.3.2	Time Series Methods	28
2.3.3	Neural Networks.....	30
2.3.4	Bayesian Statistics	31
2.3.5	Conservation Laws and Data Reconciliation.....	32
2.3.6	Data-Mining Methods.....	33

2.3.7	Heuristic Methods.....	34
2.3.8	Data Aggregation.....	35
2.3.9	Parametric and Non-Parametric Methods.....	36
2.3.10	Hybrid Methods.....	36
2.4	Introduction to Fort Knox Mine and Mill Operations	37
2.5	Methods and Materials.....	38
2.5.1	Data collection.....	38
2.5.2	Assumptions and scope	40
2.5.3	Data Preparation	40
2.5.3.1	Programming and Software	40
2.5.4	Correlation Based Sensor Data Characterization Method.....	41
2.5.4.1	Data Characterization.....	42
2.5.4.2	Description of Algorithm	44
2.5.5	Aggregation Techniques and Lags to Observe Trends.....	47
2.5.6	Use of Principal Component Analysis (PCA) in Sensor Fault Detection	49
2.6	Results.....	50
2.6.1	Correlation Based Sensor Data Characterization Method.....	50
2.6.2	Aggregation and Lags.....	51
2.6.3	Use of Principal Component Analysis (PCA) in sensor fault detection	52
2.7	Discussion.....	55

2.8	Conclusions.....	56
2.9	Acknowledgements.....	57
2.10	References.....	58
Chapter 3: Signal Processing Methods in the Detection of Industrial Sensor Calibration		
	Errors.....	63
3.1	Abstract.....	63
3.2	Introduction.....	64
3.3	Literature Review.....	66
3.4	Introduction to Pogo Mine and Mill Operations.....	67
3.4.1	Pogo Mine and Mill.....	67
3.4.2	Carbon-in-Pulp (CIP) Circuit.....	70
3.4.3	Carbon Stripping Circuit.....	71
3.5	Methods and Materials.....	75
3.5.1	Data Collection.....	75
3.5.2	Assumptions and scope.....	79
3.5.3	Data Preparation.....	80
3.5.3.1	Programming and Software.....	80
3.5.3.2	Algorithm Description.....	80
3.5.4	Application of Fast Fourier Transform (Clean vs Biased data).....	82
3.5.4.1	Fast Fourier Transform (Theoretical Background).....	82

3.5.4.2	Digital Signal Processing and Digital Filtering	86
3.6	Results.....	88
3.7	Discussion.....	91
3.8	Conclusions.....	92
3.9	Acknowledgements.....	93
3.10	References.....	94
Chapter 4:	Peak-Readings Count and Sensitivity Analysis (PRCSA), an Innovative Data-Mining Technique in the Detection of Industrial Sensor Calibration Errors.....	97
4.1	Abstract.....	97
4.2	Introduction.....	98
4.3	Literature Review.....	99
4.4	Introduction to Pogo Mine and Mill	101
4.4.1	Carbon-in-Pulp (CIP) Circuit	103
4.4.2	Carbon Stripping Circuit	104
4.5	Methods and Materials.....	108
4.5.1	Data Collection	108
4.5.2	Assumptions and scope	112
4.5.3	Data Preparation	113
4.5.3.1	Programming and Software	113
4.5.3.2	Basic Concepts	114

4.5.3.3	Algorithm Description	115
4.5.4	Peak-Readings Count Analysis (PRCA)	117
4.5.5	Peak-Readings Count and Sensitivity Analysis (PRCSA)	119
4.6	Results	120
4.6.1	Peak-Readings Count Analysis (PRCA)	120
4.6.2	Peak-Readings Count Sensitivity Analysis (PRCSA)	124
4.7	Discussion	130
4.8	Conclusions	132
4.9	Acknowledgements	133
4.10	References	133
Chapter 5:	Multiple Ratio Function Analysis with Automation (MRFAA) in the Detection of Industrial Sensor Calibration Errors	137
5.1	Abstract	137
5.2	Introduction	138
5.3	Literature Review	139
5.4	Introduction to Pogo Mine and Mill	140
5.4.1	Carbon-in-Pulp (CIP) Circuit	143
5.4.2	Carbon Stripping Circuit	144
5.5	Methods and Materials	148
5.5.1	Data Collection	148

5.5.2	Assumptions and scope	153
5.5.3	Data Preparation	154
5.5.3.1	Programming and Software	154
5.5.3.2	Matlab, Super Computing, and Fortran	155
5.5.4	Background on Concepts, Terminology, and Previous Methods	155
5.5.5	Multiple Ratio Function Analysis (MRFA).....	162
5.5.5.1	MRFA: Algorithm Description.....	168
5.5.5.2	MRFA with Multiple Tests and Automation (MRFAA)	179
5.5.5.3	MRFAA: Algorithm Description	195
5.6	Results and Validation	197
5.6.1.1	MRFAA Algorithm.....	197
5.7	Discussion.....	204
5.8	Conclusions.....	205
5.9	Acknowledgements.....	207
5.10	References.....	207
5.11	Appendix.....	210
5.11.1	Heat Ratio Test Performance.....	210
5.11.2	Max Value Test Performance	212
5.11.3	Ave Value Test Performance.....	214
5.11.4	BARNFL Ratio Test Performance	216

5.11.5	GLYFL Ratio Test Performance	218
--------	------------------------------------	-----

List of Figures	Page
Figure 1.1: A car coolant temperature sensor mechanism, and an industrial	6
Figure 1.2: A level sensor from Pogo flotation circuit.....	7
Figure 1.3: Overall accuracy of a sensor.	8
Figure 1.4: A constant and short faults.....	9
Figure 1.5: Time domain vs frequency domain.....	12
Figure 1.6: Flow sheet for a typical gold processing operation.	14
Figure 1.7: Pogo stripping circuit schematic diagram with sensor placements.	16
Figure 1.8: Various sensors and their interrelations.....	17
Figure 2.1: Robust vs linear fit.....	28
Figure 2.2: Autocorrelation plot of residuals from ARIMA (2,1,0) Model.....	29
Figure 2.3: Sample plot: smoothed data for 2 values of α (a factor).....	30
Figure 2.4: A typical Neural Network architecture.....	31
Figure 2.5: Error progression with training in an ANN.....	31
Figure 2.6: Fort Knox mine.....	37
Figure 2.7: A typical SAG mill circuit.....	39
Figure 2.8: Sensor raw data “states” in time domain.....	42
Figure 2.9: Correlation plot (cleaned data).....	43
Figure 2.10: Sensor raw data in time domain after removing outliers.....	44
Figure 2.11: Flow chart for the Matlab-based data characterization algorithm.....	45

Figure 2.12: Matlab code with various model options.....	46
Figure 2.13: State 3 fitted line plot – BP vs. HP.	46
Figure 2.14: State 3 fitted line plot – RPM vs. HP.....	46
Figure 2.15: Data trends at 60 min aggregation.	48
Figure 2.16: Data trends at 120 min aggregation.	48
Figure 2.17: Data trends at 240 min aggregation.	49
Figure 2.18: Clean vs 10% noisy/bias data.	50
Figure 2.19: Variance for clean data set (pc.cr).	53
Figure 2.20: variance for noisy data set (pc.crN).	53
Figure 2.21: Biplots for clean data set.....	54
Figure 2.22: Biplots for noisy data set.....	54
Figure 3.1: Noisy signal.	64
Figure 3.2: Time domain vs frequency domain.....	65
Figure 3.3: Vibration signal from a SAG mill operation.....	67
Figure 3.4: Pogo Mine site map.	68
Figure 3.5: Pogo mine-gold processing flow sheet.	69
Figure 3.6: CIP circuit at Pogo Mine.....	70
Figure 3.7: Pogo stripping circuit in pictures.	73
Figure 3.8: Pogo stripping circuit schematic diagram with sensor placements.	74
Figure 3.9: Strip vessel-1 heat sensor (S1).....	76

Figure 3.10: Strip vessel-2 heat sensor (S2).....	76
Figure 3.11: Heat Exchanger-1 Sensor (H1).....	77
Figure 3.12: Heat Exchanger-2 Sensor (H2).....	77
Figure 3.13: Barren flow Sensor (BARNFL).....	78
Figure 3.14: Glycol flow sensor (GLYFL).....	78
Figure 3.15: FFT analysis algorithm.	82
Figure 3.16: Time Domain vs frequency domain.....	83
Figure 3.17: Periodicity of sensor data measurements from Pogo stripping circuit.	85
Figure 3.18: Frequency responses of filters on “feed” data.	87
Figure 3.19: “Feed” data re-transferred (<i>ifft</i>) to time domain after filtering.	87
Figure 3.20: Time domain data comparison, clean vs. biased (+10%).	88
Figure 3.21: Frequency domain data comparison, clean vs. biased (+10%).	89
Figure 3.22: Time domain data comparison, clean vs. biased (-10%).	90
Figure 3.23: Frequency domain data comparison, clean vs. biased (-10%).	90
Figure 3.24: Heat sensor-H1 at Pogo stripping circuit.....	92
Figure 4.1: Pogo Mine site map.	102
Figure 4.2: Pogo Mine-gold processing flow sheet.....	103
Figure 4.3: CIP circuit at Pogo Mine.....	104
Figure 4.4: Pogo stripping circuit in pictures.....	106
Figure 4.5: Pogo stripping circuit schematic diagram with sensor placements.	107

Figure 4.6: Strip vessel-1 heat sensor (S1).....	109
Figure 4.7: Strip vessel-2 heat sensor (S2).....	109
Figure 4.8: Heat exchanger-1 sensor (H1).	110
Figure 4.9: Heat exchanger-2 sensor (H2).	110
Figure 4.10: Barren flow sensor (BARNFL).....	111
Figure 4.11: Glycol flow sensor (GLYFL).....	111
Figure 4.12: Cycles (peaks) vs thresholds in a clean set of S1 sensor data.....	114
Figure 4.13: Flow chart for the peak-readings count and sensitivity analysis (PRCSA). ..	116
Figure 4.14: Bias effect on S1-sensor data set at various thresholds.	119
Figure 4.15: Bias effect on S1-sensor data at 260°F threshold (Th_{SI}).....	121
Figure 4.16: Bias effect on S1-sensor data at 270°F threshold (Th_{SI}).....	122
Figure 4.17: Bias effect on S1-sensor data at 280°F threshold (Th_{SI}).....	123
Figure 4.18: Cumulative peak-readings count (biased).....	125
Figure 4.19: Cumulative peak-readings count (clean).	125
Figure 4.20: Cumulative no. of peaks count (biased).....	126
Figure 4.21: Cumulative no. of peaks count (clean).	126
Figure 4.22: Cumulative temperature (biased).....	127
Figure 4.23: Cumulative temperature (clean).....	127
Figure 4.24: Cumulative peak-readings count, clean vs biased ($Th_{SI}=280^{\circ}\text{F}$).....	128
Figure 4.25: Cumulative no. of peaks count, clean vs biased ($Th_{SI}=280^{\circ}\text{F}$).....	129

Figure 4.26: Cumulative Temperature, Clean vs Biased ($Th_{SI}=280^{\circ}\text{F}$).....	130
Figure 5.1: Pogo Mine site map.	141
Figure 5.2: Pogo Mine-gold processing flow sheet.....	143
Figure 5.3: CIP circuit at Pogo Mine.....	144
Figure 5.4: Pogo stripping circuit in pictures.....	146
Figure 5.5: Pogo stripping circuit schematic diagram with sensor placements.	147
Figure 5.6: Strip vessel sensors (S1 and S2).	149
Figure 5.7: Heat Exchanger sensors (H1, H2, H3, and H4).....	150
Figure 5.8: Barren flow sensor (BARNFL).....	151
Figure 5.9: Glycol flow sensor (GLYFL).....	151
Figure 5.10: Behavior of various sensors in 24 hours time span.....	152
Figure 5.11: Peaks vs thresholds in a clean set of S1 sensor data.	157
Figure 5.12: Bias effect on S1-sensor data set at various thresholds.	158
Figure 5.13: Bias effect on S1-sensor data at 280°F threshold (Th_{SI}).....	160
Figure 5.14: Cumulative peak-readings count, clean vs biased ($Th_{SI}=280^{\circ}\text{F}$).....	161
Figure 5.15: Flow chart for the peak-readings count and sensitivity analysis (PRCSA). ..	162
Figure 5.16: Various sensors and their interrelations.....	164
Figure 5.17: Ratio functions and matching.	167
Figure 5.18: Flow chart for the multiple ratio function analysis (MRFA) algorithm.	171
Figure 5.19: Ratio analysis plot for S1 and S2 peaks; <i>trunc</i> $Th_{SI}=280^{\circ}\text{F}$, bias=+2%.	176

Figure 5.20: Comparison of positive (+2%) and negative (-2%) bias effects on S1 sensor clean data set.	178
Figure 5.21: Cross-score algorithm.	180
Figure 5.22: Cross-score calculations to identify bias (clean vs biased data).	182
Figure 5.23: Classification of multiple tests.	184
Figure 5.24: True alarm vs false alarm.	186
Figure 5.25: Multiple tests in finding bias (+2%) for S1 sensor data.	189
Figure 5.26: Disadvantage with truncation threshold (<i>trunc Th</i>).	191
Figure 5.27: Effect of lowering the threshold to 260°F on peaks captured.	193
Figure 5.28: Truncation threshold vs dynamic threshold.	194
Figure 5.29: Flowchart for MRFA automation with dynamic thresholding (MRFAA).	196
Figure 5.30: Fortran interface where multiple input options are available.	197
Figure 5.31: Combined test - Alarm and TTFD performances.	200
Figure 5.32: Combined test performance.	203
Figure 5.33: Heat ratio test performance.	210
Figure 5.34: Max value test performance.	212
Figure 5.35: Ave value test performance.	214
Figure 5.36: BARNFL ratio test performance.	216
Figure 5.37: GLYFL ratio test performance.	218

List of Tables

Page

Table 1.1: Operating schedule-pressure Zadra stripping.....	15
Table 2.1: SAG mill sensors (RPM, BP, feed, HP) and data.....	39
Table 2.2: Correlation matrix (raw data).....	43
Table 2.3: State-1 baseline sensor interrelations (BSI), records: 1-6000.....	51
Table 2.4: State-2 baseline sensor interrelations (BSI), records: 6001-14000.....	51
Table 2.5: State-3 baseline sensor interrelations (BSI), records: 14001-20000.....	51
Table 2.6: Analysis of various parameters and lags at 240 min Aggregation.....	52
Table 2.7: Feed data without noise.....	52
Table 2.8: Feed data with 10% random noise.....	53
Table 3.1: Operating schedule-pressure Zadra stripping.....	71
Table 3.2: A snapshot of raw sensor data collected at 10-min average intervals.....	75
Table 3.3: Descriptive statistics for the sensor data collected.....	79
Table 4.1: Operating schedule-pressure Zadra stripping.....	105
Table 4.2: A snapshot of raw sensor data collected at 10-min average intervals.....	108
Table 4.3: Descriptive statistics for the sensor data collected.....	112
Table 4.4: Cumulative peak-readings count (clean).....	125
Table 4.5: Cumulative peak-readings count (biased).....	125
Table 4.6: Cumulative no. of peaks count (clean).....	126
Table 4.7: Cumulative no. of peaks count (biased).....	126

Table 4.8: Cumulative temperature (clean).....	127
Table 4.9: Cumulative temperature (biased).....	127
Table 4.10: Cumulative peak-readings count, clean vs biased ($Th_{S1}=280^{\circ}\text{F}$).....	128
Table 4.11: Cumulative no. of peaks count, clean vs biased ($Th_{S1}=280^{\circ}\text{F}$).....	129
Table 4.12: Cumulative Temperature, Clean vs Biased ($Th_{S1}=280^{\circ}\text{F}$).....	130
Table 5.1: Operating schedule-pressure Zadra stripping.....	145
Table 5.2: A snapshot of raw sensor data collected at 10-min average intervals.....	148
Table 5.3: Descriptive statistics for the sensor data collected.....	153
Table 5.4: Cumulative peak-readings count, clean vs biased ($Th_{S1}=280^{\circ}\text{F}$).....	161
Table 5.5: A snapshot of peaks captured for sensor S1; <i>trunc</i> $Th_{S1}=280^{\circ}\text{F}$, bias=+2%.....	173
Table 5.6: A snapshot of peaks captured for sensor S2 (clean); <i>trunc</i> $Th_{S1}=280^{\circ}\text{F}$, no bias.....	174
Table 5.7: A snapshot of S1 and S2 peaks Matching; <i>trunc</i> $Th_{S1}=280^{\circ}\text{F}$, bias=+2%.....	175
Table 5.8: S1 Bias detection, Heat ratio test ($\pm 2\%$ bias induced in S1, while S2 is clean). 187	187
Table 5.9: Multiples tests with bias identification dates.....	188
Table 5.10: Multiples tests with dynamic thresholding (90% of peak-max value).....	198
Table 5.11: Combined test - Performances at Thresholds 5 or 6.....	200
Table 5.12: Combined test - Performances at Thresholds 7 or 8.....	200
Table 5.13: Combined test True alarm performance.....	202
Table 5.14: Combined test False alarm performance.....	202

Table 5.15: Heat ratio test True alarm performance.....	211
Table 5.16: Heat ratio test False alarm performance.....	211
Table 5.17: Max value test True alarm performance.	213
Table 5.18: Max value test False alarm performance.....	213
Table 5.19: Ave value test True alarm performance.	215
Table 5.20: Ave value test False alarm performance.	215
Table 5.21: BARNFL ratio test True alarm performance.	217
Table 5.22: BARNFL ratio test False alarm performance.	217
Table 5.23: GLYFL ratio test True alarm performance.	219
Table 5.24: GLYFL ratio test False alarm performance.	219

Acknowledgements

I would like to express my sincerest appreciation and deepest gratitude towards my academic advisor, supervisor, and the committee chair, Professor Rajive Ganguli, for his support and guidance. Without his continuous encouragement, enthusiasm, optimism, and Socratic questioning, this work would hardly have been accomplished. Likewise, I owe my deepest gratitude to the members of the committee (past and present), Dr. Tathagata Ghosh, Dr. Orion Lawlor, Dr. Ronald Barry, and Dr. Seta Bogosyan, for their invaluable advice and critique throughout the course of my research.

I am deeply indebted to the Fort Knox Mine of Kinross Corporation and the Pogo Mine of Sumitomo Metal Mining Company (SMM) for their support in providing the necessary data for the project and for facilitating necessary site visits. I am specifically thankful to several professionals at Pogo Mine in this regard: Chris Kennedy, General Manager, Jason Pyecha, Chief Metallurgist, and Andrew Maxon. I also thank the two companies for supporting the Mining Engineering Research Endowment (MERE) at the College of Engineering and Mines (CEM), University of Alaska Fairbanks (UAF), without which the research undertaking would not have been possible. I would like to express my gratitude to Dr. Margaret Darrow, Department Chair, Mining and Geological Engineering (MinGeo), the deans and the staff at CEM and the Graduate School, for allowing me to utilize various academic, laboratory, and administrative resources to accomplish the project. My special thanks are due to Julene Lowdermilk, Jill Riddle, Paul Brown, and Frances Bedel, to name a few, among many other staff members. I would like to thank Carol Holz and Reija Shnoro from the UAF International Programs and Initiatives, for their continuous support and help.

I am thankful to my family members, my wife Lakshmi Prasuna, and my wonderful cat Chito, for their unconditional love, support, tolerance, and understanding throughout my stay away from home, for my academic pursuits at UAF. I am also thankful to the members of my racquetball group, Rory O'Neil, Ajay Nautiyal, and Jarkko Toivanen, for their wonderful company; playing the sport helped me maintain my health and sanity throughout my stay in Fairbanks. Last but not least, I am grateful to all my colleagues and friends for their wonderful companionship, which made my stay at UAF a memorable experience!

Chapter 1: Introduction

1.1 Background

The use of sensors to monitor industrial processes has become increasingly prevalent over the past few decades, and the mining industry is no exception. Sensors play a vital role in monitoring various process parameters, such as temperature, pressure, and weight in almost every industry in United States. The 21st century has witnessed the widespread use of sensors in all sectors of the mining industry, from drilling operations to mineral processing and recovery operations. The mining industry was quick to adapt sensor usage for its economic benefit. A recent study suggests that sensor usage in various stages of mining (for a moderately sized mine) can yield up to \$10-100 million per annum in added economic value (Buxton and Benndorf, 2013). Unfortunately, the opposite effect is true when sensors suffer from faults and provide erroneous data. According to one author, sensor faults are causing approximately 3-8% of production loss to US oil industry, ultimately leading to \$20 billion in annual losses to the US economy (Wang et al., 2009). Plant instrumentation, which includes sensors, is not inexpensive; to replace instrumentation frequently is not a viable option. It is estimated that instrumentation costs alone comprise 2-8% of the total fixed costs associated with any process plant (Narasimhan and Jordache, 2000). Moreover, frequent maintenance and sensor calibrations result in process disruptions leading to production losses totaling in the millions. In some cases, these errors can cause unsafe conditions where maintaining the industrial environment at certain limits is essential; an example is carbon monoxide (CO) monitoring systems in underground mines.

Identifying the sensor faults and fixing them through periodically scheduled (or optimally scheduled) calibration processes can dramatically improve the sensors' accuracy. Such processes can reduce equipment downtimes, increase production, and improve overall safety in the industry.

While the common types of faults or “gross errors,” such as noise, failures (flat-outs), stuck-at-faults, can easily be detected and fixed through preventive or scheduled maintenance, “calibration” related errors are difficult to impossible to identify. Unlike gross faults—which are obvious due to their high magnitudes—these kinds of faults/errors are indistinguishable if the process is particularly non-linear or non-stationary. These insidious errors can, over time, cause significant losses to industrial operators. Calibration errors are often present in the data mimicking the form of “an added offset value (bias),” to the original (true) reading.

1.1.1 Organization of the Dissertation

The following is a brief description of the organization of this dissertation. The titles of the chapters given here are concise versions of the actual ones. Each of these chapters are in the form of manuscript format outlined in the University of Alaska Fairbanks Graduate School’s “Thesis Formatting Handbook” (UAF, 2017), however, for the preparation of individual chapters, the style guide of Mining Engineering Journal, an international peer reviewed publication of Society for Mining, Metallurgy and Exploration, Inc. was followed (SME, 2017). Each chapter is prepared as a standalone paper that is publishable in the near future; with some editing to meet the specific requirements of the journal. Due to the reason, the reader should notice that certain sections or text were repeated among the chapters.

Chapter 1: Introduction

A general introduction to all the major topics presented in Chapters 2 through 5.

Chapter 2: Statistical Methods in Error Detection

Review of various statistical methods and their application in finding sensor errors with a case study on bias detection in Fort Knox SAG mill circuit sensors.

Chapter 3: Signal Processing Methods in Error Detection

Review of signal processing methods and their application in finding calibration errors with a case study focused on bias detection using Fast Fourier Transform (FFT) in Pogo stripping circuit sensors.

Chapter 4: Peak-Readings Count and Sensitivity Analysis (PRCSA) in the Detection of Calibration Errors

Review of various innovative methods developed and their application in finding calibration errors with a case study focused on bias detection with PRCSA in Pogo stripping circuit sensors.

Chapter 5: Multiple Ratio Function Analysis with Automation (MRFAA) in the Detection of Calibration Errors

Review of various innovative methods developed and their application in finding calibration errors with a case study focused on bias detection with multiple ratio function analysis (MRFA), and MRFA with automation (MRFAA); study conducted on Pogo stripping circuit sensors.

1.1.2 Objective of the Work

This research is intended to tackle the problem of “sensor calibration errors or faults,” specifically “bias,” in an industrial setting. The ultimate objective is to find practical solutions through innovative methods to identify such errors based on exploitation of sensor interrelations in a multi-sensor environment—without causing disruption to the industrial process itself.

1.1.3 Problem Statement

Sensors play a crucial role in monitoring and controlling various operations in mineral processing circuits. Accurate measurements of various physical and chemical parameters at each stage of processing are essential to achieve “overall system optimization.” Sensor errors of high magnitude like noise, short faults, and failures, are called “gross errors,” while those of lower magnitude that gradually develop are called calibration errors. The presence of these errors can cause disruption to the optimum state of the plant operation. This results in poor metal recoveries and high instrumentation costs. Gross errors can be mitigated through “calibration” process, however, errors caused by gradual calibration loss overtime are difficult to identify and require great deal of attention. The model building process for error identification in multi-sensor environments is challenging because multidimensionality of the inter-sensor relationships adds more complexity to the process. The goal of this research is to find innovative methods to recognize calibration faults, specifically bias, in a multi-sensor environment. Mineral processing is a nonlinear and non-stationary process, which adds to the complexity. Therefore, use of classical statistical methods might be disadvantageous. Exploitation of the sensor interrelations is the better choice in this context. For instance, there can be a particular “ratio” that can exist between sensor readings, and it can exhibit patterns. By capturing these “ratio statistics” for sensors in an error-free state, it may be possible to identify sensors with bias. Developing innovative methods and algorithms to identify bias in a non-invasive way is highly desired by the industries.

1.1.4 Scope of the Research

The focus of the research is limited to the detection of “calibration errors.” These errors are relatively difficult to isolate in comparison to other types of errors. The research depends chiefly on inter-sensor relations and statistics. The scope and perspective does not include the examination

of sensor hardware problems. While sensors in mineral processing circuits are examined, other industrial circuits are beyond the scope, however, the findings are generalizable or scalable to any other industrial process due to the similarities of industrial sensor networks. Few sensors in a moderately sized mineral processing circuit (carbon stripping circuit) are exploited for this research. While the findings can always be scaled up for any industrial circuits in the future, larger circuits (with hundreds of sensors) are beyond the scope of the work due to the higher dimensional complexity. Gross errors, which have been explored thoroughly by numerous researchers, are not included in this research.

1.1.5 Challenges

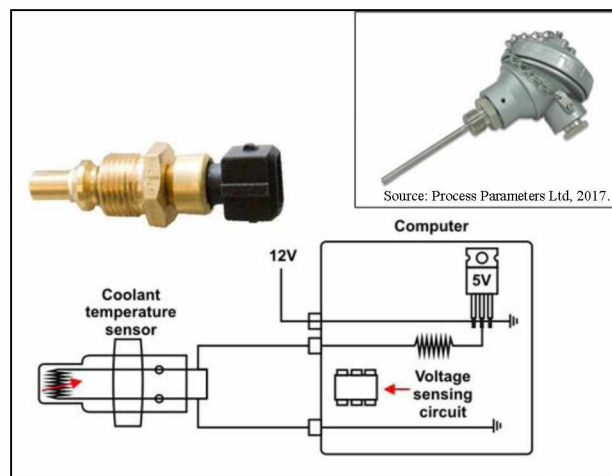
Sensor interrelations and statistics are multi-dimensional. Exploiting or capturing inter-sensor data statistics in huge circuits (with hundreds of sensors) is an immense challenge. For the purpose of the research, a circuit with eight sensors was examined. The ability to locate errors in real time, in a time span of hours or days, is beneficial for the mining industry, however, this is a huge challenge, due to the insidious and subtle nature of calibration errors. Moreover, these subtle changes are merely process fluctuations. To develop statistics sensitive enough to observe and quantify such minute data changes ($\pm 2\%$ of true reading in the case of calibration errors) that are otherwise caused by industrial processes is a difficult task. For instance, in mineral processing circuits, process parameters change frequently to cope with changing ore types and other input parameters. Mill operators (or automated control systems) must make sudden changes to process variables to cope with the incoming ore types and to achieve maximum recoveries. The data analysis algorithms developed for the research should be robust to these “operator-induced” changes and still be able to detect “calibration errors.”

1.2 Literature Review

The literature review presented in this section can be viewed as an overview for the rest of the chapters in the dissertation. An extensive and topic specific literature review can be found in each individual chapter.

1.2.1 Functionality of Sensors

According to the Institute of Electrical and Electronics Engineers (IEEE), a sensor is an electronic device that produces electrical, optical, or digital data (a signal) from a physical condition or event like pressure, temperature, flow, etc. (Institute of Electrical and Electronics Engineers, 2017). For example, a temperature sensor observes temperature and produces output—with the help of another device—in the form of a reading ($^{\circ}\text{F}$). A typical car coolant temperature sensor and the thermocouple mechanism is shown in Figure 1.1, along with an industrial grade thermocouple sensor (inset) that can measure temperatures up to 2282°F (1250°C). A level sensor from Pogo flotation circuit is shown in Figure 1.2.



Source: Adapted from Kalwinder, 2017; Process Parameters Ltd, 2017.

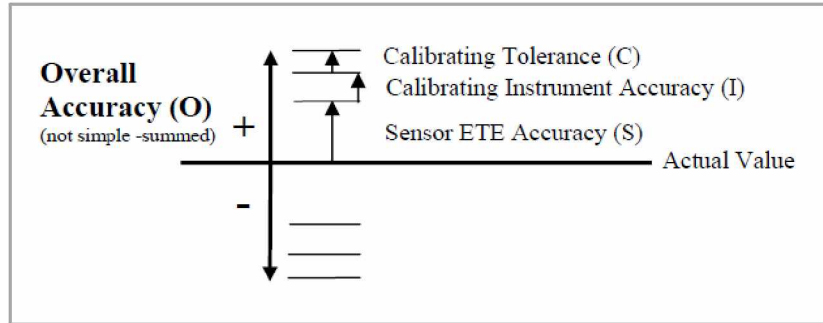
Figure 1.1: A car coolant temperature sensor mechanism, and an industrial grade thermocouple sensor (inset).



Figure 1.2: A level sensor from Pogo flotation circuit.

1.2.2 Accuracy of a Sensor and Role of Calibration

Data quality plays an important role in the decision-making process for industries. The data quality must be known and well established beyond a reasonable doubt before it can be utilized. The closeness of a measured value to the true value is called accuracy. Accuracy is the most desirable requirement for any measurements (National Institute of Instrumentation Standards, 2017a). In metrology, calibration is the process of comparing measurements to that of calibration standards established by concerned national bureaus; in the United States, this bureau is the National Institute of Instrumentation Standards (NIST). Calibration plays a vital role in maintaining the accuracy of a sensor. Overall accuracy of a sensor is the sum of the manufacturer's end-to-end (ETE) accuracy, calibrating instrument accuracy and calibrating tolerance, added to the actual reading (Figure 1.3).



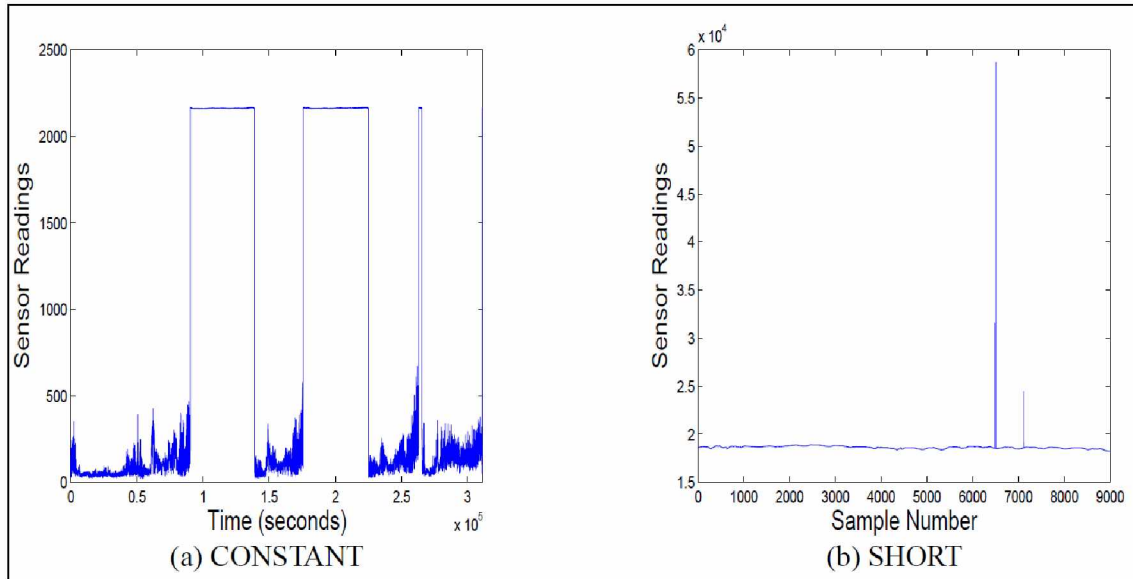
Source: Adapted from Stum, 2006.

Figure 1.3: Overall accuracy of a sensor.

It is important to realize that manufacturer's ETE accuracy of a sensor is limited to the sensor itself and is not for the transducer or other supporting gadgets involved (Stum, 2006). This may change the overall accuracy of the sensor. For the above reasons, sensors in the industrial circuits are periodically checked and calibrated, however, errors that develop between the scheduled periods of calibration are of concern because they can result in significant production losses.

1.2.3 Classification of Sensor Faults

Fault categorization or classification helps to escalate the sensor fault detection process. They can be classified into two major categories: discontinuous, such as malfunctions and random faults, and continuous, such as biases and drifts. Bias can be a positive or negative offset to the sensor's true reading (Baljak et al., 2012). Sensor faults are typically caused by hardware problems like damage, short-circuits, low battery, or are otherwise calibration related (Sharma et al., 2010). The faults also can result from the software errors. Depending on their type, sensor faults can also be classified as noise-related, short faults, constant or stuck-at faults, and calibration-related (Fang and Dobson, 2011). A constant and a short fault are shown in Figure 1.4.



Source: Adapted from Sharma et al., 2010.

Figure 1.4: A constant and short faults.

1.2.4 Sensor Faults and Impact on Operating Costs

This section reviews the impact of sensors on operating costs in terms of economic losses. Several researchers have attempted to quantify the impact of sensor faults in economic terms. For instance, Wang et al. (2009) discussed the direct and indirect losses that can be caused by sensor faults. The authors describe the heavy losses sustained at a Beijing based chemical plant in 1997 and attribute the losses incurred to sensor faults. An ethane device exploded at the plant and the losses totaled one billion Yuans (Quan-Bo and Cheng-Lin, 2006). While discussing the importance of data reconciliation and gross error detection methods, Narasimhan and Jordache (2000), describe the impact of sensor faults on overall instrumentation costs of a mineral processing plant. Instrumentation costs comprise of 2-8% of the total fixed costs of any process plant (Narasimhan and Jordache, 2000). Narasimhan and Rengaswamy (2017) attempted to quantify the annual value gained by identifying and fixing the sensor faults in a continuous stirred-tank reactor (CSTR) network. They found that “gross errors” can lead to the loss of resolution property for the

corresponding sensor network. The value of detecting these faults can be quantified through the overall loss that one will incur due to the loss in resolution property (Narasimhan and Rengaswamy, 2017). Conversely, another author noted that sensor “biases” in non-control variables are related to the loss incurred through loss of precision (Bagajewicz et al., 2004). This was a study aimed at quantifying the economic value of using a data reconciliation package to correct the faulty sensor readings. Studying a crude distillation unit, the authors found that the existing faulty sensors were causing \$7.36 million in losses. It was also found that, by using data reconciliation methods, the loss could be reduced to \$7.12 million.

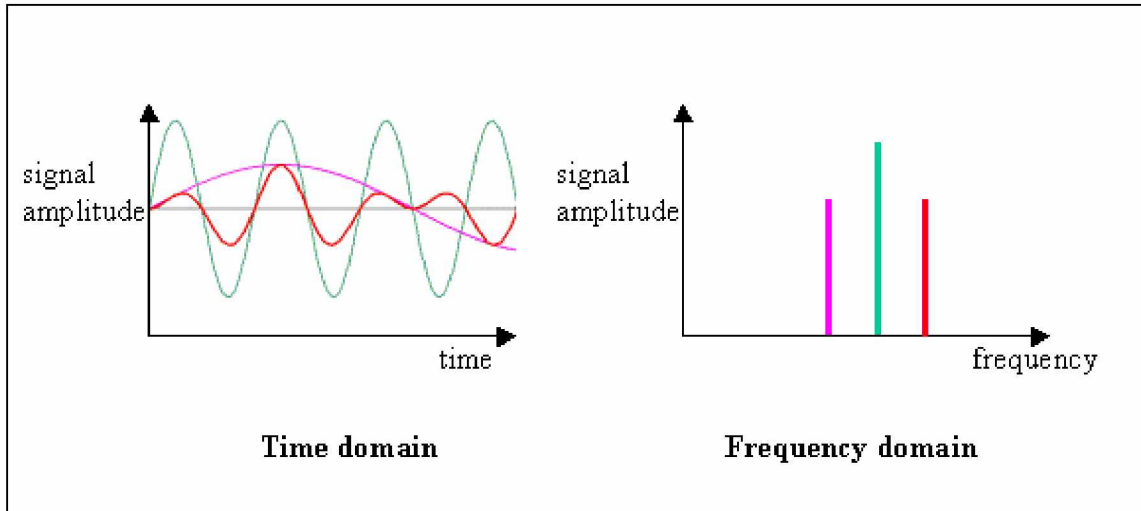
1.2.5 Fault Detection: Statistical Methods

Various statistical methods used in sensor fault detection are widely described by Sharma et al. (2010). Some of the methods briefly described are as follows. “Rule-based” methods use domain knowledge to develop heuristic rules for identifying faults. These are generally highly accurate methods, but the choice of heuristics can affect the results. “Estimation based” methods depend upon inter-sensor correlations, whereas linear “Least-squares estimation” (LLSE) techniques are used to flag faulty readings in sensor networks; however, these techniques are disadvantageous for classifying fault types. “Time series analysis” based methods are capable of finding short duration faults but prove disadvantageous to find long duration faults like noise. For sensor readings that do not exhibit periodicity, the auto regressive moving average (ARIMA) methods proves to be a better option. “Learning based” methods infer the normal sensor readings from the training data set to identify classes of faults in the test or prediction data set. Neural networks are an example of learning based methods; however, with neural nets training is time consuming. Contrary to the mainstream statistical models or classical, descriptive and inferential statistics, “Bayesian statistics” use prior knowledge. “Data Reconciliation” methods involve

adjusting the data according to conservation laws and other constraints. For data sets that follow particular statistical distribution, parametric methods are used; example: linear regression. For data sets that does not follow any distribution, non-parametric or distribution-free statistics are used; Examples: bagging and boosting. For non-linear and non-stationary processes data-mining techniques are very popular. These are generally used for extracting knowledge from huge data sets. “Heuristics” involve the theory and practical application of techniques for solving problems approximately that cannot be solved exactly (Journal of Heuristics, 2017). Hybrid methods employ any combination of the above techniques to achieve the task. Calibration errors are one exception that are hard to detect with the above methods. Without the availability of ground truth values, calibration errors become difficult to identify and even rectify. Compared to classical/parametric methods, data-mining based innovative methods seem to be the better choice to find calibration errors like bias in large industrial sensor data sets.

1.2.6 Fault Detection: Signal Processing Methods

A “signal” in communication systems is something that conveys information about the behavior or attributes of some phenomenon. Some examples are sound, video, and picture. An analog signal is a continuous signal, whereas digital signal is constructed from the discrete set of waveforms to represent the signal. An analog can be converted to digital using a converter (ADC). A signal is generally associated with noise. When the noise is filtered using digital “filters,” it is easier to study the signal. A signal can be studied by plotting its behavior against time (time domain), or against frequency, called frequency domain (Figure 1.5). Fast Fourier Transform (FFT) is a numerical algorithm that is used to convert a signal in time domain to frequency domain, or vice versa.



Source: Adapted from National Institute of Instrumentation Standards, 2017b.

Figure 1.5: Time domain vs frequency domain.

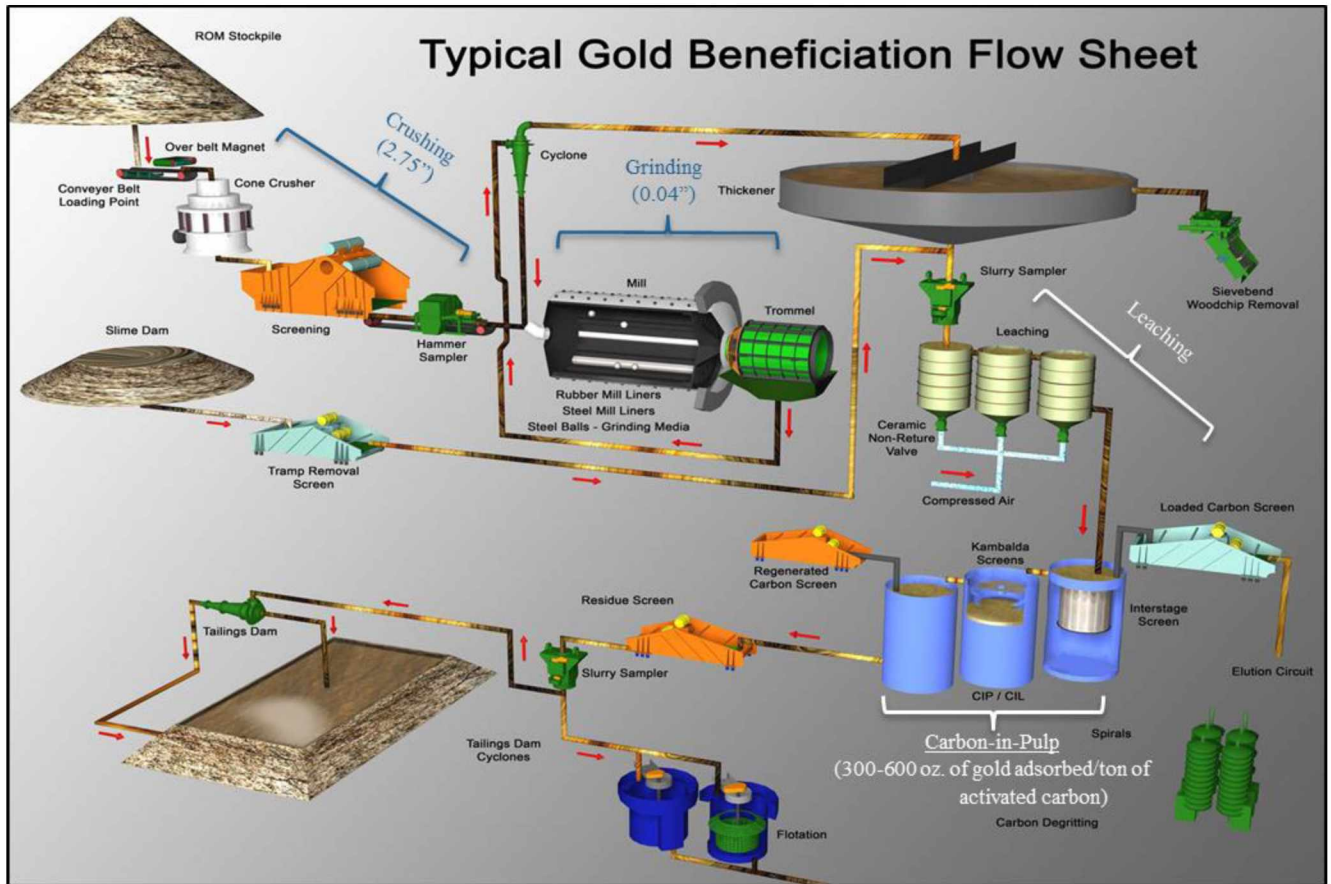
Digital signal processing (DSP) has many applications in different industries; space, medical, and military are some of them. Digital filtering, spectral analysis, speech processing, image processing, and radar processing are some of the specific applications. Industrial “process monitoring and control” is one of the areas of application too. Sensors monitor many physical parameters in industrial processing circuits. Examples include pressure sensors, temperature sensors, pH level sensors, movement or level indicating sensors, etc. Data produced from these sensors can be considered as a signal. By processing the signal one can observe the key tendencies. In the presence of bias or error, a signal’s behavior (if biased) changes when compared to the signal with no bias (clean). The differences can be observed in the form of changes in the frequency and amplitude. A literature review related to industry specific applications is presented in Chapter 3.

The most essential requirement for a signal to be analyzed successfully with FFT approach is that the signal should exhibit periodicity. This could be a serious drawback when analyzing signals related to industrial processes, which tend to be dynamic rather than periodic.

1.3 Introduction to Gold Mining, Fort Knox and Pogo

The two choices of mineral processing circuits for analysis of sensor data in this dissertation are from the gold mines located near Fairbanks, Alaska, i.e., a semi-autogenous (SAG) mill circuit from Fort Knox Mine (open-pit) and a carbon stripping circuit from Pogo Mine (underground). A typical gold mining operation is detailed in Figure 1.6. As depicted in the figure, a SAG mill is part of the grinding circuit and the stripping circuit is part of carbon-in-pulp process. Detailed descriptions of Fort Knox Mine and Pogo Mine are provided in the individual chapters.

Gold processing involves extraction of gold ore from an underground or open pit mine. The ore is broken into smaller pieces by a big cone or gyratory crusher, depending on the hardness and other characteristics of the rock. With crushing, the ore fragments are reduced to smaller pieces of approximately 2.75" (70 mm). These sizes are further reduced by a SAG mill or ball mill to as low as 0.04" (1 mm). The fine particles of ore are then sent to a "flotation" process to separate the gold bearing particles. The particles enter a batch of cyanide leaching tanks where the gold particles are separated out of the ore particles by the leaching solution. The gold particles are adsorbed by activated carbon (a type of charcoal) in a carbon-in-pulp (CIP) circuit which is then loaded into the "strip vessels" where a high heated (~280°F) solution called "elute" is circulated. An elute is a water-based solution with 1% sodium hydroxide and 0.1% sodium cyanide (Fast, 2016). The elute helps strip gold particles from carbon. The solution that is bearing gold particles is now called "pregnant leach solution" (PLS). This is sent to the electro-winning process to extract gold. The gold particles collected from the process are melted at high temperatures ultimately to pour into gold ingots.



Source: Adapted from Mulotec, 2017.

Figure 1.6: Flow sheet for a typical gold processing operation.

1.3.1 Carbon Stripping Circuit

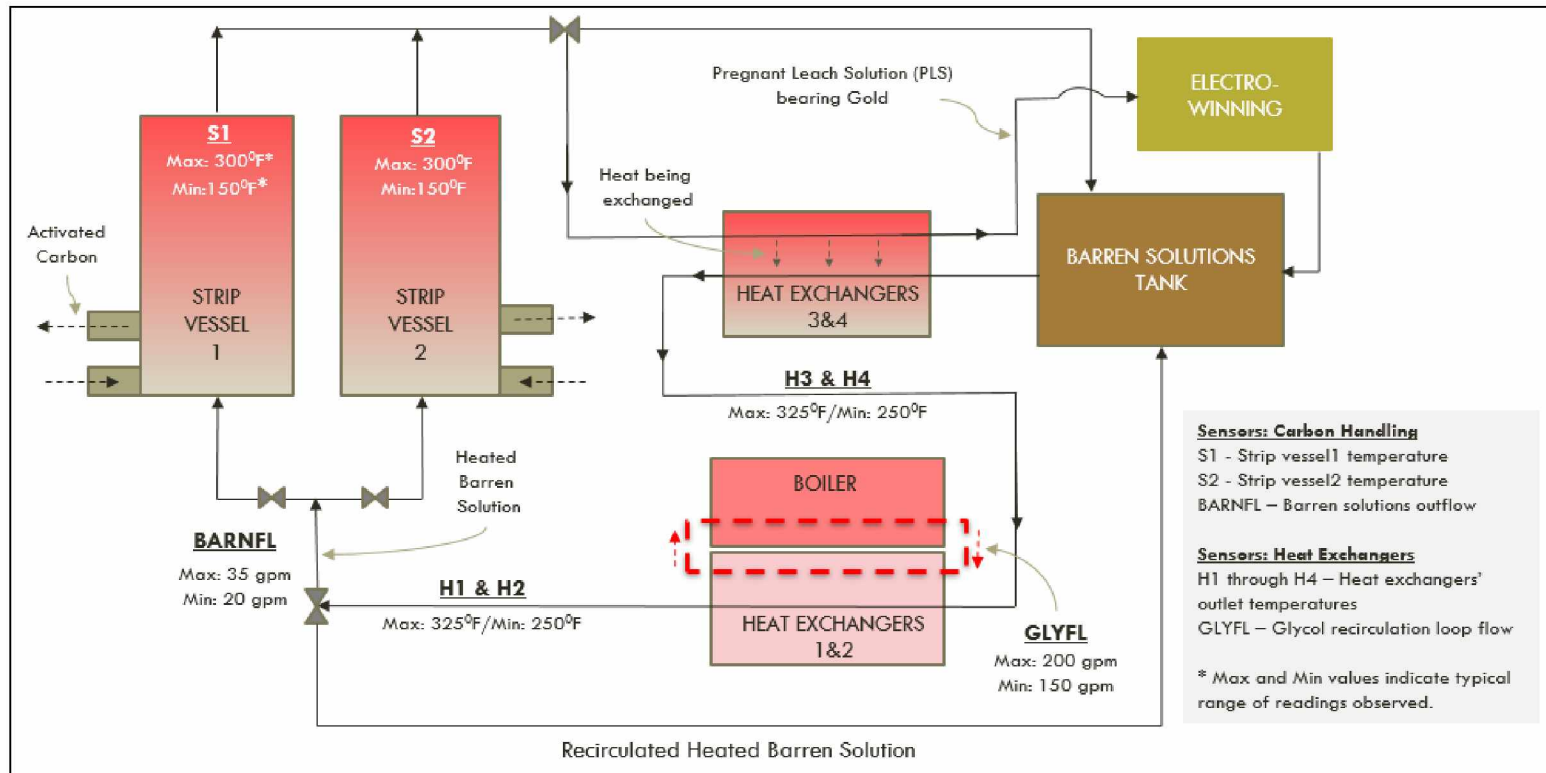
The Pogo Mine stripping circuit consists of two strip vessels that work in tandem. While one vessel (vessel-1) is being loaded with gold-bearing activated carbon, the other previously loaded (vessel-2) is operated by circulating an elute solution at approximately 280°F and 65 PSIG to liberate gold particles. The process is called “pressurized Zadra stripping.” A typical pressurized Zadra stripping cycle lasts for 11 hours and consists of the following stages: loading the vessel (1 hr), circulating elution (8 hrs), carbon cooling (1 hr), and unloading carbon from vessel ($\frac{1}{2}$ hr) (Table 1.1).

Table 1.1: Operating schedule-pressure Zadra stripping.

Operation	Solution	Time
Load Column	Transfer Water	90 minutes
Elution	0.1% NaCN, 1% NaOH	480 minutes
Carbon Cooling	Fresh Water	60 minutes
Unload Column	Transfer Water	30 minutes
TOTAL		11 hours

Source: Fast, 2016.

While the used carbon is discharged, the PLS is pumped out. The same process is repeated with strip vessel-2. The PLS, on its way out from the strip vessel, is cooled through exchanging heat with no. 3 and 4 heat exchangers. When the PLS reaches the electro-winning circuit, the gold particles are removed and the solution, now called “barren solution,” is reheated by a boiler with the aid of heat exchangers 1 and 2 and will be recirculated through strip vessels. A glycol solution is circulated between the boiler and heat exchangers as a medium of heat exchange. Sensors are strategically placed in various parts of the circuit to measure temperatures, flow rates, etc. (Figure 1.7). If any of the sensors are biased, the circuit cannot be managed at an optimum level in terms of temperature and flows. This results in poor gold recoveries. Finding and fixing faulty sensors helps improve recoveries, which is the motivation for the research. It is very important to maintain temperatures in the strip vessels at certain levels (270-280°F) for certain periods of time to maximize gold separation. In this context, monitoring the temperatures becomes crucial. Sensors S1 and S2 are two important ones in this context, which will be analyzed in this chapter. Various sensors and their placements in the circuit are shown in Figure 1.7. The sensors are given code names for simplicity. Typical interrelations between these sensors are depicted in Figure 1.8.



Source: Pogo Mine, 2016.

Figure 1.7: Pogo stripping circuit schematic diagram with sensor placements.

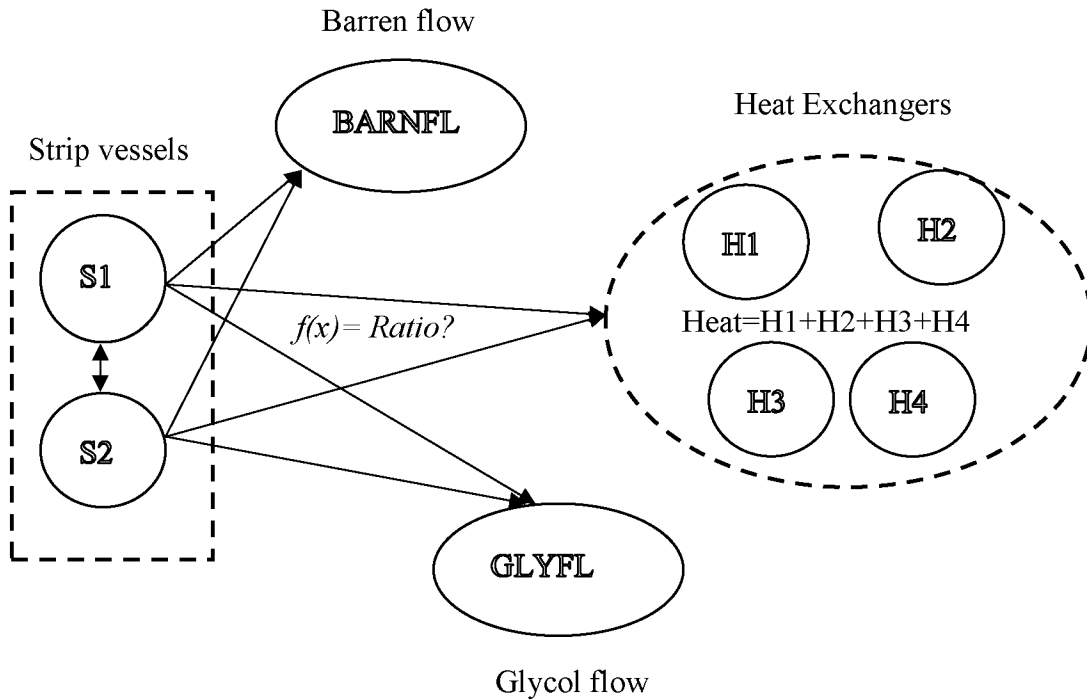


Figure 1.8: Various sensors and their interrelations.

1.4 Methods and Materials

Choosing appropriate methodology is necessary to achieve the research goal. Detailed topic or problem specific methodology is available in individual chapters. Chapters 2 through 5 are arranged to reflect the maturity of ideas as the research progressed towards achieving the goal. Detailed descriptions on the data collection, and the visual overview of various sensor data streams are available in various plots in individual chapters. Detailed results sections are available in each individual chapter.

1.5 Overview of Work Done and Conclusions

Calibration errors or biases of small quantities ($\pm 2\%$ of a sensors true reading) are hard to detect. The initial application of classical statistical methods—correlation, aggregation, PCA, etc.—in the detection of these errors, proved disadvantageous (Chapter 2). Parametric and methods

based on linear relations are less effective in dynamic environments where the process parameters or variables change dramatically (non-linear and non-stationary). For similar reasons, the sophisticated digital signal processing methods like Fast Fourier Transform (FFT), which was applied on Fort Knox sensor SAG mill data, was not effective. FFT methods heavily depend on periodicity of a sensor signal and are not effective for less periodic or dynamic processes (Chapter 3). For the above applications a 10% bias was used, which is high in magnitude. Finding bias at 2% magnitude is preferred by industry operators like Pogo Mine. Moreover, SAG mill operation is too dynamic for initial model building stages. Due to these reasons, the comparatively less dynamic carbon stripping circuit located at Pogo Mine is chosen for the study.

Stripping operation is crucial in the gold extraction process. The temperatures in the strip vessels are maintained at ideal levels (270-280°F) to maximize the extraction of gold. Hence, strip vessel temperatures are of utmost interest for the Pogo Mine. Throughout the research, strip vessel sensor S1 was chosen for bias induction experiments. Several data-mining based approaches were used on pogo strip circuit sensor data. Peak-reading count and sensitivity analysis (PRCSA) was one of them (Chapter 4). In this approach, a +2% bias was introduced in one of the sensors' data (S1). When compared to the clean data set of the sensor (S1), the number of times the biased readings crossed a preset temperature limit, i.e., Threshold (Th_{SI}), was significantly high, signaling error presence. The method requires availability of a clean set of data for the same biased sensor under study, which is highly impractical in real life situations. This is a major disadvantage using PRCSA.

The algorithms developed in Chapter 5 are based on exploiting interrelations between sensors' readings in the carbon stripping circuit. The multiple ratio function analysis (MRFA) method exploited the interrelations between strip vessel and heat exchanger sensors in the form of "ratio

functions”—these are not simple ratios (see chapter 5). When S1 is induced with +2% error, S2 and other sensors are assumed error-free. It was observed that prior to the error introduction, the ratios, “S1 bias: Heat” and “S2 clean: Heat” kept crossing each other, i.e., no ratio remained consistently higher than the other. The test is named “Heat ratio test.” However, several tests performed with the algorithm revealed that the “truncation thresholding” criteria of MRFA is disadvantageous in finding bias at certain periods of the year, specifically the last quarter of the year. A “dynamic thresholding” strategy that depends and adjusts the thresholds based on maximum values of a peak-readings is added to the algorithm. The new algorithm is fully automated in Fortran language with improved speed of execution. The algorithm is called MRFA with automation (MRFAA).

MRFAA has multiple test capabilities. Like the Heat ratio test, the ratios of S1 and S2 with barren flow (“BARNFL test”) and glycol flow (“GLYFL test”) were exploited. The maximum and average reading values of a strip vessel cycle (or peak) are also better indicators of the temperature. Due to this reason, the average and maximum values of strip vessel sensors were compared in the form of “Ave test” and “Max test.” A “Combined test” that finds the performance of all tests together is also added to MRFAA. A cross-score algorithm helps MRFAA in finding relative positions of S1 ratios to S2 ratios when plotted against time. When a test finds bias after it is induced at a particular “cross-score threshold,” it is “True alarm,” and a test fails in doing so results in “False alarm.” The time a test takes to find bias is called “time till find days” (TTFD).

The Combined test has 95% true alarm success rate. It is also observed that for 75% of the true alarms, the algorithms are findings the bias within 39.5 days. The cross-score thresholds at which this is possible are 5 and 6. Coming to the individual tests, the Heat test is the better test of all; it is able to find the bias within 33 days after induction 75% of the cases at cross-score thresholds of

5 and 6. The heat ratio, average value and max value tests showed comparable results, however BARNFL and GLYFL tests do not perform up to the mark, due to their poor response to dynamic threshold strategy. The negative bias analysis also provides similar results. Identifying bias at such low magnitudes ($\pm 2\%$) is a hard task to achieve and with the MRFAA it is identified within approximately one month span of time 75% of the time, which is valuable to the industry. For the stripping vessels with combined cycle spanning 24 hrs (one day) or more, finding the bias within the short periods of time is valuable.

Some of the disadvantages of MRFAA algorithm are as follows. At this time the algorithm is not capable of identifying bias in several sensors simultaneously. The algorithm is not capable of identifying different types of errors other than bias. Presence of relations between sensors is the fundamental requirement for the algorithm in order to be successful.

1.6 Future Work

Additional tests can be added to the algorithm through a few lines of code. With some process specific changes to the Fortran code, the algorithm's capabilities can be extended or generalized to any other industrial applications. Since the algorithms are developed with open source Fortran 95, scalability to specific industrial problems is less expensive. Finding errors in several sensors at a time is also one of the additional functionalities that can be added to the algorithm to reflect real life complex situations. The algorithm with the additional capabilities has the potential for commercialization.

1.7 Acknowledgements

The author expresses his sincere gratitude to his academic advisor Professor Rajive Ganguli, all the members of the graduate advisory committee for their help and guidance, and to the Mining

Engineering Research Endowment (MERE) established at the University of Alaska Fairbanks (UAF) for financial support. The author also would like to thank Fort Knox Mine of Kinross Corporation and Pogo Mine of Sumitomo Metal Mining Company for their support of MERE and for providing the necessary data, facilitating site visits, etc. The author extends his gratitude to College of Engineering and Mines (CEM), and, the Department of Mining and Geological Engineering (MinGeo) at UAF for providing the valuable laboratory and other academic resources to accomplish the project.

1.8 References

- Bagajewicz, M., Markowski, M., and Budek, A. 2004, "Economic value of precision in the monitoring of linear systems," *AIChE Journal*, April 2005 Vol. 51, No. 4, pp.1304-1309.
- Baljak, V., Tei, K., and Honiden, S., 2012, "Classification of faults in sensor readings with statistical pattern recognition," *SENSORCOMM 2012: The Sixth International Conference on Sensor Technologies and Applications*, Aug 19-24, 2012, Rome, Italy, pp. 270-276.
- Buxton, M., and Benndorf, J., 2013, "The use of sensor derived data in real time mine optimization: a preliminary overview and assessment of techno-economic significance," *SME Annual Meeting*, Feb. 24 - 27, 2013, Denver, CO, Preprint 13-038, pp. 5-5.
- Fang, L., and Dobson, S., 2011, "In-network sensor data modelling methods for fault detection," www.simondobson.org/softcopy/uami2013.pdf, Accessed April 2015.
- Fast, J., 2016, "Carbon stripping," *Denver Mineral Engineers, Inc.*, www.denvermineral.com/carbon-stripping/, Accessed October 2017.

- Institute of Electrical and Electronics Engineers, 2017, "IEEE P1451.6 Terms and Definitions," IEEE, <http://grouper.ieee.org/groups/1451/6/TermsDefinitions.htm>, Accessed October 2017.
- Journal of Heuristics, 2017, "Journal of Heuristics – description," <https://link.springer.com/journal/10732>, Accessed October, 2017.
- Kalwinder, K., 2017, "Inside a car – coolant temperature sensors," [www.azosensors.com/images/Article_Images/ImageForArticle_49\(1\).jpg](http://www.azosensors.com/images/Article_Images/ImageForArticle_49(1).jpg), Accessed October 2017.
- Mulotec, 2017, "Minerals processing - gold," www.multotec.com/industry-flow-sheet/minerals-processing/gold, Accessed October 2017.
- Pogo Mine, 2016, "Pogo mill process descriptions," Print medium handout by Andrew Maxon, Pogo Mill Facility, Pogo Mine, Received January 2016.
- Process Parameters Ltd, 2017, "PPL4-T Thermocouple with fixed process thread," www.processparameters.co.uk/ppl4-t-thermocouple-fixed-thread/, Accessed October 2017.
- Quan-Bo, G., and Cheng-Lin, W., 2006, "The research of process monitoring based on data fusion theory," Fault Detection, Supervision and Safety of Technical Processes, Elsevier, A Proceedings Volume from the 6th IFAC Symposium, Beijing, China, August 30-September 1, 2006, pp. 276-281.
- Narasimhan, S., and Jordache C., 2000, "*Data Reconciliation & Gross Error Detection –An Intelligent Use of Process Data*," Gulf Publishing Company, Houston, Texas, pp. 300-301.

Narasimhan, S., and Rengaswamy, R., 2017, "Characterization of value for sensor networks for process fault diagnosis," <http://folk.ntnu.no/skoge/prost/proceedings/aiche-2004/pdf/files/papers/039d.pdf>, pp.1, Accessed September 2017.

National Institute of Instrumentation Standards, 2017a, "Concepts of accuracy, precision, and statistical control," www.nist.gov/sites/default/files/documents/2017/05/09/section-2-concepts.pdf, Accessed October 2017.

National Institute of Instrumentation Standards, 2017b, "Teach tough concepts: frequency domain in measurements," www.ni.com/tutorial/13042/en/, Accessed October 2017.

Sharma, A.B., Golubchik L., and Govindan R., 2010, "Sensor faults: detection methods and prevalence in real-world datasets," *ACM Transactions on Sensor Networks*, Vol. 6, No. 3, Article 23, pp. 23.2-23.39.

SME, 2017, "Author Style Guides – Mining Engineering," www.smenet.org/SME/media/Publications-Resources/HowtoSubmittoMiningEngineering_2014.pdf, Accessed November 2017.

Stum, K., 2006, "Sensor accuracy and calibration theory and practical application," *National Conference on Building Commissioning: April 19-21, 2006*, pp. 1-15.

UAF, 2017, "UAF Thesis Formatting Handbook – Fall 2017 v.3," www.uaf.edu/files/gradsch/Handbooks/UAF-THESIS-FORMAT-HANDBOOK-Fall2016_v3_20170922_TitleIX.pdf, Accessed November, 2017.

Wang, H., Chai, T.-Y., Ding J. -L., and Brown M., 2009, "Data driven fault diagnosis and fault tolerant control: Some advances and possible new directions," *ACTA Automatica Sinica*, Vol. 35, No. 6, June 2009, pp. 739-747.

Chapter 2: Statistical Methods in the Detection of Industrial Sensor Calibration Errors

2.1 Abstract

Sensor faults are costing US industries millions of dollars each year, in direct (instrumentation) and indirect costs (production loss). For the metal mining industries, the loss occurs in terms of metal recoveries. While gross errors are easily detectable using existing classical statistical methods, “calibration errors” like “bias” are subtle and difficult to detect. In an attempt to study the behavior of bias in sensor data of Fort Knox SAG mill, several classical methods like correlation, aggregation, and principal component analysis (PCA) were applied. In the correlation based study, a 10% bias was introduced in a SAG mill’s “feed” sensor data and error trends were observed. Initially, the data was divided into several states and the correlations between various SAG mill sensors (feed, RPM, HP, etc.) in each state were captured. The values indicated that the relational changes are more drastic from state-to-state than they are predictable. For instance, the R-square values for RPM vs feed for states 1 through 3, are 0.113, 0.065, and 0.780, respectively. In this context, correlation is disadvantageous when attempting to observe subtle changes that are caused by calibration errors in highly dynamic processes like SAG mill operation. In a similar fashion (PCA was conducted on the SAG mill sensors with a focus on finding bias in “feed.” The analysis did not produce significant differences between its biased and clean sets of feed even at an error of +10% magnitude. The corresponding “loadings” and the “biplots” from the PCA did not show significant differences either. The discussion on the results with conclusions are found at the end of the chapter. The chapter also provides an overview on various statistical techniques used in the industrial sensor error detection.

