UNIVERSITY OF CINCINNATI

Date: 07/19/2005

I, Nabin Khanal,

hereby submit this work as part of the requirements for the degree of:

Master of Science

in:

Environmental Engineering

It is entitled:

Generalized Sensitivity Analysis of Water Distribution System Vulnerability to Deliberate Intrusions

This work and its defense approved by:

Chair: Dr. Steven G. Buchberger

Dr. Robert M. Clark

Dr. Walter M. Grayman

Dr. Sean A. McKenna

Dr. James G. Uber

Generalized Sensitivity Analysis of Water Distribution System Vulnerability to Deliberate Intrusions

A thesis submitted to the

Division of Graduate Studies and Research
of the University of Cincinnati

in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in the Department of Civil and Environmental Engineering
of the College of Engineering

2005

By

Nabin Khanal

B.S. in Civil Engineering,
Institute of Engineering, Tribhuvan University, NEPAL, 2001

Committee Chair: Dr. Steven G. Buchberger

Generalized Sensitivity Analysis of Water Distribution System Vulnerability to Deliberate Intrusions

ABSTRACT

This thesis presents a two-part investigation on the vulnerability of municipal water distribution systems to deliberate biochemical intrusions. In part 1, intrusions were modeled as a steady 6-hour injection of a soluble conservative contaminant into a randomly selected node on the pipe network of a small town. The propagation of the contaminant through the water distribution system was tracked with EPANET and, at the end of 72 hours, the fraction of the town's population exposed to the contaminant was estimated. This was repeated for multiple injection nodes randomly dispersed across the network. A dimensionless "Exposure Index" (EI) was introduced as a simple global measure of network vulnerability: an EI value of 0 implies that no residents are exposed to the contaminant; an EI value of 100 implies that all residents are exposed. In addition, results of the intrusion simulations were used to construct a "Zone of Influence" map which categorizes the network nodes on the basis of their exposure potential. In part 2, a Generalized Sensitivity Analysis (GSA) was performed to identify which, if any, of four dynamic network variables (base demand, storage capacity, mass loading and injection duration) had a significant influence on the percentage of population exposed to the contaminant during an intrusion. Latin Hypercube Sampling was used to set-up 1152 biochemical assault simulations at each of three injection nodes. The nodes were selected on the basis of their exposure potential (high, medium or low) as indicated on the Zone of Influence map. Based on the Kolmogorov-Smirnov *d* statistic, simulated exposure levels were found to be most sensitive to variations in base demand and mass loading. Tank capacity and injection duration tended not to be important. More work is needed, however, to identify appropriate measures of population exposure. This exercise demonstrates that GSA holds promise as a robust tool for streamlining computationally intensive network simulation experiments.

ACKNOWLEDGMENTS

First of all I would like to express my gratitude to my advisor, Dr. Steven Buchberger, for his guidance and support during my Master's degree work. Dr. Buchberger was always there to motivate and encourage me to work independently, and enhance my organizational and analytical skills. I will always appreciate his assistance for me to grow and develop as a researcher, engineer, and professional.

I would like to thank my committee members, Dr. Robert Clark, Dr. Jim Uber, Dr. Sean McKenna, and Dr. Walter Grayman for their insights and suggestions on my research. I am thankful to my senior graduate researcher Kenneth Nilsson for guiding me through the initial stages of my modeling setup, and to 'big brother' of the group, Zhiwei Li for his insightful suggestions and tips on debugging those C++ codes. I would also like to acknowledge the Sandia National Laboratories for providing the financial aid for this research through contract 60998.

I appreciate the memorable time I have with my group mates Xiaoyi Zhang and Steven Reese, and their willingness to help me with my research work. At this point I would also like to thank rest of my friends Prabhat Lamichhane, Ujjwal Gajurel, Balram Gurung, Sirish Agarwal who were supportive throughout and were there to drag me out of the occasional research blues. Lastly, with deep respect and devotion I would like to thank my parents (Narayan and Sushila) and siblings (Sushma and Suraj) for always encouraging and supporting me to fulfill what I aimed in life.

