

## ABSTRACT

Title of Document: MONITORING AND FORECASTING  
BIOSOLIDS ODORS AT PRIOR LIME AND  
POST LIME ADDITION IN WASTEWATER  
TREATMENT PLANTS

Ampun Janpengpen, Doctor of Philosophy, 2010

Directed By: Professor Gregory B. Baecher  
Department of Civil and Environmental  
Engineering

A real-time solids odor monitoring system provides an odor management feedback tool for both process control and a biosolids management program. Since higher odor levels means higher process costs, as well as greater potential for nuisance odors at land application sites, identifying the processes that contribute to these elevated levels is critical to responsible, economical, and efficient wastewater plant management and biosolids land application programs.

A real-time system is currently being utilized to monitor DC Water's 370 mgd plant in Washington, D.C. Each year, DC Water applies biosolids to over 20,000 acres of agricultural land. Nuisance odors from recycling biosolids on land may drift into surrounding neighborhoods and motivate neighboring communities to enact legislation to ban land application. Therefore, the reduction of odor emissions from biosolids recycled on field sites is a major concern.

Odors levels generated by dewatered solids and limed biosolids are measured by headspace monitoring devices in enclosed conveyance systems. Both total reduced sulfur compounds (TRS) and nitrogen (N)-containing compounds are measured with online electro-chemical sensors. The system correlates odorant levels of dewatered solids and biosolids and utilizes treatment process scenarios and various operational parameters throughout the wastewater treatment process. This study uses ordinary least squares (OLS) estimation and instrumental variable (IV) estimation to create explanatory and predictive models. Furthermore, cross-validation analyses are employed to validate both explanatory and predictive models.

Data analyses suggest that waste-activated percent solids (WAS %S) and dissolved-air flotation total solids (DAF TS) can contribute to mitigating TRS. However, all process variables at secondary sedimentation, which are gravity thickening percent solids (GT %S), gravity total solids (GT TS), and blend ratio, can contribute to increase TRS. The IV estimation indicates that % lime feeding, # centrifuges, cake percent solids (Cake %S), temperature at secondary effluent, and ambient temperature cannot directly explain TRS post-lime, but they do explain TRS levels via post-lime temperature. Additionally, cationic polymer at the secondary and dewatering process coupled with post lime temperature can contribute to increase N-containing compounds at the lime addition process. The accumulated cationic polymer inside the sludge of secondary sedimentation can also contribute to high N-containing compounds at the downstream.

MONITORING AND FORECASTING BIOSOLIDS ODORS AT PRIOR LIME  
AND POST LIME ADDITION IN WASTEWATER TREATMENT PLANTS

By

Ampun Janpengpen

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
[2010]

Advisory Committee:

Professor Gregory B. Baecher, Chair/Advisor

Associate Professor Galit Shmueli, Dean's Representative

Professor Richard H. McCuen

Associate Professor Steven A. Gabriel

Associate Professor Kaye L. Brubaker

© Copyright by  
Ampun Janpengpen  
[2010]

## Dedication

I would like to dedicate this dissertation to my wife, Chonlada Janpengpen, my son, Athithat Janpengpen, and my parents.

## Acknowledgements

Firstly, I would like to express my sincere gratitude to Professor Gregory B. Baecher, my advisor, for his suggestions and advice thorough the development of this dissertation and my entire studies at UMD, as well as his valuable instruction and comments, without a shadow of a doubt.

I wish to express my sincere thanks to Mark Ramirez, my supervisor at DC Water, who has directed my work since the first day I became as internship student, for his contribution and criticisms in completing this dissertation. I would like to thank Chris Peot, my manager, who gave me the opportunity to work at DC Water and for his encouragement to complete this dissertation. I would also like to thank each and every one of my co-workers at the Department of Wastewater Treatment for their supports.

I would like to thank Associate Professor Galit Shmueli, Professor Richard H. McCuen, Associate Professor Steven A. Gabriel, and Associate Professor Kaye L. Brubaker for serving on my dissertation committee. Their suggestions and priceless comments improved the quality of dissertation.

I would like to thank Royal Thai Army for giving me a great opportunity to study for master and doctoral degrees in the US.

Lastly, I would like to thank all of my Thai colleagues at UMD and Dr. Ravissa Suchato for their support in the last six years.

# Table of Contents

Dedication .....	ii
Acknowledgements .....	iii
Table of Contents .....	iv
List of Tables .....	vii
List of Figures .....	x
Chapter 1: Introduction .....	1
1.1 Odors in Biosolids .....	1
1.2 Research Objectives .....	3
1.3 Previous Studies and How This Study Differs .....	4
1.4 Research Hypotheses .....	7
1.5 Research Contributions .....	8
1.6 Executive Summary .....	9
1.7 Master Plan Format .....	10
Chapter 2: Background and Literature Review .....	12
2.1 Odorous Production in Lime-Stabilized Biosolids .....	12
2.1.1 Total Reduced Sulfur Compounds .....	13
2.1.2 Nitrogen (N)-Containing Compounds .....	14
2.2 Odor Measurement, Determination, and Reduction in Biosolids .....	15
2.3 Ordinary Least Squares (OLS) and Instrumental Variable (IV) Estimation .....	17
2.3.1 Ordinary Least Squares (OLS) Estimation .....	18
2.3.2 Instrumental Variable (IV) and Two-Stage Least Squares (2SLS) Estimation .....	22
Chapter 3: Operational Processes and Data Description .....	28
3.1 Site Overview .....	28
3.2 Operational Processes Data (Upstream Process Variables) .....	32
3.2.1 Ferric Chloride ( $\text{FeCl}_3$ ) Addition .....	33
3.2.2 Waste Pickle Liquid (WPL) Addition .....	33
3.2.3 Polymer Addition .....	34
3.2.4 Gravity Thickening (GT) at Primary Settling .....	35
3.2.5 Secondary Sludge Blanket Depth .....	35
3.2.6 Secondary Waste Activated Sludge .....	37
3.2.7 Dissolved Air Flotation (DAF) .....	37
3.2.8 Initial Settling Velocity (ISV) .....	38
3.2.9 Settle Sludge Volume ( $\text{SSV}_{60}$ ) .....	38
3.2.10 Concentration of Return Activated Sludge (RAS) .....	38
3.2.11 Blend Ratio .....	39
3.2.12 Temperature of secondary effluent .....	40
3.2.13 Cake Solids .....	40
3.2.14 Number of Centrifuges in Services .....	41
3.2.15 Total Flow of Dewatered Sludge .....	42
3.2.16 Percent of Lime Feeder .....	42
3.2.17 Temperature of Limed Biosolids (Post limed biosolids temperature) .....	43

3.2.18 Mixed Liquor Suspended Solids (MLSS).....	44
3.2.19 Ambient Temperature .....	44
3.3 Odor Determination and Experimental Design.....	44
3.3.1 The Importance of Odor Sample Collection.....	45
3.3.2 Real-Time Solids Odor Monitoring Systems.....	45
3.3.3 Design and Odor Monitoring Location.....	53
3.3.4 Odors Data Collection of Dewatered Solids and Biosolids .....	56
Chapter 4: Statistical Data Analysis for Biosolids Odors .....	58
4.1 Defining Modeling Purposes .....	58
4.1.1 Model Requirements.....	59
4.2 Statistical Modeling Strategy.....	65
4.2.1 Task I: Exploratory Data Analysis (Data Preparation) (Steps 1 to 3).....	65
4.2.2 Task II: Ordinary Least Squares (OLS) Estimation (Steps 4 to 5) .....	66
4.2.3 Task III: Instrumental Variable (IV) Estimation (Steps 6 to 7) .....	67
4.2.4 Task IV: Model Validation and Selection (Steps 8 to 9) .....	67
4.3 Exploratory Data Analysis (EDA) (Data Preparation) .....	71
4.3.1 Data Summaries .....	72
4.3.2 Data Visualization.....	76
4.3.3 Correlation Analysis .....	78
4.3.4 Summary Results of Exploratory Data Analysis .....	85
4.4 Additional Techniques .....	92
4.4.1 Dummy Variables .....	92
4.4.2 Interaction Variables.....	93
Chapter 5: Data Analysis, Results, and Discussion .....	95
5.1 Modeling Analysis.....	95
5.2 Explanatory Models.....	99
5.2.1 Explanatory Model for TRS at Prior-Lime.....	99
5.2.2 Explanatory Models for TRS at Post-Lime .....	115
5.2.3 Explanatory Models for N-containing Compounds.....	134
5.3 Predictive Models .....	147
5.3.1 Predictive Models for TRS at Prior-Lime ( $Y_1$ ) .....	148
5.3.2 Predictive Models for TRS at Post-lime ( $Y_2$ ).....	151
5.3.3 Predictive Models for N-containing Compounds ( $Y_3$ ).....	153
5.4 Model Validation .....	155
5.4.1 Validation of Explanatory Models of TRS Prior-Lime .....	158
5.4.2 Validation of Explanatory Models of TRS Post-Lime .....	159
5.4.3 Validation of Explanatory Models of N-containing Compounds.....	161
5.4.4 Validation of Predictive Models of TRS Prior-Lime.....	163
5.4.5 Validation of Predictive Models of TRS Post-Lime.....	164
5.4.6 Validation of Predictive Models of N-containing Compounds .....	165
5.5 Model Selection .....	166
5.6 Special Cases of Operations.....	170
5.6.1 Odor Incidents When Dewatering Processes having Problems.....	172
5.6.2 Post Lime Temperature Incidents .....	177
5.6.3 Implication of Special Cases of Operations.....	179
Chapter 6: Conclusions and Future Work.....	180



6.1Conclusions.....	180
6.2 Explanatory Models for DC Water .....	184
6.3 Predictive Models for Dissertation and DC Water .....	189
6.4 DC Water Standard Operating Procedure (SOP).....	192
6.5 Future Work .....	194
Appendix A .....	196
Appendix B .....	250
Appendix C .....	254
Appendix D .....	276
Reference .....	283

## List of Tables

Table 2.1: The range of odor thresholds of odorous production in limed biosolids.....	13
Table 2.2: Regression equations for odorant compounds.....	21
Table 4.1 Variables description.....	72
Table 4.2 Summary statistics.....	74
Table 4.3: Significant correlation coefficients between TRS prior-lime (ppm) and upstream processes.....	81
Table 4.4: Significant correlation coefficients between TRS post-lime (ppm) and upstream processes.....	81
Table 4.5: Significant correlation coefficients between N-containing compounds (ppm) and upstream processes.....	82
Table 4.6: The essential process variables relating to TRS prior-lime (ppm) during annual .....	86
Table 4.7: The essential process variables relating to TRS post-lime (ppm) during annual .....	86
Table 4.8: The essential process variables relating to N-containing compounds (ppm) during annual.....	87
Table 4.9: The essential process variables relating to TRS prior-lime (ppm) during winter.....	87
Table 4.10: The essential process variables relating to TRS post-lime (ppm) during winter.....	88

Table 4.11: The essential process variables relating to N-containing compounds (ppm) during winter.....	88
Table 4.12: The essential process variables relating to TRS prior-lime (ppm) during summer.....	89
Table 4.13: The essential process variables relating to TRS post-lime (ppm) during summer.....	90
Table 4.14: The essential process variables relating to N-containing compounds (ppm) during summer.....	91
Table 5.1: The anticipated significant upstream process variables.....	96
Table 5.2: Statistical model for TRS prior-lime (ppm) during annual period .....	100
Table 5.3: Statistical model for TRS prior-lime (ppm) during the winter period.....	107
Table 5.4: Statistical model for TRS prior-lime (ppm) during the summer period...	109
Table 5.5: Statistical model for TRS post-lime (ppm) during the annual period.....	117
Table 5.6: Statistical model for post lime temperature (°F) to create IV.....	119
Table 5.7: Hausman test for endogeneity of post lime temperature (°F) .....	121
Table 5.8: Explanatory model for TRS post-lime (ppm) with IV during annual period.....	122
Table 5.9: Explanatory model for TRS post-lime (ppm) w/o IV during winter period.....	126
Table 5.10: Explanatory model for TRS post-lime (ppm) w/ IV during winter period.....	127
Table 5.11: Explanatory model for TRS post-lime (ppm) during summer period....	131
Table 5.12: Explanatory model for N-containing (ppm) during annual period.....	135

Table 5.13: Explanatory model for N-containing (ppm) during winter period.....	141
Table 5.14: Explanatory model for N-containing (ppm) during summer period.....	144
Table 5.15: Predictive model for TRS prior-lime (ppm) during annual period.....	148
Table 5.16: Predictive model for TRS prior-lime (ppm) during winter period.....	149
Table 5.17: Predictive model for TRS prior-lime (ppm) during summer period.....	150
Table 5.18: Predictive model for TRS post-lime (ppm) during annual period.....	151
Table 5.19: Predictive model for TRS post-lime (ppm) during winter period.....	152
Table 5.20: Predictive model for TRS post-lime (ppm) during summer period.....	152
Table 5.21: Predictive model for N-containing (ppm) during annual period.....	153
Table 5.22: Predictive model for N-containing (ppm) during winter period.....	154
Table 5.23: Predictive model for N-containing (ppm) during summer period.....	154
Table 5.24: Summary of validation results.....	157

## List of Figures

Figure 1.1: Overview of research approaches.....	7
Figure 2.1: A schematic of solids handling processes at DC Water.....	18
Figure 3.1: DC Water Processes Diagram.....	31
Figure 3.2: DC Water solids handling processes diagram.....	32
Figure 3.3: Low Range Sampling System-2 (LRSS-2) and Polytron 7000.....	47
Figure 3.4 Schematic of current loops in Octan-R.....	51
Figure 3.5: Odorants concentration output from the Master Monitor.....	53
Figure 3.6: Real-Time solids odors monitoring systems.....	54
Figure 3.7: Location and devices setup inside solids processing building (SPB).....	55
Figure 3.8: Total reduced sulfur (TRS) (ppm) at prior-lime process.....	56
Figure 3.9: Total reduced sulfur (TRS) (ppm) at post-lime process.....	56
Figure 3.10: Nitrogen (N)-containing compounds (ppm).....	57
Figure 4.1: <i>K-fold</i> cross-validation.....	68
Figure 4.2: Statistical steps in data analysis and modeling process.....	71
Figure 4.3: Gravity total solids (GT TS) (lbs/day) and dissolved air flotation total solids (DAF TS) (lbs/day).....	77
Figure 4.4: Secondary blanket level SE.BL (ft), SWO.BL (ft), and SWE.BL (ft).....	77
Figure 4.5: TRS Prior-lime (ppm) VS SWE.BL (ft) During Summer Period.....	83
Figure 4.6: TRS Prior-Lime (ppm)VS Temp Sec Eff (°F) During Annual Period ....	84
Figure 5.1: Temp Sec Eff (°F) vs. Ambient Temp (°F).....	98
Figure 5.2: SWO.BL*MLSS.SWO (ft*mg/L).....	104
Figure 5.3: TRS prior-lime (ppm) vs. Temp Sec Eff (°F).....	104

Figure 5.4: Relationship between TRS prior-lime (ppm) and Temp Sec Eff (°F)....	105
Figure 5.5: DAF TS (lbs/day) during the summer.....	110
Figure 5.6: Sources of Cake percent solids (Cake %S).....	111
Figure 5.7: Amount of GT TS (lbs/day), DAF TS (lbs/day), and Blend Ratio (Pri:Sec) during summer period.....	112
Figure 5.8: Secondary blanket level (ft) during the summer period.....	114
Figure 5.9: Relationship between odor at prior-lime and post-lime process.....	116
Figure 5.10: Relationship between TRS post-lime (ppm) and post lime temp (°F)..	128
Figure 5.11: Relationship between post lime temp (°F) and TRS post-lime (ppm)..	129
Figure 5.12: Relationship between TRS post-lime (ppm) and Amb Temp (°F).....	130
Figure 5.13: Relationship between TRS post-lime (ppm) and Temp Sec Eff (°F)...	130
Figure 5.14: Relationship between N-containing (ppm) and Pol_Sec (lbs/day).....	137
Figure 5.15: Relationship between Pol_Sec (lbs/day) and WAS %S (%).....	138
Figure 5.16: Amount of Pol_DW (lbs/DTS) added in annual period.....	139
Figure 5.17: Relationship between N-containing (ppm) VS post lime temp (°F).....	141
Figure 5.18: Relationship between Pol_Sec (lbs/day) and RAS.SE (mg/L).....	143
Figure 5.19: Relationship between Pol_Sec (lbs/day) and RAS.SWO (mg/L) during summer period.....	146
Figure 5.20: Dewatering and lime stabilization process inside SPB.....	171
Figure 5.21: Odors incidents during dewatering problems.....	172
Figure 5.22: Post lime incidents.....	177

# Chapter 1: Introduction

This research focuses on total reduced sulfur (TRS) and nitrogen (N)-containing compounds, two of the main odorous components released from biosolids at wastewater treatment plants (WWTP). It is therefore significant to have a basic understanding about the sources of those compounds. This chapter provides an overview of odors in biosolids of the District of Columbia Water and Sewer Authority's (DC Water), research objectives, previous studies and differences, and research hypotheses. In addition, research contributions and a master plan format are stated.

## 1.1 Odors in Biosolids

Nuisance odors are considered on public issues associated with wastewater treatment plants (WWTP) and from recycling biosolids on land that most affect humans and cause the limitation of solids disposal options, especially for the plant employing a beneficial land application (Dague 1972; Hwang et al. 1995; USEPA 2000a; Murthy et al. 2002b; Turkmen et al. 2004; Visan 2003). Land application is widely used and is considered to be one of the most effective methods for biosolids disposal in the United States (USEPA 2000c; Chen et al. 2005). A report of the U.S. Environmental Protection Agency (USEPA) indicated that in the United States, the demand for biosolids land applications will increase from 66% in 2005 to 70% in 2010 (USEPA 1999; Oleszkiewicz and Mavinic 2002).

DC Water generates "Class B" biosolids by utilizing lime stabilization and relies on land applications for biosolids disposal (Murthy et al. 2001). Therefore,

these biosolids can cause public concerns for the acceptability of their quality. One of the biggest considerations is offensive odor. Odorous emissions at the land application sites indicated the operating issues of biosolids producers that can cause complaints and public opposition from the residents around the application sites. As a result, treatment facilities can be banned or shut down. Nonetheless, it is the duty of the biosolids producers to improve their quality by controlling and mitigating odors at the application fields (USEPA 2000a). For this reason, DC Water, through the Department of Wastewater Treatment (DWT) makes strong efforts to accomplish this objective. The best and most effective way is to minimize odors at the source (in-plant) and handle solid waste properly.

Odor mitigation in biosolids at the source can diminish odor emission from biosolids distributed for land application (Kim et al. 2002). At each step of operational treatment, in-plant processes can contribute or promote odor levels in biosolids, especially at the solids-handling process (Hwang et al. 1995; Kim et al. 2001). Both organic and inorganic compounds such as total reduced sulfur (TRS), nitrogen (N)-containing compounds (ammonia and amines), and organic fatty acids are considered the major odorous compounds emitted from biosolids, which are normally produced by heat, aeration and digestion (Dague 1972; Hwang et al. 1995; USEPA 2000a; Kim et al. 2001).

The DWT tries to maintain and improve the quality of biosolids to meet exceed the requirements of the United States Environmental Protection Agency (U.S.EPA). However, there are some complaints of odors arising from communities around land application fields. Therefore, DWT conducts a leading-edge research



program focusing on odor control in order to get support from communities. It is vital to identify and control the factors involved in odor production of limed biosolids at the source (in-plant) before transportation to land application sites. By employing real-time solids odor monitoring systems and using empirical analysis via statistical modeling, this research can identify and explain the causal of odorants production in lime-stabilized biosolids. In addition, these systems can forecast the odorants level before sending to the application sites.

### 1.2 Research Objectives

There are two overall goals of this research: to improve the treatment process by mitigating odor emitted from lime biosolids, and to develop a methodology for explaining the causes of odorants production in lime-stabilized biosolids. To achieve those goals, the following objectives are defined:

1. To quantify total reduced sulfur (TRS) concentration and nitrogen (N)-containing compounds concentration emitting from prior-lime and post-lime processes in real-time.
2. To identify the relationship between the odorants at the prior-lime and post-lime processes.
3. To determine and explain what operational process variables and conditions (in-plant) contribute to TRS and N-containing compounds.
4. To investigate and determine all plausible scenarios causing high odorants in biosolids at solids treatment processes with a real-time monitoring system.
5. To establish guidelines and strategies for practical biosolids management.

### 1.3 Previous Studies and How This Study Differs

In the last two decades, statistical modeling has been used in the environmental area including water and wastewater treatment. For instance, statistical modeling has been used: to evaluate water quality in a free flowing river (Greenberg et al. 1973); to forecast algae biomass in Lake Okeechobee by using ordinary least squares (OLS), a generalized additive model, and a regression spline model (Lamon 1995; Lamon et al. 1996; Lamon and Clyde 2000); to predict biosolids odor emitting from the wastewater treatment process by using multiple linear regression (MLR) (Vilalai 2003; Gabriel et al. 2006); to forecast the emission of dimethyl disulfide (DMDS) using MLR (Gabriel et al. 2005); to predict and connect biosolids odors between onsite and offsite by using MLR, logistic regression and discriminant analysis (Vilalai 2008).

In the field of wastewater treatment and biosolids management, especially in biosolids odor emission, there are a few studies that have employed statistical modeling to find the solutions for mitigating odorous production in biosolids and explaining the sources of odorants compounds. To the best of our knowledge, there are five studies, Vilalai (2003), Gabriel et al. (2005), Gabriel et al. (2006), Vilalai (2008), and Sekyiamah and Kim (2009).

In their studies, Vilalai (2003), Gabriel et al. (2005), and Gabriel et al. (2006) developed a model by using multiple regression analysis (MLR) to forecast and explain biosolids odor. The MLR of Vilalai (2003) and Gabriel et al. (2006) regressed the score of odors in the field (dependent variable) on the set of wastewater treatment process variables (independent variables). Both studies have found that the sludge

blanket depth, the amount of lime additions, the amount of polymer additions at dewatering and dissolve air floatation (DAF) process, and the blend ratio are statistically significant and can be included in the predicted model.

However, Gabriel et al. (2005) had used dimethyl disulfide (DMDS) as a dependent variable and regressed it on a set of sludge characteristics variables and a set of process parameters variables. The model demonstrated that oxidation-reduction potential (ORP), the blend ratio, and the number of centrifuges in services were the critical variables promoting the emission of DMDS. Sekyiamah and Kim (2009) also employed MLR but they used volatile sulfur compounds (VSCs) as the dependent variable and regressed it on a set of process variables that were measured at the primary process and at secondary process to identify the formation of VSCs. Their model indicated that food to microorganism ratio (F/M ratio), sludge blanket depth, and settle sludge volume ( $SSV_{60}$ ) are statistically significant factors in modeling VSCs.

The study of Vilalai (2008) attempted to develop models for forecasting biosolids odor at wastewater treatment plants and at the land application sites by combining two odor measurements, an analytical measurement and a sensory measurement. The overall goal of this study is to identify the relationship between biosolids odor levels at the treatment plant evaluated by the headspace approach with a portable hydrogen sulfide ( $H_2S$ ) analyzer, and an odor panel for sensory analysis and a biosolids odor level at application sites evaluated by an olfactometer. This study used MLR to develop a biosolids odor prediction model at wastewater treatment

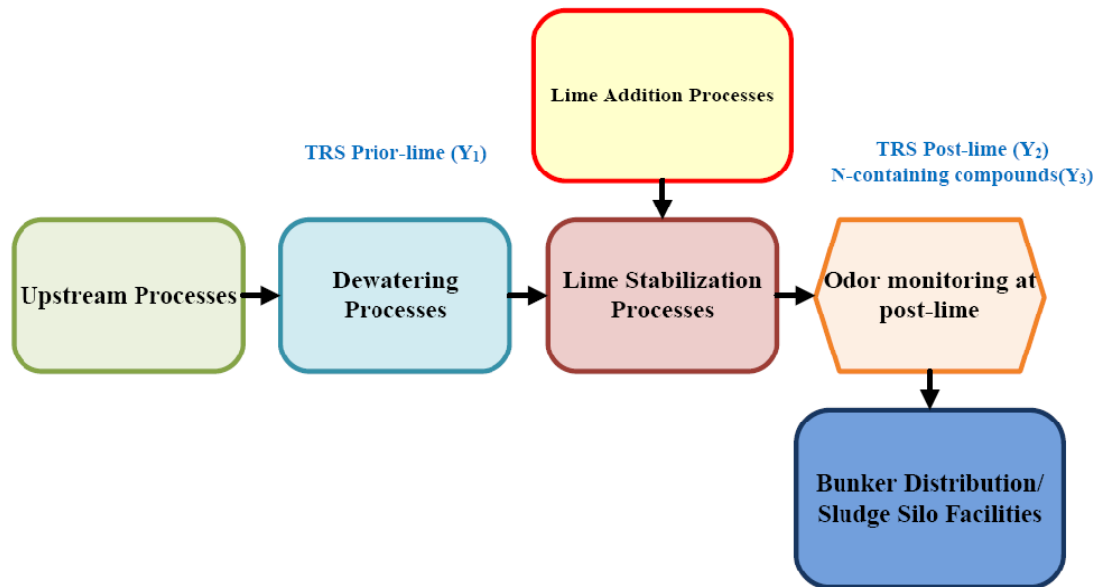
plants and used categorical analysis techniques, logistic regression, and discriminant analysis to develop models at the application site.

Additionally, all five studies collected solids samples from each step of the treatment processes or simulated the treatment processes to get the solids samples for odor data collection used in the data analysis.

The present research differs from the previous five studies in a few aspects. First, our study employs a real-time solids odor monitoring system to detect odorants directly from an actual WWTP. Second, this study uses ordinary least squares (OLS) estimation, instrumental variable (IV), and two-stages least squares (2SLS) estimation to explain the causes or sources of total reduced sulfur (TRS) and N-containing compounds, as well as to forecast those two odorants compounds at prior-lime and post-lime addition. When we employ OLS estimation in the MLR model with TRS at post-lime addition as dependent variable and all upstream processes as independent variables, some of the process variables having significant contribution to this odorous compound cannot be included in the model. Instead, they are included in the residual term, thereby losing the ability to directly quantify their effect on TRS at post-lime. To overcome this issue, we use a common method in economics called IV and 2SLS are used to identify the process variables having an indirect relationship with TRS at post-lime addition and contributing to the emissions.

The IV method can also improve the efficiency of the model by increasing the overall statistical significance and decreasing the standard error of estimates (SE). Additionally, this research focuses on explanatory modeling and predictive modeling for biosolids odor at the treatment plant, where it is believed to be the effective

solution for mitigating odor at application sites. Figure 1.1 illustrates an overview of our research approaches.



**Figure 1.1: Overview of research approaches**

#### 1.4 Research Hypotheses

From this research, there are four hypotheses to be tested:

**Hypothesis 1:** *The total reduced sulfur (TRS) at the prior-lime process positively correlates with TRS at the post-lime process.*

**Hypothesis 2:** *The causes of TRS at the post-lime process can be explained via the relation between upstream processes and the total reduced sulfur (TRS) at the prior-lime process.*

**Hypothesis 3:** *High temperature of the secondary effluent and ambient temperature can contribute to the emission of TRS and N-containing compounds.*

**Hypothesis 4:** *Dewatered solids with high percent solids content can raise the temperature of limed biosolids and promote the emission of both N-containing compounds and TRS concentration during the lime stabilization.*

### 1.5 Research Contributions

All the results from the data analysis and discussion will provide noteworthy information for DC water, and especially DWT, to improve the biosolids management program and to improve biosolids quality in terms of odor mitigation. DWT gains better understanding as to what factors in wastewater treatment processes relate to odorous emission in limed biosolids. This research, therefore, provides five contributions:

1. To provide an odor management real-time feedback tool for a process control and biosolids management program. This is the first and the largest contribution, and can be applied to any wastewater treatment plant.
2. To identify and explain what operational processes potentially promote odor production in limed biosolids.
3. To explain the causes or sources of TRS and N-containing compounds emitting from lime biosolids.
4. To suggest how to improve the wastewater treatment process in order to mitigate TRS and N-containing compounds.
5. To achieve cost savings by reducing: polymer usage, post-lime usage, quantities of biosolids generated, hauling and land application costs, and increasing the biosolids land application program benefits by the reduction of nuisance odor biosolids emissions to neighbors of land application sites.

### 1.6 Executive Summary

Odors are by far the biggest impediment to public acceptance of biosolids recycling. No matter how much value there is to a farmer, or how much carbon is sequestered, or energy saved, if a product has a malodor, the public will not welcome it into their community. The sulfur odors emitted by some biosolids are designed to warn to humans of potential pathogens. This is genetically hard wired, and difficult to overcome. When DC Water treat the biosolids, the pathogens are reduced to an acceptable safe level, but often the odor trigger is still there.

The paradigm in their profession is often that biosolids smell and that there is little they can do about it. This dissertation has found is that there are specific things DC Water can do at the plant to limit the odors, and they can even predict when odors might be generated based on a combination of process parameters. The ability to improve product quality (odors mitigation) and predict when they may occur will help reduce the nuisance complaints and therefore will increase the acceptance of biosolids. In addition, real-time solids odor monitoring system has been tested and installed equipment that serves as an early warning system for sulfur and ammonia based compounds.

This system sounds a warning and calls several individuals when an odor spike occurs. This warning allows DC Water to contact the field managers to determine if they are in a suitable location, where an odorous material will not elicit complaints. When DC Water find a product that does not match the suitability of the reuse site, DC Water take it to a more remote site, if available, and if no site is available, the material is sent to the landfill. As a result of real-time solids odor

monitoring system had been initiated and implemented, DC Water have dramatically reduced their odor complaints. DC Water have had only three odor complaints in 2009, which is a drastic reduction from a few years ago. This work is new and has not been done anywhere else in the country, to our knowledge, and had placed DC Water in a leadership position with respect to odor monitoring and prediction.

### 1.7 Master Plan Format

**Chapter 1 – Introduction.** A summary of general information about odor in biosolids and the scope of work for this dissertation as well as the research objectives are presented. The previous studies in the area of odor mitigation for biosolids and the difference between this research and those previous studies are summarized together with a brief listing. Furthermore, the research hypotheses and research contributions are stated.

**Chapter 2 – Background and Literature Review.** The background of odorants generation and analytical measurement for quantifying odor in limed biosolids are summarized. Brief descriptions of each odorous production and each measurement performed are also included.

**Chapter 3 – Operational Processes and Data Description.** This chapter provides detailed descriptions of the following: plant wastewater treatment processes, the upstream operational process variables (independent variables) considered to contribute to odorants emission, the real-time odor monitoring system, and the methodologies used to collect odor data (dependent variable).

**Chapter 4 – Statistical Data Analysis for Biosolids Odorants.** Chapter four presents methodologies used to identify odor sources and developing statistical



models as a tool for explaining and forecasting odorous productions in limed biosolids. The steps and strategies for statistical analysis are also incorporated.

**Chapter 5 – Model Analysis, Results, and Discussion.** The results from statistical models used as a tool in explaining and predicting in-plant odor concentration are demonstrated. The potential process variables that contribute to odorous emission in biosolids are identified. Also, this chapter concludes with a discussion on how those process variables can promote the two main odorous compounds.

**Chapter 6 – Conclusions and Future Work.** Conclusions derived from the statistical models in Chapter 5 are summarized. Recommendations for odor mitigation in biosolids management program and the standard operating procedure (SOP) are documented. This chapter concludes with additional recommendations for work in odor mitigation in-plant.

## Chapter 2: Background and Literature Review

Biosolids are by-products of wastewater treatment plants, and are naturally associated with odor emissions. Those emissions are known as malodorous compounds, “the volatile emissions generated from the chemical and microbial decomposition of organic nutrients. When inhaled, these odorants interact with the odor sensing apparatus (olfactory system) and the person perceives odor” (USEPA 2000b).

Many studies indicated that organic and inorganic compounds, such as total reduced sulfur (TRS), nitrogen (N)-containing compounds (ammonia and amines), and organic fatty acids, are considered the major odorous compounds released from biosolids. These compounds normally occur by the mechanism of the treatment processes (Dague 1972; Hwang et al. 1995; USEPA 2000a; USEPA 2000b; Kim et al. 2001; Murthy et al. 2001; Murthy et al. 2002a; Kim et al. 2002; Novak et al. 2002; Turkmen et al. 2004; Rosenfield and Suffet 2004; Novak et al. 2004; Chen et al. 2005). Typically, odorants production in biosolids differ by the type of sludge treatment processes and management options used (USEPA 2000a; Visan 2003; Higgins et al. 2003; Chen et al. 2005).

### 2.1 Odorous Production in Lime-Stabilized Biosolids

The District of Columbia Water and Sewer Authority (DC Water)’s advanced wastewater treatment plant (WWTP) treats an average of 370 million gallons per day (mgd) of wastewater with more than 1,200 wet tons per day of biosolids generated (DCWASA 2005). For solids handling processes, DC Water employs a lime-

stabilization process to stabilize dewatered sludge before recycling on the land application sites.

Many studies indicated that TRS, which includes dimethyl sulfide (DMS), dimethyl disulfide (DMDS), carbon disulfide (CS<sub>2</sub>), and methyl mercaptan (MM) and N-containing compounds including ammonia (NH<sub>3</sub>) and trimethylamines (TMA) are the major odorous compounds emitted from lime stabilized biosolids (Murthy et al. 2001; Kim et al. 2001; Murthy et al. 2002b; Novak et al. 2002; Kim et al. 2002; Kim et al. 2005; Schneekloth et al. 2006; Chang et al. 2005; Chen et al. 2005). However, many studies showed that lime stabilization can diminish or eliminate hydrogen sulfide (H<sub>2</sub>S) (USEPA 1985; Stuetz and Frechen 2001; Murthy et al. 2001; Kim et al. 2001), but there are other odorants concentrations that can be detected in limed biosolids. The range of odor thresholds of odorants in limed biosolids is presented in Table 2.1.

**Table 2.1: The range of odor threshold of odorous production in limed biosolids**

Substance	Compound	Formula	Odor Character	Odor Threshold
Total reduced sulfur				ppb
	Hydrogen Sulfide	H <sub>2</sub> S	Rotten eggs	0.47-1.0
	Dimethyl Sulfide	(CH <sub>3</sub> ) <sub>2</sub> S	Decayed cabbage	0.1-50.8
	Dimethyl Disulfide	(CH <sub>3</sub> ) <sub>2</sub> S <sub>2</sub>	Vegetable sulfide	0.1-346.5
	Carbon Disulfide	CS <sub>2</sub>	Rotten pumpkins	7.7
	Methyl Mercaptan	CH <sub>3</sub> SH	Sulfidy or Decayed cabbage	0.5-1.6
N-containing compounds				ppb
	Ammonia	NH <sub>3</sub>	Pungent, irritating	26.6-46,800
	Trimethylamine	(CH <sub>3</sub> ) <sub>3</sub> N	Fishy, pungent	0.2-0.8

(Hentz 1997; USEPA 2000b; Visan 2003)

### 2.1.1 Total Reduced Sulfur Compounds

**Hydrogen sulfide (H<sub>2</sub>S)** is an inorganic sulfur compound, and is the most commonly known critical odorant in wastewater treatment processes (Dague 1972;

USEPA 1985; Hentz 1997; Gostelow and Parsons 2000; Yongsrl et al. 2004). The  $\text{H}_2\text{S}$  occurs not only when an anaerobic condition is presented, but also increases its emission when the pH is lower than 9. Nevertheless, lime stabilization can raise pH levels above 9 and suppress  $\text{H}_2\text{S}$  concentration while also causing the production of the other odorants in lime stabilized biosolids (Hentz 1997; USEPA 2000b; Stuetz and Frechen 2001).

**Dimethyl disulfide (DMDS) and Dimethyl sulfide (DMS)** are organic sulfur compounds released from biosolids. DMDS and DMS derive from biodegradation of organic sulfur compounds. DMDS is a by-product of protein degradation (Hentz 1997; Geng et al. 2004). Typically, DMDS is liquid at room temperature and it will raise the emission of DMDS whenever temperature is increased (Hentz 1997).

**Methyl mercaptan (MM)** is one of the most general mercaptans emitted from odorous production in biosolids. MM has a low odor detection threshold, which can be seen in Table 2.1 (Hentz 1997; USEPA 2000b). Therefore, it can easily cause the odor complaints from the residents around the land application sites.

#### 2.1.2 Nitrogen (N)-Containing Compounds

**Ammonia ( $\text{NH}_3$ )** is the most common odorant found in lime stabilization. It derives from either aerobic or anaerobic decomposition by the force of pH, temperature and sludge characteristics (Chang and Dentel 2001; Visan 2003; North 2003). Table 2.1 indicates that  $\text{NH}_3$  has a high odor detection threshold; thus, residents can detect it more easily. The high concentration of  $\text{NH}_3$  can disguise odor caused by TRS and lead to misrepresentation of the potential odor complaints at the fields (USEPA 2000b).

**Trimethylamine (TMA) ( $\text{CH}_3\text{N}$ )** is one type of amines and the most critical odorants emitted from lime-stabilized biosolids. Many studies indicated that the fishy odor at land application sites can be characterized as TMA (Murthy et al. 2001; Novak et al. 2002; Kim et al. 2005; Schneekloth et al. 2006). Whenever the temperature rises above 104 ( $> 40^\circ\text{C}$ ) or pH rises above 12.5, the emission of TMA in lime-stabilized biosolids can dramatically rise (Kim et al. 2002). In addition, other factors promoting the emission of TMA in lime-stabilized biosolids can include: the polymer dose, the time between the addition of polymer and lime stabilization, shear conveyed on the sludge in the dewatering process, and dewatered cake solids concentration (Schneekloth et al. 2006). The presence of amines is always accompanied with the presence of ammonia (USEPA 2000b).

## 2.2 Odor Measurement, Determination, and Reduction in Biosolids

To control and mitigate odor production in biosolids, it is crucial to know how to detect, identify, and quantify them by having the appropriate methods and the proper designs (Yang and Hobson 2000). Failure to do so can cause misjudgment and inaccurate analysis. There are two measurement categories, sensory measurement and analytical measurement (Yang and Hobson 2000; Gostelow and Parsons 2000; Stuetz and Frechen 2001).

In the last decade, many studies tried to detect, identify, and quantify the odorous compound concentrations of biosolids via analytical measurement by using chemical analyzers, instruments, and methods that can provide the accurate information.

There are a number of studies trying to quantify the level and explain the source of odorants production in dewatered solids and biosolids or on-site of the wastewater treatment processes by simulating the solids treatment processes in order to get the biosolids sampling. The studies of Murthy et al. (2001), Murthy et al. (2002b), Novak et al. (2002) and Schneekloth et al. (2006) employed the headspace approach, which measure a gas sample from the container by using a portable gas detector and draw a gas sample by using tedlar bags for analysis. However, Kim et al. (2001) and Murthy et al. (2002a) studies used the biosolids sampling from each wastewater treatment process and then utilized the headspace approach with solids phase microextraction. All of those studies were analyzed by using the gas chromatography and also simulated the solids processes to get biosolids sampling.

The studies of Vilalai (2003) and Vilalai (2008) employed the headspace approach with a portable hydrogen sulfide analyzer (Jerome 631X), but in the study of Vilalai (2008), an odor panel for sensory analysis by using the gas sample from headspace tedlar bags was used. In addition, some studies employed odor dispersion models to do the following: forecast the impact of odor from solid waste landfills (Sarkar et al. 2003); identify the impact of odors from composting facilities (Williams and Servo 2005), and identify the sources of odors onsite impacting the residential communities (Voelz et al. 2006).

To the best of our knowledge, none of those above-mentioned studies quantifies the odor concentration from an actual WWTP by using a real-time solids odor monitoring system. For this reason, this research is employing the real-time solids odor monitoring system for odor data collection. However, some of those

studies used either the solids sampling from the simulated treatment processes or those from each step of the solids treatment processes, and then placed them in the containers for headspace analysis. In addition, all of those studies employed the static sampling system.

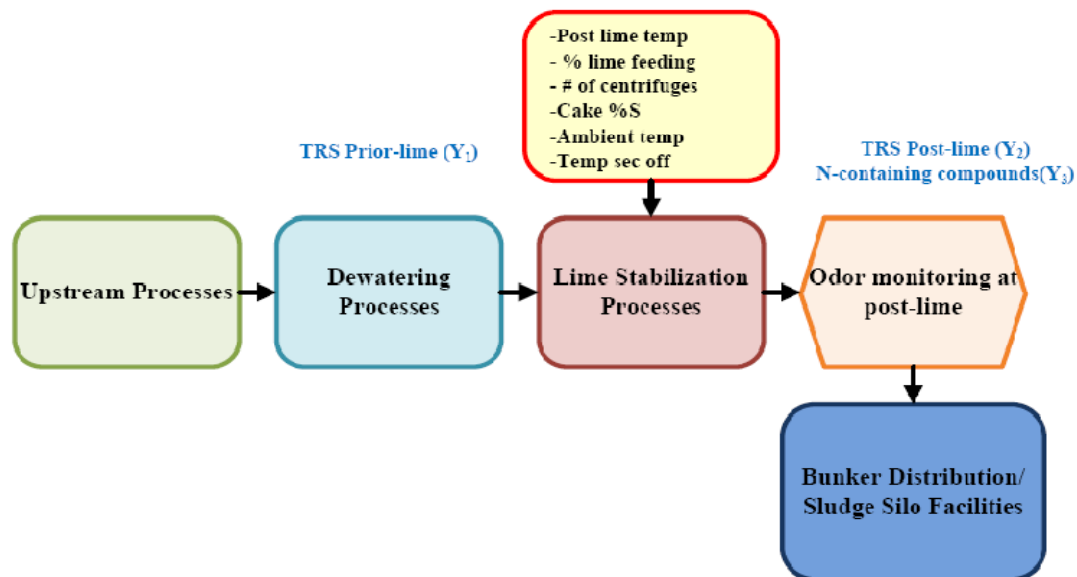
One of the most advantages of real-time solids odor system comparing to those previous studies is the new insight on the impact of daily activities in upstream treatment processes on changing the emission of biosolids odor. The information generated by real-time solids odor system can provide WWTP unprecedented control over the ways that have been done before.

### 2.3 Ordinary Least Squares (OLS) and Instrumental Variable (IV) Estimation

The ordinary least squares (OLS) estimation and instrumental variable (IV) estimation are employed to analyze the empirical data of odorants productions in dewatered solids and lime-stabilized biosolids. Many studies indicated that OLS is the most effective way to analyze empirical data (Wooldridge 2000; Wooldridge 2002; Greene 2003; Kennedy 2008). However, using only OLS estimation for creating a statistical model has some weaknesses. For instance: (i) there are measurement error issues in independent variables or omitted variables issues in approximation of causal relationships; (ii) there are correlations between the independent variables and residuals in the model. However, using IV estimation can overcome those issues (Angrist and Imbens 1995; Angrist and Krueger 2001).

DC Water uses lime stabilization after the dewatering process. Figure 2.1 presents a schematic of solids handling processes and all locations of odor monitoring systems. It was believed that lime stabilization mitigates or eliminates total reduced

sulfur (TRS) for prior lime ( $Y_1$ ) to become TRS at post lime ( $Y_2$ ) concentration (USEPA 1985; Stuetz and Frechen 2001; Murthy et al. 2001; Kim et al. 2001). Therefore, it is difficult to directly identify or explain the relationship between the upstream processes and the TRS ( $Y_2$ ) production in biosolids at the post lime process. However, by applying the knowledge of OLS and IV estimation via two-stage least squares (2SLS), these approaches can assist DC Water to identify the causes or the sources of TRS at post lime processes.



**Figure 2.1: A schematic of solids handling processes at DC Water**

### 2.3.1 Ordinary Least Squares (OLS) Estimation

Ordinary least squares (OLS) estimation is a well-known method in econometrics, especially in empirical data analysis (Wooldridge 2000; Wooldridge 2002; Greene 2003; Montgomery et al. 2006; Kennedy 2008). This method is used to estimate the unknown parameters in a linear regression model. Additionally, this method can be used in either simple linear regression (SLR) or multiple linear regression (MLR)



(Wooldridge 2000; Wooldridge 2002; Montgomery et al. 2006; Kennedy 2008). The idea of OLS is to minimize the sum of squared deviations between the true observed dependent variable values ( $Y$ ) and the corresponding predicted values ( $\hat{Y}$ ). In other words, OLS estimators minimize the sum of squared deviations (residuals) (Shmueli et al. 2007; Kennedy 2008). The following are OLS Assumptions:

Consider a regression model of the form

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \dots + \alpha_k X_{ki} + \rho_i, \text{ where } \rho_i \text{ is the error term.}$$

- (i) There is no correlation between the independent variables and the error term, i.e.,  $\text{corr}(X_{ki}, \rho_i) = 0$
- (ii) The expected value of the error term is zero, i.e.,  $E(\rho_i) = 0$
- (iii) The variance of the error term is constant (homoskedasticity), i.e.,  $E(\rho_i^2) = \sigma^2$
- (iv) The errors are independent, i.e.,  $E(\rho_i \rho_j) = 0$
- (v) The errors are normally distributed, i.e.,  $\rho \sim N(0, \sigma^2)$

According to Wooldridge (2000) and Montgomery et al. (2006), ordinary least squares (OLS) estimation method can be briefly described by the following statements:

In real world applications, and especially in biosolids treatment processes, there are more than one upstream process variable (independent variable) affecting or causing odorants production in biosolids (dependent variable).

Wooldridge (2000) stated that “the multiple regression model is still the most widely used vehicle for empirical analysis in economics and other sciences. Likewise, the method of ordinary least squares is popularly used for estimating the parameters

of the multiple regression model.” For this reason, these analyses apply OLS estimation via the MLR model to create both explanatory and predictive models for total reduced sulfur (TRS) at prior-lime ( $Y_1$ ), TRS at post-lime ( $Y_2$ ), and nitrogen (N)-containing compounds ( $Y_3$ ). In other words, the analyses run OLS regressions of three different dependent variables (TRS at prior lime, TRS at post lime, and N-containing compounds) on the set of independent variables (upstream processes). The general MLR model can be seen in the following equation (Wooldridge 2000).

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \dots + \alpha_k X_{ki} + \rho_i \quad (2.1)$$

where  $Y_i$  is the dependent variable ( $Y_i$  can be either TRS at prior lime ( $Y_1$ ), TRS at post lime ( $Y_2$ ), or nitrogen (N)-containing compounds ( $Y_3$ )),  $X_{1i}, X_{2i}, X_{3i}, \dots, X_{ki}$  are  $k$  upstream process variables included in the model,  $\alpha_0$  is the intercept,  $\alpha_1, \alpha_2, \alpha_3 \dots \alpha_k$  are the regression coefficients of any independent variable in model,  $\rho_i$  is the error or residual including any independent variables that affect  $Y_i$  but are not included in the model. As equation (2.1) presents MLR with  $k$  independent variables and an intercept, there will be  $k+1$  unknown parameters. In addition, upstream process variables ( $X_i$ ) are not always the same in each model.

In this research, there will be an OLS regression equation for each dependent variable (TRS at prior-lime, TRS at post-lime, and N-containing compounds). The structural model for each dependent variable can be seen in Table 2.2.

**Table 2.2: Regression equations for odorous compounds**

Odorants compounds	Equations
TRS at prior lime (ppm)	$Y_l = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \alpha_k X_k + \rho_i$ <p>where <math>Y_l</math> is TRS at prior lime, <math>X_1, X_2, X_3, \dots, X_k</math> are upstream process variables until dewatering processes, <math>\alpha_0</math> is the constant or intercept, <math>\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_k</math> are regression coefficients of any upstream processes, <math>\rho_i</math> is residual.</p>
TRS at post lime (ppm)	$Y_2 = \beta_0 + \beta_1 Y_1 + \beta_2 X_i + \beta_3 W_i + \varepsilon_i$ <p>where <math>Y_2</math> is TRS at the post lime process, <math>Y_1</math> is TRS at the prior lime process, <math>X_i</math> are any upstream processes variables until the lime stabilization process, <math>W_i</math> is the lime addition processes variable, <math>\beta_0</math> is the constant or intercept, <math>\beta_1, \beta_2</math>, and <math>\beta_3</math> are regression coefficients of any independent variable included in the model, <math>\varepsilon_i</math> is residual.</p>
N-containing compounds (ppm)	$Y_3 = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_k X_k + \mu_i$ <p>where <math>Y_3</math> is N-containing compounds, <math>X_1, X_2, X_3, \dots, X_k</math> are any upstream process variables including the lime stabilization process, <math>\gamma_0</math> is intercept, <math>\gamma_1, \gamma_2, \dots, \gamma_k</math> are regression coefficient of any independent variables included in the model, <math>\mu_i</math> is residual.</p>

The model for TRS prior lime is used for finding what process variables at the upstream cause TRS at prior-lime ( $Y_l$ ), because the upstream processes are the direct sources of  $Y_l$ . This research encounters a challenge with the model for TRS at post-lime ( $Y_2$ ) which includes additional process variable (lime stabilization process,  $W$ ). The problem that arises is that this variable  $W$  not only affect the dependent variable directly, but also affect one of the independent variables, namely, TRS prior lime. This occurs because DC Water employs lime stabilization to generate biosolids by maintaining pH above 12 for at least 24 hours, such that the lime addition process suppresses the level of TRS at prior-lime which then becomes TRS at post-lime (USEPA 1985; Stuetz and Frechen 2001; Murthy et al. 2001; Kim et al. 2001).

It was believed that in some of the process variables, especially at the lime addition processes, having relationships with odorants cannot be included in the regression model for TRS post-lime. Instead, those process variables are included in

the error term, and cause a correlation between the error term and some of the independent variables. Additionally, there might be some measurement errors to quantify of some the process variables at lime addition processes. For this reason, this research employs an instrumental variable (IV) coupled with two-stage least squares (2SLS) estimation. The next topic will discuss IV and 2SLS estimation.

### 2.3.2 Instrumental Variable (IV) and Two-Stage Least Squares (2SLS)

#### Estimation

Because the first goal of this research is to explain the causes or the sources of odorous productions in biosolids, the analyses must deal with the causality issue mentioned above by addressing the weaknesses of the OLS estimation method. A common technique in economics used to overcome the issues of endogeneity and measurement error of independent variables in OLS estimation of regression models is the construction of an instrumental variable (IV) (Angrist and Imbens 1995; Wooldridge 2000; Angrist and Krueger 2001; Wooldridge 2002; Kennedy 2008).

Recall that one of the assumptions underlying OLS estimation is independence of the right-hand side variables from the error term ( $\text{corr}(X_{ki}, \rho_i) = 0$ ). However, independent variables (covariate) can have correlation with error term. This correlation may occur when there are related independent variables which cannot stay in the model or when there are some measurement errors in independent variables (Angrist and Krueger 2001). Therefore, the independent variable, which is correlated with the error term, is endogenous. In contrast, the other independent variables, which are not correlated with the error, are called exogenous variables (Wooldridge 2002).

According to the studies of Angrist and Imbens (1995), Angrist and Krueger (2001), and Murray (2006), when there is correlation between error terms and one of the independent variables in the model, it indicates that one or more independent variables cannot be included in the model. This exclusion means those independent variables cannot directly explain dependent variable or they do not belong in equation and are correlated with endogenous variable. In case of the model for TRS at post lime, we can write it as follows:

$$Y_2 = \beta_0 + \beta_1 Y_1 + \beta_2 X + \beta_3 W \quad (2.2)$$

where  $Y_2$  is TRS at the post lime process,  $Y_1$  is TRS at the prior lime process,  $X$  is any upstream processes variable before the lime stabilization process, and  $W$  is any lime addition process variable included in the model.

According to equation (2.2), if the analyses employed OLS by having one of those independent variables correlated with the error term,  $\rho_i$ , it would have caused biases in the estimated regression coefficients. In addition, it would have caused our conclusions to be misleading when explaining the causes or sources of  $Y_2$ . Instead, the use of an IV and 2SLS estimation can consistently estimate these coefficients, and identify what process variables hide in the error term (Angrist and Krueger 2001; Wooldridge 2002; Greene 2003; Murray 2006; Kennedy 2008). The following example presents how IV assists to solve this issue:

Suppose there is one independent variable ( $X_i$ ) in the regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \rho_i \quad (2.3)$$

The ordinary least squares estimator is:

$$\begin{aligned}
\hat{\beta}_{OLS} &= \frac{Cov(Y, X)}{Cov(X, X)} \\
&= \frac{Cov(\beta_0 + \beta_1 X + \rho, X)}{Cov(X, X)} \\
&= \frac{\beta_1 Cov(X, X) + Cov(\rho, X)}{Cov(X, X)} \\
&= \beta_1 + \frac{Cov(\rho, X)}{Cov(X, X)} \tag{2.4}
\end{aligned}$$

For the value of  $\hat{\beta}_{OLS}$ , if there is no correlation between  $X$  and  $\rho$ ,  $\hat{\beta}_{OLS}$  will be equal to  $\beta_1$ . In other words, the OLS estimator is said to be consistent. However, when there is correlation between  $\rho$  and  $X$  ( $Cov(\rho, X) \neq 0$ ),  $\hat{\beta}_{OLS}$  does not estimate  $\beta_1$ .

According to equations 2.3 and 2.4, if  $X$  and  $\rho$  are correlated, it can make  $\hat{\beta}_{OLS}$  not to estimate  $\beta_1$ . Therefore, there is other variable that depends on the correlation between  $X$  and  $\rho$ . In other word, the effect of  $X$  on  $Y$  can be explained by using  $Z$ , which is related to  $X$ , but not  $\rho$ . The variable  $Z$  is called instrument variable (IV); the following are assumptions (conditions) for IV estimation:

(i) There is no correlation between the IV and the error term, i.e., the IV is exogenous or  $corr(Z, \rho) = 0$

(ii) There must be a high correlation between IV and independent variables replaced by that IV, i.e., IV is relevant:  $corr(Z, X) \gg 0$

Thus,

$$\hat{\beta}_{iv} = \frac{Cov(Y, Z)}{Cov(XZ)} \tag{2.5}$$

The two regression equations can assist to understand the logic of IV:

$$Y_i = \beta_0 + \beta_1 X_i + \rho_i \quad (2.6)$$

$$X_i = \alpha_0 + \alpha_1 Z_i + \theta_i \quad (2.7)$$

Referring to above two assumptions of IV and equations 2.6 and 2.7,  $\hat{\beta}_{IV}$  can estimate  $\beta_1$ :

$$\begin{aligned} \hat{\beta}_{IV} &= \frac{Cov(Y, Z)}{Cov(X, Z)} \\ &= \frac{Cov(\beta_0 + \beta_1 X + \rho, Z)}{Cov(X, Z)} \\ &= \frac{\beta_1 Cov(X, Z) + Cov(\rho, Z)}{Cov(X, Z)} \\ &= \beta_1 + \frac{Cov(\rho, Z)}{Cov(X, Z)} \\ &= \beta_1 \end{aligned} \quad (2.8)$$

As stated earlier in Section 2.2.1 and in Figure 2.1, the lime addition caused the change in TRS at prior lime to become TRS at the post lime process. Therefore, a lime addition process variable ( $W$ ), which is monitored as the temperature of the limed biosolids, can be an endogenous variable having correlation with the error. Having said that, by using the Hausman Test for endogeneity testing, it can verify whether or not  $W$  is endogenous (Wooldridge 2000; Wooldridge 2002; Greene 2003; Kennedy 2008).

To verify this, the analyses simply use the set of variables at the lime addition processes, denoted as  $Z$ , to be the IV. The set of variables at the lime addition processes are the percent of lime feeding, the temperature of the secondary effluent, the number of centrifuges in service, the ambient temperature, and the dewatered percent solids. Thus, if these analyses regress  $W$  on the set of lime addition

parameters showed in equation (2.9), then the analyses will get the residual of this regression,  $\mu$ .

$$W_i = \gamma_0 + \gamma_1 Z_i + \mu_i \quad (2.9)$$

where  $W_i$  is temperature of limed biosolids,  $Z_i$  is a set of process variables at the lime addition process,  $\gamma_0$  is an intercept,  $\gamma_1$  is regression coefficient of any process variable at the lime addition process included in the model, and  $\mu_i$  is residual.

The analyses again run the regression by using  $\mu$ , which is gotten from equation (2.9), as one of the independent variables on the right-hand side of equation (2.3). It is shown in equation (2.10).

$$Y_2 = \beta_0 + \beta_1 Y_1 + \beta_2 X_i + \beta_3 W_i + \Omega \mu \quad (2.10)$$

If the coefficient,  $\Omega$ , of  $\mu$  is not zero and is statistically significant,  $W_i$  is endogenous, otherwise it is exogenous. The fitted value of  $\hat{W}_i$  of equation (2.9), called the first-stage least squares, will be the average of the temperature of lime biosolids.

The analyses again replace  $W_i$  with  $\hat{W}_i$  in equation (2.2) and then run the regression to obtain  $\hat{Y}_2$ .

$$\hat{Y}_2 = \gamma + \omega Y_1 + \theta \hat{W}_i + \eta X_i \quad (2.11)$$

where  $\hat{Y}_2$  is the fitted value of TRS at post lime process,  $Y_1$  is TRS at the prior lime process,  $\hat{W}_i$  is the fitted value of the lime addition processes,  $X_i$  are upstream processes,  $\gamma$  is intercept, and  $\eta$ ,  $\omega$  and  $\theta$  are regression coefficients. Then, the fitted value of equation (2.11) is called the second-stage least squares prediction.

For the model of N-containing compounds ( $Y_3$ ), this analyses use all the upstream process variables as independent variables. Because many studies indicate that trimethylamine (TMA) is only detected during post lime stabilization (Kim et al.



2001; Murthy et al. 2002a; Novak et al. 2002), the analyses also include the lime-stabilizing process variables as additional independent variables.

## Chapter 3: Operational Processes and Data Description

This chapter describes in detail the in-plant wastewater treatment processes, DC Water's advanced WWTP, and also provides the details of each upstream operational process variable (independent variables) considered to contribute to odorants emission. Additionally, this chapter describes how odor monitoring systems are set up and how odor data collection used as a dependent variable is determined.

### 3.1 Site Overview

The Blue Plains Advanced Wastewater Treatment Plant (AWTP) has been a pioneer environmental warden since it started operations in 1938, as shown by its continual development in responding to the requirements, regulations, and community needs since that time. The District of Columbia Water and Sewer Authority (DC Water)'s advanced wastewater treatment plant (WWTP) treats the combination of storm water and untreated wastewater flowing from the sewers of the Washington, DC, metro area covering more than 725 square miles, which includes the District of Columbia, portions of Montgomery and Prince George's Counties in Maryland, and portions of Fairfax and Loudoun Counties in Virginia.

The treated water discharges into the Potomac River. The installed capacity treats an average of 370 million gallons per day (mgd) of wastewater, with a treatment peak flow of 740 mgd and an excess flow of 336 mgd. The peak wet-weather capacity treats 1,080 mgd, with more than 1,200 wet tons per day of biosolids generated and beneficially used (DCWASA 2005).

The wastewater treatment processes at DC Water are composed of two major processes, liquid and solids processes (Figure 3.1). The liquid processes start after the influents pass the screening process, where debris and grit, are removed and trucked to a landfill. Then, the sewage flows into primary sedimentation tanks by having about 50% of the suspended solids separated from the liquid. The primary effluent flows into the secondary treatment process, a biological process, by entering aeration tanks where microbes break down the organic matter.

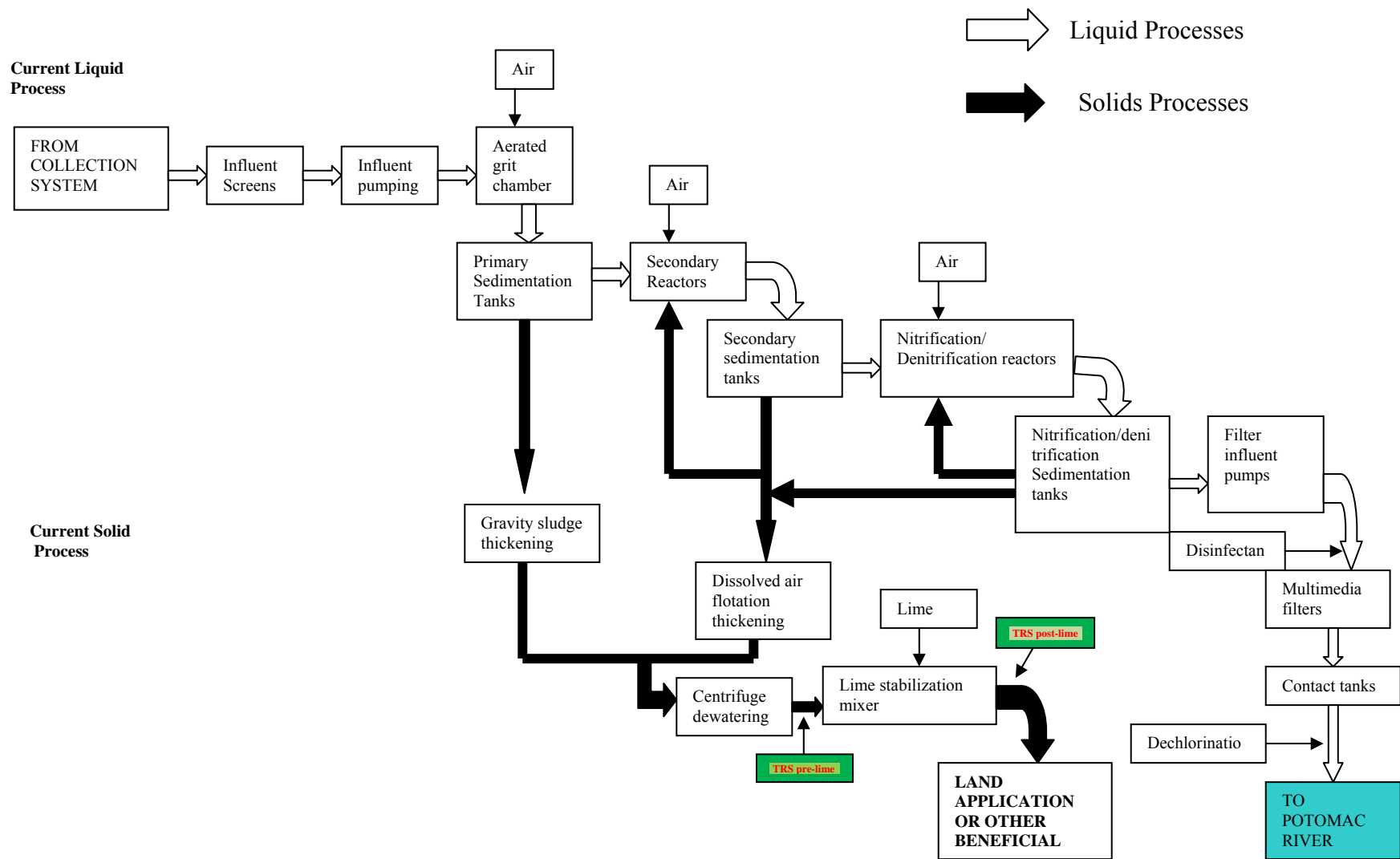
Next, the secondary effluent enters secondary sedimentation tanks, where solids are settled out and returned to the aeration tank. The overflow moves into the nitrification and denitrification tanks, and converts ammonia to nitrate and nitrate to nitrogen gas. These processes make DC Water an advanced wastewater treatment facility. The residual solids are settled out and the water is percolated through sand filters. The sand filters remove the remaining suspended solids and associated phosphorus. The water is disinfected, dechlorinated, and discharged into the Potomac River (DCWASA 2005).

In contrast, the solids handling processes for generating biosolids begins at the primary sedimentation tanks. In this tank, the solids are pumped to the gravity thickener (GT) tanks where gravity causes the solids to settle and thicken. At the secondary sedimentation, the settled solids are pumped back to secondary aeration tanks to retain a concentration of microorganisms in the aeration tanks (Vesilind 2003). The remainder of the settled solids will be pumped/wasted to dissolved air flotation (DAF) thickeners (likewise for the solids from nitrification sedimentation

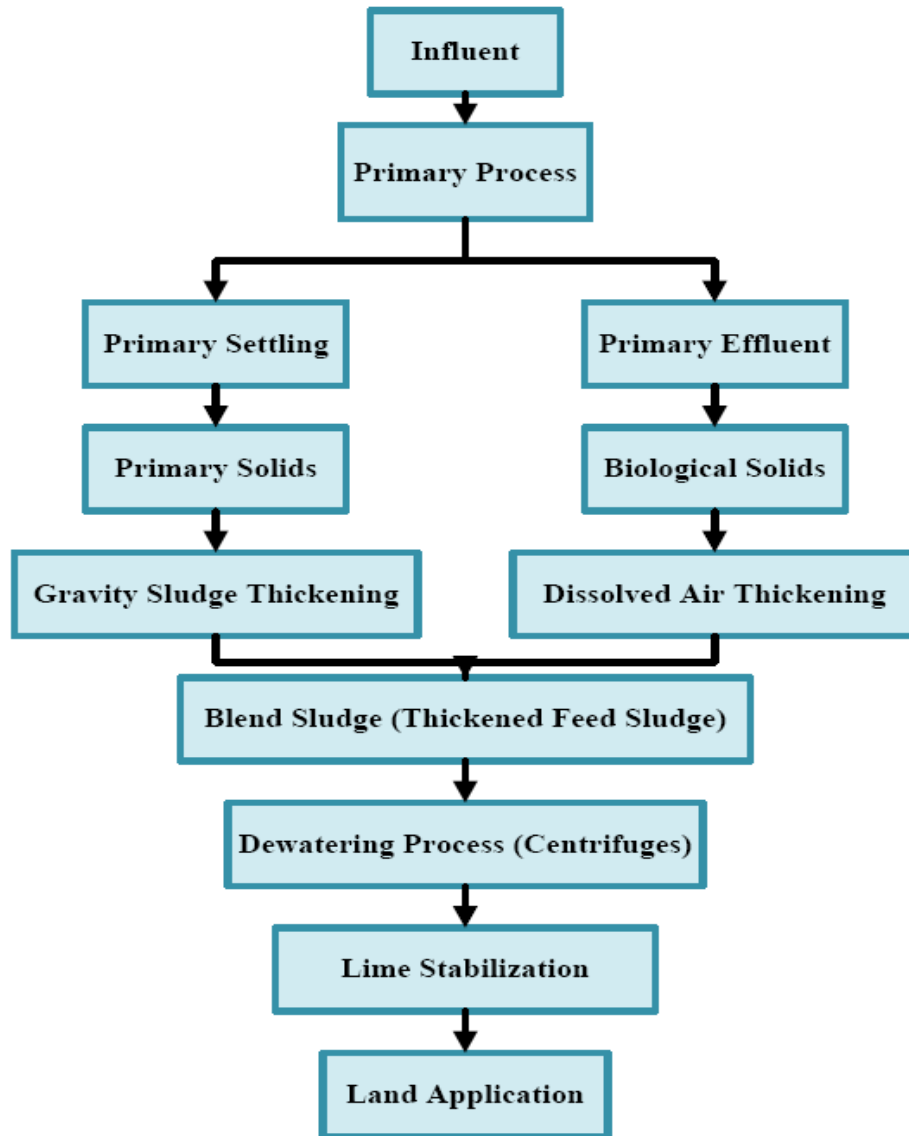
tanks), and some portion will be pumped back to the nitrification reactor with the rest sent to the secondary reactors (Peot and Ramirez 2007).

At the DAF thickeners, the combined particles settle at the bottom while air is blown in to float these particles to the surface, called flotation-thickened sludge. Then, a chained skimmer removes the surface and sends the sludge to the sludge blending tank system. In this tank, GT and DAF thickened solids are subsequently blended to form a homogeneous sludge with a ratio goal of one-to-one (50% GT solids to 50% DAF solids) (Peot and Ramirez 2007).

After the blending, the homogeneous sludge is dewatered by high-speed centrifuges, where sludge solids are separated from the water. Next, the dewatered solid (cake) goes through lime stabilization for reducing pathogens and producing biosolids. Finally, the organic biosolids are hauled out for land application. The DC Water solids handling processes diagram is presented in Figure 3.2.



**Figure 3.1: DC Water Processes Diagram (DCWASA 2005; Janpengpen 2006)**



**Figure 3.2: DC Water solids handling processes diagram (DCWASA 2005; Janpengpen et al. 2008)**

### 3.2 Operational Processes Data (Upstream Process Variables)

Since this research continues from the previous study of Janpengpen (2006), it collects all plausible operational wastewater treatment process variables from the upstream to the downstream processes for polluted analysis. The duration of operational processes data collection started simultaneously with odor data collection

in January 2009 to May 2010. For odor monitoring, however, only TRS at prior lime and post lime processes were monitored during January 2009 to April 2009.

All the below data have been collected from the daily report, the process control historian (PCH), and the process control system (PCS). The following wastewater treatment process variables (independent variables) are possibly considered the sources of odorous production in limed biosolids:

### 3.2.1 Ferric Chloride ( $\text{FeCl}_3$ ) Addition

Ferric chloride ( $\text{FeCl}_3$ ) is a flocculant and added for making tiny particles clump together so they can be removed from wastewater.  $\text{FeCl}_3$  consists of two elements, irons (Fe) and chlorine (Cl), and is by-product from a liquid of steel-making industries. There are a number of benefits for ferric chloride addition. These include, for instance, improving sludge settling, decreasing influent phosphorous concentrations, increasing dewatering performance by lowering polymer requirement for dewatering operations, and increasing BOD and total suspended solids (TSS) removal (WEF 1995(a)).

At DC Water,  $\text{FeCl}_3$  is added to the primary settling process for removing nutrients such as phosphorous in order to prevent algae growth in the Potomac River and also enhance settling performance (WEF 1996; Peot and Ramirez 2007).

### 3.2.2 Waste Pickle Liquid (WPL) Addition

Waste Pickle liquid (WPL) is a waste product from steel manufacturing for cleaning and removing mill scale (WEF 1995(a)). WPL consists of acid concentration with varying concentration of heavy metal such as chromium and nickel; however, many

other metals can be present. At DC Water, WPL is also added to the primary process for eliminating phosphorous, nutrients, and assisting to mitigate hydrogen sulfide at the upstream in wastewater system. Occasionally, DC Water utilizes either  $\text{FeCl}_3$  or WPL at the primary process.

### 3.2.3 Polymer Addition

There are three types of polymer: anionic, cationic, and non-ionic (WEF 1996).

Normally, polymers are used for sludge conditioning in separating, thickening, and dewatering; improvement of coagulation/flocculation; and expansion of alum, ferric, or calcium precipitation of phosphorous in wastewater (Churchill and Rybacki 1997; Chang et al. 2005). Cationic and other polymers are coagulant aids used in combination with other inorganic coagulants (Chang and Dentel 2001; Chang et al. 2005). The long chains of positively charged polymers can help to strengthen a floc, making it larger, faster settling and easier to filter out. Before 2004, DC Water added polymer only at the dewatering process and dissolved air floatation (DAF) processes. DC Water, however, now adds polymer in each step of the treatment processes; primary process, secondary process, nitrification process, dewatering process and dissolved air Floatation (DAF) process each improve clarification or hold sludge particles together. Only at primary process is the anionic polymer added. We, therefore, have polymer addition at each process as independent variables.

Several studies have shown that during and post lime stabilization processes, polymer added into sludge is one of the major factors contributing to odorous production in biosolids, especially amine product, e.g., Trimethyl Amine (TMA) (Kim et al. 2001; Novak et al. 2002; Kim et al. 2005; Chang et al. 2005). We



anticipate that polymer addition will have a relationship with odorants production at the post lime process.

### 3.2.4 Gravity Thickening (GT) at Primary Settling

There are four process variables, gravity thickening (GT) concentration, GT flow, GT total solids, and percent solids of GT, involved in this topic. As we stated earlier in section 3.1 (site overview) about 50% of the suspended solids separate from the liquid by GT and settle at the primary settling tank. These solids are considered as the raw sludge called total primary sludge (TPS) and are pumped to blending tanks to combine with sludge from the secondary process. DC Water tries to maintain percent solids of GT between 6-10% (Peot and Ramirez 2007).

This study, however, can calculate GT total solids used for finding blend ratio at the blending tank by employing the following formula:

$$GT\ TS(lbs / day) = \left[ \frac{GT\ flow(gallons / day) * 8.345 * GT\ \%S * 10000}{1000000} \right] \quad (3.1)$$

Note; 1% of solution is  $10^4$  mg/L and 1 mg/L is  $8.345 \times 10^{-6}$  lbs/gallon

It was believed that the more GT total solids in the blending tank, the more percent solids of dewatered solids and the more odorous production in the dewatered solids.

The details for using GT total solids to calculate blend ratio will be restated in 3.2.11.

### 3.2.5 Secondary Sludge Blanket Depth

At the secondary aeration tank, microorganisms consumed organic matter and then the wastewater flowed into the secondary settling tank to separate suspended solids

by gravity. The depth of the settled solids in this tank is called the sludge blanket depth. The settled solids at the bottom, however, will be called waste activated sludge (WAS), if they are wasted or pumped to dissolved air flotation (DAF). Additionally, some part of the settled solids (return activated sludge (RAS)) will be pumped back to the secondary aeration tank in order to balance microorganisms in the biological process (Vilalai 2003).

At DC Water, there are two sides of secondary settling tanks, namely, secondary east settling tank and secondary west settling tank. There are, however, three sources for secondary sludge blanket depth: secondary east blanket depth, secondary west even blanket depth, and secondary west odd blanket depth (DCWASA 2005; Peot and Ramirez 2007). DC Water tries to maintain and prefers secondary sludge blanket depth at 2 feet.

Some studies indicated that the greater secondary sludge blanket depth, the greater the biosolids odors at the downstream (Sekyiamah and Kim 2006; Sekyiamah and Kim 2009). The underlying principle is that higher blanket depth demonstrates more retention time for biological activated sludge in the settling tanks and can cause the expansion of anaerobic conditions, which lead to the production of total reduced sulfur (TRS) compound at the prior lime and post lime addition.

Additionally, the study of Janpengpen (2006) and Janpengpen et al. (2007) suggested that if DC Water maintained secondary sludge blanket depth higher than 1.8 feet, it can promote the higher total reduced sulfur (TRS) compounds production in dewatered solids and can also cause the higher cost for polymer addition at the dewatering process.

### 3.2.6 Secondary Waste Activated Sludge

There are three process variables involved in this topic: waste activated sludge (WAS) concentration, WAS flow, and WAS percent solids (WAS %S). WAS is biological activated sludge in the secondary sedimentation tank consisting of both sludge from the secondary process and the nitrification process. This sludge, called total waste activated sludge (TWAS), will be pumped to the dissolved air flotation (DAF) tank, and then will be pumped to the blending tank to mix with sludge from the primary process (TPS). It was believed that if operators can retain either flow or concentration of TWAS higher than TPS, they could assist to mitigate odorants production in biosolids at the post lime process.

### 3.2.7 Dissolved Air Flotation (DAF)

DAF is TWAS received from the secondary process, the nitrification process, and the denitrification process. There are two process variables involved in this topic: DAF flow and DAF total solids. DAF also relates to percent solids of WAS because the DAF total solids can be gained from the following formula:

$$\text{DAF TS}(\text{lbs} / \text{day}) = \left[ \frac{\text{DAF flow}(\text{gallons} / \text{day}) * 8.345 * \text{WAS \%S} * 10000}{1000000} \right] \quad (3.2)$$

Note; 1% of solution is  $10^4$  mg/L and 1 mg/L is  $8.345 \times 10^{-6}$  lbs/gallon

The DAF total solids can be used to calculate the blend ratio at the blending tank.

More details for using DAF total solids to calculate blend ratio will be restated again in Section 3.2.11.

### 3.2.8 Initial Settling Velocity (ISV)

The study of Sekyiamah and Kim (2009) defined initial settling velocity (ISV) as “a measure of how well a sludge sample settles in the first few minutes of the settling process and thus captures the true nature of the settling sludge (i.e., fast settling or slow settling).” This means the higher ISV, the better the sludge settling. In contrast, the lower the ISV, the poorer the sludge settling. There will be three ISV as process variables from three locations, ISV at secondary east, ISV secondary west even, and ISV at secondary west odd.

### 3.2.9 Settle Sludge Volume (SSV<sub>60</sub>)

Settle sludge volume (SSV<sub>60</sub>) has the same function as ISV, however it measures how well sludge settles in 60 minutes (Sekyiamah and Kim 2009). Contrariwise for ISV, the higher the concentration of SSV<sub>60</sub> indicates the poor sludge settle ability. There are three locations for SSV<sub>60</sub>, SSV<sub>60</sub> of secondary east, SSV<sub>60</sub> of secondary west even, and SSV<sub>60</sub> of secondary west odd.

### 3.2.10 Concentration of Return Activated Sludge (RAS)

RAS consists of settled solids at the bottom of a secondary sedimentation tank that are pumped back to a secondary aeration tank to carry on or retain an ample concentration of activated sludge in the aeration tank. In general, RAS concentrations range from 4,000 to 12,000 mg/L (Metcalf&Eddy 2003). The study of Vilalai (2008), however, indicated that high concentration of RAS can be the source of odorants in dewatered solids and biosolids because a high concentration of RAS demonstrated a

high density of settled solids at the bottom of the secondary sedimentation tank. That means these two process variables relate to each other.

Likewise for this study, we anticipate that concentration of RAS and secondary sludge blanket depth could be the sources of odorants in biosolids and also dominate the emission of odorants.

### 3.2.11 Blend Ratio

Blend ratio is the mixing between raw sludge from the primary process and activated sludge (secondary and nitrification process) from dissolved air flotation. In other words, blend ratio is the ratio between total primary sludge (TPS) and total waste activated sludge (TWAS). Referring to Sections 3.2.4 and 3.2.7, the formula of blend ratio will be:

$$\text{Blend Ratio} = \left[ \frac{\text{GT TS}(\text{lbs} / \text{day})}{\text{DAF TS}(\text{lbs} / \text{day})} \right] \quad (3.3)$$

Blend ratio in this study, therefore, indicates the ratio of dry solids of raw sludge over dry solids of activated sludge. If blend ratio is higher than 1, it indicates that there are more dry solids of GT than DAF. On the other hand, if the blend ratio is lower than 1, it indicates that there are less dry solids of GT than DAF. Additionally, if blend ratio is 1, it shows that dry solids of GT equal the dry solids of DAF.

In contrast, the study of Vilalai (2008) defined the formula of blend ratio regarding the volume of TPS and TWAS:

$$\text{Blend Ratio} = \left[ \frac{\text{Volume of TPS}(\text{gallons} / \text{day})}{\text{Total Volume (TPS+TWAS)}(\text{gallons} / \text{day})} \right] \quad (3.4)$$

Nonetheless, this study anticipates that blend ratio can be one of the key factors for contributing to odorants production in biosolids.

### 3.2.12 Temperature of secondary effluent

Temperature is a factor for microorganisms' activities, especially for microbial digestion. The study of Janpengpen et al. (2008) indicated that there were high solids content or high percent cake solids during June to September which flow temperature is about 23 °C to 28 °C, and low solids content or low percent cake solids during December to March which flow temperature is about 10 °C to 15 °C. That means, the higher the flow temperature, the higher the secondary sludge blanket depth. Additionally, in the process with temperatures more than 30 °C, sulfide production rate can be raised 7% per 1 °C (USEPA 1985; Bowker et al. 1989; Vilalai 2008).

In this study, this variable is the temperature at the secondary effluent from the secondary aeration tank to the secondary settling tank. We also anticipate that flow temperature can be one of the significant factors contributing to odorants production in biosolids. Additionally, this independent variable is an uncontrolled factor. DC Water, however, can use this information from this research to apply to a biosolids management perspective.

### 3.2.13 Cake Solids

After solids from total primary sludge (TPS) and solids from total waste activated sludge (TWAS) mix well together at the blending tank, the blended solids will be pumped to dewatering processes to separate water from solids by using high speed centrifuges. The products from each high speed centrifuge are dewatered solids.

DC Water by Department of Wastewater Treatment tries to achieve a solids content in dewatered solids more than 24%. Increasing 1% of the solids content in dewatered solids can reduce the hauling cost for land application up to million dollars a year (Peot 2007; Peot and Ramirez 2007).

Referring to Section 3.2.11, the blend ratio is one of the main factors controlling cake solids. According to experiences and operational point of views from operators at DC Water, they suggest that the more TPS in the blend, the higher the solids content in the cake solids and the higher the dewatering performance. One of the studies at DC Water mentioned that the dewatered solids with higher solids content needed more time in the lime mixing process to achieve the minimum requirements of lime stabilized biosolids because dry cake solids are more difficult to mix with lime (North 2003). In other words, the dewatered solids with high solids content have higher viscosity than dewatered solids with lower solids content (ASCE 2000).

Schneekloth et al. (2006) indicated that the dewatered solids with higher solids content generated total reduced sulfur (TRS) and trimethylamine (TMA) higher than dewatered solids with lower solids content. This research, however, anticipates that the dewatered solids with higher solids content could have an indirect relationship with odorants production in dewatered solids and biosolids.

### 3.2.14 Number of Centrifuges in Services

DC Water employs high speed centrifuges in the dewatering processes to separate water from solids. There are 14 centrifuges in total, two sides of centrifuges located at the solid processing building (SPB), west side (Sharples) and east side (Westfalia),

and each side has seven centrifuges. Each day the solids supervisor decides how many centrifuges will be run based on the inflow of wastewater, the amount of sludge in blending tank, and sludge blanket depth. However, if there are some issues concerning the machine operations, operators have to shut off centrifuges. For this reason, backup of sludge can occur in the processes by increasing retention time, and then can consequently cause high odorants emission in dewatered solids and biosolids.

To make sure that there are a sufficient number of centrifuges in service and better in dewatering performance, it is significant for DC Water to improve biosolids management program.

#### 3.2.15 Total Flow of Dewatered Sludge

Referring to Section 3.2.14, the total flow from the blending tank sent to centrifuges for dewatering on each day is called total flow of dewatered sludge. This variable should relate to other upstream processes especially flow of total primary sludge (TPS) and flow of total waste activated sludge (TWAS). It is important to manage the number of centrifuges regarding sludge volume.

#### 3.2.16 Percent of Lime Feeder

DC Water employs alkaline stabilization by using quick lime (CaO) powder, which is the production of limestone in rotary at high temperature (2000-2400 °F), and feeding in lime mixer to stabilize dewatered solids to become “Class B” biosolids (Rudolfs 1995). The percent of lime feeder is the speed of an auger feeding to dewatered solids in the lime mixer. Since one of the purposes of lime stabilization is to suppress total



reduced sulfur (TRS) in limed biosolids, this study considers that the higher the lime feeding speed, the less the TRS emission. However, increasing lime feeding speed depends on many factors, such as the required minimum pH (12), the number of centrifuges in service, the flow of dewatered solids, the solids content in cake solids, etc. We, therefore, anticipate that some of these process variables may be one of the factors mitigating odorants production in limed biosolids.