2.2 Introduction

Various statistical methods used in sensor fault detection are widely described by Sharma et al. (2010). Some of the methods briefly described are as follows. “Rule-based” methods use domain knowledge to develop heuristic rules for identifying faults. These are generally highly accurate methods, but the choice of heuristics can affect the results. “Estimation based” methods depend upon inter-sensor correlations, whereas linear “Least-squares estimation” (LLSE) techniques are used to flag faulty readings in sensor networks; however, these techniques are disadvantageous for classifying fault types. “Time series analysis” based methods are capable of finding short duration faults but prove disadvantageous to find long duration faults like noise. For sensor readings that do not exhibit periodicity, the auto regressive moving average (ARIMA) methods proves to be a better option. “Learning based” methods infer the normal sensor readings from the training data set to identify classes of faults in the test or prediction data set. Neural networks are an example of learning based methods; however, with neural nets training is time consuming. Contrary to the mainstream statistical models or classical, descriptive and inferential statistics, “Bayesian statistics” use prior knowledge. “Data Reconciliation” methods involve adjusting the data according to conservation laws and other constraints. For data sets that follow particular statistical distribution, parametric methods are used; example: linear regression. For data sets that does not follow any distribution, non-parametric or distribution-free statistics are used; Examples: bagging and boosting. For non-linear and non-stationary processes data-mining techniques are very popular. These are generally used for extracting knowledge from huge data sets. “Heuristics” involve the theory and practical application of techniques for solving problems approximately that cannot be solved exactly (Journal of Heuristics, 2017). Hybrid methods employ any combination of the above techniques to achieve the task. Calibration errors are one exception

that are hard to detect with the above methods. Without the availability of ground truth values, calibration errors become difficult to identify and even rectify. Data-mining based innovative methods seem to be the right choice to find calibration errors like bias in large industrial sensor data sets.

2.3 Literature Review

2.3.1 Estimation Based Methods

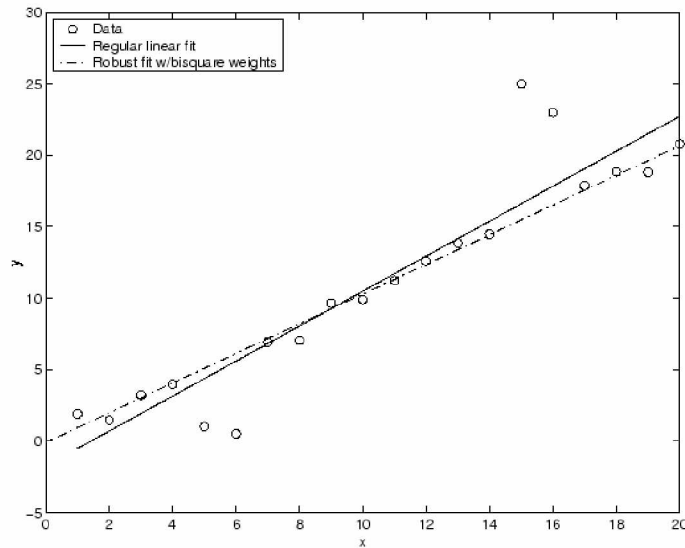
Estimation based methods depend upon inter-sensor correlations and data-fitting. Least-squares estimation (LSE) is one technique that is frequently used in flagging faulty readings in sensor networks. Due to their availability as well-established methods, linear models were extensively used in sensor fault detection in the recent past (Kusiak and Song, 2009). In the context of data-fitting, a residual or error is the difference between the observed data and expected data (see Equation (2.1)). For any given sensor data set, in order to obtain the “coefficient estimates” for a fitted line, the LSE methods minimize the summed square of residuals. When fitting the data, the underlying assumption for LSE methods is error is distributed normally (Mathworks, 2017).

$$r_i = y_i - \hat{y}_i \quad \text{(2.1)}$$

residual = data - fit error $\sim N(0, \sigma^2)$

Linear least squares estimate/fit (LLSE) is another method that assumes the errors (residuals) are distributed normally, and the variance is constant for the given data sets. A “linear model” in this context is defined as a fitted line equation for the variable of observation that has “linear” coefficients. If any of these assumptions are not true for a set of a data—for instance, a set of data with a lot of outliers, the variance is unduly large or small—the LLSE estimate could not result in a good fit. In such cases, weighted least squares (WLSE) is used to address the problem. Robust

least squares methods (RLSE) adjusts and standardizes the residuals in order to achieve a good fit (Figure 2.1). On the other hand, nonlinear least square methods build equations with nonlinear coefficients when the response data is non-linear. “Curve-fitting” methods are one example.



Source: Adapted from Mathworks, 2017.

Figure 2.1: Robust vs linear fit.

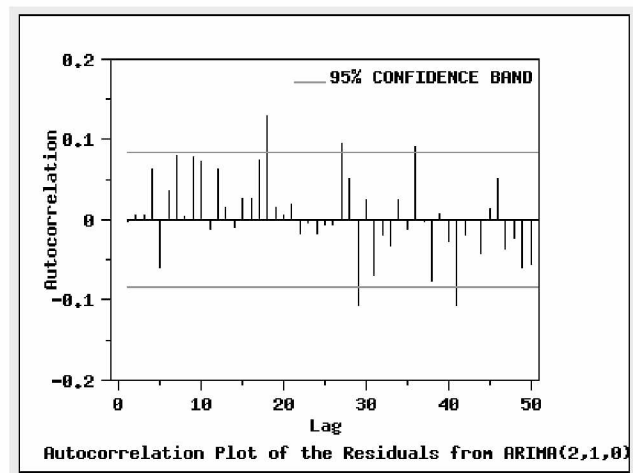
In all the above methods discussed, the “error” is found as the difference between the observed sensor data and the fit. Whether it is a single sensor or multiple sensors, these techniques are disadvantageous to classify the “types” of faults. Similarly, when multiple sensors are present, the sensors are dependent on each other, the interrelations are non-linear, and these methods are less effective.

2.3.2 Time Series Methods

A Time Series is a data-set that results from taking samples from successively, equally-spaced points in time. It is a sequence of discrete time data. It can also be defined as an ordered sequence of values of a variable of interest at equally spaced time intervals (National Institute of

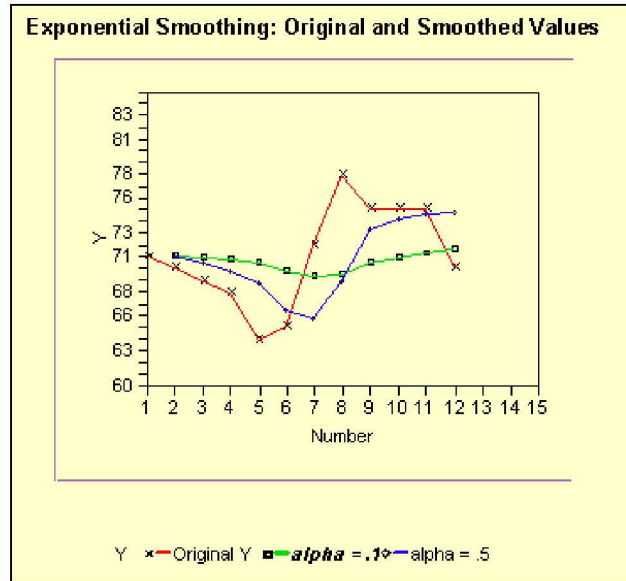
Instrumentation Standards, 2017a). According to National Institute of Standards and Technology (NIST), Time Series Analysis is used for many applications: economic forecasting, budgetary analysis, stock market analysis, process/quality control, etc.

The fitting with time series methods is done by averaging techniques like Box-Jenkins ARIMA/Multivariate models (based on non-linear least squares) or exponential smoothing methods like Holt-Winters exponential smoothing (Figures 2.2 and 2.3). In a “moving average” method, the past observations are weighted equally, whereas in “exponential smoothing,” recent observations are given more weight in forecasting than the older observations (National Institute of Instrumentation Standards, 2017a). Error in data is generally estimated by observing the difference between forecasted and actual (observed) trends of data. When the data observed is highly variable in nature—like in mineral processing industries—it is very difficult to forecast the trends accurately. This is one disadvantage in the context of this research. Time Series based methods are capable of finding short duration faults but prove disadvantageous for finding long duration faults like noise.



Source: Adapted from National Institute of Instrumentation Standards, 2017a.

Figure 2.2: Autocorrelation plot of residuals from ARIMA (2,1,0) Model.

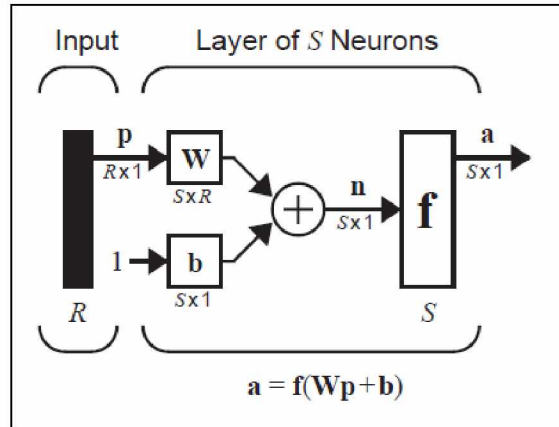


Source: Adapted from National Institute of Instrumentation Standards, 2017a.

Figure 2.3: Sample plot: smoothed data for 2 values of α (a factor).

2.3.3 Neural Networks

Neural Networks mimic the behavior of a complex non-linear system based on learning from a training data-set (Kusiak and Song, 2009). Like many non-linear methods discussed earlier, Artificial Neural Networks (ANN) are widely used to predict the sensor readings based on the training (input) data set (Figure 2.4). The difference between the neural net predicted value and the actual reading is considered as an error (Figure 2.5). Several researchers in the past used the ANNs to quantify the sensor errors. With Neural Networks, sensor faults are identified based on the magnitude of the error. Training the NNs for ever-changing process-related data is a time-consuming process; once the process changes, the NN needs to be retrained in order to predict the sensor readings. Mineral processing involves sudden changes in operating parameters to cope with changing ore types. Due to the dynamic nature of process circuits ANNs prove cumbersome for sensor error detection.



Source: Adapted from Hagan et al., 2014.

Figure 2.4: A typical Neural Network architecture.

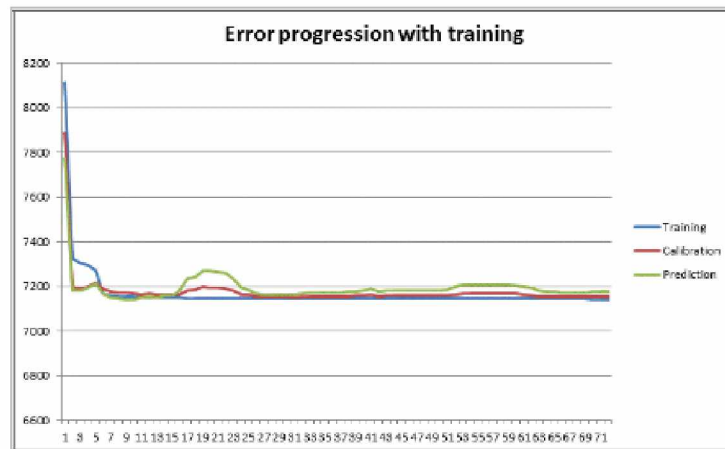


Figure 2.5: Error progression with training in an ANN.

2.3.4 Bayesian Statistics

The Bayesian approach treats population model parameters as random (not fixed) quantities. Even before looking at the current data set, prior knowledge like constraints and judgements are used (“prior distribution model”). The model developed with prior knowledge and the knowledge from the current data is called a “posterior distribution model” (National Institute of Instrumentation Standards, 2017b). Contrary to the mainstream statistical models like classical, descriptive and inferential statistics, Bayesian statistics use prior knowledge. A Bayesian belief

network was developed by Mehranbod et al. (2003) to classify faults in a sensor network. The values of process variables, their possible faults and biases were discretized into several categories. Based on the process sensor readings and their normal readings, the authors were able to calculate the probabilities for possible error types. Bayesian maximum a posteriori probability method and a hierarchical special Bayesian space-time model (HBSM) were developed by Ni (2008) to detect sensor network data faults in environment monitoring sensors. The disadvantages for Bayesian models are: there is no correct way to choose a prior model, and the choice could heavily influence the posterior model.

2.3.5 Conservation Laws and Data Reconciliation

Data Reconciliation involves adjusting the data according to conservation laws and other constraints. In the mining industry it is common to find differences between observed and actual data. Besides removing “noise” using signal processing methods and wavelet de-noising, data reconciliation can further improve the quality of the process data. Authors Tahane and Mah (1985), used weighted least squares estimation (under constraints) to reconcile the flow measurement data in a chemical process network. Hodouin (2010) developed several data reconciliation methods for mineral/metallurgical plants that use mass conservation balance laws in order to correct the observed data for errors. Several classical methods were used to estimate various statistical parameters. Hodouin (2010) had to assume that the measurement errors are Gaussian and unbiased in order to develop the methodology. If the system is highly non-linear and the variables (errors) of study does not exhibit Gaussian distributions, linear models fail. This is one reason why methods such as data reconciliation are less powerful in terms of accuracy for mineral processing circuits.

2.3.6 Data-Mining Methods

Data-mining techniques are broadly classified as: “association” based, where relation between variables is exploited; “classification” based, where attributes of each class of items are studied; “clustering” based, where one or more attributes of the classes are examined and grouped together; and “pattern recognition” based, where identifying trends or regular occurrences is used (Brown, 2012). Some other techniques that use a combination of any of these aforesaid are: “prediction,” where classification, pattern recognition and association are use (an example is forecasting of a company’s stock performance); and “decision trees,” where classification and prediction are used (an example is classification of various sensor faults). In general, real world problems may require several combinations of these techniques or entirely innovative approaches that are specific to a particular problem.

For highly non-linear and non-stationary processes, it is difficult to develop analytical models based on classical or fundamental statistical methods. Moreover, sensor validation for such processes is equally tough due to production of false alarms in excessive amounts (Kusiak and Song, 2009). Data-mining techniques are very useful in such situations. The techniques have a wide range of industrial applications, and are generally used to extract knowledge from huge data sets—industrial sensor data sets are one example. Researchers Kusiak and Song (2009) were able to develop several algorithms that are based on the clustering technique to detect sensor faults in power plant boilers. Classification techniques were used in a “cloud” based application to detect errors in a big sensor data set (Yang et al., 2015). Detailed literature reviews on various data-mining methods and their applications are found in Chapters 4 and 5.

2.3.7 Heuristic Methods

Heuristics involve the theory and practical application of techniques for solving problems approximately that cannot be solved exactly (Journal of Heuristics, 2017). Heuristic methods are useful when all conventional or exact methods fail to give optimum solutions and consume plenty of time for even trying. The solution from heuristic methods is good, but not as accurate as the solution obtained from the conventional or exact methods (detailed in the previous sections); however, it consumes less time and might be the only way the solution can be obtained. Sometimes the solution may not exist through exact methods, which could be the reason to resort to heuristics (Marti and Reinelt, 2011). There are several types of heuristic methods. “Inductive” methods find general solutions to the smaller problem, then the solution is scaled up for the bigger problem. “Reduction” methods restrict the space of solutions to good solutions only. “Constructive” methods build solutions step-by-step from scratch. In “Decomposition” methods, the original problem is broken into smaller manageable pieces for better examination. Sivakumar and Venkatesan (2015) described a Meta-heuristic approach that was used for minimizing error in localization of wireless sensor networks. Several optimization algorithms were used in this connection. A root mean square error (RMSE) was used for the performance evaluation of these algorithms. A heuristic method for error correction in parallel probe-based storage devices is presented by Varsamou and Antonakopoulos (2009).

In some cases, heuristic approaches, when used in fault detection, combine rule-based analysis and a fault detection algorithm. One such application is described by Garcia et al. (2009). The algorithm is applied to detect faulty groups of instruments in a pilot plant that consists of an inline mixer process. Where high reliability is desired, multiple sensors can be used to measure the same quantity. If one sensor output differs, it can be tagged as faulty. An algorithm developed in this

context, was used in finding a faulty sensor in a pool of four sensors that were measuring temperature (Oh, 2015). The drawback in this kind of approach is the expense of buying multiple sensors to measure the same quantity. The operational costs of such sensor networks are too high. Knowledge based systems (KBS) are another variation of the heuristic approach. The deployment of such system for detection and identification of several types of errors in an electro-mechanical actuator system, were described by Silva et al. (2012). The approach was also compared to a neural network based analysis.

2.3.8 Data Aggregation

The widespread use of sensors in industrial automation often results in large sets of data. In order to observe certain trends of data, it is essential to reduce the amount of data without losing its quality (Arku and Ganguli, 2014). Data “aggregation” is one technique that can help reduce the amount of data to be analyzed. When data is averaged in certain intervals, like 5 or 10-min, in a closely sampled data-set, like 1-min sampling, this would not impact the outcomes of an analysis significantly. Aggregation could potentially save computer memory. Experimenting with a SAG mill sensor data, Arku and Ganguli (2014) found that aggregation of data at 5-min or 10-min intervals did not give any significant difference in prediction of the SAG mill horse power. The correlations between input and output variables, and the error performance of various subsets (with the help of Neural Networks) were analyzed in the study. For this research, aggregation of data has been performed in the context of reducing noise and to observe broad underlying trends of sensor data.

2.3.9 Parametric and Non-Parametric Methods

The statistical methods that assume the given data is normally distributed and the relations are linear are called “parametric methods;” paired t-test, Pearson correlation, principal component analysis (PCA), and one-way ANOVA are some examples. In many real world problems, however, data could exhibit any kind of distribution or might be distribution-free, which warrants the use of “non-parametric” or “distribution-free” methods. Bagging, boosting, regression trees, and generalized additive models (GAMS) are some examples that can be used for highly dynamic processes (mineral processing). Bagging was proposed by Breiman (1994). Bootstrap aggregation, or bagging, is a technique that can be used with classification and regression methods to reduce the variance associated with “prediction,” thus improving the prediction process. Variance is reduced due to simple averaging (Sutton, 2005). Unlike bagging that uses simple averaging, boosting employs a weighted average of prediction results obtained from various samples. In other words, boosting is an iterative procedure that incorporates weights, as opposed to bagging (Sutton, 2005). “Decision tree” learning is one of the machine learning or data mining techniques that maps the relations between predictor and response variables. Advantages with trees are: flexibility to suit a broad range of response data (numeric, categorical, and survival), ease and robustness, ease to interpret visually, and ability to handle missing data. Thus, trees are an alternative to many traditional techniques like multiple regression, analysis of variance (ANOVA), linear discriminant analysis, and survival models (De’ath and Fabricius, 2000).

2.3.10 Hybrid Methods

Hybrid methods employ any combination of the above techniques to achieve the task. The innovative approaches undertaken in this dissertation fall under this category; they are mainly a combination of data-mining techniques.

2.4 Introduction to Fort Knox Mine and Mill Operations

Fort Knox is an open-pit gold mine located near the city of Fairbanks, Alaska. It is mined through conventional open-pit methods: drilling and blasting to liberate ore from the earth, the use of electrical excavators to dig the broken rock, and haul trucks for transporting the ore. The basic stages of mining and mill operations are similar to those of any typical gold mining operation with some differences. For instance, high grade ore is processed at mill while low grade ore is processed at a valley-fill “heap leaching” facility. The mill has a daily capacity of 49,604 tons (45,000 tonnes) (Fort Knox, 2017). Along with large volumes of low grade ore, heap leaching operations handle mineralized waste too. The location of the operation is shown in Figure 2.6.



Source: Adapted from Fort Knox, 2017; Alaska Journal, 2017.

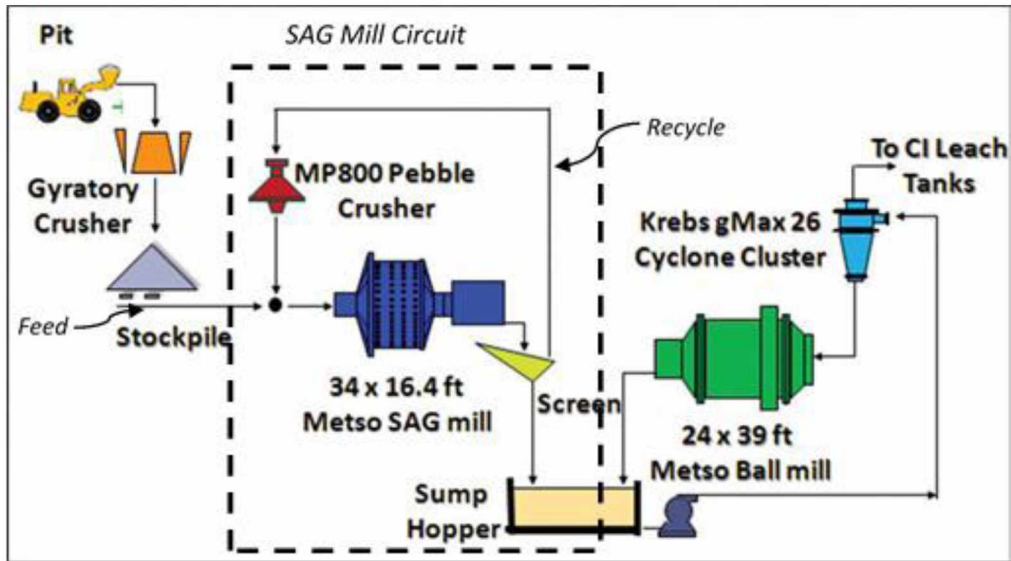
Figure 2.6: Fort Knox mine.

2.5 Methods and Materials

2.5.1 Data collection

In order to apply the proposed methodology and observe the trends for sensor readings, a semi-autogenous (SAG) grinding mill operation or simply “SAG mill” or “mill” at Fort Knox mine was chosen. The SAG mill setting is similar to that depicted in Figure 2.7. Data was collected at 1-min intervals. A snapshot of data collected and sensors of interest is summarized in Table 2.1. Various key operational parameters of SAG mill and their corresponding sensors are as follows: feed ore to mill in tonnage per hour (tph) is measured by “feed” sensor (feed is usually mixed with water before it is fed to mill to form “feed slurry”); bearing pressure of the rotating mill in pound-force per square inch (psig) is measured by “BP” sensor; the number of mill revolutions per minute (rpm) is measured by “RPM” sensor; the rate at which the work is done by the mill, i.e., horsepower (hp) is measured by “HP” sensor (1 imperial hp = 745.7 watts, and 1 watt = 1 Joule/s). After the feed is processed (ground) in the mill, some part of the product that could not meet the expected (or targeted) size distribution is recycled to feed, which is called “recycle feed” or simply “recycle” (Figure 2.7). The feed material when combined with recycle is named “feed+recycle” These two variables/parameters are also included in few analyses.

In addition to the above variables, the following variables are examined for correlations in some occasions (see data characterization section): the volume of solids in feed slurry in percentage (“solids”); percentage volume of steel balls out of mill volume (“ballvol”); noise coming from the mill in decibels (dB) (“noise”); the electric current intake of mill motors in amperes (amp) (“amps”). Approximately 20,000 records or readings (1-month worth of data) were used for the analyses described in this chapter.



Source: Powder and Bulk Solids, 2017.

Figure 2.7: A typical SAG mill circuit.

Table 2.1: SAG mill sensors (RPM, BP, feed, HP) and data.

Time of observation	Mill revolutions/min	Mill Discharge end bearing pressure	Mill new feed	Mill power
Date-Time	RPM	BP (psig)	Feed (tph)	HP (hp)
5/14/04 6:30	10.4	768.6	1848.8	13692.3
5/14/04 6:31	10.4	768.4	1843.5	13665.9
5/14/04 6:32	10.4	768.6	1907.3	13643.0
5/14/04 6:33	10.4	768.9	1833.7	13626.3
5/14/04 6:34	10.4	769.2	1827.7	13610.8
5/14/04 6:35	10.4	769.4	1850.9	13592.8
5/14/04 6:36	10.4	769.7	1804.6	13584.9
5/14/04 6:37	10.4	770.0	1844.8	13595.6
5/14/04 6:38	10.4	770.2	1865.3	13591.9
5/14/04 6:39	10.4	770.5	1841.5	13577.5
5/14/04 6:40	10.4	770.7	1843.1	13562.9
5/14/04 6:41	10.4	770.8	1865.0	13548.4

2.5.2 Assumptions and scope

Calibration errors are present in the data often in the form of an offset value (bias) added to the original or true reading; see Equation (2.2). Bias is the common error associated with sensors and often a hard one to identify, hence it is the focus of this research. Other common errors like gross errors are out of the scope of the research. For the purpose of this experimentation, it is assumed that the data collected from the Fort Knox mine is devoid of errors, hence the set is deemed as a “clean set.”

$$\text{Observed or biased reading, } x_{bias} = \text{True reading } (x_{true}) + \text{bias } (e^{bias}) \quad (2.2)$$

Bias is expressed as a percent over the true reading value. For the purpose of experimentation, bias was artificially induced in the clean data set. A +10% bias indicates 10% of the true reading added as bias, likewise -10% expresses adding -10% of the true reading. Even though 10% is slightly on the higher side, it is used in this chapter for the experimentation purpose, and to find out if it is even detectable at that magnitude. The strategy was changed in the subsequent chapters since calibration errors are subtle and could be as low as 2% over the true reading. It is also learned that identifying the bias at such low magnitudes is preferred by the industries.

2.5.3 Data Preparation

2.5.3.1 Programming and Software

The sheer volume of the data collected, the subsequent mathematical calculations and repetitions to be performed, statistical analyses to be conducted, and graphical presentation of the results, demanded the use of a mathematically intensive software package. Matlab program/package was chosen in this context. A positive (+10%) and a negative (-10%) bias were

introduced in the data, only to identify them using innovative methods described in the subsequent sections. A series of “cut-offs” or “thresholds” also were applied through Matlab algorithms to cleanse the data for unreliable or corrupted readings. The algorithm description sections have the detailed account on the process. In addition, the SPSS and R software were used to plot some descriptive statistics and to conduct principal component analysis.

2.5.4 Correlation Based Sensor Data Characterization Method

The first step in analyzing any data set is to understand the trends of various sensor readings and to exploit if there is any relation between the variables of measurement. This is called “data characterization.” Breakage of mineral from rock (ore) is essential to liberate the precious metals. Blasting operations in mines break the rock into approximately 40”-60” maximum diameter particles, called run-of-mine (ROM), which is sent to the mill for further reduction. A SAG mill is the first piece of equipment in the mill process flow that helps in reducing the ROM to the sizes of 16” diameter to as low as 75 microns, before sending the product to ball mill for further reduction (Figure 2.7). The SAG mill is generally charged (4-12% of its space) with steel balls, approximately 4” in diameter. While the mill is rotated at a certain rpm, the ROM material (ore) is fed. The action of gravity and centrifugal forces with steel balls as medium, aids in breaking the rock into smaller sizes. The bearing pressure (BP) of the mill denotes how much material is being loaded into the mill. The power consumption of the mill (HP) is dependent on all of these factors.

Different types of ROM ore require different mill parameter settings to attain the desired size reductions. These gradual changes or states of operation (simply “states”) are observed through plotting the sensor raw data (Figure 2.8). In order to observe the changes in the states through different statistics, correlation methods were used. Using correlation methods, if the change in

trends of a clean set of sensor data is observed, it can be compared to an erroneous (biased) set of data to identify bias. This is the premise that is exploited in the correlation based methods.

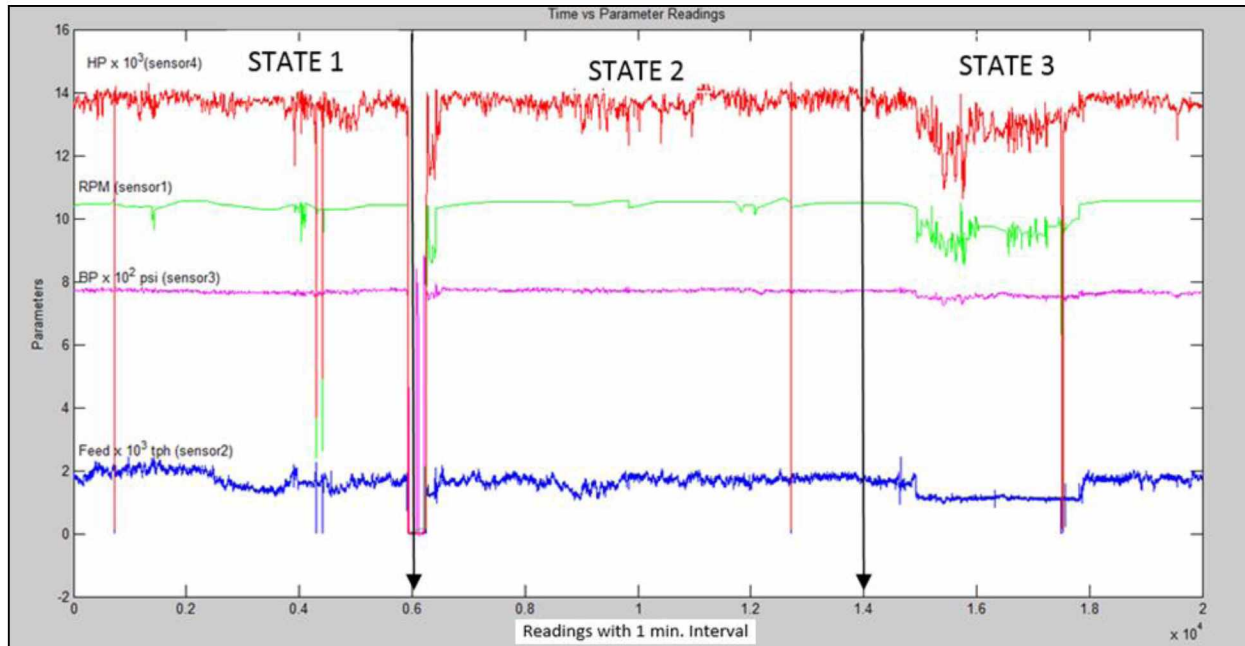


Figure 2.8: Sensor raw data “states” in time domain.

2.5.4.1 Data Characterization

Data characterization involves finding the statistical characteristics of the data and fitting them into specific distribution patterns. The characterization process can facilitate capturing statistical parameters between various combinations of sensor data variables. Initial examination of the sensor data in this context revealed the underlying distributions and the relations between various sensors (Figure 2.9). Correlation plots help to determine if the relations are reliable (“correlation” based). If it is determined such relations exist in the data set, R-squared values between the variables could be captured using Matlab statistical tools. The R-squared value (varies between 0-1) is a statistical measure of how closely the data fit to the regression line; the higher the R-squared value, the better the fit. A root mean square value (RMSE) or simply “error” indicates how far the

data are dispersed from the fitted regression line. The lesser the error, the better the “fit.” Various correlations between several variables of interest from SAG mill data (Fort Knox) are shown in Figure 2.9. The correlation values are summarized in Table 2.2. The data was cleaned for outliers and corrupted readings before it was used in the correlation plots.

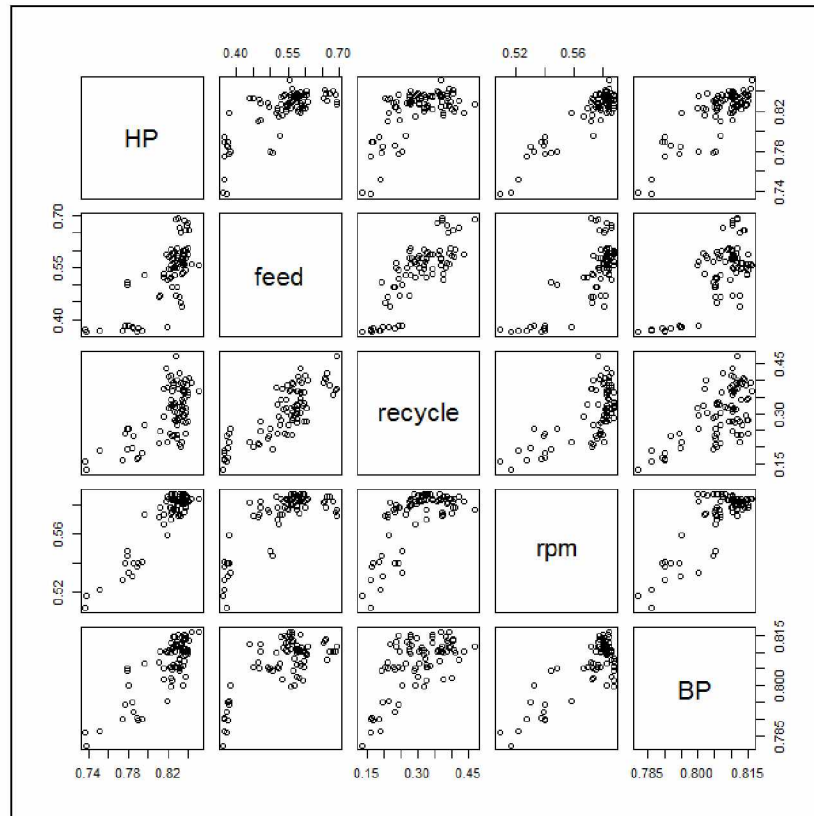


Figure 2.9: Correlation plot (cleaned data).

Table 2.2: Correlation matrix (raw data).

Correlation Matrix									
	rpm	bp	solids	noise	recycle	feed	amps	ballvol	
Correlation	rpm	1.000	.704	-.054	.067	.587	.668	.675	.238
	bp	.704	1.000	-.286	.205	.561	.607	.612	.502
	solids	-.054	-.286	1.000	-.412	-.064	.086	-.233	-.330
	noise	.067	.205	-.412	1.000	-.338	-.131	.074	.272
	recycle	.587	.561	-.064	-.338	1.000	.700	.490	.155
	feed	.668	.607	.086	-.131	.700	1.000	.650	.226
	amps	.675	.612	-.233	.074	.490	.650	1.000	.228
	ballvol	.238	.502	-.330	.272	.155	.226	.228	1.000

When plotted the Fort Knox sensor data indicated multiple “states” (Figure 2.8). If various parameters (feed, RPM, BP, etc.) show consistency for a period of time, it was called a “state” of operation. Interrelations between sensors exhibit consistency within a state and differs significantly when compared to a different state (Figure 2.8). The same plot that is cleaned for outliers is shown in Figure 2.10. The subsequent sections show the regression analysis results for various states of the process.

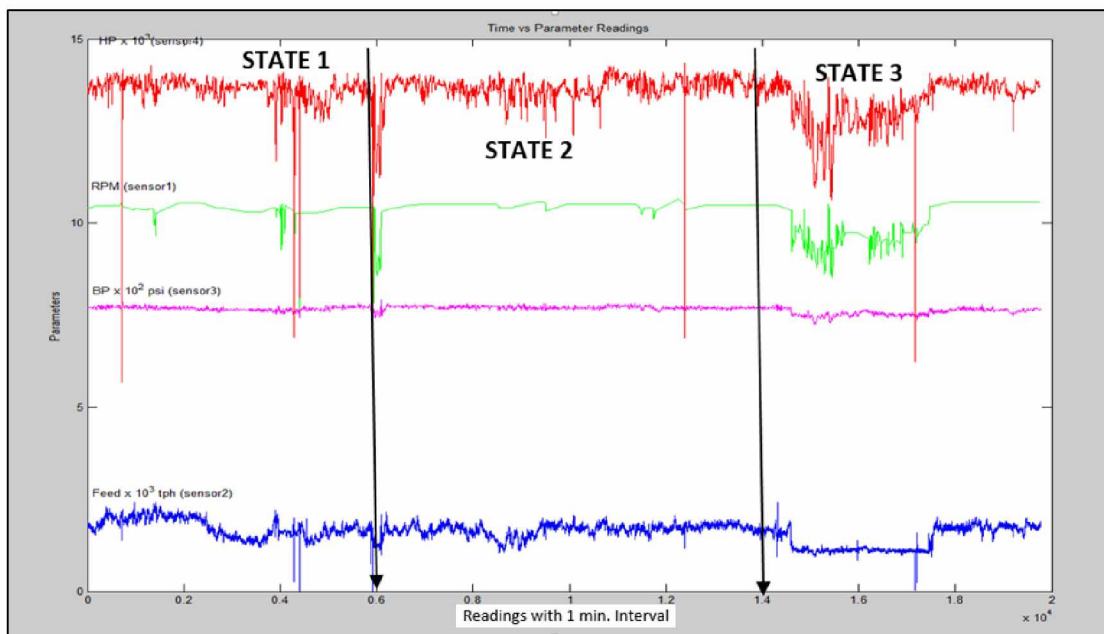


Figure 2.10: Sensor raw data in time domain after removing outliers.

2.5.4.2 Description of Algorithm

A Matlab algorithm was developed to capture and plot various R-square and RMSE statistics for each state. The algorithm in Figure 2.11 was used in the conduction of correlation based data characterization analysis. The program interface is shown in Figure 2.12. In the initial steps (data preparation), the algorithm reads various sensor data and inputs, and cleans the data for any

corrupted readings and outliers. Depending on the user input choices (based on the initial examination of the trends of data), the program divides the sensor data into different states. For each state the algorithm captures R-square and RMSE values for various sensor combinations of user interest. Once the R-square and RMSE values are captured, they are considered as the “base line statistics” (BSI), since they are from the “clean set” of the data. As an example, the “fitted line plots” for two pairs of fitted sensor measurement data are shown in Figures 2.13 and 2.14. The strategy is to compare these values to that of a future “erroneous data set” or “biased set.” At the end, the cleaned sensor data is also plotted state-by-state. This helps in the examination of state specific data trends.

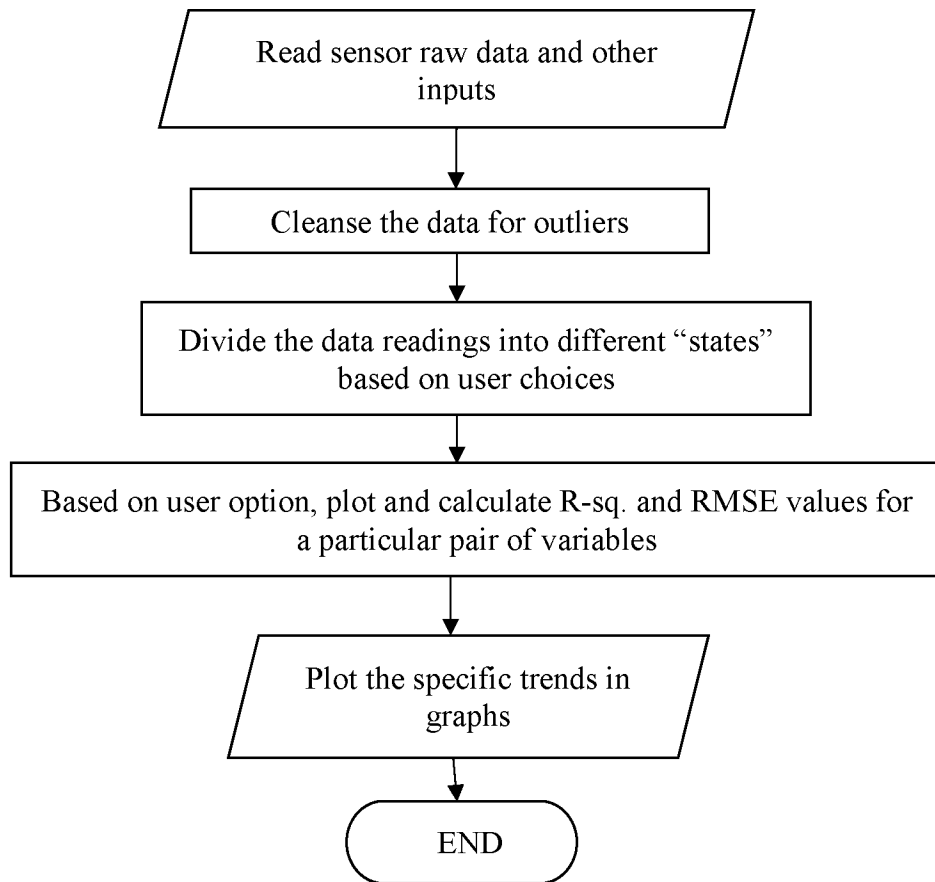


Figure 2.11: Flow chart for the Matlab-based data characterization algorithm.

```

Please input stage number:1
Please input starting cell number:1
Please input starting cell number:20000
-----
.....
Linear Regression Model Options
.....
For RPM vs HP results, enter Case number:1
For Feed vs HP results, enter Case number:2
For BP vs HP results, enter Case number:3
For RPM vs Feed results, enter Case number:4
For Feed vs BP results, enter Case number:5
For RPM vs BP results, enter Case number:6
.....
Please enter the CASE NUMBER from the above options:1

```

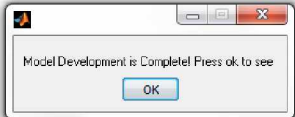


Figure 2.12: Matlab code with various model options.

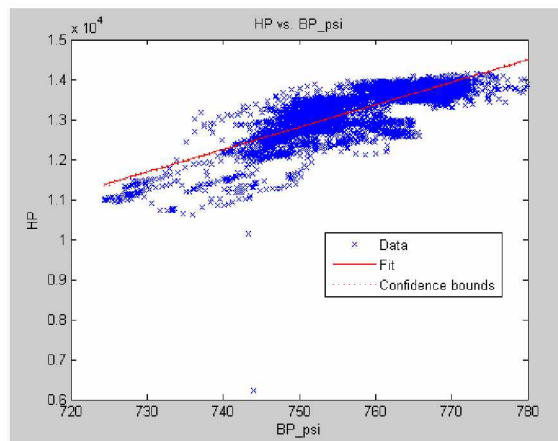


Figure 2.13: State 3 fitted line plot – BP vs. HP.

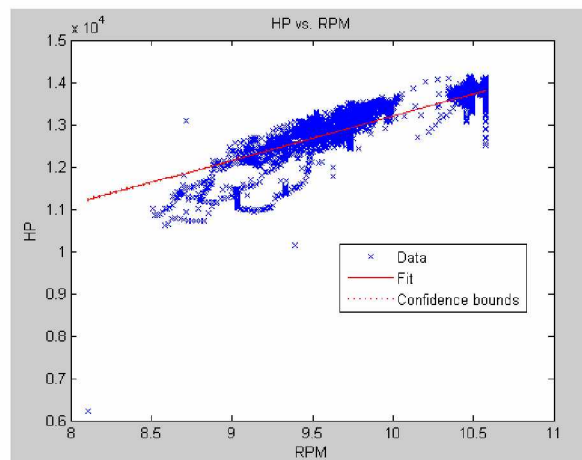


Figure 2.14: State 3 fitted line plot – RPM vs. HP.

2.5.5 Aggregation Techniques and Lags to Observe Trends

Data “aggregation” is a useful technique in observing the underlying trends of a sensor data set. Industrial sensor data exhibits frequent fluctuations and is in general associated with noise (see Figure 2.8). With aggregation, the readings for a particular interval are averaged into one value, for instance, a 60-min aggregation presents one average value for the 60 minutes’ worth of data. “States” in mineral processing circuits in general do not change minute by minute. In such cases, it would be worth observing the data for longer periods of time. Aggregation is a useful tool in that aspect. The effect of aggregation on raw sensor data is shown in Figures 2.15 through 2.17. From the figures, it is understandable that data trends are increasingly observable when aggregation is increased from 60 to 240 min average. Sometimes there might be a delay in the start of one process to another. For instance, there might be a small time gap (“lag”) when feed’s effect starts to show on mill horsepower (HP). If that lag is 30 min, then the correlation between HP reading and feed reading (observed 30 min ago) is the better estimate than the contemporary correlation estimate (R-value). Due to this reason, various correlation values at different lags were captured, just to study the effect in general. The results are provided in section 2.6.

Another technique to observe the behavior of data is “% match.” For instance, assume that an individual aggregated reading (current) of a sensor data stream (feed) is compared with its previous readings. If the previous reading is less than the current reading, a “direction” value of “1” is assigned to the current readings. If it is less than the previous one, a value “-1” is assigned. If they are equal, a value of “0” is assigned. An array that captures all these values for a sensor data stream is called “directions set.” The symbols/values 1, 0, and -1 indicate the upward or downward trends of a data stream when plotted in a graph with respect to time. If a “directions set” array is created for another sensor data stream (HP), and compared to feed array, in term of percentage of symbols

matching (“%Match”), the similarity in behavior of these two data sets can be understood. A higher matching percentage indicates highly similar behavior. Results from the “% match” analysis are provided in section 2.6.

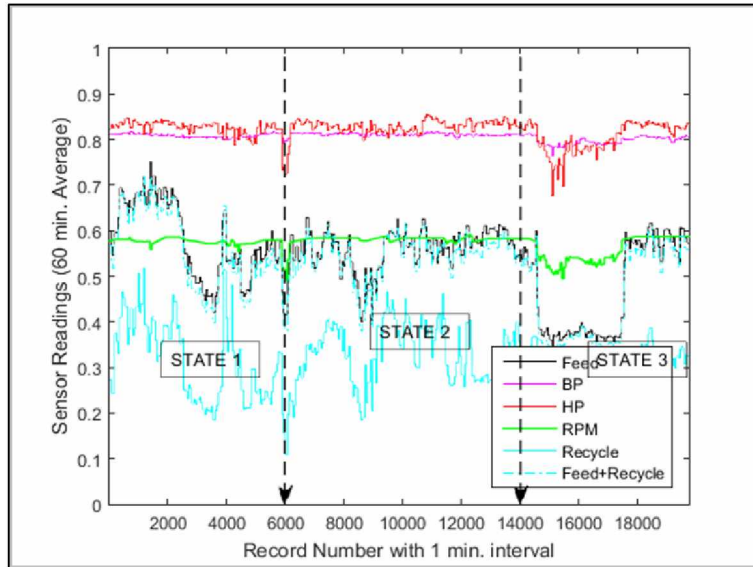


Figure 2.15: Data trends at 60 min aggregation.

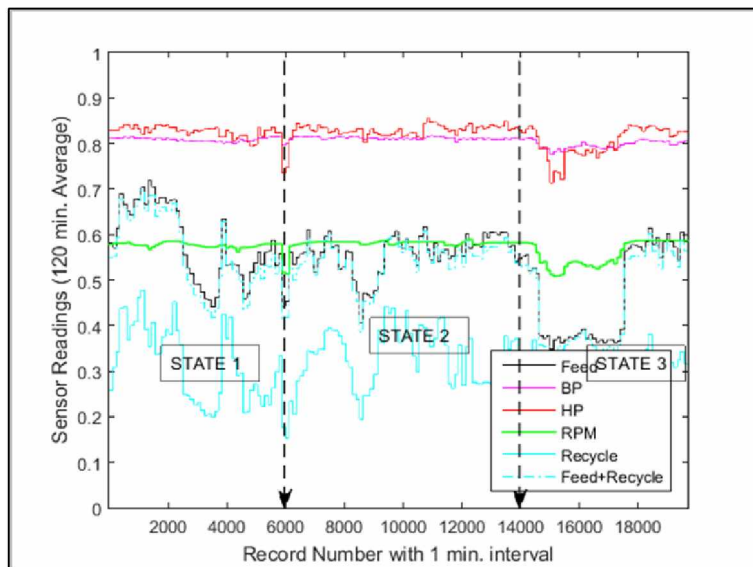


Figure 2.16: Data trends at 120 min aggregation.

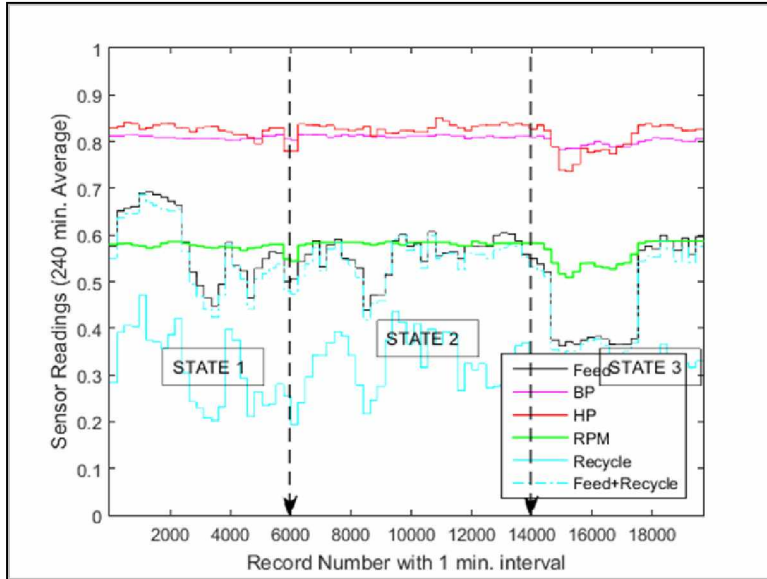


Figure 2.17: Data trends at 240 min aggregation.

2.5.6 Use of Principal Component Analysis (PCA) in Sensor Fault Detection

When the sensor data trends follow normal distribution and there exist linear relations among the variables, PCA can be used successfully to detect errors. Measurement correlations captured by PCA, residual analysis, and a sensor validity index (SVI) were used to find the sensor faults in a boiler circuit (Ricardo, 1996). The “difference” between PCA in terms of “principal component” angles for clean and faulty sensor data sets was used as the basis for sensor fault detection in a structural health monitoring system (Kerschen et al., 2004). A non-linear method called Kernel PCA was used in the fault detection process of a simulated data set in a continuous flow stirred-tank reactor (CSTR) process in another study (Cho et al., 2005).

Taking inspiration from the past research, PCA was used in analyzing the Fort Knox sensor data (SAG mill sensors). Since feed was observed to be the most influencing parameter on HP compared to other input parameters, it was selected for the analysis. Assuming the “feed” data set collected from Fort Knox as the “clean set” or a “baseline-set,” a 10% random noise was induced

into feed data to create “biased set” (Figure 2.18). The clean and biased sets are compared with each other in an attempt to detect bias presence.

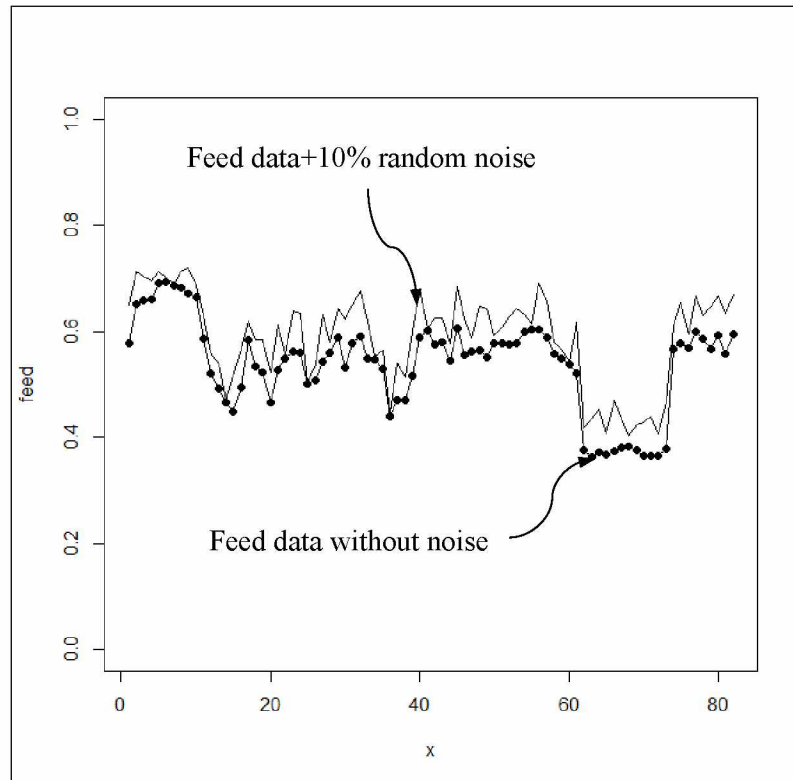


Figure 2.18: Clean vs 10% noisy/bias data.

2.6 Results

2.6.1 Correlation Based Sensor Data Characterization Method

The following are the results from the correlations (R-sq values) captured between various variables. For each “state,” the sensor interrelations (correlations) were captured, which are called base line statistics (BSI). Once the BSI were established, the future data can be compared against the BSI to find the calibration faults. The tables 2.3 through 2.5 present the results from the three process “states.”

Table 2.3: State-1 baseline sensor interrelations (BSI), records: 1-6000.

SENSORS	R-sq values				Error	Statistical Parameters		
	RPM	Feed	BP	HP	HP (RMSE)	Mean	Std	Variance
RPM	1.000	0.113	0.192	0.537	278.3	10.4	0.2	0.1
Feed	0.113	1.000	0.137	0.085	391.3	1733.8	268.4	72043.1
BP	0.192	0.137	1.000	0.230	359.0	768.2	5.1	25.5

Table 2.4: State-2 baseline sensor interrelations (BSI), records: 6001-14000.

SENSORS	R-sq values				Error	Statistical Parameters		
	RPM	Feed	BP	HP	HP (RMSE)	Mean	Std	Variance
RPM	1.000	0.065	0.163	0.371	255.6	10.5	0.2	0.0
Feed	0.065	1.000	0.012	0.088	307.9	1672.7	157.8	24909.9
BP	0.163	0.012	1.000	0.243	280.5	771.0	4.0	15.9

Table 2.5: State-3 baseline sensor interrelations (BSI), records: 14001-20000.

SENSORS	R-sq values				Error	Statistical Parameters		
	RPM	Feed	BP	HP	HP (RMSE)	Mean	Std	Variance
RPM	1.000	0.780	0.626	0.789	283.7	10.1	0.5	0.3
Feed	0.780	1.000	0.552	0.510	432.7	1412.4	311.6	97122.7
BP	0.626	0.552	1.000	0.666	357.1	758.1	9.0	80.9

2.6.2 Aggregation and Lags

The following are the correlation results (Table 2.6) from the Fort Knox data variables and corresponding “lags.” At several minutes of aggregation, 30-240 min, various correlation values were captured for various lags. For instance, lag-0 correlation means there is no delay between the variables correlated, lag-1 indicates there is one reading gap between the variables for correlation. The results populated in Table 2.6 are for 240 min aggregation. It is observed that at lag-0, there is better correlation between variables at 240 min aggregation. Likewise, at different values, “% Match” between two variables might be high at certain lag.

Table 2.6: Analysis of various parameters and lags at 240 min Aggregation.

	Feed vs HP	Feed vs BP	Feed vs Recycle	Feed+Recycle vs HP
	R-value			
Lag0	0.74715847	0.81839207	0.81825094	0.74440247
Lag1	0.70051181	0.70528901	0.70199743	0.69122573
Lag2	0.55026374	0.61064728	0.54661373	0.53810150
	% Match			
Lag0	0.53086420	0.55555556	0.74074074	0.53086420
Lag1	0.57500000	0.60000000	0.51250000	0.55000000
Lag2	0.58227848	0.48101266	0.51898734	0.48101266

2.6.3 Use of Principal Component Analysis (PCA) in sensor fault detection

The results from the PCA analysis (“loadings”) for clean and noisy data can be seen in Tables 2.7 through 2.8. Loadings indicate correlations between a component (combination of variables) to a particular variable (Example: component 1 vs feed). When compared to clean data (feed), it is observed that the loadings (component values) slightly changed due to the presence of noise in feed. The angle of the variables for highly influential components (1 and 2) also changed slightly, which is observed in biplots (Figures 2.21 and 2.22); feedN indicates noisy “feed.” These differences if quantified, could be used against certain preset limits/thresholds to identify the faulty sensors and the type of faults associated with each.

Table 2.7: Feed data without noise.

Loadings:				
	Comp. 1	Comp. 2	Comp. 3	Comp. 4
feed	-0.515	0.294	-0.549	0.590
recycle	-0.481	0.673	0.446	-0.340
rpm	-0.508	-0.405	-0.420	-0.633
BP	-0.495	-0.544	0.568	0.368

Table 2.8: Feed data with 10% random noise.

Loadings:				
	Comp.1	Comp.2	Comp.3	Comp.4
feedN	-0.509	0.348	0.656	0.436
recycle	-0.483	0.632	-0.563	-0.222
rpm	-0.513	-0.391	0.282	-0.710
BP	-0.494	-0.571	-0.417	0.506

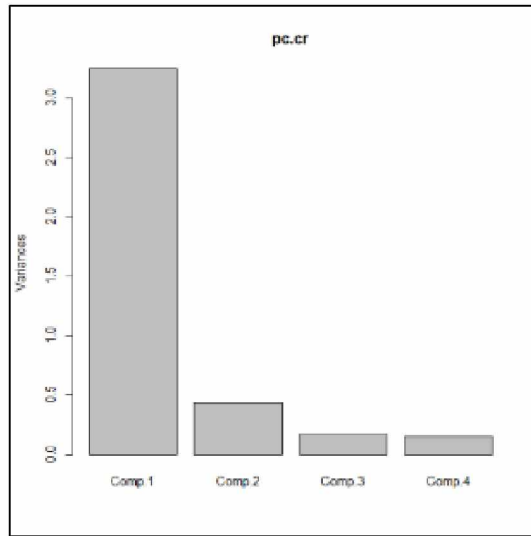


Figure 2.19: Variance for clean data set (pc.cr).

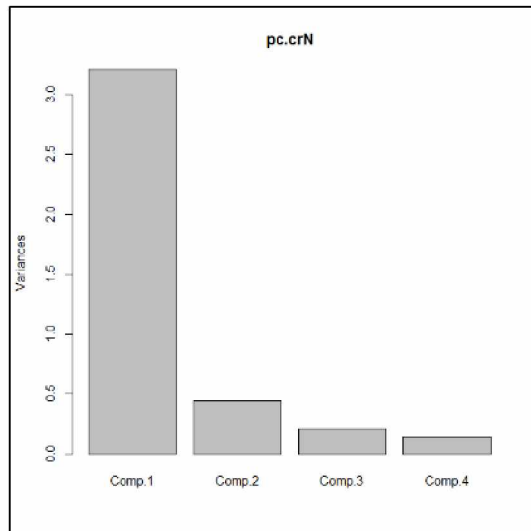


Figure 2.20: variance for noisy data set (pc.crN).

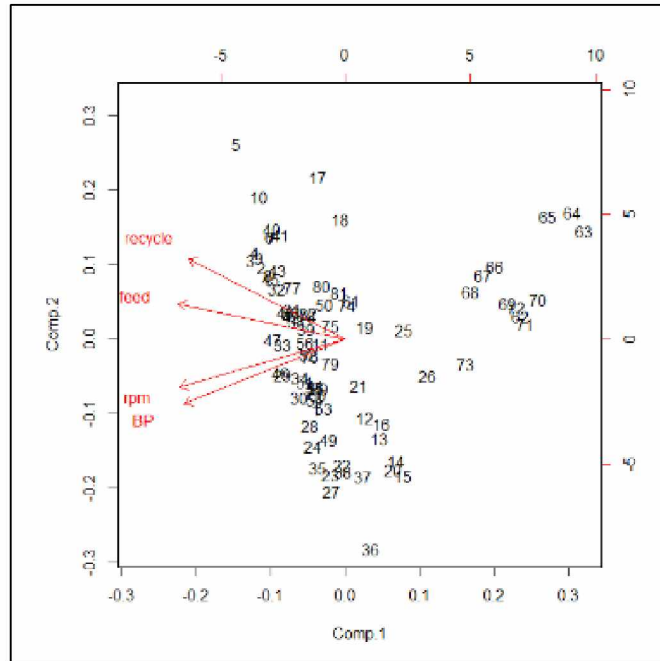


Figure 2.21: Biplots for clean data set.

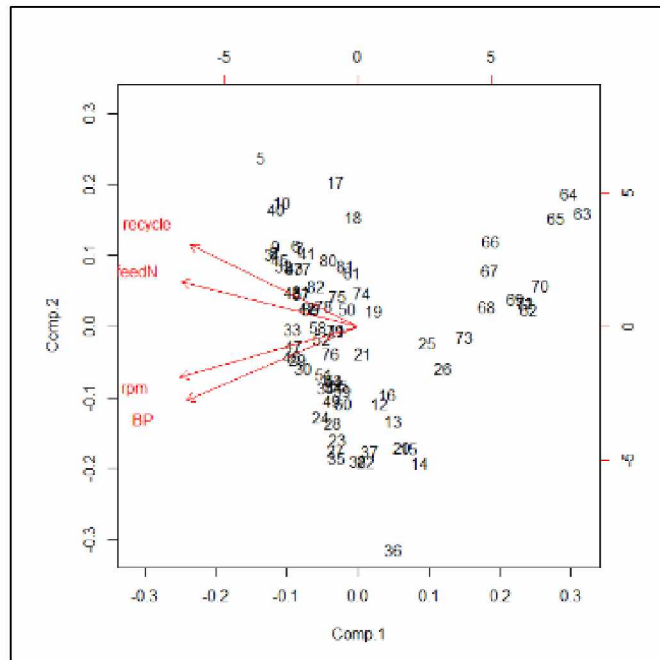


Figure 2.22: Biplots for noisy data set.

2.7 Discussion

Process “states” in industrial sensor data of a SAG mill operation were explored using correlation based methods. Baseline statistics (BSI) in the form of R-square and R-values are captured for each state of SAG mill operation. The main drawback from this kind of approach, however, is that the difference between R-square and R-values is minute from state to state, which might not even be significant at a level of $\pm 10\%$ bias over true readings. It was observed from Tables 2.3 through 2.5, that the difference between R-values of any given variable pairs is not consistent among states 1 through 3. For instance, the R-square values for RPM vs feed for states 1 through 3, are 0.113, 0.065, and 0.780, respectively. This means either there is less correlation naturally between these variables or there is a lot of variance. In general, mineral processing operations are very dynamic and the states change from time to time. This is problematic when data behavior for certain times of the year is compared to behavior for some other year (in the same time frame). There is no guarantee they might be similar. This is the reason the future data at 10% bias is not worth comparing to clean data correlations. When comparing R-square values this might become difficult, due to inconsistency. Due to the above reasons, a method that adapts to changing values is explored in the subsequent chapters. Coming to “aggregation” as a technique, it is a reliable tool that could be used to condense/comprehend the data while not compromising its quality. When it comes to detection of subtle changes like bias, however, it is not a powerful tool. Even at 10% bias, correlations cannot show the difference between biased and clean data sets, and aggregation cannot be helpful due to the averaging effect. Time delay (“lag”) and “matchings” may improve the correlation performance; however, they are not better tools to find biases of 10% magnitude over true reading.

Parametric or non-parametric, linear or non-linear, each method has its own advantages and disadvantages in dimensionality reduction and error detection processes. For instance, linear methods are sensitive to outliers, whereas non-linear or ensemble methods like neural networks can produce “fits” better than linear models, although, over-fitting might be an issue. In this case the user should know where to stop the iterations or how to optimize the process. Principal component analysis is a powerful tool when the data is normally distributed and correlations are linear between the variables. For non-linear systems these are not effective. It is obvious from observing the loadings, variance, and biplots that the differences between clean and noisy data are subtle in case of SAG mill data. Non-linear PCA methods might transform the data to linear form in order to perform the analysis. Kernel PCA and other recent non-linear methods are some improvements to the non-linear PCA methods, which could be useful. These are out of the scope for this research. It is important to know the suitability of a method for a particular data set. In the subsequent chapters, some innovative approaches that are based on signal processing, data-mining, and other combinations of various methods are discussed.

2.8 Conclusions

It is important to study and understand the basic nature, distribution, and other underlying tendencies of data before applying any methodology to achieve the goal of research. In this chapter, sensor raw data (clean) that was collected from a SAG mill operation at Fort Knox mine, Alaska, to observe various key process measurements, was analyzed. From the time domain plots, various sensors were observed to exhibit similar trends in a “state” of operation. Using correlation plots, the data behavior was observed for each state (states 1-3). Developing a Matlab based algorithm, various R-square values for the sensor pairs in each state were captured. It was planned to compare these baseline statistics (BSI) to a noisy dataset (10% bias) at a later stage, however, comparison

of BSI between the states revealed that the differences were dramatic and the subtle differences of biased and clean data sets are not worthwhile for comparison. In other words, when the differences between states are dramatic there is no chance to observe subtle differences caused by bias. A data aggregation technique was applied to observe the trends of data at 30 min, 60 min, 120 min, and 240 min averages. At different delays (“lags”), the correlations between the variables were also observed. Neither of these results showed significant differences in observing state by state difference. This means the SAG mill process is highly dynamic and the sensor errors of 10% magnitude might not be observable.

A principle component analysis was also applied to observe the difference between a noisy (10% biased) data set and clean data set of “feed”. However, the results were not definitive in identifying bias. Therefore it is concluded that descriptive statistics, correlations, aggregation, regression analysis, and PCA are useful tools in reducing the dimensionality of the sensor data when the explanatory and response variables are linearly correlated. Non-parametric methods are useful if the variables are not linearly related. Multivariate analysis and PCA could be useful tools in the fault detection process provided the variables relate to each other linearly. The PCA method is not effective with the data set due to the non-linearity of the relations between variables.