TABLE OF CONTENTS

Section **Page**

TABLE OF CONTENTS……..…I

LIST OF TABLES ……..……....IV

LIST OF FIGURES……..…V

LIST OF ABBREVIATIONS…….VII

1. INTRODUCTION…….....1

1.1 Problem Statement……....1

1.2 Study Objectives……...3

2. LITERATURE REVIEW…….….5

3. EXPERIMENTAL METHODOLOGY……8

3.1 Simulation Network……...8

3.2 Base-Case Attack…….….10

3.3 Assumptions…….…….11

3.4 Nodal Mass Calculation…….……..12

3.5 System Mass Balance…….…...13

3.6 Spatial Variability in Network Intrusion……..….16

3.7 Vulnerability Study and Zone of Influence……..19

3.8 Measure of Network Vulnerability……...22

3.9 Influences on Exposure Index……..….24

4. GENERALIZED SENSITIVITY ANALYSIS……26

4.1 Introduction…….....26

4.2 Assumptions…….27

4.3 Sampling and Simulation…….29

4.4 Node Count Calculation……...31

4.5 GSA with Kolmogorov-Smirnov (K-S) *d* statistic……...32

5. RESULTS…….36

5.1 Likelihood of Node Selection for Intrusion…….……36

5.2 Zone of Influence…….38

5.3 Exposure CDF and Exposure Index…….42

5.4 Results for deterministic variation in network variables……..45

5.5 Results for stochastic variation in network demand…….49

5.6 Generalized Sensitivity Analysis of intrusions in CH/BP ……...51

5.6.1 Average Concentration Threshold …….53

5.6.2 Nodal Mass Threshold…….67

6. SUMMARY AND CONCLUSIONS……...77

REFERENCES…….80

APPENDICES……..84

APPENDIX I……..……..85

1. EPANET input file for Base-Case Attack at node 50

2. PRPsym input file for Base-Case Attack at node 50

3. EPANET input file generated by PRPsym for Base-Case Attack at node 50

APPENDIX II……....……114

1. Program I: Percentage nodal mass consumption

2. Program II: PRP simulation of water demands

3. Program III: Latin Hypercube Sampling within EPANET execution interface

4. Program IV: Kolmogorov-Smirnov two-sample test

APPENDIX III……..……..143

1. GSA algorithm

2. Input vector sets generated by GSA program

APPENDIX IV……155

1. Exposure CDFs for 18 deterministic simulations

LIST OF TABLES

Table 1: Variables thought to be important in network assault simulation.

Table 2: Default EPANET input parameters for simulation.

Table 3: System mass distribution for Base-Case injection after 72-hr simulation.

Table 4: Percentage population associated with each network node.

Table 5: Exposure range in each zone for CH/BP Zone of Influence map.

Table 6: Four uncertain input parameters selected from dynamic variable group.

Table 7: Result matrix for chemical mass consumed at nodes.

Table 8: Percentage population exposed to chemical above mass thresholds.

Table 9: Summary statistics for percentage population exposed above mass thresholds.

Table 10: Exposure Index for different deterministic injections scenarios.

Table 11: Exposure Index for 100 Monte Carlo simulation of network demand.

Table 12a: K-S test statistics for injection made at node 17 (Red Zone).

Table 12b: K-S test statistics for injection made at node 6 (Yellow Zone).

Table 12c: K-S test statistics for injection made at node 81 (Green Zone).

Table 13: Sensitivity ranking of dynamic variables based upon K-S d statistic for average chemical concentration threshold of 2.0 mg/l.

Table 14: K-S test statistics for injection made at node 17 (Red Zone) with nodal consumed as a threshold; 72-hour simulation.

Table 15: K-S test statistics for injection made at node 17 (Red Zone) with nodal mass consumed as a threshold; 24-hour simulation.

Table 16: K-S test statistics for injection made at node 17 (Red Zone) with average concentration as a threshold; 24-hour simulation.

