ABSTRACT

Title of Document: STATISTICAL ODOR PREDICTION
MODELS FOR SUPPORTING BIOSOLIDS
ODOR MANAGEMENT

Sirapong Vilalai, Doctor of Philosophy, 2008

Directed By: Associate Professor Steven A Gabriel
Department of Civil and Environmental
Engineering

Biosolids are being beneficially recycled for agricultural purpose. Often, however, biosolids odors diminish marketability of biosolids, bring community opposition or, in the worst case, cause to ban the biosolids land application program. This dissertation is aiming to develop practical biosolids odor prediction models that can be applied for biosolids management on daily basis using the existing data available at the wastewater treatment plant and at the application sites as explanatory variables. Therefore, biosolids producer can use the plant odor predicting models to early detect and notify the hauling contractor when malodorous biosolids are anticipated. With the field odor models, malodorous products can be allocated accordingly to the appropriate sites in preventing the odor complaints from the communities.

First, biosolids odors prediction models at wastewater treatment plant were developed using linear regression analysis and categorical data analysis. Biosolids

odor was predicted in terms of detection threshold (DT) concentration and class of biosolids odor (odorous or non-odorous). Variables influencing biosolids odor levels at the plant were the percent solids and temperature of biosolids, percentage of the gravity thickener solids (GT) in the blend tank, pH of the GT solids, concentration of the return activated sludge (RAS) at the secondary process, and number of centrifuges running.

Second, simulation and sensitivity analysis were conducted on the selected biosolids odor prediction model when uncertainty in the input variables was considered. Two variables (i.e., the number of centrifuges running and the percentage of GT solids in the blend tank) were identified as decision variable that could reduce the probability of producing odorous biosolids.

Last, a biosolids odors prediction model for use at field site was developed using ordered logit model. Various variables at the field site (i.e. weather conditions, odor measurement time of the day, wind condition, temperature, and inspector odor sensitivity) were included in the analysis. Finally, variables relating to field odor levels were the biosolids odor levels (detection threshold) at the plant, temperature at the reuse site, and wind conditions.

STATISTICAL ODOR PREDICTION MODELS FOR SUPPORTING BIOSOLIDS
ODOR MANAGEMENT

By

Sirapong Vilalai

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
2008

Advisory Committee:
Associate Professor Steven A. Gabriel, Chair/Advisor
Professor Shapour Azarm, Dean's Representative
Professor Gregory B. Baecher
Associate Professor Eric A. Seagren
Dr. Danny R. Hughes

Dedication

This dissertation is dedicated to my parents Jaruwan and Somsak Vilalai and my

sister Vanuslil for their supports and patience over the past years

Acknowledgements

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The United States Environmental Protection Agency (USEPA) anticipates there will be 8.2 million dry tons of biosolids production in the United States in 2010 (USEPA, 1999). Compared with 6.9 million dry tons produced in 1998 and 7.6 million dry tons in 2005, the increasing trend of biosolids production in the U.S. raises the awareness of biosolids producers and regulators that environmentally and publicly acceptable biosolids utilization must be achieved.

1.1 What are biosolids?

Biosolids are the by-product from wastewater treatment plant that processes the domestic and industrial wastewater daily to get rid of the contaminants, i.e., organic waste, nitrogen, phosphorous etc. in wastewater prior to discharge to the receiving stream. The solids contents in wastewater are clarified, thickened, dewatered, and finally stabilized for pathogen and odors reduction. According to 40 CFR Part 503 biosolids rule, biosolids can be classified as either class A or Class B standards (USEPA, 1994). Biosolids that meet class A standard must maintain a pH of 12 for the first 72 hours and have pathogens less than 1000 fecal coliforms per gram of total solids, which is the detection level. This product can be used as commercial fertilizer. Class B biosolids are the treated product that must maintain a pH of 12 for the first two hours and a pH of 11.5 for the other 12 hours. Pathogens in class B biosolids must be less than 2,000,000 fecal coliforms per gram of total solids, which is the public health threat limit (USEPA, 2000). This product has more usage restrictions and cannot be sold as commercial fertilizer.

Regarding the nutrients in the product, biosolids return their beneficial properties back to the agricultural industry when used for soil enrichment.

1.1.1 Biosolids management: Disposal options

Due to the Ocean Dumping Ban Act (1988), millions of tons of biosolids produced in the U.S. each year have three main disposal options: land application, incineration, and surface disposal or land filling. As estimated by EPA in 1998, 41 percent of biosolids in the U.S. were disposed of by land application, 22 percent were disposed of by incineration, and 17 percent were disposed of by surface disposal and land filling (USEPA, 1999). Land application is conducted by either spreading biosolids over the application area or by injecting biosolids in liquid form into the soil to improve soil properties. In practice, biosolids are applied to agricultural farms, tree farms, forests, mine reclamation sites, and used for gardening, and landscaping. This beneficial reuse option returns the nutrient rich product to the environment and helps farmers reduce their dependence on commercial fertilizers. Additionally, biosolids applied to the land can prevent top soil erosion and retain soil moisture.

The second disposal option, incineration, uses high temperature from a combustion device to reduce biosolids to roughly 20 percent of their initial volume. Pollution from the ash and the metal left is the big issue considered in the Clean Air Act regulations (Clean air act, 1990) and the EPA 503 biosolids rule (EPA, 1994). The strict regulations on combustion devices result in a limited number of wastewater treatment plants using the incineration option or, in some cases, using this option as a second alternative when land application is not permitted (Epstein, 2002).

The last option is surface disposal and land filling. This option disregards the beneficial properties of biosolids by placing them onto the land as the final disposal. A difference between land application and surface disposal is the application rate (USEPA, 1999). The application rate for surface disposal is greater than the agronomic rate, the amount that crops need to retain nutrients for their growth (Davis, 2008). This can cause excess nutrients to contaminate ground water or surface water. Site restrictions, such as, ground water monitoring or unlined trenches dug into the ground are required in the disposal areas. In practice, attempts to prepare disposal areas to meet the land filling and surface disposal standards cost much more compared to land application (USEPA, 1999).

1.2 Problem under review: Biosolids odor

With regard to the environmental impact from the incineration option and cost of preparing and monitoring land filling or surface disposal area, land application is the most favorable disposal option. EPA estimates that land application will increase to 48% in 2010 compared to 41% in 1998 and 45% in 2005 (USEPA, 1999). In contrast, surface disposal and land filling will be reduced to 10% in 2010 compared to 17% in 1998 and 13% in 2005 (USEPA, 1999). As land application will be the major disposal option for the next few years, this research focuses on the impact of land application on communities where biosolids are applied to land, especially in terms of biosolids odor. Odor nuisance from biosolids is the biggest apparent problem with land application. Even though the USEPA ensures that odors from properly stabilized biosolids leave no threat to human health, the odor nuisances associated with land application are sometimes unbearable. Opposition to land application usually occurs at application sites located near communities or at communities close to the hauling route. Complaints can subsequently

increase restrictions on the land application program or eventually resulting ban on the land application in that county.

In 1995, the survey results by the Water Environment Federation (WEF) showed that 41% of state regulators were concerned about odors from biosolids application, 21.7% about health, 17% about nuisance, 6.5% about appearance, 8.7% about transportation, 2% about noise, and 2% about other issues (WEF, 1997). Thus, as noted earlier, biosolids odors are the most significant and urgent problem needing to be resolved by biosolids producers and associated parties.

1.2.1 Biosolids odor management: Biosolids odor predicting models

A number of studies addressed odor problems generated from wastewater treatment plants and their products. For instance, odor dispersion models were widely used to assess the odor footprint at the communities around the odor sources. Williams and Servo (2005) applied biosolids odor dispersion models to identify the impact of odors from composting facilities on the surrounding neighbors; Sarkar, Longhurst, and Hobbs (2003) used dispersion models to predict the impact of odor from solids waste landfill on neighborhood; and Voelz et. al. (2006) used odor dispersion models at a wastewater treatment plant to find the sources of odors onsite that most impact the surrounding communities. These studies considered the emission rate of the odor sources, geographical data, and meteorological data as inputs into odor dispersion models and generated an odor footprint around the sites.

At the wastewater treatment plant, researchers also investigated variables that were anticipated to influence biosolids odor emissions and suggested methods to reduce biosolids odors. Subramanian et. al. (2005) investigated the roles of process conditions on

biosolids odor emissions. Chemicals such as potassium ferrates (VI) (Luca, Idle, and Chao, 1996) and coal ash (Rynk and Goldstein, 2003) were proposed as alternatives to help reducing biosolids odor production.

In the field of environmental engineering, statistical modeling has been widely used to explain the system of interest. For example, Greenberg et al. (1973) used statistical modeling to assess water quality in terms of Dissolve Oxygen (DO) in a free flowing river system. In wastewater treatment, statistical models were also used to optimize the dewatering process in selecting polyelectrolyte type and dose for better dewatering results (Saveyn et al., 2008). Statistical methods were used to predict algae biomass measured by chlorophyll a in Lake Okeechobee, FL: using an ordinary least squares (OLS) model (Lamon, 1995), a generalized additive model (Lamon et al., 1996), and a regression spline model (Lamon and Clyde, 2000).

To the best of our knowledge, there have only been two articles (Gabriel et al., 2005 and 2006) attempting to model biosolids odor production. However, neither of them has initiated the idea of connecting the odor emitted at the wastewater treatment plant and the biosolids emitted at application sites.

In terms of predicting biosolids odor emissions using statistical models, Gabriel et al. (2006) and Vilalai (2003) have conducted research on biosolids odor monitoring and management models at the Blue Plains Wastewater Treatment plant, located in Washington, DC and operated by the District of Columbia Water and Sewer Authority (DCWASA). This idea was to develop biosolids odor forecasting models by using variables related to biosolids odor production as predictors. In the first work (Gabriel et at., 2006), models were developed to explain biosolids odors. Inspectors used their

unaided noses to quantify biosolids field odor levels and inspectors designated the odor level as normal (0), slight (3), medium (6), and high (9). The field odor forecasting models then used the average of an individual inspector's field odor scores as the response variable. Additionally, the ambient conditions at the wastewater treatment plant, such as temperature, rain fall, snow condition and process parameters on the application day and the day before were investigated to check their influence on odor production at the field sites. Significant parameters included in the final models were: the sludge blanket level, the amount of lime additions, the amount of polymer additions at dewatering and Dissolve Air Floatation (DAF) processes, and the blend ratio.

Additionally, Gabriel et al. (2005) modeled the emission of Dimethyl disulfide (DMDS), an odorous chemical compound from biosolids, using sludge characteristics, such as Oxidation Reduction Potential (ORP), sludge temperature and process parameters at the Blue Plains facility as explanatory variables. The models indicated that ORP, the blend ratio, and the number of centrifuges running were significant parameters contributing to DMDS emissions.

1.3 Difference from previous work

The research in this thesis is different from the previous research (Gabriel, et. al., 2005; Gabriel et. al., 2006; Vilalai, 2003) in that it connects biosolids odor at the Blue Plains Wastewater Treatment Plant assessed by two odor measurement techniques: sensory measurement (i.e. human olfactory) and analytical measurement (i.e. hydrogen sulfide analyzer) with biosolids odor detected at the field site assessed by field inspector (i.e. using naked nose and using machine called olfactometer). Compared to Gabriel et al. (2006) where field odors were assessed only by the field inspector's unaided nose and no

biosolids odor data were collected at Blue Plains, supplementing biosolids odor data in this study support biosolids odor assessment in various aspects. Additionally, we developed models that tied together the biosolids odor data at the plant site with the field odor data in conjunction with external factors, such as field weather conditions to explain the variation in plant and field odors from biosolids.

Compared to Gabriel et al. (2005), this research goes beyond the idea of modeling specific odorous compounds from biosolids (i.e., DMDS) using plant-specific factors. Assessing Blue Plains'biosolids odor levels by the two approaches mentioned above provides supplemental information for more realistic biosolids odor prediction models. A better biosolids odor prediction model will be a valuable tool for biosolids managers to assess and control odor-causing factors. As a result, we can reduce the possibility of distributing odorous products to field sites.

Last, the sensitivity analysis and simulation approach in this research not only bring the uncertainty that exists in real world application into the model but also identify how sensitive each independent variable on biosolids odor emission under uncertainty in daily application.

In this thesis, data from the Blue Plains Advanced Wastewater Treatment Plant were used. Blue Plains produces approximately 1200 dry tons of biosolids per day and assigns class B biosolids for land application to field sites in Maryland and Virginia. This research helps biosolids odor management at Blue Plains to build an in-house biosolids odor monitoring and forecasting model using data that contribute to odor emissions at the wastewater treatment plant as explanatory variables. In addition, knowing the original odor concentrations at Blue Plains, we investigate the external factors off the plant

contributing to the odor problem at the field site. Figure 1.1 illustrates an overview of the research.

Figure 1.1: Overview of biosolids odor modeling
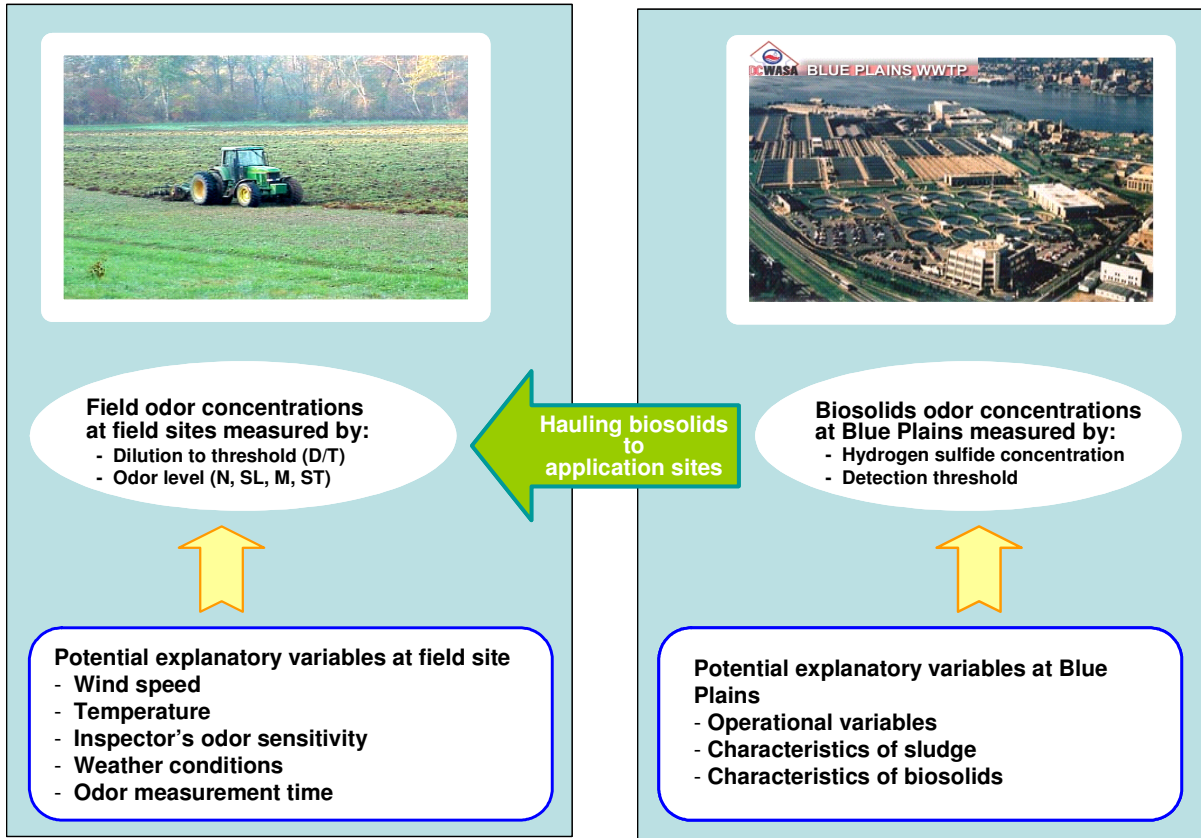
Blue Plains as well as other similar types of biosolids producers will benefit in managing odorous products to reuse sites and in controlling the odor-causing factors at the plant site. Statistical modeling techniques (i.e. linear regression, discriminant analysis, logistic regression, ordered logit model, validate model etc.) developed here provide a guideline for other wastewater treatment plants to conduct their own research using the techniques presented here.

1.4 Research Objectives

1. To identify the variables contributing to biosolids odor emissions at the plant site using odor data from sensory and analytical odor measuring techniques.

2. To develop on-site biosolids odor predicting models that can explain the production of biosolids odors.

3. To investigate the impact of the external factors, such as field condition, hauling distance, field inspectors on odor emissions at field sites.

4. To develop field site biosolids odor predicting models.

5. To investigate how the uncertainties in the predictors of odor predicting model affects the biosolids odor levels and develop probability distributions for biosolids odor levels.

1.5 Executive summary

There were two hypotheses to be tested in the study.

Hypothesis#1: There are only a key set of properties of biosolids (i.e., temperature and percent solids) and wastewater sludge (i.e., temperature, pH, ORP, and odor emissions) that influence biosolids odor production.

Hypothesis#2: Wastewater and solids operations at the plant (e.g., sludge blanket level and concentration of returned activated sludge (RAS) at the secondary process, percentage of GT solids in the blend tank, etc) contribute to the variation in biosolids odor emissions.

Results from analyses

Hypothesis # 1 was confirmed. Properties of biosolids including percent solids and temperature were significant in the selected model. Both variables are based on the final product from the solids treatment process and biosolids odor data were directly measured from these samples. In terms of properties of wastewater sludge, pH of sludge at the gravity thickener tanks (GT pH) was statistically significant and explained biosolids odor levels. Other variables, such as ORP, temperature, and hydrogen sulfide emissions were important in some subsets of the data collected. However, %GT in the blend tank and RAS east were able to explain the variation of biosolids odor better than properties of sludge in the sludge blending system. This may be due to how the data were collected. Those data from operational variables were the averages of data collected for an entire day while data for the properties of sludge were measured from samples of sludge from sludge blending system that were being collected at a single point in time. Due to the sampling procedure the data collected might not completely represent the actual properties of sludge in the sludge blending system. Another approach to improve data collection would be to have a real time monitoring of the sludge propeties in the blending system.

Hypothesis #2 was confirmed. Wastewater and solids operations at the plant (i.e., concentrations of RAS, %GT in the blend tank, and number of centrifuges in service) contributed to biosolids odor emissions as shown by significant variables in linear and categorical biosolids odor prediction models.

Contributions of this research

Various types of statistical modeling were used in this thesis to improve biosolids odor management. Using the final models in this study, biosolids odor can be described in two ways. First, odor levels can be understood as a continuous response variable (detection threshold) using linear statistical models as well as a categorical response variable (high and low biosolids odor classes) using discriminant analysis and logistic regression. Both types of models can assist biosolids managers to determine the potential odor level that a plant such as Blue Plains might produce each day. Blue Plains will benefit from these models by knowing in advance where to send the biosolids, for example, sending malodorous product to remote fields to avoid complaints or sending less smelly product to closer fields to reduce hauling costs.

In addition, the models developed not only describe biosolids odor levels but also can identify key influential factors at a plant such as Blue Plains. New significant variables identified in this study (in addition to Gabriel et al, 2005 and 2006) were the concentration of waste activated sludge at the secondary east process (RAS east), pH of gravity thickener solids (GT pH) at the gravity thickener tanks, and the percent solids and temperature of biosolids. RAS was believed to better explain the sludge conditions (e.g., anaerobic conditions) at the bottom of the secondary sedimentation tanks compared to the sludge blanket level discussed in Gabriel et al. (2005 and 2006). GT pH indicated the condition of sludge at the primary process before blending with sludge from dissolved air floatation (DAF) in the blend tank. Lastly, the percent solids and temperature of biosolids described the properties of the final product resulting from operations in the upstream processes. Other significant variables were the percentage of the gravity thickener sludge

in the blend tank (%GT in the blend tank) and the number of centrifuges running. Without the biosolids odor prediction models, it would be hard to identify the factors that jointly play significant roles in biosolids odor production. The methods in this study can aid management at Blue Plains and other advanced wastewater treatment plants in monitoring hard-to-control variables, such as the GT pH, RAS concentration on the east side, temperature and percent solids of biosolids and to make some adjustments to the controllable variables, such %GT in the blend tank and number of centrifuges running. In addition, maintenance of equipment in the sludge blending system and dewatering process must be a first priority to guarantee the optimal operation from these processes.

Another contribution from this study is the computation of the probability distribution of biosolids odor developed by simulation techniques. Blue Plains or other advanced wastewater treatment plants can assess probabilities for producing specified biosolids odor levels based on current operations at the plant. For example, based on the current operations at the Blue Plains plant, there was 49.7 percent chance of producing malodorous product (higher than 1000 ou). Management can use results from a sensitivity analysis study to control some decision variables such as number of centrifuges running or percentage of the GT solids in the blend tank in order to reduce biosolids odor levels below 100 ou on average. For instance, from a sensitivity analysis on the number of centrifuges in service, it was shown that running nine centrifuges and higher would give a probability of producing lower odor biosolids (lower than 1000 ou) at less than 50 percent.

Lastly, wastewater and solids treatment at Blue Plains and field odor observations were connected by field odor model using ordered logit regression model. It was

observed that the ordinal biosolids odor levels at the field site can be explained by the biosolids odor levels at the plant at the 1000 ou threshold, temperature at the field site, and wind speed especially when it was greater than 13 mph. This information helps the hauling contractor and field inspectors to choose the most appropriate sites to send the product.

In summary, contributions from this study are the various tools from statistical modeling, simulation, and sensitivity analysis techniques to improve biosolids odor management. At the plant, the developed models determined the potential biosolids odor produced each day as well as identified the key factors related to biosolids odor emissions. Simulation and sensitivity analysis considered the realistic aspect of uncertainty in the operation and conducted the scenario analysis for the outcomes of biosolids odor level based on different scenarios. At the field site, field odor modeling gave guidelines to the hauling contractors and field inspectors in selecting sites using predicted biosolids odor levels on that day, wind speed, and temperature. Such analyses can be implemented in other wastewater treatment plants.

1.6 Organization of thesis

The rest of this thesis is organized as follows:

Chapter 2: Literature Review summarizes articles and research related to biosolids odor production, reduction, and biosolids odor management.

Chapter 3: Data Collection describes the wastewater treatment process, data collection process, and data used in this study.

Chapter 4: Biosolids Odor Prediction Models for Wastewater Treatment Plant describes biosolids odor predicting models developed using data from the Blue Plains

Advanced Wastewater Treatment Plant. This chapter includes the modeling process, validation process, and discussion of significant predictors in the model.

Chapter 5: Simulation and Sensitivity Analysis conducts a simulation and sensitivity analysis to produce probability distributions for biosolids odors based on distributions from explanatory factors. Additionally, this chapter summarizes suggestions to wastewater treatment management on how to implement odor management strategies.

Chapter 6: Biosolids Odor Models at Field Sites analyzes the impact of field conditions, inspector odor sensitivity, and original odor at Blue Plains on odor emissions at field sites.

Chapter 7: Conclusion and Future works.

Chapter 2: Background on Biosolids Odors

With respect to its nutrient-rich properties, biosolids can greatly benefit communities if the product is recycled for agricultural purposes. Often, odors from biosolids mask the benefits. Even though the processes to produce biosolids are monitored and regulated by EPA for quality assurance, there are still complaints about odors from the farmers and residents near application sites.

The wastewater treatment plant generating biosolids is responsible for assuring the quality of biosolids before sending from the plant. EPA lists significant factors influencing biosolids odor emissions as follows: the variation in wastewater influent characteristics, type of polymer used in the process, blending primary and secondary solids before dewatering, quality of lime mixed into the product, type of lime used, and storage time (USEPA, 2000). The attempt to monitor a number of odor-causing parameters, both manageable, such as, polymer type and unmanageable, such as, the influent characteristics, is challenging.

A moderate biosolids odor is expected when all equipment in the process are functioning normally. However, in reality, biosolids management faces various operational challenges. For instance, pump breakdowns can cause a lack of polymer in the DAF and dewatering processes, pump malfunctions cause a longer retention time for blended sludge to be stored in the blend tank; an insufficient number of centrifuges in service cause an overload of sludge for each centrifuge; chemical shortages lead to insufficient lime or polymer mixed in the cake, etc. These unexpected incidents leave room for biosolids odor monitoring and forecasting models to identify important variables and give warning to biosolids producers when odorous products are anticipated.

To understand the biosolids odor problem, we first discuss the types of odorous compounds generated from biosolids.

2.1 Chemical Compounds Causing Biosolids Odors

Understanding the formation of biosolids and knowing the types of chemical compounds generated from them is helpful to biosolids odor management. Odors in wastewater treatment are typically the product of biological degradation of constituents in the wastewater under anaerobic conditions that generate various odorous compounds, which we also called odorant (Frechen, 1988). The distinction between odor and odorant is that the odorant is the chemical compound causing the odor while odor is the interpretation of odorants by human olfactory senses (Gostelow, Parsons, and Stuetz, 2001).

There are typically three groups of odorants that people recognize from biosolids odor: amine, ammonia, and compounds containing reduced sulfur.

Ammonia: ammonia is a typical odorant found from lime-stabilized biosolids. It has a high detection threshold compared to other biosolids odorants, meaning that it takes more ammonia to be detectable by human compared to amine and reduced sulfur compound (USEPA, 2000).

Reduced sulfur-containing compounds: Reduced sulfur-containing compounds from biosolids are hydrogen sulfide, dimethyl sulfide (DMS), and dimethyl disulfide (DMDS). Hydrogen sulfide is an inorganic sulfur compound noted by a rotten egg smell. It can be observed from biosolids with pH less than 9 and normally disappears after lime addition when pH is greater than 9 (USEPA, 2000). DMS and DMDS are the by-products of chemical and microbial degradation of protein.

Amine: amine compounds emitted from biosolids are methylamine, ethylamine, trimethylamine, and diethylamine. They are the result of microbial decomposition of proteins and can be detected with temperature greater than 27 degree celsius (USEPA, 2000). As mentioned above ammonia has a high DT than the other biosolids odor. However, there is occasionally evidence that amine and reduced sulfur compounds from lime-stabilized biosolids are masked by the high intensity of ammonia compound. After biosolids are diluted in the field, the real odor problem usually comes from amine and reduced sulfur compounds that are more persistent and have very low DT (USEPA, 2000).

To access the odor problem, the appropriate odor measurement method is important. Next we discuss odor measurement options.

2.2 Odor Measurement Options

To control odor, measurement methods need to be selected. There are two odor measurement techniques; sensory and analytical. The analytical approach measures the physical concentration of odorants while the sensory method measures odor perceived by human olfactory senses (Gostelow, Parsons, and Stuetz, 2001). Both measurement techniques have advantages and disadvantages. The advantages of analytical measurement are: objectivity, repeatability, and precision measurement (Gostelow, Parsons, and Stuetz, 2001). The disadvantage of analytical measurement is due to the complex nature of odor generation and odor perception that may not be able to describe by knowing concentrations of only a few odorants. Sometimes an odorant with less concentration can mask other odorants because of its lower detection threshold. The analytical measurement technique is conducted by an odor measurement device, such as

gas chromatography or H_2S analyzer. Kim et. al. (2001) applied analytical measurement techniques called Solids Phase Micro Extraction (SPME) to quantify various odorous compounds, such as TMA, dimethyle sulfide (DMS), dimethyle disulfide (DMDS), etc. using gas chromatography. This method gives precise concentrations and repeatable results. In addition, it is more convenient and inexpensive when compared with the olfactory method. However, it takes time to calibrate the machine and there must be gas chromatography equipment in house (Kim et. al., 2001).

The other odor measurement technique is sensory measurement. This method uses the human nose to assess the odor and describe it by detection to threshold, recognition to threshold, intensity, hedonic tone, and odor character. The advantage of sensory measurements is the use of the human olfactory senses to perceive odorants as a whole. The data from sensory measurement can represent what residents nearby the application sites might perceive. The disadvantages are the subjectivity of odor assessors and the cost associated with training panelists or odor evaluation fees from a professional odor evaluation company.

Next we summarize the odor causing factors found in the literature.

2.3 Odor Causing Factors and Odor Reduction Methods

In addition to factors influencing biosolids odor suggested in EPA's Biosolids Field Storage guide (USEPA, 2000), many facilities involved with biosolids production conduct their research to minimize biosolids odor problems. For example, The Montgomery County Regional Composting Facility (MCRCF) in Maryland succeeds in using wood ash in addition to lime to reduce the Volatile Organic Compounds (VOC) emissions. They also found that applying lime to the biosoids before dewatering can help

reduce biosolids odors (Biocycle, March 1999). The city of Takoma in Maryland, Wastewater Utility, that produces class A biosolids found that using a lower temperature from 130 F to 90 F in anaerobic digesters can eliminate odors (Thompson, 2004). A wastewater treatment plant in Brazil proposes the use of potassium ferrate(VI) as an alternative chemical to reduce odor mostly coming from sulfide and ammonia (Luca, et al., 1996) Sabramanian, 2005 found that shear from dewatering equipment causes protein and polymer in sludge breakdown the resulting TMA and sulfur compounds. Considering the dewatering equipment, Rynk and Goldstein (2003) found that biosolids dewatered by the plate-and-frame-press system produces less odor emission in terms of dilutions to threshold than biosolids dewatered by a solid-bowl centrifuge. Furthermore, they concluded that a new centrifuge produces greater odor than an old centrifuge.

External factors outside the plant can also contribute to adverse perceptions of biosolids application programs. For example, meteorological conditions at the field site, such as, wind speed, wind direction, cloud conditions, relative humidity, and temperature all affect odor dispersion and odor perception in nearby communities (USEPA, 2000). Calm conditions, such as warm weather and high humidity increase the possibility of odor complaints. Also, odor complaints usually happen in the early morning or at night (USEPA, 2000). Additionally, topology selection can reduce the potential of odor complaints. For instance, flat terrain with a wind speed of 8-12 mph will create moderate turbulence to dilute biosolids odor at the site whereas application in a valley will block air flow leading to odor complaints (USEPA, 2000). Rynk and Goldstein (2003), suggest to apply biosolids in the morning to let the sun dry the product and to apply as thinly as possible. EPA advises biosolids producers to prepare contingency plans to dispose of

biosolids according to odor levels. For example, incinerate malodorous product or use it for land filling, dispose of product with moderate odor to well-buffered sites or surface injection, and apply the best product to sensitive areas (USEPA, 2000). To control biosolids odor, the city of Philadelphia applied coal ash to biosolids before land application (Rynk and Goldstein, 2003). The choice of hauling routes and truck conditions can also contribute to negative impact in the communities along the hauling route. EPA suggests avoiding hauling biosolids through densely-populated area and cleaning the truck before leaving the plant and after dumping its load (USEPA, 2000).

2.4 Biosolids Odor Management: Biosolids Odor Research at Blue Plains

A number of extensive research projects have been conducted at the Blue Plains Advanced Wastewater Treatment Plant to understand the causes and characteristics of biosolids odor production. Kim et al. (2003) found that polymer addition in DAF process contributes to the production of Trimethylamine (TMA), the fishy smelling compound, from lime stabilized biosolids. Murthy et al. (2001) found that overdosing polymer results in odor emissions, especially amine compounds, over a month of storage. Lime doses at 20-25 percent of percent solids coupled with reduction of polymer addition were suggested as ways to reduce biosolids odor. Rynk and Goldstein (2003) and Aripse (2005a) found that the amount of cationic iron and aluminum left in the lime-stabilized solids appears to have a negative correlation with odor production whereas those materials have a positive correlation on unlimed biosolids. Considering the dewatering equipment, Murthy (2003) concluded that Volatile Sulfur Compounds (VSC) emitted from the centrifuge-dewatered solids is the result of shearing from the solids bowl centrifuge causing the break-down protein available for biodegradation over time.

The secondary process, which is biologically-based, is one of the major odor sources at wastewater treatment plants. Sekyiamah (2004) conducted research to find parameters contributing to odor emissions in the secondary process. His study showed that sludge blanket level in secondary sedimentation basins at Blue Plains had a positive correlation with VSC production in the secondary treatment system. As a result, he recommended reducing retention time that allows solids accumulation at the bottom of the tanks by wasting and returning sludge as quickly as possible to reduce the anaerobic condition at the bottom of sedimentation tanks. In the case of sludge characteristics, strong evidence from Aripse et al. (2005a) showed that ORP of solids in the thickening process is negatively correlated with the production of reduced sulfur compounds. This can be interpreted as the greater the ORP value the lower the reduced sulfur compound emissions.

North (2003) conducted extensive work on the adequacy of lime incorporation into biosolids. He summarized that adding insufficient lime causes biosolids odor emission over time due to ongoing microbial activity and large quantities of pathogens left in the solids. He suggested several methods for detecting inadequate lime addition, such as, low ammonia odor present in the first 30 minutes of lime mixing, high odor production after a 24-hour period, and unmixed granular lime visible in the biosolids. Additionally, it was found that the percentage of solids in dewatered sludge prior to lime stabilization influences the capacity of blending dewatered solids with lime. The higher the solids content the longer the mixing time that is needed to reach optimal lime incorporation (North, 2003). Finally, the longer mixing time and the proper lime particle size, 0.25- 2 mm, are suggested to get the optimal lime incorporation.

Chapter 3 will describe the wastewater treatment process, data sampling procedures, and data description.

Chapter 3: Data Description

This chapter provides a background on unit operations at the Blue Plains Wastewater treatment plant, which is taken to be representative to some extent, for other wastewater treatment plants. It describes the sampling procedure and the data used in developing the model. This chapter is organized as follows: first we discuss the wastewater treatment and solids handling process at Blue Plains, then the sampling procedures, and finally we describe the data.

3.1 Wastewater Treatment and Solids Handling Processes

Blue Plains is the largest advanced wastewater treatment plant in the world functioned with wastewater treatment processes and solids handling processes. The wastewater treatment processes include a preliminary process, a primary process, a secondary process, a nitrification-denitrification process, multimedia filtration, and disinfection. The solids handling processes treat solids separated from the liquid process through gravity thickeners, dissolved air flotation, a blend tank, centrifuge dewatering, and lime stabilization. Figure 3.1 give an overview of the wastewater treatment and solids handling processes at Blue Plains. Each process is described in detail below (DCWASA, 2008).

Figure 3.1: Wastewater and Solids Handling Process
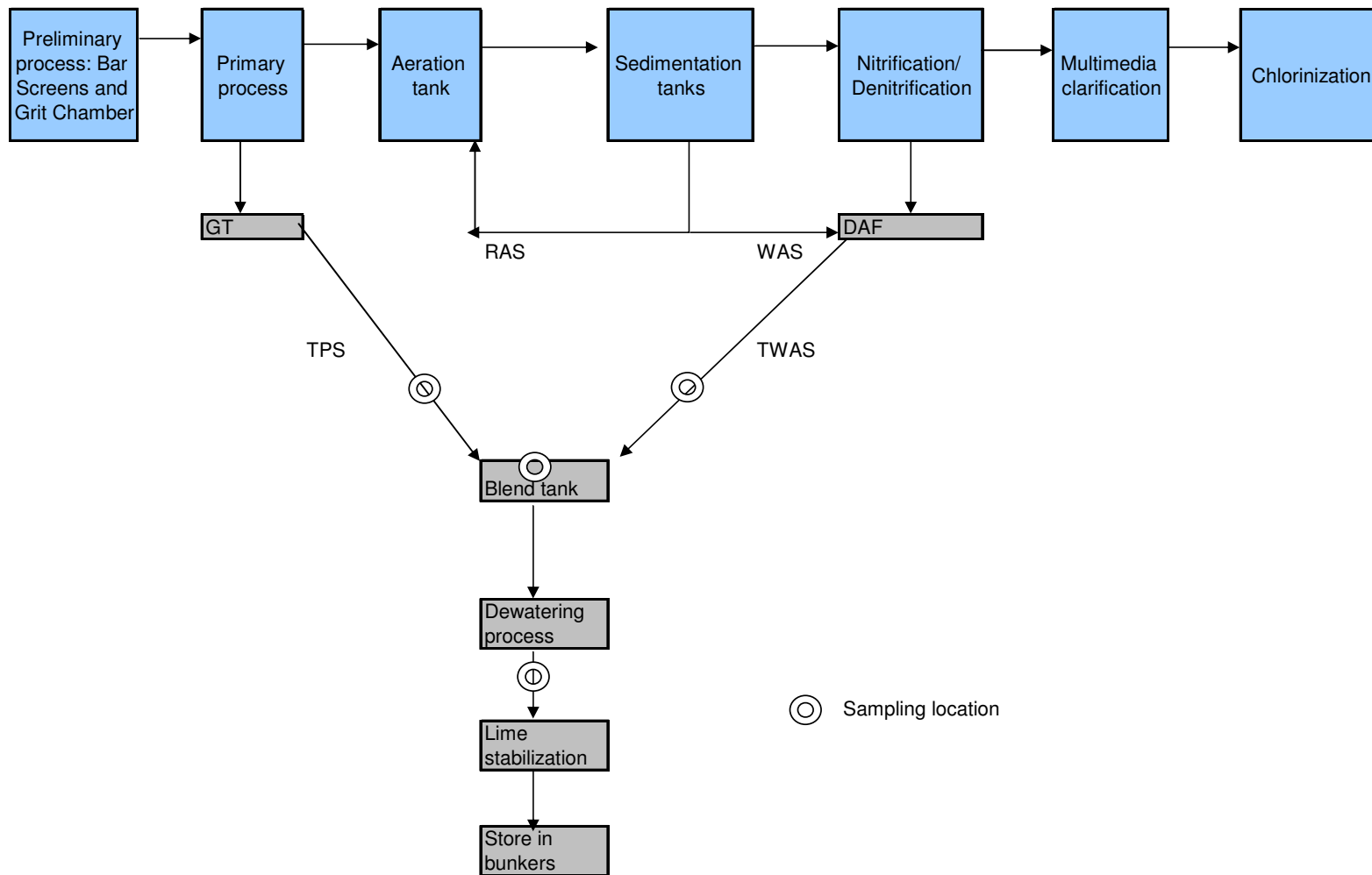
Preliminary process

Everyday approximately 370 million gallons of wastewater are processed at Blue Plains. The preliminary process employs bar screens to remove large particles such as trash and debris that can cause damage to pumps and pipelines in the subsequent processes. Iron salt ($FeCl_3$) is added for phosphorous removal before the flow proceeds to the grit chamber. In the grit chamber, grit, typically sand and silt particles, is removed to prevent abrasion in the pumps downstream and to prevent grit accumulation in the aeration tank and sludge pipe.

Primary process: primary sedimentation tanks

The primary process separates liquid and suspended solids in the incoming wastewater. The flow slows down at the primary sedimentation tanks. Organic suspended solids and chemical precipitates, such as phosphorous settle to the bottom of the tanks. Scum floats to the surface of the tanks. Then, the scum and settling solids are sent to gravity thickened tanks (GT) where solids content is increased by gravity. The thickened solids at these tanks are finally pumped to a blend tank.

Secondary process

The secondary process uses biological activity to remove organic materials left from the primary treatment. The effluent from the primary process passes to secondary aeration reactors where the flow is mixed with returned activated sludge (RAS) from the secondary sedimentation tanks and Waste Pickle Liquor (WPL), another chemical added for phosphorous removal. The anaerobic conditions in the aeration tanks increase the growth rate of microorganisms in helping to remove suspended solids, colloid carbon, and phosphorous from the stream. Then the flow passes to sedimentation tanks where

activated biological solids are separated from the liquid and go to the bottom of the tank by gravity. The level of solids built up in the sedimentation tank is called the blanket depth (BD) or sludge blanket level. The Return Activated Sludge, in the settling tank is returned to the aeration tank to maintain the microorganism concentration in the reactor. The other part of the solids, called Waste Activated Sludge (WAS), is wasted to dissolve air flotation tank.

Nitrification and denitrification

Nitrification and denitrification processes use biochemical process as to convert ammonia and organic nitrogen in the wastewater to nitrogen gas. Similar to the secondary process, some of the settling solids in the sedimentation tanks are returned to the reactors and the rest are wasted to the DAF process. The treated water then passes through filtration where fine particles and phosphorous are removed. Treated water is disinfected prior to discharge into the Potomac River.

Dissolve air flotation

The DAF process receives the solids from secondary sedimentation, nitrification, and denitrification processes. All of the solids are waste activated sludge that are difficult to settle by gravity. Polymer additions help to capture small solid particles in the tanks coupled with compressed air bubbles from the bottom of the tank to carry these small particles to the surface. The chain pad removes the floating particles that are subsequently pumped to the blend tank. Research shows that polymer addition in DAF contributes to the odor generation from the lime-stabilized biosolids (Kim, 2003).

Blend tank

The blend tank is the location where the primary sludge, sludge from gravity thickened tank, and waste activated sludge, sludge from DAF, are blended with different blend ratios depending on the operating conditions. At Blue Plains, the total sludge flow from gravity thickened tanks to the blend tank, called Total Primary Sludge (TPS), and the total sludge flow from DAF to the blend tank, called Total Waste Activated Sludge (TWAS), can be used to calculate blend ratio inside the tank. The target blend ratio is 50:50 since it facilitates dewatering equipment (private communication with DCWASA operator, 2005). However, the actual blend ratio depends on the processing conditions, such as the amount of solids available at DAF and gravity thickener to feed to the blend tank on that day. Note that the blend ratio and the retention time of blend solids in the tank are considered as factors that can cause the biosolids odor in this study (USEPA, 2000b). Blend sludge (BS) is finally transferred to the dewatering process for water removal.

Dewatering and lime stabilization

Prior to dewatering, polymer is added into blended solids to improve the dewatering capacity. Fourteen high-speed centrifuges are employed at Blue Plains to remove water content from blend solids. The better dewatering capacity of centrifuges benefits Blue Plains by reducing hauling weight and hauling cost. Lime (CaO) is added for pathogen reduction and pH increase as regulated by EPA. At the end, biosolids are stored in the bunkers and await hauling to application sites.

3.2 Laboratory Data Sampling Procedure

This section describes the sampling procedures that were conducted from April 2005 to July 2006 to assess characteristics of sludge prior to dewatering as well as to assess odor production from biosolids with and without lime addition.

Two days a week between April 2005 and July 2006, sludge samples from the GT, DAF, and BS as well as dewatered solids samples were collected and processed at the Blue Plains' laboratory except during December 2005 to January 2006 when the project was awaiting budgetary approval by DCWASA. The following diagram illustrates the sampling locations where sludge and dewatered solids samples were collected.