#### 3.2.17 Temperature of Limed Biosolids (Post limed biosolids temperature)

Post-lime temperature is byproduct of chemical reaction between quick lime (CaO) and water in dewatered solids (Rudolfs 1995; North 2003). A high temperature from lime stabilization can decrease microbial activities (North 2003). Many studies indicated that lime can raise the temperature of dewatered solids during and after lime stabilization to become limed biosolids (USEPA 2000a; Murthy et al. 2001; Kim et al. 2001; Murthy et al. 2002a; Murthy et al. 2002b; Kim et al. 2002; Novak et al. 2002; North 2003; Kim et al. 2005). According to Kim et al. (2002), lime stabilization can increase the temperature up to 104 °F and also raise pH up to 12.5.

To monitor and record temperature of limed biosolids, this research uses infrared (IR) sensors to achieve those two goals. We anticipate that temperature of limed biosolids relates to odorants production in biosolids. Nonetheless, not only can lime feeding speed cause higher temperatures in limed biosolids, but so other process variables. Therefore, this study has to identify the causes of those variables for increasing the temperature.

### 3.2.18 Mixed Liquor Suspended Solids (MLSS)

A mixed liquor suspended solid (MLSS) is generally defined as a concentration of suspended solids in a mixture raw activated sludge or settle wastewater in an aeration basin. The MLSS is normally stated in milligrams per liter (mg/L). We will employ interaction of concentration of MLSS with secondary blanket levels to find the relationship between the interaction and odorous production in solids and biosolids.

### 3.2.19 Ambient Temperature

The ambient temperature is the daily temperature at DC Water monitored by Reagan National Airport, which is located across the Potomac River from DC Water.

Referring to Sections 3.2.12 and 3.2.17, the ambient temperature is believed to have a high correlation with temperature at secondary effluent and post lime biosolids temperature. Additionally, we anticipate that ambient temperature can assist to suppress the odor production during the cooler season (fall, winter, and early spring) at the wastewater treatment plant.

## 3.3 Odor Determination and Experimental Design

The appropriate odor determination and proper design are the first and foremost steps in empirical studies. These can provide a better understanding and reliable information for both explaining the causes or sources of total reduced sulfur (TRS) and nitrogen (N)-containing compounds and for predicting those odorants. This section provides information about odor determination and the design used in this research.

### 3.3.1 The Importance of Odor Sample Collection

Odor sample collection is the most crucial part for this research. In order to obtain correct analysis of the cause or the source of odorants production in dewatered solids and biosolids from the upstream of operation facilities, we must have an appropriate odor collecting tool and proper sampling methods relating to the type of odor released from limed biosolids. Because this research focuses on the two main components, total reduced sulfur (TRS) and nitrogen containing compounds (N-containing compounds), it is necessary to find a suitable gas detector and proper measurement techniques to collect odors data because the physical and chemical properties dominate odorant emissions (Hentz 1997; USEPA 2000a). To prevent issues causing the error in data collection, real-time electro-chemical monitoring is being used in this research.

### 3.3.2 Real-Time Solids Odor Monitoring Systems

In the last decade, many studies tried to detect, identify, and quantify the odorous compound concentrations of biosolids by using a non-technical system (e.g., analytical measurement, chemical analyzers, and instrument). This study, however, employs a technical system, real-time solids odor monitoring systems. They are among the most valuable designs and systems that wastewater industry can apply in their wastewater processing systems. There are two types of gas detectors, OdaLog TRS Gas Loggers in conjunction with a dedicated sampling system (low-range sampling system-2 (LRSS-2)), available from Detection Instruments Corporation, Phoenix, AZ and Polytron 7000, available from Drager, are used to detect TRS and N-containing compounds, respectively. Both gas detectors, LRSS-2 on the left and

Polytron 7000 on the right, can be seen in Figure 3.3. By using a real-time approach via a wireless system (Octan-R), which is the best recently available technology, they can provide reliable odorants data in actual wastewater operational processes.

The LRSS-2 is a sampling system used to facilitate data collection from odor control systems. The system works with two OdaLog CEM 51s. It includes a heater that heats the sample stream to assist in removing condensable moisture. This heater can prevent changing in chemical characteristics of biosolids odor from moisture. Two external moisture removal traps further protect the OdaLogs from over-exposure to moisture. A dual-headed pump draws a sample from negative pressure situations, and delivers the sample at the proper flow rate in positive pressure situations to the OdaLogs. The LRSS contains two OdaLog gas loggers (CEM51) that can detect TRS from 0.01 ppm up to 20 ppm. (Instruments 2005).

The Polytron 7000 can detect N-containing compounds from 0 ppm to 200 ppm by having a sensor inside the Polytron 7000. The sensor responds to all N-containing compounds including ammonia ( $\text{NH}_4$ ) and amines (Dräger 2005). The Polytron 7000 is placed inside the enclosure with a pump and moisture removal.



**Figure 3.3: Low Range Sampling System-2 (LRSS-2) and Polytron 7000**

Both devices send the 4-20 mA signals to the transmitter called the Octans-R, available from Detection Instruments Corp. The Detection Instruments' Octan-R can monitor up to eight 4-20 mA current loops as shown in Figure 3.4. The output of the Octans-R is transmitted to a central location by a wireless modem. Additionally, a relay output and an isolated, low accuracy, 4-20 mA output is provided. The brief functional information of Octans-R is described as below:

**Connection:**

**Power:** A wall mount, 115 Volt input power supply connects to the Octan-R through a front panel barrel type connector. The power supply provides a regulated +15 Volts dc at a maximum of 2 amps.

**4-20 mA Inputs:** The current loop input connections are made through 6 pin Amp CST100 series connectors on the back panel. In addition, these connectors provide 5 Volts dc. Inputs must be loaded starting at position 1 and should not skip any connectors between #1 and the last one used. (The last connector used defines the "Top Channel".)

**4-20 mA Output:** A two pin Amp CST100 series connector provides an isolated, non-powered, 4-20 mA output. The connector is located on back panel. The 4-20 mA output pins are not polarity sensitive.

**Relay Output:**

A three pin Amp CST100 series connector provides connection to the output of one form C (SPDT) dry contact relay. The connector is located on the front panel.

**Setup Switches:**

Fourteen setup switches are located on the front panel. These are grouped on one 6 switch assembly (S1-1 through S1-6) and one 8 switch assembly (S2-1 through S2-8).

The switch functions are: (NOTE Switch is “ON” when lever is down.)

S1-1 Network Address 10

S1-2 Network Address 20

S1-3 Network Address 40

Network Address is the total of the address of the switches which are ON.

S1-4 Frequency Hopping Channel 1

S1-5 Frequency Hopping Channel 2

S1-6 Frequency Hopping Channel 4

Frequency Hopping Channel is the sum of the switches which are ON.

All switches (S1-4 through S1-6) ON selects the calibrate mode.

S2-1 Master Reset

S2-2 Load New Firmware Program

S2-3 Range Test (Echo received data)

S2-4 DiCom Communications Mode.

S2-5 Communications Retry Mode (Must match the other network equipment).

S2-6 Top Channel + 1

S2-7 Top Channel + 2

S2-8 Top Channel + 4

The top channel starts at 1 (S2-6 through SW2-8) OFF. Those three switches ON yields 8 active channels. The unit will sequentially scan channel 1 through the top channel.

### **Indicators:**

Front panel LED indicators are:

Green: Indicates power. It flickers when transmitting.

Amber: Lights when commands are received.

Red: One red led per channel, indicates when the specific channel is read.

### **Communication Modes:**

**DiCom Mode:** This mode simulates a DiCom module and works with any software which supports the DiCom. In this mode, the data for each channel is requested by the central computer as a separate transaction. In this mode, the network address selected by the dip switches is modified by adding the channel number to the address. i.e.: If the switches select a network address of 50, the actual data addresses are 51 (for channel 1), 52 (for channel 2)... 58 (for channel 8). Only those addresses which correspond to active channels are implemented. I.e. if the top channel is #4, then the active addresses would be 51, 52, 53, and 54.

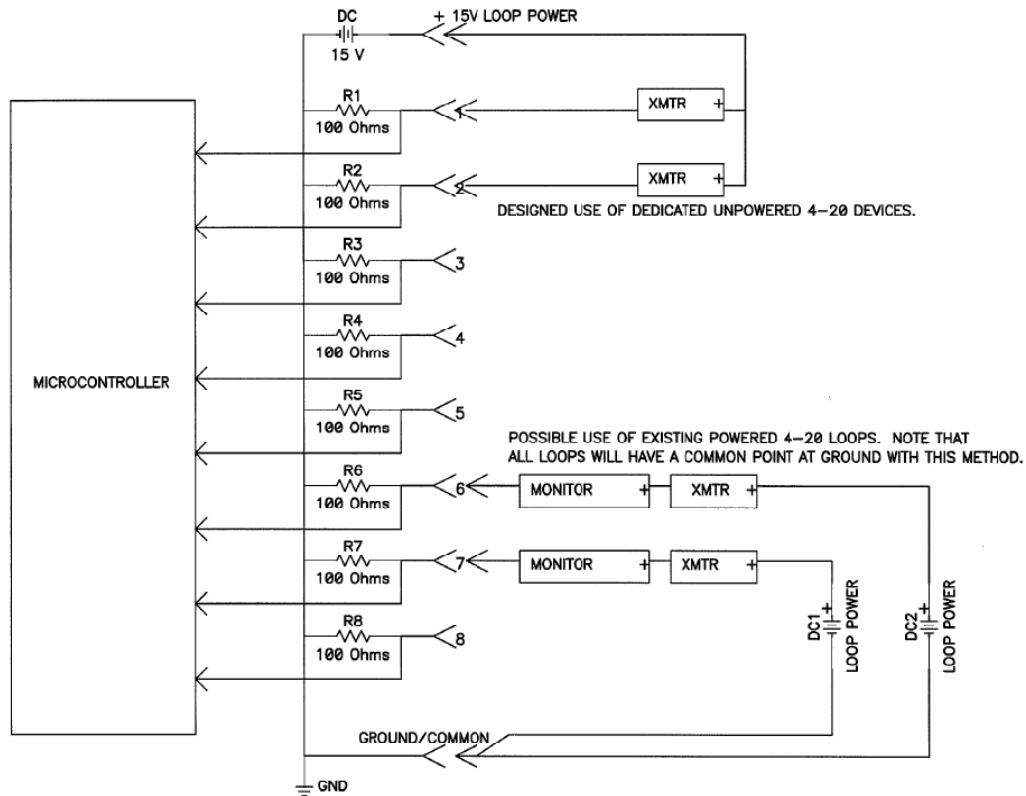
**Calibration:**

The eight 4-20 mA inputs can be calibrated with the aid of a 4.00 and 20.0 loop standard (accurate un-powered transmitter) or process meter. The calibration constants will be stored in non volatile memory.

The procedure:

- Attach a cable to the standard, or process meter that is provided with a connector that matches the 4-20 connectors on the Octan-R.
- Set a wireless modem to Network 0 and Frequency Hopping 0 by
  - Cycle power to modem on then off with S1 & 2 both ON (UP)
  - Set S2 OFF and turn on power.
  - or-
  - Use Modem AT commands.
- Connect a terminal to the modem using 9600, N, 8, 1 parameters.
- Place Octans S1-4, 5, and 6 all ON. (ON is down)
- The “top channel” selected by S2-6, 7, and 8 will be calibrated.
- Press and release the Reset Switch (S1-1)
- The corresponding red LED will slowly flash.





**Figure 3.4 Schematic of current loops in Octan-R**

The Octans –R transmits the 4-20 mA signal via a radio modem to a modem connected to a PC that imports the data into the Flexi Monitor software (available from Detection Instruments Corp). Using the Flexi Monitor software, it is possible to monitor up to 100 data nodes. A data node is defined as a single parameter at a single location. For instance, the temperature at a single point in a compost operation would be a node. Note that node numbers are assigned by the program as the unique identification of each node.

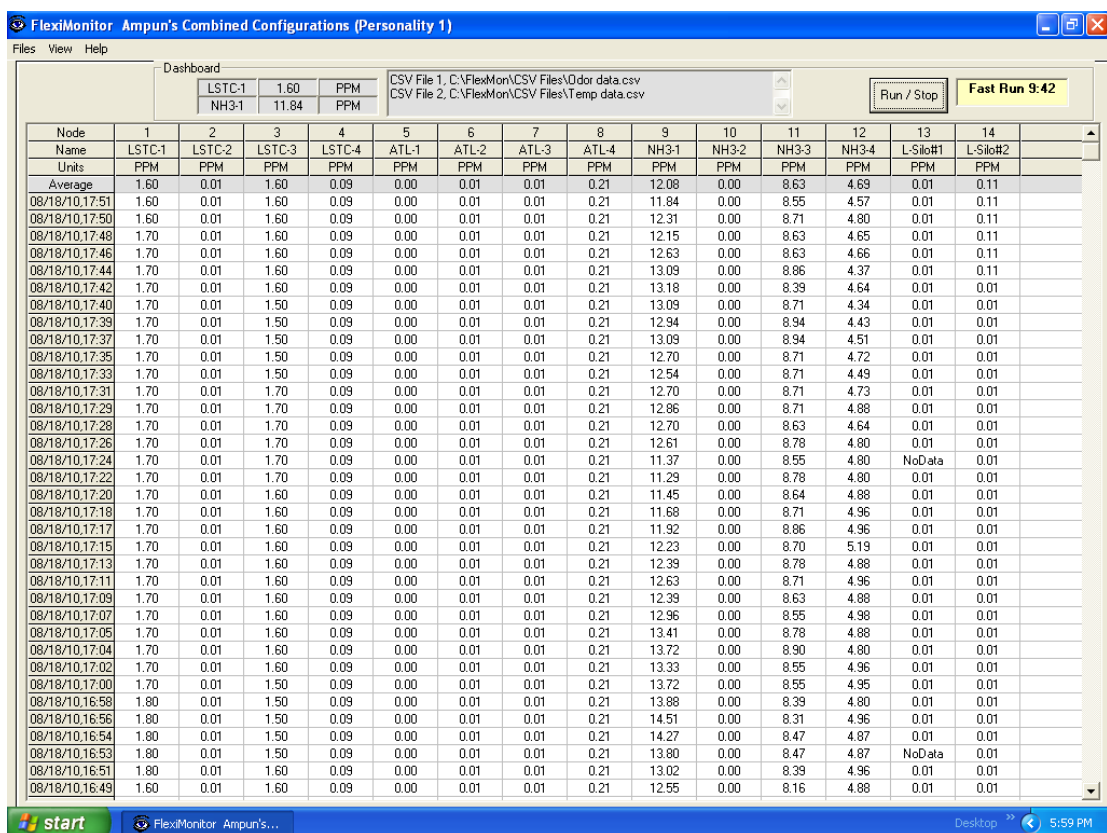
The user interface is organized into the following Menus, Pages, and windows:

- Files Menu, View Menu, and Help Menu
- Dashboard Window – Frequently needed controls and information.
- System Setup Page – Settings common to all nodes.

- Node Setup Page – Settings for each node.
- All Data Page – Last data for all nodes.
- Data Pages 1 through 10 – Current and some historic data for groups of nodes.
- Graph Page – Line graph of data from selected nodes.
- Notification Setup Page – enter name and phone number for alarm notification.
- Alarms Page – enter alarm notification information and alarm recipient information. Utility Page – Service and troubleshooting tools.

The Flexi Monitor software also allows for the generation and conversion of log files into an Excel spread sheet format by giving the site name and radio frequency. Alarm notification is provided by a dedicated phone line to any landlines and cell phones. The alarm information is user programmable and can notify recipients whenever there are high odor levels inside the solids processes. The example of the output from the Flexi Monitor is shown in Figure 3.5.

Additionally, this research attempts to monitor the temperature of lime biosolids (post lime temperature) by using an infrared temperature sensor (IR). It can monitor the temperature from 0 °F to 999 °F. This device also works with the Octan-R and the Flexi Monitor software.



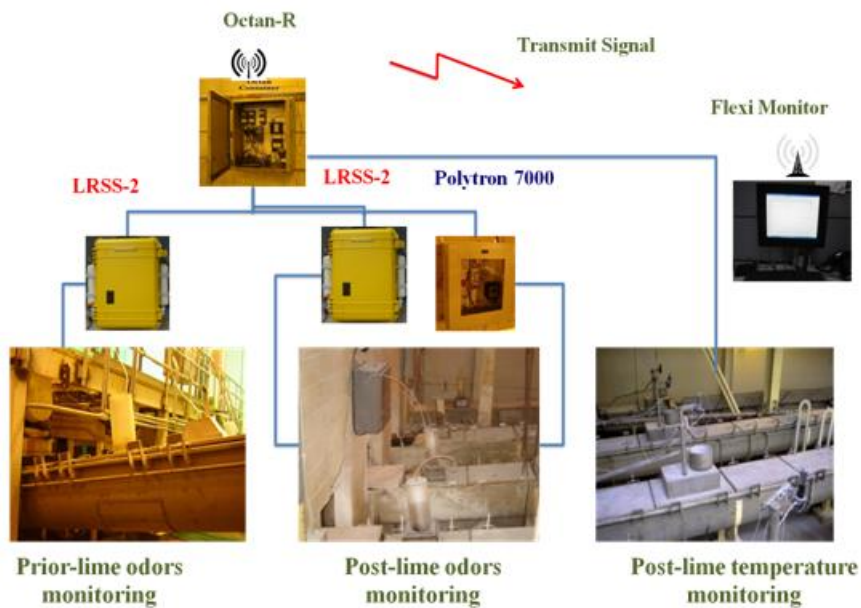
**Figure 3.5: Odorants concentration output from the Master Monitor**

### 3.3.3 Design and Odor Monitoring Location

As described in Sections 3.1 and 3.2, there are two sides of centrifuges located at the solid process facility, west side called Sharples and east side called Westfalia, and each side has seven centrifuges. At the dewatering process area, however, there are four conveyors (trains) used to transfer dewatered solids to the lime stabilization and the biosolids load-out facility (bunkers). Most of the time, operators will run two conveyors and have another two for backup.

For this research project we installed two LRSS-2 Sampling Systems at all four conveyors at the location right after the dewatering process by using two LRSS (one LRSS has two CEM 51 so each CEM51 for each conveyor) (Janpengpen 2006).

Another location is at post lime additions. At the post lime location, we installed two LRSS and two enclosures of Polytron 7000 (one enclosure has two Polytron 7000) for monitoring TRS and N-containing compounds, respectively. We also installed a temperature infrared sensor (IR) to monitor the temperature of lime-stabilized biosolids at the post lime location. All devices work with 4-20 mA cable and send the signal to the Octan-R. The Octan-R sends the 4-20 mA signal via a radio modem, to a modem connected to a PC for importation into the Flexi Monitor software. The completed installation diagram can be seen in Figure 3.6 and Figure 3.7.



**Figure 3.6: Real-Time solids odors monitoring systems**

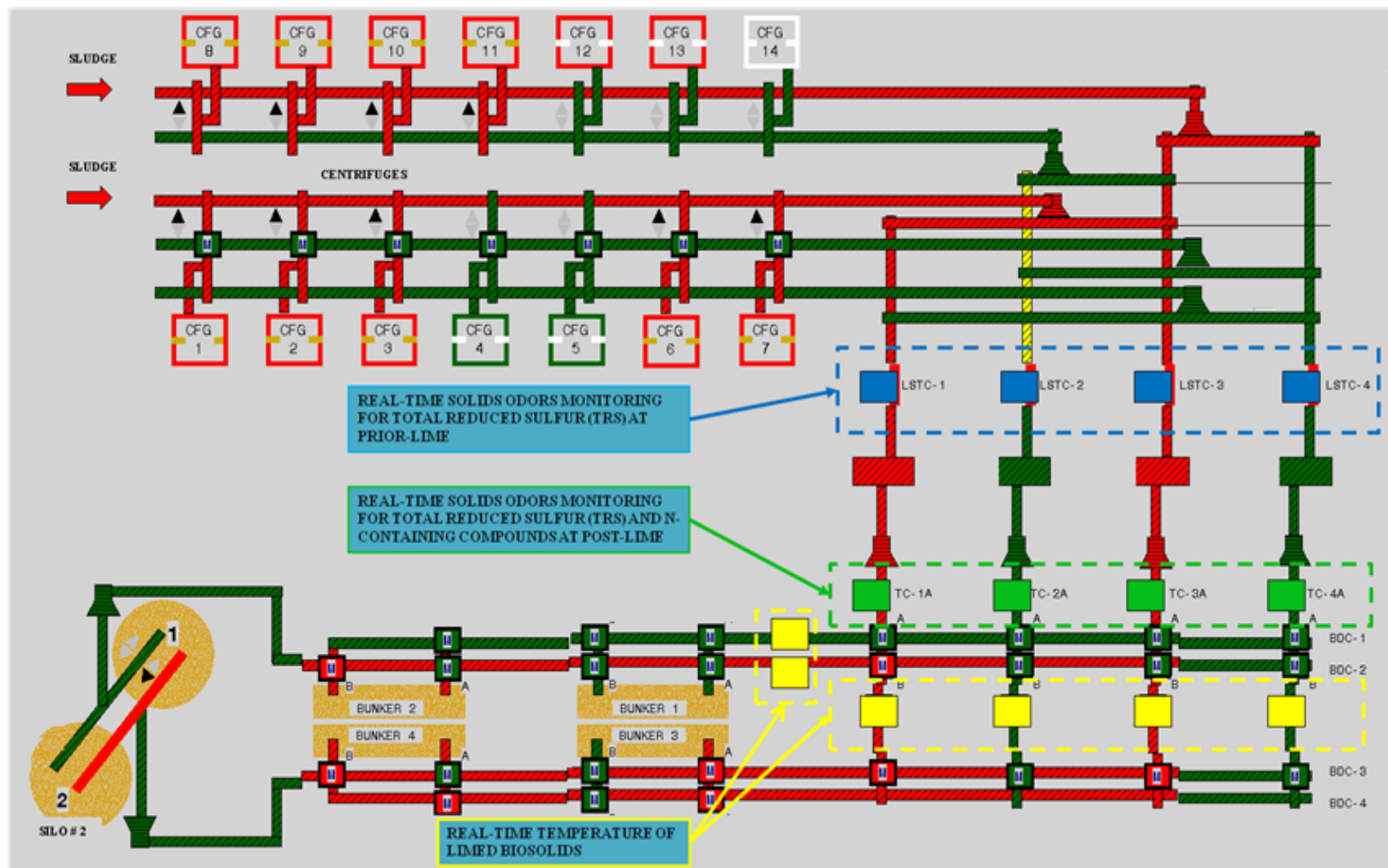
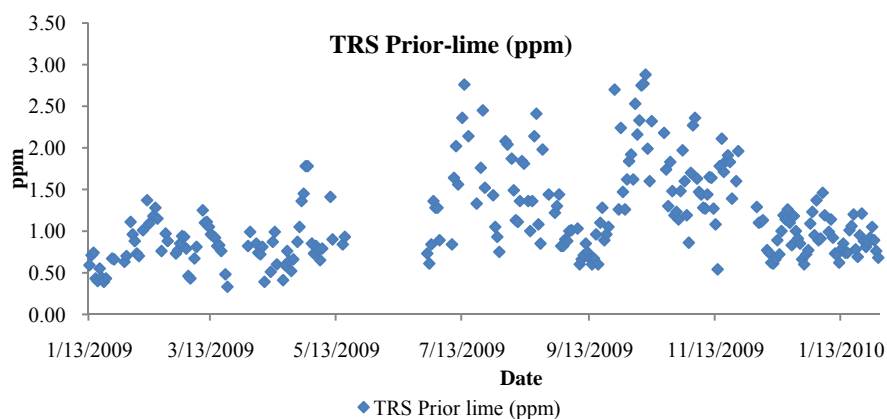


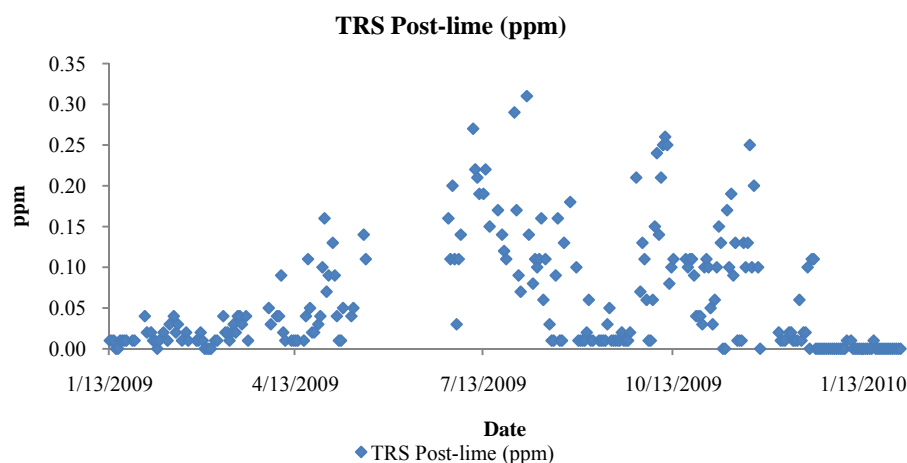
Figure 3.7: Location and devices setup inside solids processing building (SPB)

### 3.3.4 Odors Data Collection of Dewatered Solids and Biosolids

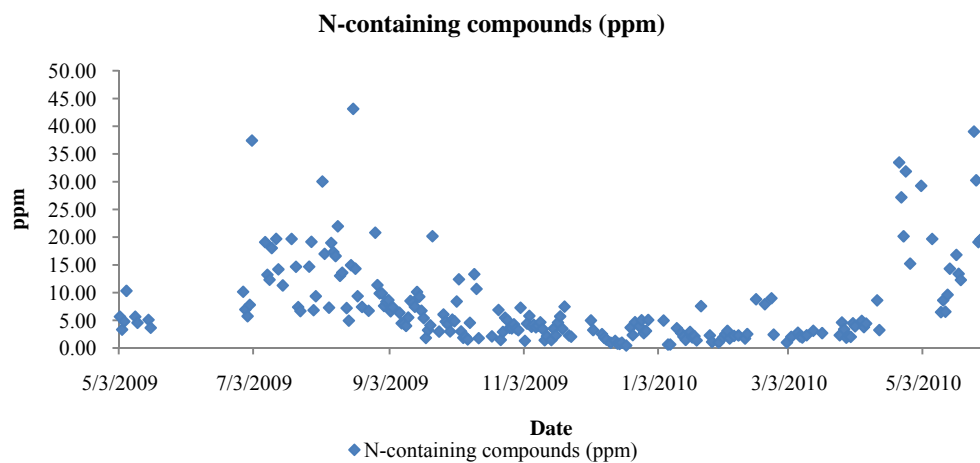
The two odorants compounds (dependent variable), total reduced sulfur (TRS) and nitrogen (N)-containing compounds, were monitored and detected simultaneously with operational process variables (January 2009 to May 2010). Only TRS, however, was monitored during January 2009 to April 2009. TRS and N-containing compounds are measured as odor volume unit, parts per million (ppm). The detailed information of those three odorants compounds can be found in Figures 3.8, 3.9, and 3.10.



**Figure 3.8: Total reduced sulfur (TRS) (ppm) at prior-lime process**



**Figure 3.9: Total reduced sulfur (TRS) (ppm) at post-lime process**



**Figure 3.10: Nitrogen (N)-containing compounds (ppm)**

## Chapter 4: Statistical Data Analysis for Biosolids Odors

In this chapter, statistical models are used to: (i) explain the sources of total reduced sulfur (TRS) and nitrogen (N)-containing compounds; and (ii) forecast those odorants levels before transporting to land application sites. The independent variables are 38 upstream process variables ( $X_1, X_2, \dots, X_{38}$ ) and the dependent variables are TRS at prior lime addition ( $Y_1$ ), TRS at the post lime addition ( $Y_2$ ), and N-containing compounds at the post lime addition ( $Y_3$ ).

To accomplish these two missions, modeling purposes are defined and then exploratory data analyses (EDA) including data summaries data visualization and correlation analysis are used. In the modeling processes, this study employs ordinary least squares (OLS) estimation and instrumental variables (IV) coupled with two-stage least squares (2SLS) estimation to analyze and to find the best explanatory and predictive models.

### 4.1 Defining Modeling Purposes

According to the stated goals and objectives, this study tries to identify and explain the sources of odorant production in dewatered solids and biosolids. In other words, we would like to explain the occurrence of odorant production in biosolids caused by the upstream process variables and also forecast the level of those odorants compounds. Empirical data analyses are used for two statistical modeling purposes: (i) modeling for explanatory purpose and (ii) modeling for predictive purposes. As a result, two types of models (an explanatory model and a predictive model) are used for each odorant compound (TRS at prior lime addition, TRS at post lime addition,



and N-containing compounds at post lime addition). Nonetheless, an explanatory model can be simultaneously used for explanatory and predictive purposes if it can offer sufficient predictive power. The details of this statement will be described more in Sections 4.1.1 and 4.2.

To establish explanatory models and predictive models, it is necessary to distinguish between those two modeling purposes (Shmueli 2010a; Shmueli and Koppius 2010). While explanatory modeling is used to elucidate the cause of occurrences or phenomena and to test a causal hypothesis, predictive modeling is used to forecast upcoming or new observations with of high precision (Gregor 2006; Shmueli and Koppius 2007; Wang et al. 2008; Shmueli and Koppius 2009; Shmueli and Koppius 2010; Shmueli 2010a; Shmueli 2010b).

To develop and get the best of those two types of models, the ordinary least squares (OLS) estimation and instrumental variable (IV) method are employed to create a model for explaining (explanatory model) the causes or sources of those three compounds. Separately, a predictive model is built for forecasting the level of those three compounds.

#### 4.1.1 Model Requirements

The model for explaining and forecasting will employ different prioritized requirements. The following statements are requirements used to select the best explanatory model in this research.

**The model with the highest adjusted  $R^2$  and the smallest standard error of estimate (SEE)**

The adjusted  $R^2$  is a modification of  $R^2$ .  $R^2$ , the coefficient of determination, can be defined as the percent of the variance in the dependent (response) variable explained exclusively or mutually by the independent (regressor) variables. In other words,  $R^2$  mirrors the number of errors made when using the regression model to estimate the value of the response variable, or  $R^2$  is the “explanatory power.” However, the problem arises when more independent variables are added in the model increasing the value of  $R^2$ . In addition,  $R^2$  does not appropriately justify the degree of freedom. To overcome this issue, the adjusted  $R^2$  can be replaced  $R^2$  (Kennedy 2008; Hyndman and Athanasopoulos 2010):

$$\text{Adjusted } R^2 = 1 - \left(1 - R^2\right) \left[ \frac{N - 1}{N - K - 1} \right] \quad (4.1)$$

where  $K$  is the number of independent (regressor) variables and  $N$  is number of observations. Many studies employed adjusted  $R^2$  value to be one of criteria to select the best statistical models (Vilalai 2003; Gabriel et al. 2005; Gabriel et al. 2006; Janpengpen 2006; Janpengpen et al. 2007; Vilalai 2008; Sekyiamah and Kim 2009). Maximizing adjusted  $R^2$  value means minimizing standard error of estimate (SEE) (Shmueli 2010a; Hyndman and Athanasopoulos 2010).

Therefore, the standard error of estimate (SEE) is a measure of forecasting accuracy for any models. In other words, it is an estimation of the dispersion of the prediction error (Winston 2004; Albright et al. 2004). For a sample size of size  $N$ , a model with  $K$  predictors, and a sum of squared errors equal to SSE, the SEE is given by:

$$SEE = \sqrt{\frac{SSE}{N - K - 1}} \quad (4.2)$$

Normally, the model with the smallest of *SEE* will be considered and selected because the *SEE* represents the level of precision of predictions gained from the regression equation. Therefore, the smaller magnitude of *SEE* is, the more accurate predictions tend to be (Albright et al. 2004).

### **All included variables are statistically significant**

The statistically significant for included variable is very important for the explanatory model. The significance level demonstrates the size of the rejection region of the null hypothesis ( $H_0$ ), which is the statement about the relation among parameters.

Therefore, the *p*-value is the smallest level of significance at which  $H_0$  would be rejected when a specified test procedure is used on a given data set. If the obtaining *p*-value is smaller than the significance level, it can be said that the  $H_0$  is false.

Choosing the level of significance is a subjective task. The conventional levels of significance are 0.01 (1%), 0.05 (5%), and 0.1(10%) (Devore 1987). For example, the significance level of 0.1, the largest significance level at which these analyses can carry out the test and fail to reject  $H_0$  is 10%. If any *p*-values are larger than 10% (such as at 15%),  $H_0$  cannot be rejected (fail to reject  $H_0$ ). If any *p*-values are smaller than or equal to 10% (such as 8%),  $H_0$  can be rejected and it is statistically significant. However, using level of significance can relate to Type I and Type II error. In general, Type I error is caused by incorrect rejecting  $H_0$  when it is true, and Type II error is caused by incorrect accepting  $H_0$  when it is false. To avoid misleading in model selection and interpretation, it is significant to employ the proper level of

significance. Therefore, these analyses consider a significance level of 0.1 as sufficient.

### **Correct sign of each independent variable coefficient**

The correct sign of each included variable is the most important requirement especially for explanatory modeling, which can be used to explain the effect of each independent variable on the dependent variable. While some independent variables might be statistically significant, if their coefficient signs do not follow the domain knowledge of environmental engineering or wastewater treatment theories, they should not be included in the model. As a result, the coefficient sign of each independent variable(s) in explanatory modeling must be explained with regard to the wastewater environment knowledge or the previous studies. A positive sign of the coefficient means that if that variable is increased, TRS or N-containing compounds increase as well. A negative sign for the coefficient means that if that variable increases, TRS and N-containing compounds in biosolids decrease.

Again, this research employs explanatory modeling to explain the relationship between the odorants production in biosolids and all upstream process variables. The explanatory model can grant insight into the significance of each upstream process variable in the model, and also provide the ranking of the importance of each upstream process variable by using the p-value and the coefficient sign (Shmueli and Koppius 2007; Shmueli and Koppius 2010; Shmueli 2010b).

To evaluate prediction performance, there are a number of criteria that can be used. We employ root mean squared error (RMSE), and Akaike's information criterion (AIC) to measure prediction performance and select the best predictive

model (Shmueli and Koppius 2007; Kennedy 2008; Shmueli and Koppius 2010; Shmueli 2010b). The RMSE is a measure of the differences between the predicted value of a model and the observed value for the thing being modeled (Albright et al. 2004; Shmueli et al. 2007), and the AIC is a measure of predictive accuracy of an estimated predictive model (Akaike 1974). In addition, these measures are computed from a holdout (validation) dataset that is obtained by partitioning the original data into training and validation sets. While the training set is used to fit the model, the holdout set is used for evaluating predictive power. The following measures are used to select the best predictive model (predictive performance) in this research.

**The model with the smallest prediction root mean squared error (RMSE)**

Prediction RMSE is one of the measures of predictive accuracy (Shmueli and Koppius 2007; Kennedy 2008; Shmueli 2010b). Prediction RMSE is similar to the standard error of estimate (SEE); however, it is calculated from the validation holdout dataset that was not used for estimating the model (Shmueli et al. 2007). Prediction RMSE is based on the differences between the predicted values of the observations in the holdout set ( $\hat{Y}$ ) and the observed values of those observations ( $Y$ ). For a holdout sample of size  $n$ , the prediction RMSE is given by:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{N}} \quad (4.3)$$

The RMSE value indicates the magnitude of the predictive power. For instance, if RMSE is 0.4, it means that the prediction errors from the model will have approximately a standard deviation of 0.4 in the units of the odorous compounds.

Therefore, the model with the smallest RMSE value is selected as the best predictive model.

### **The model with the lowest Akaike's Information Criterion (AIC)**

AIC is a measure of predictive accuracy of an estimated predictive model and is developed to approximate the predictive accuracy by forecasting new data (Akaike 1974; Foster 2002; Posada and Buckley 2004). It is a tool for model selection. AIC is a function of the maximized value of log likelihood ( $L$ ), the number of parameters ( $K$ ) (upstream process variables) in predictive model, and sum of squared error (SSE).

Assuming  $Y_1, Y_2, \dots, Y_n \sim N(\mu, \sigma^2)$ , the likelihood function is:

$$L(\mu, \sigma^2; Y_1, \dots, Y_n) = \prod_{i=1}^n f(Y_i | \mu, \sigma^2) \quad (4.4)$$

In general, the AIC is:

$$AIC = 2K - 2\ln(L) \quad (4.5)$$

$$\text{or } AIC = 2K + N[\ln(SSE)] \quad (4.7)$$

The magnitude of AIC value is used to rank the competing models and demonstrates that which one is the lowest among the models. In addition, the AIC value is a penalty for the model, which has more free parameters. Therefore, the best model is the model with the lowest AIC value and the minimum free parameter.

The model with the lowest AIC value is considered the best predictive model.

In sum, the model for explaining and forecasting will employ different prioritized requirements. Shmueli et al. (2007) stated that, “A good predictive model can have a looser fit to the data on which it is based, and a good explanatory model can have low prediction accuracy.” In other words, an explanatory model can be

created from the observed data and requires that all added variables have statistically significant and correct signs with respect to the domain knowledge of wastewater treatment theories, or previous studies. In contrast, a predictive model requires predictive accuracy, which is reflected by low RMSE and the low AIC. In addition, a good predictive model has to follow statistical theory that can make it not to fit the data, but an explanatory model can fit to the data better.

#### 4.2 Statistical Modeling Strategy

To generate an explanatory and a predictive model for biosolids odors, this research is divided into four tasks: (i) exploratory data analysis; (ii) ordinary least squares (OLS) estimation; (iii) instrumental variable (IV) estimation; (iv) model validation and selection. There are nine steps in the statistical modeling process for this research:

##### 4.2.1 Task I: Exploratory Data Analysis (Data Preparation) (Steps 1 to 3)

The analyses started monitoring both operational data and odorants data in January 2009 and completed the monitoring in May 2010. However, for odor monitoring, only TRS prior-lime and TRS post-lime addition were considered in January 2009 to April 2009. Furthermore, these analyses collected only data on the date that the wastewater treatment processes were operated continuously within 24 hours.

Many studies demonstrated that ambient temperature or inflow temperature is the major factor for odor emission rates in biosolids (Vilalai 2003; Gabriel et al. 2005; Gabriel et al. 2006; Sekyiamah and Kim 2006; Vilalai 2008). By considering on the critical season affected by the ambient temperature, these analyses partition the data into three periods: (1) annual period; (2) summer period; (3) winter period.

Summary statistics (such as frequency, mean, median, maximum value, minimum value, and standard deviation) and Box plot are employed to gain essential information on each upstream operational process variable and odorant compound production in biosolids (total reduced sulfur (TRS) and nitrogen (N)-containing compounds).

The correlation analysis is employed to identify association between variables. In particular, these analyses would like to know what upstream operational process variables have a relationship with total reduced sulfur (TRS) and nitrogen (N)-containing compounds. In the predictive model, comprehending these relationships is functional because it can assist us in identifying essential variables to use (Myatt 2007).

#### 4.2.2 Task II: Ordinary Least Squares (OLS) Estimation (Steps 4 to 5)

To create the statistical models, this research applied standard OLS estimation via multiple linear regression analysis to create both an explanatory and a predictive model for total reduced sulfur (TRS) at prior-lime ( $Y_1$ ), TRS at post-lime ( $Y_2$ ), and nitrogen (N)-containing compounds ( $Y_3$ ). Nevertheless, this step first focuses on the explanatory model.

As described in Section 2.3.1, the OLS was applied via multiple linear regression analysis to create a predictive model for total reduced sulfur (TRS) at prior-lime ( $Y_1$ ), TRS at post-lime ( $Y_2$ ), and nitrogen (N)-containing compounds ( $Y_3$ ). To obtain the predictive model, however, the data are partitioned into two data sets: a training set for estimating the model and a holdout (validation) set for evaluating the performance of this model on unobserved data (Shmueli et al. 2007).



#### 4.2.3 Task III: Instrumental Variable (IV) Estimation (Steps 6 to 7)

As described in Section 2.3.2, the Hausman test can be used to identify possible endogeneity between the post lime temperature variable and the error term of TRS post-lime regression equation (see Equations 2.9 and 2.10) (Wooldridge 2000; Wooldridge 2002; Greene 2003; Kennedy 2008). Where endogeneity is indicated, this research employs IV estimation to overcome the issue. In addition, the IV provides information in the explanatory model of step 4 that there are additional process variables indirectly causing those two odorants compounds. The IV estimation and two-stage least squares (2SLS) are powerful techniques in the econometric area and can be applied to this research (Angrist and Imbens 1995; Angrist and Krueger 2001; Wooldridge 2002; Kennedy 2008).

#### 4.2.4 Task IV: Model Validation and Selection (Steps 8 to 9)

Model Validation:

Predictive model:

To evaluate the efficiency of predictive models, Shmueli et al. (2007) suggested employing an approach common in data mining. By splitting the observed data into two data sets, a training set and a holdout (validation) set, the validation set will be used to validate the model whether or not it can precisely forecast new observations. The model which yields the smallest predicted RMSE, is then selected as the model with highest predictive power. However, when sample sizes are too small (winter and summer periods) for partitioning, using predicted RMSE is not effective validation. To overcome this issue, cross-validation analysis is employed.

*Cross validation (CV)* is a statistical validation strategy. It is a statistical procedure mainly used to assess predictive power or how precisely a predictive model can achieve it (Michaelson 1987; Hastie et al. 2008; Shmueli 2010a; Shmueli 2010b; Shmueli and Koppius 2010). Hastie et al. (2008) suggested that cross-validation is a method used to estimate prediction error. One of the simplicity methods is *K-fold* cross-validation.

The ideas of *K-fold* are to divide data evenly into  $K$  subsets. Typically,  $K$  can be either 5 or 10 subsets (Hastie et al. 2008). This research employs a 5-fold cross-validation ( $K=5$ ). Figure 4.1 illustrates how data are divided and used in cross-validation procedures.

1	2	3	4	5
validation	training	training	training	training

**Figure 4.1: *K-fold* cross-validation (Hastie et al. 2008)**

The following steps explain how to validate the models by using *K-fold* cross-validation:

1. Split data into five subsets by having one of them as a validation set (such as data set #1) and the remaining as training sets (similar to training set and validation sets in predictive models estimation).
2. Fit the model to the four data sets (data sets #2 to #5 ). At this stage, there will be four models with different coefficients from the original model that would like to be tested. In addition, the prediction error will be calculated.
3. Repeatedly fit models by having data set # 2 as the validation set, and the rest serve as training sets until having data set # 5 as validation set. At that point, there are

prediction errors of five models in cross-validation to compare with the prediction errors of the original model.

4. The RMSE will be employed for validation statistics in cross-validation. If an average RMSE of five models is lower or close to the RMSE of original model, the original is said to be validated. As a result, an explanatory model can be used for prediction. In other words, an explanatory model can offer both explanatory purpose and predictive purpose at the same time. In contrast, if an average RMSE of cross-validation is higher than the RMSE of original model, it will indicate that an explanatory model is not the best predictive model.

Explanatory model:

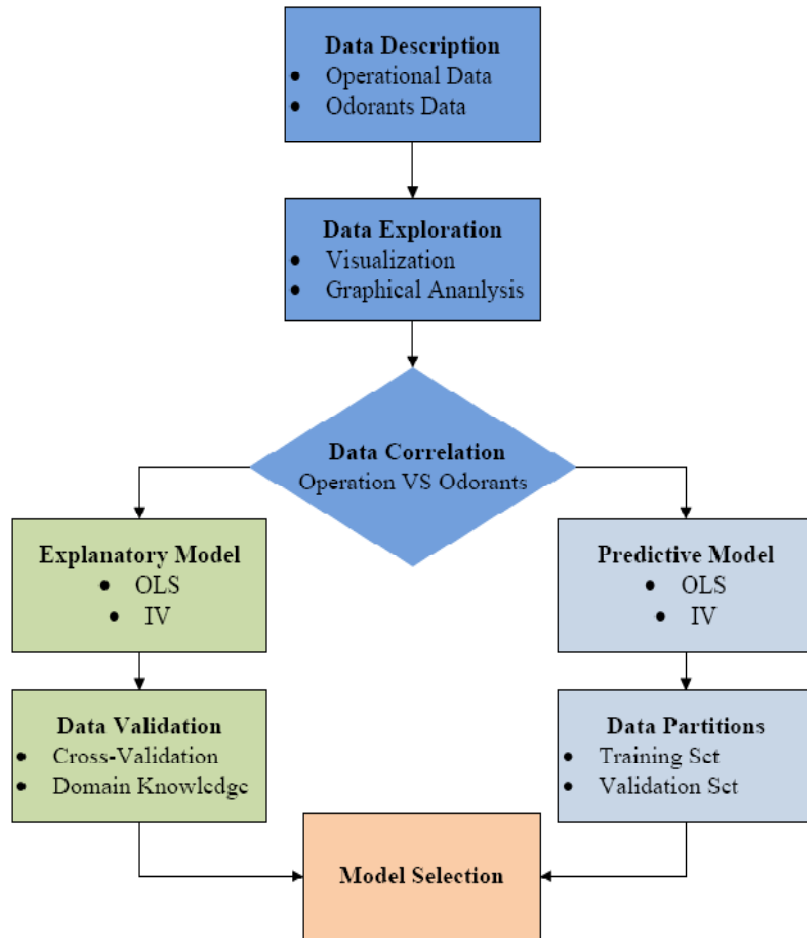
As described in Section 4.1.1, the criteria to obtain the best explanatory model for each period include: (1) a correct sign of each regression coefficient included in the model; (2) statistical significance of each independent variable included in the model; and (4) the standard error of the estimate (SEE). Although two of these criteria (a correct sign of each regression coefficient and statistical significance of each independent variable included in the model) are adequately considered as applicable information to validate the explanatory model, but this research still needs to access or measure the predictive power of the explanatory models, whether or not they can be used for prediction (Shmueli 2010a; Shmueli 2010b; Shmueli and Koppius 2010). Therefore, *K-fold* cross-validation is employed to measure and report the predictive power.

Model Selection:

To select the best model with different sets of independent variables, some studies choose the model with the lowest value of standard error of estimate (SEE) (Winston 2004). Several studies suggested adjusted  $R^2$ , the correct sign of each regression coefficient regarding the wastewater environment and the  $p$ -value to identify the best equation in a regression model (Vilalai 2003; Sekyiamah and Kim 2006; Gabriel et al. 2006; Janpengpen 2006; Janpengpen et al. 2007; Shmueli et al. 2007; Janpengpen et al. 2008; Vilalai 2008; Sekyiamah and Kim 2009).

Therefore, this research employs the standard error of the estimate (SEE),  $p$ -value of 0.1 for each independent variable, and corrects sign of each regression coefficient regarding the wastewater environment knowledge to identify the best explanatory model. Additionally, the root mean squared error (RMSE) and the Akaike's information criterion (AIC) are employed to identify the best predictive model.

The following diagram summarizes the procedures for the statistical steps in data analysis and modeling.



**Figure 4.2: Statistical steps in data analysis and modeling process**

#### 4.3 Exploratory Data Analysis (EDA) (Data Preparation)

Exploratory data analysis requires data preparation. This preparation is the first and foremost step in any data analysis (Tukey 1970; Tukey 1977). Furthermore, this process will ensure that the raw data have quality and are ready to analyze (Dasu and Johnson 2003). Performing statistical analysis without exploring and knowing the raw data can be misleading in the data interpretation and misleading in determining the logical relationships among data elements. Using exploratory data analysis can assist these analyses learn and gain value information about the raw data prior to analyzing

them. The key is that exploratory analysis can be another tool used to identify the relationship between upstream process variables and odorants production in biosolids.

As stated in Chapter 3, this section will offer data summaries (summary statistics of all variables), data visualizations (Box plots), and correlation analysis (visualizing relationships (graphs) and correlation coefficients ( $r$ )). For data displays and visualizing relationships, there will be only illustrations of some data in this chapter. The remaining plots are shown in Appendices A-1 and A-2.

#### 4.3.1 Data Summaries

Data summaries indicate summary statistics of each independent variable (process variables data) and each dependent variable (odorants data). Summary statistics data includes frequency (count), central tendency measures (mean and median), and measures of variability of the data (max, min, and standard deviation). Additionally, this section also provides variables information including abbreviation, unit, and a full description. Table 4.1 and 4.2 present variable descriptions and summary statistics of all variables, respectively.

**Table 4.1 Variables description**

Abbreviation	Unit	Full description
TRS Pri	ppm	Total Reduced Sulfur at Prior Lime Addition
TRS Post	ppm	Total Reduced Sulfur at Post Lime Addition
N-Con Comp	ppm	Nitrogen-Containing Compounds at Post Lime Addition
Cake % S	%	Percent of Dewatered Cake solids
TFW DWSL	gal	Total Flow of Dewatered Sludge
% Lime feed	%	Percent of lime feeder speed
Post limed temp	°F	Temperature of limed biosolids

**Table 4.1 Variables description (cont)**

Abbreviation	Unit	Full description
Amb Temp	°F	Ambient temperature (outside temperature)
FeCl <sub>3</sub>	gal	Ferric Chloride Addition
WPL	gal	Waste Pickle Liquid Addition
Pol_Pri	lbs/day	Polymer Addition at Primary Process
Pol_Sec	lbs/day	Polymer Addition at Secondary Process
Pol_DAF	lbs/day	Polymer Addition at Dissolved Air Flotation
Pol_DW	lbs/day	Polymer Addition at Dewatering
PolDAF+PolDW	lbs/day	Sum of Polymer at DAF and at Dewatering process
Pol_DW/DT solids	lbs/day	Polymer Addition at Dewatering Per Dry Tons of Solids
GT FW	gal	Gravity Thickener Flow
GT TS	lbs/day	Gravity Thickener Total Solids
GT %S	%	Percent Solids of Gravity Thickener
SE.BL	ft	Secondary East Blanket Depth
SWO.BL	ft	Secondary West Odd Blanket Depth
SWE.BL	ft	Secondary West Even Blanket Depth
WAS %S	%	Percent Solids of Waste Activated Sludge
DAF FW	gal	Dissolved Air Flotation Flow
DAF TS	lbs/day	Dissolved Air Flotation Total Solids
Temp	°F	Temperature at Secondary Effluent
ISV.E	ft/min	East Initial Settling Velocity
ISV.WE	ft/min	West Even Initial Settling Velocity
ISV.WO	ft/min	West Odd Initial Settling Velocity
SSV <sub>60</sub> E	ml/L	East Settled Sludge Volume
SSV <sub>60</sub> WE	ml/L	West East Settled Sludge Volume
SSV <sub>60</sub> WO	ml/L	West Odd Settled Sludge Volume

**Table 4.1 Variables description (cont)**

Abbreviation	Unit	Full description
RAS.SE	mg/L	East Secondary Return Activated Sludge
RAS.SWO	mg/L	West Odd Secondary Return Activated Sludge
RAS.SWE	mg/L	West Even Secondary Return Activated Sludge
MLSS	mg/L	Mixed Liquor Suspended Solids
SE.BL*RAS.SE	ftmg/L	Interaction between SE.BL and RAS.SE
SWE.BL*RAS.SWE	ftmg/L	Interaction between SWE.BL and RAS.SWE
SWO.BL*RAS.SWO	ftmg/L	Interaction between SWO.BL and RAS.SWO
SE.BL*MLSS.SE	ftmg/L	Interaction between SE.BL and MLSS.SE
SWE.BL*MLSS.SWE	ftmg/L	Interaction between SWE.BL and MLSS.SWE
SWO.BL*MLSS.SWO	ftmg/L	Interaction between SWO.BL and MLSS.SWE

**Table 4.2 Summary statistics**

Variables	Count	Max	Min	Mean	Median	Std. Dev.
TRS Pri	259	2.88	0.33	1.15	1	0.53
TRS Post	259	0.31	0.00	0.06	0.02	0.07
N-ConComp	259	43.1	0.45	7.7	4.8	7.65
Pol_Pri	259	1726	569	1022.84	940	203
Pol_Sec	259	1339	736	1094.05	1077	117.43
Pol_DAF	259	3422	612	1453.3	1147	731.68
Pol_DW	259	6530	1358	3406.32	3346	803.73
PolDAF+PolDW	259	8066	2026	4853.64	4582	1190.35
Pol_DW/DTS	259	25.3	4.2	10.62	10.4	2.42
Temp	259	77	50	64.44	64	6.787
SSV <sub>60</sub> E	259	455	75	245.56	240	62.64
SSV <sub>60</sub> WE	259	343	94	196.06	183	56.4



**Table 4.2 Summary statistics (cont)**

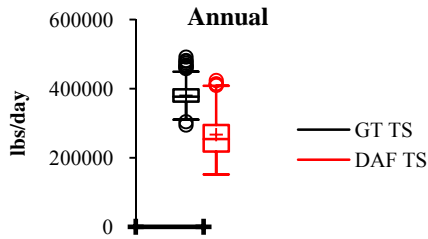
Variables	Count	Max	Min	Mean	Median	Std. Dev.
SSV <sub>60</sub> WO	259	458	141	230.98	214	69.96
RAS.SE	259	5307	2296	3479.48	3507	613.14
RAS.SWO	259	11900	2843	5622.69	5107	1897
RAS.SWE	259	11779	3390	6103.77	5494	2054.61
SE.BL*RAS.SE	259	14850	3830	8465.29	8425	2508.41
SWE.BL*RAS.SWE	259	47000	7682	17890	16464	8723.72
SWO.BL*RAS.SWO	259	54240	6635	17613	16094	8562.92
SE.BL*MLSS.SE	259	9874	2170	4622.20	4600	1453.87
SWE.BL*MLSS.SWE	259	19705	2711	7605.34	6443	3486.87
SWO.BL*MLSS.SWO	259	26731	2993	9379.96	8365	4301.38
Blend(Pri:Sec)	259	2.69	0.82	1.55	1.57	0.36
Cake %S	259	31.27	24.1	27.74	27.74	1.64
MLSS.SE	259	3657	1117	1900.73	1857	389.86
MLSS.SWE	259	4529	1291	2592.26	2526	790.43
MLSS.SWO	259	6026	1583	3002.63	2866	977.65
# Centf Serv	259	6	3	5	5	1
TFW DWSL	259	1068400	615400	827642	819552	104018.34
% Lime feed	259	18	9	13.47	13	1.6
Post limed temp	259	139.75	53.65	94.06	94.47	20.98
Amb Temp	259	87	15	54.62	55	17.5
FeCl <sub>3</sub> +WPL	259	63922	9652	22477.85	21678	6706.21
GT FW	259	1265453	852163	1134442	1138574	77738
GT TS	259	491434	294687	380182	376746	32106
GT %S	259	5.05	3.34	4.03	4	0.31
S.E.BL	259	3.4	1.6	2.41	2.4	0.47

**Table 4.2 Summary statistics (cont)**

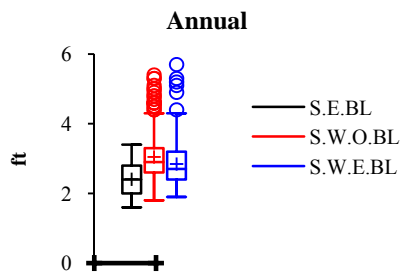
Variables	Count	Max	Min	Mean	Median	Std. Dev.
S.W.O.BL	259	5.4	1.8	3.04	2.9	0.74
S.W.E.BL	259	5.7	1.9	2.8	2.7	0.63
WAS %S	259	5.38	3.9	4.66	4.65	0.35
DAF FW	259	967292	400654	681792.12	647342	120001.62
DAF TS	259	423679	151761	267025.86	253987	62048.39
ISV.E	259	0.25	0.03	0.118	0.11	0.04
ISV.WE	259	0.33	0.060	0.173	0.18	0.06
ISV.WO	259	0.26	0.03	0.134	0.13	0.05

#### 4.3.2 Data Visualization

Data visualization is a useful and straightforward tool for representing data. It can also offer the different perspective information of each variable that numerical summaries cannot provide, such as elucidating features, measurement errors, and unreasonable data recording. This section, therefore, employs Box plots to explore each independent variable and each dependent variable. Figures 4.3 and 4.4 present examples of Box plots of annual gravity total solids (GT TS) and dissolved air flotation total solids (DAF TS), and secondary blanket level, respectively. Plots for each of the other independent variables (upstream process variables) and each dependent variable (odorous compounds) can be seen in Appendix A-1.



**Figure 4.3: Gravity total solids (GT TS) (lbs/day) and dissolved air flotation total solids (DAF TS) (lbs/day)**



**Figure 4.4: Secondary blanket level SE.BL (ft), SWO.BL (ft), and SWE.BL (ft)**

A Box plot is one of important tools in exploratory data analysis (EDA). It is used to demonstrate the shape of the distribution, its central value, and spread. The information consists of the most extreme values in the data set (i.e., maximum and minimum values), the lower and upper quartiles, and the median. In addition, the unusual observations (outliers) are shown in the Box plot.

According to the Box plots of Figures 4.3, it illustrates the mean of GT TS and DAF TS at 390,000 lbs/day and 280,000 lbs/day, respectively. The median and the mean of GT are almost the same level. The maximum and minimum values of GT TS are about 450,000 lbs/day and 310,000 lbs/day, respectively. However, the maximum and minimum values of DAF TS are 410,000 lbs/day and 150,000 lbs/day, respectively. There are outliers of GT TS above the maximum and below the minimum values, but DAF TS only has the outliers above the maximum value.

Box plot of Figure 4.4 shows information of the annual secondary blanket level at SE.BL, SWO.BL, and SWE.BL. The Box plot demonstrates that the mean of blanket level at both SWO.BL and SWE.BL are higher than SE.BL, and the values are 2.8 ft, 2.6 ft, and 2.4, respectively. The median and the mean of blanket level at SE.BL are at the same level, but the median of blanket level at SWO.BL and SWE.BL are 3.0 ft and 2.8 ft, respectively. The maximum and minimum blanket levels at SWO.BL are 4.4 ft and 1.8 ft, respectively. The maximum and minimum blanket levels at SWE.BL are 4.4 ft and 1.9 ft, respectively. However, the maximum and minimum of blanket levels at SE.BL (i.e., 3.5 ft and 1.6 ft, respectively) are lower than SWO.BL and SWE.BL. There are outliers above the maximum value of both SWO.BL and SWE.BL. However, there are no outliers on the SE.BL.

#### 4.3.3 Correlation Analysis

After gaining information from numerical summaries and data visualization, this research can start examining the relationships between odorants compounds and all upstream process variables. The most effective way to examine the relationship between a pair of variables is correlation analysis or association analysis.

Understanding the correlation between odorants production in biosolids and all upstream process variables is very useful for a number of reasons. For instance, in the predictive model, it can help to recognize which upstream process variables might be useful predictors.

To identify the relationship between dependent variable and independent variables, these analyses use two simple techniques correlation coefficients and graphical analysis. The correlation coefficient ( $r$ ) is a value indicating the strength of

the linear relationship between two variables (Albright et al. 2004; Winston 2004; Shmueli et al. 2007). This number measures the linear relationships and its value can range between -1 and 1 (Winston 2004; Myatt 2007; Shmueli et al. 2007). These analyses compute the correlation coefficient between the dependent variable and each of the independent variables. A positive sign of the correlation coefficient means that when the value of upstream process variable increases the odorant variable is likely to increase. In contrast, a negative sign of the correlation coefficient means that when the value of upstream process variable increases the odorant variable is likely to decrease.

When the value of the correlation coefficient is close to -1 or 1, it suggests that there is either a strong negative or positive correlation between the pair of variables. When the value of the correlation coefficient is close to 0, it is suggested that there is a weak correlation or no relationships between the pair of variables (Winston 2004; Myatt 2007). This research uses a statistical software package (Stat Tools 5.5 of Palisade Corporation) to run and generate the correlation coefficient ( $r$ ) between dependent and independent variables.

To decide what value of the correlation coefficient is significant and appropriate, this research employs the following rule of thumb at the 0.05 level of significance (Krehbiel 2004):

$$|r_{xy}| > 2/\sqrt{N} \quad (4.8)$$

where  $N$  is number of observations.

For instance, there are 250 data point in the annual period, so the value of the correlation coefficient to be considered as significance is  $2/\sqrt{259} = 0.12$ . In the

summer the value of the correlation coefficient to be considered as significance is  $2/\sqrt{70} = 0.24$ , and the value of the correlation coefficient to be considered as significance is  $2/\sqrt{50} = 0.28$ . Table 4.3, Table 4.4, and Table 4.5 show only the significant correlation coefficients ( $|r_{xy}| > 0.3$ ) between odorants production in biosolids and upstream process variables in each period.

**Table 4.3: Significant correlation coefficients between TRS prior-lime (ppm) and upstream processes**

<b>Annual</b>		<b>Summer</b>		<b>Winter</b>	
<b>Process Variables</b>	<b>TRS Prior lime</b>	<b>Process Variables</b>	<b>TRS Prior lime</b>	<b>Process Variables</b>	<b>TRS Prior lime</b>
WAS %S	-0.388	SWO.BL	0.373	Temp of Sec Effluent	0.498
DAF TS	-0.377	SWE.BL	0.583	Ambient Temp	0.658
Blend Ratio	0.317	RAS.SE	0.308	SWO.BL*MLSS	0.316
Temp at Sec Effluent	0.512	RAS.SWE	0.384		
Ambient Temp	0.439	RAS.SWO	0.44		
DAF FW	-0.320	Ambient	0.343		
		SWO.BL*RAS.SWO	0.46		
		SWE.BL*RAS.SWE	0.631		
		SWO.BL*MLSS.SWO	0.547		
		SWE.BL*MLSS.SWE	0.616		
		SE.BL*MLSS.SE	0.303		

**Table 4.4: Significant correlation coefficients between TRS at post lime and upstream processes**

<b>Annual</b>		<b>Summer</b>		<b>Winter</b>	
<b>Process Variables</b>	<b>TRS Post lime</b>	<b>Process Variables</b>	<b>TRS Post lime</b>	<b>Process Variables</b>	<b>TRS Post lime</b>
WAS %S	-0.416	SE.BL	0.334	WAS %S	-0.567
DAF TS	-0.394	SWE.BL	0.483	DAF TS	-0.309
Blend Ratio	0.385	RAS.SE	0.554	Blend Ratio	0.451
Temp at Sec Eff	0.544	RAS.SWE	0.372	Cake %S	0.423
Amb Temp	0.528	AmbTemp	0.306	Post Lime Temp	0.63
DAF FW	-0.336	SWE.BL*RAS.SWE	0.535	Temp of Sec Effluent	0.511
Post lime Temp	0.425	SE.BL*RAS.SE	0.474	Amb Temp	0.532
TRS Prior Lime	0.65	SWO.BL*MLSS	0.507	GT %S	0.422
		SWE.BL*MLSS	0.504	GT TS	0.331
		SE.BL*MLSS	0.586	TRS Prior	0.376
		GT %S	0.309		
		GT TS	0.316		
		TRS Prior lime	0.553		

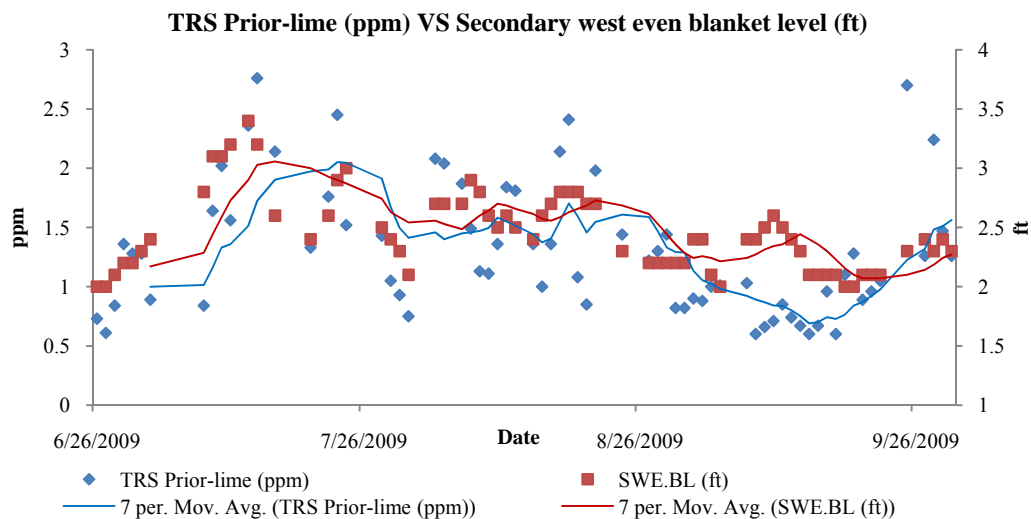
**Table 4.5: Significant correlation coefficients between N-containing compounds and upstream processes**

<b>Annual</b>		<b>Summer</b>		<b>Winter</b>	
<b>Process Variables</b>	<b>N-containing compounds</b>	<b>Process Variables</b>	<b>N-containing compounds</b>	<b>Process Variables</b>	<b>N-containing compounds</b>
Post lime Temp	0.396	WAS %S	-0.308	Post lime Temp	0.206
Pol_Sec	0.544	SWO.BL	0.339	RAS.SE	0.389
Temp at Sec Eff	0.383	SWE.BL	0.373	RAS.SWO	0.301
Amb Temp	0.516	Post lime Temp	0.243	Pol_DW/DT solids	0.296
		Pol_DW	0.344		
		Pol_DW/DT Solids	0.317		
		PolDW+PolDAF	0.343		
		SE.BL*RAS.SE	0.354		
		SWO.BL*MLSS	0.394		
		SE.BL*MLSS	0.354		
		Amb Temp	0.368		

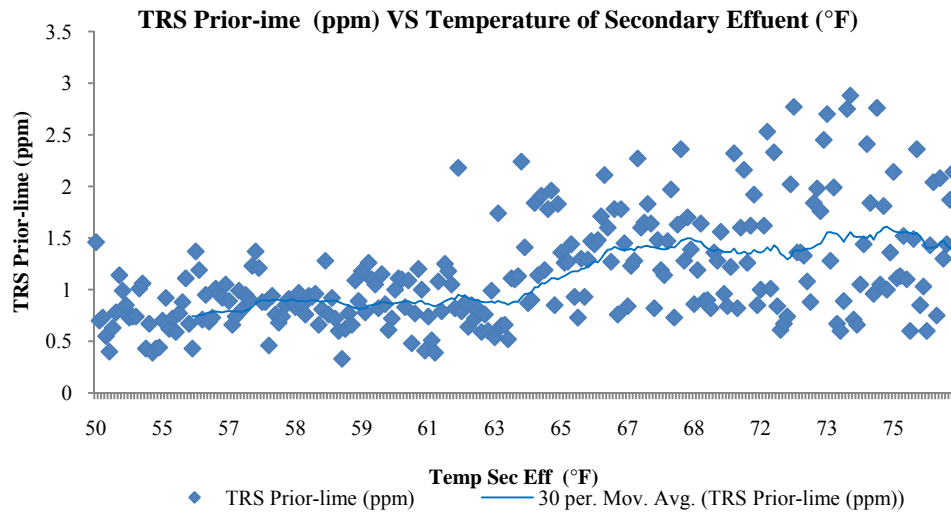


In order to observe which process variables have associations with total reduced sulfur (TRS) and nitrogen (N)-containing compounds, this research employs graphical analysis to compare with the correlation coefficients ( $r$ ). Graphical analysis is the straightforward way to find associations between odorant production compounds and each process variable by plotting two variables (TRS VS each process variable or N-containing compounds VS each process variable) in the same graph.

On the graph, there are two main axes, the X-axis, which shows the date of monitoring, and the Y-axis, which shows odorant compounds or process variables. The Y-axis consists of a left axis and a right axis, which are odorant compounds and process variables, respectively. In addition, this research also plots dependent variable on the Y-axis and each independent on the X-axis. Figures 4.5 and 4.6 illustrate the association between dependent variable and independent variable. Similar graphical analysis for the other variables can be seen in Appendix A-2.



**Figure 4.5: TRS Prior-lime (ppm) VS SWE.BL (ft) During Summer Period**



**Figure 4.6: TRS Prior-Lime (ppm)VS Temp Sec Eff (°F) During Annual Period**

The Figures 4.5 and 4.6 demonstrate the association between TRS prior-lime and two process variables (SWE.BL and Temp Sec Eff) during summer and annual periods, respectively. The Figure 4.5 indicates the strong association between TRS prior-lime and SWE.BL. When the SWE.BL increases, the TRS prior-lime tends to be increased. In addition, the correlation coefficient between these two variables is 0.583, so that the Figure 4.5 verifies the strong association between TRS prior-lime and SWE.BL.

The Figure 4.6 indicates that when the Temp Sec Eff increases, the TRS prior-lime tend to be increased. It also indicates that there is a linear relationship between TRS prior-lime and Temp Sec Eff. The correlation coefficient between TRS prior-lime and Temp Sec Eff is 0.512, so it confirms the association on Figure 4.6.

The most important information that this research learn from graphical analysis is the association between odorant compounds and process variables. Graphs can illustrate how odorant compounds change when process variables have been increased or decreased. Additionally, correlation coefficients ( $r$ ) are not enough to

verify the associations because more often some outliers can dominate the association and cause artificially high values for  $r$ .

#### 4.3.4 Summary Results of Exploratory Data Analysis

As this research stated earlier, knowing the correlation between odorant production in biosolids and all upstream process variables can assist to determine which upstream process variables are potentially significant and should be used in the explanatory and predictive models. Tables 4.6 to 4.14 show the upstream process variables considered to be essential in this research.

**Table 4.6: The essential process variables relating to TRS prior-lime (ppm) during annual**

	WAS %S	DAF TS	Blend	Cake %S	Temp Sec Eff	Ambient Temp	DAF FW	TRS Pri
Correlation Table	(%)	(lbs/day)	(Pri:Sec)	(%)	(F)	(F)	(gallons/day)	(ppm)
WAS %S	1.000							
DAF TS	0.763	1.000						
Blend(Pri:Sec)	-0.624	-0.892	1.000					
Cake %S	-0.498	-0.826	0.782	1.000				
Temp Sec Eff	-0.546	-0.629	0.492	0.667	1.000			
Ambient Temp	-0.497	-0.549	0.408	0.564	0.898	1.000		
DAF FW	0.551	0.958	-0.890	-0.856	-0.573	-0.487	1.000	
TRS Prior-lime	-0.388	-0.377	0.317	0.278	0.512	0.439	-0.320	1.000

**Table 4.7: The essential process variables relating to TRS post-lime (ppm) during annual**

	WAS %S	DAF TS	Blend	Cake %S	Post lime Temp	Temp(F)	Ambient Temp	DAF FW	TRS Post
Correlation Table	(%)	lbs/day)	(Pri:Sec)	(%)	(°F)	(°F)	(°F)	(gallons/day)	(ppm)
WAS %S	1.000								
DAF TS	0.763	1.000							
Blend(Pri:Sec)	-0.624	-0.892	1.000						
Cake %S	-0.498	-0.826	0.782	1.000					
Post lime Temp	-0.367	-0.458	0.367	0.535	1.000				
Temp Sec Eff	-0.546	-0.629	0.492	0.667	0.777	1.000			
Ambient Temp	-0.497	-0.549	0.408	0.564	0.811	0.898	1.000		
DAF FW	0.551	0.958	-0.890	-0.856	-0.427	-0.573	-0.487	1.000	
TRS Post-lime	-0.416	-0.394	0.385	0.291	0.425	0.544	0.528	-0.336	1.000

**Table 4.8: The essential process variables relating to N-containing compounds (ppm) during annual**

Correlation Table	Cake %S (%)	Post lime Temp (°F)	Pol_Sec (lbs/day)	Temp Sec Eff (°F)	Ambient Temp (°F)	N-containing (ppm)
Cake %S	1.000					
Post lime Temp	0.529	1.000				
Pol_Sec	0.166	0.194	1.000			
Temp(F)	0.640	0.773	0.216	1.000		
Ambient Temp	0.561	0.780	0.383	0.821	1.000	
N-containing	0.246	0.396	0.544	0.383	0.516	1.000

**Table 4.9: The essential process variables relating to TRS prior-lime (ppm) during winter**

Correlation Table	MLSS.SWO (mg/L)	Temp Sec Eff (°F)	Ambient Temp (°F)	S.W.O. BL*MLSS.SWO (ftmg/L)	TRS Prior-lime (ppm)
MLSS.SWO	1.000				
Temp Sec Eff	0.071	1.000			
Ambient Temp	0.142	0.813	1.000		
SWO. BL*MLSS.SWO	0.822	-0.164	-0.016	1.000	
TRS Prior-lime	0.445	0.498	0.658	0.316	1.000

**Table 4.10: The essential process variables relating to TRS post-lime (ppm) during winter**

Correlation Table	WAS %S (%)	DAF TS (lbs/day)	Blend (Pri:Sec)	Cake %S (%)	Post lime Temp (°F)	Temp Sec Eff (°F)	Ambient Temp (°F)	GT %S (%)	GT TS (lbs/day)	TRS Post-lime (ppm)
WAS %S	1.000									
DAF TS	0.704	1.000								
Blend(Pri:Sec)	-0.650	-0.835	1.000							
Cake %S	-0.565	-0.618	0.767	1.000						
Post lime Temp	-0.327	-0.175	0.450	0.512	1.000					
Temp Sec Eff	-0.237	-0.144	0.298	0.341	0.417	1.000				
Ambient Temp	-0.175	-0.153	0.246	0.338	0.566	0.813	1.000			
GT %S	-0.599	-0.474	0.787	0.667	0.443	0.327	0.196	1.000		
GT TS	-0.313	-0.228	0.603	0.491	0.537	0.331	0.248	0.831	1.000	
TRS Post-lime	-0.567	-0.309	0.451	0.423	0.630	0.511	0.532	0.422	0.331	1.000

**Table 4.11: The essential process variables relating to N-containing compounds (ppm) during winter**

Correlation Table	Post lime Temp (°F)	RAS.SE (mg/L)	RAS.SWO (mg/L)	Pol_DW/DTS (lbs/ton)	N-containing (ppm)
Post lime Temp	1.000				
RAS.SE	-0.187	1.000			
RAS.SWO	-0.340	0.757	1.000		
Pol_DW/DTS	-0.084	0.280	0.107	1.000	
N-containing	0.206	0.389	0.301	0.296	1.000

**Table 4.12: The essential process variables relating to TRS prior-lime (ppm) during summer**

Correlation Table	SWO.BL (ft)	SWE.BL (ft)	RAS.S.E (mg/L)	RAS.SWE (mg/L)	RAS.SWO (mg/L)	SWO*RAS (ftmg/L)	SWE*RAS (mg/L)	Ambient Temp (°F)	SWO.BL*MLSS (ftmg/L)	SWE.BL*MLSS (ftmg/L)	SE. BL*MLSS (ftmg/L)	TRS Prior-lime (ppm)
SWO.BL	1.000											
SWE.BL	0.545	1.000										
RAS.S.E	0.235	0.214	1.000									
RAS.SWE	-0.079	0.243	0.362	1.000								
RAS.SWO	0.156	0.158	0.176	0.657	1.000							
SWO*RAS	0.955	0.529	0.256	0.115	0.434	1.000						
SWE*RAS	0.272	0.786	0.343	0.787	0.521	0.389	1.000					
Ambient Temp	0.370	0.306	0.369	-0.065	0.123	0.370	0.152	1.000				
SWO.BL*MLSS	0.820	0.784	0.382	0.148	0.120	0.776	0.576	0.343	1.000			
SWE.BL*MLSS	0.449	0.791	0.208	0.304	0.243	0.473	0.689	0.175	0.794	1.000		
SE. BL*MLSS	0.176	0.574	0.530	0.098	-0.308	0.056	0.418	0.159	0.596	0.540	1.000	
TRS Prior-lime	0.373	0.583	0.308	0.384	0.440	0.460	0.613	0.343	0.547	0.616	0.303	1.000

**Table 4.13: The essential process variables relating to TRS post-lime (ppm) during summer**

Correlation Table	SE.BL (ft)	SWE.BL (ft)	RAS.SE (mg/L)	RAS.SWE (mg/L)	SWE*RAS (ftmg/L)	SE*RAS (ftmg/L)	Ambient Temp (°F)	SWO. BL*MLSS (ftmg/L)	SWE.BL*MLSS (ftmg/L)	SE.BL*MLSS (ftmg/L)	GT %S (%)	GT TS (lbs/day)	TRS Post-lime (ppm)
SE.BL	1.000												
SWE.BL	0.342	1.000											
RAS.SE	0.482	0.214	1.000										
RAS.SWE	-0.056	0.243	0.362	1.000									
SWE*RAS	0.175	0.786	0.343	0.787	1.000								
SE*RAS	0.912	0.317	0.788	0.117	0.263	1.000							
Ambient Temp	0.177	0.306	0.369	-0.065	0.152	0.279	1.000						
SWO.BL*MLSS	0.308	0.784	0.382	0.148	0.576	0.372	0.343	1.000					
SWE.BL*MLSS	0.018	0.791	0.208	0.304	0.689	0.092	0.175	0.794	1.000				
SE.BL*MLSS	0.777	0.574	0.530	0.098	0.418	0.780	0.159	0.596	0.540	1.000			
GT %S	-0.006	-0.005	0.537	0.582	0.351	0.256	0.020	0.082	0.188	0.161	1.000		
GT TS	0.568	0.158	0.347	0.377	0.344	0.561	-0.182	0.151	-0.034	0.452	0.321	1.000	
TRS Post-lime	0.334	0.483	0.554	0.372	0.535	0.474	0.306	0.507	0.504	0.586	0.309	0.316	1.000



**Table 4.14: The essential process variables relating to N-containing compounds (ppm) during summer**

Correlation Table	WAS %S (%)	SWO.BL (ft)	Post lime Temp (°F)	SWE.BL (ft)	SWO.BL*MLSS (ftmg/L)	Pol_DW (lbs/day)	Pol_DAF+Pol_DW (lbs/day)	Pol_DW/DTS (lbs/ton)	SE.BL*MLSS (ftmg/L)	SE*RAS (ftmg/L)	Ambient Temp (°F)	N-containing (ppm)
WAS %S	1.000											
SWO.BL	-0.304	1.000										
Post lime Temp	0.008	-0.038	1.000									
SWE.BL	-0.414	0.558	0.119	1.000								
SWO.BL*MLSS	-0.321	0.806	0.030	0.803	1.000							
Pol_DW	-0.257	0.181	0.089	0.194	0.151	1.000						
Pol_DAF+Pol_DW	-0.134	0.200	0.156	0.205	0.120	0.919	1.000					
Pol_DW/DTS	-0.241	0.130	-0.019	0.065	0.075	0.726	0.561	1.000				
SE.BL*MLSS	-0.349	0.223	0.190	0.610	0.641	0.096	0.028	-0.021	1.000			
SE*RAS	-0.389	0.234	0.179	0.395	0.422	0.030	0.035	-0.106	0.790	1.000		
Ambient Temp	-0.408	0.364	0.087	0.345	0.334	0.119	0.178	0.192	0.166	0.267	1.000	
N-containing	-0.308	0.339	0.243	0.373	0.394	0.344	0.343	0.317	0.354	0.330	0.368	1.000

#### 4.4 Additional Techniques

To gain a better understanding and to obtain more accuracy in forecasting, two techniques, dummy variables and interaction variables, are employed. In the regression model, using dummy variables and interaction variables will give us differences in intercepts and differences in slopes, respectively. Additionally, using dummy variables and interaction variables in empirical analysis can provide us with qualitative information (Wooldridge 2000).

As described in Section 4.2, there are separated statistical model for each period (annual, winter, and summer). However, if this research would like to employ annual models as the main model, the dummy variables and interaction variables techniques can assist this research to obtain the differences among seasons.

##### 4.4.1 Dummy Variables

Dummy variables are used in regression models to represent categorical information. A dummy variable ( $D$ ) is a binary variable that denotes whether an observation belongs to a certain category or not. For instance, if these analyses would like to categorize the data into two groups, the analyses can simplify it by using a 0 or a 1 for the value of the dummy variable. Basically, these analyses define a value of 0 if the condition is not met and a value of 1 if the condition is met. Given  $D_i$  as the dummy variable,  $D_i$  can be denoted as the following the condition:

$$D_i = \begin{cases} 1 & \text{if condition is present for observation } i \\ 0 & \text{otherwise} \end{cases}$$

As stated earlier, the research anticipates that including categorical independent variables via dummy variables can improve the statistical models in

Chapter 5. Equation 4.9 demonstrates how a dummy variable can be incorporated into the regression model.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \gamma D_i + \dots + \beta_k X_{ki} \quad (4.9)$$

With a dummy variable presented in Eq. (4.9), the intercept will be different for each of the two categories:

$$\begin{cases} \beta_0 + \gamma & \text{if condition is present} \\ \beta_0 & \text{otherwise} \end{cases}$$

#### 4.4.2 Interaction Variables

An interaction variable is the outcome of multiplying two or more variables. It assists these analyses to model different intercepts and different coefficients for different categories. In the wastewater treatment perspective, only one process variable occasionally cannot be used to explain the source of odor production in solids and biosolids. However, using the interaction of two variables can provide different information to clarify the odorous issues. Equation 4.10 shows how interaction variable  $X_1 X_2$  is used in a regression model.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \theta X_1 X_2 + \dots + \beta_k X_{ki} \quad (4.10)$$

As described earlier in Sections 4.4.1 and 4.4.2, these two techniques can be incorporated into the OLS equation to provide qualitative information and to allow differences in intercepts and slopes.

After establishing the statistical processes and the approaches used in this research, the next chapter will provide modeling analysis by demonstrating how to obtain the best statistical biosolids odorant model for each compound. It will also

provide the resulting models. Then, there will be a concluding discussion of the contribution in this research.

## Chapter 5: Data Analysis, Results, and Discussion

By using the advanced wastewater treatment plant (WWTP) of the District of Columbia Water and Sewer Authority (DC Water) as a case study, this chapter presents data analysis and the results of developing statistical models for odor. There are two types of models, explanatory and predictive. As described in Chapter 4, each type of model employs different requirements to select the best explanatory and predictive models.

This research presents general statistical models for each type, which are annual models. However, to gain a better understanding of odorous phenomena and to obtain precise odorous forecasts, this research studies and presents two additional statistical models: winter, and summer. Therefore, each odorants compound consists of three seasonal statistical models: an annual statistical model, a winter model, and a summer model.