LIST OF FIGURES

Figure 1: All-pipes network model of Cherry Hill/Brushy Plains WDS.

Figure 2: Base-Case injection represented in total system demand time series plot.

Figure 3a: System mass distribution for network with CSTR tank mixing.

Figure 3b: System mass distribution for network with LIFO tank mixing.

Figure 3c: System mass distribution for network with FIFO tank mixing.

Figure 4: 45 network nodes selected for injection in CH/BP network.

Figure 5: An illustrative example of different types of Exposure CDFs.

Figure 6: Flow Chart for GSA with K-S *d* statistic.

Figure 7: An illustration of K-S two sample test.

Figure 8: Likelihood of a node being selected for intrusion.

Figure 9: Zone of Influence map for CH/BP water distribution network.

Figure 10: Exposure CDF for Base-Case intrusion in CH/BP.

Figure 11: Effect of total simulation period on the Exposure CDF for Base-Case attack made on CH/BP.

Figure 12: Exposure CDF for Base-Case intrusion in CH/BP network with stochastic water demand.

Figure 13: CDF for uniform distribution of Base Demand.

Figure 14a: Variation in Sensitivity Ranking with respect to average concentration threshold value when injection is made at node 17 (Red zone).

Figure 14b: Variation in Sensitivity Ranking with respect to average concentration threshold value when injection is made at node 6 (Yellow zone).

Figure 14c: Variation in Sensitivity Ranking with respect to average concentration threshold value when injection is made at node 81 (Green zone).

Figure 15a: Highest and Lowest ranked dynamic variables, for injection made at node 17 lying in Red zone and threshold of 2.0 mg/l.

Figure 15b: Highest and Lowest ranked dynamic variables, for injection made at node 6 lying in Yellow zone and threshold of 2.0 mg/l.

Figure 15c: Highest and Lowest ranked dynamic variables, for injection made at node 81 lying in Green zone and threshold of 2.0 mg/l.

Figure 16: Variation in Sensitivity Ranking with respect to nodal mass consumed threshold value when injection is made at node 17 (Red zone).

Figure 17: Highest and Lowest ranked dynamic variables, for nodal mass threshold of 216 gram, injection made at node 17 (Red zone).

Figure 18: Variation in Sensitivity Ranking with respect to and average concentration and nodal mass consumed threshold for 24-hour simulation (injection node 17).

LIST OF ABBREVIATIONS

ADR: Advection-Dispersion-Reaction

CDF: Cumulative Distribution Function

CH/BP: Cherry Hill/Brushy Plains

CSTR: Continuously-Stirred Tank Reactor

EI: Exposure Index

EPA: Environmental Protection Agency

FIFO: First In First Out

GSA: Generalized-Sensitivity Analysis

K-S: Kolmogorov-Smirnov

LIFO: Last In First Out

NC: Node Count

PRP: Poisson Rectangular Pulse

SDWA: Safe Drinking Water Act

TEVA: Threat Ensemble Vulnerability Assessment

WDS: Water Distribution System

1. INTRODUCTION

1.1 Problem Statement

Water distribution systems are designed to supply water for basic domestic purposes and for commercial and industrial uses. Most municipal water supply systems also provide flows necessary to fight fires. Water distribution networks are closed systems and, hence, they tend to be more secure than raw water sources. Nonetheless, water distribution networks are susceptible to various types of threats. Conceivable threats include the physical destruction of facilities or equipment, airborne release of hazardous chemicals stored onsite, sabotage of Supervisory Control and Data Acquisition instruments, and the introduction of chemical, biological, or radiological contaminants into the water supply (Murray *et al.,* 2004).

The potential ramifications that arise from contamination of a water distribution system can be severe because (except for residual disinfectant) no contingency exists to provide a barrier, network monitoring is quite limited, and the travel time between source and consumers can be quite short (Walski *et al*., 2003). Early warning devices like sensors can be used to detect contaminants at critical points of the network, but due to the spatial diversity of the drinking water network it is almost impossible to install sensors all around the network. As a complement to sensors, careful and realistic engineering analyses of a pipe network can help to identify service zones most vulnerable to attack, assess the extent of consumer risk, and prioritize plans for risk reduction.

