To the Graduate Council:

I am submitting herewith a thesis written by Patrick Andrew Dobbs entitled “Risk Analysis of Decentralized Wastewater Design Flows.” I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Biosystems Engineering.

John R. Buchanan

Dr. John R. Buchanan, Major Professor

We have read this thesis
and recommend its acceptance:

Chris D. Cox

Dr. Chris D. Cox

John S. Tyner

Dr. John S. Tyner

Accepted for the Council:

Carolyn Hodges

Vice Provost and Dean
of the Graduate School

Risk Analysis of Decentralized Wastewater Design Flows

A Thesis
Presented for the
Master of Science
Degree in
Biosystems Engineering
The University of Tennessee, Knoxville

Patrick Andrew Dobbs
May 2007

Dedication

This thesis is dedicated to my excellent family for encouraging me to succeed and for being great fun.

Acknowledgements

Professors and educators thank you for communicating scientific knowledge. Dr. Cox and Dr. Tyner, thank you for the expertise that you both brought to this project. This document is intended to model the variation of the views that you both expressed. Dr. Buchanan, thank you for finding the time to allow me to be your graduate student. Committee members, thank you for the time that all of you dedicated to this project. The many hours that you spent discussing this project with me are hours that enhanced my education beyond expectations. The time that you spent reading and writing emails and reading and editing my writing has led to an educational experience that I was not sure existed. As my committee, you made me feel like I was part of team.

Thank you to my family and friends for listening to me talk about this project and for the encouragement and faith that you offered. Thank you to God for listening and answering my prayers about finishing this project.

Abstract

Decentralized wastewater treatment systems are often designed at flows of either 284 L/person/d (75 gal/person/d) or 568 L/bedroom/d (150 gal/bedroom/d). Water use data suggest that designing systems at these flow rates can lead to overly conservative designs. A study quantifying the risk of failure (exceeding a system design flow) was needed to create a design basis for future systems. The objectives of the study were to quantify the risk of failure of decentralized system design flows depending on the number of residences served by a system and to develop new guidelines for design flows of cluster systems based on quantifiable research. Data sets were from Consolidated Utility District of Rutherford County, Tennessee and contain water use information from July 2005 through July 2006 for seven subdivisions (636 residences) served by cluster systems. Water use was adjusted to wastewater production in each data set using a factor of 80 percent, and from each data set, probability distributions of average monthly flows and monthly peaking factors were made to model the variance due to residences and months, respectively. Monte Carlo simulations were conducted to simulate monthly flow distributions for differing numbers of residences, which were evaluated for risk of exceeding differing design flows. For subdivisions with thirty or more three-bedroom residences, the results show that a design flow of 25552 L/month/residence (225 gal/d/residence) limits the yearly risk of exceeding a month's design flow to less than one percent. The results of this study can be used to design future cluster systems in similar regions.

Table of Contents

Chapter 1: Introduction

Background

Decentralized Treatment vs. Centralized Treatment

Decentralized wastewater systems treat and dispose of wastewater at or near the site of wastewater generation. Decentralized systems are used as an alternative to centralized wastewater treatment systems. Centralized systems serve large densely populated areas and are composed of sanitary sewers leading to a high volume treatment facility. Decentralized systems serve areas lacking centralized treatment facilities, areas that have low density populations; these systems often serve a single residence or a group of residences. Systems serving a group of residences are often referred to as cluster systems. Decentralized systems exist in a variety of forms. A common configuration for a decentralized wastewater system serving a single residence or business is a "conventional" system, which is composed of a septic tank and a subsurface infiltration gallery made up of field lines. Advanced technologies have led to other forms of decentralized collection, treatment, and dispersal, such as low pressure pipe systems, mound systems, drip systems, sand filters, and aerobic treatment units. An example of an advanced collection system is the cluster system, which is formed of several residences, each with a septic tank and pump tank. The septic tank provides primary treatment before wastewater flows into the pump tank, which pumps the wastewater to the secondary treatment stage, which is often a sand filtration system. After secondary treatment, cluster systems typically dispose of wastewater using drip lines in a disposal field.

Health, Safety, and Environmental Impacts

Wastewater treatment benefits society by protecting human health and the Earth's environment. Untreated wastewater can contain pathogens that can cause many diseases, such as cholera and typhoid fever. These diseases are very rare in the United States because of wastewater sanitation practices. Untreated wastewater contains chemical pollutants (nutrients) that lead to increased algae growth, which can lead to decreased dissolved oxygen. Nitrates from wastewater and other sources can pollute drinking water supplies and lead to methemoglobinemia (blue baby syndrome), which causes decreased oxygen levels in the blood of infants leading to suffocation (U.S. EPA, 2002). Drinking water supplies can also be contaminated by the aforementioned pathogens and other nutrients. Decentralized wastewater treatment is a method of protecting human health and the environment in areas that centralized wastewater treatment is unavailable.

Modern Decentralized Wastewater Treatment History

Decentralized wastewater systems serve 25% of the population and 40% of new development in the United States (Hogye et al., 2001). The U.S. Environmental Protection Agency (U.S. EPA) recognized a growing need for wastewater disposal facilities and recommended to Congress that decentralized wastewater systems be used as a long term solution for wastewater treatment (U.S. EPA, 1997). The increased number of small communities that are served by decentralized wastewater systems and the growth in decentralized

technologies have led to a need for research to validate the design methods for these systems (Siegrist, 2001).

Design Flows

A fundamental step in wastewater treatment system design is the determination of wastewater flow, which should be determined either from existing data or estimated from a data set of a similar treatment system (Metcalf & Eddy, 1979). Knowledge of the wastewater flow creates a cost effective design by both minimizing the initial system costs and preventing future costs due to system failure.

Design flows for centralized wastewater treatment systems and drinking water supply systems are often designed on a per capita basis using data from existing centralized or water supply systems. Cluster systems are often designed using an expected average per residence flow, which is dependent on having a large enough number of residences connected to the system for the average to be consistently achieved.

Design flows for decentralized wastewater treatment systems range from 284 (Perkins, 1989) to 380 liters per person per day (Imhoff et al., 1989) (75 to 100 gallons per person per day). Decentralized systems are often built for residences without the knowledge of the exact number of occupants; so, many states have developed guidelines for design flows based on either the number of bedrooms in a residence or the floor area of the residence. Tennessee, amongst many other states, uses a standard design flow of 568 liters per bedroom per day (150 gallons per bedroom per day) (Tennessee, 2006). An important note is that

design flows can vary between states, and some larger decentralized wastewater systems, like cluster systems, are designed at lower per unit flows.

In contrast to the design flows, the U.S. EPA recently published average wastewater flows that range from 189 to 265 liters per person per day (50 to 70 gallons per person per day) (U.S. EPA, 2002). These values are based on measured flows from studies of hundreds of residences; however, most regulatory agencies have not adopted these as design values.

Data vs. Design

A comparison can be made between the U.S. EPA expected wastewater flow and Tennessee's required design flow. Using a typical three bedroom residence as an example, most states require a design flow of 1700 L/d (450 gal/d). The U.S. EPA data of 265 liters per person per day and average household size of 2.7 persons suggest an expected flow of 716 L/d (190 gal/d). For most states, the required design flow is approximately 2.4 times larger than the expected flow of wastewater.

This calculation offers insight into the conservative design of systems. Conservatism is needed because with a single residence system, the system must work for above average conditions. Particularly, not every residence is an average residence; some residences may have only one occupant, while others could have five or ten occupants. Also, not every person uses the average flow, the EPA reports a range of standard deviations from flow studies with an average standard deviation of 150 L/person/d (40 gal/person/d) (U.S. EPA 2002). Variation of this magnitude is often cited as a reason for over designing a

system; however, without accurate data pertaining to particular system designs (a system serving thirty residences may not need the same design flow as a system serving one residence), it is unknown whether a safety factor of 2.4 is necessary. Additionally, advanced treatment systems may not operate satisfactorily if under loaded.

Need

A seemingly large safety factor causes concerns about whether decentralized system design flows are too conservative, and previous research suggests that design flows are too conservative (Berkowitz, 2001; Sievers and Miles, 2001). The current design methods succeed in limiting the risk of design flow exceedance, but it is not known to what degree the risk is limited. New knowledge is needed to quantify failure risk of current design flows. The term risk of failure is used to indicate the probability of exceeding the design flow. This definition of failure is much more conservative than definitions put forth by most state regulatory agencies, which typically involve sewage surfacing in a drainage field, sewage backing up into a residence, or sewage entering a nearby waterway.

The risk of failure of a design flow is dependent on the number of residences connected to a system. The risk of applying a decreased design flow on a cluster system should be less than the risk of applying the same design flow to a single residence system, because multiple residences are served by a cluster system, and it is not expected that all of the residences will meet or

exceed the design flow concurrently. In contrast, if a single residence system exceeds the design flow, then the system fails.

Purpose

The objective of this study was to quantify the risk of failure of decentralized system design flows depending on the number of residences served by a system. The current baseline flow for a decentralized system serving a single three bedroom residence is 1700 L/d (450 gal/d). The risk of failure as a function of the number of residences using a system will show when the 1700 L/d design flow can be decreased to an equally effective design flow. The goal is to develop new guidelines for decentralized system design flows based on quantifiable research.

Preliminary Methods

Data Background

To achieve the goal, a multi-step plan was developed with the first step being data collection. The data were collected from Consolidated Utility District (CUD) of Rutherford County, Tennessee. CUD provides water to and manages twenty decentralized cluster systems, which serve subdivisions ranging from approximately 30 to 115 residences. The collected data are from seven of these systems, which service residences ranging in size from three to five bedrooms and 1200 to 3000 square feet. The data are from July 2005 through July 2006 and contain customer identification numbers, dates, and corresponding monthly water usages in tens of gallons. Since the data are in monthly increments,

system failure will be defined as any measured flow exceeding the design flow on a per month basis.