2.9 Acknowledgements

The author expresses his sincere gratitude to his academic advisor Professor Rajive Ganguli, all the members of the graduate advisory committee for their help and guidance, and to the Mining Engineering Research Endowment (MERE) established at the University of Alaska Fairbanks (UAF) for financial support. The author also would like to thank Fort Knox Mine of Kinross Corporation and Pogo Mine of Sumitomo Metal Mining Company for their support of MERE and for providing the necessary data, facilitating site visits, etc. The author extends his gratitude to

College of Engineering and Mines (CEM), and, the Department of Mining and Geological Engineering (MinGeo) at UAF for providing the valuable laboratory and other academic resources to accomplish the project.

2.10 References

Arku, D., and Ganguli, R., 2014, "Investigating utilization of aggregated data: does it compromise information gleaning?" *Mining Engineering*, 2014, Vol. 66, No. 6, pp. 60-65.

Alaska Journal, 2017, "2016 is a milestone year for miners," www.alaskajournal.com/sites/alaskajournal.com/files/lead/11ak-mining-fort-knox-aerial.jpg, Accessed October 2017.

Breiman, L., 1994, "Bagging Predictors," <http://statistics.berkeley.edu/sites/default/files/tech-reports/421.pdf>, pp. 1-19. Accessed December 2015.

Brown, M., 2012, "Data mining techniques," IBM, www.ibm.com/developerworks/library/ba-data-mining-techniques/, Accessed October 2017.

Cho, J., Lee, J., Choi, S., Lee, D., and Lee, I., 2005, "Fault identification for process monitoring using kernel principal component analysis," *Chemical Engineering Science*, Vol. 60, pp. 279 – 288, <https://doi.org/10.1016/j.ces.2004.08.007>.

De'ath, G., and Fabricius, K., 2000, "Classification and regression trees: a powerful yet simple technique for ecological data analysis," *Ecology*, 81(11), pp. 3178–3192.

Fort Knox, 2017, "Fort Knox gallery," <http://fb.kinross.com/operations/operation-fort-knox-alaska-usa.aspx>

- Garcia, R., Perez, C., Rolle, J., and Pazos, A., 2009, “A Sensor FDI Strategy for Safety Critical Systems,” IEEE, <http://ieeexplore.ieee.org/document/5347064/>, Accessed October 2017.
- Hodouin, D., 2010, “Process Observers and Data Reconciliation Using Mass and Energy Balance Equations,” *Advanced Control and Supervision of Mineral Processing Plants*, Springer London, Chapter 2, pp. 15-83.
- Journal of Heuristics, 2017, “Journal of Heuristics – description,” <https://link.springer.com/journal/10732>, Accessed October, 2017.
- Kerschen, G., Boe, P., Golinval, J., and Worden, K., 2004, “Sensor validation using principal component analysis,” *Institute of Physics Publishing Smart Materials and Structures*, Vol. 14, number 1, pp. 36–42, <http://iopscience.iop.org/article/10.1088/0964-1726/14/1/004/meta>. Accessed October 2017.
- Kusiak, A., and Song, Z., 2009, “Sensor fault detection in power plants,” *Journal of Energy Engineering*, December, 2009, No. 127, pp. 127-135, DOI: 10.1061/(ASCE)0733-9402(2009) 135:4 (127).
- Mathworks, 2017, “Least-squares fitting,” www.mathworks.com/help/curvefit/least-squares-fitting.html?requestedDomain=www.mathworks.com, Accessed September 2017.
- Ricardo, D., Qin, S., Edger, F., and McAvoy, T., 1996, “Identification of faulty sensors using principal component analysis,” *Process Systems Engineering, AIChE Journal*, October 1996 Vol. 42, No. 10, pp. 2797-2812.
- Sutton, C., 2005, “*Classification and Regression Trees: Bagging, and Boosting*,” *Handbook of Statistics*, Vol. 24, pp. 303-329.

- Marti, R., and Reinelt, G., 2011, "The linear ordering problem, exact and heuristic methods in combinatorial optimization," Springer-Verlag, Berlin, Heidelberg, DOI: 10.1007/978-3-642-16729-4 2, pp. 17-40.
- Mehranbod, N., Soroush, M., Piovoso, M., and Ogunnaike, B., 2003, "Probabilistic model for sensor fault detection and identification," *AIChE Journal*, July 2003, Vol. 49, No. 7, pp. 1787-1802.
- National Institute of Instrumentation Standards, 2017a, "Introduction to time series analysis," www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm, Accessed October 2017.
- National Institute of Instrumentation Standards, 2017b, "*How can Bayesian Methodology be Used for Reliability Evaluation?*" NIST/SEMATECH e-Handbook of Statistical Methods, www.itl.nist.gov/div898/handbook/apr/section1/apr1a.htm, Accessed October 2017.
- Ni, K., 2008, "Sensor Network Data Faults and their Detection Using Bayesian methods," Ph.D. dissertation, University of California, Los Angeles.
- Oh, W., 2015, "A simple sensor fault detection algorithm," International Conference on Computer Science, Data Mining & Mechanical Engineering, ICCDMME'2015, April 20-21, 2015, Bangkok, Thailand, pp.135-137.
- Sivakumar, S., and Venkatesan, R., 2015, "Meta-heuristic approaches for minimizing error in localization of wireless sensor networks," *Applied Soft Computing Journal*, <http://dx.doi.org/10.1016/j.asoc.2015.05.053>, Accessed October 2017.

- Sharma, A.B., Golubchik L., and Govindan R., 2010, "Sensor Faults: Detection Methods and Prevalence in Real-World Datasets," *ACM Transactions on Sensor Networks*, Vol. 6, No. 3, Article 23, pp. 23.2-23.39.
- Silva, J., Saxena, A., Balaban, E., and Goebel, K., 2012, "A knowledge-based system approach for sensor fault modeling, detection and mitigation," Elsevier, <http://dx.doi.org/10.1016/j.eswa.2012.03.026>, Accessed October 2017.
- Tahane, A., and Mah, R., 1985, "Data reconciliation and gross error detection in chemical process networks," *American Statistical Association, Technometrics*, November 1985, Vol. 27, No. 4, pp. 409-422.
- Varsamou, M., and Antonakopoulos, T., 2009, "A heuristic method for error correction in parallel probe-based storage devices," 6th International Multi-Conference on Systems, Signals and Devices, March 23-26, 2009, Djerba, Tunisia, <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4956774>, Accessed October 2017.
- Yang, C., Liu, C., Zhang, X., and Nepal, S., 2015, "Time efficient approach for detecting errors in big sensor data on cloud," *IEEE Transactions on Parallel and Distributed Systems*, February, 2015, Vol. 26, No. 2, pp. 329-339, DOI: 10.1109/TPDS.2013.2295810.

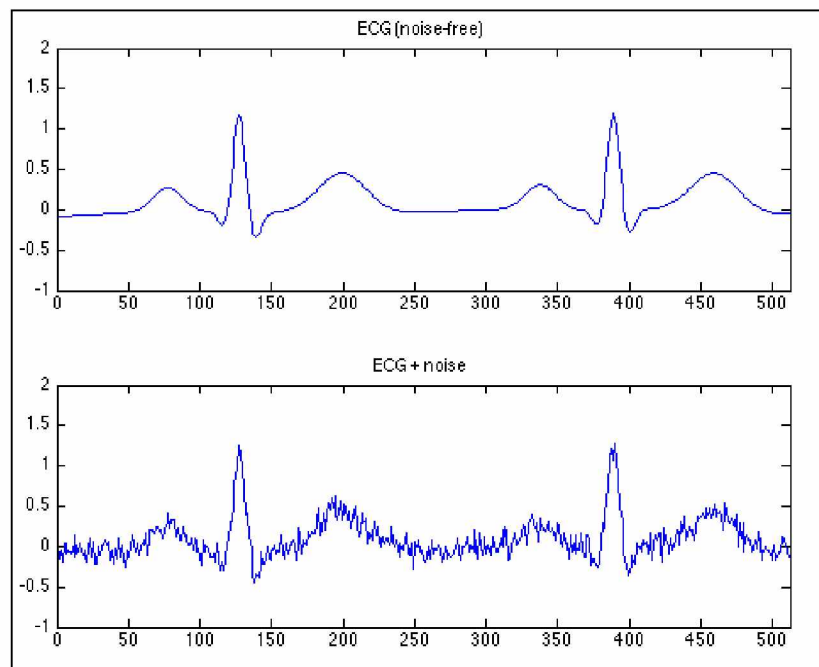
Chapter 3: Signal Processing Methods in the Detection of Industrial Sensor Calibration Errors

3.1 Abstract

In the field of communication systems, a “signal” is something that conveys information about the behavior or attributes of some physical phenomenon (sound, picture, or a sensor reading of temperature). Signal processing techniques have wide variety of applications in many industries in general; digital filtering, spectral analysis, image processing, and process control, to name a few. The techniques are powerful tools in analyzing data due to their ability to transform the data into more manageable forms. Fast Fourier Transform (FFT) is one of those techniques that can convert sensor data (signal) in “time domain” to “frequency domain” where data trends are observed. Sensors became vital part of process monitoring in many industries, including mining and mineral processing industries; however, each year sensor faults are causing the industry millions of dollars in direct (instrumentation) and indirect costs (production loss). For mining industry it is production lost in terms of metal recoveries. Errors of higher magnitude like short faults and failures (gross-errors) can be detected using classical statistical methods; however, their subtle forms like calibration errors are hard to impossible to identify. These are the errors like bias that develop over time. In this research, sensor signals from a carbon stripping circuit (Pogo Mine, Alaska), which plays a key role in recovering gold, were studied. A sensor that monitors the temperature of a strip vessel (S1) was studied using FFT analysis. Ten percent (10%) positive and negative biases were introduced into a clean dataset to create an artificially biased set. Then, the biased and clean sets were transformed into frequency domain using Matlab based FFT tools/algorithms. The goal is to find if bias can be identified through signal processing techniques, like FFT. This chapter describes the methodology, with findings and conclusions.

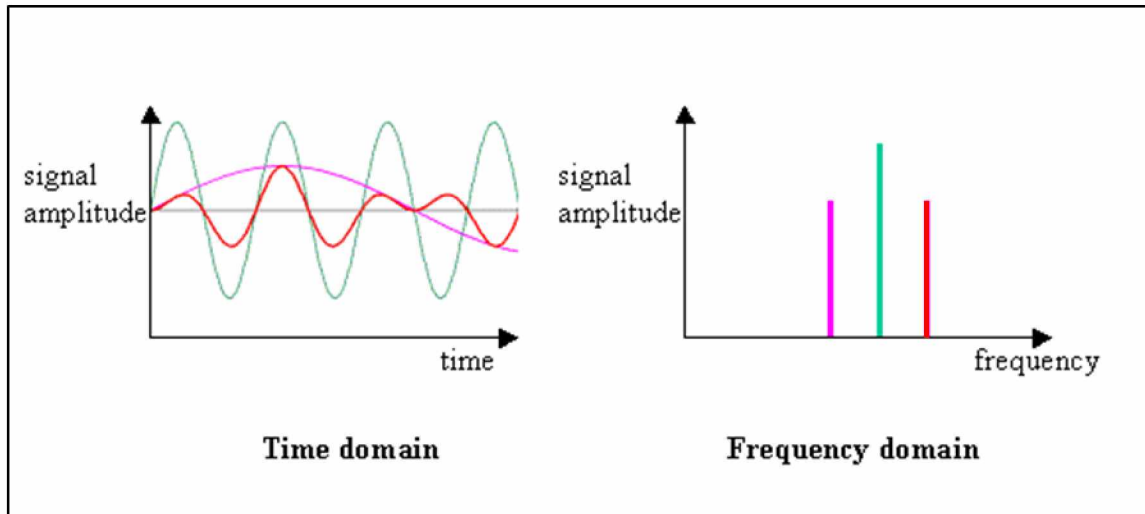
3.2 Introduction

A “signal” in communication systems is something that conveys information about the behavior or attributes of some phenomenon. Some examples are sound, video, and picture. An analog signal is a continuous signal, whereas a digital signal is constructed from the discrete set of waveforms to represent the signal. An analog signal can be converted to digital using an analog to digital converter (ADC). A signal is generally associated with “noise” (Figure 3.1). When the noise is filtered using digital “filters,” it is easier to study the signal. A signal is studied by plotting its behavior against time, i.e., time domain, and against frequency, i.e., frequency domain (Figure 3.2). For instance an EKG signal of a heart when noise is removed (Figure 3.1) can convey essential information about heart behavior with clarity. Fast Fourier Transform (FFT) is a numerical algorithm that is used to convert a signal in time domain to frequency domain, and vice versa.



Source: Adapted from Universite Laval, 2017.

Figure 3.1: Noisy signal.



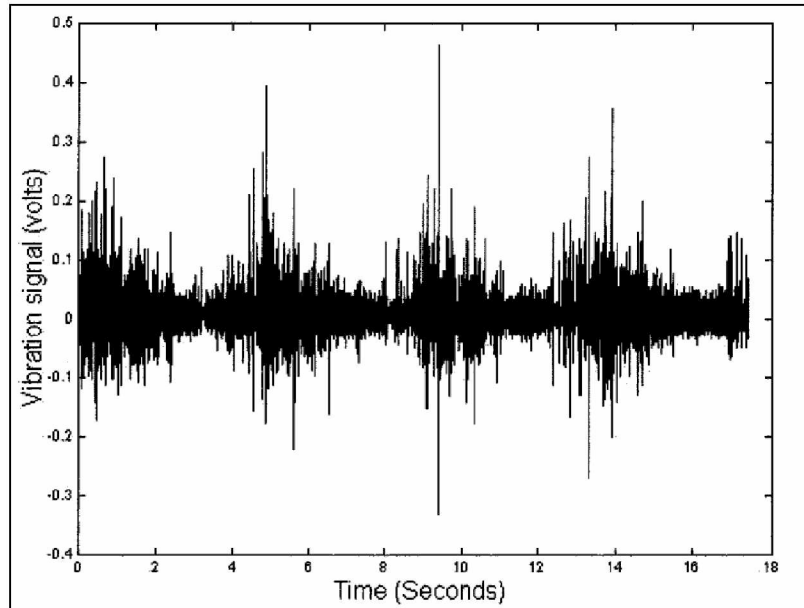
Source: Adapted from National Institute of Instrumentation Standards, 2017.

Figure 3.2: Time domain vs frequency domain.

Digital signal processing (DSP) has a lot of applications in different industries; space, medical, and industrial are some areas. Digital filtering, spectral analysis, speech processing, image processing, and radar processing are some of the area-specific applications. Industrial processing monitoring and control is one of the areas of application. Sensors monitor many physical parameters in industrial processing circuits. Examples include pressure sensors, temperature sensors, pH level sensors, movement or level indicating sensors, etc. Data produced from these sensors are considered as a signals since they convey certain information. Key tendencies of a signal are observed by “processing,” i.e., signal processing. In the presence of bias or error, a signal’s behavior (biased) changes when compared to the signal with no bias (clean). The differences can be observed in the form of changes in the frequency and amplitude. Some industry specific applications are described below.

3.3 Literature Review

A fault isolation algorithm for a light rail vehicle suspension system was developed by Wei et al. (2013), in order to isolate suspension mechanism related faults. In the process, the sensor data from the suspension system were fused and a Kalman filter was applied to obtain the “residuals.” Kalman filter, which is also known as linear quadratic estimation (LQE), is an algorithm that estimates unknown variables with improved accuracy by observing input data with inaccuracies (like noise). The FFT analysis was applied to convert the residual data into frequency domain, then compared to the known fault frequency trends to isolate the fault and determine its type. Tipsuwanporn et al. (2013) developed an FFT-based algorithm to detect faults in a compressor. The paper is aimed at studying compressor current signal with the help of FFT analysis. The frequency domain comparison of a biased compressor current signal with fault free (healthy) signal helped in the detection of the fault. Another mineral industry specific example is the study conducted by Spencer et al. (1999), where SAG mill operational parameters like feed weight, density, rotational speed and power are observed from the vibration signal registered by an accelerometer attached to the mill (Figure 3.3). In another research project, signal processing methods like Hilbert transform and wavelet transform are used to study a SAG mill acoustics signal to identify a characteristic signal that was used to control a coal bunker level (Kang et al., 2006). From the above studies, it is understood that the most essential requirement for a signal to be analyzed successfully with the FFT approach is that the signal should exhibit periodicity. This is a serious drawback when analyzing signals related to industrial processes which tend to be dynamic rather than periodic.



Source: Adapted from Spencer et al., 1999.

Figure 3.3: Vibration signal from a SAG mill operation.

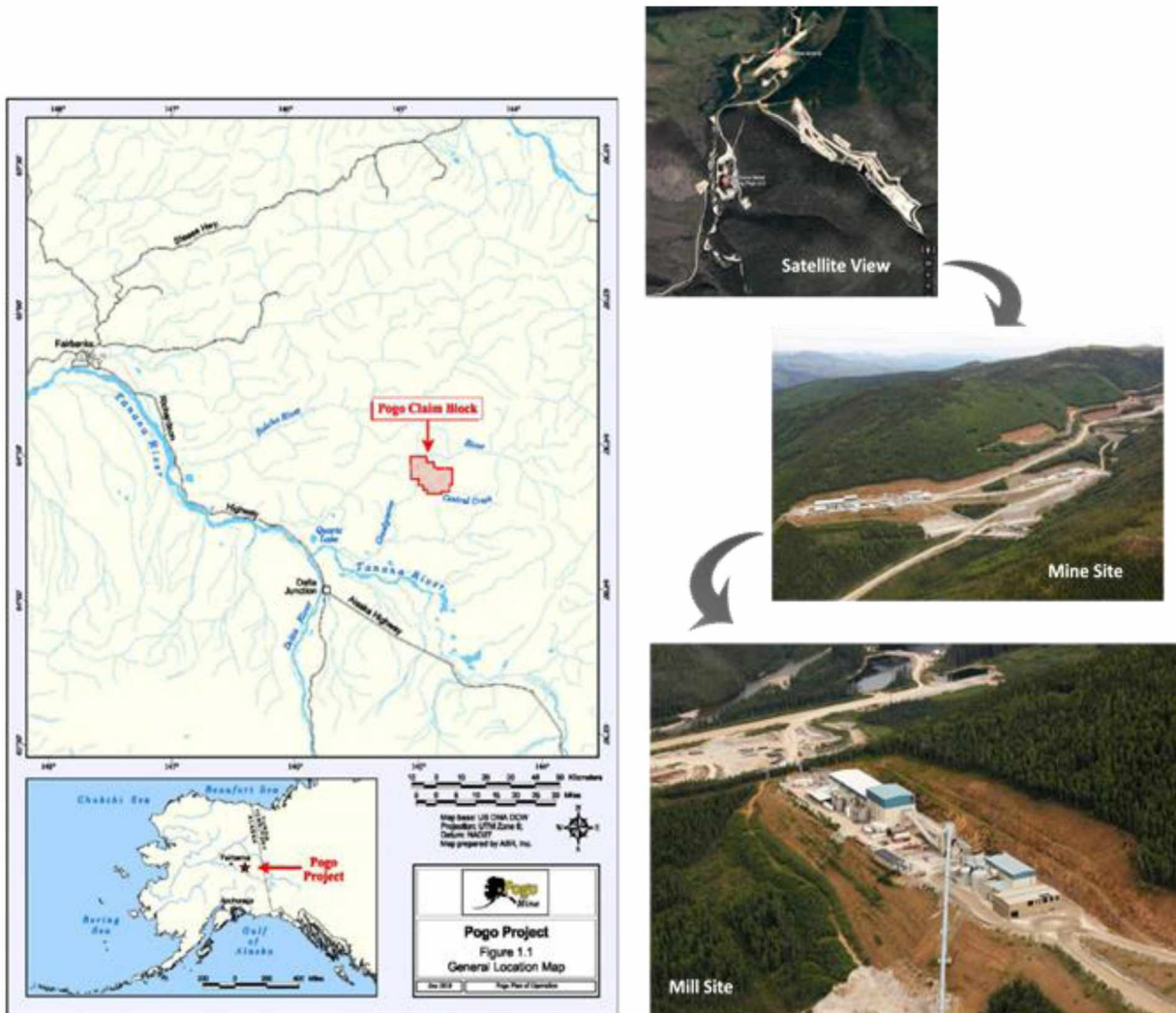
3.4 Introduction to Pogo Mine and Mill Operations

Except for the initial experimentation with the data, which was obtained from Fort Knox mine, the main body of research methodology was developed using the data collected from a carbon stripping circuit at the gold processing facility (mill) of Pogo Mine, Alaska. Due to the reason, the Pogo Mine and its mill operations are described in detail.

3.4.1 Pogo Mine and Mill

Pogo Mine is one of the biggest gold producers in Alaska (Figure 3.4). Located on the Goodpaster River, 38 miles (61 km) north of Delta Junction in east-central Alaska, the nearest city to Pogo is Fairbanks, located approximately 70 miles (112 km) northwest of the mine property (Konigsmann et al., 2017). Pogo is an underground operation. Elevations on the property range from 1,299 ft (396 m) on the Goodpaster River to over 4,003 ft (1,220 m) on the top of Pogo Ridge,

an east-west trending ridge. The climate is classified as sub-Arctic with cold, dry winters and relatively mild summers (Konigsmann, 2017).

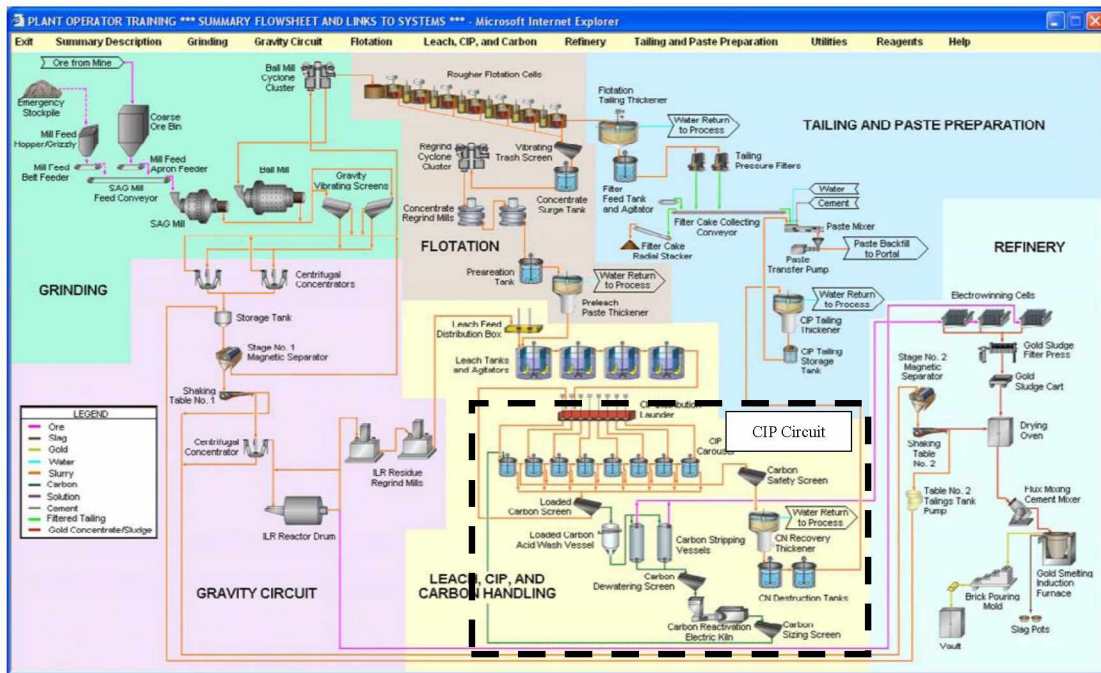


Source: Adapted from Pogo Mine, 2017.

Figure 3.4: Pogo Mine site map.

Pogo Mill processes up to 3,500 tons of ore daily. The process flowsheet is shown in Figure 3.5. The Pogo plant’s process flow mainly is comprised of a crusher, semi-autogenous (SAG) mill, ball mill, flotation circuit, leaching tanks, carbon-in-pulp (CIP) circuit (described in detail in section 3.4.2), stripping circuit, and an electro-winning circuit. The Pogo plant is a closed-circuit

operation; all the water used is recirculated and cleaned before it is released into a pond to minimize environmental impact. Up to 67% of the gold is recovered through the processing facility. A brief description of the process follows (Pogo Mine, 2017). Ore is fed to a conventional SAG/ball mill grinding circuit. Gravity recovery is the technique employed throughout the process. The ball mill circulating load is screened at one (1) mm, and the undersize is fed to two 48-inch centrifugal concentrators operating in parallel. The primary gravity concentrate is then fed to the intensive cyanidation circuit with the leach solutions reporting directly to electro-winning for final gold recovery. Primary gravity tailings are returned to the grinding circuit via the cyclone feed pump-box while the intensive leach residues are reground and pumped into the flotation concentrate leach circuit. Primary gravity gold concentrates are intensively leached on a batch basis; the typical leach residence times are approximately 14 hours. Grinding cyclone overflow reports to a sulphide rougher flotation circuit, which produces a 10% weight concentrate.



Source: Adapted from Pogo Mine, 2016.

Figure 3.5: Pogo mine-gold processing flow sheet.

The concentrate is then reground to 80% passing 10 microns using stirred media detritors prior to being leached in a conventional cyanidation circuit followed by gold recovery in an eight cell carousel CIP circuit. The particles then enter a stripping circuit. A more detailed description of the process can be found in subsequent sections.

3.4.2 Carbon-in-Pulp (CIP) Circuit

The purpose of the CIP process is to allow the gold previously dissolved in the leach tanks to be adsorbed by activated carbon—activated carbon is a form of charcoal that has a large number of low volume pores that can adsorb fine particles.



Source: Adapted from Pogo Mine, 2016.

Figure 3.6: CIP circuit at Pogo Mine.

During the CIP process, the gold particles are slowly adsorbed onto the carbon particles, and eventually will be extracted in the “stripping” process. The CIP circuit is designed to allow

adequate time for the absorption process. The particle-bearing slurry spends 30 minutes in each CIP tank. The Pogo CIP circuit has eight such tanks (Figure 3.6). The slurry spends four hours in the circuit, and approximately 300 to 600 ounces of gold is adsorbed per ton of activated carbon used.

3.4.3 Carbon Stripping Circuit

The Pogo stripping circuit in pictures is shown in Figure 3.7. The circuit consists of two strip vessels that work in tandem. While one vessel (vessel-1) is being loaded with gold-bearing activated carbon, the other previously loaded vessel (vessel-2) is operated by circulating a solution called ‘elute’ at approximately 280°F and 65 PSIG to liberate gold particles. An elute is a water-based solution with 1% sodium hydroxide and 0.1% sodium cyanide (Fast, 2016). The process is called “pressurized Zadra stripping.” A typical pressurized Zadra stripping cycle lasts for 11 hours and consists of the following stages: loading the vessel (1 hr), circulating elution (8 hrs), carbon cooling (1 hr), and unloading carbon from vessel ($\frac{1}{2}$ hr) (Table 3.1).

Table 3.1: Operating schedule-pressure Zadra stripping.

Operation	Solution	Time
Load Column	Transfer Water	90 minutes
Elution	0.1% NaCN, 1% NaOH	480 minutes
Carbon Cooling	Fresh Water	60 minutes
Unload Column	Transfer Water	30 minutes
TOTAL		11 hours

Source: Fast, 2016.

While the used carbon is discharged, the “pregnant leach” solution is pumped out. The same process is repeated with strip vessel-2. The pregnant leach solution (PLS), on its way out from the strip vessel, is cooled through exchanging heat with no. 3 and 4 heat exchangers. When the PLS

reaches the electro-winning circuit, the gold particles are removed and the solution, now called “barren solution,” is reheated by a boiler with the aid of heat exchangers 1 and 2 and will be recirculated through strip vessels. A glycol solution is circulated between the boiler and heat exchangers as a medium of heat exchange. Sensors are strategically placed at various parts of the circuit to measure temperatures, flow rates, etc. (Figure 3.8). If any of the sensors are biased, the circuit cannot be managed at an optimum level in terms of temperature and flows. This results in poor gold recoveries. Finding and fixing faulty sensors helps improve recoveries, which is the motivation for this research. It is very important to maintain temperatures in the strip vessels at certain levels (270-280°F) for certain periods of time to maximize gold separation. In this context, monitoring the temperatures becomes crucial. Sensors S1 and S2 are two important ones in this context, which will be analyzed in this chapter. The sensors are given code names for simplicity, which are detailed in Figure 3.8.

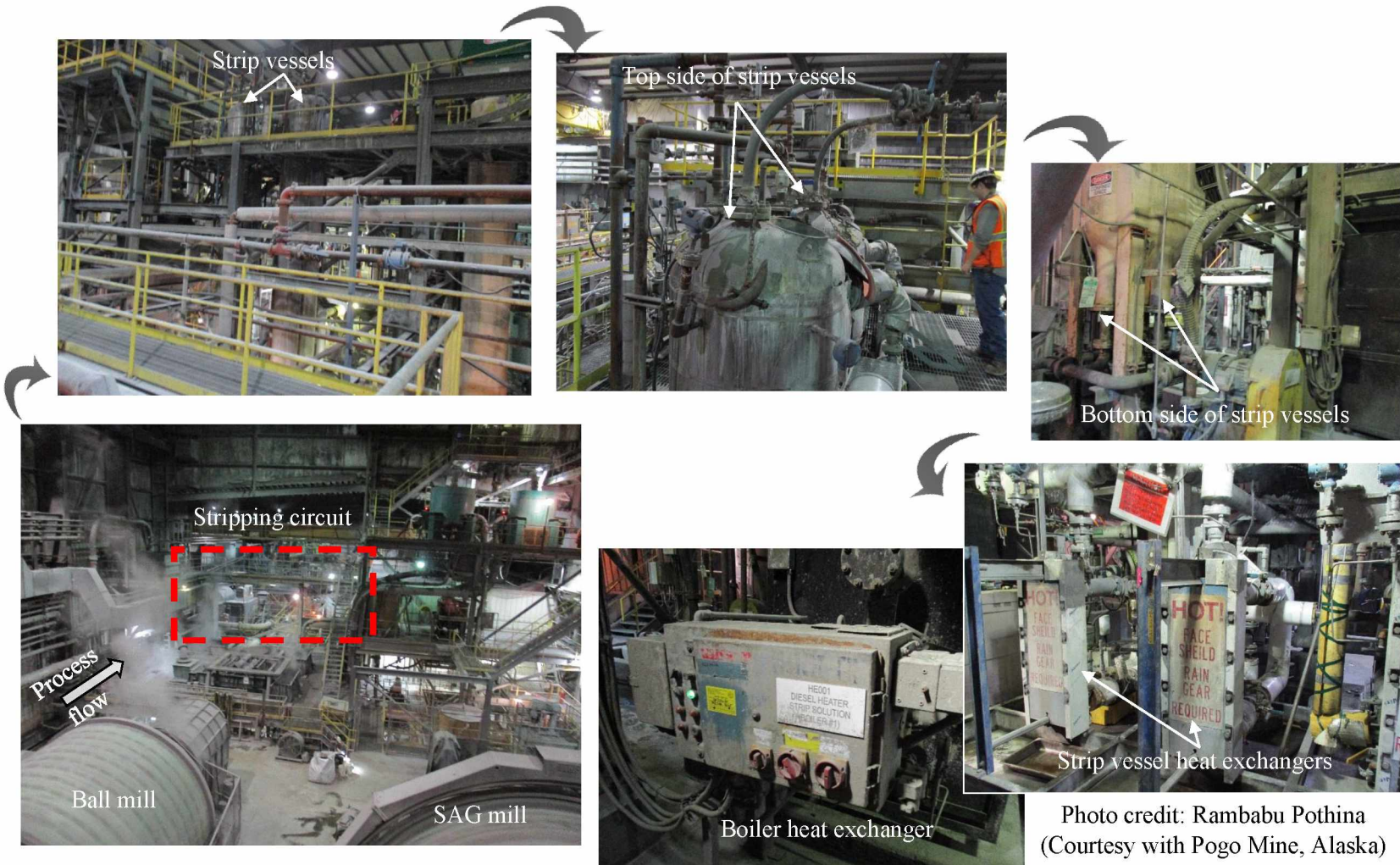
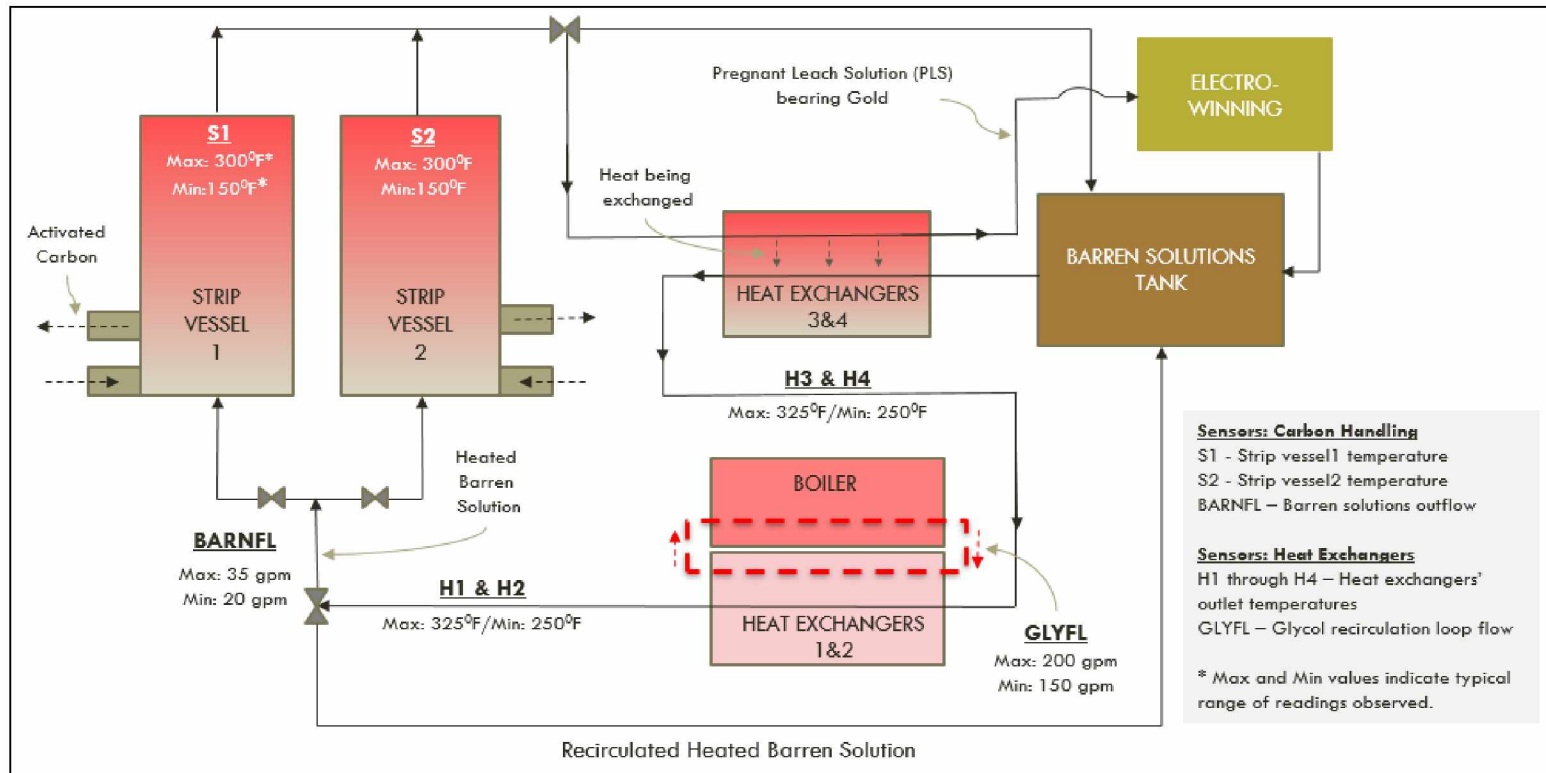


Photo credit: Rambabu Pothina (Courtesy with Pogo Mine, Alaska)

Figure 3.7: Pogo stripping circuit in pictures.



Source: Pogo Mine, 2016.

Figure 3.8: Pogo stripping circuit schematic diagram with sensor placements.

3.5 Methods and Materials

3.5.1 Data Collection

The raw sensor data was collected at 10-min average intervals from the Pogo mill database. Various sensors of interest and their data readings can be seen in the Table 3.2. The data was collected for a period of nine (9) months; Jan 1, 2015 through September 31, 2015. The visual format of various sensor raw data streams with close-up views can be seen in (Figures 3.8 through 3.13). It should be noted that sensors H3 and H4 are not part of the study in this chapter. Various descriptive statistics for the sensor data are given in Table 3.3. The “data cleansing cut-off value ($Th_{cleanse}$)” column in the Table 3.3 refers to the cut-off value based on which undesired and corrupted data were removed (cleaned), which is explained in the data preparation section.

Table 3.2: A snapshot of raw sensor data collected at 10-min average intervals.

Record #	Time	Strip vessel sensors		Barren flow sensor	Heat exchanger sensors		Glycol flow sensor
		S1 (°F)	S2 (°F)	BARNFL (GPM)	H1 (°F)	H2 (°F)	GLYFL (GPM)
1	1/1/15 12:00 AM	90.5	261.6	35.4	299.4	300.7	204.0
2	1/1/15 12:10 AM	90.4	261.6	35.4	291.7	293.0	203.0
3	1/1/15 12:20 AM	90.3	261.6	35.3	284.3	285.4	200.0
4	1/1/15 12:30 AM	90.1	233.9	35.3	277.3	278.3	201.1
5	1/1/15 12:40 AM	90.0	219.2	35.2	270.5	271.5	202.3
6	1/1/15 12:50 AM	89.8	216.9	35.2	264.1	265.1	204.6
7	1/1/15 1:00 AM	89.7	215.3	35.2	258.0	258.8	204.4
8	1/1/15 1:10 AM	89.6	204.2	35.1	252.1	252.9	205.0
9	1/1/15 1:20 AM	89.4	192.4	35.1	246.5	247.3	201.9
10	1/1/15 1:30 AM	89.3	189.8	35.0	241.3	242.0	201.0

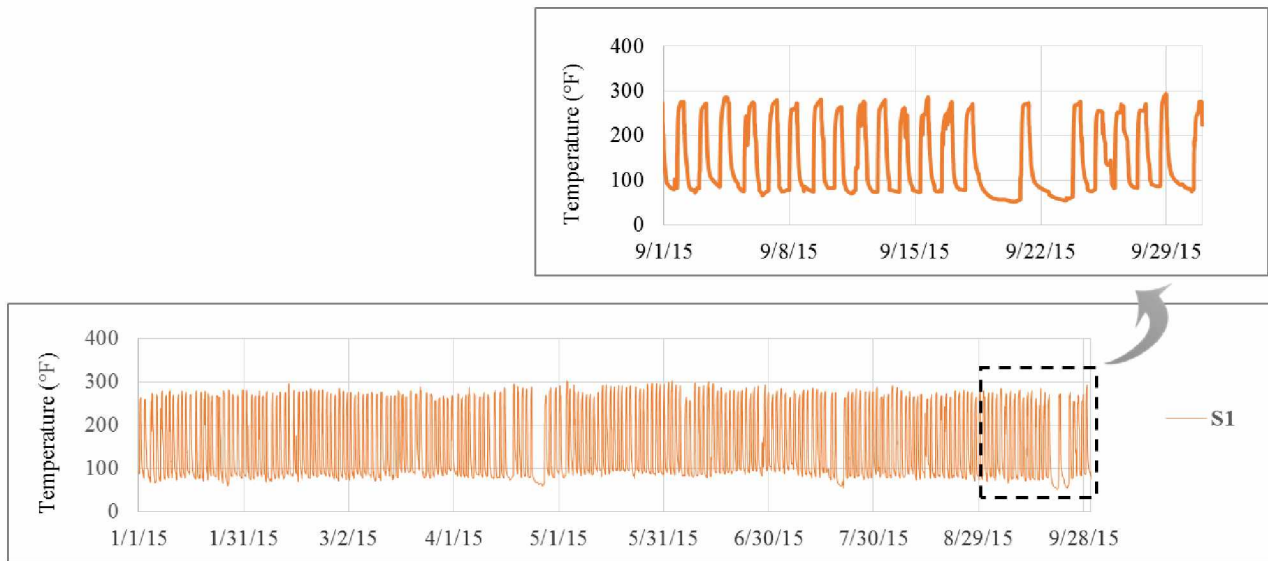


Figure 3.9: Strip vessel-1 heat sensor (S1).

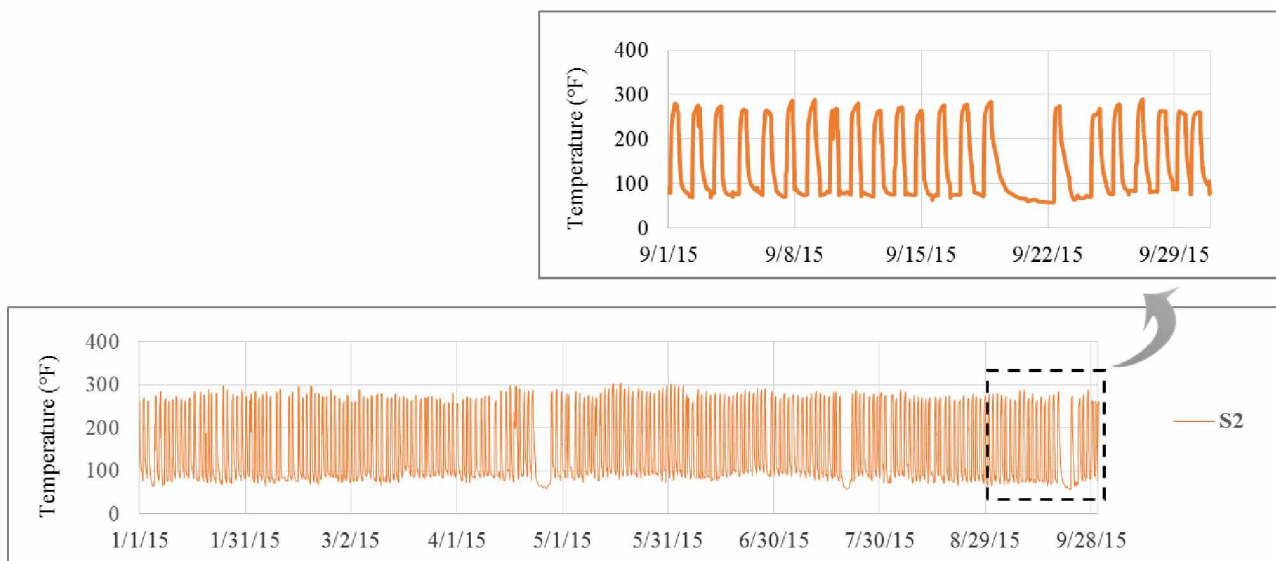


Figure 3.10: Strip vessel-2 heat sensor (S2).

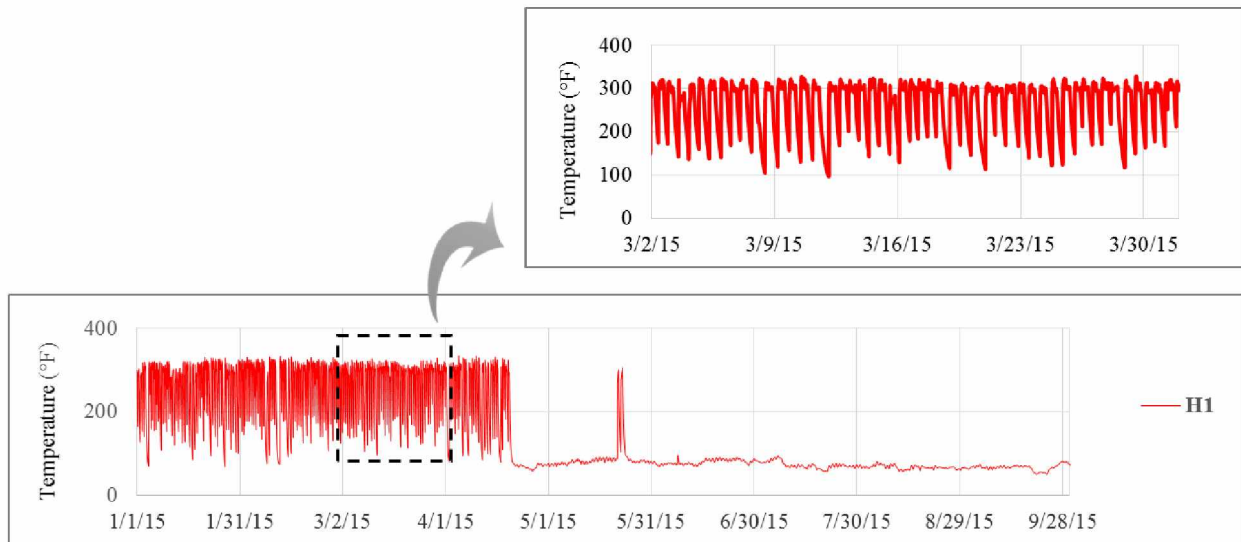


Figure 3.11: Heat Exchanger-1 Sensor (H1).

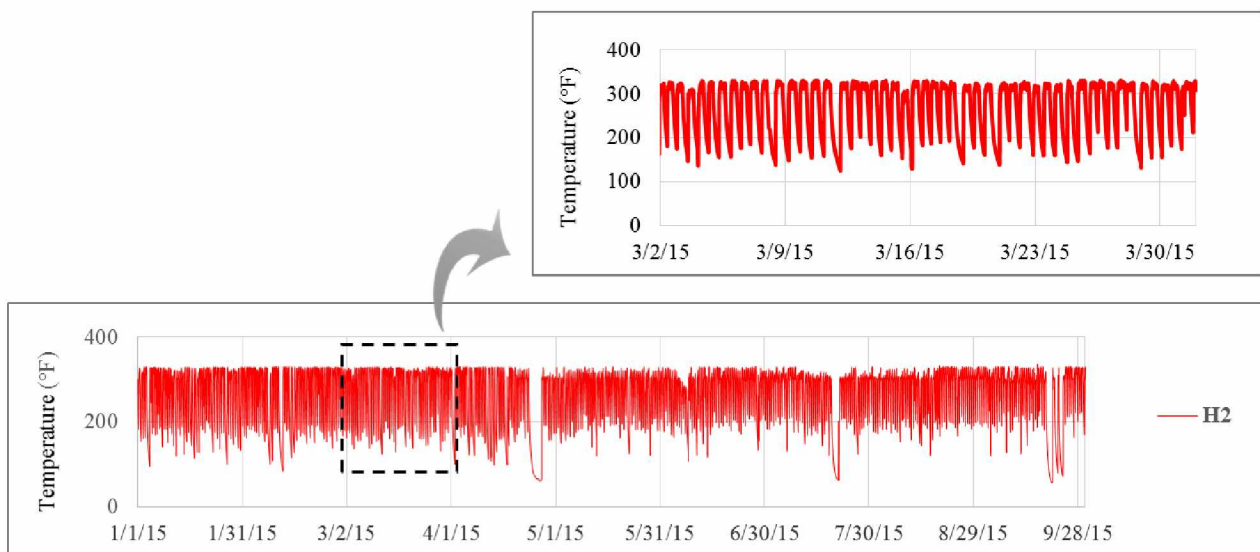


Figure 3.12: Heat Exchanger-2 Sensor (H2).

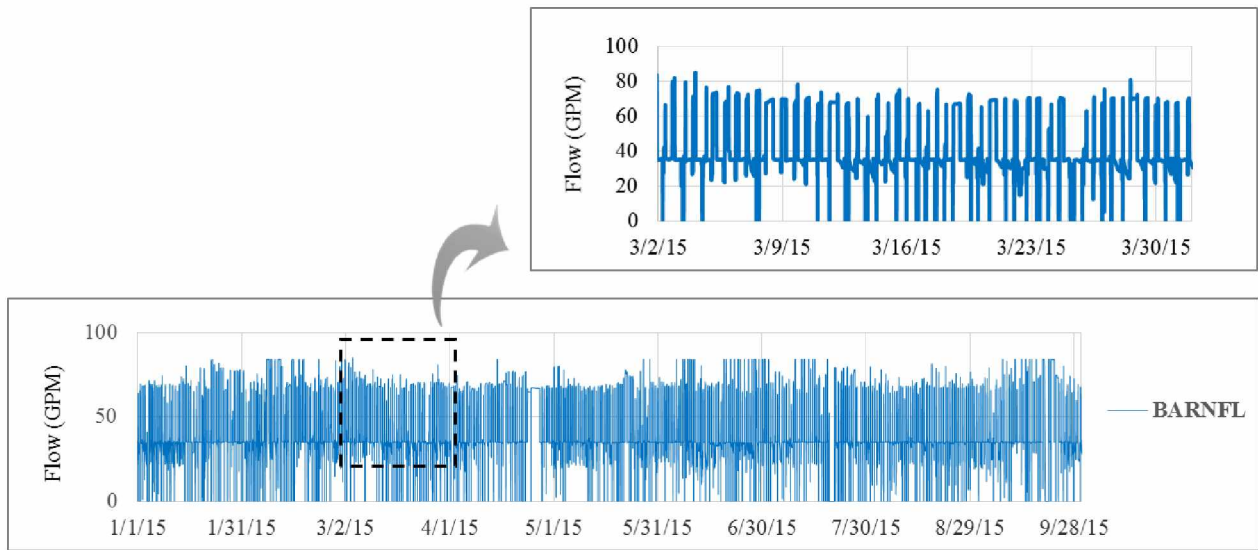


Figure 3.13: Barren flow Sensor (BARNFL).

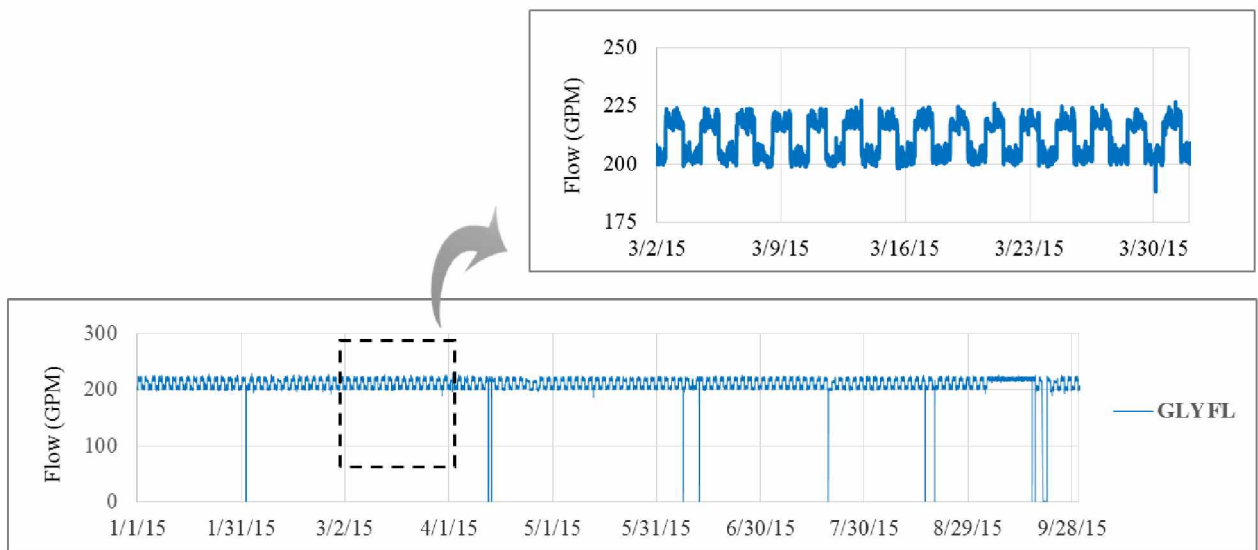


Figure 3.14: Glycol flow sensor (GLYFL).

Table 3.3: Descriptive statistics for the sensor data collected.

Sensor	Measuring parameter/variable	Unit	Maximum	Minimum	Mean	Median	Standard deviation	Data cleansing cut-off value ($Th_{cleanse}$)
S1	Strip vessel-1 temperature	°F	302.7	52.1	150.1	104.9	78.3	150
S2	Strip vessel-2 temperature	°F	304.3	57.7	148.8	105.2	77.6	150
H1	Heat exchanger-1 outlet temperature	°F	334.0	50.0	144.9	82.7	96.7	250
H2	Heat exchanger-2 outlet temperature	°F	336.2	56.3	263.1	296.0	63.6	250
BARNFL	Barren solution flow	GPM	302.7	0.0	40.4	35.0	17.9	20
GLYFL	Glycol flow	GPM	229.7	0.0	209.8	214.6	18.2	150

3.5.2 Assumptions and scope

Calibration errors are present in the data often in the form of an offset value (bias) added to the original or true reading; see Equation (3.1). Bias is the common error associated with sensors and often a hard one to identify. Other common errors like gross errors are out of the scope of this research. Gross errors are dramatically high or low errors, which are easily detected through common statistical techniques. For the purpose of this experimentation, it is assumed that the data collected from the Pogo mine is devoid of errors, i.e., the set is deemed a ‘clean set.’

$$\text{Observed or biased reading, } x_{bias} = \text{True reading } (x_{true}) + \text{bias } (e^{bias}) \quad (3.1)$$

Bias is expressed as a percentage over the true reading value. For the purpose of experimentation, bias is artificially induced in the clean data set. A +2% bias indicates 2% of the true reading added as bias to the true reading itself; likewise -2% expresses adding -2% of the true reading; see Equation (3.1). Calibration errors are subtle and can be as low as 2% over the true reading. Identifying the bias at such low magnitudes is preferred by the industries, specifically

Pogo Mine, which sees the value in it. Only S1 and S2 sensors were analyzed with the methods described, due to the important role of these sensors in monitoring the temperature in the crucial process of stripping, i.e., gold stripping from activated carbon. Only one sensor with bias at a time was analyzed in this study, using other sensors' interrelations. Identifying the bias in multiple sensors at a time in a multi-sensor environment is out of the scope for this research. Due to the space constraints and to avoid redundancy, the results from S1 analysis and validation are only presented in this chapter; S2 analysis and validation also demonstrated similar trends.

3.5.3 Data Preparation

3.5.3.1 Programming and Software

The sheer volume of data collected, the subsequent mathematical calculations/repetitions to be performed, statistical analyses to be conducted, and graphically presentation of the results, demands for the usage of a mathematically intensive software package. Matlab was chosen to meet these needs. Matlab has an excellent signal processing tool box that has functionality to perform various operations, like FFT, inverse FFT, etc. Matlab also has an excellent plotting capability that was utilized for the research. A positive (+2%) and a negative (-2%) bias were introduced in the data, only to identify them using signal processing methods described in the subsequent sections. The algorithm description session has the detailed account on the process.

3.5.3.2 Algorithm Description

The algorithm was developed using Matlab's "signal processing toolbox," which has a variety of Fourier Transform related functions. The function '*fft*' can perform Fourier Transform, whereas

'*ifft*' can perform the inverse of the operation. A flowchart of the algorithm is shown in Figure 3.15.

Step 1: The algorithm reads all the sensor data (clean and biased sets for a random week) and other necessary inputs like sampling frequency, filtering frequency, etc. A biased data set in this context, should be created by inducing noise in the middle of the week.

Step 2: The data is plotted in the time domain for visual examination. Then the data is converted into the frequency domain using FFT tools. Plotting the data in frequency domain provides the opportunity to summarize the whole dataset into few frequencies; examination for noise becomes easier in this format. Based on the frequencies related to noise, the filters' "cut-off" frequencies are chosen.

Step 3: Digital filtering is applied to remove the noise from the biased data using low-pass filters at the chosen cut-off values.

Step 4: Biased and clean data are plotted in frequency domain to observe and detect the presence of bias. The data from frequency domain is then converted back to time domain.

Step 5: The data is reexamined in time domain. This is when effects of filtering on biased data can be observed. Filtering helps attenuates the data set for undesired noise.

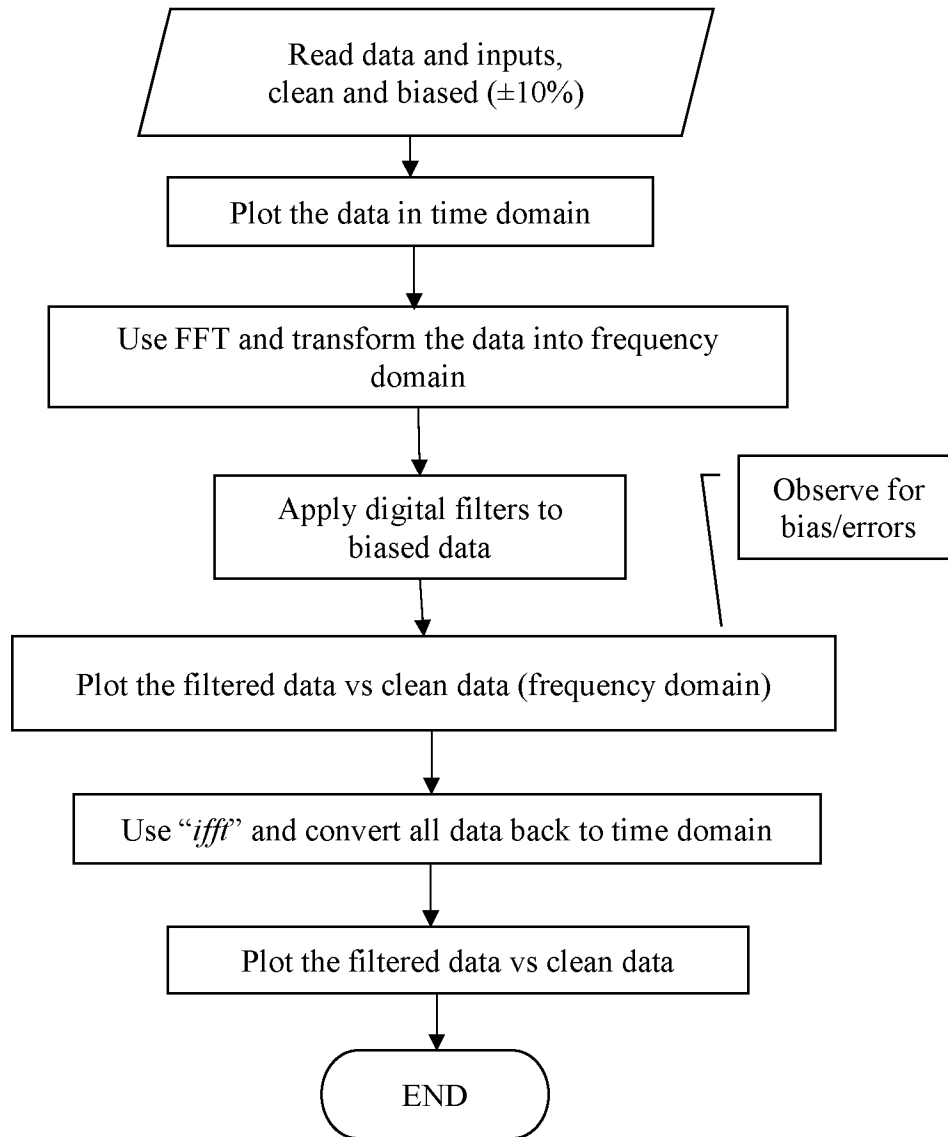


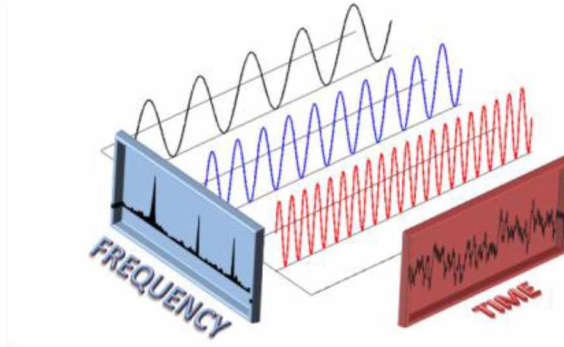
Figure 3.15: FFT analysis algorithm.

3.5.4 Application of Fast Fourier Transform (Clean vs Biased data)

3.5.4.1 Fast Fourier Transform (Theoretical Background)

The Fourier Transform shows that any signal or waveform in time domain can be re-written as the sum of sinusoidal functions in frequency domain (Figure 3.16). The Fourier transform

decomposes a function of time (a signal) into the constituent frequencies. Fourier Transform can also represent the wave as a mathematical expression (Fouriertransform.com, 2017).



Source: Adapted form Bocconi Students Investment Club, 2017.

Figure 3.16: Time Domain vs frequency domain.

The Fourier Transform of a time domain function $g(t)$ is defined by:

$$\mathcal{F}\{g(t)\} = G(f) = \int_{-\infty}^{\infty} g(t)e^{-2\pi ift} dt \quad (3.2)$$

Where

f is frequency,

i is an imaginary number,

t is time, and

$G(f)$ is the frequency or spectrum function.

The time domain function $g(t)$ can also be obtained by inverse Fourier Transform of frequency function, $G(f)$. The time domain function and its frequency domain transform is called a “Fourier Transform Pair.”

$$\mathcal{F}^{-1}\{G(f)\} = \int_{-\infty}^{\infty} G(f)e^{2\pi ift} df = g(t) \quad (3.3)$$

Even though periodicity of the sensor data over a nine-month span is quite rare, specifically in mineral processing circuits, a week's worth of data usually exhibits a certain level of periodicity (Figure 3.17). Due to this reason, the periodicity of a typical week's worth of data was analyzed. A positive and negative 10% randomized error was introduced into the S1 data to create a "biased set." Using FFT, the data was transformed to frequency domain. The raw data was aggregated to 5 min intervals before it was input into the algorithm. A sampling time (T_s) of 300 seconds (five minutes) was chosen due to the fact that each reading is the average of data observed for five minutes.

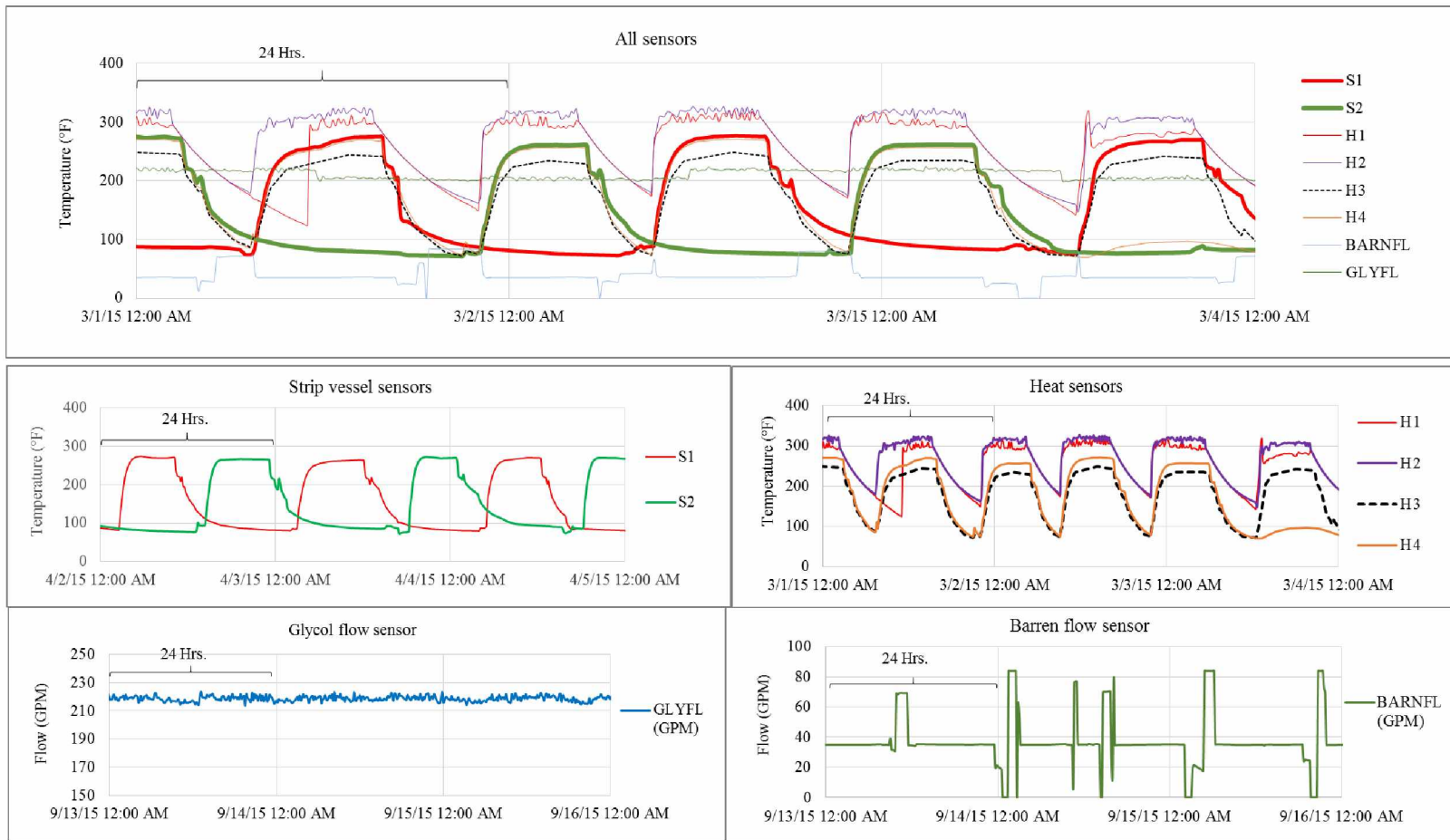


Figure 3.17: Periodicity of sensor data measurements from Pogo stripping circuit.