For each seasonal period, we present and discuss only the best statistical model. Nonetheless, the remaining candidate statistical models for each odorant compound will be presented in Appendix B.

### 5.1 Modeling Analysis

As mentioned in the previous chapter, we employ ordinary least squares (OLS) estimation as well as use an instrumental variable (IV) and 2SLS estimation to create explanatory and predictive models by using the data from DC Water's advanced WWTP. We follow the 9 steps of statistical strategies (Section 4.2) by starting from step 4. Prior to starting, we state the anticipated significant upstream process

variables, which are assumed to contribute to those three odor compounds, obtained from steps 1 to 3:

**Table 5.1: The anticipated significant upstream process variables**

<b>Annual</b>		<b>Summer</b>		<b>Winter</b>	
<b>TRS (ppm)</b>	<b>N-con (ppm)</b>	<b>TRS (ppm)</b>	<b>N-con (ppm)</b>	<b>TRS (ppm)</b>	<b>N-con (ppm)</b>
WAS %S	Post lime temp	GT %S	WAS %S	WAS %S	Post lime temp
DAF TS	Pol_Sec	GT TS	SWO.BL	DAF TS	RAS.SE
Blend Ratio	Temp Sec Eff	Post lime temp	SWE.BL	Blend Ratio	RAS.SWO
Cake %S	Ambient temp	Ambient temp	Post lime temp	Cake %S	Pol_DW/dry solids
Temp Sec Eff		SE.BL	Pol_DW	GT %S	
Ambient temp		SWE.BL	Pol_DW/dry solids	GT TS	
DAF FW		SWO.BL	PolDW+PolDAF	Post lime temp	
Post lime temp		RAS.SE	SE.BL*RAS.SE	Temp of Sec	
GT %S		RAS.SWE	SWO.BL*MLSS	Ambient temp	
		RAS.SWO	SE.BL*MLSS	SWOBL*MLSS	
		SE.BL*MLSS	Ambient Temp		
		SWE.BL*MLSS			
		SWO.BL*MLSS			
		SE.BL*RAS			
		SWE.BL*RAS			
		SWO.BL*RAS			

After employing exploratory data analysis (EDA) and correlation analysis, one significant result showed an important relationship between temperature at secondary effluent and ambient temperature. Figure 5.1 illustrates the relationship between temperature at secondary effluent and ambient temperature. According to Figure 5.1, even when the ambient temperature was freezing (lower than 32 °F) during the winter period, the temperature at the secondary effluent was about 48 °F. In other words, the temperature of wastewater will not be that much lower during winter, nor will it be that much higher during summer.

We also compared the correlation coefficient ( $r$ ) between temperatures at the secondary effluent and ambient temperature during winter and summer periods. We found the following correlations:

*Annual*: Temp at secondary effluent VS Ambient temp ( $r = 0.898$ )

*Winter*: Temp at secondary effluent VS Ambient temp ( $r = 0.813$ )

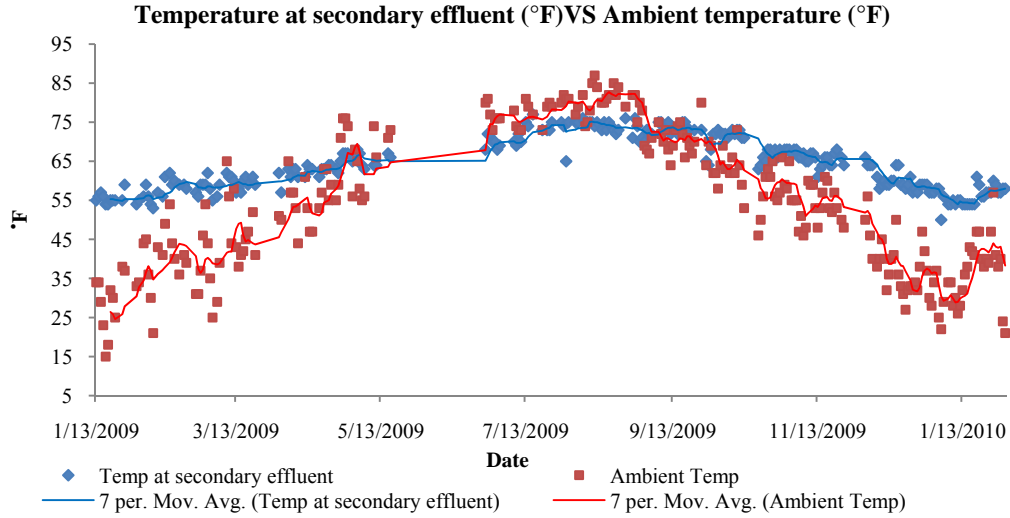
*Summer*: Temp at secondary effluent VS Ambient temp ( $r = 0.285$ )

The above information is crucial because it demonstrates the relationship between ambient temperature and temperature at the secondary effluent especially during summer period. In other words, it does not matter that the ambient temperature in summer will be higher or warmer, because it will not affect emission of odors in dewatered solids and biosolids. Conversely, during the winter period, ambient temperature has more of an effect on temperature at the secondary effluent and can assist in mitigating the emission in dewatered solids and biosolids. Furthermore, total reduced sulfur (TRS) at prior-lime ( $Y_1$ ) and at post-lime ( $Y_2$ ) have weak correlation coefficients ( $r$ ) with temperatures at the secondary and ambient temperature in the summer period:

( $Y_1$ ):  $r = 0.15$  and  $r = 0.34$ , respectively

( $Y_2$ ):  $r = -0.024$  and  $r = 0.306$ , respectively

In contrast, during annual and winter periods, the correlation coefficient ( $r$ ) of ( $Y_1$ ) and ( $Y_2$ ) with both temperatures is very high, which is above 0.6.



**Figure 5.1: Temp Sec Eff (°F) vs. Ambient Temp (°F)**

As stated in Section 4.5, employing dummy and interaction variables, can assist in quantifying the differences between winter periods and summer periods.

$$\text{Given; } D_T = \begin{cases} 1 & \text{Winter} \\ 0 & \text{Summer} \end{cases}$$

where;  $D_T$  is dummy variable.

This research, therefore, anticipates obtaining the different slopes and intercepts between the winter and summer periods.

For instance, the difference between winter and summer can be seen by writing out the regression models for the separate seasons:

$$\text{Annual : } Y_1 = \beta_0 + \beta_1(Temp) + \beta_2(D_T * Temp) + \beta_3(D_T)$$

$$\text{Winter : } Y_1 = (\beta_0 + \beta_3) + (\beta_1 + \beta_2)(Temp)$$

$$\text{Summer : } Y_1 = \beta_0 + \beta_1(Temp)$$

We also include interaction variables between secondary blanket level and mix liquor suspended solids (MLSS) and secondary blanket level and return activated sludge (RAS) concentration. Many studies cited in Chapter 1, including the previous



study by Janpengpen (2006), indicated that the secondary blanket level was one of the essential process variables contributing to TRS at prior lime. Additionally, by using the product of interaction between secondary blanket level and MLSS, and interaction between secondary blanket level and RAS, can provide more information about the sources of TRS at prior-lime and post-lime addition. In other words, if there was normally a high blanket level at the secondary settling tank, it would indicate that there was a high concentration of MLSS and RAS. But, this level is not necessarily indicated because the low blanket level can have higher concentration than the high blanket level, as the suspended solids are settling better or they are more compacted.

Therefore, there are six additional process variables added to this model:

Secondary east blanket level\*Mix liquor suspended solids (SE.BL\*MLSS)

Secondary west odd blanket level\*Mix liquor suspended solids (SWO.BL\*MLSS)

Secondary west even blanket level\*Mix liquor suspended solids (SWE.BL\*MLSS)

Secondary east blanket level\*Return activated sludge (SE.BL\*RAS)

Secondary west odd blanket level\*Return activated sludge (SWO.BL\*RAS)

Secondary west even blanket level\*Return activated sludge (SWE.BL\*RAS)

The unit for each above interaction is ft\*mg/L, which is a mass unit.

## 5.2 Explanatory Models

### 5.2.1 Explanatory Model for TRS at Prior-Lime

This section will employ the ordinary least square estimation to gain the best explanatory model for TRS prior-lime ( $Y_I$ ).

### Case I: Annual Model for TRS Prior-lime (January 2009 to January 2010)

**Table 5.2: Statistical model for TRS prior-lime (ppm) during annual period**

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.6466	0.4181	0.4066	0.405120486		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	5	29.83147587	5.966295174	36.3527	< 0.0001	
Unexplained	253	41.52301988	0.164122608			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	-2.023725624	0.680571888	-2.9736	0.0032	-3.364033558	-0.68341769
WAS %S	-0.388862351	0.094880373	-4.0984	< 0.0001	-0.575718316	-0.202006385
Temp Sec Eff	0.05838969	0.005826328	10.0217	< 0.0001	0.046915408	0.069863972
SWO. BL*MLSS. SWO	2.6907E-05	7.71915E-06	3.4857	0.0006	1.1705E-05	4.2109E-05
GT %S	0.164807583	0.085661938	1.9239	0.0555	-0.003893737	0.333508902
D <sub>T</sub> *Temp	0.007622134	0.001137213	6.7025	< 0.0001	0.005382523	0.009861745

The best explanatory model for TRS at prior-lime ( $Y_I$ ) in the annual model is shown in Table 5.2. All independent variables or process variables including interaction variables have correct signs of wastewater treatment and follow the theories of environmental engineering. The model explains about 42% of variation in  $Y_I$  and all process variables on the right-hand side are statistically significant. Let us interpret the variables in this model.

**Waste activated percent solids (WAS %S):** This variable has a negative coefficient, implying that high percent solids of WAS can mitigate  $Y_I$  concentration. In other words, by holding all the other variables fixed, changing WAS %S by 1% of solids can cause an average change in  $Y_I$  of -0.39 ppm. Similarly, the correlation coefficient ( $r$ ) between  $Y_I$  and WAS %S also has a negative sign, which is -0.39. This information strongly supports the model in Table 5.2.

As described in Section 3.2.6, WAS %S is the biologically activated solids at the secondary settling tank. In other words, these solids are aerated before passing through the secondary sedimentation process, and then pass through the dissolved air flotation (DAF) tank. In fact, there is not only just WAS % S represented at this point because WAS %S can be used to form dissolved air flotation total solids (DAF TS) by associating it with the waste activated (WAS) flow, which is described by the following formula:

$$\text{DAF TS}(\text{lbs} / \text{day}) = \left[ \frac{\text{DAF flow}(\text{gallons} / \text{day}) * 8.43 * \text{WAS \%S} * 10000}{1000000} \right]$$

Furthermore, the correlation coefficients ( $r$ ) between  $Y_I$  and DAF TS and between  $Y_I$  and DAF flow are -0.377 and -0.322, respectively. However, DAF TS cannot be included in the model because DAF TS is the product between WAS %S and DAF flow, so there is multicollinearity. These confirm that WAS %S can assist in decreasing  $Y_I$  in dewatered solids.

**Gravity thickener percent solids (GT %S):** This variable has a positive coefficient, in contrast to the negative WAS %S coefficient. This difference implies that high percent solids of GT can contribute to increased  $Y_I$  concentration in dewatered solids. In other words, by holding all the other variables fixed, changing GT %S by 1% of solids can cause an average change in  $Y_I$  of 0.16 ppm. Unlike WAS %S, GT %S is the raw settling sludge at primary settling tank as stated in Section 3.2.4. Therefore, it is easier for GT %S to cause septicity in sludge passing to the blending tank since they are not aerated solids. Additionally, a higher GT %S indicates a higher proportion of food source for microorganisms that can contribute to a greater

production of odors in dewatered solids and biosolids as well (Peot and Ramirez 2007).

Therefore, to decrease the level of  $Y_l$ , this model suggests that DC Water should maintain GT %S at a lower level. However, using only this information to draw a suggestion is inadequate and can be misleading. As DC Water ultimately prefers the dewatered cake with high solids content (Cake %S), it was believed that if there were high GT %S and high GT flow at the blending tank, they could provide high percent solids in dewatered solids. This statement can be confirmed by the following formula:

$$\begin{aligned} \text{Blend Ratio} &= \left[ \frac{\text{GT TS}(\text{lbs} / \text{day})}{\text{DAF TS}(\text{lbs} / \text{day})} \right] \\ &= \left[ \frac{\text{GT flow}(\text{gallons} / \text{day}) * 8.34 * \text{GT \%S} * 10000 / 1000000}{\text{DAF flow}(\text{gallons} / \text{day}) * 8.34 * \text{WAS \%S} * 10000 / 1000000} \right] \end{aligned}$$

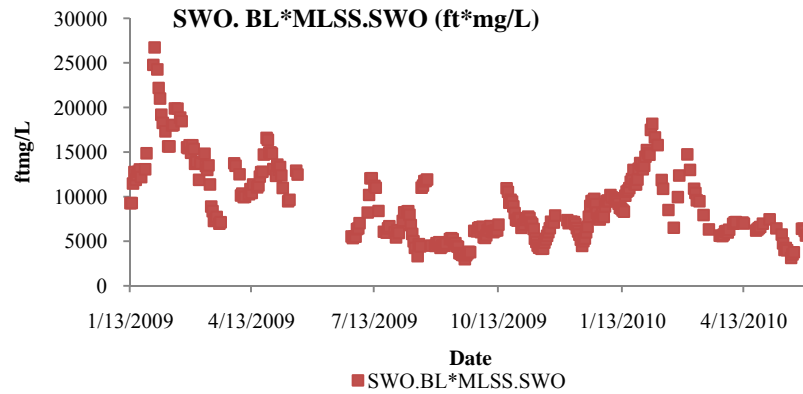
Practically, it was believed that having high GT %S and high GT TS could increase the dewatering process performance because they are both raw sludge without any chemical addition. However, the model suggests that operators should keep GT %S lower. Maintaining the ultimate requirement of high cake %S at the final product can be seen as a tradeoff. If WAS %S and DAF flow are increased, they can obtain the same cake %S at the final product.

**SWO.BL\*MLSS:** This process variable is new information for the secondary sedimentation process. According to Sekyiamah and Kim (2006), Janpengpen (2006), Janpengpen (2007), and Sekyiamah and Kim (2009), the greater the secondary sludge blanket depth, the greater the biosolids odors at the downstream. In other words, a

higher secondary blanket level can promote TRS concentration of dewatered solids and biosolids.

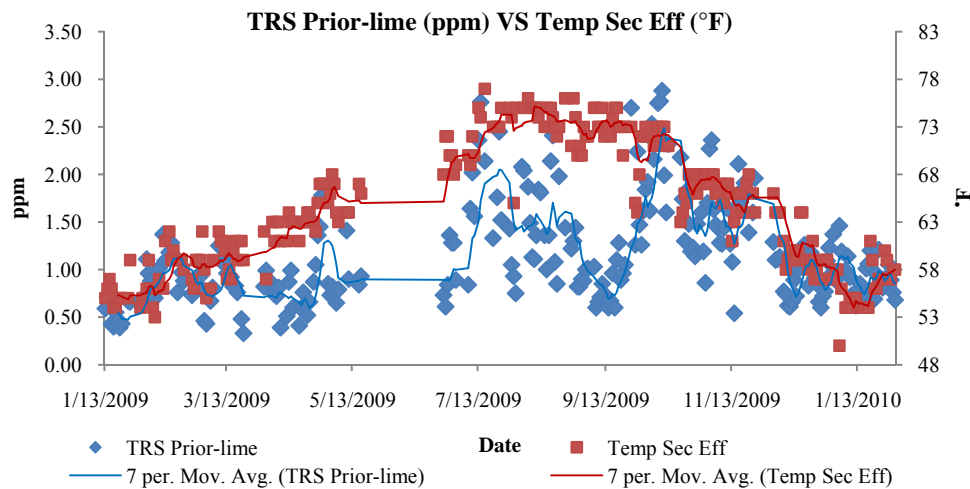
We investigate a different aspect, because using only a blanket level does not provide ample information. A high secondary blanket level sometimes does not indicate the high odors downstream; instead, a low secondary blanket level more often indicates high odors at downstream. That means the poor settling suspended solids at the secondary sedimentation process can show an error reading for a high number of secondary blanket level. In contrast, the better settling suspended solids can compact more easily and can give the low secondary blanket level. For this reason, we use the mass unit (ft\*mg/L), which is the product between height and concentration.

In the model, this process variable has a positive coefficient sign, implying that high mass of SWO.BL\*MLSS can contribute to an increase of  $Y_I$  concentration in dewatered solids. By holding all the other variables fixed, a change in SWO.BL\*MLSS by 1 ft\*mg/L can cause an average change in  $Y_I$  of  $2.6907 \times 10^{-5}$  ppm. While the statistical significance of coefficients is relatively low, the graph in Figures 5.2 indicates that an average mass of SWO.BL\*MLSS is approximately 10,000 ft\*mg/L. Therefore, if there are any changes, for instance, from 10,000 ft\*mg/L to 15,000 ft\*Mg/L, it will significantly affect  $Y_I$  concentration (e.g.,  $5,000 \times 2.6907 \times 10^{-5} = 0.13$  ppm).

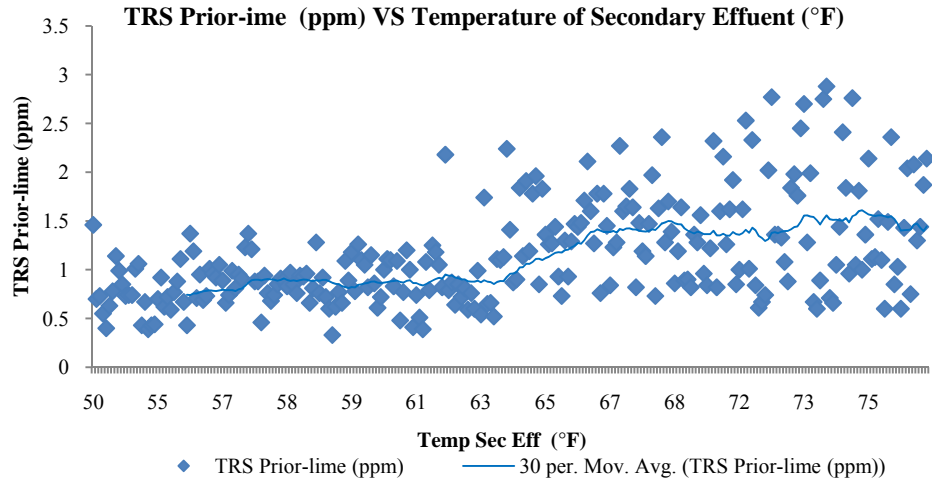


**Figure 5.2: SWO.BL\*MLSS.SWO (ft\*mg/L) (January 2009 to May 2010)**

**Temperature at the secondary effluent (Temp Sec Eff):** This is one of significant variables causing the changes in TRS prior-lime ( $Y_I$ ) (see Figures 5.3 and 5.4). Both Figures indicate that increasing in temperature of effluent at the secondary process can contribute to level of  $Y_I$  at the downstream after dewatering process. Figure 5.4 demonstrates the trend of  $Y_I$  when Temp sec Eff increases.



**Figure 5.3: TRS prior-lime (ppm) vs. Temp Sec Eff (°F)**



**Figure 5.4: Relationship between TRS prior-lime (ppm) and Temp Sec Eff (°F)**

Additionally, the correlation coefficient ( $r$ ) between temperature at the secondary effluent and  $Y_I$  is 0.512, which is a strong correlation. Furthermore, this correlation is higher than the correlation between ambient temperature and  $Y_I$ , which is just 0.439. Temp Sec Eff has a positive coefficient sign indicating that increasing temperature at the secondary effluent can accelerate the  $Y_I$  concentration in dewatering solids. By holding all the other variables fixed, changing Temp Sec Eff by one degree Fahrenheit (1 °F) can cause an average change in  $Y_I$  of 0.06 ppm.

In short, temperature at the secondary effluent dramatically contribute to increase  $Y_I$ . For this reason, we use a dummy variable, as stated in Section 5.1 in the annual model of  $Y_I$ , to distinguish the difference between the winter and the summer.

**Temp\*  $D_T=1$ :** The variable  $D_T=1$  means it is a winter period. This variable is interaction variable between Temp Sec Eff and dummy variable. It implies that temperature at the secondary effluent can naturally assist in decreasing  $Y_I$  concentration during the winter period. By holding all the other variables fixed, if it is

the winter period, it can cause an average change in  $Y_I$  of 0.008 ppm. The regression equations for winter and summer periods can be rewritten as follows:

Winter:

$$\hat{Y}_I = -2.0161 - 0.3889(\text{WAS \%S}) + 0.0584(\text{Temp Sec Eff}) + 2.6907 \cdot 10^{-5}(\text{SWO.BL} \cdot \text{MLSS}) + 0.1648(\text{GT \%S})$$

Summer:

$$\hat{Y}_I = -2.0237 - 0.3889(\text{WAS \%S}) + 0.0584(\text{Temp Sec Eff}) + 2.6907 \cdot 10^{-5}(\text{SWO.BL} \cdot \text{MLSS}) + 0.1648(\text{GT \%S})$$

In both above regression equations, as in the output from Table 5.2, the intercepts are negative. These intercepts can lead to misinterpretation and raise the question of whether both equations can be applied to any wastewater treatment plants without exceptional weather conditions (e.g., Alaska and Bangkok). These two equations can definitely be employed in any weather conditions. As described in Section 5.1, these equations verify that even though ambient temperature is below freezing point (i.e., 15 °F), temperature at the secondary effluent is 48 °F. In addition, even if it is the case that all independent variables are close to zero, for instance, the wastewater treatment plant is shut down<sup>1</sup>, the temperature at the secondary effluent or temperature of the wastewater never becomes zero. Therefore, the negative intercept does not contradict domain knowledge and does not extrapolate into ranges of independent variable values that do not exist. To give a better understanding, as well as to offer information about an annual explanatory model, the next two cases, (case II: winter period and case III: summer period), are provided.

---

<sup>1</sup> When the wastewater treatment plant is shut down, all the treatment processes are stopped, so that those process data will not be used in the analyses.



## CASE II: Winter Model for TRS Prior-lime (January 2009 to March 2009)

The correlation analysis indicates that there are four process variables having a relatively high positive correlation with TRS prior-lime ( $Y_I$ ), which are MLSS.SWO ( $r = 0.445$ ), SWO.BL\*MLSS ( $r = 0.316$ ), temperature at the secondary effluent ( $r = 0.498$ ), and ambient temperature ( $r = 0.658$ ). Table 5.3 presents the statistical model output for the winter period.

**Table 5.3: Statistical model for TRS prior-lime (ppm) during the winter period**

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.7476	0.5589	0.5401	0.173475978		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares		p-Value	
Explained	2	1.792274003	0.896137002	29.7780	< 0.0001	
Unexplained	47	1.414413997	0.030093915			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	-0.283899873	0.15693682	-1.8090	0.0768	-0.599616031	0.031816285
MLSS.SWO	0.000110959	3.02443E-05	3.6688	0.0006	5.01151E-05	0.000171803
Ambient Temp	0.015049725	0.002427339	6.2001	< 0.0001	0.010166549	0.019932902

The model in Table 5.3 is the best explanatory model for TRS prior-lime ( $Y_I$ ) during the winter period and consists of two upstream process variables, which are mix liquor suspended solids at secondary west odd (MLSS.SWO) and ambient temperature (Ambient Temp). Both process variables including intercepts are statistically significant. The explanatory power of this model suggests that about 56% of variation in  $Y_I$  can be explained by those two process variables during the winter period.

**MLSS.SWO:** This process variable has a positive coefficient, implying that increasing MLSS.SWO concentration can contribute to an increase in  $Y_I$  concentration. By holding all the other variables fixed, changing MLSS.SWO by 1 mg/L can cause an average change in  $Y_I$  of  $1.1 \times 10^{-4}$  ppm. This information verifies that if operators keep more concentration of MLSS then it can contribute to production of  $Y_I$  in dewatered solids.

**Ambient temperature (Amb Temp):** This variable, known as the outside temperature, is an uncontrollable variable, but it has relatively high correlation ( $r = 0.658$ ) with the TRS prior-lime ( $Y_I$ ) during the winter period. Like Temp Sec Eff process variable, ambient temperature has a positive coefficient. This implies that ambient temperature can naturally assist in decreasing  $Y_I$  concentration during the winter period. By holding all the other variables fixed, changing ambient temperature by 1 °F can cause an average change in  $Y_I$  of 0.015 ppm.

Therefore, this information can assist the DC Water biosolids management section in cost saving for the land application program. That means DC Water can diminish the payment of hauling costs during the winter period because the contractors can distribute biosolids to any land application site that is close to a wastewater treatment plant. In other words, DC Water can transport biosolids to nearby residents or the most sensitive areas during the winter period. However, it is still necessary for DC Water to have backup application sites in case that there are any odor incidents occurred.

### Case III: Summer Model for TRS Prior-Lime (June 2009 to September 2009)

**Table 5.4: Statistical model for TRS prior-lime (ppm) during the summer period**

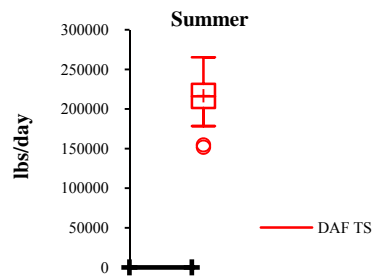
Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.7688	0.5911	0.5591	0.366718913		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	5	12.44056186	2.488112372	18.5013	< 0.0001	
Unexplained	64	8.606896711	0.134482761			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	4.869238295	2.045571422	2.3804	0.0203	0.782739605	8.955736985
DAF TS	-6.29232E-06	2.51552E-06	-2.5014	0.0149	-1.13177E-05	-1.26698E-06
Cake %S	-0.237370938	0.055436952	-4.2818	< 0.0001	-0.348118981	-0.126622895
SWE.BL	0.63003452	0.148290202	4.2487	< 0.0001	0.333790785	0.926278254
RAS.SWO	0.000596106	0.000132375	4.5031	< 0.0001	0.000331655	0.000860556
GT %S	0.21122279	0.145349169	1.4532	0.1511	-0.079145554	0.501591135

Table 5.4 represents the best explanatory model of TRS prior-lime ( $Y_I$ ) during the summer period with  $R^2$  of 0.59. All five process variables included in the model are statistically significant, and these include: dissolved air flotation total solids (DAF TS), cake percent solids (Cake %S), secondary west even blanket level (SWE.BL), return activated sludge at secondary west odd (RAS.SWO), and gravity thickener percent solids (GT %S). This means that about 59% of variation in TRS prior-lime ( $Y_I$ ) can be explained by those five process variables.

**Dissolved air flotation total solids (DAF TS):** This process variable has a negative coefficient, implying that high DAF TS can assist mitigating  $Y_I$  concentration. As described in Case I, DAF TS is a product of DAF flow and WAS %S, and its unit is lbs/day. The following is the formula for DAF TS:

$$\text{DAF TS (lbs / day)} = \left[ \frac{\text{DAF flow (gallons)} * 8.43 * \text{WAS \%S} * 10000}{1000000} \right]$$

By holding all the other variables fixed, changing DAF TS by 1 lbs/day can cause an average change in  $Y_I$  of  $-6.292 \times 10^{-6}$  ppm. Although the coefficient of DAF TS is minimal and is seem practically insignificant to affect  $Y_I$  concentration, note that the average daily of DAF TS is about 220,000 lbs/day (see Figure 5.5) and DAF TS is in fact very vital for decreasing  $Y_I$ . Therefore, if DC Water maintains a high level of DAF TS (i.e., higher than GT TS), it will effectively mitigate  $Y_I$  concentration.

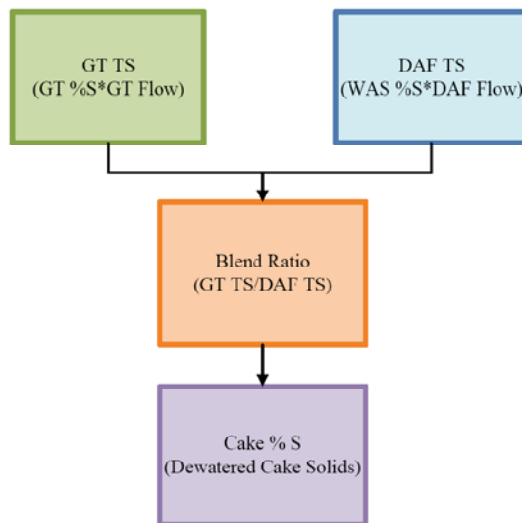


**Figure 5.5: DAF TS (lbs/day) during the summer**

In addition, the DAF TS is the total solids gained from WAS %S, which consist of biologically activated solids, and DAF flow, which consists of flotation-thickened sludge blown by air. As a result, DAF TS is made up of aerated total solids. Unlike GT TS, it contains raw total solids gained from a gravity thickener of primary process, which is more septic as explained in Case I. It has been previously understood that if there are GT TS higher than DAF TS then they could increase the dewatering performance. In other words, centrifuges can easily separate solids from water. For this reason, DC Water generally maintains GT TS higher than DAF TS.

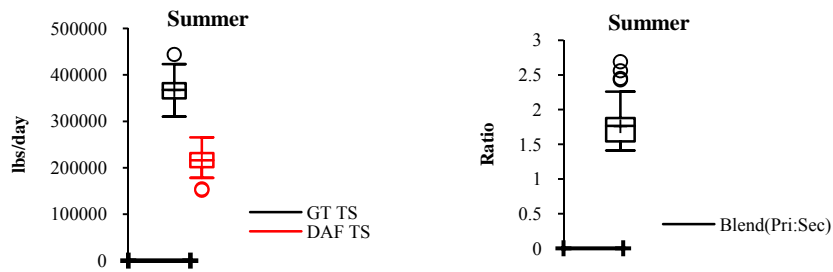
**Gravity thickener percent solids (GT %S):** This process variable has a positive coefficient, implying that high GT %S can increase  $Y_I$  concentration. Like GT %S of Case I, it is the raw settling sludge at the primary settling tank and can be considered a critical source of  $Y_I$ . By holding all the other variables fixed, changing GT %S by 1% of solids can cause an average change in  $Y_I$  of 0.21 ppm. As described in Case I, it is easy for GT %S to become septic because it is a raw sludge without any aeration. This variable is a vital variable used to calculate GT TS in blending tank. Therefore, this variable directly relates to blend ratio and ultimately relates to cake %S in dewatered solids and biosolids.

**Cake percent solids (Cake %S):** Cake %S consists of dewatered solids before they pass through the lime stabilization process. This is one of the most essential process variables because Cake %S relates to other process variables that can contribute to TRS prior-lime ( $Y_I$ ). Before discussing this further, this research will demonstrate the sources of Cake %S and other process variables, i.e., GT %S, GT Flow, GT TS, WAS %S, DAF Flow, DAF TS, and Blend Ratio, that have relationships with Cake %S.



**Figure 5.6: Sources of Cake percent solids (Cake %S)**

According to the model, Cake %S has a negative coefficient, implying that high Cake %S can decrease  $Y_I$  concentration. Technically, to gain a higher Cake %S or to gain more dry cake downstream, DC Water maintains an amount of GT TS higher than that amount of DAF TS in the blending tanks. In other words, operators generally keep the Blend Ratio (ratio between GT TS and DAF TS as shown in Figure 5.6) more than 1, which is to say there are GT TS more than DAF TS in the blend. As a result, there is more raw sludge than aerated sludge (GT TS = GT %S\*GT Flow and DAF TS = WAS %S\*DAF Flow). This can cause Cake %S to have a positive coefficient sign with  $Y_I$ .



**Figure 5.7: Amount of GT TS (lbs/day), DAF TS (lbs/day), and Blend Ratio (Pri:Sec) during summer period**

According to this model, however, Cake %S has a negative coefficient sign because the DAF TS, which has a negative coefficient sign, is also included into the model. The model suggests that DC Waster must maintain DAF TS higher than GT TS to mitigate  $Y_I$  concentration. That means if DC Water keeps DAF TS higher then it will cause a lower Blend Ratio or have more biological activated sludge (i.e., aerated sludge). The mass balance below illustrates how to gain Cake %S:

$$\text{Total solids (Dry cake or Cake \%S)} = \text{GT TS} + \text{DAF TS}$$

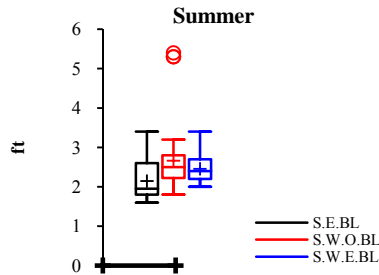
Increasing DAF TS or decreasing the Blend Ratio does not mean a decrease in Cake %S. Ultimately, DC Water needs a dryer cake or high Cake %S for the final product. Therefore, to maintain that goal, operators must have more DAF TS than GT TS. To increase DAF TS, operators can increase either WAS %S or DAF Flow.

Vilalai (2008) stated that “ A low percent solid indicates septic conditions of the sludge that is usually retained in the tanks for a long time and making it hard to remove the water content. However, a high percent solid was corresponding to a high dewatering capacity of the centrifuges on that day. Sufficient centrifuges running can reduce the dewatering load on each centrifuge as well as reduce the retention time of sludge in the blend and sedimentation tank.” Thus, this comment supports the negative coefficient sign in this model.

In sum, Cake %S is the product between DAF TS and GT TS. Therefore, the negative coefficient of Cake %S in the model is caused by increasing of DAF TS to minimize  $Y_I$  concentration and to maintain higher Cake %S. In addition, a negative coefficient of Cake %S means longer retention at sedimentation and greater septicity.

**Secondary west east blanket level (SWE.BL):** According to the correlation coefficient ( $r$ ), SWE.BL has the highest  $r$  ( $r = 0.583$ ) among those three secondary blanket variables (i.e., SE.BL ( $r = 0.026$ ), SWO.BL ( $r = 0.373$ ), and SWE.BL( $r = 0.583$ )). In fact, as shown in Figure 5.8, the blanket level at both sides of the secondary west has blanket depth higher than the blanket level at secondary east. DC Water usually tries to split flows to each side of secondary settling evenly. However, there were some constructions and some mechanical problems at the aerated tanks of the east side during May 2009 to July 2009; therefore, more flows went to the

secondary west side and caused a higher blanket level at the west side. Both SWE.BL and SWO.BL have almost the same average blanket level, therefore SWE.BL can be represented by the blanket level at the secondary west side.



**Figure 5.8: Secondary blanket level (ft) during the summer period**

In the model, only SWE.BL can be included into the right-hand side of the regression equation, and it is also statistically significant to  $Y_I$ . SWE.BL has a positive coefficient, implying that a high level of SWE.BL can contribute to the increasing of  $Y_I$  concentration in dewatered solids. By holding all the other variables fixed, changing SWE.BL by 1 ft can cause an average change in  $Y_I$  of 0.63 ppm, which is very high.

Many studies indicated that the greater the blanket level, the greater the production of  $Y_I$  concentration inside dewatered solids and biosolids. In addition, the higher secondary west blanket level couples with longer retention time at the bottom of the settling tank, indicating more septicity inside the sludge passing to the dewatering process. Therefore, it is crucial for DC Water to properly maintain the secondary west blanket level so as not to exceed 1.8 ft (Janpengpen 2006). Furthermore, as described in Case II, maintenance and mechanical problems also play important roles to cause backup or a higher secondary west blanket level. To avoid those issues leading to high  $Y_I$  concentration in dewatered solids, DC Water must



have effective management plans with functional organization from the upper level to the bottom level. The details of the standard operating procedure (SOP) will be stated more explicitly in Chapter 6.

**Return activated sludge of secondary west odd (RAS.SWO):** This variable has a positive coefficient, implying that high RAS.SWO can contribute to increased  $Y_I$  concentration. By holding all the other variables fixed, changing RAS.SWO by 1 mg/L can cause an average change in  $Y_I$  of 0.0006 ppm. However, an average value of RAS.SWO is about 4,000 mg/L. Therefore, any changes in RAS.SWO can significantly affect  $Y_I$  concentration. In addition, correlation coefficient ( $r$ ) between RAS.SWO and  $Y_I$  is the highest among those three return activated sludge (i.e.,  $r = 0.44$ ).

As described in Section 3.2.10, a high concentration of RAS.SWO can be the source of odorants in dewatered solids and biosolids because a high concentration of RAS.SWO demonstrates a high density of settled solids at the bottom of the secondary sedimentation tank. In addition, this variable is not present in the statistical model of Case II (winter period) because the properties of RAS.SWO relate to the temperature of wastewater, especially the temperature at the secondary process. Therefore, high RAS.SWO concentration coupled with high temperature can contribute to an increased  $Y_I$  concentration. In other words, there are more septic conditions in RAS concentration during the summer period than the winter period.

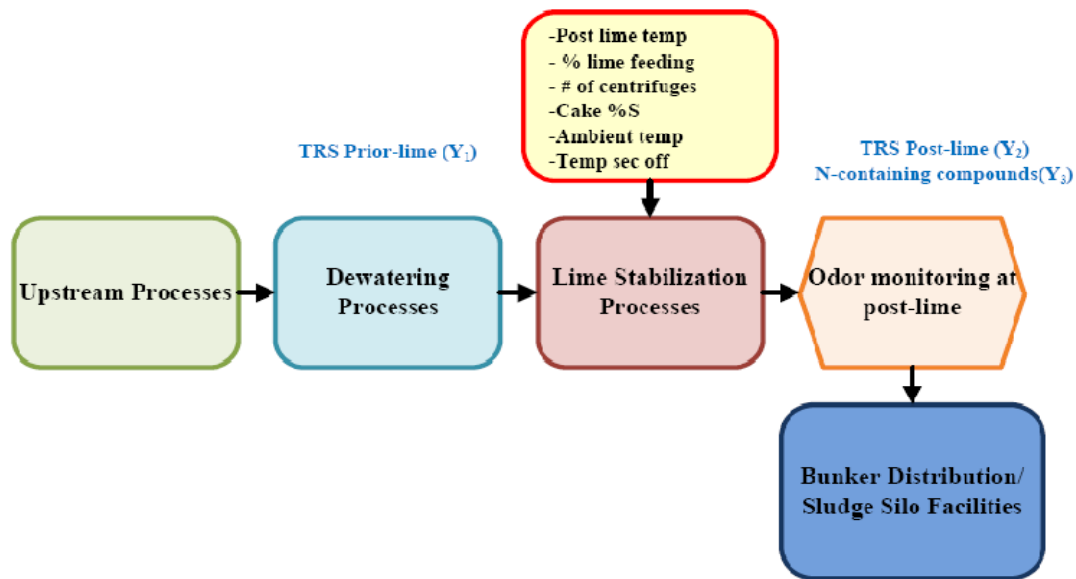
### 5.2.2 Explanatory Models for TRS at Post-Lime

In this section, ordinary least squares (OLS) estimation, instrumental variable (IV), and two stage least squares (2SLS) estimation are employed to gain the best

explanatory model for TRS post-lime ( $Y_2$ ). Additionally, dummy variables and interaction terms will be used in this analysis as well. There are three cases to be discussed, which include, Case IV, Case V, and Case VI.

As described in Chapter 3, DC Water employs a lime stabilization process to transform dewatered solids to biosolids. Theoretically, limed addition suppresses TRS at prior-lime ( $Y_1$ ) concentration. In other words, lime stabilization changes TRS at prior-lime ( $Y_1$ ) to TRS at post-lime ( $Y_2$ ) as stated in one of hypotheses in this research. In addition, this research defines that greater  $Y_1$  concentration can cause the greater of  $Y_2$  concentration after lime addition process as well.

Before presenting results and discussions for  $Y_2$  in all those three cases, Figure 5.9 represents processes diagram and how  $Y_2$  relates to  $Y_1$ .



**Figure 5.9: Relationship between odor at prior-lime and post-lime process**

#### Case IV: Annual Model for TRS Post-lime ( $Y_2$ )

##### Ordinary least squares (OLS) estimation

We first use ordinary least squares (OLS) estimation in multiple linear regression (MLR) to estimate TRS post-lime ( $Y_2$ ). At this process, there are four additional process variables, including percent lime feeding rate, number of centrifuges in service, dewatered cake percent solids, and post lime temperature. Furthermore, there are two variables carrying over from dewatering process, which include temperature at secondary effluent and ambient temperature, and are to be added at lime stabilization process as well. We know that the lime stabilization process can suppress  $Y_1$  to  $Y_2$ , so we will regress  $Y_2$  on the upstream process variables and those six additional variables at the lime stabilization process to get statistical model for  $Y_2$ , as shown in Table 5.5.

**Table 5.5: Statistical model for TRS post-lime (ppm) during the Annual period**

	Multiple	R-Square	Adjusted	StErr of		
<i>Summary</i>	R		R-Square	Estimate		
	0.7396	0.5470	0.5381	0.04624257		
	Degrees of	Sum of	Mean of	F-Ratio	p-Value	
<i>ANOVA Table</i>	Freedom	Squares	Squares			
Explained	5	0.653374066	0.130674813	61.1094	< 0.0001	
Unexplained	253	0.541008945	0.002138375			
	Coefficient	Standard	t-Value	p-Value	Confidence Interval 95%	
<i>Regression Table</i>		Error			Lower	Upper
Constant	-0.140766718	0.024538068	-5.7367	< 0.0001	-0.189091616	-0.09244182
Blend Ratio	0.03898684	0.009223219	4.2270	< 0.0001	0.020822773	0.057150906
<i>Post lime temp</i>	0.000275202	0.000194641	1.4139	0.1586	-0.00010812	0.000658525
SE. BL*MLSS	1.01345E-05	2.10161E-06	4.8222	< 0.0001	5.99558E-06	1.42734E-05
(DT = 1)	-0.025740412	0.008102973	-3.1767	0.0017	-0.041698283	-0.009782541
TRS Prior-lime	0.070068507	0.005924878	11.8262	< 0.0001	0.058400142	0.081736872

The model from Table 5.5 is an explanatory model for TRS post-lime ( $Y_2$ ). There are five variables, which are Blend Ratio, Post lime temperature, SE.BL\*MLSS, ( $D_T=1$ ), and TRS prior-lime ( $Y_1$ ), included in the model to explain  $Y_2$ . In other words, about 55% of variation in  $Y_2$  can be explained by those five variables, and all of them are statistically significant. In addition, all coefficient signs are correct based on earlier discussion.

Note that, only one of six additional process variables at lime stabilization process, namely post lime temperature (post lime temp), is statistically significant, has a correct coefficient sign, and can be included in the model. In contrast, the other five variables, (% lime feeding, # centrifuges, Cake %S, temperature at secondary effluent, and ambient temperature) are not statistically significant and have incorrect coefficient sign with respect to domain knowledge of wastewater treatment. In other words, they cannot be included in the model. Instead, they are included in the residual or error term of the regression equation on  $Y_2$ , or they cannot directly explain  $Y_2$ .

In order to overcome this issue, to gain a better understanding of the sources of  $Y_2$ , and to identify those variables hidden in the error term, we employ an instrumental variable (IV) as a way to address those statements. The post lime temp is a byproduct of the reaction between lime feeding rate and dewatered solids, so we regress post lime temp on the variables at lime stabilization process considered to contribute to temperature increase of limed biosolids.

**Table 5.6: Statistical model for Post lime temperature (°F) to create IV**

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.8238	0.6787	0.6724	12.01149963		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	5	77112.82734	15422.56547	106.8962	< 0.0001	
Unexplained	253	36501.85918	144.2761232			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	-30.00763808	18.57902221	-1.6151	0.1075	-66.5968821	6.581605932
% of lime feed	0.537839971	0.506377945	1.0621	0.2892	-0.459413059	1.535093002
Cake %S	0.931279837	0.620707391	1.5004	0.1348	-0.291131868	2.153691541
# Cent Serv	1.795134369	1.183741575	1.5165	0.1306	-0.536108301	4.126377039
Temp Sec Eff	0.700633801	0.283943932	2.4675	0.0143	0.141438937	1.259828665
Amb Temp	0.683685206	0.099159961	6.8948	< 0.0001	0.488401086	0.878969326

The model from Table 5.6 confirms that all those five variables at lime stabilization can explain about 68 % of variation in Post lime temp ( $\hat{W}$ ), and they are all statistically significant with correct coefficient signs according to wastewater treatment theory.

**Percent lime feeder (%lime feed):** Lime used in lime stabilization process at DC Water is quick lime (CaO) powder. It is fed in lime mixer to stabilize dewatered solids by maintaining pH above 12 for up to 2 hours. DC Water, however, maintains pH above 12.2 for up to 24 hours to produces so-called Class B biosolids. The percent of lime feeder is the speed of an auger feeding to dewatered solids in the lime mixer. Adding lime causes not only changes in pH but also in temperature. Therefore, % lime feeding is one of the variables expected to cause changes in post lime temp. In addition, % lime feeding has a positive coefficient, implying that high % lime feeding

can contribute to increase post lime temp. By holding all the other variables fixed, changing % lime feeding by 1 % can cause an average change in  $\hat{W}$  of 0.54 °F.

**Cake percent solids (Cake %S):** This variable can provide an answer to one of the hypotheses in Chapter 1. Cake %S has positive coefficient, implying that high Cake %S or dryer cake can contribute to increase post lime temp more easily than less Cake %S or wet cake. In other words, with the same of % lime feeding, the higher Cake %S can give higher temperature than lower cake %S. By holding all the other variables fixed, changing Cake %S by 1 % can cause an average change in  $\hat{W}$  of 0.93 °F.

**Number Centrifuges in service (# Cent Serv):** This variable has a positive coefficient, implying that a high # of centrifuges in service can contribute to increase post lime temperature. By holding all the other variables fixed, increasing # Cent Serv by 1 centrifuge can cause an average increase in  $\hat{W}$  of 1.79 °F. Section 5.6 will further discuss this variable and how it affects post lime temperature in the case that there are a lower # of centrifuges in service, but operators still keep the same % lime feeding rate.

**Temp Sec Eff and Amb Temp:** Naturally and theoretically, these two variables are the main variables affecting post lime temp. These two variables have positive coefficients implying that high Temp Sec Eff and Amb Temp can contribute to increase  $\hat{W}$ . By holding all the other variables fixed, changing Temp Sec Eff by 1 °F can cause an average change in  $\hat{W}$  of 0.7 °F. Similarly, holding all the other variables fixed, changing Amb Temp by 1 °F can cause an average change in  $\hat{W}$  of 0.68 °F, which is very close to Temp Sec Eff.

To verify that all of the above five variables are included in the error term of the regression equation of  $Y_2$ , we employ a Hausman Test for endogeneity testing, as described in Section 2.3.2. Therefore, we first use the residuals from  $\hat{W}$  as an independent variable in the regression equation of  $Y_2$ . This means that we regress  $Y_2$  on Blend Ratio, post lime Temp, SE.BL\*MLSS, TRS prior-lime, and on the residual of post lime temp. If the estimated model indicates that residual of post lime temp is statistically significant and the coefficient is not zero, it thus verifies that  $\hat{W}$  is endogenous.

**Table 5.7: Hausman test for endogeneity of post lime temp (°F)**

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.7397	0.5471	0.5381	0.046239526		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	5	0.653445296	0.130689059	61.1241	< 0.0001	
Unexplained	253	0.540937715	0.002138094			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	-0.215889163	0.022422524	-9.6282	< 0.0001	-0.26004774	-0.171730586
Blend(Pri:Sec)	0.030665326	0.009416107	3.2567	0.0013	0.012121388	0.049209263
Post lime Temp	0.001113986	0.000206498	5.3947	< 0.0001	0.000707311	0.00152066
SE.BL*MLSS	1.04811E-05	2.12246E-06	4.9382	< 0.0001	6.30111E-06	1.4661E-05
<b><u>Residual of Post lime temp</u></b>	<b><u>-0.00108089</u></b>	0.000339678	-3.1821	<b><u>0.0016</u></b>	-0.00174985	-0.000411936
TRS Pri	0.061771759	0.006324019	9.7678	< 0.0001	0.049317331	0.074226186

The output in Table 5.7 indicates that the residual of post lime temp is statistically significant. Therefore, post lime temp ( $\hat{W}$ ) is endogenous. As a result, all five variables at lime stabilization (% lime feeder, Cake %S, # centrifuges in service, temp at secondary effluent, and ambient temperature) can be defined as the instrumental variable (IV) in this analysis.

After we verified that post lime temp ( $W$ ) is endogenous with the error term of the regression equation of  $Y_2$ , the fitted value of post lime temp ( $\hat{W}$ ), which is obtained using the estimated model shown in Table 5.6, is used as an independent variable in the regression equation of  $Y_2$ . In other words, we again regress  $Y_2$  on Blend ratio, SE.BL\*MLSS, TRS prior-lime, and predict post lime temp ( $\hat{W}$ ). The new result after replacing IV into the model of Table 5.5 is presented in Table 5.8:

**Table 5.8: Explanatory model for TRS post-lime (ppm) with IV during Annual period**

	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
<i>Summary</i>	0.7421	0.5507	0.5418	0.046055898		
	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
<i>ANOVA Table</i>						
Explained	5	0.657733134	0.131546627	62.0168	< 0.0001	
Unexplained	253	0.536649878	0.002121146			
	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
<i>Regression Table</i>					Lower	Upper
Constant	-0.177903592	0.034821302	-5.1090	< 0.0001	-0.246480135	-0.109327048
Blend(Pri:Sec)	0.034349804	0.009725666	3.5319	0.0005	0.015196226	0.053503382
SE.BL*MLSS	1.06533E-05	2.07664E-06	5.1301	< 0.0001	6.56364E-06	1.4743E-05
<b><u>Predicted Post lime temp</u></b>	0.000708878	0.000351361	2.0175	0.0447	1.69139E-05	0.001400842
(DT = 1)	-0.015346898	0.0107471	-1.4280	0.1545	-0.036512074	0.005818278
TRS Pri	0.065067996	0.006662666	9.7661	< 0.0001	0.051946642	0.07818935

Consequently, we can compare output of Table 5.5 to output of Table 5.8. The output of Table 5.8 indicates that according to domain knowledge of wastewater treatment, all independent variables, including the predicted post lime temp, are statistically significant with the correct coefficient sign. Furthermore, all independent variables in Table 5.8 can explain 55 % of variation in TRS post –lime ( $Y_2$ ).



However, using IV improves the adjusted  $R^2$  by about 0.4% (i.e., the adjusted  $R^2$  of Table 5.8 = 54.18% and adjusted  $R^2$  of Table 5.5 = 53.81%). Although using IV has less of an impact on adjusted  $R^2$ , using IV can provide valuable information, especially on the missing relationship between  $Y_1$  and  $Y_2$  with the process variables at lime stabilization processes. For this reason, employing an IV can provide us more information and increase the efficiency of not only the explanatory but also the predictive model. The following is a discussion of the results in Table 5.8.

**Blend Ratio (Blend (Pri:Sec)):** As described in all cases of Section 5.2.1, and especially in Figure 5.6, Blend Ratio is a ratio between GT TS and DAF TS. Hence, if Blend Ratio is larger than 1, it indicates that there are more GT TS than DAF TS. In the model shown in Table 5.8, Blend Ratio has a positive coefficient, implying that high Blend Ratio can contribute to increasing  $Y_2$ . In other words, if Blend Ratio is larger than 1, it is most likely to cause high  $Y_2$  concentration after lime stabilization process. By holding all the other variables fixed, changing Blend Ratio by 1 can cause an average change in  $Y_2$  of 0.03 ppm.

To mitigate  $Y_2$  concentration and to maintain high Cake %S at the downstream process, Section 5.2.1 suggests that DC Water ought to increase DAF %S and DAF flow in order to increase DAF TS. As a result, DC Water can still produce high Cake %S with less  $Y_2$  in biosolids.

**SE.BL\*MLSS:** This variable is an interaction between secondary east blanket level (SE.BL) and mix liquor suspended solids (MLSS). As in Case I and Case II, this variable has a positive coefficient, implying that high interaction of SE.BL\*MLSS can contribute to increase  $Y_2$  concentration. By holding all the other variables fixed,

changing SE.BL\*MLSS by 1 ft\*mg/L can cause an average change in  $Y_2$  of  $1.0653 \times 10^{-5}$  ppm. Even though the coefficient of SE.BL\*MLSS is a small number, an average amount of SE.BL\*MLSS is approximately 5000 ft\*mg/L. Therefore, any changes in SE.BL\*MLSS will impact  $Y_2$  concentration.

As described in Case I of Section 5.2.1, SE.BL\*MLSS is mass unit. Using only information of secondary blanket depth is not quite adequate. However using an interaction can provide another dimension in terms of mass product. A lower blanket level with a higher concentration of MLSS indicates that there is more impact on  $Y_2$  than a high blanket level with a lower concentration of MLSS. In addition, high SE.BL\*MLSS coupled with longer retention time of sludge at sedimentation tank can cause more septicity in sludge. As a result, it can cause more production of  $Y_2$  concentration at the downstream of the process.

To avoid high  $Y_2$  concentration after lime stabilization process, DC Water should closely monitor to assure that all equipments function properly regardless of natural incident (e.g., rain storm, snow storm, etc.). More importantly, DC Water must have a functional management plan to deal with all kinds of issues.

**The predicted post lime temperature:** This variable is a predicted value of post lime temp in Table 5.6. It represents all five process variables at lime stabilization that cannot be included in the regression model of TRS post-lime ( $Y_2$ ). Instead, all those variables are included in the error term of  $Y_2$ . Therefore, all five process variables are called instrument variables (IV). The predicted post lime temp has a positive coefficient, implying that high predicted post lime temp can contribute to increase  $Y_2$

concentration. By holding all the other variables fixed, changing fitted post lime temp by 1 °F can cause an average change in  $Y_2$  of 0.0007 ppm.

This variable has less impact on  $Y_2$  relative to coefficient value. However, it is important in terms of explanatory information. In addition, it can verify that high cake %S directly relates to post lime temp rather than to  $Y_2$  concentration. In other words, cake %S has an indirect affect on  $Y_2$  concentration.

**Dummy variable when it is winter ( $D_T=1$ ):** This variable is used to represent the difference between winter and summer period by giving differences in intercepts.

Winter model:

$$\hat{Y}_2 = -0.19325 + 0.03435(\text{Blend Ratio}) + 0.00071(\text{Fitted Post lime temp}) + 1.06533 \times 10^{-5}(\text{SE.BL} \times \text{MLSS}) + 0.06507(\text{TRS Prior lime})$$

Summer model:

$$\hat{Y}_2 = -0.1779 + 0.03435(\text{Blend Ratio}) + 0.00071(\text{Fitted post lime temp}) + 1.06533 \times 10^{-5}(\text{SE.BL} \times \text{MLSS}) + 0.06507(\text{TRS Prior lime})$$

**TRS Prior-lime ( $Y_1$ ):** Originally, this variable is the source of TRS post-lime ( $Y_2$ ).

Having said that, it is nonetheless crucial to identify any upstream process variables relating to  $Y_1$ . In this model,  $Y_1$  has a positive coefficient, implying that high  $Y_1$  can contribute to increase  $Y_2$  concentration. This statement can firmly verify Hypothesis # 1 in Chapter 1, which states that there is a high relationship between  $Y_1$  and  $Y_2$ . In other words, part of the causes or the sources of  $Y_2$  can be explained through  $Y_1$ . By holding all the other variables fixed, changing  $Y_1$  by 1 ppm can cause an average change in  $Y_2$  of 0.07 ppm.

As stated in contributions of this analysis, mitigating  $Y_1$  is most likely to have an impact on decreasing of  $Y_2$  concentration. However, not only  $Y_1$  should be considered as the main process variable contributing to increase  $Y_2$  concentration. There are also a number of process variables that can cause  $Y_2$ , as described in previous statements.

**Case V: Winter Model for TRS Post-lime ( $Y_2$ ) (January 2009 to March 2009)**

**Table 5.9: Explanatory model for TRS post-lime (ppm) w/o IV during winter period**

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.7953	0.6325	0.6085	0.008055422		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	3	0.005137068	0.001712356	26.3887	< 0.0001	
Unexplained	46	0.002984932	6.48898E-05			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	0.248256568	0.05510444	4.5052	< 0.0001	0.137337084	0.359176052
WAS %S	-0.052073517	0.010238671	-5.0860	< 0.0001	-0.072682893	-0.031464142
<u>Post lime Temp</u>	0.000304345	9.06485E-05	3.3574	<b>0.0016</b>	0.000121879	0.000486811
TRS Prior-lime	0.016791306	0.005024577	3.3418	0.0017	0.006677357	0.026905254

As in the previous case, we start the analysis with standard ordinary least squares (OLS) to obtain an explanatory model. There are three process variables, which are WAS %S, Post lime temp, and TRS prior-lime ( $Y_1$ ), to be included in the model. All three variables can explain about 63.25 % of variation in TRS post-lime ( $Y_2$ ), and all of them are statistically significant. However, we can employ an instrument variable (IV) and replace post lime temp (see Table 5.9), to obtain the model shown in Table 5.10.

**Table 5.10: Explanatory model for TRS post-lime (ppm) with IV during winter**

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.7583	0.5750	0.5472	0.008663018		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	3	0.004669797	0.001556599	20.7414	< 0.0001	
Unexplained	46	0.003452203	7.50479E-05			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	0.26473271	0.065204705	4.0600	0.0002	0.133482447	0.395982973
WAS %S	-0.05501787	0.011670193	-4.7144	< 0.0001	-0.07850875	-0.03152699
<u>Predicted Post lime temp</u>	0.000276997	0.000147635	1.8762	<b>0.0670</b>	-2.01772E-05	0.00057417
TRS Prior-lime	0.017966929	0.005835715	3.0788	0.0035	0.006220245	0.029713614

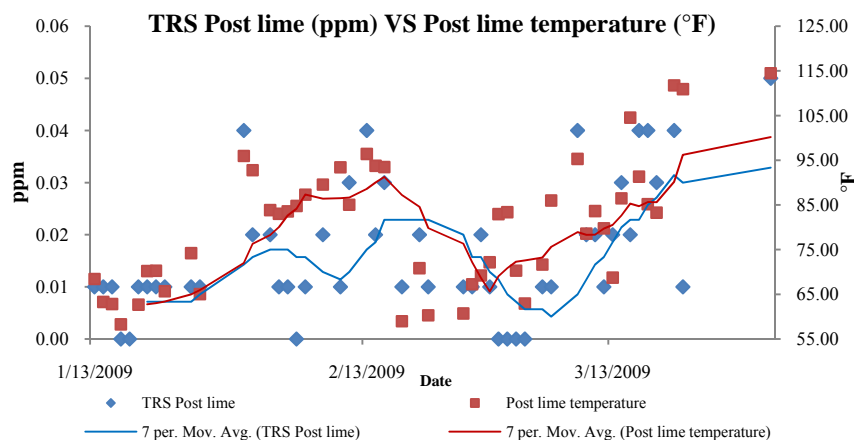
The output in Table 5.10 is the result of a regression model with an instrumental variable (IV). All of the three variables including the predicted post lime temp are still statistically significant with all coefficient signs correct according to domain knowledge of wastewater treatment. However, all three process variables can explain about 57.5% of variation in TRS post-lime ( $Y_2$ ), which is less than the model in Table 5.9 (i.e., 63.25%). However, both models have almost identical coefficient values for each of the independent variables.

However, we use the model of Table 5.9 to represent statistical model for  $Y_2$  during winter period due to a smaller standard error of estimates (SEE) (i.e., 0.0081 and 0.0087 for Table 5.9 and Table 5.10, respectively).

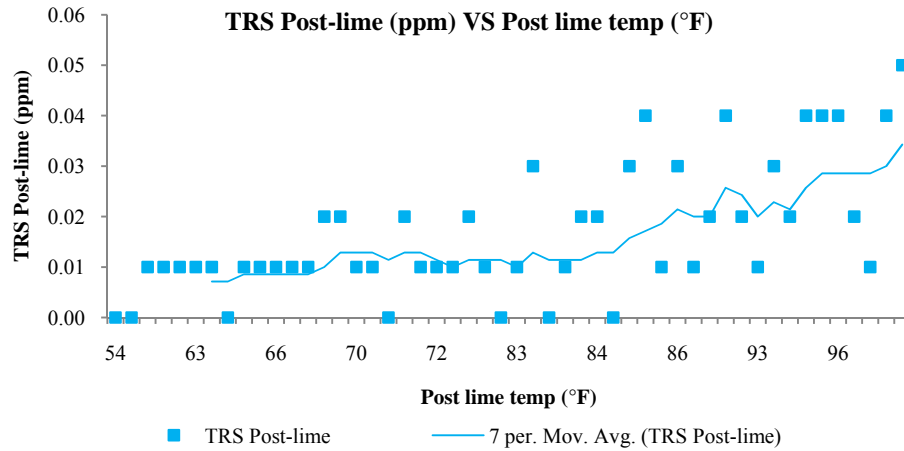
**Waste activated sludge percent solids (WAS %S):** As described in Case I of Section 5.2.1, WAS %S is biological activated sludge or aerated solids. Therefore, WAS %S can assist with mitigating TRS post-lime ( $Y_2$ ) concentration. Furthermore, it can be confirmed by this model that WAS %S has a negative coefficient, implying

that high WAS %S can help to decrease ( $Y_2$ ) concentration. By holding all the other variables fixed, changing WAS %S by 1 percent can cause an average change in  $Y_2$  of -0.05 ppm. Therefore, it is considered that WAS %S is one of significant process variables to mitigate not only TRS prior-lime ( $Y_1$ ) but also TRS post-lime ( $Y_2$ ).

**Post lime temp:** This process variable is a result of physical and chemical reaction from lime stabilization that is believed to cause the change of TRS prior-lime ( $Y_1$ ) becoming TRS prior-lime ( $Y_2$ ). As described in Case IV, post lime temp is used to represent all process variables at the lime stabilization process that cannot be included in the right hand-side of  $Y_2$ . According to the above output, post lime temp has a positive coefficient with  $Y_2$  and also has high correlation with  $Y_2$  ( $r = 0.63$ ), as shown in Figures 5.10 and 5.11.



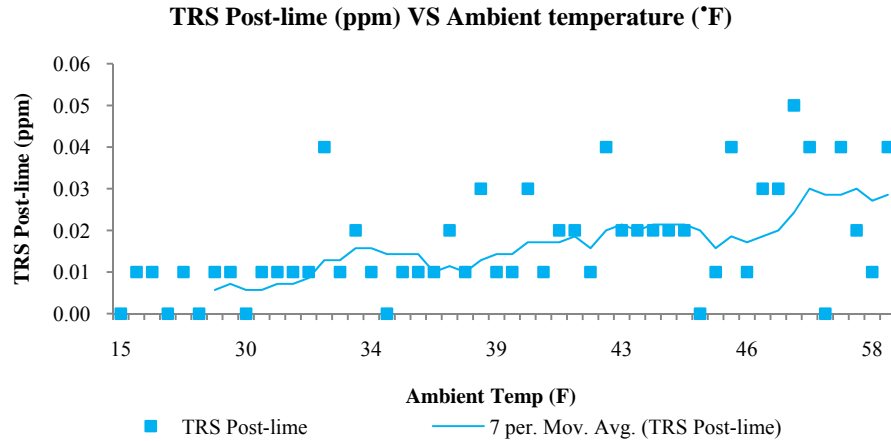
**Figure 5.10: Relationship between TRS post-lime (ppm) and post lime temp (°F)**



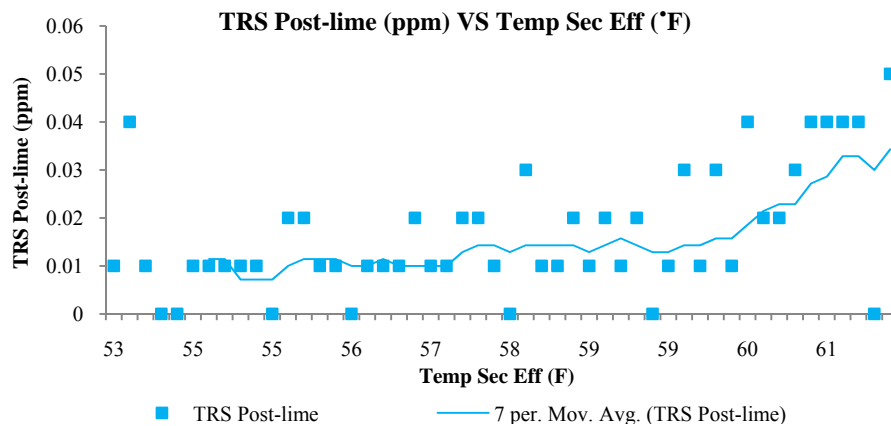
**Figure 5.11: Relationship between post lime temp (°F) and TRS post-lime (ppm)**

The positive coefficient implies that high post lime temp during winter can contribute to increase  $Y_2$  concentration. However, high post lime temp can be explained by another aspect. The high temperature at lime stabilization can decrease microbial activities. In other words, adequate high temperature at post-lime has a pasteurizing effect on dewatered solids. By holding all the other variables fixed, changes post lime temp by 1 °F can cause an average change in  $Y_2$  of 0.0003 ppm. With regard to the coefficient value comparing winter and annual period, the coefficient of winter period is less significant than the annual period in affecting  $Y_2$  concentration (i.e., annual = 0.001 and winter = 0.0003). Therefore, it is confirmed that both ambient and temperature at secondary effluent naturally help to decrease  $Y_2$  concentration during winter period.

In other words, it is most likely to have low  $Y_2$  concentration during winter. This can be verified by visually examining the relationship between ambient temperature and  $Y_2$  ( $r = 0.532$ ) and the relationship between temperature at secondary effluent and  $Y_2$  ( $r = 0.511$ ) as shown in Figures 5.12 and 5.13, respectively.



**Figure 5.12: Relationship between TRS post-lime (ppm) and Amb Temp (°F)**  
**(January 2009 – March 2009)**



**Figure 5.13: Relationship between TRS post-lime (ppm) and Temp Sec Eff (°F)**  
**(January 2009 – March 2009)**

**TRS prior-lime ( $Y_1$ ):** Like Case IV, TRS prior-lime ( $Y_1$ ) is the original concentration of TRS prior-lime ( $Y_2$ ). During winter period,  $Y_1$  has positive coefficient, implying that high  $Y_1$  can contribute to increase  $Y_2$  concentration. By holding all the other variables fixed, changing  $Y_1$  by 1 ppm can cause an average change in  $Y_2$  of 0.02 ppm. This information confirms Hypothesis #1 in Chapter 1. Apparently,  $Y_1$  is considered as one of the main sources causing higher  $Y_2$  concentration. Therefore,



mitigating  $Y_1$  is an effective direction to decrease  $Y_2$  concentration after the lime stabilization process.

**Case VI: Summer Model for TRS Post-lime ( $Y_2$ ) (June 2009 to September 2009)**

**Table 5.11: Explanatory model for TRS post-lime (ppm) during summer period**

Summary	Multiple	R-Square	Adjusted	StErr of		
	R		R-Square	Estimate		
	0.7407	0.5486	0.5208	0.055351281		
ANOVA Table	Degrees of	Sum of	Mean of	F-Ratio	p-Value	
	Freedom	Squares	Squares			
Explained	4	0.242032462	0.060508116	19.7496	< 0.0001	
Unexplained	65	0.19914468	0.003063764			
Regression Table	Coefficient	Standard	t-Value	p-Value	Confidence Interval 95%	
		Error			Lower	Upper
Constant	-0.452628502	0.09391427	-4.8196	< 0.0001	-0.64018825	-0.265068754
RAS.SE	6.19441E-05	1.75389E-05	3.5318	0.0008	2.69165E-05	9.69718E-05
SW.E. BL*MLSS S.W.E	1.34304E-05	5.48136E-06	2.4502	0.0170	2.48339E-06	2.43775E-05
GT TS	6.35973E-07	2.63697E-07	2.4118	0.0187	1.09334E-07	1.16261E-06
TRS Prior-lime	0.042975773	0.015835827	2.7138	0.0085	0.011349443	0.074602104

Table 5.11 represents the best explanatory model for TRS post-lime ( $Y_2$ ) during the summer (June 2009 to September 2009). There are four process variables in the model: return activated sludge at secondary east (RAS.SE), interaction of SWE.BL\*MLSS, gravity thickener total solids (GT TS), and TRS prior-lime. All four variables are statistically significant and have correct coefficient signs corresponding to domain knowledge of wastewater treatment. Furthermore, about 55% of the variation in TRS post-lime ( $Y_2$ ) can be explained by those four variables.

However, post lime temperature cannot be included in this model. Referring to Table 4.4, which illustrates correlation outputs between significant process variables and TRS post-lime ( $Y_2$ ), there is no correlation between post lime temp and

$Y_2$ . Furthermore, correlation between ambient temperature and  $Y_2$  is not quite strong as well ( $r = 0.306$ ). Therefore, instrumental variables (IV) will not be used in this analysis. Instead, only standard ordinary least squares (OLS) will be employed to obtain the best explanatory model.

**Return activated sludge at secondary east (RAS.SE):** Like Case III of the summer model for TRS prior-lime ( $Y_1$ ), the activated sludge has a positive coefficient, implying that high RAS.SE can contribute to increase  $Y_2$  concentration. By holding all the other variables fixed, changing RAS.SE by 1 mg/L can cause an average change in  $Y_2$  by 0.000062 ppm.

As described in Section 3.2.10 and Case III of Section 5.2.1, high concentration of RAS.SE can be the source of odorants in dewatered solids and biosolids because a high concentration of RAS.SE demonstrates a high density of settled solids at the bottom of the secondary sedimentation tank. Again, this variable is not present in the statistical model of Case V (winter period) similar to the discussion in Case III of Section 5.2.1. Therefore, high RAS.SE concentration coupled with high temperature can contribute to increase  $Y_2$  concentration. In other words, there are more septic conditions in RAS concentration during summer than winter.

**Interaction of SWE.BL\*MLSS:** This variable has a positive coefficient, implying that high SWE.BL\*MLSS can contribute to increase  $Y_2$  concentration. By holding all the other variables fixed, changing SWE.BL\*MLSS by 1 ft\*mg/L can cause an average change in  $Y_2$  of 0.000013 ppm. While the coefficient value makes SWE.BL\*MLSS seem to have less impact on  $Y_2$  concentration, the average value of

SWE.BL\*MLSS is 10,000 ft\*mg/L. Therefore, even small changes in SWE.BL\*MLSS can have a significant effect on  $Y_2$  concentration.

Like the discussion in Case I, Case II, and Case IV, this information verifies that if operators keep the blanket level high coupled with more concentration of MLSS, it can contribute to production of  $Y_2$  in biosolids. Again, using SWE.BL\*MLSS as an interaction can provide another dimension to explain the potential of secondary blanket level and concentration of MLSS impacting to  $Y_2$  concentration. In addition, high SWE.BL\*MLSS together with longer retention time of sludge at sedimentation tank can cause more septicity in sludge. As a result, this can cause more production of  $Y_2$  concentration at the downstream of the process. Additionally this analysis can offer more information regarding management perspective in Section 5.6 and in Chapter 6.

**Gravity thickener total solids (GT TS):** GT TS is a product between GT %S and GT Flow as described in Case III and Figure 5.6. It has positive coefficient, implying that high GT TS can contribute to increase  $Y_2$  concentration as expected. By holding all the other variables fixed, changing GT TS by 1 lb can cause the change in  $Y_2$  of  $6.4 \times 10^{-7}$  ppm.

$$GT\ TS(lbs / day) = \left[ \frac{GT\ flow(gallons / day) * 8.43 * GT\ \%S * 10000}{1000000} \right]$$

The above formula illustrates how to calculate GT TS. As GT %S and GT flow are the main process variables to form GT TS. Therefore, GT TS is raw total solids. In other words, GT TS can potentially contribute to high concentration of  $Y_2$ . In addition, if there are higher GT TS than DAF TS in blending tank, it will cause a blend Ratio higher than 1. Consequently, this can cause a high  $Y_2$  concentration at

lime stabilization. Practically, operators prefer higher GT TS than DAF TS in blend because it is believed that higher GT TS can provide higher Cake %S and easily to dewater.

In summary, it is mandatory for operators to maintain GT TS amounts that are not higher than DAF TS amounts (i.e., Blend Ratio = 1) without lower Cake %S at the final product. In other words, operators can compensate by increasing the DAF TS amount (i.e., increasing WAS %S and DAF flow) in order to attain the same Cake %S.

### 5.2.3 Explanatory Models for N-containing Compounds

Unlike TRS post-lime ( $Y_2$ ), our model for N-containing compounds ( $Y_3$ ) employs only ordinary least squares (OLS) estimation to obtain the best explanatory model for all three cases: Case VII (annual period), Case VIII (winter period), and Case IX (summer period). Many studies stated that N-containing compounds (i.e., ammonia and amines) can be found during and after lime stabilization process (Murthy et al. 2001; Kim et al. 2001; Murthy et al. 2002b; Novak et al. 2002; Kim et al. 2002; Kim et al. 2005; Schneekloth et al. 2006). However, this analysis identifies any potential process variables contributing to N-containing compound from upstream to the lime stabilization processes.

### Case VII: Annual Model for N-containing compounds ( $Y_3$ )

**Table 5.12: Explanatory model for N-containing (ppm) during annual period**

<i>Summary</i>	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.6533	0.4268	0.4154	5.852027964		
<i>ANOVA Table</i>	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	4	5124.994883	1281.248721	37.4128	< 0.0001	
Unexplained	201	6883.492491	34.2462313			
<i>Regression Table</i>	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	-21.03751731	4.36162886	-4.8233	< 0.0001	-29.63793632	-12.43709829
(DT = 1)	-2.556762953	1.202867381	-2.1256	0.0348	-4.928620767	-0.184905139
Post lime Temp	0.077582307	0.029715833	2.6108	0.0097	0.018987542	0.136177072
Pol_Sec	0.01445052	0.001583192	9.1275	< 0.0001	0.011328723	0.017572316
Pol_DW/DT S	0.497290039	0.179679273	2.7677	0.0062	0.142991889	0.851588189

According to the model in Table 5.12, there are three process variables, which are polymer at secondary process (Pol\_Sec), polymer of dewatering per dry ton of solids (Pol\_DW/DTS), and post lime temperature (Post lime temp), and one dummy variable for winter ( $D_T=1$ ). All are statistically significant and have correct coefficient signs. This model suggests that about 43% of variation in N-containing compound can be explained by these four independent variables.

As described in Chapter 2, an N-containing compound (i.e., ammonia and amines) is produced by decomposition of nitrogen containing compounds under both aerobic and anaerobic conditions. Generally, amino acid in protein is the original source for ammonia. Bacteria known as ammonification remove or break down these amino acids to produce ammonia. Furthermore, amines are produced by microbial decomposition by relating to decarboxylation of amino acids and proteins degradation

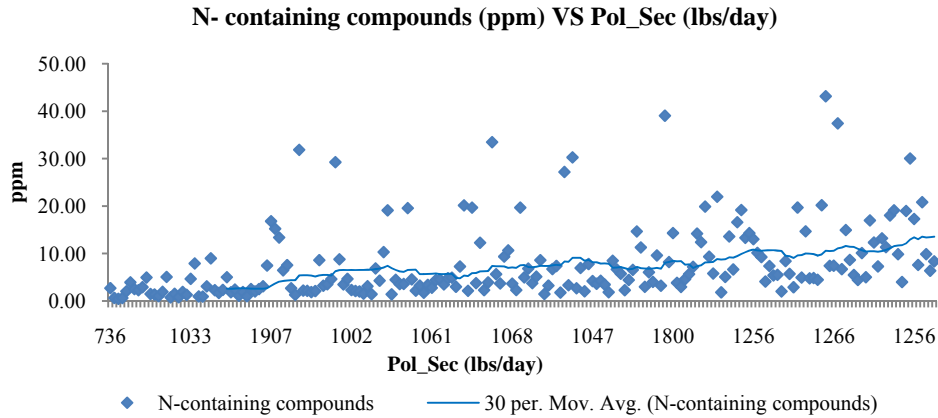
(Novak et al. 2002; Visan 2003). However, the concentration of ammonia and amines emitted mainly depend on pH, temperature, characteristics of solids.

The reason to restate how ammonia and amines can be formed in dewatered solids and biosolids in the previous paragraph is to use this statement as the main knowledge or theory leading to the next discussion of all process variables. In addition, the statement in the previous paragraph can be used to support the output of the explanatory models for N-containing compound ( $Y_3$ ).

**Polymer addition at secondary process (Pol\_Sec):** DC Water adds cationic polymers at the secondary process to enhance sludge coagulant performance by making suspended solids coagulate easily then faster settling at secondary sedimentation tank. The cationic polymers used by DC Water are copolymer of acrylamide and the methyl chloride quat of dimethylaminoethacrylate (a cationic ester of acrylic acid). However, many studies indicate that cationic polymer or cationic polyelectrolyte, which consists of cationic monomer and amino acid base, can be the source of N-containing compounds for lime stabilization process (Kim et al. 2001; Chang and Dentel 2001; Chang et al. 2005).

The study of Chang et al. (2005) indicates that copolymerizing acrylamide (AM) and acryloyloxyethyltrimethylammonium chloride (AETAC), which is quaternized from dimethylaminoethylacrylate, are significant sources of N-containing compounds (e.g., amine, trimethylamines) at lime addition process. Therefore, this statement supports and verifies the model of Table 5.12 that presents essential result.

According to correlation analysis and visual inspection, it can be verified that Pol\_Sec can contribute to increase N-containing compound ( $Y_3$ ) ( $r = 0.54$ ).

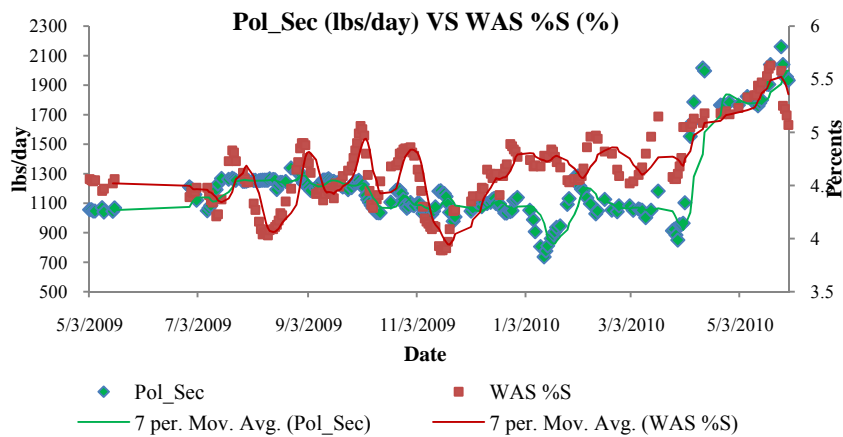


**Figure 5.14: Relationship between N-containing (ppm) and Pol\_Sec (lbs/day)**

Furthermore, the output of Table 5.12 indicates that Pol\_Sec has a positive coefficient, implying that high Pol\_Sec amount added at secondary process can contribute to increase  $Y_3$  concentration. By holding all the other variables fixed, changing Pol\_Sec by 1 lb/day can cause an average change in  $Y_3$  of 0.014 ppm.

Theoretically, when the pH is higher than 9.8 caused by the lime addition process (e.g., pH above 12), the ammonium inside the sludge can change to become ammonia ( $\text{NH}_3$ ). In addition, high post lime temperature (e.g., above 104 °F) can maximize the emission of N-containing compounds (Kim et al. 2002). Therefore, one of the reasons that the Pol\_Sec is one of process variables causing higher  $Y_3$  is cationic polymers at secondary carry over from secondary process to lime stabilization process. In other words, when operators add those polymers to enhance settling performance, they can stay inside settling solids and those solid will be pumped to blend tank, then are passed through the dewatering process and lime stabilization. When there is a lime addition process with pH above 12 coupled with high post lime temperature (e.g., above 104 °F), these can be cause high value of  $Y_3$  in biosolids.

Additionally, this discussion can be supported by Figure 5.15, which illustrates the relationship between polymer at the secondary process and waste-activated sludge percent solids (WAS %S) ( $r = 0.565$ ). Results from Figure 5.15 can verify that cationic polymer at secondary process can stay inside sludge and it can cause high  $Y_3$  concentration by polymer degradation, especially at lime stabilization when high pH is occurred.



**Figure 5.15: Relationship between Pol\_Sec (lbs/day) and WAS %S (%)**

To avoid and control high concentration of  $Y_3$  concentration, DC Water can alternatively use another cationic polymer that can prevent degradation. In addition, operators have to assure that amount of polymers added at secondary process does not overdose.

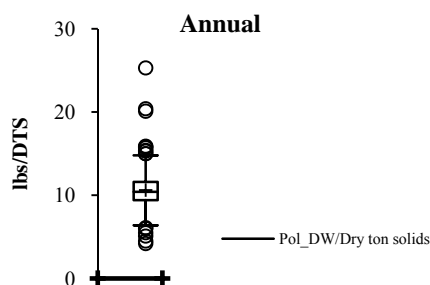
**Polymer at Dewatering per dry ton solids (Plo\_DW/DTS):** Like polymer at the secondary process, polymer in the dewatering process is a cationic polymer and has the same properties and chemical structure. This polymer is added to the dewatering process to enhance thickening and to increase dewatering performance. This research, however, analyzes this polymer by calculating the usages per total dry solids



produced. This calculation can provide more accurate information than can be found by only employing the amount of polymer added to dewatering process.

According to model output in Table 5.12, Pol\_DW/DTS has a positive coefficient, implying that high Pol\_DW/DTS amounts added at dewatering process can contribute to increase  $Y_3$  concentration. By holding all the other variables fixed, changing Pol\_DW/DTS by 1 lb/DTS can cause an average change in  $Y_3$  of 0.5 ppm. Additionally, the previous statement of polymer at secondary confirms that cationic polymer can cause high  $Y_3$  concentration. That means that excessive amount of polymers or overdose polymers added by operators can considerably contribute to high  $Y_3$  concentration at after lime addition process.

Furthermore, an average amount of polymer for DC Water to add to dewatering process is about 12 lbs/DTS as shown in Figure 5.16. Therefore, it is essential for DC Water to consider and assure that operators do not overdose amount of polymers. Case VIII and Case IX will present amount of polymer at dewatering used during winter and summer period.