CUD has evaluated data from January 2004 through July 2005 (19 months) for all of the cluster systems that it manages. The average flows for each subdivision range from 394 to 980 L/d/residence (104 to 259 gal/d/residence). For the wastewater systems, CUD uses a design flow of 757 L/d/residence (200 gal/d/residence). Based upon the data across all of the cluster systems that CUD manages, which show an average flow of 610 L/d/residence (161 gal/d/residence) and a median flow of 560 L/d/residence (148 gal/d/residence), the design flow appears to work in most cases, but three of the twenty subdivisions have averages during this data collection period exceeding the 757 L/d/residence (200 gal/d/residence) design flow. Analysis of the new data (July 2005 through July 2006) will provide more information concerning the design flows of these systems and the associated risks.

Project Limitations

Project limitations occur due to the kind of data being used. Since the data was from a water utility, the data shows the amount of water a residence uses in a month. Two limitations result from this. First, the data represent the amount of water a particular residence consumes and not the amount of wastewater that a residence is producing. Water that is used for lawn watering or car washing does not enter the wastewater treatment system, but is counted in the data. This will cause any results from this project to slightly err on the side of a larger design flow.

The second limitation is that the data is for monthly water use; therefore, the smallest time frame for failure analysis is one month. This ignores that a system could potentially exceed its design flow for half of a month and then be under its design flow for the other half of the month and still appear to be operating within the design. The project can only make statements about any system on a monthly basis. Additional studies will have to be performed to find information relating to daily risk of system failure.

The project is also limited in application because all of the data is from one county in Tennessee. The information from this project should only be applied to areas that are considered similar to the data collection region.

Data Formatting

The data provided by CUD are in the form of a spaced text file for each system (subdivision) containing useful information about the billing date, water usage, and customer being billed. The space text file was formatted to a tab delimited text file that only contains the aforementioned useful information. The billing dates for each customer are used to assign a month to each water usage, indicating the month that the water usage occurred. Months were assigned to usages by calculating the number of days between billing dates and then discerning which month contains the majority of the days in the billing period. An example of this is a customer receives a bill on the fifth of November, and the previous bill was issued on the fifth of October; the bill indicates that the usage is for November, but looking at the date on the previous bill, one can deduce that the water usage on the bill is for the month of October. Performing this check is

important because some data sets have billing dates at the beginning of month, while others have billing dates at the end of months, and as seen a billing date at the beginning of a month indicates a previous months water usage.

After assigning a month to each water usage, the usage was normalized for the month that it has been assigned. For a given customer, the water usage was normalized for each month by dividing the water usage by the number of days in the billing period to get an average daily flow for the month, and then multiplying this flow by the number of days in the month to get a monthly flow. Monthly flows were calculated in this manner because billing periods are not exactly one month in length.

The data formatting process was dependent on the information for the current billing date and the previous billing date. The data at the beginning of the data set, July 2005, were only used for the billing date information and not the flow information because without a previous billing date the time period of the flow was unknown. Another important note was that not all residences had data for the entire data collection period of July 2005 through July 2006. This occurs because occupants could move into or out of residences during this time period resulting in a partial data set. The partial data sets were used, but as mentioned the flows from the beginning of these data sets were not used.

Preliminary Statistical Analysis

The first part of the analysis was to identify important descriptive statistics for each subdivision. These included means, modes, medians, ranges, standard deviations, maximums, and minimums. Another part of the analysis was to test

for variances in flows due to months. Differences in months were tested by Analysis of Variance (ANOVA) using Statistical Analysis Software (SAS, 2004). When differences exist amongst the months, a decision was made about how to handle this variation when assessing risk. SAS was used to identify statistically similar data sets. The similar (equal variances) data sets will be combined to create a larger data set for risk analysis.

A Randomized Block Design (RBD) was used to test for differences within the months. This experimental design was used to control for expected variation in one factor when testing for variation in another factor. In this study, variation was expected to exist between different residences. One residence may have only two occupants, while another home may have four occupants. The home with more occupants is expected to use more water; this variation is expected and is controlled for by making it the blocking factor in the RBD. If differences were found within the months, then an attempt was made to combine the data sets from each subdivision in an effort to decrease the number of calculations.

For testing whether differences exist between subdivisions, a strip-plot experimental design, sometimes referred to as an RBD strip-plot, was used. This design is used when large experimental units exists with multiple blocking factors within the experimental unit. The factors that were blocked or controlled for variation in this analysis are the months and residences, since variation is expected to exist for both. Controlling for these variations allowed for a more accurate test of whether true differences existed between subdivisions.

Outlier Identification

Outliers occur in the data sets for several practical reasons. High outliers can occur because a residence could water their grass, have a water line burst, or could be filling a swimming pool. Low outliers could occur if occupants of a residence go on vacation or and leave the residence vacant. Low outliers were removed using cited materials later in the project discussion.

High outliers were identified in each data set by calculating a winter average flow and standard deviation. The winter months are used in order to avoid the seasonally high water use of the summer due to outdoor activities, such as washing cars and watering lawns. The winter average flow was calculated using the flows from the months of November, December, January, February, and March. The high outlier criterion for each data set was three standard deviations above the winter average. The high outlier criteria ranged from 37000 to 49000 L/month (approximately 10000 to 13000 gal/month) depending on the subdivision. Different high outlier criteria were used for each subdivision due to the expected variability among high water uses for different subdivisions; wealthier subdivisions with larger homes are expected to use more water than other subdivisions. An example is that wealthier subdivisions tend to have more lawn irrigation systems, oversized bath tubs, and social events; all result in higher water demand.

In summary, outliers were identified for each data set based on a low criterion of 2840 L/month and a high criterion that was calculated by finding three standard deviations from the winter mean for each subdivision.

Probability Distribution

Probability distributions were developed for the individual subdivision data sets. The distributions describe the likelihood (probability) of a particular flow occurring. For each data set probability distribution functions were found using Crystal Ball software, which is an Excel add-on (Crystal Ball, 2006). The software automatically calculates the probability of a particular data point (flow) occurring, by finding the number of times a flow in a particular range (bin) occurs and then dividing that by the total number of flows. This information is plotted, and several different probability distributions are fitted to the data using statistical parameters.

Each distribution that is fit to the data set was evaluated using three different goodness-of-fit tests. A goodness-of-fit test mathematically measures how well the probability distribution fits the data. Crystal Ball performs the Chi-square, Kolmogorov-Smirnov, and Anderson-Darling goodness-of-fit tests.

The Chi-square test is the classic goodness-of-fit test. It breaks the distribution down into regions of equal probability, and then compares the number of data points occurring in a probability region to the number of expected data points for that region. Effectively, this is evaluating the differences in the vertical distances between the data and the distribution that has been fit to the data (Crystal Ball, 2000; Stanford and Vardeman, 1994). Typically, a Chi-square value greater than 0.5 indicates a good fit. One limitation of this test is that it requires a large number of data points to be valid. It is based on a summation of the measures of fit for each probability region, and this could cause a close fit in

a couple of regions and a poor fit in other regions to sum to an apparently good fit (Crystal Ball, 2000).

The Kolmogorov-Smirnov test measures the largest vertical distance between the cumulative distribution of the data and the cumulative distribution that has been fit to the data (Stanford and Vardeman, 1994). Usually, a value less than 0.03 indicates a good fit (Crystal Ball, 2000). This test tends to be most sensitive at the center of the distribution; so, if the tails of the distribution are a concern, then this test is not the best test to use.

The Anderson-Darling test is a modification of the Kolmogorov-Smirnov test that weights the differences at the tails of the distribution more than the differences at the middle of the distribution. Generally, a value less than 1.5 indicates a good fit(Crystal Ball, 2000; Stanford and Vardeman, 1994). The Anderson-Darling goodness-of-fit test was thought to be the most useful goodness-of-fit test for this project since the probability distributions were expected to have relatively large standard deviations and therefore extended tails.

In addition to using goodness-of-fit tests to find a probability distribution that closely fits the data, one should visually check that the selected distribution matches the data set. A distribution could result in an acceptable goodness-of-fit statistic, but could not actually fit the distribution well. This could happen by a distribution closely fitting a majority of the data, but then not fitting an important section of the data, particularly higher or lower data points (*i.e.* the tails of the distribution).

Monte Carlo Simulation

A Monte Carlo simulation is a process that uses pseudo-random numbers to predict the result of a model. The name Monte Carlo refers to Monte Carlo, Monaco, which is known for its casinos, which house games of chance, such as roulette, craps, slot machines, and poker, which are based on random processes. Historically, Monte Carlo simulations are associated with providing critical information to the Manhattan Project for the development of the first nuclear weapons.

In a model, a Monte Carlo simulation (MCS) randomly selects variable values; however, this process is not truly random and is correctly referred to as a pseudo-random process. For a given variable, a probability distribution must be defined. The distribution describes the likelihood of every possible value for that variable. The MCS then randomly, based on the probability distribution of the data, selects values for the variable; this is why the process is pseudo-random because only values defined by the distribution can occur and the values with higher probabilities will be selected more often.

To get reliable results from the MCS, many trials (thousands) must be performed. Each trial generates a value for each variable in the simulation and recomputes the model based on these new variables. Thousands of trials are performed to find the likelihood (probability distribution of the results) associated with each result of the model.

The MCS was used to model the wastewater produced by a single home or group of homes. The probabilities associated with the results of the models are synonymous with the risks of exceeding the design flow. The objective of this study was to quantify the risk of failure of decentralized system design flows depending on the number of residences served by a system.

Chapter 2: Risk Analysis of Decentralized Wastewater Design Flows

Introduction

Decentralized wastewater systems serve 25% of the population and 40% of new development in the United States (Hogye et al., 2001). The U.S. Environmental Protection Agency (U.S. EPA) recognized a growing need for wastewater disposal facilities and recommended to Congress that decentralized wastewater systems be used as a long term solution for wastewater treatment (U.S. EPA, 1997). The increased number of small communities that are served by decentralized wastewater systems and the growth in decentralized technologies have led to a need for research to validate the design methods for these systems (Siegrist, 2001).