Figure 3.2: sampling locations at Solids Handling Process.
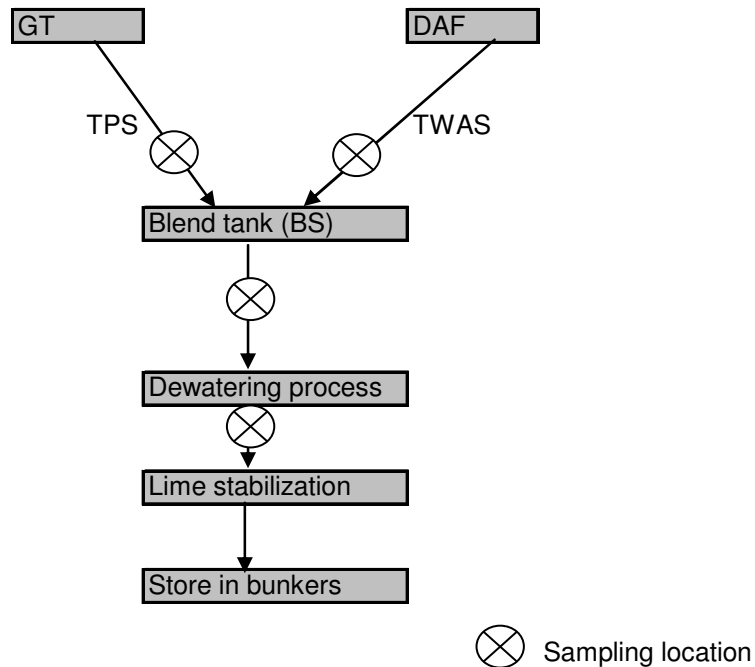
Sample collection: one-gallon samples of sludge each from GT, DAF, and BS were collected from sludge sampling sinks located at the solids processing building.

Feeding rates (gpm) of sludge from the GT and from the DAF pumped into the blend tank were recorded to assess the blending condition in the blend tank prior to dewatering. At the dewatering process, two gallons of pre-limed dewatered cake were collected from the conveyer. All samples were transferred to the Blue Plains' laboratory for processing.

Sample processing and measurements: At the laboratory, two 400-gram sludge samples from each location were placed into two 1-liter Teflon jars, using one jar as duplicate. The sludge samples were stirred occasionally before measuring pH, ORP, and temperature (C) from each sample using a pH/conductivity meter (Accumet AR-50, Fisher Scientific, Pittsburgh, PA). After that, all Teflon jars were closed for 30 minutes. Then, the H_2S concentration (parts per million volume: ppmv) in the headspace of each jar was measured using a Jerome 631x Hydrogen Sulfide Analyzer (Arizona Instrument, Phoenix, AZ).

The dewatered solids sample was first analyzed for the percent solids content using a Computrac Max 2000 XL Moisture/Solids Analyzer (Arizona Instrument, Phoenix, AZ). Then a mixture was prepared by adding lime by 15 percent of solids content. (Note: 15-percent is the actual target rate applied in the process, (Ramirez, 2005)). Two 1000-gram pre-limed samples were placed into two mixing bowls, five-quart heavy-duty dough mixers (Kitchen Aid Heavy Duty Mixer, model K5SS; St. Joseph, MI). Lime was added by 15 percent of the solids content into one bowl. The other bowl receiving no lime was the control. Both mixer bowls were turned on for three minutes. The bowl with lime addition was used to simulate the actual lime addition in the lime mixing process onsite and the bowl without lime addition was used for comparison purposes. After mixing, both samples were allowed to cool down to room temperature,

29

approximately two hours, then two 400-gram samples from each bowl were transferred to two 1-liter Teflon jars. The jars were closed for 30 minutes, after which the headspace H_2S concentration was measured by the Jerome 631X.

Air sample collections for sensory evaluation: The following procedures were used to draw headspace air samples from each Teflon jar into a bag and the equipment used in the procedures are shown in Figure 3.3.

Figure 3.3: Illustration of the vacuum chamber, Tedlar bag, and air pump (SKC, 2006)

First the 10-liter tedlar bag (Environmental Sampling Supply, Oakland, Ca) was connected to the sample valve inside a vacuum chamber. Then, the tedlar bag valve was opened and the vacuum chamber closed. The Teflon jar containing the dewatered sludge sample was connected to the vacuum chamber by the sample line. The pump was turned on at a rate of two liters per minute to create a vacuum inside the chamber for three minutes. With the negative air pressure inside the vacuum chamber, the air sample was drawn into the sample bag to about half full. This odorous air sample was to prepare the

sample bag prior to taking the actual odorous air sample. Then the bag was empty and the air sample was drawn into the bag again for five minutes or until the bag was 80 percent full. This air sample was used for odor evaluation. At a five minute mark, the sampling valve was closed, and the tedlar bag was labeled for delivery. These air sampling procedures were applied to each sample (with lime and without lime addition). When both sample bags were ready, they were packed into a shipping box and sent to an odor evaluation company by overnight shipping for sensory measurements.

At the end of a sampling day, both jars, with and without lime addition, were placed inside an incubator for 24 hours. After 24 hours, the H_2S measurement and the air sampling procedures again were run on both limed and unlimed samples and finally the air-sample box was shipped to an odor evaluation company.

3.3 Data Collection

We collected data from two locations, at Blue Plains Advanced Wastewater Treatment Plant and at biosolids application sites (field sites). Data collected from each location can be categorized either as dependent variables or independent variables. A dependent variable or response variable has values we are interested in explaining on its relationship to independent variables.

At the Blue Plains plant, dependent variable is the odor data generated from biosolids. Independent variables are the variables expected to help describe the variation in biosolids odor concentrations. These dependent variables were obtained from operational data at the secondary process, the sludge blending system, from characteristics of sludge in the sludge blending system and from characteristics of biosolids. Figure 3.4 illustrates the Blue Plains data.

| Dependent variables | Independent variables |

Dependent variables

| **Biosolids odor data**
Sensory measurement
Analytical measurement |

Independent variables

| **Operational data from secondary processes** |

| **Operatoinal data from the sludge blending system** |

| **Characteristics of sludge in the sludge blending system** |

| **Characteristics of biosolids** |
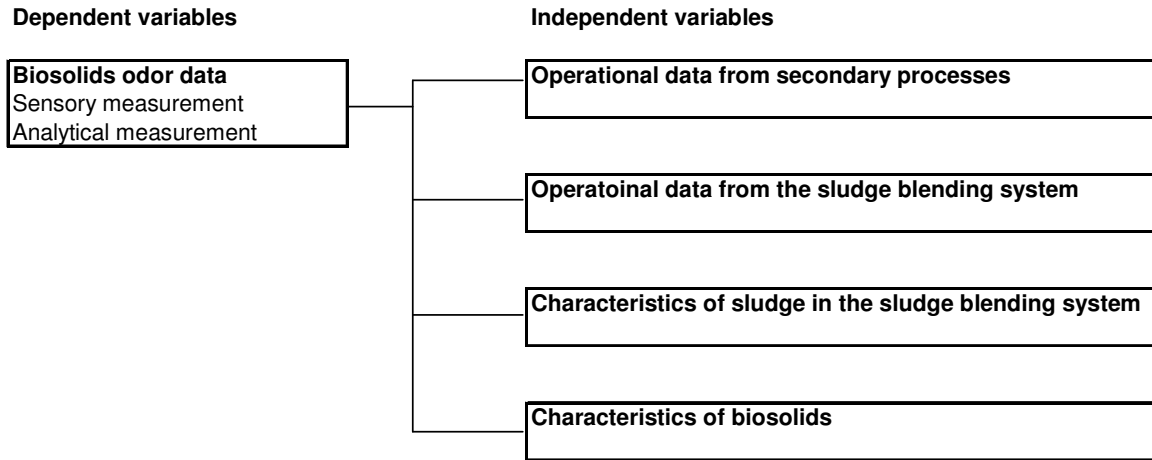
Figure 3.4: Dependent and independent variables at Blue Plains

3.3.1 Blue Plains' Biosolids Odor Data

Two types of odor measurement techniques were employed on dewatered solids samples with lime and without lime at three hours and at 24 hours after the dewatered solids samples were processed at the DCWASA laboratory. Table 3.1 summarizes the choice of dependent variables available in this study. All related to biosolids odor levels but obtained through two approaches. The first method, sensory measurement, used a set of experienced odor assessors (i.e., an odor panel), normally composed of six to ten individuals, to assess the biosolids odor information and describe the biosolids odor by detection threshold, recognition threshold, intensity, odor character, and hedonic tone. The second odor measurement technique was analytical measurement. We used the Jerome 631X, a handheld device, to measure hydrogen sulfide (H_2S) concentration in parts per million volume (ppmv).

Table 3.1: Blue Plains' biosolids odor data

Biosolids odor data	
Sensory measurement	**Analytical measurement**
Detection threshold	Hydrogen sulfide concentration
Recognition threshold	
Intensity	
Hedonic tone	
Odor character	

A.1 Odor data from sensory measurement

Detection threshold (DT;odor unit): Detection threshold is the diluted odor concentration where half of the panelists correctly detect the odor sample. Basically, in each round a panelist sniffs three air samples randomly presented from the sample ports of the device called an olfactometer.

Figure 3.5: Sensory measurement through the olfactometer (McGinley, 2005)

Among the three air samples presented, only one sample contains diluted odorous air while the other two samples contain odor-free air. The panelist is then forced to select one sample having a different odor from the other samples and notifies the test administrator as identifying a detection or a guess (McGinley, M. C., et al., 2002). This approach is called triangular forced-choice (ASTM E679-91). The test proceeds with a higher odor concentration in each round until the panelist correctly detects an odor sample. This approach is called an ascending concentration series. The test administrator

then summarizes the DT of the particular sample by averaging the individual thresholds identified by each panelist to determine the detection threshold level where half of the panelists correctly detect the odor sample. The unit of detection threshold is odor units (ou) or odor units per unit volume.

Recognition threshold (**RT;odor unit**): The RT is the diluted odor level where half of the panelists can recognize the character of the odor sample presented. The method for obtaining the RT is similar to that for the DT by using the same procedure as detection threshold but goes beyond the DT limit until the panelist can correctly recognize and describe the odor's character, i.e., a smell like something from one of the three samples presented (McGinley et al., 2000). At that level, the RT is recorded. The test administration then averages the RT of all the panelists and uses that average value as the RT of that particular odor sample. The unit of detection threshold is also odor units (ou) or odor units per unit volume.

Intensity (**ppm butanol**): Intensity is the relative strength of the odor above the recognition threshold (suprathreshold) as described in ASTM E544-75(1988). Using a continuous flow of standard odorant (butanol) from the olfactometer, the assessor compares the observed intensity of the odor sample to a specific concentration level of the standard odorant (n-butanol) from an olfactometer device. The result is given in as parts per million (ppm) of butanol (n-butanol).

Hedonic tone: Using a scale from 10 to -10 the hedonic tone measures the odor level from pleasantness (10) to unpleasantness (-10) of the odor with level zero as neutral. This value is subjective relying on experience and memories of odor of the assessors (McGinley et al., 2000).

Odor character: Odor character is described in terms of taste, sensation, and odor descriptors. Taste is divided into four categories; salty, sweet, bitter, and sour. Sensation is divided into eight categories; itching, tingling, warm, burning, pungent, sharp, cool, and metallic. Odor description can be categorized into seven recognized descriptors; vegetable, fruity, floral, medicinal, fishy, offensive, and earthy. The assessors choose each category (taste, sensation, and odor descriptors) and determine their intensities on a 0 to 5 scale (McGinley et al., 2000).

A.2 Odor data from analytical measurement

Hydrogen Sulfide (H_2S) ***concentration***: H_2S is the typical odorous compound found at domestic wastewater treatment plants. With a rotten smell, H_2S is the product of the decomposition of reduced sulfur reduction bacteria to reduce sulfate to H_2S gas as shown in the following chemical reaction.

$$SO_4^{-2} + \text{organic matter} \xrightarrow[bacteria]{anaerobic} S^{-2} + H_2O + CO_2 \tag{3.1}$$

$$S^{-2} + 2H^+ \longrightarrow H_2S \tag{3.2}$$
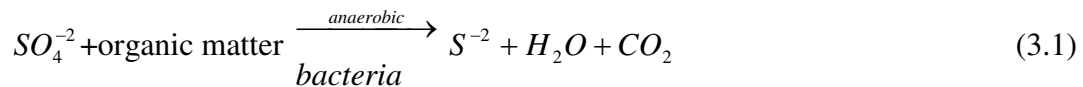
Since biosolids are the result of the processing of wastewater solids, the odor-causing substance separated from treated wastewater should continue releasing H_2S as well as other odorous compounds after sludge are dewatered. As suggested by Bowker et al. (1989) determining the conditions that favor production of H_2S should lead to the possibility of reducing production of other odorous compounds from wastewater. In this study, H_2S was selected to represent one of the major odorous compounds from biosolids. Knowing how to control H_2S emitted from biosolids would allow for controlling biosolids odor to some extent. We used the Jerome 631-x hydrogen sulfide

analyzer to measure H_2S concentration at the headspace of dewatered solids samples with and without lime after three and 24 hours. The detection range of this machine is between 0.03 – 50 parts per million volume.

3.3.2 Independent variables

Independent variables are variables that help explain the variation in the response or dependent variables, i.e., biosolids odors. The variables listed below are the anticipated odor-causing parameters according to the findings from previous researches and input from experts in the field (Kim et. al., 2004).

Operational data at Blue Plains

Operational data at Blue Plains were acquired from the Process Control History database (PCH), which are operating data recorded daily at the Blue Plains facility by DCWASA operators and stored values online by a DCWASA supervisor. This database uses the average of process data recorded in each shift during a day. The operating data from PCH that we selected for study were: 1) the operational data at the secondary process, 2) operational data at the sludge blending system.

Chemical additions change the properties of sludge and dewatered solids as mentioned in the literature review. The key chemicals added into the processes, such as polymer, iron salt (FeCl3), waste pickle liquid (WPL) were considered. However, due to the change in the chemical storage and feeding system and the retirement of a solid processing supervisor in 2005, none of the chemical usage parameters were consistently recorded until late 2006. Therefore, we could not include these chemical data in this study due to the reliability of data.

B.1 Operational data at the secondary process

Sludge blanket level (BD; in feet): Sludge blanket level measures the depth of solids built up in the secondary sedimentation tank. In the sedimentation tank, the wastewater solids settle down by gravity to the bottom of the tank. The operator could return the activated sludge (return activated sludge; RAS) back to the aeration tank to maintain the concentration of microorganisms in the reactor or waste the activated sludge to the DAF process (waste activated sludge; WAS). The decision depends on the operating conditions of the wastewater treatment process and solids handling process on that day, such as whether there is sufficient activated sludge at the reactor or a high sludge blanket level in the sedimentation tanks. Anaerobic conditions at the bottom of the sedimentation tanks and odor emissions at the secondary treatment process are considered to have a strong correlation with sludge blanket depth (Sekyiamah, 2005). The hypothesis is that the higher blanket level indicates a greater anaerobic condition in the waste activated sludge prior to waste to DAF and greater odor emissions from biosolids. At Blue Plains, there are three secondary treatment locations; East Side, West Odd, and West Even. The data collected covers the sampling period from April 2005 to July 2006, during which Blue Plains operated the East Side every month. Conversely, the West Even side was operated from April 2005 to October 17, 2005 and the West Odd side was operated from October 14, 2005 to July 2006.

Concentration of Return Activated Sludge volume (RAS; mg/l): RAS is the concentration of return activated sludge (mg/l) from the secondary sedimentation tank to the secondary reactor to maintain the concentration of microorganisms. The data available in this study were from both East Side RAS and West Side RAS. The high

concentration of RAS indicates the high density of the solids at the bottom of the secondary sedimentation tank that are finally wasted to DAF. High RAS concentration coupled with high sludge blanket level are assumed to influence odor emissions from biosolids due to the greater anaerobic conditions at the bottom of the sedimentation tank. The greater concentrations of RAS are expected to be positively correlated with biosolids odor production.

Flow temperature: The flow temperature can signal potential biological activity in the process. Up to 30 degree celsius, increase of one degree Celsius could expedite the sulfide production rate in wastewater seven percent (Bowker, et al. 1989). As shown by experiments at Blue Plains, low temperatures can slow microbial digestion during the winter as compared to the summer. We used flow temperature from the primary effluent to investigate seasonal effects. The temperature at this point should roughly indicate the biological activity that was occurring in the secondary process and subsequent processes.

B.2 Operational aspect of the sludge blending system

Total Waste Activated Sludge (TWAS(gal); gallons/day): Total waste activated sludge is the total gallons of waste activated sludge (DAF sludge) pumped from DAF tanks to the blend tank on that day. At the DAF, waste activated sludge from the secondary sedimentation tanks is mixed with polymer prior to pumping it into the blend tank. Total waste activated sludge is included in the study to determine if the total amount of waste activated sludge in the blend tank accounts for the biosolids odor production. Also, the additional flow rate data (gpm) from WAS to the blend tank on the sampling day was recorded to assess the blending conditions right before sludge samples were collected.

Total Primary Sludge (TPS(gal); gallons/day): Total primary sludge is the total gallons of sludge from the gravity thickener tanks (GT sludge) pumped into the blend tank on that day. Thickened primary solids are mostly organic solids that are biodegradable by microorganisms in the DAF sludge. This parameter is included to observe when the amount of organic material in the blend tank accounts for biosolids odor production. Also, the additional flow rate data (gpm) from the primary sludge (TPS) to the blend tank on the sampling day were recorded to assess the blending conditions right before sludge samples were collected.

Blend ratio (Blend ratio; %): Blend ratio measures a ratio of total primary sludge (gal/day) in the blend tank on that day. As illustrated, in the following formula:

Blend ratio = (total gallons of primary sludge fed to the blend tank per day) / (sum of total gallons of primary sludge and total gallons of waste activated sludge fed into blend tank per day).   (3.1)

In the sludge blending system, the total flow from the gravity thickener, mostly organic materials that are the food source for microorganisms, and the total flow from DAF, the microorganism, are combined in the blend tank. The blend ratio determines the proportion of food in the blend tank before dewatering and lime stabilizing. A high blend ratio means that there is sufficient food source for microbial activity which is expected to yield high biosolids odor emissions. In contrast, a low blend ratio means that there is insufficient food source with respect to microorganisms leading to low odor levels.

Additional data on the percent of primary sludge in the blend tank (TPS% in blend) is calculated from the flow rate of TPS (gpm) over the sum of flow rate from TPS and

TWAS using the data collected on the sampling day to assess the blending condition right before sludge samples were collected.

Percent of Gravity thickened solids in blend tank (GT% in blend; %): This parameter calculates the ratio of actual solids content (percent solids) of primary sludge (GT) in the blend tank. The percent solids of the GT sludge sample and DAF sludge sample were analyzed at the DCWASA laboratory daily. The formula for GT% in blend is the following:

GT% in blend = (percent solids of GT) / (sum of percent solids of GT sludge and percent solids of DAF). (3.2)

This parameter shows the proportion of actual percent solids of primary sludge with respect to the sum of the actual percent solids from DAF and GT.

Number of centrifuges running (#centrifuge in service): The number of centrifuges running indicates the amount of sludge dewatered per day and can indicate the dewatering load on each centrifuge as well as dewatering capacity. Managing the amount of centrifuges running with respect to the sludge volume needing to be dewatered is the key. In daily operations, there is always the chance that there will be more blend sludge needing to be dewatered than there are centrifuges available. Overloading centrifuges leads to poor dewatered solids with high water content left in biosolids. Consequently, prolonged odor emissions are expected when this product is applied to the field site.

There were 14 centrifuges for dewatering process including seven new centrifuges installed in the year 2005. During the data collection, there was not a large number of centrifuges operating due to the limitation of the dewatering process. The reason was that a new set of centrifuges needed to be calibrated to fit with the solids operation at Blue

Plain (e.g. adjusted torque and chemical addition, etc.) while the old set of centrifuges also needed a major maintenance.

Characteristics of sludge and biosolids

Data collected on the sampling day are laboratory data and processing data obtained by sampling. These data include the characteristics of sludge in the blend system, properties of solids that were generating odor emissions.

B.3 Characteristics of sludge in the sludge blend system

The PCH data are mostly related to manageable parameters for each unit operation at Blue Plains. However, no parameters in the PCH database could directly provide information about sludge characteristics before dewatering. This type of data is crucial in understanding characteristics of sludge prior to dewatering and lime stabilization. Therefore, sludge samples from the gravity thickener, DAF, and blend tank were collected every sampling day to measure four characteristics; pH, oxidation reduction potential, concentration of H_2S from headspace, and temperature.

pH (GT pH, DAF pH, and BS pH): describe the acidity of the solution measured by the concentration of hydrogen ions (Reynolds et al., 1996). This can be illustrated by

$$pH = \log \frac{1}{[H^+]}$$

(3.3)

where a pH value of seven is considered neutral, below seven is acidic, and higher than seven is alkaline. The chemical and biological activities in sludge and dewatered solids are believed to be influenced by the pH of the sludge sample.

Oxidation Reduction Potential (GT ORP, DAF ORP, and BS ORP): ORP describes the oxidation state and reduction state of the sludge. ORP can be used as a mean to measure anaerobic conditions in a sample. Kim et al. (2002) noted that a lower

ORP value represents a greater anaerobic state in wastewater sludge which contributes to the rise of reduced sulfur compound production. As a result, low ORP values for sludge could indicate high biosolids odors.

Concentration of H_2S measured by a Jerome meter (JRM) in the headspace of the sludge sample (GT JRM, DAF JRM, and BS JRM; parts per million volume, ppmv): The generation of H_2S at the wastewater treatment plant is mostly the result of sulfur reduction bacteria to reduce sulfate or sulfur-containing matter to sulfide. The concentration of H_2S from sludge samples can be used as a mean to measure how septic the sludge conditions are before dewatering. These conditions are believed to influence odor production from biosolids after the dewatering process.

Temperature of sludge (GT T, BS T, and DAF T, Celsius): As mentioned before, microorganisms and their biological activity are temperature-dependent. Therefore, the temperature of sludge in the blend system should contribute to their biological activity. This parameter is included to monitor how sludge temperature in the blend system affects odor production.

B.4 Characteristics of biosolids

The properties of dewatered solids are expected to influence odor emissions. The relevant measurements include percent solids and temperature of solids in a sample.

Percent Solids (%): This variable is defined as the solids percentage by weight after the dewatered solids sample is baked in an oven to get rid of water content compared to the total weight before baking. This parameter indicates characteristics of sludge prior to dewatering on that day and/or can indicate the dewatering capacity of the centrifuge itself operating on that day. North (2003) found that dewatered solids with a

high percentage of solids need a longer time for the lime mixer to incorporate lime effectively into the solids, otherwise, odor will persist after lime stabilization. As lime was mixed sufficiently in the laboratory for three minutes, this should not be the case. Our assumption about the contribution of percent solids to odor emissions is that the greater the percent solids the lower the odor production as measured after 24 hours. This is due to less water content left in the mix to sustain microbial activity over time.

Sample Temperatures after three and 24 hours (Degrees Celsius): The temperatures of dewatered solids sampled both with without lime addition were recorded as we measured H_2S concentrations and collected air samples after three and 24 hours to investigate the effect of temperature on odor production. The assumption is that higher temperatures produce greater odor emissions.

In chapter four, we discuss biosolids odor prediction models at wastewater treatment plants. We identified key variables associated with biosolids odor emissions. Two types of models were developed; biosolids odor prediction in terms of a continuous odor variable and a discrete odor variable.

Chapter 4: Biosolids Odor Prediction Models at Wastewater Treatment Plants

In this chapter, we develop biosolids odor prediction models for operations at the Blue Plains wastewater treatment plant using detection threshold of biosolids as response variable. Missing observation techniques were used to prepare data for analysis and a variety of functional forms in addition to linear form were applied to the potential independent variables to investigate their relationship with biosolids odor levels. Then, linear statistical models were developed from three different groups of data sets: 1. high and moderate to low (MTL) odor groups, 2. summer and non-summer odor groups, and 3. full set of data. The results are the final biosolids odor prediction models at the Blue Plains facility. We also classified odor data into high or low odor groups and applied categorical data analysis techniques, logistical regression and discriminant analysis, to develop statistical models. It is anticipated that these models will provide tools to biosolids manager to more easily make decisions based on the discrete odor prediction result on any particular day.

Summary of major results found

Among models developed from different groups of data sets, the model for full set of data performed best on validation sets based on the mean absolute error (MAE). The model explained 40% of the variation in response variable. Variables in the model for full set of data were variables related to biosolids properties (i.e., percent solids and temperature of biosolids), operational parameters (i.e., return activated sludge concentration at secondary process, percentage of GT solids in the blend tank, and the

number of centrifuges running) and a variable related to GT solids properties (i.e., pH of GT solids). This indicates that biosolids odor is influenced by the properties of wastewater as well as operation at the plant.

For classification models, logistic regression and discriminant functions were used to classify biosolids odor to either odorous (class 1) or non odorous (class 0). This approach can assist biosolids management to decide appropriate reuse site according to the predicted class of biosolids. A number of variables were tried on both types of models to find the variables that best classified class of biosolids odor. Best models from both types were compared on the 20 validation sets. The model that performed best should have lower number of misclassification in class 1 and (if possible) lower number of overall misclassification rate. Misclassification in class 1 means we predict the actual odorous biosolids as non odorous. This can mislead the management to send odorous biosolids to sensitive area. Misclassification in class 0, on the other hand, means the biosolids with actual non-odorous type is classified as odorous. Thus, management possibly sends them to remote sites resulting in unnecessary hauling cost. The model that performed best on validation sets was from discriminant functions. It composes of six variables similar to the model for full set of data. (i.e., the percent solids and temperature of biosolids, return activated sludge concentration, percentage of GT solids in the blend tank, the number of centrifuges running and pH of GT sludge).

In the end, when unequal misclassification costs were taking into account, an unequal misclassification cost equation was developed. Unequal misclassification costs mean the cost associated with misclassifying one class of biosolids odor is different from misclassifying the other class. In this context, we considered misclassification biosolids

odor class 1 costs DCWASA more than class 0 since adverse attitude from neighbor to biosolids management program can cause DCWASAS to find alternative reuse sites that possibly need more cost and time to deliver the product. We assigned five misclassification cost ratios (cost of misclassification class1: class 0) at 3:1, 5:1 7:1, 9:1, and 10:1 to the equation. An average cost with respect to each ratio was used to identify the best classification model. According to the validation result and discussion with DCWASSA personnel, the discriminant functions with six explanatory variables still performed best to classify biosolids odor levels.

Next, we start this chapter with the objectives of statistical model and how it can be used in odor management purposes.

4.1 Introduction

A benefit of biosolids odor prediction models at wastewater treatment plants is to allow better management of those plant-specific variables that can lead to high levels of biosolids odor. In particular, these models will allow biosolids managers either monitor or control the odor causing variables before the malodorous product is produced. In this study, we developed statistical models for odor prediction based on data mentioned in Chapter 3.

To analyze the data, time series methods were considered. However, due to the lack of frequency of biosolids odor data samplings (they were collected once a day and mostly two days a week on Monday and Tuesday) an insufficient amount of information was collected so no time series analysis was done.

Generally, statistical models can be used for two purposes: prediction and explanation. In prediction, models are used to ascertain future outcomes of the dependent

variable (Shmueli, Patel, and Bruce, 2007). The performance of predictive models can be measured by the ability to predict future levels of the dependent variable. One approach is to split the data into a training set, a validation set, and, a test set. These concepts are very popular in data mining, where there are more observations and variables in the data set. The goal is to choose the model that best predicts values in the validation or test set.

In explanatory modeling, the preferred model is the model that fit training data best. No validation set or test set is needed. The estimated coefficients, coefficients' signs of independent variables, domain knowledge, and supporting theories help explain the impact of independent variables on the response variable in question.

The purpose of the statistical biosolids models in this thesis is to predict levels of biosolids odors. There are two types of odor response variables: those that can take on any continuous value and those that are categorical (e.g. high, medium, and low odor levels.). We present results on both types of models.

Observational data at the Blue Plains wastewater treatment plant were used to explain and predict biosolids odor production. There are many odor-causing factors at the plant as discussed in literature review.

Next, we describe the data used to develop the statistical models.

4.2 Data used

This section summarizes the data used to develop biosolids odor prediction models at Blue Plains. Table 4.1 provides abbreviations, units, and descriptions of data collected in this study. Table 4.2 presents summary statistics of data collected.

Table 4.1: Data abbreviations and descriptions

Name	unit	Data description
Biosolids odor data		
JWOL3	ppm	H2S concentration measured from headspace of sample without lime addition at 3rd hour
JWL3	ppm	H2S concentration measured from headspace of sample with lime addition at 3rd hour
JWOL24	ppm	H2S concentration measured from headspace of sample without lime addition at 24th hour
JWL24	ppm	H2S concentration measured from headspace of sample with lime addition at 24th hour
DTWOL3	OU	DT concentration measured from headspace of sample without lime addition at 3rd hour
DTWL3	OU	DT concentration measured from headspace of sample with lime addition at 3rd hour
DTWOL24	OU	DT concentration measured from headspace of sample without lime addition at 24th hour
DTWL24	OU	DT concentration measured from headspace of sample with lime addition at 24th hour
Sampling data		
% solid	%	Percent solids of dewatered solids sample
W/O lime 3 T	Celsius	Temperature of sample without lime addition at 3rd hour
W/O lime 24 T	Celsius	Temperature of sample with lime addition at 3rd hour
With lime 3 T	Celsius	Temperature of sample without lime addition at 24th hour
With lime 24 T	Celsius	Temperature of sample with lime addition at 24th hour
GT T	Celsius	Temperature of sludge from gravity thickener
DAF T	Celsius	Temperature of sludge from DAF
BS T	Celsius	Temperature of sludge from blend tank
GT pH	-	pH of sludge from gravity thickener
DAF pH	-	pH of sludge from DAF
BS pH	-	pH of sludge from blend tank
GT ORP	mv.	ORP of sludge from gravity thickener
DAF ORP	mv.	ORP of sludge from DAF
BS ORP	mv.	ORP of sludge from blend tank
GT JRM	ppm	H2S concentration from headspace of sludge from gravity thickener
DAF JRM	ppm	H2S concentration from headspace of sludge from DAF
BS JRM	ppm	H2S concentration from headspace of sludge from blend tank
TPS	gal/min	Primary sludge (Gravity thickener sludge) feeding rate into blend tank
TWAS	gal/min	Waste Activate Sludge (DAF sludge) feeding rate into blend tank
# centrifuge running	-	Number of centrifuge running when samples were collected
% TPS IN BLEND	%	Percent of TPS in blend tank when sample were collected

Table 4.1: Data abbreviations and descriptions (continued)

Name	unit	Data description
PCH data		
BD east	feet	Average sludge blanket level at secondary sedimentation east side
BD west odd	feet	Average sludge blanket level at secondary sedimentation west odd side
BD west even	feet	Average sludge blanket level at secondary sedimentation west even side
GT% in blend	%	Average percent of (% solids of GT sludge/ (% solids of GT + % solids of DAF))
TWAS (gal)	gal	Total primary sludge (Gravity thickener sludge) fed into blend tank on that day
TPS (gal)	gal	Total Waste Activate Sludge (DAF sludge) fed into blend tank on that day
RAS west (mg/l)	mg/l	Average Return Activated Sludge concentration at secondary east reactor
RAS east (mg/l)	mg/l	Average Return Activated Sludge concentration at secondary west reactor
BD east d-1	feet	Average sludge blanket level at secondary sedimentation east side on the previous day
BD west odd d-1	feet	Average sludge blanket level at secondary sedimentation west odd side on the previous day
BD west even d-1	feet	Average sludge blanket level at secondary sedimentation west even side on the previous day
GT % in blend d-1	%	Average percent of (% solids of GR sludge/ (% solids of GT + % solids of DAF)) on the previous day
TWAS (gal) d-1	gal	Total primary sludge (Gravity thickener sludge) fed into blend tank on that day on the previous day
TPS (gal) d-1	gal	Total Waste Activate Sludge (DAF sludge) fed into blend tank on that day on the previous day
RAS west (mg/l) d-1	mg/l	Average Return Activated Sludge concentration at secondary east reactor on the previous day
RAS east (mg/l) d-1	mg/l	Average Return Activated Sludge concentration at secondary west reactor on the previous day

Table 4.2: Summary statistics

Name	unit	Mean	Std. Dev.	Minimum	Maximum	Count
BIOSOLIDS ODOR DATA						
JWOL3		0.2669	0.1355	0.0703	0.7833	77
JWL3		1.4135	0.9964	0.0022	3.4500	77
JWOL24		9.89	13.31	0.12	50.00	77
JWL24		1.137	1.381	0.017	6.933	77
DTWOL3		1270.33	1336.71	177.00	8200.00	76
DTWL3		1119.79	935.31	170.00	6500.00	76
DTWOL24		11230.88	15548.11	370.00	74798.00	72
DTWL24		1740.28	2258.20	240.00	14000.00	72
DTWOL3 (St.Croix)		1633.20	1510.12	220.00	8200.00	50
DTWL3 (St.Croix)		1195.40	1125.63	170.00	6500.00	50
DTWOL24 (St.Croix)		6708.30	9183.84	370.00	53000.00	47
DTWL24 (St.Croix)		1408.09	2271.22	240.00	14000.00	47
DTWOL3 (OdorS)		572.50	347.63	177.00	1386.00	26
DTWL3 (OdorS)		974.38	328.03	385.00	1794.00	26
DTWOL24(OdorS)		19733.32	20931.84	1166.00	74798.00	25
DTWL24(OdorS)		2364.80	2139.28	451.00	8694.00	25
Sampling data						
% solid		25.701	2.801	18.600	35.710	77
Room T (24)		25.403	3.126	15.000	31.800	76
W/O lime T		23.703	3.208	15.500	30.500	75
W/O lime 24 T		24.680	2.963	18.000	32.000	66
With lime T		24.896	3.024	16.000	30.000	75
With lime 24 T		24.408	2.891	18.000	32.000	66
GT T		20.634	4.527	11.300	28.350	52
BS T		22.321	4.653	12.550	30.350	52
DAF T		21.334	3.997	12.700	27.700	52
GT pH		5.7173	0.3088	4.8200	6.2500	77
BS pH		6.2558	0.2831	5.2050	6.6750	77
DAF pH		6.6142	0.3397	5.5300	7.2200	77
GT ORP		-97.16	65.26	-268.40	30.15	50
BS ORP		-137.29	58.37	-253.15	4.95	50
DAF ORP		-190.48	66.46	-281.70	-27.55	50
GT JRM		1.098	1.558	0.143	12.433	77
BS JRM		0.5239	0.3031	0.1500	1.3333	77
DAF JRM		0.7023	0.5370	0.1417	2.9333	77
BLEND RATIO		1.807	1.509	0.421	9.818	51
TPS		789.78	170.95	400.00	1309.00	51
TWAS		594.61	278.60	132.00	1517.00	51
# centrif running		6.675	1.922	3.000	11.000	77
% TPS IN BLEND		0.5854	0.1370	0.2965	0.9076	51
PCH data						
BD east		2.3052	0.9950	0.0000	4.7000	77
BD west odd		1.100	1.453	0.000	5.000	77
BD west even		0.904	1.044	0.000	3.600	77
GT% in blend		53.78	23.83	8.16	100.00	64
TWAS (gal)		997719.25	1378011.56	0.00	9559220.00	68
TPS (gal)		958019.72	1104222.43	113100.00	9333000.00	67
Blend ratio		0.4942	0.2167	0.0998	1.0000	67
RAS west (mg/l)		4591.82	1686.87	1300.00	8000.00	33
RAS east (mg/l)		7239.87	3152.23	1350.00	19560.00	76
BD east d-1		2.4959	0.8043	1.3000	4.7000	74
BD west odd d-1		1.367	1.613	0.000	5.000	73
BD west even d-1		1.000	1.085	0.000	3.300	76
GT% in blend d-1		55.91	17.88	8.16	100.00	63
TWAS (gal) d-1		878717.00	1087561.17	0.00	9559220.00	70
TPS (gal) d-1		891532.38	332174.32	100700.00	1291310.00	65
Blend ratio d-1		0.5087	0.1630	0.0998	0.9184	65
RAS west (mg/l) d-1		4849.68	1901.71	1300.00	7800.00	31
RAS east (mg/l) d-1		7184.16	3111.53	1150.00	15500.00	77

4.3 Data Preparation and Techniques used with missing observations

As is the case with initial data sets, it is common to find incomplete data for certain variables related to input errors, missing observations, typos, etc. To overcome these problems, we conducted a screening process by first running scatter plots on all the variables to identify outliers. For missing observations, we applied three techniques to missing data (Maddala, 1977).

- Use an **average value of the adjacent observations.** If adjacent observations of the missing data value were available, such as the data on the day before and after, we took an average of those available data to fill in for the missing values. The variables we applied this technique to were the sludge blanket level at secondary west and east on day d-0 and d-1, for seven observations each.

- Use **regression model to estimate the missing values.** If there was an index variable that showed a correlation with the variable of interest, we used this index variable as a independent variable to estimate the value of the missing observation. The variables we applied this technique to were sludge temperatures from GT, DAF, and BS. Flow temperature at the secondary process was used as a independent variable (index variable) in a regression model to estimate the missing value of sludge temperatures. The models developed explained temperatures at GT, DAF, and BS with adjusted R-squared values of 0.92, 0.69, and 0.82, respectively.

- Use an **average of all observations** for that variable. In the case that no index variable and adjacent values were available, the averages of all

existing observations of that variable were used. Since there were few observations from GT% in blend, TPS, and TWAS on day d-0 and d-1 this technique was applied for these variables.

In summary, the screening process and missing observation techniques were conducted to prepare data for analysis. The missing observation techniques were applied to facilitate the variable selection process where the complete data set was required to run variable selection techniques, such as backward, forward, or stepwise selection.

Analysis

Three approaches were used to develop biosolids odor prediction models. First, different functional forms were checked to assess the relationship between odor and potential odor-causing variables. Second, linear regression modeling was used was used to predict and explain biosolids odor concentrations for three different groups of data sets: summer and non-summer sets, high (H) and moderate-to-low (MTL) odor scores, and the full data set. Third, a discriminant analysis and logistic regression were used to classify odor data into groups, high (H) and moderate-to-low MTL odor groups. Figure 4.1 illustrates the Blue Plains' biosolids odor analysis scheme. In what follows, we comment on each of these approaches.
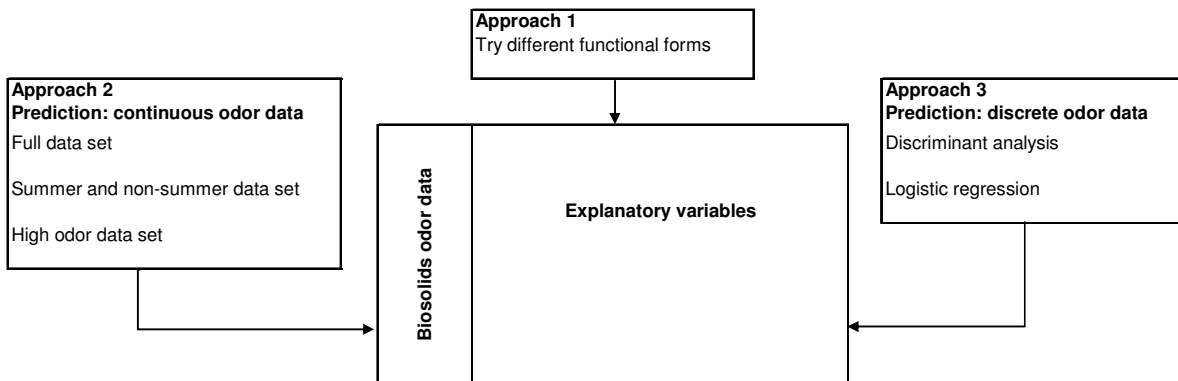
Figure 4.1: Blue Plains' biosolids odor analytical diagram

4.4 Approach 1: Different Functional Forms

Due to the simplicity of a linear model, this approach was the starting point for the analysis. The relationship between a response variable and an independent variable is not always best described by a linear functional form, for example, as in the case of an exponential growth rate or exponential decay rate of microorganisms.

The transformation of independent variables can be use to transform a nonlinear regression relation existing between a response variable and an independent variable to a linear relation (Kutner, 2005). Thus, the following functional forms were also applied to more than 30 predictive variables using DTWL24 as the response variable:

1. Linear: DTWL24 vs independent variable

2. Log-linear: log(DTWL24) vs. log(independent variable)

3. Semilog: DTWL24 vs. log(independent variable)

4. Square-root: DTWL24 vs. $\sqrt{\text{independent variable}}$)

5. Square: DTWL24 vs. (independent variable)^2

6. Reciprocal: DTWL24 vs. 1/(independent variable)

R squared and p-values when applying linear regression on functional forms mentioned above are presented in Table 4.3. Variables showing R squared values greater than 0.1 were the temperature of biosolids (With lime 24T) for all function forms and concentration of waste activated sludge at east side (RAS east) for linear, squared, and reciprocal forms. In general, transformations of independent variables did not significantly improve the explanatory power of these variables.

Table 4.3: Regression results when applying different functional forms to independent variables (one at a time)

Name	Function # 1 Linear R-square	p-value	Function # 2 Log R-square	p-value	Function # 3 Semi-log R-square	p-value	Function # 4 Square-root R-square	p-value	Function # 5 Square R-square	p-value	Function # 6 Reciprocal R-square	p-value
% solid	0.02	0.2	0.015	0.28	0.02	0.21	0.02	0.21	0.02	0.02	0.02	0.2
With lime 24 T	**0.15**	**0.0004**	**0.24**	**0.0001**	**0.14**	**0.0006**	**0.14**	**0.0005**	**0.16**	**0.0003**	**0.16**	**0.0003**
GT T	0.003	0.61	0.00001	0.99	0.0018	0.71	0.0025	0.66	0.005	0.53	0.005	0.53
BS T	0.02	0.25	0.004	0.56	0.0122	0.33	0.014	0.29	0.02	0.18	0.02	0.18
DAF T	0.02	0.19	0.01	0.31	0.02	0.22	0.02	0.2	0.02	0.17	0.02	0.16
GT pH	0.05	0.05	0.03	0.09	0.04	0.06	0.04	0.05	0.05	0.04	0.05	0.04
BS pH	0.04	0.05	0.04	0.07	0.04	0.06	0.04	0.06	0.05	0.04	0.05	0.04
DAF pH	0.00001	0.97	0.00001	0.95	0.00001	0.97	0.00001	0.97	0.00001	0.97	0.00001	0.97
GT ORP	0.02	0.22	0.06	0.02	0.07	0.01	0.03	0.1	0.0019	0.71	0.0019	0.71
BS ORP	0.009	0.4	0.02	0.2	0.04	0.07	0.01	0.29	0.0019	0.7	0.0019	0.7
DAF ORP	0.01	0.34	0.01	0.28	0.03	0.08	0.01	0.31	0.0087	0.41	0.0087	0.41
GT JRM	0.00001	0.98	0.02	0.15	0.01	0.32	0.003	0.63	0.0019	0.7	0.0019	0.7
BS JRM	0.02	0.22	0.01	0.24	0.01	0.26	0.02	0.24	0.02	0.18	0.02	0.18
DAF JRM	0.02	0.21	0.04	0.06	0.01	0.25	0.02	0.21	0.01	0.33	0.01	0.33
# centrif running	0.0007	0.81	0.005	0.53	0.00001	0.81	0.0007	0.81	0.0008	0.36	0.0008	0.8
BD east	0.0069	0.47	0.02	0.25	0.008	0.44	0.007	0.46	0.0069	0.47	0.0069	0.47
BD west	0.00017	0.71	0.0015	0.74	0.001	0.75	0.0015	0.73	0.0024	0.67	0.0024	0.07
GT% in blend	0.05	0.05	0.03	0.1	0.06	0.22	0.05	0.03	0.03	0.12	0.03	0.13
RAS west (mg/l)	0.00001	0.97	0.004	0.56	0.0018	0.73	0.0005	0.84	0.0008	0.81	0.0008	0.81
RAS east (mg/l)	0.05	0.04	0.033	0.11	0.06	0.02	0.06	0.03	0.02	0.16	0.02	0.16
BD east d-1	0.004	0.56	0.007	0.45	0.002	0.64	0.003	0.6	0.0062	0.49	0.0062	0.49
BD west d-1	0.014	0.29	0.007	0.46	0.006	0.47	0.01	0.38	0.02	0.16	0.02	0.16
GT% in blend d-1	0.047	0.05	0.009	0.39	0.022	0.19	0.03	0.11	0.08	0.0097	0.08	0.0097
RAS west (mg/l) d-1	0.06	0.47	0.004	0.57	0.011	0.36	0.008	0.42	0.005	0.53	0.0091	0.53
RAS east (mg/l) d-1	**0.11**	**0.002**	0.05	0.04	0.07	0.26	0.09	0.0055	**0.14**	**0.0008**	**0.14**	**0.0008**

Of the candidate variables collected, Table 4.4 summarizes just those variables that showed a correlation coefficient greater than 0.2 in absolute value with DTWL24.