The key variables in simulating network assaults can be broadly categorized into two groups, namely, *static* variables and *dynamic* variables. Static variables refer to properties of the network which ordinarily are not affected by human behavior. For example, pipe size is a static variable. In contrast, dynamic variables refer to features of the network that are controlled by the behavior of people. In a municipal pipe network, these people may include consumers, utilities, or assailants. There is some degree of uncertainty associated with all static and dynamic variables.

Table 1 lists many of the static and dynamic variables which are likely to be important in simulating a network assault. For purposes of simulation, the static variables can be considered deterministic. On the other hand, the behavior-driven dynamic variables (consumer demands, utility operations, and assault logistics) can be rather unpredictable. The stochastic nature of these dynamic variables should be incorporated into the simulation process to more accurately mimic the system behavior and to better understand the network response to potential intrusions.

Table 1: Variables thought to be important in network assault simulation.

Sources of Uncertainty or Variability in Simulating Network Attacks			
Dynamic Variables			**Static Variables**
Consumer	**Utility**	**Assailant**	**Deterministic**
End user type	Network operation	Attack duration	Mixing in tanks
Water demand	(source blending)	Attack timing	Pipe roughness
Exposure	(hydrant flushing)	Attack location	Pipe size (D and L)
Dose response	(pump schedule)	Attack loading	Transport (ADR)
	(tank operation)	Contaminant type	Skeletonization

Note: ADR = Advection-Dispersion-Reaction

In an early application of Monte Carlo techniques to water distribution systems, Nilsson *et al*. (2005) combined stochastic water demands with advective contaminant transport to simulate a biochemical assault launched from a *single* node in a small municipal pipe network. Nilsson's work suggests that, with thoughtful planning, desk-top simulations may offer a cost-effective surrogate for expensive field work. The problem considered in this thesis extends Nilsson's analysis to include attack nodes located through out the entire pipe network.

There is, however, a potential downside. Users can be overwhelmed by the amount of data generated during the simulation exercise. Hence, there is a need to summarize network simulation results in concise and intuitive ways for effective communication with water utilities. Further, considering the large number of computational permutations, it is desirable to determine which of the network variables are most important. The question is: which of the variables listed in Table 1 exert the largest influence on the simulation results and, consequently, deserve the closest scrutiny?

1.2 Study Objectives

This thesis demonstrates how uncertainties in key network variables affect the vulnerability of a water distribution system subject to deliberate intrusions and, further, identifies which of these variables are most critical for Monte Carlo simulations. There are two objectives. 1) To develop and demonstrate use of a dimensionless 'Exposure Index' as a simple intuitive measure of network vulnerability. Extending the work of Nilsson *et al*. (2005), the aim here is to examine the sensitivity of network vulnerability

to changes in base demands, to assumptions about tank mixing, and to variability in the location, timing, duration and intensity of the chemical assault. 2) To evaluate the suitability of Generalized Sensitivity Analysis (GSA) as a statistical tool for effectively identifying and ranking the dynamic network variables that are most significant for Monte Carlo simulations of network intrusions.

2. LITERATURE REVIEW

Water distribution systems are designed using heuristic criteria, but for the past few decades computerized simulation models have been the standard practice for designing water distribution systems (Gargano and Pianese, 2000). Implementation of the Safe Drinking Water Act (SDWA) has made network models an essential tool for predicting the system's ability to comply with water quality standards (Buchberger and Wells, 1996). Modeling techniques, in conjunction with monitoring data could be used to identify the source of the threat, where contaminants may move in the system and to suggest remediation approaches (Clark and Buchberger, 2004).