A fundamental step in the design of a wastewater treatment system is the determination of wastewater flow, which should be determined either from existing data or estimated from a data set of a similar treatment system (Metcalf & Eddy, 1979). Design flows for decentralized wastewater treatment systems range from 284 (Perkins, 1989) to 380 L/person/d (Imhoff et al., 1989) (75 to 100 gal/person/d). Due to the variability in the number of occupants a residence could house, most states have developed guidelines for design flows based on either the number of bedrooms in a residence or the floor area of the residence. Tennessee, amongst many other states, uses a standard design flow of 568 L/bedroom/d (150 gal/bedroom/d) (Tennessee, 2006). An important note is that design flows can vary between states, and multi-residence decentralized wastewater systems are typically designed at lower flows per unit. In contrast to

the design flows, the U.S. EPA recently published average wastewater flows that range from 189 to 265 L/person/d (50 to 70 gal/person/d) (U.S. EPA, 2002). The U.S. EPA also reports an average household size of 2.7 people (U.S. EPA, 2002).

A comparison can be made between the U.S. EPA expected wastewater flow and Tennessee's required design flow. Using a typical three bedroom residence as an example, most states require a design flow of 1700 L/d (450 gal/d). The U.S. EPA data of 265 liters per person per day and average household size of 2.7 persons suggest an expected flow of 716 L/d (190 gal/d). For most states, the required design flow is approximately 2.4 times larger than the expected flow of wastewater.

This calculation offers insight into the conservative design of systems. Conservatism is needed because with a single residence system, the system must work for above average conditions. Particularly, not every residence is an average residence; some residences may have only one occupant, while others could have five or ten occupants. Also, not every person uses the average flow, the EPA reports a range of standard deviations from flow studies with an average standard deviation of 150 L/person/d (40 gal/person/d) (U.S. EPA 2002). Variation of this magnitude is often cited as a reason for over designing a system; however, without accurate data pertaining to particular system designs (a system serving thirty residences may not need the same design flow as a system serving one residence), it is unknown whether a safety factor of 2.4 is

necessary. Additionally, advanced treatment systems may not operate satisfactorily if under loaded.

A seemingly large safety factor causes concerns about whether decentralized system design flows are too conservative, and previous research suggests that design flows are too conservative (Berkowitz, 2001; Sievers and Miles, 2001). This creates a need for a study quantifying risk of failure for current design flows. The term risk of failure is used to indicate the risk of exceeding the system design flow. This definition of failure is much more conservative than definitions put forth by most state regulatory agencies; failure typically means sewage surfacing in a drainage field, sewage backing up into a residence, or pollution of a nearby waterway.

The risk of failure of a design flow is dependent on the number of residences connected to a system. The risk of using a small design flow on a cluster system should be less than the risk of applying the same design flow to a single residence system, because multiple residences using a cluster system are not expected to meet or exceed the design flow concurrently. A system serving a single residence will fail if the design flow of that residence is exceeded; however, a system serving ten residences will not fail as long as the total flow from all ten residences is less than the system design flow. A residence on a system with ten other residences could exceed its per residence design flow, but if the other residences do not meet or exceed their design flows then the system will not fail. Systems serving a large number of residences can be designed for

an expected average flow per residence instead of an expected maximum flow per residence.

The objective of this study was to quantify the risk of failure of decentralized system design flows depending on the number of residences served by a system. The goal was to develop new guidelines for decentralized system design flows of cluster systems based on quantifiable research.

Methods

Data Background

To achieve the goal, a multi-step plan was developed with the first step being data collection. The data were collected from Consolidated Utility District (CUD) of Rutherford County, Tennessee. CUD provides water to and manages twenty decentralized cluster systems, which serve subdivisions ranging from approximately 30 to 115 residences. The collected data were from seven of these systems, which service residences ranging in size from three to five bedrooms and 1200 to 3000 square feet. The data were from July 2005 through July 2006 and contained customer identification numbers, dates, and corresponding monthly water usages in tens of gallons. Since the data were in monthly increments, system failure was defined as any measured flow exceeding the design flow on a per month basis.

CUD has evaluated data from January 2004 through July 2005 (19 months) for all of the cluster systems that it manages. The average flows for each subdivision ranged from 394 to 980 L/d/residence (104 to 259 gal/d/residence). For the wastewater systems, CUD uses a design flow of 757

L/d/residence (200 gal/d/residence). Based upon the data across all of the cluster systems that CUD manages, which show an average flow of 610 L/d/residence (161 gal/d/residence) and a median flow of 560 L/d/residence (148 gal/d/residence), the design flow appears to work in most cases, but three of the twenty subdivisions had averages during this data collection period exceeding the 757 L/d/residence (200 gal/d/residence) design flow. Analysis of the new data (July 2005 through July 2006) provided more information concerning the design flows of these systems and the associated risks.

Project Limitations

This project had limitations due to the kind of data being used. Since the data was from a water utility, the data is monthly water usage not wastewater production. The first limitation was that the data is water usage, and an estimate had to be made to relate it to wastewater production; the estimate used was 80 percent of water used becomes wastewater (Crites and Tchobanoglous, 1998; Metcalf and Eddy, 2003). The second limitation was that the data was for monthly water use; therefore, the smallest time frame for failure analysis was one month. This ignores that a system could potentially exceed its design flow for half of a month and then be under its design flow for the other half of the month and still appear to be operating within the design. The project will only make statements about any system on a monthly basis. Additional studies will have to be performed to find information relating to daily risk of system failure. The project was also limited in application because all of the data was from one

county in Tennessee. The information from this project should only be applied to areas that are considered similar to the data collection region.

Outlier Identification

Outliers occur in the data sets for several practical reasons. High outliers can occur because a residence could water their grass, have a water line burst, or could be filling a swimming pool. Low outliers could occur if occupants of a residence go on vacation or and leave the residence vacant. Low outliers were removed using cited materials later in the project discussion.

High outliers were identified in each data set by calculating a winter monthly average flow and standard deviation. The winter months were used in order to avoid the seasonally high water use of the summer due to outdoor activities, such as washing cars and watering lawns. The winter average flow was calculated using the flows from the months of November, December, January, February, and March. The high outlier criterion for each data set was three standard deviations above the winter average. The high outlier criteria ranged from 37000 to 49000 L/month (approximately 10000 to 13000 gal/month) depending on the subdivision. Different high outlier criteria were used for each subdivision due to the expected variability among high water uses for different subdivisions; wealthier subdivisions with larger homes were expected to use more water than other subdivisions. An example is that wealthier subdivisions tend to have more lawn irrigation systems, oversized bath tubs, and social events; all result in higher water demand.

High outliers, typically in excess of 57000 L/month/residence (15000 gal/month/residence), were eliminated prior to the adjustment of the water usage data to wastewater production data. The later adjustment of water usage to wastewater production would not be effective for outlier data; outlier data would require a different estimated adjustment in order to be included in the data sets, such an adjustment would include an additional level of complexity and an unreasonable amount of estimation.

Analysis of Variance

Analysis of Variance (ANOVA) testing was performed to find similar data sets and similar months within data sets using Statistical Analysis Software (SAS). First each data set was evaluated separately to find months with similar flow distributions. A Randomized Block Design (RBD) experiment was conducted blocking on individual residences because water usage was expected to vary between residences. The treatment was months, which was used to test for differences in flows. ANOVA was conducted using mixed models (SAS, 2004; Saxton, 2006) and least squares means were separated using Tukey's significant differences test. Months were different ($P<.01$) for all data sets. Based on the least squares means some months were similar to other months, but to avoid any confusion resulting from combining different sets of months for different subdivisions, no months were combined, and the data sets were left unchanged.

Data set (subdivision) differences were also tested in an effort to combine similar data sets, and hopefully generate one large data set from the seven

separate data sets for risk analysis. An RBD split-plot was conducted with subdivision being the treatment and months and residences being blocks. ANOVA was conducted using mixed models (SAS, 2004; Saxton, 2006) and least square means were separated using Tukey's significant differences test. Subdivisions were different ($P<.001$). The subdivisions were therefore not combined, and the data sets were left unchanged.

Intermediate results of statistics are presented in Tables 1 and 2. Table 1 presents the means and standard deviations for the data sets after outlier elimination. Table 2 presents the mean separation letter groupings from the ANOVA of the subdivisions.

Risk Analysis

To analyze the amount of risk associated with a design flow, the sources of variability leading to the risk must be identified. Three sources of variability were identified in each data set: varying water use among residences, varying water use with season (months), and random variation. A method of risk analysis was developed that simulated the two main types of variability due to residence and time.

First a method of modeling the variability due to different residences was developed using Microsoft Excel (Microsoft, 2003) and Crystal Ball, an Excel add-on for risk analysis and simulation, (Crystal Ball 7, 2006). An average monthly flow was calculated for each residence in a data set (subdivision) by summing the monthly flows and dividing by the number of months resulting in an average monthly flow in units of L/month. A probability distribution was fit to the

Table 1: Means and standard deviations from data set after outlier elimination

Subdivision	Average Monthly Flow (Water Use)	Average Monthly Flow (Water Use)	Standard Deviation	Standard Deviation
	(L/month)	(gal/month)	(L/month)	(gal/month)
1	17100	4530	6830	1800
2	18100	4780	9830	2600
3	18800	4960	8890	2350
4	15500	4100	6630	1750
5	22800	6030	10420	2750
6	18600	4920	6580	1740
7	17600	4640	8370	2210

Table 2: Subdivision ANOVA mean separation letter groupings

Observation	Subdivision	Letter Grouping	Number of Bedrooms
1	5	A	4-5
2	6	AB	3
3	7	ABC	3
4	3	B	3
5	2	BC	3
6	1	BC	3-4
7	4	C	3

means to represent the likelihood of a residence having a certain mean. The distribution of the means in a subdivision was used to model the variability among the residences.

Crystal Ball has the capability to fit probability distributions to data sets and calculate goodness-of-fit tests for each distribution. Using Crystal Ball, the means from each data set were fit to distributions. The Anderson-Darling goodness-of-fit test was used to select a distribution that best represents the means. The Anderson-Darling test is a modification of the Kolmogorov-Smirnov test, which measures the largest vertical distance between the cumulative distribution of the data and the cumulative distribution that has been fit to the data. The Kolmogorov-Smirnov test is most sensitive at the center of the probability distribution; the Anderson-Darling test offers a modification that weights the differences at the tails of the distribution more than the differences at the middle of the distribution (Crystal Ball, 2000, Stanford and Vardeman, 1994). The higher sensitivity of Anderson-Darling at the tails was desired because information about the risk is in the upper tail of the distributions.