Table4.4: Correlation coefficients between DTWL24 and potential independent variables of all functional forms greater than 0.2

	Functional forms					
Predictors	linear	log	reciprocal	squared	square-root	log-log
Withlime 24 T	**0.39**	**0.38**	**-0.37**	**0.40**	**0.39**	**0.49**
GT pH	0.22	0.21	-0.21	0.23	0.22	
BS pH	0.22	0.21	-0.20	0.22	0.22	0.22
GT % in blend d-1	0.22			0.29		
RAS east d-1	**0.35**	0.25		**0.36**	**0.30**	
RAS east* GT% d-1	**0.42**					

Temperature of biosolids sample (Withlime 24 T): The temperature of the biosolids sample at 24 hours showed a moderate correlation with odor emissions for all functional forms considered. The log-log transformation (the functional form # 2 on previous page) gave the highest correlation coefficient with a value of 0.49, followed by the square transformation (the functional form # 5) with the a value of 0.4. All the remaining functions including linear, log, square-root and reciprocal revealed roughly the same correlation coefficient values in absolute value, 0.38 and 0.39.

pH of sludge samples (GT pH and BS pH): pH of sludge samples from GT and BS showed weak correlation with DT on all functions applied. As a result, a simple functional form, linear relationship, was considered as an appropriate choice.

GT % in blend d-1: weak positive correlation exists in linear and squared functional forms of GT% in blend d-1.

RAS east on day d-1 (RAS east d-1): RAS east on day d-1 showed moderate correlation for linear (0.35), squared (0.36), and square-root (0.3) functional forms, and even greater when we checked the interaction between RAS east d-1 and GT% in blend d-1 (RAS east*GT% d-1) showing moderate correlation coefficient (0.42).

4.5 Approach 2: Predict Continuous Odor Levels

Collection of one yearly worth of data in this research included several events, such as extremely high and low ambient temperatures at Blue Plains, various characteristics of wastewater loading into the plant, change in operations with respect to seasons, or machinery malfunction, etc. Consequently, we divided data into different subsets to facilitate the analysis. Data were reorganized into the following groups:

1) High odor emissions (H) and moderate-to-low (MTL) odor emissions sets which we wanted to investigate what causes high odor biosolids;

2) Summer and non-summer sets which we wanted to observe possible impacts of temperature on microorganism activity, daily operations at Blue Plains, properties of wastewater, and most importantly odor emissions.

3) Full data set which we want to investigate all data as a whole.

Next, we describe H and MTL groups.

High (H) and moderate-to-low (MTL) odor emissions groups

An understanding of what influences unusually high biosolids odor emissions will assist wastewater managers in avoiding odor incidents. To this end, values of the response variable, DTWL24 greater than 2000 ou were selected as the point to divide biosolids odor observations into a high odor group (H) and a moderate to low (MTL) group. The value of 2000 was determined by observing biosolids odor generation for the entire odor data collected in this study taking odor concentration at the 75 percentile or greater as high odor. Figure 4.2 illustrates the DTWL24 distribution.

In 18 of 77 observations or 22 percent, the DT level was greater than 2000 ou. High odor concentrations mostly happened in the summer and spring seasons. (Summer: May to August, Fall: September to December, and Spring: February to April).
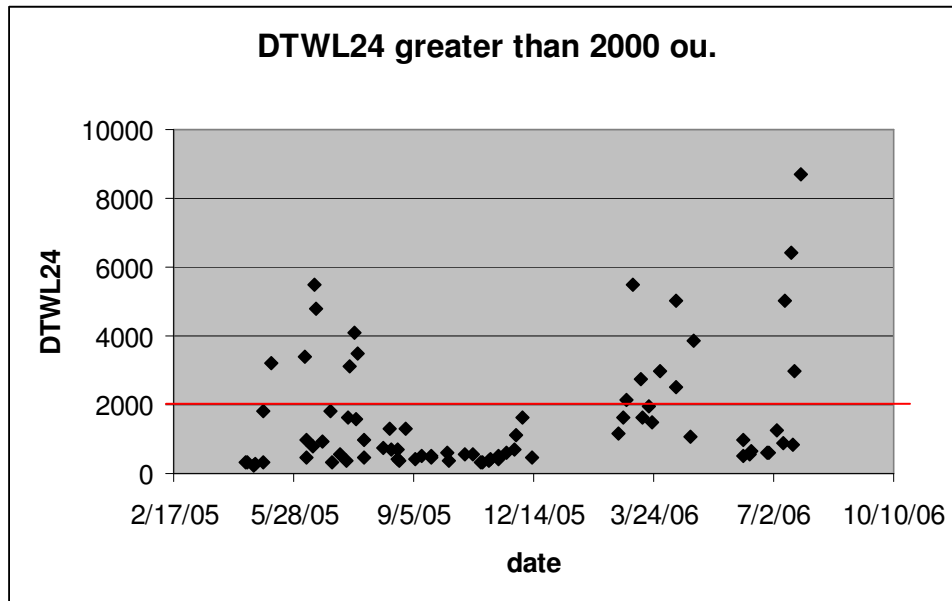
Figure 4.2: DTWL24 vs. date, high and moderate-to-low groups

Summer and non-summer odor emission groups

Seasonality is one of the factors believed to influence biosolids odor levels. The rationale is that the change of temperature could have impacts on biological activity. In addition, seasonality may be important in explaining the patterns of biosolids odor emissions caused by the various types of wastewater treated at the plant. For example, the summer may have more wastewater from industry rather than from residential uses.

Data were divided into summer (May to August 2005 and May and July 2006[1]) and non-summer (the rest of the data) to observe the effects of ambient temperature on

---

[1] Data was collected until July 2006 due to project scheduling reasons.

microbial activity and ultimately on biosolids odors. The number of observations from the summer and non-summer groups were 38 and 39, respectively.

Figure 4.3 illustrates the odor emissions from biosolids collected on both groups. It should be noted that the odor level greater than 2000 ou were existed on both groups.
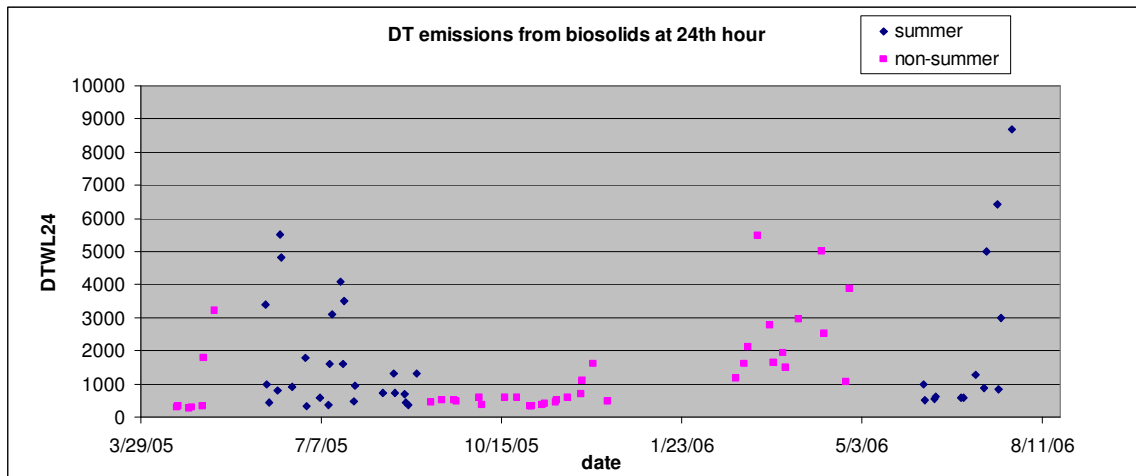
Figure 4.3: DT odor emissions vs. date, summer and non-summer groups

Full data set using all data

Using all 73 observations for modeling is preferred to disaggregating the data into small subsets. The reason is that the full data set provides a thorough coverage that includes a variety of events related to temperature, operation, or type of wastewater at Blue Plains. Using the full data set also increases explanatory power of the model indicated by degrees of freedom (number of observations minus number of predictors including intercept). Dummy variables and interaction variables techniques were also applied to find specific incidents in wastewater operation. Figure 4.4 illustrates all DT data collected in this study.
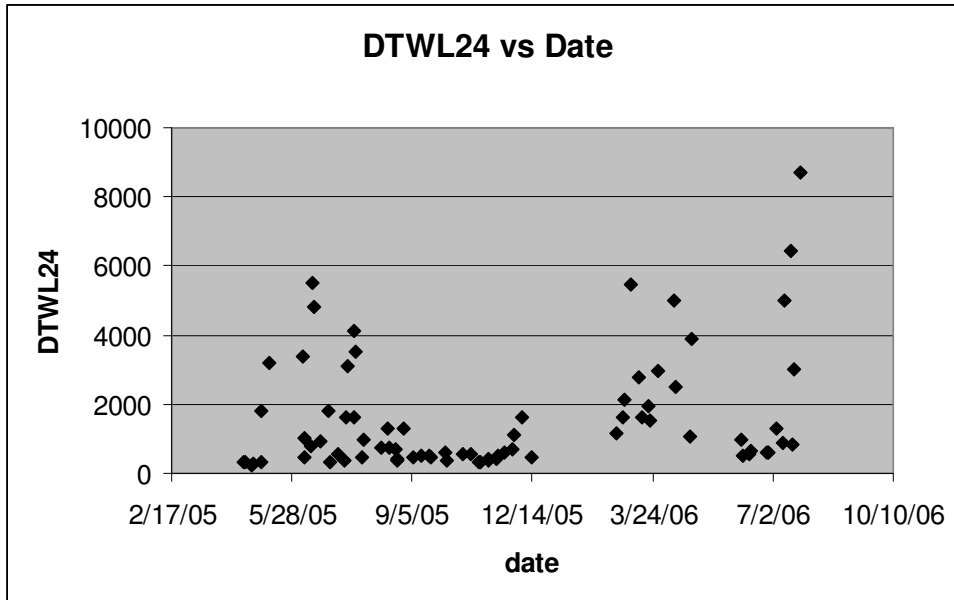
Figure 4.4: DT odor emissions vs. date all data set

Next, we describe the modeling strategy: variable selection method and model selection criteria.

4.5.1 Modeling Strategy

If a small number of potential independent variables are of interest, we can try all possible combinations of these variables to formulate all possible models. For example, if we have four variables under consideration, we can try 2^4-1 = 15 different subsets to check the significance of each variable in each subset and carefully investigate each model to select the appropriate model(s). However, numbers of potential independent variables in this study were greater than that so we decided to use an automatic searching technique to find smaller subsets that were manageable.

In particular, we used stepwise regression techniques (Albright, 2003) to find the significant variables. Stepwise regression starts with no variables in the model. Then, from a selected set of data, the program searches for the variable that explains the

variation in the mean of the response variable best to include by using the following criteria: 1) reduction in sum of squares error and 2) statistical significance of variables added in or removed from the model by t-statistic and F-statistic.

At each round, with respect to the explanatory variables already added into the model, the program searches for the other variables left in the data set that can improve the explanatory power of the existing model. If previously entered variables become insignificant, it will be excluded from the model.

Of the selected models from the stepwise regression procedure, the following criteria are used as a guideline to search for candidate models.

Significance of variables added into the model: The variables added into the model must be statistically significant at a p-value less than 0.1 or at most 0.2. A P-value of 0.2 means one could reject the null hypothesis that the coefficient of the observed variable is equal to zero with 80% confidence or with only a 20% chance of being wrong. This p-value, 0.2, may be higher than the most popular p-value of significance, such as 0.001 or 0.01 but since our study involved various uncontrolled parameters at Blue Plains the probability of being wrong at 20% is still applicable.

Interpretation: The coefficient sign of the explanatory variable(s) in the model should be explainable either by theory, previous research, or empirical study.

Degrees of freedom: The model as a whole should have a high degree of freedom given by (n-k-1), where n is the number of observations, k is number of the estimators in the model, and 1 is used for the intercept of the model.

Explanatory power: The model should be able to explain at least 50% of the variation in the mean of the response variable. 50% explanatory power may seem too

low. However, with respect to the nature of exploratory study that includes a number of uncontrollable variables from daily wastewater treatment plant operation 50 percent explanatory power from the developed model is sufficient here. We consider both regression R^2 square and adjusted R^2.

R^2 is the proportion of variation that is explained by a statistical model. R^2 can be defined by

$$R^2 = 1 - \frac{SSE}{SST} \qquad \text{(Devore, 1995)} \qquad (4.1)$$

where

SST = the total variation in the response variable

SSE = the variation in the response variable unexplained by the model

$$SSE = Y_i - \hat{Y}_i$$

$Y_i - \hat{Y}_i$ = the difference between the actual and predicted value (Y_i is the actual value of the response variable at observation i and $\hat{Y}_i$ is the predicted value of the response variable by the model at observation i)

$$SST = \Sigma(Y_i - \bar{Y})^2 \quad (\bar{Y} = \text{the mean of the response variable})$$

Since R^2 will always increase as the number of explanatory variable in the model increase, the adjusted R^2 is used to adjust this error.

$$\text{Adjusted } R^2 = 1 - \left(\tfrac{n-1}{n-p}\right) \frac{SSE}{SST} \qquad (4.2)$$

where

n = number of the observations used for estimation

P = number of the variables to be estimated

The following sections introduce the best models from each data set and describe the contribution of independent variables in the models.

4.5.2 Results

According to modeling strategy, stepwise variable selection technique was conducted to reduce number of potential independent variables and identify significant ones to add. We selected the best model for each data followed the selection criteria. The best models for each set were as follows.

A. Models for High and Moderate to Low Odor Data Sets

To initially explore the biosolids odor data in the H group and MTL group data, hypothesis tests were conducted on the interesting variables presumably contribute to high odors. A summary of hypothesis tests were reported here. This was a necessary first step before developing statistical models.

Summary statistics and hypothesis testing (t-test)

We conducted hypothesis tests as to whether these variables from the H group had higher odor values than in the MTL group on the average. See Table 4.5

Table 4.5: Hypothesis testing of data in H group and MTL group

	Average		Hypothesis testing
Variables	H group	MTL group	whether H group > MTL group
DTWL24	4193.00	785.29	Yes at 1% significance level
DTWOL24	23226.39 ou	7790.72 ou	Yes at 1% significance level
DTWL3	1491 ou	1012 ou	Yes at 5% significance level
DTWOL3	1478.33	1205.62	No at 1% significance level
% solids	24.90%	25.90%	No at 10% significance level
Withlime 24 T	25.76 c	23.51 c	Yes at 1% significance level
GT% in blend d-1	58.84%	54.74%	Yes at 1% significance level
RAS east d-1	8588 mg/l	6650 mg/l	Yes at 1% significance level
# of observations	59	18	

The followings describe the results of the variables being tested.

Odor data: Most of the odor variables from the H group were higher on average than the MTL group. An exception is the DT level from biosolids without lime addition at 3 hours (DTWOL3). This can be described that high odor products were the result of the dewatered sludge originally containing high odor.

Percent solids (% solids): Average %solids from the H and MTL groups were 24.9% and 25.9%, respectively. The H group had a lower percent solids than MTL group at the 10% significance level. This coincides with our assumption that the higher the percent solids of the biosolids samples the lower the odor production, and vice versa.

Temperature of sample (Withlime 24 T): The hypothesis that the H group had higher sample temperatures was accepted at the 1% significance level. This supports the assumption that the higher the biosolids temperature the higher the biosolids odor concentration.

 Percent of GT sludge in the blend tank on day d-1 (GT% in blend d-1): The average values for GT% in blend d-1 for the H and MTL groups were 58.84% and 54.74%. A hypothesis test of whether GT% in blend d-1 from the H group was higher than the MTL group was accepted at the 1% significance level. This suggests that

odorous biosolids in H group was influenced by higher percent of gravity thickener solids in blend tank compared to MTL group.

Concentration of return activated sludge on day d-1 (RAS east d-1): RAS east d-1 for the H group was higher than for the MTL group on average with average values of 8588 mg/l and 6650 mg/l respectively. A hypothesis test of whether RAS east d-1 from the H group was higher than the MTL group was accepted at the 1% significance level.

The hypothesis tests helps to identify potential odor causing variables for the H and MTL groups. We took this information into account in selecting key variables. Next, we present the statistical model for predicting odor levels for the data set of high odors.

High odor prediction model

The high odor model is present in Table 4.6. It composes of the concentration of RAS on the east side and the percent of GT solids in the blend tanks contributed to biosolids odor production for observations corresponding to 2000 ou and greater. This implies that both variables should be monitored and controlled closely since they influence the production of odorous biosolids.

Table 4.6: Selected high biosolids odor model

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.8301	0.6890	0.6475	974.8102897		

ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	2	31578831.49	15789415.74	16.6160	0.0002	
Unexplained	15	14253826.51	950255.1009			

Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95% Lower	Upper
Constant	-2407.921036	1170.701623	-2.0568	0.0575	-4903.212477	87.37040589
RAS east (mg/l)	0.482701939	0.10467512	4.6114	0.0003	0.259592202	0.705811677
GT% in blend d-1	45.92401117	12.80800996	3.5856	0.0027	18.62438417	73.22363817

The model explains 65 % of the variation in DTWL24 with two independent variables. Both variables have positive coefficient signs implying that high

concentrations of RAS on the east side and high percentage of solids of sludge from GT in the blend tank promote high biosolids odor emissions. High RAS indicates a high concentration of sludge at the bottom of the secondary sludge sedimentation tank resulting in septic conditions before wasting to the blend tank. A high percentage of GT sludge in blend indicates a greater proportion of food source available for microorganisms in the blend tank while also contributes to producing highly odorous biosolids.

None of models developed in the MTL data set was promising. The resulting models had low explanatory power (i.e., less than 30 percent for adjusted R square) or counterintuitive coefficient signs

Next, we introduce the summer and non-summer models.

B. Models for summer and non-summer data sets

To investigate the seasonality effects, odor prediction models for summer and non-summer periods were developed. The best model for the summer data is presented below in Table 4.7.

Table 4.7: Model for summer data

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.7010	0.4915	0.4298	1502.255344		

ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	4	71975218.12	17993804.53	7.9733	0.0001	
Unexplained	33	74473446.94	2256771.119			

Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95% Lower	Upper
Constant	-3419.370219	3553.393811	-0.9623	0.3429	-10648.80429	3810.063848
% solid	-184.7192295	83.40847392	-2.2146	0.0338	-354.4150457	-15.0234134
With lime 24 T	145.4531122	95.11811208	1.5292	0.1357	-48.0661419	338.9723663
GT% in blend d-1	73.5270584	18.55704114	3.9622	0.0004	35.77247433	111.2816425
RAS east (mg/l) d-1	0.271978535	0.085171219	3.1933	0.0031	0.098696387	0.445260682

The model for the summer data consists of two variables similar to the high odor model described above; GT% in blend d-1 and RAS east d-1. Two additional variables added into the model for summer data were percent solids and temperature of the biosolids. Both variables are related to the characteristics of the biosolids sample. We would expect the contribution of GT% in blend and RAS east d-1 on a sample with extremely high odor (H group) were more than the contributions of percent solids and temperature of biosolids samples as seen in previous section where percent solids and temperature of biosolids samples were not selected into High odor prediction model. However, in the summer season, properties of the biosolids samples relative to percent solids and temperature were also contributing to biosolids odor production.

During the summer season, operational variables such as GT% in blend d-1 and RAS east d-1 need monitoring to yield acceptable biosolids odor levels. Proper management of the sludge blanket level and the retention time of sludge in the sedimentation tank are still mandatory as well. Also a high GT% in blend, indicating a greater amount of food source in the blend tank, should be monitored to maintain the proper blend ratio.

Percent solids in the summer data was a result of operations at the Blue Plains facility. A low percent solids indicates septic conditions of the sludge that is usually retained in the tanks for a long time making it is hard to remove the water content (Ramirez, 2007). On the other hand, a high percent solids 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 tanks. Therefore, proper maintenance and operation on centrifuge is important to maintain acceptable biosolids odor.

Lastly, the temperature of the biosolids sample after 24 hours here corresponds to the ambient temperature where we store the biosolids. Biosolids management should be aware of the influence of the ambient temperature after the biosolids are produced either in storage or at the field site.

Next we introduce the non-summer model.

Model for non-summer data

The model for the non-summer data set consists of the following independent variables: biosolids temperature, pH of GT, and RAS east d-1 as shown in Table 4.8.

Table 4.8: Model for non-summer data

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.7175	0.5148	0.4720	922.6216635		

ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	3	30705275.26	10235091.75	12.0239	< 0.0001	
Unexplained	34	28941844.95	851230.7339			

Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95% Lower	Upper
Constant	-19740.19491	4425.368303	-4.4607	< 0.0001	-28733.62534	-10746.76447
With lime 24 T	132.2463375	55.28152523	2.3922	0.0224	19.90076136	244.5919136
GR PH	2934.235307	799.7337654	3.6690	0.0008	1308.980754	4559.489861
RAS east (mg/l) d-1	0.110385174	0.05070062	2.1772	0.0365	0.007349119	0.21342123

The model for the non-summer data has two variables that are similar to the model for the summer: RAS east d-1 and sample temperature. It does not include % solids and GT% in blend d-1 but includes pH at GT instead. We demonstrate scatter plots and correlation coefficients between some independent variables and DTWL24 during summer and non-summer periods below in Figure 4.5, 4.6, and 4.7.

Percent solids: From Figure 4.5, it can be seen that the percent solids of biosolids showed no correlation with DTWL24 for the non-summer data and showed moderate

correlation (-0.3) for the summer period. The range of percent solids in the non-summer data was from 18 to 28 while in the summer data the range was from 20 to 36 percent. A hypothesis test of whether percent solids of biosolids during summer was greater than non-summer was accepted at 1% significance level. The greater percent solids during the summer period may be due to better operations in the dewatering process. As a result, it contributed to odor emissions only in the summer model.
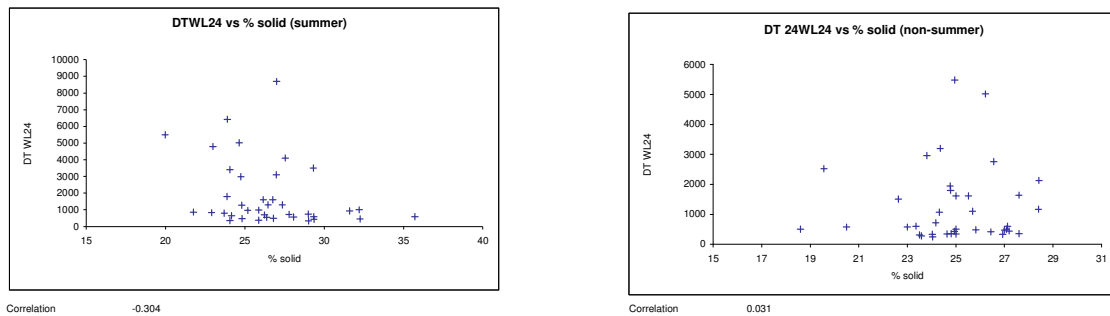
Figure 4.5: Scatter plot DTWL24 vs Percent solids

pH of sludge from GT: pH of sludge from GT was significant in the non-summer data. Figure 4.6 shows a graph of the relevant variables with the correlation coefficient between GT pH and DTWL24 of 0.59 in the non-summer data with pH values varying between 5.4 to 6.3. In the summer season, pH GT showed a weak correlation of 0.24 and the pH values varied from 4.8 to 6. A hypothesis test of whether pH of GT during summer period was greater than the non-summer period was accepted at 1% significance level. However, we found that for the selected summer model, GT% in blend d-1 was significant while pH GT was rejected. The rationale can be that both variables were explaining the same variation in the response variable but GT% in blend d-1 explained better than pH GT.
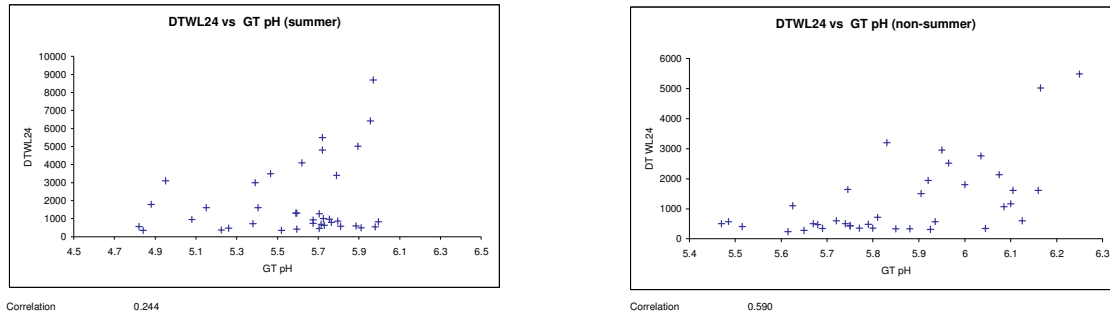
Figure 4.6: Scatter plot DTWL24 vs GT pH

GT% in blend d-1: GT% in blend d-1 was not significant in the non-summer model and showed no correlation with DT. GT% in blend d-1 was significant in the summer model showing a moderate correlation coefficient of 0.337 with DT during the summer season as indicated in Figure 4.7. A hypothesis test of whether the percentage of GT in the blend tank during the summer period was greater than the non-summer period was accepted at 1 % significance level. This should be the result of the operations at the blend tank during summer. As a result, the greater the amount of GT solids in the blend tank the greater the contribution to the odor emission in the summer model.
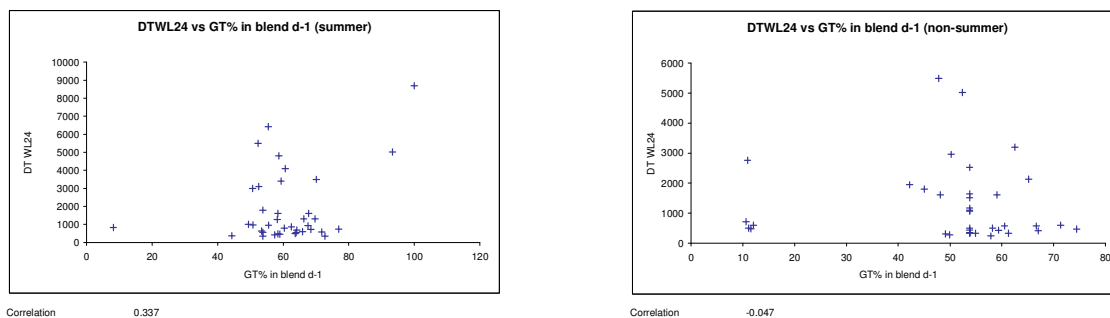
Figure 4.7: Scatter plot DTWL24 vs GT % in blend

Next, we introduce a model developed on full data set.

C. Models based on the full data set

Using the entire data set, the best model includes six independent variables from the summer and the non-summer models plus the number of centrifuges running on any given day. The selected model is shown in Table 4.9.

Table 4.9: Biosolids odor prediction model on full data set

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate		
	0.6726	0.4525	0.4055	1308.058108		

ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value	
Explained	6	98970986.22	16495164.37	9.6406	< 0.0001	
Unexplained	70	119771121	1711016.014			

Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95% Lower	Upper
Constant	-12936.39243	3615.958781	-3.5776	0.0006	-20148.19482	-5724.590045
% solid	-141.5438774	56.06522807	-2.5246	0.0139	-253.362449	-29.72530592
With lime 24 T	256.6562437	55.3648653	4.6357	< 0.0001	146.2345017	367.0779857
GR PH	1828.693101	523.5315344	3.4930	0.0008	784.5423802	2872.843822
# centrif running	-200.7211861	90.77327715	-2.2112	0.0303	-381.7627788	-19.67959348
GT% in blend d-1	31.34905799	9.919003998	3.1605	0.0023	11.56622831	51.13188767
RAS east (mg/l) d-1	0.15918685	0.049518402	3.2147	0.0020	0.060425511	0.257948188

For the resulting full model, the number of centrifuges in service is the only new variable added into the model beyond what was already included into models described above. A negative coefficient sign on the variable "number of centrifuges running" supports the assumption that a sufficient number of centrifuges can reduce the sludge retention time in the blend tank and secondary sedimentation basins. Overall, the model explains roughly 41 percent of the DT variation.

One thing to be mentioned here is that the concentration of returned activated sludge at the east side (RAS east (mg/l)) was statistically significant for all selected models. This was reasonable since it represented the concentration of solids at the bottom of the secondary sedimentation tanks where we assumed that higher concentrations should be related to the higher potential of anaerobic conditions in the tanks,

consequently leading to odor emissions. Moreover, this variable (RAS east) better explains the anaerobic conditions of sludge at the bottom of the tank than the height of the sludge blanket level shown to be significant in Gabriel, et al., 2005 and 2006.The reason is that the blanket level may only represent the height of solids loosely dispersed in the tank and thus it was not significant in this study.

4.5.3 Validation

To validate selected models, we created 20 validation sets, each of which contained 15 observations (20 % of full data) that were randomly selected by the statistical software XLMiner (http://xlminer.com). For each validation set, four models were selected; the full data set model, the summer model, the non-summer model, and the high odor model. The Mean Absolute Error (MAE) was used to evaluate errors between predicted DT and actual DT values because it was less sensitive to large forecast errors compared to Mean Squared Error (MSE). The Mean Absolute Errors (MAE) for each validation set was recorded and is presented in Table 4.10. At the bottom of Table 4.10, average MAEs from the 20 validation sets are provided.

Table 4.10: Mean Absolute Error (MAE) on validation sets

	MAE full model	MAE non summer model	MAE summer model	MAE high odor model
set 1	**1109**	1253	1632	1986
set 2	**884**	1202	1078	2172
set 3	**755**	1096	1143	2258
set 4	924	**669**	1157	2678
set 5	1111	**955**	1426	2431
set 6	902	**806**	1278	2401
set 7	831	**653**	927	2717
set 8	**731**	757	1119	2714
set 9	684	**639**	754	2106
set 10	940	**776**	1062	2170
set 11	**963**	1293	1252	2122
set 12	918	**814**	1089	2630
set 13	998	**887**	1483	2316
set 14	**820**	1141	1420	2155
set 15	**678**	974	1166	2490
set 16	**1308**	1806	1540	2158
set 17	**1191**	1232	1213	2529
set 18	**697**	714	967	2024
set 19	**1381**	1490	1437	2668
set 20	**781**	784	955	2752
Average	**930**	997	1205	2374

Note: Number in **bold character** represents the model with lowest SSE in the data set

On average the full data model outperformed the rest of the selected models with the lowest average MAE. The non-summer model was the second with the lowest MAE in 8 of 20 validation sets. The summer and high odor models were the third and fourth, respectively. The full data model had the greatest number of independent variables and this may be the reason that it performed best

Another approach to minimize the validation bias was to retain a few observations (i.e., seven observations from the collected data) for a test set and use the remaining observations (i.e., 70 observations) to re-estimate the four selected models. Thus, we used different sets of observations for the test set and the estimation set. Ten test sets were created in this study. For each test set, the coefficients for the independent variables for

the four selected models were re-estimated and they were used to predict the DT levels. The error between predicted and actual DT levels was computed using the Mean Absolute Error (MAE). The lower the MAE values the better the ability that model can predict the DT levels. Table 4.11 presents the MAE values from re-estimated models for the test sets.

Table 4.11: MAE values from re-estimated models on the test sets

	MAE full model	MAE summer model	MAE non-summer model	MAE high-odor model
Test set 1	1422.75	1623.57	1469.54	**1112.26**
Test set 2	**1525.81**	1778.85	1718.99	1940.36
Test set 3	**1704.33**	1888.99	1941.68	2164.12
Test set 4	1054.67	**819.48**	1164.41	1185.51
Test set 5	**1016.25**	1154.17	1018.54	1471.96
Test set 6	**913.38**	1088.61	1380.30	1433.42
Test set 7	**805.27**	1090.31	965.91	997.05
Test set 8	1282.32	1083.79	**865.51**	1000.41
Test set 9	**1250.81**	1342.96	1349.11	1731.23
Test set 10	708.29	711.56	**669.23**	867.30
Average	**1168.39**	1258.23	1254.32	1390.36

Note: Numbers in bold represent the model with the lowest MAE for the data set in question

From Table 4.11, the model for the full data set had the lowest MAE on average. Therefore, the model for full data still performed best for the test sets.

Next we introduce new approach to analyze biosolids odor data when they are categorical as opposed to continuous-valued.

4.6 Approach 3:Prediction Classification

Typically, some of basic information that biosolids managers want to know each day is whether the biosolids produced today will be odorous. This information helps biosolids managers to decide appropriate sites for biosolids distribution taking into

account other considerations as well, i.e., population density, availability of field space, etc. as discussed in Sahakij (2007).

Biosolids odor data collected in this study are continuous-valued response variables which if needed can be converted to binary variables. For example, the detection threshold data can be classified to either high biosolids odor or low biosolids odor which can be more practical. DTWL24 greater than 1000 ou was regarded as a relatively high biosolids odor level and the odor level at 1000 ou and below was regarded as low biosolids odor. Figure 4.8 illustrated DTWL24 data collected in this dissertation. There were 33 observations (43%) having DT levels greater than 1000 ou and 44 observations (57%) equal to or lower than 1000 ou.
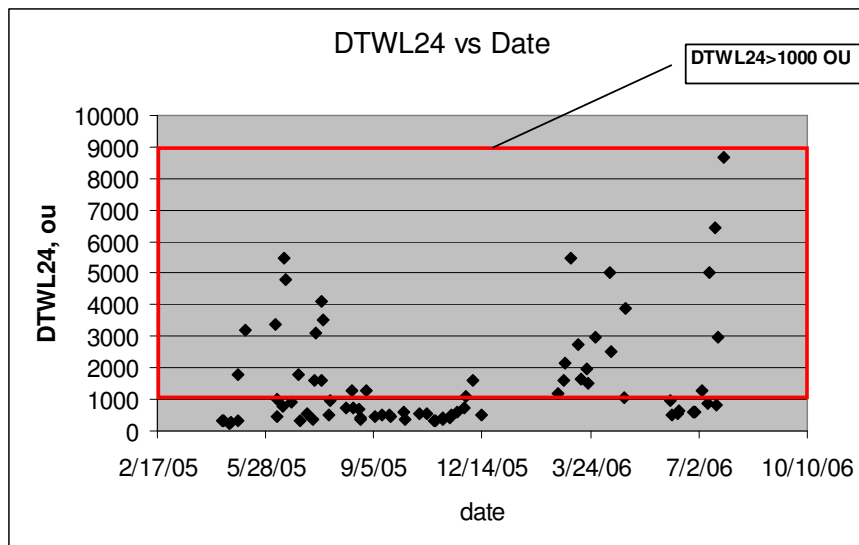
Figure 4.8: Graph showing DT levels of data collected greater than 1000 ou

This section provides alternative tools to categorize biosolids odor into groups that can support the decision making process for biosolids managers

4.6.1 Model Choices

Violation of ordinary least square (OLS) assumptions

Using a linear statistical model with ordinary least squares (OLS) is not appropriate when the response variable is categorical for the following reasons: (Shmueli, Patel, and Bruce, 2007)

1. When using binary response variables, the outcome of the OLS model can give predicted values other than zero and one.

2. The normality assumption for the residuals is violated due to binary response variables.

3. The assumption of constant variance of the residuals is violated due to the binary response variables.

Logistic regression (LGR) and discriminant analysis (DA) are two models widely used in categorical data analysis (Lee, et al., 2006). Discriminant analysis is a classification technique proposed by Fisher (1936) to classify set of explanatory variables into classes. DA is developed under the basic assumption that set of explanatory variables in the model follows a multivariate normal distribution. The advantages of discriminant analysis are parsimony and robustness of the model even for a small data set (Shmueli, Patel, and Bruce, 2007). We can find DA application in various fields; finance and bankruptcy (Altman, 1968; Lee et al., 1997), business failure (Deakin, 1972; Bardos, 1998), marketing research (Kim et al., 2000), investment (Trevino and Daniels, 1995), and credit scoring models (Desai et al., 1996; Martell and Fitts, 1981; and Reichert et al., 1983).

Logistic regression models the probability of success in terms of log of the odds (i.e., log of probability of success over probability of failure) which can be explained by

independent variables in the same fashion as linear regression models. Underlining basic assumptions (e.g., normality of residual) are not required under LGR. LGR is still robust even when the basic assumption of normality of residuals in discriminant analysis is not met (Harrell and Lee, 1985). The applications of LGR can be found in various fields; predicting firm bankruptcy (Flagg et. al., 1991; Laitinen, 2000), identifying customer behavior in business (Kay et al., 2000), market segmentation (Suh et al, 1999), and credit scoring models (Joanes (1993); Laitinen (1999); Westgaard and Van, (2001); Wiginton (1980))

Next, we give a brief introduction to discriminant analysis and show examples of how it works in assigning response variable into class.

4.6.2 Discriminant Analysis

In discriminant analysis, linear classification functions are linear functions of independent variables used to classify an observation to one of the two or more groups. It is developed for each group of interest to find the line or plane (i.e., more than two independent variables) that has equal statistical distance (Mahalanobis distance) from the two group means.

In particular, let X represent a vector of p independent variables in a discriminant function with $X = [x_1, x_2, ..., x_p]$ and $\overline{X}$ represent a vector of average independent variables in a discriminant function with $\overline{X} = [\overline{x}_1, \overline{x}_2, ..., \overline{x}_p]$

Thus, the statistical or Mahalanobis distance can be expressed as

$$D(X, \overline{X}) = \sqrt{(X - \overline{X})^T S^{-1} (X - \overline{X})} \qquad (4.3)$$

where

$$(X - \bar{X})^T S^{-1} (X - \bar{X}) = [(x_1 - \bar{x}_1), (x_2 - \bar{x}_2), ..., (x_p - \bar{x}_p)] S^{-1} \begin{bmatrix} (x_1 - \bar{x}_1) \\ (x_2 - \bar{x}_2) \\ ... \\ (x_p - \bar{x}_p) \end{bmatrix}$$

S^{-1} = inverse of the covariance matrix between the p-variables, which is a symmetric positive definite square matrix with dimension p

For each classification function, the classification score of each observation is computed by coefficients and corresponding values of independent variables from the classification function (Shmueli, Patel, and Bruce, 2007).

Let

z_{1i} = classification score for observation i in group 1 and z_{2i} = classification score for observation i in group 0

a_{10} = intercept for classification function 1

a_{20} = intercept for classification function 0

a_{1m} = weight on independent variable m of classification function for group 1 $m=$ 1,2,...,p

a_{2m} = weight on independent variable m of classification function for group 0 $m=$ 1,2,...,p

Then, classification functions with respect to group 1 and group 0 are

$$z_{1i} = a_{10} + a_{11}x_{1i} + a_{12}x_{2i} + ... + a_{1p}x_{pi} \tag{4.4}$$

$$z_{2i} = a_{20} + a_{21}x_{1i} + a_{22}x_{2i} + ... + a_{2p}x_{pi} \tag{4.5}$$

Consequently, each observation is assigned its group according to the highest classification score. In what follows we give an example of discriminant analysis.

Example: Discriminant analysis for a biosolids odor prediction model.

To better understand the discriminant analysis, an example of how to compute the classification score is presented here. Suppose data for independent variables are available and we want to classify biosolids odor levels as either: high odor (class 1) or low odor (class 0) using XLminer, the data mining add-in for Excel software. The output from discriminant analysis calculated by XLminer is

Table 4.12: Sample of classification function from XLminer

Classification Function		
Variables	Classification Function	
	1	0
Constant	-84.1146927	-78.9614105
% solid	2.85912609	3.06298923
With lime 24 T	3.68128753	3.25805616

From Table 4.12, the classification function of class 1 is

$$z_{1i} = -84.115 + 2.859 * \%solids + 3.681 * withline24T$$

and the classification function of class 2 is

$$z_{2i} = -78.961 + 3.063 * \%solids + 3.258 * withline24T$$

If we know that %solids on a particular sample is 23.58 % and the sample temperature is 18.5 Celsius, the classification score of class 1 is 51.40 and the classification score of class 0 is 53.54. Thus, according to classification scores, odor level from this sample is classified into class 0, low odor group.

In XLminer, we evaluate performance of classification functions by 1) a classification confusion matrix and 2) an error report as shown in Table 4.13.

Table 4.13: Sample of classification confusion matrix and error report from XLminer

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	25	5
0	14	29

Error Report			
Class	# Cases	# Errors	% Error
1	30	5	16.67
0	43	14	32.56
Overall	73	19	26.03

The classification confusion matrix reports the actual number of observations for class 1 and class 0 from the data set and the prediction results for class 1 and class 0 performed by the classification functions. For example, on row one in the classification confusion matrix, there were 25 observations that were correctly classified into class 1 and 5 observations that were wrongly classified as class 0 instead of class 1. Analogously, on row two there were 14 observations that were wrongly classified into class 1 instead of class 0 and there were 29 observations that were correctly classified into class 0.

The error report summarizes the percentages of error in class 1, in class 0, and the overall error rate. From Table 4.13, of the actual 30 observations in class 1 in the data set, five observations were wrongly classified (16.67%). Analogously, of the 43 observations from class 0, 14 observations were wrongly classified (32.56%). Lastly, for the total performance of these classification functions, 19 observations in total from both class 0 and class 1 were wrongly classified (26.03%). Next, we introduce the logistic regression.

4.6.3 Logistic Regression

Unlike discriminant analysis, logistic regression does not have an underlying assumption of multivariate normal distribution between the independent variables. Logistic regression can be used in the same way as OLS, such as predicting future values, explaining the contribution of independent variables on response variable, applying dummy variables techniques, etc. The difference is that logistic regression models the discrete response variable in terms of a probability.

Assuming there are two classes, logistic regression would model the probability of being in class 1 over the probability of being in class 0 in terms of log of the odds (logit).

In particular, let P = probability of being in class 1 and $1 - P$ = probability of being in class 0

The odds of being in class 1 over class 0 is then

$$\text{Odds} = \frac{P}{1-P} \tag{4.6}$$

Then, log(odds) can be explained by a linear function of independent variables

$$\text{Log (odds)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_q x_q \tag{4.7}$$

or

$$\text{Odds} = e^{\beta_{0+} + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_q x_q} \tag{4.8}$$

In addition, the success probability can be illustrated in the form

$$P = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_q x_q}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_q x_q}} \tag{4.9}$$

Example: Logistic regression for a biosolids odor prediction model.