**Figure 5.16: Amount of Pol\_DW (lbs/DTS) added in annual period**

**Dummy variable for winter ( $D_T=1$ ):** This variable is included in the model to provide different interceptor for winter and summer periods. The dummy variable ( $D_T=1$ ) has a negative coefficient, implying that the winter period can naturally assist to

mitigate  $Y_3$  concentration. Holding all the other variables fixed, when it is the winter season,  $Y_3$  is lower by an average of -2.6 ppm. Here are the models for the winter and summer periods:

Winter period:

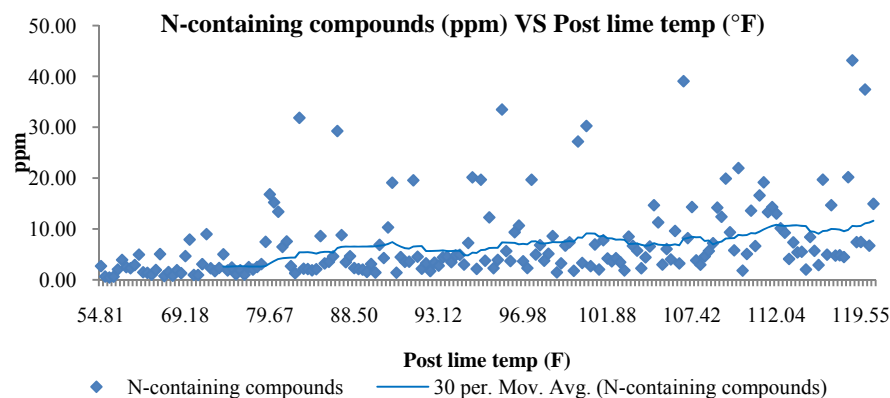
$$\hat{Y}_3 = -23.59427 + 0.07758(\text{Post lime temp}) + 0.014451(\text{Pol\_Sec}) + 0.49729(\text{Pol\_DW/DTS})$$

Summer period:

$$\hat{Y}_3 = -21.03751 + 0.07758(\text{Post lime temp}) + 0.014451(\text{Pol\_Sec}) + 0.49729(\text{Pol\_DW/DTS})$$

**Post lime temperature (Post lime temp):** As stated in the discussion in Sections 5.2.1 and 5.2.2, post lime temp is a byproduct of the chemical reaction between quick lime and dewatered solids. However, many studies (see Chapters 2 and 3) confirm that post lime temp at lime stabilization process is the key to increasing the emission of N-containing compounds ( $Y_3$ ). Theoretically, whenever the temperature is above 80 °F (27 °C), N-containing compounds ( $Y_3$ ) can easily be volatile. Figure 5.17 presents the relationship between  $Y_3$  and post lime temp. This indicates that the average of post lime temp was about 95 °F in 2009.

Therefore, as long as DC Water employs a lime addition process to stabilize dewatered solids, N-containing compounds ( $Y_3$ ) are always present. However, Figure 5.17 illustrates that during winter the post lime temperature noticeably drops to about 55 °F. As a result,  $Y_3$  concentration is lower when compared to summer. This information can confirm that it is most likely to have lower  $Y_3$  concentration during the winter period.



**Figure 5.17: Relationship between N-containing (ppm) VS post lime temp (°F)**

Post lime temp has a positive coefficient, implying that high post lime temperature can contribute to increase emission of  $Y_3$ . By holding all the other variables fixed, changes post lime temp by 1 °F can cause an average change in  $Y_3$  of 0.08 ppm.

**Case VIII: Winter Model for N-containing compounds ( $Y_3$ ) (December 2009 to March 2010)**

**Table 5.13: Explanatory model for N-containing (ppm) during winter period**

	Multiple	R-Square	Adjusted	StErr of		
Summary	R		R-Square	Estimate		
	0.5228	0.2733	0.2344	1.654450718		
	Degrees of	Sum of	Mean of	F-Ratio	p-Value	
ANOVA Table	Freedom	Squares	Squares			
Explained	3	57.6562163	19.21873877	7.0213	0.0004	
Unexplained	56	153.283602	2.737207179			
	Coefficient	Standard	t-Value	p-Value	Confidence Interval 95%	
Regression Table		Error			Lower	Upper
Constant	-7.724369172	2.350857946	-3.2858	0.0018	-12.43370353	-3.015034809
Post lime Temp	0.046262438	0.018160085	2.5475	0.0136	0.009883415	0.08264146
RAS.S.E	0.001230162	0.000385603	3.1902	0.0023	0.000457706	0.002002618
Pol_DW/DTS	0.197215755	0.109774633	1.7966	0.0778	-0.02268926	0.417120771

There are three process variables in this model: post lime temperature, return activated sludge at secondary east (RAS.SE), and polymer at dewatering process per dry ton solids (Pol\_DW/DTS). All process variables are statistically significant and have correct coefficient signs with N-containing compounds ( $Y_3$ ) with regard to domain knowledge of wastewater treatment.

The model, however, suggests that only 27.3 % of variation in N-containing compounds ( $Y_3$ ) can be explained by these three process variables. Like Case VII, two of those process variables, which are post lime temp and Pol\_DW/DTS, are included in an explanatory model.

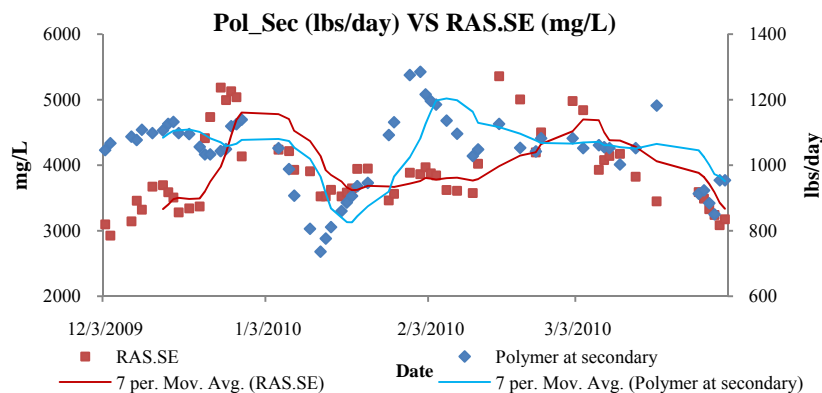
**Post lime temperature (Post lime temp):** As described in Case VII, this variable is one of the key process variables that can contribute to increase N-containing compounds ( $Y_3$ ) concentration. In this model, post lime temp has a positive coefficient, implying that high post lime temp can contribute to increase  $Y_3$  concentration. By holding all the other variables fixed, changing post lime temp by 1 °F can cause an average change in  $Y_3$  of 0.05 ppm. However, the post lime temp is insignificant to impact emission of  $Y_3$  because ambient temperature naturally assist to prevent volatile in limed biosolids. This means that DC Water can distribute biosolids to sensitive application site or close to resident areas.

**Polymer at dewatering per dry ton solids (Pol\_DW/DTS):** As stated in Case VII, Pol\_DW/DTS can be considerably anticipated the source of N-containing compounds ( $Y_3$ ). In this model, Pol\_DW/DTS has a positive coefficient sign with  $Y_3$ . This implies that high Pol\_DW/DTS amounts added in dewatering process can contribute to increase  $Y_3$ . By holding all the other variables fixed, changing Pol\_DW/DTS by 1

lb/DTS can cause an average change in  $Y_3$  by 0.2 ppm. This is another piece of information that confirms that polymer at dewatering significantly contributes to emission of  $Y_3$ .

**Return activated sludge at secondary east (RAS.SE):** This process variable has positive coefficient, implying that a high concentration of RAS.SE can contribute to increase  $Y_3$ . By holding all the other variables fixed, changing the concentration of RAS.SE by 1 mg/L can cause an average change in  $Y_3$  of 0.0012 ppm.

This research tries to identify how this process variable relates to  $Y_3$ . The exploratory data analysis (EDA) provided vital information. Since DC Water adds cationic polymers at secondary process, these polymers can accumulate inside the sludge at the sedimentation tank. Therefore, there are most likely relationships between cationic polymer and process variables at secondary sedimentation process (e.g., RAS.SE, RAS.SWO, RAS.SWE, SE.BL, SWE.BL, SWO.BL, etc). Considering correlation coefficient ( $r$ ) and visualization analysis (graphs), we have found the following crucial information. There is positive correlation ( $r$ ) between polymer at secondary process and RAS.SE ( $r = 0.201$ ).



**Figure 5.18: Relationship between Pol\_Sec (lbs/day) and RAS.SE (mg/L)**

Therefore, EDA confirms the relationship between polymer at secondary and RAS.SE. As a result, this information can clarify why RAS.SE is included in the model to explain N-containing compounds ( $Y_3$ ). Due to polymers at secondary process that accumulate inside sludge of RAS.SE, some of this sludge will be wasted to dissolved air floatation (DAF) and blending tank, respectively. The accumulated polymers can be possibly carried over from secondary process to lime stabilization process. When they reach lime addition and the condition is right, these can cause the emission of  $Y_3$  as stated in Case VII about polymer at secondary process.

In short, RAS.SE does not directly contribute to increase  $Y_3$  concentration. Instead, the accumulated cationic polymers inside the sludge cause and contribute to increase emissions of  $Y_3$ . Again, it is mandatory that operators do not overdose polymer at secondary process in order to avoid higher emission of  $Y_3$ .

**Case IX: Summer Model for N-containing compounds ( $Y_3$ ) (June 2009 to September 2009)**

**Table 5.14: Explanatory model for N-containing (ppm) during summer period**

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.5678	0.3224	0.2787	6.554906995		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	4	1267.621144	316.9052861	7.3756	< 0.0001	
Unexplained	62	2663.941954	42.96680571			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	-24.1874781	9.012854084	-2.6837	0.0093	-42.2039167	-6.171039498
S.W.O.BL	2.481853264	1.079104403	2.2999	0.0248	0.324754299	4.63895223
Post lime Temp	0.137201887	0.070421295	1.9483	0.0559	-0.003568277	0.27797205
Pol_DW/DTS	0.745779805	0.267703264	2.7858	0.0071	0.210648605	1.280911005
S.E. BL*MLSS	0.001147703	0.000474962	2.4164	0.0186	0.000198267	0.002097138

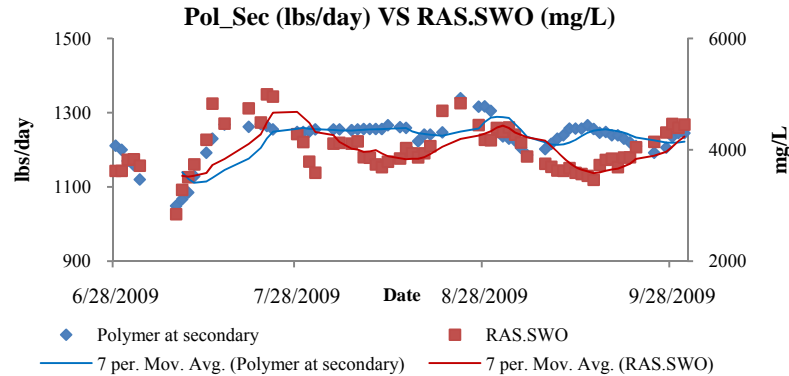
There are four process variables in the model shown in Table 5.14: secondary west odd blanket level (SWO.BL), post lime temperature, polymer at dewatering per dry ton solids (Pol\_DW/DTS), and interaction of SE.BL\*MLSS. All four variables are statistically significant and have correct coefficient sign with regard to domain knowledge of wastewater treatment. The model, however, suggests that only about 32% of variation in N-containing compounds ( $Y_3$ ) can be explained by these four process variables.

As stated in Case VII and Case VIII, it is confirmed that post lime temp, Pol\_DW/DTS can contribute to increase of  $Y_3$ . However, SWO.BL and SE.BL\*MLSS are new process variables found to affect on  $Y_3$ .

**Secondary west odd blanket level (SWO.BL) and Interaction of SE.BL\*MLSS):**

Both process variables have positive coefficient, implying that high blanket level of SWO.BL and high mass of SE.BL\*MLSS can contribute to increase  $Y_3$ . By holding all the other variables fixed, changing blanket level of SWO.BL by 1 ft and changing mass of SE.BL\*MLSS by 1 ft\*mg/L can cause average change in  $Y_3$  of 2.48 ppm and 0.001 ppm, respectively.

As shown in Case VIII, both SWO.BL and SE.BL\*MLSS are process variables at secondary process. Therefore, they could relate to cationic polymer at secondary process for some extents, and they do not directly affect on  $Y_3$  as stated in Case VIII. Correlation coefficient ( $r$ ) and visualization analysis show that there are no correlation between polymer and SWO.BL and between polymer and SE.BL\*MLSS. However, there is strong correlation between polymer and RAS.SWO ( $r = 0.498$ ) as shown in Figure 5.19.



**Figure 5.19: Relationship between Pol\_Sec (lbs/day) and RAS.SWO (mg/L) during summer period**

Figure 3.1 in Chapter 3 represents DC Water treatment process and illustrates the relationship between SWO.BL and RAS.SWO. Inside secondary west odd sedimentation, there is secondary west odd blanket level (i.e., settling concentrated sludge). Part of this settling concentrated sludge will be wasted to dissolved air flotation (DAF) and another part, which is returns activated sludge and (RAS) will be pumped back to secondary aerated tank to maintain microorganisms. Therefore, this can confirm that RAS.SWO is part of concentrated sludge at SWO.BL.

We, however, cannot find any relationships or any links between polymer at secondary and SE.BL\*MLSS to support why SE.BL\*MLSS can be included in the model. However, there is a possibility that SE.BL\*MLSS can relate to cationic polymer at secondary process even though it does not show any correlation. Like RAS.SWO, SE.BL is settling concentrated sludge at secondary east sedimentation tank, and MLSS is concentration of total suspended solids inside secondary east aerated tank that can overflow to secondary east sedimentation. Therefore, it is possible that any amount of cationic polymer added to the process can accumulate inside the sludge or mass of SE.BL\*MLSS.



In summary, we can draw conclusions of how SWO.BL and SE.BL\*MLSS relate to increasing of N-containing compounds ( $Y_3$ ). Similar to Case VIII of RAS.SE and RAS.SWO, the accumulated cationic polymers inside the sludge of SWO.BL and SE.BL\*MLSS can be carried over to the downstream. When these accumulated polymers pass through lime addition with pH above 9.8 and temperature above 80 °F, they can easily drive  $Y_3$  emission from biosolids.

### 5.3 Predictive Models

As in Section 5.2, there are three cases (i.e., annual, winter, and summer) of predictive models for each odorant compound. In total, there are nine predictive models. As described in Section 4.1.1, there are two requirements or two criteria employed to select the best predictive model: (i) model with the smallest root mean squared error (RMSE) on holdout set and (ii) model with the lowest AIC.

To gain the best explanatory model for each case, we employ the software XLMiner to create the predictive models. The most unique advantage of XLMiner is that it can build a predictive model using a training set, as well as provide accessible predictive power measure on a holdout (validation) set, simultaneously. XLMiner conventionally splits data 60/40 (i.e., 60% of the data for training data set and 40% for a validation set).

Splitting data can lead to an ineffective model if there are not ample data to split and to create a model; for instance, there are about 55 data points during the winter and summer periods. Therefore, the predictive model for the annual period of each odorant compound (i.e., Case I, Case IV, and Case VII) will have both a training set and a validation set. In contrast, there will only be a training set for the winter and

summer period of each odorant compound. Nevertheless, to assess the predictive power of each predictive model in winter and summer period, we employ cross-validation analysis. The results and discussions of cross-validation will be described again in Section 5.4 (Model Validation). Additionally, this section will present only the best predictive model for each case. The other candidate predictive models will be presented in Appendix B.

### 5.3.1 Predictive Models for TRS at Prior-Lime ( $Y_I$ )

#### Case I: Predictive model for annual period of TRS prior-lime ( $Y_I$ )

**Table 5.15: Predictive model for TRS prior-lime (ppm) during annual period**

Input variables	Coefficient	Std. Error	p-value	SS		
Constant term	2.5260975	1.1519859	0.0298818	202.00839	Residual df	148
GT FW	-1.85E-06	4.6E-07	9.826E-05	2.2398329	Multiple R-squared	0.435544
WAS %S	-0.4256331	0.1177421	0.0004111	7.1748996	Std. Dev. estimate	0.4030708
SE.BL*RAS.SE	-6.049E-05	2.092E-05	0.0044149	0.5851115	Residual SS	24.044977
Temp Sec Eff	0.0365479	0.0077	4.84E-06	1.7907034	AIC	835.6
SE. BL*MLSS	0.00012	3.214E-05	0.0002694	2.2039928		
(Temp*DT = 1)	0.0075206	0.0014197	4.2E-07	4.558979		

Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
24.04575862	0.393870444	-0.002239371

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
14.60075105	0.374689	0.01894817

This is the best predictive annual model for TRS prior-lime ( $Y_I$ ). There are six process variables, which consist of gravity thickener flow (GT FW), waste activated sludge percent solids (WAS %S), interaction of SE.BL\*RAS.SE, temperature at secondary effluent (Temp Sec Eff), interaction of SE.BL\*MLSS, and interaction of TempSecEff\* $D_T=1$ , included in this predictive model.

Additionally, the model provides the lowest validation RMSE (RMSE = 0.375) and the lowest AIC (AIC=835.6). For the annual model we have both RMSE

of training data set and RMSE of validation set (i.e., 0.3939 and 0.3747, respectively).

The RMSE of validation set is lower than RMSE of training set ( $0.3747 < 0.3939$ ).

Verifying that the annual predictive model for TRS prior-lime ( $Y_I$ ) does not overfit the training data. Furthermore, this model is parsimonious compared to another candidate models in Appendix B.

## Case II: Predictive model for winter period of TRS prior-lime ( $Y_I$ )

**Table 5.16: Predictive model for TRS prior-lime (ppm) during winter period**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.02792609	0.2463826	0.9102731	32.256512
SE.BL	0.33644986	0.1638628	0.0460234	0.1855095
RAS.SE	-0.00014061	4.723E-05	0.004717	0.2603906
Amb Temp	0.01455642	0.002347	1.7E-07	1.1900514
S.WO. BL*MLSS	0.00002952	7.36E-06	0.0002328	0.3524687
SE.BL*MLSS	-0.00009392	5.758E-05	0.1099693	0.0694777

Residual df	44
Multiple R-squared	0.6417519
Std. Dev. estimate	0.1615824
Residual SS	1.14879
AIC	16.94

The model in Table 5.16 is the best predictive model for winter period of TRS prior-lime ( $Y_I$ ). There are five process variables, which include secondary east blanket level (SE.BL), return activated sludge at secondary east (RAS.SE), ambient temperature (Amb Temp), interaction of SWO.BL\*MLSS, and interaction of SE.BL\*MLSS. The model provides the smallest RMSE (0.1515) and the smallest AIC (16.94).

### Case III: Predictive model for summer period of TRS prior-lime ( $Y_I$ )

**Table 5.17: Predictive model for TRS prior-lime (ppm) during summer period**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	3.64457941	2.236974	0.10825169	119.7866516
GT %S	0.25450373	0.14822559	0.09089159	1.04515743
DAF TS	-0.00000629	0.0000025	0.01453312	0.57240433
Cake %S	-0.25128782	0.05612963	0.00003242	3.94246984
S.W.E.BL	0.58212888	0.15188292	0.00029498	4.15345144
RAS.S.W.O	0.00057573	0.00013253	0.00005182	2.72708106
Temp Sec Eff	0.0228566	0.01737705	0.19316527	0.23004428

Residual df	63
Multiple R-squared	0.6020018
Std. Dev. estimate	0.3646449
Residual SS	8.3768501
AIC	160.783

This model in Table 5.17 is the best predictive model for summer period of TRS prior-lime ( $Y_I$ ). There are six process variables, including gravity thickener percent solids (GT %S), dissolved air flotation total solids (DAF TS), cake percent solids (Cake %S), secondary west even blanket level (SWE.BL), return activates sludge at secondary west odd (RAS.SWO), and temperature at secondary effluent (Temp Sec Eff). The model provides the smallest AIC (160.783) and the smallest RMSE (0.346).

### 5.3.2 Predictive Models for TRS at Post-lime ( $Y_2$ )

#### Case IV: Predictive model for annual period of TRS post-lime ( $Y_2$ )

**Table 5.18: Predictive model for TRS post-lime (ppm) during annual period**

Input variables	Coefficient	Std. Error	p-value	SS		
Constant term	-0.2130059	0.0455867	6.55E-06	0.5896361	Residual df	149
TRS Pri	0.0659544	0.0087952	0	0.3409052	Multiple R-squared	0.5780572
Blend(Pri:Sec)	0.0305947	0.0137123	0.0271624	0.0351426	Std. Dev. estimate	0.0492163
S.E. BL*MLSS S.E	1.207E-05	2.76E-06	2.296E-05	0.0287148	Residual SS	0.3609146
Fitted Post lime temp	0.0011472	0.0004799	0.0180696	0.0845526	AIC	-253.954
(DT = 1)	-0.020767	0.0142643	0.1475329	0.0051341		

Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
0.360914657	0.04825434	1.75401E-05

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
0.202561964	0.044132881	-0.01066482

The model in Table 5.18 is the best annual predictive for TRS post-lime ( $Y_2$ ). There are five process variables including TRS prior-lime, blend ratio, interaction of SE.BL\*MLSS, fitted post lime temperature, and dummy variable when it is winter ( $D_T=1$ ). This model provides the smallest validation RMSE (RMSE = 0.044133) and the smallest AIC (AIC=-253.954). Similar to Case I, this research can access predictive power of this model by comparing RMSE of training set and RMSE of validation set. Output of validation set suggests that the model for annual period of TRS post-lime ( $Y_2$ ) does not overfit the training data (RMSE of validation is less than RMSE of training set: 0.044132<0.048254).

**Case V: Predictive model for winter period of TRS post-lime ( $Y_2$ )****Table 5.19: Predictive model for TRS post-lime (ppm) during winter period**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.2867726	0.0590205	1.535E-05	0.013778
TRS Pri	0.0116767	0.0066711	0.0870302	0.0011469
WAS %S	-0.0593135	0.0105097	1.12E-06	0.0032587
S.W.O.BL	0.0077809	0.0029327	0.0110519	0.0001113
RAS.S.W.E	-0.0000026	1.47E-06	0.0848898	0.0003808
Fitted Post lime temp	0.0003485	0.0001882	0.0707076	0.0002332

Residual df	44
Multiple R-squared	0.6317336
Std. Dev. estimate	0.0082449
Residual SS	0.0029911
AIC	-280.624

The model in Table 5.19 is the best predictive model for the winter period of TRS post-lime ( $Y_2$ ). There are five process variables including TRS prior-lime, waste activated percent solids (WAS %S), secondary west odd blanket level (SWO.BL), return activated sludge at secondary west even (RAS.SWE), and fitted post lime temperature. This model provides the smallest RMSE (0.00773) and the smallest AIC (-280.624).

**Case VI: Predictive model for summer period of TRS post-lime ( $Y_2$ )****Table 5.20: Predictive model for TRS post-lime (ppm) during summer period**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.17571084	0.0467413	0.0003674	0.5742228
TRS Pri	0.05332687	0.0140151	0.0003161	0.1351165
RAS.S.E	0.00006615	1.628E-05	0.0001322	0.0718594
SWO.BL*RAS	-0.00001047	2.74E-06	0.0002975	0.0063159
SWO.BL*MLSS S	0.00001758	4.47E-06	0.0002067	0.0437984

Residual df	65
Multiple R-squared	0.5827369
Std. Dev. estimate	0.0532175
Residual SS	0.1840869
AIC	-110.464

The model in Table 5.20 has four process variables including TRS prior-lime, return activated sludge at secondary east (RAS.SE), interaction of SWO.BL\*RAS, and interaction of SWO.BL\*MLSS. This model provides the smallest RMSE (0.0513) and the smallest AIC (-110.464).

### 5.3.3 Predictive Models for N-containing Compounds ( $Y_3$ )

#### Case VII: Predictive model for annual period of N-containing compounds ( $Y_3$ )

**Table 5.21: Predictive model for N-containing (ppm) during annual period**

Input variables	Coefficient	Std. Error	p-value	SS		
Constant term	-26.581068	5.9264584	1.701E-05	7671.8359	Residual df	118
Blend(Pri;Sec)	3.75422859	1.8682782	0.0467703	161.21564	Multiple R-squared	0.4360306
Pol_Sec	0.01555401	0.0018966	0	2413.6448	Std. Dev. estimate	6.0208659
Pol_DAF+Pol_DW	0.00136673	0.0008619	0.1154837	283.43039	Residual SS	4277.5977
Pol_DW/DTS	0.47099641	0.2808827	0.0962203	126.53569	AIC	1732.4
(DT = 1)	-3.48496151	1.1686184	0.0034784	322.38025		

Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
4277.597855	5.873393985	2.66803E-05

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
2709.280652	5.7480438	1.074029007

There are four process variables, which include the blend ratio, polymer at secondary process (Pol\_Sec), sum of polymer at dissolved air flotation and polymer at dewatering (Pol\_DAF+Pol\_DW), and polymer at dewatering per dry ton solids (Pol\_DW/DTS), and one dummy variable for winter ( $D_T=1$ ). This model is considered to be the best predictive model for N-containing compounds ( $Y_3$ ) during annual period. The model gives the smallest validation RMSE (5.74804) and the smallest AIC (1732.4).

Nevertheless, RMSE of validation set of this model verifies that model for annual period of N-containing compounds ( $Y_3$ ) is validated. Because the RMSE of validation set is less than RMSE of training set ( $5.74804 < 5.8734$ ), the model is likely not overfitting the training data.

**Case VIII: Predictive model for winter period of N-containing compounds ( $Y_3$ )****Table 5.22: Predictive model for N-containing (ppm) during winter period**

Input variables	Coefficient	Std. Error	p-value	SS		
Constant term	-8.95781898	2.249253	0.0002054	460.76184	Residual df	54
Post lime Temp	0.06529341	0.0172074	0.0003759	8.9302416	Multiple R-squared	0.4354948
RAS.S.W.O	0.00096194	0.0002004	0.000013	32.835461	Std. Dev. estimate	1.4849654
Pol_DW/DTS	0.19674	0.1061232	0.0692212	16.826542	Residual SS	119.07661
SWO.BL*MLSS	-0.00037296	9.627E-05	0.0002917	19.832117	AIC	296.786
SE.BL*MLSS	0.00049883	0.0002021	0.0167575	13.438837		

This model is the best predictive model for winter period of N-containing compounds ( $Y_3$ ). There are five process variables included in the model. These process variables are post lime temperature, return activated sludge at secondary west odd (RAS.SWO), polymer at dewatering per dry ton solids (Pol\_DW/DTS), interaction of SWO.BL\*MLSS, and interaction of SE.BL\*MLSS. In addition, this model provides the smallest RMSE (1.4088) and smallest AIC (296.786).

**Case IX: Predictive model for summer period of N-containing compounds ( $Y_3$ )****Table 5.23: Predictive model for N-containing (ppm) during summer period**

Input variables	Coefficient	Std. Error	p-value	SS		
Constant term	-18.11300278	8.9057169	0.04618162	8600.490234	Residual df	63
Post lime Temp	0.12375169	0.07252789	0.09288455	232.6981964	Multiple R-squared	0.264612982
Pol_DW/DTS	0.83004338	0.27406377	0.00355824	407.0157471	Std. Dev. estimate	6.77439213
SE.BL*MLSS	0.00140841	0.00047668	0.00439817	400.6286926	Residual SS	2891.220459
					AIC	539.952

This model is the best predictive model for summer period of -containing compounds ( $Y_3$ ). There are three process variables, which are post lime temperatures, polymer at dewatering per dry ton solids (Pol\_DW/DTS), and interaction of SE.BL\*MLSS. This model also provides the smallest RMSE (6.5691) and the smallest AIC (539.952).



The next section will present how to validate both explanatory models of Section 5.2 and predictive models of Section 5.3. Furthermore, the next section also accesses predictive power of explanatory models and uses the results to compare the predictive models and determine whether or not those explanatory models can be use as predictive models.

#### 5.4 Model Validation

To validate both explanatory and predictive models, we employ cross-validation analysis. Cross-validation (like a validation set) is mainly used to access or estimate how precisely a predictive model will perform in practice, but it can be useful when t measuring and reporting the predictive power of all explanatory models. Evidently, explanatory models are considered to be valid because all independent variables included in models are statistically significant and have correct coefficient sign regarding to domain knowledge of wastewater treatment (theory-driven). But, they need to access predictive power regardless of whether or not they can be used as practical predictive models.

**Cross-validation (CV) analysis:** cross-validation is one of many validation strategies used to validate predictive models. The idea of this analysis is to test the model by using data that are not used to fit the model, similar to validation set in Section 5.3.

Section 4.2 describes how to employ cross-validation analysis, and Sections 5.4.1 to 5.4.6 will present validation of explanatory models and validation of predictive models. Table 5.24 summarizes the validation results between the root mean squared error (RMSE) of training data and RMSE of validation data for both explanatory and predictive models in each period. Therefore, the discussion in

Sections 5.4.1 to 5.4.6 will refer to Table 5.24. However, the details how to obtain RMSE of training set and validation set for each model are illustrated in Appendix C.

**Table 5.24: Summary of validation results**

Model	Explanatory Models		Predictive Models	
	RMSE of Training	RMSE of Cross-Validation	RMSE of Training	RMSE of Cross-Validation
Annual TRS Prior-lime ( $Y_1$ )	0.4004	0.41233	0.39387	0.374689
Winter TRS Prior-lime ( $Y_1$ )	0.168191	0.176497	0.151578	0.184483
Summer TRS Prior-lime ( $Y_1$ )	0.350651	0.384728	0.345933	0.382111
Annual TRS Post-lime ( $Y_2$ )	0.045519	0.046633	0.048254	0.044133
Winter TRS Post-lime ( $Y_2$ )	0.008331	0.008954	0.007734	0.013933
Summer TRS Post-lime ( $Y_2$ )	0.053358	0.059449	0.0512817	0.058819
Annual N-containing ( $Y_3$ )	5.78057	6.02966	5.873394	5.7480438
Winter N-containing ( $Y_3$ )	1.59835	1.72006	1.408762	1.55932
Summer N-containing ( $Y_3$ )	6.305579	7.38086	6.56906	7.26188

#### 5.4.1 Validation of Explanatory Models of TRS Prior-Lime

##### **Validation of explanatory model of TRS prior-lime ( $Y_I$ ) during annual period:**

The results of the RMSE of the training data and validation data (see Table 5.24) confirm that the explanatory model of TRS prior-lime ( $Y_I$ ) during annual period can be used as a predictive model and does not overfit the training data. Because the difference between RMSE of original model and RMSE of validation is practically insignificant (i.e.,  $0.412327 - 0.400400476 = 0.011927$ ), the model is said to be validated. In other words, the explanatory model of  $Y_I$  during annual period is said to have predictive power. Therefore, this model can be used for a prediction. As a result, this explanatory model can also be used for prediction.

In addition, the magnitude of the RMSE of validation is 0.412, so it means that the prediction errors from the explanatory model of  $Y_I$  during annual period will have approximately a standard deviation of 0.412 ppm of  $Y_I$ .

##### **Validation of explanatory model of TRS prior-lime ( $Y_I$ ) during winter period:**

The results of the RMSE of training data and validation data (see Table 5.24) confirm that the explanatory model of TRS prior-lime ( $Y_I$ ) during the winter period can be used as a predictive model. Because the difference between the RMSE of the original model and RMSE of validation is practically insignificant (i.e.,  $0.176497 - 0.168191201 = 0.008306$ ), the model is said to be validated. In other words, the explanatory model of  $Y_I$  during winter period is said to verify well or to have predictive power. Therefore this model can be used as a practical prediction. Furthermore, this model is said to serve the two goals of this research, which are explaining and forecasting. The magnitude of the RMSE of validation is 0.176, so it

means that the prediction errors from the explanatory model of  $Y_I$  during winter period will have approximately a standard deviation of 0.176 ppm of  $Y_I$ .

**Validation of explanatory model of TRS prior-lime ( $Y_I$ ) during summer period:**

The RMSE of training data and the RMSE of validation (see Table 5.24) indicate that the explanatory model of the TRS prior-lime ( $Y_I$ ) during the summer period has slightly greater differences (i.e.,  $0.384728 - 0.350651 = 0.034077$ ) than in the previous two periods. However, the difference in RMSE is the difference between predicted values and actual values of  $Y_I$  the summer. That means that a RMSE of 0.034077 equals mean square error (MSE) of 0.001161. In other words, an average of the errors between predicted values and actual values of  $Y_I$  is 0.001161 ppm, which is practically small. Therefore, this difference is insignificant.

As a result, the model is shown to be validated. In other words, the explanatory model of  $Y_I$  during the summer period is said to be verified or to have predictive power. Therefore, this model can be used prediction. In addition, the magnitude of the RMSE of validation is 0.385, so it means that the prediction errors from the explanatory model of  $Y_I$  during summer period will have approximately a standard deviation of 0.385 ppm of  $Y_I$ .

#### 5.4.2 Validation of Explanatory Models of TRS Post-Lime

**Validation of explanatory model of TRS post-lime ( $Y_2$ ) during annual period:**

The results of Table 5.24 verifies that there a minimal difference between RMSE of training data and RMSE of validation data (i.e.,  $0.046633 - 0.045519 = 0.001114$ ). Therefore, the model is said to be validated and can be used as predictive model or practical prediction. In other words, an explanatory model of TRS post-lime

( $Y_2$ ) during the annual can provide not only explaining function but also predicting function as well. In addition, the magnitude of the RMSE of validation is 0.047, so it means that the prediction errors from the explanatory model of  $Y_2$  during annual period will have approximately a standard deviation of 0.047 ppm of  $Y_2$ .

**Validation of explanatory model of TRS post-lime ( $Y_2$ ) during winter period:**

The difference between RMSE of training data and RMSE of validation data in Table 5.24, which is 0.000623 (0.008954-0.008331), verifies that an explanatory model of TRS post-lime ( $Y_2$ ) during winter is validated. Therefore, we can verify that the model has strong predictive power. In addition, the magnitude of the RMSE of validation data is 0.009, so it means that the prediction errors from the explanatory model of  $Y_2$  during winter period will have approximately a standard deviation of 0.009 ppm of  $Y_2$ .

**Validation of explanatory model of TRS post-lime ( $Y_2$ ) during summer period:**

According to the RMSE of training data and RMSE of validation data (see Table 5.24), there is a small difference in the value of the RMSE (0.00609). Therefore, the model is said to be validated and can be successfully verified. In other words, this model can be practically used as predictive model. In addition, the magnitude of the RMSE of validation is 0.059, so it means that the prediction errors from the explanatory model of TRS post-lime ( $Y_2$ ) during summer period will have approximately a standard deviation of 0.059 ppm of  $Y_2$ .

### 5.4.3 Validation of Explanatory Models of N-containing Compounds

#### **Validation of explanatory model of N-containing ( $Y_3$ ) during annual period:**

The result of RMSE of training data and validation data (see Table 5.24) illustrate that the difference of RMSE is about 0.24909 (6.02966 - 5.78057). Based on our judgment, this RMSE value is acceptable. Because the RMSE of 0.249088 equals MSE of 0.062045, it indicates that there are no significant differences between predicted values and actual value of N-containing compounds ( $Y_3$ ) during annual period. In other words, an explanatory model of  $Y_3$  works very well in predictive function. Therefore, the model is said to be validated and can be used as prediction. The magnitude of the RMSE of validation is 6.03, so it means that the prediction errors from the explanatory model of  $Y_3$  during annual period will have approximately a standard deviation of 6.03 ppm of  $Y_3$ .

#### **Validation of explanatory model of N-containing ( $Y_3$ ) during winter period:**

The difference between the RMSE of training data and the RMSE of validation data suggests that an explanatory model of N-containing compounds ( $Y_3$ ) during the winter period is validated. Because the difference value of RMSE is 0.12171 (1.72006 – 1.59835), there is an insignificant difference between the original model and validation model. Therefore, the explanatory model of  $Y_3$  during the winter period is not only the best explanatory model but also the most practical predictive model. In addition, the magnitude of the RMSE of validation is 1.72, so it means that the prediction errors from the explanatory model of  $Y_3$  during winter period will have approximately a standard deviation of 1.72 ppm of  $Y_3$ .

**Validation of explanatory model of N-containing ( $Y_3$ ) during summer period:**

According to the RMSE of training data and RMSE of validation (see Table 5.24), the difference between both values is 1.07528 ( $7.38086 - 6.30558$ ). This is the biggest difference of RMSE in validation section. It suggests that the model is not quite validated and that it might be overfitting the training data. In other words, an explanation of N-containing compounds ( $Y_3$ ) during the summer period cannot serve as practical prediction or it has low predictive power. Therefore, this model is used only for explanatory purpose. In addition, the magnitude of the RMSE of validation is 7.381, so it means that the prediction errors from the explanatory model of  $Y_3$  during summer period will have approximately a standard deviation of 7.381 ppm of  $Y_3$ .

The next three sections present the validation of predictive models. As mentioned in Section 5.3, the annual model of each odorous compound contains ample data to split into two parts, training data set and holdout (validation) data set. Therefore, the fit model from training data set can be validated by the hold out data set called validation data set. As a result, if the RMSE of validation set smaller than the RMSE of training data set, the model is said to be validated.

However, there are inadequate data points during winter and summer to split into two parts. Therefore, to validate those predictive models during the winter and summer, a cross-validation analysis similar to the validation of all explanatory models in Sections 5.4.1 to 5.4.3 is employed.



#### 5.4.4 Validation of Predictive Models of TRS Prior-Lime

##### **Validation of predictive model of TRS prior-lime ( $Y_I$ ) during annual period:**

The RMSE outputs (Table 5.24) illustrate that the annual predictive model of TRS prior-lime ( $Y_I$ ) has the RMSE of validation set smaller than the RMSE of the training set ( $0.374689 < 0.39387$ ). Therefore, the annual predictive model of  $Y_I$  is said to be validated. The magnitude of the RMSE of validation is 0.375, so it means that the prediction errors from the predictive model of  $Y_I$  during annual period will have approximately a standard deviation of 0.375 ppm of  $Y_I$ .

##### **Validation of predictive model of TRS prior-lime ( $Y_I$ ) during winter period:**

According to Table 5.24, the difference between the RMSE of training model and validation model of winter predictive model of TRS prior-lime ( $Y_I$ ) is 0.032905, which is very small. Therefore, this model is said to be validated and to verify well. In other words, this model can be fully used as a predictive model of  $Y_I$  during the winter period. The magnitude of the RMSE of validation is 0.184, so it means that the prediction errors from the predictive model of  $Y_I$  during winter period will have approximately a standard deviation of 0.184 ppm of  $Y_I$ .

##### **Validation of predictive model of TRS prior-lime ( $Y_I$ ) during summer period:**

The difference between the RMSE of training data and the RMSE of the validation data is 0.036178 ( $0.382111 - 0.345933$ ) (see Table 5.24). This difference is insignificant. In other words, RMSE of original model and RMSE of validation model are close together. Therefore, the model is said to be validated and can be used as a predictive model in this research. In addition, the magnitude of the RMSE of validation is 0.382, so it means that the prediction errors from the predictive model of

TRS prior-lime ( $Y_1$ ) during summer period will have approximately a standard deviation of 0.382 ppm of  $Y_1$ .

#### 5.4.5 Validation of Predictive Models of TRS Post-Lime

##### **Validation of predictive model of TRS post-lime ( $Y_2$ ) during annual period:**

Like the annual predictive model of the TRS prior-lime ( $Y_1$ ), the annual predictive model of the TRS post-lime ( $Y_2$ ) has both a training set and validation set, in which the RMSE of validation set is smaller than RMSE of training set ( $0.044133 < 0.04825$ ). Therefore, the model is said to be validated and can be employed as a predictive model in this research. The magnitude of the RMSE of validation is 0.044, so it means that the prediction errors from the predictive model of  $Y_2$  during annual period will have approximately a standard deviation of 0.044 ppm of  $Y_2$ .

##### **Validation of predictive model of TRS post-lime ( $Y_2$ ) during winter period:**

According to the output of the RMSE, the difference between the RMSE of the training data and the RMSE of validation data is 0.006199 (see Table 5.24), which is of minimal value. Therefore, there is a practically insignificant difference between the RMSE of original model and the  $RMSE_v$  of validation. The summer predictive model of TRS post-lime ( $Y_2$ ) is said to be validated and to verify successfully. The magnitude of the RMSE of validation is 0.014, so it means that the prediction errors from the predictive model of  $Y_2$  during winter period will have approximately a standard deviation of 0.014 ppm of  $Y_2$ .

##### **Validation of predictive model of TRS post-lime ( $Y_2$ ) during summer period:**

The summer predictive model of TRS post-lime ( $Y_2$ ) has the RMSE of training data smaller than the RMSE of the validation data (see Table 5.24) ( $0.0512817 <$

0.058819). However, the difference is practically insignificant (0.007537). In other words, the model is said to be validated and can be used for practical prediction in this research. In addition, the magnitude of the RMSE of validation is 0.059, so it means that the prediction errors from the predictive model of  $Y_2$  during summer period will have approximately a standard deviation of 0.059 ppm of  $Y_2$ .

#### 5.4.6 Validation of Predictive Models of N-containing Compounds

##### **Validation of predictive model of N-containing ( $Y_3$ ) during annual period:**

Like the annual predictive model of the TRS prior-lime and the annual predictive model of TRS post-lime, the annual predictive model of N-containing compounds ( $Y_3$ ) has a training data set to fit the model, and also has validation set to validate the model as well. The outputs in Table 5.24 indicate that the RMSE of validation set is smaller than the RMSE of training set ( $5.7480438 < 5.873394$ ). Therefore, the annual predictive model of  $Y_3$  is said to be validated and to verify well. In addition, the magnitude of the RMSE of validation is 5.748 so it means that the prediction errors from the predictive model of  $Y_3$  during annual period will have approximately a standard deviation of 5.748 ppm of  $Y_3$ .

##### **Validation of predictive model of N-containing ( $Y_3$ ) during winter period:**

The results of RMSE indicate that the RMSE of the validation data is higher than the RMSE of the training data (see Table 5.24), which is  $1.55932 > 1.408762$ . However, the difference is practically small or insignificant (0.150558). Therefore, the winter predictive model of N-containing compounds ( $Y_3$ ) is said to be validated. The magnitude of the RMSE of validation is 1.559 so it means that the prediction

errors from the predictive model of  $Y_3$  during winter period will have approximately a standard deviation of 1.559 ppm of  $Y_3$ .

**Validation of predictive model of N-containing ( $Y_3$ ) during summer period:**

The difference between the RMSE of the validation data and the training data for summer predictive model of N-containing compounds ( $Y_3$ ) in Table 5.24 is 0.69282 (7.26188 – 6.56906), which is the biggest difference among those predictive models. This suggests that the summer predictive model of  $Y_3$  is invalidated. The magnitude of the RMSE of validation is 7.262 so it means that the prediction errors from the predictive model of  $Y_3$  during summer period will have approximately a standard deviation of 7.262 ppm of  $Y_3$ .

The next section presents the best explanatory and predictive models for each odorant compound in each period. However, explanatory models can be predictive models if they have higher predictive powers than those of the original predictive models.

5.5 Model Selection

This section first presents the best explanatory models for each odorous compounds based on model requirements of Section 4.1. More importantly, all the best explanatory models are theory-driven with regard to domain knowledge of wastewater treatment. Selection of the best predictive models, however, we compare the validation results of both explanatory and predictive models to identify the best predictive models. As we employ cross-validation to access predictive power of all the best explanatory models, some of them can be the best predictive models if they

have more predictive powers than those of other predictive models. The details for all explanatory and predictive models are shown in Appendix D.

#### **A. The best explanatory model for TRS prior-lime ( $Y_1$ )**

Annual explanatory model of TRS prior-lime:

$$\hat{Y}_1 = -2.02373 - 0.38889(\text{WAS \%S}) + 0.0584(\text{TempSecEff}) + 2.691 \times 10^{-5}(\text{SWO.BL*MLSS}) + 0.1648(\text{GT\%S}) + 0.0076(\text{TempSecEff}*D_T) \quad (5.1)$$

Winter explanatory model of TRS prior-lime:

$$\hat{Y}_1 = -0.2839 + 0.00011(\text{MLSS.SWO}) + 0.01504(\text{AmbTemp}) \quad (5.2)$$

Summer explanatory model of TRS prior-lime:

$$\hat{Y}_1 = 4.8692 - 6.2923 \times 10^{-6}(\text{DAF TS}) - 0.23734(\text{Cake \%S}) + 0.630(\text{SWE.BL}) + 0.0006(\text{RASS.SWO}) + 0.2112(\text{GT \%S}) \quad (5.3)$$

#### **B. The best explanatory model for TRS post-lime ( $Y_2$ )**

Annual explanatory model of TRS post-lime:

$$\hat{Y}_2 = -0.1779 + 0.0344(\text{Blend Ratio}) + 0.0007(\text{Predicted Post lime temp}) + 1.0653 \times 10^{-5}(\text{SE.BL*MLSS}) - 0.0154(D_T = 1) + 0.0651(\text{TRS Prior lime}) \quad (5.4)$$

Winter explanatory model of TRS post-lime:

$$\hat{Y}_2 = 0.2647 - 0.0550(\text{WAS \%S}) + 0.00028(\text{Predicted Post lime temp}) + 0.01387(\text{TRS Prior lime}) \quad (5.5)$$

Summer explanatory model of TRS post-lime:

$$\hat{Y}_2 = -0.4526 + 6.194 \times 10^{-5}(\text{RAS.SE}) + 1.343 \times 10^{-5}(\text{SWE.BL*MLSS}) + 6.3597 \times 10^{-7}(\text{GT TS}) + 0.04298(\text{TRS Prior}) \quad (5.6)$$

### C. The best explanatory model for N-Containing compounds ( $Y_3$ )

Annual explanatory model of N-containing compounds:

$$\hat{Y}_3 = -21.03751 - 2.55676(D_T = 1) + 0.07758(\text{Post lime temp}) + 0.014451(\text{Pol\_Sec}) \\ + 0.49729(\text{Pol\_DW/Dry ton solids}) \quad (5.7)$$

Winter explanatory model of N-containing compounds:

$$\hat{Y}_3 = -7.7244 + 0.0463(\text{Post lime temp}) + 0.0012(\text{RAS.SE}) \\ + 0.1972(\text{Pol\_DW/DT solids}) \quad (5.8)$$

Summer explanatory model of N-containing compounds:

$$\hat{Y}_3 = -24.1875 + 2.4819(\text{SWO.BL}) + 0.1372(\text{Post lime temp}) \\ + 0.7458(\text{Pol\_DW/DT solids}) + 0.0011(\text{SE.BL*MLSS}) \quad (5.9)$$

### D. The best predictive model for TRS prior-lime ( $Y_I$ )

Annual predictive model of TRS prior-lime

$$\hat{Y}_1 = 2.5261 - 1.85 * 10^{-6}(\text{GTFW}) - 0.426(\text{WAS\%S}) - 6.05 * 10^{-5}(\text{SE.BL*RAS.SE}) \\ + 0.0366(\text{TempSecEff}) + 0.0001(\text{SE.BL*MLSS}) + 0.0075(\text{Temp Sec Eff}*D_T) \quad (5.10)$$

Winter predictive model of TRS prior-lime

$$\hat{Y}_1 = 0.02793 + 0.33645(\text{SE.BL}) - 0.00014(\text{RAS.SE}) + 0.01456(\text{Amb Temp}) \\ + 0.00003(\text{SWO.BL*MLSS}) - 0.00009(\text{SE.BL*MLSS}) \quad (5.11)$$

Summer predictive model of TRS prior-lime

$$\hat{Y}_1 = 3.6446 + 0.2545(\text{GT\%S}) - 0.000006(\text{DAF TS}) - 0.25129(\text{Cake\%S}) \\ + 0.58213(\text{SWE.BL}) + 0.00058(\text{RAS.SWO}) + 0.0229(\text{TempSecEff}) \quad (5.12)$$

### **E. The best predictive model for TRS post-lime ( $Y_2$ )**

Annual predictive model of TRS post-lime

$$\hat{Y}_2 = -0.21301 + 0.065954(\text{TRS Prior}) + 0.030595(\text{Blend Ratio}) + 1.207 \times 10^{-5}(\text{SE.BL*MLSS}) + 0.00035(\text{Fitted Post lime}) - 0.02077(D_T=1) \quad (5.13)$$

Winter predictive model of TRS post-lime

$$\hat{Y}_2 = 0.28677 + 0.01168(\text{TRS Prior}) - 0.059314(\text{WAS \%S}) + 0.00778(\text{SWO.BL}) - 0.000003(\text{RAS.SWE}) + 0.00035(\text{Fitted Post lime}) \quad (5.14)$$

Summer predictive model of TRS post-lime

$$\hat{Y}_2 = -0.17571 + 0.05333(\text{TRS Prior}) + 0.00007(\text{RAS.SE}) - 0.00001(\text{SWO.BL*RAS}) + 0.000018(\text{SWO.BL*MLSS}) \quad (5.15)$$

### **F. The best predictive model for N-containing compounds ( $Y_3$ )**

Annual predictive model of N-containing compounds

$$\hat{Y}_3 = -26.5811 + 3.75423(\text{Blend Ratio}) + 0.01555(\text{Pol_Sec}) + 0.00137(\text{Pol_DAF+Pol_DW}) + 0.471(\text{Pol_DW/DTsolids}) - 3.485(D_T=1) \quad (5.16)$$

Winter predictive model of N-containing compounds

$$\hat{Y}_3 = -8.9578 + 0.0653(\text{Post lime temp}) + 0.001(\text{RAS.SWO}) + 0.197(\text{Pol_DW/solids}) - 0.0004(\text{SWO.BL*MLSS}) + 0.001(\text{SE.BL*MLSS}) \quad (5.17)$$

Summer predictive model of N-containing compounds

$$\hat{Y}_3 = -18.113 + 0.12375(\text{Post lime temp}) + 0.83004(\text{Pol_DW/DTsolids}) + 0.00141(\text{SE.BL*MLSS}) \quad (5.18)$$

In general, both explanatory and predictive models mostly have the same included process variables on the right-hand side. Table 5.24 verifies that all predictive models are useful for setting the benchmark of good prediction. In

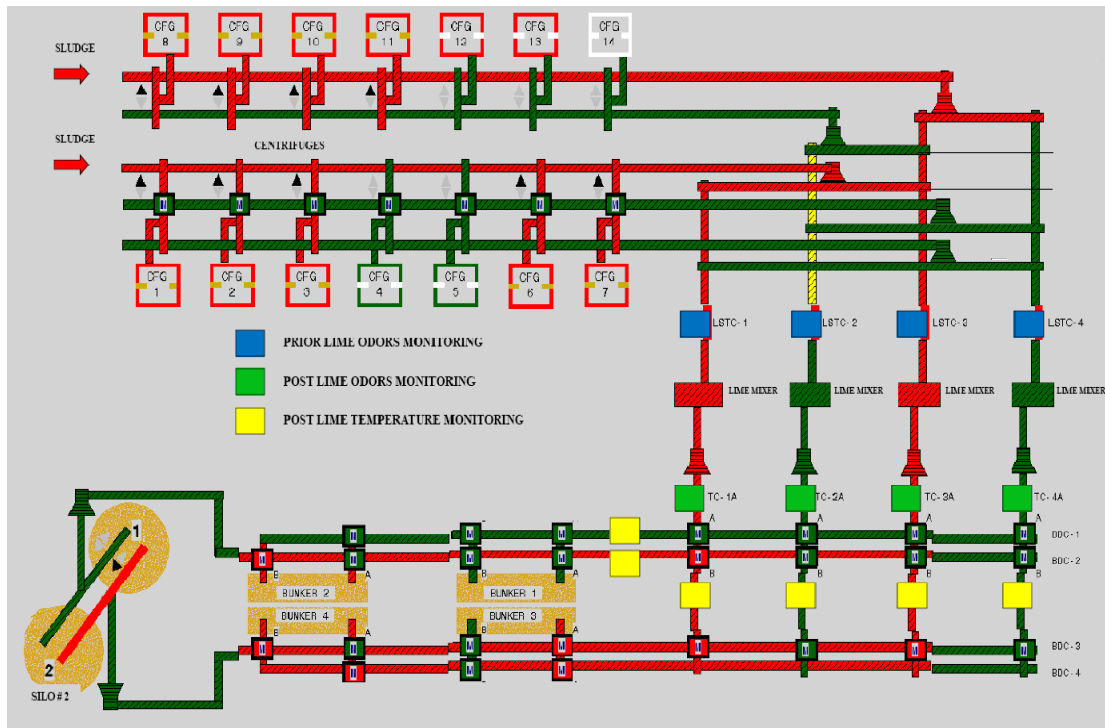
addition, RMSE of validation models for all predictive models are smaller than RMSE of validation models for predictive models. However, all explanatory models except summer explanatory model of N-containing compounds can be used for prediction, because the differences between validations RMSE and training RMSE of explanatory models are insignificant.

The next section presents and discusses feedback gained from real-time solids odors monitoring. The information includes unusual or incident events that caused odors to spike in both the dewatering process (prior-lime addition) and lime stabilization process (post-lime addition).

#### 5.6 Special Cases of Operations

Real-time feedback of odors incidents is one of the objectives in this research. As described in Section 3.3.3, real-time solids odors monitoring systems are installed inside solids processing building (SPB) at prior-lime and post-lime addition. The entirety of the discussions in this section refer to Figure 5.20.



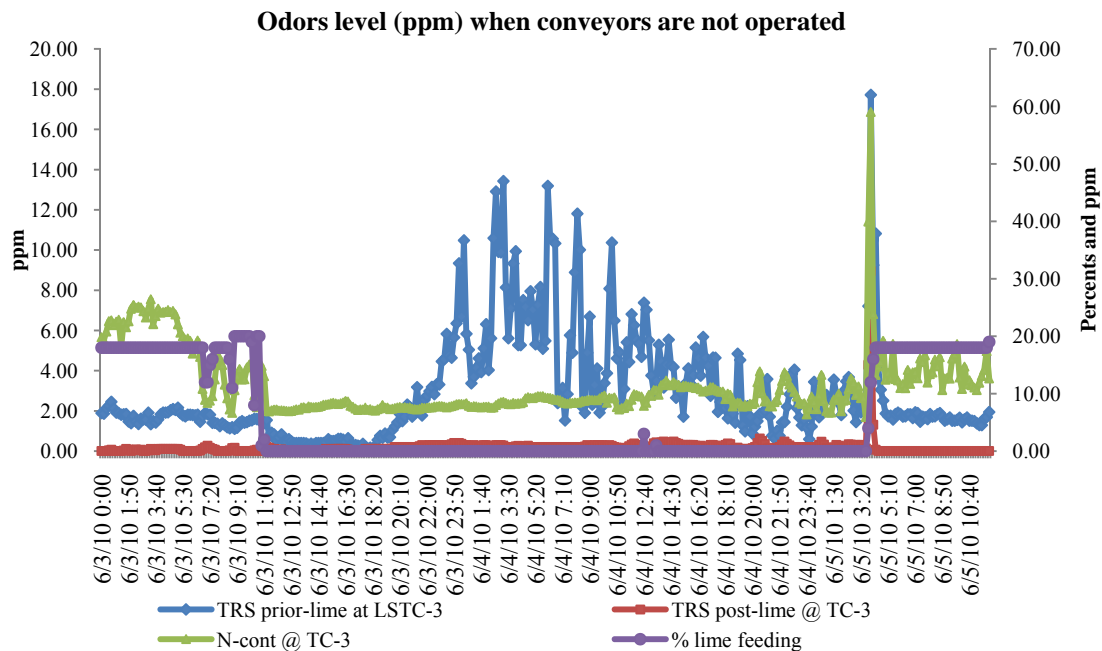


**Figure 5.20: Dewatering and lime stabilization process inside SPB**

Inside SPB, there are two sides of centrifuges, west side (Sharples) and east side (Westfalia), and each side has seven centrifuges as shown in Figure 5.20. At the dewatering process area, however, there are four conveyors (trains) used to transfer dewatered solids to the lime stabilization and the biosolids load-out facility (bunkers and silos). Most of the time, operators will run two conveyors and have another two for backup. The following are the plausible conditions that cause odors incidents and post lime temperature incidents.

In this case, if there are any problems in dewatering and lime stabilization at solids processing building (SPB), real-time solids odors monitoring can provide decisive information and evidence to prove that the high odors incidents have occurred. For instance, Figure 5.20 illustrates activities of conveyor # 3 on June 3<sup>rd</sup>,

2010 and the following are feedback information and discussions gained from real-time odors monitoring;



**Figure 5.21: Odors (ppm) incidents during dewatering problems**

#### 5.6.1 Odor Incidents When Dewatering Processes having Problems

According to Figure 5.21, the blue line represents TRS prior-lime, red line represents TRS post-lime, green line represents N-containing compounds, and purple line represents % lime feeding at lime addition process. Conveyor # 3 has operated continuously until 10:40 AM, when it was out of service. The % lime feeding had completely stopped, which indicates no dewatered sludge fed or passed lime mixer. However, in case conveyor # 3 had suddenly stopped regarding to any unexpected circumstances, there might be some limed biosolids left over inside conveyor between LSTC-3 and TC-3.

After conveyor # 3 has been stopped for several hours, there are a few incidents have been occurred inside conveyor # 3.

**High TRS prior-lime has been occurred (blue line)**

Before 10:40 AM, TRS prior-level was about 1.8 ppm, which was an usual level. However, when conveyor # 3 had stopped working, it started dropping to almost 0 ppm because no sludge moved passed this point. But, there were some solids still inside conveyor and those solids were dewatered solids. This took about 8 hours (10:40 AM to 18:20 PM); TRS prior-lime started climbing up and reached the highest level at 14 ppm for several points. When conveyor # 3 was off for a day and a half, the level of TRS prior-lime was between 4 ppm and 10 ppm.

This evidence verifies that whenever the conveyor has been stopped for several hours, solids left over inside conveyor can become septic. In addition, when operators suddenly started putting conveyor # 3 back in service, the TRS prior-lime level jumped from 2 ppm to 18 ppm within the first 10 minutes of operating, as shown in Figure 5.21. This circumstance is crucial because all the old solids that had sat inside conveyor # 3 right after centrifuges until the TRS prior-lime gas detector location (LSTC-3) are high septicity solids.

As soon as the screw conveyor starts pushing all those old solids pass the gas detector, these can cause high TRS prior-lime for several minutes. Then, TRS prior-lime will be back to normal. This finding is one of the most practical contributions of this research and valuable of real-time feedback information. Ignorance of this information can cause a significant issue for DC Water. Because those amounts of solids are the most septic solids, they can still cause high odors of TRS post-lime (red

line), even they are limed by a lime stabilization process with pH above 12.

Therefore, operators have to aware that if they start operating any conveyor that contains old solids left inside, they have to make sure that they feed an adequate % of lime to those solids.

In addition, operators have to report to Solids Foreman and General Foreman on that day to be aware of this circumstance. The Solids Foreman has to know what bunker of those old solids went to so that he can inform the Foreman at the sludge loading station to avoid that bunker. Therefore, the Foreman at sludge loading station can assign a contractor to send those solids to composting sites or remote sites. However, if operators or Foremen are unaware of these incidents, those solids can provoke complaints of high odor and lead to rejection from inspectors. Consequently, DC Water has to pay extra money for a contractor to haul biosolids to another remote location away from the residents.

#### **High TRS post-lime has been occurred (red line)**

If there is an incident happening, it will cause high odors not only the TRS prior-lime but also the TRS post-lime. The TRS post-lime has a relationship with the TRS prior-lime, which can be confirmed by Figure 5.21 when conveyor # 3 was restarted on June 5<sup>th</sup> at 03:20 AM. However, when conveyor # 3 was out of service on June 3<sup>rd</sup>, there was no high TRS post lime occurring. This can be clarified by process diagram of Figure 5.20.

The locations of all TRS post-lime and N-containing compounds gas detectors are located after lime mixers and before being dropped in all four bunkers distribution facilities. For this reason, all dewater solids passing through all these gas detectors are

limed by the lime mixer. In other words, all solids have been passed through the lime stabilization process. As a result, there are no high odors incidents of TRS post-lime that occurred when the conveyor # 3 was out of service.

However, when operators started putting conveyors # 3 back in service, the TRS post-lime jumped from 0.1 to almost 9 ppm. This phenomenon can be explained again by using Figure 5.20 and some of the previous discussion. As stated earlier in this section, any old dewatered solids after lime mixer were limed biosolids. But, the main concern was of the existence of any old dewatered solids between after centrifuges and before lime mixers. Those old dewatered solids were septic solids and they were unknown how many left in conveyor # 3. Despite being limed by lime mixer, the high pH cannot completely suppress all of the TRS because of the high septic solids and high TRS prior-lime. For this reason, the TRS post-lime was high after starting the conveyor # 3.

Like TRS prior-lime, the incident of the TRS post-lime will not last long. After about 10 minutes, level of TRS post-lime goes back to normal level, which is about 0.1 ppm. This incident suggests that when operators started using a conveyor that was out of service for some period of time (like this case only one day and a half), the leftover dewatered solids inside that conveyor can cause high TRS post-lime in spite of high % of limes were fed in lime mixer. Therefore, operators, Foreman, and General Foreman have to be made aware of this incident. All operators, the Foreman, and the General Foreman have to know what bunker received these biosolids. Therefore, as soon as it can be identified, operators have to change the destination of new biosolids output to another bunker distribution and inform

Foreman of loading station about all the incidents and what bunker distribution receive those high odor biosolids.

**High N-containing compounds has been occurred (green line)**

As in the TRS prior-lime and the TRS post-lime, right after conveyor # 3 was stopped, the N-containing level was dropped from 15 ppm to about 7 ppm. However, an average N-containing concentration constantly maintains about 7 ppm. This may be caused by the limed biosolids that were leftover inside conveyor # 3. An important incident occurred when operators started using conveyor # 3 again.

Similar to TRS prior-lime and TRS post-lime, the N-containing concentration jumped from 10 ppm to about 60 ppm within about 10 minutes. This happened when operators start putting conveyor # 3 back on service, and the lime mixer will be work as soon as the lime feeder sensor can detect dewatered solids; otherwise, it will be whiteout<sup>2</sup>. Therefore, the operator cannot start lime speed before starting the conveyor. However, when the operators started conveyor # 3 and then started the lime feeder, there were some old dewatered solids left inside. For this reason, if operators increase the speed % of lime when there is a small amount of dewatered solid that sensor can be detected, it will cause high N-containing concentration by swiftly changing of pHs.

Nevertheless, the level of N-containing concentration will be back to normal within 10 minutes, just like other odorant compounds. It is necessary for operators to add more lime at the beginning of the startup of the conveyor. They have to ensure that there is an ample lime amount added to those dewatered solids that are left inside

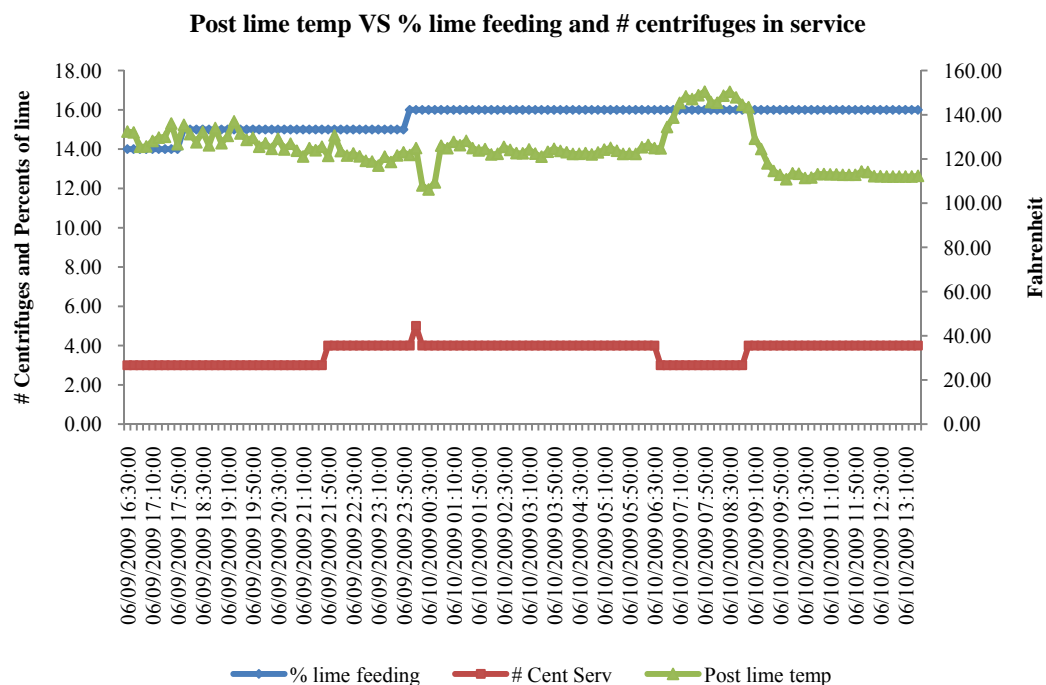
---

<sup>2</sup> Whiteout happens when the lime feeder starts feeding lime to the lime mixer without dewatered solids passing through the mixer.

the conveyer. But, it is difficult to avoid high N-containing compounds concentration incident because changing pH rapidly can force biosolids to emit N-containing compound, especially ammonia (NH<sub>3</sub>).

### 5.6.2 Post Lime Temperature Incidents

Referring to Figure 5.20, there are six post lime temperature sensors installed on each conveyor and on two bunker distribution conveyor (BDC-1 and BDC-2), and they are located after TRS post-lime and N-containing compounds gas detectors. The following are considered to cause high and low of post lime temperature incidents.



**Figure 5.22: Post lime incidents**

Figure 5.22 illustrates significant information on low and high post lime incidents. The operators put 3 centrifuges from Sharples to conveyor # 1 on June 9<sup>th</sup>, 2009 at 16:30 PM. Before post lime incidents happened, operators employed 14 % of lime feeding and then increased it to 15% in the next 2 hours at 18:50 PM. They

maintained 15 % of lime feeding for 5 hours and then changed to 16% at 23:50 PM. They kept 16% of lime feeding all day on June 10<sup>th</sup>. In the meantime, they have put one more centrifuge on service at 21:50 PM, so total number was 4 centrifuges. They added one more centrifuge to become 5 centrifuges, but it was just about 10 minutes before it went back to 4 centrifuges again.

The post lime temp incidents began declining when operators put more centrifuges to conveyor # 1, but they maintained the same % of lime feeding. Another incident of post lime dropping was when operators added one more centrifuge to become 5 centrifuges and increase one percent more. Both incidents suggest that whenever operators add number centrifuges in service and keep the same amount of lime, it can cause dropping on post lime temperature. Conversely, when operators lowered number of centrifuges but maintained the same % of lime feeding, this could cause high post lime temperature incident. This suggests that operators overdose the amount of lime to dewatered solids. As the model of post lime temp has found that # centrifuge in service is one of independent variables contributing to increase of post lime temp, it has become necessary for operators to avoid high temperature incidents like those in this case.

Apparently, an overdose % of lime can make DC Water waste a high amount of lime and will cost in lime stock. In addition, DC Water has to pay more for hauling costs because adding more amounts of lime can increase the weight of biosolids. Therefore, contractors have to use more trucks to haul biosolids. In other words, contractors' trucks haul not only biosolids but also haul extra lime with them. Once



again, this feedback is essential information similar to the feedback of solids odors monitoring.

### 5.6.3 Implication of Special Cases of Operations

Real-time solids odors feedback is a vital approach that any wastewater industry can duplicate and apply to their processes. The first and foremost benefit of contribution from this research is that it is cost effective. Real-time solids odors monitoring verifies that odors incidents will be happened when operators start a conveyor that has been out of service for a long time. Therefore, the DC Water by Biosolids Management Section can inform contractors to send those septic biosolids to composting sites or remote areas without any complaints or rejection from inspectors.

But, without that vital information and awareness of those odors incidents, and if contractors send septic biosolids to application sites where are surrounded by residents and randomly inspect by inspectors, DC Water can get complaints from residents and septic biosolids will be denied by inspectors. As a result, DC Water has to pay extra money for contractors to put applied biosolids back on a truck and haul them back to the plant or send them to composting sites. That can be a lot of works and place a strain budgets. Therefore, real-time solids odors feedback can grant the decisive contributions and are cost-effective for DC Water. To implement real-time solids odors feedback to solids handling process and biosolids management program, however, we will provide more details in Chapter 6.

The next chapter provides conclusions and suggestions for future research, especially regarding how DC Water can deal with any odor incidents or improve solids handling process in order to mitigate biosolids odors.

## Chapter 6: Conclusions and Future Work

In this dissertation, statistical analyses were employed to identify process variables influencing odorous compounds emitted from dewatered solids and biosolids. Additionally, they were used to develop statistical models to explain relationship between biosolids odor production and upstream process variables. By using DC Water as the case study, any wastewater treatment plants with similar type of unit operations can apply the decisive conclusion from this dissertation to improve their biosolids management program. Furthermore, the standard operating procedure (SOP) and future work are stated.

### 6.1 Conclusions

This dissertation focuses on two missions: (i) identifying and explaining associations between upstream process variables and three odorous compounds, which are the total reduced sulfur at prior-lime addition (TRS prior-lime), the total reduced sulfur at post-lime addition (TRS post-lime), and nitrogen containing compounds at post lime addition (N-containing compounds); and (ii) predicting odorants levels of those three compounds before hauling biosolids to land application sites. Additionally, this dissertation will provide practical conclusions about odors of real-time solids monitoring systems to benefit the biosolids management program of DC Water.

First, data analyses have found that all process variables at secondary process (i.e., secondary blanket level, return activated sludge, gravity thickening solids, waste activated sludge, dissolved air flotation, and blend ratio) associate with both TRS at prior-lime and TRS at post-lime. During the annual period, WAS %S, DAF Flow, and

DAF TS can assist DC Water to mitigate TRS at prior lime and post lime addition. In other words, increasing DAF TS in blend ratio can help to mitigate odors in dewatered solids and biosolids at the downstream process. In addition, temperature of flow at secondary effluent and ambient temperature can naturally assist to decrease both TRS at prior-lime and post-lime. Conversely, temperature at the secondary effluent, ambient temperature, and post lime temperature have positive correlation with N-containing compounds. This means that all three temperatures contribute to increases of N-containing compounds emissions in biosolids at the downstream processes. However, cationic polymers at the secondary and cationic polymers in the dewatering process are the main process variables that have strong association with N-containing compounds.

During the winter period, ambient temperature is a key variable. It has a strong association with all three odorants compounds. Again, WAS %S and DAF TS have negative correlations with TRS. On the contrary, Blend Ratio, GT %S, and GT TS are positively correlated with TRS. As during the annual period, polymer addition at secondary process and polymer addition at dewatering process also dominate N-containing compounds during the winter; only polymer at the dewatering process has a correlation with N-containing compounds. In addition, RAS.SE and RAS.SWO have positive correlation with N-containing compounds. These suggested that accumulated polymer added at secondary process can be carried over by sludge from the secondary process to downstream process and can cause the emission of N-containing compounds.

During the summer period, all process variables at the secondary sedimentation process coupled with ambient temperature dominate the emission of TRS prior-lime and TRS post-lime. Those process variables at secondary sedimentation are SE.BL, SWO.BL, SWE.BL, RAS.SE, RAS.SWO, and RAS.SWE. Nevertheless, using only the secondary blanket level does not offer ample information to explain association. Therefore, the other six interaction variables at secondary sedimentation including the interaction of SE.BL\*RASS, interaction of SWE.BL\*RASS, interaction of SWO.BL\*RASS, interaction of SE.BL\*MLSS, interaction of SWE.BL\*MLSS, and interaction of SWO.BL\*MLSS have a strong relationship with TRS prior-lime and TRS post-lime. Like N-containing compounds during the annual period, polymer addition at secondary process and polymer addition at dewatering process have strong positive correlation with N-containing compounds during the summer period.

In addition, the following conclusions can be clarified all four hypotheses:

**Hypothesis 1:** *The total reduced sulfur (TRS) at the prior-lime process positively correlates with TRS at the post-lime process.* This hypothesis was confirmed by the results of correlation between TRS at prior-lime and TRS at post-lime process. Additionally, the results of the statistical models in Chapter 5, demonstrate that the higher TRS at prior-lime, the higher are TRS levels at post-lime.

**Hypothesis 2:** *The causes of TRS at the post-lime process can be explained via the relation between upstream processes and the total reduced sulfur (TRS) at the prior-lime process.* This hypothesis can be verified by the outputs of all explanatory models in Chapter 5, especially in Sections 5.2.1 and 5.2.2. The explanatory models

of TRS at prior-lime confirm that any upstream process variables included in the models of TRS at prior-lime also have high correlation with TRS at post-lime. However, some of them cannot be included in explanatory models of TRS at post-lime. To explain relationship between upstream process variables and TRS at post-lime, therefore, we can explain by using relationship between upstream process variables and TRS at prior-lime process.

**Hypothesis 3:** *High temperature of the secondary effluent and ambient temperature can contribute to the emission of TRS and N-containing compounds.* The correlation analysis and statistical models in Chapter 5 for both TRS compounds and N-containing compounds confirmed this hypothesis. All statistical models indicated that both secondary effluent and ambient temperatures are among the significant process variables to contribute to the emission of TRS compounds and N-containing compounds. As a result, there are more chances to have higher emission of those odorous compounds during summer periods compared to winter periods. Furthermore, the ambient temperature can affect the temperature at secondary effluent as well. However, although winter periods have below-freezing temperatures, the temperature at secondary effluent is always warmer. In contrast, while summer periods have an average temperature of above 90 °F, the temperature at secondary effluent is always cooler by 20 °F.

**Hypothesis 4:** *Dewatered solids with high percent solids content can raise the temperature of limed biosolids and promote the emission of both N-containing compounds and TRS concentration during the lime stabilization.* By employing a correlation analysis, we found that there are high positive correlations between cake

percent solids and the post-lime temperature of biosolids. Statistical models of post-lime temperature of biosolids especially indicated that cake percent solids (Cake %S) is one of the process variables contributing to high temperatures in limed biosolids. In other words, by measuring the temperature of limed biosolids with the same amount of lime feeding, the dewatered cake with high percent solids shows a higher temperature than the dewatered cake with low percent solids.

## 6.2 Explanatory Models for DC Water

For explanatory and predictive models, ordinary least squares (OLS) estimation, instrumental variable (IV), and two-stages least squared (2SLS) were used to create both explanatory and predictive models. In addition, IV estimation was found to be an effective statistical tool to be used to clarify the relationship between lime stabilization processes and TRS post-lime. The following are the best models;

### **A. Explanatory models of TRS prior-lime**

Annual explanatory model of TRS prior-lime:

$$\hat{Y}_1 = -2.02373 - 0.38889(\text{WAS \%S}) + 0.0584(\text{TempSecEff}) + 2.691 \times 10^{-5}(\text{SWO.BL*MLSS}) + 0.1648(\text{GT\%S}) + 0.0076(\text{TempSecEff*D}_T) \quad (6.1)$$

Winter explanatory model of TRS prior-lime:

$$\hat{Y}_1 = -0.2839 + 0.00011(\text{MLSS.SWO}) + 0.01504(\text{AmbTemp}) \quad (6.2)$$

Summer explanatory model of TRS prior-lime:

$$\hat{Y}_1 = 4.8692 - 6.2923 \times 10^{-6}(\text{DAF TS}) - 0.23734(\text{Cake \%S}) + 0.630(\text{SWE.BL}) + 0.0006(\text{RASS.SWO}) + 0.2112(\text{GT \%S}) \quad (6.3)$$

WAS %S and DAF TS are the key factors in mitigating TRS prior-lime. Increasing WAS %S and DAF TS inside the blending tank automatically decreases the blend ratio and decreases the percent solids (cake %S) in biosolids. DC Water, however, can maintain the same target of high cake %S by increasing more WAS %S and DAF Flow. As a result, DC Water not only gains higher cake %S but also mitigates TRS after the dewatering process. In addition, employment of dummy variables can provide the differences between winter and summer season by giving differences interceptors.

All the above models suggest that process variable at the secondary sedimentation (i.e., SWE.BL, SWO.BL\*MLSS, MLSS.SWO, and RAS.SWO) are sources that can contribute to increase TRS at prior-lime. To practically avoid high TRS prior-lime concentrations at the downstream location, operators have to control secondary blanket levels to not exceed 1.8 ft and ensure that all equipment (e.g., pumps stations, centrifuges in service, etc) functions properly in order to prevent longer retention times of settling sludge. In addition, mechanical problems are one of the factors that cause backup and longer retention times of secondary blanket inside sedimentation tanks. As a result, there are more septicities, high biosolids odors production at downstream location, and lower cake %S.

It is mandatory for operators to handle all equipment properly by following all maintenance rules of equipment use. In addition, it is necessary for all foremen to communicate with upper level management by reporting any unusual circumstances that can cause equipment to be out of service. However, there can be uncontrollable circumstances, such as rain and snow storms, that can cause high secondary blanket

levels, high concentrations of mixed liquor suspended solids (MLSS), and high concentrations of return activated solids (RAS). Therefore, these can contribute to TRS prior-lime, as well.

Assessing the predictive power of all three explanatory models by using cross-validation analyses verified that all three explanatory models have been validated and can be used for prediction.

### **B. Explanatory models of TRS post-lime**

Annual explanatory model of TRS post-lime:

$$\hat{Y}_2 = -0.1779 + 0.0344(\text{Blend Ratio}) + 0.0007(\text{Predicted Post lime temp}) + 1.0653 * 10^{-5}(\text{SE.BL*MLSS}) - 0.0154(D_T = 1) + 0.0651(\text{TRS Prior lime}) \quad (6.4)$$

Winter explanatory model of TRS post-lime:

$$\hat{Y}_2 = 0.2647 - 0.0550(\text{WAS \%S}) + 0.00028(\text{Predicted Post lime temp}) + 0.01387(\text{TRS Prior lime}) \quad (6.5)$$

Summer explanatory model of TRS post-lime:

$$\hat{Y}_2 = -0.4526 + 6.194 * 10^{-5}(\text{RAS.SE}) + 1.343 * 10^{-5}(\text{SWE.BL*MLSS}) + 6.3597 * 10^{-7}(\text{GT TS}) + 0.04298(\text{TRS Prior}) \quad (6.6)$$

Post lime temperature model:

$$\hat{W} = -30.00764 + 0.53784(\% \text{lime feed}) + 0.93128(\text{Cake \%S}) + 1.79513(\# \text{ Cent Serv}) + 0.70063(\text{Temp Sec Eff}) + 0.68369(\text{Amb Temp}) \quad (6.7)$$

All information gained from models of TRS post-lime is decisive for DC Water. The models suggested that TRS prior-lime has a strong relationship with TRS post-lime. In other words, higher TRS of dewatered solids at the prior lime addition process can cause higher TRS of limed biosolids at the post lime addition process.



The instrumental variable (IV) provides practical information to explain the relationship between the lime addition process and TRS post-lime. The IV estimation clarifies that six process variables at the lime addition process cannot be included in the model to explain TRS at post-lime except post lime temperature. Instead, all five process variables, which are percent of lime feeding (% lime feeding), number of centrifuges in service (# centrifuges), Cake %S, temperature of flow at secondary effluent, and ambient temperature, are included in the error term of TRS post-lime model. In addition, these five process variables are statistically significant in the model of post lime temperature. Therefore, they can explain TRS at post-lime via post lime temperature.

The post lime temperature has no direct effect on the TRS post-lime even though it has a strong correlation with the TRS post-lime ( $r = 0.63$ ). Instead, five process variables at the lime addition process have the potential to increase TRS at post-lime, but they cannot directly explain the TRS at post-lime because they were not statistically significant when they were included in model of TRS at post-lime. Therefore, the employment of IV can provide more information and increase the efficiency of explanatory modeling. Like the TRS prior-lime, cross-validation analysis confirms that all three explanatory models of the TRS post-lime are said to be validated and can be used for prediction.

### **C. Explanatory models of N-containing compounds**

Annual explanatory model of N-containing compounds:

$$\hat{Y}_3 = -21.03751 - 2.55676(D_T = 1) + 0.07758(\text{Post lime temp}) + 0.014451(\text{Pol\_Sec}) + 0.49729(\text{Pol\_DW/Dry ton solids}) \quad (6.8)$$

Winter explanatory model of N-containing compounds:

$$\hat{Y}_3 = -7.7244 + 0.0463(\text{Post lime temp}) + 0.0012(\text{RAS.SE}) + 0.1972(\text{Pol\_DW/DT solids}) \quad (6.9)$$

Summer explanatory model of N-containing compounds:

$$\hat{Y}_3 = -24.1875 + 2.4819(\text{SWO.BL}) + 0.1372(\text{Post lime temp}) + 0.7458(\text{Pol\_DW/DT solids}) + 0.0011(\text{SE.BL*MLSS}) \quad (6.10)$$

Cationic polymer at the secondary process (Pol\_Sec) and the cationic polymer at dewatering per dry ton solids (Pol\_DW/DT solids) coupled with post lime temperature are the sources of emission of N-containing compounds. Changing the pH of limed biosolids (pH >9.8) and increasing the post lime temperature (post lime temp > 80 °F) are the factors that contribute to increase emission of N-containing compounds, especially ammonia. The models also verified that cationic polymer agents for thickening and dewatering were the sources of N-containing compounds.

The process variables at the secondary sedimentation process, which are RAS.SE, SWO.BL, and SE.BL\*MLSS, have no direct effect on N-containing compounds, but the accumulated cationic polymers inside the sludge of RAS and the sludge of the secondary blanket level cause correlations between those process variables and N-containing compounds. In addition, they can contribute to increase the emission of N-containing compounds at the lime addition process. This is the decisive information gained from these three explanatory models of N-containing compounds.

The Equations 6.9 and 6.10 verified that the higher secondary blanket level and higher concentration of RAS not only contribute to increases in the TRS prior-

lime and TRS post-lime, but also contribute to increases in the N-containing compounds at the downstream. In other words, the higher secondary blanket level, the more polymers were added into the process. Consequently, the accumulated polymer in secondary sedimentation sludge can contribute to increases in the emission of N-containing compounds. Therefore, proper management of all process variables at the secondary sedimentation process is mandatory for DC Water.

Similar to TRS prior-lime and TRS post-lime, cross-validation analyses confirms that only annual and winter explanatory models of N-containing compounds are valid and can be used for prediction. The summer explanatory model cannot be used for prediction.