The means from each data set were fit to logistic probability distributions. Again, the logistic distributions were chosen because the Anderson-Darling goodness-of-fit test indicated that this type of distribution was a good fit (Crystal Ball, 2000, Stanford and Vardeman, 1994). The logistic distribution was fit using two parameters, a mean (μ) and a scaling factor (α). The distribution is of the form:

$$f(x) = \frac{z}{\alpha(1+z)^2}$$

where

$$z = e^{\left(\frac{\mu - x}{\alpha}\right)}$$

The term *x* represents the data point at *f(x)*. Each data set (subdivision) has a different logistic distribution based on the mean and scaling factor, and these distributions were used to represent the variability between residences. See Figure 1 for an example of a typical logistic distribution from the data.

The next step in the risk analysis was to develop a procedure to model the variability due to time represented by different months. For each residence, each monthly flow was represented by a peaking factor, which was calculated by $PF_{ij} = \frac{MF_{ij}}{AMF_i}$, where *PF* = peaking factor, *MF* = monthly flow, *AMF* = average monthly flow, *i* = residence, and *j* = month. For example, if residence one (*i* = 1) has a January (*j* = 1) flow of 20000 L/month (*MF* = 20000) an average monthly flow (*AMF*) of 22000 L/month, then the peaking factor for residence one in January ($PF_{1,1}$) was 0.91. An average month had a peaking factor of one, while an above average month had a peaking factor greater than one, and a below average month had a peaking factor less than one. The variability within each month in a data set was represented by a distribution of peaking factors. Using the same technique for selecting the distributions as used with the average monthly flows, distributions were selected for each month in each data set.

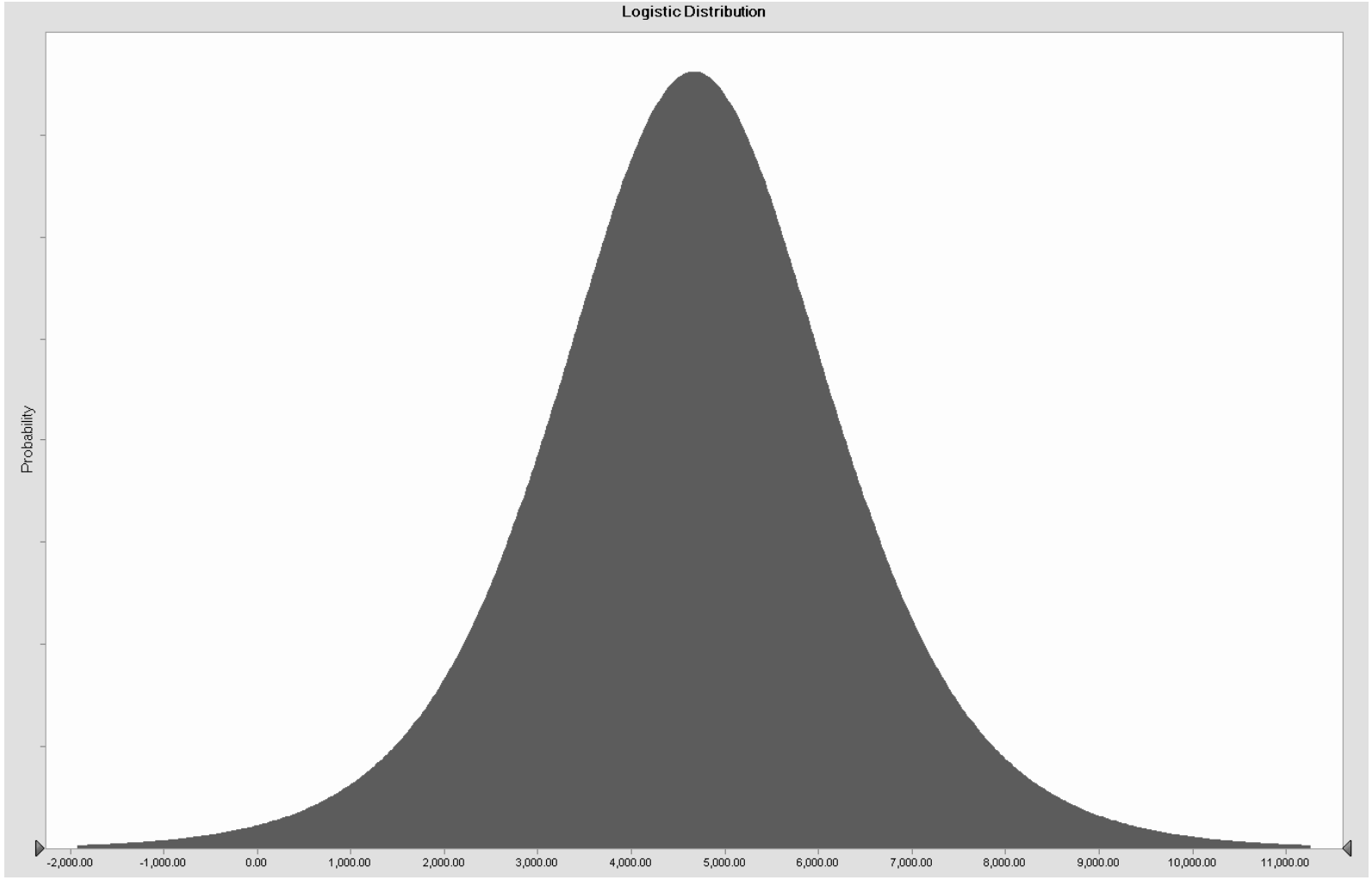

Figure 1: Example of logistic distribution

Crystal Ball fit several probability distributions to the data. The logistic distribution was chosen based on the goodness-of-fit indicated by the Anderson-Darling test. While the logistic distribution was not always the best fit, it was consistently a good fit for all the months in the data sets. Using a logistic distribution for each month greatly simplified the amount of computation involved in the risk analysis simulation by only requiring one subroutine performing calculations for one distribution type, instead of multiple subroutines for multiple distribution types.

No attempt was made to directly quantify and model the random variation in flows. How the Monte Carlo simulations handled the random variation will be addressed in the simulation discussion.

To analyze the risk of failure of different design flows, a Monte Carlo simulation (MCS) was performed. In a model, a MCS randomly selects variable values; however, this process is not truly random and is correctly referred to as a pseudo-random process. For a given variable, a probability distribution must be defined. The distribution describes the likelihood of every possible value for that variable. The MCS then randomly, based on the probability distribution, selects values for the variable. MCS uses the random variables to perform any calculation in a model that involves the variables. MCS performs these calculations thousands of times to develop a probability distribution of the output of the model. The distribution of the model output can then be used to evaluate the risk associated with the output, for example, the risk of the output exceeding a design value.

A simulation was performed for each data set to analyze the risk of failure of different design flows and different numbers of residences. The distribution variables used for each data set were the twelve (monthly) distributions of the peaking factors, and the distribution of average monthly flows. The distributions from the data were adjusted from the logistic distribution that was fit to each variable. Peaking factor distributions were adjusted by setting a condition that the distributions can only simulate values greater than zero. If a value less than or equal to zero occurs during the random simulation a new random value is selected until a value greater than zero results for the variable of interest. A second model boundary prevented the simulation of average monthly flows less than 3400 L/month (900 gal/month). This corresponded to a low outlier criterion associated with the idea that if a monthly flow less than 3400 L occurs, then a residence is unoccupied (Crites and Tchobanoglous, 1998). The distributions of the average monthly flows were also adjusted from water usage to wastewater production by multiplying the flow values by 80 percent (Crites and Tchobanoglous, 1998; Metcalf and Eddy, 2003). Crites and Tchobanoglous cite a range of 60 to 80 percent of water used becomes wastewater, and Metcalf and Eddy cite a range of 60 to 90 percent with the qualifying statement that higher percentages correspond to northern states in cold weather, and lower percentages correspond to the semiarid southwestern states, which use extensive landscape irrigation (Crites and Tchobanoglous, 1998; Metcalf and Eddy, 2003). The 80 percent value was chosen because it is a conservative value that is reasonable for the data region. The 90 percent value was not

chosen because, based on the comments made by Metcalf and Eddy, it would be too conservative.

For each data set, the MCS first sampled from the average monthly flow distribution. The MCS sampled values from the logistic distribution using the inverse transformation technique, which sets a random number, between zero and one, equal to the output, *f(x)*, of the logistic function and then solved the logistic function for the input variable, *x*. After the average monthly flow was sampled, a peaking factor for each month was sampled. Each peaking factor was then multiplied by the average monthly flow to get a flow for each month. This process introduced the only component of the model that accounts for any random behavior in the data. By sampling the monthly peaking factors independently, the average of the peaking factors for a year does not always equal one. The method used to calculate peaking factors for each residence resulted in an average of one, but by independently sampling a peaking factor for each month, a part of the randomness was maintained due to the expectation that a residence will not over time have the same peaking factors for every month.

The flows in each month, found by multiplying the average monthly flow by the peaking factors, were added together for each residence served by a decentralized system to find the flow of the system in each month. The result was a distribution of the system flow for each month. The simulated output distributions can be used to assess risk of system failure.

A validation test was performed to determine if the model was accurately simulating the behavior of the data set. To validate the model, subdivision one was simulated and the output was compared to the original data set. The root mean squared error (RMSE) was calculated for each set of monthly flows. The maximum RMSE observed for any month was approximately 970 L/residence (250 gal/residence); this error was not of great significance because a decentralized system has extra capacity in septic tanks and pump tanks, which could easily store this flow. The majority of errors were due to the prediction of higher, more conservative, flows than observed in the data sets used for the simulation.

In summary, high outliers were removed from the data. A logistic distribution of the average monthly flows was developed for each data set and for monthly peaking factors in each data set. The average monthly flow and peaking factor distributions were then sampled during a MCS that consists of 10000 trials (10000 trials are used because no significant changes occurred in the output distribution by increasing the number of trials). In each trial, each residence had a sampled average monthly flow that was multiplied by a sampled peaking factor for each month to get monthly flows; the monthly flows were summed to get a system flow for each month. The MCS resulted in distributions of monthly system flows which were related to the risk of system failure in a given month.