Suppose that we obtain the following:

Log (odds) $= -4.529 - 0.157 * \%SOLIDS + 0.342 * with\lim e24T$

And, we know that a % solid for a particular sample was 30 % and the sample temperature was 30 Celsius.

Then the probability of being in class 1 can be computed by

Log (odds) $= -4.529 - 0.157 * \%SOLIDS + 0.342 * with\lim e24T = 1.02116$

and

$$P = \frac{e^{-4.529 - 0.157 * \%SOLIDS + 0.342 * with\lim e24T}}{1 + e^{-4.529 - 0.157 * \%SOLIDS + 0.342 * with\lim e24T}}$$

Thus,

$$P = \frac{e^{1.02116}}{1 + e^{1.02116}} = 0.735$$

Consequently, a logistic regression model would predict that there is a 74 percent chance that the collected sample would be in class 1 and a 26 percent chance that it would be in class 0.

Next, we present the results from classification functions and logistic regression models as applied to biosolids odor data.

4.6.4 Results

To reduce time spent in the variable selection process, we developed categorical models based on significant variables identified in Section 4.5 (Approach 2: Predict Continuous Odor Levels). Those variables were

- % solids

- With lime 24 T

- GT % in blend d-1

- RAS east (mg/l) d-1

- TPS d-1

- GT pH

- BS JRM

- # of centrifuges running

- Interaction RAS east d-1 and GT % in blend d-1 (RAS east * GT d-1)

- Interaction RAS east d-1 and TPS d-1 (RAS east * TPS d-1)

Discriminant analysis results

Table 4.14 summarizes models that classified biosolids odor levels into a high odor group and a moderate-to-low (MTL) group. The top two models were model D1 and model D2. Model D1 uses % solids and temperature of the biosolids sample as independent variables. Model D2 also includes these independent variables plus % of GT solids in blend tank, RAS east d-1, GT pH, and # of centrifuge running as independent variables. Model D2 uses the same independent variables as the full data set model in Section 4.5. Considering the performance of both models in terms of misclassification rates, model D1 outperformed D2 by misclassifying 21% in class 1 (7 observations) compared to 24% for D2 (8 observations). Models D3 and D4 also performed moderately well in terms of the error rate in class 1 with 24% and 27% respectively.

Table 4.14: Discriminant analysis results

Parameter in function	Discriminant function name														
	D 1	D 2 [1]	D 3 [2]	D 4 [3]	D 5	D 6	D 7	D 8	D 9	D 10	D 11	D 12 [4]	D 13	D 14	D 15
% solids	x	x		x	x		x	x	x	x					
With lime 24 T	x	x	x	x	x	x	x	x		x	x				
GT% in blend d-1		x		x		x						x			
RAS east d-1		x	x	x						x		x		x	
TPS d-1														x	
GT pH		x	x			x									
BS JRM								x							
# of centrifuges running		x													
RAS east*GT% d-1					x				x		x		x		
RAS east*TPS d-1						x									x
Performance															
% error in class 1	**21%**	24%	24%	27%	27%	27%	27%	30%	30%	33%	33%	33%	33%	36%	39%
% error in class 0	25%	**20%**	27%	20%	25%	25%	32%	30%	34%	23%	23%	32%	32%	32%	39%
overall	23%	**22%**	26%	23%	26%	26%	30%	30%	32%	27%	27%	32%	32%	34%	39%
number of miss-classification in class 1	**7**	8	8	9	9	9	9	10	10	11	11	11	11	12	13
number of miss-classification in class 0	11	**9**	12	9	11	11	14	13	15	10	10	14	14	14	17
overall	18	**17**	20	18	20	20	23	23	25	21	21	25	25	26	30

Note:
[1] this function used same variables as full data set model in approach 2
[2] this function used same variables as non-summer model in approach 2
[3] this function used same variables as summer model in approach 2
[4] this function used same variables as high odor model in approach 2
x represents variable in the function

Model D1 can be shown as follows.

Classification function of class 1

$$z_{1i} = -74.671 + 2.784 * \%solids + 3.062 * withline24T \qquad (4.10)$$

Classification function of class 2

$$z_{2i} = -70.219 + 2.942 * \%solids + 2.710 * withline24T \qquad (4.11)$$

Model D2 can be shown as follows.

Classification function of class 1

$$z_{1i} = -330.515 + 2.647 * \%solids + 4.811 * withline24T + 78.475GTpH - 2.121centrif$$

$$+0.377 * GT\%inblendd - 1 + 0.001RASeastd - 1 \qquad (4.12)$$

Classification function of class 0

$$z_{2i} = -310.994 + 2.849 * \%solids + 4.401 * withline24T + 76.381GTpH - 2.019centrif$$

$$+0.341 * GT\%inblendd - 1 + 0.001RASeastd - 1 \qquad (4.13)$$

Variables	Classification Function	
	1	0
Constant	-330.514648	-310.993683
% solid	2.64700723	2.84917426
With lime 24 T	4.8105731	4.40281057
GR PH	78.47462463	76.38134003
# centrif running	-2.12150145	-2.01898265
GT% in blend d-1	0.3767345	0.34115428
RAS east (mg/l) d-1	0.00124968	0.00103495

Logistic regression results

Similar to discriminant analysis, for logistic regression we used those significant variables identified at the beginning of subsection 4.6.4 in the variable selection process. Table 4.15 summarizes the performance of the resulting LGR models. It can be seen that models L1, L2, and L3 performed best with the same misclassification rate in class 1 of

33%. However, for the misclassification rate for class 0, model L1 outperformed models L2 and L3 with a lower value of 18% compared to 20%, and 23% for models L2 and L3 respectively.

Table 4.15: Logistic regression results

Parameter in function		L 1[3]	L 2[1]	L 3	L 4	L 5	L 6	L 7	L 8	L 9	L 10[2]	L 11	L 12[4]	L 13	L 14	L 15
	Logistic regression model name															
% solids		x	x	x		x	x		x	x				x		
With lime 24 T		x	x	x	x	x	x	x	x	x	x					
GT% in blend d-1		x	x	x									x			
RAS east d-1		x	x						x			x	x	x		
TPS d-1												x				
GT pH			x					x				x				
BS JRM										x *						
# of centrifuges running			x *													
RAS east*GT% d-1					x	x								x	x	
RAS east*TPS d-1								x								x
Performance																
% error in class 1		**33%**	33%	33%	39%	39%	39%	39%	36%	36%	42%	52%	52%	52%	58%	70%
% error in class 0		**18%**	20%	23%	16%	16%	18%	20%	18%	18%	16%	23%	23%	25%	20%	11%
overall		25%	26%	27%	26%	26%	27%	29%	26%	26%	27%	35%	35%	36%	36%	36%
number of miss-classification in class 1		**11**	11	11	13	13	13	13	12	12	14	17	17	17	19	23
number of miss-classification in class 0		**8**	9	10	7	7	8	9	8	8	7	10	10	11	9	5
overall		19	20	21	20	20	21	22	20	20	21	27	27	28	28	28

Note:
[1] this function used same variables as full data set model in approach 2
[2] this function used same variables as non-summer model in approach 2
[3] this function used same variables as summer model in approach 2
[4] this function used same variables as high odor model in approach 2
x   represents variable in the function
*   represents varaibles with p-value greater than 0.2

The best logistic regression model included four variables: % solids, with lime 24T, GT% in blend d-1, and RAS east d-1. All independent variables in the model were statistically significant at 10 % significant levels and had the correct coefficient's sign. The following is the equation of model L 1, the best logistic regression model

$$Y = -7.308 - 0.202x_1 + 0.355x_2 + 0.037x_3 + 0.0002x_4 \qquad (4.14)$$

where

Y = Log (odds) of the observation being in the high odor group

x_1 = percent solids of biosolids sample

x_2 = temperature of biosolids sample

x_3 = percent of GT solids in blend tank

x_4 = concentration of RAS at the east side

Thus, we see that

$$P = \frac{e^{-7.308 - 0.202x_1 + 0.355x_2 + 0.037x_3 + 0.002x_4}}{1 + e^{-7.308 - 0.202x_1 + 0.355x_2 + 0.037x_3 + 0.002x_4}} \qquad (4.15)$$

where

P = probability of the observation being in the high odor group

4.6.5 Comparing Results from DA and LRG

Using results in Figures 4.12 and 4.13, we compared the performance of the best classification function D1 with the best logistic regression model L1. We found that models D1 and D2 (the top two models from DA) outperformed models L1 and L2 (the top two models from LR) in terms of lower % misclassification rate in class 1 with 21% and 24% respectively compared to 33% of models L1 and L2. Misclassification in class 1

is important since it means that the model predicts low odor biosolids while the actual biosolids odor level is high. Given these results, biosolids managers could mistakenly assign odorous biosolids to sensitive application sites leading to odor incident around those areas.

4.6.6 Validation

The 20 data sets that we used to validate the models in subsection 4.5.3 were used here to validate our classification models. Table 4.16 illustrates how well models D1, D2, L1, and L2 performed on each validation set. Table 4.16 includes the sum of correct predictions, number of observations that were correctly predicted in class 1 and class 0, and the number of observations that were wrongly predicted in class 1 and class 0 performed by each model. In terms of overall correct prediction (correct observations for both class 1 and class 0), model D1 performed best on 13 of 20 validation sets while model L1 performed best on 10 of 20 validation sets. However, four of these sets from model L1 including validation sets 2, 10, 12, and 14 had the same number of correct predictions as model D1.

Therefore, regarding the overall performance, model D1 outperformed models D2 and L1 in terms of 1) overall prediction (highest correct prediction), 2) prediction on class 1 (highest correct prediction for class 1), and most importantly 3) misclassification for class 1 (lowest in misclassification in class 1).

Table 4.16: Validation results on categorical models

	Actual observation in class 1	Actual observation in class 0	Model D 1					Model D 2					Model L 1					Model L2				
			Sum of correctly predict	Correctly predict class 1	Correctly predict class 0	Error class 1	Error class 0	Sum of correctly predict	Correctly predict class 1	Correctly predict class 0	Error class 1	Error class 0	Sum of correctly predict	Correctly predict class 1	Correctly predict class 0	Error class 1	Error class 0	Sum of correctly predict	Correctly predict class 1	Correctly predict class 0	Error class 1	Error class 0
set 1	6	9	9	4	5	2	4	11	6	5	1	3	11	4	7	2	2	10	6	4	2	3
set 2	7	8	11	5	6	2	2	12	6	6	1	2	11	4	7	3	1	12	6	6	1	2
set 3	7	8	13	7	6	0	2	14	7	7	0	1	3	7	6	0	2	13	7	6	1	1
set 4	1	14	10	1	9	0	5	11	10	1	0	4	11	1	10	0	4	11	10	1	0	4
set 5	4	11	11	4	7	0	4	11	7	4	0	4	9	3	6	1	5	11	7	4	0	4
set 6	6	9	10	3	7	3	2	10	7	3	3	2	10	2	8	4	1	10	7	3	3	2
set 7	5	10	10	2	8	3	2	12	9	3	2	1	11	2	9	3	1	12	9	3	2	1
set 8	3	12	12	2	10	1	2	12	10	2	1	2	11	0	11	3	1	11	10	1	2	2
set 9	3	12	13	3	10	0	2	13	11	2	1	1	12	2	10	1	2	13	11	2	1	1
set 10	5	10	11	4	7	1	3	13	8	5	0	2	11	2	9	3	1	12	8	4	1	2
set 11	8	7	10	7	3	1	4	10	5	5	3	2	11	6	5	2	2	10	5	5	3	2
set 12	5	10	10	3	7	2	3	7	5	2	3	5	10	2	8	3	2	7	5	2	3	5
set 13	4	11	11	3	8	1	3	13	9	4	0	2	14	4	10	0	1	13	9	4	0	2
set 14	5	10	12	4	8	1	2	12	9	3	2	1	12	3	9	2	1	12	9	3	2	1
set 15	10	5	13	8	5	2	0	13	4	9	1	1	10	5	5	5	0	10	4	6	4	1
set 16	7	8	13	7	6	0	2	11	5	6	1	3	12	7	5	0	3	11	5	6	1	3
set 17	3	12	11	2	9	1	3	8	7	1	2	5	12	2	10	1	2	8	7	1	2	5
set 18	6	9	9	4	5	2	4	11	7	4	2	2	10	2	8	4	1	11	7	4	2	2
set 19	7	8	11	5	6	2	2	10	5	5	2	3	10	3	7	4	1	9	5	4	3	3
set 20	5	10	12	3	9	2	1	12	9	3	2	1	11	2	9	3	1	11	9	2	3	1
	107	193	222	81	141	26	52	226	146	80	27	47	212	63	159	44	34	217	146	71	36	47

4.6.7 Validation: Unequal Misclassification Costs

Previously, we assume the cost of misclassifying class 1 (odorous) to class 0 (non-odorous), called misclassification type 1, was equal to the cost of misclassifying class 0 to class 1, called misclassification type 2. In reality, these costs are not the same. Misclassification type 1 causes distributing the odorous biosolids to sensitive application sites resulting in complaints from community. On the other hand, misclassification type 2 causes hauling the biosolids with low odor to remote sites resulting in unnecessary hauling costs (e.g. gas and fewer trips to deliver biosolids per day). Therefore, an equation to measure the total misclassification cost for a particular model with respect to the unequal misclassification costs was developed in equation 4.16.

$$M_i = C_{10}*(n_{10}/N_1) + C_{01}*(n_{01}/N_0) \tag{4.16}$$

where

M_i = the total misclassification cost for model i

C_{10} = the cost associated with misclassifying class 1 to class 0 (type 1)

C_{01} = the cost associated with misclassifying class 0 to class 1 (type 2)

n_{10} = the number of observations misclassified as class 0

N_1 = the total actual observations in class 1

n_{01} = the number of observations misclassified as class 1

N_0 = the total actual observations in class 0

This equation combines the total costs (M_i) for misclassification in types 1 and 2 from model i on a particular data set when the misclassification costs are not equal. From the equation, (n_{10} / N_1) and (n_{01} / N_0) are the probabilities of misclassification

types 1 and 2, respectively and C_{10} and C_{01} are the associated costs for misclassification types 1 and 2, respectively.

In this study, C_{10} is assumed to be greater than C_{01} since applying odorous biosolids to sensitive communities would lead to the opposition to the biosolids as well as the ban on field application in the area. Five ratios of misclassification costs 3:1, 5:1. 7:1, 9:1, and 10:1 were selected to assess how the selected models perform under unequal misclassification costs. The ratio 3:1, for example, means it costs three times more to misclassify type 1 than type 2.

On the selected models, we applied equation 4.16 to all classification results from 20 validation sets. For example, on a particular validation set if model L1 misclassified two from six of biosolids odor in class 1 (misclassification type 1) and misclassified two from nine of biosolids odor in class 0 (misclassification type 2), then (n_{10} / N_1) was equal to $(2/6)$ and (n_{01} / N_0) was equal to $(2/9)$ on this validation set.

For a misclassification cost ratio 3:1 $(C_{10} : C_{01})$, a misclassification cost for model L1 on this validation set is

$$M_i = C_{10}*(n_{10} / N_1) + C_{01}*(n_{01} / N_0)$$

$$= 3*(2/6) + 1*(2/9)$$

$$= 1.22$$

Averages misclassification costs from 20 validation sets for models, L1, L2, D1, and D2 were summarized in Table 4.17. For instance, at a ratio of 3:1 an average misclassification cost for model D1 from 20 validation sets was 0.98 and an average misclassification cost for model L1 at the same ratio was 1.39.

Figures 4.9, plot the average misclassification cost for each model at different ratios. It appears that model D1 and D2 still outperformed model L1 and L2 when we assigned unequal misclassification costs. Note that now model L2 surpassed model L1 when unequal misclassification costs were considered.

Table 4.17: Average misclassification costs from selected models relative to various ratios

Ratio C10:C01	Average cost			
	D 1	D 2	L 2	L 1
3:1	0.98	1.02	1.24	1.39
5:1	1.46	1.54	1.91	2.21
7:1	1.94	2.06	2.58	3.02
9:1	2.42	2.58	3.24	3.83
10:1	2.66	2.83	3.58	4.24
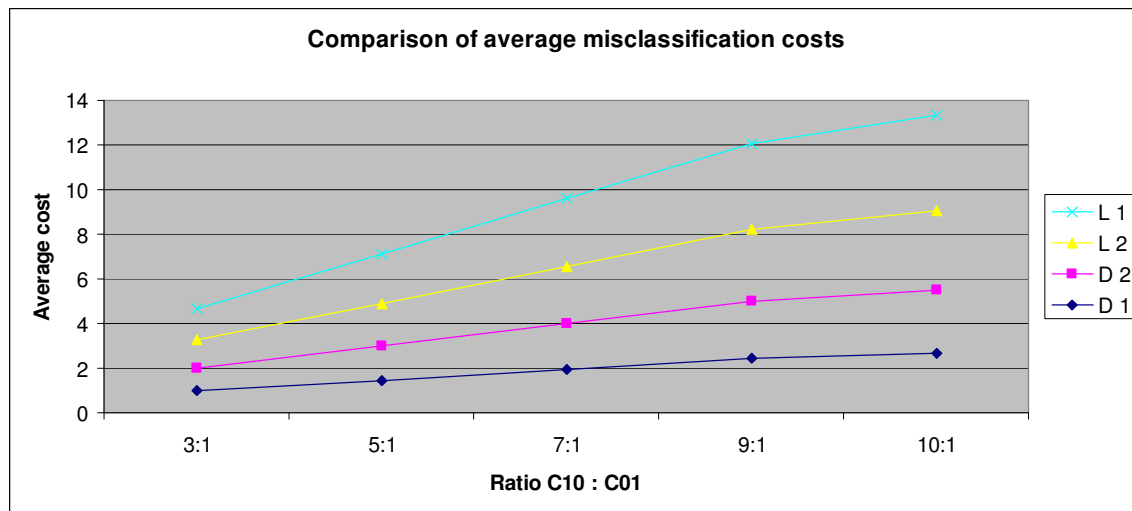
Figure 4.9: Plot of average misclassification costs from selected models

In summary, the discriminant functions D1 performed best in both the training and the validation sets followed by D2. Two variables for biosolids properties; temperature and % solids were significant in the selected models for approaches 2 (continuous biosolids odor case) and 3 (categorical case) in this chapter. In the end, after the final results were discussed with DCWASA personnels (C. Peot and M Ramirez,

Personal communication, June 27, 2008), they suggested that model D2 (six variables) was more appropriate to use in practice than the model D1 since it included six variables, four from operational variables and two from characteristics of biosolids, that may be better to monitor the fluctuation in daily operation than the two variables identified in the model D1. Also, the average misclassification costs from both models were close.

In the next chapter we run what-if analysis on the selected model from approach 2 considering uncertainty that commonly exists in real-world system. We simulate the potential outcomes from the biosolids odor prediction model taking probability distributions of predictors in the model as inputs. In addition, we investigate impacts of each independent variable in the model on biosolids odor production when uncertainty is involved. The end goal of this analysis is to produce a probability distribution for biosolids odor levels.

Chapter 5: Simulation and Sensitivity Analysis

In this chapter, we show results from Monte Carlo simulations to understand how the uncertainty in the independent variables from previously identified models affect biosolids odor production. As a consequence, the biosolids odor profile at Blue Plains is shown and sensitivity analysis is performed to assess the significance of each independent variable in the model.

5.1 Monte Carlo Simulation

The study of uncertainty in the model assists biosolids managers in gaining insights into biosolids odor production. Additionally, the sensitivity of odor levels to independent variables is important for biosolids management and to this end Monte Carlo Simulation was used. In particular, the statistically-derived functions that predicted biosolids odor levels were used with the independent variables' values taken to be random variables.

To start the simulation procedure, one must define values of the independent variables to use either by employing theoretical probability distributions derived from historical data or using expert judgment. For instance, from historical data in the years 2005-2006 the sludge blanket level data (at the east side) was best fit by a log-normal distribution as illustrated in Figure 5.1
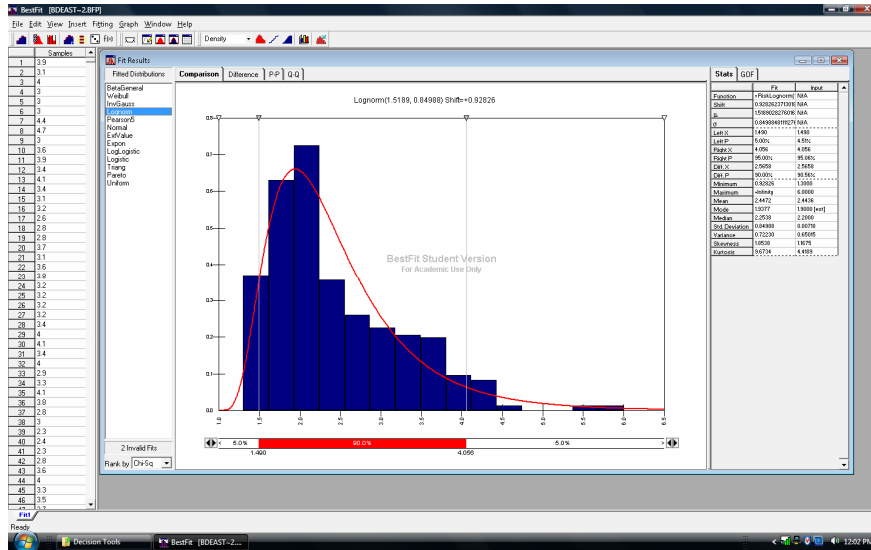
Figure 5.1: Sludge blanket level data years 2005-2006 fitted by log-normal distribution

After appropriate probability distributions for the independent variables were selected, biosolids odor prediction model previously mentioned was used to create the biosolids odor profile or probability distribution for odor levels. Figure 5.2 illustrates an example of the probability distribution of the predicted DT generated by this simulation.
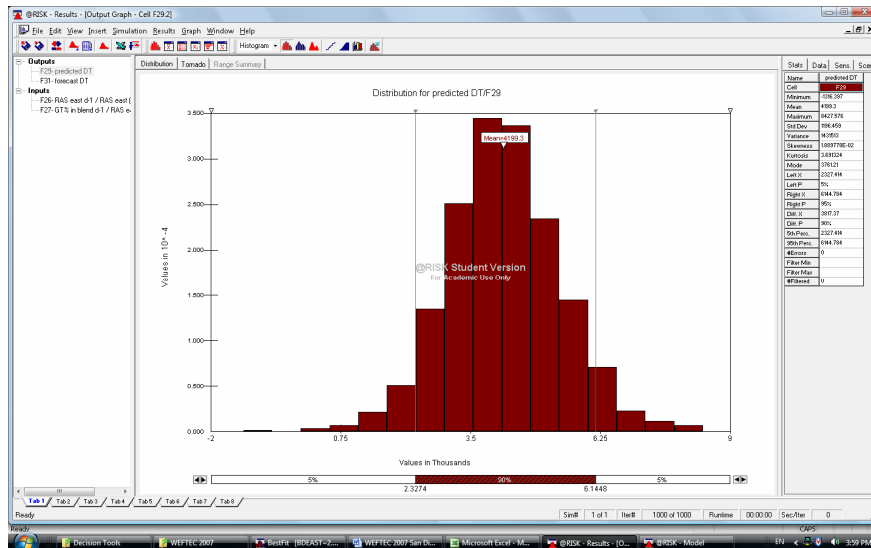
Figure 5.2: A sample of predicted DT distribution generated by simulations

In this thesis, @Risk, a decision analysis software from Palisade Corporation (http://www.palisade.com), was used to perform the simulation analysis. The simulation process was conducted on the full data model discussed in Section 4.5 (Approach 2: Predict Continuous Odor Levels Values) from the previous chapter. DT levels from Monte Carlo Simulation iterations were stored in order to generate an evaluated biosolids odor probability distribution.

The full data set model was as follows:

$$Y = -12936.4 - 141.544X_1 + 256.656X_2 + 1828.693X_3 - 200.721X_4 + 31.349X_5 + 0.159X_6$$

$$(5.1)$$

X_1 = percent solids

X_2 = with lime 24 T

X_3 = GT pH

X_4 = number of centrifuges

X_5 = percent of GT in blend tank d-1

X_6 = RAS east d-1

Next, we present theoretical probability distributions that best fit the historical data for the independent variables just mentioned.

5.2 Assigning Probability Distribution to Independent Variables

Another part of the Palisade software, BestFit, was used to find the best fitting distribution for each of the independent variables. More than 30 theoretical probability distributions were considered (e.g., normal, lognormal, Pearson, etc.). BestFit ranked the distributions based on the Chi-squared goodness of fit.

5.2.1 Chi-Squared Statistic

Chi-squared goodness of fit test is a statistic comparing the predicted number of observations from a specified probability distribution with the actual number of observations from pre-specified bins[2]. The test statistic shown in equation 5.2 follows a chi-squared distribution (Palisade, 2001).

$$\chi^2 = \sum_{i=1}^{K} \frac{(N_i - E_i)^2}{E_i}$$

(5.2)

where

N_i = the observed number of samples in the ith bin

E_i = the expected number of samples in the ith bin

K = number of bins

5.2.2 Probability Distribution Selection Criteria

A chi-squared goodness-of-fit test was used as one of selection criteria. @Risk ranked how well the specified distributions in the software fit the data by a chi-squared statistics. The best distribution was chosen based on fitting capability and also the properties of the distribution that generated data in a plausible range, e.g., nonnegative values.

As opposed to using just the collected data (77 observations), realistic data for any input variables from the PCH database in the years 2005 and 2006 (if available) were

---

[2] Collected data were divided into classes or bins to compare the data in each bin with data from a specified distribution.

used to select the best fit distribution. Those variables were percent solids, percent of GT in blend tank d-1, and RAS east d-1. Probability distributions and parameters of these variables were derived from more than 500 observations truncated at the maximum and minimum values with respect to the collected data used to develop the biosolids odor prediction model. Additional observations from the PCH database helped in selecting the best probability distribution corresponding to variations in real data. The rest of the input variables were derived from the collected data at hand.

Table 5.1 summarizes selected theoretical distributions of independent variables in the model.

Table 5.1: Theoretical probability distributions on input variables in full data set model

Independent variable	Distribution	Parameters
RAS east d-1	BetaGeneral	Alpha1 = 4.4555, Alpha2 = 8.3279, Min = 0.000, and Max = 19588
T sample	BetaGeneral	Alpha1 = 38.339, Alpha2 = 49.343, Min = 0.000, and Max = 54.9751
GT pH	Weilbull	Alpha = 24.353 and Beta = 58454
%GT in the blend d-1	Weilbull	Alpha = 3.4522 and Beta = 58.033
% Solids of sample	Lognormal	Mean = 26.226 and Standard deviation = 2.9617
# of centrifuges running	Discrete	See Table 5.2

Three continuous probability distributions (i.e., BetaGeneral, Weilbull, and lognormal) and one discrete probability distribution were used in the simulation. Descriptions of the distributions selected are provided below.

5.2.3 BetaGeneral Distribution

BetaGeneral distribution in @Risk is defined by "RiskBetaGeneral(α_1, α_2, min, max)". Typically, the Beta distribution has minimum value 0, maximum value 1, and shape parameters α_1 and α_2. BetaGeneral distribution is similar to the standard Beta distribution; however, its minimum and maximum values are user-specified.

The probability density function of BetaGeneral distribution is given as

$$f(x) = \frac{(x - min)^{\alpha_1 - 1}(max - x)^{\alpha_2 - 1}}{B(\alpha_{1,}, \alpha_2)(max - min)^{\alpha_1 + \alpha_2 - 1}}$$

(5.3)

where

α_1 = continuous shape parameter $\alpha_1 > 0$

α_2 = continuous shape parameter $\alpha_2 > 0$

B = the Beta Function

BetaGeneral distribution showed best fit to variables temperature of biosolids sample and concentration of returned activated sludge. The goodness-of-fit results are illustrated in Figures 5.3 and 5.4.
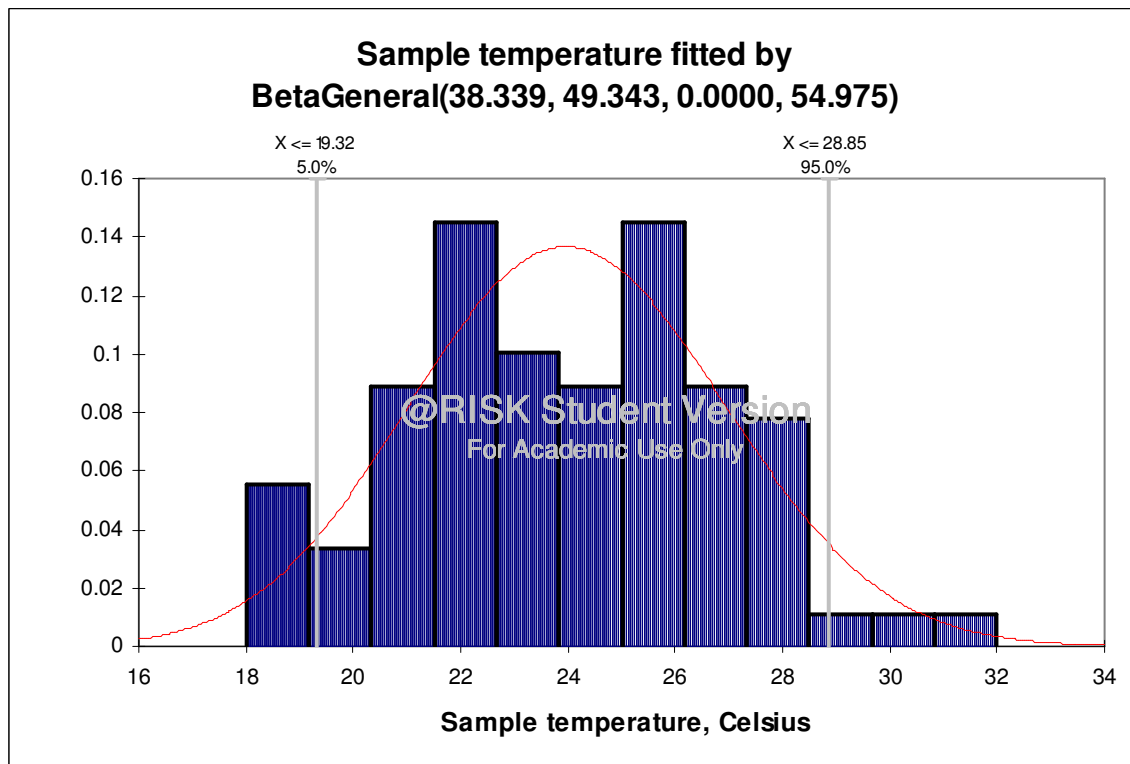
Figure 5.3: Sample temperature (T sample) fitted by BetaGeneral distribution
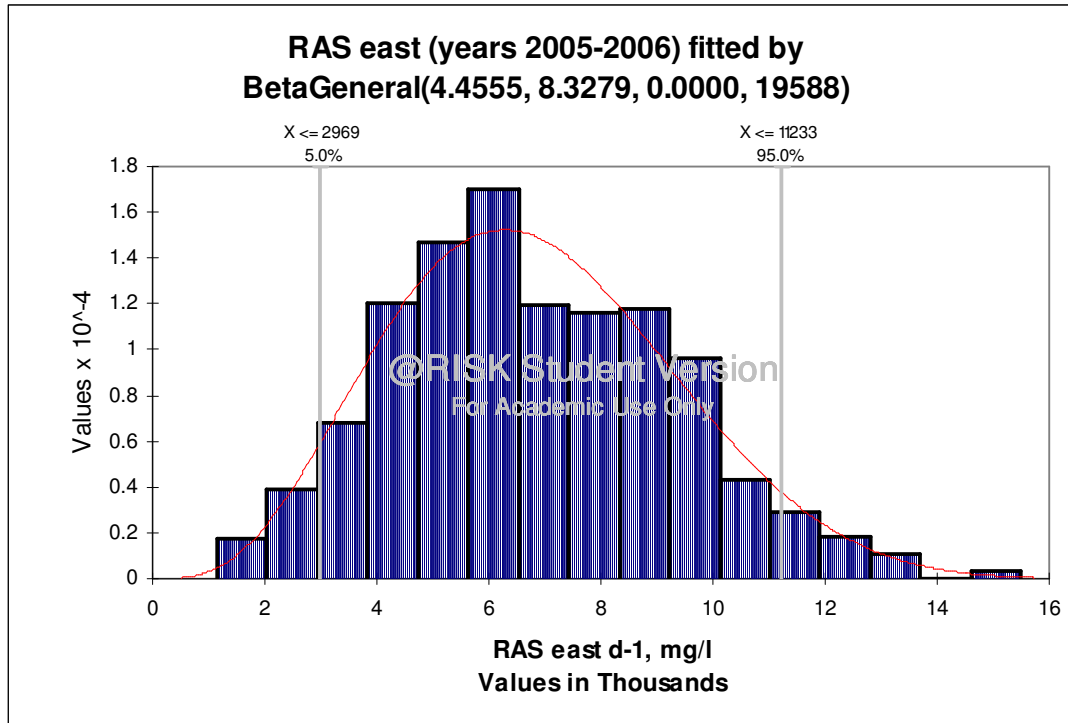
Figure 5.4: RAS east d-1 fitted by Beta (generalized) distribution

5.2.4 Weibull Distribution

The Weibull distribution in @Risk is defined by RiskWeibull(α, β). @Risk generates a Weibull distribution with shape parameter α and scale parameter β with the probability density function given as

$$f(x) = \frac{\alpha x^{\alpha-1}}{\beta^{\alpha}} e^{-(x/\beta)^{\alpha}}$$

(5.4)

where

α = continuous shape parameter $\alpha > 0$

β = continuous scale parameter $\beta > 0$

Domain of x = $0 \le x \le +\infty$

In terms of the chi-squared statistic, the Weibull distribution provided the best fit for the variables %GT in the blend tank and GT pH. Additionally, it generates only positive values which are appropriate to the nature of the input variables. Figures 5.5 and 5.6 illustrate the Weibull distribution fitted to %GT in the blend tank and GT pH.
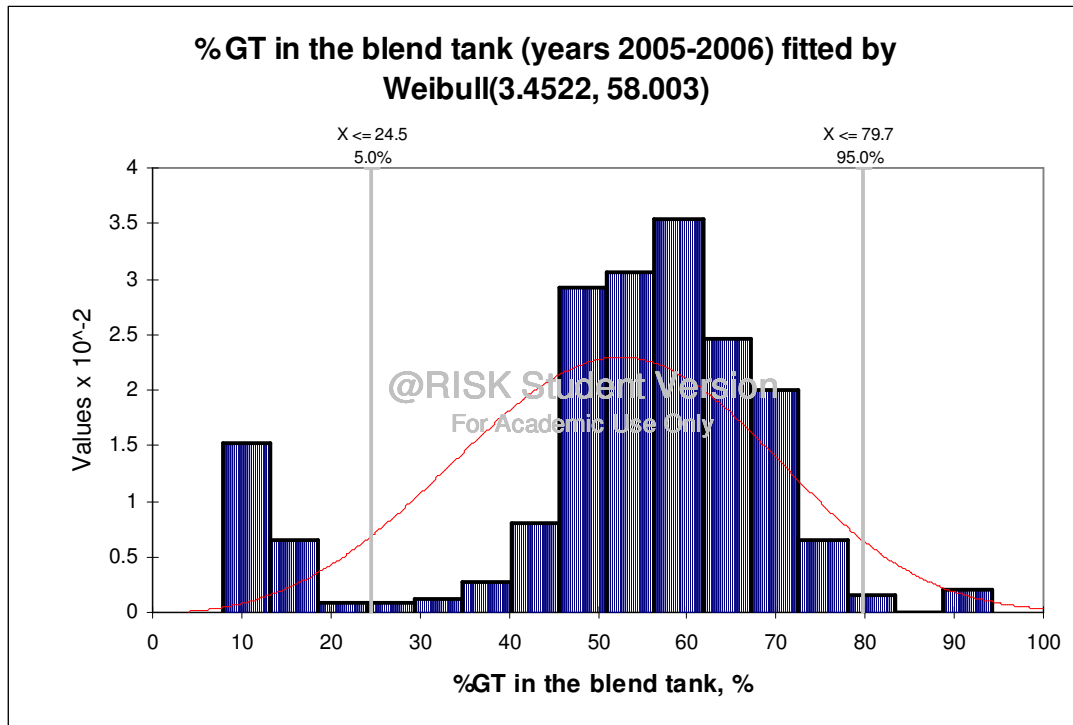
Figure 5.5: % GT in blend tank d-1 fitted by Weibull distribution
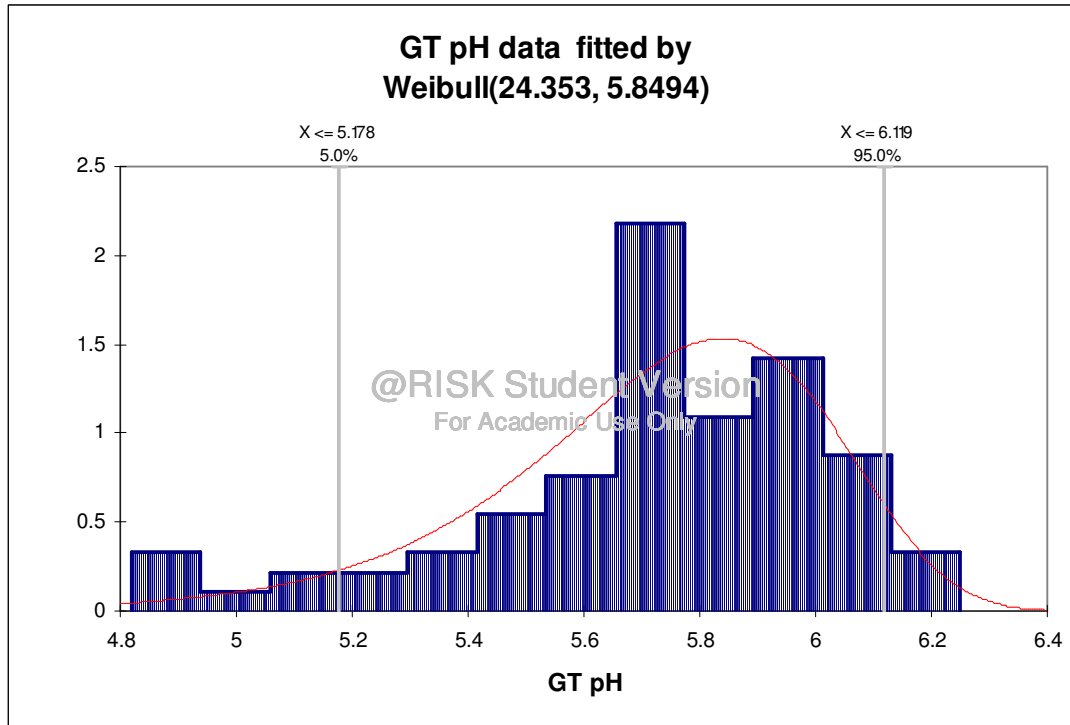
Figure 5.6: GT pH fitted by logistic distribution

5.2.5 Lognormal Distribution

If a random variable x has log normal distribution, then y = log(x) is normally distributed. Also, the values for a log normal distribution range from $0 \leq x < +\infty$. The log normal distribution is defined by two parameters the mean (μ) and the standard deviation (σ). In @Risk, the lognormal distribution is usually defined by RiskLognorm(μ, σ) and a shift factor, where the factor moves the distribution either to the left (+) or to the right (-) to fit the data. The probability density function for a log normally distributed random variable x is as follows.

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma'} e^{-\frac{1}{2}\left[\frac{\ln x - \mu'}{\sigma'}\right]^2}$$

(5.5)

where

$\mu =$ mean $\qquad\qquad \mu > 0$

$\sigma^2 =$ variance $\qquad\qquad \sigma > 0$

$$\mu' \equiv \ln\left[\frac{\mu^2}{\sqrt{\sigma^2 + \mu^2}}\right]$$

$$\sigma' \equiv \sqrt{\ln\left[1 + \left(\frac{\sigma}{\mu}\right)^2\right]}$$

The probability distribution of the variable "percent solids" was best fitted by a lognormal distribution as shown in Figure 5.7.
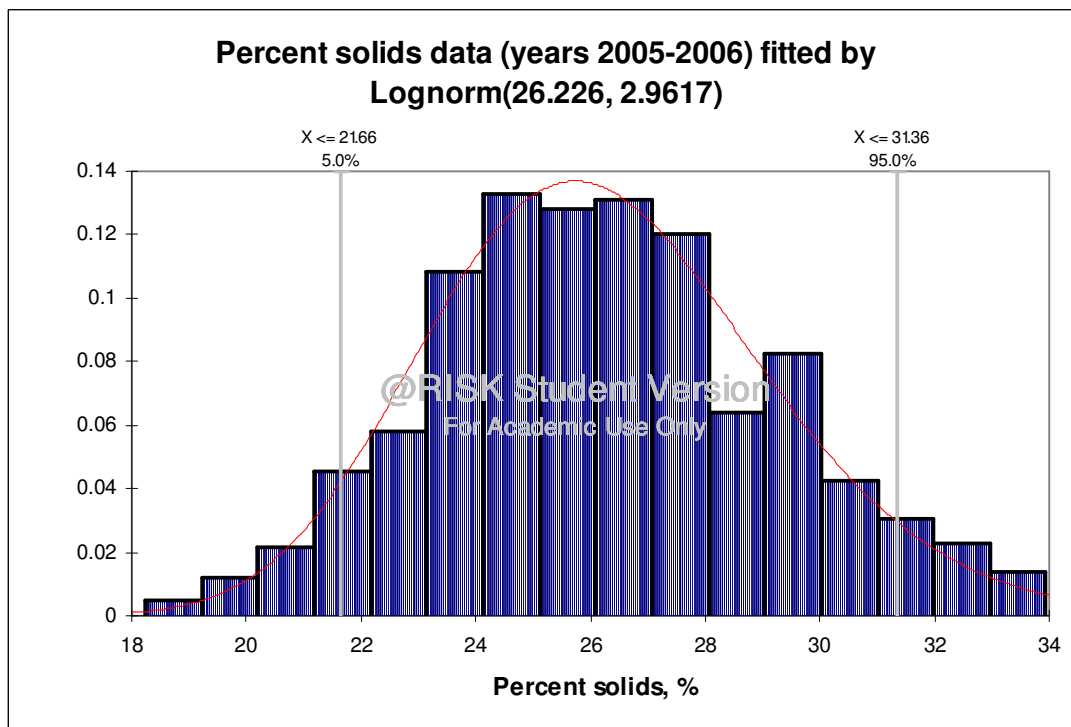
Figure 5.7: Percent solids fitted by log normal distribution

5.2.6 Discrete Distribution

Uncertainty in the number of centrifuges running was fitted by a discrete distribution. The probabilities associated with the number of centrifuges operating are presented in Table 5.2; these probabilities were derived from collected data (years 2005-2006).