The driving force behind the hydraulic dynamics occurring in water distribution systems can be attributed to the consumer water demand (Walski *et al*. 2003). It is unrealistic to study a water distribution system with a deterministic approach, which assumes that the system is functioning properly with purely conventional demand conditions only. A probabilistic approach that allows distribution systems to be studied under a wide variety of loading conditions provides a more complete description of the system's behavior and constitutes the necessary approach for an objective evaluation of the system's hydraulic reliability (Gargano and Pianese, 2000).

Buchberger and Wu (1995) proposed that residential water demands occur as a non-homogeneous Poisson-Rectangular-Pulse (PRP) process, thus providing a stochastic model for indoor residential water demands as a starting point for quantifying the instantaneous temporal and spatial variability of flow through municipal water

distribution systems. Since then, results from field studies and research in Latina, Italy (Guercio *et al*. 2001), Cincinnati, Ohio (Buchberger *et al.* 2003) and Valencia, Spain (Garcia *et al*. 2004) have corroborated the validity of the PRP model for residential water demands in distribution networks.

The focus on water infrastructure security and consumer health impacts has increased since the Bioterrorism Act was introduced in 2002 (Bioterrorism Act, 2002). Propato and Uber (2004) performed vulnerability analysis of a network which systematically and quantitatively assessed distribution system vulnerability against microbial intrusions. Nilsson *et al*. (2005) studied the network vulnerability under influence of biochemical intrusion using Monte Carlo techniques to simulate the dynamic behavior of a network demand. EPA is developing the Threat Ensemble Vulnerability Assessment (TEVA) program, a probabilistic framework for assessing the vulnerability of a water utility to a large range of contamination attacks (Murray *et al*. 2004).

The uncertainties in dynamic network variables can be simulated using the Monte Carlo technique which can provide detailed knowledge of the system vulnerability. Knowing which input variables exert the most influence on the model outputs i.e., sensitivity analysis helps the analyst validate and verify the computer model and also identify those input variables that may require further investigation (Shortencarier and Helton, 1999).

Generalized Sensitivity Analysis (GSA), a technique developed by Spear and Hornberger (1980) within the context of environmental quality studies, uses sampling-based sensitivity analysis technique which can be used in network intrusion simulation, to identify the dynamic variables that have greatest influence on the total consumer exposure. In GSA, a multi-parameter Monte Carlo study is performed, sampling parameters from statistical distribution functions (Saltelli et al. 2004). James *et al*. (1996), Makino *et al*. (2001), and Cox and Whitehead (2005) have demonstrated the application of GSA in field of ground water modeling, nuclear waste disposal, and dynamic water quality modeling respectively. GSA will be used to perform sensitivity analysis of Monte Carlo simulation results using the Kolmogorov-Smirnov (K-S) *d* statistic in this study.

3. EXPERIMENTAL METHODOLOGY

3.1 Simulation Network

The all-pipes network model of Cherry Hill/Brushy Plains (CH/BP) water distribution system (WDS) lying in South Central Connecticut region is used as a network to conduct the hypothetical chemical intrusions. The network comprises of 103 links, 88 demand nodes, a pump, and a tank. The average base demand of system is 1227 Lpm and the mean residence time in the tank is 89 hours. Figure 1 shows the all-pipe model of CH/BP water distribution system.

The same network was used by Nilsson *et al*. (2005), but their study was limited to single intrusion (at node 50) and only 55 hr of simulation time. Multiple intrusion locations and an extended simulation time of 72 hr are used in this study to compile a more complete picture of network vulnerability at the CH/BP site.