Results and Discussion

For each number of residences simulated, an output distribution was obtained for every month; from the distributions, output percentile information

was obtained that described the risks of exceeding the monthly design flow of the system. The risks at a particular design flow from every month were quantified from the output distributions. The risks from every month were summed to find the yearly risk of a system exceeding the monthly design flow one or more times.

Figure 2 shows the results of the risk analysis for data set (subdivision) one. The plot illustrates how the yearly risk associated with a particular design flow decreases as the number of residences being served by a decentralized system increases. For a design flow of 22712 L/month/residence, the yearly risk is limited to less than one percent when the system is serving 15 or more residences. The risk curves have an initially steep decrease as residences are added to a system; this behavior is expected because as the number of residences increases, the likelihood of each residence simultaneously exceeding a specified design flow decreases. The risk curves' asymptotic behavior shows the differences in risk associated with design flows approach zero as the number of residences increases, meaning that at some number of residences, increasing or decreasing the design flow does not significantly change the risk. For this subdivision using a 22712 L/month/residence design flow (200 gal/d/residence), the risk is less than one percent starting at approximately ten homes.

Figure 3 illustrates the results for each subdivision composed of three bedroom residences at a design flow of 22712 L/month/residence (6000 gal/month/residence). Each curve (subdivision) exhibits similar behavior, but some curves correspond to higher risks; this is a result of the population demographics of a subdivision. For example, a subdivision with residences

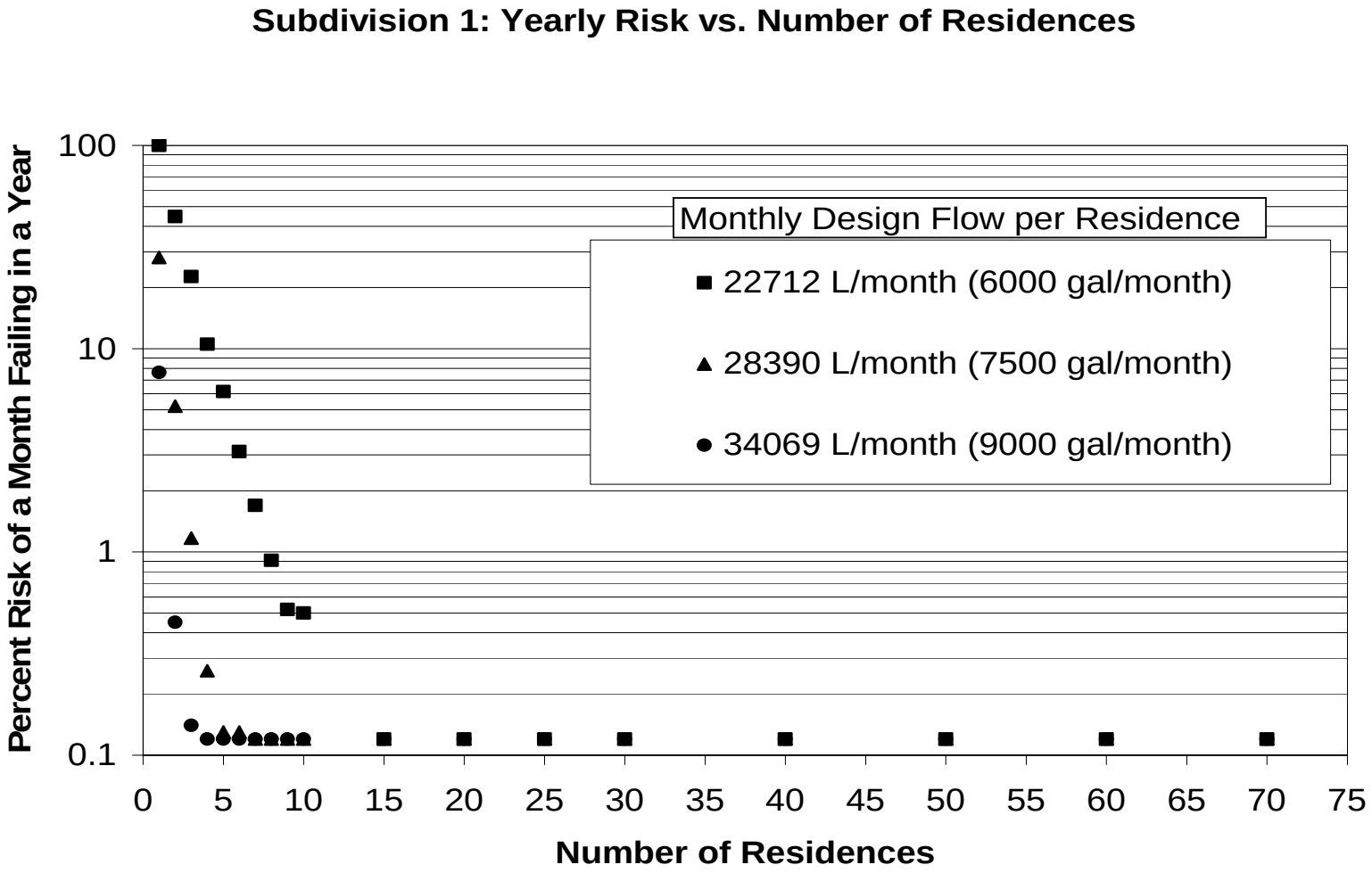

Figure 2: Risk of a single month failing in a year vs. number of residences

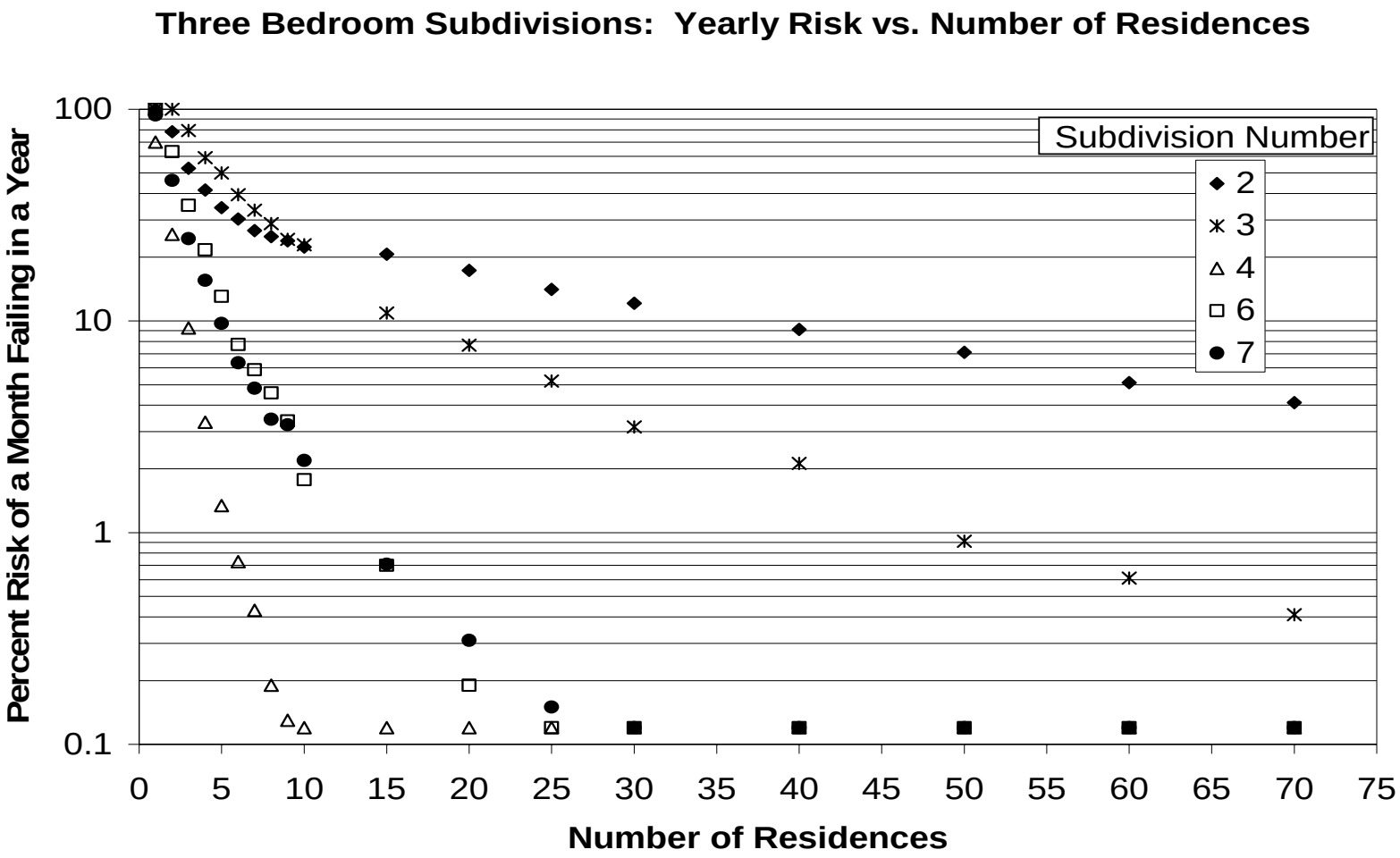

Figure 3: Three bedroom subdivisions: yearly risk vs. number of residences for 22712 L/month design flow

occupied by families with children is expected to produce more wastewater than subdivisions occupied by singles or couples without children. For a decentralized system serving 40 residences, the risk for most subdivisions is less than one percent. Subdivision two (diamonds) has higher risk values than the other subdivisions, with risk above one percent for a system serving 70 residences. If subdivision two is designed using 28390 L/month/residence (250 gal/d/residence), the risk reaches a level below one percent for as little as 15 residences.