Table 5.2: Probability associated with number of centrifuges running

# of centrifuges	Probability
3	0.03
4	0.09
5	0.16
6	0.29
7	0.13
8	0.10
9	0.09
10	0.10
11	0.01
	1

5.2.7 Correlation between Independent Variables

If there are dependencies between input variables in the model, they can influence outcomes of simulation runs (Clemen and Reilly, 2001). Among independent variables in the model, correlations existed in the following pairs:

%GT in the blend tank d-1 and % solids (correlation coefficient 0.22): the more the GT solids in the blend tank the greater the percent solids of biosolids after dewatering as shown by the positive coefficient. Research at DCWASA found that ratio of primary solids and secondary solids over 1.3 (primary: secondary) influences the higher % solids of the dewatered solids (Janpengpen, 2008).

%GT in the blend tank d-1 and the number of centrifuges running (correlation coefficient -0.29): the negative correlation between %GT in the blend tank and the

number of centrifuges running is reasonable since the larger the number of centrifuges to remove blend sludge from the blend tank the lower the amount of GT solids retained in the blend tank as shown by the negative coefficient.

Number of centrifuges running and RAS east d-1 (correlation coefficient -0.28): similarly, the negative correlation between RAS east d-1 and the number of centrifuges running is reasonable since the greater the number of centrifuges available for dewatering helps in removing solids stored in the secondary sedimentation tanks resulting in lower sludge blanket levels and lower RAS concentrations.

Even though moderately weak correlations were found between the independent variables mentioned above, they were still incorporated into the simulation process to better simulate real operation at the Blue Plains plant. The rest of the correlations between independent variables in the model were ignored since there was no reasonable explanation to their values. Next we present the results from the simulation runs.

5.3 Simulation results

Summary statistics of simulation with and without using correlations between input variables and summary statistics of the collected data are shown in Table 5.3.

Table 5.3: Summary statistics of simulation runs

	Comparison of summary statistics		
	Simulation result (without correlation)	Data collection	Simulation result (with correlation)
Maximum	5648	8694	5169
Minimum	-3319	240	-4382
Mean	**1398**	**1582**	**1367**
Median	1449	825	1398
Standard deviation	1214	1697	1284
Skewness	-0.19	-	-0.09

Due to weak correlation coefficients, the outputs from simulations showed a little difference between these two biosolids odor profiles. Figure 5.8 compares biosolids odor profiles from simulations with and without correlation effects. As illustrated, the biosolids odor distribution with correlation is slightly to the left side of the one without correlation. The average DT level from simulation with correlation (1367 ou) was lower than the average DT level from simulation without correlation (1380 ou). Maximum and minimum values from simulation with correlation were also lower than their counterpart. To keep simulation runs realistic, from this point on all analyses are based on simulation with correlation effect.
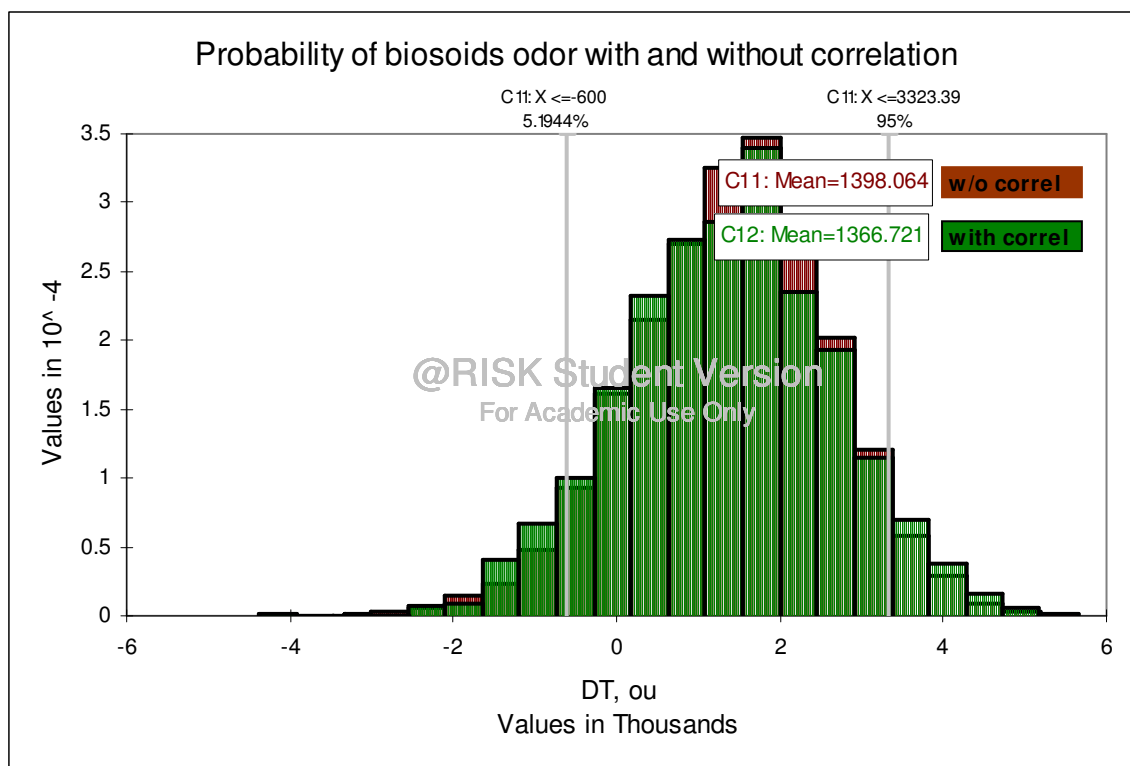
Figure 5.8: Comparison of biosolids odor's probability distributions with and without correlation

Compared with the summary statistics from the data collected in Table 5.3, the results from simulation with correlation had a lower average DT (1367 ou vs. 1582 ou) and lower standard deviation (1284 ou vs. 1697 ou) but higher median (1398 ou vs. 825 ou) than the results from the data collected. The mean and median values of simulated DT data were close (1367 ou vs. 1398 ou respectively) indicating that the simulated DT data were evenly spread. Compared with the mean and median values of the collected data (1582 ou and 825 ou respectively), the difference between these two values was influenced by extreme values from the data collected. The benefit from simulation (if appropriated probability distributions of input variables were defined) was more information associated with biosolids odor distribution.

Figure 5.9 provides graphical information on the probability distribution of biosolids odors determined at a convergence percentage of 1.5[3].

---

[3] A convergence percentage of 1.5 means the simulation would stop if the relative change in these statistics (mean, standard deviation, and percentile) is less than 1.5 percent (Palisade, 2001)
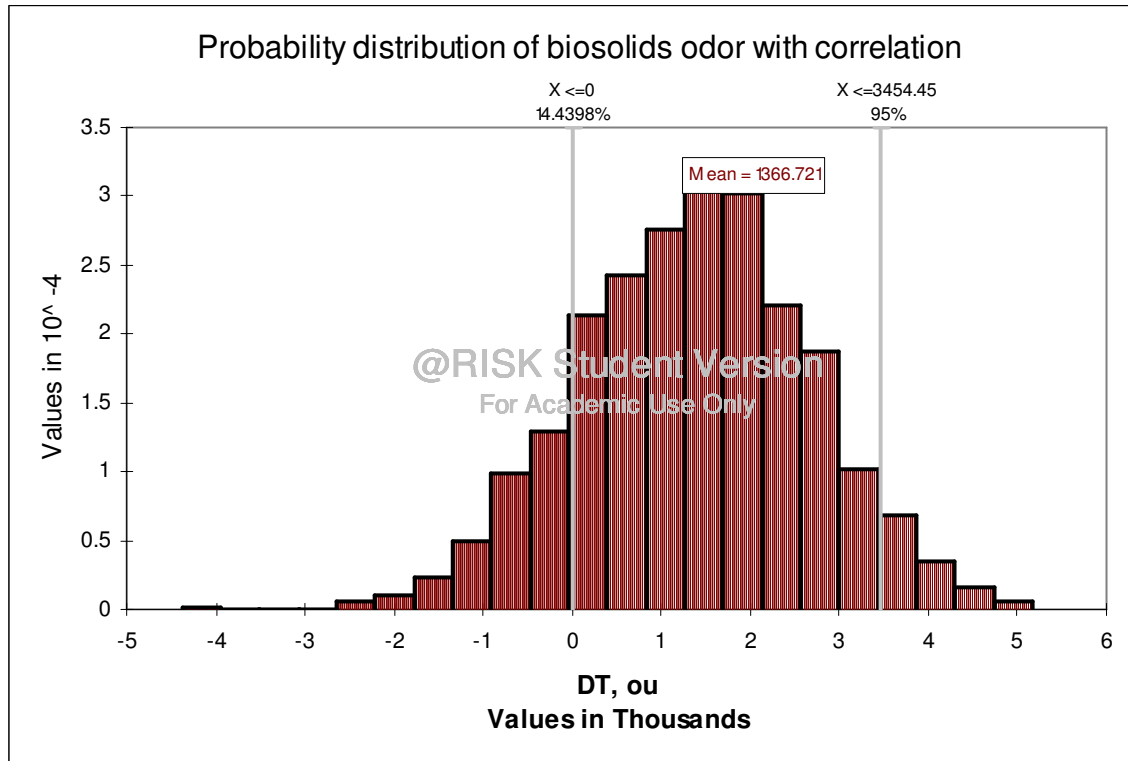
Figure 5.9: Biosolids odor profile at Blue Plains

In Figure 5.9, however, there is a 14.43 % probability of negative biosolids odor levels generated from the simulation which does not represent actual biosolids odor levels at the plant. Therefore, the selected biosolids odor prediction model (equation 5.1) was re-estimated in order to remove the negative biosolids odor values. This was done by applying a log transformation to the dependent and independent variables in the model. The re-estimated model was as follows:

$$\log Y = -11.846 - 1.737 \log X_1 + 4.268 \log X_2 + 4.525 \log X_3 - 0.496 \log X_4$$

$$+ 0.362 \log X_5 + 0.278 \log X_6$$

(5.6)

where

Y = DTWL24

X_1 = percent solids

X_2 = with lime 24 T

X_3 = GT pH

X_4 = number of centrifuges

X_5 = percent of GT in blend tank d-1

X_6 = RAS east d-1

The re-estimated model explains 35 percent (adjusted R-squared) of the variation in DT. Independent variables in the model were statistically significant at the 10 percent level except for log (number of centrifuges) which was statistically significant at the 14 percent level.

Next, to perform simulation run we converted equation 5.6 from predicting log(DTWL24) to DTWL24 as follows.

$$\log Y = -11.846 - 1.737\log X_1 + 4.268\log X_2 + 4.525\log X_3 - 0.496\log X_4$$
$$+ 0.362\log X_5 + 0.278\log X_6$$

so that

$$e^{\log Y} = e^{-11.846 - 1.737\log X_1 + 4.268\log X_2 + 4.525\log X_3 - 0.496\log X_4 + 0.362\log X_5 + 0.278\log X_6}$$

or

$$Y = e^{-11.846}\, e^{-1.737\log X_1}\, e^{4.268\log X_2}\, e^{4.525\log X_3}\, e^{-0.496\log X_4}\, e^{0.362\log X_5}\, e^{0.278\log X_6}$$

$$Y = e^{-11.846}\, X_1^{-1.737}\, X_2^{4.268}\, X_3^{4.525}\, X_4^{-0.496}\, X_5^{0.362}\, X_6^{0.278} \tag{5.7}$$

Using the re-estimated model shown in equation 5.7, we applied the same defined probability distributions to the independent variables and performed simulation runs (at a convergence percentage of 1.5). The probability distribution from the re-estimated model is illustrated below.
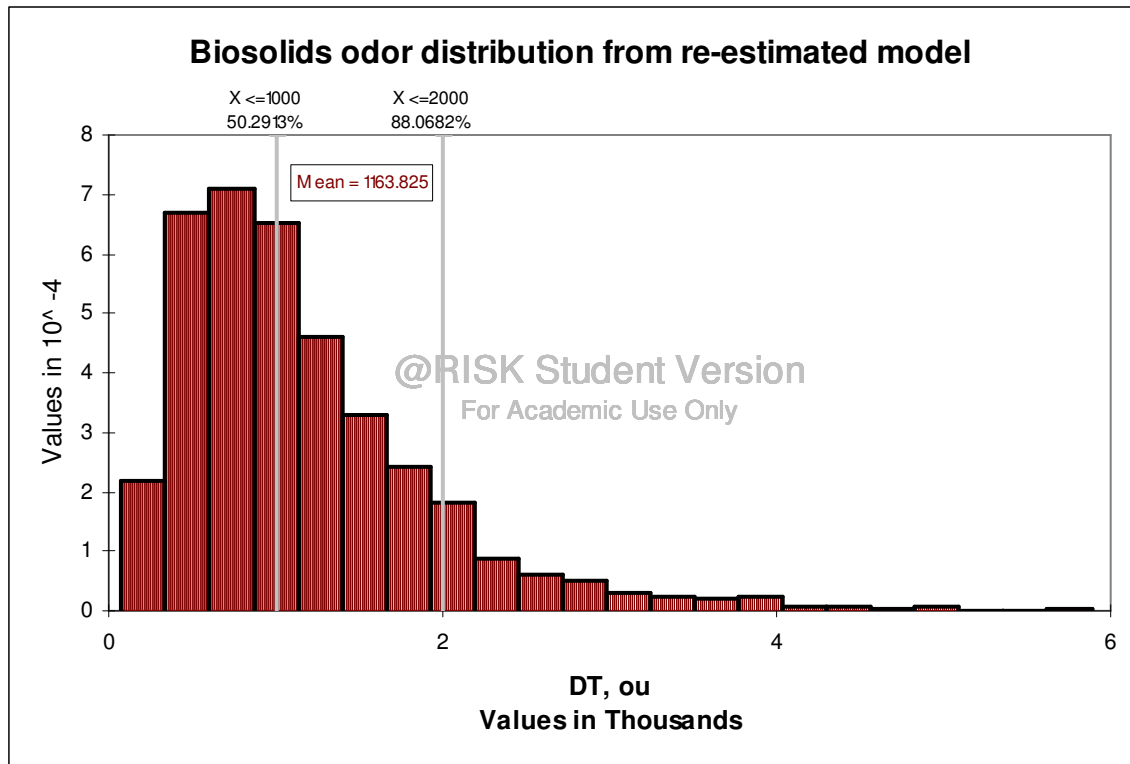
Figure 5.10: Biosolids odor distribution from re-estimated model

Now, the probability distribution of biosolids odor using equation 5.7 has a long tail distribution with no negative DT levels. Simulation results based on odor data and input data collected between April 2005 and July 2006 are summarized as follows:

1. Simulated biosolids odor levels at 24 hours emitted from a 400-gram sample being processed as described in Chapter 3 had an average odor level of 1164 ou, a maximum odor level of 5886 ou , and a minimum odor level of 76 ou. Mean and maximum values of biosolids odor generated from the re-estimated model were close to

the previous model while the more realistic minimum value of 76 ou was obtained from the new equation.

2. From the biosolids odor distribution, there was a 49.71% chance that the odor level would be greater than or equal to 1000 ou. (Note: a DT level of 1000 ou was the point where we discriminated moderate-to-low odor level (MTL) and High odor level (H) in Chapter 4's categorical data analysis).

3. In addition, there was a 11.93% chance that Blue Plains would generate product with a DT level greater than 2000 ou on any particular day.

In summary, a biosolids odor distribution benefits DCWASA by providing information on the probability that a particular event, such as high biosolids odor (DT>2000 OU) would happen. Compared to summary statistics from the collected data (e.g., a sample average, standard deviation, etc.), simulation results provide much more information regarding an overview of biosolids odor distribution at the plant based on the unit operations. Knowing the probability of producing malodorous product, management can the adjust processes accordingly to avoid odor problems at the reuse site. Next, we investigate sensitivity of two controllable variables in the model on biosolids odor generation.

5.4 Sensitivity Analysis

There are three types of sensitivity analysis (Saltelli et al, 2000); screening, local sensitivity analysis, and global sensitivity analysis. The screening method is typically a preliminary method to identify the most influential output variables. The computation is simple and mostly can not gauge the relative importance of variables. The local approach looks at the local impact of one variable on the variation of output variable (e.g., by

computing the partial derivative of the output function) while keeping other input variables at their nominal values. However, this approach does not take into account the information of the distribution of input values. Global sensitivity analysis estimates the variation for output variables by the uncertainty of the input variables (i.e., uncertainty is assigned by distribution of input variables). Carliboni et al. (2007), Borgonovo et al. (2003), Saltelli et al. (2000 and 2006), Zador et al. (2005) discuss aspects of global sensitivity analysis.

Global sensitivity analysis was used in this section. Two types of independent variables exist in the full data model (equation 5.1) controllable and uncontrollable ones. Controllable variables considered here are: 1) the number of centrifuges running, and 2) the percentage of GT solids in the blend tank. These variables are useful in biosolids odor management as levels that can be adjusted to improve odor levels based on varying inputs.

The rest of the variables, percent solids, temperature of biosolids, pH of GT solids, and concentration of returned activated solids at secondary east process cannot be directly controlled.

For example, the percent solids of biosolids is the results of different unit operations (e.g., how DCWASA operates percentage of GT solids in the blend tank or a number of centrifuges running) or pH of GT solids is influenced by the characteristics of the coming wastewater and chemical additions that are difficult to control. Thus, monitoring the levels of these variables is important relative to avoiding high odor levels.

Next we present results from sensitivity analysis to check how the number of centrifuges operating affects the distribution of biosolids odor levels.

5.4.1 Sensitivity Analysis on the number of centrifuges running

Since the number of centrifuges in operation is a controllable decision variable, we investigated the impact of operating different number of centrifuges (i.e., 3, 5, 7, 9 and 11 centrifuges) while leaving the other independent variables as they were (i.e., as random variables). We conducted nine simulation runs, each of which had a fixed number of centrifuges starting, i.e., 3, 4, 5, 6, 7, 8, 9, 10, and 11.[4] Summary statistics including average biosolids odor levels from each simulation runs are presented in Table 5.4 and Figure 5.11. The average DT levels from simulation runs decreased as more centrifuges were in operation, i.e. 1660 ou for three centrifuges, 1440 ou from for four centrifuges, 1091 ou for seven centrifuges, and 963 ou for nine centrifuges. The lowest biosolids odor level under normal operations at Blue Plains was 872 ou corresponding to running 11 centrifuges.

Table 5.4: Summary statistics of simulations at different # of centrifuges running

# of centrifuge as decision variable

Name	# of centrifuge operating								
	3	4	5	6	7	8	9	10	11
Mean	1660	1440	1289	1177	1091	1021	963	914	872
Minimum	123	107	96	88	81	76	72	68	65
Maximum	6035	5233	4685	4281	3966	3712	3501	3323	3170

---

[4] Based on the historical data, the average number of centrifuges operating was 7 with a maximum of 11 and a minimum of 3.
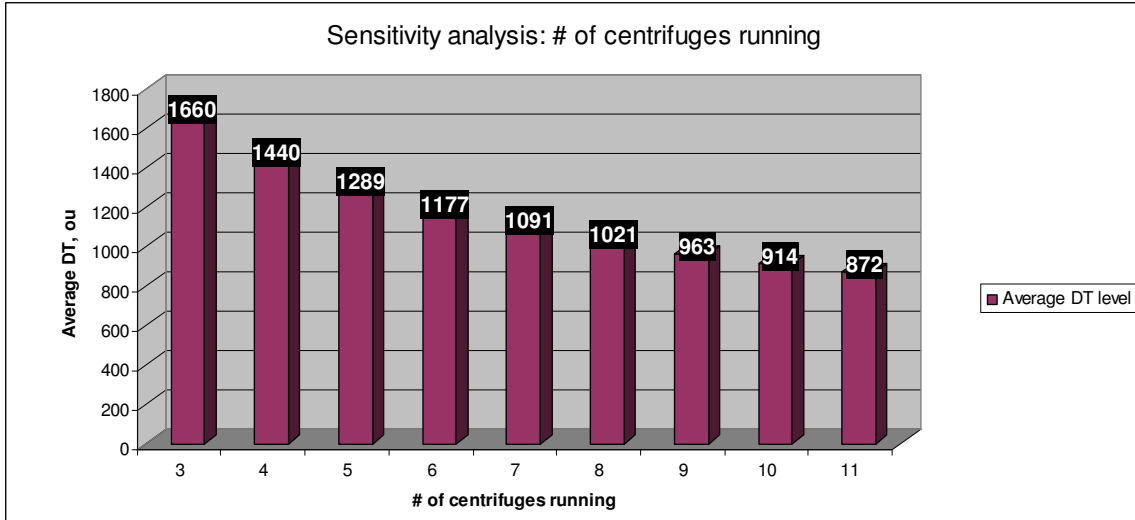
Figure 5.11: Average DT levels at different number of centrifuges running

Figure 5.12 demonstrates biosolids odor distribution as three centrifuges are in service. It shows that if DCWASA ran only three centrifuges, an average odor level emitted from biosolids at 24 hours would be 1660 ou. Average odor levels when three centrifuges were running were higher than an average odor level of 1164 ou when the number of centrifuges running was stochastic in the previous section (Figure 5.10). The probability distribution also indicated that there was a 30% chance that the odor level would be higher than 2000 ou[5]. This is due to the limited capacity in dewatering process resulting in the chance of longer sludge retention time in the different unit operation (e.g., the secondary sedimentation tanks, the DAF tanks, and the blend tank). Thus, the more septic conditions in these processes result in the higher odor levels.

---

[5] The point represents an extremely odorous odor level.
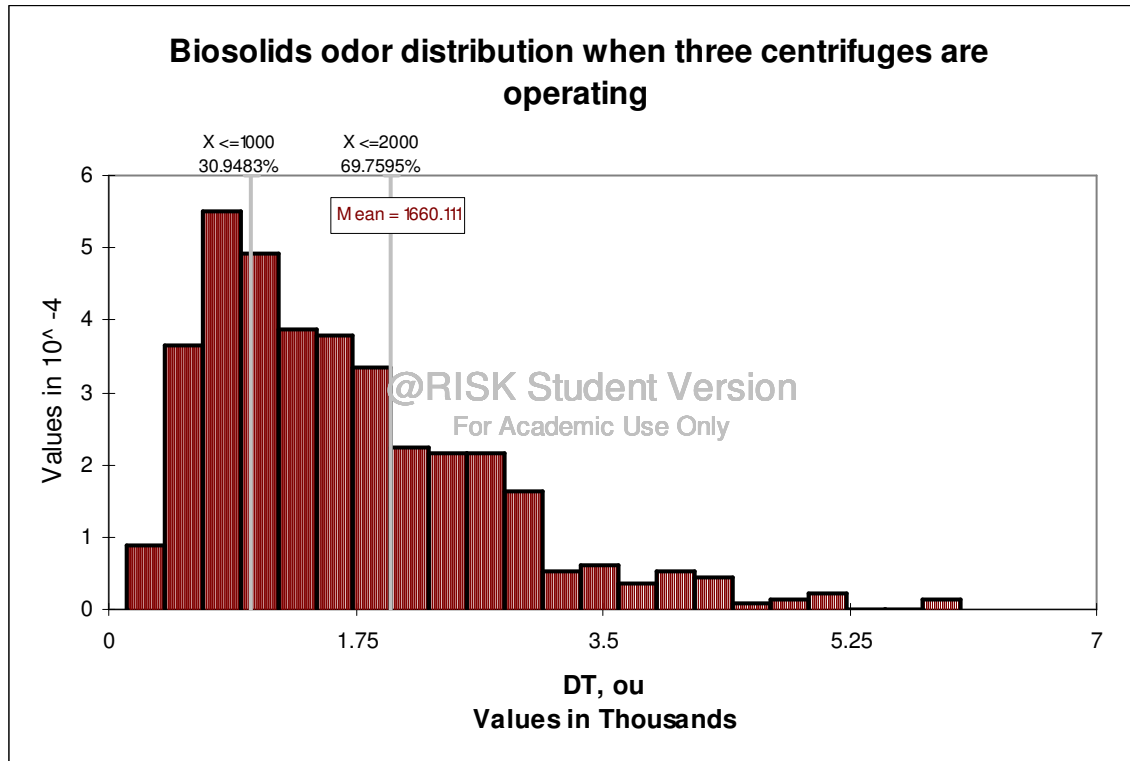
Figure 5.12: Biosolids odor profile at Blue Plains when three centrifuges are operating

Following on this analysis but with greater number of centrifuges in operation, we see the following:

- If DCWASA ran seven centrifuges which was the average number of centrifuges in service at the plant, an average odor level emitted from biosolids at 24 hours would be 1091 ou, lower than the key 2000 ou but still higher than 1000 ou. Also, there would be a 9% chance that the biosolids odor level would be greater than 2000 ou.

- If DCWASA ran nine centrifuges, an average odor level emitted from biosolids at 24 hours would be 963 ou which below the key 2000 ou and even less than 1000

ou. In addition, there would be 6% and 40% chances that odor levels higher than 2000 ou and 1000 ou would have been produced, respectively.

- If DCWASA ran 11 centrifuges, an average odor level emitted from biosolids at 24 hours would be 872 ou. There would be only 4% and 32% chances that the product leaving the plant would have an odor level greater than 2000 ou and 1000 ou, respectively.

Comparison of biosolids odor profiles between the minimum (three centrifuges), maximum (11 centrifuges) and the one in between (seven centrifuges) is illustrated in Figure 5.13.

Comparison of biosolids odor profiles at different number of centrifuges running
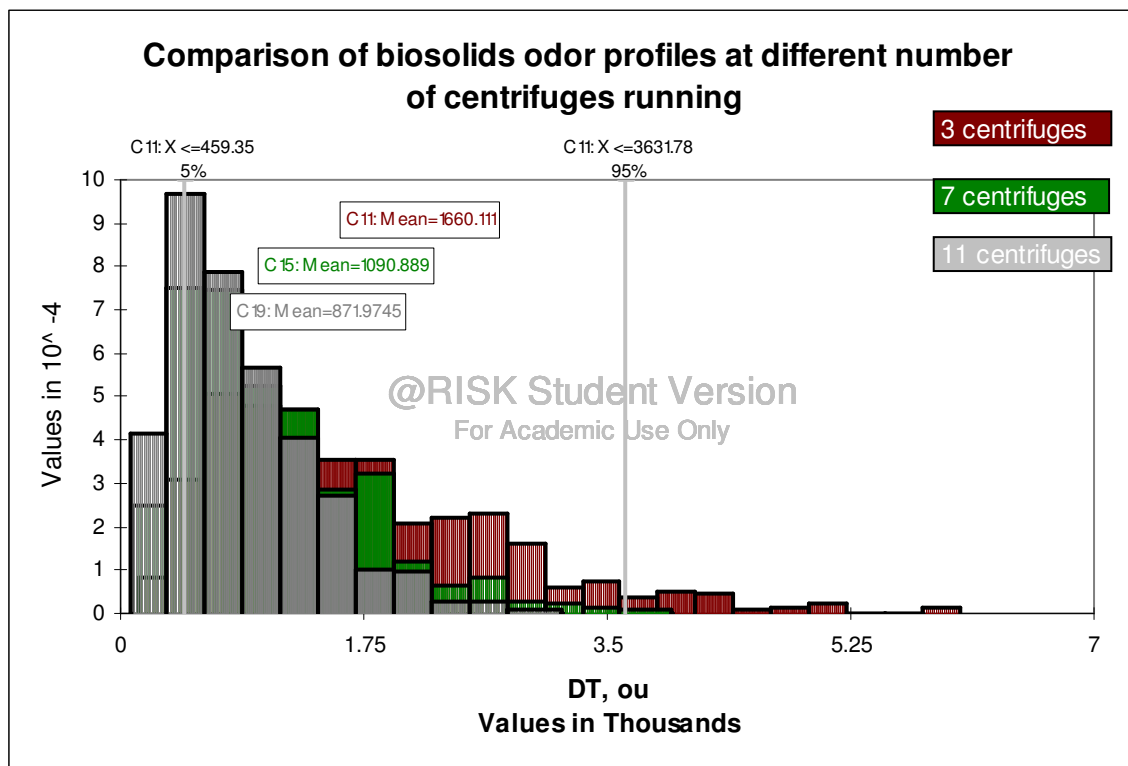
Figure 5.13: Comparison of biosolids odor profiles with respect to number of centrifuges running

Figure 5.13 shows that the probability of producing odorous biosolids reduces as an additional centrifuge is in operation (a biosolids odor distribution shifts to the left side). If biosolids odor levels at 1000 ou and higher cause odor problem at the reuse site, operating seven centrifuges still give a 47% probability of odor problem while operating eight and nine centrifuges give a 43% and 40%, respectively. Therefore, according to the sensitivity analysis results operating nine centrifuges on average is suggested due to the average odor level of 963 ou (below 1000 ou) and the 40% chance (less than 50%) of producing odor level higher than 1000 ou.

In conclusion, the greater the number of centrifuges in operation result in the lower the chance of odorous biosolids. Essentially, running nine centrifuges and higher gives the chance of producing odorous biosolids lower than 50 percent.

Next, we conduct sensitivity analysis on another decision variable, % GT in the blend tank.

5.4.2 Sensitivity Analysis on Percentage of Solids from GT in the blend tank

Another variable that can be controlled is percentage of solids from the gravity thickener process in the blend tank. Seven simulation runs were conducted based on a fixed percentage of gravity thickener solids in the blend tank while leaving other variables in model stochastic. In particular, %GT in the blend tank was fixed at 20%, 30%, 40%, 50%, 60%, 70%, and 80%. Results from simulation are provided in Table 5.5 and Figure 5.14.

Table 5.5: Summary statistics of simulations at different percentages of GT solids in the blend tank

%GT as decision variable

Name	20 %GT	30 %GT	40 %GT	50 %GT	60 %GT	70 %GT	80 %GT
Mean	840	973	1080	1171	1251	1322	1388
Minimum	82	95	105	114	122	129	135
Maximum	5732	6637	7365	7984	8528	9017	9464
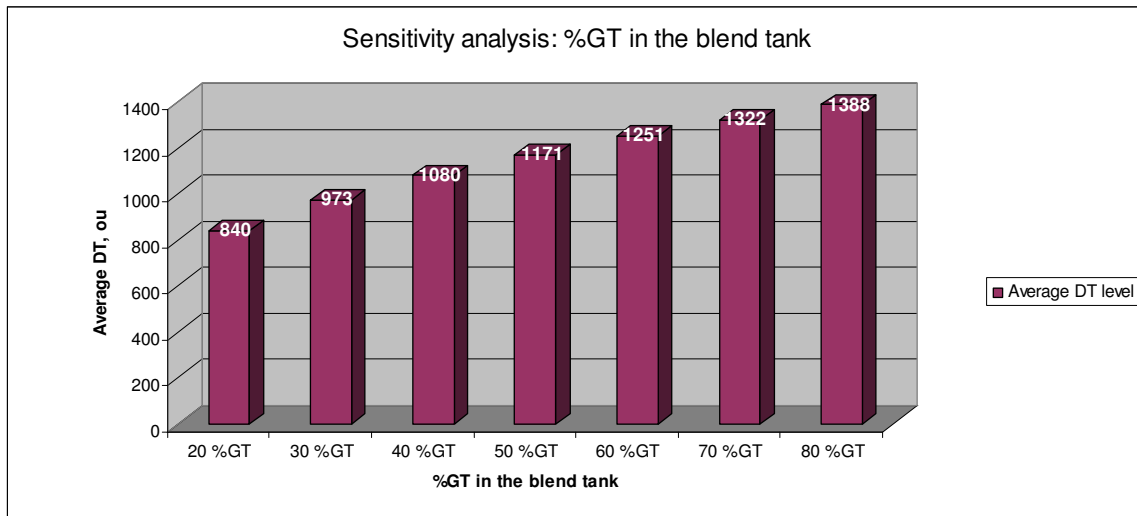
Figure 5.14: Average DT levels at different percentages of GT solids in the blend tank

Results showed that on average with the higher percentage of GT solids in the blend tank resulted in the higher DT levels. As seen in Table 5.5, at 20% GT in the tank average DT level would be 840 ou; at 50% the average odor level went up to 1171ou and at 80% the average odor level went up to 1388 ou. Comparison of biosolids odor profiles at 20%, 50%, and 80% of GT solids in the blend tank is shown in Figure 5.15.
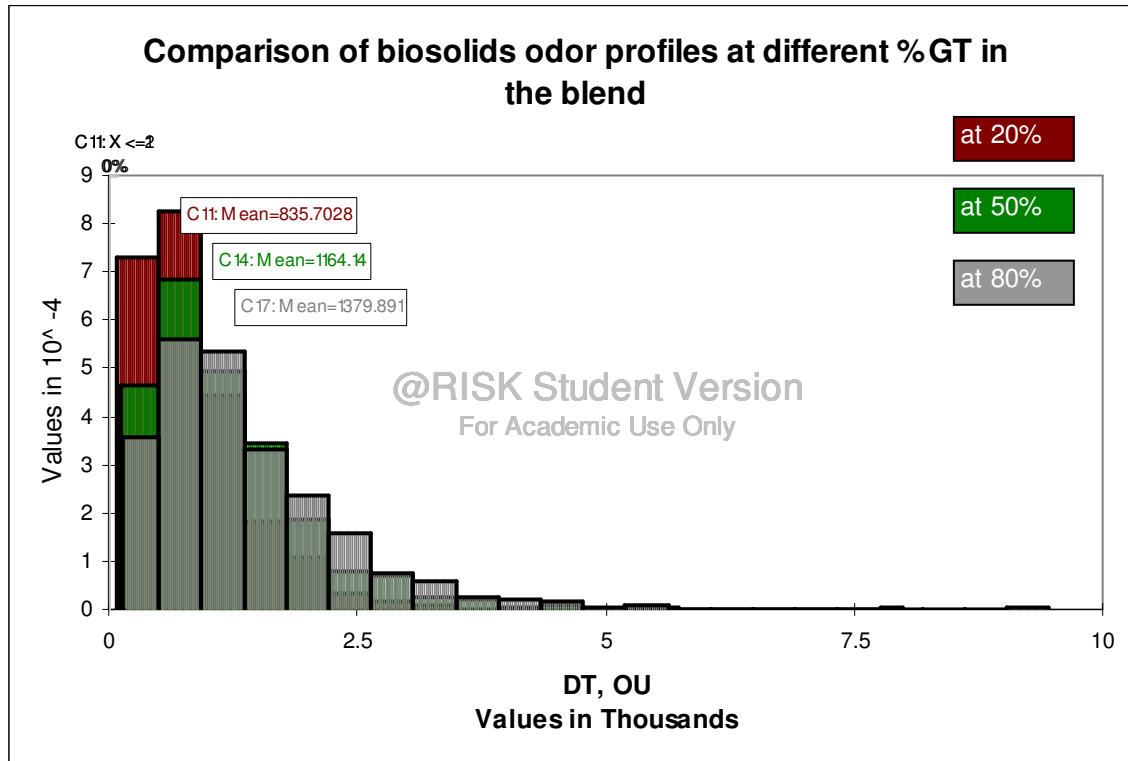
Figure 5.15: Biosolids odor profile from different percentage of GT solids in the blend tank

From simulation results, at 20%of GT in the tank, the probability of producing odor greater than 2000 ou and 1000 ou is 4.6% and 30% respectively. Therefore, it is less likely for DCWASA to produce odorous biosolids if they can control percentage of GT at 20%. Compared to 80% of GT in the tank, the probabilities of producing odor greater than 2000 ou and 1000 ou are 18.66% and 59% respectively. It is more likely to produce odorous biosolids if DCWASA keeps percentage of GT in the blend tank at 80%. In practice, DCWASA aims to maintain the percentage of GT solids in the tank at 50% to facilitate dewatering process. Thus, keeping %GT at 20% is practically too low and at 80% is also affected biosolids odor production.

Result from simulation run at 50%GT in the tank showed that there is 47.33% (below 50%) chance for DT higher than 1000 ou compared to 52.4% (above 50%) at 60%GT in the tank. Therefore, at 50%GT levels it is less likely to produce odorous product and at the same time still facilitating dewatering process. Maintaining percentage of GT solids in the blend tank at 50% but not lower than 40% should compromise both operational and odor reduction aspects.

5.5 Scenario Analysis Results

The next analysis performed was to vary both the number of centrifuges operating as well as the fixed percentage of GT in the blend tank. In all, 45 scenarios were tried, varying the number of centrifuges operating equal to 3, 4, 5, 6, 7, 8, 9, 10, or 11 and to the %GT equal to 20%, 40%, 50%, 60%, or 80%. All other independent variables were left at actual distributions. Table 5.6 summarizes the resulting average DT levels generated from the simulation runs and Figure 5.16 illustrates averages DT levels generated from each scenario. As Figure 5.16 shows, high DT levels from biosolids happen when there is a high percentage of GT solids in the blend tank and few centrifuges running. As percentage of GT solids in the tank is lowered coupled with more centrifuges operating, DT levels of biosolids produced are also lower.

Table 5.6: Average DT level generated from simulation runs with respect to fixed number of centrifuge running and fixed percentage of GT solids in blend tank

# centrif	% GT in the blend tank				
	20%	40%	50%	60%	80%
3	1207	1551	1682	1797	1994
4	1047	1345	1458	1558	1729
5	937	1205	1306	1395	1548
6	856	1100	1193	1274	1414
7	793	1020	1105	1181	1310
8	743	954	1034	1105	1226
9	700	900	976	1042	1157
10	665	854	926	989	1098
11	634	815	883	944	1047
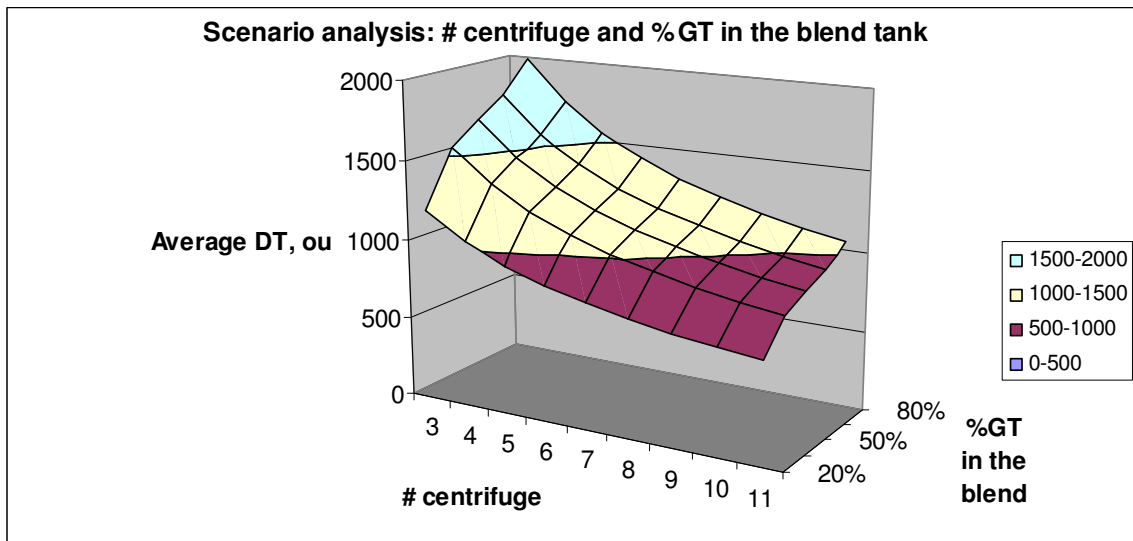
Figure 5.16: Average DT level generated from simulation runs with respect to fixed number of centrifuge running and fixed percentage of GT solids in blend tank

From Table 5.6, scenarios that generated an average DT greater than or equal to 1000 ou were highlighted. For example, if biosolids are desired with DT levels below 1000 ou on average and only nine centrifuges can be in service, we must maintain percentage of GT solids at 50% and below. In practice, however, maintaining percentage of GT solids less than 30% is undesirable for dewatering process that prefers the target blending ratio at 50% and also the higher percentage of GT solids in the tank contributes

to the higher biosolids odor according to the analysis in previous section. Therefore, we should consider the operating target of %GT in blend tank between 40% - 60%.

As a result, the followings are suggested to biosolids manager at Blue Plains.

- If we are able to run ten centrifuges and more, it guarantees that the final product will have DT level less than 1000 ou on average as seen on Table 5.6 when running eight or nine centrifuges and maintaining %GT in the blend tank between 40% - 60%.

- In contrast, if we are able to run five centrifuges or less, there is a high probability that biosolids odor level on that day would be higher than 1000 ou on average considering the most likely range of % GT in blend tank (between 40% and 60%). As a result, the distribution plan for odorous biosolids should be prepared.

- If we are able to run from six to nine centrifuges, biosolids odor levels slightly above 1000 ou would be expected. Maintaining % GT in blend tank below 60% when running nine centrifuges or below 50% when running eight centrifuges would reduce the average odor levels below 1000 ou.

In conclusion, after observing various simulated scenarios that can occur at Blue Plains running at least nine centrifuges on a daily basis is suggested. Such a strategy promotes producing biosolids with acceptable odor levels below 1000 ou. The optimum amount of centrifuges running could be decided once a tradeoff between centrifuge operating cost (i.e., between $65 - $209/ dry ton (Sahakij (2007)) and odor information from scenario analysis is taken into consideration.

In this chapter, we presented the probability distribution of biosolids odor when uncertainties in input variables were involved. We demonstrated how two key decision

variables, the number of centrifuge running and the percentage of GT solids in the blend

tank, influence biosolids odor emissions. Last, results of a scenario analysis with respect

to a fixed number of centrifuges running and a fixed percentage of GT solids in the blend

tank were presented.

In the next Chapter, we present the method to connect biosolids odor data at the

Blue Plains plant and the field sites.

Chapter 6: Field Odor Modeling

6.1 Introduction

The field odor modeling scheme in Figure 6.1 illustrates the connection between the biosolids odor levels at the Blue Plains plant and the biosolids odor levels at application sites. The biosolids odor emissions at the wastewater treatment plant are the results of the biosolids odor-causing factors at the plant. The biosolids odor emissions at reuse sites are the consequences of the original biosolids odor levels produced at WWT plant and external factors, such as the weather conditions at the field.

Further study to investigate impacts of external factors beyond the WWT plant on biosolids odor levels at field sites is important. This information would help hauling contractors and Blue Plains biosolids managers to choose appropriate reuse sites as well as develop contingency plans when malodorous products at Blue Plains are anticipated.