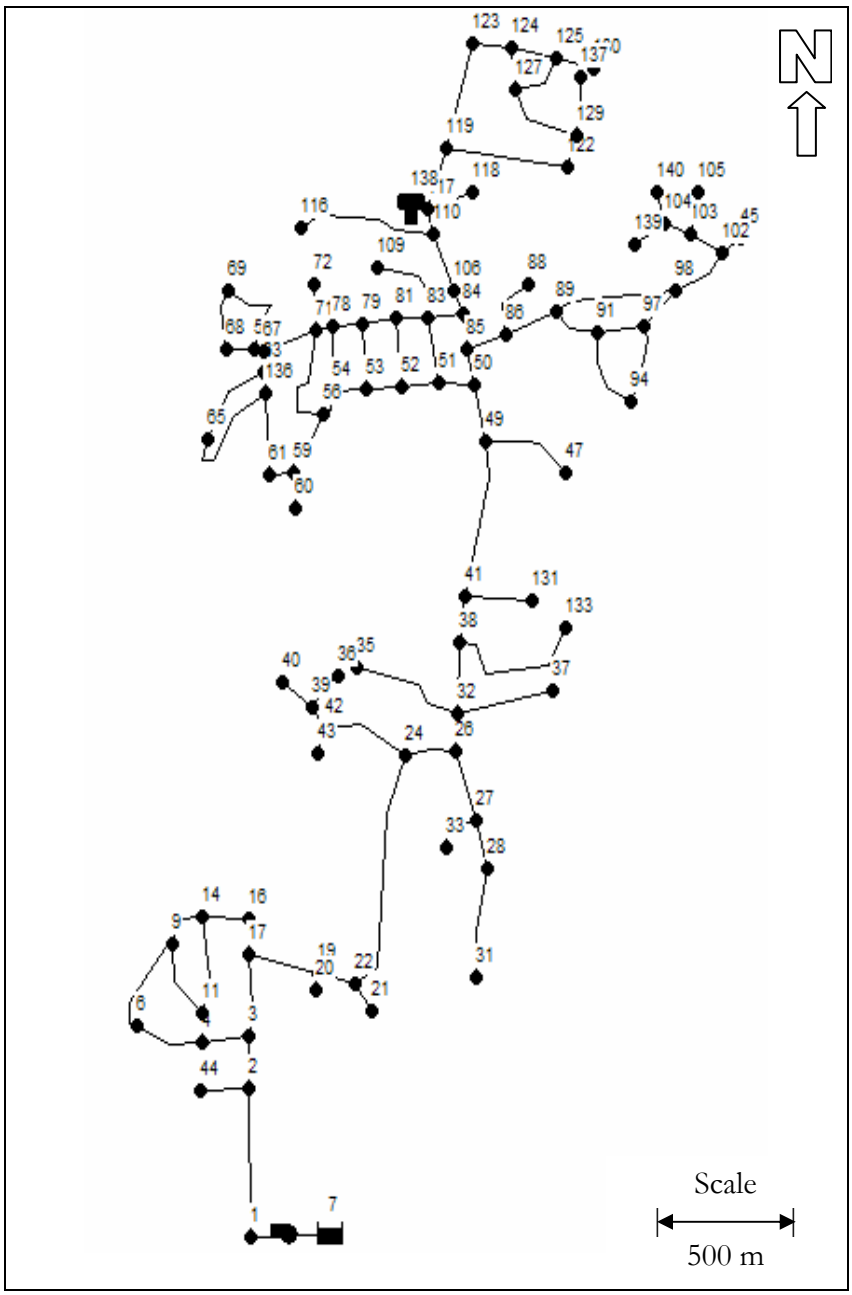

Figure 1: All-pipes network model of Cherry Hill/Brushy Plains WDS

3.2 Base-Case Attack

A Base-Case attack on the network is defined as the injection of 7200 g of a conservative chemical at a steady rate of 20 g/min, starting at 09:00 hour and lasting for 6 hours during the 72-hour simulation period. A conservative chemical is used for the injection, as it represents the worst-case contamination scenario. The mass booster source was used in EPANET to simulate the chemical injection into the network. Figure 2 shows the Base-Case injection in the time series plot of total system demand for the CH/BP network.