Figure 4 differs from Figure 3 because it includes subdivisions one and five. Subdivision one contains three and four bedrooms residences, and subdivision five has four and five bedroom residences. Subdivision one's data (diamonds) are similar to the three bedroom subdivisions' data from Figure 3. Subdivision five illustrates the major concern with subdivisions containing larger residences having increased water usage and wastewater production. Again, it is important to note that the wastewater production data is estimated from water usage data. A subdivision with large residences like subdivision five could possibly produce wastewater that is less than 80 percent of the water usage, due to extensive use of lawn irrigation systems. Research suggests that wastewater production can range from 60 to 90 percent of water usage (Crites and Tchobanoglous, 1998; Metcalf and Eddy, 2003). Subdivision five can be designed at a flow of 28390 L/month/residence (250 gal/d/residence) and achieve a risk less than one percent for systems serving 30 or more residences. The results show that subdivisions with larger residences should have higher

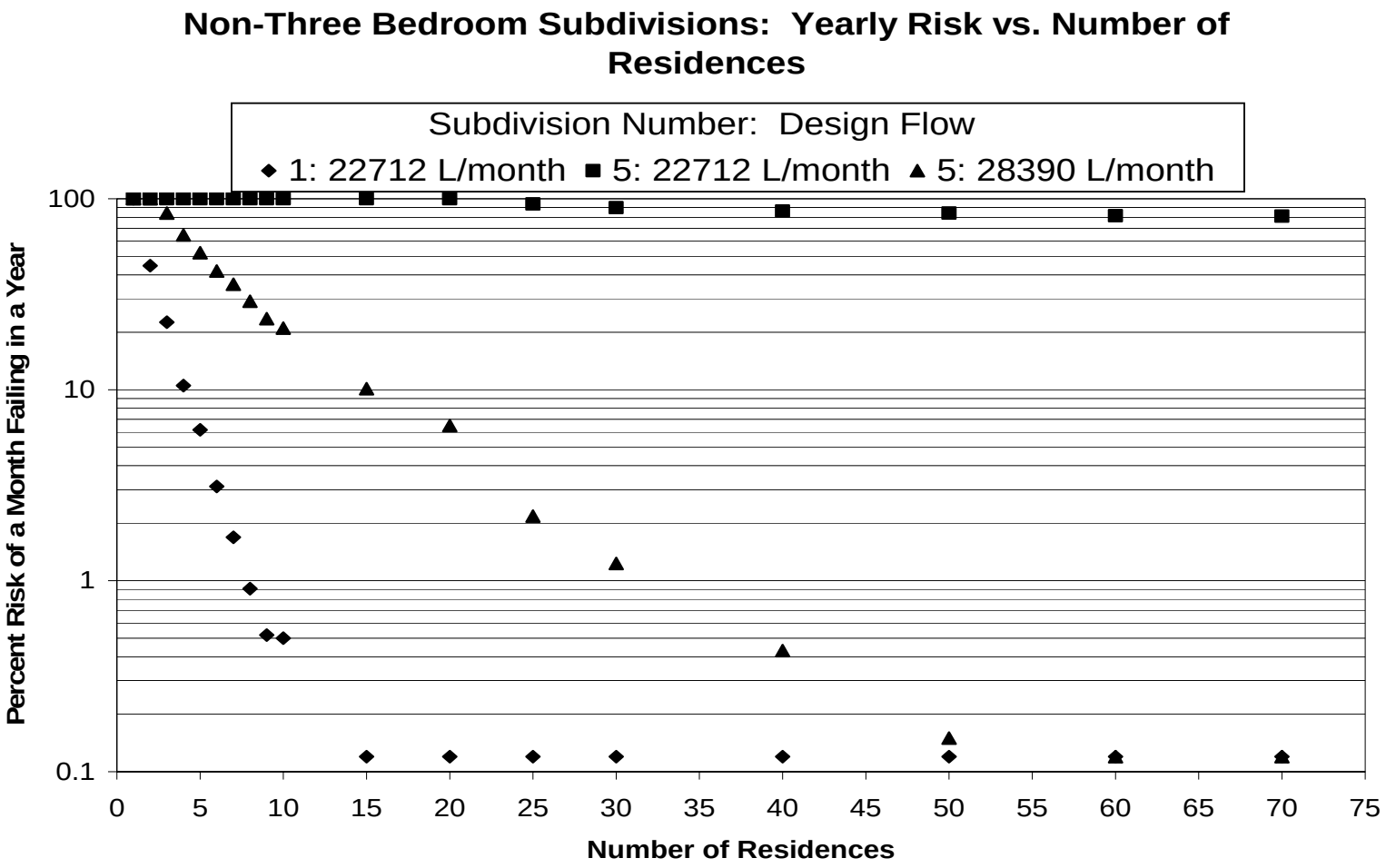

Figure 4: Non-three bedroom subdivisions: yearly risk vs. number of residences at specified design flows

design flows, but the analysis only included one subdivision with five bedroom residences; so, further research should be conducted with subdivisions containing large (five bedroom) residences to see if design flows of this magnitude are often required to achieve a one percent yearly risk of failure.

Conclusion

Decentralized wastewater system design flows can be more accurately determined by performing risk analyses on current data. When designing a decentralized system for more than thirty three–bedroom residences, a design flow of 25552 L/month/residence (225 gal/d/residence) decreases the yearly risk of exceeding the monthly design flow to less than one percent; some three bedroom subdivisions with thirty residences can be designed at 22712 L/month/residence (200 gal/d/residence) and still limit risk to less than one percent, but this design would not be recommended without proof that the new subdivision would have a similar population to an existing subdivision designed at the same flow. For decentralized wastewater systems with thirty or more five–bedroom residences, a design flow of 28390 L/month/residence (250 gal/d/residence) is suggested by the results. Again, since only one subdivision with five bedroom residences was included in this study, further research should be done with this type of subdivision to discern if the results are typical.

The objective of the study was to quantify the failure risk of decentralized system design flows depending on the number of residences served by a system. The goal of the study to develop guidelines for decentralized system design flows is achieved with the recommendation that subdivisions with thirty or

more three bedroom residences be designed at 25552 L/month/residence, and subdivisions with thirty or more five bedroom residences should be designed at 28390 L/month/residence. Again, refer to the project limitations section before using any value from the figures or recommendations for design purposes; also, note that an assumption is made that 80 percent of water used becomes wastewater (Crites and Tchobanoglous, 1998; Metcalf and Eddy, 2003).

Chapter 3: Conclusion

Water flow data can be used to evaluate the risk of design flows for decentralized systems. Design flow risks can aid in performance based design of wastewater systems by providing information about actual wastewater production of residences. Risk analysis of the data for this project indicate that a design flow of 25552 L/month/residence (225 gal/d/residence) will limit risk to less than one percent for systems serving thirty or more three bedroom residences; the recommendations of this project are based on the assumption that 80 percent of the water used becomes wastewater (Crites and Tchobanoglous, 1998; Metcalf and Eddy, 2003). Systems designed to serve large (five bedroom) residences should be designed at a larger flow (28390 L/month/residence), but more research is needed to confirm this design flow for other five bedroom subdivisions.

Other research should be performed to further the knowledge of decentralized wastewater system design flow risk. A study using wastewater data from individual residences over a period of years would be ideal because this would eliminate the need for any assumption about the amount of water usage that becomes wastewater. Studies incorporating both water use data from individual residences and total system wastewater flow data would also be valuable or possibly a study analyzing daily risk based on daily wastewater production data.

In conclusion, the current design method for cluster systems used by CUD of 22712 L/month/residence (200 gal/d/residence) is a design with excessive risk

and should be adjusted based on research. Systems built for five bedroom subdivisions should be designed at the higher flow previously mentioned. The cluster systems managed by CUD probably work due to the extra capacity in the systems due to septic tanks and pump tanks serving each residence, but the amount of wastewater being applied to the soil should be closely monitored to ensure that it does not exceed the design criteria.

References

Anderson, D.L., and R. L. Siegrist. 1989. The performance of ultra-low-volume flush toilets in Phoenix. Journal of the American Water Works Association. 81(3):52-57.

Berkowitz, S. J. 2001. Hydraulic performance of subsurface wastewater drip systems. p. 583-592. *In* K. Mancl (ed.) On-Site Wastewater Treatment: Proc. 9th Nat. Symp. Individual and Small Community Sewage Systems, Fort Worth, TX. 11-14 Mar. 2001. ASAE, St. Joseph, MI.

Brown and Caldwell. 1984. Residential water conservation projects. Research report 903. U.S. Department of Housing and Urban Development, Washington D.C.

Burton, D. J., F. H. Harned, B. J. Lesikar, J. F. Prochaska, and R. J. Sucheki. 2001. Design principles for drip irrigation disposal of highly treated on-site wastewater effluent. p. 606-616. *In* K. Mancl (ed.) On-Site Wastewater Treatment: Proc. 9th Nat. Symp. Individual and Small Community Sewage Systems, Fort Worth, TX. 11-14 Mar. 2001. ASAE, St. Joseph, MI.

Crites, R., G. Tchobanoglous. 1998. Small and decentralized wastewater management systems. McGraw-Hill, Boston, MA.

Crystal Ball. 2000. Crystal Ball 2000 user manual. Decisioneering Inc., Denver, CO.

Crystal Ball. 2006. Crystal Ball 7.2.2 user manual. Decisioneering Inc., Denver, CO.

Hogye, S., A. R. Rubin, and J. Hudson. 2001. Development of EPA guidelines for management of onsite/decentralized wastewater systems. p. 470-478. *In* K. Mancl (ed.) On-Site Wastewater Treatment: Proc. 9th Nat. Symp. Individual and Small Community Sewage Systems, Fort Worth, TX. 11-14 Mar. 2001. ASAE, St. Joseph, MI.

Imhoff, Klaus R., M. Olthof, P. A. Krenkel. 1989. Wastewater Characterization. p. 88. *In* Handbook of Urban Drainage and Wastewater Disposal. John Wiley & Sons, New York, NY.

Mayer, Peter W., W. B. DeOreo, D. M. Lewis. 2000. Seattle home water conservation study: the impacts of high efficiency plumbing fixture retrofits in single-family homes. Submitted to Seattle Public Utilities and the U.S. EPA by Aquacraft, Inc. Water Engineering and Management, Boulder, CO.

Metcalf & Eddy, Inc. 1979. Chapter 2: Wastewater Flowrates. *In* G. Tchobanoglous (ed.) Wastewater Engineering: Treatment Disposal Reuse. 2nd ed. McGraw-Hill, Inc. New York, NY.

Metcalf & Eddy, Inc. 2003. Wastewater engineering: treatment and reuse. McGraw-Hill, Boston MA.