### 6.3 Predictive Models for Dissertation and DC Water

#### **D. Predictive models of TRS prior-lime**

Annual predictive model of TRS prior-lime

$$\hat{Y}_1 = 2.5261 - 1.85 * 10^{-6} (GTFW) - 0.426 (WAS\%S) - 6.05 * 10^{-5} (SE.BL * RAS.SE) + 0.0366 (TempSecEff) + 0.0001 (SE.BL * MLSS) + 0.0075 (Temp Sec Eff * D_r) \quad (6.11)$$

Winter predictive model of TRS prior-lime

$$\hat{Y}_1 = 0.02793 + 0.33645 (SE.BL) - 0.00014 (RAS.SE) + 0.01456 (Amb Temp) + 0.00003 (SWO.BL * MLSS) - 0.00009 (SE.BL * MLSS) \quad (6.12)$$

Summer predictive model of TRS prior-lime

$$\hat{Y}_1 = 3.6446 + 0.2545 (GT\%S) - 0.000006 (DAF TS) - 0.25129 (Cake\%S) + 0.58213 (SWE.BL) + 0.00058 (RAS.SWO) + 0.0229 (TempSecEff) \quad (6.13)$$

### E. Predictive models of TRS post-lime

Annual predictive model of TRS post-lime

$$\hat{Y}_2 = -0.21301 + 0.065954(\text{TRS Prior}) + 0.030595(\text{Blend Ratio}) + 1.207 \times 10^{-5}(\text{SE.BL*MLSS}) + 0.00035(\text{Fitted Post lime}) - 0.02077(D_T=1) \quad (6.14)$$

Winter predictive model of TRS post-lime

$$\hat{Y}_2 = 0.28677 + 0.01168(\text{TRS Prior}) - 0.059314(\text{WAS \%S}) + 0.00778(\text{SWO.BL}) - 0.000003(\text{RAS.SWE}) + 0.00035(\text{Fitted Post lime}) \quad (6.15)$$

Summer predictive model of TRS post-lime

$$\hat{Y}_2 = -0.17571 + 0.05333(\text{TRS Prior}) + 0.00007(\text{RAS.SE}) - 0.00001(\text{SWO.BL*RAS}) + 0.000018(\text{SWO.BL*MLSS}) \quad (6.16)$$

### F. Predictive models of N-containing compounds

Annual predictive model of N-containing compounds

$$\hat{Y}_3 = -26.5811 + 3.75423(\text{Blend Ratio}) + 0.01555(\text{Pol_Sec}) + 0.00137(\text{Pol_DAF} + \text{Pol_DW}) + 0.471(\text{Pol_DW/DTsolids}) - 3.485(D_T=1) \quad (6.17)$$

Winter predictive model of N-containing compounds

$$\hat{Y}_3 = -8.9578 + 0.0653(\text{Post lime temp}) + 0.001(\text{RAS.SWO}) + 0.197(\text{Pol_DW/solids}) - 0.0004(\text{SWO.BL*MLSS}) + 0.001(\text{SE.BL*MLSS}) \quad (6.18)$$

Summer predictive model of N-containing compounds

$$\hat{Y}_3 = -18.113 + 0.12375(\text{Post lime temp}) + 0.83004(\text{Pol_DW/DTsolids}) + 0.00141(\text{SE.BL*MLSS}) \quad (6.19)$$

All predictive models of TRS prior-lime, TRS post-lime and N-containing compounds were validated. Therefore, DC Water can employ these models to

forecast daily concentrations of those odorants compounds before hauling biosolids to the application sites.

All of the above conclusions verify that a real-time solids odor monitoring system can be applied to any type of wastewater treatment industry that would like to improve biosolids quality in terms of odor reduction. Furthermore, this system can practically offer a cost-effective biosolids management program. For example, when odor incidents have occurred due to mechanical problems in the dewatering process of DC Water, feedback information sent from real-time solids odors, monitoring can identify high odors caused by septic of old dewatered solids or septic of old biosolids. Operators can identify a destination of old biosolids, so those septic biosolids do not mix with other biosolids. Nevertheless, lacking feedback information from real-time solids odor monitoring can require DC Water to increase expenditure for hauling those biosolids back to the plants or back to composting sites in the event that they get rejected by inspectors.

Additionally, real-time solids odor system can provide feedback information and respond to high post lime temperature incidents. High post lime temperature incidents means that the operators have overdosed the % lime feeding to dewatered solids to gain limed biosolids with a pH above 12. For example, when operators operate a fewer number of centrifuges (e.g., 3 centrifuges in service), they still maintain the % lime feeding at the same rate as when 5 centrifuges are in service. For this reason, a smaller number of centrifuges running with a high % of lime feeding can cause a high post lime temperature incident. Therefore, real-time solids odor

system can send feedback to operators to adjust % lime feeding to the proper speed. However, they have to ensure that the pH of lime biosolids is higher than 12.

As a result, employment of a real-time solids odor monitoring system by DC Water can save budgets in hauling costs and lime addition costs. Furthermore, it can help DC Water to save costs in the cationic polymer addition at both the secondary process and the dewatering process as well.

#### 6.4 DC Water Standard Operating Procedure (SOP)

Standard Operating Procedure (SOP) is one of the contributions generated from this research. By using the District of Columbia Water and Sewer Authority (DC Water)'s advanced wastewater treatment plant (WWTP) as the case study, this SOP can be applied directly to DC Water.

This SOP covers odor incidents occurring at both the prior lime and the post lime addition processes. The SOP will be associated with centrifuges dewatering operation, sludge cake conveyance system, dewatered sludge loading facilities (DSLFF), and DSLFF lime feed system. A real-time solids odor system evaluates the treatment plant solids for elevated emissions of odorants emitted from dewatered solids and limed biosolids. The odor monitoring system measures when either increased or decreases solids odors are present. This measurement can be related to either process changes in the upstream processes or chemical dose changes within the WWTP.

The following procedures should be used by operators of solids process treatment:

1. When centrifuge dewatering systems and cake conveyance systems are stopped by mechanical problems, operators have to inform the Solids Foreman and the General

Foreman as soon as possible, as well as record the date, time, and conveyor number that had been operated before the mechanical problems took place.

2. After centrifuge dewatering systems and cake conveyance systems are stopped, operators have to randomly sample the old lime biosolids remaining inside of the conveyor at the location beyond the lime mixer to ensure that those limed biosolids have a pH above 12.

3. When the problem due to centrifuge dewatering and cake conveyance systems are fixed, the operators should operate blank conveyors using screw conveyors to push any leftover dewatered solids passing to the lime mixer. More importantly, the Solids Foreman has to ensure that there is adequate amount of lime feeding because these dewatered solids are high septic solids.

4. After running a blank conveyor, operators have to randomly sample the solids at the point beyond the lime mixer to ensure pH above 12.

5. Operators have to record the number that bunker receives the left over biosolids, and then inform the Solids Foreman of all necessary information.

6. The Solids Foreman has to inform the General Foreman of all information gained from operators and also submit a daily report log sheet. In addition, the Solids Foreman has to communicate with the Foreman at the loading station to ensure that the old leftover biosolids are hauled to the composting site or applied at remote sites.

7. The General Forman has to inform the Biosolids Manager of any incidents.

Therefore, the Biosolids Manager should communicate with all inspectors to be aware of odors complaints from those biosolids.

8. When a high post lime temperature has occurred, operators have to sample lime biosolids to measure pH. If the pH is higher than 12.2, they have to decrease the lime feeding speed until a pH of 12.2 is reached.

9. Normally, operators try to maintain the same operating system as the previous shift. However, when the number centrifuges in service decreases, they still keep the same lime feeding speed. For this reason, high post lime temperature incidents can occur.

In sum, real-time odors solids monitoring is a useful and decisive system that can practically assist DC Water to improve the treatment process, to improve biosolids quality, and to reduce process costs.

### 6.5 Future Work

Recommendations of future work are as follows:

1. According to the explanatory models of N-containing compounds, DC Water has to quantify the proper doses of cationic polymer in the secondary process that can provide the best settling performance.
2. Optimizing of the lime feeding speed regarding Cake %S can also be done by DC Water. The results in Section 5.2 indicate that an increase in the percent of cake solids (Cake %S) can contribute to an increase in post lime temperature. Since DC Water produces “class B” biosolids by using lime stabilization, temperatures caused by chemical reactions can have a pasteurizing effect on microorganisms. With the same of lime feeding speed, therefore, a higher percent of cake solids (Cake %S) can provide a high post lime temperature more easily than a lower Cake %S.



3. Referring to # 8 of SOP and # 2 of future work, high post lime temperatures are not only caused by an overdose of % lime feeding but also by a higher percent of cake solids (Cake %S). Therefore, it is necessary for DC Water to install real-time Cake %S monitoring or real-time density of solids monitoring to measure the quality of dry cake and measure Cake %S.

4. Real-time pH monitoring or online pH monitoring should also be done by DC Water. This can provide insight and accurate information of limed biosolids.

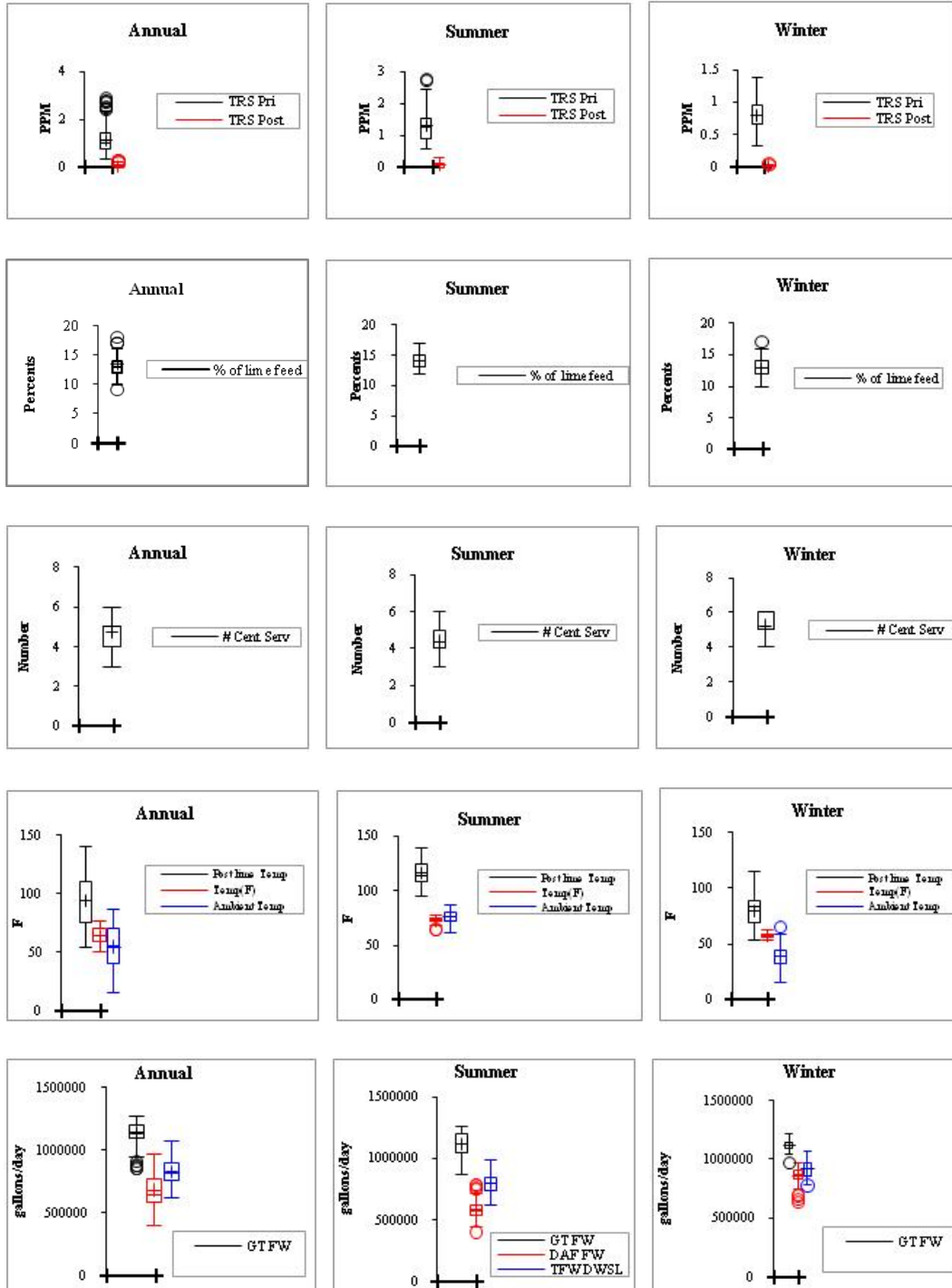
Therefore, DC Water can employ this feedback information to ensure that the pH of limed biosolids is always above 12.2 and can reduce the cost of the lime addition process.

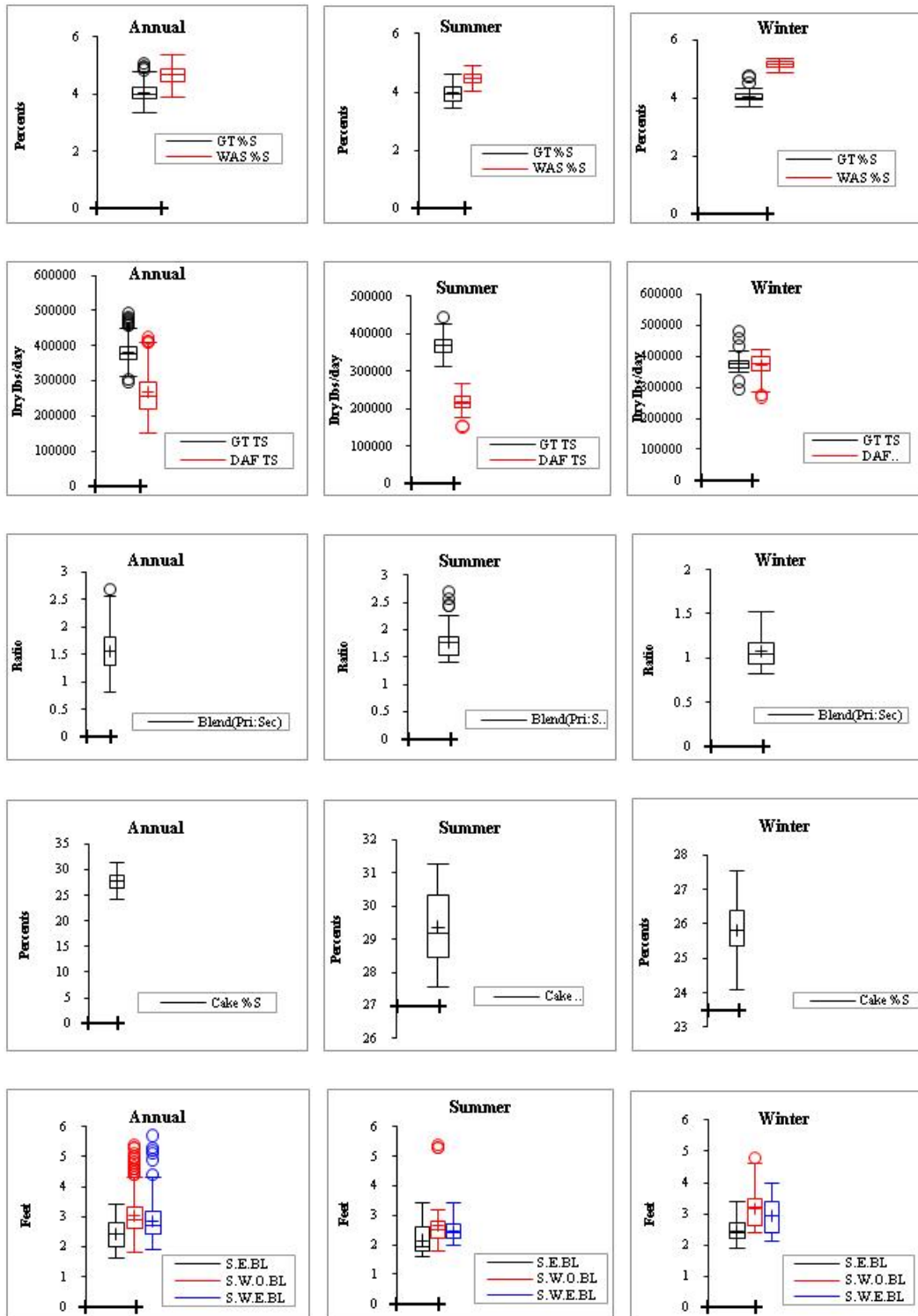
5. The employment only ordinary least squares (OLS) might not be explained some of unobserved dependent variables data. The censored regression model can be another approach to create statistical model in data analyses because the real-time solids odor system observed odorous compounds only if it above cutoff level. For instance, the minimum level of TRS and N-containing compounds that can be detected by the system is 0.01 ppm. Thus, any odors level below 0.01 ppm will be censored at zero. The familiar censored regression model is the tobit model.

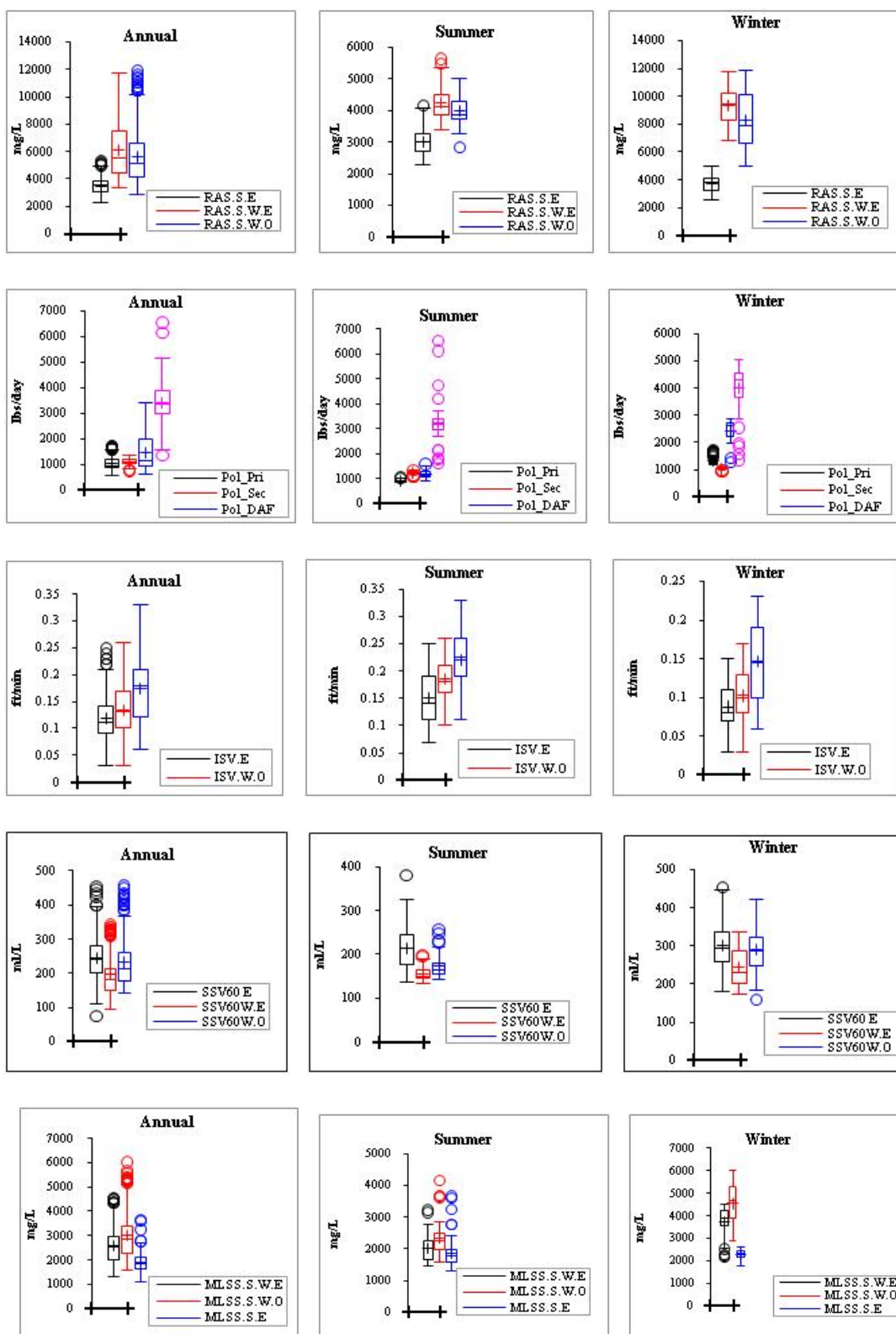
# Appendix A

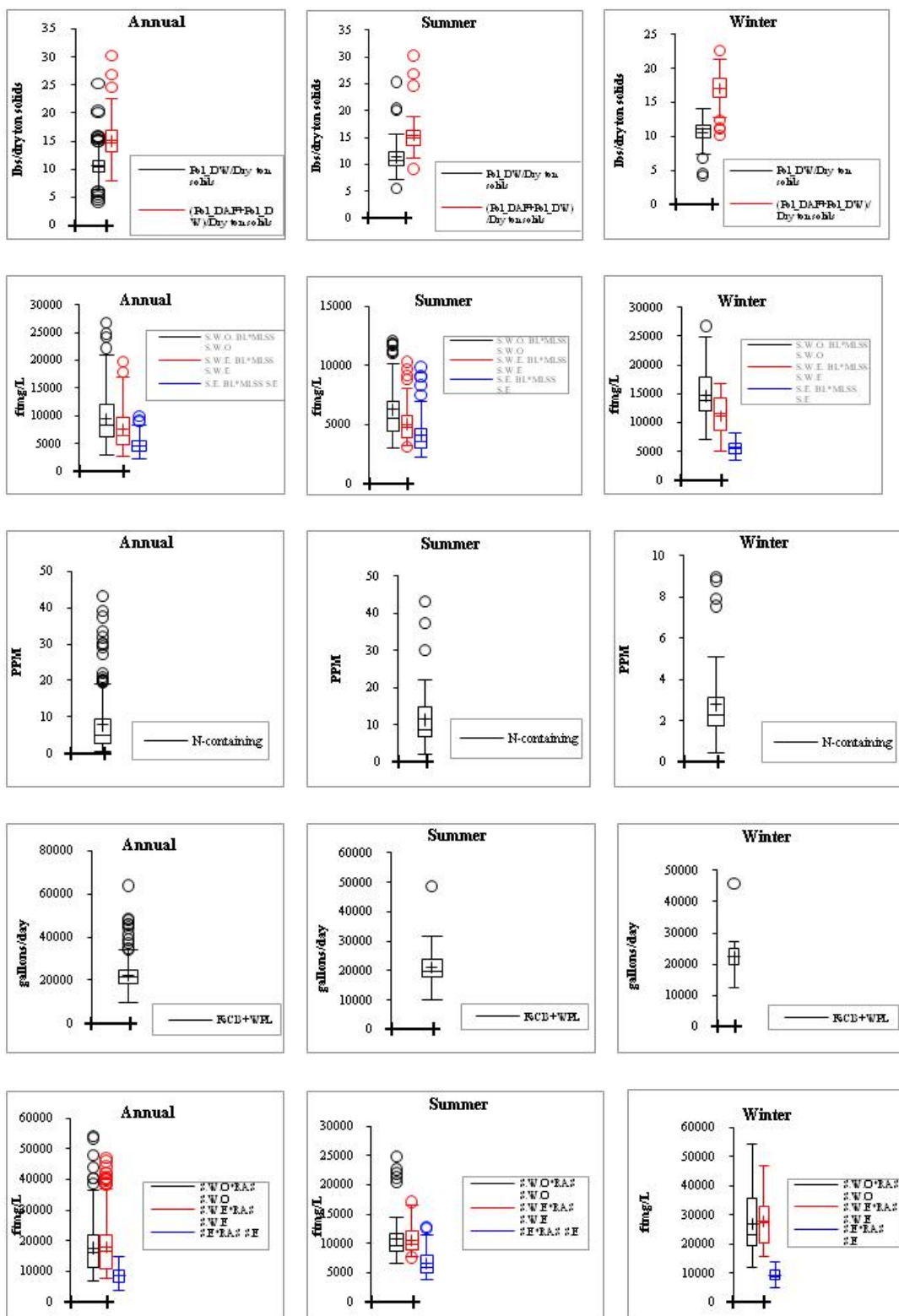
## Exploratory Data Analysis (EDA)

Appendix A-1: Box plot of dependent and independent variables

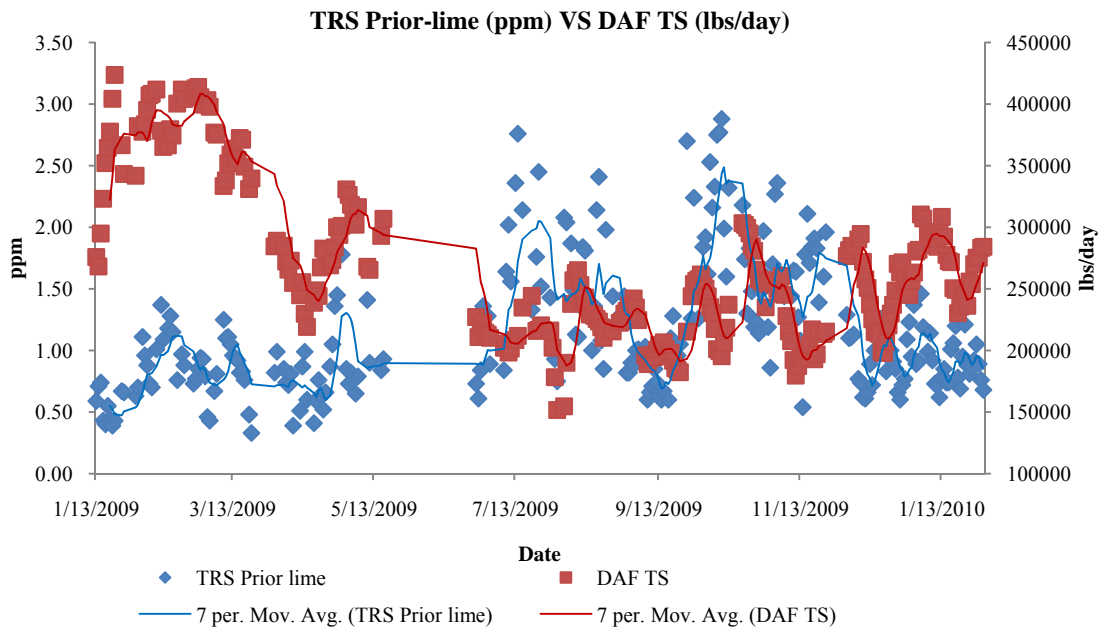
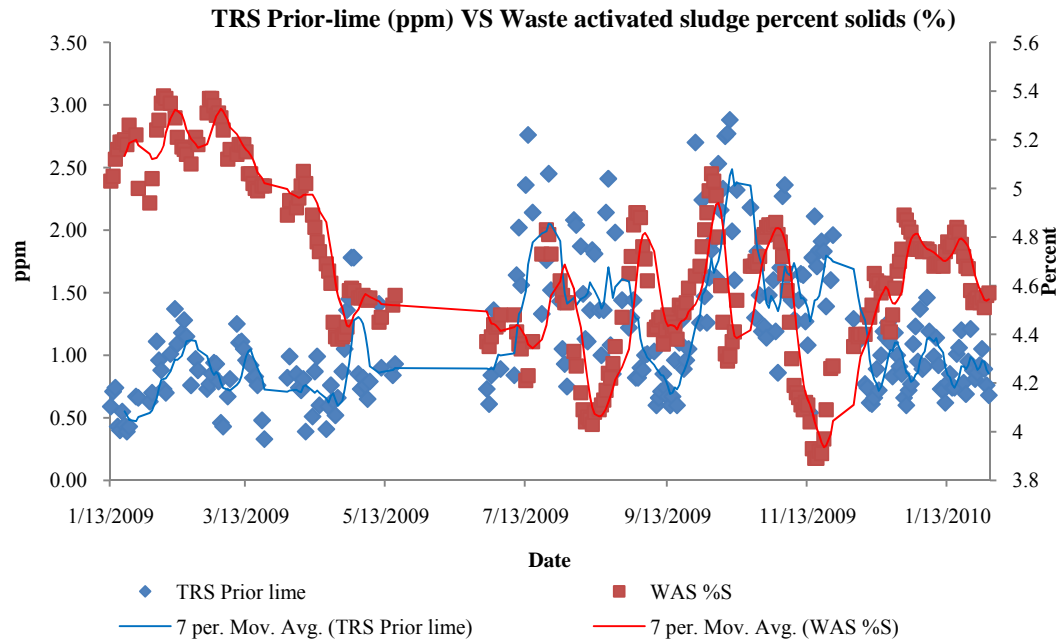


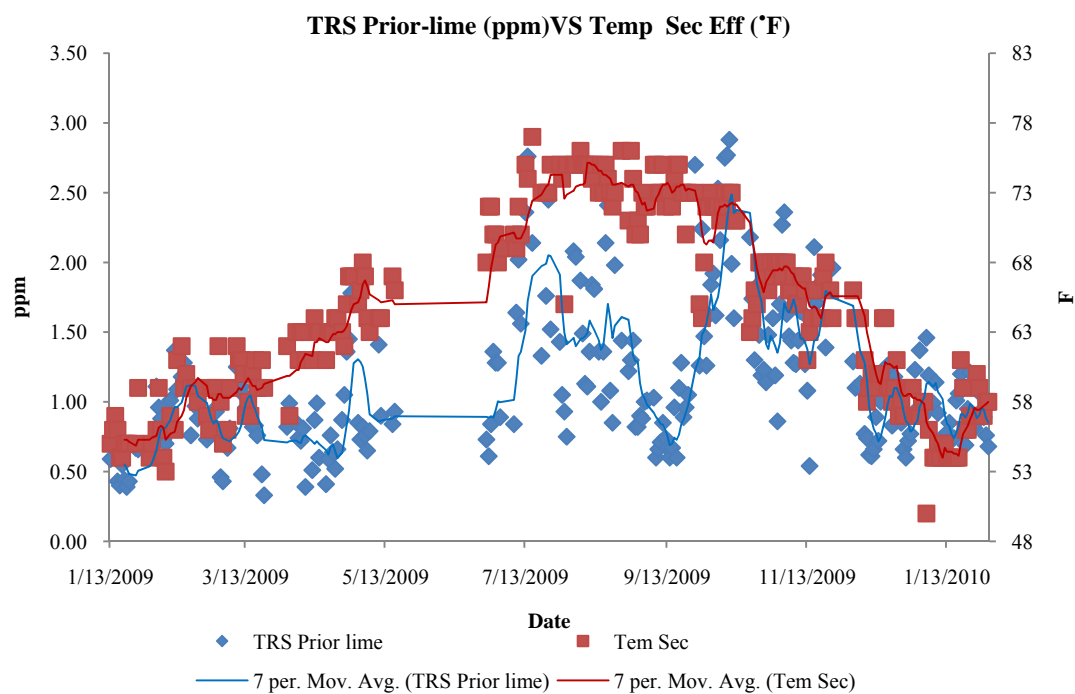
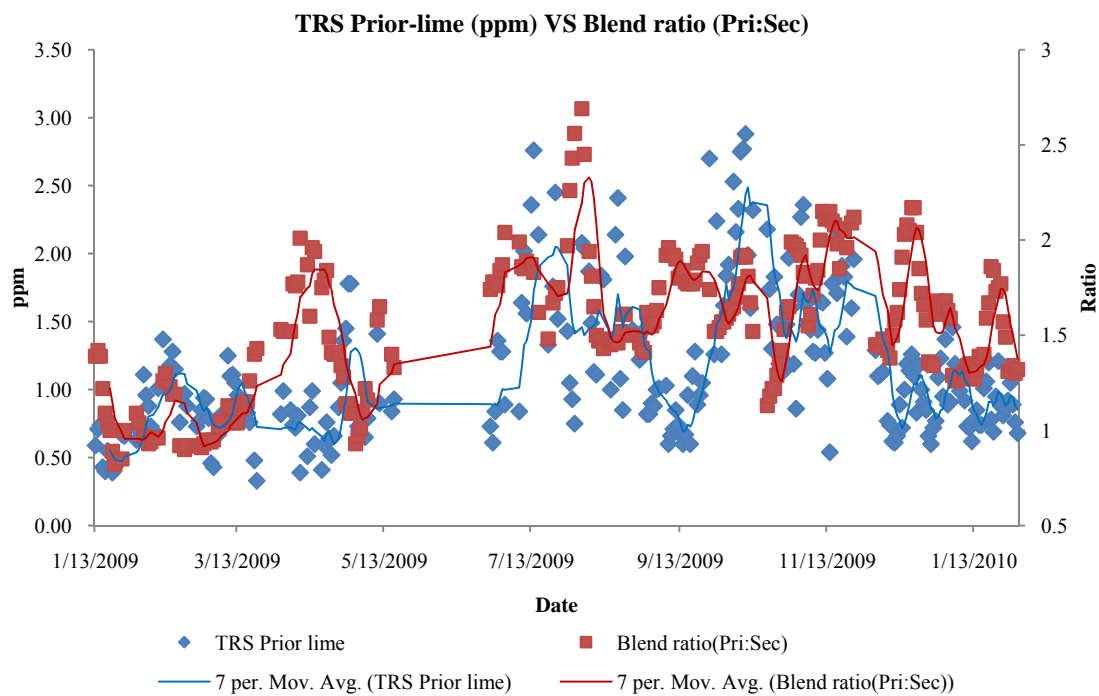


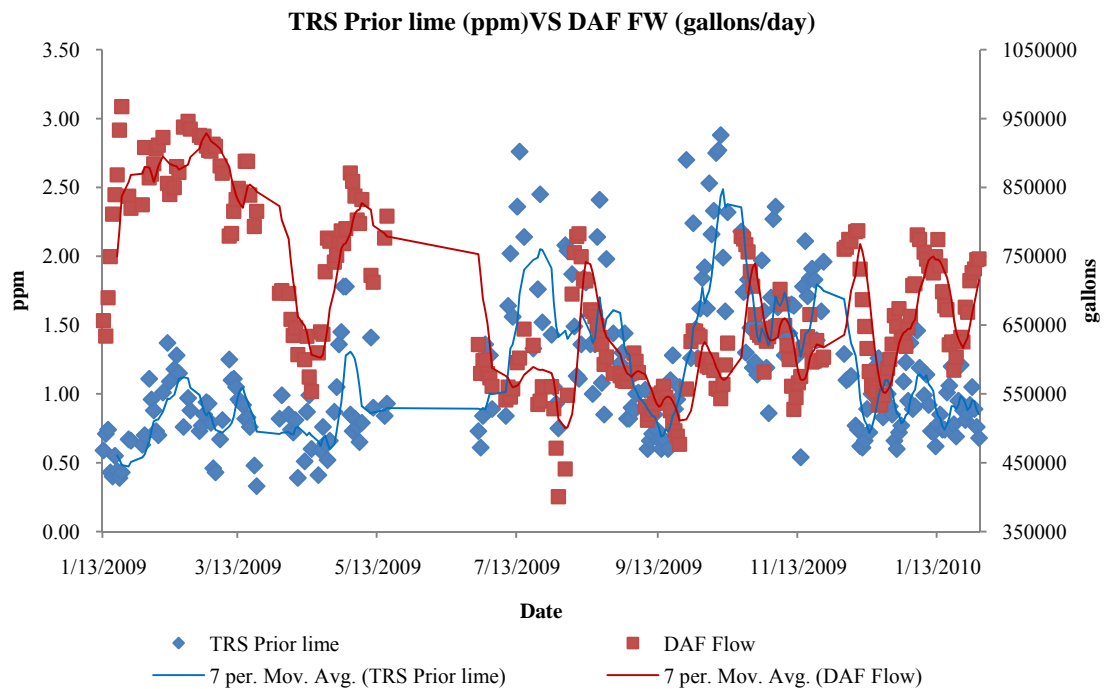
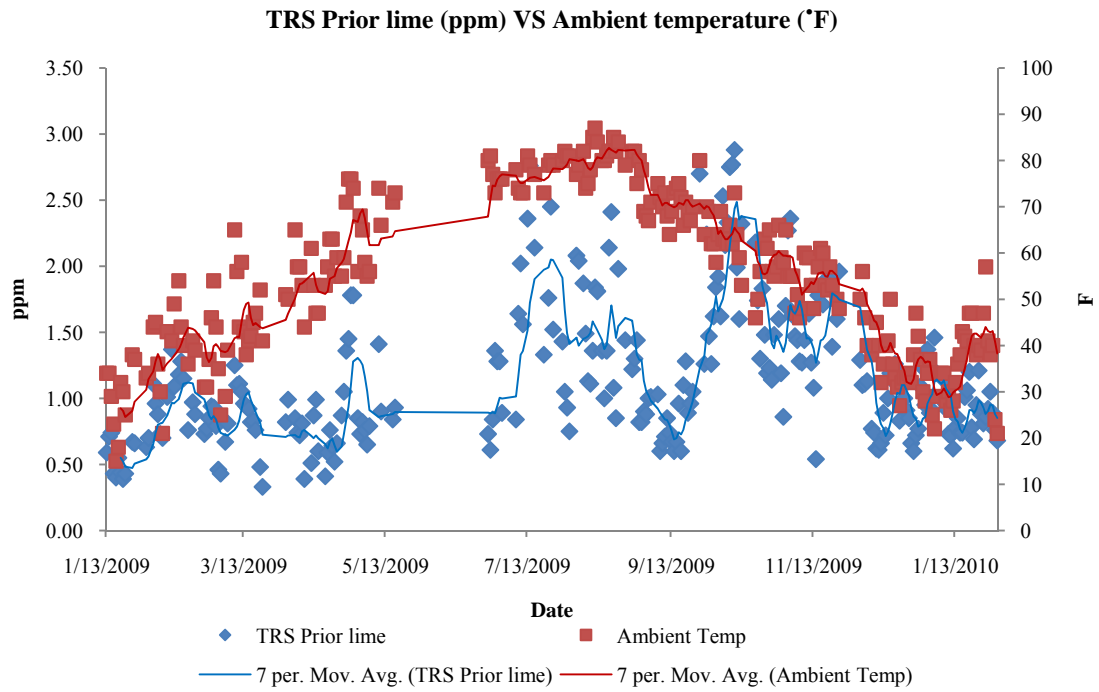




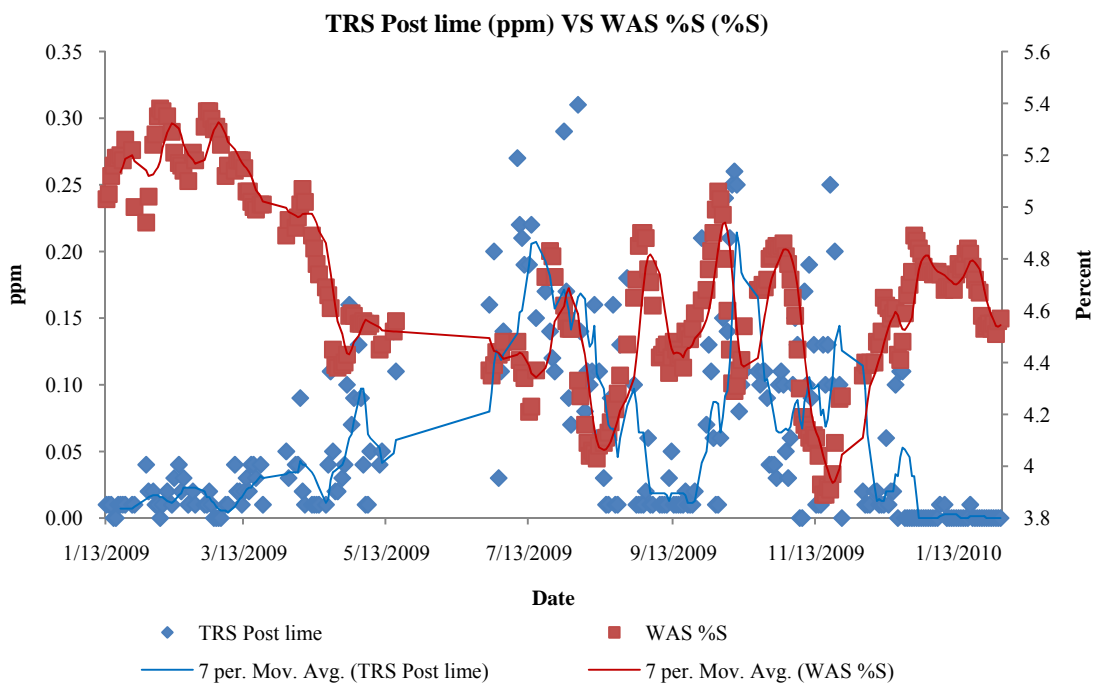
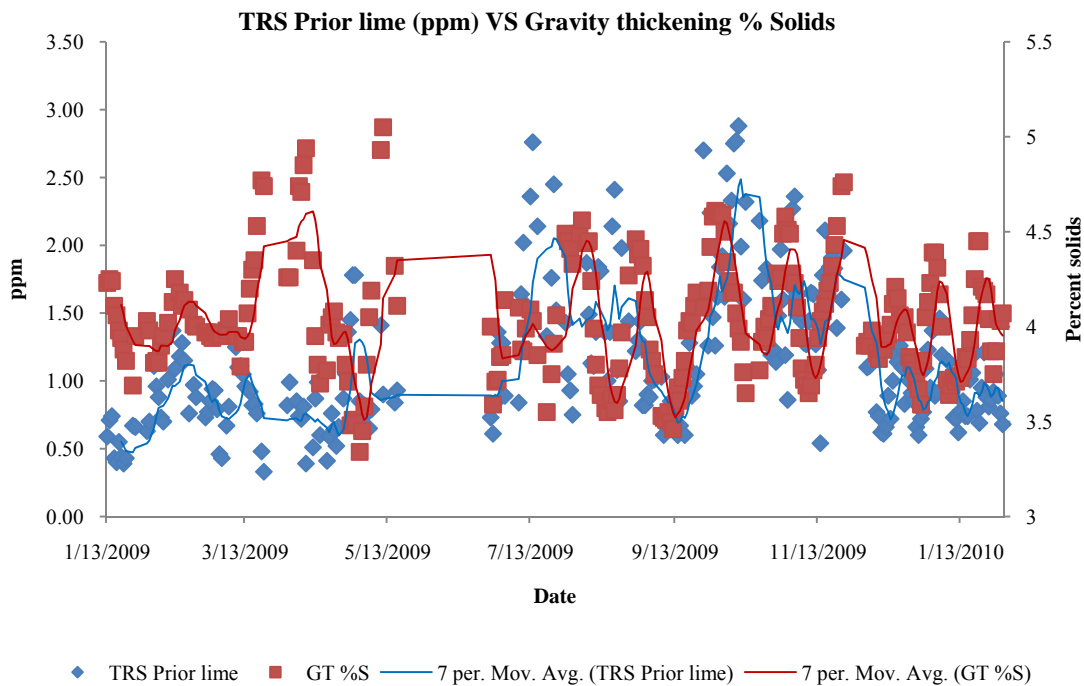
## Appendix A-2: Visualization analysis for the significant correlations

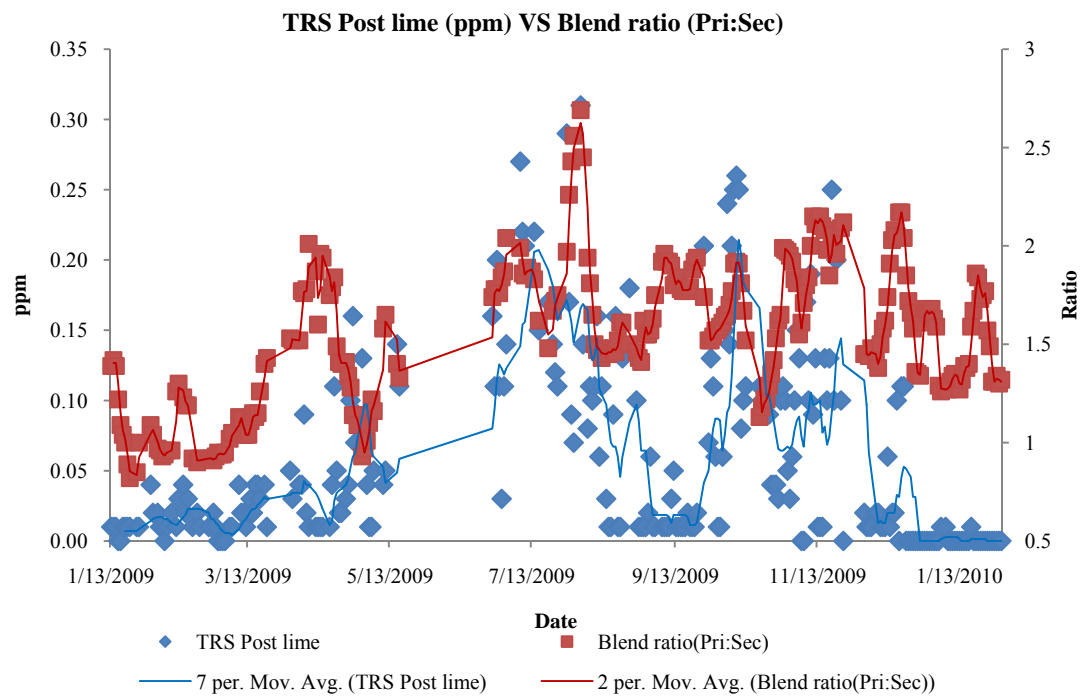
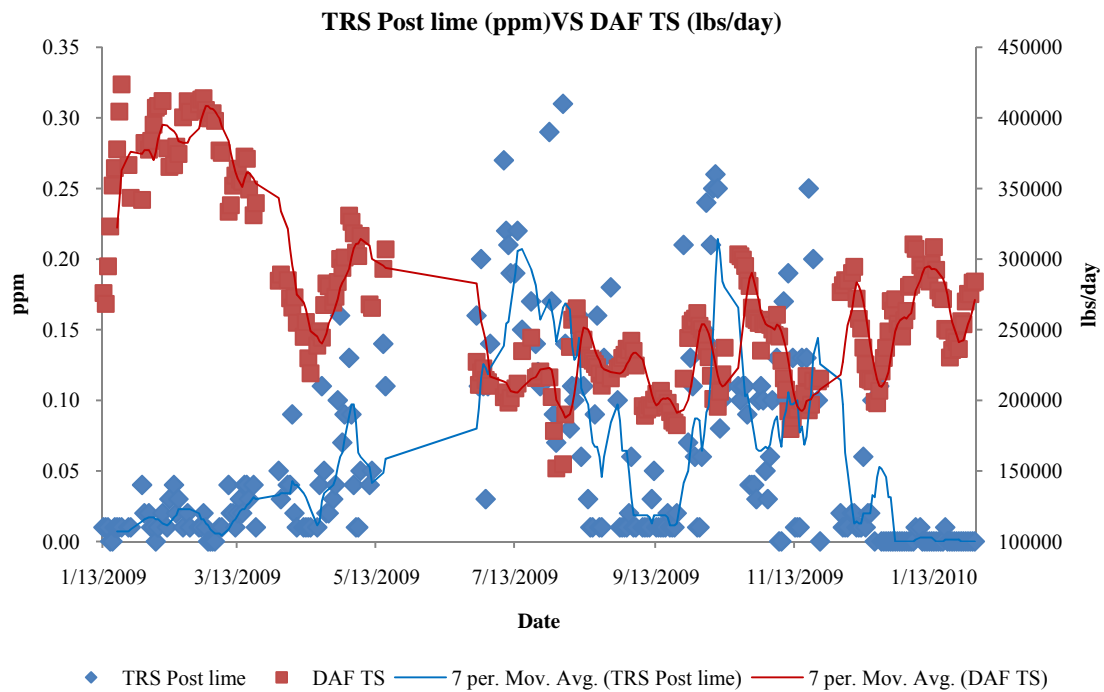


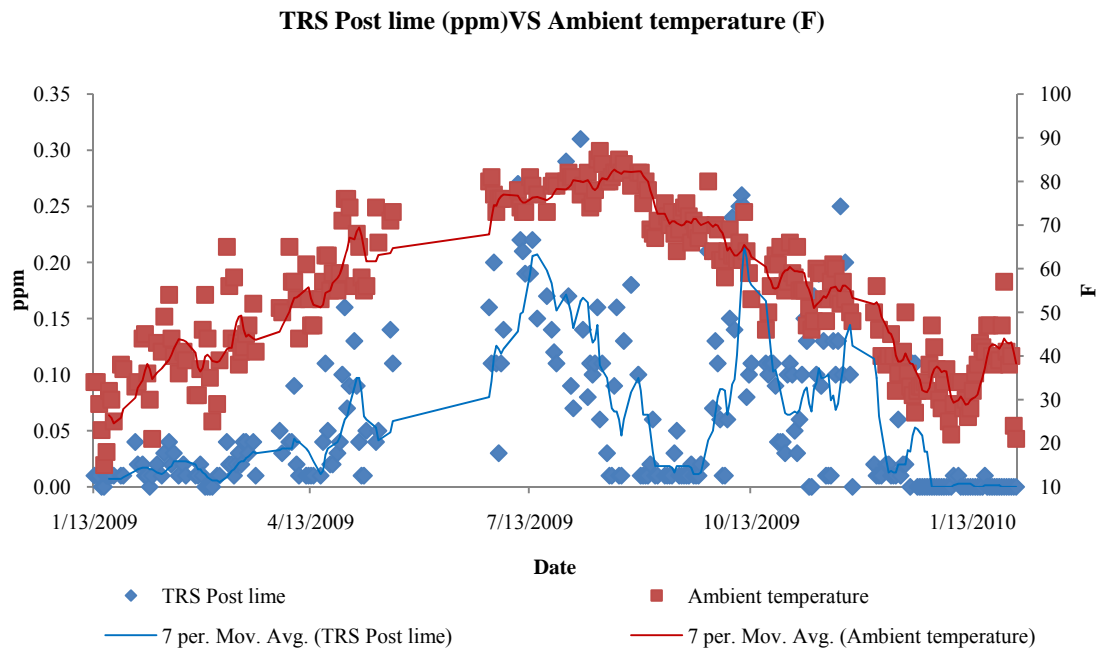
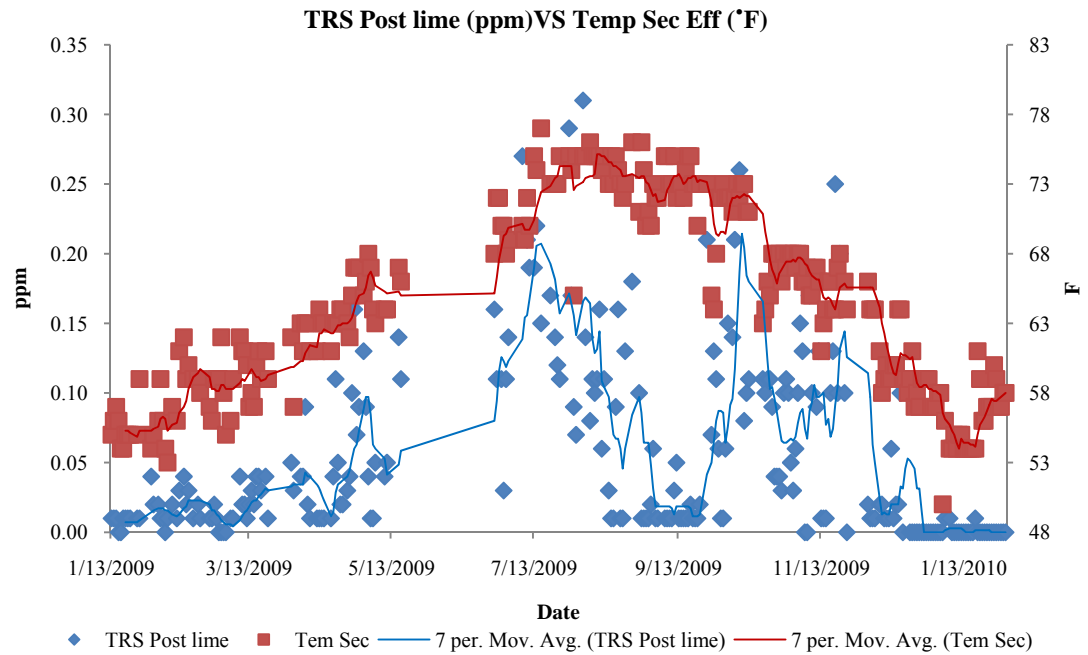


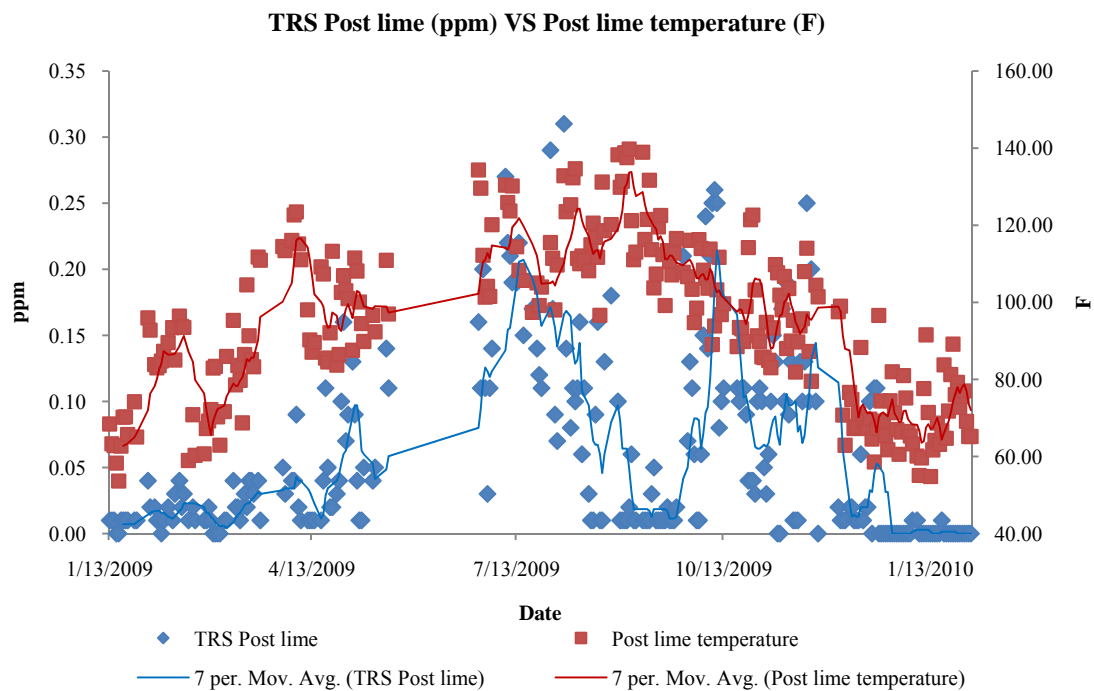
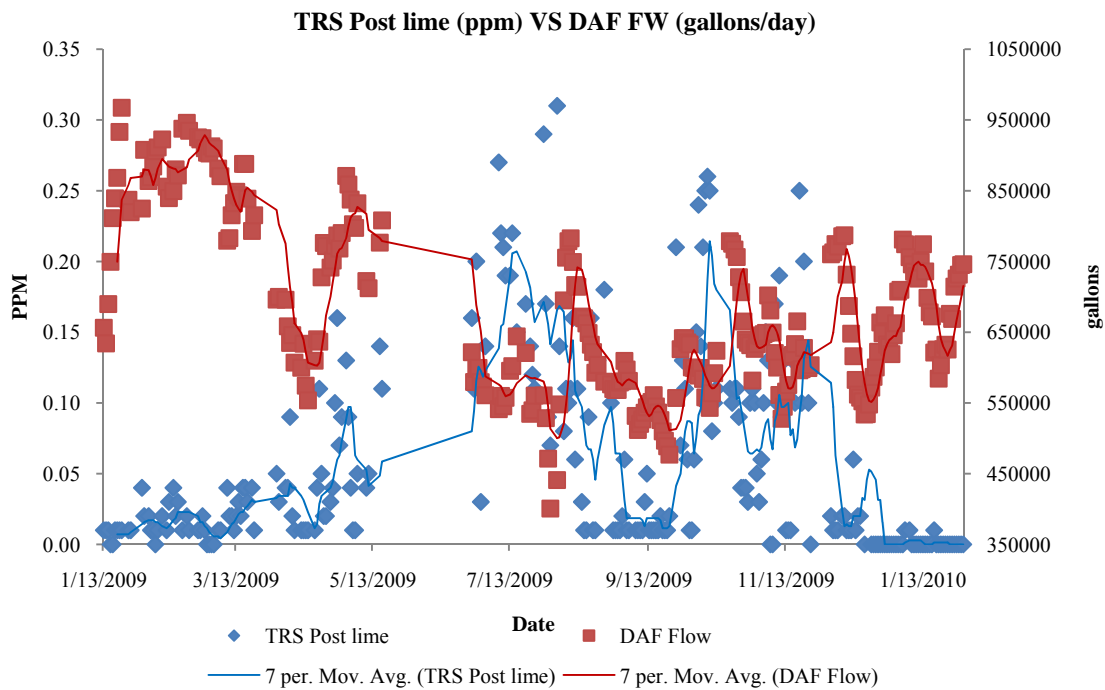


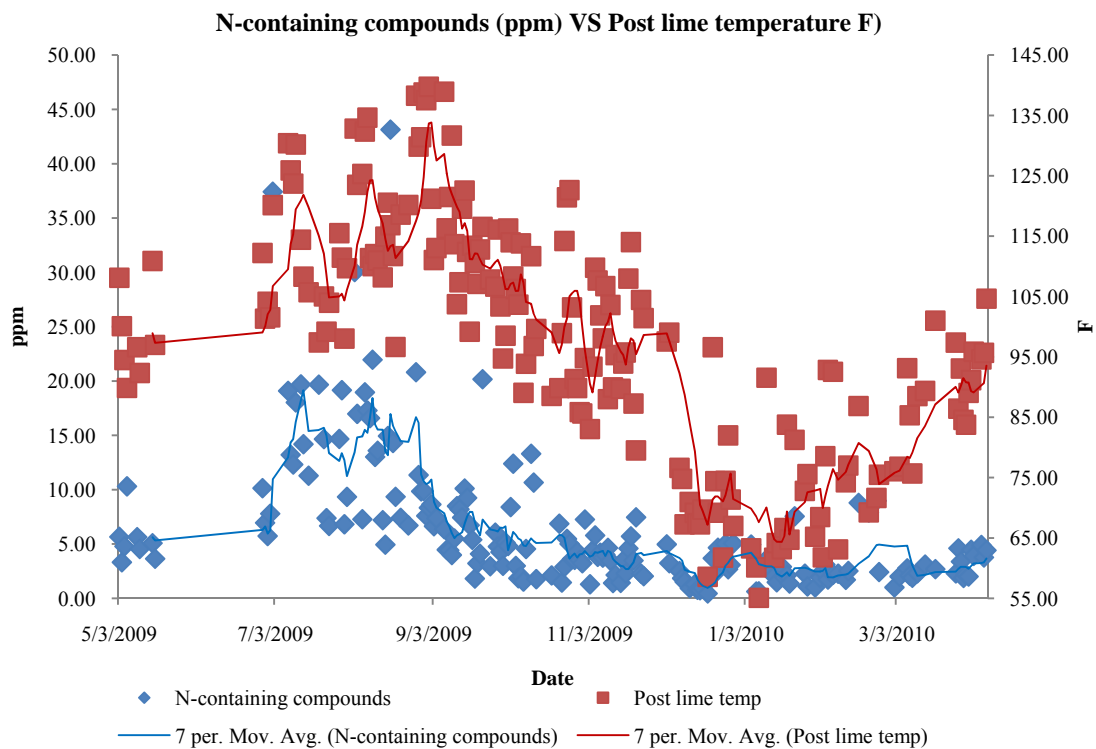
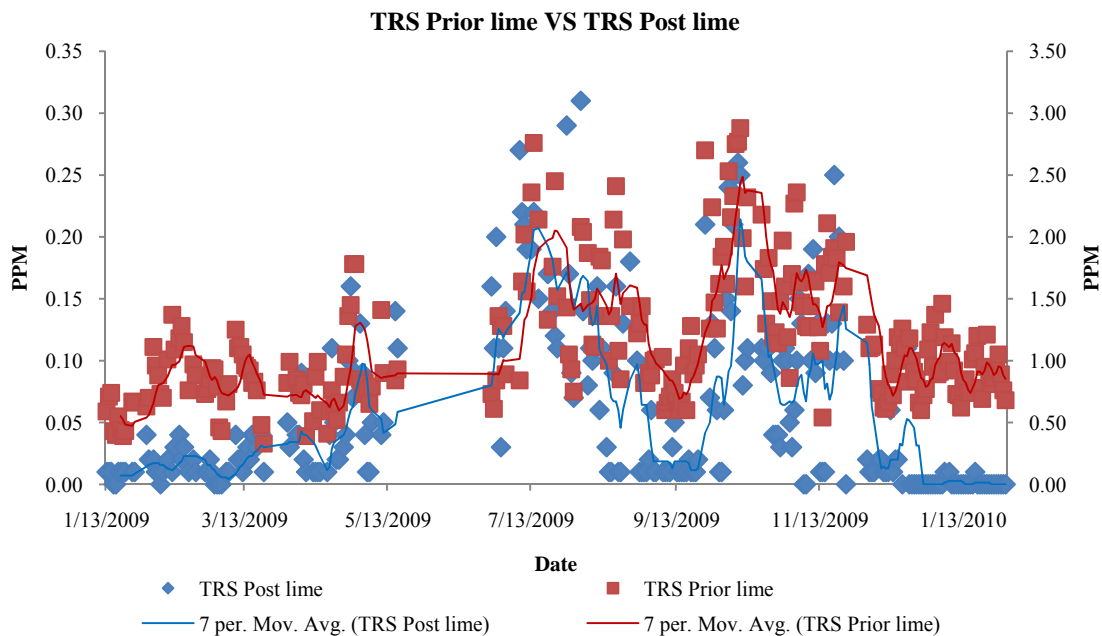


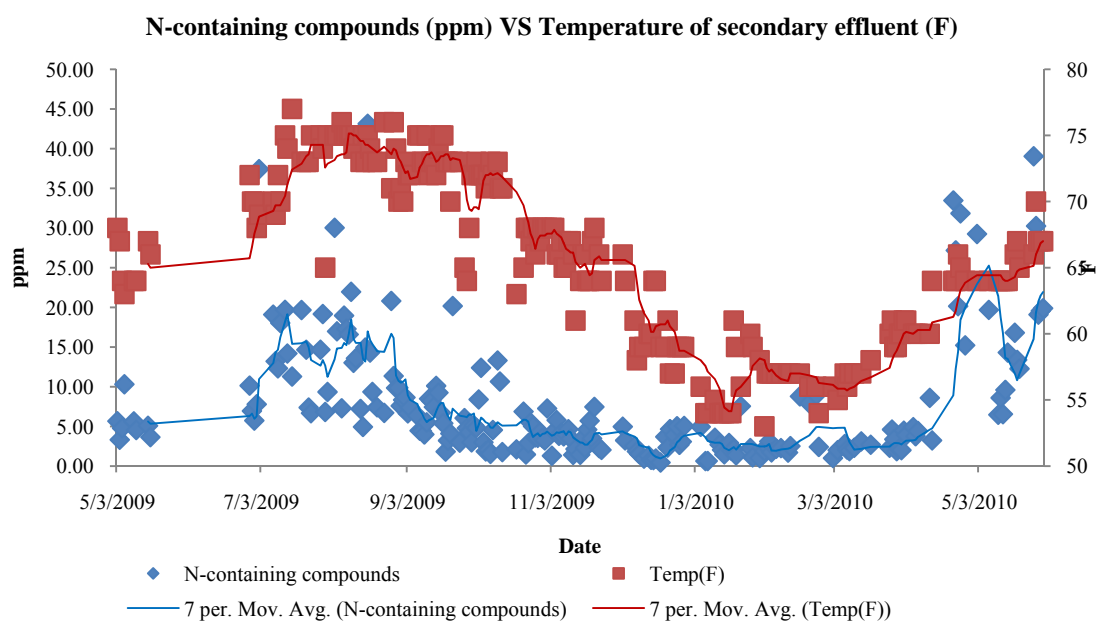
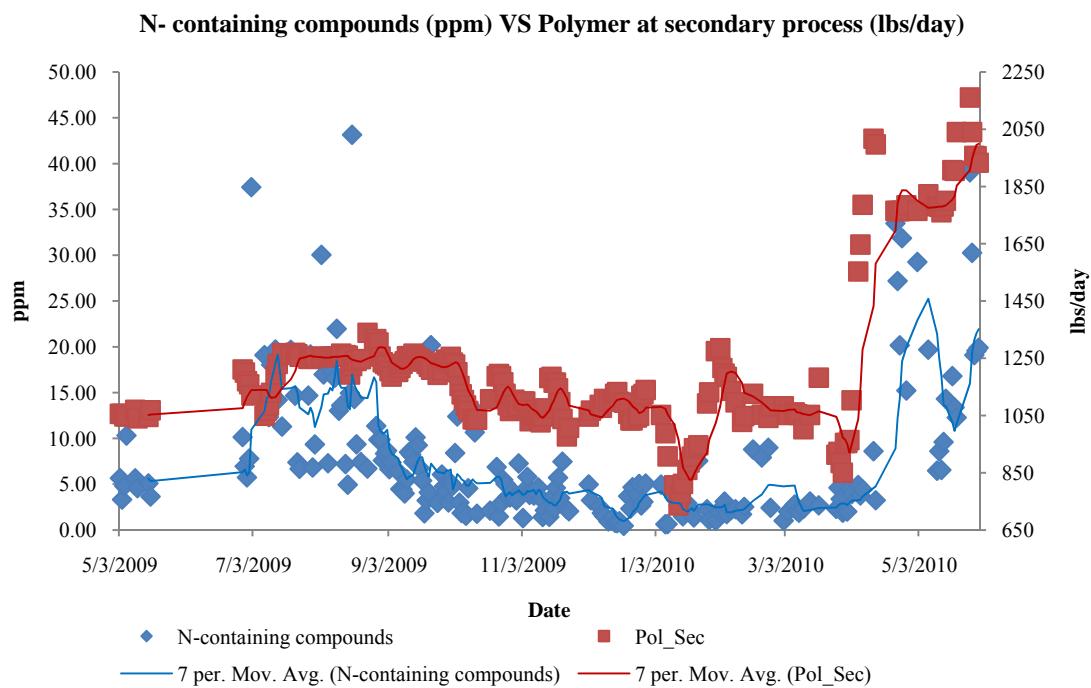


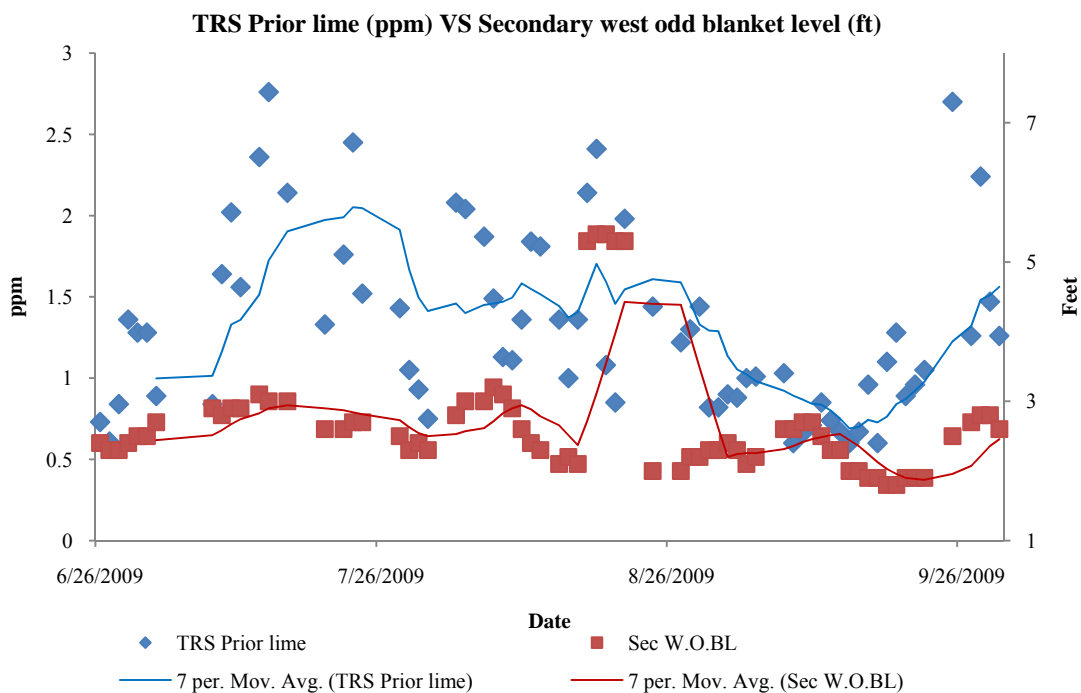
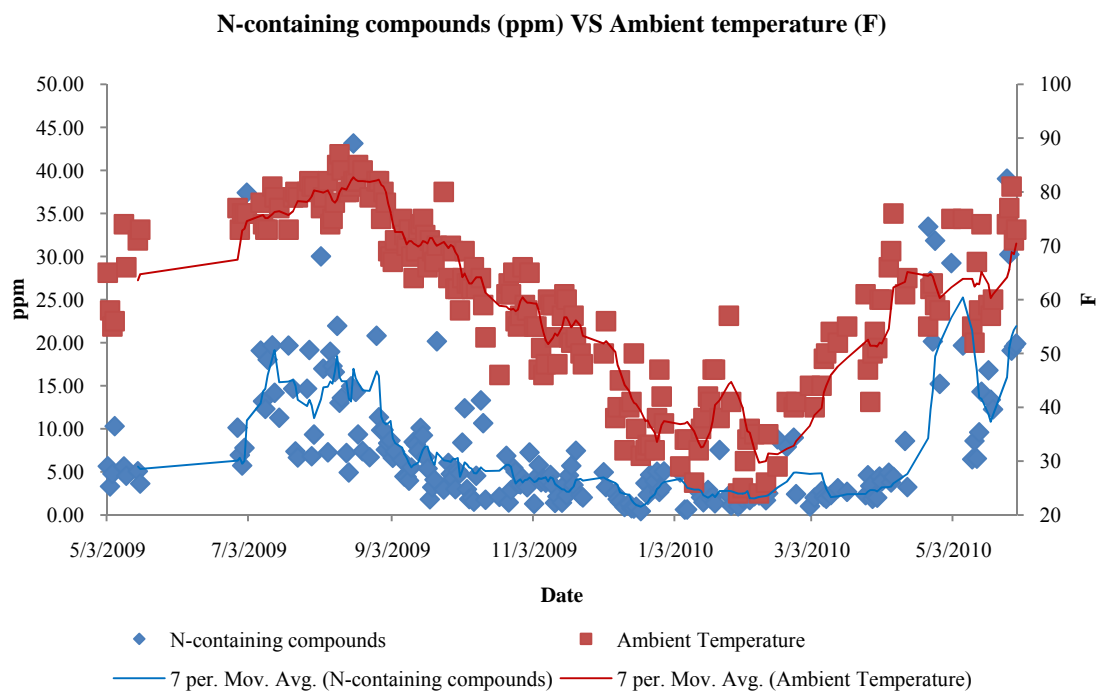


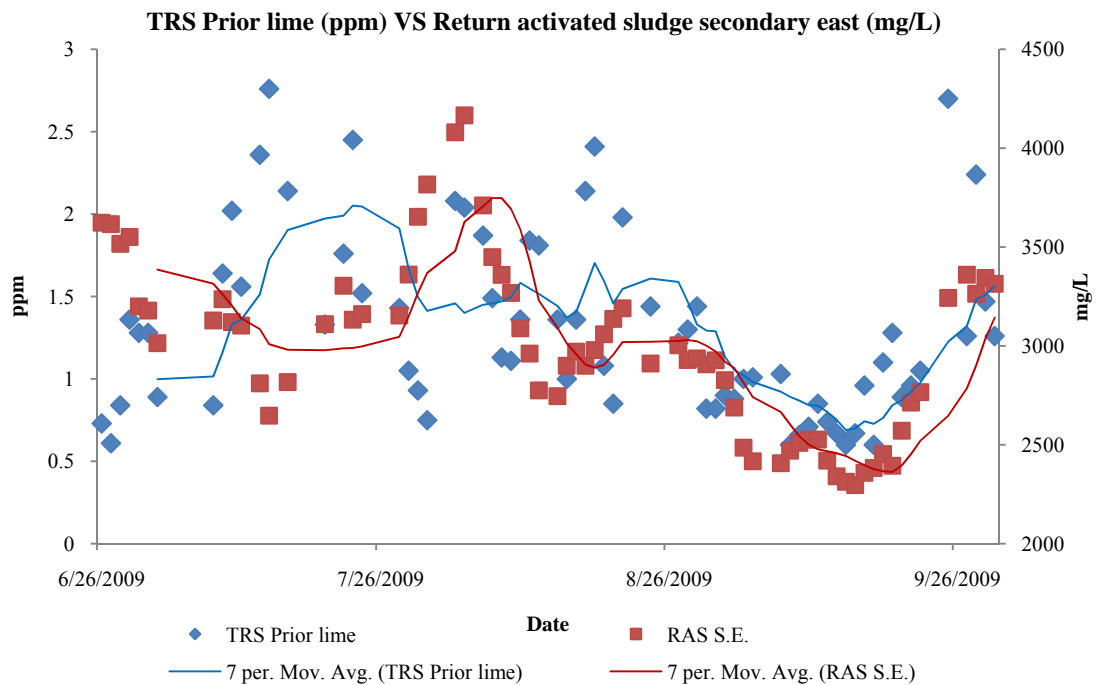
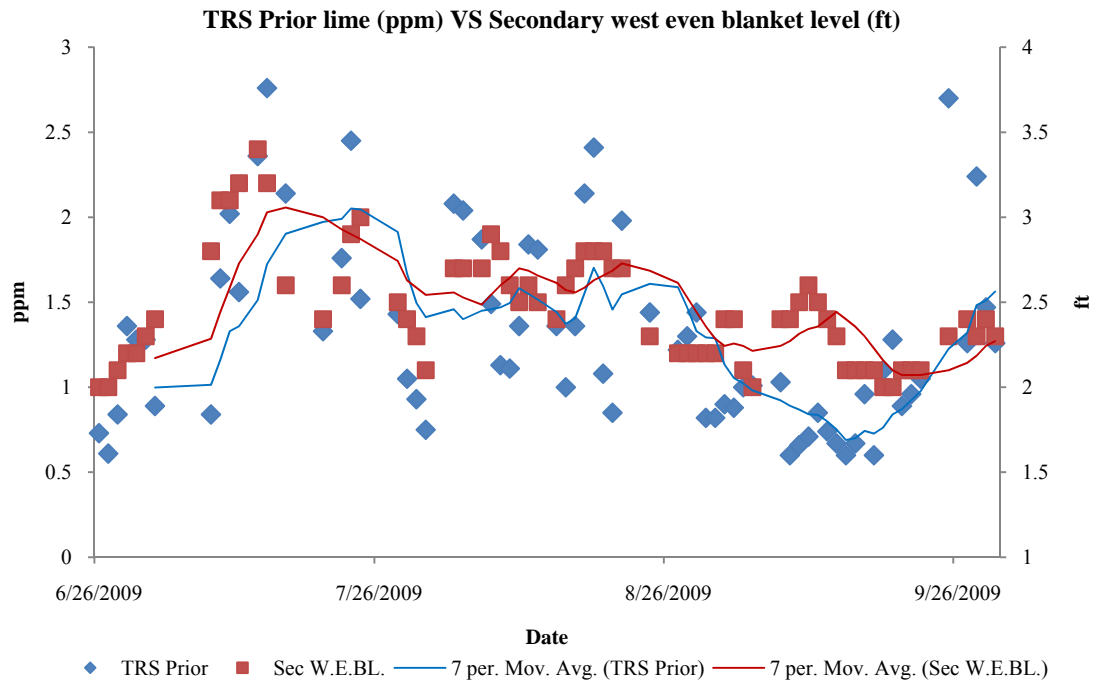




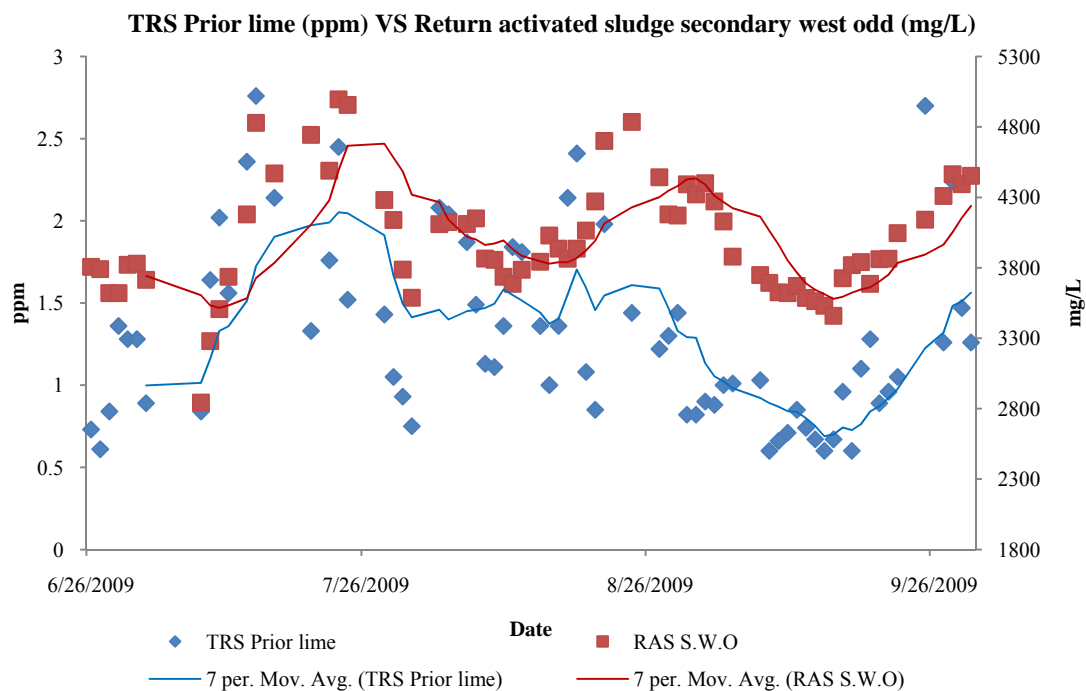
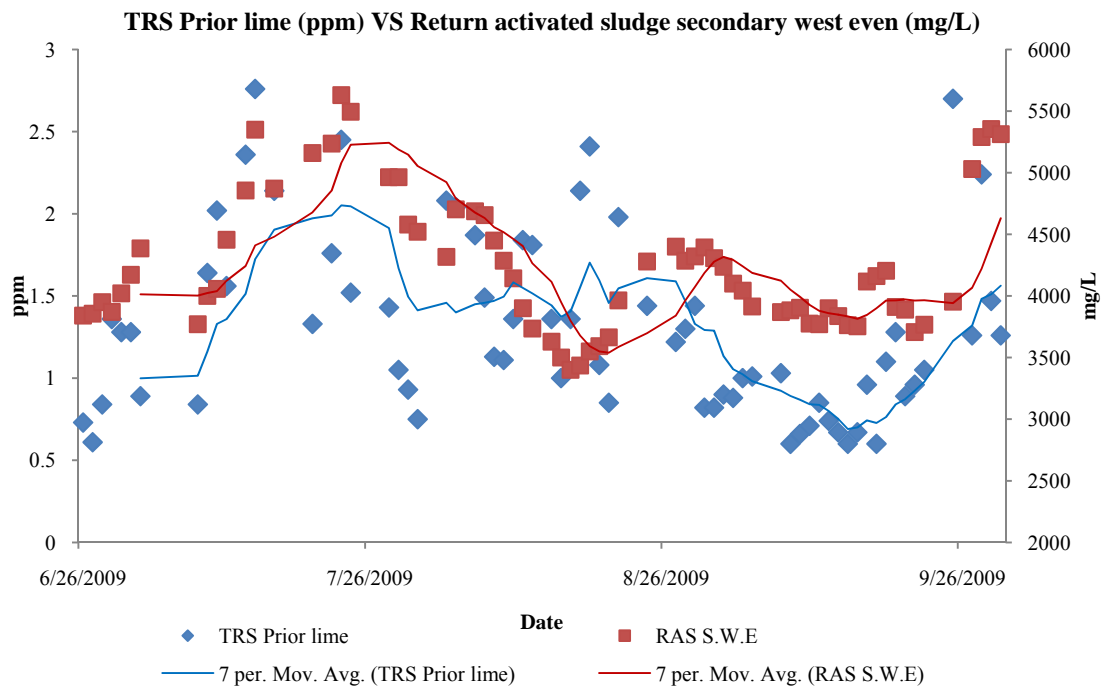