3.5.4.2 Digital Signal Processing and Digital Filtering

A “digital signal” is a discrete form of the original electrical signal or data reading from a sensor. In contrast, “analog” signal represents a continuous form. In this context, any processing performed on a digital signal, like transformation (FFT) or noise filtering (simply “filtering”), is called digital signal processing (DSP). A filter in this context should be understood as a “digital filter” that uses mathematical formulae or logic to isolate noise from a digital signal. The digital signal in the form of a “text file” or any other accepted format was read using the Matlab algorithm developed and the digital filters were applied through the programming tools (specifically, signal processing tool box).

Two low-pass filters (band-1 and band-2) of different cut-off frequencies (f_c) were applied to the bias-free data set to remove unwanted noise from the data. The cut-off frequencies for band-1 and band-2 are 0.0001 Hz, 0.0003 Hz, respectively (1 Hz = 1 cycle/s). A low-pass filter (LPF) is a digital filter that passes frequencies lower than a certain preset cut-off value and attenuates the signal higher than the cut-off value. The choice of these filters are based on the observation of the data in frequency domain. When the data set is converted to frequency domain, the unwanted frequencies can be identified. Cleaning the data for unwanted noise improves the quality of the data. The effect of filtering on a “feed” sensor data set from the Fort Knox mine’s SAG mill operation is shown in Figures 3.17 and 3.18. Obviously, the data after filtering is more refined and reliable for future use. After applying these low-pass filters, the biased data set created for stripping circuit sensor (S1) of Pogo mill—with induced negative and positive 10% bias—was compared to its clean set. The results are presented in section 3.6. The following figure shows filtering effects on noise of a SAG mill “feed” sensor.

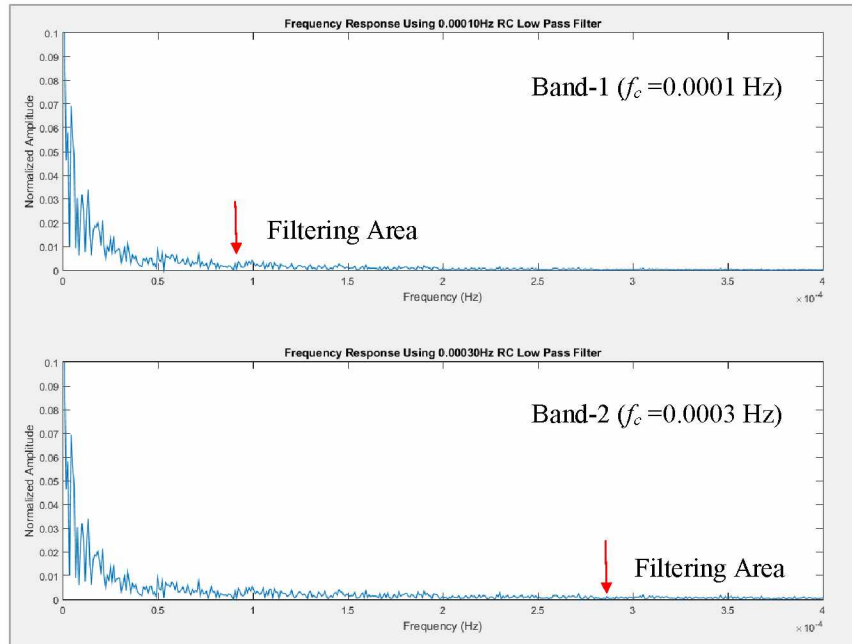


Figure 3.18: Frequency responses of filters on “feed” data.

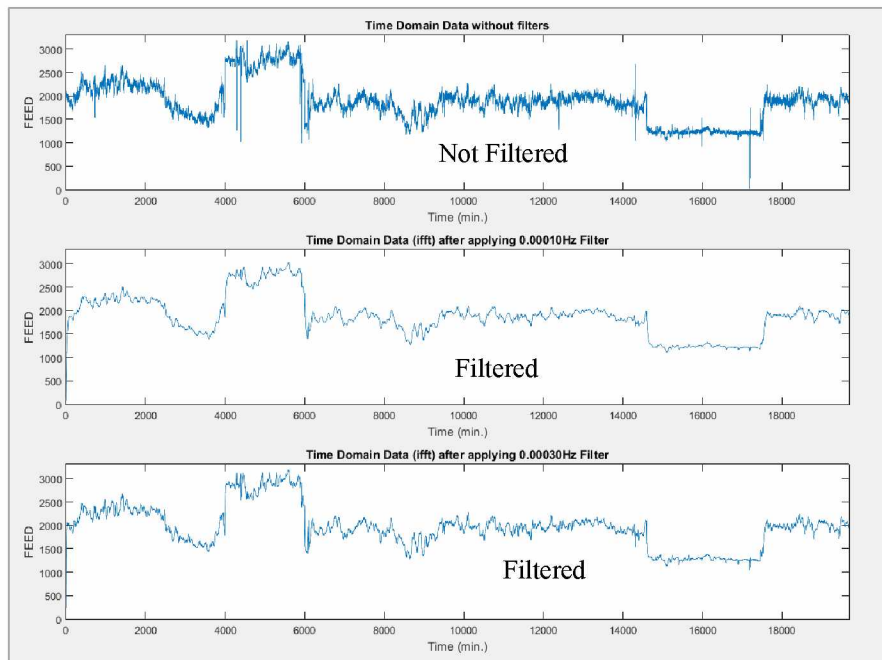


Figure 3.19: “Feed” data re-transferred (*ifft*) to time domain after filtering.

3.6 Results

The following figures show the effect of +10% bias that was induced in the middle of a random week's worth of data (S1 sensor). For viewing convenience, the time domain data in seconds is converted to hours (Figure 3.20). The time domain observations reveal that there is an observable difference around the 80th hour (or after three days), due to the presence of bias. In order to know if it is observable in frequency domain, the data was converted to frequency domain. This is an important aspect to observe, since the time domain difference for a week is easy to notice; however, in a span of a year, the frequencies of sensor data cycles differ a lot for mineral processing operations. If the difference in frequencies is not high in magnitude, it is difficult to observe.

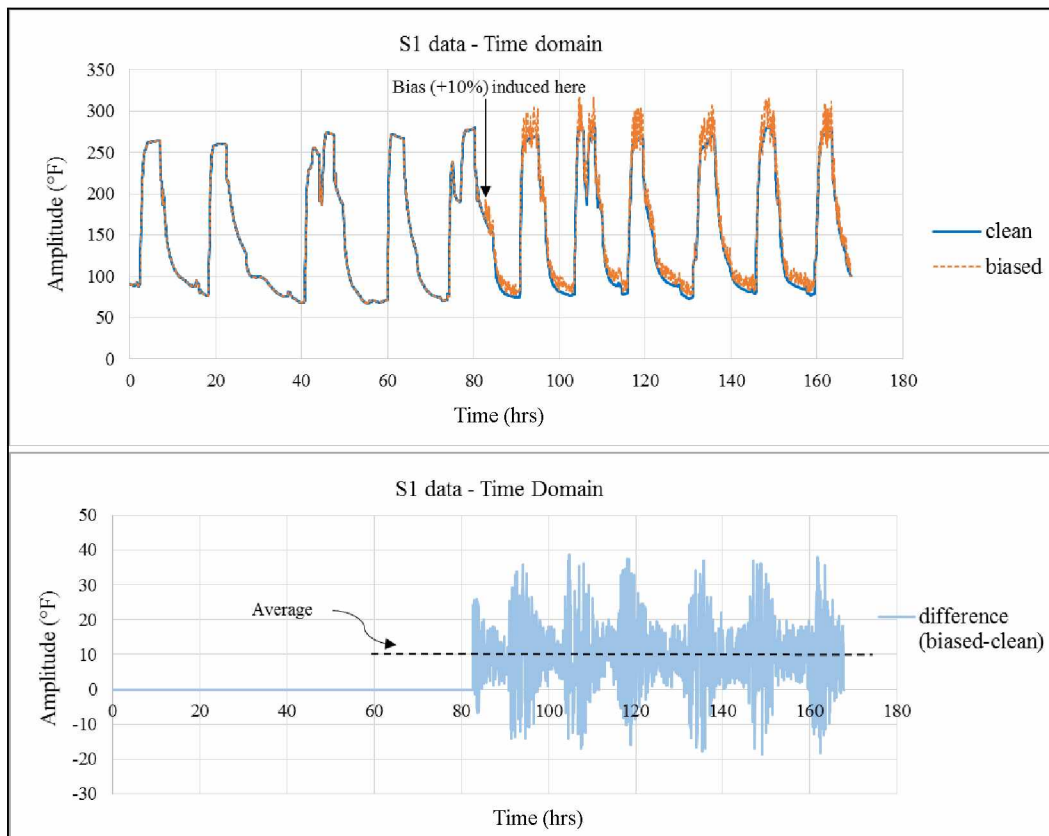


Figure 3.20: Time domain data comparison, clean vs. biased (+10%).

It should be noticed that when the data in time domain is converted to frequency domain, it is normalized, with the amplitudes and frequencies falling between the values of 0 and 1. In frequency domain (Figure 3.21), it appears that the “difference” is very subtle. Several other random weeks of observations, also provided the same result. An error of 10% is very high in magnitude—Pogo mine prefers an error/bias as low as 2% over the true reading—and if it cannot be observed clearly, FFT could be disadvantageous for the stripping circuit data. Results from the 10% negative bias analysis are plotted in Figures 3.21 and 3.22.

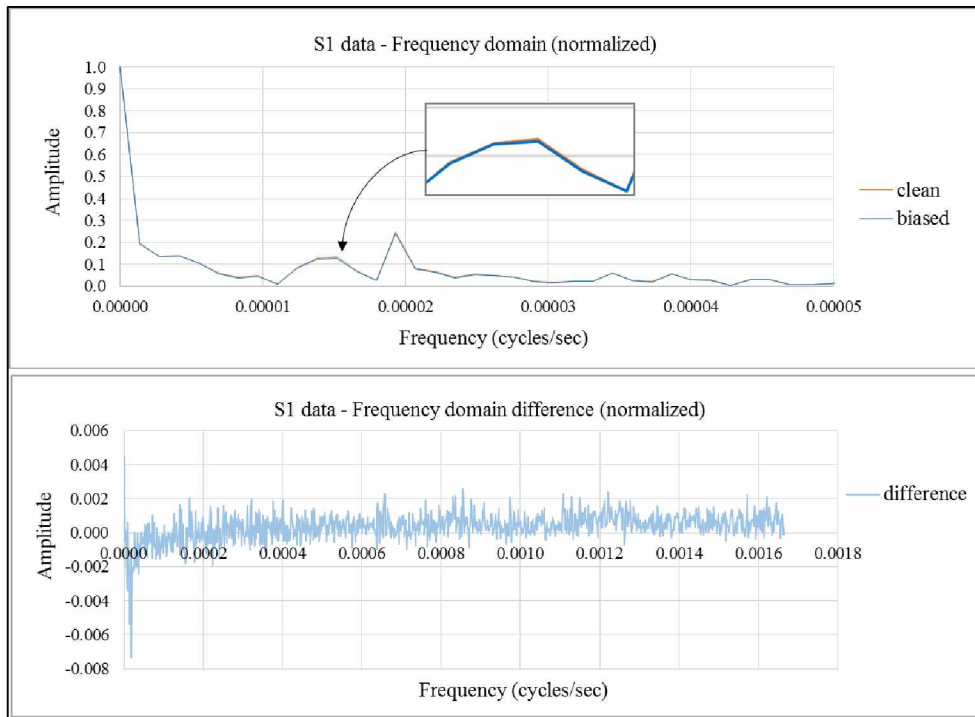


Figure 3.21: Frequency domain data comparison, clean vs. biased (+10%).

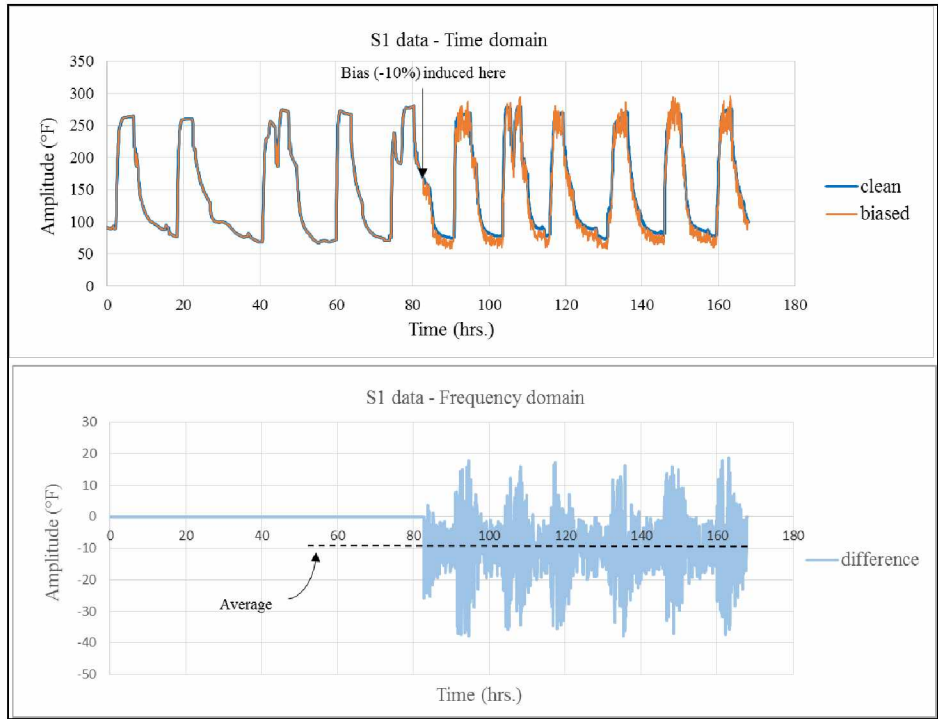


Figure 3.22: Time domain data comparison, clean vs. biased (-10%).

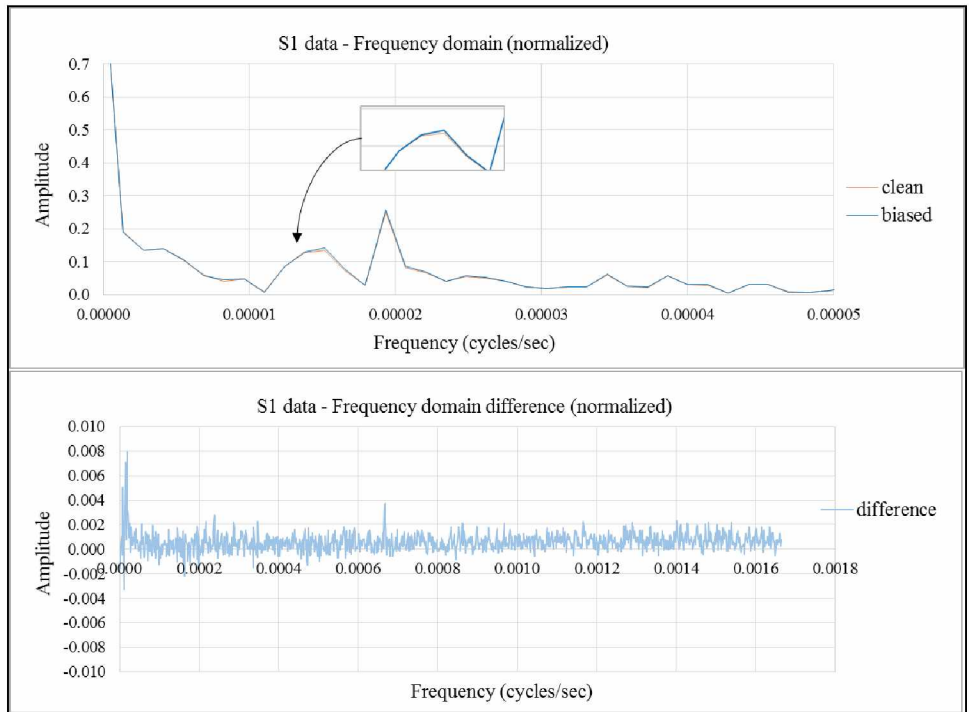


Figure 3.23: Frequency domain data comparison, clean vs. biased (-10%).

3.7 Discussion

From the results it is observed that conversion of sensor readings from time domain to frequency domain gives the opportunity to observe the data trends in manageable form. This is an excellent way to comprehend the data behavior from long periods of time (Pogo strip circuit data set is nine-months long). Observation of sensor data in frequency domain also assists in choosing proper digital filters for eliminating or attenuating noise. In this context, filtering data for noise can be considered as an excellent tool to improve data quality and reliability. Coming to the bias observation, unfortunately, the data sets at +10% and -10% bias when compared to clean sets did not register major differences; specifically, when observed in frequency domain. The difference is obvious in time domain, but is not noticeable in frequency domain.

Unfortunately, the week-by-week analysis of the strip circuit temperature sensor (S1) data did not yield consistent results. This is due to the fact that the data did not exhibit periodicity in terms of frequency and amplitude, which are the most essential requirements for FFT analysis. Revisiting the close-up windows from Figures 3.8 and 3.9 reveals that the periodicity of S1 and S2 readings is not consistent. The close-up view of heat sensor between the months April-May 2015 reveals the same (Figure 3.24). FFT method might be suitable for cyclical data, for instance, the acoustic signals coming from the rotating SAG mill. In this case, due to the continuous rotation of the SAG mill, the acoustics register a periodic signal. If the goal is to observe the signal coming from steel balls, it can be observed as a unique frequency in the frequency domain.

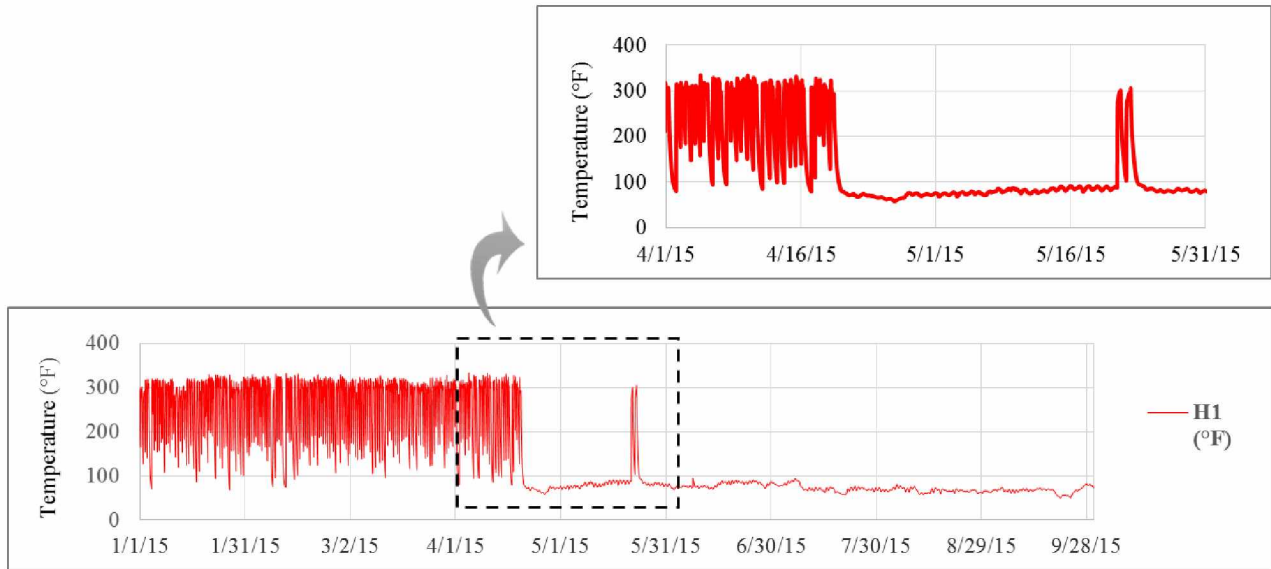


Figure 3.24: Heat sensor-H1 at Pogo stripping circuit.

A study conducted by Spencer et al. (1999) is one example where SAG mill operational parameters like feed weight, density, rotational speed, and power are observed from the vibration signal registered by an accelerometer attached to the mill (Figure 3.3). It can also be observed from the figure how periodic the signal is for four revolutions of the mill. In another study, signal processing methods like Hilbert transform and wavelet transform were used to study a SAG mill acoustics signal to identify a characteristic signal that was used to control a coal bunker level (Kang et al., 2006). In these studies the signals exhibited high periodicity whereas in the case of the stripping circuit, unfortunately, the characteristic is not consistent. This proved disadvantageous for FFT application.

3.8 Conclusions

Signal processing techniques have applications in many industries in general. The techniques are powerful tools for analyzing data due to their ability to transform data into more manageable forms. FFT is one of those tools that is capable of converting any given sensor data set in time

domain (signal) into few observable frequencies (frequency domain). Signal processing methods are very effective tools where data exhibit periodicity and stationarity. Unfortunately, the tools are less effective where the cyclical frequency of sensor data changes very often. In the case of industrial process circuits and parameters, the periodicity assumption is violated due to their dynamic nature. For this reason, the FFT methods are less effective in observing the data trends and bias over longer periods of time. A similar conclusion was reached through the examination of a temperature sensor data (S1) in a strip circuit at the Pogo Mine, Alaska. Week-by-week observation of the data in time domain and frequency domain revealed that the data did not exhibit periodicity to the extent required by signal processing methods (FFT). The goal of the research was to find if the artificially induced bias of 10% magnitude (positive or negative) was observable compared to a clean set of data. The difference observed for S1 data stream was too subtle. Hence, it is concluded that the FFT tools are less effective in achieving the task in this particular case. The methods could be effective to some extent for data observed for shorter periods of time that exhibit consistent periodicity. The only advantage of the application of FFT methods for highly dynamic processes such as in mineral processing circuits is in refining (filtering) the data to more reliable forms by the reduction of noise. Digital filtering is a good tool in that respect.

3.9 Acknowledgements

The author expresses his sincere gratitude to his academic advisor Professor Rajive Ganguli, all the members of the graduate advisory committee for their help and guidance, and to the Mining Engineering Research Endowment (MERE) established at the University of Alaska Fairbanks (UAF) for financial support. The author also would like to thank Fort Knox Mine of Kinross Corporation and Pogo Mine of Sumitomo Metal Mining Company for their support of MERE and for providing the necessary data, facilitating site visits, etc. The author extends his gratitude to

College of Engineering and Mines (CEM), and, the Department of Mining and Geological Engineering (MinGeo) at UAF for providing the valuable laboratory and other academic resources to accomplish the project.

3.10 References

Bocconi Students Investment Club, 2017, “Using Fourier Transform to detect large orders _ part 1,” www.bsic.it/using-fourier-transform-detect-large-orders-part-1/, Accessed September 2017.

Fast, J., 2016, “Carbon stripping,” Denver Mineral Engineers, Inc., www.denvermineral.com/carbon-stripping/, Accessed October 2017.

Konigsmann, E., Abols, J., Pyecha, J., and Boudreau, T., 2017, “The inline leach reactor installation at Teck- Pogo Inc. Pogo Mine,” pp.1-11, <http://www.gekkos.com/documents/>

Kang, E., Guo, Y., Du, Y., and Zhao, L., 2006, “Acoustic vibration signal processing and analysis in ball mill,” Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21 - 23, 2006, Dalian, China, pp. 6690-6693, <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1714378>, DOI: 10.1109/WCICA.2006.1714378, Accessed October 2017.

National Institute of Instrumentation Standards, 2017, “Concepts of accuracy, precision, and statistical control,” www.nist.gov/sites/default/files/documents/2017/05/09/section-2-concepts.pdf, Accessed October 2017.

Pogo Mine, 2016, “Pogo Mill Process Descriptions,” Print medium handout by Andrew Maxon, Pogo Mill Facility, Pogo Mine, Received January 2016.

Spencer, S., Campbell, J., Weller, K., and Liu, Y., 1999, "Acoustic emissions monitoring of SAG mill performance," IEEE, <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=791509>, DOI: 10.1109/IPMM.1999.791509.

Tipsuwanporn, V., Leawsoong, M., Numsomran, A., and Wongratanapornkul, C., 2013, "Fault detection in compressor using fft algorithm," Proceedings of the World Congress on Engineering and Computer Science 2013 Vol I, WCECS 2013.

Universite Laval, 2017, "Example 2: ECG signal denoising with the SASS algorithm," http://eeweb.poly.edu/iselesni/sass/SASS_toolbox/html/Example2.html, Accessed May 2017.

Wei, X., Kun Guo, K., Liu, H., and Jia, L., 2013, "Fault isolation for light rail vehicle suspension system based on multi-sensor information fusion," IEEE, <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6561560>, Accessed October 2017.

Chapter 4: Peak-Readings Count and Sensitivity Analysis (PRCSA), an Innovative Data-Mining Technique in the Detection of Industrial Sensor Calibration Errors

4.1 Abstract

Sensor calibration errors are causing millions of dollars in losses to the mining and mineral processing industries, and due to their subtlety, they are hard to identify. It is very difficult to develop analytical models based on classical statistical methods for highly non-linear and non-stationary processes such as mineral processing operations. Data-mining methods are particularly effective in such cases. In this chapter, an innovative method developed to detect calibration errors (“bias”) in the sensors of a carbon stripping circuit at Pogo Mine, Alaska is described, i.e., peak-readings count and sensitivity analysis (PRCSA). For the purpose of the experimentations with PRCSA, a 2% bias was artificially introduced in one of the strip vessel sensors (S1). Several “thresholding” strategies were used to filter out undesired data. A strip vessel sensor’s cycle or “peak” is defined as the continuous rise and maintenance of its temperature above a certain desired value. If the number of readings (nr_p) for each peak of a particular sensor data stream above a certain “threshold” (Th) is captured, it is expected that in the presence of bias, the variable nr_p changes dramatically. If this change is captured in terms of certain percentage statistics, bias identification is possible. This is the key principle that was exploited with PRCSA. It is observed that peak-readings count of biased S1 sensor data—at a threshold (Th_{S1}) value of 280°F and a Th_{nr} value greater than 30—increased to 353% when compared to its clean data set (96%). This increment is caused by the presence of bias. Likewise, the cumulative number of peaks and temperatures were increased to 153% and 156%, respectively, for the biased set when compared to the clean set (108% and 109%, respectively). The high increase in percentage change values in these statistics indicate the presence of bias. Similar results were observed for negative bias. The

important finding from the analysis is that the sensitivity of bias at certain temperature thresholds is high. For the stripping circuit sensors, it is 280°F (“effective threshold”). Analysis based on Sensor-S2 is not presented for space constraints; however, it demonstrated the similar trends. The main disadvantage of using PRCSA is that the availability of a “clean data set” for a biased sensor data stream for comparison is highly impractical in real-life situations. The PRCSA algorithm’s performance is presented at the end of the chapter.

4.2 Introduction

Usage of sensors in monitoring industrial processes has become increasingly prevalent in the recent past, and the mining industry is no exception. In fact, a recent study finds that usage of sensors in various stages of mining operations—for a moderately sized mine—can create millions of dollars in economic value; usage of sensors in mineral processing operations can yield \$10-100 million in added economic value annually (Buxton and Benndorf, 2013). The opposite effect is true, however, when sensors suffer from faults and produce erroneous data. For instance, it was observed that sensor faults are causing approximately 3-8% production loss to the US oil industry, resulting in \$20 billion in annual losses to US economy (Wang et al., 2009). Identifying the sensor faults, and fixing them through a calibration process can dramatically improve the sensors’ accuracy. Such processes can reduce the equipment downtimes, increase production, and improve overall safety in the industry. The common type of faults, like noise, failures (flat-outs), stuck-at-faults, etc. (gross errors), can easily be detected and fixed through preventive maintenance and calibration processes; however, calibration bias related errors are hard to identify. These are the errors that are insidious, creep-up over time, and often times indistinguishable if the process is particularly non-linear or non-stationary. Data-mining techniques are of significant use in such situations. Calibration errors are present in the data often in the form of an added offset value (bias)

to the original (true) reading. For the purpose of this research, calibration errors are those that occur in the form of a bias. Innovative techniques that are based on data-mining concepts are explored in this chapter in the pursuit of developing methodology towards finding solutions. Other standard errors were thoroughly explored by the previous researchers and the methodology is well-established to a large extent. Hence they are out of the scope for this research undertaking.

Sensors that monitor various operations in the carbon stripping circuit in the Pogo Mine, Alaska, are chosen for the study. The methods are applied on each sensor individually. Multiple sensor relations are not exploited and are out of scope for this chapter. The methodology and concepts described in this chapter are the basis for the successive innovative methods presented in Chapter 5.

4.3 Literature Review

Data-mining techniques are broadly classified as: “association” based, where relation between variables is exploited; “classification” based, where attributes of each class of items are studied, “clustering” based, where one or more attributes of the classes are examined and grouped together; and “pattern recognition” based, where identifying trends or regular occurrences is used (Brown, 2012). There are some other techniques that use a combination of any of the above. The “prediction” based techniques use classification, pattern recognition, and association (an example is forecasting of a company’s stock performance). The “decision trees” use classification and prediction together (an example is, classification of various sensor faults). In general, real world problems might require several combinations of these techniques or entirely innovative approaches that are specific to a particular problem.

Data-mining techniques have a wide range of industrial applications, and are generally used for extracting knowledge from huge data sets—industrial sensor data sets are one example. For highly non-linear and non-stationary processes, however, it is difficult to develop analytical models based on classical or fundamental statistical methods. Moreover, sensor validation for such processes is equally tough due to production of false alarms in excessive amounts (Kusiak and Song, 2009). Researchers Kusiak and Song (2009) were able to develop several algorithms based on a clustering technique to detect sensor faults in power plant boilers. Classification techniques were used in a “cloud” based application to detect errors in a big sensor data set (Yang et al., 2015). A data-mining approach that employed “rough set” theory along with the artificial neural networks (ANN) was used by Hou et al. (2006) to identify sensor faults in a heating, ventilating and air conditioning (HVAC) system. The rough set theory is effective in the classification of uncertain and incomplete information, and any data set that met the criteria could be improved by using the rough set methodology. Historical performance data of a HVAC system in a building was used in the study. In addition, authors were able to describe several algorithms that were used in this connection. Using “decision-trees algorithms,” Baljak et al. (2012) developed a methodology to classify sensor faults. The authors based their classification on continuity and frequency of occurrence of a fault.

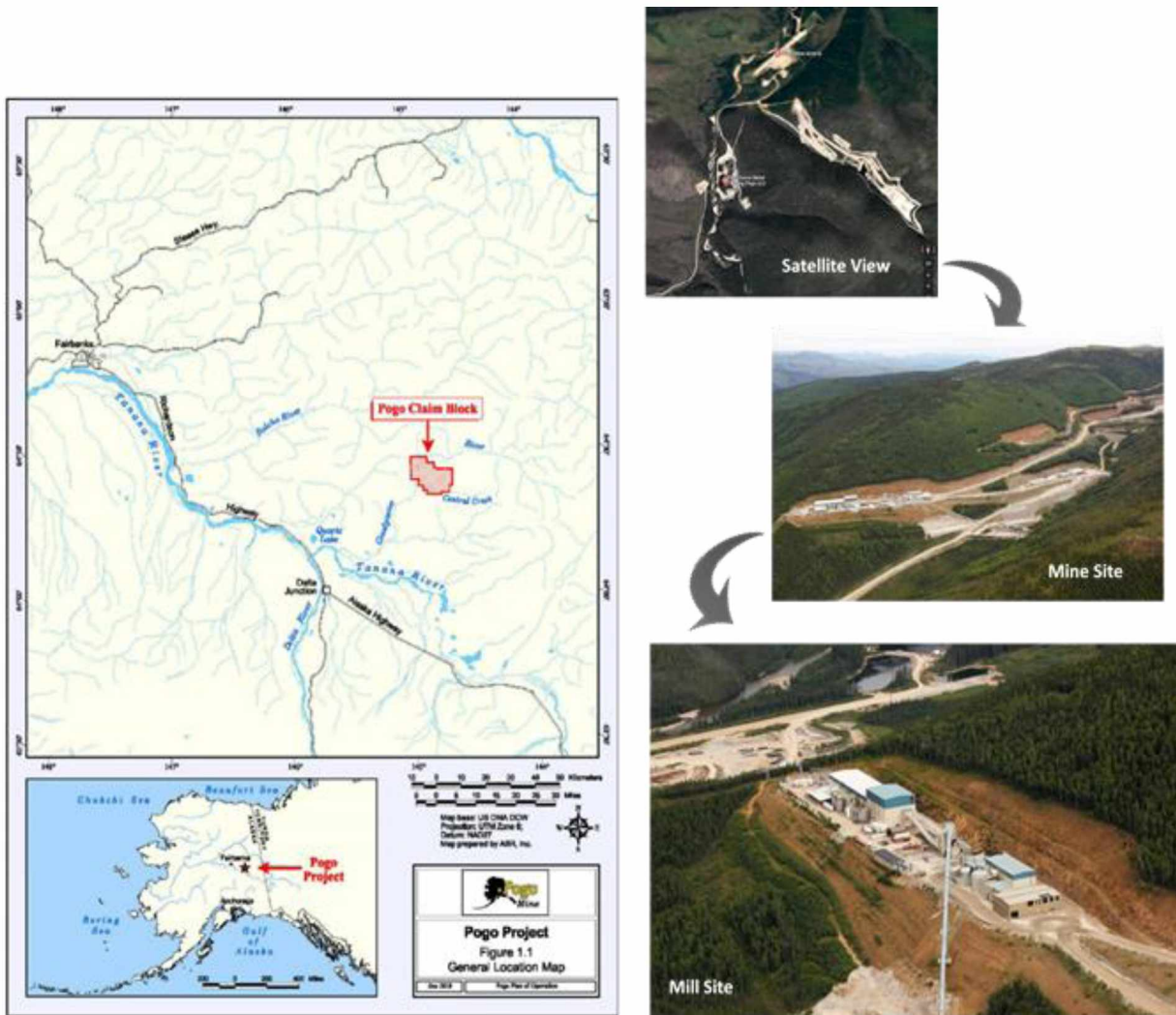
The bulk of the literature is devoted to data-mining approaches for finding gross-errors in the sensor data; thus, the motivation in this chapter is to develop innovative methodology to identify calibration-related errors. Approaches that employ a combination of data-mining methods, i.e., classification, decision-trees, and pattern recognition, are used in the process. Mineral processing is an energy intensive process and accounts for 39% of the energy consumed in mineral

production (U.S. Department of Energy, 2000). The U.S. Department of Energy (DOE) identified sensors' improvements as one of the opportunities to save energy in the mineral industry.

4.4 Introduction to Pogo Mine and Mill

The sensor data that was used for experimentation in this chapter is collected from Pogo Mine's mill facility. Pogo Mine is a major gold producer in Alaska (Figure 4.1). Located on the Goodpaster River, 38 miles (61 km) north of Delta Junction in east-central Alaska, the nearest city to Pogo is Fairbanks, located approximately 70 miles (112 km) northwest of the mine property (Konigsmann et al., 2017). Pogo is an underground operation. Elevations on the property range from 1,299 ft (396 m) on the Goodpaster River to over 4,003 ft (1,220 m) on the top of Pogo Ridge, an east-west trending ridge. The climate is classified as sub-Arctic with cold, dry winters and relatively mild summers (Konigsmann, 2017).

Pogo Mill processes up to 3,500 tons of ore daily. The process flowsheet is shown in Figure 4.2. The Pogo plant's process flow mainly is comprised of a crusher, semi-autogenous (SAG) mill, ball mill, floatation circuit, leaching tanks, carbon-in-pulp CIP circuit (described in detail in section 4.4.1), stripping circuit, and an electro-winning circuit. The Pogo plant is a closed-circuit operation; all the water used is recirculated and cleaned before it is released into a pond to minimize environmental impact. Up to 67% of the gold is recovered through the processing facility. A brief description of the process follows (Pogo Mine, 2017).

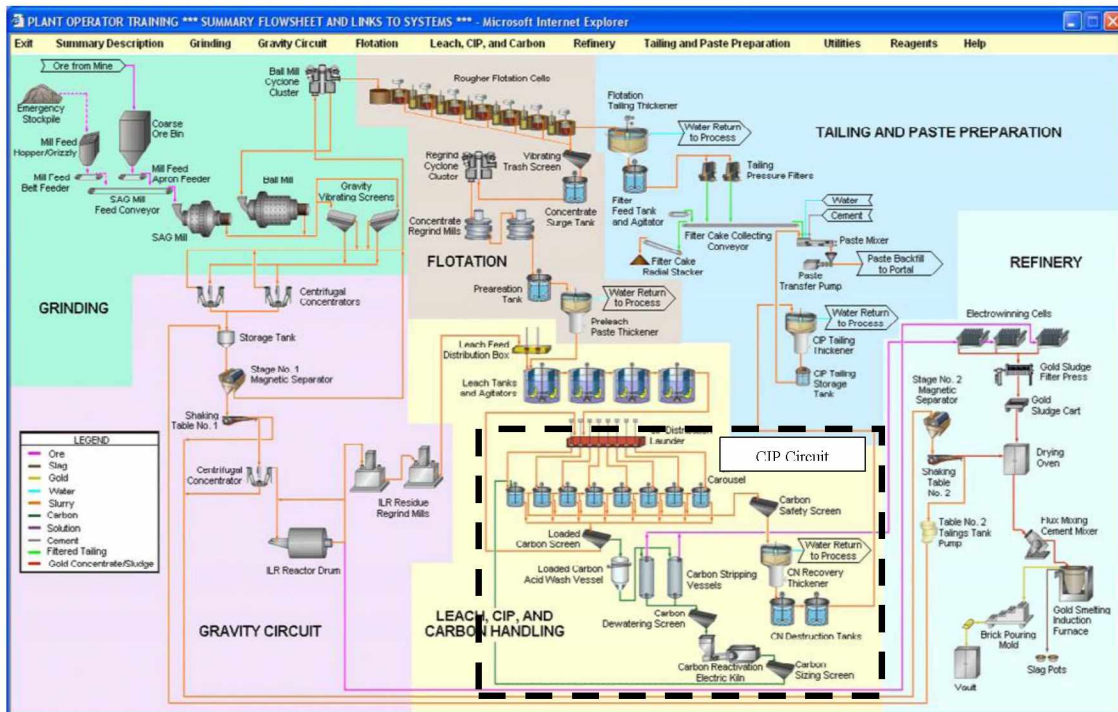


Source: Adapted from Pogo Mine, 2017.

Figure 4.1: Pogo Mine site map.

Ore is fed to a conventional SAG/ball mill grinding circuit. Gravity recovery is the technique employed throughout the process. The ball mill circulating load is screened at one (1) mm, and the undersize is fed to two 48-inch centrifugal concentrators operating in parallel. The primary gravity concentrate is then fed to the intensive cyanidation circuit with the leach solutions reporting directly to electro-winning for final gold recovery. Primary gravity tailings are returned to the grinding circuit via the cyclone feed pump-box while the intensive leach residues are reground and pumped into the flotation concentrate leach circuit. Primary gravity gold concentrates are

intensively leached on a batch basis; the typical leach residence times are approximately 14 hours. Grinding cyclone overflow reports to a sulphide rougher flotation circuit, which produces a 10% weight concentrate. The concentrate from the flotation circuit is then reground to 80% passing 10 microns using stirred media detritors, prior to being leached in a conventional cyanidation circuit followed by gold recovery in an eight cell carousel CIP circuit. The particles then enter a stripping circuit. A detailed description of the process can be found in the subsequent sections.



Source: Adapted from Pogo Mine, 2016.

Figure 4.2: Pogo Mine-gold processing flow sheet.

4.4.1 Carbon-in-Pulp (CIP) Circuit

The purpose of the CIP process is to allow the gold previously dissolved in the leach tanks to be adsorbed by activated carbon—activated carbon is a form of charcoal that has a large number of low volume pores that help in the adsorption of the fine particles. During the CIP process, the

gold particles are slowly adsorbed onto the carbon particles, and eventually extracted in the subsequent “stripping” process. The CIP circuit is designed to allow adequate time for the absorption process. The particle-bearing slurry spends 30 minutes in each CIP tank. The Pogo CIP circuit has eight such tanks (Figure 4.3). The slurry spends four hours in the circuit, and approximately 300 to 600 ounces of gold is adsorbed per ton of activated carbon used.



Source: Adapted from Pogo Mine, 2016.

Figure 4.3: CIP circuit at Pogo Mine.

4.4.2 Carbon Stripping Circuit

The Pogo stripping circuit in pictures is shown in Figure 4.4. Pogo stripping circuit consists of two strip vessels that work in tandem. While one vessel (vessel-1) is being loaded with gold bearing activated carbon, the other previously loaded vessel (vessel-2) is operated by circulating a solution called “elute.” The temperature and pressure of elute is maintained approximately at 280°F and 65 PSIG, respectively, to facilitate maximum liberation of gold particles. An elute is a

water based solution with 1% sodium hydroxide and 0.1% sodium cyanide (Fast, 2016). The process is called “pressurized Zadra stripping.” A typical pressurized Zadra stripping cycle lasts for 11 hours and consists of the following stages: loading the vessel (1 hr), circulating elution (8 hrs), carbon cooling (1 hr), and unloading carbon from the vessel ($\frac{1}{2}$ hr) (Table 4.1).

Table 4.1: Operating schedule-pressure Zadra stripping.

Operation	Solution	Time
Load Column	Transfer Water	90 minutes
Elution	0.1% NaCN, 1% NaOH	480 minutes
Carbon Cooling	Fresh Water	60 minutes
Unload Column	Transfer Water	30 minutes
TOTAL		11 hours

Source: Fast, 2016.

While the used carbon is discharged, the “pregnant leach” solution is pumped out. The same process is repeated with strip vessel-2. The pregnant leach solution (PLS), on its way out from the strip vessel, is cooled off by heat exchangers 3 and 4 (Figure 4.5). When the PLS reaches the electro-winning circuit, the gold particles are removed and the solution, now called “barren solution,” is reheated by a boiler with the aid of heat exchangers 1 and 2. The reheated barren solution is then recirculated through strip vessels. A glycol solution is circulated between the boiler and heat exchangers as a medium of heat exchange. Sensors are strategically placed at various parts of the circuit to measure temperatures, flow rates, etc. It is very important to maintain temperatures in the strip vessels at certain levels (270-280°F) for certain periods of time to maximize gold separation. A false “optimal temperature” will result in either poor gold recoveries or higher costs. Thus, monitoring the temperatures, and identifying sensor errors became crucial. S1 and S2, the temperature sensors for strip vessels 1 and 2, respectively, are the two important sensors in this context, and therefore are the focus of this chapter.

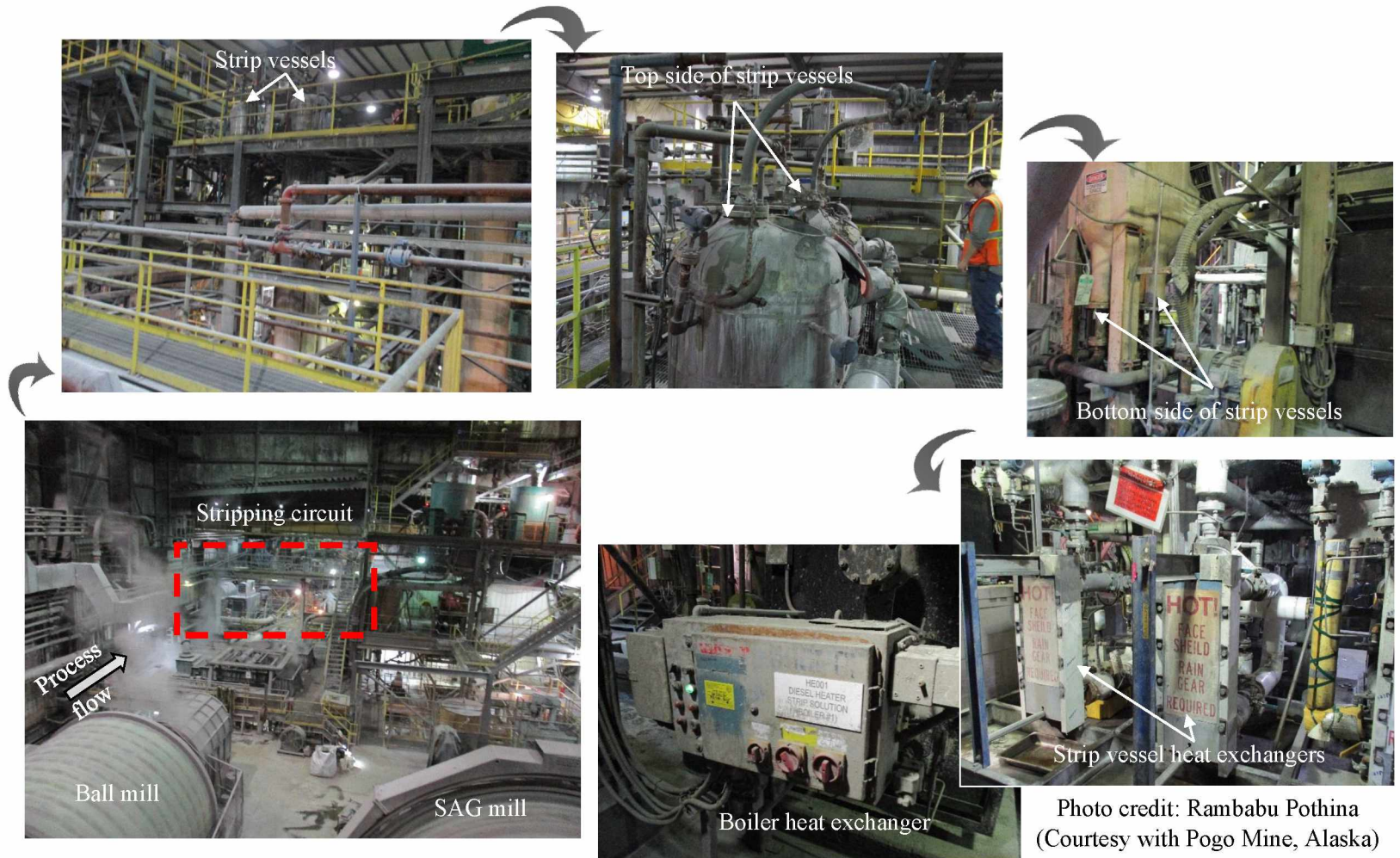
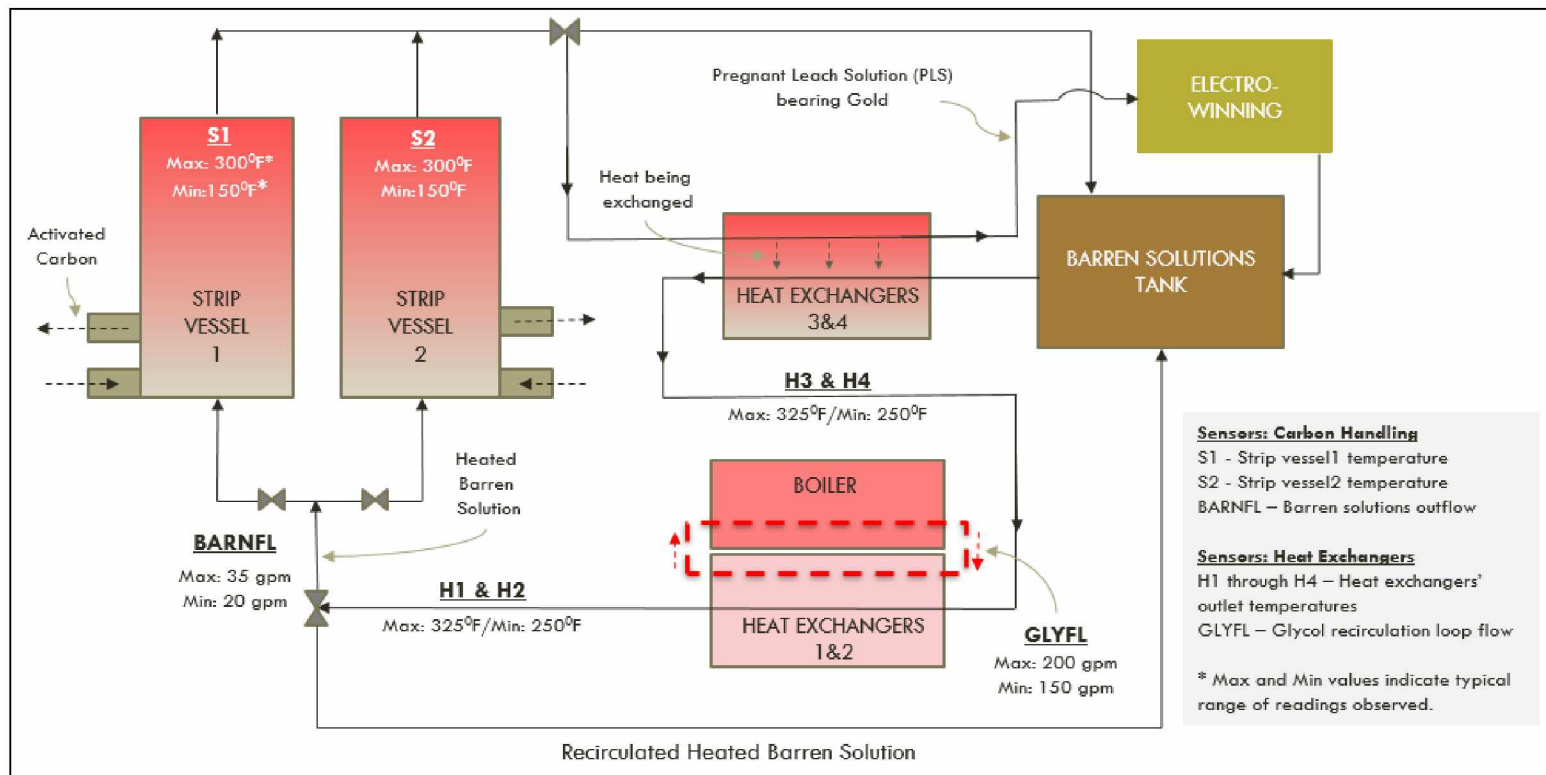


Figure 4.4: Pogo stripping circuit in pictures.



Source: Pogo Mine, 2016.

Figure 4.5: Pogo stripping circuit schematic diagram with sensor placements.

4.5 Methods and Materials

The methodology involves the following stages: data collection, data preparation, description of algorithms, and a detailed account on peak-readings count and sensitivity analyses.

4.5.1 Data Collection

The raw sensor data was collected at 10-min average intervals from the Pogo mill database. Various sensors of interest and their data readings is summarized in Table 4.2. A 10-min average or aggregation helps save computer memory and the results are not significantly affected at that level of aggregation (Arku and Ganguli, 2014). The data was collected for a period of nine (9) months: Jan 1, 2015 through September 31, 2015. The visual format of various sensor raw data streams with close-up views can be seen in Figures 4.6 through 4.11. It should be noted that sensors H3 and H4 are not part of the study in this chapter. Various descriptive statistics for the sensor data are given in Table 4.3. The “data cleansing cut-off value ($Th_{cleanse}$)” column in Table 4.3, refers to the cut-off value based on which undesired and corrupted data were removed (cleaned). This process is explained in the data preparation section.

Table 4.2: A snapshot of raw sensor data collected at 10-min average intervals.

Record #	Time	Strip vessel sensors		Barren flow sensor	Heat exchanger sensors		Glycol flow sensor
		S1 (°F)	S2 (°F)	BARNFL (GPM)	H1 (°F)	H2 (°F)	GLYFL (GPM)
1	1/1/15 12:00 AM	90.5	261.6	35.4	299.4	300.7	204.0
2	1/1/15 12:10 AM	90.4	261.6	35.4	291.7	293.0	203.0
3	1/1/15 12:20 AM	90.3	261.6	35.3	284.3	285.4	200.0
4	1/1/15 12:30 AM	90.1	233.9	35.3	277.3	278.3	201.1
5	1/1/15 12:40 AM	90.0	219.2	35.2	270.5	271.5	202.3
6	1/1/15 12:50 AM	89.8	216.9	35.2	264.1	265.1	204.6
7	1/1/15 1:00 AM	89.7	215.3	35.2	258.0	258.8	204.4
8	1/1/15 1:10 AM	89.6	204.2	35.1	252.1	252.9	205.0
9	1/1/15 1:20 AM	89.4	192.4	35.1	246.5	247.3	201.9
10	1/1/15 1:30 AM	89.3	189.8	35.0	241.3	242.0	201.0

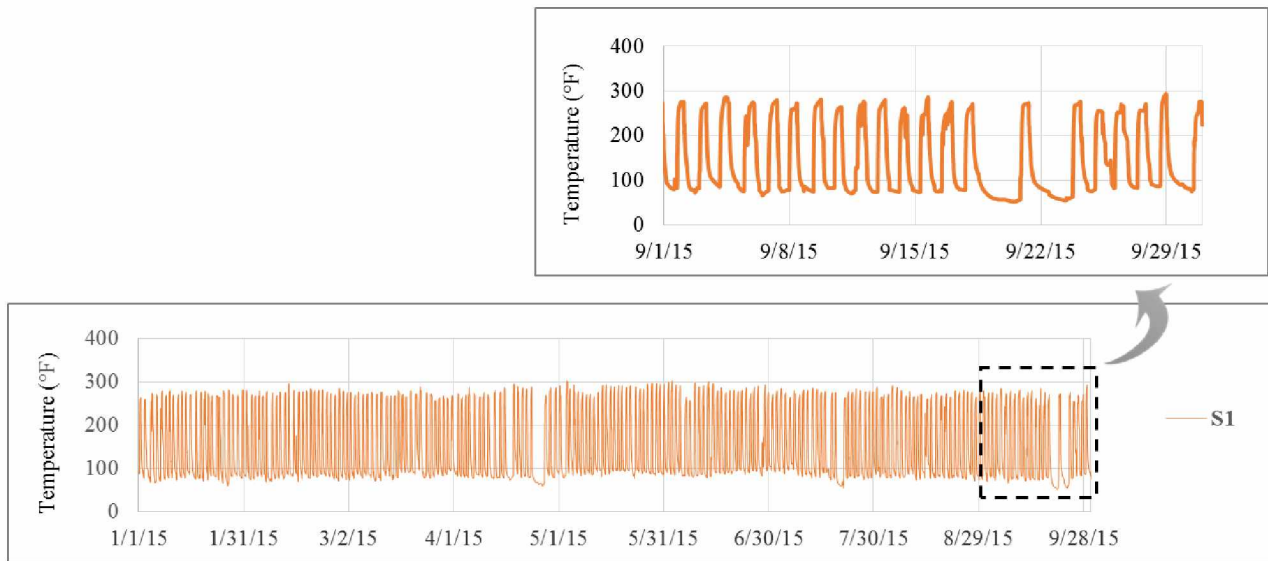


Figure 4.6: Strip vessel-1 heat sensor (S1).

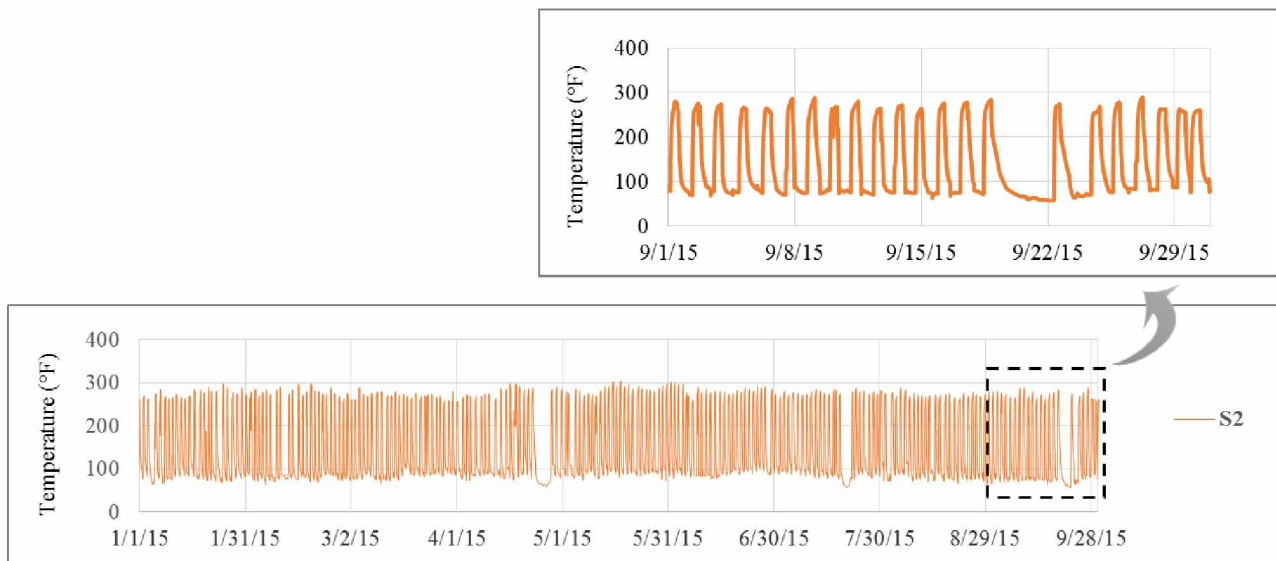


Figure 4.7: Strip vessel-2 heat sensor (S2).

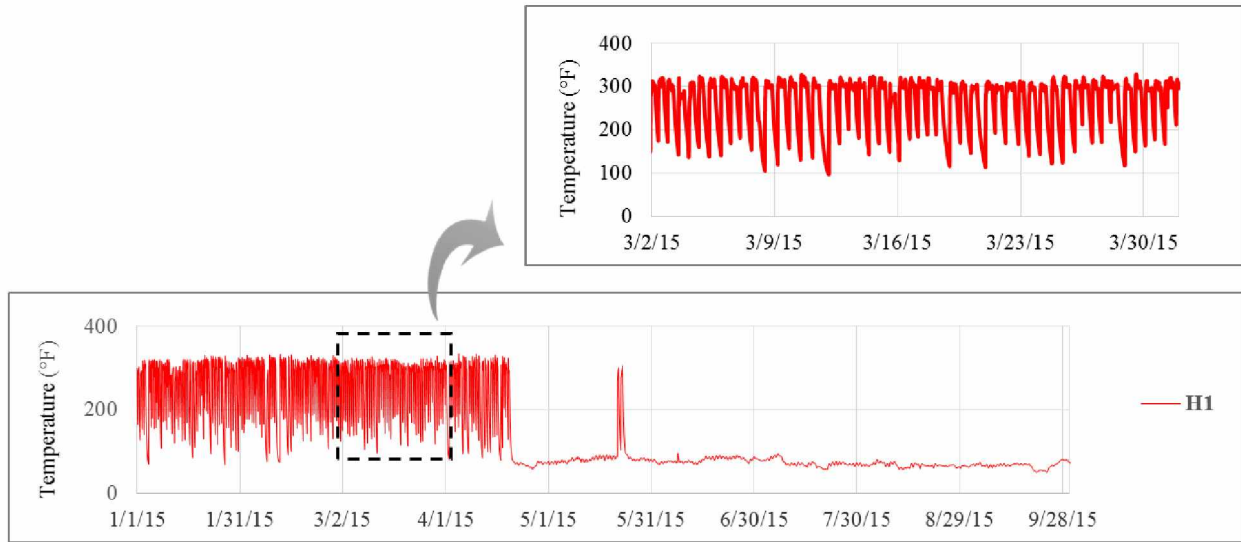


Figure 4.8: Heat exchanger-1 sensor (H1).

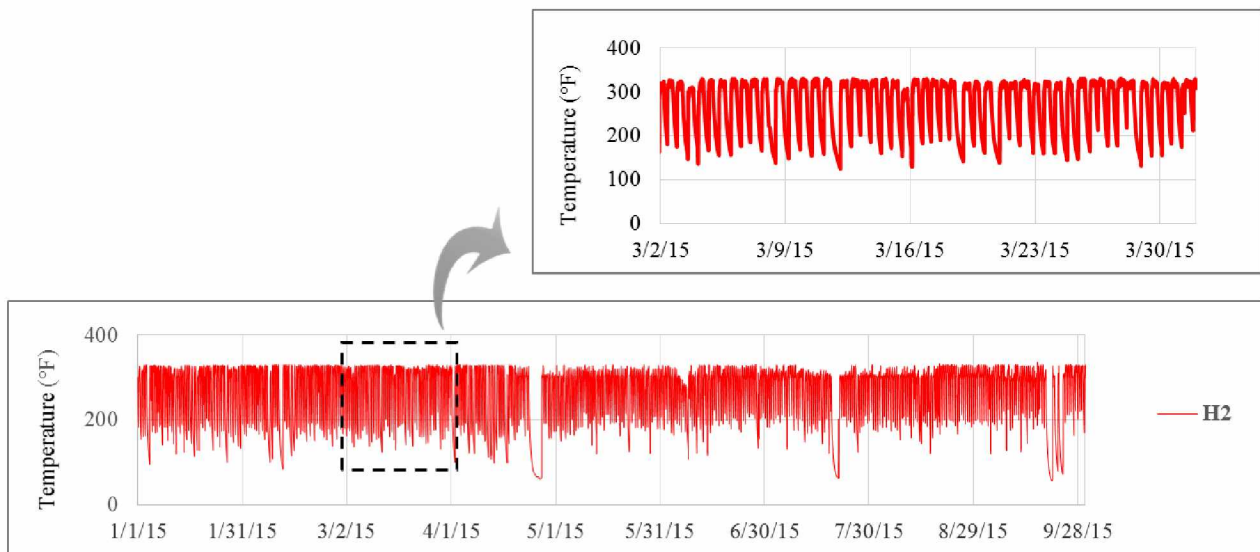


Figure 4.9: Heat exchanger-2 sensor (H2).

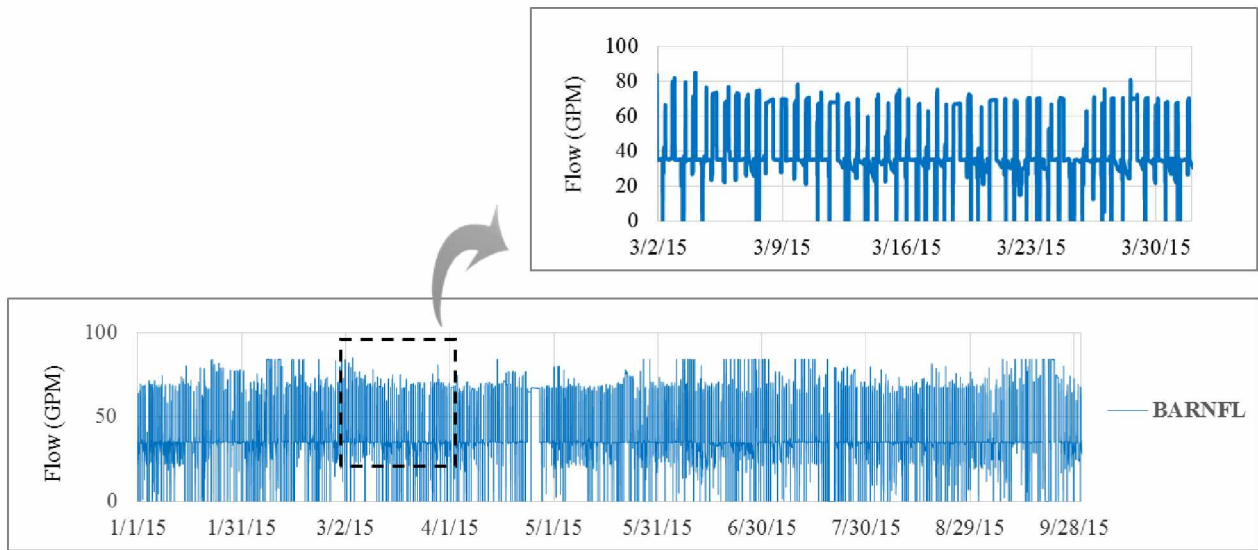


Figure 4.10: Barren flow sensor (BARNFL).

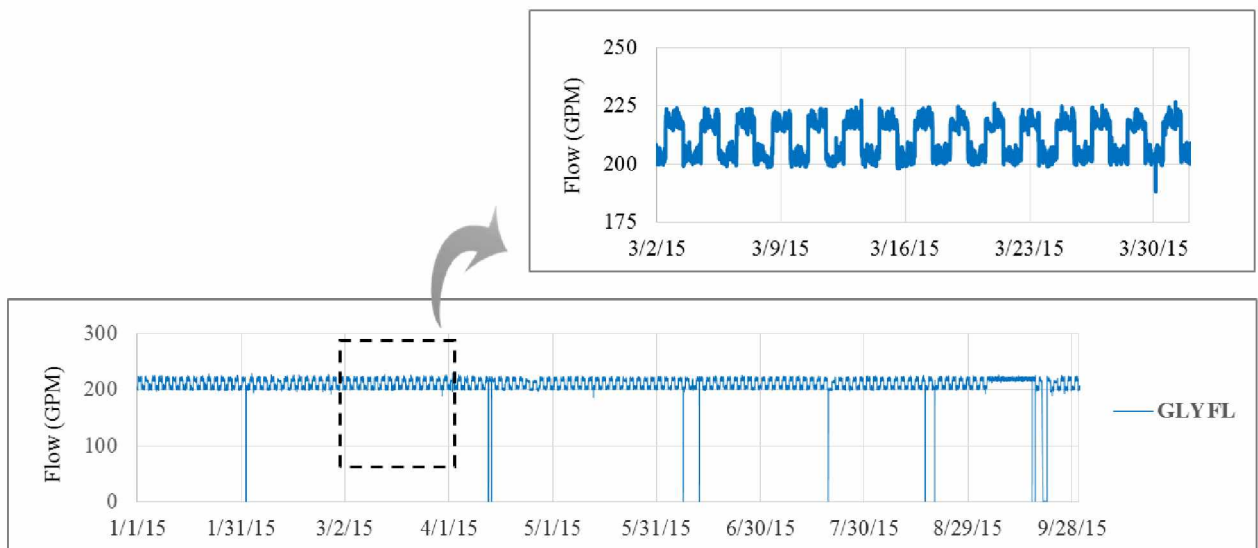


Figure 4.11: Glycol flow sensor (GLYFL).