Microsoft. 2003. Microsoft Excel XP. Microsoft Corporation, Redmond, WA.

Perkins, Richard J. 1989. Waste Process. p. 43-44. *In* Onsite Wastewater Disposal. Lewis Publishers, Inc., Chelsea, MI.

SAS. 2004. SAS 9.1.3 help and documentation. SAS Institute Inc., Cary, NC.

Saxton, A. 2006. Design and analysis macro for SAS. University of Tennessee, Knoxville, TN.

Siegrist, R. L. 2001. Advancing the science and engineering of onsite wastewater systems. p. 1-10, *In* K. Mancl (ed.) On-Site Wastewater Treatment: Proc. 9th Nat. Symp. Individual and Small Community Sewage Systems, Fort Worth, TX. 11-14 Mar. 2001. ASAE, St. Joseph, MI.

Sievers, D. M., and R. J. Miles. 2001. Performance of three drip irrigation disposal systems in a Karst sinkhole plain. p. 593-600. *In* K. Mancl (ed.) On-Site Wastewater Treatment: Proc. 9th Nat. Symp. Individual and Small Community Sewage Systems, Fort Worth, TX. 11-14 Mar. 2001. ASAE, St. Joseph, MI.

Stanford, John L. (ed.), and Stephen B. Vardeman (ed.), 1994. Methods of experimental physics: Statistical methods for physical science volume 28. Academic Press Inc., San Diego, CA.

Tennessee Department of Environment and Conservation Division of Ground Water Protection. Chapter 1200-1-6 Regulations to Govern Subsurface Sewage Disposal Systems. 1200-1-6. TDEC DGWP, Nashville, TN.

U.S. EPA. 1997. Response to Congress on Use of Decentralized Wastewater Treatment Systems. EPA/832/R-97/001b. U.S. EPA, Washington, D.C.

U.S. EPA. 2002. Onsite Wastewater Treatment Systems Manual. EPA/625/R-00/008. U.S. EPA, Washington, D.C.

Watson, J. T., and C. L. McEntyre. 2004. Peer reviewed guidelines for wastewater subsurface drip distribution. p. 068-072. *In* R. Cooke (ed.) On-Site Wastewater Treatment X Conference Proceedings, Sacramento, CA. 21-24 Mar. 2004. ASAE, Sacramento, CA

Appendices

Appendix I: Literature Review

Background/Current State of Technology

Research has identified, “the basis and need for advancing the science and engineering of onsite wastewater systems to secure their necessary and appropriate status as a component of a sustainable wastewater infrastructure,” (Siegrist, 2001). The term onsite is synonymous with decentralized in reference to wastewater treatment; an example is The Consortium of Institutes for Decentralized Wastewater Treatment (CIDWT), which is often referred to as “The Onsite Consortium” and has a web address of www.onsiteconsortium.org.

Decentralized systems serve 25% of the population, and 40% of new development in the U.S. utilizes decentralized systems (Hogye et al., 2001; Siegrist, 2001). The percentage of the population using decentralized technologies provides a basis for the need of increased research to validate current design methods.

The United States Environmental Protection Agency (U.S. EPA) recognized the need for decentralized technologies in 1997 by publishing its Response to Congress on Use of Decentralized Wastewater Treatment Systems (U.S. EPA, 1997). The U.S. EPA discusses a need for improving decentralized management techniques by improving design methods. The U.S. EPA concluded, “adequately managed decentralized wastewater treatment systems can be a cost effective and a long-term option for meeting public health and water quality goals, particularly for small, suburban, and rural areas,” (U.S. EPA 1997). The recent past has resulted in a push for research in the area of

decentralized technologies, particularly by government agencies such as the U.S. EPA.

The U.S. EPA has suggested that decentralized wastewater treatment can be improved by developing performance based requirements; examples of methods to define performance requirements are, “characterizing wastewater flows and pollutant loads, evaluating site conditions, and defining performance and design boundaries,” (U.S. EPA, 2002). The U.S. EPA encourages a shift from prescriptive management techniques to performance based management; prescriptive management sets forth a set of regulations that all systems must meet; however, performance based management requires that systems be designed in a logical scientific manner (U.S. EPA, 2002).

Water Usages

Anderson and Siegrist performed a water usage study finding a range of 5000 to 25000 gallons per month per residence (Anderson and Siegrist, 1989). The study measured water usages for residences in Phoenix, Arizona, by acquiring billing data from the local utility provider for an 18 month period (Anderson and Siegrist, 1989). Often water usage studies present data on a flow per person basis because indoor water flow is being observed and population statistics are available. A study conducted by Brown and Caldwell of 210 residences presents water usages that range from 57.3 to 73.0 gallons per person per day (Brown and Caldwell, 1984).

Another recent water usage study was performed by collecting one month of data by continuously monitoring flow for two weeks in two seasons at over

1100 residences from 12 cities (most of the cites in the study are in the western U.S.). This study found a water usage range of 57.1 to 83.5 gallons per person per day (Mayer et al., 1999). The study noted that all measured water use is indoor water use from flow meters attached to all of the water fixtures in each residence.

The U.S. EPA presents research of an average water usage of 68.6 gallons per person per day; additionally, the U.S. EPA estimates average daily wastewater flow for residences built before 1994 to be 50 to 70 gallons per person per day and 40 to 60 gallons per person per day for homes built after 1994 due to the Energy Policy Act requiring low flow water fixtures (U.S. EPA, 2002). The range of average wastewater flows observed by the U.S. EPA is similar but slightly less than the average water usage cited by the U.S. EPA; this is expected due to some outdoor water use (watering the lawn) and indoor water use (drinking water) that would not enter the wastewater stream.

Crites and Tchobanoglous present a method for calculating household water use based on 10 gal for dishwashing, 25 gal for laundry, and 5 gal for miscellaneous uses, and personal use of 2 gal for drinking and cooking, 3 gal for oral hygiene, 14 gal for bathing, and 16 gal for toilet flushing. The resulting equation is:

Flow, gal/home/d = 40 gal/home/d + 35 gal/person/d * (persons/home) (Crites and Tchobanoglous, 1998). For a home with three persons, the resulting indoor water use is 145 gal.

Wastewater Flows

Metcalf and Eddy present wastewater flow data showing that as the number of occupants in a house increases the per capita flow decreases (Metcalf and Eddy, 2003). A brief example is the comparison of the per capita flow of a one person household, 75 – 130 gal/d, to a three person household, 54 – 70 gal/d (Metcalf and Eddy, 2003). Another source cites per capita wastewater flow in terms of newer and older homes with newer homes having a range of 40 – 100 gal/person/d and older homes having a range of 30 – 80 gal/person/day. (Crites and Tchobanoglous,1998).

Metcalf and Eddy also present information for estimating wastewater production based on water use. A range of 60 – 90 percent of water use in the U.S. becomes wastewater. 90 percent corresponds to northern states during cold weather, and the lower percentages correspond to the semiarid southwestern states (Metcalf and Eddy, 2003). Crites and Tchobanoglous present a range of 60 – 80 percent of water use (Crites and Tchobanoglous, 1998)

Design Guidelines

A fundamental step in the design of a wastewater treatment system is the determination of the flow of wastewater, which should be determined either from existing data or estimated from a data set of a similar treatment system (Metcalf & Eddy, 1979; Burton et al., 2001; Watson and McEntyre, 2004). Design flows for decentralized wastewater treatment systems range from 284 (Perkins, 1989) to 380 liters per person per day (Imhoff et al., 1989) (75 to 100 gallons per

person per day). Decentralized systems are often built for residences without the knowledge of the exact number of occupants; so, many states have developed guidelines for design flows based on either the number of bedrooms in a residence or the floor area of the residence. Tennessee, amongst many other states, uses a standard design flow of 568 liters per bedroom per day (150 gallons per bedroom per day) (Tennessee, 2006). An important note is that design flows can vary between states, and some types of decentralized wastewater systems like cluster systems are designed at lower flows. Tennessee's standard design flow is 150 gallons per day per bedroom, which is based on 2 people per bedroom and 284 liters per person per day (75 gallons per person per day) (Tennessee, 2006). The primary exception to this is for cluster systems, which can be designed at flows of 200gal/d/residence.

Design Flow and Expected Flow

Design flows are often in excess of two times the amount of expected wastewater. Experiments performed show several instances that wastewater flows do not reach design flows (Berkowitz, 2001; Sievers and Miles, 2001).

Risk Analysis

Risk analysis in the field of wastewater treatment has primarily focused on risks posed to human health (Crites and Tchobanoglous, 1998; Metcalf and Eddy, 2003; U.S. EPA, 2002). Risk analysis focusing on human health evaluates the likelihood of human contact with wastewater components and the magnitude of the negative effects. The U.S. EPA though has recently requested systems be

designed on a performance basis, which will require risk analyses of design criteria and components failing (U.S. EPA, 2002).

Appendix II: Guide to Procedures

Data Formatting

1. Open spaced delimited text file in Excel. Excels data import window will open providing the opportunity to define the columns of data. Be sure that the customer id, month, day, and flow columns are all clearly marked before clicking finish.
2. Delete all columns that are not customer id, month, day, and flow.
3. Download and install the J-walk conditional row delete add-in for Excel from j-walk.com.
4. In the customer id column, open the J-walk add-in (it is found under Tools). Upon opening the add-in, the column for customer id should be selected, and the add-in will inquire the condition for deleting rows. Select “Equal to” and then 0. Click ok. This will delete all rows that do not contain data.
5. The data is now sorted by customer id and from the most recent month to the furthest past month.
6. Create a column that has the total number of days in each month. For example if the month in row 5 is January, then the corresponding number in this row will be 31 (*i.e.* January has 31 days). A sample of the Excel code to do this follows where column E contains months.

   =IF(OR(E3=1,E3=3,E3=5,E3=7,E3=8,E3=10,E3=12),31,IF(E3=2,28,30))

   The preceding code discerns the number of days the month in column E using conditional logic.