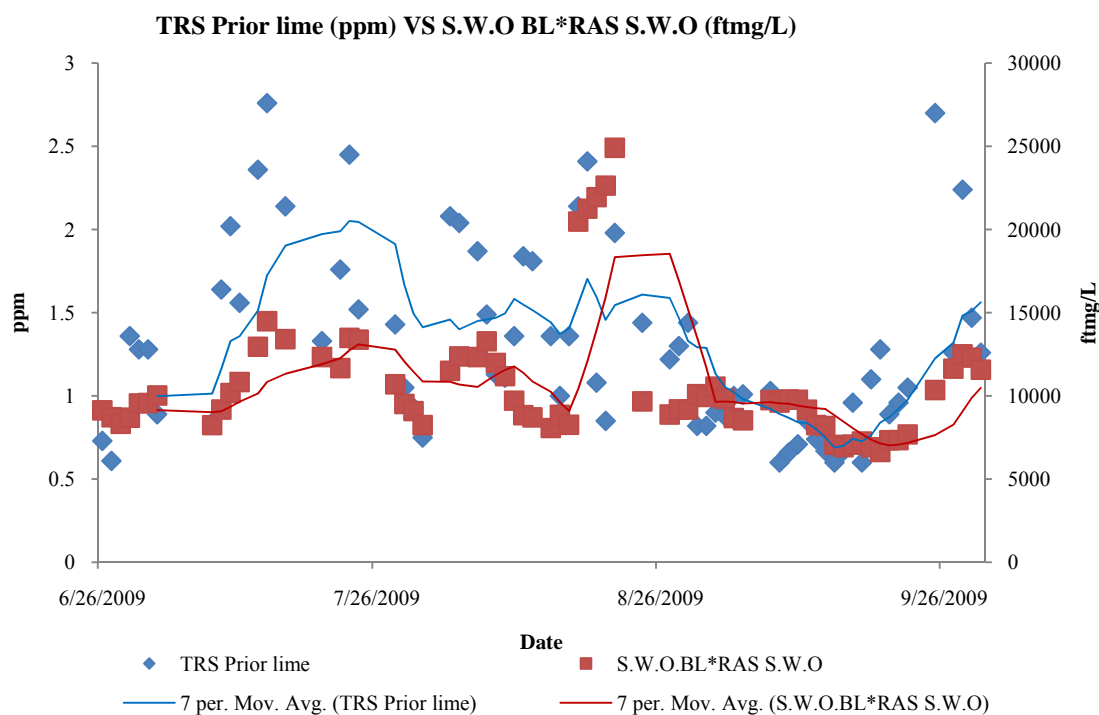
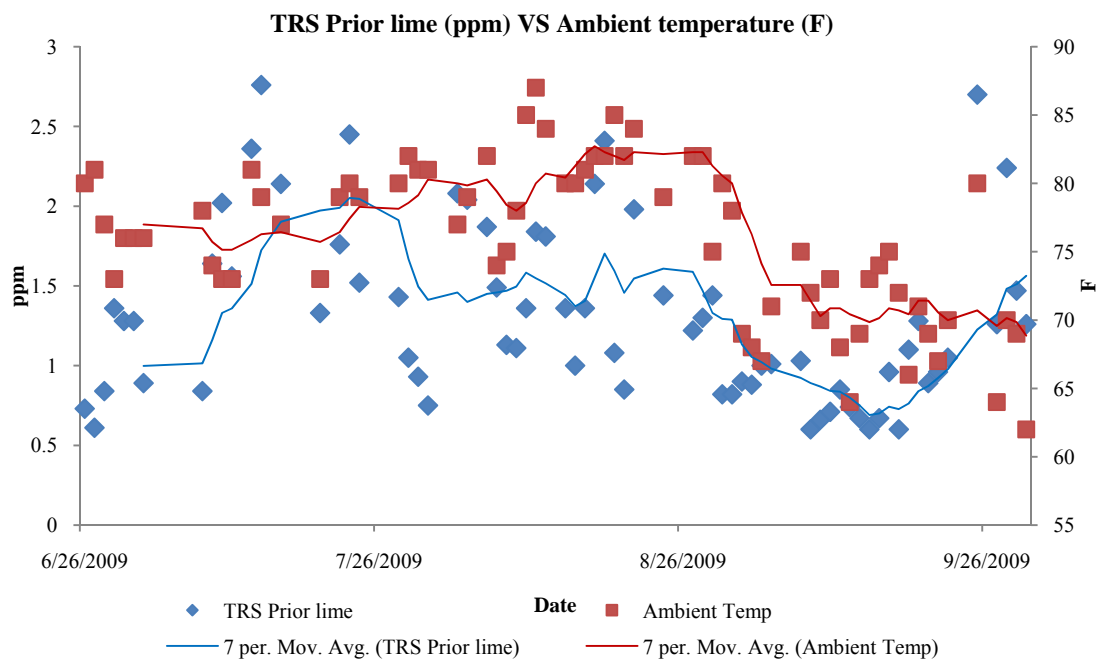


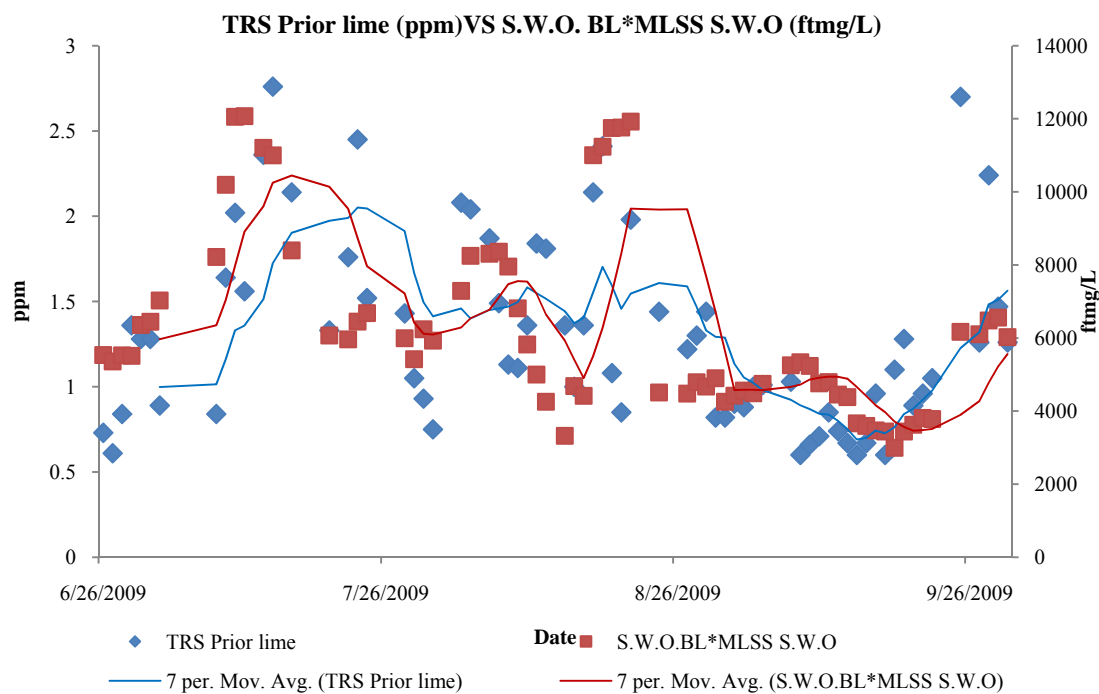
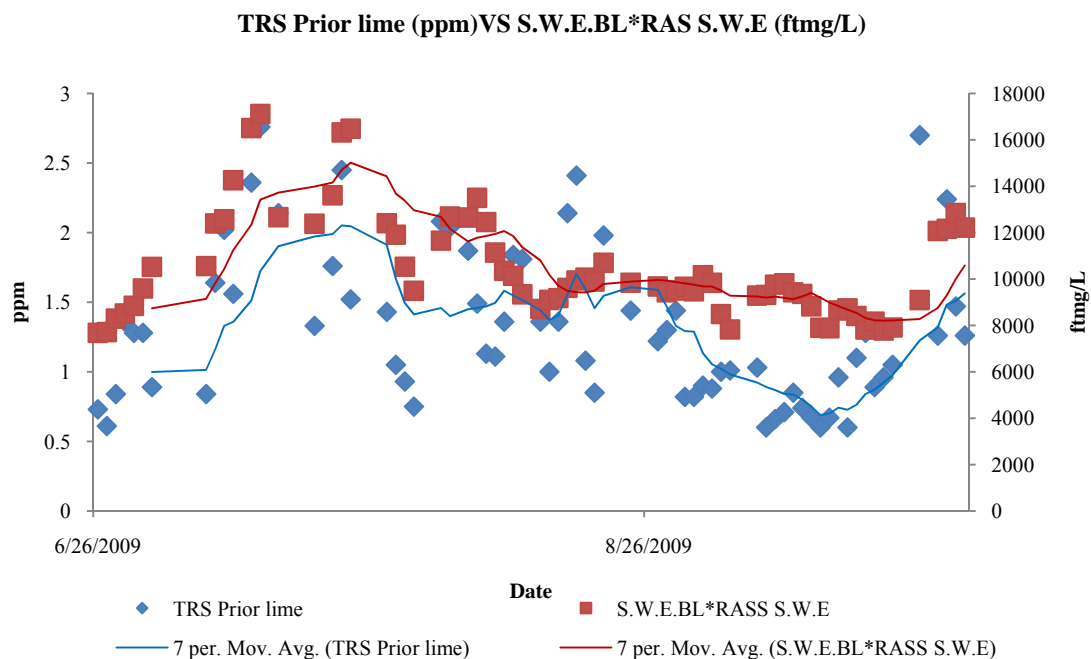


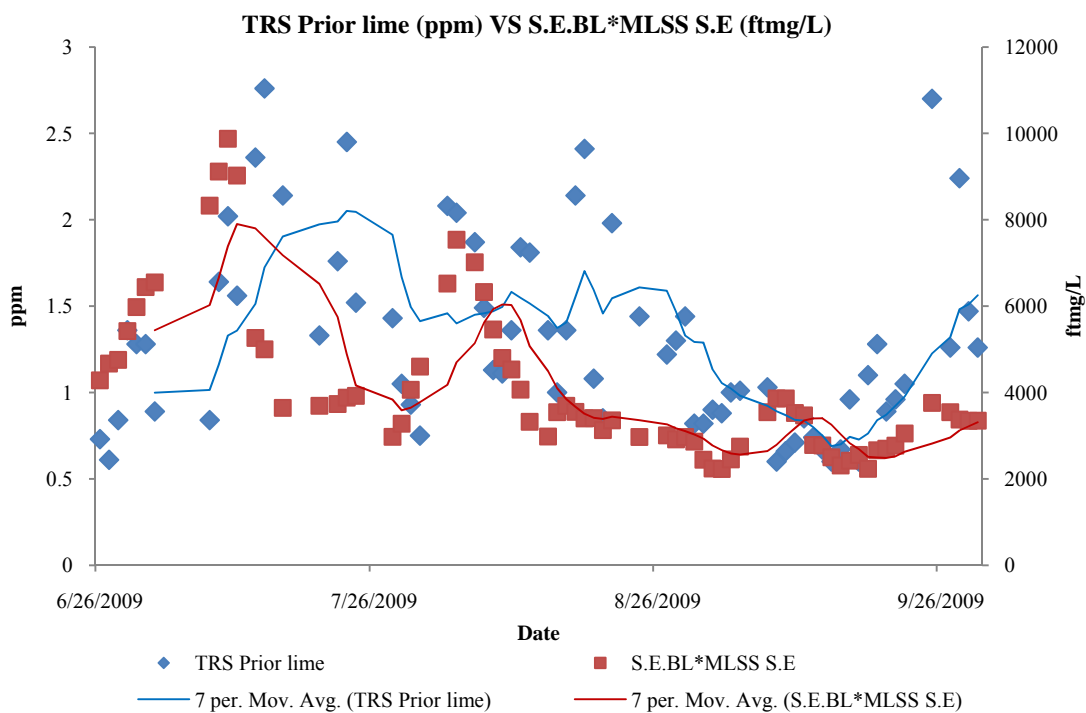
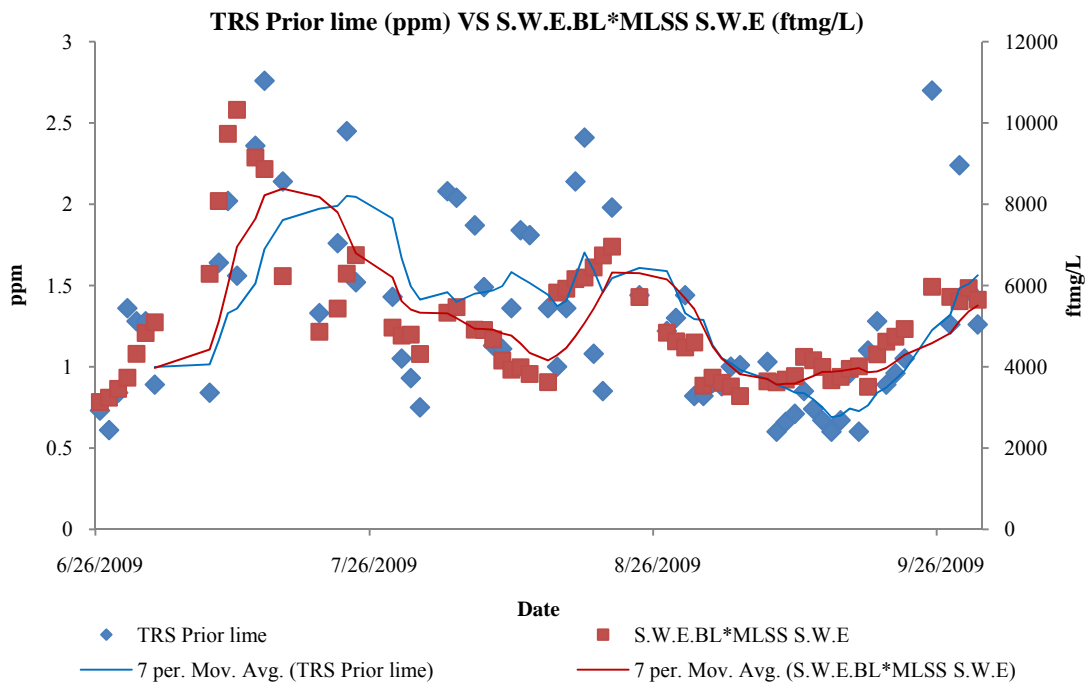


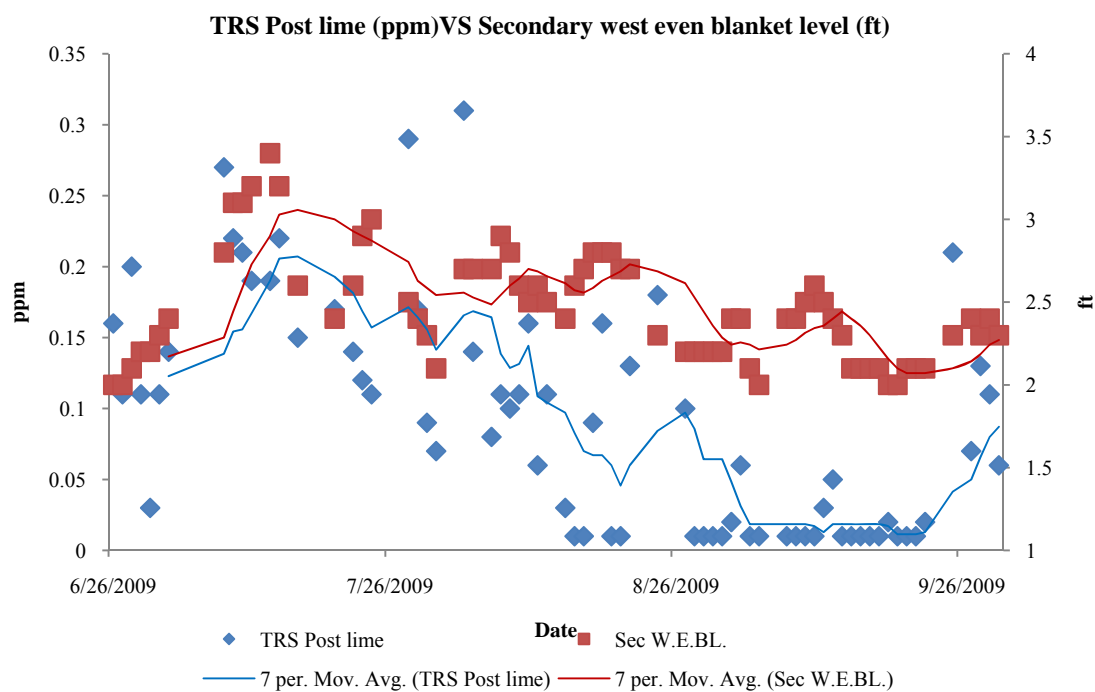
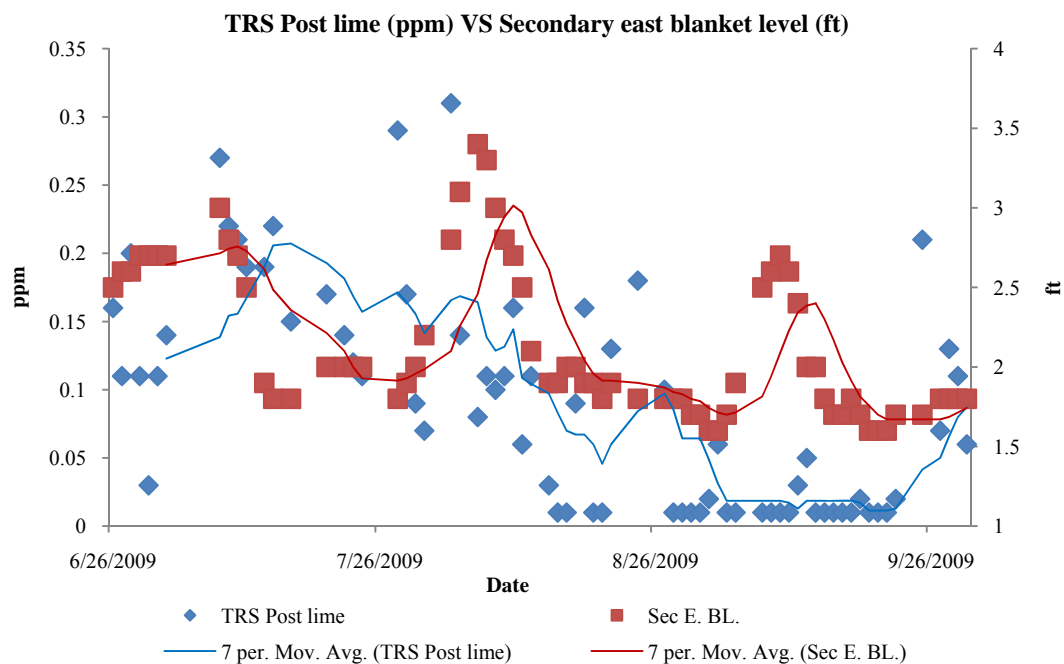


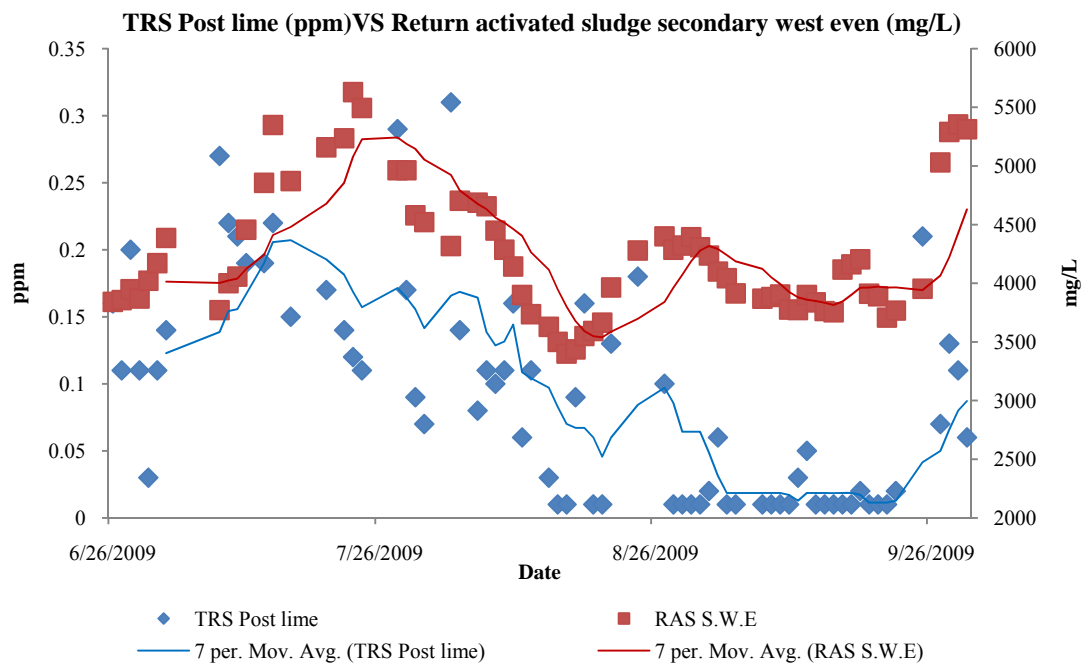
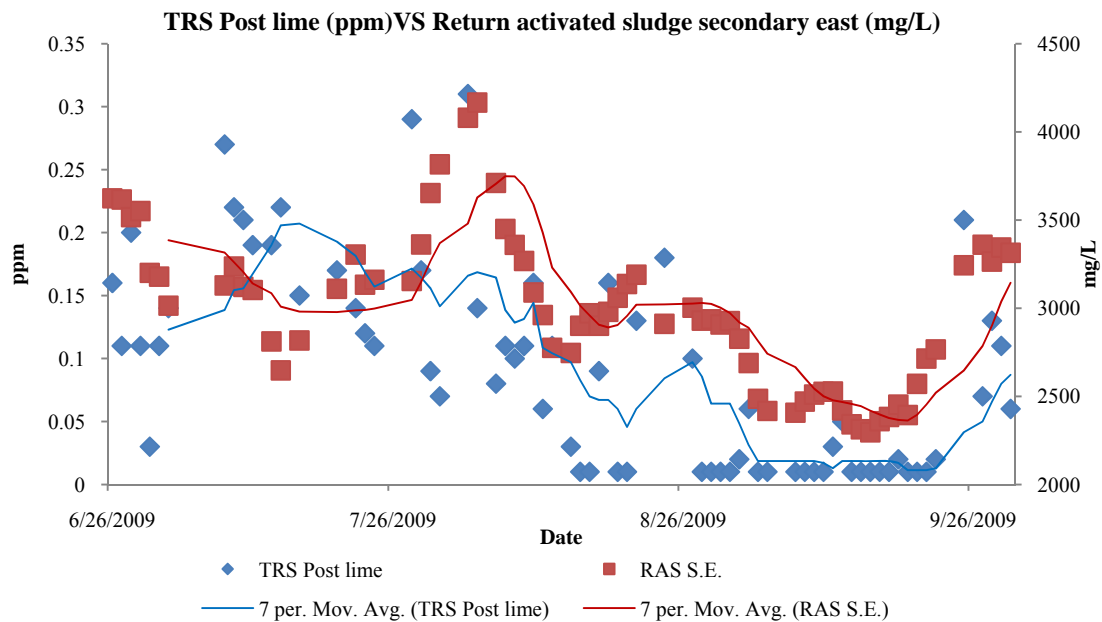


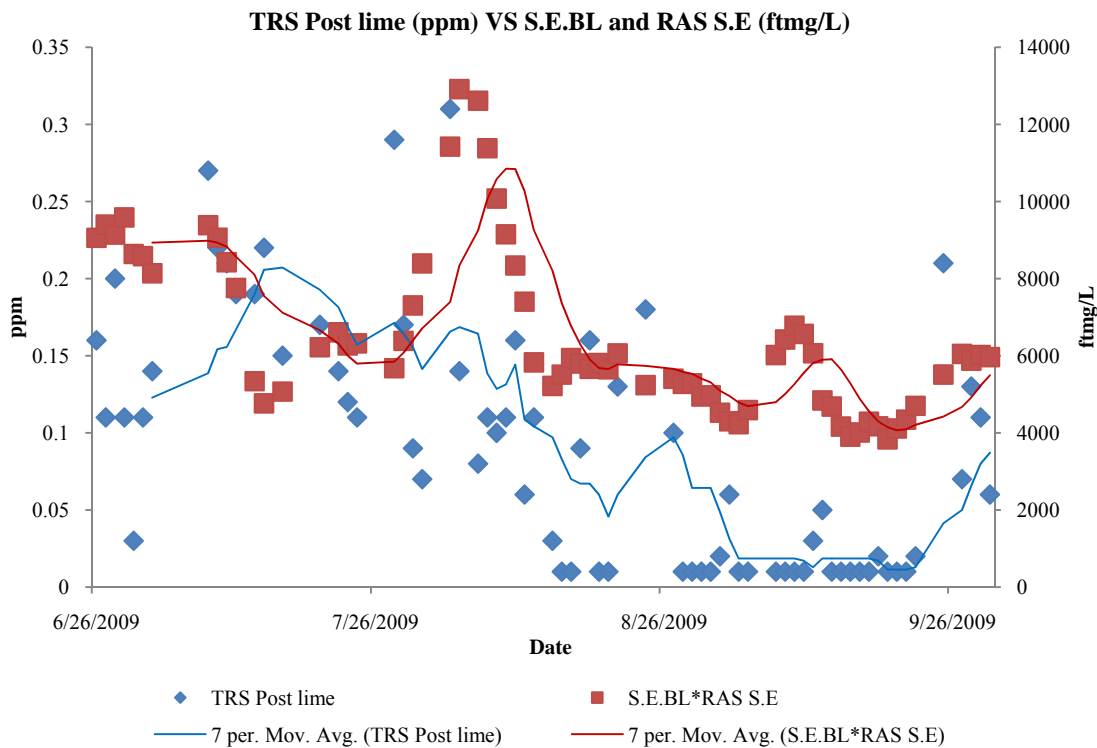
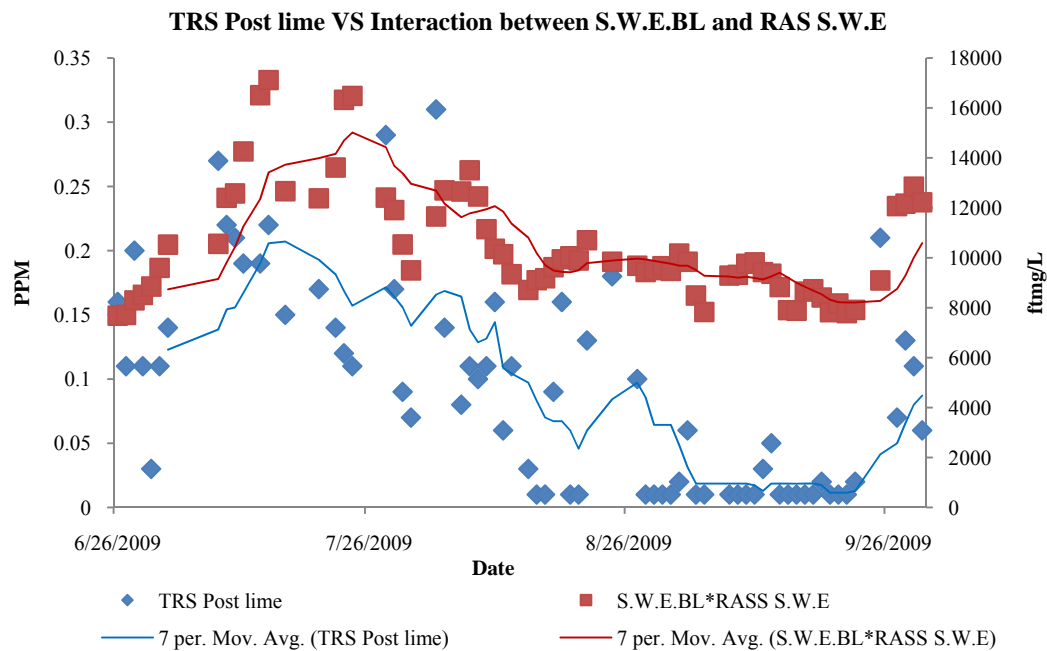


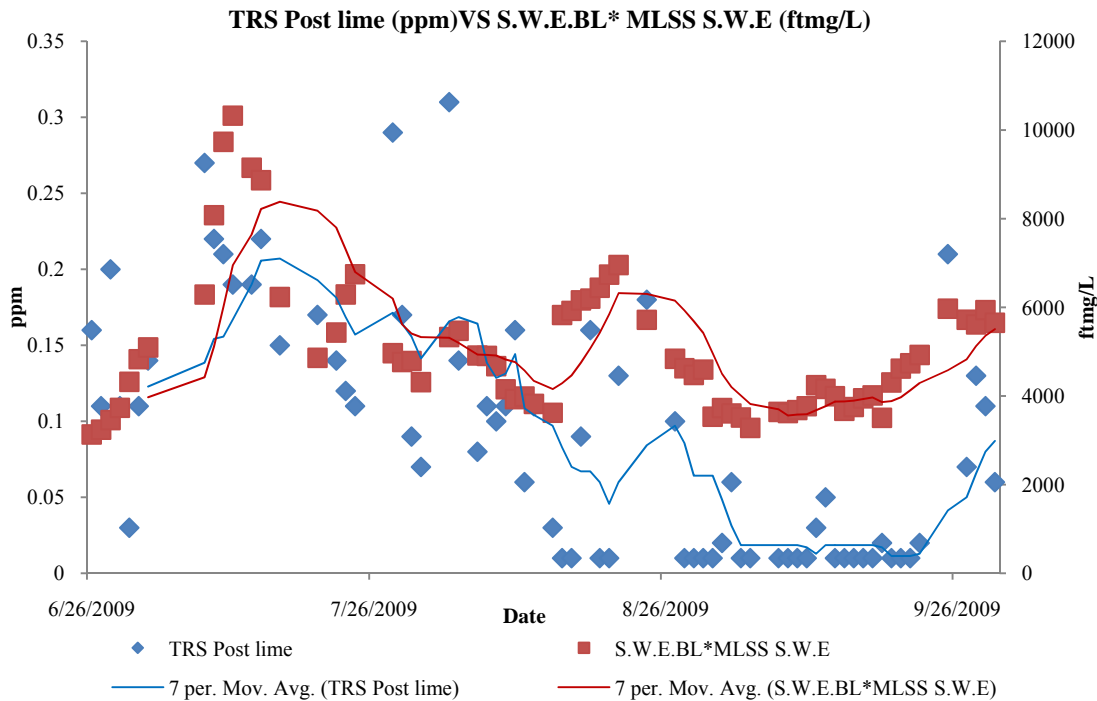
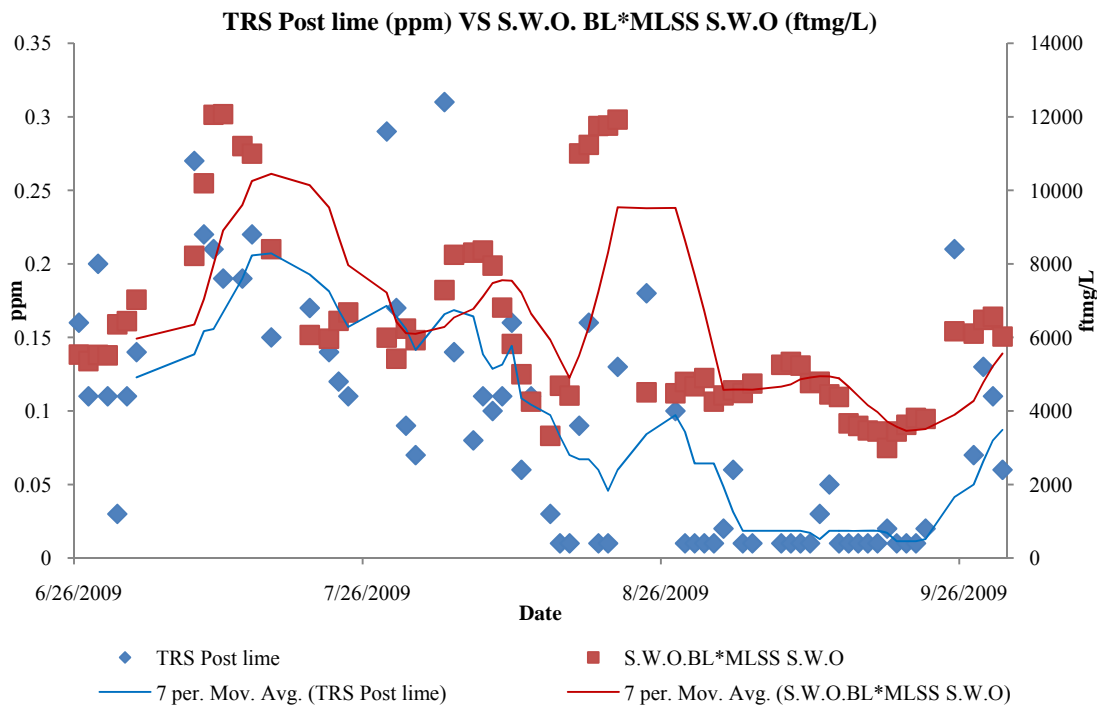




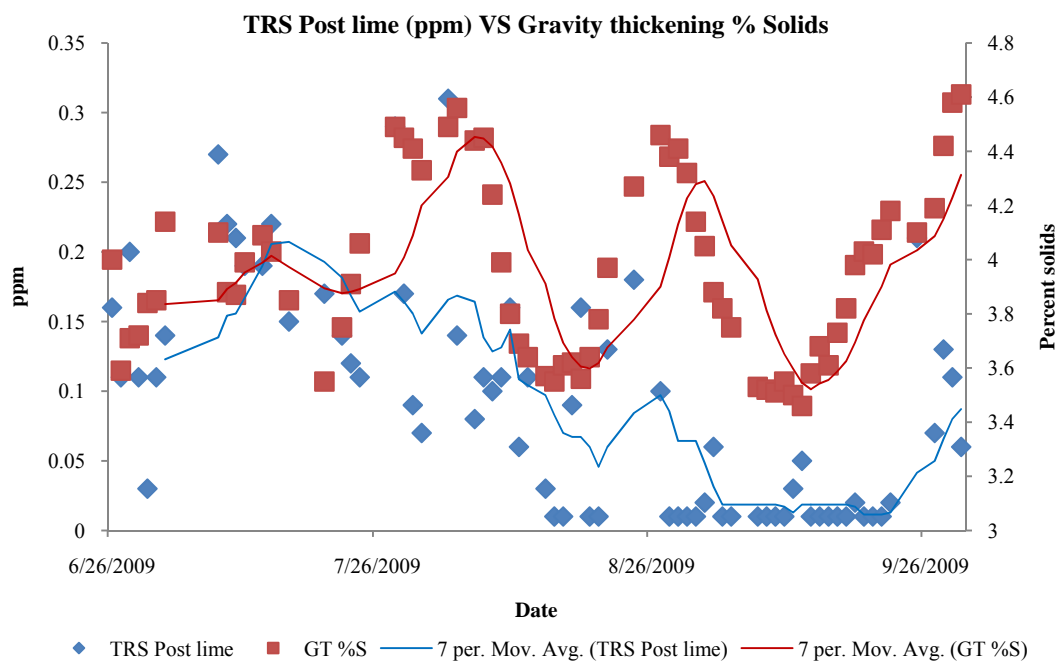
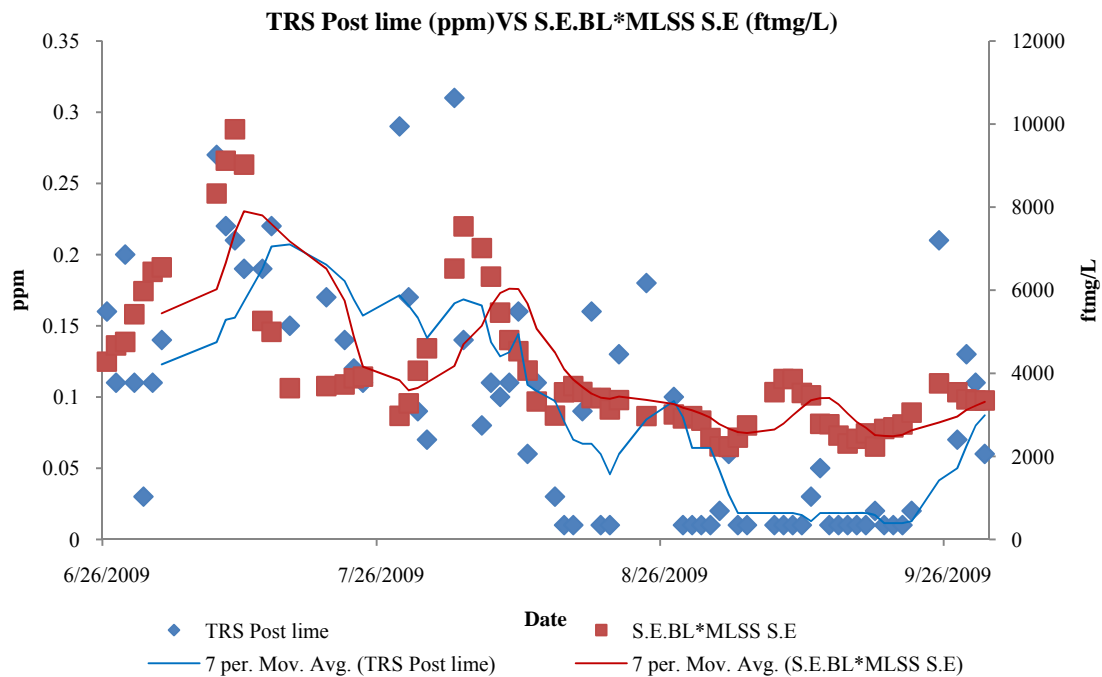


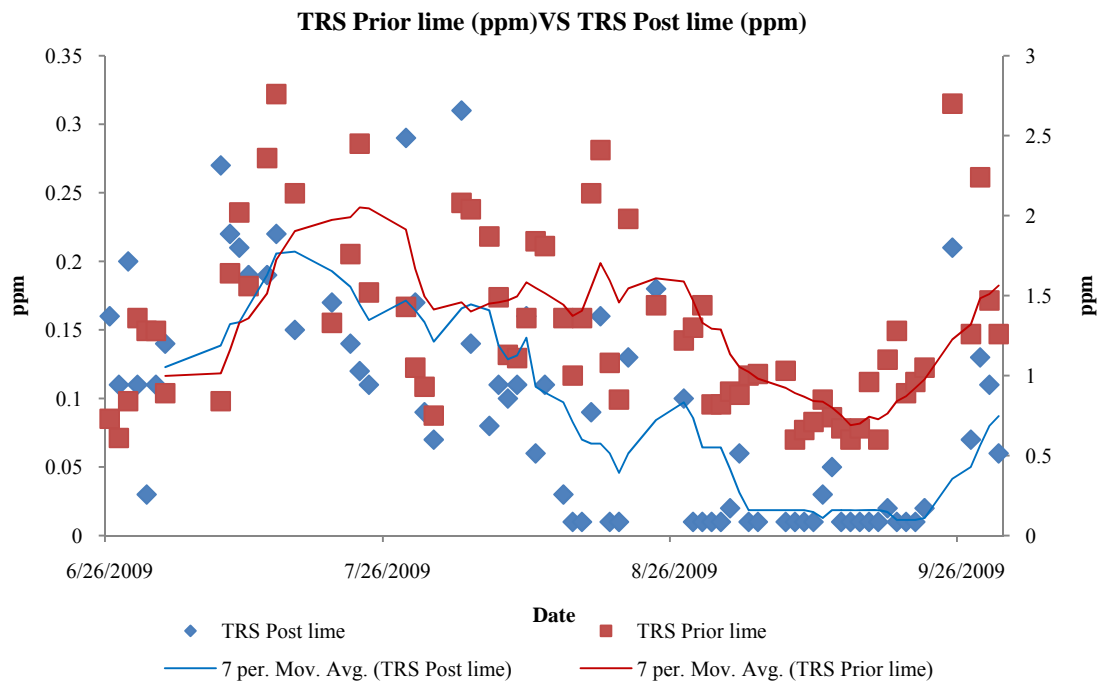
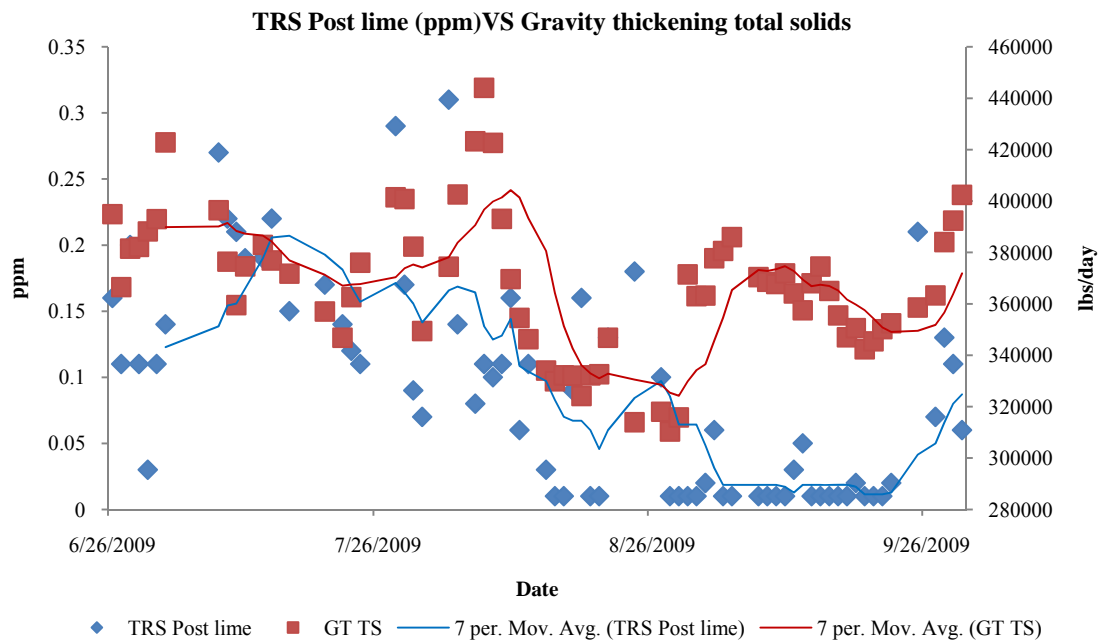


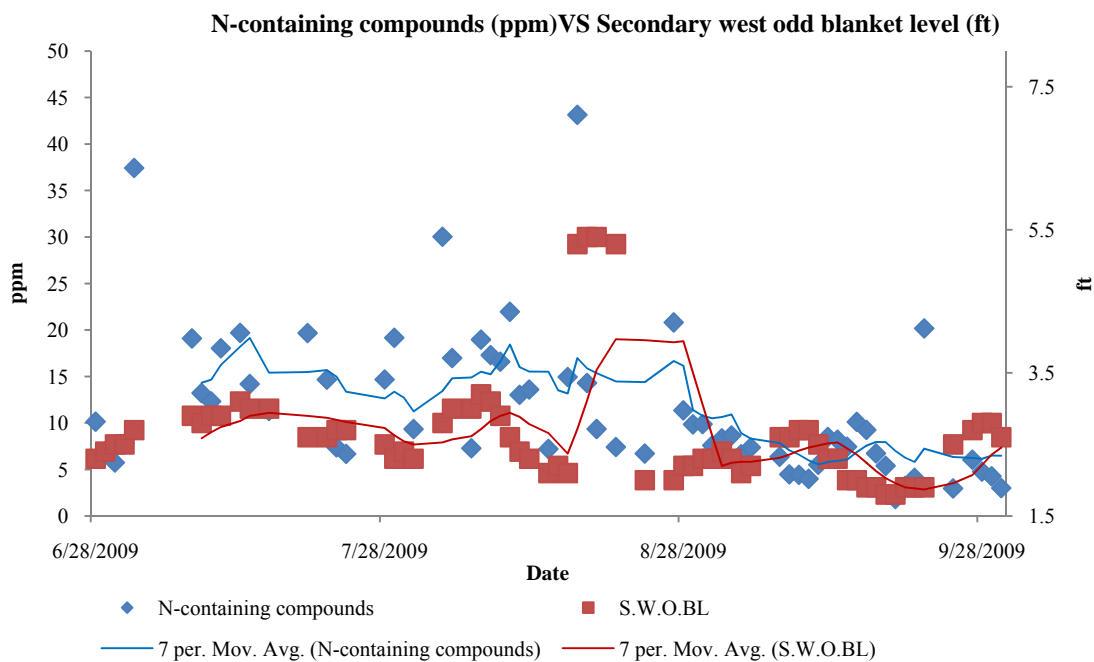
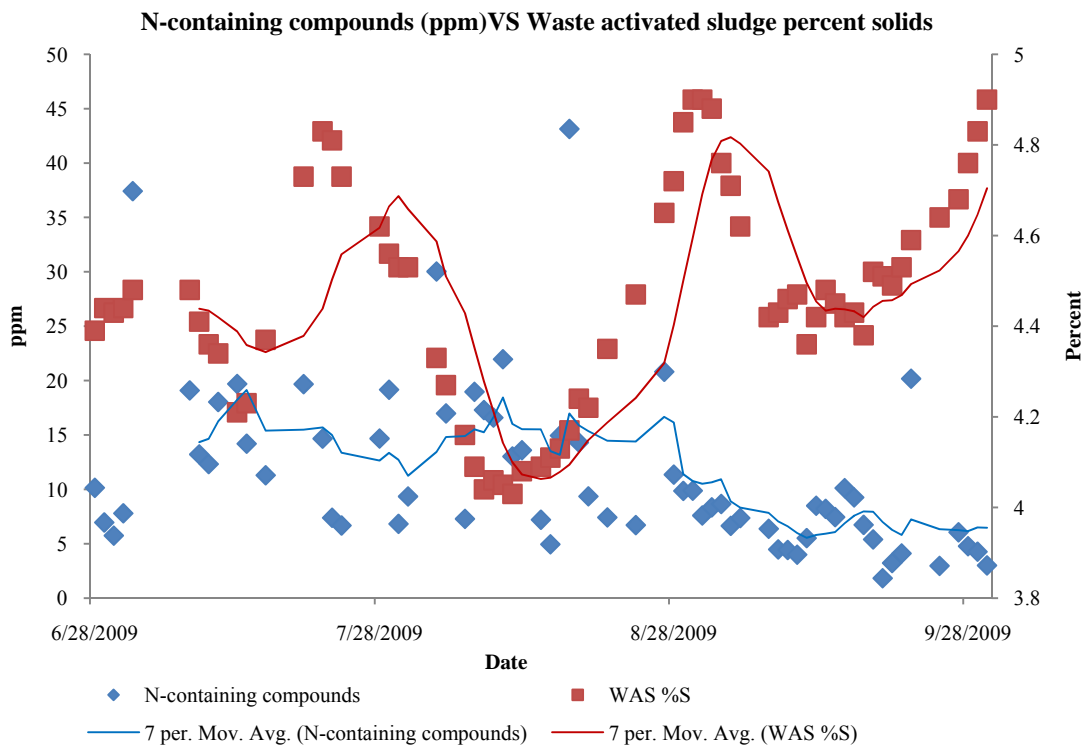


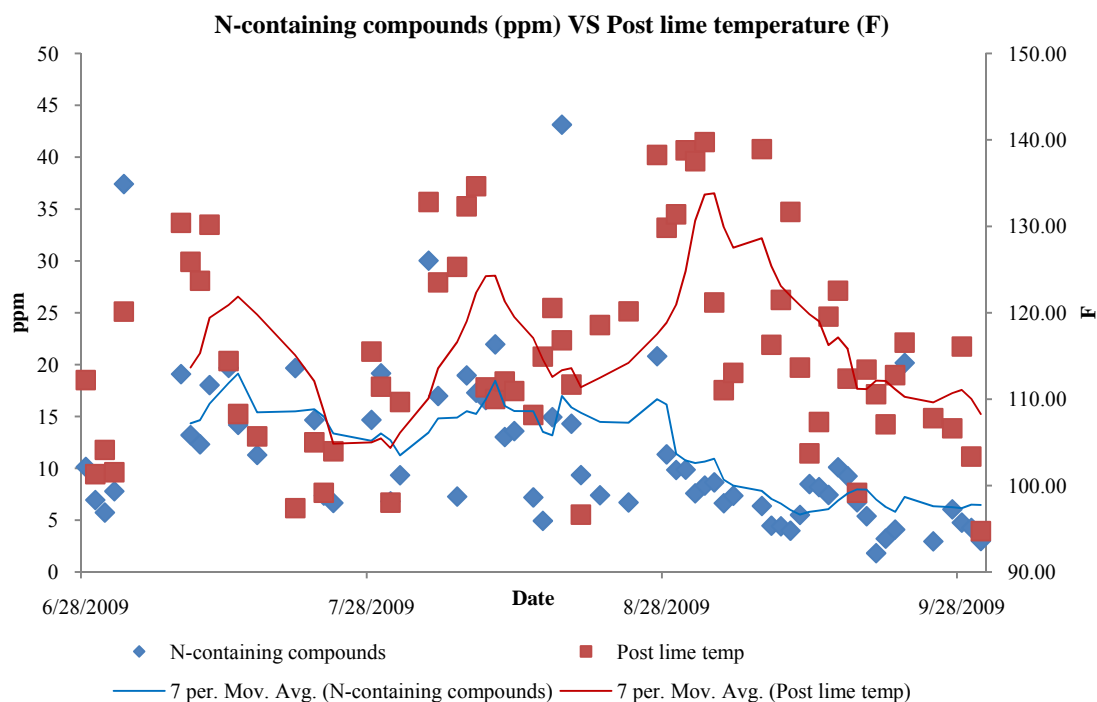
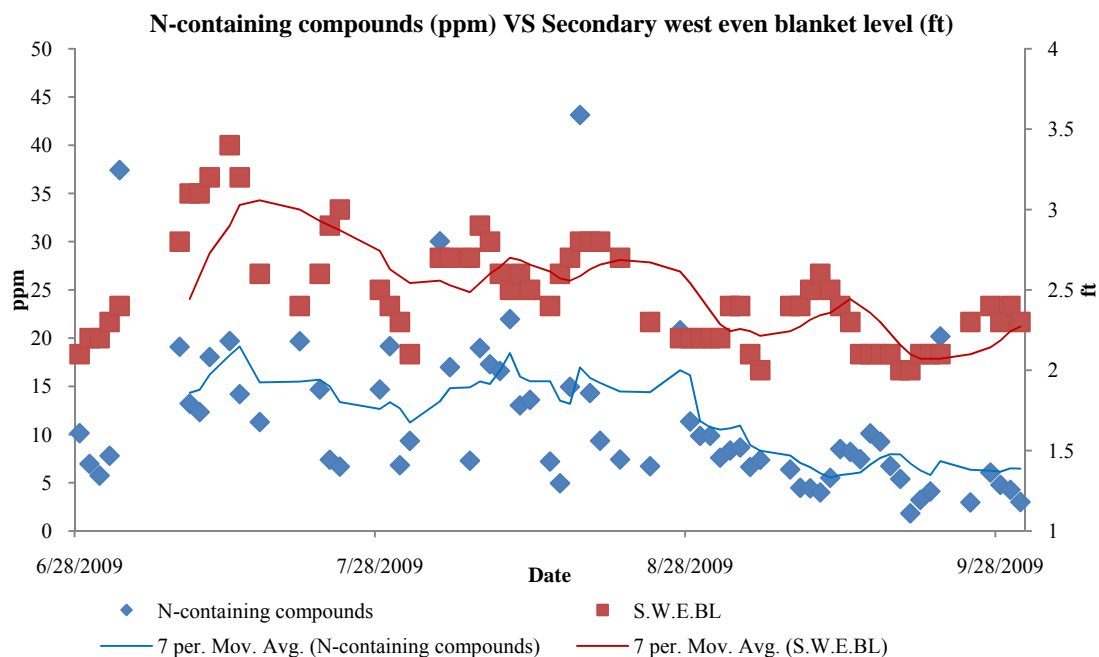




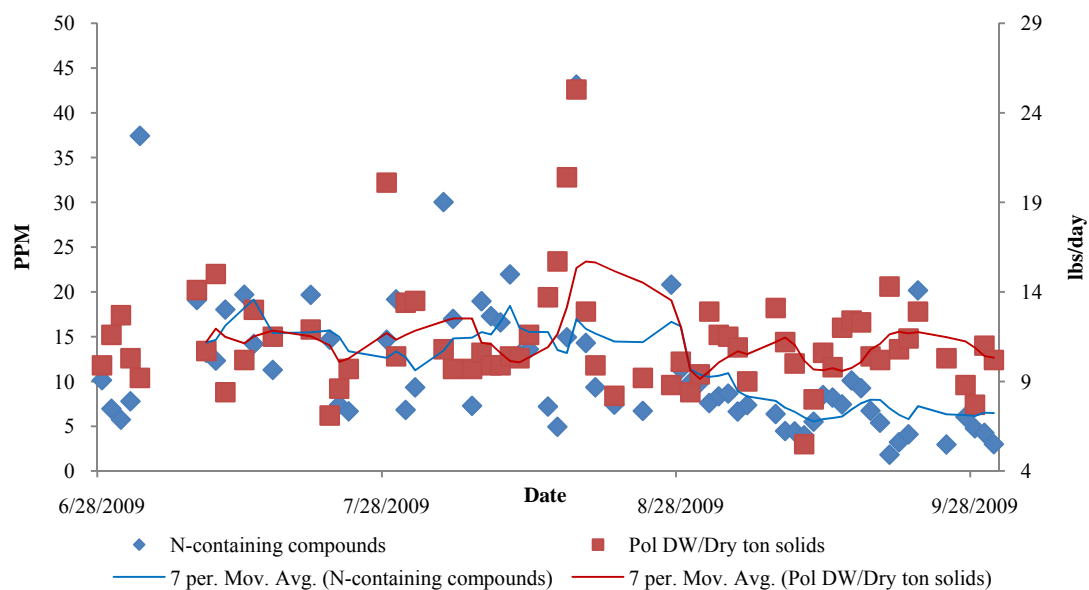




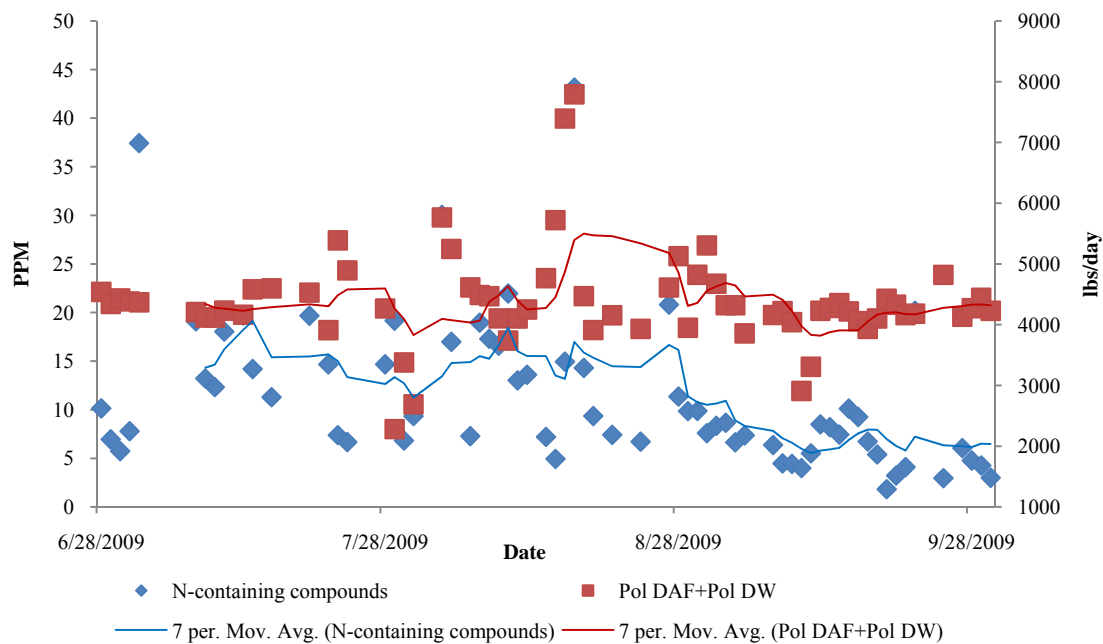


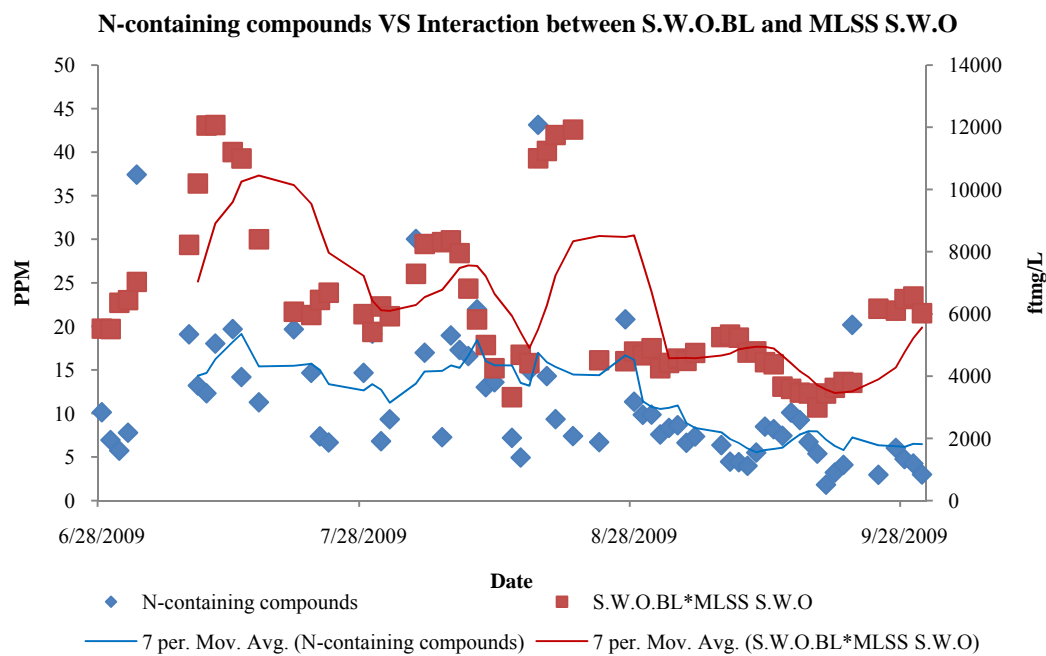
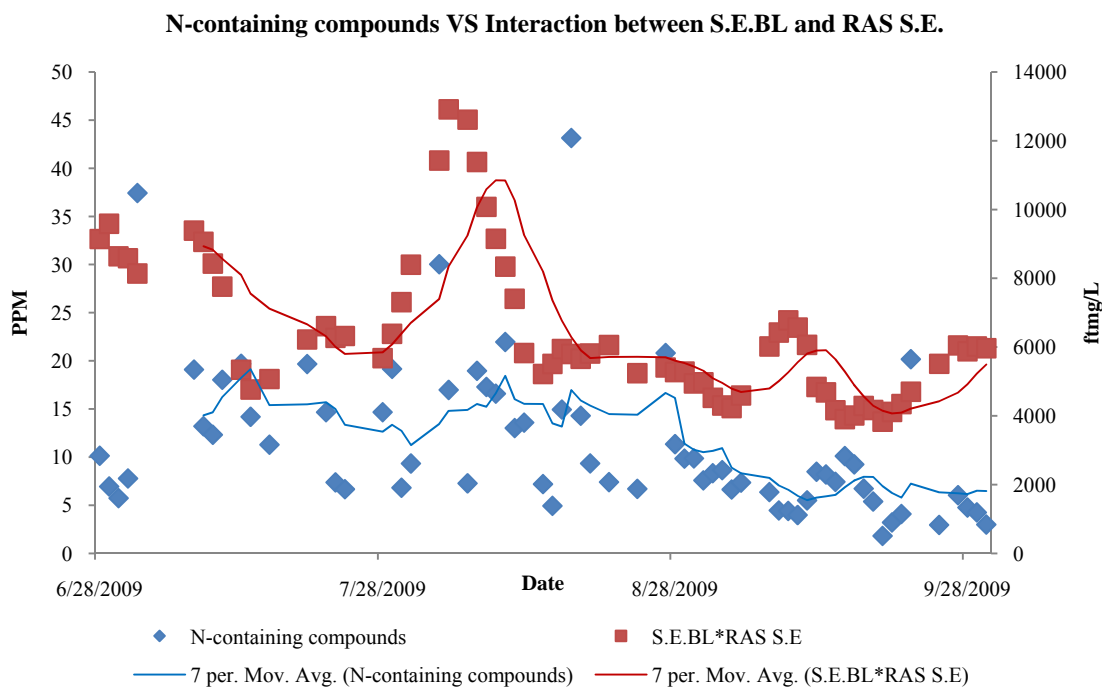


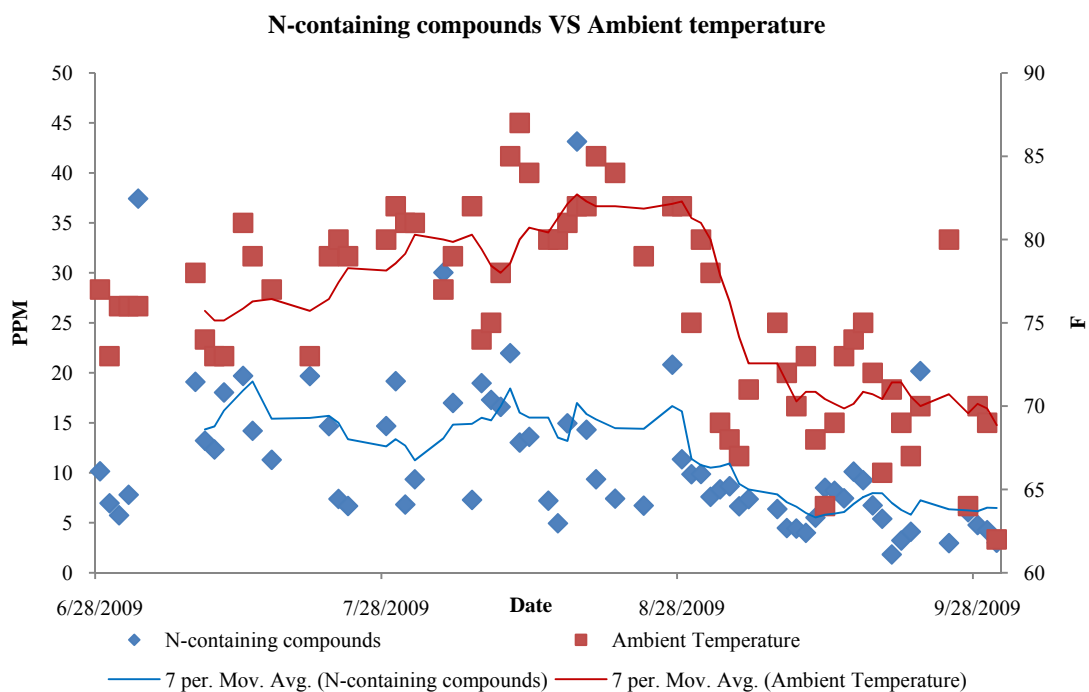
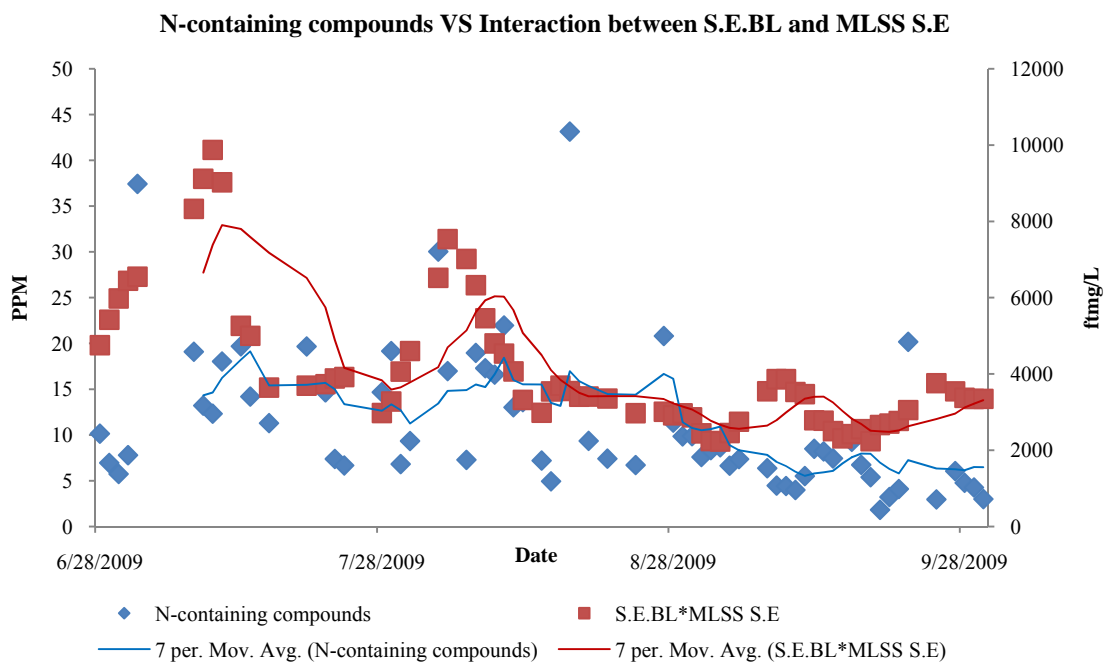
**N-containing compounds VS Polymer at dewatering process**

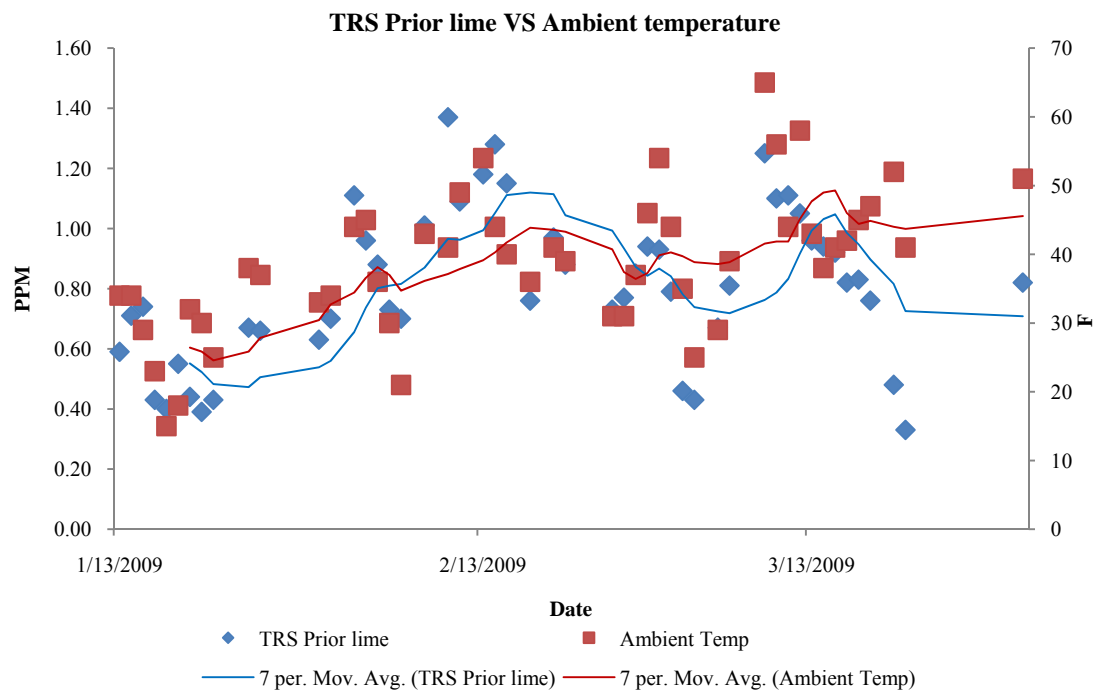
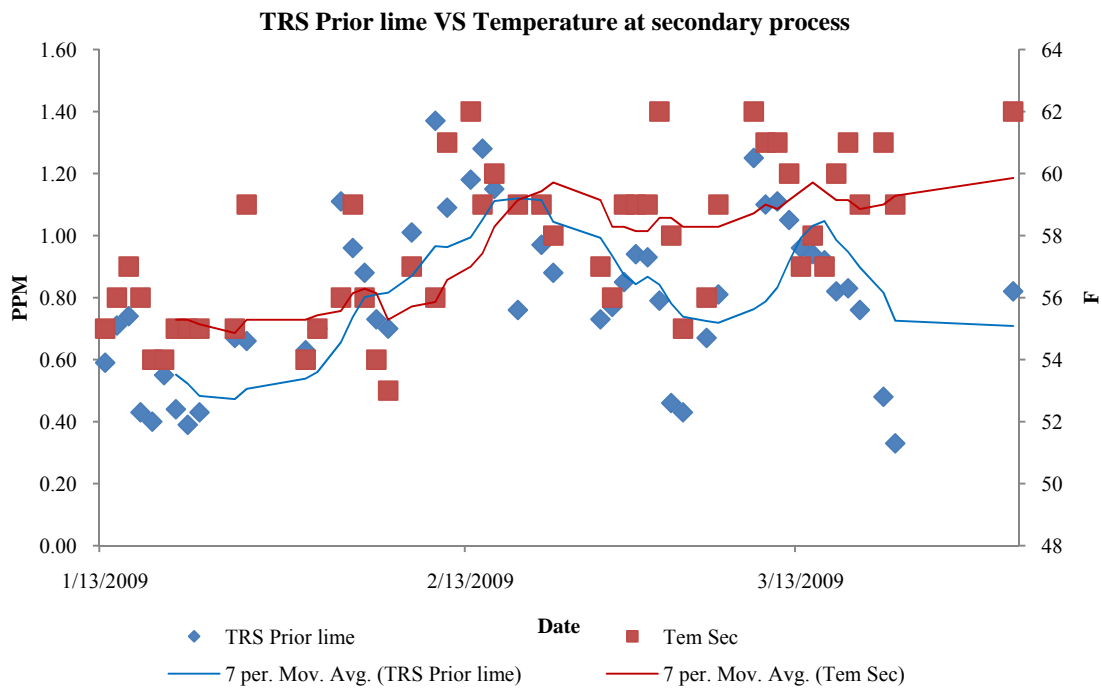


**N-containing compounds VS Sum of Polymer at DAF & Polymer at DW**

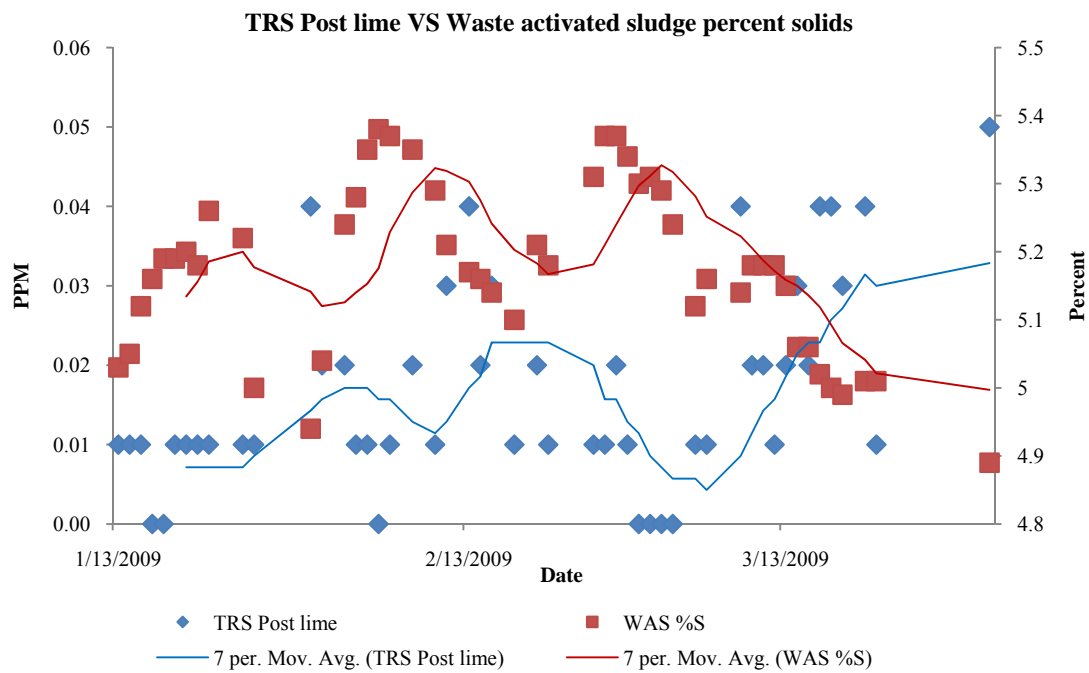
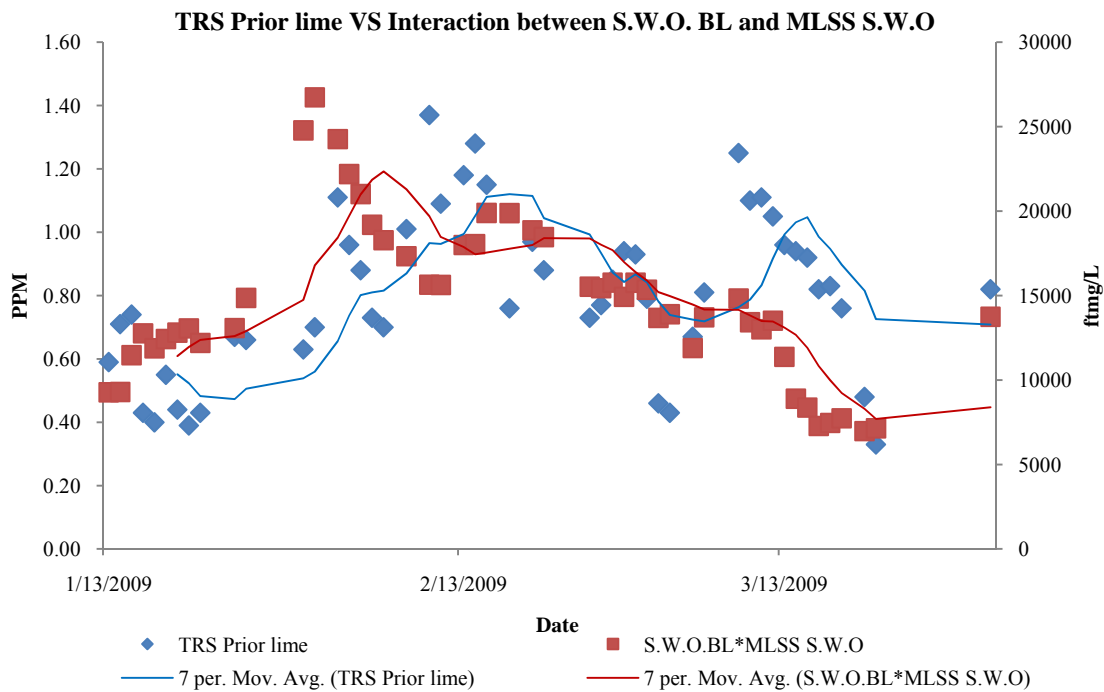


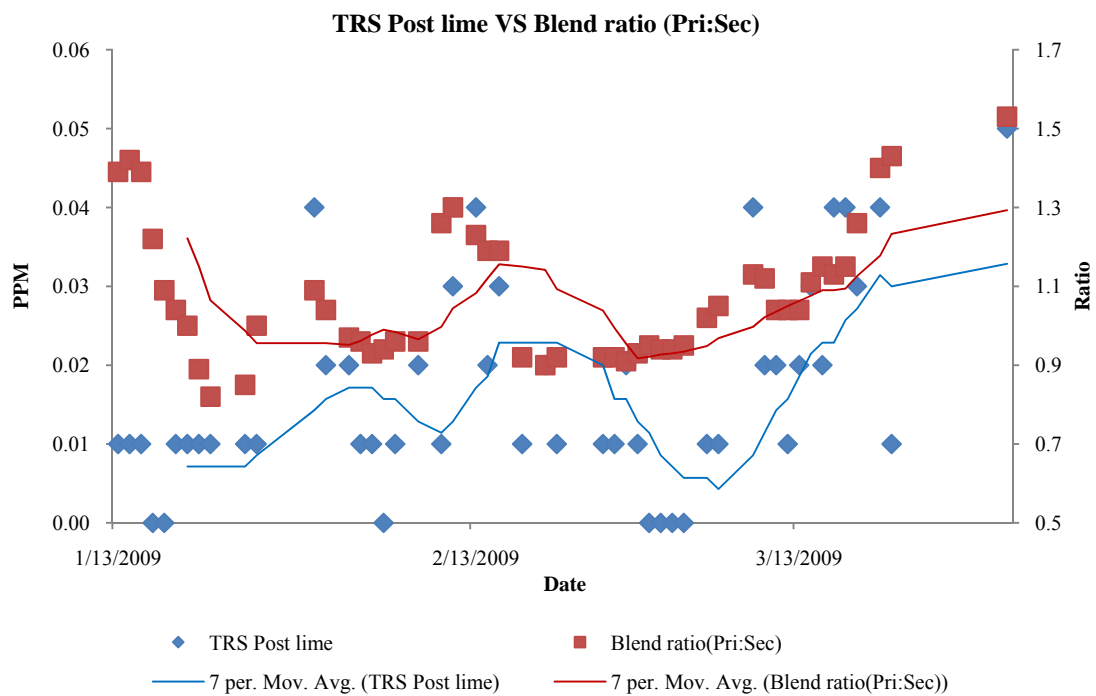
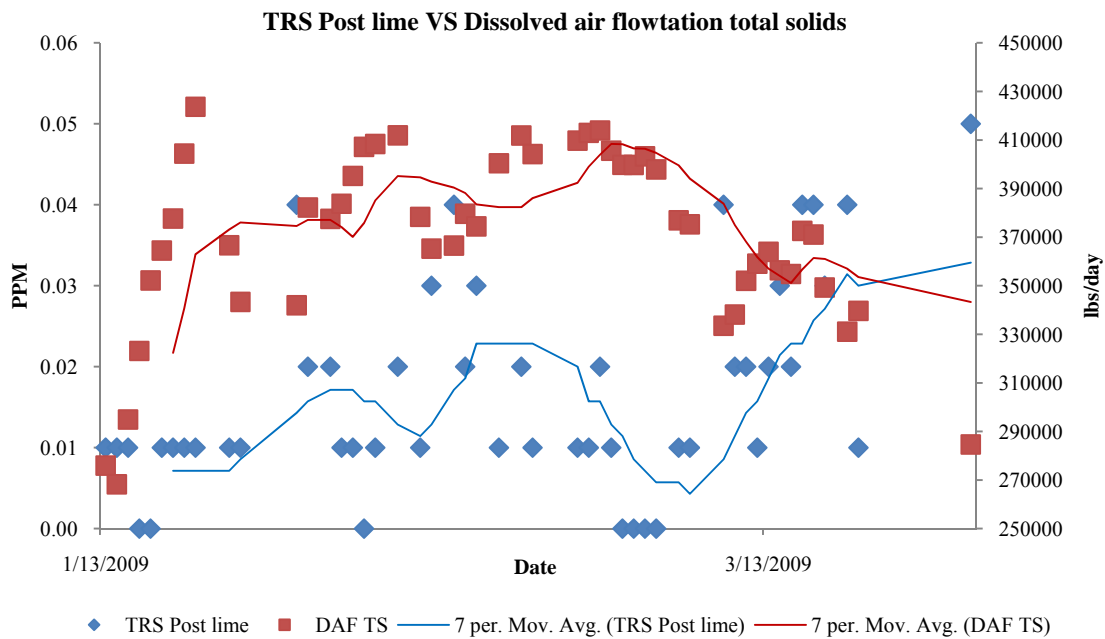


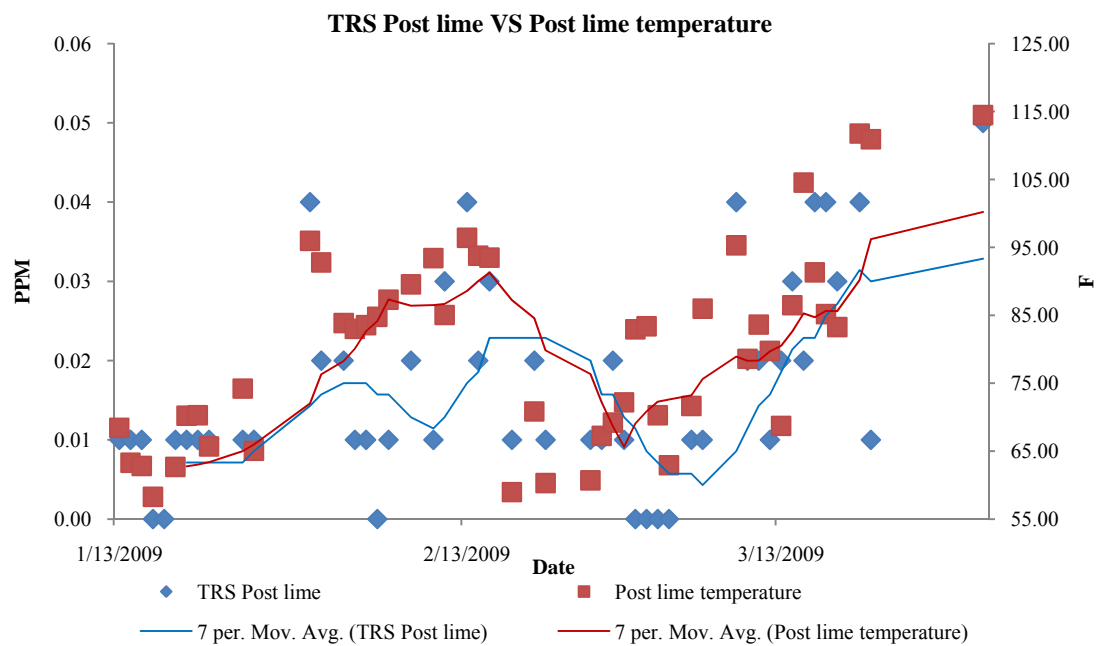
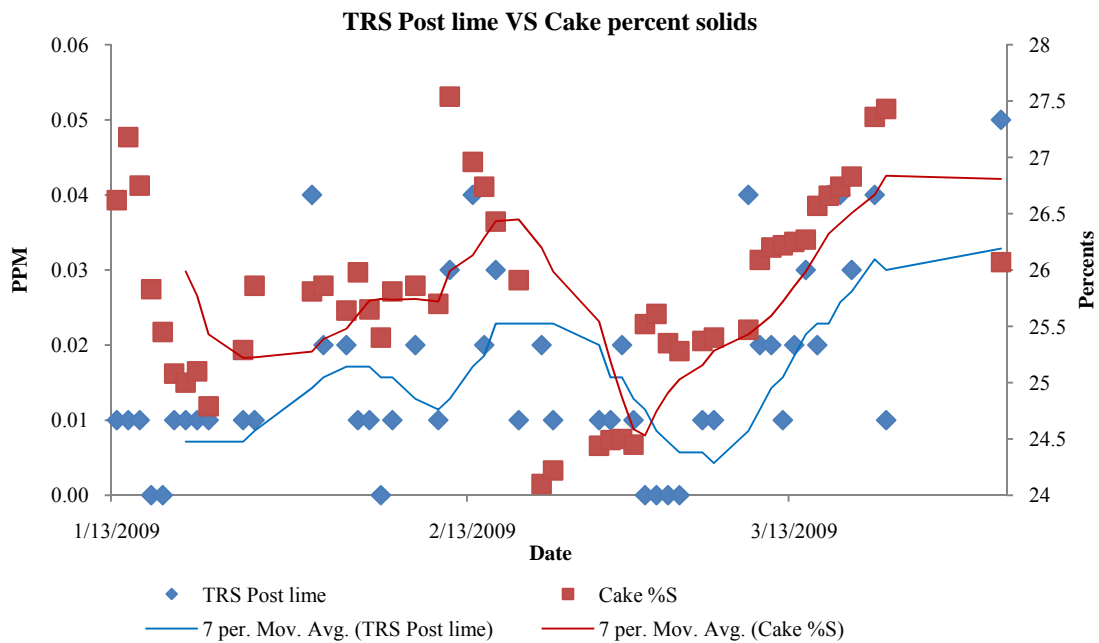