Table 4.3: Descriptive statistics for the sensor data collected.

Sensor	Measuring parameter/variable	Unit	Maximum	Minimum	Mean	Median	Standard deviation	Data cleansing cut-off value ($Th_{cleanse}$)
S1	Strip vessel-1 temperature	°F	302.7	52.1	150.1	104.9	78.3	150
S2	Strip vessel-2 temperature	°F	304.3	57.7	148.8	105.2	77.6	150
H1	Heat exchanger-1 outlet temperature	°F	334.0	50.0	144.9	82.7	96.7	250
H2	Heat exchanger-2 outlet temperature	°F	336.2	56.3	263.1	296.0	63.6	250
BARNFL	Barren solution flow	GPM	302.7	0.0	40.4	35.0	17.9	20
GLYFL	Glycol flow	GPM	229.7	0.0	209.8	214.6	18.2	150

4.5.2 Assumptions and scope

Calibration errors are present in data often in the form of an offset value (bias) added to the original or true reading; see Equation (4.1). Bias is the common error associated with sensors and often a hard one to identify; hence it is the interest of this research. Other common errors like gross errors are out of the scope of this research. For the purpose of this experimentation, it is assumed that the data collected from the Pogo Mine is devoid of errors, hence the set is deemed as a “clean set.”

$$\text{Observed or biased reading, } x_{bias} = \text{True reading } (x_{true}) + \text{bias } (e^{bias}) \quad (4.1)$$

Bias is expressed as a percentage over the true reading value. It is artificially induced in the clean data set. A +2% bias indicates 2% of the true reading was added as bias to the true reading. Likewise -2% expresses adding -2% of the true reading; see Equation (4.1). Calibration errors are subtle and can be as low as 2% over the true reading. Identifying the bias at such low magnitudes is preferred by the industries. Pogo Mine specifically requested that errors be detected at the $\pm 2\%$

level. Only S1 and S2 sensors are analyzed with the methods described. Bias is introduced only in one sensor (S1) at a time in this project. Identifying the bias in several sensors at a time in a multi-sensor environment is out of the scope for this research. Due to the space constraints and to avoid redundancy, the results from the S1-based analysis and validation only are presented in this chapter.

4.5.3 Data Preparation

4.5.3.1 Programming and Software

The sheer volume of the data collected, the subsequent mathematical calculations and repetitions to be performed, statistical analyses to be conducted, and graphical presentation of the results, demand for the usage of a mathematically-intensive software package. Matlab was selected as the software of choice. Matlab has various discipline-specific toolboxes that can be added according to the need, i.e., optimization, signal processing, simulations, neural nets, etc.

The goal of the project was to detect an artificially introduced bias (+2% or -2%). For the purpose of conducting bias identification tests using algorithms, the bias was introduced at a random time period in S1. Once the bias was introduced, all following data for the sensor also contained the bias. The assumption was that once a bias occurred, the sensor remained “uncalibrated” until the error was detected. Thus, if a -2% bias was injected into S1 on March 12, all S1 data starting March 12 was corrupted with a -2% bias. Algorithms were designed to detect the bias as soon as possible. The location of the error was systematically varied from Jan 1, 2015 through September 31, 2015. A series of cut-offs or thresholds were applied to cleanse the data of unreliable or possibly corrupted readings. The algorithm description section has the detailed account of the process.

4.5.3.2 Basic Concepts

Cleansing Threshold (Th_{cleans}):

The first process in the data preparation is to remove or clean undesired or corrupted data. Usually these are sensor data readings below a certain minimum value (“cleansing threshold”) that reflects when a piece of equipment is down for maintenance or idling. The last column of Table 4.3 provides these values, which were used as cleansing thresholds for each sensor data stream.

Peak (P) and Threshold (Th):

For the purposes of this research, a strip vessel cycle or “peak” is defined as the continuous rise and maintenance of a sensor’s temperature above a certain desired value. According to Pogo Mine, the desirable range for the strip vessels’ operational temperature is between 270°F and 280°F. The data stream is filtered for the temperature sensors using a “threshold (Th)” value to define a cycle or peak. As described previously, the strip vessel temperatures are cyclical in nature. A peak is assumed to start (peak start-time) when the temperature rises above the threshold. It is assumed to have ended when the temperature goes below the threshold (peak end-time) (Figure 4.12).

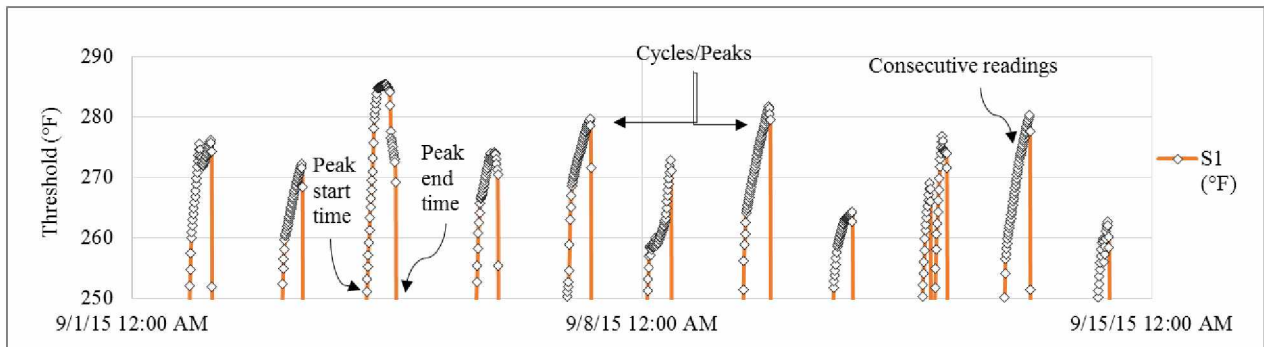


Figure 4.12: Cycles (peaks) vs thresholds in a clean set of S1 sensor data.

Unlike a traditional “cycle,” which is typically associated with positive and negative components, a peak, for the purpose of the research, can be understood as a truncated form of a cycle (Figure 4.12). The readings below the threshold are ignored in this research. Figure 4.12 shows the appearance of peaks above the 250°F threshold.

4.5.3.3 Algorithm Description

Step 1: The algorithm reads the sensor raw data (readings) along with the corresponding time of observations (Table 4.2). All other input parameters that include various thresholds and cut-offs are also read simultaneously.

Step 2: The data are cleaned for undesired or corrupted values based on the given cleansing cut-off values ($Th_{cleanse}$).

Step 3: The algorithm creates a biased data set by adding the bias at the given percentage.

Step 4: The algorithm reads each and every measurement of the chosen sensor and checks them against the threshold (Th_{sensor}). If it is for S1, the threshold is Th_{S1} . As described previously, if a reading is above the Th for a continuous period of time (t), the algorithm identifies it as a peak (P) and captures the number of readings that constitute a peak (nr_p) along with the number of peaks (nP) simultaneously. While the peaks are captured, the cumulative temperature per peak ($T_{peak-cum}$) and the average temperature per peak ($T_{peak-Ave}$) are also calculated. The variables $cum\ nr_p$, $T_{peak-cum}$, and $cum\ T$ (cumulative temperature) have proven to be good indicators of bias.

Step 5: The results from the algorithm are written to an excel spreadsheet where plotting of various graphs is performed.

The variable nr_p for the clean data set exhibits certain trend (Figure 4.12). In the presence of bias, the variable nr_p will change dramatically. This is the underlying principle that is exploited in peak-readings count analysis (PRCA) and peak-readings sensitivity analysis (PRSA)—together called, peak readings count and sensitivity analysis (PRCSA)—which is explained in detail in the corresponding sections. The flow chart for the algorithm is shown in Figure 4.13.

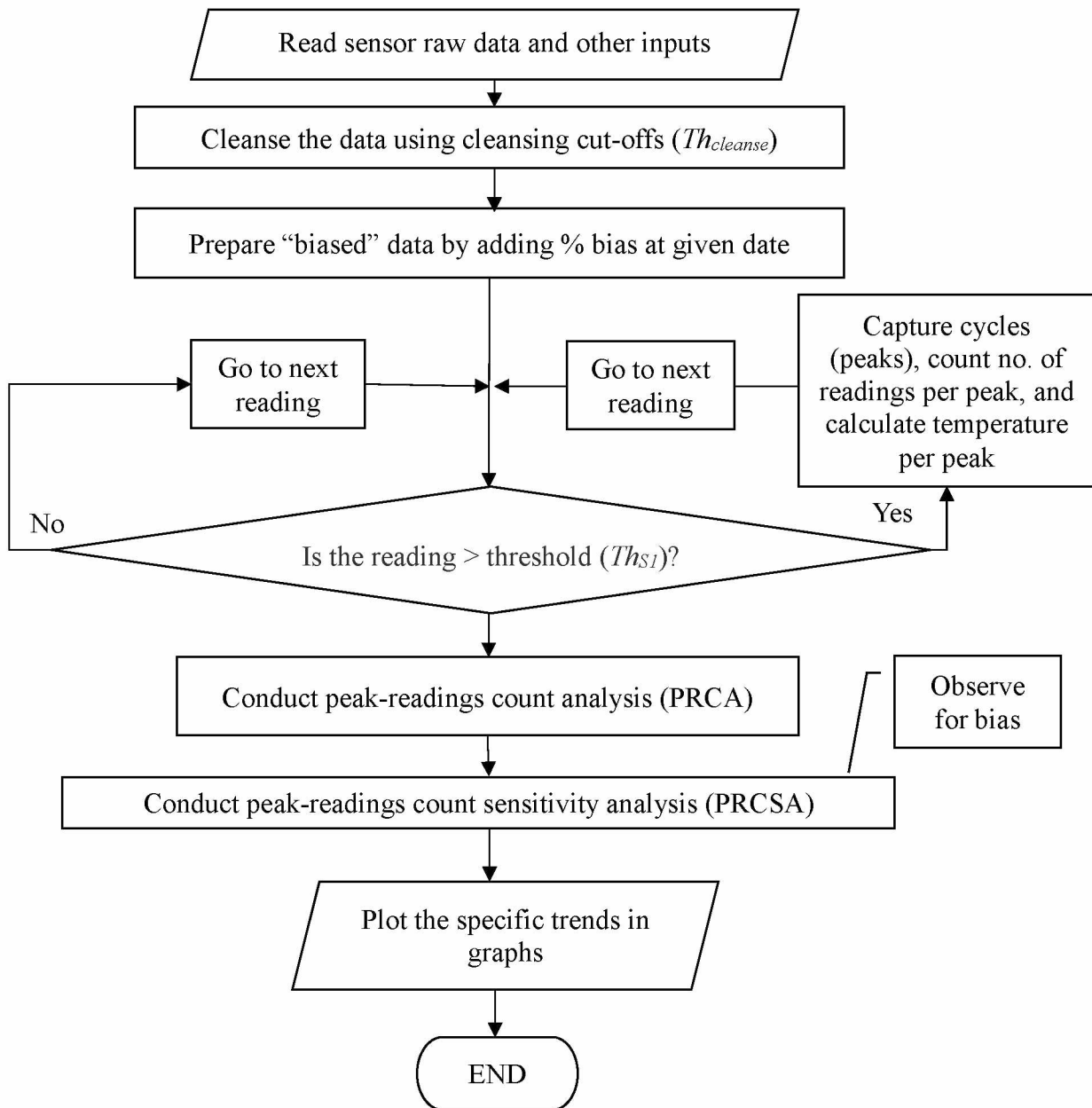


Figure 4.13: Flow chart for the peak-readings count and sensitivity analysis (PRCSA).

4.5.4 Peak-Readings Count Analysis (PRCA)

In order to find the bias present in sensor data using PRCA, the following assumptions are made and procedure is used.

(i) Assume the sensor data collected is devoid of errors. In an industrial setting, this can be achieved by start with a proper calibration. For the purpose of this analysis, it is assumed that the data was collected from calibrated sensors. The rationale is if the behavior of clean data is observed in terms of patterns, it can be compared with biased data to detect errors. This means that the clean set serves as a baseline to find bias.

(ii) Create another set of data by inducing bias on a known date, just to mimic the “biased data set.” Compare the biased set with the clean set. The change in the number of readings per peak of S1 data is obvious from the example shown in Figure 4.14. In the example, a 10 % bias was introduced on 5/17/15 12:40 PM and the bias stayed to the end of the set, i.e., 9/30/15.

(iii) Equations (4.2), (4.3) and (4.4) can be used to calculate the percentage change in peak-readings count, percentage change in cumulative number of peaks, and percentage change in cumulative temperature per peak, respectively, for a preset threshold (Th). The latter two measures are good indicators of bias.

$$cum\ nr_p\ change\% = \frac{cum\ nr_p\ biased - cum\ nr_p\ clean}{cum\ nr_p\ clean} \quad (4.2)$$

where

cum nr_p change% is percentage change in cumulative number of peak-readings count, *cum nr_p biased* is cumulative number of readings per peak when bias is present, and *cum nr_p clean* is cumulative number of readings per peak when data is clean.

$$cum\ nP\ change\% = \frac{cum\ nP\ biased - cum\ nP\ clean}{cum\ nP\ clean} \quad (4.3)$$

where

cum nP change% is percentage change in cum number of peaks, *cum nP biased* is cumulative number of peaks when bias is present, and *cum nP clean* is cumulative number of peaks when data is clean.

$$cum\ T\ change\% = \frac{cum\ T\ biased - cum\ T\ clean}{cum\ T\ clean} \quad (4.4)$$

where

cum T change% is cumulative percent change in temperature, *cum T biased* is cumulative temperature of all peaks combined when bias is present, and *cum T clean* is cumulative temperature of all peaks combined for clean data.

From the example depicted in Figure 4.14, the change in the number of peaks (*cum nP*) — “before the bias is induced” plus “after the bias is induced”—and the reading-counts (*cum nr_p*) is quite evident at a *Th_{SI}* of 270°F. The *cum nP* value (past the threshold) for the clean set of data is nine, whereas for biased data, it is twelve. Specifically, there is no percentage change in the *nP* value in either of the sets “before the bias is introduced,” whereas after the bias is introduced in one set, the number of peaks for that biased set increased by four. This is the clear indication of

positive bias, whereas the reduced number of cum nr_p indicates negative bias. The findings and graphs from the PRCA are presented in the results section.

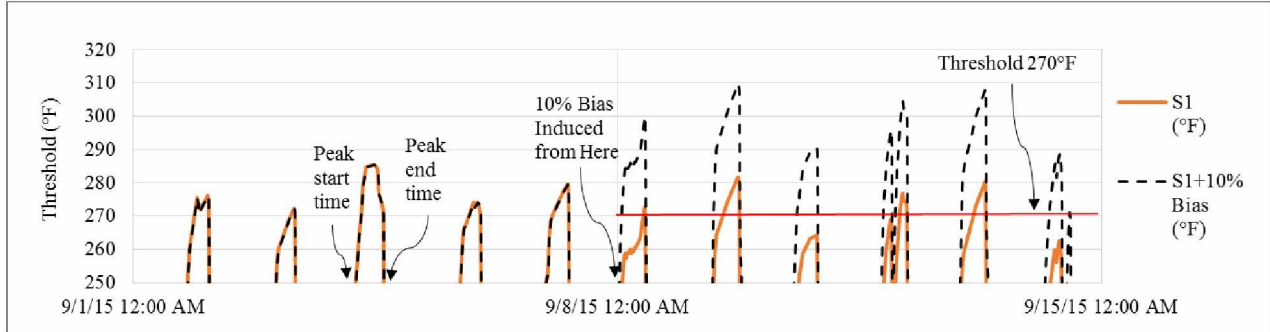


Figure 4.14: Bias effect on S1-sensor data set at various thresholds.

4.5.5 Peak-Readings Count and Sensitivity Analysis (PRCSA)

It was observed that the percentage change in the total number of peak-reading counts ($cum nr_p change\%$) and peak cumulative temperatures ($cum T change\%$), are more sensitive to certain “thresholds” than others. The change is too high for the positive bias and too low for the negative bias. For instance, from the Figure 4.14, it can be observed that the difference between cumulative number of peaks in biased and clean data sets at a threshold of 270°F (four readings) is more than the difference registered at a threshold of 260°F (zero readings). At a threshold of 280°F, the difference is five which is even higher than the differences registered for previous thresholds. This means that the sensitivity of the difference ($cum nP change\%$) is higher at $Th=280^\circ F$ —compared to other thresholds. Identifying the threshold (Th_{sensor}) that is more sensitive to bias can make the entire process much easier to perform. The sensitivity analysis is aimed at finding such threshold values. Likewise, peaks that have readings more than a preset number, i.e., number of peak-readings threshold (Th_{nr}), exhibited more sensitivity to bias. Using Equations (4.2), (4.3), and (4.4)

at different Th_{nr} and Th_{sensor} values, bias was found in terms of sensitivity statistics. The findings are presented in the results sections.

The PRCSA is achieved using the following steps.

(i) Find a threshold at which there is a greater percentage of change between biased and the clean data sets in terms of these three key statistics, i.e., cumulative readings-count (*cum nr_p change%*), cumulative no. of peaks (*cum nP change%*) and cumulative temperature (*cum T change%*).

(ii) Establish a threshold for the no. of readings per peak (Th_{nr}) and observe all the three key statistics from (i) again. Increase the Th_{nr} (30<40<50) and observe the same. The intent is to find at what values of Th_{nr} all three statics are highly sensitive compared to other values.

(iii) For the thresholds (Th_{nr}) established in step (ii), conduct sensitivity analysis. From the experimentation, the sensitivities at a value of 30 are found higher. This threshold is called as effective threshold (*eff Th_{nr}*).

4.6 Results

4.6.1 Peak-Readings Count Analysis (PRCA)

The analysis is basically a comparison between the clean data set and biased data set (+2% bias) for sensor S1. The results from the analysis are shown in Figures 4.15 through 4.17. In the figures, the number of readings per peak (nr_p) values are represented as flat lines rather than points to show where the step change in the values occur. It should also be noted that the values are aggregated to 1-hour averages to observe the broad trends.

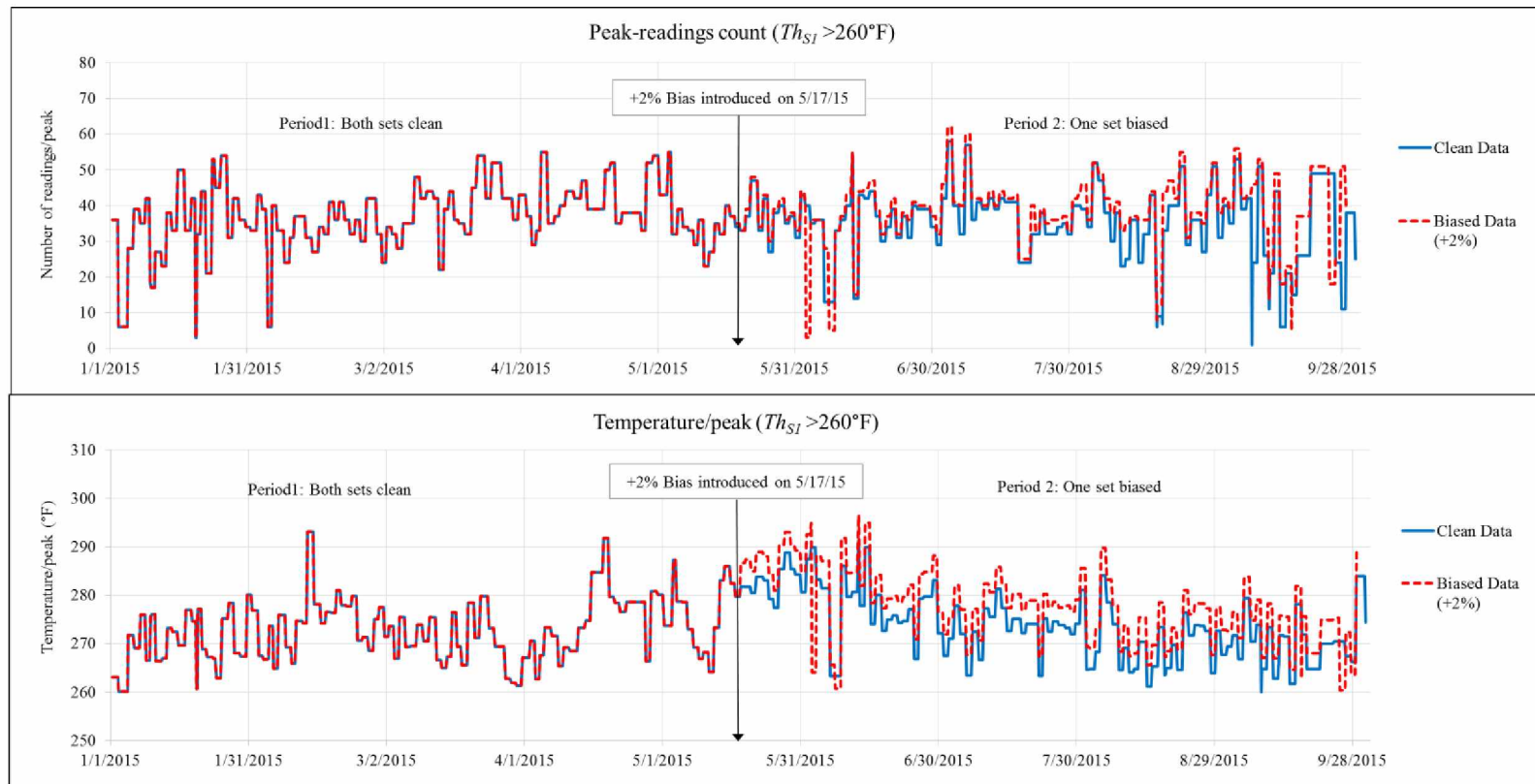


Figure 4.15: Bias effect on S1-sensor data at 260°F threshold (Th_{S1}).

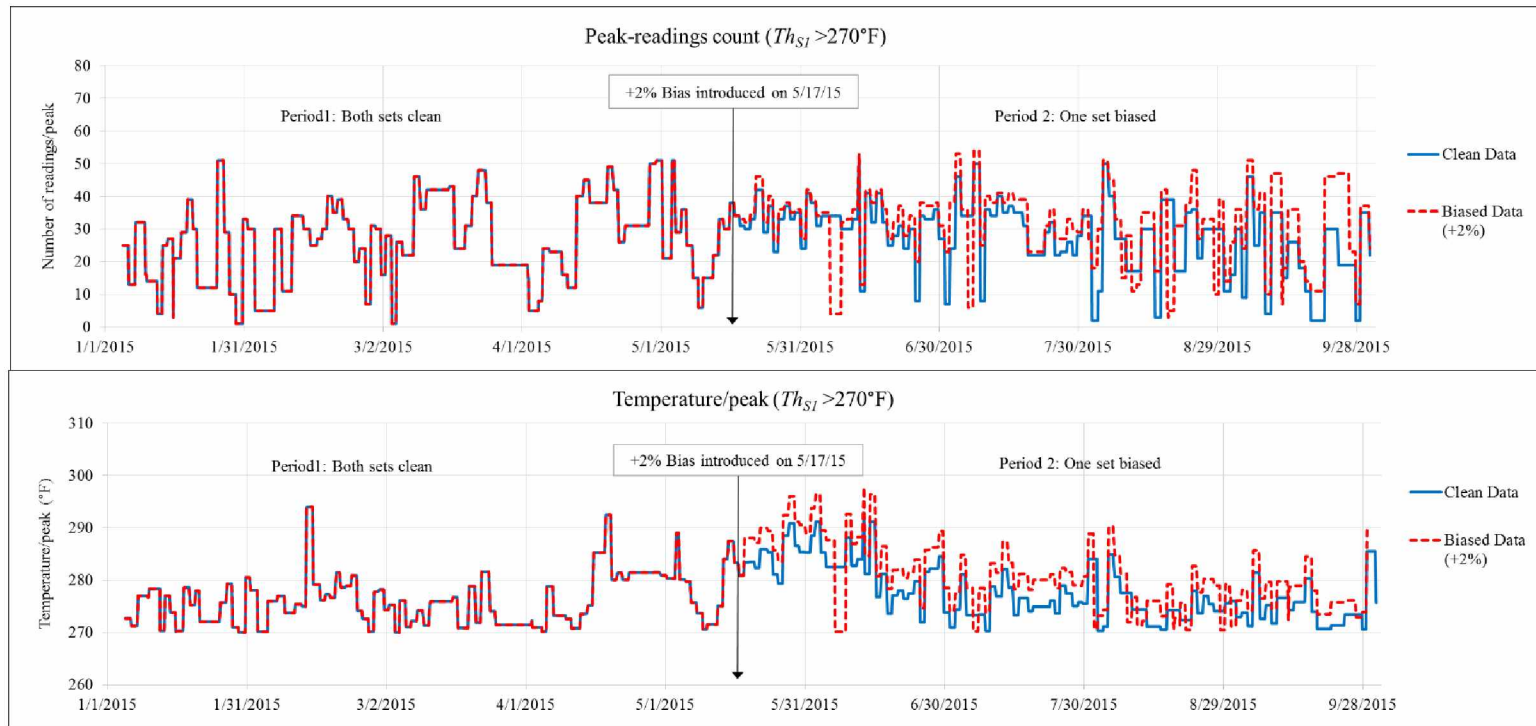


Figure 4.16: Bias effect on S1-sensor data at 270°F threshold (Th_{S1}).

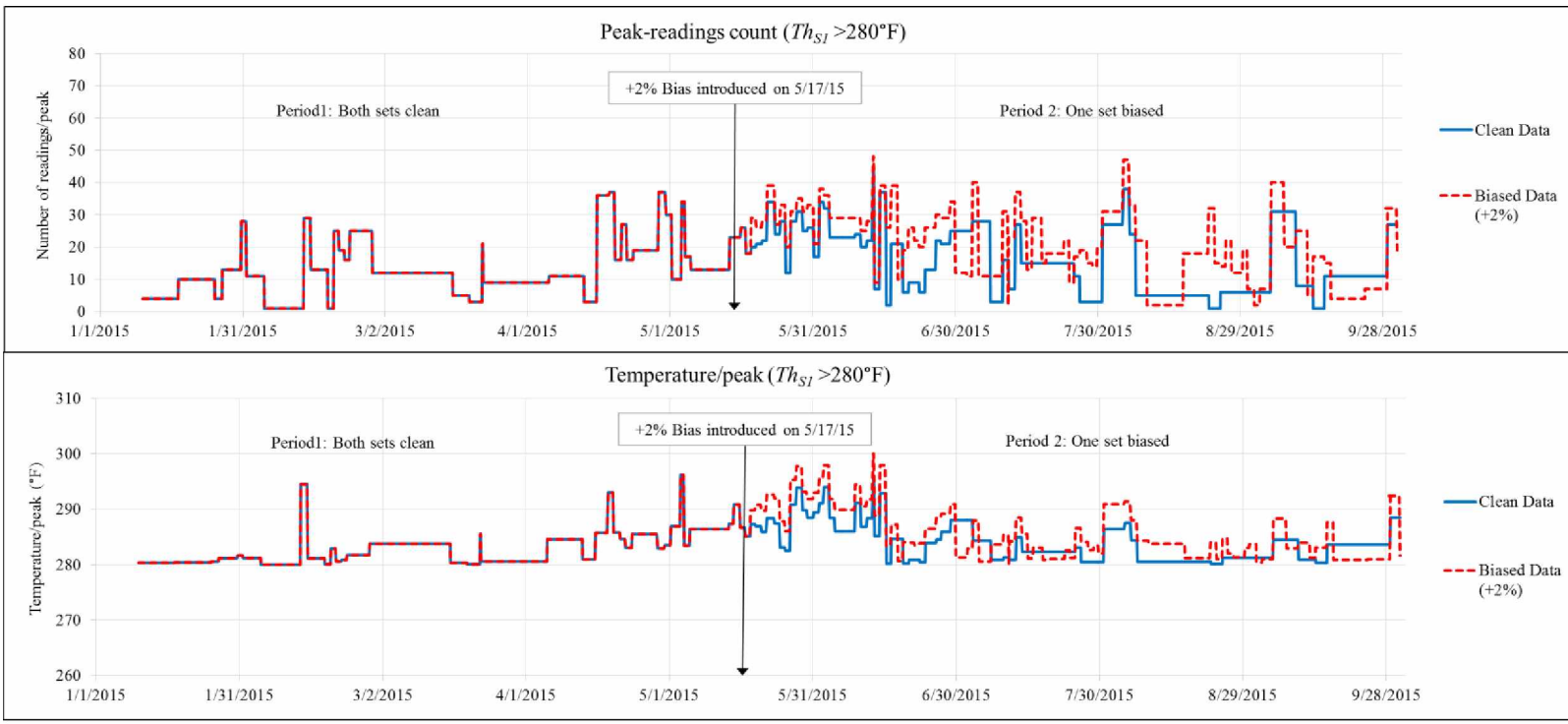


Figure 4.17: Bias effect on S1-sensor data at 280°F threshold (Th_{S1}).

4.6.2 Peak-Readings Count Sensitivity Analysis (PRCSA)

It can be observed from the Tables 4.4 through 4.9 and Figures 4.18 through 4.23 that the sensitivities of peak-readings counts and corresponding cumulative temperatures are high at 280°F threshold. The reader should notice that the figures and tables arranged in pairs convey the same information (Example: Table 4.4 and Figure 4.19).

The peak-readings count sensitivity analysis (PRCSA) is conducted on the results obtained from the peak-reading count analysis (PRCA). The results are as follows. The results that constitute the effective threshold for S1 (*eff Th_{S1}*) are summarized in Tables 4.10 through 4.12 and shown in Figures 4.24 through 4.26. Again, the reader should notice that the figures and tables arranged in pairs convey the same information (Example: Table 4.10 and Figure 4.24).

CLEAN DATA USED



Table 4.4: Cumulative peak-readings count (clean).

Sensor-S1 readings threshold (°F)	Threshold for 'no. of readings/peak'	Cum peak-readings count		% Change
		Period 1	Period 2	
Th_{S1}	Th_{nr}	$cum\ nr_P$	$cum\ nr_P$	$cum\ nr_P\ change\%$
260	30	3479	3271	-6
270	30	1301	1584	22
280	30	144	282	96

BIASED DATA USED



Table 4.5: Cumulative peak-readings count (biased).

Sensor-S1 readings threshold (°F)	Threshold for 'no. of readings/peak'	Cum peak-readings count		% Change
		Period 1	Period 2	
Th_{S1}	Th_{nr}	nr_P	nr_P	$nr_P\ change\%$
260	30	3514	3875	10
270	30	1180	2072	76
280	30	144	653	353

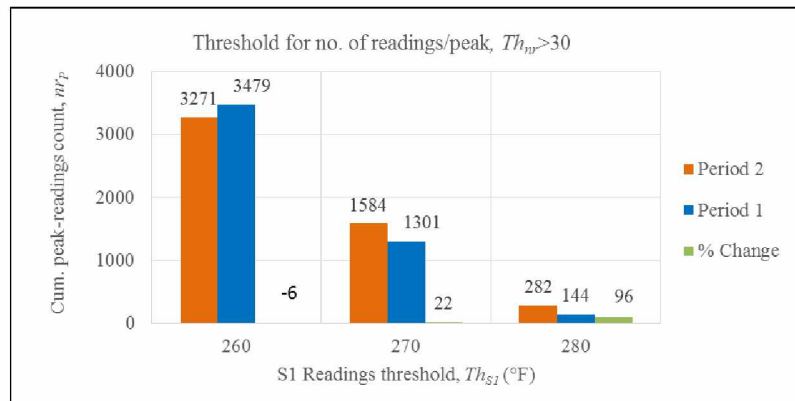


Figure 4.19: Cumulative peak-readings count (clean).

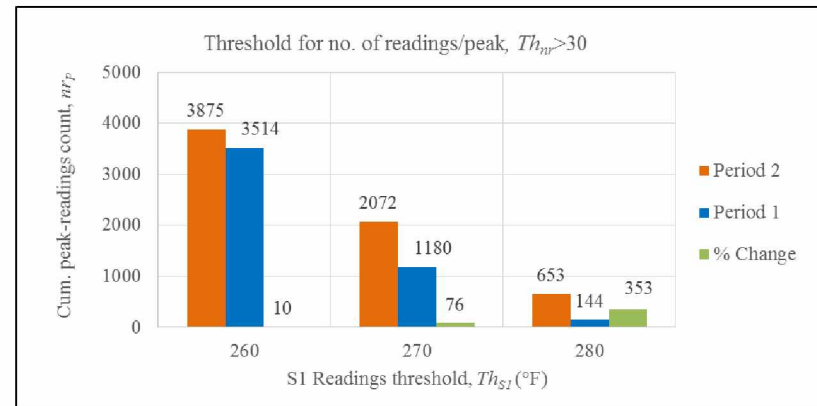


Figure 4.18: Cumulative peak-readings count (biased).

Note: Period-1 denotes both sets clean, while Period-2 denotes one set still clean while the other is +2% biased (see Figure 4.14)

CLEAN DATA USED



Table 4.6: Cumulative no. of peaks count (clean).

Sensor-S1 readings threshold (°F)	Threshold for 'no. of readings/peak'	Cum peak-readings count		% Change
		Period 1	Period 2	
Th_{SI}	Th_{nr}	$cum nrP$	$cum nrP$	$cum nPchange\%$
260	30	3919	3815	-3
270	30	2323	2572	11
280	30	665	969	46

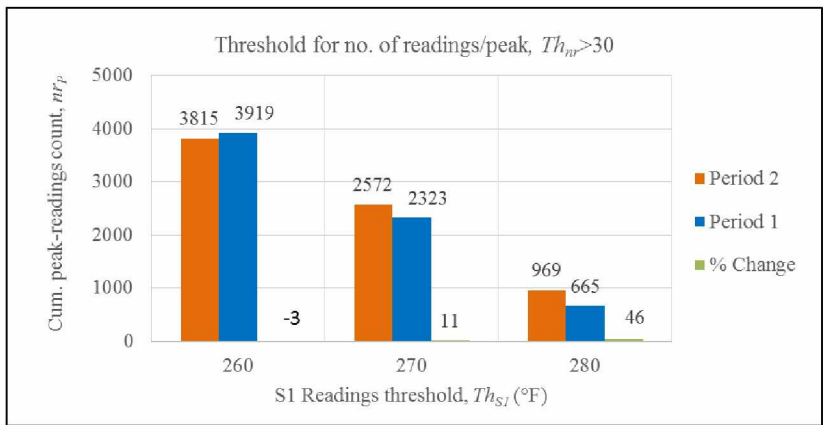


Figure 4.21: Cumulative no. of peaks count (clean).

BIASED DATA USED



Table 4.7: Cumulative no. of peaks count (biased).

Sensor-S1 readings threshold (°F)	Threshold for 'no. of readings/peak'	Cum peak-readings count		% Change
		Period 1	Period 2	
Th_{SI}	Th_{nr}	$cum nP$	$cum nP$	$cum nPchange\%$
260	30	3919	4239	8
270	30	2258	2934	30
280	30	658	1667	153

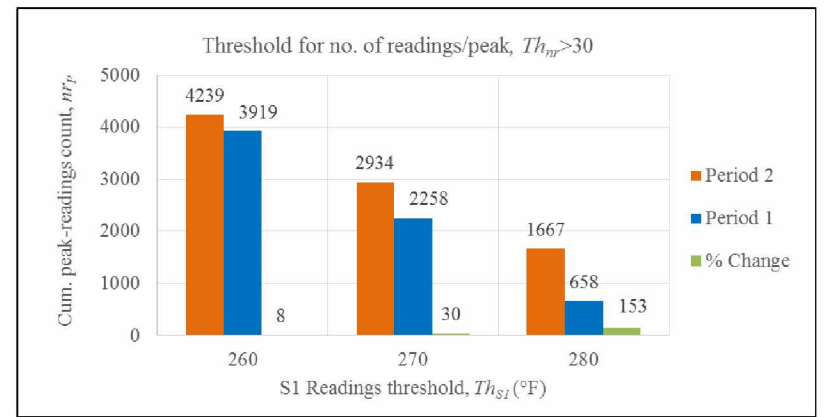


Figure 4.20: Cumulative no. of peaks count (biased).

Note: Period-1 denotes both sets clean, while Period-2 denotes one set still clean while the other is +2% biased (see Figure 4.14)

CLEAN DATA USED



Table 4.9: Cumulative temperature (clean).

Sensor-S1 readings threshold (°F)	Threshold for 'No. of readings/peak'	Cum temperature (x1000)		% Change
		Period 1	Period 2	
Th_{S1}	Th_{nr}	$cum T$	$cum T$	$cum T$ change%
260	30	1070	1049	-2
270	30	645	719	11
280	30	190	279	47

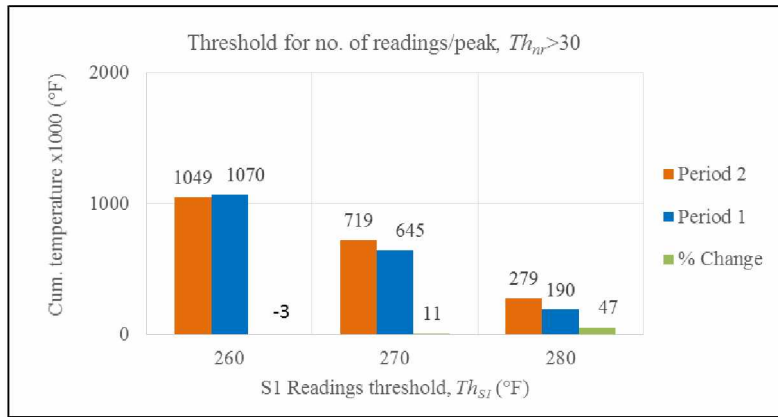


Figure 4.23: Cumulative temperature (clean)

BIASED DATA USED



Table 4.9: Cumulative temperature (biased).

Sensor-S1 readings threshold (°F)	Threshold for 'no. of readings/peak'	Cum temperature (x1000)		% Change
		Period 1	Period 2	
Th_{S1}	Th_{nr}	$cum T$	$cum T$	$cum T$ change%
260	30	1070	1181	10
270	30	627	830	32
280	30	188	481	156

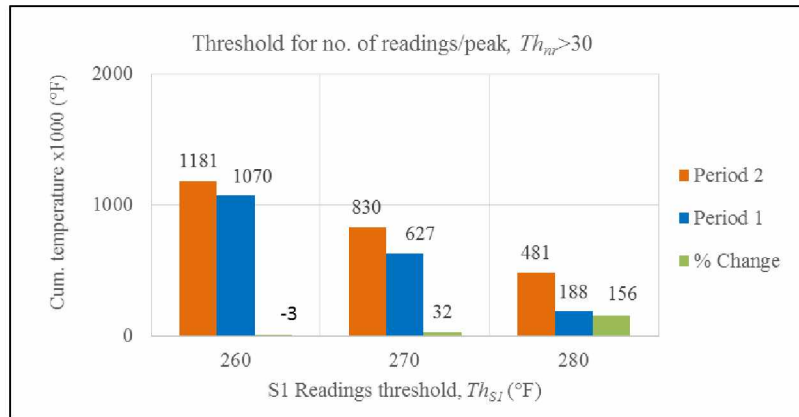


Figure 4.22: Cumulative temperature (biased)

Note: Period-1 denotes both sets clean, while Period-2 denotes one set still clean while the other is +2% biased (see Figure 4.14)

Table 4.10: Cumulative peak-readings count, clean vs biased ($Th_{SI}=280^{\circ}\text{F}$).

Data-set	Sensor-S1 readings threshold ($^{\circ}\text{F}$)	Threshold for 'no. of readings/peak'	Cum peak-readings count		% Change
			Period 1	Period 2	
	Th_{SI}	Th_{nr}	$cum\ nr_p$	$cum\ nr_p$	$cum\ nr_p\ change\%$
Clean	280	30	144	282	96
Biased	280	30	144	653	353

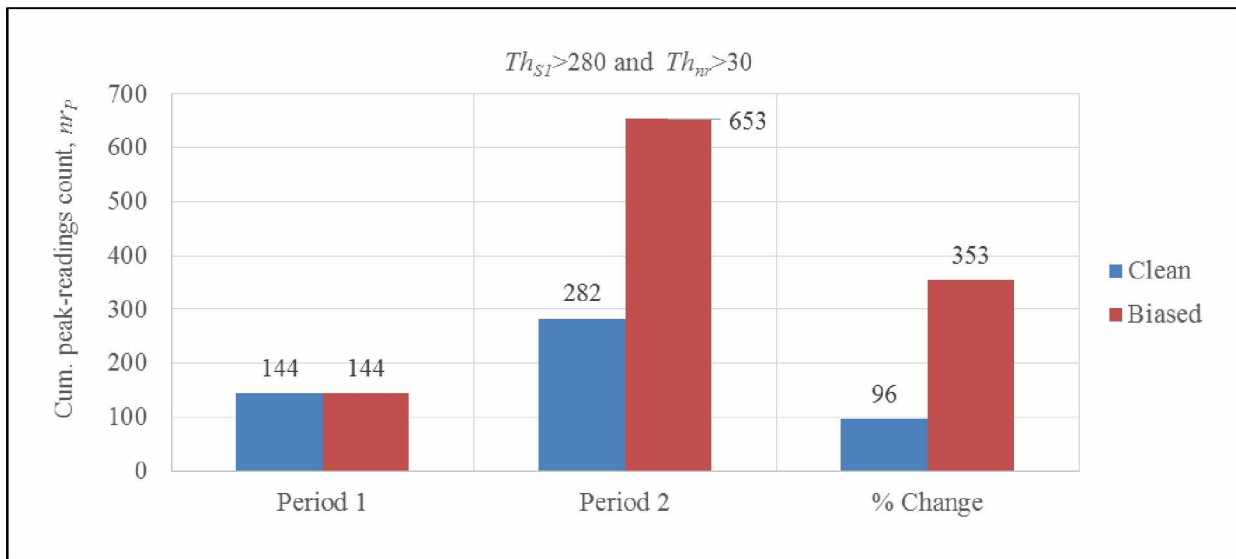


Figure 4.24: Cumulative peak-readings count, clean vs biased ($Th_{SI}=280^{\circ}\text{F}$).

Table 4.11: Cumulative no. of peaks count, clean vs biased ($Th_{SI}=280^{\circ}\text{F}$).

Data-set	Sensor-S1 readings threshold ($^{\circ}\text{F}$)	Threshold for 'no. of readings/peak'	Cum no. of peaks		% Change
			Period 1	Period 2	
	Th_{SI}	Th_{nr}	$cum\ nP$	$cum\ nP$	$cum\ nP\ change\%$
Clean	280	30	665	969	46
Biased	280	30	658	1667	153

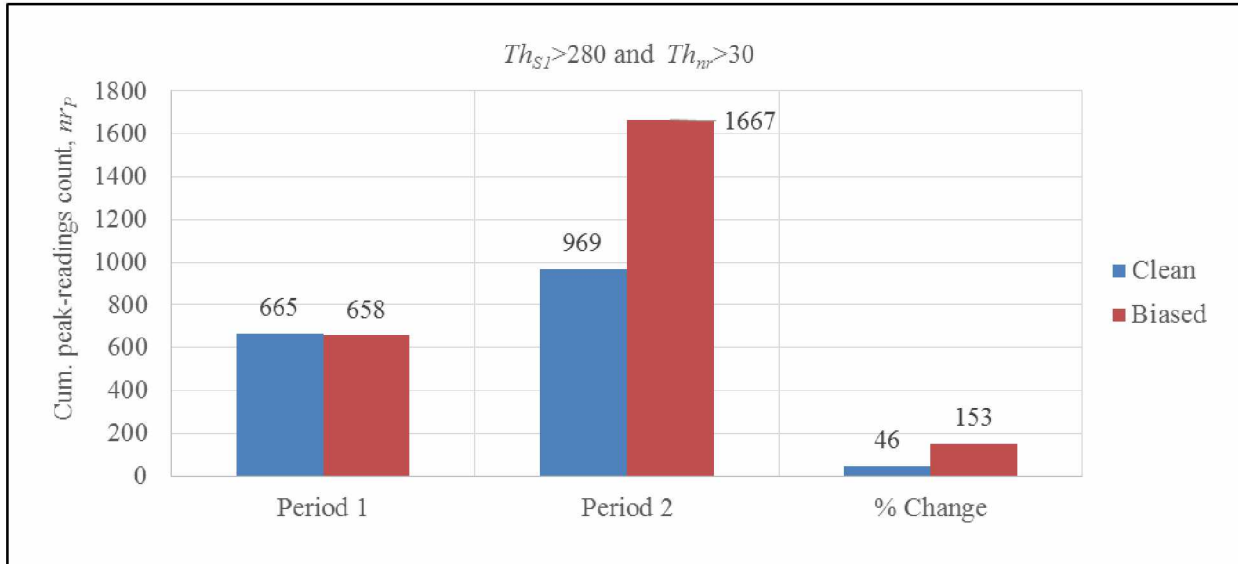


Figure 4.25: Cumulative no. of peaks count, clean vs biased ($Th_{SI}=280^{\circ}\text{F}$).

Table 4.12: Cumulative Temperature, Clean vs Biased ($Th_{SI}=280^{\circ}F$).

Data-Set	Sensor-S1 readings threshold ($^{\circ}F$)	Threshold for 'no. of readings/peak'	Cum temperature (x1000)		% Change
			Period 1	Period 2	
	Th_{SI}	Th_{nr}	$cum T$	$cum T$	$cum T$ change%
Clean	280	30	190	279	47
Biased	280	30	188	481	156

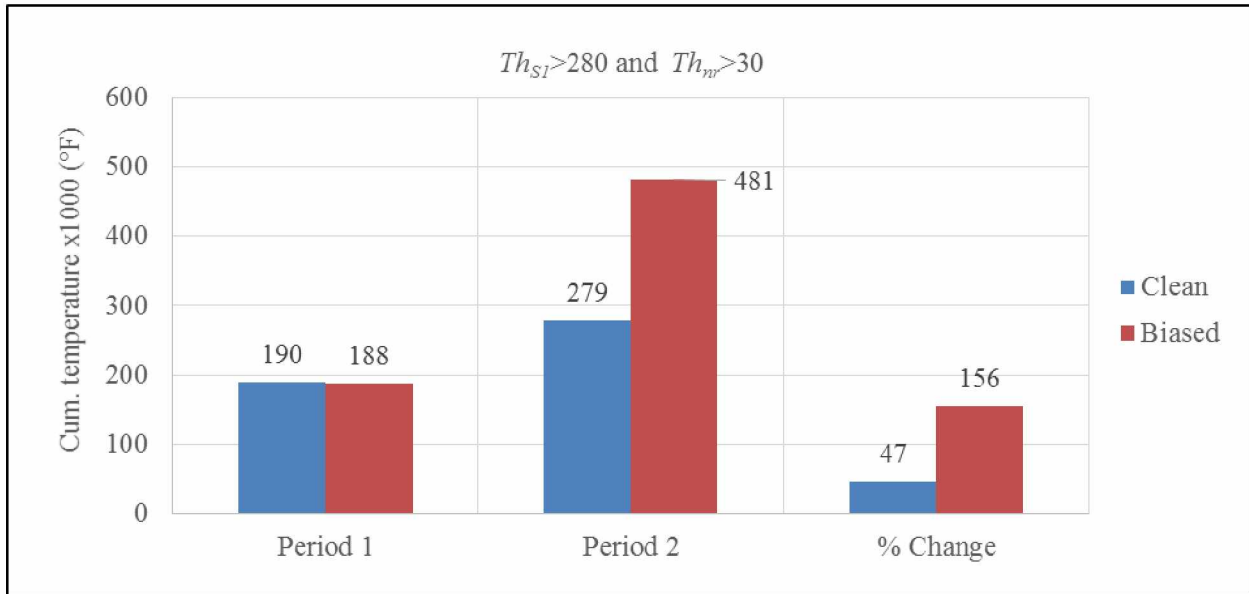


Figure 4.26: Cumulative Temperature, Clean vs Biased ($Th_{SI}=280^{\circ}F$).

4.7 Discussion

It can be observed from the figures that the cumulative number of peaks ($cum nP$) are increasingly sensitive to increasing temperature thresholds ($260^{\circ}F < 270^{\circ}F < 280^{\circ}F$). In order to ensure the difference exist between clean and biased data sets in the presence of bias, cumulative temperature per peak (T_p) values were compared. There is dramatic change in T_p in the presence of bias. Similar trends were observed when $cum nP$ values of clean and biased data sets were compared. For the purpose of the analysis, the plots (Figures 4.15 through 4.17) were divided into two parts, i.e., Part-1 which consists of two clean data sets before the bias was introduced (on

5/17/15), and Part-2 which consists of one clean set and one biased set. In this arrangement, the bias effect in terms of change in the peak-readings count, number of peaks, and cumulative temperatures can be observed clearly.

In Figures 4.15 through 4.17, the biased set has registered a higher number of nr_p or nP values than the clean set; however, the sensitivities varied. Visually, it is easily observed that at 280°F threshold, biased data had more readings count, peaks count, and cumulative temperature counts than the clean data set when compared to other thresholds, i.e., 270°F and 280°F; however, one observation seems anomalous. Even though the cumulative number of peaks for biased data in some places seems higher than that of clean data, the cumulative temperature per peak values seemed to be dipping low. This is because there were many peak-readings registered to make up for that cumulative temperature value. Dividing the cumulative temperature value (sum of all readings in a peak) with the high number of readings make the temperature per peak value appear very low. This seems inconsistent, and contrary to the rise in overall temperature in the biased set (compared to clean set); however, this phenomena indicates the presence of more number of readings per peak. In this context, it can be observed that cumulative temperature of a peak (*cum T*) is a good measure of sensitivity rather than temperature per peak value (T_p). This is the reason the measure was chosen for the sensitivity analysis (PRCSA).

It was observed from Table 4.10 and Figure 4.24 that the change in peak-readings count for biased data—at a Th_{SI} value of 280°F and Th_{nr} value greater than 30 has increased to 353% (Part-2) when compared to clean data (96%) in Part-1. This sudden increase could be attributed to the presence of bias. Likewise, the cumulative peaks and the temperatures increased to 153% and 156%, respectively for biased Part-2. For clean data they stayed at 108% and 109%, respectively (see Tables 4.11 and 4.12 and their corresponding Figures 4.25 and 4.26). The high increase in the

percentage change indicates the presence of bias. Similar results were observed for negative bias analysis. Sensor (S2) data results are not presented for space constraints; however, the results demonstrated similar trends.

4.8 Conclusions

Calibration errors or biases of small magnitude ($\pm 2\%$ of a sensor's true reading) are hard to detect through classical statistical methods that are otherwise applied to detect gross errors. On the other hand, sophisticated digital signal processing methods like Fast Fourier Transform (FFT) depend heavily on periodicity of a sensor signal. Data-mining methods seem to be a better choice by contrast due to their flexibility and ability to work on large data sets; however, innovative methods that are tailored to tackle calibration-related errors like bias are needed. In this connection, peak-readings count analysis (PRCA) and peak-readings count sensitivity analysis (PRCSA) were developed to observe effects of bias on a clean set of sensor data (S1 data). While PRCA captures various key statistics like number of readings in a peak, number of peaks, cumulative temperature for the biased and clean sets, PRCSA tests the sensitivity of these statistics at certain thresholds. The sensitivity statistics showed that there is a dramatic increase in some key statistics, i.e., peak-readings count, cumulative number of peaks and cumulative temperature values for biased data, when compared to clean data. The change in these statistics is more drastic, particularly where number of consecutive peak-readings is greater than 30. This is a key finding from the analyses that is worth exploring more on future data. Chapter 5 provides a detailed account on such ventures.

Identification of sensor biases in a short span of time after they occur is important. If they could be identified within a shorter span of time (few weeks), corrective actions could be taken immediately. Obtaining a clean set of data from frequent calibration processes might cause disruption to the production process and is therefore not practical for industrial operators. This is

one major disadvantage when using PRCSA method, due to the requirement of clean set of data for the sensor of bias observation. Exploiting sensor interrelations might be viable option. In this chapter, the methods are applicable to a single sensor scenario only.

4.9 Acknowledgements

The author expresses his sincere gratitude to his academic advisor Professor Rajive Ganguli, all the members of the graduate advisory committee for their help and guidance, and to the Mining Engineering Research Endowment (MERE) established at the University of Alaska Fairbanks (UAF) for financial support. The author also would like to thank Fort Knox Mine of Kinross Corporation and Pogo Mine of Sumitomo Metal Mining Company for their support of MERE and for providing the necessary data, facilitating site visits, etc. The author extends his gratitude to College of Engineering and Mines (CEM), and, the Department of Mining and Geological Engineering (MinGeo) at UAF for providing the valuable laboratory and other academic resources to accomplish the project.

4.10 References

- Arku, D., and Ganguli, R., 2014, "Investigating utilization of aggregated data: does it compromise information gleaning?" *Mining Engineering*, 2014, Vol. 66, No. 6, pp. 60-65.
- Baljak, V., Tei, K., and Honiden, S., 2012, "Classification of faults in sensor readings with statistical pattern recognition," *SENSORCOMM 2012: The Sixth International Conference on Sensor Technologies and Applications*, Aug 19-24, 2012, Rome, Italy, pp. 270-276.
- Brown, M., 2012, "Data mining techniques," IBM, www.ibm.com/developerworks/library/ba-data-mining-techniques/, Accessed October 2017.

Buxton, M., and Benndorf, J., 2013, "The use of sensor derived data in real time mine optimization: a preliminary overview and assessment of techno-economic significance," SME Annual Meeting, Feb 24 - 27, 2013, Denver, CO, Preprint 13-038, pp. 5-5.

Fast, J., 2016, "Carbon stripping," Denver Mineral Engineers, Inc., www.denvermineral.com/carbon-stripping/, Accessed October 2017.

Hou, Z., Lian, Z., Yao, Y., and Yuan, X., 2006, "Data mining based sensor fault diagnosis and validation for building air conditioning system," Elsevier, DOI:10.1016/j.enconman.2005.1.010.

Konigsmann, E., Abols, J., Pyecha, J., and Boudreau, T., 2017, "The inline leach reactor installation at Teck-Pogo Inc. Pogo Mine," www.gekkos.com/documents/033TheInLineLeachReactorInstallationAtTeckPogIncPogoMine.pdf, pp.1-11, Accessed October 2017.

Kusiak, A., and Song, Z., 2009, "Sensor fault detection in power plants," Journal of Energy Engineering, December, 2009, No. 127, pp. 127-135, DOI: 10.1061/(ASCE)0733-9402(2009)135:4 (127).

Pogo Mine, 2016, "Pogo Mill Process Descriptions," Print medium handout by Andrew Maxon, Pogo Mill Facility, Pogo Mine, Received January 2016.

Pogo Mine, 2017, "Who we are," <http://pogominealaska.com/>, Accessed October 2017.

U.S. Department of Energy, 2000, "Mineral processing technology road map," U.S. Department of Energy, www.eere.energy.gov/industry/mining/, Accessed May 2006.

Wang, H., Chai, T., Ding J., and Brown M., 2009, "Data driven fault diagnosis and fault tolerant control: some advances and possible new directions," *ACTA Automatica Sinica*, Vol. 35, No. 6, June 2009, pp. 739-747.

Yang, C., Liu, C., Zhang, X., and Nepal, S., 2015, "Time efficient approach for detecting errors in big sensor data on cloud," *IEEE Transactions on Parallel and Distributed Systems*, February, 2015, Vol. 26, No. 2, pp. 329-339, DOI: 10.1109/TPDS.2013.2295810.

Chapter 5: Multiple Ratio Function Analysis with Automation (MRFAA) in the Detection of Industrial Sensor Calibration Errors

5.1 Abstract

Sensor calibration errors are causing millions of dollars in losses in the mining industry, and due to their subtlety, are hard to identify - “bias” is one of such error. In this chapter, an innovative data-mining approach, multiple ratio function analysis with automation (MRFAA) that was used in the detection of bias ($\pm 2\%$) in the strip vessel sensor (S1) of a carbon stripping circuit in the Pogo Mine is described. Several data-mining based innovative methods developed and their corresponding algorithms that exploit the sensor interrelations (“ratio functions”) also are described. At first, a single test—S1 to Heat ratio—algorithm, multiple ratio function analysis (MRFA), was developed. The algorithm has disadvantages in finding bias when strip vessels’ temperatures ran below a certain threshold, i.e., “truncation threshold.” A “dynamic thresholding strategy” with the addition of automated tests improved MRFA performance. The resultant algorithm (MRFAA) was capable of conducting multiple tests along with a “Combined test.” The Combined test has 95% success rate of finding +2% bias after it was induced (True alarms). The time taken by all the tests together to find the artificially induced bias, i.e., time till find days (TTFD), was within 39.5 days for 75% of the time, at the cross-score threshold of 5 and 6. Among the individual tests, the “Heat ratio test” is the better test; it is able to find the induced bias within 33 days for 75% of the time, at the cross-score thresholds of 5 and 6. The heat ratio, “average value” and “maximum value” tests showed comparable results; however, BARNFL and GLYFL ratio tests did not perform well due to their poor response to the dynamic thresholding strategy. The negative bias analysis also produced similar results overall. Identifying bias at such lower

magnitudes ($\pm 2\%$) is a hard task to achieve, and with the MRFAA algorithm, it is identified within approximately one month span of time for 75% of the time, which is very valuable to the industry.

5.2 Introduction

Usage of sensors in monitoring industrial processes has become increasingly prevalent in the recent past, and the mining industry is no exception. In fact, a recent study finds that usage of sensors in various stages of mining operations—for a moderately sized mine—can create millions of dollars in economic value; usage of sensors in mineral processing operations can yield \$10-100 million in added economic value annually (Buxton and Benndorf, 2013). The opposite effect is true, however, when sensors suffer from faults and produce erroneous data. For instance, it was observed that sensor faults are causing approximately 3-8% production loss to the US oil industry, resulting in \$20 billion in annual losses to US economy (Wang et al., 2009). Identifying the sensor faults, and fixing them through a calibration process can dramatically improve the sensors' accuracy. Such processes can reduce the equipment downtimes, increase production, and improve overall safety in the industry. The common type of faults, like noise, failures (flat-outs), stuck-at-faults, etc. (gross errors), can easily be detected and fixed through preventive maintenance and calibration processes; however, calibration bias related errors are hard to identify. These are the errors that are insidious, creep-up over time, and often times indistinguishable if the process is particularly non-linear or non-stationary. Data-mining techniques are of significant use in such situations. Calibration errors are present in the data often in the form of an added offset value (bias) to the original (true) reading. For the purpose of this research, calibration errors are those that occur in the form of a bias. Innovative techniques that are based on data-mining concepts are explored in this chapter in the pursuit of developing methodology towards finding solutions. Other

standard errors were thoroughly explored by the previous researchers and the methodology is well-established to a large extent. Hence they are out of the scope for this research undertaking.

Sensors that monitor various operations in the carbon stripping circuit in the Pogo Mine, Alaska, were chosen for the study. The multi-sensor environment was exploited in the hope of finding induced bias in strip vessel temperature sensor S1.

5.3 Literature Review

Data-mining techniques are broadly classified as: “association” based, where relation between variables is exploited; “classification” based, where attributes of each class of items are studied, “clustering” based, where one or more attributes of the classes are examined and grouped together; and “pattern recognition” based, where identifying trends or regular occurrences is used (Brown, 2012). There are some other techniques that use a combination of any of the above. The “prediction” based techniques use classification, pattern recognition, and association (an example is forecasting of a company’s stock performance). The “decision trees” use classification and prediction together (an example is, classification of various sensor faults). In general, real world problems might require several combinations of these techniques or entirely innovative approaches that are specific to a particular problem.

Data-mining techniques have a wide range of industrial applications, and are generally used for extracting knowledge from huge data sets—industrial sensor data sets are one example. For highly non-linear and non-stationary processes, however, it is difficult to develop analytical models based on classical or fundamental statistical methods. Moreover, sensor validation for such processes is equally tough due to production of false alarms in excessive amounts (Kusiak and Song, 2009). Researchers Kusiak and Song (2009) were able to develop several algorithms based

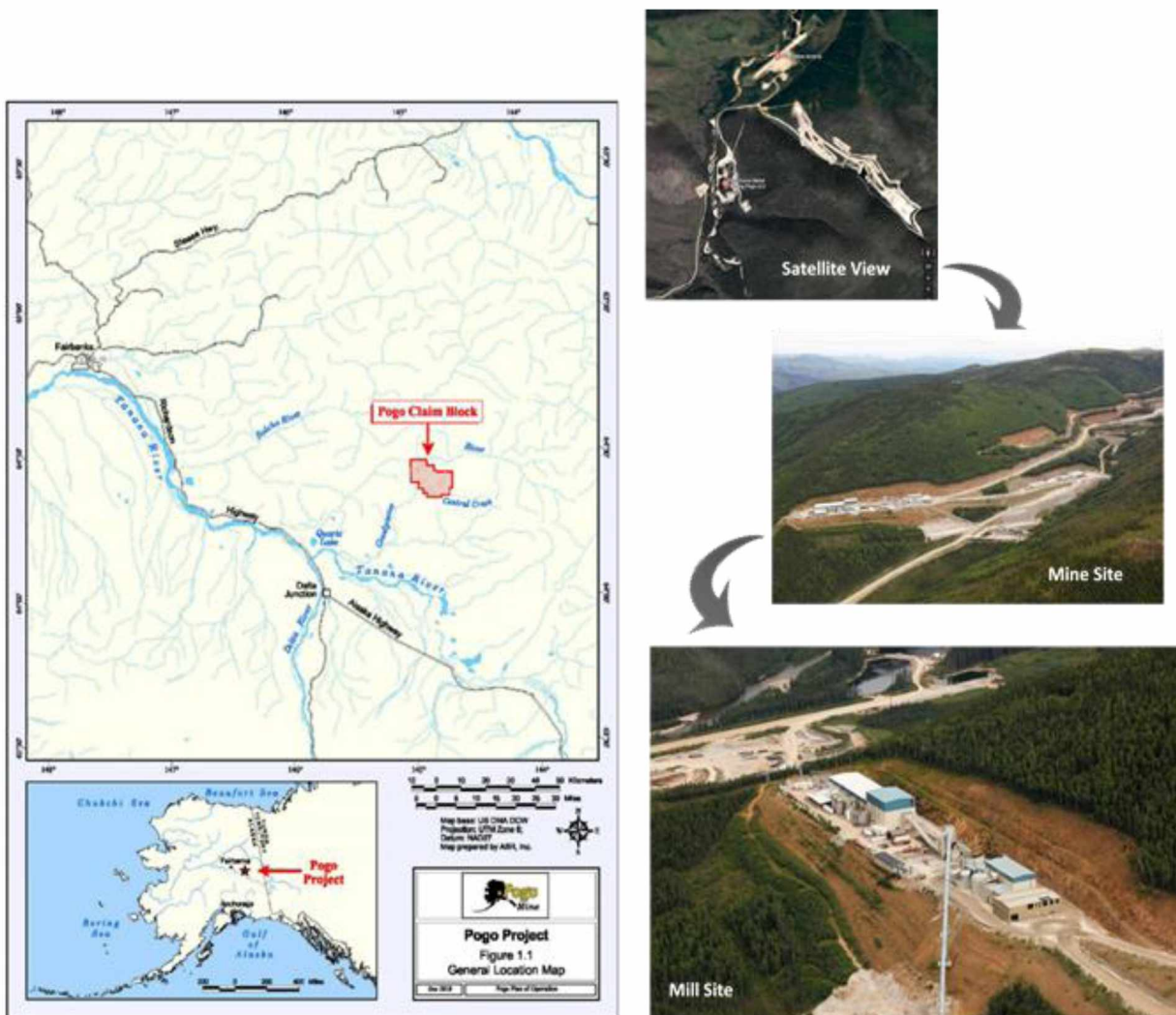
on a clustering technique to detect sensor faults in power plant boilers. Classification techniques were used in a “cloud” based application to detect errors in a big sensor data set (Yang et al., 2015). A data-mining approach that employed “rough set” theory along with the artificial neural networks (ANN) was used by Hou et al. (2006) to identify sensor faults in a heating, ventilating and air conditioning (HVAC) system. The rough set theory is effective in the classification of uncertain and incomplete information, and any data set that met the criteria could be improved by using the rough set methodology. Historical performance data of a HVAC system in a building was used in the study. In addition, authors were able to describe several algorithms that were used in this connection. Using “decision-trees algorithms,” Baljak et al. (2012) developed a methodology to classify sensor faults. The authors based their classification on continuity and frequency of occurrence of a fault.

The bulk of the literature is devoted to data-mining approaches for finding gross-errors in the sensor data; thus, the motivation in this chapter is to develop innovative methodology to identify calibration-related errors. Approaches that employ a combination of data-mining methods, i.e., classification, decision-trees, and pattern recognition, are used in the process. Mineral processing is an energy intensive process and accounts for 39% of the energy consumed in mineral production (U.S. Department of Energy, 2000). The U.S. Department of Energy (DOE) identified sensors’ improvements as one of the opportunities to save energy in the mineral industry.

5.4 Introduction to Pogo Mine and Mill

The sensor data that was used for experimentation in this chapter is collected from Pogo Mine’s mill facility. Pogo Mine is a major gold producer in Alaska (Figure 5.1). Located on the Goodpaster River, 38 miles (61 km) north of Delta Junction in east-central Alaska, the nearest city to Pogo is Fairbanks, located approximately 70 miles (112 km) northwest of the mine property

(Konigsmann et al., 2017). Pogo is an underground operation. Elevations on the property range from 1,299 ft (396 m) on the Goodpaster River to over 4,003 ft (1,220 m) on the top of Pogo Ridge, an east-west trending ridge. The climate is classified as sub-Arctic with cold, dry winters and relatively mild summers (Konigsmann, 2017).