7. Create a column that calculates the number of days in the billing period for the flows in the data set. The code for this is =IF(A3<>A4,0,F3+G4-F4), where column A is customer ids. The code operates on the condition that the data in the next row (row 4) is from the same residence as the data in row 3. If the data in the next row is not from the same residence as the data in the current row, then a zero is input because the number of days for the billing period cannot be calculated without knowledge of the current and previous billing dates. When the condition is met the day (F3) of the current bill is added to the days in the previous month (G4) and then the day of the previous bill is subtracted (F4). An example is if the current bill arrived on July 7 and the previous bill on June 5. The 7 days in July plus the 30 days in June minus the 5 days in June on the previous bill results in 7+30-5 = 32 days. This is the length of time for flow in the row with the July 7 date.
8. Create a column that contains the month that the majority of each billing period occurred. This is done by calculating whether the majority of the days in the billing period occurred in the current month or the previous month. Example code is IF(F3/H3>0.5,E3,E4), where F3 is the day of the billing period, H3 is the number of days in the billing period, E3 is the current month, and E4 is the previous month. An example is if the bill arrived on July 7 and the billing period is 32 days long, then the majority of the billing period occurred in June, which would correspond to E4 from the code.

9. Create a column to calculate the monthly flow based on the based on the days in the billing period and the month of the flow. This requires two steps. First create a column of days corresponding to the month column created in step 8. This can be done in the exact same way the day column was created in step 6. The flows from the data set are in tens of gallons. To create the column of monthly flows based on the length of the period and the month of the flows, multiply the flows by 10, divide by the billing period length, and multiply by the days in the month. The result in the monthly flow for the month of the billing period. An example is flow(from data) * 10 / billing period (step 7) * days in billing month (step 9 part A).
10. Create column to identify low flows deemed to low to be contributing wastewater. Use the column from step 9 containing the flow per month to create an if statement that inputs zeros for any values less than 900 gal/month.
11. Copy and paste the customer id column, the column of months from step 8, and the monthly flow column from step 10 into a new worksheet.
12. Based on the zeros in the column of the monthly flows, use J-walk add-in to delete these rows. This removes all exceedingly low flow data from the data set.

High Outlier Removal

1. Sort the three columns of customer id, month, and flows by months in ascending order.

2. Copy and paste flows from November, December, January, February, and March into a column. These are the winter flows.
3. Calculate the average and standard deviation of the winter flows. Calculate the high outlier criterion, which is three standard deviations above the average.
4. Copy and paste the customer id, month, and flows. Sort by flows in ascending order.
5. Create a column with a conditional statement that inputs zeros for flows that are higher than the high outlier criterion. Since the list is sorted by flows in ascending order all of the zeros will appear at the end of the list.
6. Copy and paste the customer ids, months, and flows that are not high outlier data points into a new worksheet.

Monte Carlo Simulation

1. Sort customer id, months, and flows by customer id in ascending order.
2. Count the number of data points each residence has using the following code: =IF(A3<>A2,1,D2+1), where column A is the customer id and column D is the column that the counting occurs (*i.e.* column D is where the code belongs).
3. Sum the flows for each residence using the following code: =IF(D3>D2,C3+E2,C3), where column D is the number of data points, column C is the column of flows, and column E is the column that the summing occurs (*i.e.* column E is where the code belongs).

4. Calculate the average monthly flow by dividing the column of sums (E) by the column of data points (D). Use the following code: =IF(E3>E4,E3/D3,0). This code will result in only one average monthly flow for each residence. The rest of the rows for each residence will be zeros.
5. Create a column that has average monthly flows corresponding to every data point using the following code: =IF(F3=0,G4,F3), where column F is the column from step 4 and column G is where the code belongs. This will copy the average monthly flow for each residence to all of the data points for the residence.
6. Copy and paste the customer id column, month column, flow column, and column from step 5. Sort these columns by month in ascending order.
7. Calculate peaking factors by dividing the flow column by the average monthly flow column.
8. Copy the customer id column corresponding to the average monthly flow column from step 4 and the column from step 4. Paste these two columns into a new worksheet. Use J-walk add-in to delete the rows corresponding to zeros in the column from step 4.
9. Copy the resulting two columns from step 8 and paste into the worksheet being used in step 7. This is a list of residences and average monthly flows.
10. Create a column of the number of residences desired for simulation. Start by using 70. This column should contain the numbers 1 to 70.

11. Create a column of expected average monthly flows for each residence. It does not matter what this number is as long as it is just a number and not a formula. Use 6000 gal/month. This column should have 70 entries, one for each residence from step 10.
12. Create columns for January through December with ones as the values in each of these cells corresponding to the 70 residences. A matrix should be visible now that has residence numbers in the left most column, average monthly flows in the next column, and peaking factors (ones) in the next twelve columns corresponding to the months.
13. If Crystal Ball is not open at this point, save and open the file in Crystal Ball. Click on the first cell of the average monthly flow column from step 11.
14. Define the probability distribution used to sample the average monthly flows. This is done by either clicking the left most icon on the Crystal Ball tool bar or by going to Define → Define Assumption. Click Fit in the Define Assumption dialog. Click range of data to select a range of data from the Excel spreadsheet. Select the list of average monthly flows from step 9. Hit enter twice to accept the data range and the default assumption options. A new dialogue window opens showing the distributions and fit statistics that correspond to the data. Select a distribution with an acceptable fit. The logistic distribution seemed to work well. After selecting a distribution the next dialogue window shows the distribution and the fit parameters. Click to expand the window to view the

bounds of the distribution. Enter 900 as the lower bound of the distribution. Press ok to accept this assumption. As a default the cell now turns green.

15. Copy the assumption (green cell) from step 14 to the rest of the cells in that column. Copy by selecting the Copy Data icon on the Crystal Ball tool bar or by selecting Define → Copy Data. Highlight the non-green cells in the column, then select Paste Data from the Crystal Ball tool bar or Define → Paste Data. The highlighted cells will all turn green and be assumption variables for the average monthly flow.
16. Select the first cell for the column corresponding to January; this cell is next to the first green cell from the average monthly flow column. Define the assumption for January using the same method from step 14. The data that should be selected are the peaking factors corresponding to the month of January; this is easy since the peaking factors have already been sorted by months. Again the logistic distribution works well. This time set the lower bound of the distribution to zero.
17. Copy the assumption from step 16 and paste it into the rest of the cells in the column.
18. Perform steps 16 and 17 for the remaining months.
19. Copy and paste the column of residence numbers (1-70). Paste the column to the right of the December column.
20. Create columns for January through December next to the column in step 19.

21. In the month columns from step 20 multiply the average monthly flow for each residence by the peaking factor for that month. Remember when clicking and dragging the formulas to keep the average monthly flow constant for each residence. This results in a matrix of monthly flows for each residence.
22. Copy and paste the column of residence numbers (1-70). Paste the column to the right of the December column with monthly flows from step 21.
23. Create columns for January through December next to the column in step 22.
24. Decide what size (number of residences) systems are of interest. A suggestion is to use 1 through 10, 15, 20, 25, 30, 40, 50, 60, and 70. For the rows corresponding to this number of residences input a formula that sums the cells of monthly flows and divides by the number of residences (sum January flows and divide by the number of residence corresponding to the number of flows in the sum). This yields the system flow on a per residence basis. The monthly flows that are being summed are the flows in the columns from step 21. Sum down columns to avoid mixing data from different months. This step should result in values only in the rows corresponding to the number of residences of interest (*i.e.* 1 through 10, 15, 20, 25, 30, 40, 50, 60, and 70). Every month in these rows should have a value.

25. The values of monthly flow per residence from step 24 can be modified using a factor to predict the amount of water usage that will become wastewater. The factor used is 80%. This factor can be multiplied now to all of the monthly flow per residence values or it can be taken into account later when analyzing the simulation output. Multiplying now is useful if 80% is the only factor being used. If a set of factors are used, it is easier to address the issue in the analysis of the output.
26. Select the cell that corresponds to the first monthly flow for January (step 24). Define this cell as a forecast by clicking the forecast icon (third from left on Crystal Ball tool bar) or by selecting Define → Define Forecast. The forecast dialogue will open. Expand the forecast dialogue to view all of the options. In the Forecast Window tab select show window "When simulation stops". In the Precision, Filter, and Auto Extract tabs make sure nothing is selected. Click ok.
27. Copy the forecast cell from step 26 and paste it into all of the cells containing values for the monthly flows per residence.
28. Select the run preferences icon from the Crystal Ball tool bar or select Run → Run Preferences. In the Trials tab input the number of trials to be 10000. Check (select) the "Stop on calculation errors" option. In the sampling tab select Monte Carlo; do not select "Use same sequence of random numbers". In the Speed tab select Extreme speed and select "Suppress chart windows (fastest)". In the Options tab only select "Warn if insufficient memory", "Show control panel", and "Leave control panel open

on reset". Deselect all other options. In the Statistics tab select "Probability below a value" and "10%, 90%, etc." Do not select anything else. Click ok.

29. Click the Start Simulation icon (play button) or select Run → Start Simulation. Immediately minimize Excel; this will increase the speed of the simulation. The simulation should take approximately 20 seconds.
30. On the control panel select Analyze → Extract Data. In the Data tab, select Forecasts All and Assumptions None. Also select only the check box for Percentiles. When clicking this box a dialogue window will open. In this window, select the custom setting and input the percentiles of interest. Suggested percentiles of interest are 99.99 to 99.9 in increments of 0.01 then 99.9 to 99 in increments of 0.1 then 99 to 80 in increments of 1. In the options tab select Current workbook and New sheet. Input a name and starting cell for the new sheet. Check the boxes for Include labels and AutoFormat. Click ok.
31. Note that 99.99 percentile corresponds to 0.01 percent risk of failure.
32. In the new worksheet containing the percentiles and monthly flows per residences, create a method of inputting a design value and finding the associated risk from the percentile values for each set of residences. Use conditional logic.
33. Plot the risk against the number of residences using a log scale.

Vita

Patrick Andrew Dobbs began teaching recitation and laboratory classes for the Engineering Fundamentals Division of the College of Engineering at the University of Tennessee in 2003. He has worked on projects studying decentralized wastewater design, infiltration under pervious concrete, automated greenhouse sprayer systems, and audio detection of insects. He received a B.S. in Biosystems Engineering in 2005 and an M.S. in Biosystems Engineering in 2007, both from the University of Tennessee.