In this chapter, an analysis is done on those factors in the field that might additionally affect biosolids odors for the products produced at the Blue Plains. Inspectors with a variety of field odor levels in their field odor observations as well as with sufficient number of observations in their data set were selected. An ordered logit model, a categorical model dealing with an ordinal response variable, was used as a tool to understand relationships between the field data and the Blue Plains' odor data. We begin by discussions the field data and relevant assumptions.
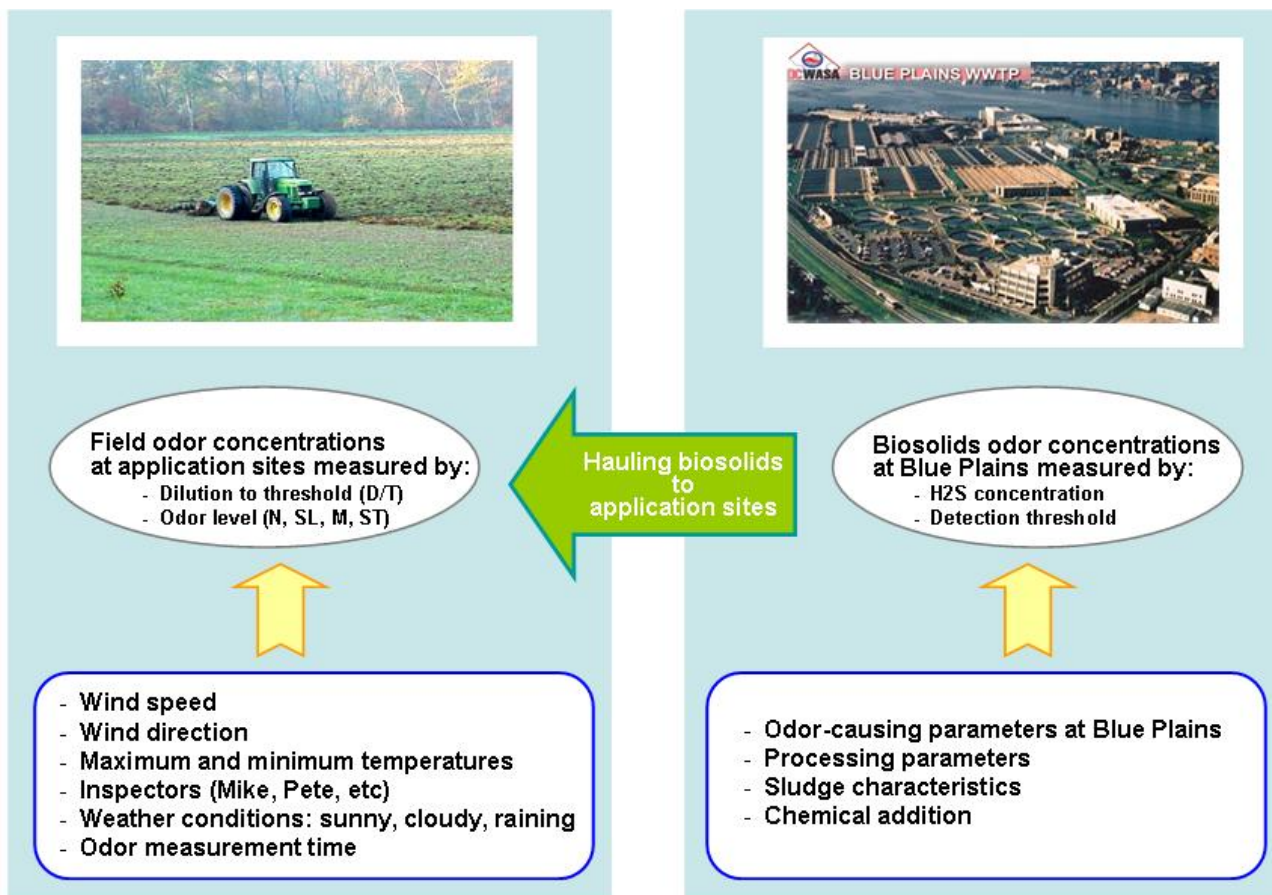
Figure 6.1: Field odor modeling scheme

6.2 Field Data

Data used include: 1) the field odor data considered as response variables, 2) Blue Plains' biosolids odor data representing the initial biosolids odor level, 3) the field conditions for their impacts on the biosolids field odor emissions, and 4) the inspector odor sensitivity data representing the health and capabilities of the inspectors' odor perceptions. These data are described below.

Dependent variables   **Independent variables**

```
                                    ┌──────────────────────────────────┐
                                    │ Biosolids odor data at Blue Plains│
                                    └──────────────────────────────────┘
┌──────────────────┐
│ Field odor data  │                ┌──────────────────────────────────┐
│ D/T              │───────────────│ Field conditions                  │
│ Odor level       │                └──────────────────────────────────┘
└──────────────────┘
                                    ┌──────────────────────────────────┐
                                    │ Inspector sensitivity data        │
                                    └──────────────────────────────────┘
```
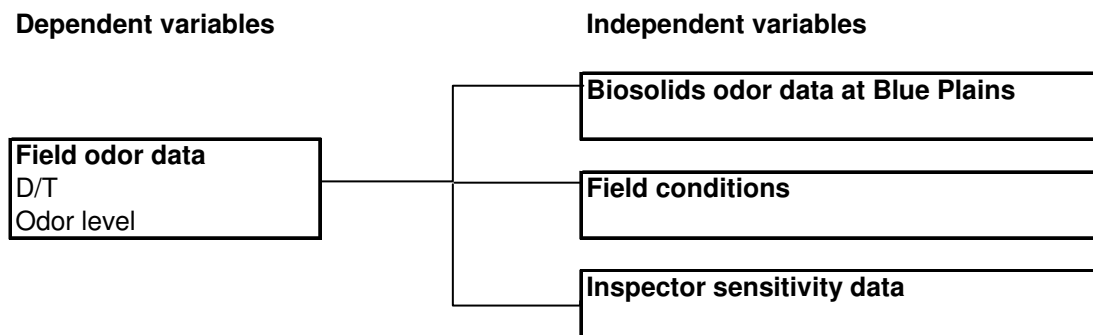
Figure 6.2: Dependent and independent variables at field site

6.2.1 Field Odor Data

For each weekday, Maryland Environmental Service (MES) assigns field inspectors to record field data that include the field odor concentrations and the field conditions at the application sites. Field inspectors record two types of biosolids odor data: the detection to threshold (D/T) ratio and the field odor level (OL).

Detection to threshold (D/T): Detection to threshold is the dilution ratio of the odor-free air needed to mix with the odorous ambient air until the odorous ambient air is no longer detectable. The dilution to threshold (D/T) is obtained through a device called a field olfactometer (St.Croix Sensory, Lake Elmo, MN); see Figure 6.3.

Figure 6.3: Illustration of olfactometer showing the adjustable dilution odor level (McGinley, 2005)

The field inspectors sniff the ambient air samples through a field olfactometer. The olfactometer has six adjustable dilution dials with the levels 2, 4, 7, 15, 30, and 60. Each dial represents the ratio of the volume of odor-free air (carbon-filtered air) over the volume of odorous air (ambient air). Therefore, the ratio of 60 is the most diluted ratio and 2 is the least diluted ratio. The field inspector sniffs the ambient odorous air through a field olfactometer in descending order in terms of the diluted air samples starting from highest dilution 60 to the lower dilution ratio until the field inspector detects the ambient odor. At that dilution level, the field inspector records the D/T level at that field site.

In practice, as stated in the MES field inspector standard operating procedure (MES, 2005), the field inspector chooses the odor measurement location at the down-wind edge of the application site where the receptors, such as houses and schools are in proximity prior to performing field odor measurements. Additionally, an inspector odor sensitivity test is conducted for MES inspectors monthly to monitor their personal dilution-to-threshold.

Field odor level (Sniff test): Field inspectors use their unaided noses to quantify the strength of ambient odorous air at the application sites and designate the corresponding odor level by none (0), slight (3), moderate (6), or strong (9). Odor levels are recorded at the same location as the D/T measurement. This odor measurement technique is used by MES field inspectors before the field olfactometer was applied.

Additional field odor characteristics, such as odor description and hedonic tone are also recorded in the MES database. Odor description explains what the odor smells like and the hedonic tone describes the pleasantness and unpleasantness of the odor. For modeling purposes, we use only the D/T ratio.

Next, assumptions on factors contributing to the field odor readings are discussed.

6.2.2 Assumptions on Factors Contributing to the Field Odor Readings

Assumptions on factors contributing to the field odor readings at the application sites are:

1) **Biosolids odor concentration at the plant:** The initial biosolids odor levels emitted at Blue Plains was represented by the detection threshold of biosolids sample taken at 24 hours in the laboratory at Blue Plains plant (DTWL24). The retention time that biosolids were left in the bunker before being trucked out is also an important factor. For this case study, we assume that biosolids were typically trucked out on the same day or at most the next day. The DT of limed samples at 24 hours represents the biosolids odor level that could be emitted at field site regardless of external factors, such as weather conditions.

2) **Field conditions at the application site**: External factors such as the field conditions and geography of the site are expected to change the concentration of the odor. The field conditions at the reuse sites, such as wind speed, temperature, field geography, measurement time of day, season, and relative humidity are mostly anticipated to influence the biosolids odor production at the field sites. These data were recorded by field inspectors when land application took place.

Since the locations are predetermined and also due to EPA regulations concerning application of biosolids to sites that have a slope less than seven percent to prevent run-off and erosion of top soil (Evanylo, 1999), the geography of the field site should have fewer effects on the field odor data. As a result, we disregard the geographical data of the field sites in this study. Here, we apply the field data collected by field inspectors to investigate their impacts on the inspectors' field odor perceptions instead. The external factors used in this study were:

The high, low, and average temperatures at the field site: (HT, LT, and AVGT): The maximum and minimum temperatures (in Fahrenheit) when the biosolids application took place were recorded. Since these data didn't represent the highest and lowest temperatures for an entire day, we used the average temperature to represent the temperature while the field inspectors collected odor data.

Wind speed (WINDS: miles per hour (mph)**):** The wind speeds as the field inspector collected the field odor measurements were recorded.

Dummy wind speed (WINDD) dummy variable on moderate wind speed at 7 mph was created. Seven miles per hour is an approximate average wind speed from the

data set. WINDD = 1 when WINDS is greater than seven miles per hour and zero, otherwise.

Wind direction: While wind direction data were available, since it was stated in the field inspector standard operation procedure (MES, 2005) to conduct field odor measurements at the down-wind edge of field site, the wind direction was disregarded in this study.

Weather conditions: Weather conditions were recorded as S= sunny, PC = partly cloudy, C = cloudy, H = hazy, SLT = sleet, P = precipitation, F = fog, R = rain, SN = snow. Few observations of H, SLT, and SN (e.g., 3, 4, 5 observations respectively) were recorded during April 2005 to July 2006, the time period we collected data at Blue Plains. Thus, we did not consider these categories in the data set. For the other categories left, we reorganized them into three groups based on similar weather conditions. We then presented them as three dummy variables as follows:

S = 1 if sunny conditions were present, and 0 otherwise.

CLOUD = 1 if the weather was either cloudy or partly cloudy, and 0 otherwise.

RP = 1 if rain or precipitation were present, and 0 otherwise.

Of the field data set we collected, category S accounted for 1834 observations, category C for 1543 observations, and category R for 149 observations.

Time of day: The time of day is assumed to be a significant factor in explaining odor levels. As mentioned in Chapter 2, odor complaints usually happened in the early morning and evening when the weather is calm (USEPA, 2000). We assume that the morning time before 10:00 is the time when the soil starts getting heat from the sunlight

and the ground dries. Further, the time between 10:00 to 14:00 may represent the period when the ground receives the heat directly and gets warm, and the time after 14:00 represents the time when all the surface is heated up for a certain period.

Since temperature influences odor emissions, the time of day should have an effect on odor perception. We set up dummy variables for the time of day as follows:

M = 1 equal to measurement in the morning time (before 10:00).

Otherwise M = 0

N = 1 equal to measurement at noon time (between 10:00 and 14:00).

Otherwise N = 0

AF = 1 equal to measurement in the afternoon (after 14:00).

Otherwise AF = 0

3) **The field inspector's odor sensitivity**: Field odor data were quantified using the field inspector's olfactory sense. The individual field inspector's odor sensitivities varied by their personal odor perception capabilities and their physical conditions while measuring odor. These factors directly impact the inspector's field odor rating. Almost every month, field inspector olfactory threshold is checked by an odor sensitivity testing (Lay and McGinley, 2004; ASTM, 1997). The procedure is similar to how to obtain the DT in the laboratory, by using a triangular force choice procedure. In this procedure, the field inspector sniffs the presentation of three odor pens randomly, including one with a known odor level and the other two pens being odor-free. When the field inspector detects the correct pen, the detection thresholds of the field inspectors are recorded. The sensitivity score start from 2 (the least sensitive nose) to 15 (the most sensitive nose).

We apply this information to an explanatory variable to represent the olfactory conditions of the field inspectors for that month.

4) **The ambient conditions along the hauling route**: As biosolids are hauled to application sites, the ambient conditions, such as wind speed, temperature, weather conditions may also affect the level of biosolids odor. In practice, the biosolids on the hauling truck are covered to prevent the odor from affecting the communities along the hauling route. Thus, the ambient condition may have less impact compared to other factors. Also it is difficult to track the weather conditions along the hauling route which varied by county, and state. Thus, the impact of ambient conditions along the hauling route were initially considered but eventually disregarded in this study.

Next, we present an ordered logit model used as a tool to study the relationship between the field data and the odor data at the Blue Plains plant.

6.3 Methodology

Unlike the odor data at the wastewater treatment plant, which were continuous-valued, the field odor data from MES were categorical. The Ordinary Least Square (OLS) approach, which was used to estimate the coefficients of the explanatory variables in the plant odor model, would not be appropriate in this context due to the violation of the basic assumption of OLS as mentioned in Section 4.6.1 (i.e. violation of normality and constant variance of residual). Applying OLS approach to the categorical response variable could lead to biased estimators because of the nonlinear relationship nature of the categorical data, and the violations of the basic assumption of the regression model (Mckelvey and Zavoina, 1975).

An ordered logit model is suggested as an appropriate approach to estimate the nonlinear relationship between the ordinal responses and the exploratory variables as compared to a multinomial logit model or a probit model which does not take the ordinal nature of the response variable into account (Greene, 1997). Previous applications of ordered logit model have been in residential satisfaction (Lu, 1999), in firm allocation decisions (Dijk and Pellenbarg, 2000), and applied economics (Robson and Bennett, 2000; Kramer, 1996).

6.3.1 Ordered Logit Model

What follows is an explanation of an ordered logit model (Agresti, 2002) but adapted to the problem at hand.

Cumulative probability

First, let

x be a vector of explanatory variables (a vector of wind speed, temperature, and biosolids odor level at Blue Plains)

Y be the vector of actual odor levels

j be the site odor level designated on an ordinal scale, e.g., 0, 3, 6, and 9 from the field odor levels or 2, 4, 7, 15, 30, or 60 from the D/T levels.

Then, if $P(Y \leq j|x)$ is the cumulative probability that the actual odor level is less than or equal to the j th odor level given a set of explanatory parameters x.

then

$$P(Y \leq j|x) = P_1(x) + ... + P_j(x), \, j = 1,...,J.$$ (6.1)

where

$P_j(x)$ is the probability that the actual odor level falls into the category j given a

set of explanatory parameters x.

Cumulative logits

Logit or the log of the odds ratio of $P(Y \leq j|x)$ can be represented by

$$\text{logit}\left[P(Y \leq j|x)\right] = \log\frac{P(Y \leq j|x)}{1 - P(Y \leq j|x)} = \log\frac{P_1(x) + ... + P_j(x)}{P_{j+1}(x) + ... + P_J(x)}, j = 1,...,J - 1 \quad (6.2)$$

A model that uses all cumulative logits as a linear function of independent

variables is of the form

$$\text{logit}\left[P(Y \leq j|x)\right] = \alpha_j + \beta'x, j = 1,...,J - 1 \quad\quad\quad (6.3)$$

where

$\beta'x$ is the effect of the explanatory vector x on the log of the odds ratio when the

response variable falls into category j or below. $\beta'x$ has no subscript j so the effect of x

remains the same for all response values.

However, α_j is the intercept that varies by the level j of the ordered response

level.

Figure 6.4 illustrates the odds ratio model for four category responses with a

single explanatory parameter x. Each curve follows the same logistic function shape with

a common β varied as a function of x. The response variable on each curve is a binary

outcome either $Y \leq j$ or $Y > j$. For a fixed x, the curves follow the cumulative

probabilities, with the probability of $P(Y \leq 1)$ being lowest.
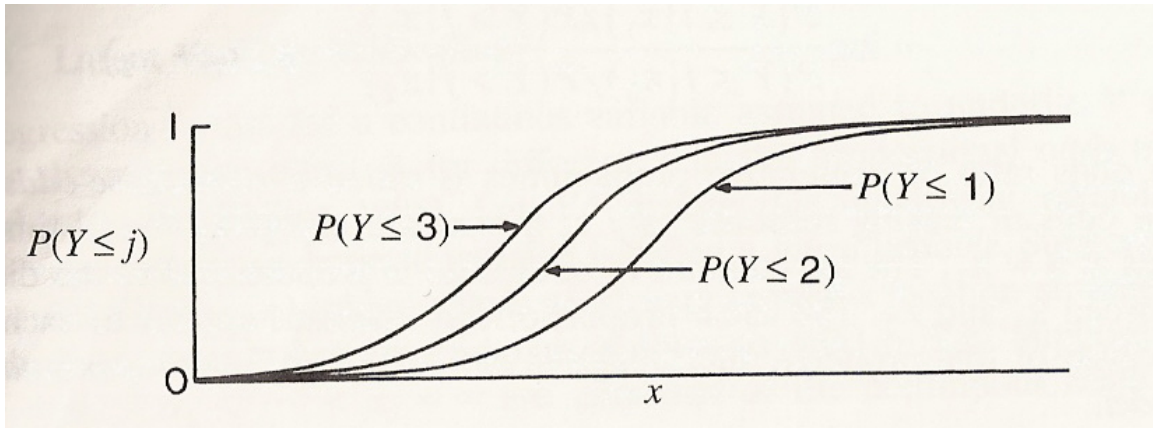
Figure 6.4: The cumulative probabilities in proportional odds model (Agresti, 1996)

Interpretation of the ordered logit model

To see how the ordered logit model can be used in the field odor modeling context, we use the field odor level measure as a response variable. (For simplicity compared to D/T, the odor level measure, an odor measurement by unaided nose given the odor level value 0, 3, 6, or 9, is used as an example). First, let

$P_1 + P_2 + P_3 + P_4 = 1$, where

$P_1 =$ probability that odor level is 0

$P_2 =$ probability that odor level is 3

$P_3 =$ probability that odor level is 6

$P_4 =$ probability that odor level is 9

Then,

$$\text{logit}[P_1] = \alpha_1 + \beta'x \quad \text{or} \quad P_1 = \frac{\exp(\alpha_1 + \beta'x)}{1 + \exp(\alpha_1 + \beta'x)}$$

$$\text{logit}[P_1 + P_2] = \alpha_2 + \beta'x \quad \text{or} \quad P_1 + P_2 = \frac{\exp(\alpha_2 + \beta'x)}{1 + \exp(\alpha_2 + \beta'x)}$$

As a result, $(P_1 + P_2) - P_1 = P_2 = (\alpha_2 + \beta'x) - (\alpha_1 + \beta'x)$

$$= \alpha_2 - \alpha_1$$

Note that the vector $\beta'x$ remains the same as the odor level changes. The intercept α_j tells the actual probability that the odor level will fall into the category j.

Also,

$$\text{logit}[P_1 + P_2 + P_3] = \alpha_3 + \beta'x \quad \text{or} \quad P_1 + P_2 + P_3 = \frac{\exp(\alpha_3 + \beta'x)}{1 + \exp(\alpha_3 + \beta'x)}$$

and $P_3 = (P_1 + P_2 + P_3) - (P_1 + P_2) = (\alpha_3 + \beta'x) - (\alpha_2 + \beta'x)$

$$= \alpha_3 - \alpha_2$$

Therefore, $\quad 1 - (P_1 + P_2 + P_3) = P_4$

All other things being equal, one unit change on the predictor x_l affects the odds of making response variable equal to P_j by $\exp^{\beta_l}$.

Table 6.1 illustrates an example of the prediction results from the ordered logit model as applied to the field odor level response variable. In this example, the model used odor level at Blue Plains (DTWL24), wind speed, and lowest temperature as explanatory variables. With these values, it can be seen that the field odor level will fall in the level 0 with a probability of 0.59, level 3 with a probability of 0.22, odor level 6 with a probability of 0.17, and odor level 9 with a probability of 0.030. Thus, the most likely field odor level is 0.

Table 6.1: Example of predicting results from the ordered logit model

			Explanatory variables			Predicted value		
Date	Field name	Actual field odor level	DTWL24	wind Speed	LT	Predicted field odor level	Prob.	cummulative Prob.
4/19/2005	DWCLE-1	0	310	8	49	at 0	0.59	0.59
						at 3	0.22	0.80
						at 6	0.17	0.97
						at 9	0.03	1

Next, we describe how to select an inspector for further analysis and show the distribution of their D/T data[6] after merging with the Blue Plains' odor data.

6.4 Field Odor Data Distribution

Seven of the twelve field inspectors who collected field data between April 2005 and July 2006 (3596 field observations in total) were selected for field odor analysis since they had a sufficient number of observations in this period. The resulting 77 onsite odor observations were matched with the inspector field data in two ways: first the plant data and field site data were matched by observation date, and second the plant odor data were matched with the field data on the next day. For example, if we have the plant odor data on 4/17/06, it was matched with the field data on 4/17/06 as well as 4/18/06. The rationale of the first way was the assumption that the field inspector observed the field odor from the biosolids produced on the same day. The second format was based on the assumption that the field inspector observed the field odor from the product produced the day before. The second assumption related to a retention time, normally no more than one day for biosolids to sit in the bunker[7].

6.4.1 Inspector Selection Criteria

All things being equal inspector with more diverse D/T scores were preferred to be able to better study what factors impacted the change of odor from one level to another. Data from three inspectors had insufficient variety of scores due to the less

---

[6] D/T distribution is important since few or sparse data in some classes can cause the problem of coefficient estimation (Agresti, 1996).
[7] The bunker is the location where lime stabilized biosolids are stored before being trucked out of the plant.

variety so the scores from these inspectors were removed. The scores from the remaining

four selected inspectors, inspectors D, G, J, and P[8] are illustrated in Figures 6.5 and 6.6.

D/T levels 2 and 4 are considered as low values. Most D/T data from inspector D

were lower than 7 (88% in Figure 6.5 and 90% in Figure 6.6) while most D/T data of

inspector P were at 7 and higher (72% and 81%, respectively). The difference in their

field odor data can be due to the differences in the field locations, the weather conditions,

the biosolids applied to their sites, or the inspector odor sensitivity.

Approximately 50 percent of inspector G's data are in low values (62% and 50%,

respectively). The rest of inspector G's data are mostly at 7 (23 % and 30%,

respectively). Most D/T data from inspector J were in low D/T levels (99% and 98%,

respectively). During the inspector selection step, inspector J's data showed diversity in

their D/T data. However, after merging with Blue Plains' odor data his data were limited

to just two categories. This could be problematic since D/T from inspector J's data had no

high levels. Combining his data with other inspectors should minimize this problem.

---

[8] For privacy purpose, we used the letter instead of their full name to identify inspectors' field information.

Figure 6.5: Match the field odor data (D/T) with the onsite odor data on the same day (d-0)
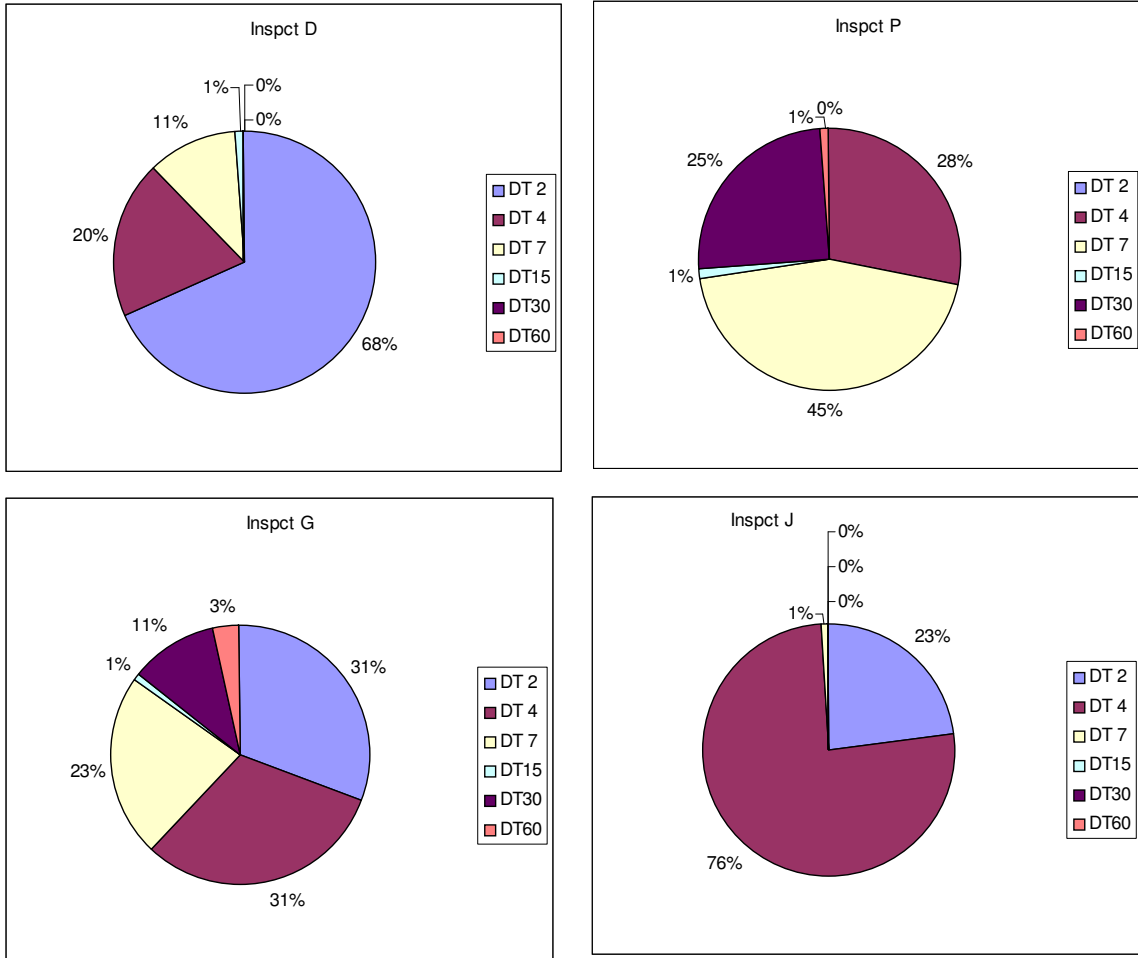
Figure 6.6: Match the field odor data (D/T) with the on-site odor data one day before (d-1)
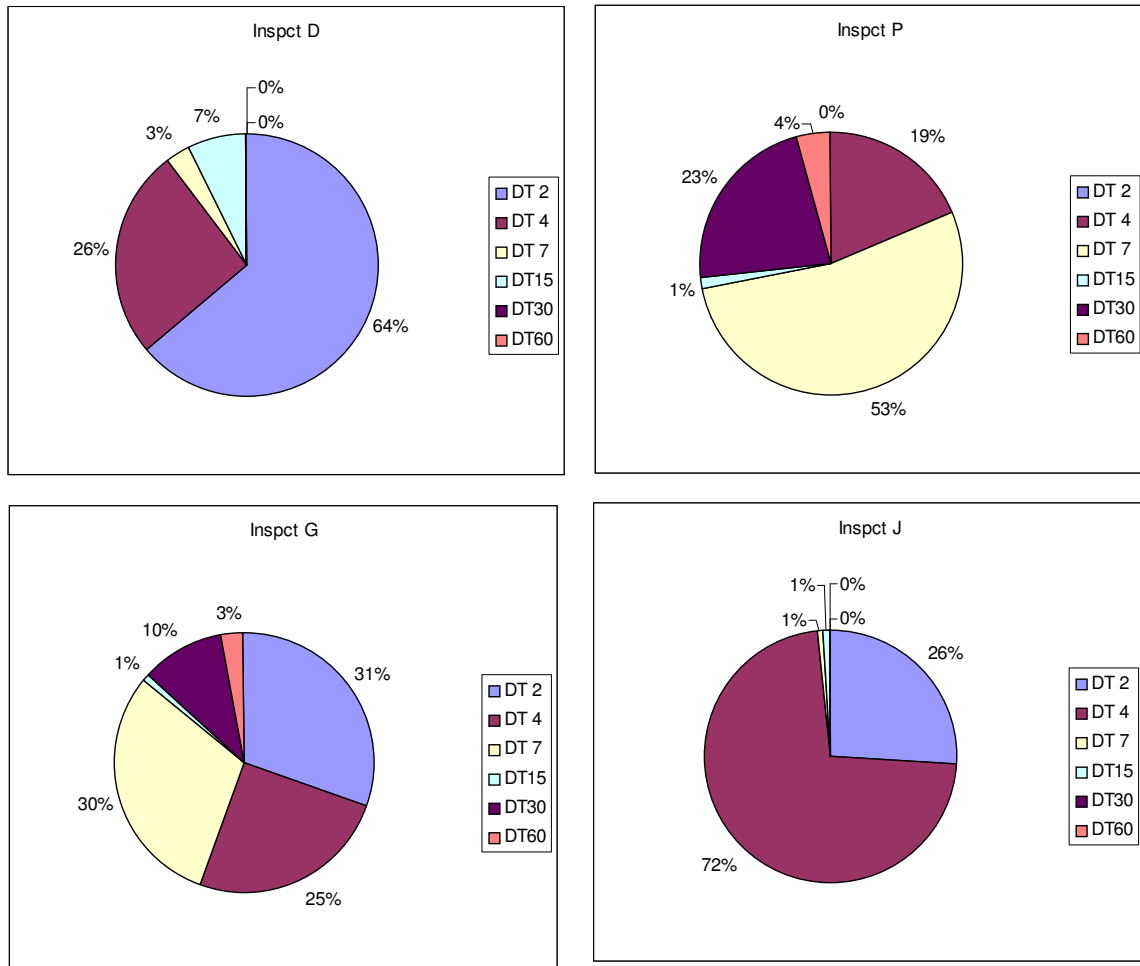
Sensitivity score

Sensitivity scores of selected inspectors between April 2005 and July 2006 are shown in Table 6.2. These scores were used to indicate how sensitive inspectors were on the field odor for each corresponding month. The higher the sensitivity score represented the more sensitive the inspector was on field odor perceptions. For the month that the score was missing, the scores from adjacent months were averaged.

Table 6.2: Selected inspectors' sensitivity score

Month	Inspector's odor sensitivity score			
	D	G	J	P
April-05	5	8.5	5	8
May-05	5	8.5	5	8
June-05	5.5	5.5	5.5	9
July-05	5	7	8	6
August-05	4	7	7.5	5.5
September-05	4.5	7	8.5	7
October-05	5.5	7	7.5	9.5
November-05	9	2	5.5	4.5
December-05	7.5	8.5	5.5	11
January-06				
February-06	9.5	13	8.5	
March-06	7	12	12	11
April-06	7	7	12	6
May-06				
June-06	4.5	7	15	9
July-06	4.5	7	15	9
Average	**6.0**	**7.6**	**8.6**	**8.0**

6.4.2 Regrouping Inspectors' Field Odor Data

As seen from the inspectors' odor distribution in Figures 6.5 and 6.6, certain D/T levels have few or no observations. For example, in the d-0 data set (Figure 6.5) there were no D/T data values > 7 from inspector J when we merged his data with the Blue Plains' database. Also, only one percent of the D/T data from inspector D had D/T values greater than 7. These issues of sparse data can cause problems in estimation of the odds ratios, in performance of computational algorithms, and in asymptotic approximation of chi-squared statistics (Agresti, 2002).

In many states including Connecticut (Reg. 22a-174-23), Illinois (Title 35, Subtitle B, Chap.1, Part 245), Kentucky (Reg. 401 KAR 53:010), Nevada (NAC 445B.22087), and Wyoming (Ch.2, Sec. 11), D/T levels of 7 or higher represent a violation (Hamel and McGinley, 2004). Therefore, to avoid sparsity problems and be consistent with these sorts of regulations, we re-classified field odor data (only D/T) into

fewer classes According to suggestions from Mr. Mark Ramirez, a biosolids process engineer at DCWASA, and Mr. Al Razik from MES, we reclassified D/T levels in four different ways as shown below data with the new variable denoted as DTRG (regrouped diction threshold) (M. Ramirez and A. Razik, Personal, May, 23, 2007).

DTRG1

- If D/T = 2 or 4 then DTRG1 =1
- If D/T = 7 or 15 then DTRG1 =2
- If D/T > 15 then DTRG1 =3

DTRG2

- If D/T = 2, 4, or 7 then DTRG2 =1
- If D/T = 15 then DTRG2 =2
- If D/T > 15 then DTRG2 =3

DTRG3

- If D/T = 2 then DTRG3 =1
- If D/T = 4, 7 then DTRG3 =2
- If D/T= 15 or 30 then DTRG3 =3
- If D/T > 30 then DTRG3 =4

DTRG4

- If D/T = 2 or 4 then DTRG4 =1
- If D/T > 4 then DTRG4 =0

Regrouped field odor data from the four inspectors were combined to two data sets, the combined inspectors' data on d-o and d-1, to increase observations especially in

some classes where the data were still sparse. Distributions of regrouped D/T were provided in Tables 6.3 and 6.4.

Table 6.3: Regrouped inspectors' field odor data (d-0)

Count of DTRG (d-0)							
DTRG1	Total	DTRG2	Total	DTRG3	Total	DTRG4	Total
1	314	1	397	1	129	0	317
2	85	2	2	2	268	1	130
3	48	3	48	3	41	Grand Total	447
Grand Total	447	Grand Total	447	4	9		
				Grand Total	447		

Table 6.4: Regrouped inspectors' field odor data (d-1)

Count of DTRG (d-1)							
DTRG1	Total	DTRG2	Total	DTRG3	Total	DTRG4	Total
1	271	1	346	1	123	0	119
2	85	2	10	2	223	1	271
3	34	3	34	3	38		
Grand Total	390	Grand Total	390	4	6	Grand Total	390
				Grand Total	390		

Even after combining scores from individual inspectors, the regrouped D/T data still showed only a few observations for some groups. For example, for the DTRG2 variable level group on level two, there were only two and ten observations for the d-o and d-1 sets, respectively. For the DTRG3 variable, level four, there were only nine and six observations on d-0 and d-1 sets, respectively.

Next, we investigate the relationships between field odor observations and influential factors using an ordered logit model.

6.5 Simple Ordered Logit Model

Factors assumed to have influences on field odor measurements are: 1) the odor measurement time of the day, 2) the weather conditions, 3) the wind conditions, 4) the field temperature, 5) the inspector's odor sensitivity score, and 6) the biosolids odor

levels from the Blue Plains plant. Table 6.5 summarizes the factors of interest, their abbreviations, and their descriptions.

Table 6.5: Field data description

Abbreviation	Description
Odor measurement time of the day	
M	M = 1 if inspector measured field odor before 10 am
N	N = 1 if inspector measured field odor between 10 am and 14pm
AF	N = 1 if inspector measured field odor after 14pm
Weather conditions	
S	S = 1 if recorded weather condition = sunny, else 0
CLOUD	CLOUD = 1 if recorded weather conditon = cloudy, else 0
RP	RP = 1 if there was rain or precipitation, else 0
Wind conditions	
WINDS	Wind speed
WINDD	Dummy wind speed (WINDD = 1 if wind speed greater than 7 mph, else 0)
Temperature	
HT	High temperature
LT	Low temperature
AVGT	Average temperature
Inspector odor sensitivity score	
ODORSD	Inspector's odor sensitivity score
Odor at Blue Plains	
odor concentration from Jerome meter	
JWOL3	data from sample without lime at 3 hours
JWL3	data from sample with lime at 3 hours
JWOL24	data from sample without lime at 24 hours
JWL24	data from sample with lime at 24 hours
DT level from odor panel	
DTWOL3	data from sample without lime at 3 hours
DTWL3	data from sample with lime at 3 hours
DTWOL24	data from sample without lime at 24 hours
DTWL24	data from sample with lime at 24 hours

A simple ordered logit model was used to assess the contributions of these factors on the variation of the DTRG 1, 2, 3, and 4. We checked their contributions through two data sets: the matched data on the same day (d-0 set) and on the next day (d-1 set). For each data set, the investigation was conducted individually on each inspector's data (four inspectors) as well as on the combined inspectors' database.

SAS software was used to conduct an ordered logit model specifically proc logistic. The following is the sample code that regress DTRG1 (regrouped DT#1) on N (odor measurement at noon time). The option "descending" is necessary to arranged D/T data to appropriate order (UCLA, 2008). Otherwise, the estimated coefficient's sign would be in the opposite direction.

```
proc logistic descending;
TITLE 'MODEL 1.1';
model DTRG1 =  N;
run;
```

Part of SAS output corresponding to the SAS code above is presented below.

```
                                 MODEL 1.11        15:20 Thursday, September 4,
2008 1298

                      Analysis of Maximum Likelihood Estimates

                                        Standard         Wald
           Parameter      DF    Estimate    Error    Chi-Square    Pr > ChiSq

           Intercept 3     1     -1.9734    0.1823     117.1922       <.0001
           Intercept 2     1     -0.6801    0.1438      22.3688       <.0001
           N               1     -0.3890    0.2062       3.5605       0.0592 (1)

                             Odds Ratio Estimates

                              Point          95% Wald
                   Effect    Estimate    Confidence Limits

                   N           0.678      0.452      1.015      (2)
```

.

The highlighted area number 1 under "Analysis of Maximum Likelihood Estimates" shows the estimated coefficient of -0.3890 of variable N, the dummy field odor measurement at noon, as DTRG1 is regressed on variable N. This means that if field odor was measured at noon (N =1), the odds of getting a higher odor is equal to $e^{-0.3890}$ or 0.678 compared to when the odor measurements is not noon (N = 0), which is also indicated in the highlighted area number (2).

On each simple ordered logit model tried, the estimated coefficient and significance value (Pr>chq) are presented in Tables 6.6, 6.7, and 6.8. The numbers of observations in each class of the regrouped D/T when we performed the analysis are summarized at the top of each table.