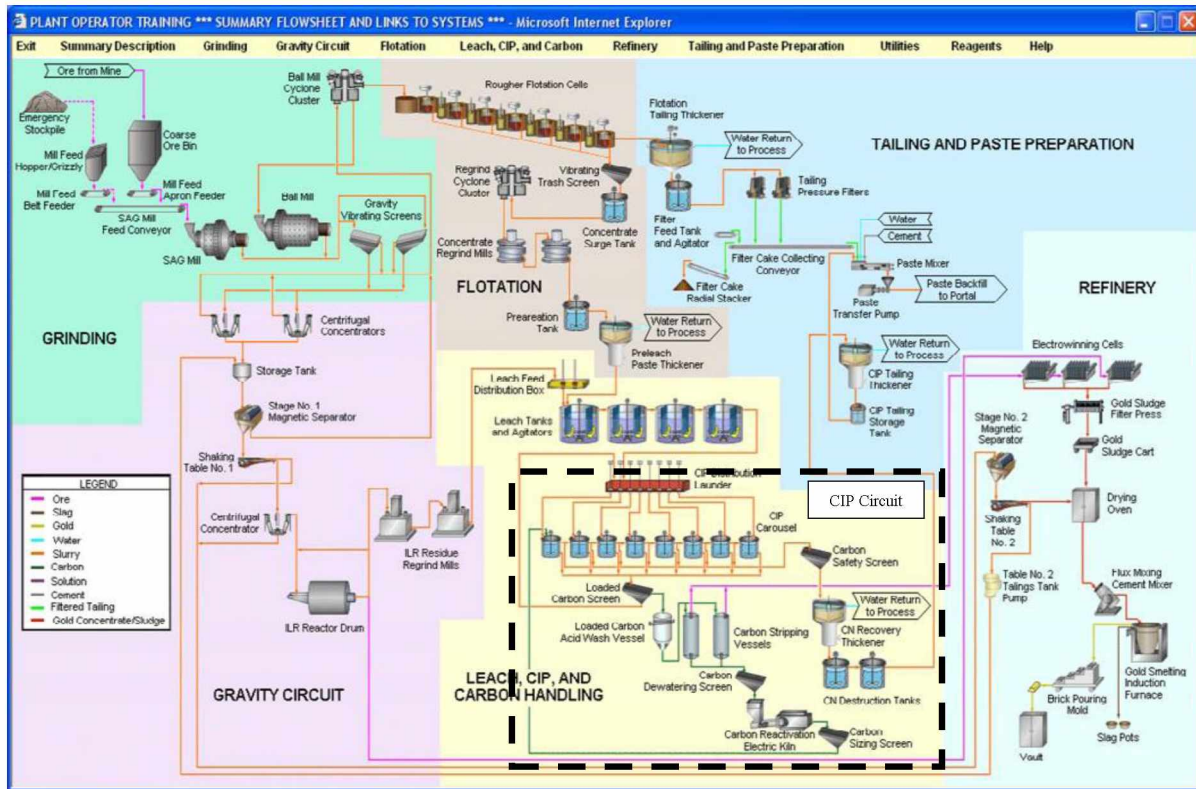
Source: Adapted from Pogo Mine, 2017.

Figure 5.1: Pogo Mine site map.

Pogo Mill processes up to 3,500 tons of ore daily. The process flowsheet is shown in Figure 5.2. The Pogo plant's process flow mainly is comprised of a crusher, semi-autogenous (SAG) mill,

ball mill, floatation circuit, leaching tanks, carbon-in-pulp CIP circuit (described in detail in section 5.4.1), stripping circuit, and an electro-winning circuit. The Pogo plant is a closed-circuit operation; all the water used is recirculated and cleaned before it is released into a pond to minimize environmental impact. Up to 67% of the gold is recovered through the processing facility. A brief description of the process follows (Pogo Mine, 2017).

Ore is fed to a conventional SAG/ball mill grinding circuit. Gravity recovery is the technique employed throughout the process. The ball mill circulating load is screened at one (1) mm, and the undersize is fed to two 48-inch centrifugal concentrators operating in parallel. The primary gravity concentrate is then fed to the intensive cyanidation circuit with the leach solutions reporting directly to electro-winning for final gold recovery. Primary gravity tailings are returned to the grinding circuit via the cyclone feed pump-box while the intensive leach residues are reground and pumped into the flotation concentrate leach circuit. Primary gravity gold concentrates are intensively leached on a batch basis; the typical leach residence times are approximately 14 hours. Grinding cyclone overflow reports to a sulphide rougher flotation circuit, which produces a 10% weight concentrate. The concentrate from the floatation circuit is then reground to 80% passing 10 microns using stirred media detritors, prior to being leached in a conventional cyanidation circuit followed by gold recovery in an eight cell carousel CIP circuit. The particles then enter a stripping circuit. A detailed description of the process can be found in the subsequent sections.



Source: Adapted from Pogo Mine, 2016.

Figure 5.2: Pogo Mine-gold processing flow sheet.

5.4.1 Carbon-in-Pulp (CIP) Circuit

The purpose of the CIP process is to allow the gold previously dissolved in the leach tanks to be adsorbed by activated carbon—activated carbon is a form of charcoal that has a large number of low volume pores that help in the adsorption of the fine particles. During the CIP process, the gold particles are slowly adsorbed onto the carbon particles, and eventually extracted in the subsequent “stripping” process. The CIP circuit is designed to allow adequate time for the absorption process. The particle-bearing slurry spends 30 minutes in each CIP tank. The Pogo CIP circuit has eight such tanks (Figure 5.3). The slurry spends four hours in the circuit, and approximately 300 to 600 ounces of gold is adsorbed per ton of activated carbon used.



Source: Adapted from Pogo Mine, 2016.

Figure 5.3: CIP circuit at Pogo Mine.

5.4.2 Carbon Stripping Circuit

The Pogo stripping circuit in pictures is shown in Figure 5.4. Pogo stripping circuit consists of two strip vessels that work in tandem. While one vessel (vessel-1) is being loaded with gold bearing activated carbon, the other previously loaded vessel (vessel-2) is operated by circulating a solution called “elute.” The temperature and pressure of elute is maintained approximately at 280°F and 65 PSIG, respectively, to facilitate maximum liberation of gold particles. An elute is a water based solution with 1% sodium hydroxide and 0.1% sodium cyanide (Fast, 2016). The process is called “pressurized Zadra stripping.” A typical pressurized Zadra stripping cycle lasts for 11 hours and consists of the following stages: loading the vessel (1 hr), circulating elution (8 hrs), carbon cooling (1 hr), and unloading carbon from the vessel ($\frac{1}{2}$ hr) (Table 5.1).

Table 5.1: Operating schedule-pressure Zadra stripping.

Operation	Solution	Time
Load Column	Transfer Water	90 minutes
Elution	0.1% NaCN, 1% NaOH	480 minutes
Carbon Cooling	Fresh Water	60 minutes
Unload Column	Transfer Water	30 minutes
TOTAL		11 hours

Source: Fast, 2016.

While the used carbon is discharged, the “pregnant leach” solution is pumped out. The same process is repeated with strip vessel-2. The pregnant leach solution (PLS), on its way out from the strip vessel, is cooled off by heat exchangers 3 and 4 (Figure 5.5). When the PLS reaches the electro-winning circuit, the gold particles are removed and the solution, now called “barren solution,” is reheated by a boiler with the aid of heat exchangers 1 and 2. The reheated barren solution is then recirculated through strip vessels. A glycol solution is circulated between the boiler and heat exchangers as a medium of heat exchange. Sensors are strategically placed at various parts of the circuit to measure temperatures, flow rates, etc. It is very important to maintain temperatures in the strip vessels at certain levels (270-280°F) for certain periods of time to maximize gold separation. A false “optimal temperature” will result in either poor gold recoveries or higher costs. Thus, monitoring the temperatures, and identifying sensor errors became crucial. S1 and S2, the temperature sensors for strip vessels 1 and 2, respectively, are the two important sensors in this context, and therefore are the focus of this chapter.

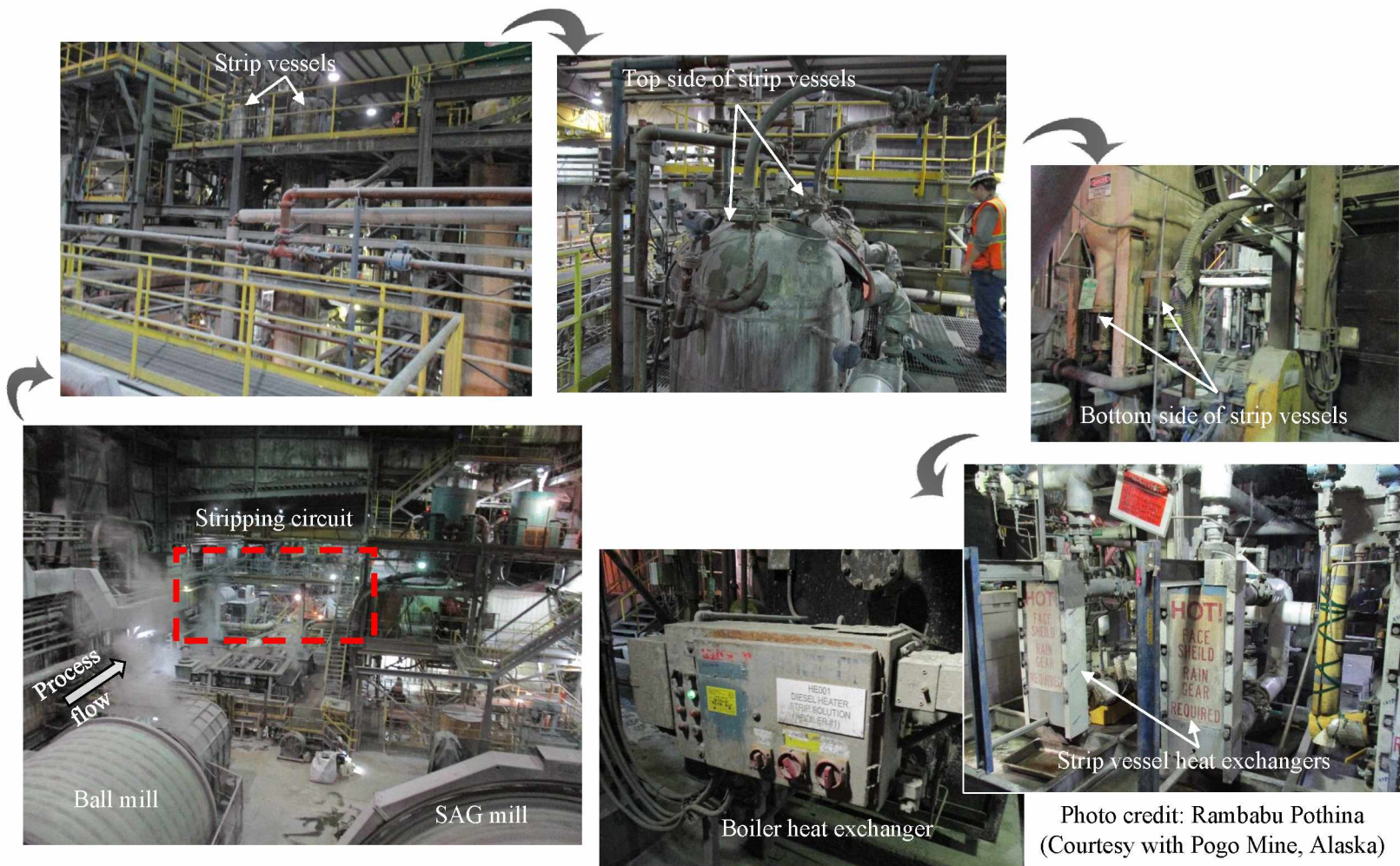
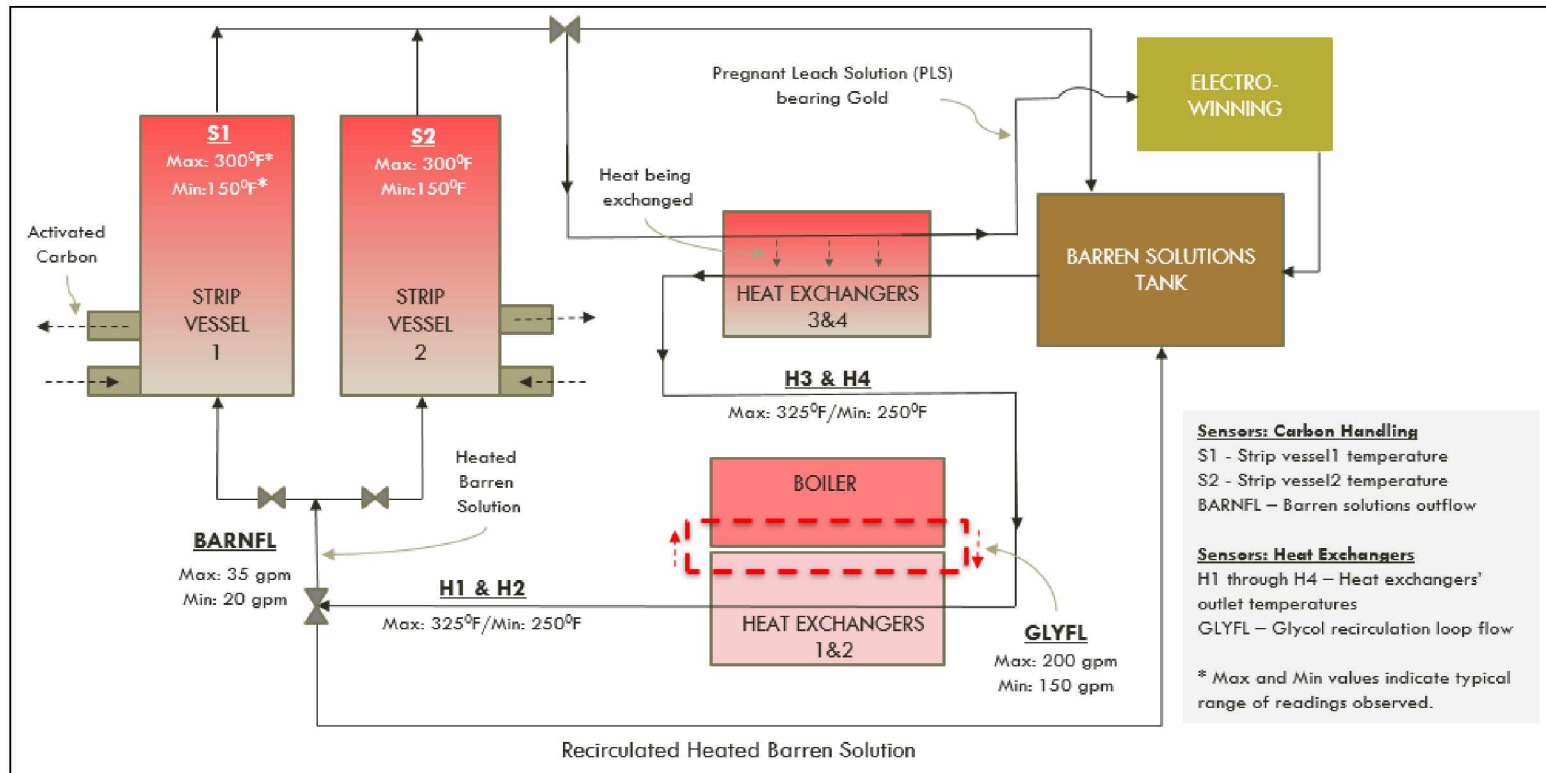


Photo credit: Rambabu Pothina (Courtesy with Pogo Mine, Alaska)

Figure 5.4: Pogo stripping circuit in pictures.



Source: Pogo Mine, 2016.

Figure 5.5: Pogo stripping circuit schematic diagram with sensor placements.

5.5 Methods and Materials

The description of research methodology in this chapter is achieved in the following stages: data collection, data preparation, background on various concepts, terminology and previous methods, and the description of various algorithms developed.

5.5.1 Data Collection

The raw sensor data was collected at 10-min average intervals from the Pogo mill database. Various sensors of interest and their data readings are summarized in Table 5.2. To facilitate the observation of data trends throughout the year, data was collected for a period of twelve (12) months: Jan 1, 2015 through December 31, 2015. The visual format of various sensor raw data streams with close-up views can be seen in Figures 5.6 through 5.10. Various descriptive statistics for the sensor data are given in Table 5.3. The “data cleansing cut-off value ($Th_{cleanse}$)” column in Table 5.3 refers to the cut-off value based on which undesired and corrupted data were removed (cleaned). This process is explained in the data preparation section.

Table 5.2: A snapshot of raw sensor data collected at 10-min average intervals.

Reading #	Time	Strip vessel sensors		Barren flow sensor	Heat exchanger sensors				Glycol flow sensor
		S1 (°F)	S2 (°F)	BARNFL (GPM)	H1 (°F)	H2 (°F)	H3 (°F)	H4 (°F)	GLYFL (GPM)
1	1/1/15 12:00 AM	91	262	35	299	301	239	104	204
2	1/1/15 12:10 AM	90	262	35	292	293	237	104	203
3	1/1/15 12:20 AM	90	262	35	284	285	236	103	200
4	1/1/15 12:30 AM	90	234	35	277	278	228	102	201
5	1/1/15 12:40 AM	90	219	35	271	272	208	101	202
6	1/1/15 12:50 AM	90	217	35	264	265	196	100	205
7	1/1/15 1:00 AM	90	215	35	258	259	180	96	204
8	1/1/15 1:10 AM	90	204	35	252	253	185	92	205
9	1/1/15 1:20 AM	89	192	35	247	247	182	91	202
10	1/1/15 1:30 AM	89	190	35	241	242	176	90	201

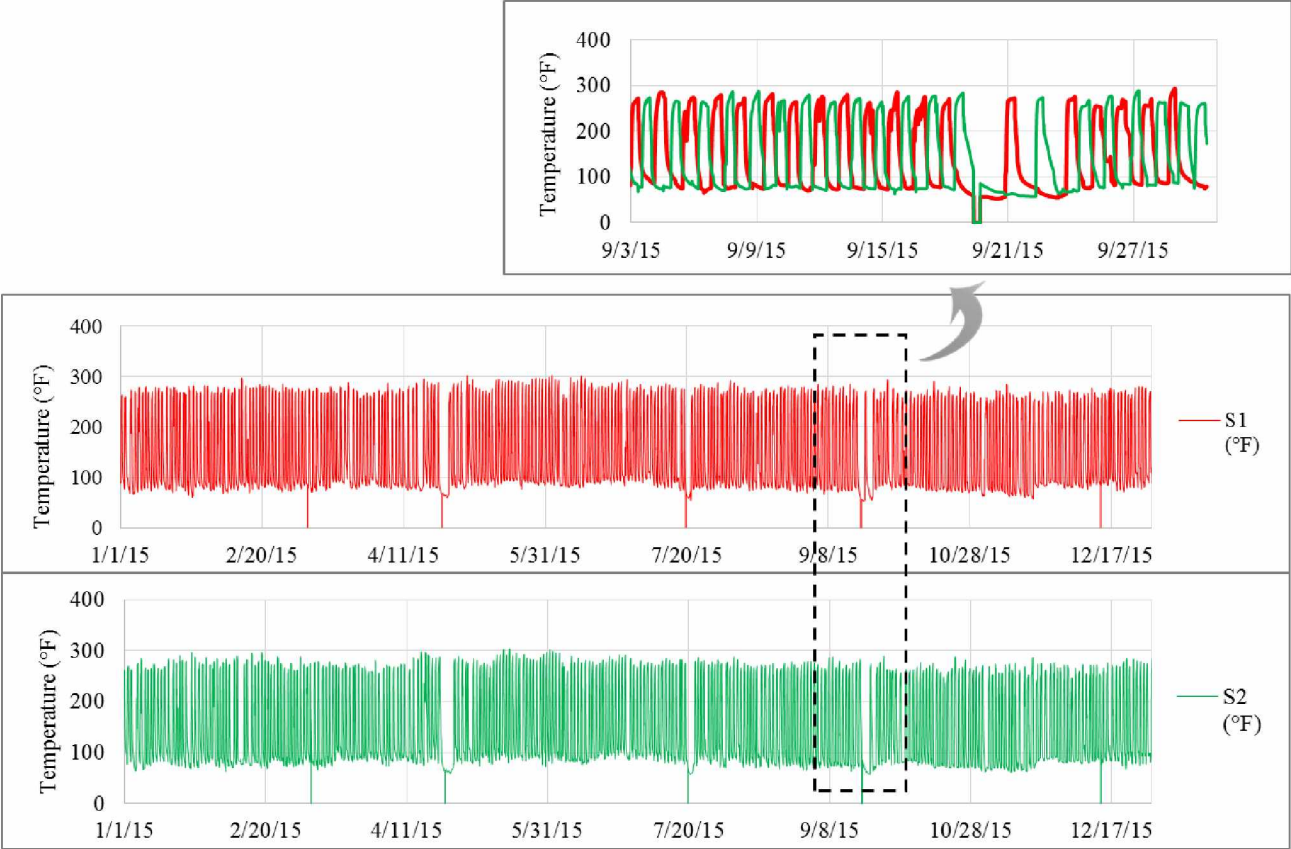


Figure 5.6: Strip vessel sensors (S1 and S2).

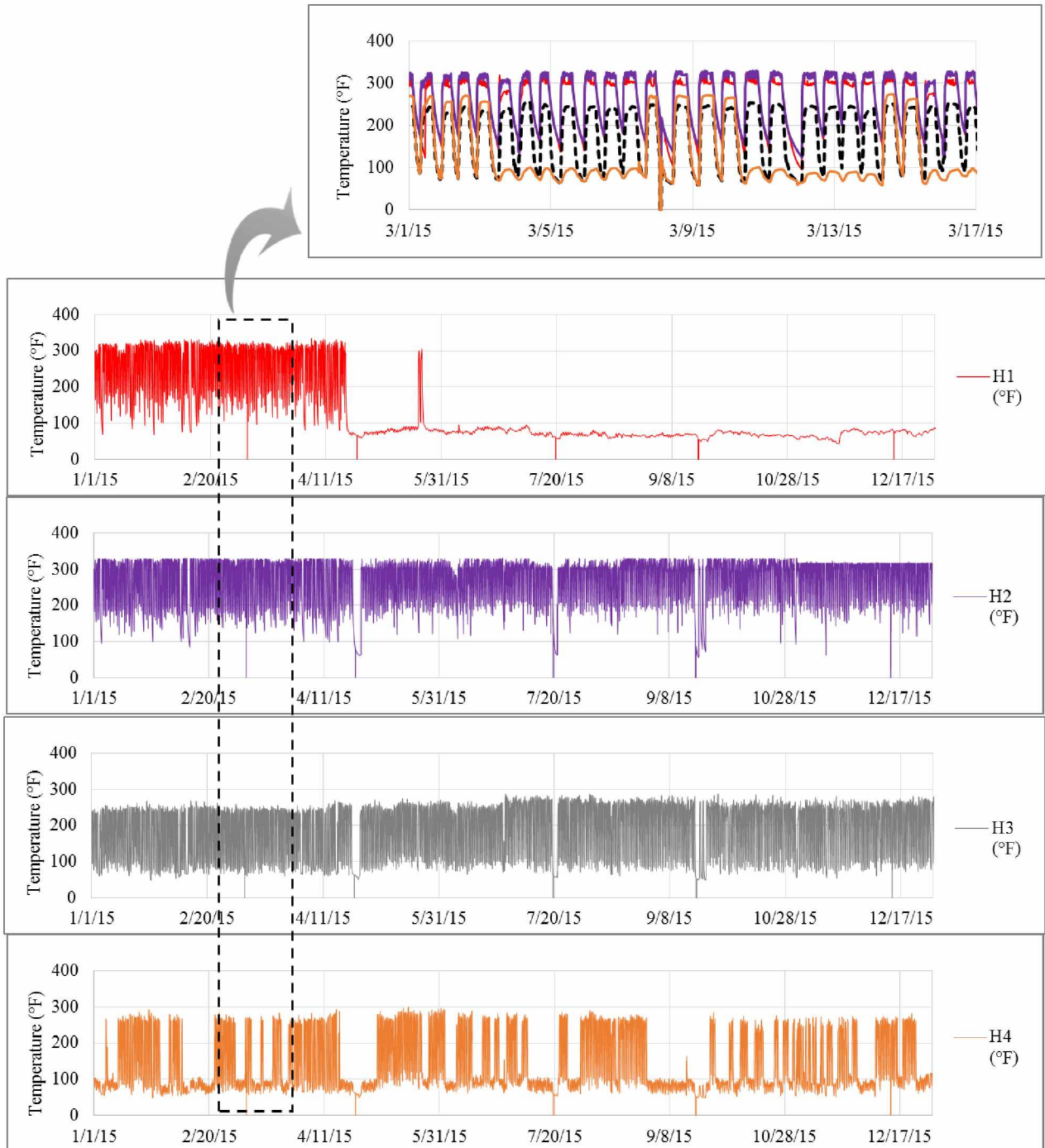


Figure 5.7: Heat Exchanger sensors (H1, H2, H3, and H4).

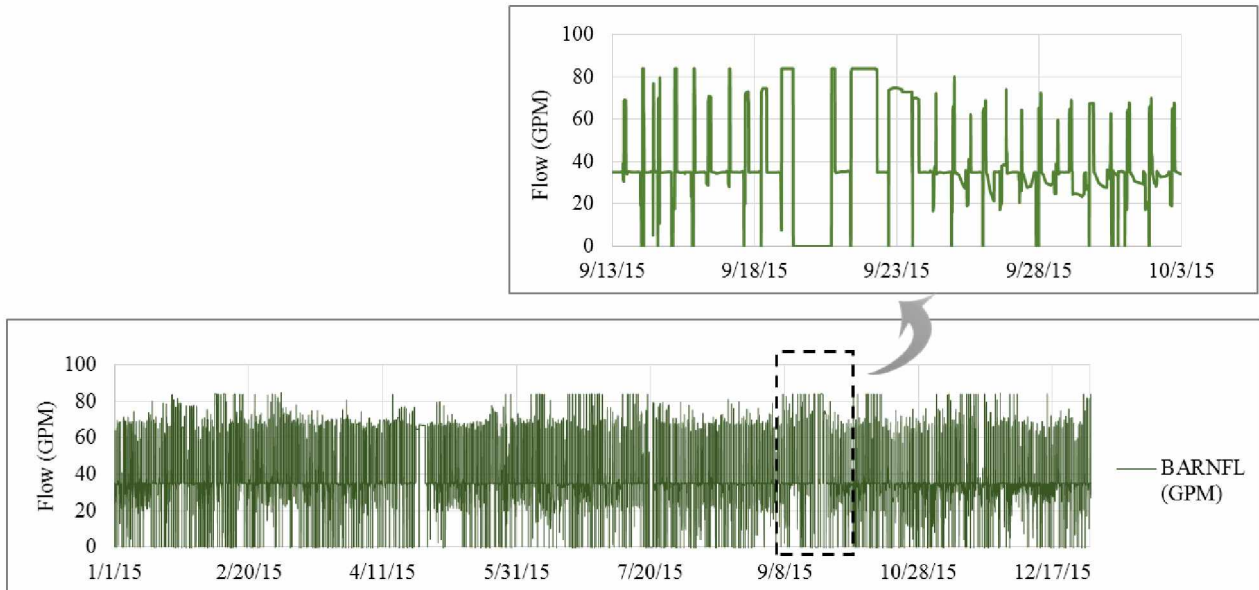


Figure 5.8: Barren flow sensor (BARNFL).

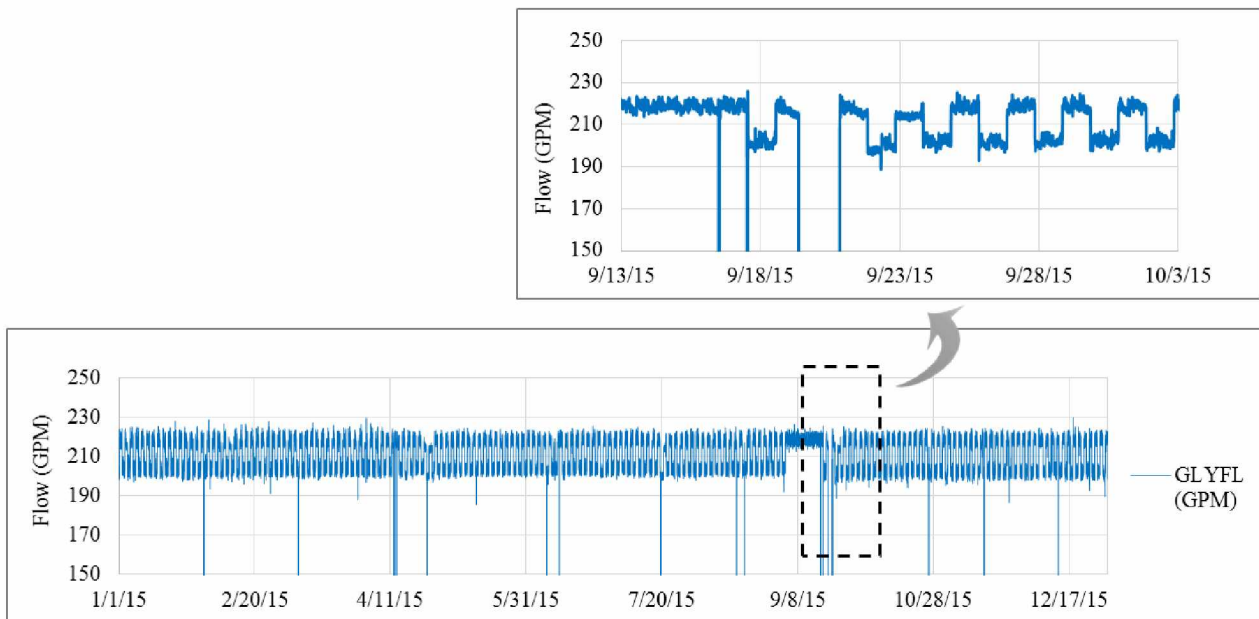


Figure 5.9: Glycol flow sensor (GLYFL).

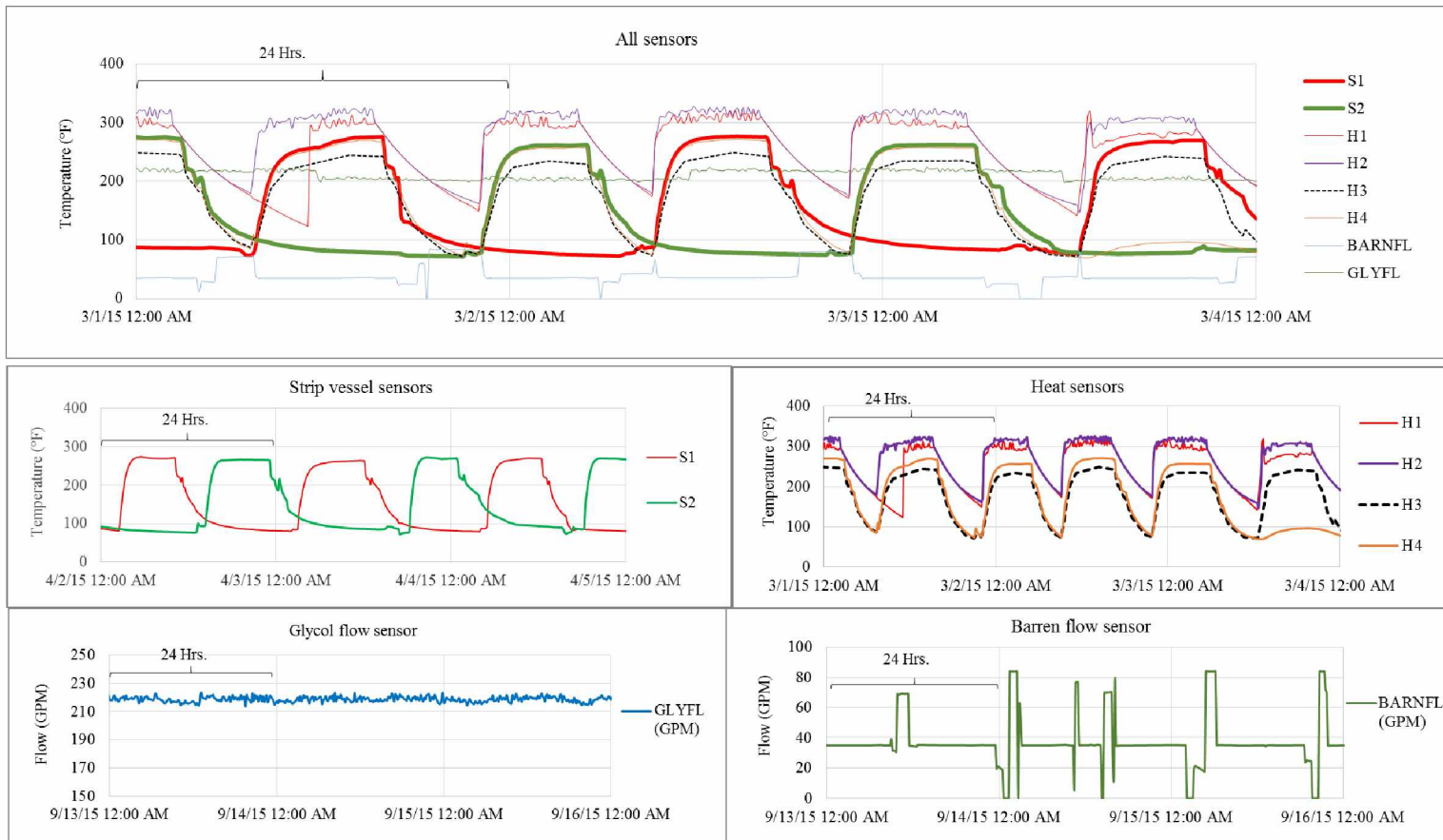


Figure 5.10: Behavior of various sensors in 24 hours time span.

Table 5.3: Descriptive statistics for the sensor data collected.

Sensor	Measuring parameter/variable	Unit	Maximum	Minimum	Mean	Median	Standard deviation	Data cleansing cut-off value ($Th_{cleanse}$)
S1	Strip vessel-1 Temperature	°F	302.7	52.1	149.4	104.1	78.2	150
S2	Strip vessel-2 Temperature	°F	304.3	57.7	148.1	104.2	77.6	150
H1	Heat Exchanger-1 outlet Temperature	°F	334.0	42.9	126.1	77.1	89.8	250
H2	Heat Exchanger-2 outlet Temperature	°F	336.2	56.3	266.4	297.9	60.2	250
H3	Heat Exchanger-3 outlet Temperature	°F	287.6	48.4	191.4	223.5	71.0	250
H4	Heat Exchanger-4 outlet Temperature	°F	299.5	48.5	144.5	99.0	79.2	250
BARNFL	Barren Solution Flow	GPM	84.0	0.0	39.6	35.0	17.3	20
GLYFL	Glycol Flow	GPM	230.0	0.0	209.7	214.6	18.4	150

5.5.2 Assumptions and scope

Calibration errors are often present in the data in the form of an offset value (bias) added to the original or true reading; see Equation (5.1). Bias is the common error associated with sensors and often a hard one to identify. Other common errors, like gross errors, are out of the scope of this research. Gross errors are dramatically high or low in value (magnitude), which are easily detected through common statistical techniques. For the purpose of the experimentation, it is assumed that the data collected from the Pogo Mine is devoid of errors, i.e., the set is deemed a “clean set.”

$$\text{Observed or biased reading, } x_{bias} = \text{True reading } (x_{true}) + \text{bias } (e^{bias}) \quad (5.1)$$

Bias is expressed as a percentage over the true reading value. It was artificially induced in the clean data set. A +2% bias indicates 2% of the true reading was added as bias to the true reading.

Likewise -2% expresses adding -2% of the true reading; see Equation (5.1). Calibration errors are subtle and can be as low as 2% over the true reading. Identifying the bias at such low magnitudes is preferred by the industries. Pogo Mine specifically requested that errors be detected at the $\pm 2\%$ level. Only S1 and S2 sensors are analyzed with the methods described. Bias is introduced only in one sensor (S1) at a time in this project. Identifying the bias in several sensors at a time in a multi-sensor environment is out of the scope of this research. Due to the space constraints and to avoid redundancy, the results from S1-based analysis and validation only are presented in this chapter.

5.5.3 Data Preparation

5.5.3.1 Programming and Software

The sheer volume of the data collected, the subsequent mathematical calculations and repetitions to be performed, statistical analyses to be conducted, and graphical presentation of the results, required the usage of a mathematically-intensive software package. Matlab was selected as the software of choice. Matlab has various discipline-specific toolboxes that can be added according to the need, i.e., optimization, signal processing, simulations, neural nets, etc.; however, at a later stage, the Matlab program was replaced with Fortran to improve on speed of execution.

The goal of the project was to detect an artificially-introduced bias ($+2\%$ or -2%). For the purpose of conducting bias identification tests using algorithms, the bias was introduced at a random time period in S1. Once the bias was introduced, all following data for the sensor also contained the bias. The assumption was that once a bias occurred, the sensor remained “uncalibrated” until the error was detected. Thus, if a -2% bias was injected into S1 on March 12, all S1 data starting March 12 was corrupted with a -2% bias. Algorithms were designed to detect the bias as soon as possible. The location of the error was systematically varied from Jan 1, 2015

to December 31, 2015. A series of cut-offs or thresholds were applied to cleanse the data of unreliable or possibly corrupted readings. The algorithm description section has the detailed account of the process.

5.5.3.2 Matlab, Super Computing, and Fortran

Matlab's parallel computing tool box is utilized in improving the execution speed of various algorithms, specifically MRFA and MRFAA. In order to execute a single experiment or test (iteration), Matlab-based algorithms required 5-10 minutes. It is a considerably long time, and for this reason, multiple tests were consuming even longer durations of time. As a result, the number of possible experiments became limited. The supercomputing center at Computing Systems (RCS) of the Geophysical Institute, University of Alaska Fairbanks, was utilized for a brief period of time. Utilizing the multiple cores at the center improved the algorithm speed of execution considerably. At this stage, the Matlab program was recoded into the Fortran 95 program. The built-in support for parallelization (dynamic allocation and vectorization) allows the Fortran-based programs to perform highly memory-intensive tasks. Functions and subroutines can be built in support of the main program, and can be organized into "modules" that work as a single program when called. After recoding, the algorithm (MRFAA) speed improved tremendously and each test (iteration) was completing in seconds.

5.5.4 Background on Concepts, Terminology, and Previous Methods

The following text serves as a refresher on various basic concepts, terminology, and previous methodology applied in connection with the topics of the chapter. Due to the dynamic and non-stationary nature of the sensor data collected in the mineral processing circuits, the preliminary data characterization methods and the subsequent signal processing methods (FFT) applied on Fort

Knox and Pogo Mine data sets did not yield definitive results in terms of finding bias. Consequently, innovative methods that are based on data-mining concepts were attempted. Data-mining and machine learning techniques are widely popular for gaining insights on the behavior of various parameters when the data sets are particularly large. The stripping circuit sensor data collected from the Pogo mill is very large (200,000 readings a year per sensor, if collected in one-min intervals) and meets that qualification. The reason for choosing the stripping circuit for algorithm development is also due to the fact that SAG mill data is too dynamic to handle for any preliminary scientific experimentation; specifically, in the initial model building stages. The algorithms developed from this point forward are based on the Pogo stripping circuit sensor data.

Cleansing Threshold (Th_{cleans}):

The first process in the data preparation is to remove or clean undesired or corrupted data. Usually these are sensor data readings below a certain minimum value (“cleansing threshold”) which reflects where a piece of equipment is down for maintenance or idling. The last column of Table 5.3 provides these values, which are used as cleansing thresholds for each sensor data stream.

Peak (P) and Threshold (Th):

For the purposes of this research, a strip vessel cycle or “peak” is defined as-the continuous rise and maintenance of a sensor’s temperature above a certain desired value. According to Pogo Mine, the desirable range for the strip vessels’ operational temperature is between 270°F and 280°F. The data stream is filtered for the temperature sensors using a “threshold (Th)” value to define a cycle or peak. As described previously, the strip vessel temperatures are cyclical in nature. A peak is assumed to start (peak start-time) when the temperature rises above the threshold. It is

assumed to have ended when the temperature goes below the threshold (peak end-time) (Figure 5.11).

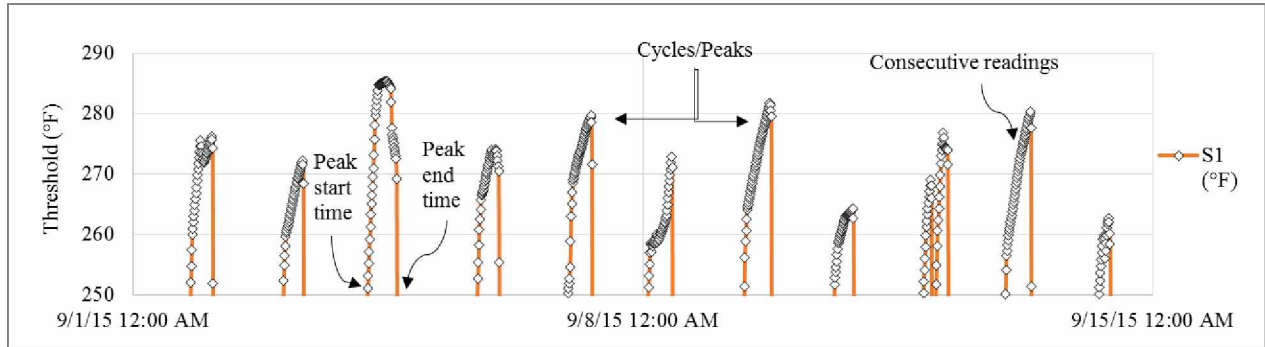


Figure 5.11: Peaks vs thresholds in a clean set of S1 sensor data.

Unlike a traditional “cycle,” which is typically associated with positive and negative components, a peak, for the purpose of this research, can be understood as a truncated form of a cycle (Figure 5.11). The readings below the threshold are ignored in this research. Figure 5.11 shows the appearance of peaks above the 250°F threshold.

Effect of Bias on Peak Statistics, and PRCA:

If the number of readings (nr_p) for each peak of a particular sensor data stream above a certain threshold (Th) is captured, it is expected that in the presence of bias, the variable nr_p will change dramatically (Figure 5.12). If this change is captured in terms of certain percentage statistics, bias identification is possible. This is the key principle that was exploited in peak-readings count analysis (PRCA) and peak-readings sensitivity analysis (PRSA)—together referred to as peak-readings count and sensitivity analysis (PRCSA). For example, in Figure 5.12, after the bias (+10%) is introduced (9/8/15 12:00 AM), the number of peaks (nP) in the S1 data stream with bias are seven at a threshold of 270°F. For the same conditions, the number of peaks (nP) value for the

S1 clean set is three. When compared to the clean S1 stream, there is 133% change in nr^P value for biased S1. The same effect is true for the constituents of a peak, i.e., the number of readings (nr_p).

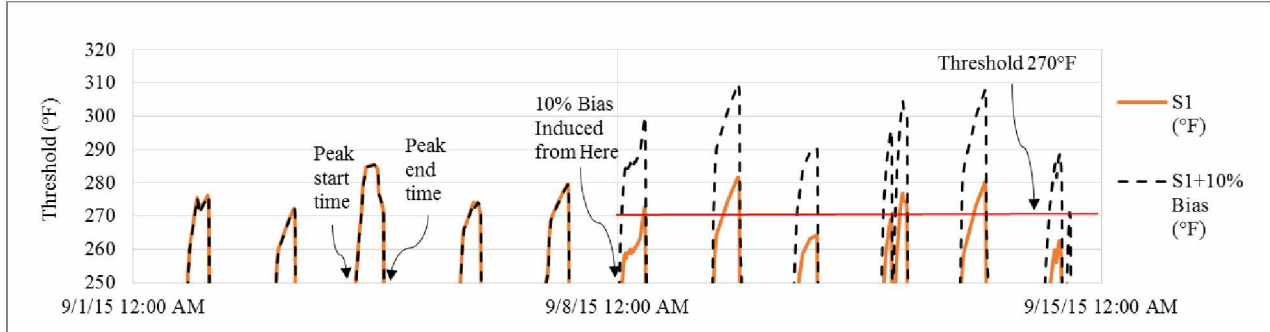


Figure 5.12: Bias effect on S1-sensor data set at various thresholds.

PRSA:

After the peaks are captured by the PRCA algorithm for each data set (clean and biased) based on the threshold (Th), the cumulative temperature per peak ($T_{peak-cum}$) and the average temperature per peak ($T_{peak-Ave}$) values are calculated simultaneously. In addition, the following statistics are calculated for each data set: the cumulative number of peaks ($cum nr_p$), and the cumulative temperature of all peaks ($cum T$), which prove to be effective indicators of bias. While the PRCA method helps capture these statistics, the PRSA method observes changes in these statistics (“sensitivity”) when bias is present in one of the data sets.

Peak Sensitivity and PRCSA:

As indicated in the previous sections, the PRCA and PRSA methods together are called PRCSA. At certain thresholds of temperature (Th) and number of readings per peak (Th_{nr}), the percentage change (sensitivity) in the above-mentioned statistics ($cum nr_p$, $T_{peak-cum}$, and $cum T$) is high due to the presence of bias. This is called “peak sensitivity,” which is the concept upon which the PRCSA and multiple ratio function analysis (MRFA) algorithms (Section 5.5.5) are modelled.

The results from the PRCA algorithm for one of the experiments conducted on the S1 data set is plotted in Figure 5.13. From the figure, it can be observed that at a threshold of 280°F, the number of readings per peak (m_p) for biased data stream (+2%) visibly departs (redline) from that of the clean data stream (blue line). A similar trend is observed for the temperature per peak values.

Results from the sensitivity analysis (PRSA) conducted on the same experiment can be seen in Table 5.4 (a bar chart representation is available in Figure 5.14). It can be observed that at a Th_{SI} value of 280°F and Th_{nr} of 30, the percentage change from part-1 (both sets clean) to part-2 (one set biased) is high: 96% to 353%. This change is high when compared to any other thresholds: Th_{SI} of 260 through 280, and Th_{nr} of 30 through 50. The change can also be observed visually in Figure 5.13. The PRCSA algorithm in the form of flowchart is available in Figure 5.15.

The major drawback of PRCSA is its inability to function without a clean data set (from calibrated sensors) for comparison. In real life industrial situations, this is a highly impractical, invasive, and time-consuming process. In most of the industrial circuits, some calibrated sensors and some biased sensors co-exist at a given point of time. The MRFA algorithm is modelled to work in such practical situations.

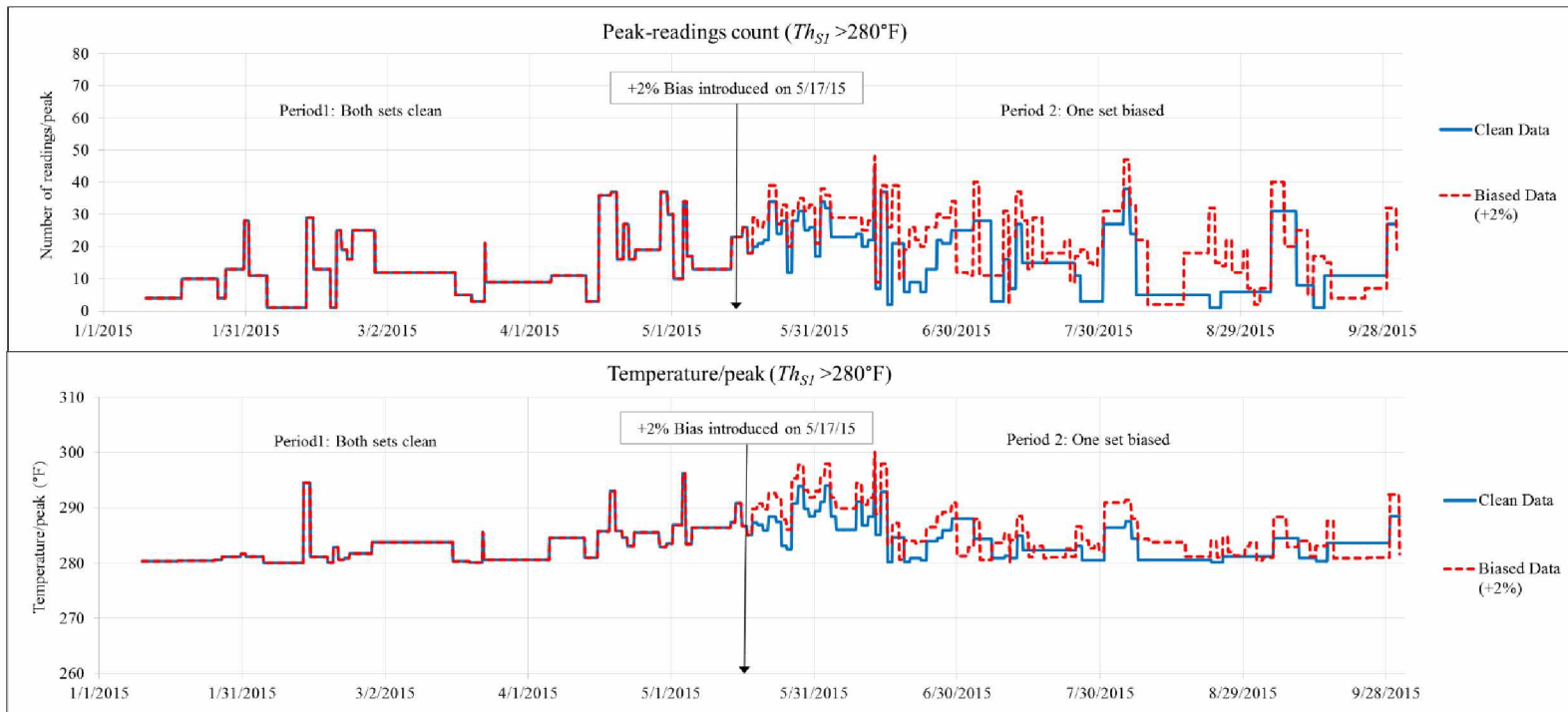


Figure 5.13: Bias effect on S1-sensor data at 280°F threshold (Th_{S1}).

Table 5.4: Cumulative peak-readings count, clean vs biased ($Th_{SI}=280^{\circ}\text{F}$).

Data-set	Sensor-S1 readings threshold ($^{\circ}\text{F}$)	Threshold for 'no. of readings/peak'	Cum peak-readings count		% Change
	Th_{SI}	Th_{nr}	Period 1	Period 2	
			$cum\ nr_P$	$cum\ nr_P$	$cum\ nr_P\ change\%$
Clean	280	30	144	282	96
Biased	280	30	144	653	353

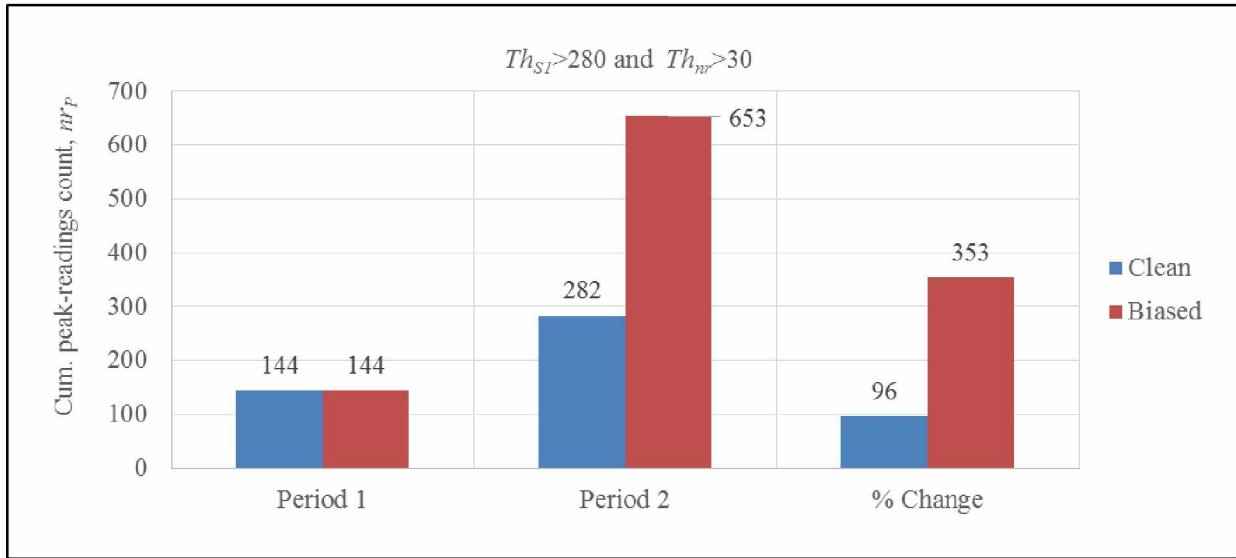


Figure 5.14: Cumulative peak-readings count, clean vs biased ($Th_{SI}=280^{\circ}\text{F}$).

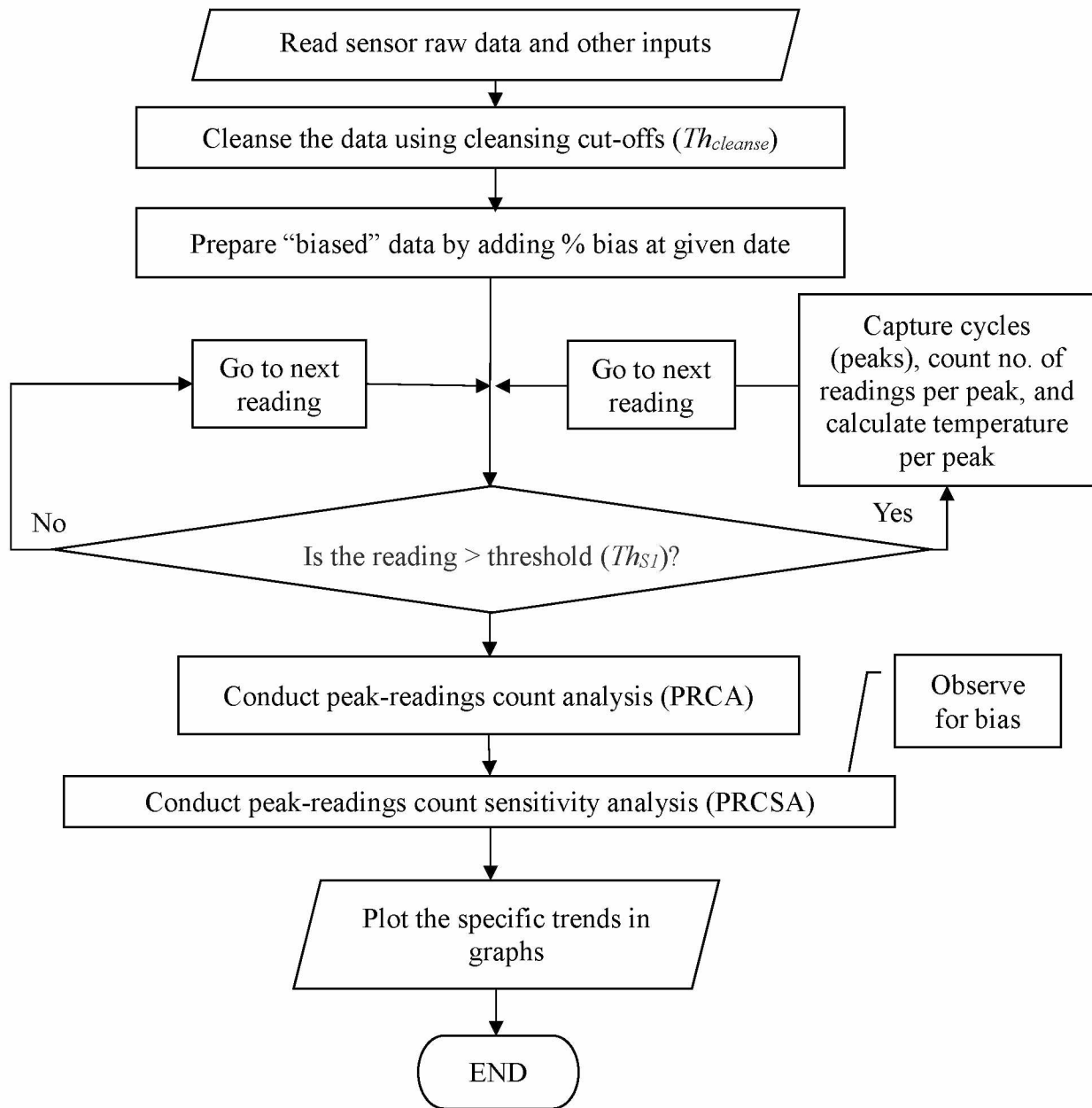


Figure 5.15: Flow chart for the peak-readings count and sensitivity analysis (PRCSA).

5.5.5 Multiple Ratio Function Analysis (MRFA)

For highly non-linear and non-stationary processes like mineral processing operations, the application of classical statistical methods prove disadvantageous. The Pogo carbon stripping circuit is an example where several dynamic processes were observed for this study. Although

peak-readings count and sensitivity analysis (PRCSA) proved promising, there are drawbacks. The method requires comparisons between the corrupted data set and a clean data set. In real life, only one data stream would be present. To mitigate these disadvantages, the interrelations between sensors were exploited in a multiple ratio function analysis (MRFA), an innovative data-mining or hybrid method. For instance, in the carbon stripping circuit, the sensors S1 and S2 measure temperatures of strip vessels 1 and 2, respectively. The heat to the vessels is supplied by the circulation of a hot “barren solution” through a series of four heat exchangers and their corresponding boilers. Sensors H1, H2, H3, and H4 measure the outlet temperatures of these four heat exchangers, respectively (Figure 5.5). From the set up, it can be understood that there exists a relationship between the heat supplied by the heat exchangers and the heat retained by the strip vessels through the medium (barren solution). This means the measurements of S1 and S2 are in some way related to the measurements of H1 through H4. In fact, the sum of all the heat sensor measurements, simply called “Heat,” directly affects the temperature measurements of S1 and S2. In other words, the S1 and S2 measurements are a function of “Heat.” In a similar fashion, barren solutions flow is related to strip vessel functionality. Glycol flow through the boilers is also indicative of the heat exchange process. A higher glycol flow would indicate more “heat exchange” and heating of barren flow, thus affecting S1 and S2. Given the process, one could expect a relationship between BARNFL and GLYFL sensors, and strip vessel sensors (S1 and S2). Apart from these relations, there is also a relationship between the strip vessel sensors, S1 and S2, due to the strip vessels operating in tandem (Figure 5.10).

A schematic diagram for various sensor interrelations in the stripping circuit is provided in Figure 5.16. In the above context, the research explored the following aspects in the stripping circuit:

- (i) Existence of relations between the sensor measurements.
- (ii) Possibility to express the relations mathematically.
- (iii) Existence of the “difference” between sensor relations in error-free state (calibrated) and erroneous state (biased).
- (iv) Quantification of the “difference” in terms of statistics.
- (v) Identification of bias based on the quantification of difference.
- (vi) Limitations on the “identification of bias” process; magnitude of error, location, etc.

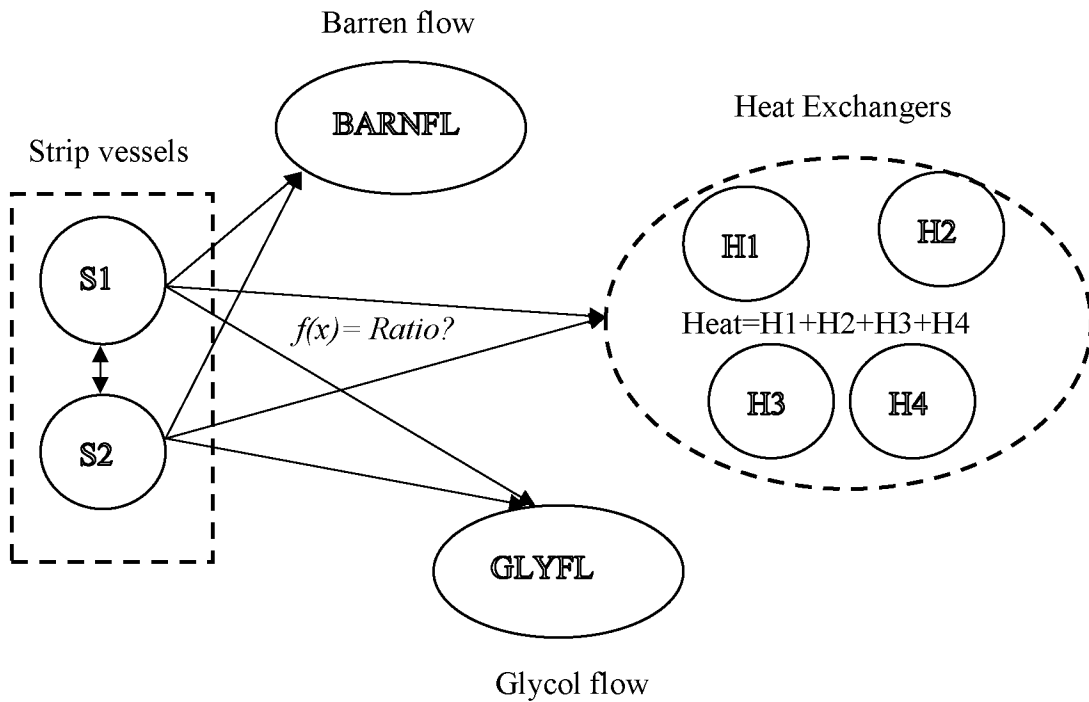


Figure 5.16: Various sensors and their interrelations.

Ratio Function:

Several parametric (correlation, aggregation, and PCA) and non-parametric techniques (neural nets), used in the past in exploring the sensor interrelations were proved unsuccessful. This led to a new approach (MRFA), in which it is suspected that there may be a “ratio” based relation between various processes in the stripping circuit, so as between sensor readings. A simple “ratio analysis” is used in many of the industrial optimization applications and even in financial analysis; however, the ratio function that is explored in this new data-mining based approach is slightly complicated, and is explained in detail in the subsequent sections. Due to the dynamic nature of the stripping circuit processes, it is anticipated that a modified form of ratio function can become a better tool in describing the sensor interrelations. The relation in the form of a “ratio function” is defined in Equation (5.2); S1 and Heat are chosen for illustration purpose.

$$Ratio^* = f[T_{peak-ave}(Heat), T_{peak-ave}(S1)] \quad (5.2)$$

For sensor S2, the relation is expressed as follows:

$$Ratio^* = f[T_{peak-ave}(Heat), T_{peak-ave}(S2)] \quad (5.3)$$

* This is not a “simple ratio,” since several conditions have to be met--as detailed below--in order for the function to exist. Due to this reason, “ratio function” is used throughout the text instead of simple ratio.

The following are some of the features of the ratio function (Figure 5.17).

(i) The cycle of S2 starts only after S1 concludes its cycle since the two stripping vessels are used alternatively. Therefore, cycles cannot be compared in the same time span. Statistics from a

S1 cycle are compared to statistics from a S2 cycle only if the S2 cycle started within 24 hours of the end of S1 cycle, i.e., the ratio function exists only between two peaks of S1 and S2 if captured within a 24 hour span. This process is called “matching forward.” For instance, ratios computed from the sensor data collected during a S1 peak is matched to ratios computed from sensor data collected during a S2 peak that appears 24 hours after S1 occurs (forward of S1 when data is viewed as a chart). Given the context, the term “matching” when used in the subsequent sections should be understood as “matching forward,” since “matching backward” was not used at all.

(ii) The peaks are truncated cycle forms resulting from using a preset “truncation threshold” (*trunc Th*). The *trunc Th* value is modelled where the ratio function sensitivity to bias is high—which is explained in the subsequent sections.

(iii) Based on the peak-start and peak-end times of S1 and S2, the corresponding heat exchanger sensor peaks (H1 through H4) are captured and the total “heat” is calculated.

(iv) The average temperature ($T_{peak-ave}$) value for each and every peak is calculated.

(iv) Lastly, the $T_{peak-ave}$ ratios of S1 to heat and S2 to heat are calculated.

(v) All the peaks are at least 1-hour in duration. This is to avoid very small peaks, which are a distraction to the study.