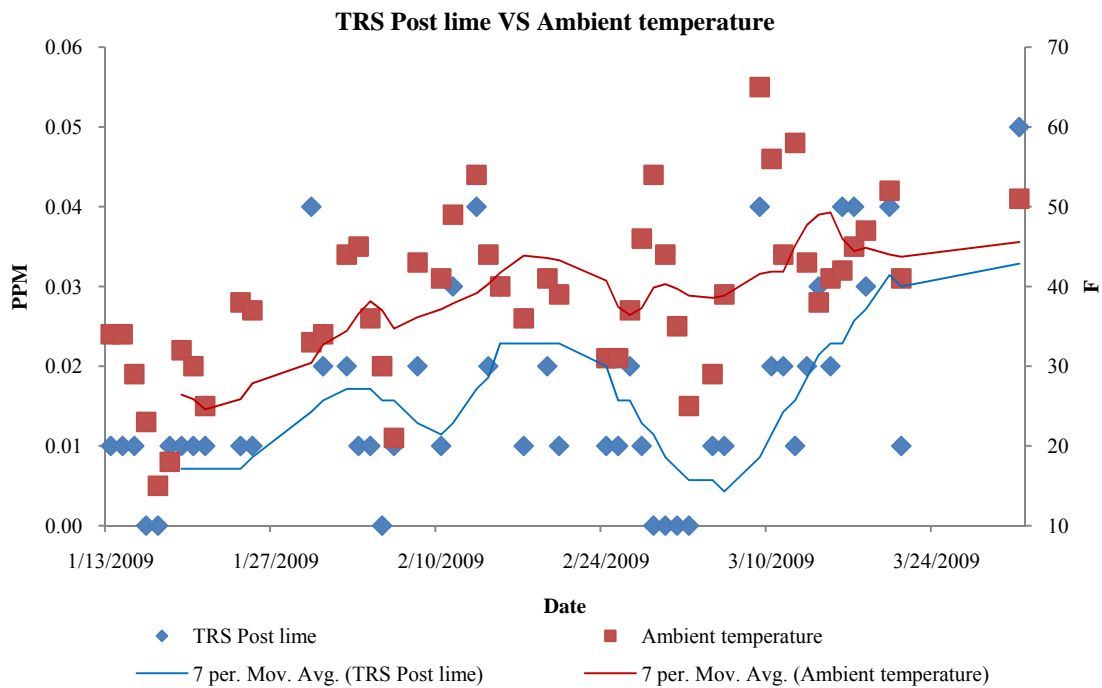
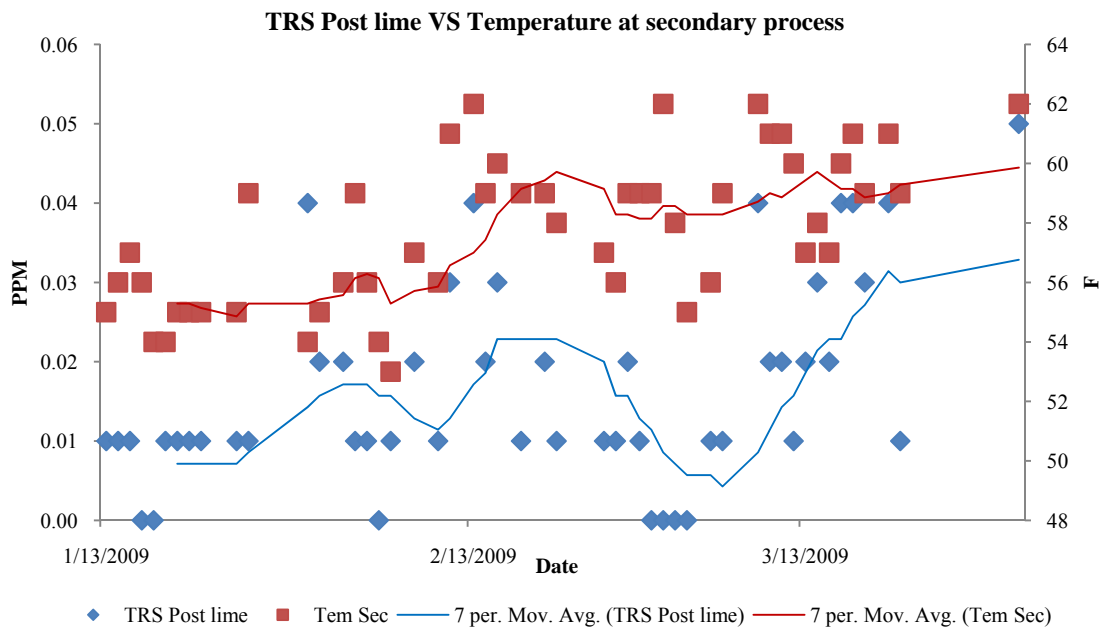


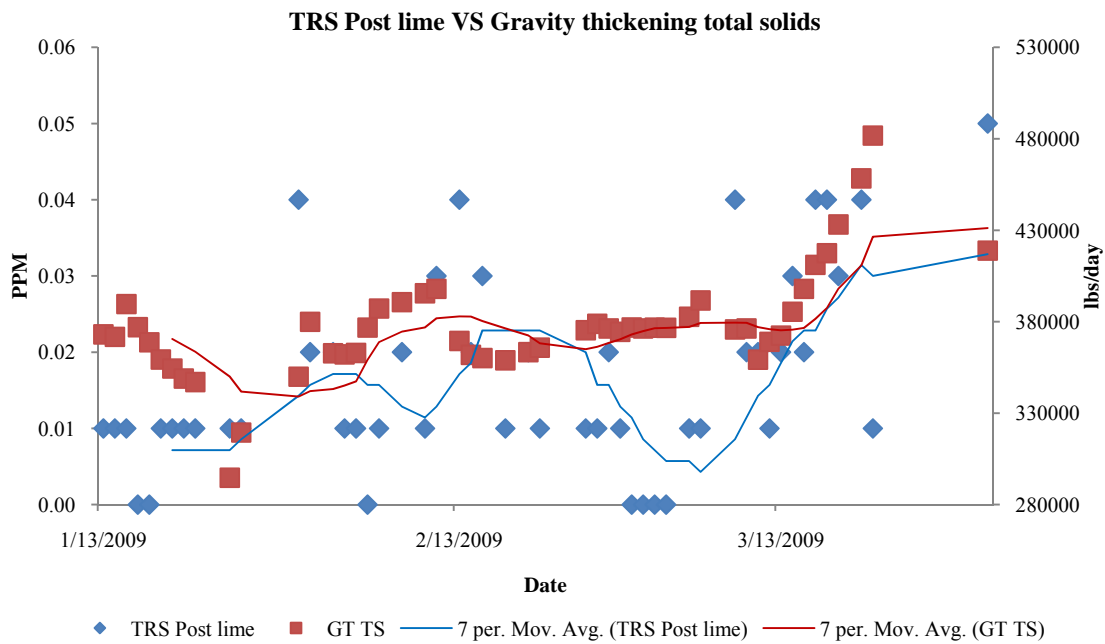
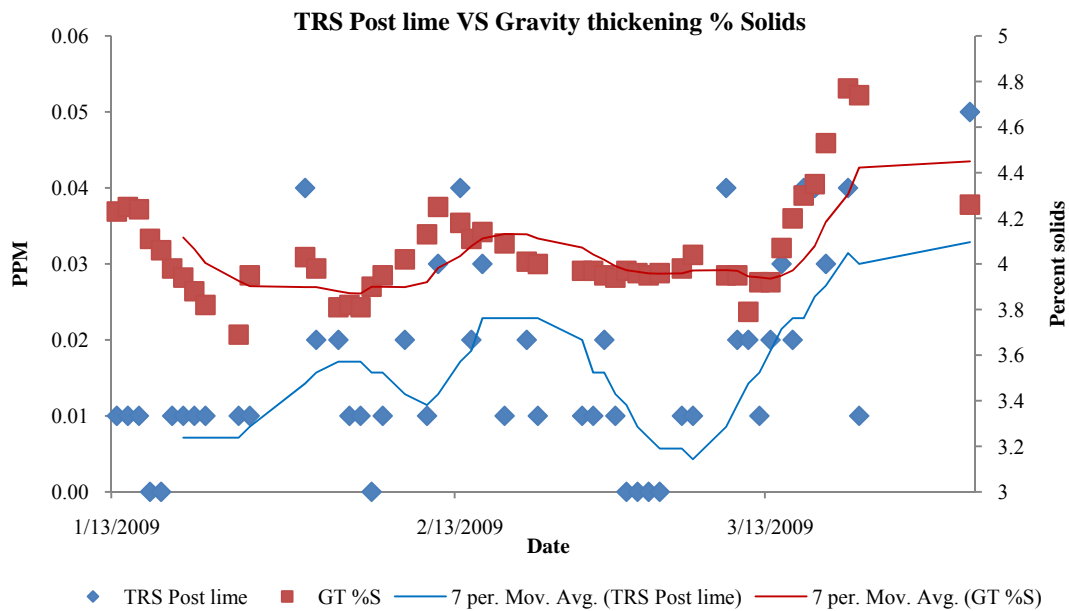


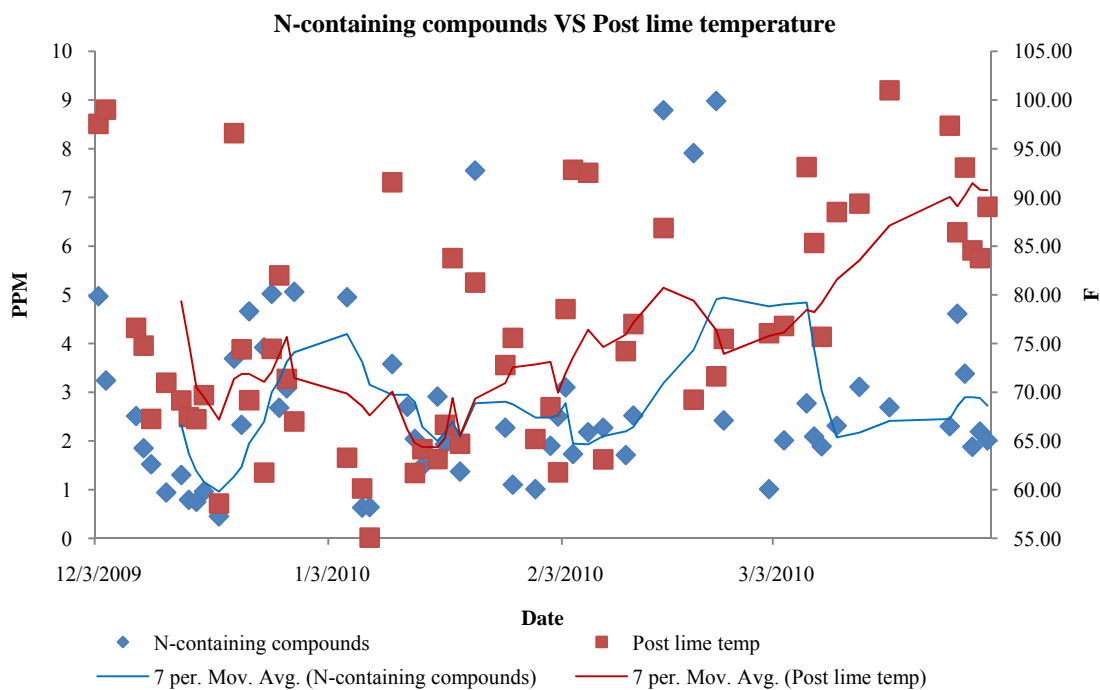
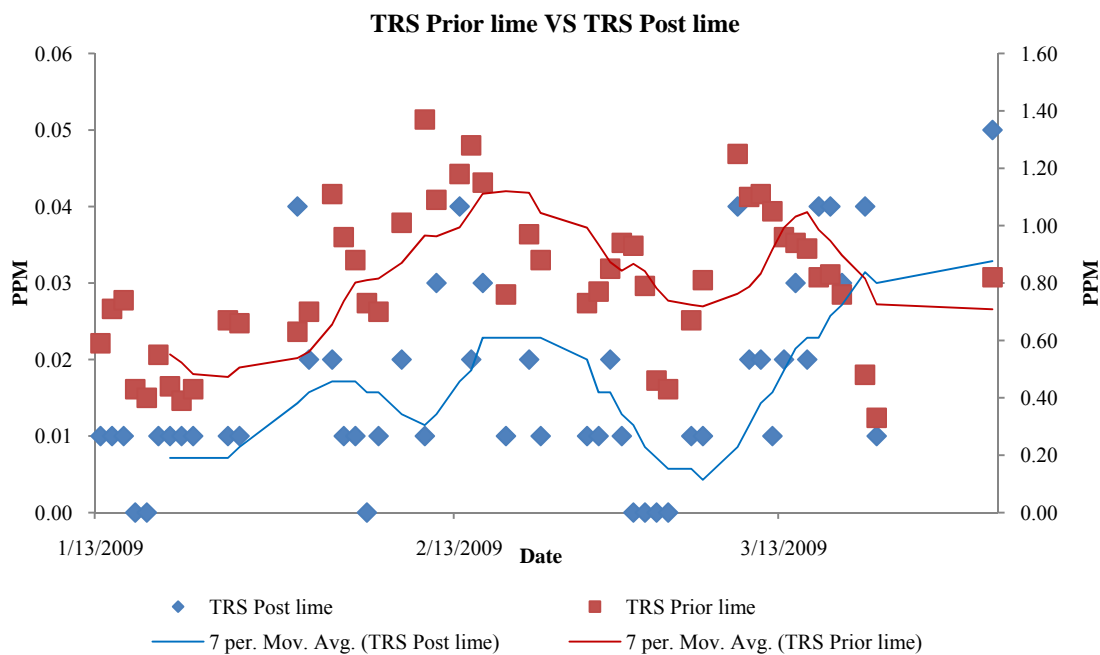


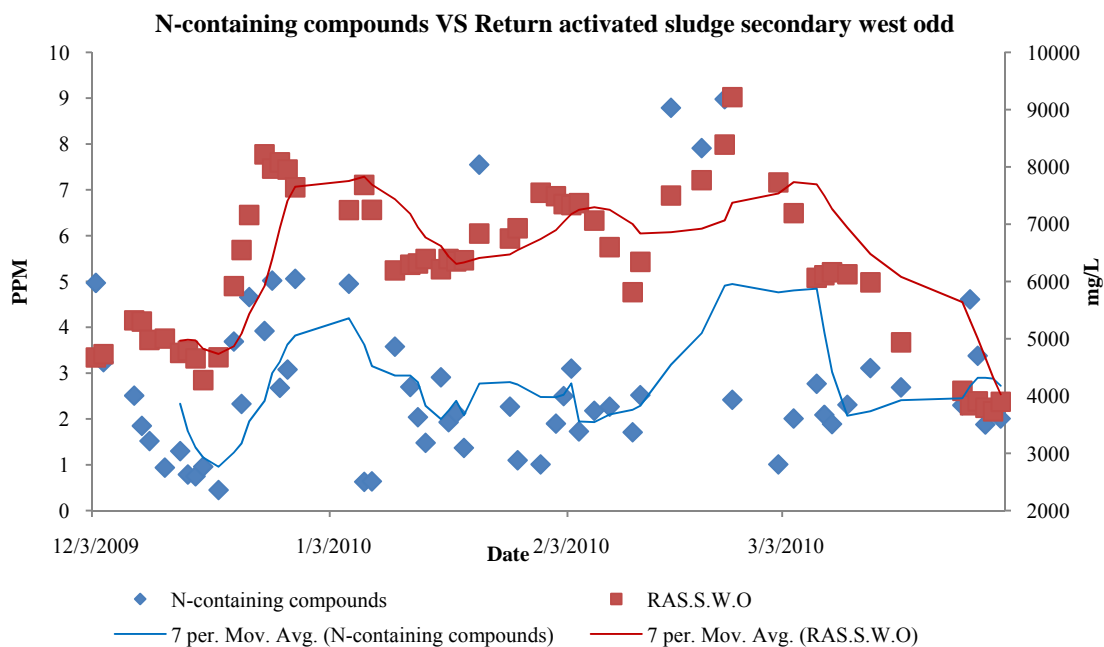
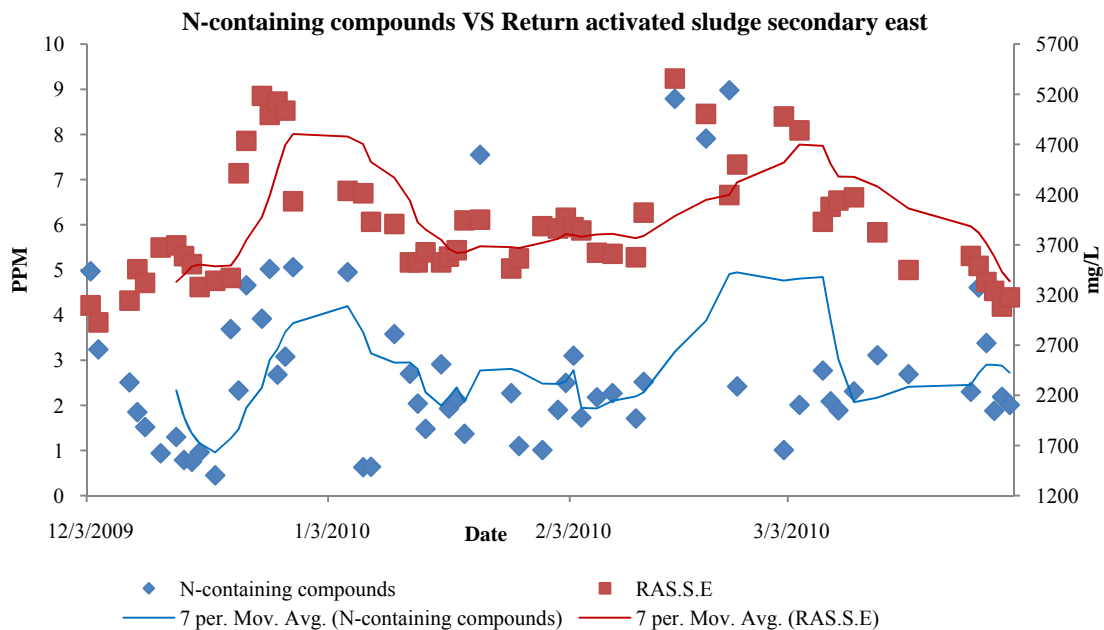


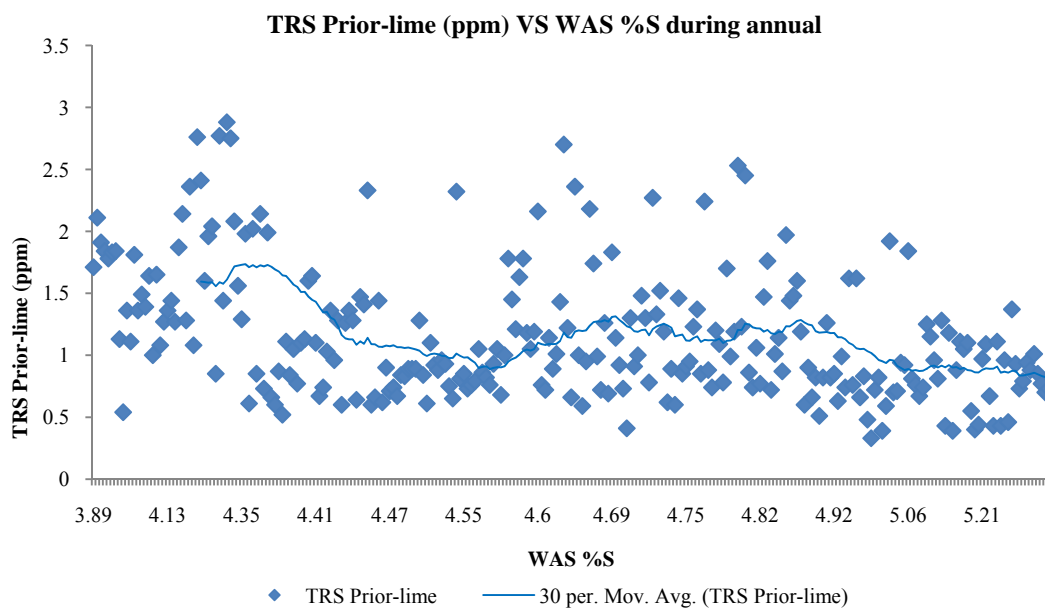
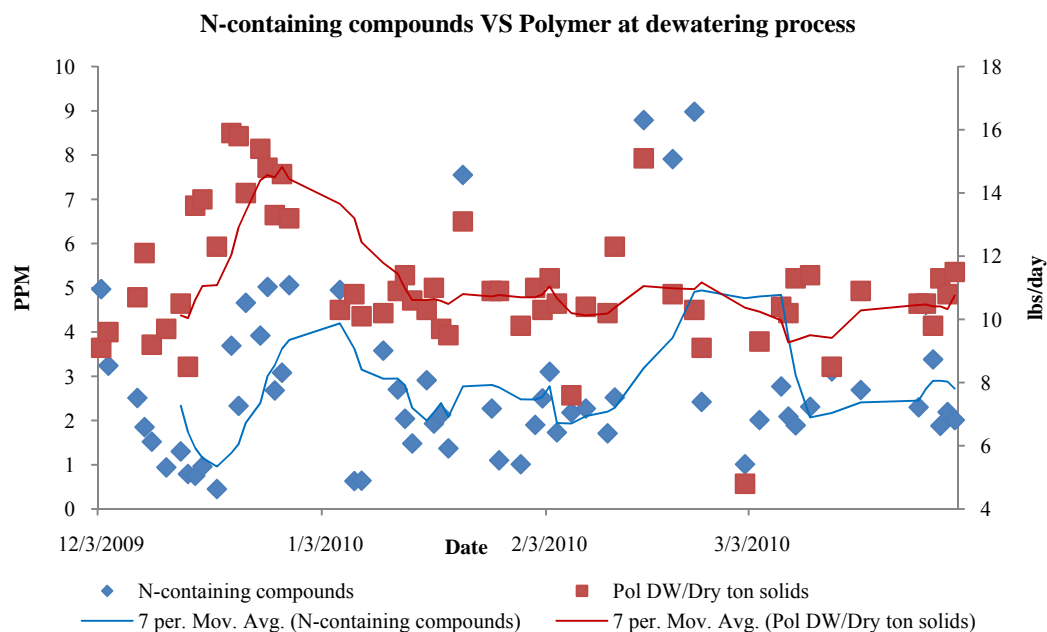




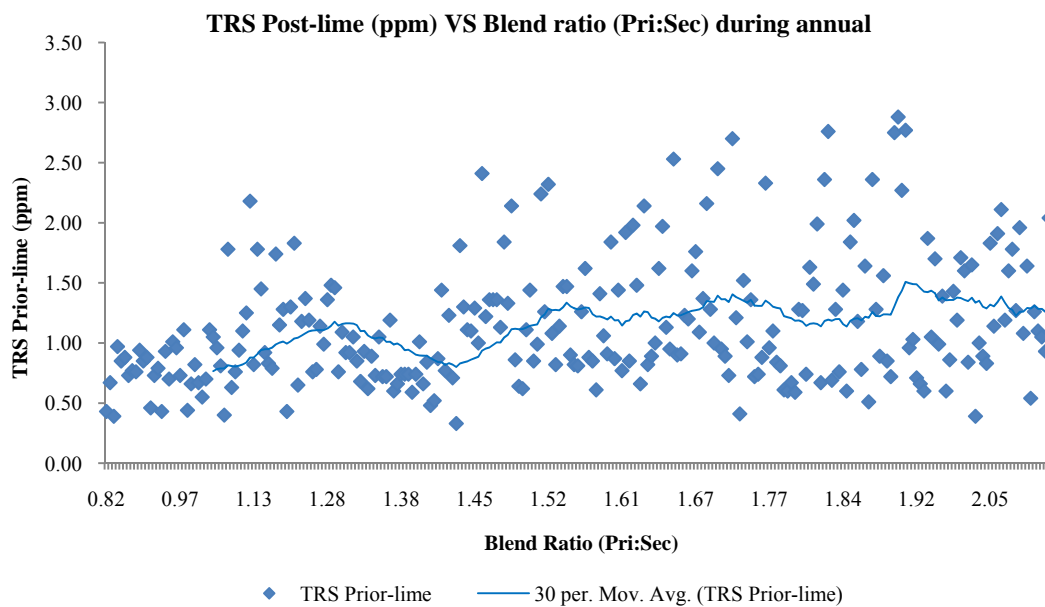
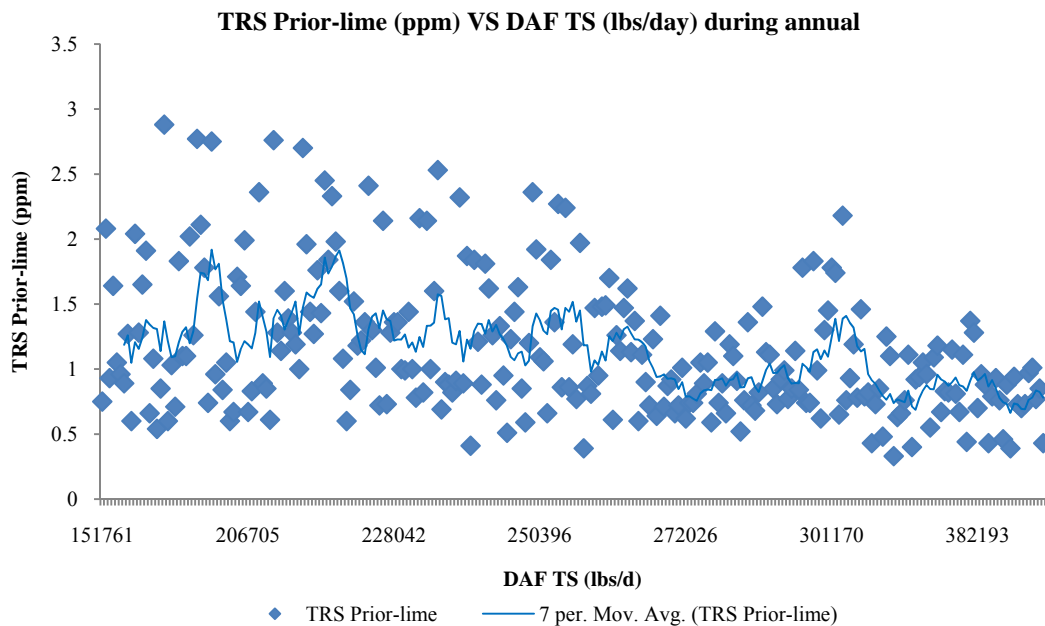


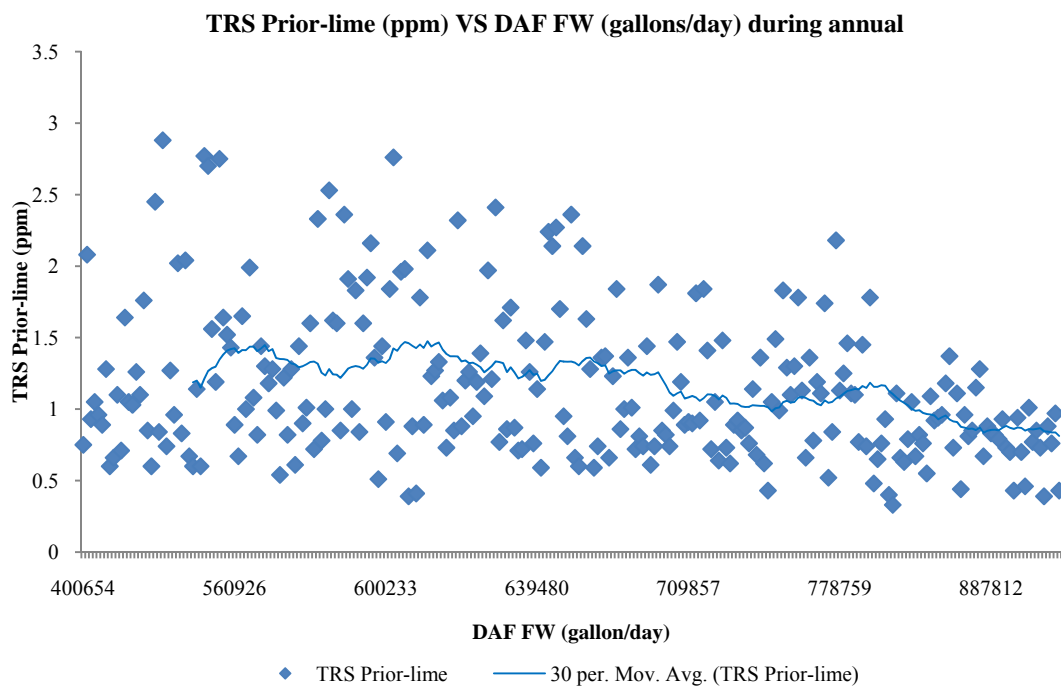
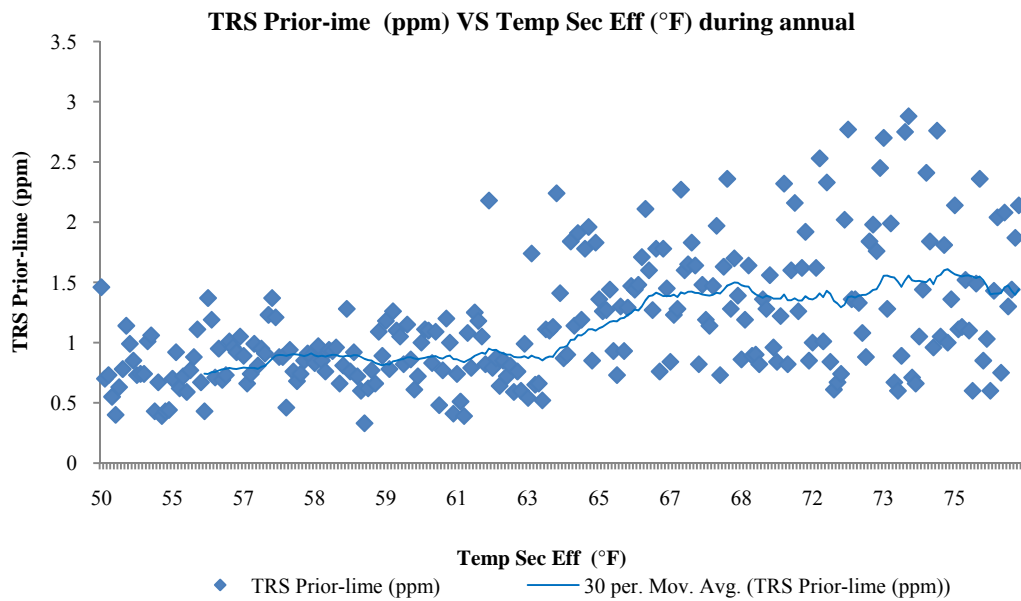


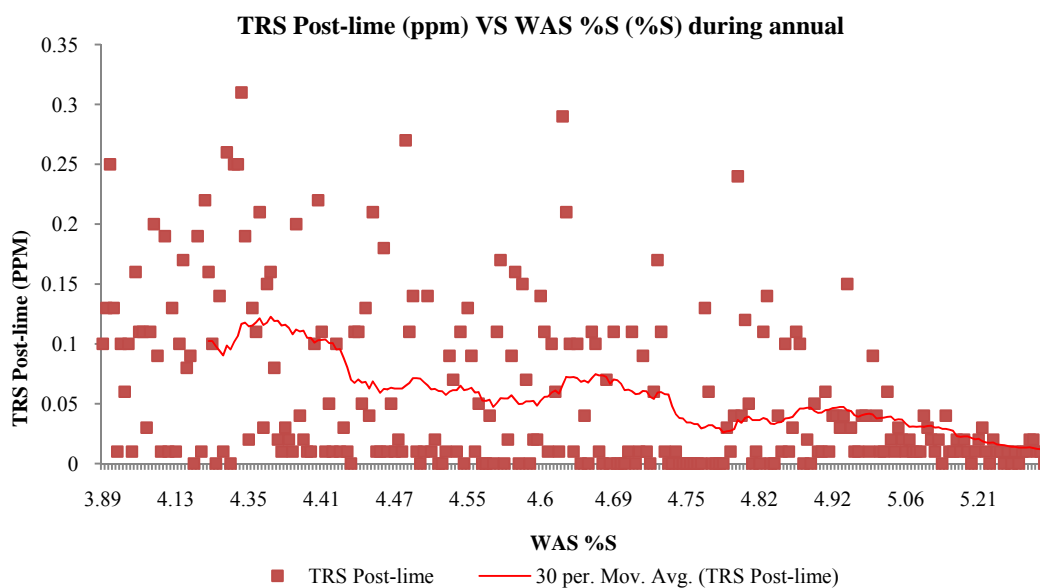
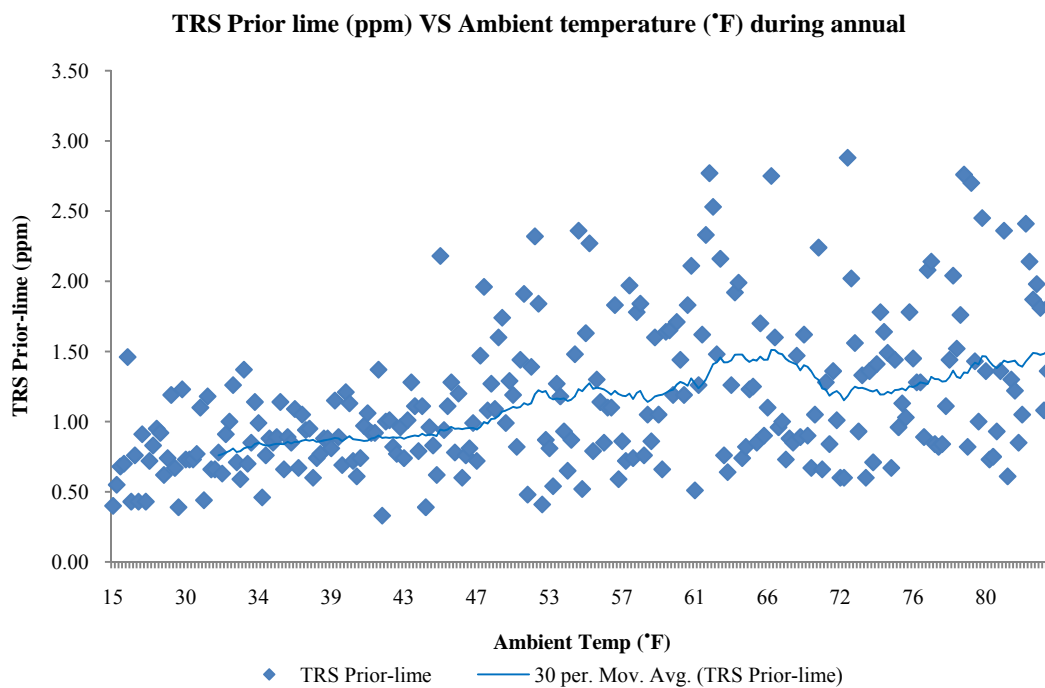


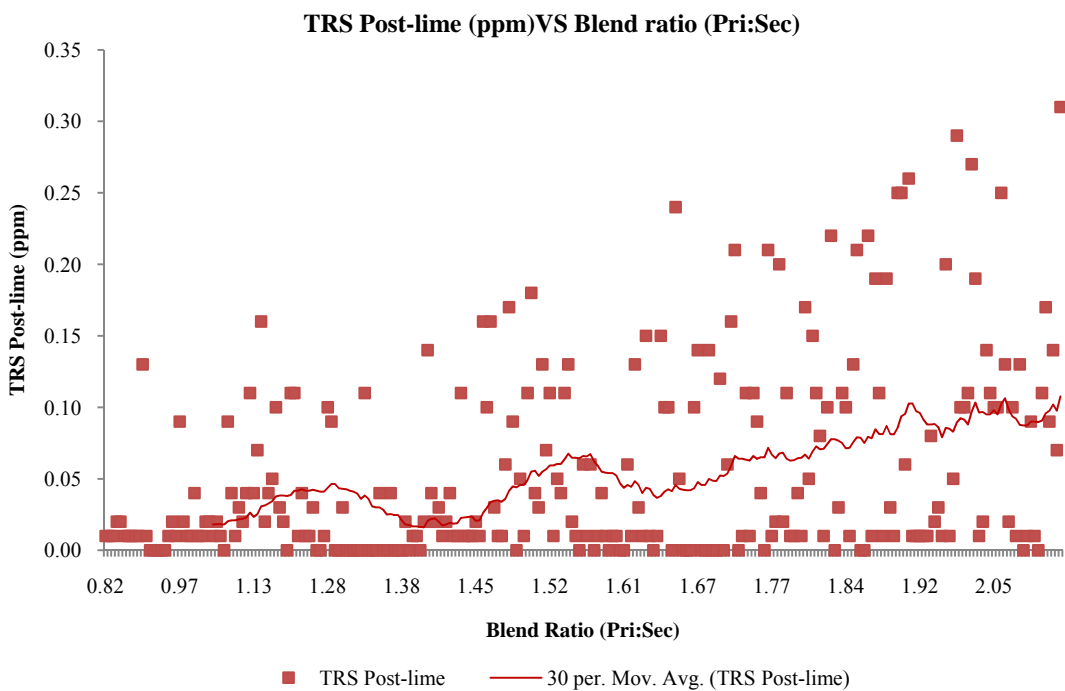
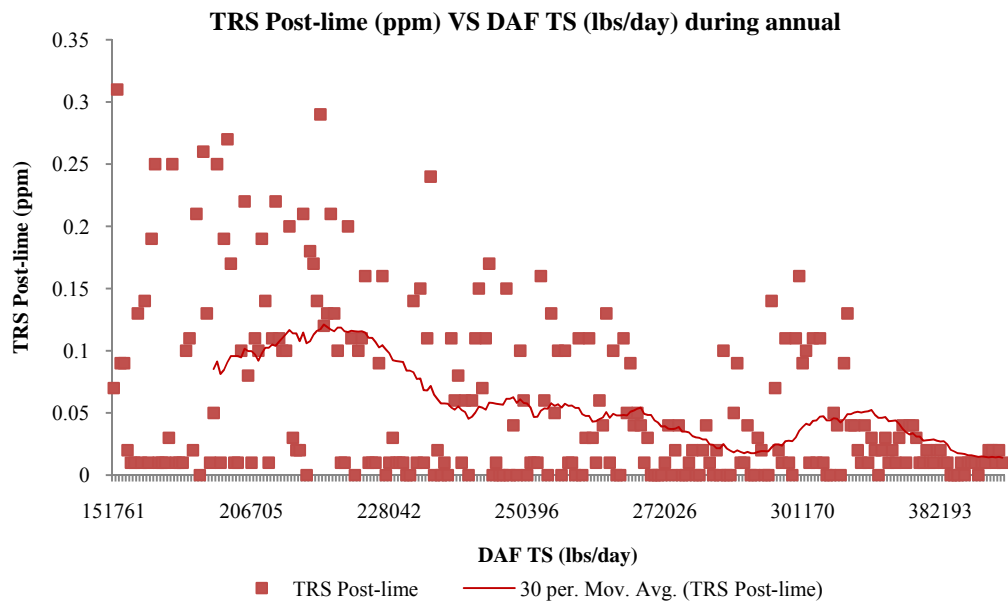


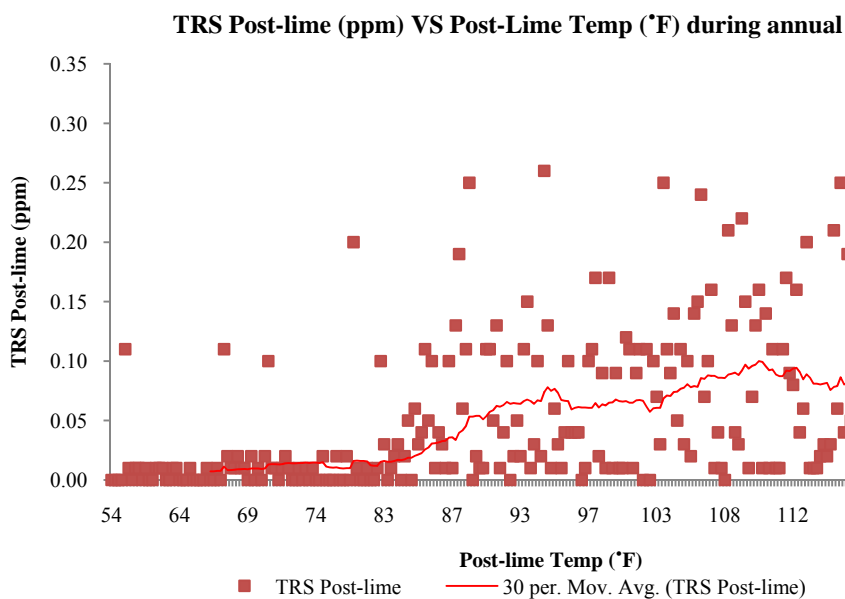
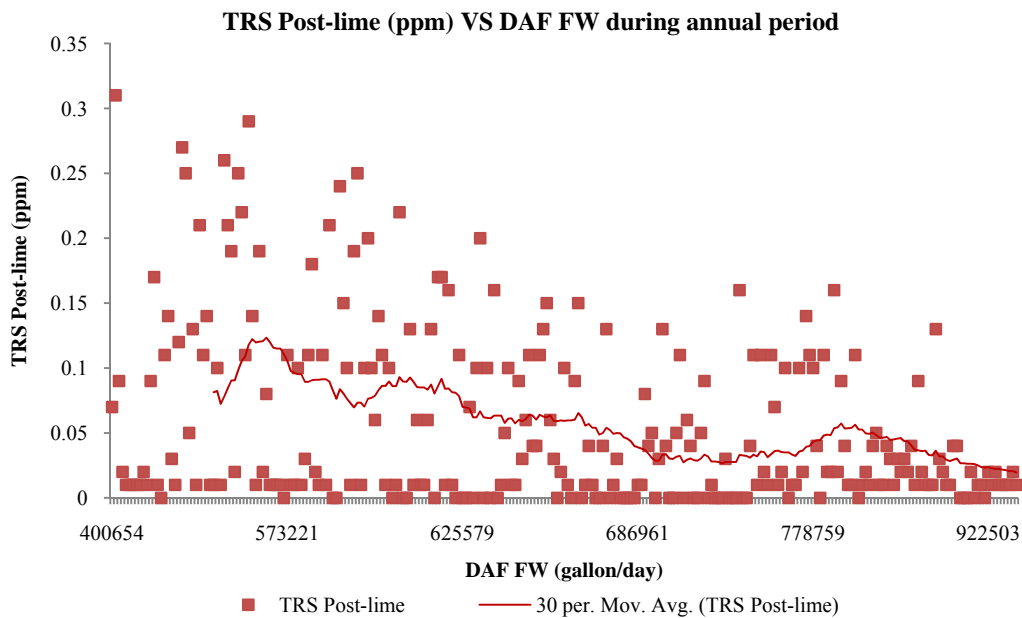


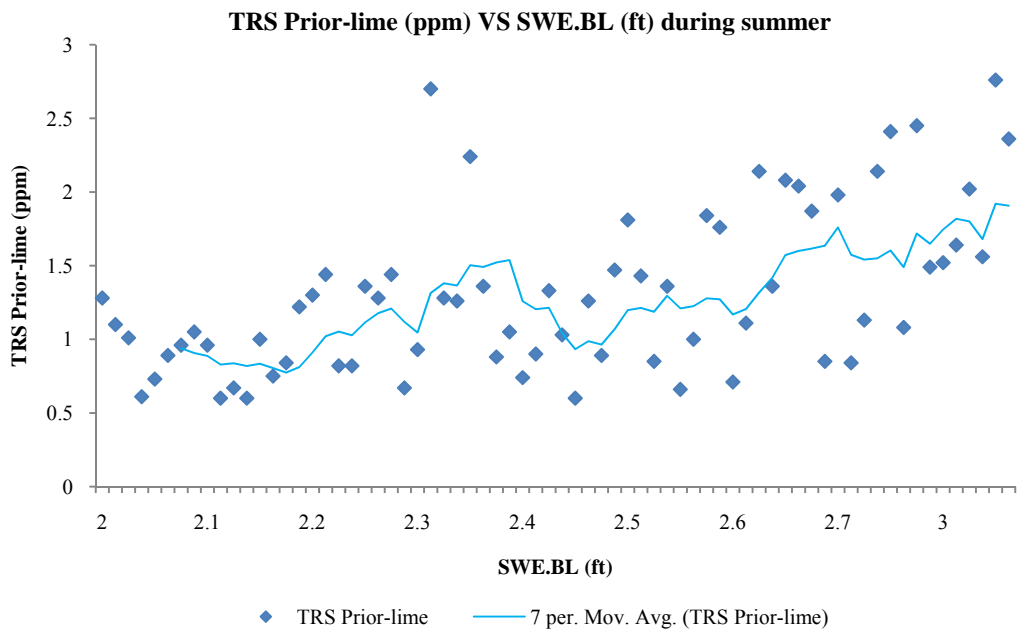
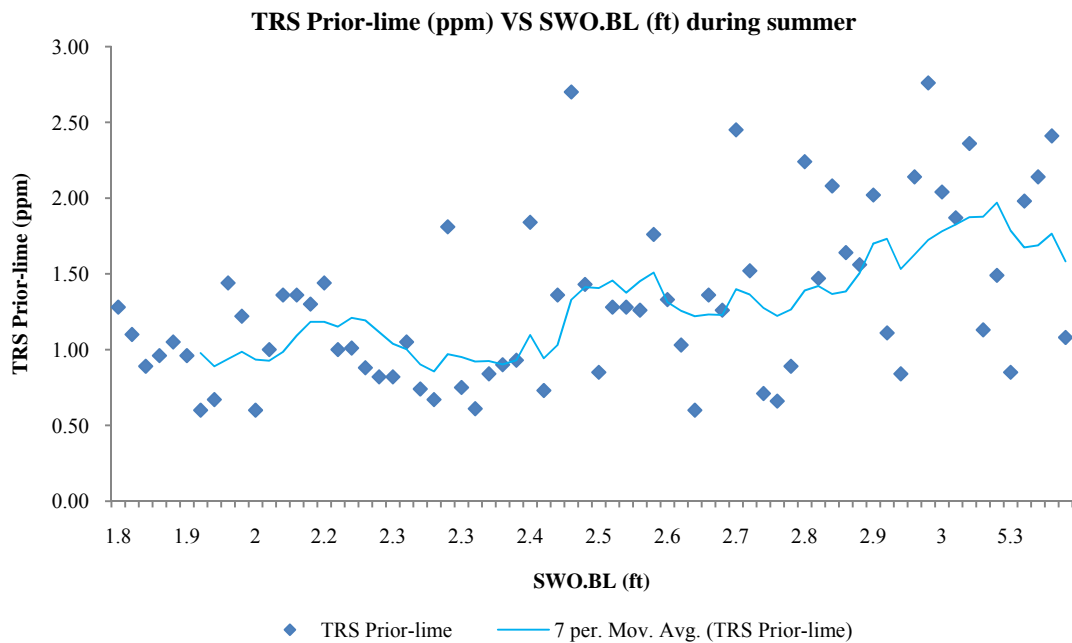


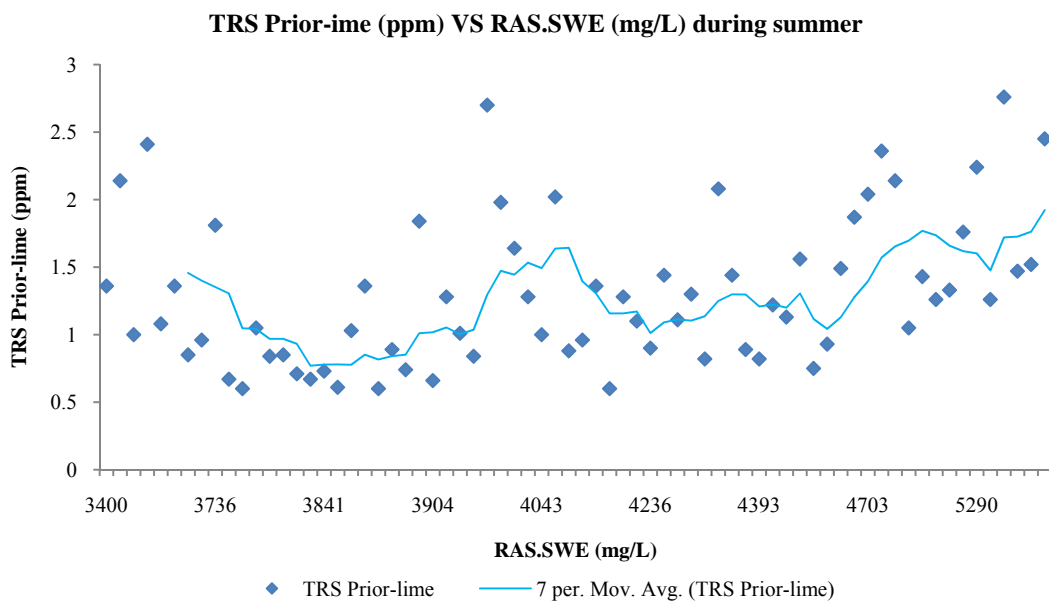
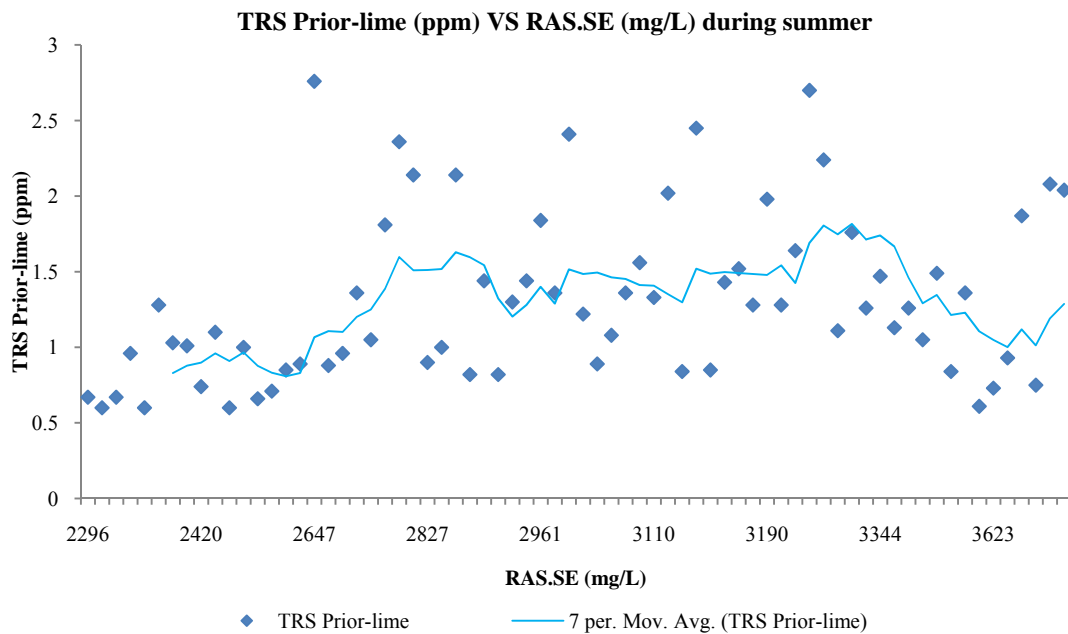




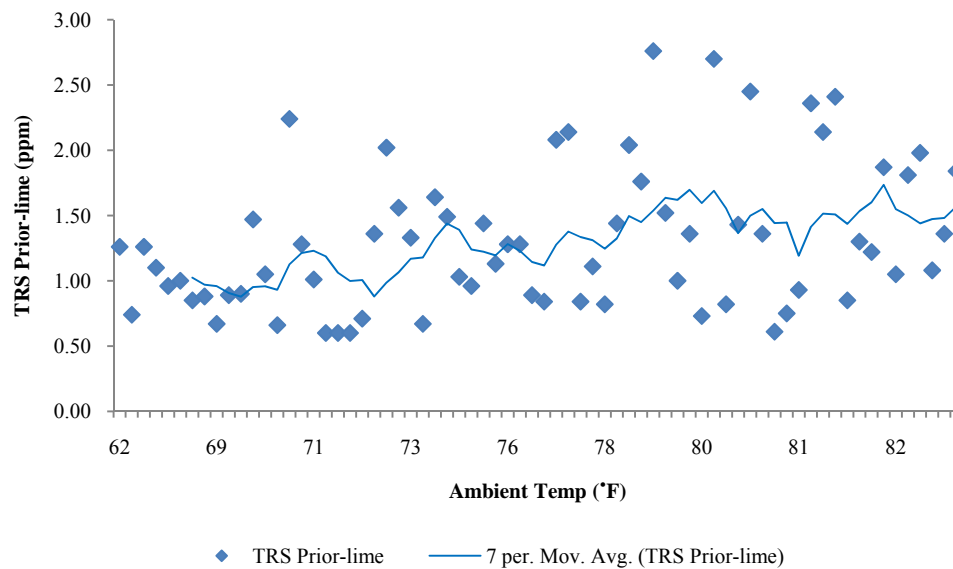




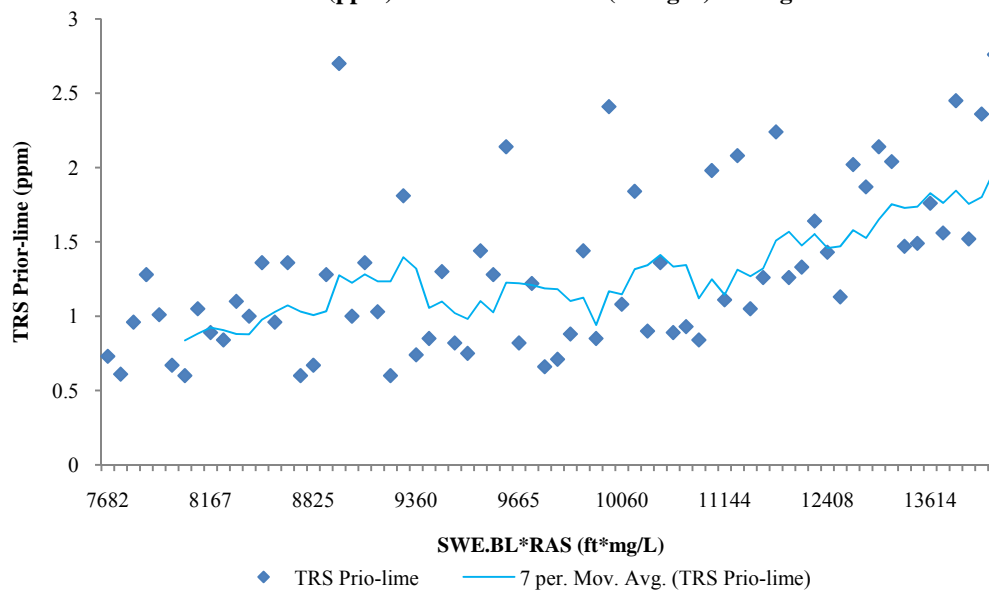




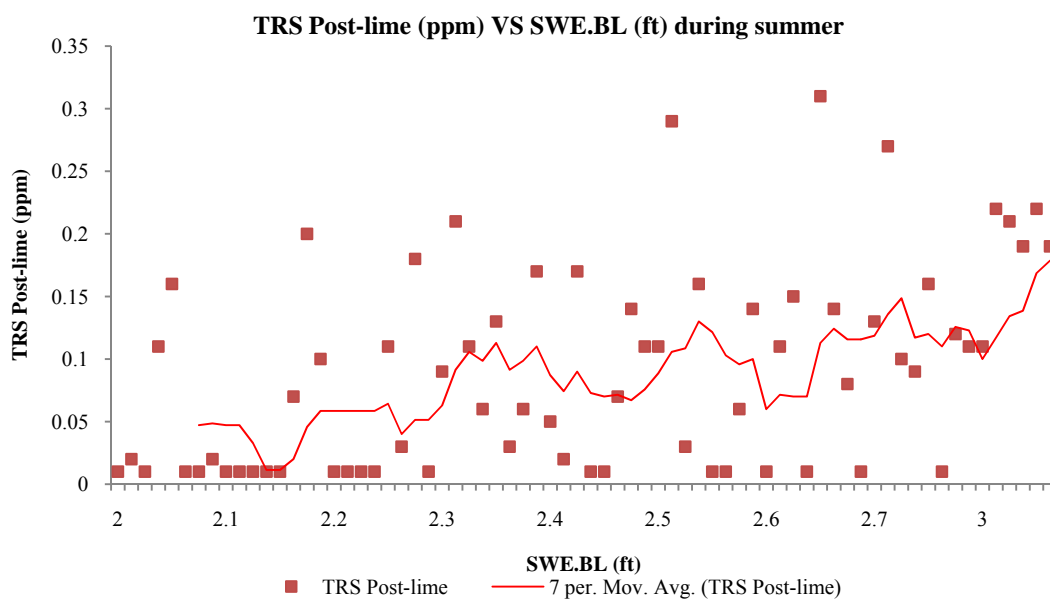
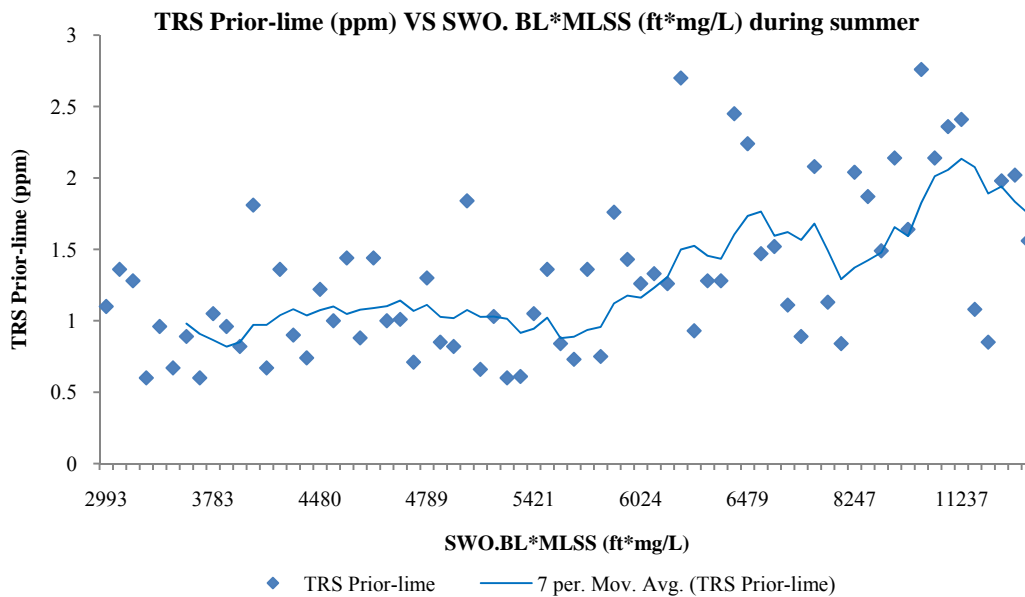
**TRS Prior-line (ppm) VS Amb Temp (°F) during summer**

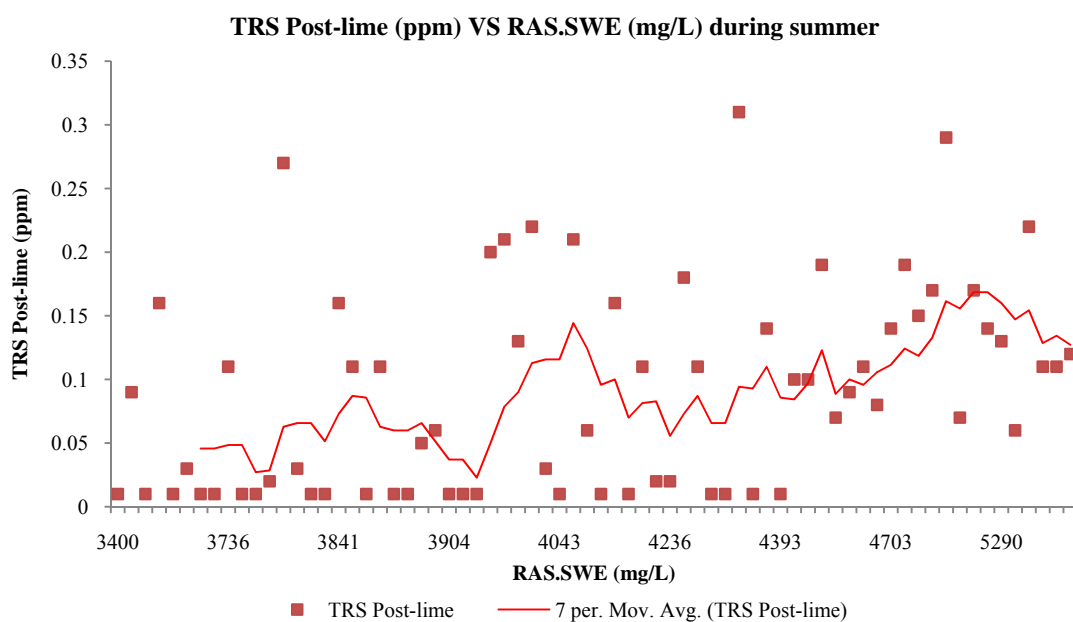
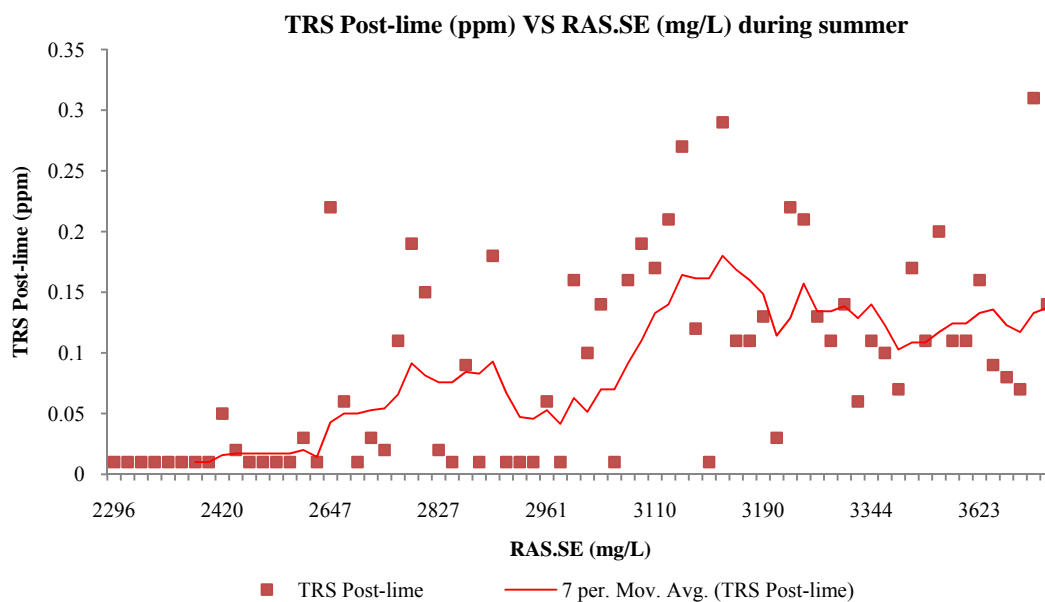


**TRS Prior-line (ppm) VS SWE.BL\*RAS (ft\*mg/L) during summer**

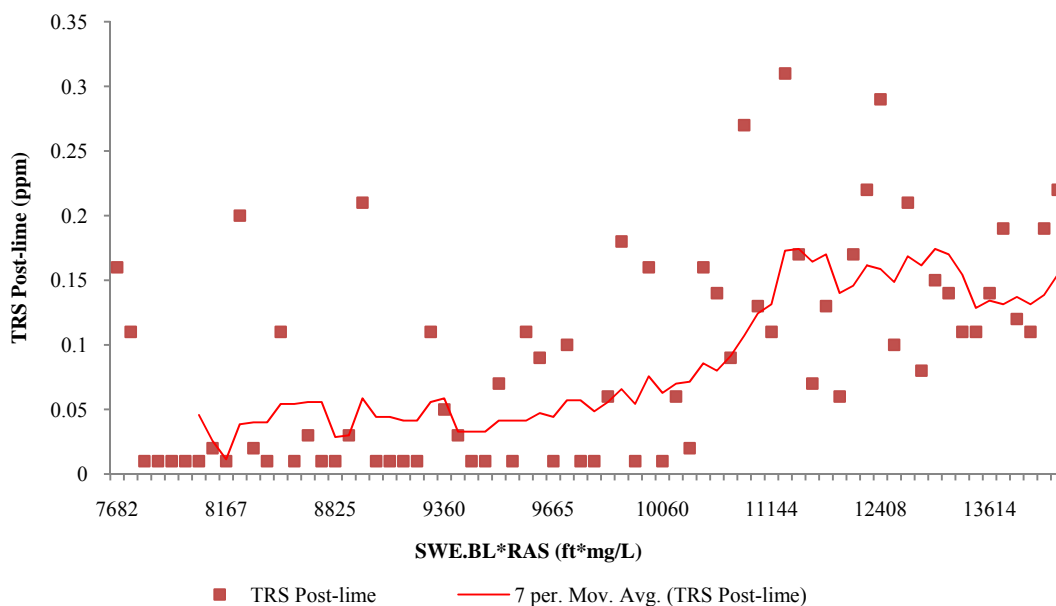




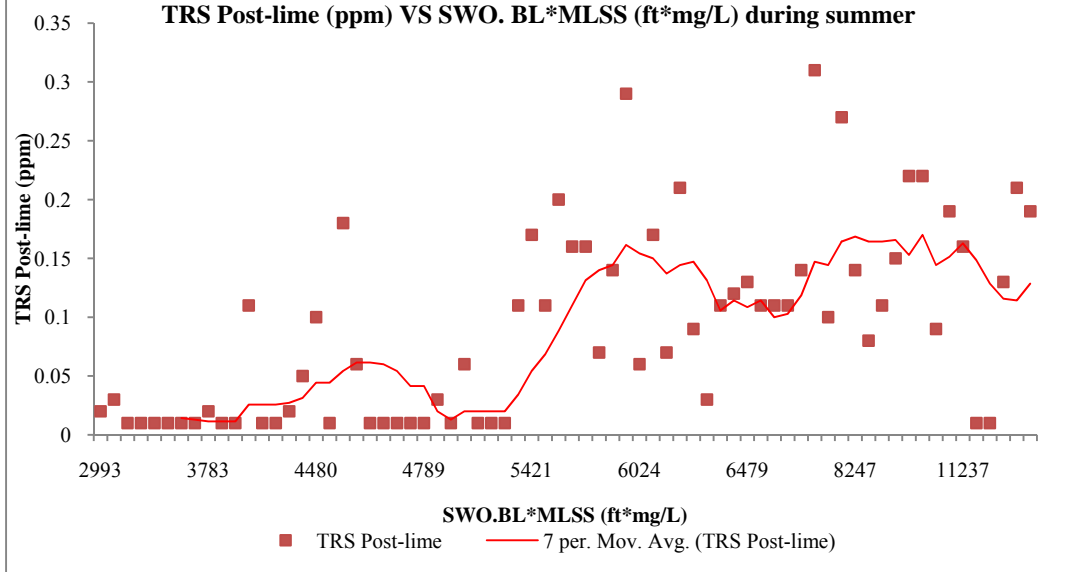


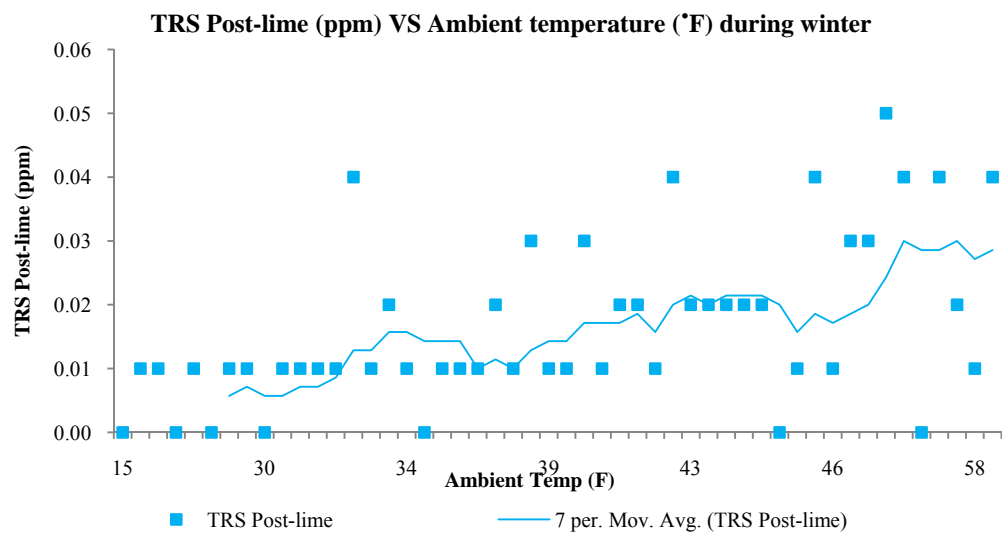
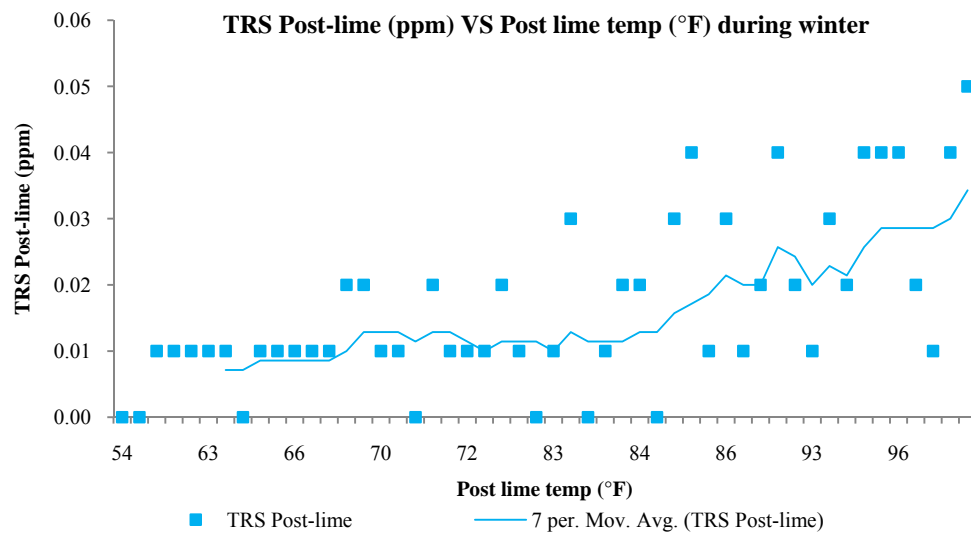


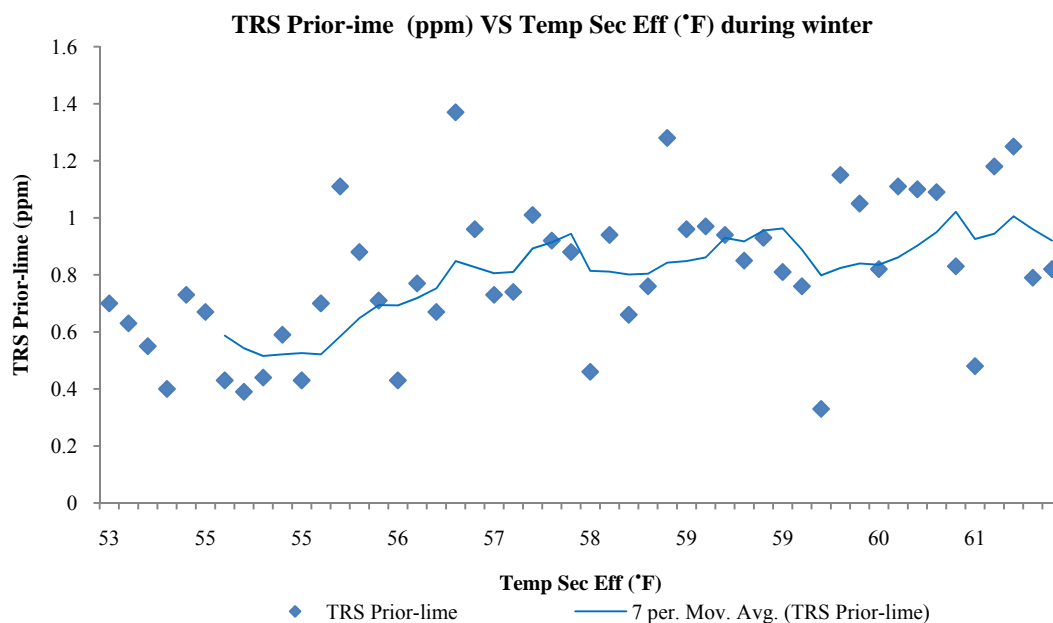
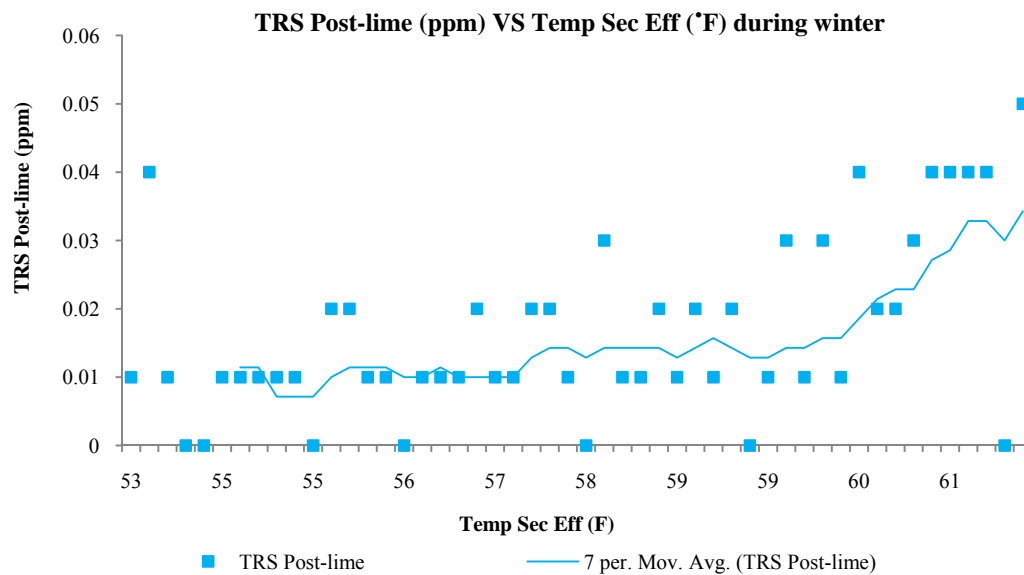
**TRS Post-lime (ppm) VS SWE.BL\*RAS (ft\*mg/L) during summer**

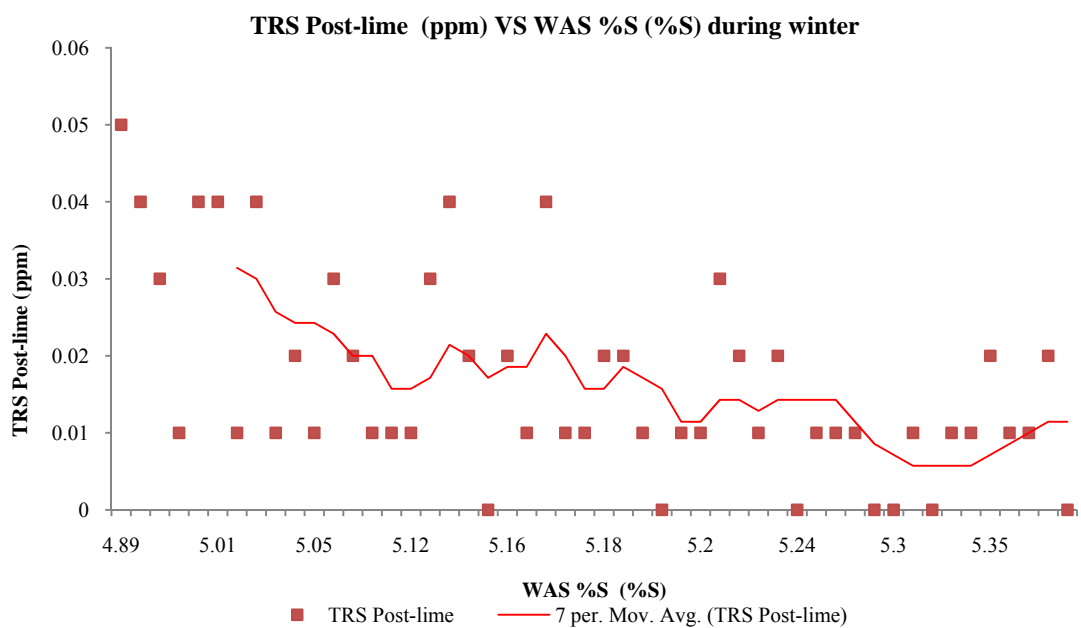
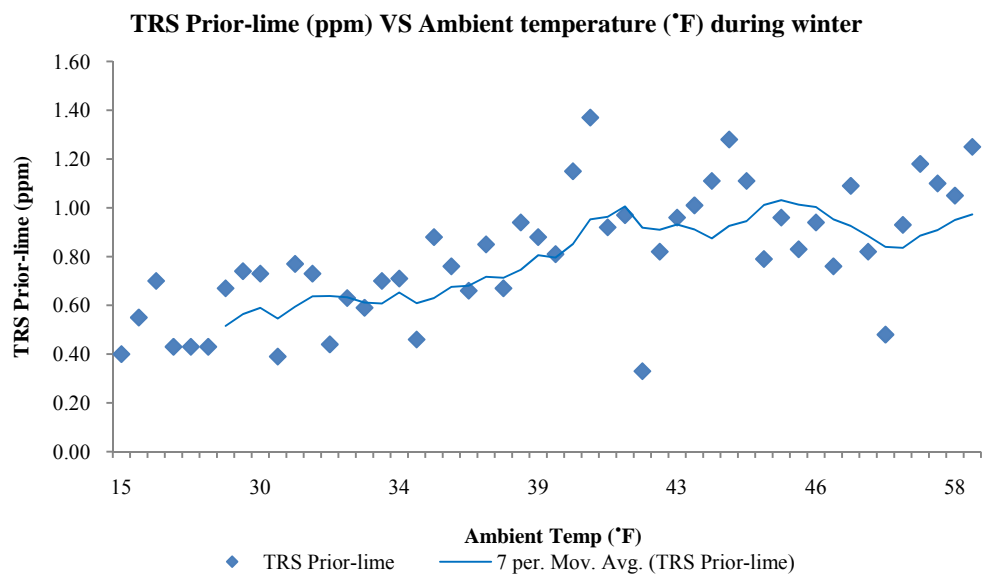


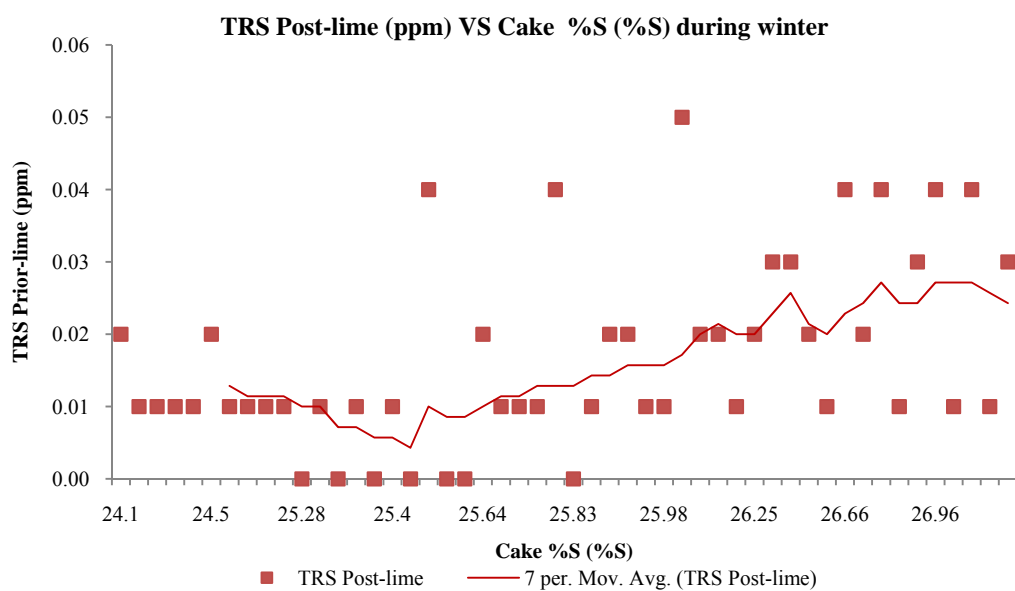
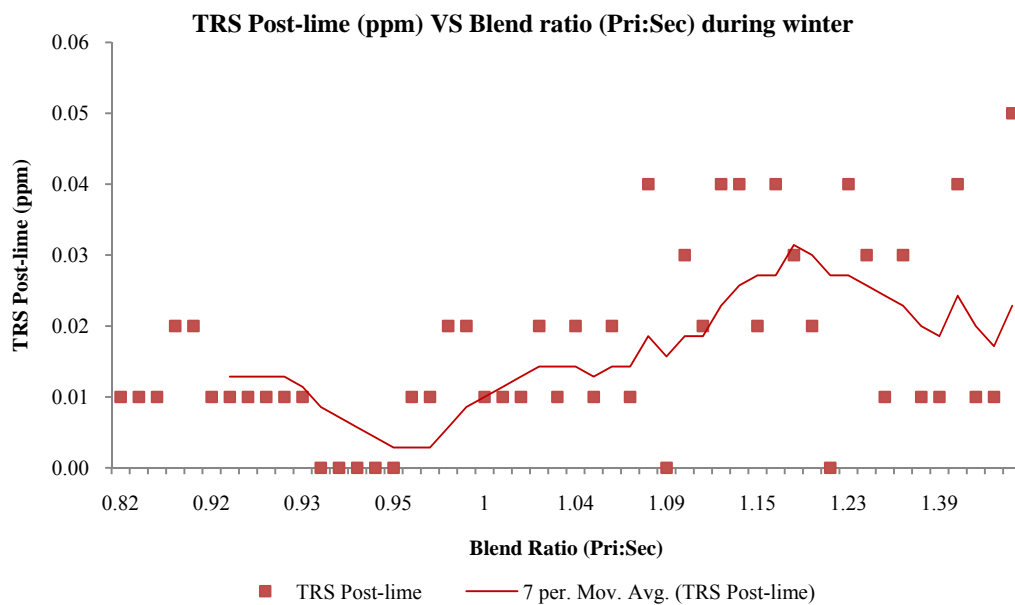
**TRS Post-lime (ppm) VS SWO. BL\*MLSS (ft\*mg/L) during summer**