Table 6.6: Simple ordered logit model on the same day (d-0) and on the next day (d-1)

	Combined inspectors d-0								Combined inspectors d-1							
	DTRG1		DTRG2		DTRG3		DTRG4		DTRG1		DTRG2		DTRG3		DTRG4	
Number of observations in each level	1	314	1	397	1	129	0	315	1	271	1	346	1	123	0	296
	2	85	2	2	2	268	1	130	2	85	2	10	2	223	1	116
	3	46	3	46	3	41			3	34	3	34	3	38		
					4	7							4	6		
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
Time of day																
M	0.23	0.32	0.33	0.29	X	X	0.28	0.21	X	X	0.62	0.14	-0.27	0.2389	0.17	0.48
N	-0.39	0.06	0.88	0.00	X	X	-0.38	0.06	-0.27	0.22	X	X	X	X	-0.31	0.15
AF	0.31	0.21	X	X	X	X	0.24	0.33	0.28	0.2867	-0.59	0.0975	0.36	0.158	-0.28	0.29
Weather condition																
S	-0.35	0.10	X	X	0.20	0.30	-0.43	0.04	0.27	0.22	-0.30	0.344	0.28	0.15	0.25	0.25
CLOUD	0.47	0.02	X	X	-0.16	0.39	0.53	0.01	-0.33	0.14	X	X	-0.27	0.17	-0.34	0.12
RP	-0.57	0.27	X	X	X	X	-0.47	0.35	X	X	X	X	-0.42	0.411	X	X
Wind condition																
WINDS	0.13	<0.0001	-0.09	0.01	0.12	<0.0001	0.15	0.00	0.06	0.08	-0.02	0.34	0.03	0.25	0.05	0.09
WINDD	X	X	X	X	0.28	0.15	X	X	0.41	0.0616	-0.33	0.3037	0.31	0.12	0.367	0.1
Temperature																
HT	X	X	-0.01	0.22	-0.004	0.45	-0.01	0.30	X	X	X	X	-0.009	0.1618	-0.01	0.38
LT	0.01	0.24	X	X	-0.01	0.21	0.01	0.24	0.01	0.1166	-0.01	0.2945	-0.01	0.17	0.01	0.15
AVGT	X	X	-0.01	0.36	-0.01	0.30	X	X	X	X	X	X	-0.01	0.14	X	X
Odor sensitivity score																
ODORT	0.06	0.11	0.01	0.41		X	0.05	0.19	X	X	X	X	0.11	0.0007	X	X
Odor level at Blue Plains																
JWOL3	X	X	-1.35	0.10	X	X	X	X	X	X	-2.52	0.02	0.65	0.4052	-0.037	0.23
JWL3	X	X	X	X	X	X	0.08	0.46	X	X	-0.16	0.37	0.13	0.2332	X	X
JWOL24	0.01	0.16	-0.02	0.03	-0.01	0.12	0.01	0.27	0.01	0.31	-0.01	0.23	X	X	X	X
JWL24	0.14	0.04	X	X	-0.10	0.11	0.17	0.02	0.12	0.12	X	X	0.09	0.20	0.13	0.08
DTWOL3	-0.00015	0.12	X	X	0.00020	0.01	-0.0002	0.07	-0.00010	0.23	0.00	0.0625	-0.00020	0.09	-0.0002	0.11
DTWL3	-0.00010	0.46	X	X	0.00030	0.02	X	X	0.00010	0.404	0.00	0.0909	X	X	X	X
DTWOL24	0.00001	0.19	0.00	0.18	-0.00001	0.05	0.0001	0.20	0.00000	0.25	X	X	X	X	0.0000	0.31
DTWL24	0.00004	0.27	X	X	-0.00001	0.43	0.0001	0.09	0.00010	0.03	0.00	0.4797	X	X	0.0001	0.02

X interested variable is insignificant

Table 6.7: Simple ordered logit model on the next day by inspector

Time frame d-1	Inspct D								Inspct G							
	DTRG1		DTRG2		DTRG3		DTRG4		DTRG1		DTRG2		DTRG3		DTRG4	
Number of observations in each level	**1**→86 **2**→10 **3**		**1**→89 **2**→7 **3**		**1**→61 **2**→28 **3**→7 **4**		**0**→86 **1**→10		**1**→58 **2**→33 **3**→14		**1**→90 **2**→1 **3**→14		**1**→32 **2**→58 **3**→12 **4**→3		**0**→58 **1**→47	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
Time of day																
M	X	X	X	X	-0.70	0.18	X	X	0.58	0.15	0.46	0.42	0.32	0.43	0.57	0.14
N	-0.64	0.35	-1.16	0.18	X	X	-0.64	0.34	-0.41	0.29	-0.68	0.25	X	X	-0.30	0.48
AF	1.96	**0.01**	2.43	**0.01**	1.13	**0.02**	1.96	**0.01**	X	X	X	X	X	X	X	X
Weather condition																
S	1.92	**0.01**	1.48	**0.07**	X	X	1.91	**0.01**	0.56	**0.17**	0.55	0.38	X	X	0.51	0.22
CLOUD	-1.55	**0.03**	-1.16	**0.15**	X	X	-1.55	**0.03**	-0.78	**0.08**	-0.63	0.36	X	X	-0.82	**0.06**
RP	X	X	X	X	X	X	X	X	X	X	X	X	-1.62	0.19	X	X
Wind condition																
WINDS	0.11	0.27	X	X	0.05	0.45	0.11	0.27	0.04	0.47	0.13	**0.12**	0.12	0.04	X	X
WINDD	-0.87	0.42	X	X	X	X	-0.86	0.42	1.01	**0.03**	X	X	1.45	0.00	0.87	0.07
Temperature																
HT	X	X	-0.03	0.28	0.04	**0.01**	X	X	0.02	**0.15**	0.04	**0.09**	X	X	0.01	0.35
LT	X	X	-0.03	0.32	-0.06	**0.00**	X	X	0.02	**0.09**	0.02	**0.20**	0.01	0.40	0.02	0.14
AVGT	X	X	-0.34	0.28	-0.05	**0.00**	X	X	0.02	**0.11**	-0.02	**0.13**	X	X	0.02	0.22
Odor sensitivity score																
ODORT	-0.63	0.11	-0.87	**0.13**	-0.40	**0.02**	-0.63	0.10	X	X	0.08	0.45	0.13	**0.07**	X	X
Odor level at Blue Plains																
JWOL3	X	X	X	X	4.90	**0.00**	X	X	X	X	7.60	**0.00**	X	X	-2.07	0.20
JWL3	0.45	0.29	0.76	**0.15**	X	X	0.45	0.29	0.24	0.21	0.36	0.23	0.39	**0.05**	0.23	0.24
JWOL24	X	X	-0.07	0.30	-0.04	**0.07**	X	X	0.05	**0.00**	0.08	**0.00**	0.05	**0.00**	0.02	0.16
JWL24	X	X	X	X	-0.28	**0.16**	X	X	0.37	**0.01**	0.43	**0.02**	0.53	**0.00**	0.33	**0.02**
DTWOL3	0.0002	0.25	0.0002	0.34	0.0002	0.34	0.00020	0.24	-0.0005	**0.08**	X	X	-0.0008	**0.01**	-0.00070	**0.02**
DTWL3	X	X	0.0003	0.43	-0.0009	**0.07**	X	X	X	X	X	X	X	X	X	X
DTWOL24	X	X	0.0000	0.47	0.0000	**0.08**	X	X	0.0000	**0.12**	0.0000	**0.02**	0.0000	**0.05**	X	X
DTWL24	0.0003	**0.06**	0.0002	0.32	-0.0001	0.31	0.00030	**0.06**	0.0002	**0.07**	0.0001	0.46	0.0001	**0.16**	0.00030	**0.04**

Table 6.7: Simple ordered logit model on the next day by inspector (continued)

Time frame d-1	Inspct P								Inspct J							
	DTRG1		DTRG2		DTRG3		DTRG4		DTRG1		DTRG2		DTRG3		DTRG4	
Number of observations in each level	**1**	14	**1**	58	**1**		**0**	14	**1**	113	**1**	114	**1**		**0**	111
	2	40	**2**	1	**2**	53	**1**	57	**2**	2	**2**	1	**2**	84	**1**	2
	3	20	**3**	20	**3**	18			**3**		**3**		**3**	1		
					4	3							**4**			
	Coeff.	**P-value**	Coeff.	**P-value**	Coeff.	**P-value**	Coeff.	**P-value**	Coeff.	**P-value**	Coeff.	**P-value**	Coeff.	**P-value**	Coeff.	**P-value**
Time of day																
M	-0.72	0.16	2.15	**0.04**	-2.17	**0.04**	X	X	X	X	X	X	X	X	X	X
N	0.76	**0.10**	-1.25	**0.03**	1.24	**0.03**	X	X	X	X	X	X	-0.46	0.30	X	X
AF	X	X	X	X	X	X	-0.47	0.48	X	X	X	X	0.87	**0.23**	X	X
Weather condition																
S	0.78	**0.10**	-0.61	0.25	0.69	0.19	1.46	**0.07**	X	X	X	X	X	X	X	X
CLOUD	-0.70	**0.13**	0.36	0.49	-0.44	0.39	-1.66	**0.04**	X	X	X	X	X	X	X	X
RP	X	X	X	X	X	X	X	X	X	X	X	X	1.97	0.26	X	X
Wind condition																
WINDS	0.33	0.22	X	X	X	X	0.80	**0.03**	X	X	X	X	-0.12	**0.16**	X	X
WINDD	0.42	0.45	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Temperature																
HT	X	X	X	X	X	X	X	X	X	X	X	X	-0.02	**0.15**	X	X
LT	0.01	0.47	X	X	X	X	0.02	0.38	X	X	X	X	X	X	X	X
AVGT	X	X	X	X	X	X	X	X	X	X	X	X	-0.02	0.28	X	X
Odor sensitivity score																
ODORT	X	X	-0.11	0.40	0.11	0.38	X	X	X	X	X	X	0.12	**0.06**	X	X
Odor level at Blue Plains																
JWOL3	1.60	0.39	-2.31	0.27	2.26	0.26	X	X	X	X	X	X	-2.73	**0.05**	X	X
JWL3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
JWOL24	-0.02	0.37	0.06	**0.06**	-0.06	**0.07**	X	X	X	X	X	X	-0.03	**0.04**	X	X
JWL24	-0.17	0.27	0.77	**0.03**	-0.79	**0.03**	X	X	X	X	X	X	X	X	X	X
DTWOL3	0.0006	**0.04**	0.00	**0.02**	0.0006	**0.02**	X	X	X	X	X	X	-0.0003	**0.06**	X	X
DTWL3	0.0015	**0.00**	0.00	**0.01**	0.0013	**0.00**	0.00190	**0.08**	X	X	X	X	X	X	X	X
DTWOL24	0.0000	0.24	0.00	**0.05**	0.0000	**0.05**	X	X	X	X	X	X	0.0000	0.19	X	X
DTWL24	X	X	X	X	X	X	0.00020	0.34	X	X	X	X	X	X	X	X

149

Table 6.8: Simple ordered logit model on the same day by inspector

Time frame d-0	Inspct D								Inspct G							
	DTRG1		DTRG2		DTRG3		DTRG4		DTRG1		DTRG2		DTRG3		DTRG4	
Number of observations in each level	**1** 80 / **2** 11 / **3**		**1** 90 / **2** 1 / **3**		**1** 62 / **2** 28 / **3** 1 / **4**		**0** 80 / **1** 11		**1** 72 / **2** 27 / **3** 18		**1** 99 / **2** / **3** 18		**1** 36 / **2** 63 / **3** 13 / **4** / **5**		**0** 72 / **1** 45	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
Time of day																
M	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
N	X	X	X	X	X	X	X	X	X	X	X	X	0.43	0.22	X	X
AF	X	X	X	X	-1.12	0.16	X	X	X	X	X	X	-0.71	**0.08**	X	X
Weather condition																
S	0.93	0.21	X	X	X	X	1.08	0.15	X	X	0.54	0.33	X	X	X	X
CLOUD	-1.21	**0.08**	X	X	X	X	-0.85	0.25	X	X	X	X	0.33	0.40	X	X
RP	X	X	X	X	X	X	X	X	-0.99	0.22	X	X	-0.88	0.18	-0.83	0.30
Wind condition																
WINDS	0.41	**0.00**	X	X	0.24	**0.00**	0.41	**0.00**	0.11	**0.04**	0.09	0.24	0.09	0.12	0.15	**0.02**
WINDD	1.80	**0.01**	X	X	1.40	**0.00**	1.79	**0.01**	0.91	**0.07**	1.75	**0.10**	0.64	0.14	0.98	**0.07**
Temperature																
HT	-0.06	**0.00**	X	X	-0.04	**0.01**	-0.06	**0.00**	0.01	0.35	0.03	**0.10**	X	X	X	X
LT	-0.06	**0.01**	X	X	-0.05	**0.01**	-0.06	**0.01**	0.01	0.31	0.03	**0.08**	0.01	0.36	X	X
AVGT	-0.06	**0.01**	X	X	-0.04	**0.01**	-0.06	**0.01**	0.03	0.32	0.03	**0.08**	X	X	X	X
Odor sensitivity score																
ODORT	-0.15	**0.03**			-0.08	**0.02**	0.55	**0.00**	-0.02	0.25	-0.02	0.28	X	X	X	X
Odor level at Blue Plains																
JWOL3	X	X	X	X	X	X	X	X	4.42	**0.00**	10.82	**0.00**	4.07	**0.01**	1.19	0.48
JWL3	0.52	**0.09**	X	X	X	X	0.71	**0.03**	-0.15	0.45	X	X	X	X	-0.18	0.35
JWOL24	-0.04	0.24	X	X	-0.07	**0.01**	-0.04	0.30	0.06	**0.00**	0.09	**0.00**	0.07	**<0.0001**	0.03	**0.01**
JWL24	0.48	**0.05**	X	X	X	X	0.55	**0.03**	X	X	0.15	0.32	0.24	**0.05**	X	X
DTWOL3	-0.0010	**0.07**	X	X	-0.0002	0.37	-0.0009	0.12	X	X	0.00020	0.24	-0.00040	**0.06**	-0.00030	0.24
DTWL3	X	X	X	X	-0.0002	0.39	X	X	-0.00030	0.36	X	X	-0.00040	**0.17**	-0.00060	0.16
DTWOL24	X	X	X	X	-0.0001	**0.09**	X	X	0.00001	**0.01**	0.00001	**0.00**	0.00001	**0.00**	0.00001	**0.08**
DTWL24	0.0003	**0.02**	X	X	X	X	0.0004	**0.02**	0.00020	**0.06**	0.00030	**0.01**	0.00030	**0.00**	X	X

Table 6.8: Simple ordered logit model on the same day by inspector (continued)

Time frame d-0	Inspct P								Inspct J							
	DTRG1		DTRG2		DTRG3		DTRG4		DTRG1		DTRG2		DTRG3		DTRG4	
Number of observations in each level	**1** 29		**1** 74		**1** 1		**0** 25		**1** 133		**1** 134		**1** 31		**0** 134	
	2 46		**2** 1		**2** 74		**1** 72		**2** 1		**2** 1		**2** 103		**1** 1	
	3 273		**3** 27		**3** 27				**3** 1		**3** 1		**3**			
					4 1								**4** 1			
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
Time of day																
M	0.37	0.34	-0.44	0.36	-0.55	0.25	2.01	**0.01**	X	X	X	X	X	X	X	X
N	-0.56	0.11	X	X	-0.35	0.39	-0.65	0.14	X	X	X	X	-1.03	**0.01**	X	X
AF	-0.53	0.21	X	X	X	X	-0.71	0.14	X	X	X	X	X	X	X	X
Weather condition																
S	0.45	0.23	0.36	0.41	0.32	0.44	0.49	0.32	X	X	X	X	-1.12	**0.02**	X	X
CLOUD	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
RP	-1.82	**0.01**	X	X	X	X	-1.58	**0.04**	X	X	X	X	2.32	**0.00**	X	X
Wind condition																
WINDS	-0.60	**0.01**	-0.41	**0.10**	-0.43	**0.09**	-0.91	**0.00**	X	X	X	X	0.08	**0.03**	X	X
WINDD	-1.48	**0.00**	-1.16	**0.03**	-1.11	**0.03**	-1.73	**0.00**	X	X	X	X	X	X	X	X
Temperature									X	X						
HT	0.01	0.19	0.01	0.44	0.01	0.37	0.02	0.16	X	X	X	X	-0.01	0.23	X	X
LT	0.01	0.42	X	X	X	X	0.02	0.14	X	X	X	X	-0.02	0.19	X	X
AVGT	0.01	0.24	X	X	X	X	0.02	0.11	X	X	X	X	-0.02	0.17	X	X
Odor sensitivity score																
ODORT	0.03	**0.08**	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Odor level at Blue Plains																
JWOL3	1.08	0.38	X	X	X	X	1.46	0.40	X	X	X	X	-2.53	**0.02**	X	X
JWL3	0.22	0.23	0.35	0.15	0.36	0.14	0.17	0.49	X	X	X	X	X	X	X	X
JWOL24	X	X	-0.05	**0.03**	-0.04	**0.06**	0.04	**0.08**	X	X	X	X	0.02	0.16	X	X
JWL24	X	X	-0.42	**0.05**	-0.44	**0.04**	0.49	**0.04**	X	X	X	X	0.17	0.14	X	X
DTWOL3	0.00300	**0.06**	0.0004	**0.03**	0.0005	**0.01**	X	X	X	X	X	X	-0.0004	**0.00**	X	X
DTWL3	0.00050	**0.06**	0.0004	0.12	0.0006	**0.05**	0.00060	0.20	X	X	X	X	-0.0008	**0.00**	X	X
DTWOL24	-0.00001	0.39	-0.0001	**0.02**	-0.0001	**0.03**	0.00001	0.11	X	X	X	X	0.0001	**0.02**	X	X
DTWL24	-0.00010	0.24	-0.0005	**0.02**	-0.0004	**0.02**	0.00010	0.40	X	X	X	X	X	X	X	X

For each data set (individual inspectors or combined), significances of all pertinent variables DTRG1, DTRG2, DTRG3, and DTRG4 were checked. Full analysis of all inspectors can be seen in Tables 6.7 and 6.8.

DTRG1 is preferred to DTRG2 since according to the odor rule in many states D/T 7 should not be regrouped into the same class as D/T 2 and D/T 4, which are considered as low or acceptable D/T levels.

DTRG3 is also problematic since it has four classes. For the combined inspectors data at high D/T levels such as DTRG3 = 4, there were only a few observations in this class relative to other classes, i.e., seven observations in the d-0 data set and six observations in the d-1 data. It is even worse when analysis was performed individually for each inspector. There was no observation in some classes for inspectors D, P, and J on both the d-1 and d-0 data sets. Therefore, we selected only DTRG1 and DTRG4 for further analysis based on the number of observations in each class and the state odor regulation.

As shown in Table 6.6, we lost a great amount of data after the field odor data and the Blue Plains' odor data were merged. Another approach was conducted in order to observe only the relationship between the field odor data (DTRG1 and DTRG4) and field conditions. This approach gains a great number of observations (2000 vs. 500). Results of simple ordered logit model on unmerged field data set are showed in Table 6.9.

Table 6.9: Simple ordered logit model on field data 2005-2006 (unmerged set)

	Combined inspectors			
	DTRG1		DTRG4	
Number	**1**	1322	**0**	1322
of	**2**	450	**1**	618
observati	**3**	166		
ons in				
	Coeff.	**P-value**	Coeff.	**P-value**
Time of day				
M	X	X	X	X
N	-0.16	0.10	-0.17	0.08
AF	0.22	0.05	0.23	0.04
Weather conditions				
S	-0.25	0.01	-0.26	0.01
CLOUD	0.32	0.001	0.33	0.001
RP	-0.46	0.06	-0.46	0.07
Wind conditions				
WINDS	0.09	<0.0001	0.10	<0.0001
WINDD	0.18	0.07	0.19	0.05
Temperature				
HT	0.01	0.15	0.004	0.17
LT	0.01	0.001	0.01	0.001
AVGT	0.01	0.011	0.01	0.013

X   variable of interest is insignificant at 10 percent significance level

For further discussion, we call the result based on this analysis the analysis on the unmerged set.

6.5.1 Results from Simple Ordered Logit Model

The results from analyses on the merged data sets (d-1 and d-0, Table 6.6) and unmerged data set (the field data, Table 6.9) are summarized below.

Time of the day: Field odor measurement before 10 am (M) is insignificant in the unmerged data set and in the d-1 data set. The dummy variables N and AF are significant for all data sets. N is negatively correlated with field odor measurement meaning that if field odor is measured between 10 am to 14 pm it tends to yield lower D/T levels. However, there is no strong reason to support a negative coefficient of N variable. In contrast, a positive correlation for the variable AF is reasonable since odor measurement

after 14pm represents an effect of heating the ground that encourages higher biosolids odor emission rates compared to the time when the ground is cold.

Weather conditions: CLOUD is strongly significant at less than a 5 percent level on the unmerged data set and the d-0 data set. This variable had a positive coefficient meaning that if field odor measurements were taken in cloudy weather conditions, the observed D/T level is likely to be at a higher field odor level compared to if no clouds were present. CLOUD in the d-1 data set is also negatively significant but with a higher p-value (at a 14 % level)

S and RP are significantly negative for the unmerged and the d-0 data sets. These results are counter-intuitive. We expected sunny weather conditions to heat up the ground and faster odor emission rates. Precipitation is also expected to encourage higher odor emission rates since it provides humidity to microorganisms to produce more odorous gas. Applying multiple variables in an ordered logit model could possibly provide more explanation.

Wind speed: Positive coefficients on the wind related variables WINDS and WINDD are significant for all data sets at the 5% level or below. On one hand, higher wind speeds disperse and dilute odor from the field sites to the nearby area (negative coefficient). On the other hand, moderate wind speed can transport odor to other areas (positive coefficient). Further analysis of what wind speed the odor will disperse or transport odor to nearby area should be conducted.

Temperature: HT, LT, and AVGT are significantly positive in the unmerged data set (Table 6.9). However, only LT is significantly positive in both merged data sets. Thus, LT (the lowest temperature) should be used for further analysis.

Odor levels at Blue Plains: DTWL24 and JWL24 are positively associated with the field odor data for both d-1 and d-0 data sets (Table 6.6). The rest of the BP odor variables, such as DTWOL24 and JWOL24 are also significantly positive. DTWL24 will be proceeded to further analysis in multiple factors' works as stated in the assumption. DTWL24 showed a positive sign but with a small coefficient value (i.e., when DTRG1 regresses on DTWL24, it shows the coefficient values of 0.00004 on the d-0 set and 0.0001 on the d-1 set shown in Table 6.6). This is due to the large values of the DTWL24 variable ranging from 240 ou to 8000 ou compared to DTRG1 values (1, 2, and, 3). Therefore, the dummy variable DTWL24D1 (DTWL24D1 = 1 when DTWL24 > 1000 ou and zero, otherwise) was created to use in next section.

Inspector's odor sensitivity score: inspector's odor sensitivity score has a counter-intuitive coefficient sign. It is positively correlated with the field odor data on d-1 set and insignificant on the d-0 data set. We expect a negative coefficient's sign on this variable. Thus, an inspector with a high inspector's odor sensitivity score, which represents the sensitive nose, often detects odor at a low D/T level and an inspector with a low sensitivity score, insensitive nose, often detect the same odor at a higher level. This will be further discussed when we run analysis on multiple variables.

Next, combined effects of interested factors on field odor data were tried using multiple ordered logit model.

6.6 Ordered Logit Model: Multiple Variables

Among all variables tried, variables related to wind speed (i.e., wind speed (WINDS) and a dummy wind speed (WINDD)) consistently showed a positive relationship with field odor levels. This is, however, counterintuitive since at a high wind

speed, odor should be dispersed from farm sites as mentioned in Hamilton and Carlson (2008). This article describes that at a high wind speed above 13 mph coupled with strong solar radiation during the day time excellent air dispersion conditions would be anticipated. At a wind speed below 13 mph without a heavy overcast sky, moderately good air dispersion condition would be anticipated. Only with a heavy overcast sky during the day time would there be anticipated moderately poor to poor air dispersion conditions.

Therefore, a dummy variable WINDD13 (dummy wind speed at 13 mph) was created where WINDD13 = 1 if a wind speed is greater than 13 mph and zero, otherwise. An interaction variable HWIND (high wind speed), which is a product of WINDS*WINDD13 (HWIND = WINDS*WINDD13) was also created where HWIND equals zero if a wind speed is below 13 mph and equals a corresponding wind speed when it is higher above 13 mph. This variable was created to investigate the effect of a unit increase in miles per hour for a wind speed above 13 mph at the field. Therefore, a negative coefficient is expected for this interaction variable since at a high wind speed the odor should be diluted from the field site.

The odor level at the Blue Plains plant (DTWL24), representing an original odor level of the product applied to the field sites, was included in all models tried. Results from the previous section showed that DTWL24 was positive and significant in explaining the field odor levels. As discussed in previous section, a dummy variable DTWL24D1 (a dummy variable of DTWL24) was created to minimize the low estimated coefficients of DTWL24 in Tables 6.6 and 6.7. When DTWL24>1000 ou, a dummy variable DTWL24D1 = 1 and zero, otherwise. The threshold of 1000 ou came from the

point where we distinguished a high odor level (H) from a moderate-to-low odor level (MTL) as described in Chapter 4 and is also used here to distinguish classify odor level of biosolids applied to the field site.

Next, we describe multiple factors expected to influence field odor measurements and present different models used in the analysis.

6.6.1 Assumptions when Multiple Factors Influence Field Odor Measurements

Based on an assumption that there is more than one factor that could influence field odor measurements, an ordered logit model is developed using multiple factors associated with the field odor data. Four variables: DTWL24D1 (a dummy variable of DTWL24), WINDS (a wind speed), HWIND (an interaction variable at high wind speed), and LT (the lowest temperature) are the basic variables included in all models tried due to their significance from the previous section and their relationship to the field odor dispersion from the literature (USEPA, 2000). DTWL24D1 represents original biosolids odor levels at the Blue Plains plant. WINDS and HWIND are wind-related variables associated with odor dispersion where HWINDS represents an additional effect of a wind speed higher than 13 mph. Last, LT represents the lowest daily field temperature influencing field odor dispersion and ongoing biological activities of odor-causing microorganisms in biosolids.

In addition to these variables, the following are additional anticipated ones.

1. **Inspector's odor sensitivity score** (ODORT): ODORT is included in the model to check whether the sensitivity score of an inspector can determine his/her field

odor measurements. Thus, DTRG1 was modeled as a function of DTWL24, WINDS, HWIND, LT, and ODORT for the d-1 and d-0 data sets

2. **Weather conditions** (S (sunny), RP (rain and precipitation), and CLOUD): weather conditions during the field odor measurements were checked separately for their impacts on field odor measurements. Specifically,

DTRG1 was modeled as a function of DTWL24, WINDS, HWIND, LT, and S for the d-1 and d-0 data sets;

DTRG1 was modeled as a function of DTWL24, WINDS, HWIND, LT, and RP for the d-1 and d-0 data sets;

DTRG1 was modeled as a function of DTWL24, WINDS, HWIND, LT, and CLOUD for the d-1 and d-0 data sets

3. **Field odor measurement time of the day** (M (morning), N (noon), and AF (afternoon)): This variable is to check if the time that the field odor measurement was taken had an impact on the odor levels. The following models are tried.

DTRG1 was modeled as a function of DTWL24, WINDS, HWIND, LT, and M for the d-1 and d-0 data sets;

DTRG1 was modeled as a function of DTWL24, WINDS, HWIND, LT, and N for the d-1 and d-0 data sets;

DTRG1 was modeled as a function of DTWL24, WINDS, HWIND, LT, and AF for the d-1 and d-0 data sets.

Logistic regression using SAS was implemented to conduct the ordered logit model based on multiple variables. Stepwise regression technique was used for variable

selection based on a probability to enter in the model at 0.4 (SLENTRY = 0.4) and a probability to stay in the model at 0.3 (SLSTAY = 0.3).

6.6.2 Results: Multiple Variables

The findings from multiple data analysis showed that four variables were significant for all models that we tried: DTWL24D1 (a dummy variable for biosolids odor level at the plant), WINDS (wind speed), HWIND (an interaction variable for wind speed higher than 13 mph), and LT (the lowest temperature). Variables that were not strongly related to field odor levels were variables related to weather conditions (S (sunny), RP (rain and precipitation), and CLOUD), variables related to field odor measurement time of the day (M (morning), N (noon), and AF (afternoon)), and inspector's odor sensitivity score (ODORT). The followings are the details of multiple data analysis

Table 6.10 summarizes results from all models mentioned in the previous section; Models were tested on two data sets: d-1 and d-0 data sets. The heading score test at the top of table is a test statistic to check whether the proportional odds assumption of the model, the assumption of common slope (coefficient) on all levels of response variable, was valid. If a score test is insignificant (p-value > 0.2), a corresponding ordered logit model is valid.

The AIC heading to the right at the top of table stands for Akaike Information Criterion (AIC) (Akaike, 1974 and 1976). AIC is normally used to compare models. All things being equal, models with smaller AIC are preferred. AIC is defined as follows:

$$AIC = -2 \log L + 2((k-1) + s) \qquad (6.4)$$

where

L = the maximum likelihood of the fitted model

K = the total number of response levels

s = the number of predictors in the model

Table 6.10: Results from testing ordered logit models on multiple variables

					Score test	AIC
DTWL24D1	**HWIND**	WINDS	LT	**ODORT**	Pr> chiSq	
Inspector's odor sensitivity score						
d-0	x	**-0.11	**0.23	**0.016	**0.10	0.90	686.93
d-1	**0.88	**-0.15	**0.19	*0.014	x	0.25	600.80
Weather condition						
DTWL24D1	**HWIND**	WINDS	LT	**S**		
d-0	0.32	**-0.10	**0.22	0.01	-0.3	0.34	690.77
d-1	**0.88	**-0.15	**0.19	*0.01	X	0.25	600.80
DTWL24D1	**HWIND**	WINDS	LT	**RP**		
d-0	0.33	**-0.10	**0.22	0.01	X	0.57	690.75
d-1	**0.88	**-0.15	**0.19	*0.01	X	0.25	600.80
DTWL24D1	**HWIND**	WINDS	LT	**CLOUD**		
d-0	0.34	**-0.10	**0.22	0.01	*0.40	0.50	689.33
d-1	**0.86	**-0.15	**0.18	*0.01	-0.26	0.31	601.44
Time of the day						
DTWL24D1	**HWIND**	WINDS	LT	**M**		
d-0	0.33	**-0.10	**0.22	0.01	X	0.57	690.75
d-1	**0.88	**-0.15	**0.19	*0.01	X	0.25	600.80
DTWL24D1	**HWIND**	WINDS	LT	**N**		
d-0	0.32	**-0.10	**0.22	*0.01	*-0.35	0.71	689.95
d-1	**0.87	**-0.16	**0.19	*0.01	-0.26	0.25	601.51
DTWL24D1	**HWIND**	WINDS	LT	**AF**		
d-0	0.33	**-0.10	**0.22	0.01	X	0.57	690.75
d-1	**0.88	**-0.15	**0.19	*0.01	X	0.25	600.80

**significant level < 0.05
* significant level < 0.10
X   insignificant variable

For all models, the d-1 data set has lower AIC score and is thus preferred over the
d-0 data set. In addition, DTWL24 and LT are more significant in the d-1 data set than in
the d-0 data set in almost all models. Thus, in general, the merged d-1 data set fits the
models better than the d-0 data set.

A dummy DTWL24 (DTWL24D1) was strongly significant at less than the 5% level for all models in the d-1 data set and had a positive effect on the field odor level. This indicated that biosolids odor level especially at 1000 ou and higher at the plant influenced field odor levels.

The lowest temperature (LT) was also strongly significant for all models in the d-1 data (< 10% level) and was positively correlated with field odor levels. This is reasonable since temperature at the field site was related to the microorganism activity and biosolids odor emission rate.

An interaction of high wind speed (HWIND) and wind speed (WINDS) were both strongly significant for all models on both data sets and at a significance level less than 5%. WINDS was positively associated with DTRG1 interpreted that the higher the wind speed the higher the field odor observation. However, at a wind speed above 13 mph, HWIND showed a negative effect with DTRG1 meaning that an additional mile per hour of a wind speed above 13 mph results in lower DTRG1 levels.

Inspector's sensitivity score (ODORT) was insignificant in the d-1 data set and it was disregarded in the d-0 data set due to a counterintuitive coefficient sign.

For the weather conditions, S (sunny) was insignificant on the d-1 data set. RP (rain and precipitation) was also insignificant for both data sets. CLOUD was insignificant at 10% level and showed a negative coefficient for the d-1 set.

Concerning the odor measurement time of the day, M (morning), and AF (afternoon) were insignificant for both data sets. N (noon) was negatively associated with DTRG1 for both sets meaning that the field odor measurement around noon time results

in lower odor levels. This is counterintuitive and has no supporting theory. N was removed for further analysis.

Next, we present the final model incorporating significant variables to explain field odor levels.

6.6.3 Final ordered logit model

From multiple data analysis results, four variables were identified as significant for all models: DTWL24D1, WINDS, HWIND, and LT. Additional analyses were conducted to check whether an average temperature (AVGT) and a highest temperature (HT) could replace the lowest temperature (LT). Consequently, LT was replaced by HT in one model and AVGT in another model keeping DTWL24D1, WINDS, and HWIND as they were. Ordered logit models with respect to these two variables were built. Results from stepwise variables selection showed that both variables, AVGT and HT, were insignificant in the models. Thus, LT was the temperature variable that was kept.

Additional analyses were also conducted for HWIND. It appeared that at wind speeds higher than 13 mph, an additional mile per hour of wind speed resulted in the likelihood of observing lower field odor levels. Thus, we wanted to assess the significance of the same interaction variable HWINDS but at the different wind speed thresholds of 5, 7, and 10 mph on field odor levels. The results showed that a threshold of a wind speed above 13 mph in HWIND was more statistically significant compared to a threshold of 5, 7, and 10 mph. The details of analysis are the following.

After changing the cut off value to 5 mph, the estimated coefficients' signs of WINDS (wind speed) and HWIND (an interaction variable of high wind speed) from the ordered logit model were reversed. The resulting model showed a negative coefficient for

WINDS and a positive coefficient for HWIND (at 5 mph) while the rest of variables had the same coefficient signs as they were. This is counterintuitive since at wind speed above 5 mph the odor should be dispersed faster resulting in the negative coefficient sign for HWIND. Thus, we rejected an interaction variable HWIND at 5 mph.

At 7 mph, both HWINDS and WINDS had correct coefficient signs, a negative coefficient and a positive coefficient, respectively. However, at this threshold (a wind speed above 7 mph) HWIND was less significant (p-value less than 10 percent) compared to HWINDS at a wind speed above 13 mph (p-value less than 5 percent).

At 10 mph, both HWINDS and WINDS had correct coefficient signs and were statistically significant at a 5 percent level. The AIC of this model was compared to the one with HWIND above 13 mph. With the same number of independent variables in the model, the model with HWIND above 10 mph model had a higher AIC value than the model with HWIND above 13 mph at the values of 601.865 and at 600.801, respectively. This indicated a better fit for the latter model.

In conclusion, a threshold of a wind speed above 13 mph in HWIND showed a better fit compared to a threshold of 5, 7, and 10 mph.

Finally, DTWL24D1, WINDS, HWIND (at above 13 mph), and LT are selected to describe field odor measurements. The SAS output of the ordered logit model using the selected independent variables is shown below.

```
                The LOGISTIC Procedure

                  Model Information
        Data Set                   WORK.FIELDUSE2
        Response Variable          DTRG1
        Number of Response Levels  3
        Number of Observations     387
        Model                      cumulative logit
        Optimization Technique     Fisher's scoring
                    Response Profile

          Ordered                      Total
           Value        DTRG1        Frequency
```

```
                                1                3                34
                                2                2                85
                                3                1                268
       Probabilities modeled are cumulated over the lower Ordered Values.


                              Model Convergence Status

                   Convergence criterion (GCONV=1E-8) satisfied.


                  Score Test for the Proportional Odds Assumption
                       Chi-Square        DF      Pr > ChiSq

                         5.3546           4         0.2528

                           The LOGISTIC Procedure

                             Model Fit Statistics

                                                  Intercept
                                       Intercept     and
                       Criterion         Only     Covariates

                       AIC             624.008      600.801
                       SC              631.925      624.552
                       -2 Log L        620.008      588.801


                    Testing Global Null Hypothesis: BETA=0

              Test                Chi-Square        DF      Pr > ChiSq

              Likelihood Ratio      31.2073          4        <.0001
              Score                 29.7153          4        <.0001
              Wald                  30.1006          4        <.0001

NOTE: All effects have been entered into the model.

                    Analysis of Maximum Likelihood Estimates

                                        Standard        Wald
       Parameter      DF    Estimate      Error     Chi-Square    Pr > ChiSq

       Intercept 3     1    -4.8448      0.6029      64.5674        <.0001
       Intercept 2     1    -3.2331      0.5677      32.4305        <.0001
       DTWL24D1        1     0.8763      0.2314      14.3453        0.0002
       HWIND           1    -0.1548      0.0422      13.4888        0.0002
       WINDS           1     0.1869      0.0422      19.6022        <.0001
       LT              1     0.0143      0.00766      3.4745        0.0623
                           The LOGISTIC Procedure

                             Odds Ratio Estimates

                               Point          95% Wald
                 Effect       Estimate    Confidence Limits

                 DTWL24D1       2.402      1.526      3.780
                 HWIND          0.857      0.789      0.930
                 WINDS          1.205      1.110      1.309
                 LT             1.014      0.999      1.030


       Association of Predicted Probabilities and Observed Responses

              Percent Concordant     66.3    Somers' D    0.342
              Percent Discordant     32.1    Gamma        0.348
              Percent Tied            1.6    Tau-a        0.159
              Pairs                 34782    c            0.671
```

From the SAS output, an equation associated with the field odor level in terms of

DTRG1 is

$$\text{logit}\left[P(DTRG1 \le j|x)\right] = \alpha_j + 0.8763 \text{ DTWL24D1} - 0.1548 \text{ HWIND} + 0.1869$$
$$\text{WINDS} + 0.0143 \text{ LT} \qquad\qquad (6.5)$$

where

$\alpha_3 = -4.8448$

$\alpha_2 = -3.2331$

$j = 1,..., J - 1$

Descriptions for each independent variable in the model are as follows.

DTWL24D1: a coefficient of 0.8763 for DTWL24D1 means that on any given day if DTWL24 is greater than 1000 ou (DTWL24D1 = 1), the odds of getting higher DTRG1 is 2.04 ($e^{0.8763}$) that relative to the day when DTWL24 less than or equal to 1000 ou (DTWL24D1 = 0) given that other variables in the model were held constant.

HWIND: a negative coefficient of -0.1548 of HWIND means that on any given day if a wind speed is greater than 13 mph (HWIND = 1), faster wind by one more mile gives the odds of getting lower DTRG1 of 0.857 ($e^{-0.1548}$) relative to a wind speed below 13 mph (HWIND = 0), all else being equal held.

WINDS: a positive coefficient of 0.1869 of WINDS means that a unit increase in wind speed gives the odds of getting higher DTRG1 of 1.205 ($e^{0.1869}$) given other variables in the model are held constant.

LT: a positive coefficient of 0.0143 of LT means a unit increase in the recorded lowest temperature gives the odds of getting higher DTRG1 of 1.014 ($e^{0.0143}$) given other variables in the model are held constant.

6.7 Conclusions

In this chapter, another type of statistical model, the ordered logit model, was used to assess the relationship between an ordinal response variable for field odor levels

and the potential explanatory variables including: the odor measurement time of the day, the weather conditions, the wind conditions, the field temperature, the inspector's odor sensitivity score, and the biosolids odor level at the plan. Four variables were identified as best describing the odor measurement at the field site:

1) A dummy variable related to the biosolids odor level of 1000 ou at the plant (DTWL24D1) where DTWL241 equals 1 if DTWL24 > 1000 ou and zero, otherwise.

2) A wind speed (WINDS) and an interaction variable of a wind speed and a dummy wind speed at 13 mph (HWIND). Combined effects of these variables show that when wind speed is below 13 mph the higher the wind speed the higher the field odor observation. At wind speed above 13 mph, the interaction variable HWIND shows a negative coefficient meaning that a unit increase for a wind speed above 13 mph would result in the lower field odor observation.

3) A lowest temperature shows a positive effect on field odor measurement. The higher the lowest temperature at the site results in the higher the field odor observation.

Chapter 7 Conclusions and Future Works

In this dissertation, several statistical analyses were applied to the data at the Blue Plains plant and at application sites that no one has done before. They were used to identify variables influencing biosolids odor emissions and to develop statistical functions to explain the biosolids odor production. DCWASA and other wastewater treatment plants with similar types of unit operations can benefit from using the findings in this dissertation to improve their biosolids management. For instance, this can be done by collecting data related to biosolids odor production identified in this study to develop the in-house biosolids odor prediction model, using similar statistical approaches to identify and confirm the odor-causing variables, investigating the impact of uncertainty on the developed model using simulation analysis, etc.

In this dissertation, the introduction, literature review, and data collection were discussed in Chapter 1, 2, and 3, respectively. The analyses of data collected were discussed in Chapter 4, 5, and 6. Findings in this study are summarized below.

In Chapter 4, biosolids odor prediction models at the plant were developed with respect to continuous and categorical odor data. In the continuous odor modeling approach, DT data (detection threshold) was used as the response variable. Different subsets of collected data: high (H) and moderate-to-low (MTL) sets, summer and non-summer sets, and the full data set were created to find explanatory variables that best explained biosolids odor emissions for different subsets. Finally, models based on the high odor data set (at DT levels greater than 2000 ou), summer and non summer data sets, and the full data sets were developed. The performances of those models were compared on 20 randomly selected validation sets and 20 test sets. A model based on the full data

set outperformed the rest with the lowest mean absolute error (MAE). The full data set model was composed of six variables from three groups of data: the characteristics of biosolids (i.e., the percent solids and temperature of biosolids sample), the characteristic of wastewater sludge (i.e., the pH of sludge from the gravity thickener tank), and operational variables at the plant (i.e., the percentage of the gravity thickener solids in the blend tank, the concentration of returned activated sludge at the secondary process, and the number of centrifuges in service).

In addition to the continuous model, categorical biosolids odor prediction models were developed. Benefits from categorical models include appropriateness in terms of classifying biosolids as either malodorous or regular biosolids odor. In particular, this is the information biosolids management would like to know whether today they are going to produce odorous product.

The DT data were classified into two groups: a high odor group (DT>1000 ou) and moderate-to-low odor group (DT<=1000 ou). A number of logistic regression models and discriminant analysis models were developed for categorical response variable (DT). The performances of the best models from two approaches were compared using the validation sets. The model from discriminant analysis was best.

In reality, the cost of misclassifying odorous biosolids as non-odorous (Type 1) and the cost of misclassifying non-odorous biosolids as odorous (Type 2) are not equal. Mistakenly sending odorous product to sensitive areas can cause adverse effects on the community. Therefore, a new equation considering unequal misclassification costs was developed to analyze the performance of categorical models. The best models from two approaches were compared again using a developed equation at the ratio of

misclassification cost of 3:1[9], 5:1, 7:1, 9:1 and 11:1, where the first is the cost for misclassification type 1 and the latter is the cost for misclassification type 2. The discriminant analysis model performed best again on the validation sets. The best categorical model had six variables similar to the best continuous biosolids odor prediction model. These variables were 1) the percent solids and 2) temperature of biosolids sample, 3) the pH of sludge from the gravity thickener tank, 4) the percentage of the gravity thickener solids in the blend tank, 5) the concentration of returned activated sludge at the secondary process, and 6) the number of centrifuges in service.

In summary, both continuous and categorical models can be used in a daily biosolids odor management to gain information on the potential biosolids odor. The continuous model gives the output in terms of the predicted DT levels and the categorical model gives the output as predicted class of biosolids odor (high or normal odor level)

Chapter 5 involved a simulation and sensitivity analysis on the selected model, which are the full data set model. Since uncertainty exists in most real-world operations, the simulation analysis was conducted on a selected biosolids odor prediction model. The probability distributions for independent variables in the full data set model were defined by @Risk based on the 77 observations of data used to develop the model as well as additional PCH data, if available, to more closely match historical data. Also, odor distributions taking into account correlation between input variables were developed. It was found that on average biosolids at the Blue Plains plant had an odor level of 1164 ou, which was lower than the average value of 1582 from the collected data. In addition, the

---

[9] It cost three times more for misclassification in type 1 than type 2.

chance to produce odorous biosolids with odor level greater than 2000 ou would be 12 percent according to the current operations at the plant.

Sensitivity analysis was also conducted for the simulation models. In particular, the number of centrifuges in service and the percentage of gravity thickener solids in the blend tank were varied with the remaining variables stochastic to see the resulting probability distributions. The results suggested that more centrifuges running could reduce the likelihood of high odor and reducing the percentage of the GT solids in the blend tanks to 50% or less could also maintain biosolids odor at an acceptable level.

Finally, the results form an analysis for various scenarios corresponding to the two decisions variables mentioned above recommended that running at least nine centrifuges and maintaining the percentage of the gravity thickener sludge in the blend tank between 40 % and 60 % would lead to biosolids with odor levels below 1000 ou.

Chapter 6 investigated the connection between biosolids odor data at the plant and the field odor data. Four inspectors were selected for analysis based on diversity in their field odor data and a sufficient number of observations after matching their field data with the Blue Plains data. Results from the ordered logit model found that biosolids odor levels at Blue Plains when higher than 1000 ou, the recorded lowest temperature at the field site, the wind speed, all had positive effects on the field odor level while an interaction variable of wind speed at 13 mph had a negative effect.

Limitations

The following are limitations that were encountered as part of the data collection and modeling efforts.

Limited biosolids odor observation at the plant: For cost reasons, biosolids odor data at the Blue Plains plant were limited to 77 observations. As the field data are collected almost everyday except weekend, limited biosolids odor observations at the plant results in unused information from field data that can not be matched with the odor at Blue Plains for analysis.

Missing chemical addition data: During the data collection step, there were no reliable data on chemical additions for the processes (e.g. amount of polymer, ferric chloride, and lime addition, additions). These data are important since the amount of chemical additions can change the characteristics of wastewater and wastewater solids that eventually become the source of biosolids odor production. Missing this piece of information is assumed to be one of the reasons that the final biosolids odor prediction model showed a lower explanatory power than expected. Only after this study, these variables are recorded and stored as standard operation procedure at the Blue Plains plant.

Future works

Findings in this dissertation can be implemented by first collecting data needed to run biosolids odor prediction models and reporting the prediction results to biosolids manager on a daily basis. For the data already available in the PCH database (i.e., the concentration of return activated sludge of secondary sedimentation tank, the GT% in the blend tank, the number of centrifuges running, temperature and percent solids of biosolids), DCWASA should acquire these data specifically for the biosolids odor prediction purpose. For variables, such as the pH of GT solids that have not been collected daily, DCWASA should assign operators to collect this data and input into PCH

database. Finally, the predicted biosolids odor level and class should be reported to the biosolids manager at the beginning of the day.

Design of experiments (DOE) to evaluate influencing factors contributing to biosolids odor emissions from the real unit operations at Blue Plains was considered. However, due to the high volumes of wastewater (370 million gallons) and wastewater solids (1300 dry tons of biosolids) needed to be processed every day it was difficult to run a controlled experiment (e.g., control or adjust the variables related to biosolids odor emissions) on the on-going wastewater treatment operation at Blue Plains. Therefore, DOE was not used.

Due to the change in operation at Blue Plains, model should be calibrated. As most of data at Blue Plains now are available in PCH database, the data, such as chemical additions that were not available during the data collection in this study and the change in the operation, such as different types of chemical added into the process can be further used to investigate the effects to these missing information and the change in operation on biosolids odor production as well as to calibrate the current models.

Lastly, DCWASA should implement new odor measurement techniques, such as real-time biosolids odor data acquisition and data monitoring for keys variables identified in this study (i.e., sludge blanket level, RAS concentration, blending ratio, number of centrifuges running etc.). This will allow DCWASA to be able to monitor biosolids odor levels at the plant and observe any variations in key factors in a timely manner.

Sensitivity analysis on the impacts of two decision variables identified in this study (i.e., number of centrifuges running and percentage of GT solids in the blend tank) on biosolids odor emission could also be conducted. DCWASA could adjust these key

factors and investigate the corresponding biosolids odor emissions. Thus, resources could be allocated accordingly to obtain the optimal performance for these key factors (e.g., maintenance).