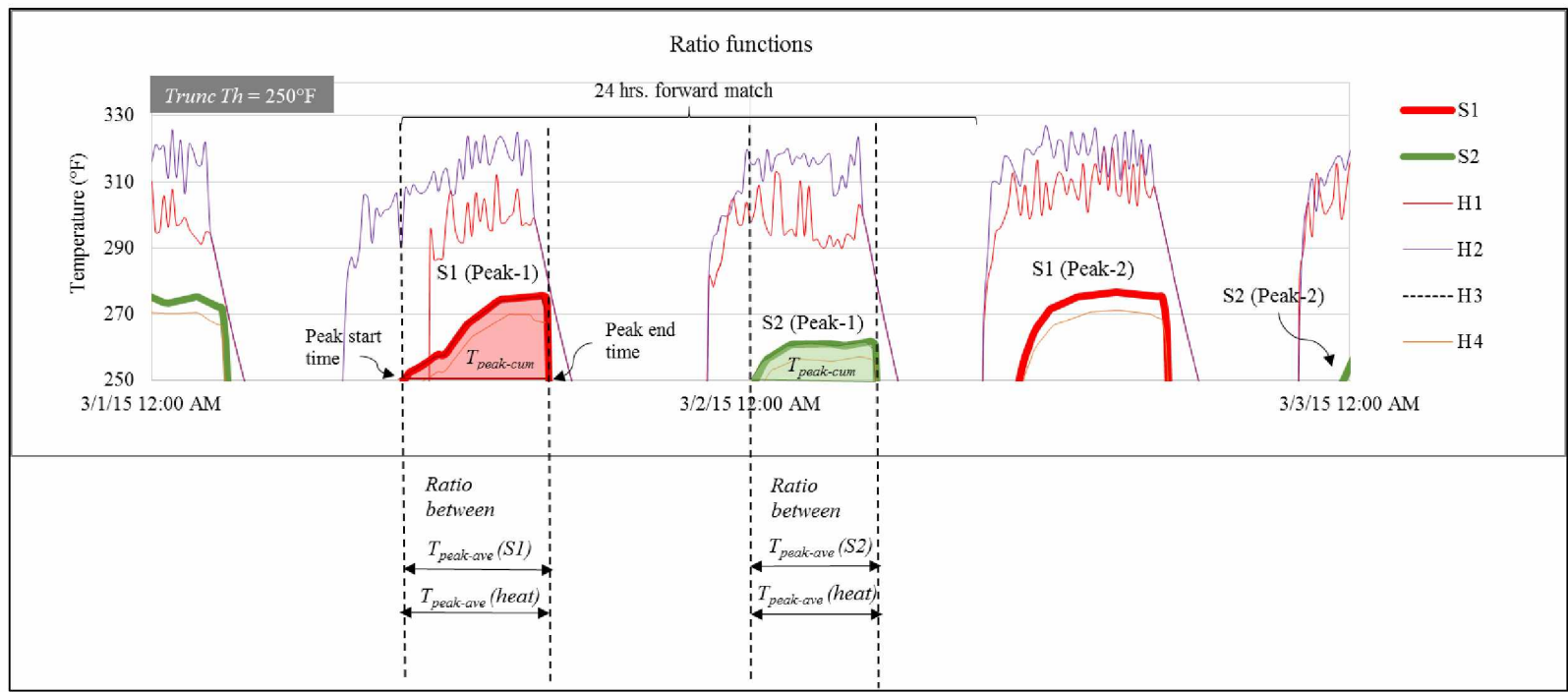


Figure 5.17: Ratio functions and matching.

Revisiting Figure 5.16, multiple ratio function relations can be exploited between various sensors in order to identify a biased sensor. For instance, if bias is induced in S1 sensor data—while the other sensors are kept well-calibrated—the ratios between S1 to all other sensors and S2 to all other sensors can be compared to find bias in S1. Since S1 and S2 are tandem processes, in an error-free state, the “ratio functions” they form with other sensors show similarities. In the presence of bias, the “ratio functions” a biased sensor (S1) forms with other sensors should show significant departure from that of a sensor with no bias (S2). This approach is clearly meritorious compared to PRCSA, where a clean set of data—for baseline statistics—is always required from the same biased sensor for comparison (highly impractical and a disruption to industrial production process). In that context, MRFA algorithm is more practical for industry applications. MRFA is designed as a single test that only exploits ratio relations between strip vessel sensors (S1 and S2) and Heat. The ratios of S1 and S2 with barren flow and glycol flow are exploited with other tests that are added to MRFA at later stages—to produce a more efficient MRFAA algorithm.

5.5.5.1 MRFA: Algorithm Description

The algorithm is based on the concept of exploring ratio function relations between strip vessel sensors (S1 and S2) and Heat. For the purpose of the algorithm description, S1 is assumed to have bias. The flowchart is given in Figure 5.18.

Step 1: Reads sensor data sets with thresholds and other input data.

Step 2: Removes all the undesired or corrupted data by applying $Th_{cleanse}$.

Step 3: Creates the biased data set for S1 by adding $\pm 2\%$ bias (depends on the user choice) to the clean set of S1 at the chosen date—bias stays in all data from that point forward.

Step 4: Using a truncation threshold (*trunc Th*) the S1 biased data set is cleaned; for instance, at a *trunc Th* of 280°F, all the S1 readings below that value are deleted. The intention is to study only the readings above this threshold, since the sensitivity of peak ratios to bias at this threshold is high. Then, the peaks of S1 data set that are at least one-hour in duration are captured; smaller peaks are less significant and are of distraction to the study. For the same duration (start-end times) of S1 bias peaks, the algorithm captures the peaks for heat exchanger sensors individually (Figure 5.17). The total “Heat” of all the heat exchanger sensors together for each peak of S1 bias data is calculated. Then, the cumulative temperature per peak ($T_{peak-cum}$) and average temperature per peak ($T_{peak-ave}$) values are calculated; these values are truncated at *Trunc Th* to reduce their relative size, i.e., to make the S1 to heat ratio a small comparable value with respect to other sensor ratios. The ratio of S1 bias $T_{peak-ave}$ to Heat $T_{peak-ave}$ is calculated for each peak of S1. The whole process is repeated for S2 immediately. The above mentioned statistics along with peak-max values captured by the program are found in Table 5.5 for S1 and Table 5.6 for S2.

Revisiting Table 5.5, the S1 bias data statistics populated can be understood as below. For instance, the peak number-2 (highlighted) was captured from 1/26/15 5:00 pm through 1/26/15 7:10 pm. It lasted for 2 hrs. 10 min. (2.17 hrs.), and there were 13 readings consecutively captured (each 10 min duration), which are above 280°F. The maximum value of the peak-readings is 281.88°F. The truncated cumulative temperature is calculated as follows:

Cumulative sum of temperature of peak-2 readings= 3655.2°F

No. of peaks captured (nP) =13

Truncated cumulative temperature ($T_{peak-cum}$) = $3655.2 - nP * trunc Th$

$$= 3655.2 - (13 \times 280)$$

$$= 3655.2 - 3640$$

$$= 15.2$$

Truncated average temperature ($T_{peak-ave}$) = $T_{peak-cum}/nP$

$$= 15.2/13$$

$$= 1.17$$

The calculations for the S2 peaks were done in similar fashion (see Table 5.6).

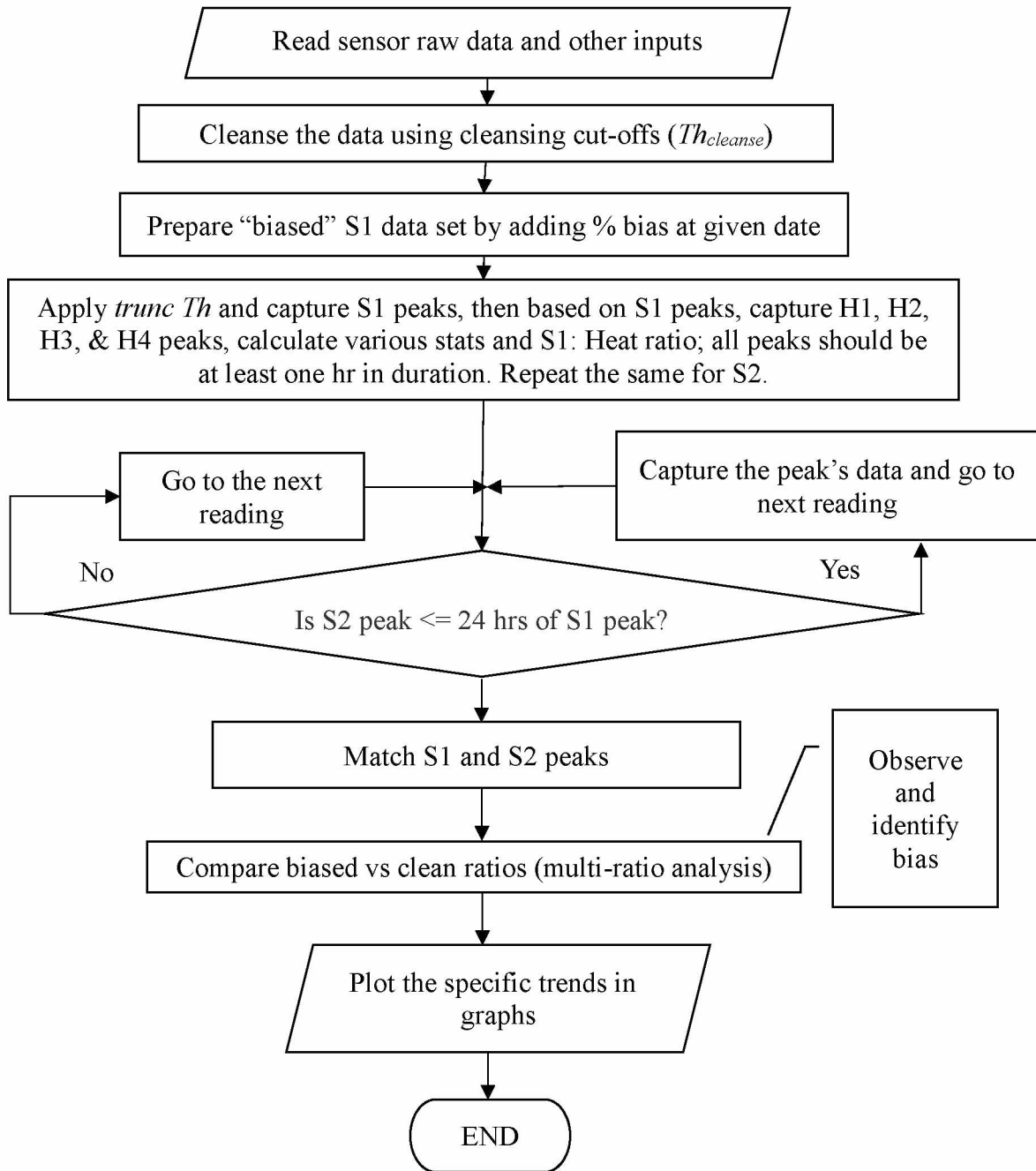


Figure 5.18: Flow chart for the multiple ratio function analysis (MRFA) algorithm.

In a similar manner, the total “Heat” of all the heat exchanger sensor readings that fall in the same peak start-end periods of S1 and S2 are calculated. This means, for S1 peaks there are corresponding heat peaks and statistics, and for S2 peaks, there are corresponding heat peaks and statistics. These are in the form of Tables 5.5 and 5.6. Then, the ratio of “S1bias” to Heat peaks and “S2clean” to Heat peaks in terms of $T_{peak-ave}$ are calculated. This is clearly depicted in Figure 5.17.

Step 5: In this step, the S2 clean data peaks that occur within 24 hours (matching forward) period of each of S1 peaks are captured. From Figure 5.10, it can be observed that after strip vessel sensor S1 cycle is completed, the S2 cycle starts. Only the peaks that are matched are retained for “ratio analysis.” Sometimes it is possible that within 24 hours of S1 peak, two or more peaks of S2 occur; this could be due to small process fluctuations. In this situation, the sum of all those peak temperatures is taken into account. As an example to demonstrate “matching” process, a matching peak is located for the S1 peak that occurred at 2/12/15 6:30 PM (see highlighted text in Table 5.5), and the match is found in the corresponding Table 5.6 (S2 clean peaks), i.e., on 2/13/15 7:10 AM.

Step 6: In the sixth step, the algorithm populates the matching S1 and S2 peak data and statistics into an Excel spreadsheet (see Table 5.7). From the table, it can be observed that S2clean: Heat ratio is obtained on a clean set of data. On the contrary, S1bias: Heat statistic is obtained on a bias set. The date from which the bias is introduced is marked in the same table (from 32nd peak onwards). Comparison of these ratios in a plot--in terms of trends and behaviors--can provide the clues to where the induced bias is located. It should be noted that this comparison process is not automated for the algorithm yet. The plot with comparison of these two ratios is available in Figure 5.19.

Table 5.5: A snapshot of peaks captured for sensor S1; $trunc Th_{S1}=280^{\circ}\text{F}$, $bias=+2\%$.

S1 Bias peaks or cycles captured (+2% induced on 6/17/15)							
Peak #	PeakStartTime	Peak EndTime	Peak width (hrs.) W_{peak}	Cum temp/ peak ($^{\circ}\text{F}$) $T_{peak-cum}$	no. of peaks nP	Peak ave temp ($^{\circ}\text{F}$) $T_{peak-ave}$	Peak max temp ($^{\circ}\text{F}$) $T_{peak-max}$
1	1/17/15 11:20 AM	1/17/15 1:00 PM	1.67	4.04	10	0.40	280.68
2	1/26/15 5:00 PM	1/26/15 7:10 PM	2.17	15.20	13	1.17	281.88
3	1/30/15 3:50 PM	1/30/15 8:30 PM	4.67	46.84	28	1.67	281.99
4	1/31/15 5:40 PM	1/31/15 7:30 PM	1.83	13.10	11	1.19	281.48
5	2/12/15 6:30 PM	2/12/15 11:20 PM	4.83	421.44	29	14.53	297.09
6	2/14/15 7:00 AM	2/14/15 9:10 AM	2.17	14.77	13	1.14	281.61
7	2/19/15 2:20 AM	2/19/15 6:30 AM	4.17	72.37	25	2.89	285.19
8	2/20/15 5:00 AM	2/20/15 8:10 AM	3.17	11.30	19	0.59	280.97
9	2/21/15 9:20 AM	2/21/15 12:00 PM	2.67	13.77	16	0.86	281.53
10	2/22/15 11:30 AM	2/22/15 3:40 PM	4.17	43.42	25	1.74	282.74
11	2/27/15 6:30 AM	2/27/15 8:30 AM	2.00	45.39	12	3.78	285.84
12	3/22/15 9:10 AM	3/22/15 12:40 PM	3.50	116.55	21	5.55	288.83
13	3/22/15 2:30 PM	3/22/15 4:00 PM	1.50	5.32	9	0.59	281.00
14	4/5/15 10:30 AM	4/5/15 12:20 PM	1.83	50.41	11	4.58	287.17
15	4/15/15 11:30 AM	4/15/15 5:30 PM	6.00	206.87	36	5.75	287.14
16	4/18/15 2:40 AM	4/18/15 8:50 AM	6.17	481.93	37	13.03	295.51
17	4/19/15 8:00 AM	4/19/15 10:40 AM	2.67	92.93	16	5.81	290.25
18	4/20/15 4:30 PM	4/20/15 9:00 PM	4.50	124.92	27	4.63	286.47
19	4/21/15 7:40 PM	4/21/15 10:20 PM	2.67	48.71	16	3.04	284.33
20	4/23/15 4:40 AM	4/23/15 7:50 AM	3.17	105.04	19	5.53	288.73
21	4/28/15 11:20 AM	4/28/15 5:30 PM	6.17	108.68	37	2.94	286.63
22	4/30/15 12:40 AM	4/30/15 5:40 AM	5.00	105.33	30	3.51	286.90
23	5/1/15 8:00 AM	5/1/15 9:40 AM	1.67	69.25	10	6.92	291.45
24	5/3/15 7:10 AM	5/3/15 12:50 PM	5.67	551.06	34	16.21	302.26
25	5/4/15 2:40 AM	5/4/15 5:30 AM	2.83	58.52	17	3.44	284.66
26	5/5/15 7:40 AM	5/5/15 9:50 AM	2.17	83.64	13	6.43	294.04
27	5/13/15 11:10 AM	5/13/15 3:00 PM	3.83	169.62	23	7.37	292.89
28	5/14/15 11:40 AM	5/14/15 3:30 PM	3.83	249.83	23	10.86	295.00
29	5/15/15 7:10 PM	5/15/15 11:30 PM	4.33	174.03	26	6.69	290.95
30	5/16/15 9:10 PM	5/17/15 12:10 AM	3.00	92.02	18	5.11	289.24

Table 5.6: A snapshot of peaks captured for sensor S2 (clean); $trunc\ Th_{SI}=280^{\circ}\text{F}$, no bias.

S2-peaks captured for clean data							
Peak #	PeakStartTime	Peak EndTime	Peak width (hrs.) W_{peak}	Cum temp/ peak ($^{\circ}\text{F}$) $T_{peak-cum}$	no. of peaks nP	Peak ave temp ($^{\circ}\text{F}$) $T_{peak-ave}$	Peak max temp ($^{\circ}\text{F}$) $T_{peak-max}$
1	1/6/15 8:20 PM	1/7/15 12:40 AM	4.33	79.22	26	3.05	285.09
2	1/16/15 12:10 PM	1/16/15 3:00 PM	2.83	51.00	17	3.00	284.55
3	1/20/15 10:50 AM	1/20/15 2:00 PM	3.17	89.45	19	4.71	290.54
4	1/21/15 10:20 AM	1/21/15 12:40 PM	2.33	13.01	14	0.93	281.27
5	1/24/15 4:20 PM	1/24/15 8:10 PM	3.83	352.42	23	15.32	297.72
6	1/25/15 7:00 PM	1/25/15 11:10 PM	4.17	179.07	25	7.16	288.53
7	1/29/15 9:40 PM	1/30/15 2:40 AM	5.00	155.06	30	5.17	288.27
8	2/2/15 5:10 PM	2/2/15 9:50 PM	4.67	112.52	28	4.02	285.30
9	2/5/15 9:30 PM	2/5/15 11:40 PM	2.17	44.30	13	3.41	284.78
10	2/8/15 6:00 PM	2/8/15 9:50 PM	3.83	56.61	23	2.46	284.76
11	2/9/15 7:40 PM	2/9/15 11:40 PM	4.00	157.37	24	6.56	290.22
12	2/13/15 7:10 AM	2/13/15 11:10 AM	4.00	56.02	24	2.33	283.44
13	2/15/15 1:40 AM	2/15/15 5:40 AM	4.00	320.56	24	13.36	295.79
14	2/18/15 1:20 PM	2/18/15 8:30 PM	7.17	609.96	43	14.19	297.76
15	2/19/15 5:20 PM	2/19/15 7:30 PM	2.17	10.70	13	0.82	281.38
16	2/24/15 6:10 AM	2/24/15 11:30 AM	5.33	193.57	32	6.05	288.68
17	3/4/15 5:50 AM	3/4/15 10:20 AM	4.50	197.10	27	7.30	289.39
18	3/30/15 9:20 PM	3/30/15 10:40 PM	1.33	8.41	8	1.05	281.57
19	4/13/15 10:40 AM	4/13/15 2:50 PM	4.17	113.95	25	4.56	287.17
20	4/14/15 9:30 PM	4/15/15 2:50 AM	5.33	69.81	32	2.18	284.88
21	4/16/15 2:00 AM	4/16/15 7:00 AM	5.00	361.39	30	12.05	298.37
22	4/17/15 3:40 PM	4/17/15 10:00 PM	6.33	557.58	38	14.67	297.85
23	4/18/15 2:40 PM	4/18/15 10:00 PM	7.33	422.74	44	9.61	291.74
24	4/20/15 3:20 AM	4/20/15 5:10 AM	1.83	37.80	11	3.44	285.87
25	4/21/15 7:00 AM	4/21/15 10:00 AM	3.00	65.70	18	3.65	286.88
26	4/22/15 1:00 PM	4/22/15 5:00 PM	4.00	119.09	24	4.96	288.66
27	4/27/15 11:10 PM	4/28/15 2:30 AM	3.33	45.90	20	2.30	284.46
28	4/29/15 6:50 AM	4/29/15 10:20 AM	3.50	108.87	21	5.18	290.06
29	4/30/15 6:20 PM	4/30/15 8:50 PM	2.50	31.39	15	2.09	284.15
30	5/4/15 6:30 PM	5/4/15 8:10 PM	1.67	25.99	10	2.60	284.55

Table 5.7: A snapshot of S1 and S2 peaks Matching; $trunc Th_{SI}=280^{\circ}\text{F}$, $bias=+2\%$.

S1bias peaks					S2clean peaks				
Peak#	PeakStartTime	Peak max temp (°F) $T_{peak-max}$	Peak ave temp (°F) $T_{peak-ave}$	S1 bias:Heat ratio) $(T_{peak-ave})$	Peak#	PeakStartTime	Peak max temp (°F) $T_{peak-max}$	Peak ave temp (°F) $T_{peak-ave}$	S2:Heat ratio) $(T_{peak-ave})$
1	2/12/15 6:30 PM	297.09	14.53	0.412	1	2/13/15 7:10 AM	283.44	2.33	0.071
2	2/14/15 7:00 AM	281.61	1.14	0.032	2	2/15/15 1:40 AM	295.79	13.36	0.383
3	2/19/15 2:20 AM	285.19	2.89	0.100	3	2/19/15 5:20 PM	281.38	0.82	0.028
4	4/15/15 11:30 AM	287.14	5.75	0.159	4	4/16/15 2:00 AM	298.37	12.05	0.425
5	4/18/15 2:40 AM	295.51	13.03	0.424	5	4/18/15 2:40 PM	291.74	9.61	0.352
6	4/19/15 8:00 AM	290.25	5.81	0.208	6	4/20/15 3:20 AM	285.87	3.44	0.103
7	4/20/15 4:30 PM	286.47	4.63	0.123	7	4/21/15 7:00 AM	286.88	3.65	0.110
8	4/21/15 7:40 PM	284.33	3.04	0.094	8	4/22/15 1:00 PM	288.66	4.96	0.142
9	4/28/15 11:20 AM	286.63	2.94	0.126	9	4/29/15 6:50 AM	290.06	5.18	0.187
10	4/30/15 12:40 AM	286.90	3.51	0.134	10	4/30/15 6:20 PM	284.15	2.09	0.103
11	5/4/15 2:40 AM	284.66	3.44	0.130	11	5/4/15 6:30 PM	284.55	2.60	0.108
12	5/5/15 7:40 AM	294.04	6.43	0.207	12	5/5/15 10:40 PM	281.53	0.79	0.000
13	5/13/15 11:10 AM	292.89	7.37	0.252	13	5/14/15 1:00 AM	292.51	7.13	0.238
14	5/14/15 11:40 AM	295.00	10.86	0.311	14	5/15/15 2:50 AM	303.07	14.51	0.396
15	5/15/15 7:10 PM	290.95	6.69	0.212	15	5/16/15 6:50 AM	291.44	6.07	0.215
16	5/16/15 9:10 PM	289.24	5.11	0.175	16	5/17/15 7:50 AM	304.27	18.88	0.510
17	5/18/15 12:10 AM	292.31	7.29	0.229	17	5/18/15 1:10 PM	290.97	6.74	0.230
18	5/19/15 1:30 AM	291.82	6.94	0.225	18	5/19/15 5:40 PM	295.92	9.61	0.306
19	5/20/15 6:50 AM	288.02	5.92	0.195	19	5/20/15 9:50 PM	285.32	3.94	0.126
20	5/23/15 12:20 AM	289.53	7.42	0.235	20	5/23/15 12:30 PM	286.28	4.86	0.149
21	5/24/15 3:00 AM	284.76	3.10	0.106	21	5/24/15 6:50 PM	284.24	2.18	0.077
22	5/25/15 6:50 AM	283.78	2.52	0.095	22	5/25/15 7:40 PM	283.30	1.98	0.059
23	5/27/15 2:00 PM	297.06	13.89	0.419	23	5/28/15 8:30 AM	287.64	3.48	0.000
24	5/28/15 8:50 PM	296.98	9.82	0.307	24	5/29/15 8:50 AM	293.95	7.57	0.148
25	5/29/15 8:40 PM	295.50	8.46	0.273	25	5/30/15 12:30 PM	297.94	11.30	0.234
26	5/31/15 4:00 AM	295.65	9.43	0.278	26	5/31/15 1:50 PM	302.54	14.95	0.329
27	6/1/15 6:20 AM	299.76	11.08	0.317	27	6/1/15 5:50 PM	299.30	12.31	0.391
28	6/2/15 6:00 AM	302.68	13.99	0.376	28	6/2/15 8:00 PM	295.93	9.14	0.358
29	6/3/15 10:40 AM	294.32	8.44	0.249	29	6/4/15 2:20 AM	296.38	9.81	0.254
30	6/10/15 1:10 AM	289.54	6.83	0.288	30	6/10/15 7:10 PM	285.99	3.25	0.000
31	6/12/15 12:30 PM	294.16	13.95	0.570	31	6/13/15 9:50 AM	281.66	1.25	0.213
32	6/17/15 8:10 PM	281.28	0.64	0.026	32	6/18/15 9:20 AM	0.00	0.00	0.000
33	6/18/15 8:00 PM	285.98	4.08	0.119	33	6/19/15 10:30 AM	282.43	1.32	0.097
34	6/20/15 12:10 AM	287.15	3.99	0.119	34	6/20/15 5:10 PM	0.00	0.00	0.000
35	6/21/15 5:40 AM	284.88	3.55	0.096	35	6/21/15 5:10 PM	289.88	5.88	0.055
36	6/22/15 6:30 AM	286.43	3.90	0.125	36	6/22/15 9:00 PM	283.80	2.63	0.000
37	6/23/15 11:50 AM	292.65	6.51	0.192	37	6/24/15 3:30 AM	0.00	0.00	0.000
38	6/25/15 6:20 PM	294.00	8.43	0.239	38	6/26/15 7:30 AM	287.91	4.84	0.035
39	6/26/15 7:50 PM	296.58	9.20	0.260	39	6/27/15 8:10 AM	291.44	8.45	0.000
40	6/28/15 6:40 PM	300.07	10.93	0.410	40	6/29/15 7:10 AM	285.45	3.04	0.174
41	7/2/15 8:20 AM	285.22	2.89	0.109	41	7/2/15 7:30 PM	0.00	0.00	0.084
42	7/4/15 7:50 PM	281.00	0.57	0.024	42	7/5/15 10:00 AM	0.00	0.00	0.000
43	7/9/15 7:50 PM	287.86	5.37	0.189	43	7/10/15 3:10 PM	0.00	0.00	0.194
44	7/11/15 2:20 AM	286.83	4.22	0.145	44	7/11/15 7:30 PM	285.93	2.76	0.127
45	7/12/15 11:10 AM	292.24	8.52	0.303	45	7/13/15 3:10 AM	0.00	0.00	0.234

Matched

Biased data

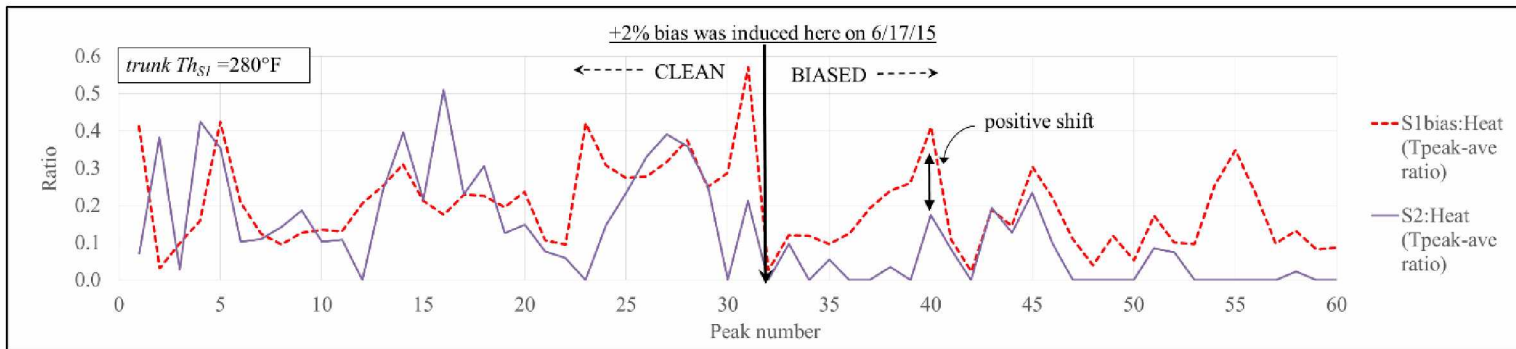


Figure 5.19: Ratio analysis plot for S1 and S2 peaks; $trunc Th_{SI} = 280^{\circ}\text{F}$, bias = +2%.

From Figure 5.19, it can be observed that after the bias is introduced on 6/17/15 (32nd matched peak), there is a positive shift of S1 bias ratios with respect to S2 clean ratios. Before the introduction of bias, the ratios crossed at regular intervals—which reflects the tandem nature of the underlying processes. The positive shift indicated positive bias. In a similar fashion, when the negative bias (-2%) is introduced, negative shift is observed. In case of MRFA algorithm, the process of identifying bias is manual after the completion of plotting process. Due to this reason, it was very time consuming to conduct multiple experiments on multiple bias induction dates. Identification of bias on a particular date after induction is important and with MRFA, it cannot be achieved automatically. The duration between the bias induction date and bias identification date is called “time till find days” (TTFD). Lower values are highly desired. The MRFA algorithm needs to be improved to achieve this automatically, which is described in the subsequent sections.

After conducting several Heat ratio tests at different thresholds to find bias in S1—based on S2 clean Heat ratios—it was observed that at a truncation threshold (*trunc Th*) of 275°F, the S1 sensor “drift” has more sensitivity to bias. This is called “effective threshold” (*eff Th*). Drift or shift is where biased sensor ratios (S1_{bias}: Heat) grow apart from the clean sensor ratios (S2_{clean}: Heat). When observed graphically, at the bias of +2%, the drift is upwards, whereas at -2% it is downwards (Figure 5.20); the shift is observed at 45th peak. For comparison purposes, a figure with no bias (S1 and S2 are clean) is also embedded within Figure 5.20 (see middle plot). In the absence of bias, there is no shift or drift observed in the plot.

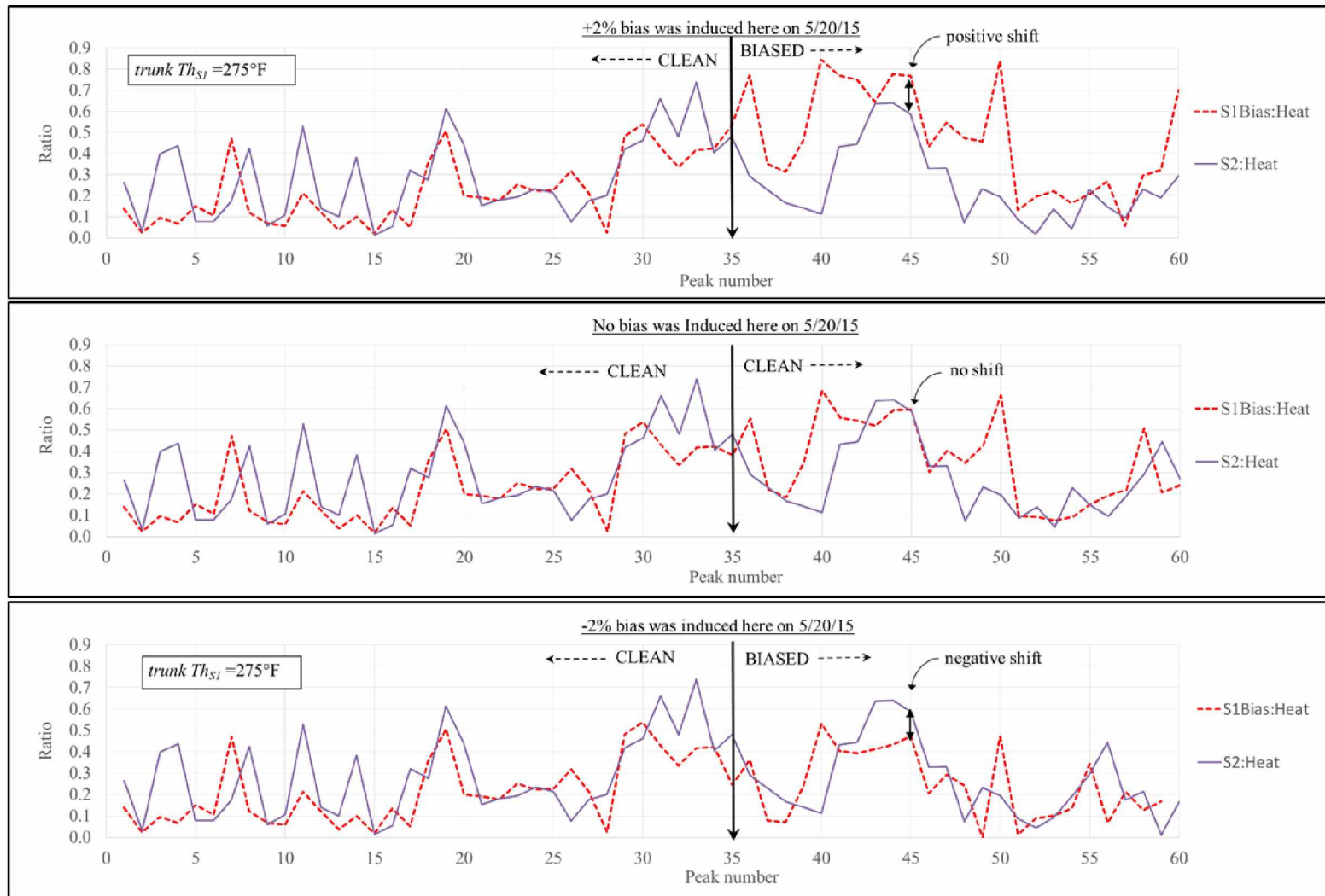


Figure 5.20: Comparison of positive (+2%) and negative (-2%) bias effects on S1 sensor clean data set.

5.5.5.2 MRFA with Multiple Tests and Automation (MRFAA)

In order to compensate for the drawbacks of the MRFA algorithm described in the previous section, several changes were made to the algorithm as outlined below. The improved algorithm is more reliable in terms of finding bias and is named “MRFAA.”

(i) A “cross-score” algorithm was developed in order to assist the tests to detect bias automatically after it is introduced in the data.

(ii) Multiple tests were added that feature the cross-score algorithm, i.e., ratio tests such as, Heat, BARNFL, GLYFL, and value tests such as Ave value test and Max value test.

(iii) A “dynamic thresholding” strategy is included to compensate the disadvantages of “truncation threshold” strategy of MRFA algorithm.

Cross-Score Algorithm:

In order to automate the bias detection process from MRFA, a “cross-score algorithm” was developed (Figure 5.21) to score the relative position of the “S1 bias to Heat ratio” with respect to “S2 clean to Heat ratio” (Figure 5.22). The figure demonstrates the scoring mechanism of the algorithm with the aid of a Heat test. For instance, for each of the peaks matched for S1 and S2, if S1 ratio value is greater than that of S2 ratio value at any given peak number, the algorithm assigns a score of 1 to that peak. If S2 ratio value is greater instead, the score assigned is -1. For example, at the 39th peak captured (and matched), S1 ratio is relatively higher than the S2 ratio. This means, for peak 39, the algorithm assigns a “score” of 1. The “cross-score” value for the same peak is the sum of the “scores” from the past 10 peaks including peak number 39, i.e., four (4). It is noticeable that the cross-score value at a given peak number is a moving sum, because the algorithm applies

the same strategy for each peak number (matched). From Figure 5.22 it is noticeable that cross-scores are low values (2) before the bias is introduced and turned out as high values (4) after the bias is introduced.

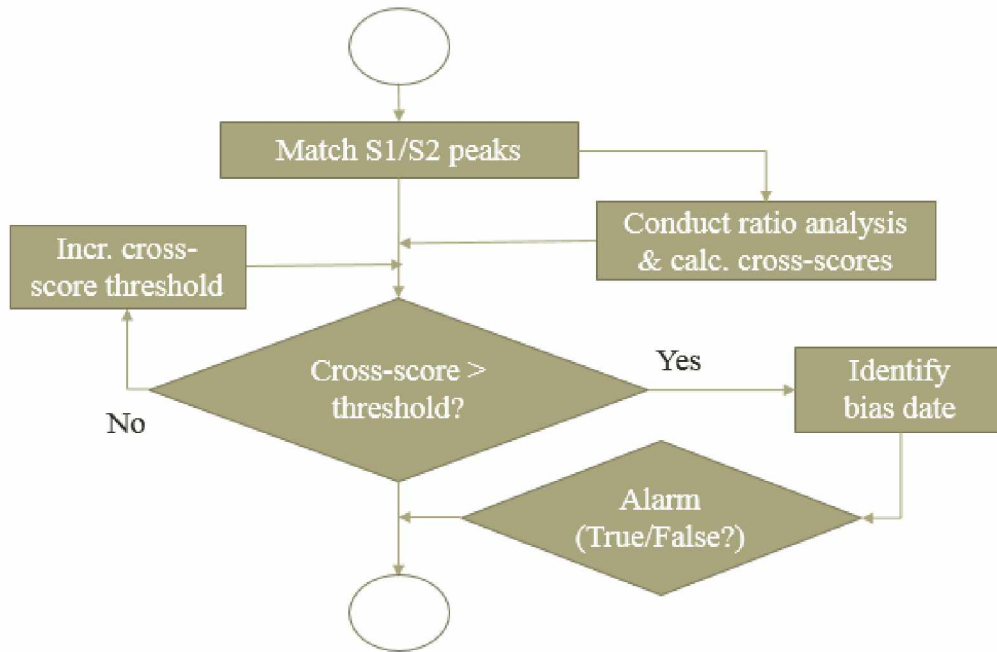


Figure 5.21: Cross-score algorithm.

Cross-Score Threshold Strategy:

From the previous section, it is understood that each of the peak numbers that are produced from the matching process is assigned with a cross-score value. A “cross-score threshold” (*cross-score Th*) in this context is a preset value that ranges between -10 to 10. At a given peak number, if the last 10 cross-scores observed were negative, a maximum score of -10 is possible. On the contrary, if all the 10 cross-scores observed are positive, a maximum score of 10 is possible. Any combinations other than these will result in cross-scores that range in between -10 and 10. The algorithm uses this following strategy to identify bias. If the cross-scores of peaks are continuously

showing negative values, it is the indication of negative bias, whereas if the cross-scores are continuously positive values, it indicates the presence of positive bias. In general, higher scores (\pm) implicate that S1 bias sensor ratios are drifting off from that of S2. Normally S1 and S2 should regularly cross each other to keep the cross-scores at low levels (0-2), due to the fact that these two sensors observe the processes that are tandem and identically cyclical.

Starting from the first peak (number) matched (in fact 10th peak to have at least 10 cross-scores for summing), the algorithm checks the cross-score of each subsequent peak against the lowest cross-score threshold (-10). If the cross-score at a peak is greater than the threshold, the algorithm captures the peaks start time as the “bias date of identification,” and calculates TTFD. Then, the threshold is increased to find the TTFD again. The threshold at which the TTFD value is minimum, is the desired threshold for that test. The cross-score algorithm is shown in Figure 5.21.

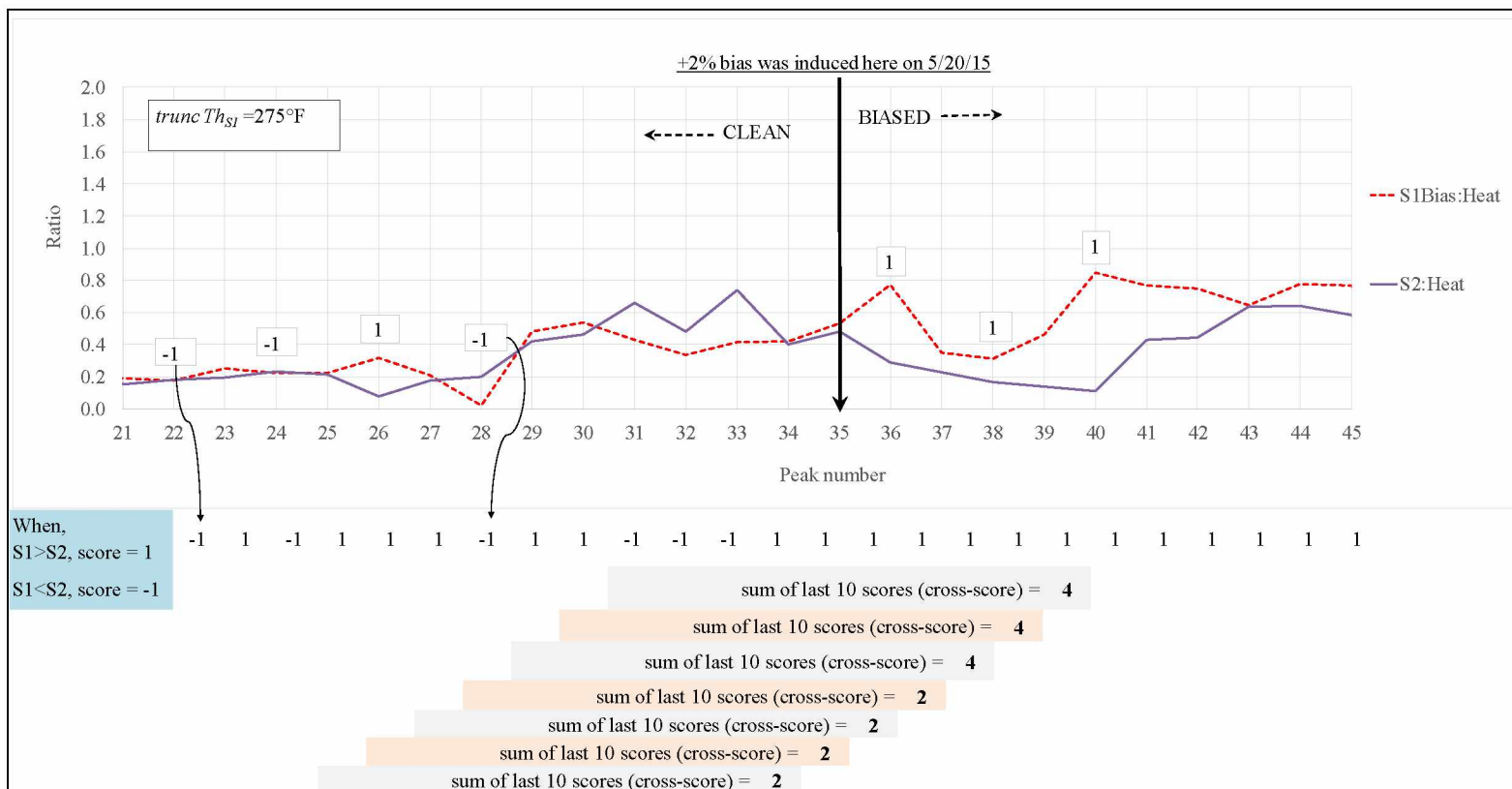


Figure 5.22: Cross-score calculations to identify bias (clean vs biased data).

Multiple Tests:

The MRFA algorithm is equipped with a single test, i.e., “Heat ratio test.” In order to exploit the ratio relations between the other sensors and to improve the TTFD performance of the algorithm, several other tests were included. If the Heat part of the ratio test is replaced with barren flow values it is called “BARNFL ratio test,” and if it is replaced with glycol flow values, the test is now a “GLYFL ratio test.” For any industrial process, the average and maximum values tend to be good indicators of change. Therefore, another class of tests, “value” tests, were created. Unlike the ratio tests that used data from two or more sensors, value tests utilized the average and maximum values of the S1 and S2 sensor peaks. In the “average value” test, peak-ave of S1 was compared to the peak-ave of S2. If the peak-ave of S1 was greater than the peak-ave of S2, the score of 1 was assigned to the peak. If not, the score was -1. In the “max value” test, the peak-max of S1 was compared to the peak-max of S2. If the peak-max of S1 was greater than the peak-max of S2, the score of 1 was assigned to the peak. If not, the score was -1. Class of tests are depicted in Figure 5.23.

In order to simplify the names of the five tests, they are called: Heat test, BARNFL test, GLYFL test, Ave test, and Max test. All tests are capable of identifying induced bias (on a particular date) with a cross-score threshold strategy, which was explained previously. The MRFA equipped with the new capabilities is named “improved MRFA algorithm.”

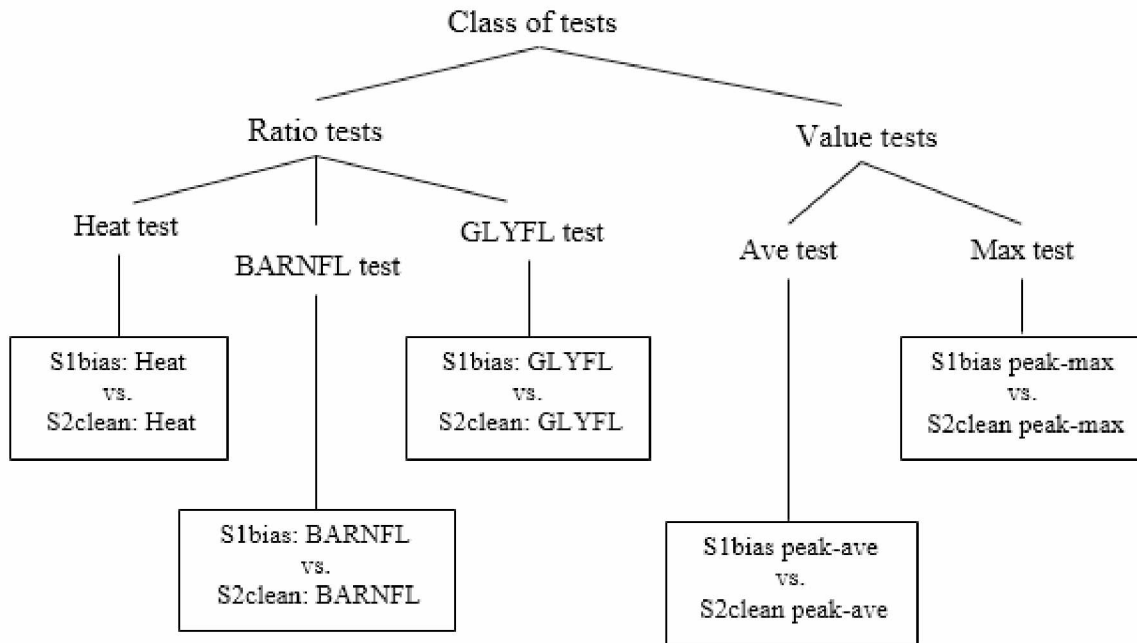


Figure 5.23: Classification of multiple tests.

True Alarm vs False Alarm:

In a typical experiment, bias is introduced on a random day and it stays until the end of the year from that day, mimicking the real life industrial situation. A test designed in this context needs to find the bias after it occurs (or introduced in this case). Ideally it should find the bias, on the day it was introduced, which is difficult due to the process variabilities and the algorithms' logical response. If a test finds bias after it is introduced, however, the result is a "success" and the alarm designated is "True alarm" (value=1). When a test identifies a date as bias introduction date even before the date bias is introduced, it is called "False alarm" or "failure" (value=0). The TTFD value is an absolute value and only indicates the difference in days between bias "introduction date" and the "finding date." Only the alarm mechanism indicates if a test is True (desired) or False (not desired) at a particular cross-score threshold value.

The success rate of a test is dependent on how good it is in producing true alarms. For instance, it can be observed from Figure 5.24 that a Heat test finds bias after four days (TTFD) from the day bias is introduced at a cross-score threshold of 6. It also finds bias within 19 days (TTFD) before the introduction day at a cross-score threshold of 3, which is not desired. Thus, it is understood that cross-score threshold 6 has a better success rate in finding bias for that particular test. Results from a single Heat test are summarized in Table 5.8. The ideal cross-score threshold at which a test can yield minimum TTFD values can be found through observation of TTFD values for multiple experiments, i.e., introduction of bias on multiple dates at multiple thresholds. From Table 5.8, it can be observed that when +2% bias is introduced on 5/20/15 (day 140 of the year, see highlighted row), at a *trunc th* of 275°F, the minimum TTFD (8.81 days) occurs at a cross-score threshold of 5 (several cross-score thresholds were applied). At a threshold of 3, the test was able to identify bias in 5.54 days; however, the alarm for this test is false (0) indicating that the bias was found earlier than the introduction date, which is not desired. Multiple tests have multiple TTFD performances that can be observed from Table 5.9. The visual representation of multiple bias tests for a bias introduction date of 6/1/15 is found in Figure 5.25.

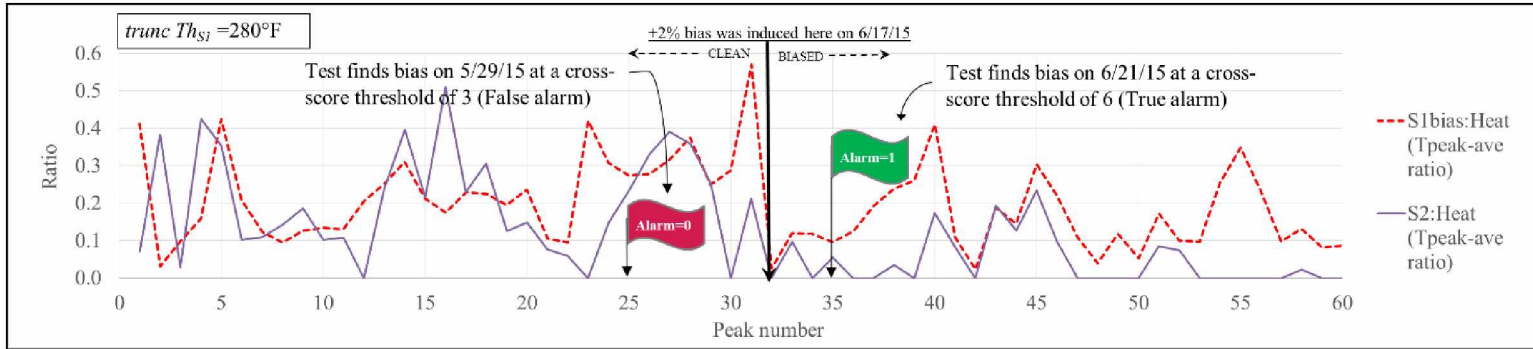


Figure 5.24: True alarm vs false alarm.

Table 5.8: S1 Bias detection, Heat ratio test ($\pm 2\%$ bias induced in S1, while S2 is clean).

Test-1												
Heat Ratio Test												
Day#	S1Bias start dt.	S1Bias end dt.	Induced bias %	Trunc threshold	S1Bias identified dt.	S1S2 cross-score	Cross-score threshold	Time till finding-Hrs.	Time till finding-Days (TTFD)	Alarm type		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	4/28/15 9:40 AM	0	-1	518:20:00	21.60	0	True alarm (1) Bias identified after it occurred in 9 days, a desired value!	
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	4/28/15 9:40 AM	0	0	518:20:00	21.60	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/5/15 7:20 AM	2	1	352:40:00	14.69	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/5/15 7:20 AM	2	2	352:40:00	14.69	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/14/15 11:00 AM	4	3	133:00:00	5.54	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/14/15 11:00 AM	4	4	133:00:00	5.54	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/28/15 7:30 PM	6	5	211:30:00	8.81	1		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/28/15 7:30 PM	6	6	211:30:00	8.81	1		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/29/15 7:10 PM	8	7	235:10:00	9.80	1		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/29/15 7:10 PM	8	8	235:10:00	9.80	1		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/31/15 3:00 AM	10	9	267:00:00	11.13	1		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	2	275	5/31/15 3:00 AM	10	10	267:00:00	11.13	1		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-12	2186:50:00	91.12	0		False Alarm (0) Bias identified in 22 days before it occurred! not desired!
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-11	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-10	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-9	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-8	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-7	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-6	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-5	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-4	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-3	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-2	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	2/17/15 9:10 PM	-2	-2	2186:50:00	91.12	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	4/28/15 9:40 AM	0	-1	518:20:00	21.60	0		
140	5/20/15 12:00 AM	12/30/15 12:00 AM	-2	275	4/28/15 9:40 AM	0	0	518:20:00	21.60	0		

Table 5.9: Multiples tests with bias identification dates.

					Test-1	Test-2	Test-3	Test-4	Test-5
					Heat ratio test	Max value test	Ave value test	BARNFL ratio test	GLYFL ratio test
Day#	S1Bias start dt	S1Bias end dt	Induced bias%	Trunk thres hold (°F)	S1Bias identified dt.	S1Bias identified dt.	S1Bias identified dt.	S1Bias identified dt.	S1Bias identified dt.
7	1/7/15 12:00 AM	12/30/15 12:00 AM	2	275	2/12/15 6:20 PM	2/20/15 1:50 AM	2/12/15 6:20 PM	2/20/15 1:50 AM	2/20/15 1:50 AM
7	1/7/15 12:00 AM	12/30/15 12:00 AM	2	275	2/12/15 6:20 PM	2/20/15 1:50 AM	2/12/15 6:20 PM	2/20/15 1:50 AM	2/20/15 1:50 AM
7	1/7/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	4/28/15 9:10 AM	4/29/15 9:20 PM	4/29/15 9:20 PM	4/29/15 9:20 PM
7	1/7/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	4/28/15 9:10 AM	4/29/15 9:20 PM	4/29/15 9:20 PM	4/29/15 9:20 PM
7	1/7/15 12:00 AM	12/30/15 12:00 AM	2	275	5/1/15 7:20 AM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM
7	1/7/15 12:00 AM	12/30/15 12:00 AM	2	275	5/1/15 7:20 AM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM
14	1/14/15 12:00 AM	12/30/15 12:00 AM	2	275	2/12/15 6:20 PM	2/20/15 1:50 AM	2/12/15 6:20 PM	2/20/15 1:50 AM	2/20/15 1:50 AM
14	1/14/15 12:00 AM	12/30/15 12:00 AM	2	275	2/12/15 6:20 PM	2/20/15 1:50 AM	2/12/15 6:20 PM	2/20/15 1:50 AM	2/20/15 1:50 AM
14	1/14/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	4/28/15 9:10 AM	4/29/15 9:20 PM	4/29/15 9:20 PM	4/29/15 9:20 PM
14	1/14/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	4/28/15 9:10 AM	4/29/15 9:20 PM	4/29/15 9:20 PM	4/29/15 9:20 PM
14	1/14/15 12:00 AM	12/30/15 12:00 AM	2	275	5/1/15 7:20 AM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM
14	1/14/15 12:00 AM	12/30/15 12:00 AM	2	275	5/1/15 7:20 AM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM
21	1/21/15 12:00 AM	12/30/15 12:00 AM	2	275	2/17/15 8:50 PM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM
21	1/21/15 12:00 AM	12/30/15 12:00 AM	2	275	2/17/15 8:50 PM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM
21	1/21/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	4/28/15 9:10 AM	4/29/15 9:20 PM	4/29/15 9:20 PM	4/29/15 9:20 PM
21	1/21/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	4/28/15 9:10 AM	4/29/15 9:20 PM	4/29/15 9:20 PM	4/29/15 9:20 PM
21	1/21/15 12:00 AM	12/30/15 12:00 AM	2	275	5/1/15 7:20 AM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM
21	1/21/15 12:00 AM	12/30/15 12:00 AM	2	275	5/1/15 7:20 AM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM
28	1/28/15 12:00 AM	12/30/15 12:00 AM	2	275	2/17/15 8:50 PM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM
28	1/28/15 12:00 AM	12/30/15 12:00 AM	2	275	2/17/15 8:50 PM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM
28	1/28/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	4/28/15 9:10 AM	4/29/15 9:20 PM	4/29/15 9:20 PM	4/29/15 9:20 PM
28	1/28/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	4/28/15 9:10 AM	4/29/15 9:20 PM	4/29/15 9:20 PM	4/29/15 9:20 PM
28	1/28/15 12:00 AM	12/30/15 12:00 AM	2	275	5/1/15 7:20 AM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM
28	1/28/15 12:00 AM	12/30/15 12:00 AM	2	275	5/1/15 7:20 AM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM
35	2/4/15 12:00 AM	12/30/15 12:00 AM	2	275	2/17/15 8:50 PM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM
35	2/4/15 12:00 AM	12/30/15 12:00 AM	2	275	2/17/15 8:50 PM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM	2/20/15 1:50 AM
35	2/4/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	2/23/15 5:00 PM	2/23/15 5:00 PM	2/23/15 5:00 PM	2/23/15 5:00 PM
35	2/4/15 12:00 AM	12/30/15 12:00 AM	2	275	2/20/15 1:50 AM	2/23/15 5:00 PM	2/23/15 5:00 PM	2/23/15 5:00 PM	2/23/15 5:00 PM
35	2/4/15 12:00 AM	12/30/15 12:00 AM	2	275	2/23/15 5:00 PM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM
35	2/4/15 12:00 AM	12/30/15 12:00 AM	2	275	2/23/15 5:00 PM	4/29/15 9:20 PM	5/1/15 7:20 AM	5/1/15 7:20 AM	5/1/15 7:20 AM

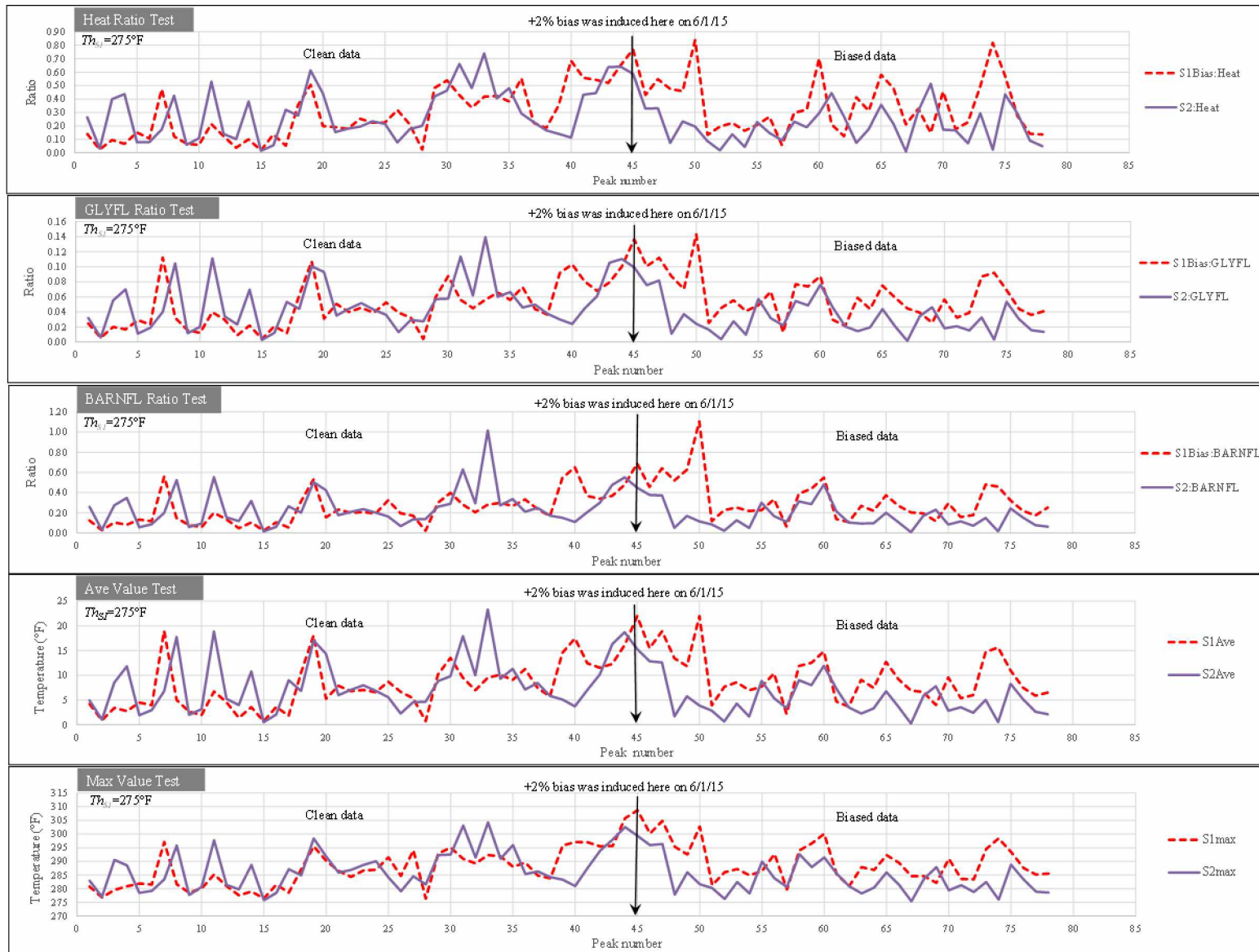


Figure 5.25: Multiple tests in finding bias (+2%) for S1 sensor data.

The “improved MRFA” algorithm that resulted from the inclusion of cross-score algorithm, and multiple tests with automatic detection capabilities is similar to the MRFAA algorithm (Figure 5.29) except for one aspect. It continues to use a “truncation threshold” (*trunc Th*) strategy, as opposed to the “dynamic threshold” strategy of MRFAA.

Disadvantage with Truncation Threshold (*trunc Th*) Strategy:

The only disadvantage with the improved MRFA is related to the choice of truncation threshold value. Several experimentations with the *trunc Th* value have revealed that it was highly sensitive to bias at a value of 275°F (effective threshold). However, at this threshold the algorithm was unable to capture peaks for the last quarter of the year; specifically after September, 2015. This is due to the fact that the strip vessel temperatures (observed by S1 and S2 for the time) often ran below 275°F. In the absence of S1 and S2 peaks—eliminated by truncation threshold strategy—the cross-score algorithm for various tests failed to detect bias. From the Figure 5.26 it is evident that no peaks were captured around the time in order to identify bias that was introduced in the month of November 2015. This means that the algorithm is not capable of calculating ratio statistics and cross-scores in order to detect bias. The disadvantage is mitigated through dynamic thresholding strategy, which is explained in the subsequent section.

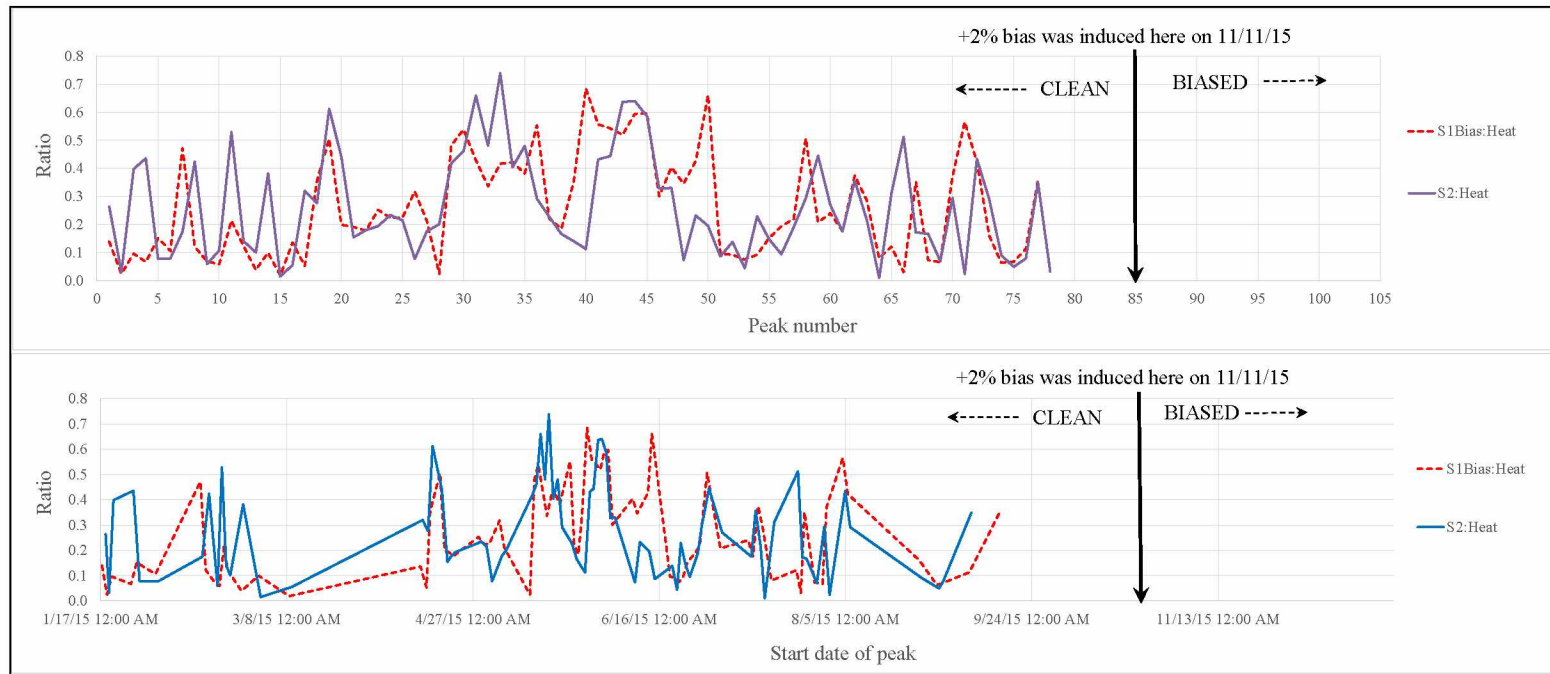


Figure 5.26: Disadvantage with truncation threshold ($trunc Th$).

Dynamic Thresholding:

In order to observe the effect of a reduced truncation threshold value (from 275°F to 260°F) in capturing peaks when bias is introduced in the November month of 2015, an experiment was conducted (Figure 5.27). It is noticed that more peaks were brought into the “matching” process, making the bias detection possible for that time frame. However, it is observed that lowering the *trunc th* has a compromising effect on the sensitivity of cross-scores with respect to the thresholds and ratios in finding bias. This is compensated by dynamically adjusting the *trunc Th* based on the maximum value of a peak. The new technique that dynamically adjusts the *trunc Th* value to improve the TTFD performance of a test is named, “dynamic thresholding” (*dyn Th*). The resultant improved algorithm is called MRFAA, and is Fortran based. Unlike *trunc th* value, a *dyn Th* value is not a fixed one. For the purpose of the study, a 90% of the maximum reading value of any given peak (peak-max) is chosen as the dynamic threshold for that peak. For instance, if the maximum value of the readings of a S1 peak is 290°F, then the 90% of the value, i.e., 261°F, is established as the *dyn th*. The strategy provides flexibility in adjusting the threshold to cope with the varying high values of a data-set. The strategy is depicted in Figure 5.28. From the figure, it can be observed that the S2 peak with a peak-max value of 261°F was not captured at a truncation threshold of 265°F. However, it was captured with dynamic thresholding strategy when the threshold is adjusted to 90% of the peak-max value (234.9°F).