## Appendix B

### Candidate Predictive Models

#### Candidate predictive models for TRS Prior-lime during annual period

Input variables	Coefficient	Std. Error	p-value	SS		
Constant term	3.78029752	1.4408797	0.0096248	202.00839	Residual df	146
GT FW	-0.0000018	4.6E-07	0.0001561	2.2398329	Multiple R-squared	0.4435422
DAF FW	-0.00000027	3.8E-07	0.4866484	4.8521428	Std. Dev. estimate	0.4029367
WAS %S	-0.36707926	0.1269891	0.004432	2.9992158	Residual SS	23.704268
S.E*RAS S.E	-0.00006307	2.099E-05	0.0031292	0.2927136	AIC	835.902
Temp(F)	0.01676291	0.0156893	0.2870912	1.4338033		
S.E. BL*MLSS S.E	0.00012443	3.282E-05	0.0002184	2.3982768		
(DT = 1)	-1.58757186	1.1441728	0.167395	4.1339955		
(Temp*DT = 1)	0.03034921	0.0165762	0.0691559	0.5442494		

#### Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
23.70498814	0.391069568	0.002094211

#### Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
14.9141416	0.3786888	0.025865828

#### Candidate predictive models for TRS Prior-lime during winter period

Input variables	Coefficient	Std. Error	p-value	SS		
Constant term	-0.34992853	0.3481922	0.3206583	32.256512	Residual df	42
SE.BL	0.13709851	0.0776971	0.0849154	0.1855095	Multiple R-squared	0.6637187
SWE.BL	0.15937284	0.0997161	0.117481	0.1424843	Std. Dev. estimate	0.1602341
RAS.SE	-0.00025692	7.184E-05	0.000894	0.1179105	Residual SS	1.0783491
RAS.S.W.E	0.00010569	4.945E-05	0.038432	0.0696217	AIC	17.755
RAS.SWO	-0.00005363	3.362E-05	0.1181219	0.6136962		
Amb Temp	0.01550165	0.0028505	2.54E-06	0.9606406		
SWO.BL*MLSS	0.00001072	8.75E-06	0.227713	0.038476		

Total sum of squared errors	RMS Error	Average Error
1.078349604	0.146857046	-8.85967E-05



### Candidate predictive models for TRS Prior-lime during summer period

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	4.24493027	2.25407672	0.06443998	119.7866516
GT %S	0.34499928	0.1557508	0.03049928	1.04515743
DAF TS	-0.00000629	0.00000249	0.01413679	0.57240433
Cake %S	-0.27148885	0.05699603	0.00001212	3.94246984
S.E.BL	-0.18433464	0.12166237	0.13490374	0.53699291
S.W.O.BL	0.05794982	0.06435493	0.37141037	2.28315639
S.W.E.BL	0.58860487	0.18309699	0.00208868	3.01883602
RAS.S.W.O	0.00043885	0.00015557	0.00645189	1.32945597
Temp(F)	0.02846573	0.01752483	0.10946672	0.3448959

Residual df	61
Multiple R-squared	0.6211376
Std. Dev. estimate	0.3615561
Residual SS	7.9740901
AIC	161.334

Total sum of squared errors	RMS Error	Average Error
7.974149236	0.337515063	0.000896779

### Candidate predictive models for TRS Post-lime during annual period

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.1366453	0.0959672	0.1565875	0.5896361
TRS Pri	0.0642399	0.0090024	0	0.3409052
WAS %S	-0.0143538	0.0158713	0.3672595	0.0308247
Blend(Pri:Sec)	0.0238281	0.0156281	0.1294674	0.0112432
S.E. BL*MLSS S.E	1.243E-05	2.79E-06	1.656E-05	0.0354587
Predicted Post lime temp	0.0011477	0.0004802	0.0181056	0.0740025
(DT = 1)	-0.0185868	0.0144752	0.201132	0.0039986

Residual df	148
Multiple R-squared	0.5803763
Std. Dev. estimate	0.0492464
Residual SS	0.358931
AIC	-253.378

#### Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
0.358931009	0.04812155	1.23389E-05

#### Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
0.202991198	0.044179616	-0.010980631

### Candidate predictive models for TRS Post-lime during winter period

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.2873996	0.0586773	1.409E-05	0.013778
TRS Pri	0.0155411	0.0073356	0.0399427	0.0011469
WAS %S	-0.059228	0.0104484	1.06E-06	0.0032587
S.W.O.BL	0.0054113	0.0034922	0.1285761	0.0001113
RAS.S.E	4.45E-06	3.61E-06	0.2243733	2.316E-05
RAS.S.W.E	-0.0000035	1.64E-06	0.0383005	0.0005438
Fitted Post lime temp	0.0002881	0.0001934	0.1435917	0.0001491

Residual df	43
Multiple R-squared	0.6443044
Std. Dev. estimate	0.0081967
Residual SS	0.002889
AIC	-280.342

Total sum of squared errors	RMS Error	Average Error
0.002889029	0.007601354	-3.71074E-05

### Candidate predictive models for TRS Post-lime during summer period

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.26951048	0.0971871	0.0072656	0.5742228
TRS Pri	0.05598473	0.0141994	0.0002023	0.1351165
GT TS	0.0000003	2.7E-07	0.2753301	0.0483704
RAS.S.E	0.00005985	1.723E-05	0.0009265	0.0401524
S.W.O.*RAS S.W.O	-0.00000947	2.88E-06	0.0016372	0.0037671
S.W.O. BL*MLSS S.W.O	0.00001602	4.68E-06	0.0010804	0.0331014

Residual df	64
Multiple R-squared	0.5904834
Std. Dev. estimate	0.0531315
Residual SS	0.1806694
AIC	-109.776

Total sum of squared errors	RMS Error	Average Error
0.180905401	0.0508366	-0.001831578

### Candidate predictive models for TRS N-containing during annual period

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-22.0163326	5.2132139	4.734E-05	7671.8359
Blend(Pri:Sec)	3.20669842	1.8477393	0.085246	161.21564
Pol_Sec	0.01565635	0.0019075	0	2413.6448
Pol_DW/Dry ton solids	0.67670655	0.2507101	0.0079635	315.74268
(DT = 1)	-3.50139284	1.175985	0.0035229	325.45297

Residual df	119
Multiple R-squared	0.424013
Std. Dev. estimate	6.0590572
Residual SS	4368.7485
AIC	1734.74

#### Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
4368.748524	5.935641746	-4.44754E-06

#### Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
2767.081113	5.8090353	1.228680016

### Candidate predictive models for TRS N-containing during winter period

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-9.55229282	2.4299016	0.0002467	460.76184
Post lime Temp	0.06700663	0.0174857	0.0003387	8.9302416
RAS.S.W.E	0.0002351	0.0003521	0.5072567	15.127789
RAS.S.W.O	0.00078726	0.0003302	0.0207413	22.687103
Pol_DW/Dry ton solids	0.20741624	0.107864	0.0598676	13.947087
S.W.O. BL*MLSS	-0.00040159	0.0001059	0.0003821	17.771725
S.W.O				
S.E. BL*MLSS S.E	0.00054344	0.0002138	0.0139949	14.392384

Residual df	53
Multiple R-squared	0.440203
Std. Dev. estimate	1.4926454
Residual SS	118.08348
AIC	298.284

Total sum of squared errors	RMS Error	Average Error
118.0834882	1.402875	-2.48804E-05

### Candidate predictive models for TRS N-containing during summer period

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-27.06305313	10.35056591	0.01123803	8600.490234
Post lime Temp	0.14050624	0.07313239	0.05937511	232.6981964
Pol_DW/DT solids	0.95264608	0.28560275	0.00145262	407.0157471
S.W.E*RAS	0.00088235	0.0005305	0.1013918	354.8959961
S.W.E. BL*MLSS	-0.00053535	0.0008497	0.53102034	2.54366541
S.E. BL*MLSS	0.00117053	0.00058802	0.05100794	178.9925232

Residual df	61
Multiple R-squared	0.299154857
Std. Dev. estimate	6.72092009
Residual SS	2755.416748
AIC	540.727

Total sum of squared errors	RMS Error	Average Error
2755.416946	6.412926482	5.49249E-05

## Appendix C

### Cross-Validation Outputs for Explanatory and Predictive Models

#### Validation of explanatory model of TRS prior-lime ( $Y_I$ ) during annual period:

RMSE of explanatory model of TRS prior-lime ( $Y_I$ ) during annual

Total sum of squared errors	RMS Error	Average Error
41.52302014	0.400400476	-2.8491E-05

Cross-validation output of explanatory model of TRS prior-lime ( $Y_I$ ) during annual

Validation of Sub Data	$SSE\left(\sum_{i=1}^n (\hat{\sigma}_i)^2\right)$
Sub Data #1	12.251103
Sub Data #2	10.425415
Sub Data #3	8.268828
Sub Data #4	6.5313182
Sub Data #5	6.556915
$RMSE\left[\frac{\sum_{i=1}^n (\hat{\sigma}_i)^2}{n_v}\right]^{\frac{1}{2}}$	0.412327

Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-1.6136272	0.72598064	0.02735128	258.3632507
GT %S	0.14644508	0.08846668	0.09940994	0.45670193
WAS %S	-0.34105667	0.10004644	0.00078778	6.73980379
Temp(F)	0.05063755	0.00635092	0	4.43061018
S.W.O. BL*MLSS S.W.O	0.00002098	0.00000847	0.01412733	0.45927268
Interaction(Temp(F),DT = 1)	0.00698714	0.00120843	0	4.9686327

Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-2.26721835	0.74412119	0.00262292	283.7602539
GT %S	0.21727198	0.09693497	0.02609211	1.17859936
WAS %S	-0.47557858	0.10251185	0.00000625	12.54754448
Temp(F)	0.0637417	0.00627967	0	7.59533358
S.W.O. BL*MLSS S.W.O	0.00003385	0.00000823	0.00005716	1.29712641
Interaction(Temp(F),DT = 1)	0.00861129	0.0012379	0	7.60559511

### Validation set of sub data #3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-2.19273925	0.7769196	0.00524386	283.8193359
GT %S	0.1408702	0.09562531	0.14226775	0.78024191
WAS %S	-0.37850687	0.10985524	0.00069321	10.76668167
Temp(F)	0.06202289	0.0064058	0	8.58029747
S.W.O. BL*MLSS S.W.O	0.00002577	0.00000879	0.00376367	0.73114675
Interaction(Temp(F),DT = 1)	0.00790401	0.0012604	0	6.51773691

### Validation set of sub data #4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-2.05498695	0.76959652	0.00820056	274.9548035
GT %S	0.13708495	0.09891956	0.16733675	0.41286451
WAS %S	-0.3637116	0.11028954	0.00115232	8.68247128
Temp(F)	0.05880776	0.00658696	0	7.48102474
S.W.O. BL*MLSS S.W.O	0.00002608	0.00000852	0.0025115	0.75766402
Interaction(Temp(F),DT = 1)	0.00754293	0.00130786	0	5.79257345

### Validation set of sub data #5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-1.88215089	0.7927559	0.01852934	274.2867737
GT %S	0.18540752	0.1004192	0.06631388	1.10388517
WAS %S	-0.38281474	0.10901658	0.00055002	10.42708206
Temp(F)	0.05498722	0.00703591	0	5.42697954
S.W.O. BL*MLSS S.W.O	0.00002614	0.00000926	0.00522917	0.68040162
Interaction(Temp(F),DT = 1)	0.00695778	0.00135184	0.00000058	4.61981249

### Validation of explanatory model of TRS prior-lime ( $Y_I$ ) during winter period:

RMSE of explanatory model of TRS prior-lime ( $Y_I$ ) during winter

Total sum of squared errors	RMS Error	Average Error
1.414413998	0.168191201	-4.8632E-06

Cross-validation output of explanatory model of TRS prior-lime ( $Y_I$ ) during winter

Validation of Sub Data	$SSE \left( \sum_{i=1}^n (\hat{\sigma}_i)^2 \right)$
Sub Data #1	0.3644875
Sub Data #2	0.4124564
Sub Data #3	0.4718112
Sub Data #4	0.1024846
Sub Data #5	0.2063163
RMSE $\left[ \frac{\sum_{i=1}^n (\hat{\sigma}_i)^2}{n_v} \right]^{\frac{1}{2}}$	0.176497

#### Validation set of sub data #1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.18278824	0.17557548	0.30459338	26.0822506
MLSS.S.W.O	0.00009329	0.00003483	0.01096709	0.38285899
Ambient Temp	0.0144603	0.00241948	0.00000063	1.02684629

#### Validation set of sub data #2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.33320093	0.17411365	0.06341894	28.25761032
MLSS.S.W.O	0.00010966	0.00003274	0.00187273	0.5890727
Ambient Temp	0.01685336	0.00308603	0.00000333	0.83179593

#### Validation set of sub data #3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.30440477	0.16736732	0.07704589	23.34783936
MLSS.S.W.O	0.00012037	0.00003114	0.00043298	0.49998513
Ambient Temp	0.0140189	0.00270923	0.00000817	0.70086968

#### Validation set of sub data #4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.31880328	0.18328953	0.09028897	26.16306114
MLSS.S.W.O	0.00011026	0.00003411	0.00258114	0.54152602
Ambient Temp	0.01608912	0.00289334	0.00000243	1.10104394

#### Validation set of sub data #5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.28039205	0.18534093	0.13881466	25.2969017
MLSS.S.W.O	0.00011925	0.00003746	0.0029478	0.5333761
Ambient Temp	0.0140128	0.0026769	0.00000677	0.90042037

#### Validation of explanatory model of TRS prior-lime ( $Y_I$ ) during summer period:

RMSE of explanatory model of TRS prior-lime ( $Y_I$ ) during summer

Total sum of squared errors	RMS Error	Average Error
8.606915865	0.350650739	-0.000520196

Cross-validation output of explanatory model of TRS prior-lime ( $Y_l$ ) during summer

Validation of Sub Data	$SSE\left(\sum_{i=1}^n (\hat{\sigma}_i)^2\right)$
Sub Data #1	4.1260648
Sub Data #2	1.5412812
Sub Data #3	0.9498785
Sub Data #4	2.1607459
Sub Data #5	1.5831385
$RMSE \left[ \frac{\sum_{i=1}^n (\hat{\sigma}_i)^2}{n_v} \right]^{\frac{1}{2}}$	0.384728

Validation set of sub data #1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	4.07743692	1.94832778	0.04146343	89.38504028
GT %S	0.26842105	0.14143457	0.06349725	0.62512231
DAF TS	-0.0000061	0.0000023	0.01064226	0.29687461
Cake %S	-0.21396916	0.05633853	0.00039623	2.81380486
S.W.E.BL	0.65783286	0.14763661	0.0000472	3.47191978
RAS.S.W.O	0.00052445	0.00012028	0.00006484	1.81935406

Validation set of sub data #2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	5.04576254	2.35608649	0.03712338	98.20903015
GT %S	0.21614027	0.16556452	0.19770664	1.66917765
DAF TS	-0.00000713	0.0000029	0.01764224	0.191725
Cake %S	-0.24067767	0.06244401	0.00033204	3.3757422
S.W.E.BL	0.59291601	0.17019744	0.00103768	3.38233209
RAS.S.W.O	0.00064084	0.00014621	0.0000601	2.73143172

Validation set of sub data #3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	4.22544765	2.5977447	0.11011249	97.75928497
GT %S	0.24293453	0.17708831	0.17624164	0.79603672
DAF TS	-0.0000049	0.00000321	0.13311797	1.09552646
Cake %S	-0.23080185	0.0708054	0.00201015	2.65591264
S.W.E.BL	0.6284048	0.17228112	0.00063117	3.79931736
RAS.S.W.O	0.00060447	0.00015615	0.00031493	2.31934333

Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	5.50553226	2.24062419	0.01751743	96.04921722
GT %S	0.13731353	0.15825543	0.38972023	0.7122215
DAF TS	-0.0000083	0.00000272	0.00359829	0.3311477
Cake %S	-0.25889418	0.06114365	0.00009809	2.8430779
S.W.E.BL	0.7385568	0.16910659	0.00006329	3.38752317
RAS.S.W.O	0.00071623	0.0001548	0.0000266	2.90575624

Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	5.76188707	2.3824172	0.01926773	97.8914566
GT %S	0.19348839	0.17812788	0.28258508	0.64679593
DAF TS	-0.00000532	0.00000307	0.08942114	0.59359938
Cake %S	-0.25242814	0.06192075	0.00016359	4.17119551
S.W.E.BL	0.545165	0.17498171	0.00303869	2.39170408
RAS.S.W.O	0.00050806	0.0001687	0.00406826	1.31987321

**Validation of explanatory model of TRS post-lime ( $Y_2$ ) during annual period:**

RMSE of explanatory model of TRS post-lime ( $Y_2$ ) during annual

Total sum of squared errors	RMS Error	Average Error
0.536649945	0.045519309	1.5313E-05

Cross-validation output of explanatory model of TRS post-lime ( $Y_2$ ) during annual

Validation of Sub Data	$SSE\left(\sum_{i=1}^n(\hat{\sigma}_i)^2\right)$
Sub Data #1	0.089403
Sub Data #2	0.131075
Sub Data #3	0.082616
Sub Data #4	0.161287
Sub Data #5	0.09885125
$RMSE\left[\frac{\sum_{i=1}^n(\hat{\sigma}_i)^2}{n_v}\right]^{\frac{1}{2}}$	0.046633

Validation set of sub data #1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.17249058	0.04005617	0.00002591	0.56347823
TRS Pri	0.05940421	0.0081301	0	0.27357507
Blend(Pri:Sec)	0.04141361	0.01080362	0.00016903	0.04600448
S.E. BL*MLSS S.E	0.00001127	0.00000249	0.00001029	0.03614472
Predicted Post lime temp	0.00058982	0.0004062	0.14804594	0.04762288
(DT = 1)	-0.01856329	0.01253622	0.140233	0.00491618



### Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.14034623	0.04019396	0.00058991	0.66243529
TRS Pri	0.07183505	0.00739948	0	0.4751215
Blend(Pri:Sec)	0.03204269	0.01082543	0.00344667	0.02459708
S.E. BL*MLSS S.E	0.00000948	0.00000231	0.00005901	0.02404776
Predicted Post lime temp	0.00037159	0.00040197	0.35637438	0.03996858
(DT = 1)	-0.02328327	0.01223534	0.05847597	0.00737496

### Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.17193806	0.04018134	0.00002897	0.71089232
TRS Pri	0.06862083	0.00749449	0	0.4581989
Blend(Pri:Sec)	0.03543651	0.01114014	0.00169922	0.03318158
S.E. BL*MLSS S.E	0.00001073	0.00000235	0.00000872	0.03970831
Predicted Post lime temp	0.00062234	0.0003991	0.12047368	0.05431617
(DT = 1)	-0.01969476	0.01236859	0.11287638	0.00571597

### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.18605813	0.03598293	0.00000052	0.64445651
TRS Pri	0.06051498	0.00690788	0	0.37281477
Blend(Pri:Sec)	0.0332864	0.01034553	0.00150751	0.03339111
S.E. BL*MLSS S.E	0.00001032	0.00000218	0.00000396	0.03520977
Predicted Post lime temp	0.00082622	0.0003587	0.02228056	0.04415446
(DT = 1)	-0.00800728	0.01105537	0.46973065	0.00098626

### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.21242036	0.03891534	0.0000001	0.67951494
TRS Pri	0.06475236	0.00751115	0	0.4407638
Blend(Pri:Sec)	0.027857	0.01138921	0.01530887	0.0345581
S.E. BL*MLSS S.E	0.00001136	0.00000232	0.00000191	0.03632364
Predicted Post lime temp	0.00110495	0.00040492	0.00692021	0.06784793
(DT = 1)	-0.00824984	0.01202761	0.49356091	0.00103447

### Validation of explanatory model of TRS post-lime ( $Y_2$ ) during winter period:

RMSE of explanatory model of TRS post-lime ( $Y_2$ ) during winter

Total sum of squared errors	RMS Error	Average Error
0.003470104	0.008330791	3.36304E-07

Cross-validation output of explanatory model of TRS post-lime ( $Y_2$ ) during winter

Validation of Sub Data	$SSE\left(\sum_{i=1}^n(\hat{\sigma}_i)^2\right)$
Sub Data #1	0.001345
Sub Data #2	0.0005749
Sub Data #3	0.0007973
Sub Data #4	0.0006215
Sub Data #5	0.0006698
RMSE $\left[\frac{\sum_{i=1}^n(\hat{\sigma}_i)^2}{n_v}\right]^{\frac{1}{2}}$	0.008954

Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.35590735	0.06511521	0.0000036	0.01089
TRS Pri	0.01316214	0.00736791	0.08245696	0.00103313
WAS %S	-0.07263096	0.01170774	0.00000037	0.00372412
Predicted Post lime temp	0.00033971	0.00018251	0.07087398	0.00022411

Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.28483838	0.06856921	0.00019186	0.0133225
TRS Pri	0.01904802	0.00824111	0.02665034	0.00075858
WAS %S	-0.05898215	0.01236141	0.00003014	0.00238985
Predicted Post lime temp	0.00027905	0.00023751	0.24776015	0.00011185

Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.23391238	0.06808797	0.0015066	0.009
TRS Pri	0.01501349	0.00782772	0.06307148	0.00090056
WAS %S	-0.04970548	0.01207143	0.00021358	0.00214793
Predicted Post lime temp	0.0003561	0.00020584	0.09220044	0.00022654

Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.2549901	0.07362349	0.00139398	0.0119025
TRS Pri	0.01581494	0.00720472	0.03468985	0.00113892
WAS %S	-0.05391682	0.01333292	0.00026518	0.00213385
Predicted Post lime temp	0.00035929	0.00019929	0.0797931	0.00025875

Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.27706403	0.07007175	0.00034472	0.01024
TRS Pri	0.0141349	0.00745182	0.06589467	0.000645
WAS %S	-0.05704153	0.01249792	0.00005645	0.00271819
Predicted Post lime temp	0.00030121	0.0001985	0.13789558	0.00018015

**Validation of explanatory model of TRS post-lime ( $Y_2$ ) during summer period:**

RMSE of explanatory model of TRS post-lime ( $Y_2$ ) during summer

Total sum of squared errors	RMS Error	Average Error
0.1992954	0.053358009	-0.00146334

Cross-validation output of explanatory model of TRS post-lime ( $Y_2$ ) during summer

Validation of Sub Data	$SSE\left(\sum_{i=1}^n(\hat{\sigma}_i)^2\right)$
Sub Data #1	0.0447402
Sub Data #2	0.0235883
Sub Data #3	0.0537814
Sub Data #4	0.0354636
Sub Data #5	0.089821
$RMSE\left[\frac{\sum_{i=1}^n(\hat{\sigma}_i)^2}{n_v}\right]^{\frac{1}{2}}$	0.059449

Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.44319564	0.10259682	0.0000722	0.40800714
TRS Pri	0.03106434	0.01930756	0.11380704	0.08748905
GT TS	0.00000064	0.00000029	0.03396073	0.036032
RAS.S.E	0.00005766	0.00001948	0.00465641	0.02871277
S.W.E. BL*MLSS S.W.E	0.0000163	0.00000605	0.00955092	0.02238384

Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.38431886	0.11294568	0.00130585	0.49971607
TRS Pri	0.03852218	0.01850338	0.04238814	0.11791269
GT TS	0.00000042	0.00000031	0.187894	0.02401834
RAS.S.E	0.00006714	0.00002007	0.00155109	0.03720866
S.W.E. BL*MLSS S.W.E	0.00001455	0.00000637	0.02654988	0.018326

### Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.52127182	0.10180665	0.00000466	0.45359996
TRS Pri	0.03666688	0.01643579	0.03011162	0.08889505
GT TS	0.00000092	0.00000029	0.00260831	0.06315182
RAS.S.E	0.00004924	0.00001957	0.0150793	0.0215973
S.W.E. BL*MLSS S.W.E	0.0000152	0.00000564	0.00948295	0.02147508

### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.3784458	0.10768083	0.00093356	0.51111603
TRS Pri	0.04646393	0.01923585	0.01933589	0.1165152
GT TS	0.00000042	0.0000003	0.15914239	0.02468498
RAS.S.E	0.00005988	0.00001946	0.00335647	0.03185174
S.W.E. BL*MLSS S.W.E	0.0000151	0.00000796	0.06344637	0.01182423

### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.55654055	0.09961392	0.00000086	0.42874998
TRS Pri	0.05663443	0.01600486	0.00086791	0.13205446
GT TS	0.00000081	0.00000028	0.00567107	0.05321962
RAS.S.E	0.00007967	0.00001958	0.00016444	0.04230213
S.W.E. BL*MLSS S.W.E	0.00000756	0.0000052	0.15213044	0.00501677

### Validation of explanatory model of N-containing ( $Y_3$ ) during annual period:

RMSE of explanatory model of N-containing ( $Y_3$ ) during annual

Total sum of squared errors	RMS Error	Average Error
6883.492491	5.780571952	-1.22428E-07

Cross-validation output of explanatory model of N-containing ( $Y_3$ ) during annual

Validation of Sub Data	$SSE\left(\sum_{i=1}^n(\hat{\sigma}_i)^2\right)$
Sub Data #1	1652.7688
Sub Data #2	1912.2285
Sub Data #3	1075.4709
Sub Data #4	1067.507
Sub Data #5	1781.5208
$RMSE\left[\frac{\sum_{i=1}^n(\hat{\sigma}_i)^2}{n_v}\right]^{\frac{1}{2}}$	6.02966

### Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-19.622757	4.86293697	0.00008452	10303.92773
Post lime Temp	0.07458825	0.0335843	0.02776915	1283.177368
Pol_Sec	0.01291948	0.00172118	0	1796.430664
Pol_DW/Dry ton solids	0.57217956	0.1876936	0.00269452	404.8916626
(DT = 1)	-2.5766778	1.32084501	0.05284127	126.22789

### Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-16.2872715	4.86356354	0.001012	9745.15625
Post lime Temp	0.06765807	0.03113895	0.03126813	1578.484619
Pol_Sec	0.01537322	0.0017107	0	2476.460938
Pol_DW/Dry ton solids	0.10818642	0.22018027	0.62384957	28.11370277
(DT = 1)	-3.28561211	1.30156195	0.01256507	205.2554169

### Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-23.9482098	4.87384892	0.00000219	9862.475586
Post lime Temp	0.08358901	0.03405543	0.01517948	1958.323853
Pol_Sec	0.01589398	0.0019803	0	2277.056396
Pol_DW/Dry ton solids	0.57066441	0.20060802	0.00502713	408.4032288
(DT = 1)	-2.46164536	1.38035035	0.07642537	116.5712051

### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-20.56046677	5.00239992	0.00006302	9510.042969
Post lime Temp	0.07813048	0.03458369	0.02522304	1480.271973
Pol_Sec	0.01387061	0.00179438	0	2081.130127
Pol_DW/Dry ton solids	0.5049749	0.19583386	0.0108214	332.0267334
(DT = 1)	-2.63143563	1.40931237	0.0637063	126.9289475

### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-23.68843651	4.86697245	0.00000269	9360.46875
Post lime Temp	0.08061138	0.03301869	0.01572089	1255.030029
Pol_Sec	0.01436236	0.00166382	0	2333.821777
Pol_DW/Dry ton solids	0.65051007	0.2057998	0.00188162	403.297821
(DT = 1)	-1.88220775	1.31691253	0.15487963	65.92568207

### Validation of explanatory model of N-containing ( $Y_3$ ) during winter period:

RMSE of explanatory model of N-containing ( $Y_3$ ) during winter

Total sum of squared errors	RMS Error	Average Error
153.283602	1.598351244	8.41104E-06

Cross-validation output of explanatory model of N-containing ( $Y_3$ ) during winter

Validation of Sub Data	$SSE\left(\sum_{i=1}^n(\hat{\sigma}_i)^2\right)$
Sub Data #1	26.692416
Sub Data #2	86.005759
Sub Data #3	27.874329
Sub Data #4	24.72281
Sub Data #5	12.221732
$RMSE\left[\frac{\sum_{i=1}^n(\hat{\sigma}_i)^2}{n_v}\right]^{\frac{1}{2}}$	1.72006

Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-8.67798805	2.77228069	0.00309896	366.9155273
Post lime Temp	0.04666851	0.02160704	0.03627175	6.41247129
RAS.S.E	0.00171669	0.00048464	0.00095297	52.78307724
Pol_DW/Dry ton solids	0.12164825	0.15839763	0.44659367	1.77487302

Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-7.97980642	1.90004337	0.00012812	316.1106873
Post lime Temp	0.05575303	0.01547796	0.00079839	14.02059746
RAS.S.E	0.00102902	0.00030925	0.00177737	28.93017006
Pol_DW/Dry ton solids	0.20569523	0.08942139	0.02623189	8.66257

Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-7.64728355	2.64262199	0.00590016	351.4877319
Post lime Temp	0.05059863	0.02087076	0.01951065	13.33826637
RAS.S.E	0.00108753	0.00045359	0.02081246	23.21550751
Pol_DW/Dry ton solids	0.20893879	0.12382349	0.09860579	8.15744972

#### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-5.83759308	2.9796977	0.05645397	385.2200317
Post lime Temp	0.03368986	0.02094302	0.11484692	1.90995574
RAS.S.E	0.00102817	0.00048245	0.03870164	18.60653687
Pol_DW/Dry ton solids	0.19057095	0.12236442	0.12653881	7.20480633

#### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-8.17711163	2.99088979	0.00898032	427.929657
Post lime Temp	0.0447272	0.02302514	0.05848901	3.74225926
RAS.S.E	0.00133645	0.00045322	0.00509161	38.37176514
Pol_DW/Dry ton solids	0.21644875	0.12876609	0.09986003	9.10062504

#### Validation of explanatory model of N-containing ( $Y_3$ ) during summer period:

RMSE of explanatory model of N-containing ( $Y_3$ ) during summer

Total sum of squared errors	RMS Error	Average Error
2663.941954	6.305579091	9.99169E-06

Cross-validation output of explanatory model of N-containing ( $Y_3$ ) during summer

Validation of Sub Data	$SSE \left( \sum_{i=1}^n (\hat{\sigma}_i)^2 \right)$
Sub Data #1	439.50598
Sub Data #2	729.74109
Sub Data #3	890.94924
Sub Data #4	325.55432
Sub Data #5	1264.2154
$RMSE \left[ \frac{\sum_{i=1}^n (\hat{\sigma}_i)^2}{n_v} \right]^{1/2}$	7.38086

#### Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-27.06731415	10.05653286	0.00970362	7884.617188
Post lime Temp	0.16368711	0.07886629	0.04321032	254.7974091
S.W.O.BL	3.33703351	1.31698322	0.0145297	672.5153809
Pol_DW/Dry ton solids	0.65633905	0.31046861	0.03962643	168.1104431
S.E. BL*MLSS S.E	0.00090351	0.00051596	0.08618593	142.2775574

### Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-18.59605789	10.53304672	0.08384097	6439.206055
Post lime Temp	0.07114482	0.08645198	0.41461176	64.78304291
S.W.O.BL	3.41354465	1.25588226	0.00910953	603.7216187
Pol_DW/Dry ton solids	0.76503158	0.27967173	0.00870369	294.0295715
S.E. BL*MLSS S.E	0.00094656	0.0005523	0.09300707	123.7935715

### Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-27.3769207	9.38144398	0.00530276	6929.20166
Post lime Temp	0.15548363	0.07412478	0.04112214	209.8670044
S.W.O.BL	2.55813932	1.00905693	0.0144811	452.0293579
Pol_DW/Dry ton solids	0.86416513	0.25813276	0.00157206	408.4367981
S.E. BL*MLSS S.E	0.00093856	0.00048112	0.05681286	139.97052

### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-26.46740913	10.66418266	0.01662069	7103.172852
Post lime Temp	0.14385492	0.08223621	0.0866329	228.150528
S.W.O.BL	2.60166717	1.35376203	0.06057769	431.5637512
Pol_DW/Dry ton solids	0.88973743	0.35106117	0.01457956	299.8216858
S.E. BL*MLSS S.E	0.00109781	0.00054213	0.04844886	201.2860565

### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-16.06728363	10.06443787	0.11682028	6110.551758
Post lime Temp	0.13237143	0.07266811	0.07462461	172.5416107
S.W.O.BL	0.30035964	1.16687644	0.79794419	33.03540039
Pol_DW/Dry ton solids	0.27445114	0.32346112	0.40029529	19.70939636
S.E. BL*MLSS S.E	0.00198342	0.00057752	0.00121768	443.5253296

### Validation of predictive model of TRS prior-lime ( $Y_l$ ) during annual period:

RMSE of predictive model of TRS prior-lime ( $Y_l$ ) during annual

Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
24.04575862	0.393870444	-0.002239371

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
14.60075105	0.374689	0.01894817



### Validation of predictive model of TRS prior-lime ( $Y_I$ ) during winter period:

RMSE of predictive model of TRS prior-lime ( $Y_I$ ) during winter

Total sum of squared errors	RMS Error	Average Error
1.148790121	0.151577711	-5.03604E-05

Cross-validation output of predictive model of TRS prior-lime ( $Y_I$ ) during winter

Validation of Sub Data	SSE $\left(\sum_{i=1}^n (\hat{\sigma}_i)^2\right)$
Sub Data #1	0.3237439
Sub Data #2	0.353964
Sub Data #3	0.4505517
Sub Data #4	0.2728825
Sub Data #5	0.3005489
RMSE $\left[\frac{\sum_{i=1}^n (\hat{\sigma}_i)^2}{n_v}\right]^{\frac{1}{2}}$	0.184483

Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.05874452	0.26596686	0.82651347	26.0822506
S.E.BL	0.39635867	0.17752446	0.03226065	0.13340688
RAS.S.E	-0.00010847	0.00005187	0.04404362	0.20917533
Ambient Temp	0.01460141	0.00237059	0.00000054	1.04730177
S.W.O. BL*MLSS S.W.O	0.00002326	0.00000798	0.00626148	0.14711303
S.E. BL*MLSS S.E	-0.00010965	0.00006458	0.09867546	0.07318256

Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.024817	0.28173894	0.93032581	28.25761032
S.E.BL	0.36226723	0.17402887	0.04497369	0.10342547
RAS.S.E	-0.0001413	0.00005133	0.00940696	0.25676274
Ambient Temp	0.01652887	0.00306331	0.00000526	0.92458194
S.W.O. BL*MLSS S.W.O	0.00002676	0.00000739	0.00094172	0.26472694
S.E. BL*MLSS S.E	-0.00009881	0.00006305	0.12629554	0.06086689

Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.3450534	0.28262055	0.23052068	23.34783936
S.E.BL	0.03905203	0.19073832	0.83899492	0.07386089
RAS.S.E	-0.00015266	0.00004979	0.00423589	0.19477518
Ambient Temp	0.01340984	0.00264309	0.00001379	0.71560788
S.W.O. BL*MLSS S.W.O	0.00002741	0.00000794	0.001501	0.37596649
S.E. BL*MLSS S.E	-0.00000117	0.00006666	0.9861173	0.00000731

#### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.05228124	0.28810164	0.85707772	26.16306114
S.E.BL	0.51677704	0.20745333	0.01777985	0.2161732
RAS.S.E	-0.00020528	0.00006691	0.00421264	0.25451833
Ambient Temp	0.0139587	0.00285562	0.00002397	1.02538681
S.W.O. BL*MLSS S.W.O	0.00003868	0.00000965	0.00031553	0.33422691
S.E. BL*MLSS S.E	-0.00015753	0.00007244	0.03670711	0.1379492

#### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.06478707	0.27497694	0.81514949	25.2969017
S.E.BL	0.39898172	0.18055588	0.03395838	0.27613273
RAS.S.E	-0.00013506	0.00005194	0.01368234	0.1584762
Ambient Temp	0.01389134	0.00244962	0.00000231	0.94445848
S.W.O. BL*MLSS S.W.O	0.00003553	0.00000931	0.00054928	0.28743148
S.E. BL*MLSS S.E	-0.00011853	0.00006263	0.06697404	0.09368367

#### Validation of predictive model of TRS prior-lime ( $Y_I$ ) during summer period:

RMSE of predictive model of TRS prior-lime ( $Y_I$ ) during summer

Total sum of squared errors	RMS Error	Average Error
8.376898669	0.345933492	0.000806436

#### Cross-validation output of predictive model of TRS prior-lime ( $Y_I$ ) during summer

Validation of Sub Data	SSE $\left(\sum_{i=1}^n (\hat{\sigma}_i)^2\right)$
Sub Data #1	4.3884027
Sub Data #2	1.4793374
Sub Data #3	0.8173868
Sub Data #4	2.1343542
Sub Data #5	1.4011608
RMSE $\left[\frac{\sum_{i=1}^n (\hat{\sigma}_i)^2}{n_v}\right]^{\frac{1}{2}}$	0.382111

#### Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	1.99286532	2.15535164	0.35969868	89.38504028
GT %S	0.32071081	0.13971643	0.02602754	0.62512231
DAF TS	-0.00000592	0.00000223	0.01071688	0.29687461
Cake %S	-0.22334045	0.05488417	0.00017108	2.81380486
S.W.E.BL	0.63002801	0.14397073	0.00006313	3.47191978
RAS.S.W.O	0.00047777	0.00011903	0.00020424	1.81935406
Temp(F)	0.03264007	0.01618798	0.04926579	0.3666088

### Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	3.80242658	2.61268878	0.15194422	98.20903015
GT %S	0.2439815	0.1672018	0.15089032	1.66917765
DAF TS	-0.00000701	0.0000029	0.01936081	0.191725
Cake %S	-0.25533095	0.06375225	0.00021004	3.3757422
S.W.E.BL	0.53017664	0.17932306	0.00477476	3.38233209
RAS.S.W.O	0.00063097	0.00014621	0.000077	2.73143172
Temp(F)	0.02378284	0.02177845	0.28015789	0.16890419

### Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	3.45153379	2.85146546	0.23191586	97.75928497
GT %S	0.27435952	0.18401551	0.14238228	0.79603672
DAF TS	-0.00000499	0.00000323	0.12914681	1.09552646
Cake %S	-0.24057478	0.07264329	0.0017466	2.65591264
S.W.E.BL	0.61030275	0.17527737	0.00105691	3.79931736
RAS.S.W.O	0.00058842	0.00015878	0.0005361	2.31934333
Temp(F)	0.01461799	0.02160706	0.50188059	0.07161708

### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	4.53050661	2.46948028	0.07263999	96.04921722
GT %S	0.17034969	0.16224793	0.29889911	0.7122215
DAF TS	-0.00000816	0.00000273	0.00430942	0.3311477
Cake %S	-0.26696989	0.0618054	0.00007602	2.8430779
S.W.E.BL	0.68980116	0.17699552	0.00029539	3.38752317
RAS.S.W.O	0.00069723	0.00015627	0.00004758	2.90575624
Temp(F)	0.01715119	0.01816996	0.34983608	0.12120697

### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	4.46453381	2.52744675	0.08355424	97.8914566
GT %S	0.271346	0.18456767	0.14790787	0.64679593
DAF TS	-0.00000571	0.00000305	0.06707462	0.59359938
Cake %S	-0.27932385	0.06412439	0.00006745	4.17119551
S.W.E.BL	0.46213657	0.18271741	0.01469889	2.39170408
RAS.S.W.O	0.00049042	0.00016744	0.00514943	1.31987321
Temp(F)	0.02959743	0.02074874	0.16007392	0.29010886

### Validation of predictive model of TRS post-lime ( $Y_2$ ) during annual period:

#### RMSE of predictive model of TRS post-lime ( $Y_2$ ) during annual

Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
0.360914657	0.04825434	1.75401E-05

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
0.202561964	0.044132881	-0.01066482

### Validation of predictive model of TRS post-lime ( $Y_2$ ) during winter period:

#### RMSE of predictive model of TRS post-lime ( $Y_2$ ) during winter

Total sum of squared errors	RMS Error	Average Error
0.002991057	0.007734413	5.7367E-06

### Cross-validation output of predictive model of TRS post-lime ( $Y_2$ ) during winter

Validation of Sub Data	$SSE\left(\sum_{i=1}^n(\hat{\sigma}_i)^2\right)$
Sub Data #1	0.00128333
Sub Data #2	0.0005978
Sub Data #3	0.0006556
Sub Data #4	0.0004325
Sub Data #5	0.0008373
$RMSE\left[\frac{\sum_{i=1}^n(\hat{\sigma}_i)^2}{n_v}\right]^{\frac{1}{2}}$	0.013933

### Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.32699418	0.06337424	0.00001066	0.01089
TRS Pri	0.00738912	0.0073573	0.32231644	0.00103313
WAS %S	-0.07040627	0.0111922	0.00000036	0.00372412
S.W.O.BL	0.00658568	0.00301163	0.03573798	0.00012092
RAS.S.W.E	-0.00000107	0.00000164	0.51984566	0.00009679
Predicted Post lime temp	0.0004827	0.00019315	0.01744486	0.00036236

### Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.29091328	0.06500649	0.00008141	0.0133225
TRS Pri	0.01590706	0.00790952	0.05229409	0.00075858
WAS %S	-0.05938454	0.01176874	0.00001498	0.00238985
S.W.O.BL	0.00866769	0.00338111	0.01495392	0.00015024
RAS.S.W.E	-0.00000316	0.00000172	0.07443172	0.00032639
Predicted Post lime temp	0.00028796	0.00023487	0.22861192	0.00010807

### Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.25665611	0.0668454	0.00051161	0.009
TRS Pri	0.0131534	0.00784007	0.10257355	0.00090056
WAS %S	-0.05159394	0.01195453	0.00012982	0.00214793
S.W.O.BL	0.0069372	0.00345101	0.0523977	0.00002944
RAS.S.W.E	-0.00000301	0.00000168	0.08176975	0.00038913
Predicted Post lime temp	0.00028863	0.000217	0.19234742	0.00012528

### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.27095062	0.07413749	0.00086046	0.0119025
TRS Pri	0.01276927	0.00745891	0.09601482	0.00113892
WAS %S	-0.05699927	0.01311402	0.00011872	0.00213385
S.W.O.BL	0.00666221	0.00342684	0.06019244	0.00003138
RAS.S.W.E	-0.00000221	0.00000171	0.20570485	0.00029115
Predicted Post lime temp	0.00038243	0.00022267	0.09498502	0.00022371

### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.29385769	0.0648398	0.00006886	0.01024
TRS Pri	0.00870764	0.00720426	0.23512533	0.000645
WAS %S	-0.05923011	0.01162259	0.00001289	0.00271819
S.W.O.BL	0.0099088	0.0033464	0.00555675	0.00013102
RAS.S.W.E	-0.00000344	0.00000158	0.03631308	0.00048337
Predicted Post lime temp	0.00029762	0.00019941	0.14478281	0.0001465

### Validation of predictive model of TRS post-lime ( $Y_2$ ) during summer period:

RMSE of predictive model of TRS post-lime ( $Y_2$ ) during summer

Total sum of squared errors	RMS Error	Average Error
0.18408697	0.0512817	-2.54089E-05

Cross-validation output of predictive model of TRS post-lime ( $Y_2$ ) during summer

Validation of Sub Data	SSE $\left(\sum_{i=1}^n (\hat{\sigma}_i)^2\right)$
Sub Data #1	0.0423658
Sub Data #2	0.0190541
Sub Data #3	0.0369723
Sub Data #4	0.0607363
Sub Data #5	0.0830447
RMSE $\left[\frac{\sum_{i=1}^n (\hat{\sigma}_i)^2}{n_v}\right]^{1/2}$	0.058819

### Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.15217446	0.05064283	0.00411292	0.40800714
TRS Pri	0.0445551	0.01701416	0.0115932	0.08748905
RAS.S.E	0.00005832	0.00001774	0.00183213	0.05219924
S.W.O*RAS S.W.O	-0.00001126	0.00000284	0.00022691	0.00592907
S.W.O. BL*MLSS S.W.O	0.00001994	0.00000475	0.00010888	0.04781482

### Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.16354738	0.0551777	0.00460674	0.49971607
TRS Pri	0.05352212	0.01643313	0.00200534	0.11791269
RAS.S.E	0.00006432	0.00001864	0.00113201	0.05483054
S.W.O*RAS S.W.O	-0.00001097	0.00000401	0.00858096	0.00223765
S.W.O. BL*MLSS S.W.O	0.00001758	0.00000521	0.00143093	0.03675106

### Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.15509735	0.05388933	0.0058314	0.45359996
TRS Pri	0.04675478	0.01476082	0.00259685	0.08889505
RAS.S.E	0.00005998	0.00001871	0.00233164	0.0550212
S.W.O*RAS S.W.O	-0.00001086	0.00000282	0.00032887	0.00545227
S.W.O. BL*MLSS S.W.O	0.00001902	0.00000464	0.00014726	0.04874091

### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.15551254	0.05381637	0.0056498	0.51111603
TRS Pri	0.06100557	0.01684879	0.00067532	0.1165152
RAS.S.E	0.00005723	0.00001817	0.00273287	0.0506326
S.W.O*RAS S.W.O	-0.00001112	0.00000362	0.00343471	0.00527889
S.W.O. BL*MLSS S.W.O	0.00001887	0.0000065	0.00543266	0.02553549

### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.25734371	0.04963407	0.00000371	0.42874998
TRS Pri	0.05534116	0.01433386	0.00031953	0.13205446
RAS.S.E	0.0000958	0.00001828	0.00000303	0.07674418
S.W.O*RAS S.W.O	-0.00000891	0.0000026	0.00121893	0.00684106
S.W.O. BL*MLSS S.W.O	0.00001351	0.00000439	0.00336995	0.02159626

### Validation of predictive model of N-containing ( $Y_3$ ) during annual period:

RMSE of predictive model of N-containing ( $Y_3$ ) during annual

Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
4277.597855	5.873393985	2.66803E-05

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
2709.280652	5.7480438	1.074029007

### Validation of predictive model of N-containing ( $Y_3$ ) during winter period:

RMSE of predictive model of N-containing ( $Y_3$ ) during winter

Total sum of squared errors	RMS Error	Average Error
119.0766169	1.408762	-6.44517E-06

Cross-validation output of predictive model of N-containing ( $Y_3$ ) during winter

Validation of Sub Data	$SSE \left( \sum_{i=1}^n (\hat{\sigma}_i)^2 \right)$
Sub Data #1	14.668108
Sub Data #2	62.897187
Sub Data #3	26.616443
Sub Data #4	19.595084
Sub Data #5	22.111341
$RMSE \left[ \frac{\sum_{i=1}^n (\hat{\sigma}_i)^2}{n_v} \right]^{\frac{1}{2}}$	1.55932

### Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-9.81142616	2.85579443	0.00136669	369.4326172
Post lime Temp	0.06332319	0.02140568	0.00511761	6.00860834
RAS.S.W.O	0.00101248	0.0002354	0.00010255	35.63976669
Pol_DW/Dry ton solids	0.2625013	0.15465687	0.09721694	12.61915493
S.W.O. BL*MLSS S.W.O	-0.0003762	0.00011232	0.00174716	18.76857376
S.E. BL*MLSS S.E	0.00051965	0.00025964	0.05199272	10.40822792

### Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-8.49135685	1.97599804	0.00009711	314.4035645
Post lime Temp	0.06656916	0.01517155	0.00007295	14.69402504
RAS.S.W.O	0.00068846	0.00018424	0.00054484	15.54042721
Pol_DW/Dry ton solids	0.23146655	0.09126958	0.0149245	16.41931725
S.W.O. BL*MLSS S.W.O	-0.00028005	0.00008549	0.00208786	8.60968781
S.E. BL*MLSS S.E	0.00042406	0.00018343	0.02564154	7.84639883

### Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-9.28103447	2.40089297	0.00037822	351.4877319
Post lime Temp	0.06863222	0.01933898	0.00096802	13.33826637
RAS.S.W.O	0.00090472	0.00022084	0.00018695	25.94114113
Pol_DW/Dry ton solids	0.20327428	0.11520455	0.08492578	12.05634594
S.W.O. BL*MLSS S.W.O	-0.00036639	0.00010924	0.00169637	11.31712341
S.E. BL*MLSS S.E	0.0005694	0.00022035	0.01333429	14.83060837

### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-8.22279835	2.83878851	0.00596708	385.2200317
Post lime Temp	0.06318911	0.02068965	0.00390898	1.90995574
RAS.S.W.O	0.00103742	0.00025323	0.00018702	20.26739883
Pol_DW/Dry ton solids	0.14690812	0.11965648	0.22637787	10.19570541
S.W.O. BL*MLSS S.W.O	-0.00037767	0.00011681	0.00238487	18.78800011
S.E. BL*MLSS S.E	0.00039199	0.0002383	0.10743925	6.4921627

### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-8.97561359	2.73662305	0.00209303	427.929657
Post lime Temp	0.06612473	0.02099786	0.00301314	3.74225926
RAS.S.W.O	0.00116515	0.00023693	0.0000139	34.76968765
Pol_DW/Dry ton solids	0.13857614	0.12525623	0.27487558	15.7131958
S.W.O. BL*MLSS S.W.O	-0.00047319	0.00011195	0.00028454	23.69659615
S.E. BL*MLSS S.E	0.00058495	0.0002423	0.02020729	14.01468182

### Validation of predictive model of N-containing ( $Y_3$ ) during summer period:

RMSE of predictive model of N-containing ( $Y_3$ ) during summer

Total sum of squared errors	RMS Error	Average Error
2891.220447	6.569059597	-2.58462E-06

Cross-validation output of predictive model of N-containing ( $Y_3$ ) during summer

Validation of Sub Data	$SSE_v \left( \sum_{i=1}^n (\hat{\sigma}_i)^2 \right)$
Sub Data #1	364.00717
Sub Data #2	558.8245
Sub Data #3	901.81471
Sub Data #4	380.04765
Sub Data #5	1328.5441
$RMSE_v \left[ \frac{\sum_{i=1}^n (\hat{\sigma}_i)^2}{n_v} \right]^{1/2}$	7.26188



### Validation set of sub data # 1

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-18.7894726	10.54402828	0.08107539	7612.352051
Post lime Temp	0.13057736	0.08470499	0.12974936	197.126236
Pol_DW/Dry ton solids	0.88159162	0.32226831	0.00870095	339.9533691
S.E. BL*MLSS S.E	0.0013454	0.00054733	0.01762695	320.2048035

### Validation set of sub data # 2

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-15.4128695	10.51010227	0.14865653	6710.99707
Post lime Temp	0.0947485	0.08838063	0.28874421	107.9253998
Pol_DW/Dry ton solids	0.91947818	0.28426504	0.00214007	502.4208984
S.E. BL*MLSS S.E	0.00125902	0.00053792	0.02320474	253.1344299

### Validation set of sub data # 3

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-20.9177246	9.50623322	0.03242452	6929.20166
Post lime Temp	0.1405312	0.07779657	0.07687722	209.8670044
Pol_DW/Dry ton solids	0.95484078	0.26915956	0.00085619	528.3829346
S.E. BL*MLSS S.E	0.00119138	0.00049555	0.0199573	235.6619568

### Validation set of sub data # 4

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-20.19968605	10.42854691	0.0585268	7103.172852
Post lime Temp	0.12863187	0.08407332	0.13244841	228.150528
Pol_DW/Dry ton solids	1.01137185	0.3546719	0.00635135	429.9729004
S.E. BL*MLSS S.E	0.00134697	0.00054067	0.01616286	321.4047546

### Validation set of sub data # 5

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-14.9698105	9.03118992	0.10366684	6110.551758
Post lime Temp	0.1296912	0.07124355	0.07468709	172.5416107
Pol_DW/Dry ton solids	0.26243696	0.3170732	0.41178161	12.15908432
S.E. BL*MLSS S.E	0.00201614	0.00055807	0.00070218	481.6195679

## Appendix D

### The Best Explanatory and Predictive Models

#### A. The best explanatory model for TRS prior-lime ( $Y_I$ )

##### The best annual explanatory models of TRS prior-lime ( $Y_I$ )

	Multiple	R-Square	Adjusted	StErr of		
Summary	R		R-Square	Estimate		
	0.6466	0.4181	0.4066	0.405120486		
	Degrees of	Sum of	Mean of	F-Ratio	p-Value	
ANOVA Table	Freedom	Squares	Squares			
Explained	5	29.83147587	5.966295174	36.3527	< 0.0001	
Unexplained	253	41.52301988	0.164122608			
	Coefficient	Standard	t-Value	p-Value	Confidence Interval 95%	
Regression Table		Error			Lower	Upper
Constant	-2.023725624	0.680571888	-2.9736	0.0032	-3.364033558	-0.68341769
WAS %S	-0.388862351	0.094880373	-4.0984	< 0.0001	-0.575718316	-0.202006385
Temp Sec Eff	0.05838969	0.005826328	10.0217	< 0.0001	0.046915408	0.069863972
SWO. BL*MLSS.	2.6907E-05	7.71915E-06	3.4857	0.0006	1.1705E-05	4.2109E-05
SWO	0.164807583	0.085661938	1.9239	0.0555	-0.003893737	0.333508902
GT %S	0.164807583	0.085661938	1.9239	0.0555	-0.003893737	0.333508902
D <sub>T</sub> *Temp	0.007622134	0.001137213	6.7025	< 0.0001	0.005382523	0.009861745

##### The best winter explanatory models of TRS prior-lime ( $Y_I$ )

	Multiple	R-Square	Adjusted	StErr of		
Summary	R		R-Square	Estimate		
	0.7476	0.5589	0.5401	0.173475978		
	Degrees of	Sum of	Mean of		p-Value	
ANOVA Table	Freedom	Squares	Squares			
Explained	2	1.792274003	0.896137002	29.7780	< 0.0001	
Unexplained	47	1.414413997	0.030093915			
	Coefficient	Standard	t-Value	p-Value	Confidence Interval 95%	
Regression Table		Error			Lower	Upper
Constant	-0.283899873	0.15693682	-1.8090	0.0768	-0.599616031	0.031816285
MLSS.SWO	0.000110959	3.02443E-05	3.6688	0.0006	5.01151E-05	0.000171803
Ambient Temp	0.015049725	0.002427339	6.2001	< 0.0001	0.010166549	0.019932902

## The best summer explanatory models of TRS prior-lime ( $Y_1$ )

Summary	Multiple	R-Square	Adjusted	StErr of		
	R		R-Square	Estimate		
	0.7688	0.5911	0.5591	0.366718913		
ANOVA Table	Degrees of	Sum of	Mean of	F-Ratio	p-Value	
	Freedom	Squares	Squares			
Explained	5	12.44056186	2.488112372	18.5013	< 0.0001	
Unexplained	64	8.606896711	0.134482761			
Regression Table	Coefficient	Standard	t-Value	p-Value	Confidence Interval 95%	
		Error			Lower	Upper
Constant	4.869238295	2.045571422	2.3804	0.0203	0.782739605	8.955736985
DAF TS	-6.29232E-06	2.51552E-06	-2.5014	0.0149	-1.13177E-05	-1.26698E-06
Cake %S	-0.237370938	0.055436952	-4.2818	< 0.0001	-0.348118981	-0.126622895
SWE.BL	0.63003452	0.148290202	4.2487	< 0.0001	0.333790785	0.926278254
RAS.SWO	0.000596106	0.000132375	4.5031	< 0.0001	0.000331655	0.000860556
GT %S	0.21122279	0.145349169	1.4532	0.1511	-0.079145554	0.501591135

## B. The best explanatory model for TRS post-lime ( $Y_2$ )

### The best annual explanatory models of TRS post-lime ( $Y_2$ )

Summary	Multiple	R-Square	Adjusted	StErr of		
	R		R-Square	Estimate		
	0.7421	0.5507	0.5418	0.046055898		
ANOVA Table	Degrees of	Sum of	Mean of	F-Ratio	p-Value	
	Freedom	Squares	Squares			
Explained	5	0.657733134	0.131546627	62.0168	< 0.0001	
Unexplained	253	0.536649878	0.002121146			
Regression Table	Coefficient	Standard	t-Value	p-Value	Confidence Interval 95%	
		Error			Lower	Upper
Constant	-0.177903592	0.034821302	-5.1090	< 0.0001	-0.246480135	-0.109327048
Blend(Pri:Sec)	0.034349804	0.009725666	3.5319	0.0005	0.015196226	0.053503382
SE.BL*MLSS	1.06533E-05	2.07664E-06	5.1301	< 0.0001	6.56364E-06	1.4743E-05
<b>Predicted Post lime temp</b>	0.000708878	0.000351361	2.0175	0.0447	1.69139E-05	0.001400842
(DT = 1)	-0.015346898	0.0107471	-1.4280	0.1545	-0.036512074	0.005818278
TRS Pri	0.065067996	0.006662666	9.7661	< 0.0001	0.051946642	0.07818935

### The best winter explanatory models of TRS post-lime ( $Y_2$ )

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.7583	0.5750	0.5472	0.008663018		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	3	0.004669797	0.001556599	20.7414	< 0.0001	
Unexplained	46	0.003452203	7.50479E-05			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	0.26473271	0.065204705	4.0600	0.0002	0.133482447	0.395982973
WAS %S	-0.05501787	0.011670193	-4.7144	< 0.0001	-0.07850875	-0.03152699
<u>Predicted Post lime temp</u>	0.000276997	0.000147635	1.8762	<b>0.0670</b>	-2.01772E-05	0.00057417
TRS Prior-lime	0.017966929	0.005835715	3.0788	0.0035	0.006220245	0.029713614

### The best summer explanatory models of TRS post-lime ( $Y_2$ )

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.7407	0.5486	0.5208	0.055351281		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	4	0.242032462	0.060508116	19.7496	< 0.0001	
Unexplained	65	0.19914468	0.003063764			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	-0.452628502	0.09391427	-4.8196	< 0.0001	-0.64018825	-0.265068754
RAS.SE	6.19441E-05	1.75389E-05	3.5318	0.0008	2.69165E-05	9.69718E-05
SW.E. BL*MLSS S.W.E	1.34304E-05	5.48136E-06	2.4502	0.0170	2.48339E-06	2.43775E-05
GT TS	6.35973E-07	2.63697E-07	2.4118	0.0187	1.09334E-07	1.16261E-06
TRS Prior-lime	0.042975773	0.015835827	2.7138	0.0085	0.011349443	0.074602104

### C. The best explanatory model for N-Containing compounds ( $Y_3$ )

#### The best annual explanatory models of N-containing ( $Y_3$ )

<i>Summary</i>	Multiple	R-Square	Adjusted	StErr of		
	R		R-Square	Estimate		
	0.6533	0.4268	0.4154	5.852027964		
<i>ANOVA Table</i>	Degrees of	Sum of	Mean of	F-Ratio	p-Value	
	Freedom	Squares	Squares			
Explained	4	5124.994883	1281.248721	37.4128	< 0.0001	
Unexplained	201	6883.492491	34.2462313			
<i>Regression Table</i>	Coefficient	Standard	t-Value	p-Value	Confidence Interval 95%	
		Error			Lower	Upper
Constant	-21.03751731	4.36162886	-4.8233	< 0.0001	-29.63793632	-12.43709829
(DT = 1)	-2.556762953	1.202867381	-2.1256	0.0348	-4.928620767	-0.184905139
Post lime Temp	0.077582307	0.029715833	2.6108	0.0097	0.018987542	0.136177072
Pol_Sec	0.01445052	0.001583192	9.1275	< 0.0001	0.011328723	0.017572316
Pol_DW/DT solids	0.497290039	0.179679273	2.7677	0.0062	0.142991889	0.851588189

#### The best winter explanatory models of N-containing ( $Y_3$ )

<i>Summary</i>	Multiple	R-Square	Adjusted	StErr of		
	R		R-Square	Estimate		
	0.5228	0.2733	0.2344	1.654450718		
<i>ANOVA Table</i>	Degrees of	Sum of	Mean of	F-Ratio	p-Value	
	Freedom	Squares	Squares			
Explained	3	57.6562163	19.21873877	7.0213	0.0004	
Unexplained	56	153.283602	2.737207179			
<i>Regression Table</i>	Coefficient	Standard	t-Value	p-Value	Confidence Interval 95%	
		Error			Lower	Upper
Constant	-7.724369172	2.350857946	-3.2858	0.0018	-12.43370353	-3.015034809
Post lime Temp	0.046262438	0.018160085	2.5475	0.0136	0.009883415	0.08264146
RAS.S.E	0.001230162	0.000385603	3.1902	0.0023	0.000457706	0.002002618
Pol_DW/Dry ton solids	0.197215755	0.109774633	1.7966	0.0778	-0.02268926	0.417120771

### The best summer explanatory models of N-containing ( $Y_3$ )

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.5678	0.3224	0.2787	6.554906995		
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	4	1267.621144	316.9052861	7.3756	< 0.0001	
Unexplained	62	2663.941954	42.96680571			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	-24.1874781	9.012854084	-2.6837	0.0093	-42.2039167	-6.171039498
S.W.O.BL	2.481853264	1.079104403	2.2999	0.0248	0.324754299	4.63895223
Post lime Temp	0.137201887	0.070421295	1.9483	0.0559	-0.003568277	0.27797205
Pol_DW/DTS	0.745779805	0.267703264	2.7858	0.0071	0.210648605	1.280911005
S.E. BL*MLSS S.E	0.001147703	0.000474962	2.4164	0.0186	0.000198267	0.002097138

### D. The best predictive model for TRS prior-lime ( $Y_I$ )

#### The best annual predictive model for TRS prior-lime ( $Y_I$ )

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	2.52609754	1.1519859	0.0298818	202.00839
GT FW	-0.00000185	4.6E-07	9.826E-05	2.2398329
WAS %S	-0.42563313	0.1177421	0.0004111	7.1748996
S.E*RAS S.E	-0.00006049	2.092E-05	0.0044149	0.5851115
Temp(F)	0.03654794	0.0077	4.84E-06	1.7907034
S.E. BL*MLSS S.E	0.00011997	3.214E-05	0.0002694	2.2039928
(Temp*DT = 1)	0.00752061	0.0014197	4.2E-07	4.558979

#### The best winter predictive model for TRS prior-lime ( $Y_I$ )

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.02792609	0.2463826	0.9102731	32.256512
SE.BL	0.33644986	0.1638628	0.0460234	0.1855095
RAS.SE	-0.00014061	4.723E-05	0.004717	0.2603906
Amb Temp	0.01455642	0.002347	1.7E-07	1.1900514
S.WO. BL*MLSS	0.00002952	7.36E-06	0.0002328	0.3524687
SE.BL*MLSS	-0.00009392	5.758E-05	0.1099693	0.0694777

**The best summer predictive model for TRS prior-lime ( $Y_1$ )**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	3.64457941	2.236974	0.10825169	119.7866516
GT %S	0.25450373	0.14822559	0.09089159	1.04515743
DAF TS	-0.00000629	0.0000025	0.01453312	0.57240433
Cake %S	-0.25128782	0.05612963	0.00003242	3.94246984
S.W.E.BL	0.58212888	0.15188292	0.00029498	4.15345144
RAS.S.W.O	0.00057573	0.00013253	0.00005182	2.72708106
Temp Sec Eff	0.0228566	0.01737705	0.19316527	0.23004428

**E. The best predictive model for TRS post-lime ( $Y_2$ )**

**The best annual predictive model for TRS post-lime ( $Y_2$ )**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.2130059	0.0455867	6.55E-06	0.5896361
TRS Pri	0.0659544	0.0087952	0	0.3409052
Blend(Pri:Sec)	0.0305947	0.0137123	0.0271624	0.0351426
S.E. BL*MLSS	1.207E-05	2.76E-06	2.296E-05	0.0287148
Fitted Post lime temp	0.0011472	0.0004799	0.0180696	0.0845526
(DT = 1)	-0.020767	0.0142643	0.1475329	0.0051341

**The best winter predictive model for TRS post-lime ( $Y_2$ )**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	0.2867726	0.0590205	1.535E-05	0.013778
TRS Pri	0.0116767	0.0066711	0.0870302	0.0011469
WAS %S	-0.0593135	0.0105097	1.12E-06	0.0032587
S.W.O.BL	0.0077809	0.0029327	0.0110519	0.0001113
RAS.S.W.E	-0.0000026	1.47E-06	0.0848898	0.0003808
Fitted Post lime temp	0.0003485	0.0001882	0.0707076	0.0002332

**The best summer predictive model for TRS post-lime ( $Y_2$ )**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-0.17571084	0.0467413	0.0003674	0.5742228
TRS Pri	0.05332687	0.0140151	0.0003161	0.1351165
RAS.S.E	0.00006615	1.628E-05	0.0001322	0.0718594
S.W.O*RAS S.W.O	-0.00001047	2.74E-06	0.0002975	0.0063159
S.W.O. BL*MLSS S.W.O	0.00001758	4.47E-06	0.0002067	0.0437984

**F. The best predictive model for N-containing compounds ( $Y_3$ )**

**The best annual predictive model for N-containing compounds ( $Y_3$ )**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-26.581068	5.9264584	1.701E-05	7671.8359
Blend(Pri:Sec)	3.75422859	1.8682782	0.0467703	161.21564
Pol_Sec	0.01555401	0.0018966	0	2413.6448
Pol_DAF+Pol_DW	0.00136673	0.0008619	0.1154837	283.43039
Pol_DW/DT solids (DT = 1)	0.47099641 -3.48496151	0.2808827 1.1686184	0.0962203 0.0034784	126.53569 322.38025

**The best winter predictive model for N-containing compounds ( $Y_3$ )**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-8.95781898	2.249253	0.0002054	460.76184
Post lime Temp	0.06529341	0.0172074	0.0003759	8.9302416
RAS.S.W.O	0.00096194	0.0002004	0.000013	32.835461
Pol_DW/DT solids	0.19674	0.1061232	0.0692212	16.826542
SWO.BL*MLSS	-0.00037296	9.627E-05	0.0002917	19.832117
SE.BL*MLSS	0.00049883	0.0002021	0.0167575	13.438837

**The best summer predictive model for N-containing compounds ( $Y_3$ )**

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	-18.11300278	8.9057169	0.04618162	8600.490234
Post lime Temp	0.12375169	0.07252789	0.09288455	232.6981964
Pol_DW/DT solids	0.83004338	0.27406377	0.00355824	407.0157471
SE.BL*MLSS	0.00140841	0.00047668	0.00439817	400.6286926



## Reference

- AKAIKE, H. (1974) A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, 19, 716-723.
- ALBRIGHT, S. C., WINSTON, W. L. & ZAPPE, C. (2004) *Data Analysis for Manager with Microsoft Excel*, Belmont, CA, Brooks/Cole-Thomson Learning.
- ANGRIST, J. D. & IMBENS, G. W. (1995) Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association*, 90, 431-442.
- ANGRIST, J. D. & KRUEGER, A. B. (2001) Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives*, 15, 69-85.
- ASCE (2000) Conveyance of Residuals from Water and Wastewater Treatment. Reston, VA, Environmental and Water Resources Institute of the ASCE.
- BOWKER, R. P. G., SMITH, J. M. & WEBSTER, N. A. (1989) *Odor and Corrosion Control in Sanitary Sewerage Systems and Treatment Plants*, SciTech Publishing Incorporated.
- CHANG, J.-S., ABU-ORF, M. & DENTEL, S. K. (2005) Alkylamine odors from degradation of flocculant polymers in sludge. *Elsevier*, 39, 3369-3375.
- CHANG, L.-L. & DENTEL, S. K. (2001) Aerobic and anaerobic biodegradability of a flocculant polymer. *Water Science and Technology*, 44, 461-468.
- CHEN, Y., HIGGINS, M. J., MURTHY, S. N., TOFFEY, W. E. & FOSTER, D. J. (2005) Roles of methanogens on volatile organic sulfur compound production in anaerobically digested wastewater biosolids. *Water Science and Technology*, 52, 67-72.
- CHURCHILL, R. J. & RYBACKI, R. L. (1997) Polymer Addition Improves Waste Water Treatment Process. *Water and Sewage Work*, 20, 10-12, 14-17.
- DAGUE, R. R. (1972) Fundamental of odor control. *Water Pollutant Control*, 44, 12.
- DASU, T. & JOHNSON, T. (2003) *Exploratory Data Mining and Data Cleaning*, Hoboken, NJ, A John Wiley & Sons, Inc., Publication.
- DCWASA (2005) *Blue Plains Advanced Wastewater Treatment Plant*. Department of Wastewater Treatment, Washington, D.C.

- DRÄGER (2005) Dräger Polytron 7000. Lubeck, Germany, Dräger Safety AG & Co. KGaA.
- FOSTER, M. R. (2002) Predictive Accuracy as an Achievable Goal of Science. *Philosophy of Science*, 69, S124-S134.
- GABRIEL, S., VILALAI, S., APRISE, S., KIM, H., MCCONNELL, L. L., TORRENT, A., PEOT, C. & RAMIREZ, M. (2005) Prediction of Dimethyl Disulfide Levels from Biosolids Using Statistical Modeling. *J. of Environ. Sci. Health*, 40, 2009-2025.
- GABRIEL, S., VILALAI, S., PEOT, C. & RAMIREZ, M. (2006) Statistical modeling to forecasting odor levels of biosolids applied to reuse sites. *J. of Environmental Engineering*, 132, 479-488.
- GENG, A. L., CHEN, X. G., GOULD, W. D., NG, Y. L., YAN, R., LEE, C. C. & LIANG, D. T. (2004) Removal of odorous sulfur-containing gases by new isolate from activated sludge. *Water Science and Technology*, 50, 291-297.
- GOSTELOW, P. & PARSONS, S. A. (2000) Sewage treatment works odour measurement. *Water Science and Technology*, 41, 33-40.
- GREENBERG, M. R., CAREY, G. W., ZOBLER, L. & HORDON, R. M. (1973) A Statistical Dissolve Oxygen Model for a Free-Flowing River System. *J. of American Statistical Association*, 68, 279-283.
- GREENE, W. H. (2003) *Econometric Analysis*, Saddle River, N.J., Prentice Hall.
- GREGOR, S. (2006) The Nature of Theory in IS. *MIS Quarterly*, 30, 611-642.
- HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. (2008) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer.
- HENTZ, L. H. (1997) The Chemical, Biological, and Physical Origins of Biosolids Emissions: A Review. IN WEF (Ed.) *WEF Odor and VOC Conference*. Houston, TX, Water Environmental Federation.
- HIGGINS, M. J., YAROSZ, D. P., CHEN, Y., MURTHY, S. N., MAAS, N. A., COONEY, J. R. & GLINDEMANN, D. (2003) Mechanisms of volatile sulfur productions in digested biosolids. IN WEF (Ed.) *Water Environment Federation and AWWA Annual Biosolids and Residuals*. Baltimore, MD, Water Environment Federation.
- HWANG, Y., MATSUO, T., HANAKI, K. & SUZUKI, N. (1995) Identification and Quantification of Sulfur and Nitrogen Containing odorous Compounds in Wastewater. *Wat. Res.*, 29.

- HYNDMAN, R. J. & ATHANASOPOULOS, G. (2010) *Forecasting: principles and practice*. Melbourne, Australia.
- INSTRUMENTS, D. (2005) *Low Range Sampling System-2 (LRSS-2)*. Phoenix, AZ, Detection Instruments Corporation.
- JANPENG PEN, A. (2006) Real-time monitoring and forecasting of odor after dewatering in wastewater treatment. *Department of Civil and Environmental Engineering*. College Park, MD, University of Maryland.
- JANPENG PEN, A., BAECHER, G. B., KIM, H., PEOT, C. & RAMIREZ, M. (2007) Monitoring and Forecasting of Odour after Dewatering in Wastewater Treatment Plant. IN AETT (Ed.) *12th European Biosolids & Organics Resources Conference & Workshop*. Manchester, U.K., Aqua Enviro Technology Transfer (AETT).
- JANPENG PEN, A., BAECHER, G. B., PEOT, C. & RAMIREZ, M. (2008) An evaluation of dewatered sludge percent solids at District of Columbia Water and Sewer Authority (DCWASA). IN AETT (Ed.) *13th European Biosolids & Organics Resources Conference & Workshop*. Manchester, U.K., Aqua Enviro Technology Transfer (AETT).
- KENNEDY, P. (2008) *A Guide to Econometrics*, Malden, MA, Blackwell Publishing.
- KIM, H., MCCONNELL, L. L., RAMIREZ, M., ABU-ORF, M., CHOI, H. L. & PEOT, C. (2005) Characterization of odors from limed biosolids treated with nitrate and anthraquinone. *Journal of Environmental Science and Health*, 40, 139-149.
- KIM, H., MURTHY, S., MCCONNELL, L. L., PEOT, C., RAMIREZ, M. & STRAWN, M. (2002) Characterization of wastewater and solids odor using solids phase microextraction at a large wastewater treatment plant. *Water Science and Technology*, 46, 9-16.
- KIM, H., MURTHY, S., PEOT, C., RAMIREZ, M., STRAWN, M. & MCCONNELL, L. L. (2001) Examination of mechanisms for odor compound generation during lime stabilization. *74th Annual Water Environment Federation Technical Exhibition and Conference*. Atlanta, GA, Water Environment Federation.
- KREHBIEL, T. C. (2004) Teaching Brief Correlation Coefficient Rule of Thumb. *Decision Sciences Journal of Innovative Education*, 2, 97-100.
- LAMON, E. C. (1995) A Regression Model for the Prediction of Chlorophyll a in Lake Okeechobee, Florida. *Lake and Reservoir Management*, 11, 283-290.

- LAMON, E. C. & CLYDE, M. A. (2000) Accounting for Model Uncertainty in Prediction of Chlorophyll a in Lake Okeechobee. *Journal of Agricultural, Biological, and Environmental Statistics*, 5, 297-322.
- LAMON, E. C., RECKHOW, K. H. & HARVENS, K. E. (1996) Using Generalized Additive Model for Prediction of Chlorophyll a in Lake Okeechobee, Florida. *Lake & Reservoir: Research and Management*, 2, 37-46.
- METCALF&EDDY, I. (2003) *Wastewater Engineering: Collection, Treatment, and Disposal*, New York, NY, McGraw-Hill.
- MICHAELSEN, J. (1987) Cross-Validation in Statistical Climate Forecast Models. *Journal of Climate and Applied Meteorology*, 26, 1589-1600.
- MONTGOMERY, D. C., PECK, E. A. & VINING, G. G. (2006) *Introduction to Linear Regression Analysis*, Hoboken, NJ, John Wiley & Sons, Inc.
- MURRAY, M. P. (2006) Avoiding Invalid Instruments and Coping with Weak Instruments. *Journal of Economic Perspectives*, 20, 111-132.
- MURTHY, S. N., PEOT, C., NORTH, J., NOVAK, J. & HIGGINS, M. (2002b) Characterization and control of residual sulfur odors from lime stabilized and digested biosolids. IN WEF (Ed.) *16th Annual Residual and Biosolids Management Conference*. Austin, Texas, Water Environment Federation.
- MURTHY, S. N., SADICK, T., BAILEY, W., PEOT, C., TOLBERT, D. & STRAWN, M. (2001) Odor Mitigation from Lime Stabilized Biosolids. IN WEF (Ed.) *Residuals and Biosolids Management Conference*. San Diego, California, Water Environment Federation.
- MURTHY, S. N., SADICK, T., KIM, H., MCCONNELL, L. L., PEOT, C., BAILEY, W. & NOVAK, J. (2002a) Mechanisms for odour generation during lime stabilization. *IWA Biennial Conference*. Melbourne, Australia, IWA.
- MYATT, G. J. (2007) *Making sense of Data (A Practical Guide to Exploratory Data Analysis and Data mining)*, Hoboken, NJ, A John Wiley & Sons, Inc., Publication.
- NORTH, J. M. (2003) An Evaluation of Methods for Quantifying Lime Incorporation into Mechanically Dewatered Sludge. *Department of Civil and Environmental Engineering*. College Park, MD, University of Maryland, College Park.
- NOVAK, J., GLINDEMANN, D., MURTHY, S. N., GERWIN, S. C. & PEOT, C. (2002) Mechanisms for generation and control of trimethylamine and dimethyl disulfide from lime stabilized biosolids. IN WEF (Ed.) *Odors and*

*Toxic Air Emissions 2002 Conference*. Albuquerque, NM, Water Environment Federation.

NOVAK, J. T., ADAMS, G., CHEN, Y.-C., ERDAL, Z., ROBERT H. FORBES, J., GLINDEMANN, D. M., HARGREAVES, J. R., LAWRENCE H. HENTZ, J., HIGGINS, M. J., MURTHY, S. N., WITHERSPOON, J. & CARD, T. (2004) Odor generation patterns from anaerobically digested biosolids. IN WEF (Ed.) *WEF/A&WMA odors and Air Emission 2004*. Bellevue, WA, Water Environment Federation.

OLESZKIEWICZ, J. A. & MAVINIC, D. S. (2002) Wastewater Biosolids: An Overview of processing, treatment, and Management. *Environmental Engineering Science*, 1, 75-88.

PEOT, C. (2007) Personal Communication in Solids Processes. Washington, DC.

PEOT, C. & RAMIREZ, M. (2007) Personal Communication. Washington, DC.

POSADA, D. & BUCKLEY, T. R. (2004) Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches over Likelihood Ratio Tests. *Systematic Biology*, 53, 793-808.

ROSENFELD, P. E. & SUFFET, I. H. (2004) Understanding odorants associated with compost, biomass facilities, and the land application of biosolids. *Water Science and Technology*, 49, 193-198.

RUDOLFS, W. (1995) *Lime handling application and storage in treatment process*, Arlington, VA, National Lime Association.

SARKAR, U., LONGHURST, P. J. & HOBBS, S. E. (2003) Community Modeling: a tool for correlating estimates of exposure with perception of odour from municipal solids waste (MSW) landfills. *J. Environ. Manage.*, 68, 133-140.

SCHNEEKLOTH, E., MURTHY, S. N. & NOVAK, J. T. (2006) Factors Contributing to Trimethylamine Generation from Limed and Polymer Conditioned Sludge. *Odors and Air Emission*. Water Environmental Federation.

SEKYIAMAHA, K. & KIM, H. (2006) Reduction in biosolids odor by managing solids inventory in the secondary activated sludge treatment system. *J of Odor Research and Engineering*, 5.

SEKYIAMAHA, K. & KIM, H. (2009) Biosolids odor reduction by solids inventory management in the secondary activated sludge treatment system. *Water Science and Technology*, 59, 241-247.

- SHMUELI, G. (2010a) Personal Communication on Cross-Validation Analysis. Washington, DC.
- SHMUELI, G. (2010b) To Explain or To Predict? *Statistical Science*, *forthcoming*.
- SHMUELI, G. & KOPPIUS, O. (2007) Predictive vs. Explanatory Modeling in IS Research. *Conference on Information Systems & Technology (Best Paper Award)*. Seattle, WA.
- SHMUELI, G. & KOPPIUS, O. (2009) The Challenge of Prediction in information Systems Research. *Working Paper RHS 06-099*. College Park, MD, Smith School of Business, University of Maryland.
- SHMUELI, G. & KOPPIUS, O. (2010) Predictive Analytics in Information Systems Research. *MIS Quarterly (Under review)*.
- SHMUELI, G., PATEL, N. R. & BRUCE, P. C. (2007) *Data mining for business intelligence*, Hoboken, NJ, A John Wiley & Sons, Inc., Publication.
- STUETZ, R. & FRECHEN, F.-B. (2001) *Odour in wastewater treatment measurement, modeling and control*, London, UK, IWA.
- TUKEY, J. W. (1970) *Exploratory Data Analysis*, Reading, MA, Addison-Wesley.
- TUKEY, J. W. (1977) *Exploratory Data Analysis*, Reading, MA, Addison-Wesley.
- TURKMEN, M., DENTEL, S. K., CHIU, P. C. & HEPNE, S. (2004) Analysis of sulfur and nitrogen odorants using solids-phase microextraction and GC-MS. *Water Science and Technology*, 40, 115-120.
- USEPA (1985) *Odor and Corrosion Control in Sanitary Sewerage Systems and Treatment Plants*, Cincinnati OH, EPA.
- USEPA (1999) *Biosolids Generation, Use, and Disposal in the United State*. Washington, D.C., EPA.
- USEPA (2000a) *Biosolids and Residuals Management Fact Sheet Odor Control in Biosolids Management*. Washington, D.C., EPA.
- USEPA (2000b) *Guide to Field Storage of Biosolids*. Washington, D.C., EPA.
- USEPA (2000c) *Biosolids Technology Fact Sheet Land Application of Biosolids*. IN AGENCY, U. S. E. P. (Ed.). Washington, DC, USEPA.
- VESILIND, P. A. (2003) *Wastewater Treatment Plant Design*. *Water Environment Federation*.

- VILALAI, S. (2003) Forecasting Odor Levels for Biosolids Product Based on Ambient Conditions. *MS Thesis, Department of Civil and Environmental Engineering*. University of Maryland, College Park.
- VILALAI, S. (2008) Statistical Odor Prediction Models for Supporting Biosolids Odor Management. *Ph.D. Thesis, Department of Civil and Environmental Engineering*. College Park, University of Maryland, College Park.
- VISAN, M. M. (2003) Mechanism of Odour Production in Dewatered Biosolids. *MS Thesis, Department of Civil and Environmental Engineering*. Ottawa, Ontario, Carleton University.
- VOELZ, L. D., THOMPSON, J. K., GUADES, R. J., ABRI-SAMARA, S., SCHIRIPA, R., TORRES, E. M., DILLON, C. D., AHN, T. & RAINE, T. J. (2006) Odor Modelling for Overall Odor Control Planning. IN WEF/AWWA (Ed.) *Odors and Air Emissions*. Hartford, CT, Water Environment Federation.
- WANG, S., JANK, W. & SHMUELI, G. (2008) Explaining and Forecasting Online Auction Prices and Their Dynamics Using Functional Data Analysis. *Journal of Business & Economic Statistics*, 26, 144-160.
- WEF (1995(a)) *Odor Control in Wastewater Treatment Plants*, Alexandria, VA and ASCE, New York, NY, Water Environment Federation.
- WEF (1996) *Operation of Municipal Wastewater Treatment Plants*, Alexandria, VA, Water Environment Federation.
- WILLIAMS, T. O. & SERVO, S. (2005) Odor Modeling at a Biosolids Composting Facility. *Biocycle*, 46, 25.
- WINSTON, L. W. (2004) *Operation Research Applications and Algorithms*, Belmont, CA., Thomson- Brooks/Cole.
- WOOLDRIDGE, J. M. (2000) *Introductory Econometrics: A Modern Approach*, South Western College Publishing, Thomson Learning.
- WOOLDRIDGE, J. M. (2002) *Econometric analysis of cross section and panel data*, Cambridge, M.A., The MIT Press.
- YANG, G. & HOBSON, J. (2000) Odour nuisance-advantages and disadvantages of a quantitative approach. *Water Science and Technology*, 41, 97-106.
- YONGSLRL, C., VOLLERTSEN, J. & HVLTVED-JACONSEN, T. (2004) Hydrogen sulfide emission in sewer networks: a two-phase modeling approaching to the sulfur cycle. *Water Science and Technology*, 50, 161-168.