In addition, with real-time data collection, DCWASA could gain insight into the causes of biosolids odor fluctuations with respect to other operational parameters as well as characteristics of sludge at the plant. .Lastly, this increasing amount of data will enable DCWASA to conduct more complete research on the field odor modeling.

Appendix: Data used

date	observation #	DTWL24	% solid	With lime 24 T	GT T	BS T	DAF T
04/18/05	1	310	23.50	22.50	18.65	20.70	20.43
04/19/05	2	330	26.93	22.50	18.92	20.95	20.60
04/25/05	3	240	24.04	18.50	17.52	19.60	19.66
04/26/05	4	280	23.58	18.50	18.92	20.95	20.60
05/02/05	5	330	24.03	19.50	18.13	20.19	20.07
05/03/05	6	1800	24.78	19.50	19.00	21.03	20.66
05/09/05	7	3200	24.35	20.00	20.40	22.38	21.60
06/06/05	8	3400	24.05	23.50	21.53	23.48	22.37
06/07/05	9	1000	32.21	23.50	22.14	24.07	22.78
06/08/05	10	450	32.26	22.00	23.62	25.51	23.78
06/13/05	11	790	23.70	22.00	23.62	25.51	23.78
06/14/05	12	5500	19.97	22.00	23.62	25.51	23.78
06/15/05	13	4800	22.98	22.00	24.49	26.35	24.37
06/21/05	14	930	31.59	22.00	22.40	24.32	22.96
06/28/05	15	1800	23.88	22.00	24.14	26.01	24.13
06/29/05	16	340	29.01	23.00	24.14	26.01	24.13
07/06/05	17	570	28.06	21.00	24.14	26.01	24.13
07/11/05	18	380	25.88	26.00	24.49	26.35	24.37
07/12/05	19	1610	26.16	26.00	25.36	27.19	24.96
07/13/05	20	3100	26.98	27.80	24.49	26.35	24.37
07/18/05	21	4100	27.54	28.20	24.49	26.35	24.37
07/19/05	22	1600	26.76	28.60	24.93	26.77	24.66
07/20/05	23	3500	29.31	27.60	24.75	26.60	24.54
07/25/05	24	480	24.83	28.50	24.75	26.60	24.54
07/26/05	25	960	25.18	30.10	24.75	26.60	24.54
08/10/05	26	740	29.00	25.50	25.65	27.70	25.85
08/16/05	27	1300	27.35	25.50	27.15	28.65	27.35
08/17/05	28	720	27.78	25.80	25.85	27.85	25.90
08/22/05	29	690	26.23	24.00	26.05	27.35	26.80
08/23/05	30	430	29.35	23.50	25.15	26.10	25.40
08/24/05	31	350	24.05	22.60	23.75	25.90	24.60
08/29/05	32	1300	26.46	25.80	23.30	24.90	23.10
09/06/05	33	440	27.19	21.30	23.80	25.80	23.90
09/12/05	34	500	27.08	22.00	23.20	24.60	22.80
09/19/05	35	500	18.60	24.20	24.35	24.95	24.60
09/20/05	36	470	27.00	22.60	24.95	26.05	25.10
10/03/05	37	600	27.13	21.50	21.35	22.00	21.30
10/04/05	38	350	24.80	21.00	25.00	23.40	22.65
10/17/05	39	570	23.00	21.50	19.25	20.30	18.10
10/24/05	40	570	20.50	18.00	18.80	20.15	17.95
10/31/05	41	340	25.00	20.50	15.85	18.30	16.30
11/01/05	42	340	24.63	19.00	17.80	18.45	17.10
11/07/05	43	350	27.60	21.00	19.45	19.50	18.85
11/08/05	44	410	26.44	22.50	17.85	19.05	17.80
11/14/05	45	430	24.94	23.70	19.35	19.85	18.95
11/15/05	46	500	25.00	25.50	19.50	20.40	19.30
11/21/05	47	600	23.35	19.00	16.60	17.10	15.15
11/28/05	48	710	24.17	27.80	18.80	18.85	18.05
11/29/05	49	1100	25.68	26.10	20.95	21.80	21.15
12/05/05	50	1610	25.00	24.30	12.85	14.85	12.75
12/13/05	51	480	25.82	23.60	11.30	13.30	14.00
02/22/06	52	1170	28.40	24.80	12.50	14.20	15.25
02/27/06	53	1610	25.51	24.80	11.80	12.55	12.70
03/01/06	54	2132	28.42	26.20	15.50	17.40	19.80
03/06/06	55	5486	24.95	26.10	13.65	16.85	19.20
03/13/06	56	2756	26.56	27.80	17.75	22.35	22.75
03/15/06	57	1638	27.59	26.40	15.95	18.45	19.60
03/20/06	58	1947	24.76	26.80	13.85	16.85	17.25
03/22/06	59	1507	22.62	27.10	15.15	17.90	20.35
03/29/06	60	2958	23.80	25.60	18.00	19.70	22.05
04/11/06	61	5018	26.21	26.20	17.45	18.55	18.00
04/12/06	62	2522	19.57	25.70	19.45	20.40	20.55
04/24/06	63	1066	24.32	24.70	19.45	20.35	19.25
04/26/06	64	3874	23.62	22.60	17.05	18.25	16.90
06/06/06	65	979	25.87	25.00	22.85	24.40	23.30
06/07/06	66	495	26.80	25.50	23.20	26.10	23.60
06/12/06	67	539	26.38	23.00	19.95	22.45	20.40
06/13/06	68	638	24.15	23.00	22.15	25.00	22.45
06/27/06	69	583	35.71	24.00	24.40	27.85	25.00
06/28/06	70	594	29.34	23.00	24.20	25.05	24.25
07/05/06	71	1274	24.80	21.50	25.50	28.00	25.50
07/10/06	72	867	21.75	27.00	23.95	25.95	23.90
07/11/06	73	5018	24.64	27.50	25.35	28.00	26.05
07/17/06	74	6426	23.90	32.00	26.35	29.70	27.00
07/18/06	75	825	22.90	27.00	28.35	30.35	27.70
07/19/06	76	2990	24.75	27.00	26.25	28.65	26.60
07/25/06	77	8694	27.00	26.00	25.00	28.25	25.15

observation #	GT PH	BS PH	DAF PH	GT ORP	BS ORP	DAF ORP	GT JRM
1	5.93	6.44	6.67	-110.74	-154.52	-212.20	0.71
2	5.88	6.39	6.59	-110.74	-154.52	-212.20	0.82
3	5.62	6.38	6.37	-110.74	-154.52	-212.20	0.14
4	5.65	6.29	6.39	-110.74	-154.52	-212.20	0.38
5	5.85	6.39	6.39	-110.74	-154.52	-212.20	0.37
6	6.00	6.47	6.56	-110.74	-154.52	-212.20	0.32
7	5.83	6.36	6.15	-110.74	-154.52	-212.20	0.67
8	5.79	6.26	6.51	-110.74	-154.52	-212.20	1.04
9	5.73	6.27	6.45	-110.74	-154.52	-212.20	0.68
10	5.71	6.27	6.48	-110.74	-154.52	-212.20	1.13
11	5.77	6.46	6.67	-110.74	-154.52	-212.20	0.92
12	5.72	6.44	6.52	-110.74	-154.52	-212.20	1.20
13	5.72	6.34	6.52	-110.74	-154.52	-212.20	0.92
14	5.68	6.36	6.73	-110.74	-154.52	-212.20	1.14
15	4.88	5.40	5.93	-110.74	-154.52	-212.20	2.97
16	4.84	5.21	5.86	-110.74	-154.52	-212.20	2.42
17	4.82	5.64	6.01	-110.74	-154.52	-212.20	0.62
18	5.23	5.83	5.79	-110.74	-154.52	-212.20	0.58
19	5.15	5.62	5.67	-110.74	-154.52	-212.20	0.62
20	4.95	5.57	5.53	-110.74	-154.52	-212.20	0.88
21	5.62	6.24	6.60	-110.74	-154.52	-212.20	4.08
22	5.41	6.17	6.37	-110.74	-154.52	-212.20	1.67
23	5.47	6.67	6.58	-110.74	-154.52	-212.20	0.65
24	5.26	5.76	6.09	-110.74	-154.52	-212.20	0.50
25	5.08	5.85	6.11	-110.74	-154.52	-212.20	0.80
26	5.68	6.17	6.52	-110.74	-154.52	-212.20	0.92
27	5.59	6.60	6.70	-267.00	-215.90	-249.85	6.18
28	5.38	6.36	6.71	-268.40	-207.40	-227.45	12.43
29	5.72	6.33	6.78	-160.20	-182.65	-237.40	1.15
30	5.60	6.33	6.85	-155.40	-194.70	-246.80	0.78
31	5.52	6.31	7.00	-161.35	-193.35	-251.00	1.40
32	5.60	6.14	6.87	-135.80	-169.10	-243.25	1.16
33	5.75	6.07	6.68	-115.25	-161.40	-214.05	1.19
34	5.67	5.93	6.93	-112.10	-128.40	-228.30	0.71
35	5.47	6.26	6.71	-119.30	-150.45	-202.10	1.18
36	5.68	6.22	6.66	-128.90	-148.60	-195.35	0.95
37	5.72	6.16	6.67	-108.45	-140.80	-200.10	1.10
38	5.80	6.40	6.82	-143.05	-172.05	-229.45	0.28
39	5.94	6.23	6.85	-132.25	-180.25	-272.25	0.44
40	5.49	6.19	6.91	-113.60	-161.60	-262.20	0.60
41	5.69	6.23	6.91	-93.85	-164.70	-215.50	0.89
42	6.05	6.62	7.22	-91.10	-168.65	-224.45	1.30
43	5.77	6.32	6.59	-43.75	-90.90	-164.50	1.21
44	5.52	6.16	6.91	-110.74	-154.52	-212.20	0.51
45	5.75	6.50	7.21	-110.74	-154.52	-212.20	0.91
46	5.74	6.27	6.96	-110.74	-154.52	-212.20	0.69
47	6.13	6.23	6.73	-110.74	-154.52	-212.20	0.48
48	5.81	6.37	6.77	-110.74	-154.52	-212.20	0.67
49	5.63	5.95	6.37	-110.74	-154.52	-212.20	0.59
50	6.16	6.31	6.56	-110.74	-154.52	-212.20	0.31
51	5.79	5.94	6.18	-105.30	-163.95	-232.75	0.39
52	6.10	6.49	7.03	-66.15	-120.90	-159.35	0.34
53	6.11	6.22	6.78	-78.10	-128.10	-197.60	0.40
54	6.08	6.62	6.88	-31.30	-110.35	-164.00	0.36
55	6.25	6.45	6.81	-14.25	-74.50	-151.15	0.52
56	6.04	6.67	6.80	-15.30	-62.20	-131.25	1.67
57	5.75	6.51	6.89	-110.74	-154.52	-212.20	1.62
58	5.92	6.50	6.96	-61.45	-105.95	-166.30	0.83
59	5.91	6.42	6.79	-28.65	-79.05	-139.00	0.92
60	5.95	6.56	6.62	-98.95	-147.20	-158.00	0.67
61	6.17	6.68	7.19	15.15	-89.60	-162.00	0.75
62	5.97	6.56	7.14	-86.70	-125.70	-207.55	0.96
63	6.09	6.48	6.92	-122.85	-160.75	-214.30	0.37
64	6.02	6.59	7.01	-37.30	-118.00	-179.30	1.01
65	5.76	6.21	6.48	-79.20	-113.00	-139.50	0.93
66	5.91	6.26	6.57	-73.30	-120.80	-161.30	1.38
67	5.98	6.42	6.62	-137.40	-170.80	-232.85	0.74
68	5.73	6.37	6.68	-124.80	-176.70	-220.80	0.65
69	5.81	6.18	6.64	-125.80	-196.95	-260.90	0.26
70	5.89	6.21	6.61	-107.60	-183.70	-238.00	0.18
71	5.71	6.24	6.79	-141.25	-195.45	-271.35	0.66
72	5.80	6.21	6.68	-113.30	-180.05	-235.70	0.73
73	5.90	6.29	6.76	-145.10	-195.55	-252.40	0.91
74	5.96	6.35	6.75	-122.55	-168.15	-231.65	1.17
75	6.00	6.35	6.68	-140.00	-174.55	-230.15	0.46
76	5.39	5.98	6.36	-200.25	-253.15	-281.70	0.82
77	5.97	6.36	6.45	-170.30	-198.40	-241.85	0.63

observation #	BS JRM	DAF JRM	# centrif running	BD east	BD west	GT% in blend	TWAS (gal)
1	0.59	0.79	7.00	4.70	2.60	54.95	756400
2	0.69	0.23	6.00	3.00	3.60	68.66	385100
3	0.18	0.24	5.00	3.20	2.30	49.91	1002000
4	0.61	2.48	5.00	3.20	2.20	53.97	912000
5	0.30	0.92	6.00	4.00	1.30	45.02	1129902
6	0.44	2.12	6.00	2.90	1.30	90.28	95598
7	0.20	0.98	3.00	2.30	1.70	53.78	881103
8	0.87	1.68	6.00	1.90	1.60	49.36	881659
9	0.54	0.77	6.00	2.20	1.90	58.94	741100
10	0.31	1.02	5.00	2.30	1.50	64.51	699400
11	0.35	0.74	5.00	2.10	1.80	52.31	967800
12	0.48	1.62	4.00	3.20	1.70	58.65	1050900
13	0.39	0.71	7.00	1.90	1.40	59.14	941900
14	0.33	0.86	6.00	2.20	2.30	66.22	968000
15	0.18	0.27	5.00	3.00	1.80	53.78	881103
16	0.99	0.27	4.00	3.00	1.80	53.78	881103
17	0.16	0.15	6.00	2.90	2.10	66.71	1195800
18	0.23	0.42	5.00	1.80	2.00	67.78	1099100
19	0.46	0.44	6.00	2.10	2.10	52.56	846900
20	0.27	0.71	5.00	2.10	2.50	63.89	772400
21	0.24	0.25	7.00	1.90	2.60	58.38	753000
22	0.88	1.01	7.00	2.50	2.60	70.08	659000
23	0.22	0.48	5.00	2.30	1.70	98.09	617000
24	0.18	0.24	6.00	2.20	2.30	55.50	986220
25	0.29	0.18	5.00	1.90	2.10	64.41	931910
26	0.23	0.52	6.00	2.10	1.80	95.84	686000
27	0.37	1.50	6.00	1.80	1.70	68.43	723000
28	0.35	0.36	6.00	1.80	1.20	53.78	881103
29	0.40	0.66	6.00	2.50	1.00	57.42	1177000
30	0.56	0.61	6.00	3.10	1.20	72.77	1013000
31	0.79	0.56	6.00	3.10	1.50	95.22	95859
32	0.38	0.58	6.00	2.00	1.20	56.37	913500
33	0.36	0.45	6.00	1.70	1.50	53.78	881103
34	0.20	0.76	3.00	2.40	1.80	53.78	881103
35	0.20	0.41	6.00	1.70	1.90	74.42	565000
36	0.26	0.87	6.00	1.60	2.00	18.86	715000
37	0.86	0.74	4.00	1.80	3.10	53.78	881103
38	0.30	0.34	4.00	2.20	2.10	53.78	881103
39	0.35	0.19	10.00	3.60	5.20	100.00	881103
40	0.60	0.54	5.00	3.60	2.40	57.42	779900
41	1.07	0.27	10.00	3.50	3.30	53.78	881103
42	0.40	0.24	4.00	3.50	2.50	53.78	881103
43	0.32	1.32	5.00	3.30	3.30	67.03	898870
44	0.39	0.27	9.00	4.10	2.70	63.99	913210
45	0.51	0.42	8.00	1.80	2.30	11.14	970000
46	0.70	0.85	8.00	1.80	2.30	13.10	887000
47	0.45	0.32	8.00	1.90	1.90	12.67	695400
48	0.49	0.30	8.00	1.60	1.90	53.78	633780
49	0.65	0.43	6.00	1.60	2.90	53.78	659160
50	0.30	0.17	10.00	2.00	2.10	10.41	799300
51	0.42	0.34	9.00	2.20	2.40	10.90	719000
52	0.50	0.62	4.00	3.20	1.30	54.79	636000
53	1.03	1.70	7.00	2.00	3.50	49.93	658250
54	0.81	2.93	7.00	2.20	3.90	11.39	7814000
55	1.23	1.09	7.00	3.50	2.50	11.26	668000
56	0.33	0.81	9.00	2.90	2.60	54.46	598800
57	0.49	0.95	10.00	4.20	2.10	57.98	821000
58	1.33	0.70	10.00	4.30	2.30	37.52	977620
59	0.47	0.58	10.00	3.50	2.40	48.30	887400
60	1.08	1.50	7.00	2.20	1.00	68.58	657420
61	1.11	1.30	4.00	3.30	2.10	53.78	632600
62	1.25	1.15	9.00	3.30	3.20	11.05	677000
63	1.28	0.71	11.00	2.30	3.00	10.86	930030
64	0.35	0.82	10.00	2.20	1.40	61.07	861340
65	0.84	0.40	9.00	1.80	1.90	63.65	649900
66	0.53	0.87	6.00	2.00	2.30	58.83	619000
67	0.92	0.65	8.00	2.20	2.60	53.38	677900
68	0.78	1.06	10.00	2.20	1.60	48.54	837200
69	0.45	0.54	7.00	2.10	1.80	65.85	568000
70	0.62	0.24	8.00	2.50	2.30	66.98	795100
71	0.16	0.33	7.00	2.60	5.00	67.24	627350
72	0.15	0.14	9.00	3.30	5.00	93.36	101021
73	0.60	0.26	5.00	3.00	3.30	62.16	804030
74	0.33	0.38	8.00	2.30	3.50	8.16	9559220
75	0.27	0.23	9.00	1.80	4.80	50.65	955460
76	0.64	0.35	8.00	1.60	1.80	57.01	867000
77	0.53	0.21	6.00	1.40	1.80	13.31	576000

observation #	TPS (gal)	Blend ratio	RAS west (mg/l)	RAS east (mg/l)	BD east d-1	BD west d-1	GT% in blend d-1
1	558900	0.42	6080	9700	4.40	2.90	49.14
2	634791	0.62	6180	8080	4.70	2.60	54.95
3	923000	0.48	5480	7180	3.20	2.50	57.90
4	928000	0.50	4580	19560	3.20	2.30	49.91
5	1066080	0.49	4700	10000	3.40	1.30	61.29
6	1053291	0.92	5650	9150	4.00	1.30	45.02
7	958020	0.52	6100	7700	3.00	1.70	62.53
8	995440	0.53	7200	4900	1.90	1.60	59.30
9	816800	0.52	6850	6000	1.90	1.60	49.36
10	970100	0.58	8250	10350	2.20	1.90	58.94
11	946200	0.49	5750	11900	2.50	1.80	60.35
12	988800	0.48	4800	11300	2.10	1.80	52.31
13	1073800	0.53	5750	9950	3.20	1.70	58.65
14	966000	0.50	5400	9600	1.90	1.90	67.54
15	958020	0.52	6150	7300	3.10	2.10	53.78
16	958020	0.52	6050	11550	3.00	1.80	53.78
17	718500	0.38	7250	13450	4.20	3.30	53.78
18	829000	0.43	4900	8750	1.90	2.30	44.33
19	832700	0.50	6600	8750	1.80	2.00	67.78
20	827800	0.52	6750	6550	2.10	2.10	52.56
21	841000	0.53	3300	9000	1.70	2.60	60.59
22	846000	0.56	4500	8750	1.90	2.60	58.38
23	992000	0.62	4850	9400	2.50	2.60	70.08
24	962560	0.49	4400	5800	2.10	2.10	58.41
25	997730	0.52	5300	3400	2.20	2.30	55.50
26	9333000	0.93	6850	6850	2.20	1.60	76.93
27	934000	0.56	5080	4400	1.90	1.40	69.71
28	958020	0.52	5960	1860	1.80	1.70	68.43
29	968000	0.45	4840	2700	2.90	1.10	64.15
30	951000	0.48	3380	2940	2.50	1.00	57.42
31	943750	0.91	4440	5740	3.10	1.20	72.77
32	1076100	0.54	1600	3580	2.40	1.80	66.22
33	958020	0.52	7550	4250	1.90	2.20	53.78
34	958020	0.52	5600	5750	2.00	1.30	53.78
35	1077000	0.66	6850	6600	2.30	2.00	58.22
36	113200	0.14	5200	3650	1.70	1.90	74.42
37	958020	0.52	11000	5000	2.10	3.10	71.37
38	958020	0.52	3900	1350	1.80	3.10	53.78
39	125200	1.00	5355	7250	3.30	3.50	66.66
40	761500	0.49	5355	5100	3.60	2.30	60.53
41	958020	0.52	7650	9050	2.90	3.30	53.78
42	958020	0.52	3750	4500	3.50	3.30	53.78
43	1149780	0.56	5355	5100	2.20	3.70	53.78
44	1167190	0.56	4500	7550	3.30	3.30	67.03
45	123100	0.11	4050	4450	1.90	2.20	59.39
46	116600	0.12	5150	3500	1.80	2.40	11.14
47	129980	0.16	4600	4050	1.50	2.20	12.05
48	128544	0.17	4700	8200	2.10	1.70	10.61
49	132200	0.17	5200	7950	1.60	1.90	53.78
50	135176	0.14	4580	2860	2.20	2.10	48.13
51	157300	0.18	3680	5280	1.80	2.40	11.59
52	1165000	0.65	5500	4550	1.50	1.30	53.78
53	1135540	0.63	5355	2750	1.80	2.70	59.06
54	984600	0.11	7100	6350	2.20	4.60	65.21
55	118900	0.15	4550	10600	3.30	3.10	47.84
56	831400	0.58	7250	9000	2.20	3.90	10.90
57	1178000	0.59	6350	13350	4.20	3.50	53.78
58	1002740	0.51	8000	9750	3.50	3.00	42.19
59	1042900	0.54	5100	9450	4.30	2.10	53.78
60	1117250	0.63	5355	5355	2.30	1.00	50.26
61	1085800	0.63	5440	6680	2.60	2.20	52.39
62	113100	0.14	4980	6220	3.30	2.10	53.78
63	124218	0.12	3360	6180	2.60	3.70	53.78
64	1311540	0.60	3840	5660	2.40	2.50	61.50
65	1196900	0.65	1950	4400	1.60	1.80	50.74
66	1162000	0.65	1800	4850	1.80	1.90	63.65
67	1139000	0.63	2450	8800	1.30	1.00	63.96
68	1166100	0.58	2600	6800	2.20	2.60	53.38
69	1198000	0.68	1300	11000	2.30	3.10	71.79
70	1280851	0.62	2900	9750	2.10	1.80	65.85
71	1013800	0.62	3450	6800	2.00	3.80	58.16
72	1136950	0.92	5650	8800	3.50	4.60	62.51
73	1026170	0.56	5250	5950	3.30	5.00	93.36
74	1060159	0.10	6750	12400	2.00	4.30	55.48
75	1064091	0.53	5150	8950	2.30	3.50	8.16
76	1114000	0.56	3750	5850	1.80	4.80	50.65
77	127200	0.18	3250	10800	1.60	1.40	100.00

observation #	TWAS (gal) d-1	TPS (gal) d-1	RAS west (mg/l) d-1	RAS east (mg/l) d-1
1	919600	749400	5960	3900
2	756400	558900	6080	9700
3	797200	881440	5340	10600
4	1002000	923000	5480	7180
5	698860	963990	5650	8150
6	1129902	1066080	4700	10000
7	550090	980910	4000	8150
8	704851	711460	7800	5650
9	881659	995440	7200	4900
10	741100	816800	6850	6000
11	1121800	985200	4600	11400
12	967800	946200	5750	11900
13	1050900	988800	4800	11300
14	912000	940000	5550	8650
15	881103	958020	7050	1300
16	881103	958020	6150	7300
17	881103	958020	4900	11800
18	997500	880200	5150	8600
19	1099100	829000	4900	8750
20	846900	832700	6600	8750
21	875000	915000	6750	11300
22	753000	841000	3300	9000
23	659000	846000	4500	8750
24	933490	962310	7350	5900
25	986220	962560	4400	5800
26	488000	900000	3650	6150
27	543000	838000	5780	5760
28	723000	934000	5080	4400
29	1179120	921160	5420	3560
30	1177000	968000	4840	2700
31	1013000	951000	3380	2940
32	873400	1101300	2580	6780
33	881103	958020	8100	3800
34	881103	958020	6200	4350
35	844100	1251700	6400	4700
36	565000	1077000	6850	6600
37	616000	1243000	9050	1700
38	881103	958020	11000	5000
39	1074580	1235770	5355	4850
40	962000	999000	4700	5150
41	881103	958020	5355	6350
42	881103	958020	7650	9050
43	881103	958020	3600	8900
44	898870	1149780	5355	5100
45	977879	1291310	5400	4350
46	970000	123100	4050	4450
47	699000	129500	5600	3400
48	888700	132520	4400	3850
49	633780	128544	4700	8200
50	102397	136851	5720	4080
51	631000	134500	5280	5520
52	881103	958020	4600	5600
53	444100	1208400	5355	3750
54	307970	926200	1860	1150
55	840000	1226000	4350	9200
56	862000	100700	7800	13400
57	881103	958020	6600	13200
58	396000	363200	6550	13250
59	1020990	1060800	5355	7900
60	630400	927770	5750	7100
61	734880	1076330	7480	9300
62	632600	1085800	5440	6680
63	881103	958020	7160	6440
64	604730	943080	3100	2220
65	867250	1236300	5355	5900
66	649900	1196900	1950	4400
67	632050	1186259	2500	6400
68	677900	1139000	2450	8800
69	931000	1249000	2950	5850
70	568000	1198000	1300	11000
71	726400	1125590	6250	10900
72	758545	1031771	6400	5950
73	101021	1136950	5650	8800
74	831770	1061180	6300	15500
75	9559220	1060159	6750	12400
76	955460	1064091	5150	8950
77	676000	123700	1600	6500

178

Bibliography

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*, John Wiley & Sons, Inc., New York.

Agresti A. (2002). *Categorical Data Analysis, second edition*, John Wiley&Sons, Inc., Hoboken, New Jersey.

Albright, S. C. (2003). *Learning Statistics with StatTools,* Palisade Corporation, Inc., NY.

Altman, E. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankrupcy." *J. Finance*, 23(4), 589-609.

Arispe, S., Kim, H., McConnell, L.L., Ramirez, M., and Torrents, A. (2005a). "Analysis of Sulfides, Amines, Mercapans, and Volatile Fatty Acids of Wastewater and Sludge Using Solid Phase Microextraction under Static Condition." *J. Chromatogr*, (submitted).

American Society for Testing and Materials (ASTM), (Reapproved 1997). "Standard Practice for Defining and Calculation Individual and Group Sensory Thresholds from Forced-Choice Data Sets of Intermediate Size." *E1432-91,* approved august 15, 1991, and published October 1991

American Society for Testing and Materials (ASTM) (Reapproved 1988). "Standard Practice for Suprathreshold Odor Intensity Measurement." *E544-75*, Philadelphia, PA: 1988.

American Society for Testing and Materials (ASTM) (1991). "Standard Practice for Determination of Odor and Taste Threshold by a Forced-Choice Ascending Concentration Series Method of Limits." *E679-91*, Philadelphia, PA: 1991.

Bardos, M. (1998). "Detecting the Risk of Company Failure at the Banque de France." *J. Banking Finance*, 22(10), 1405-1419.

Borgonovo, E., Apostolakis, G.E., Tarantola, S., and Saltelli, A. (2003). "Comparison of Global Sensitivity Analysis Techniques and Important measures in PSA." *Reliability Engineering&System Safety*, 76, 175-185.

Bowker, R. P.G., Smith, J.M., and Webster, N.A. (1989). *Odor and Corrosion Control in Sanitary Sewerage Systems and Treatment Plants*, SciTech Publishing, Incorporated.

Cariboni, J., Gatelli, D., and Saltelli, A. (2007) "The Role of Sensitivity Analysis in Ecological Modelling." *Ecological Modeling*, 203, 167-182.

Clemen, R. T., and Reilly, T. (2001). *Making hard decisions with DecisionTools*, Duxbury, CA.

Davis, J. (2008). "Fact Sheet # 25: Making Decision about Application rates." CAFO Fact Sheet Series. Retrieved February 17, 2008, from http://www.lpes.org/cafo/25FS_Rates.pdf

DCWASA, (2008). "DCWASA Biosolids Management Program." Retrieved January 15, 2008, from http://www.dcwasa.com/education/biosolids_brochure.pdf

Deakin, E. (1972). "A Discriminant Analysis of Predictors of Business Failure." *J. Accounting Res.*, Spring1972, 167-179.

Desai, V. S., Crook, J. N., and Overstreet, G. A. (1996). "A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment." *European J. Oper Res.*, 95, 24-37.

Devore, J. L. (1995). *Probability and Statistics for Engineering and the Sciences,* Duxbury Press 4th edition, Belmont, CA.

Dijk, J.V.and Pellenbarg, P.H. (2000). "Firm Relocation Decisions in the Netherlands: An Ordered Logit Approach." *Regional Science*, 79(2), 191-219

Epstein, E. (2002). *Land Application of Sewage Sludge and Biosolids*, Lewis Publishers, Boca Raton, Fl.

Evanylo, G.K. (1999). "Agricultural Land Application of Biosolids in Virginia: Managing Biosolids for Agricultural Use." *Virginia Cooperative Extension*, Publication number 452-303.

Fisher, R. A. (1936). "The Use of Multiple Measurements in Taxonomic Problems." *Ann. Eugen*, 7, 179-188.

Frenchen, F.B. (1988). "Odour Emissions and Odour Control at Wastewater Treatment Plants in West Germany." *Water Sci. Technol.*, 20, 261-266.

Gabriel, S.A., Vilalai, S., Aprise, S., Kim, H., McConnell, L.L., Torrents, A., Peot, C., and Ramirez, M. (2005). "Prediction of Dimethyl Disulfide Levels from Biosolids Using Statistical Modeling." *J. Environ. Sci. Health*, 40, 2009-2025.

Gabriel, S.A., Vilalai, S., Peot, C., and Ramirez, M. (2006). "Statistical Modeling to Forecasting Odor Levels of Biosolids Applied to Reuse Sites." *J. Environ. Eng.*, May 2006, 479-485.

Gostelow, P., Parsons, S.A., and Stuetz, R.M. (2001). "Odour Measurements for Sewage Treatment Works." *Wat. Res.*, 35(3), 579-597.

Greenberg, M.R., Carey, G.W., Zobler, L., and Hordon, R.M. (1973). "A statistical Dissolved Oxygen Model for a Free-Flowing River System." *J. of American Statistical Association*, 68(342), 279-283.

Greene, W.H. (1997). *Econometric Analysis, 3rd edition,* Uppper Saddle River, New Jersy: Prentice Hall.

Hamel, K.C., Walters, L., Sulerud, C., McGinley, M.A. (2004). "Land Application Odor Control Case Study." *Waster environment federation residuals and biosolids management conference*, Salt Lake City, UT

Hamilton, D.W., and Carlson, J.D. (2008). "Movement of Odors Off-Farm." Oklahoma Cooperative Extension Fact Sheet.

Harrel, F. E., and Lee, K. L. (1985). *A Comparison of the Discrimination of Discriminant Analysis and Logistic Regression under Multivariate Normality.* Biostatistics: Statistics in Biomedical, Public Health, and Environmental Science, Amsterdam: Elsevier Science Publisher B.V., 333-343.

Janpengpen, A., Baecher, G., Peot, C., and Ramirez, M. (2008). "An Evaluation of Dewatered Sludge Percent Solids at District of Columbia Water and Sewer Authority (DCWASA)." *13th European Biosolids Conference*, Mancherter, UK.

Joanes, D. N. (1993). "Rejecting Inference Applied to Logistic Regression for Credit Scoring." *IMA J. Math. Appl. Bus. Ind.*, 5, 35-43.

Johnston, A., Smith, A. (2001). "Environmental performance." *Water Environ. Manage.*, 15(1), 40-45.

Kay, O. W., Warde, A., and Martens, L. (2000). "Social Differentiation and the Market for Eating out in the UK." *Int. J. Hosp. Manage.*, 19(2), 173-190.

Kim, H., Murthy, S., McConnell, L.L., Peot, C., Ramirez, M., and Strawn, M. (2001). "Characterization of Wastewater and Solids Odors Using Solid Phase Microextraction at a Large Wastewater Treatment Plant." *IWA Biosolids Conference*, Acapulco, Mexico.

Kim, H., Murthy, S., McConnell, L.L., Peot, C., Ramirez, M., and Strawn, M. (2002). "Characterization of Wastewater and Solids Odor Using Solids Phase Microextraction at a Large Wastewater Treatment Plant." *Water Sci. Technol.* 46 (10), 9-16.

Kim, H., Murthy, S.N., McConnell, L., Peot, C., Ramirez, M., and Strawn, M. (2003). "Examination of Mechanisms for Odor Compound Generation during Lime Stabilization." *Water Environ. Res.*, 75, 121-125.

Kim, J. C., Kim, D. H., Kim, J. J., Ye, J. S., and Lee, H. S. (2000). "Segmenting the Korean Housing Market Using Multiple Discriminant Analysis." *Construct. Manage. Econ.*, 18, 45-54.

Kramer, B. (1996). "An Ordered Logit Model for the Evaluation of Dutch Non-Life Insurance Companies." *De Economist* , 144(1). 79-91.

Kuthner, M., Nachtsheim, C.J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Model, 5th edition,* McGraw-Hill/Irwin

Laitinen, T., and Kankaanpaa, M. (1999). "Comparative Analysis of Failure Prediction Methods: The Finnish Case." *European Account. Rev.*, 8(1), 67-92.

Laitinen, T., and Laitinen, E. K. (2000). "Bankruptcy Prediction: Application of the Taylor's Expansion in Logistic Regression." *Int. Rev. of Finance. Analysis*, 9, 327-349.

Lamon, E.C. III. (1995). "A Regression Model for the Prediction of Chlorophyll *a* in LakeOkeechobee, Florida." *Lake and Reservoir Management*, 11(4), 283-290.

Lamon, E.C. III, Reckhow, K.H., and Havens, K.E. (1996). "Using Generalized Additive Models for Prediction of Chlorophyll a in Lake Okeechobee, Florida." *Lakes & Reservoirs: Research and Management.* 2(1-2): 37-46.

Lamon, E.C. III and Clyde, M.A. (2000). "Accounting for Model Uncertainty in Prediction of Chlorophyll *a* in Lake Okeechobee." *Journal of Agricultural, Biological, and Environmental Statistics*. 5(3), 297-322.

Lay, A. M. and Mcginley, C.M. (2004). "A Nasal Chemosensory Performance Test for Odor Inspectors." *Water Environment Federation. Odors and Air Emissions 2004*, Bellevue, WA: 18-21 April 2004

Lee, H., Jo, H., and Han, I. (1997). "Bankruptcy Prediction Using Case-based Reasoning, Neural Networks, and Discriminant Analysis." *Expert System Appl.*, 13(97-108).

Lee, T.-S., Chiu, C.-C., Chou, Y.-C., and Lu, C.-J. (2006). "Mining the Customer Credit Using Classification and Regression Tree and Multivariate Adaptive Regression Splines." *Computational Statistics & Data Analysis*, 50(4), 1113-1130.

Lu, M. (1999). "Determinant of Residential Satisfaction: Ordered Logit vs Regression Model." *Growth and Change*, 30(2), 264-287

Luca, S. J. d., Idle, C. N., and Chao, A. C. (1996). "Quality Improvement of Biosolids by Ferrate(VI) Oxidation of Offensive Odour Compounds." *Water Sci. Technol.*, 33(3), 119-130.

Maddala, G. S. (1977). *Econometrics*, McGraw-Hill, Inc, New York.

Martell, T. F., and Fitts, R. L. (1981). "A Quadratic Discriminant Analysis of Bank Credit Card User Characteristics." *J. Econ Bus.*, 33, 153-159.

Maryland Environmental Service (MES). (2005). "Standard Operating Procedure- Odor Monitoring."

McGinley, C.M., and McGinley, M.A. (2002). "Odor Testing Biosolids for Decision Making." *WEF Specialty Conference*, Austin, TX

McGinley, M. A., McGinley, C. M., 2000, "Olfactomatics: Applies Mathematics for Odor Testing". *WEF Odor/VOC 2000 Specialty Conference*, Cincinnati, OH

McGinley, M.A., and McGinley, C.M. (2005). "Measuring Composting Odors for Decision Making." *U.S.Composting Council 2005 Annual Conference*, San Antonio, TX

Mckelvey, R.D, and Zavoina, W. (1975). "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *J. Math. Sociology*, 4:103-120

Morris, G., Firth, C. (2004). "From sewage to salmon." *CIWEM National Conference*, Aqua Enviro, wake field, UK.

Murthy, S. (2003). "Effect of Dewatering and Digestion Practices on Volatile Sulfur Compound Production." *WEF/AWWA/CWEA Joint Residuals and Biosolids Management Conference*, Baltimore, MD.

Murthy, S., Sadick, T., Bailey, W., Peot, C., Strawn, M., and Tolbert, D. (2001). "Odor Mitigation from Lime Stabilized Biosolids." *WEF Biosolids Conference 2001*, San Diego, CA.

North, J. M, (2003). "An Evaluation of Methods for Quantifying Lime Incorporation into Mechanically Dewatered Sludge." MS.thesis, Department of Civil and Environmental Engineering, University of Maryland, College Park, Maryland.

Palisade Corporation. (2001). *Guide to Using @Risk: Risk Analysis and Simulation Add-In for Microsoft Excel*, Palisade Corporation, Inc, NY.

Public Law 100-688. (1988). "The Ocean Dumping Ban Act of 1988."

Reynolds, T.D., Paul, A.R. (1996). *Unit Operations and Processes in Environmental Engineering,* PWS Publishing Co., Boston MA.

Richert, A. K., Cho, C. C., and Wagner, G. M. (1983). "An Examination of the Conceptual Issues Involved in Developing Credit-Scoring Models." *J.Bus. Econ. Statist.*, 1, 101-114.

Robson, P.J.A. and Bennett, R.J. (2000). "The Use and Impact of Business Advice by SMEs in Britain: an Empirical Assessment Using Logit and Ordered Logit Models." *Applied Economics*, 32, 1675-1688

Rynk, R., and Goldstein, N. (2003). "Reducing Odor Impacts at Land Application Sites." *Bisolids Management Trends*(April 2003), 54-58.

Sahakij, P. (2007). "Multiobjective Optimization Models for Distributing Biosolids to Reuse Fields." Ph.D. Thesis, Department of Civil and Environmental Engineering, University of Maryland, College Park, Maryland.

Sarkar, U., Longhurst, P. J., and Hobbs, S. E. (2003). "Community Modeling: a tool for correlating estimates of exposure with perception of odour from municipal solid waste (MSW) landfills." *J. Environ. Manage..*, 68, 133-140.

Saltelli, A., Chan, K., and Scott, E.M. (2000). *Sensitivity Analysis*, John Wiley & Sons.

Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F. (2006). "Sensitivity Analysis Practices: Strategies for Model-Based Inference." *Reliability Engineering&System Safety*, 91, 1109-1125.

Saveyn, H., Curvers, D., Thas, O., and Van der Meeren, P. (2008). "Optimization of Sewage Sludge Conditioning and Pressure Dewatering by Statistical Modelling" *Water Res*, 43(4/5), 1061-1074.

Schifman, S.S., McLaughlin, B., Katul, G.G., and Nagle, H.T. (2005). "Eulerian-Lagrangian Model for Predicting Odor Dispersion Using Instrumental and Human Measurements." *Sensors and Actuator B:Chemical*, 106(1), 122-127

Sekyiamah, K. H. (2004). "Identification of Seasonal Variations in Volatile Sulfur Compound Production and Release in the Secondary Treatment System of a Wastewater Treatment Plant." MS thesis, Department of Civil and Environmental Engineering, University of Maryland, College Park, Maryland.

Shmueli, G., Patel, N. R., and Bruce, P. C. (2007). *Data Mining for Business Intelligence*, John Wiley & Sons, Inc., New Jersey.

SKC (2006). "The SKC Vac-U-Chamber", retrieved on January, 20, 2006, from http://www.skcinc.com/prod/231-939.asp

Subramanian, R., Novak, J. T., Murthy, S., Glindemann, D., and North, J. (2005). "Investigating the Role of Process Conditions in Wastewater Sludge Odor Generation." *WEFTEC 2005 Annual Conference*, Washington, DC.

Suh, E. H., Noh, K. C., and Suh, C. K. (1999). "Customer List Segmention Using the Combined Response Model." *Expert System Appl.*, 17(2), 89-97.

Thompson, D. (2004). "Biosolids that don't stink-Using Temperature-Phased Reduction to Tackle the Odor Issue." *Water and Environ. & Technol.*, 16(7), 54-56.

Trevino, L. J., and Daniels, J. D. (1995). "FDI Theory and Foreign Direct Investment in the United States: A comparison of Investors and Non-investors." *International Business*, 4, 177-194.

U.S. Environmental Protection Agency (EPA). (1994). "Plain English Guide to the EPA Part 503 Biosolids Rule." *Rep. No.EPA/832/R-93/003*, U.S. Environmental Protection Agency, Office of Wastewater Management, Washington, D.C

U.S. Environmental Protection Agency (EPA). (1999). "Biosolids Generation, Use, and Disposal in the United States." *Rep. No.EPA/530/R-99/009*, U.S. Environmental Protection Agency, Office of Solids Waste, Washington, D.C

U.S. Environmental Protection Agency (EPA). (2000). "Guide to Field Storage of Biosolids." *Rep. No.EPA/832/B-00/007*, U.S. Environmental Protection Agency, Office of Wastewater Management, Washington, D.C
UCLA. (2008). "SAS Annotated Output Ordered Logistic Regression." UCLA Academic technology service. Retrieved August 22, 2008, from http://www.ats.ucla.edu/stat/sas/output/sas_ologit_output.htm

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, Maryland.

Voelz, L. D., Thompson, J. K., Gaudes, R. J., Abi-Samra, S., Schirripa, R., Torres, E. M., Dillon, C. D., Ahn, T., and Raine, T. J. (2003). "Odor Modeling for Overall Odor Control Planning." *WEF/AWWA ODORS AND AIR EMISSIONS*, Hartford, Connecticut.

WEF, 1997, "National Outlook-State Beneficial Use of Biosolids Activities." *Water Environmental Federation*, Washington, D.C.

Westgaard, S., and Wijst, N. V. D. (2001). "Default Probabilities in a Corporate Bank Portfolio:a Logistic Model Approach." *European J. Oper. Res.*, 135(2), 338-349.

Wiginton, J. C. (1980). "A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior." *J. Financial Quant. Anal.*, 15, 757-770.

Williams, T. O., and Servo, S. (2005). "Odor Modeling at a Biosolids Composing Facility." *Biocycle*, 46(3), 25.

Zador, J., Zsely, I.G., Turanyi, T., Ratto, M., Tarantola, S., and Saltelli, A. (2005). "Local and Global Uncertainty Analyses of a Methane Flame Model." *J. Phys. Chem.* , 109, 9795-9807.