Figure 5.27: Effect of lowering the threshold to 260°F on peaks captured.

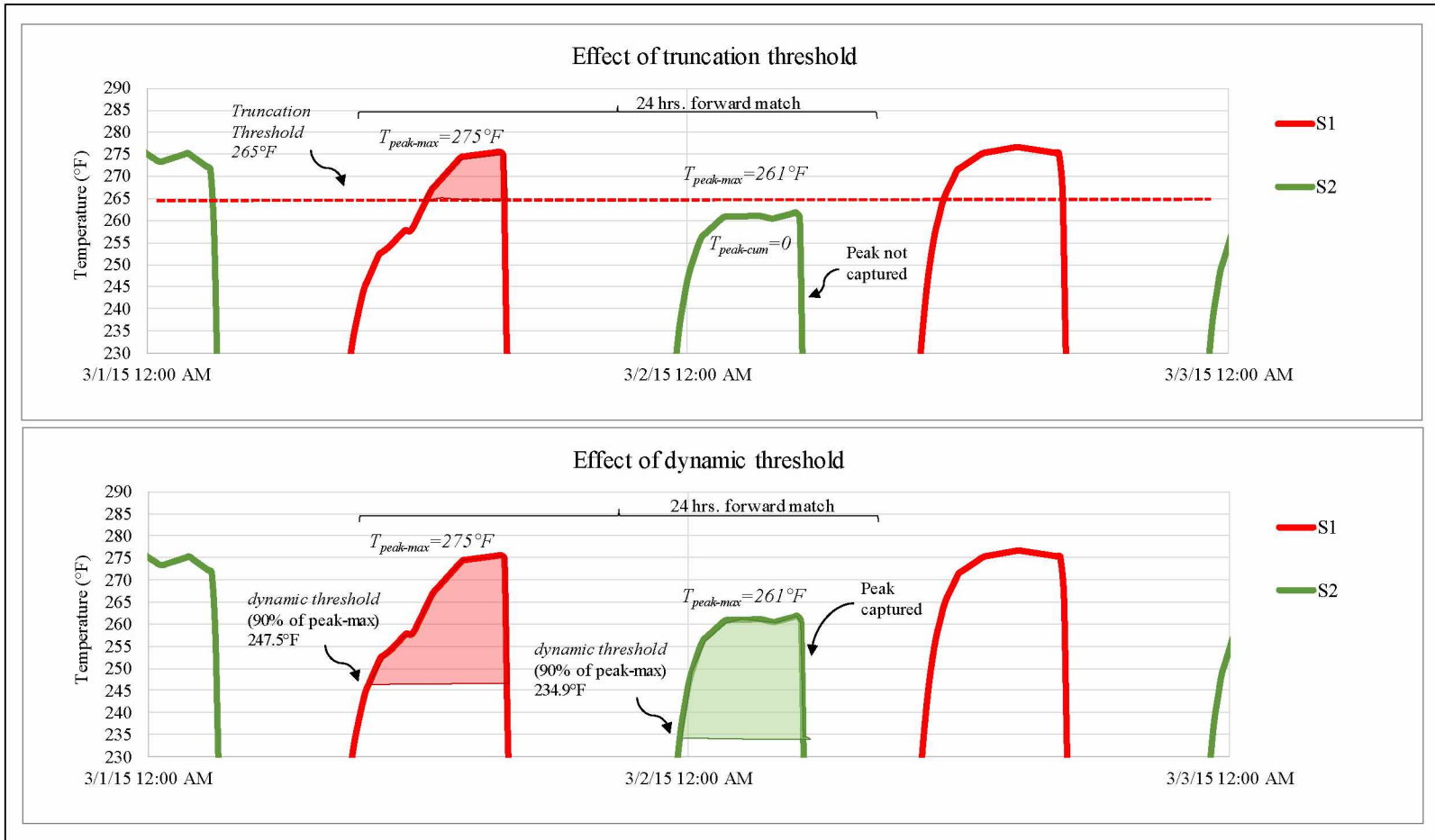


Figure 5.28: Truncation threshold vs dynamic threshold.

5.5.5.3 MRFAA: Algorithm Description

The flowchart for the algorithm is shown in the Figure 5.29. The following are some salient features of the MRFAA algorithm. Since the algorithm is an extension of the MRFA algorithm, there is no exclusive description provided. The only additional feature of the algorithm, i.e., dynamic thresholding was described in the previous section. MRFAA is a Fortran-based algorithm.

(i) Fully automated algorithm, right from the choice of date for bias induction to the writing of results to the output file of choice (text or Excel).

(ii) Ability to select different thresholds; can switch between dynamic and truncation threshold if necessary. A combination of these threshold is also possible. A snapshot of the Fortran interface used is shown in Figure 5.30.

(iii) Ability to add an automated new test by adding a few lines of code.

(iv) Ability to validate the user choice of test by calculating performance statistics for the success (true alarms) and failure rates (false alarms).

(v) Ability to identify either positive or negative bias.

(vi) When compared to previous Matlab-based algorithms MRFAA has better speed of execution.

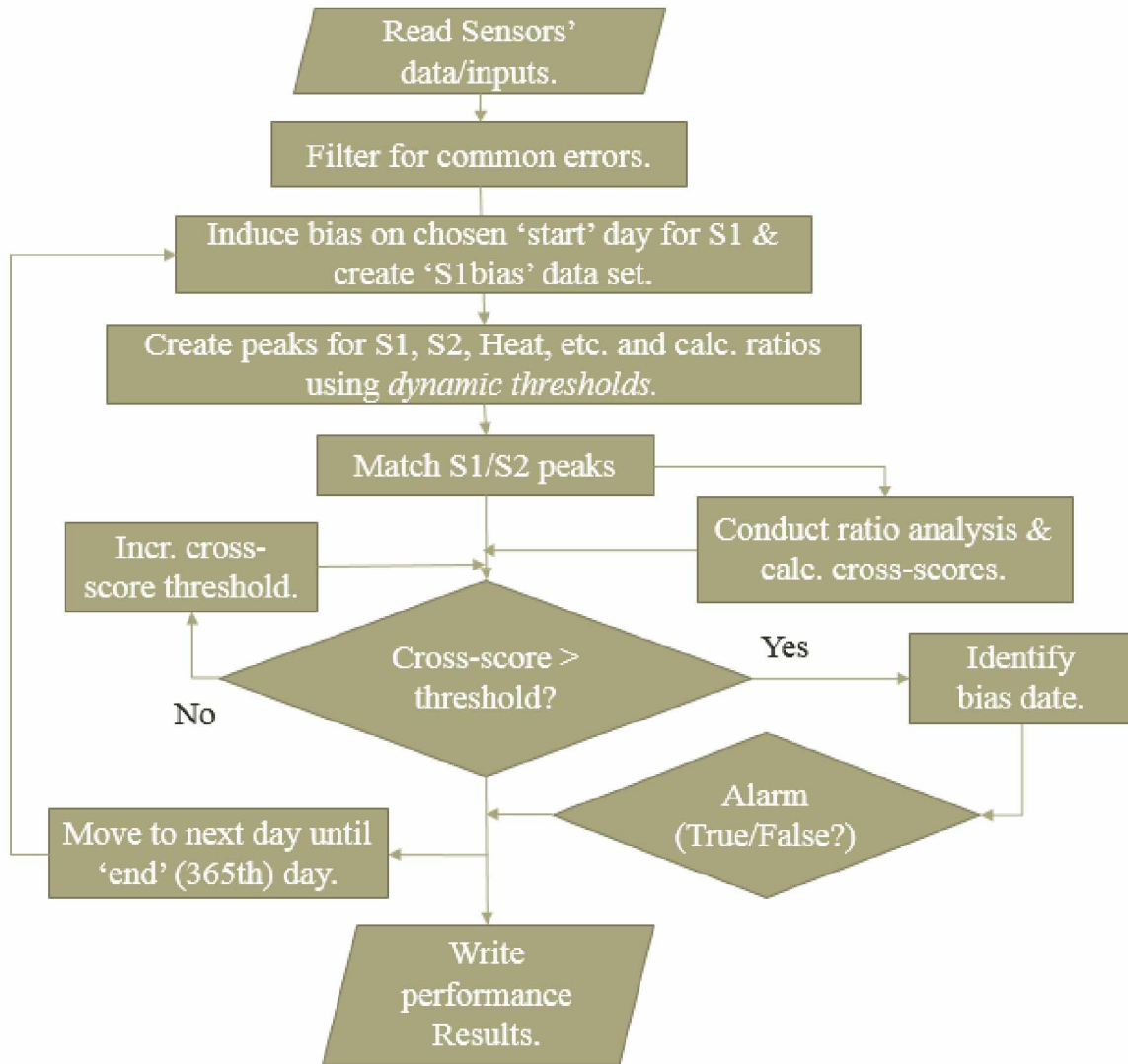


Figure 5.29: Flowchart for MRFA automation with dynamic thresholding (MRFAA).

```

!=====initialize=====
RunNum=0
s1BiasEndDt=42368.0000000104d+0      !... 12/30/15 12:00 AM (end date of bias induction in sensor s1)
s1BiasPCT=2.0                          !... Bias percentage for sensor s1
s2BiasPCT=0                             !...or Trunk threshold
cutoffTemp=220.0

S1maxcut=600                            !.. S1/S2 (f)
S1mincut=150                             !.. H1 (f)
H1maxcut=500                             !.. H2 (f)
H1mincut=150                             !.. H3 (f)
H2maxcut=600                             !.. H4 (f)
H2mincut=150
H3maxcut=400
H3mincut=0
H4maxcut=400
H4mincut=0
BARNFLmaxcut=200                         !.. GPM
BARNFLmincut=20                          !.. GPM
GLYFLmaxcut=600
GLYFLmincut=150

Percnt1=0.90                             !..this is percentile of 'max peak value', used in dyn thld
dynThrld_desn=1                          !...to keep dyn threshold, choice=1, if not, defaults to cutoffTemp/Trunk threshold
RowsToAllocate= 52188                    !...How many rows max is your data-set?!51586(2016) !52188(2015)

!=====READ INPUT DATA=====
call ReadInputData (RowsToAllocate,&
                   Row, DTIME, S1, S2, BARNFL, H1, H2, GLYFL, H3, H4,totalRows)

!=== TESTS =====
decision=3
!if (decision==3) then
!----3 tests-----
print *, '====3 TESTS===='
do i =42005,42365
    RunNum=RunNum+1
    print *, 'RunNum=',RunNum
    s1BiasStartDt=db1e(i)
    numTests=3
    call sensors (RunNum,s1BiasStartDt,s1BiasEndDt,s1BiasPCT,s2BiasPCT,cutoffTemp, &
                 Row, DTIME, S1, S2, H1, H2, H3, H4,totalRows,&
                 S1maxcut,S1mincut,H1maxcut,H1mincut,&
                 H2maxcut,H2mincut,H3maxcut,H3mincut,H4maxcut,H4mincut,Percnt1,dynThrld_desn,numTests)
end do

```

Figure 5.30: Fortran interface where multiple input options are available.

5.6 Results and Validation

This section presents results from various experimentations with MRFAA algorithm along with validation.

5.6.1.1 MRFAA Algorithm

Using MRFAA, multiple bias identification tests were conducted by inducing bias ($\pm 2\%$) on each day of the year (365 days of the year) in the S1 sensor data set. Then each test is validated by capturing TTFD values. The performance of each test in terms of success (True alarms) and failure rates (False alarms) is also captured (see Table 5.10).

Table 5.10: Multiples tests with dynamic thresholding (90% of peak-max value).

Day#	S1Bias start dt	S1Bias end dt	Induced bias %	Truncation threshold (T _{trunc} Th _{ST})	Cross-score threshold (cross-score Th _{ST})	Test-1		Test-2		Test-3		Test-4		Test-5		Test-6	
						Heat ratio test		Max value test		Ave value test		BARNFL ratio test		GLYFL ratio test		Combined test	
						Time till finding-days (TTFD)	Alarm type	Time till finding-days (TTFD)	Alarm type	Time till finding-days (TTFD)	Alarm type	Time till finding-days (TTFD)	Alarm type	Time till finding-days (TTFD)	Alarm type	Time till finding-days (TTFD)	Alarm type
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-10	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-9	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-8	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-7	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-6	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-5	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-4	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-3	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-2	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	-1	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	1	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	2	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0	78.76	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	3	74.69	0	74.69	0	76.38	0	74.69	0	39.75	0	74.69	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	4	74.69	0	74.69	0	76.38	0	74.69	0	39.75	0	74.69	0
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	5	42.42	1	7.54	0	74.69	0	365.00	0	38.56	0	42.42	1
92	4/2/15 12:00 AM	12/30/15 12:00 AM	2	255	6	42.42	1	7.54	0	74.69	0	365.00	0	38.56	0	42.42	1

The results from the automatic detection of the bias using MRFAA algorithm are given below. In fact, various tests were conducted with a combination of *dyn Th*, i.e., 90% of the peak-max value of any matched peak, and a *trunc Th* of 255°F (the bigger of the two is selected as the dynamic threshold for a peak). In this context, the *trunc Th* is only used to prevent the dynamic threshold from capturing peaks with too low of readings (<255°F); too low values can compromise the sensitivity of “ratios” to bias. For each threshold (ranges between -10 to 10), bias was introduced at the first day of the year (2015) and the test captures bias identification date and the TTFD for that threshold. The process was repeated for 365 days of the year. When a test (cross-score algorithm) completely failed to identify bias, a value of 365 was assigned as TTFD. This was to symbolically indicate the test attempted to detect bias by checking all the peaks that were matched (representing the whole year) and failed. A snapshot of various tests and results are presented in Table 5.10. A +2% and -2% bias were introduced in two separate cases, in each and every day of the year 2015 for S1 data to observe the effect on TTFD values. The success rates for the “alarm” tests and their corresponding TTFD performances are shown in Figure 5.31. The results shown in Figure 5.31 are also summarized in tabular form for reader’s convenience. The alarm and TTFD performances at cross-score thresholds 5 or 6 are summarized in Table 5.11. The summary for threshold 7 or 8 is available in Table 5.12.

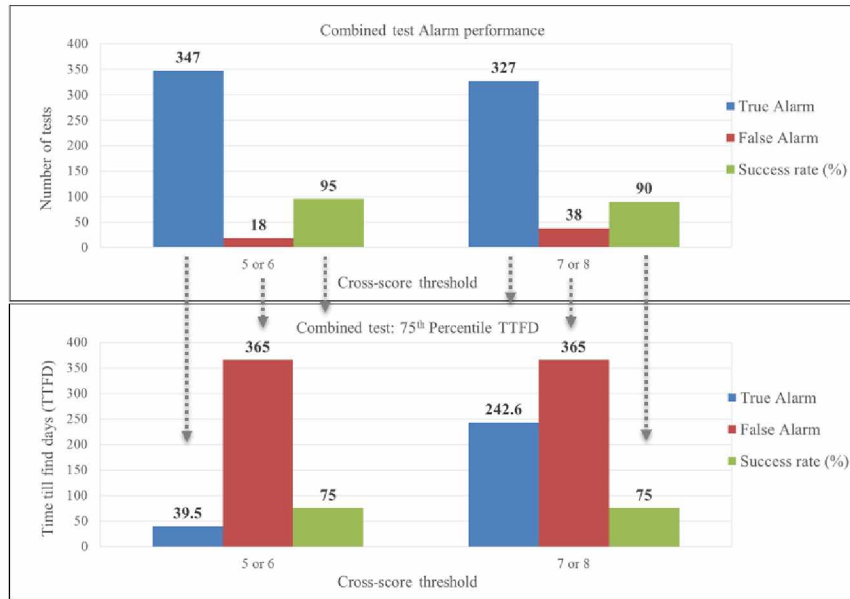


Figure 5.31: Combined test - Alarm and TTFD performances.

Table 5.11: Combined test - Performances at Thresholds 5 or 6.

Cross-score thresholds	Alarm Performance	TTFD Performance (days)
	5 or 6	5 or 6 (75% of alarms)
No. of True alarms	347	39.5 ⁺ (260.3)
No. of False alarms	18	365* (13.5)
Total tests	365	273.8 (75% of 365)
Success (True alarm) rate	95%	

⁺75% of the 347 True alarms (260.3) have detected bias within 39.5 days (TTFD) at the thresholds, 5 or 6.

*75% of the 18 False alarms (13.5) have completely failed to detect bias, i.e. took 365 days (TTFD) at the thresholds, 5 or 6.

Table 5.12: Combined test - Performances at Thresholds 7 or 8.

Cross-score thresholds	Alarm Performance	TTFD Performance (days)
	7 or 8	7 or 8 (75% of Alarms)
No. of True alarms	327	242.6 ⁺ (245.3)
No. of False alarms	38	365* (28.5)
Total tests	365	273.8 (75% of 365)
Success (True alarm) rate	90%	

⁺75% of the 327 True alarms (245.3) have detected bias within 242.6 days (TTFD) at the thresholds, 7 or 8.

*75% of the 38 False alarms (28.5) have completely failed to detect bias, i.e. took 365 days (TTFD) at the thresholds, 7 or 8.

The results from the analysis where +2% bias was inducted in S1 were presented in this section. However, it should be noted that similar trends were observed for -2% bias. The tests with minimum TTFD values and corresponding cross-score thresholds can be observed in Table 5.10.

A “Combined test” is designed to capture the minimum TTFD value of all tests together at a particular threshold. For instance, it can be observed from Table 5.10 that when bias is induced on 4/2/15, the algorithm that can find it with success (alarm=1) at a minimum TTFD (42.42 days) is Heat test at a cross-score threshold of 5 or 6. The Combined test is also capable of providing the success (true alarms) and failure (false alarm) rates of all the tests together.

The true alarm performance of the Combined test can be found in Table 5.13. In 75% of cases, the algorithms find the bias within 39.5 days (see highlighted text in the table). The results can be visually observed in Figure 5.32 .The Combined test false alarm performance in terms of the percentages of cases a test is yielding false alarms is found in Table 5.14. From the table it is observed that when all the tests fail, 75% of the time the TTFD value is 271 days at a threshold of 4. The performance of all the other tests can be found in the appendix of this chapter.

Table 5.13: Combined test True alarm performance.

Cross-score Threshold	5th percentl_True	25th percentl_True	50th percentl (Median)_True	75th percentl_True	95th percentl_True	Alarm (True)
-10	365.0	365.0	365.0	365.0	365.0	1
-9	365.0	365.0	365.0	365.0	365.0	1
-8	365.0	365.0	365.0	365.0	365.0	1
-7	365.0	365.0	365.0	365.0	365.0	1
-6	365.0	365.0	365.0	365.0	365.0	1
-5	365.0	365.0	365.0	365.0	365.0	1
-4	365.0	365.0	365.0	365.0	365.0	1
-3	365.0	365.0	365.0	365.0	365.0	1
-2	365.0	365.0	365.0	365.0	365.0	1
-1	365.0	365.0	365.0	365.0	365.0	1
0	365.0	365.0	365.0	365.0	365.0	1
1	365.0	365.0	365.0	365.0	365.0	1
2	365.0	365.0	365.0	365.0	365.0	1
3	7.2	365.0	365.0	365.0	365.0	1
4	7.2	365.0	365.0	365.0	365.0	1
5	2.6	8.5	17.4	39.5	63.1	1
6	2.6	8.5	17.4	39.5	63.1	1
7	9.0	62.1	152.3	242.6	365.0	1
8	9.0	62.1	152.3	242.6	365.0	1
9	9.4	18.9	41.2	365.0	365.0	1
10	9.4	18.9	41.2	365.0	365.0	1

Table 5.14: Combined test False alarm performance.

Cross-score Threshold	5th percentl_False	25th percentl_False	50th percentl (Median)_False	75th percentl_False	95th percentl_False	Alarm (False)
-10	18.3	90.5	180.8	271.0	343.2	0
-9	18.3	90.5	180.8	271.0	343.2	0
-8	18.3	90.5	180.8	271.0	343.2	0
-7	18.3	90.5	180.8	271.0	343.2	0
-6	18.3	90.5	180.8	271.0	343.2	0
-5	18.3	90.5	180.8	271.0	343.2	0
-4	18.3	90.5	180.8	271.0	343.2	0
-3	18.3	90.5	180.8	271.0	343.2	0
-2	18.3	90.5	180.8	271.0	343.2	0
-1	18.3	90.5	180.8	271.0	343.2	0
0	18.3	90.5	180.8	271.0	343.2	0
1	18.3	90.5	180.8	271.0	343.2	0
2	18.3	90.5	180.8	271.0	343.2	0
3	18.3	90.5	180.7	271.0	365.0	0
4	18.3	90.5	180.7	271.0	365.0	0
5	365.0	365.0	365.0	365.0	365.0	0
6	365.0	365.0	365.0	365.0	365.0	0
7	308.9	365.0	365.0	365.0	365.0	0
8	308.9	365.0	365.0	365.0	365.0	0
9	365.0	365.0	365.0	365.0	365.0	0
10	365.0	365.0	365.0	365.0	365.0	0

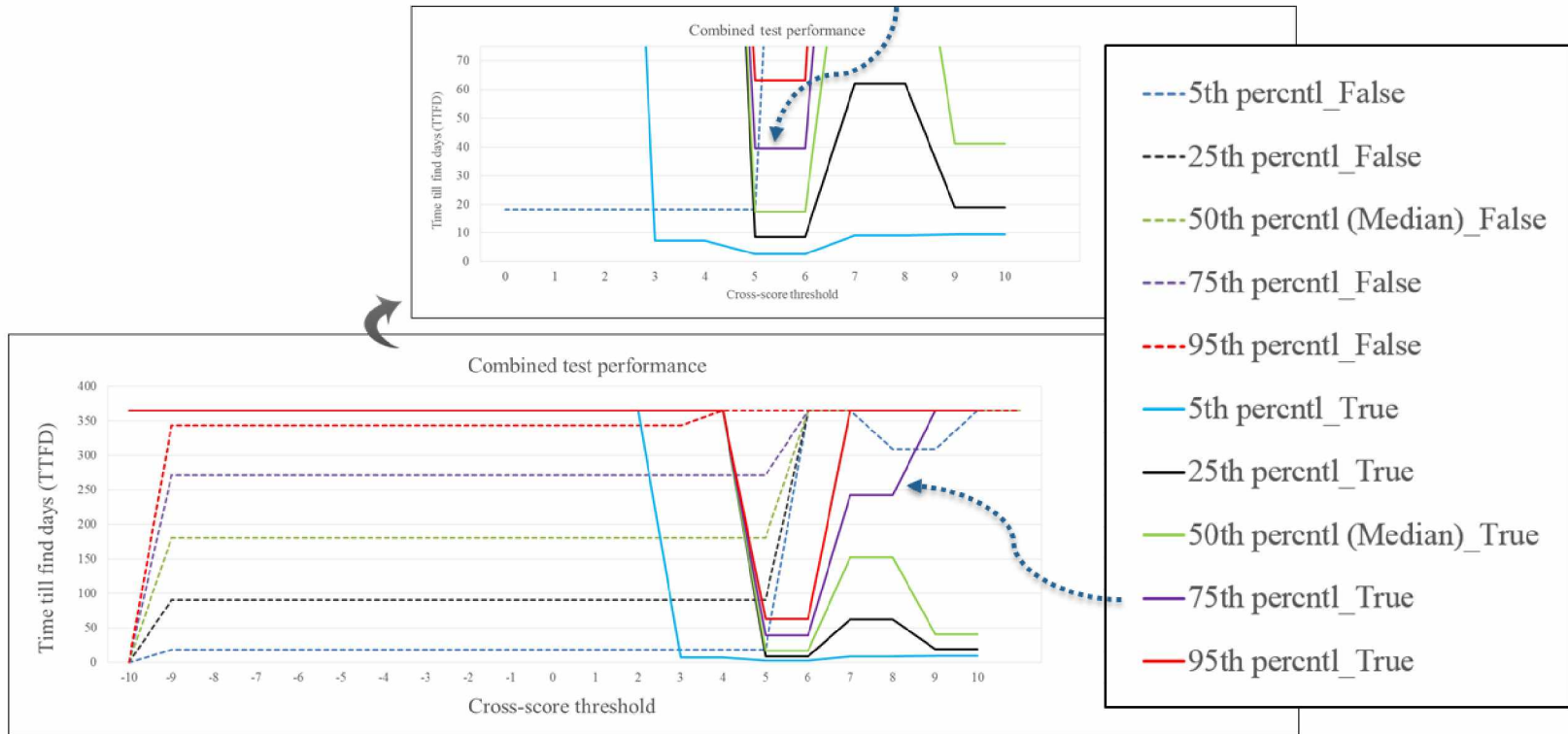


Figure 5.32: Combined test performance.

5.7 Discussion

Various algorithms described in the chapter are based on exploiting ratios that depend on sensor interrelations. The multiple ratio function analysis (MRFA) introduced is based on exploiting the strip vessel sensor (S1 and S2) relations with combined “Heat” of the heat exchanger sensors all together. The main disadvantage with this algorithm is that it cannot achieve the bias finding process automatically, which is time consuming and a major hindrance for conducting multiple tests on multiple days. The MRFA algorithm’s efficiency is mainly dependent on the sensitivity of ratios at higher temperatures, i.e., between 270 through 280°F. The truncation threshold value was designed based on that premise and eliminates peaks (and ratios) that are less sensitive to bias. The disadvantage with establishing truncation threshold at higher values is that it eliminates some peaks that are essential in identifying bias; for instance, the peaks from the last quarter of the year 2015. The MRFA is chiefly a heat ratio test. A dynamic threshold strategy adjusts the threshold to the maximum value of the peak-readings, and compensates for the disadvantages of truncation threshold; a 90% of peak-maximum value is selected in this research. The Fortran-based MRFAA algorithm employs dynamic thresholding strategy and hence, is more efficient than MRFA.

It can be observed from the Combined test results that all the tests together have a success (True alarms) rate of 95% at the cross-score thresholds of 5 or 6. Out of 365 tests conducted by inducing bias on each day of the year, 347 tests showed true alarm performance, and 18 tests failed. In those “True” alarms, 75% of the cases, all tests together are finding the bias within 39.5 days, and in 63 days for 95% of the cases. Out of 18 “False” alarms, 75% of the cases the tests completely failed (TTFD=365). At the cross-score thresholds of 7 or 8, the true alarm success rate is 90% (327) out of 365 tests; however, the TTFD (242.6 days for 75% of the cases) performance at these

thresholds is poor when compared to the thresholds at 5 or 6. Given that a combined cycle of strip vessel sensors (S1 and S2) takes more than a day, identifying the bias within a month span of time is very valuable for the industries. When all the tests fail (False alarms) in 95% of the cases, it is when the cross-scores are low which is observed from the high TTFD values, i.e., >340 days. This means S1 and S2 are exhibiting perfect cyclical behavior, and due to the cancelling effect, the cross-score values stay low in between the values -2 to 2. If the cross-scores starts to sway towards positive high values (>4) or negative low values (<-4), the algorithms detects bias. Coming to the individual tests, the Heat test was capable of detecting bias within 33 days after induction for 75% of the cases, and, within 55 days for 95% of the cases at cross-score thresholds of 5 and 6. The “max value test” identified bias in 44 days for 75% of the cases at thresholds of 7 and 8. The average value test can find bias in 65 days for 50% of the cases. This could be attributed to the minimizing effect of the “average” on peak-readings and bias, which is disadvantageous for the sensitivity at very lower magnitudes of bias (2%). Average test may be better suited for errors of higher magnitude but for the subtle errors like bias, it might be disadvantageous. The BARNFL and GLYFL tests did not show promise, and are less sensitive to the dynamic thresholding strategy. The ratios of flows to strip vessel temperatures are very minute and produced low cross-score values. At such low cross-scores, the algorithm is less effective and sensitive to thresholds in finding bias.

5.8 Conclusions

Identification of bias at 2% is crucial for Pogo and valuable for the optimum operation of the strip vessels. It can be concluded from the results that bias as low as 2% in magnitude can be identified in a reasonable amount of time with MRFAA algorithm. The algorithms developed in this chapter are based on exploiting interrelations between sensors’ readings in the carbon stripping

circuit, in the form of ratio function relations. The MRFA algorithm is capable of performing single heat ratio test with manual analysis of the results. A cross-score algorithm is the key feature of MRFA that is capable of finding the bias induction date with a cross-score thresholding mechanism. To exploit the relations between S1 and S2 with other sensors readings, i.e., BARNFL and GLYFL, additional ratio tests were added (BARNFL ratio test and GLYFL ratio test). The maximum and average values of the strip vessel cycles play an important role in observing the temperature trends. Due to this reason, “average value” and “maximum value” tests that can compare peak-max and peak-ave values for S1 and S2, respectively, were added. Several tests with the improved MRFA algorithm revealed that the truncation thresholding criteria is disadvantageous in finding bias at certain periods of the year, specifically, the last quarter of the year. A dynamic thresholding strategy that depends on peak-max values (90%) was added to the MRFA algorithm to produce a fully automated multi-test capable MRFAA algorithm.

A “Combined test” was created to capture the time till find days (TTFD) of all the tests together. The results from the test proved that all the tests together have a high success rate (95%) in terms of True alarms at the cross-score thresholds of 5 or 6. From the Combined test results, it can be observed that 75% of the cases the tests are findings the bias within 39.5 days. Regarding the individual tests, the Heat test is the most effective, which is capable of identifying the bias within 33 days after induction 75% of the time at cross-score thresholds of 5 and 6. The tests that used GLYFL and BARNFL were not very effective. Although Combined test results at the thresholds of 7 and 8 have the True alarm success rate of 90%, it had poor TTFD performance (242.6 days for 75% cases). In this context, it can be concluded that at cross-score thresholds of 5 or 6 all the tests together have high chance of detecting bias for any future data set. A -2% random error was detected with a similar success rate. Additional tests can be added in the future to

improve the TTFD performance of the MRFAA algorithm. It is also scalable and generalized to any industry due to its flexibility and speed.

At this time the algorithm is not capable of identifying bias in several sensors at the same time in a multi-sensor environment. The algorithm is also not capable at this time of identifying different types of errors other than bias. Presence of relations between sensors is the fundamental requirement for the algorithm in order to be successful.

5.9 Acknowledgements

The author expresses his sincere gratitude to his academic advisor Professor Rajive Ganguli, all the members of the graduate advisory committee for their help and guidance, and to the Mining Engineering Research Endowment (MERE) established at the University of Alaska Fairbanks (UAF) for financial support. The author also would like to thank Fort Knox Mine of Kinross Corporation and Pogo Mine of Sumitomo Metal Mining Company for their support of MERE and for providing the necessary data, facilitating site visits, etc. The author extends his gratitude to College of Engineering and Mines (CEM), and, the Department of Mining and Geological Engineering (MinGeo) at UAF for providing the valuable laboratory and other academic resources to accomplish the project.

5.10 References

Baljak, V., Tei, K., and Honiden, S., 2012, "Classification of faults in sensor readings with statistical pattern recognition," SENSORCOMM 2012: The Sixth International Conference on Sensor Technologies and Applications, Aug 19-24, 2012, Rome, Italy, pp. 270-276.

- Brown, M., 2012, "Data mining techniques," IBM, www.ibm.com/developerworks/library/ba-data-mining-techniques/, Accessed October 2017.
- Buxton, M., and Benndorf, J., 2013, "The use of sensor derived data in real time mine optimization: a preliminary overview and assessment of techno-economic significance," SME Annual Meeting, Feb 24 - 27, 2013, Denver, CO, Preprint 13-038, pp. 5-5.
- Fast, J., 2016, "Carbon stripping," Denver Mineral Engineers, Inc., www.denvermineral.com/carbon-stripping/, Accessed October 2017.
- Hou, Z., Lian, Z., Yao, Y., and Yuan, X., 2006, "Data mining based sensor fault diagnosis and validation for building air conditioning system," Elsevier, DOI:10.1016/j.enconman.2005.1.010.
- Konigsmann, E., Abols, J., Pyecha, J., and Boudreau, T., 2017, "The inline leach reactor installation at Teck-Pogo Inc. Pogo Mine," www.gekkos.com/documents/033TheInLineLeachReactorInstallationAtTeckPogIncPogoMine.pdf, pp.1-11, Accessed October 2017.
- Kusiak, A., and Song, Z., 2009, "Sensor fault detection in power plants," Journal of Energy Engineering, December, 2009, No. 127, pp. 127-135, DOI: 10.1061/(ASCE)0733-9402(2009)135:4 (127).
- Pogo Mine, 2016, "Pogo mill process descriptions," Print medium handout by Andrew Maxon, Pogo Mill Facility, Pogo Mine, Received January 2016.
- Pogo Mine, 2017, "Who we are," <http://pogominealaska.com/>, Accessed October 2017.

U.S. Department of Energy, 2000, "Mineral processing technology road map," U.S. Department of Energy, www.eere.energy.gov/industry/mining/, Accessed May 2006.

Wang, H., Chai, T., Ding, J., and Brown, M., 2009, "Data driven fault diagnosis and fault tolerant control: some advances and possible new directions," *ACTA Automatica Sinica*, Vol. 35, No. 6, June 2009, pp. 739-747.

Yang, C., Liu, C., Zhang, X., and Nepal, S., 2015, "Time efficient approach for detecting errors in big sensor data on cloud," *IEEE Transactions on Parallel and Distributed Systems*, February, 2015, Vol. 26, No. 2, pp. 329-339, DOI: 10.1109/TPDS.2013.2295810.

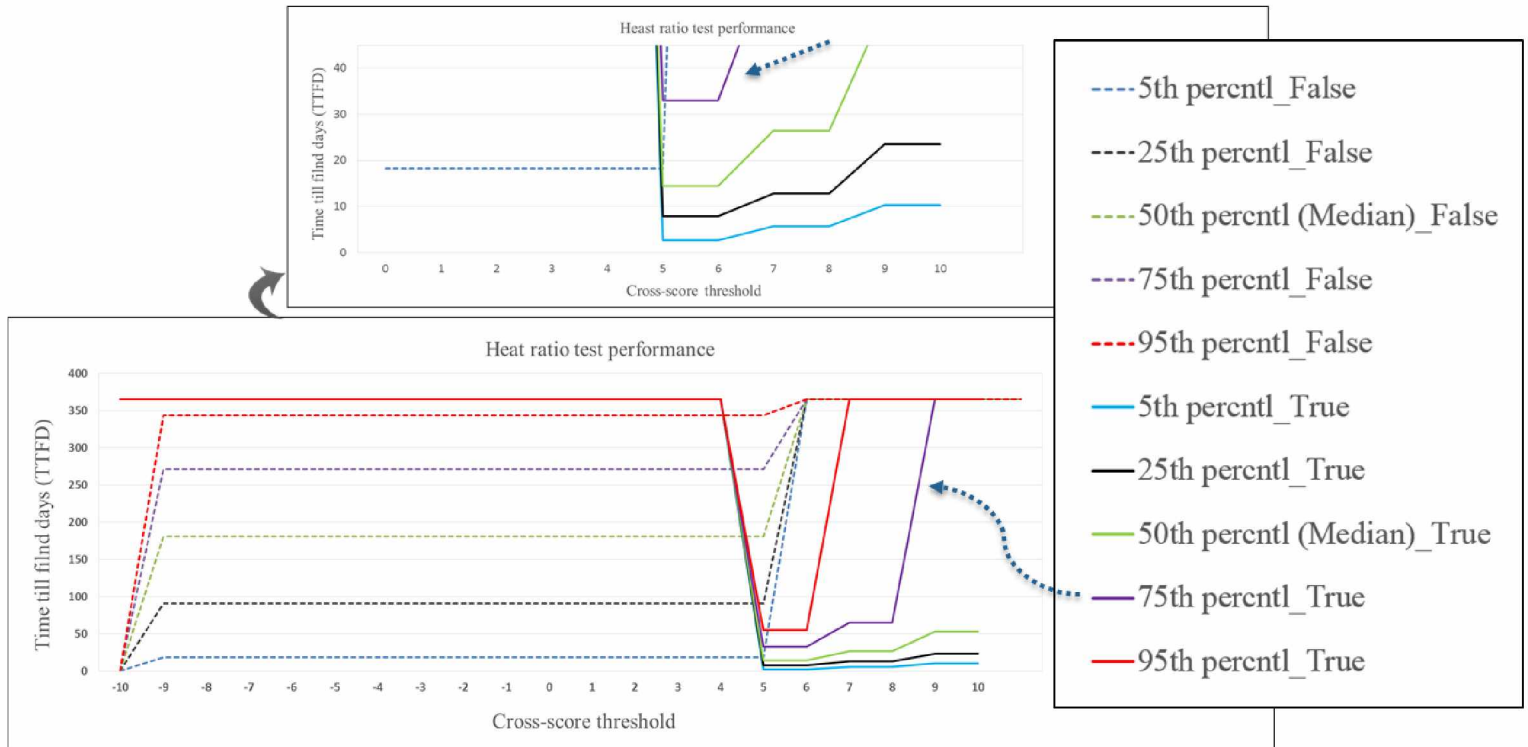


Figure 5.33: Heat ratio test performance.

Table 5.15: Heat ratio test True alarm performance.

Cross-score Threshold	5th percentl_True	25th percentl_True	50th percentl (Median)_True	75th percentl_True	95th percentl_True	Alarm (True)
-10	365.0	365.0	365.0	365.0	365.0	1
-9	365.0	365.0	365.0	365.0	365.0	1
-8	365.0	365.0	365.0	365.0	365.0	1
-7	365.0	365.0	365.0	365.0	365.0	1
-6	365.0	365.0	365.0	365.0	365.0	1
-5	365.0	365.0	365.0	365.0	365.0	1
-4	365.0	365.0	365.0	365.0	365.0	1
-3	365.0	365.0	365.0	365.0	365.0	1
-2	365.0	365.0	365.0	365.0	365.0	1
-1	365.0	365.0	365.0	365.0	365.0	1
0	365.0	365.0	365.0	365.0	365.0	1
1	365.0	365.0	365.0	365.0	365.0	1
2	365.0	365.0	365.0	365.0	365.0	1
3	365.0	365.0	365.0	365.0	365.0	1
4	365.0	365.0	365.0	365.0	365.0	1
5	2.6	7.9	14.4	32.9	55.2	1
6	2.6	7.9	14.4	32.9	55.2	1
7	5.7	12.8	26.4	64.8	365.0	1
8	5.7	12.8	26.4	64.8	365.0	1
9	10.3	23.5	53.2	365.0	365.0	1
10	10.3	23.5	53.2	365.0	365.0	1

Table 5.16: Heat ratio test False alarm performance.

Cross-score Threshold	5th percentl_False	25th percentl_False	50th percentl (Median)_False	75th percentl_False	95th percentl_False	Alarm (False)
-10	18.3	90.5	180.8	271.0	343.2	0
-9	18.3	90.5	180.8	271.0	343.2	0
-8	18.3	90.5	180.8	271.0	343.2	0
-7	18.3	90.5	180.8	271.0	343.2	0
-6	18.3	90.5	180.8	271.0	343.2	0
-5	18.3	90.5	180.8	271.0	343.2	0
-4	18.3	90.5	180.8	271.0	343.2	0
-3	18.3	90.5	180.8	271.0	343.2	0
-2	18.3	90.5	180.8	271.0	343.2	0
-1	18.3	90.5	180.8	271.0	343.2	0
0	18.3	90.5	180.8	271.0	343.2	0
1	18.3	90.5	180.8	271.0	343.2	0
2	18.3	90.5	180.8	271.0	343.2	0
3	18.2	90.4	180.7	270.9	343.1	0
4	18.2	90.4	180.7	270.9	343.1	0
5	365.0	365.0	365.0	365.0	365.0	0
6	365.0	365.0	365.0	365.0	365.0	0
7	365.0	365.0	365.0	365.0	365.0	0
8	365.0	365.0	365.0	365.0	365.0	0
9	365.0	365.0	365.0	365.0	365.0	0
10	365.0	365.0	365.0	365.0	365.0	0

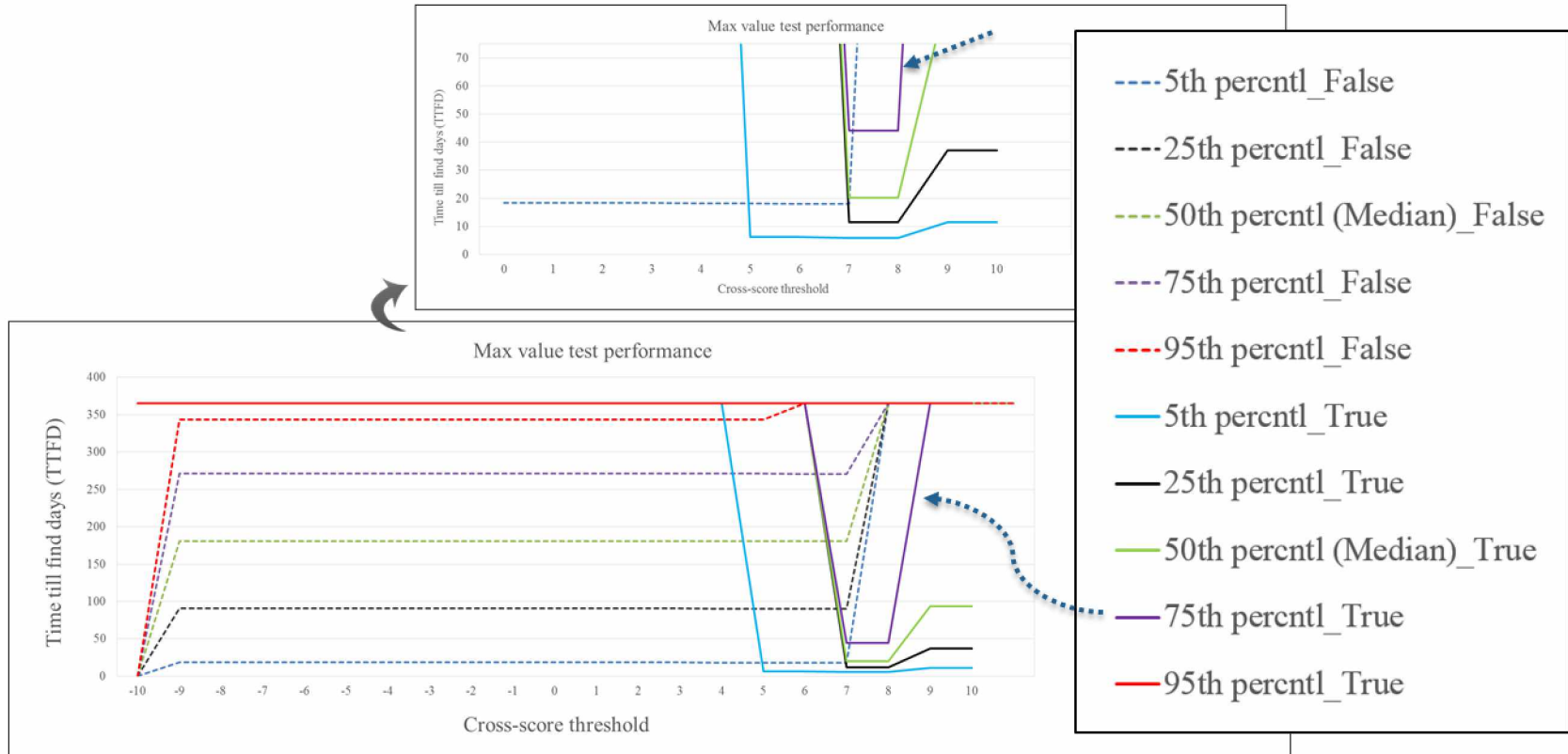


Figure 5.34: Max value test performance.

Table 5.17: Max value test True alarm performance.

Cross-score Threshold	5th percentl_True	25th percentl_True	50th percentl (Median)_True	75th percentl_True	95th percentl_True	Alarm (True)
-10	365.0	365.0	365.0	365.0	365.0	1
-9	365.0	365.0	365.0	365.0	365.0	1
-8	365.0	365.0	365.0	365.0	365.0	1
-7	365.0	365.0	365.0	365.0	365.0	1
-6	365.0	365.0	365.0	365.0	365.0	1
-5	365.0	365.0	365.0	365.0	365.0	1
-4	365.0	365.0	365.0	365.0	365.0	1
-3	365.0	365.0	365.0	365.0	365.0	1
-2	365.0	365.0	365.0	365.0	365.0	1
-1	365.0	365.0	365.0	365.0	365.0	1
0	365.0	365.0	365.0	365.0	365.0	1
1	365.0	365.0	365.0	365.0	365.0	1
2	365.0	365.0	365.0	365.0	365.0	1
3	365.0	365.0	365.0	365.0	365.0	1
4	365.0	365.0	365.0	365.0	365.0	1
5	6.2	365.0	365.0	365.0	365.0	1
6	6.2	365.0	365.0	365.0	365.0	1
7	5.9	11.5	20.3	44.2	365.0	1
8	5.9	11.5	20.3	44.2	365.0	1
9	11.4	37.0	93.4	365.0	365.0	1
10	11.4	37.0	93.4	365.0	365.0	1

Table 5.18: Max value test False alarm performance.

Cross-score Threshold	5th percentl_False	25th percentl_False	50th percentl (Median)_False	75th percentl_False	95th percentl_False	Alarm (False)
-10	18.3	90.5	180.8	271.0	343.2	0
-9	18.3	90.5	180.8	271.0	343.2	0
-8	18.3	90.5	180.8	271.0	343.2	0
-7	18.3	90.5	180.8	271.0	343.2	0
-6	18.3	90.5	180.8	271.0	343.2	0
-5	18.3	90.5	180.8	271.0	343.2	0
-4	18.3	90.5	180.8	271.0	343.2	0
-3	18.3	90.5	180.8	271.0	343.2	0
-2	18.3	90.5	180.8	271.0	343.2	0
-1	18.3	90.5	180.8	271.0	343.2	0
0	18.3	90.5	180.8	271.0	343.2	0
1	18.3	90.5	180.8	271.0	343.2	0
2	18.3	90.5	180.8	271.0	343.2	0
3	18.2	90.4	180.7	270.9	343.1	0
4	18.2	90.4	180.7	270.9	343.1	0
5	18.1	90.3	180.5	270.8	365.0	0
6	18.1	90.3	180.5	270.8	365.0	0
7	365.0	365.0	365.0	365.0	365.0	0
8	365.0	365.0	365.0	365.0	365.0	0
9	365.0	365.0	365.0	365.0	365.0	0
10	365.0	365.0	365.0	365.0	365.0	0

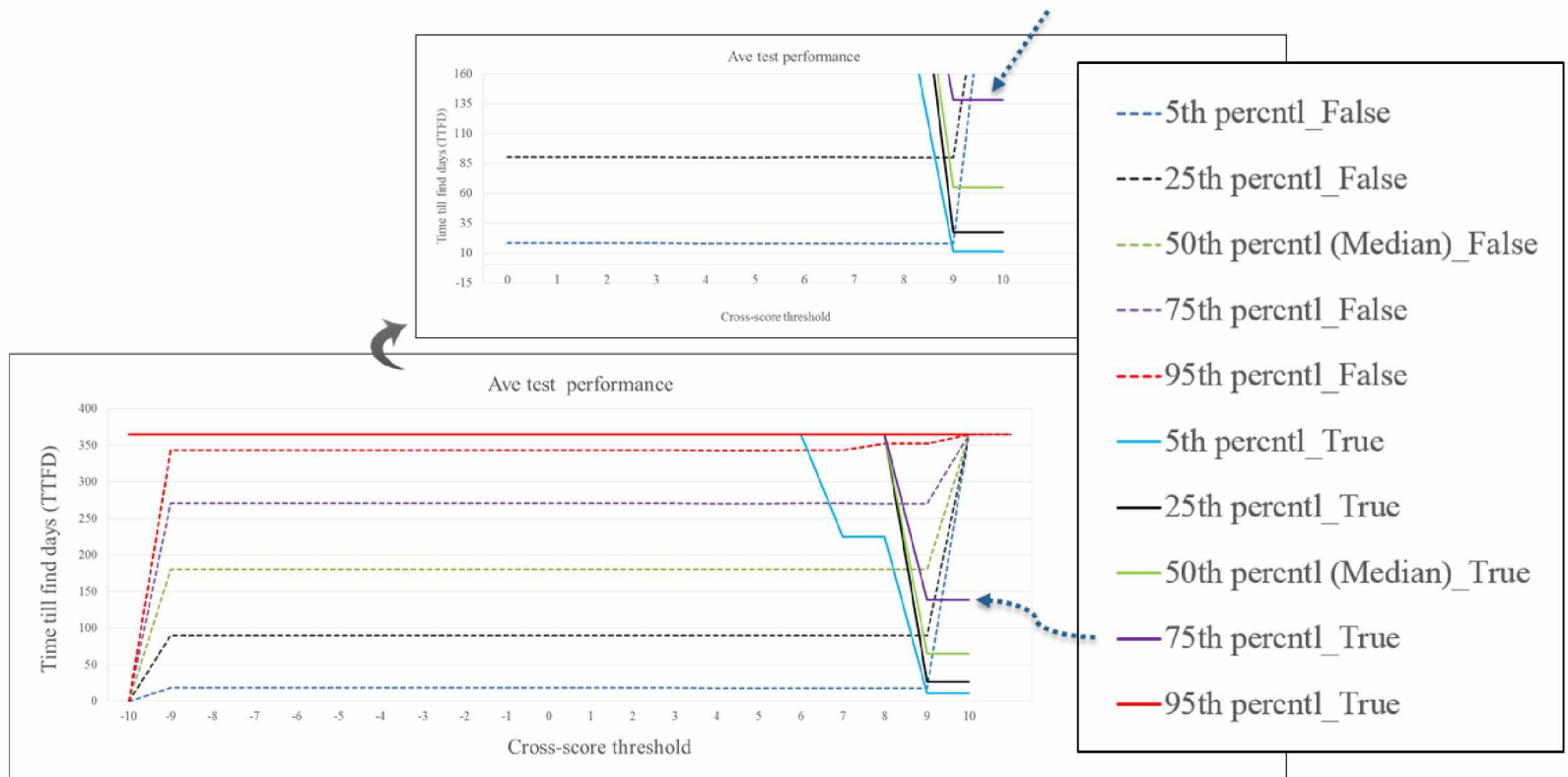


Figure 5.35: Ave value test performance.

Table 5.19: Ave value test True alarm performance.

Cross-score Threshold	5th percentl_True	25th percentl_True	50th percentl (Median)_True	75th percentl_True	95th percentl_True	Alarm (True)
-10	365.0	365.0	365.0	365.0	365.0	1
-9	365.0	365.0	365.0	365.0	365.0	1
-8	365.0	365.0	365.0	365.0	365.0	1
-7	365.0	365.0	365.0	365.0	365.0	1
-6	365.0	365.0	365.0	365.0	365.0	1
-5	365.0	365.0	365.0	365.0	365.0	1
-4	365.0	365.0	365.0	365.0	365.0	1
-3	365.0	365.0	365.0	365.0	365.0	1
-2	365.0	365.0	365.0	365.0	365.0	1
-1	365.0	365.0	365.0	365.0	365.0	1
0	365.0	365.0	365.0	365.0	365.0	1
1	365.0	365.0	365.0	365.0	365.0	1
2	365.0	365.0	365.0	365.0	365.0	1
3	365.0	365.0	365.0	365.0	365.0	1
4	365.0	365.0	365.0	365.0	365.0	1
5	365.0	365.0	365.0	365.0	365.0	1
6	365.0	365.0	365.0	365.0	365.0	1
7	224.9	365.0	365.0	365.0	365.0	1
8	224.9	365.0	365.0	365.0	365.0	1
9	11.2	27.2	65.2	138.4	365.0	1
10	11.2	27.2	65.2	138.4	365.0	1

Table 5.20: Ave value test False alarm performance.

Cross-score Threshold	5th percentl_False	25th percentl_False	50th percentl (Median)_False	75th percentl_False	95th percentl_False	Alarm (False)
-10	18.3	90.5	180.8	271.0	343.2	0
-9	18.3	90.5	180.8	271.0	343.2	0
-8	18.3	90.5	180.8	271.0	343.2	0
-7	18.3	90.5	180.8	271.0	343.2	0
-6	18.3	90.5	180.8	271.0	343.2	0
-5	18.3	90.5	180.8	271.0	343.2	0
-4	18.3	90.5	180.8	271.0	343.2	0
-3	18.3	90.5	180.8	271.0	343.2	0
-2	18.3	90.5	180.8	271.0	343.2	0
-1	18.3	90.5	180.8	271.0	343.2	0
0	18.3	90.5	180.8	271.0	343.2	0
1	18.3	90.5	180.8	271.0	343.2	0
2	18.3	90.5	180.8	271.0	343.2	0
3	17.9	90.1	180.4	270.6	342.8	0
4	17.9	90.1	180.4	270.6	342.8	0
5	18.2	90.4	180.7	270.9	343.1	0
6	18.2	90.4	180.7	270.9	343.1	0
7	17.9	90.1	180.3	270.6	352.5	0
8	17.9	90.1	180.3	270.6	352.5	0
9	365.0	365.0	365.0	365.0	365.0	0
10	365.0	365.0	365.0	365.0	365.0	0

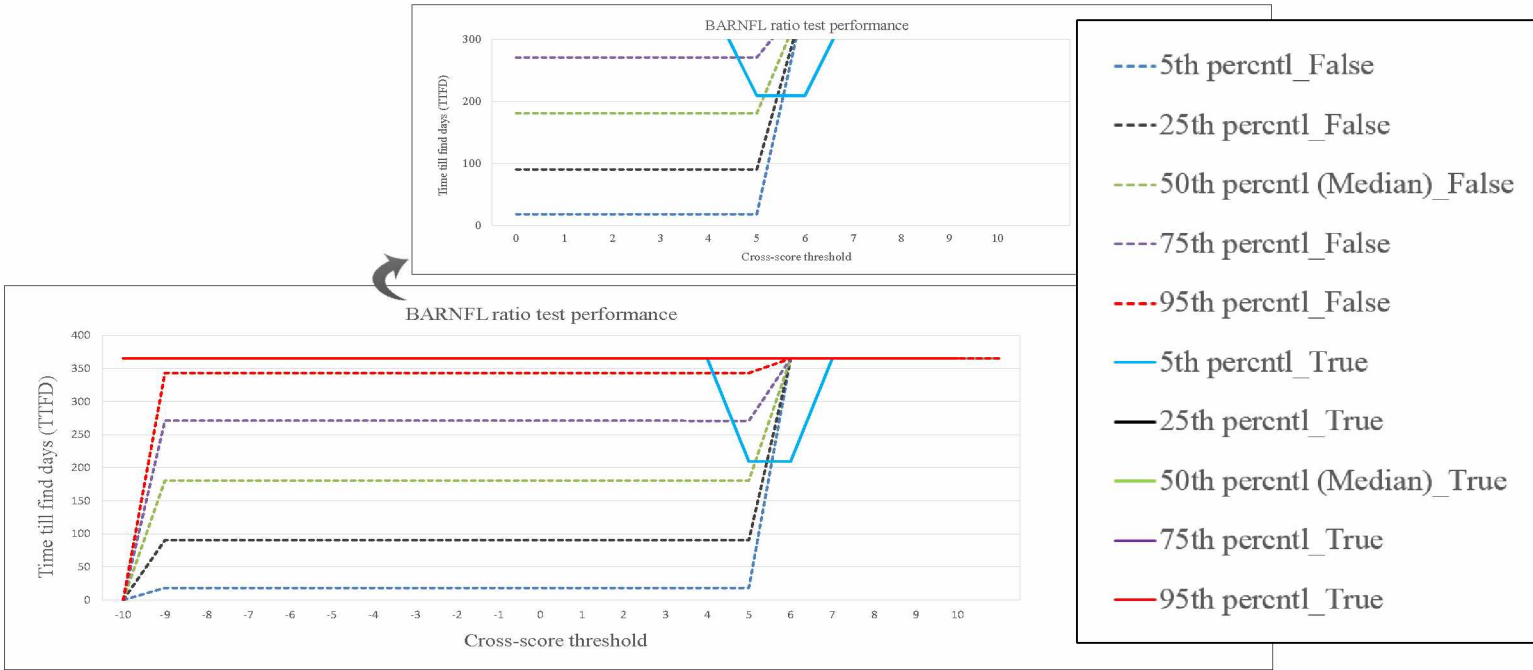


Figure 5.36: BARNFL ratio test performance.

Table 5.21: BARNFL ratio test True alarm performance.

Cross-score Threshold	5th percentl_True	25th percentl_True	50th percentl (Median)_True	75th percentl_True	95th percentl_True	Alarm (True)
-10	365.0	365.0	365.0	365.0	365.0	1
-9	365.0	365.0	365.0	365.0	365.0	1
-8	365.0	365.0	365.0	365.0	365.0	1
-7	365.0	365.0	365.0	365.0	365.0	1
-6	365.0	365.0	365.0	365.0	365.0	1
-5	365.0	365.0	365.0	365.0	365.0	1
-4	365.0	365.0	365.0	365.0	365.0	1
-3	365.0	365.0	365.0	365.0	365.0	1
-2	365.0	365.0	365.0	365.0	365.0	1
-1	365.0	365.0	365.0	365.0	365.0	1
0	365.0	365.0	365.0	365.0	365.0	1
1	365.0	365.0	365.0	365.0	365.0	1
2	365.0	365.0	365.0	365.0	365.0	1
3	365.0	365.0	365.0	365.0	365.0	1
4	365.0	365.0	365.0	365.0	365.0	1
5	209.3	365.0	365.0	365.0	365.0	1
6	209.3	365.0	365.0	365.0	365.0	1
7	365.0	365.0	365.0	365.0	365.0	1
8	365.0	365.0	365.0	365.0	365.0	1
9	365.0	365.0	365.0	365.0	365.0	1
10	365.0	365.0	365.0	365.0	365.0	1

Table 5.22: BARNFL ratio test False alarm performance.

Cross-score Threshold	5th percentl_False	25th percentl_False	50th percentl (Median)_False	75th percentl_False	95th percentl_False	Alarm (False)
-10	18.3	90.5	180.8	271.0	343.2	0
-9	18.3	90.5	180.8	271.0	343.2	0
-8	18.3	90.5	180.8	271.0	343.2	0
-7	18.3	90.5	180.8	271.0	343.2	0
-6	18.3	90.5	180.8	271.0	343.2	0
-5	18.3	90.5	180.8	271.0	343.2	0
-4	18.3	90.5	180.8	271.0	343.2	0
-3	18.3	90.5	180.8	271.0	343.2	0
-2	18.3	90.5	180.8	271.0	343.2	0
-1	18.3	90.5	180.8	271.0	343.2	0
0	18.3	90.5	180.8	271.0	343.2	0
1	18.3	90.5	180.8	271.0	343.2	0
2	18.3	90.5	180.8	271.0	343.2	0
3	18.2	90.4	180.7	270.9	343.1	0
4	18.2	90.4	180.7	270.9	343.1	0
5	365.0	365.0	365.0	365.0	365.0	0
6	365.0	365.0	365.0	365.0	365.0	0
7	365.0	365.0	365.0	365.0	365.0	0
8	365.0	365.0	365.0	365.0	365.0	0
9	365.0	365.0	365.0	365.0	365.0	0
10	365.0	365.0	365.0	365.0	365.0	0

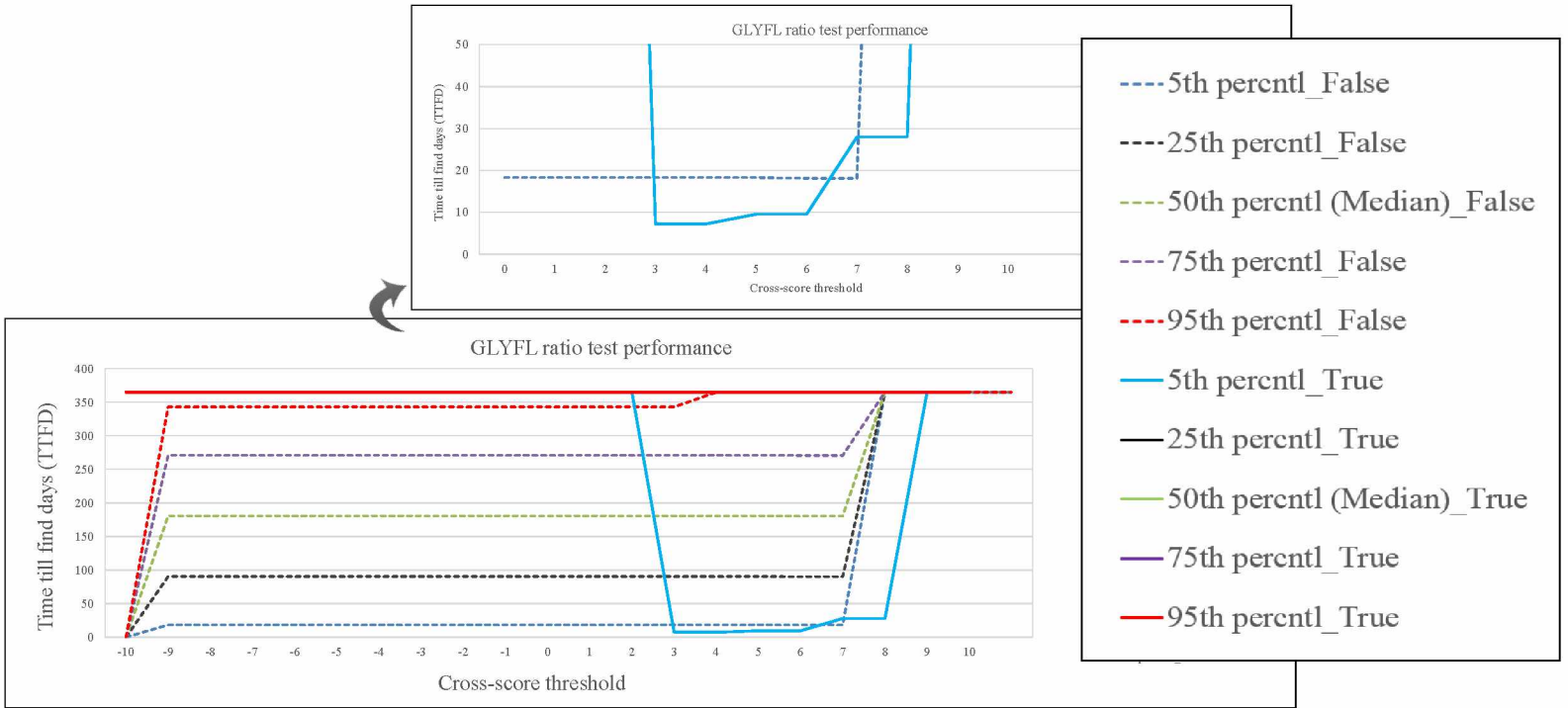


Figure 5.37: GLYFL ratio test performance.

Table 5.23: GLYFL ratio test True alarm performance.

Cross-score Threshold	5th percentl_True	25th percentl_True	50th percentl (Median)_True	75th percentl_True	95th percentl_True	Alarm (True)
-10	365.0	365.0	365.0	365.0	365.0	1
-9	365.0	365.0	365.0	365.0	365.0	1
-8	365.0	365.0	365.0	365.0	365.0	1
-7	365.0	365.0	365.0	365.0	365.0	1
-6	365.0	365.0	365.0	365.0	365.0	1
-5	365.0	365.0	365.0	365.0	365.0	1
-4	365.0	365.0	365.0	365.0	365.0	1
-3	365.0	365.0	365.0	365.0	365.0	1
-2	365.0	365.0	365.0	365.0	365.0	1
-1	365.0	365.0	365.0	365.0	365.0	1
0	365.0	365.0	365.0	365.0	365.0	1
1	365.0	365.0	365.0	365.0	365.0	1
2	365.0	365.0	365.0	365.0	365.0	1
3	7.2	365.0	365.0	365.0	365.0	1
4	7.2	365.0	365.0	365.0	365.0	1
5	9.6	365.0	365.0	365.0	365.0	1
6	9.6	365.0	365.0	365.0	365.0	1
7	28.0	365.0	365.0	365.0	365.0	1
8	28.0	365.0	365.0	365.0	365.0	1
9	365.0	365.0	365.0	365.0	365.0	1
10	365.0	365.0	365.0	365.0	365.0	1

Table 5.24: GLYFL ratio test False alarm performance.

Cross-score Threshold	5th percentl_False	25th percentl_False	50th percentl (Median)_False	75th percentl_False	95th percentl_False	Alarm (False)
-10	18.3	90.5	180.8	271.0	343.2	0
-9	18.3	90.5	180.8	271.0	343.2	0
-8	18.3	90.5	180.8	271.0	343.2	0
-7	18.3	90.5	180.8	271.0	343.2	0
-6	18.3	90.5	180.8	271.0	343.2	0
-5	18.3	90.5	180.8	271.0	343.2	0
-4	18.3	90.5	180.8	271.0	343.2	0
-3	18.3	90.5	180.8	271.0	343.2	0
-2	18.3	90.5	180.8	271.0	343.2	0
-1	18.3	90.5	180.8	271.0	343.2	0
0	18.3	90.5	180.8	271.0	343.2	0
1	18.3	90.5	180.8	271.0	343.2	0
2	18.3	90.5	180.8	271.0	343.2	0
3	18.3	90.5	180.7	271.0	365.0	0
4	18.3	90.5	180.7	271.0	365.0	0
5	18.1	90.3	180.6	270.8	365.0	0
6	18.1	90.3	180.6	270.8	365.0	0
7	365.0	365.0	365.0	365.0	365.0	0
8	365.0	365.0	365.0	365.0	365.0	0
9	365.0	365.0	365.0	365.0	365.0	0
10	365.0	365.0	365.0	365.0	365.0	0