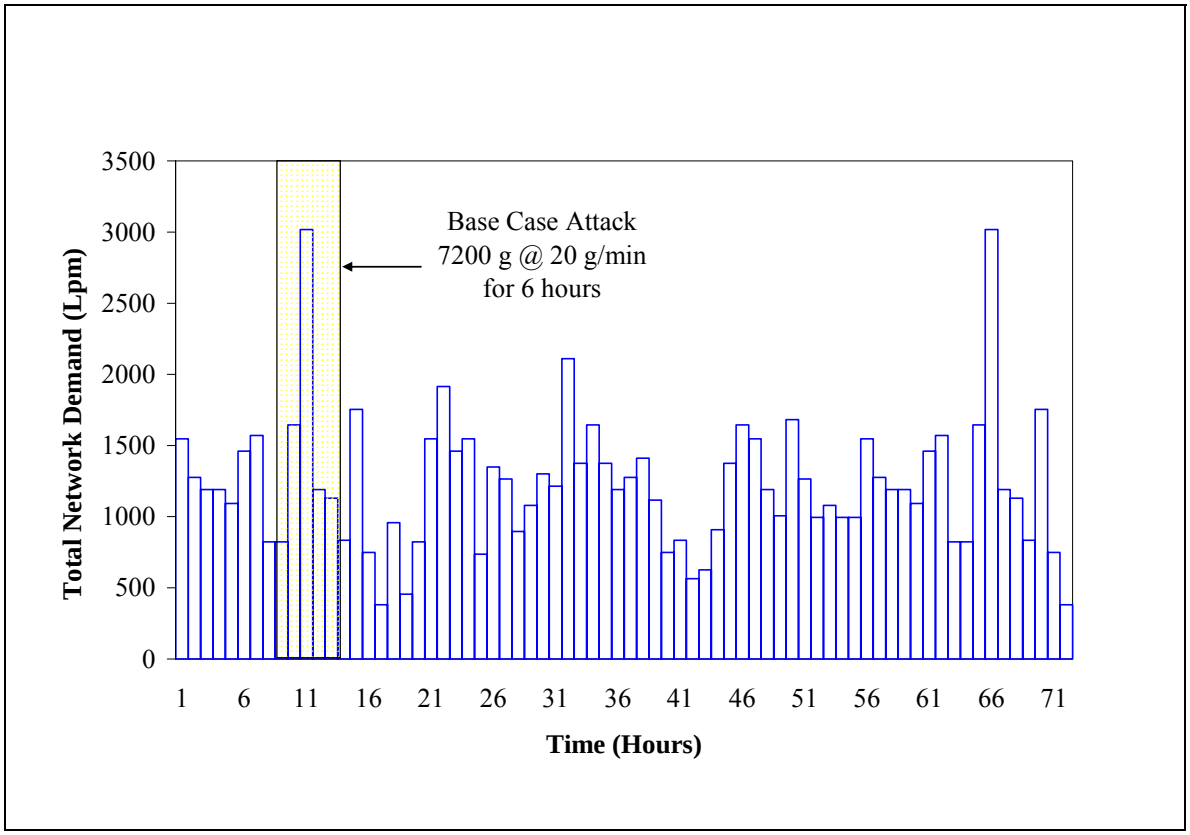

Figure 2: Base-Case injection represented in total system demand time series plot.

3.3 Assumptions

Various assumptions were made before the simulation was carried out. It was assumed that the percentage of population associated with each node was equivalent to the percentage of the base demand at that node. An assumption was made that all network nodes have equal probability of being selected for an injection irrespective of population associated with the nodes. This assumption holds true for simulation of network for maximum vulnerability and examination of this assumption is illustrated in section **5.1 Likelihood of node selection for intrusion**, of this study.

It was also assumed that the attack remained undetected throughout the 72-hour simulation period and any member of the population can be exposed at only one node, the one with which they are associated. A threshold value was selected for chemical mass consumed at node, which was assumed to determine the node vulnerability. For the Base-Case attack of 7200g of chemical, if a node received more than 36 g i.e. 0.5% of chemical mass within the 72-hour simulation period then the node was assumed to be vulnerable i.e. associated with consumer exposure. Since, the contaminant is a hypothetical chemical; there is no information about the minimum dose that causes consumer sickness or mortality. So, the arbitrarily selected 0.5% threshold value for chemical mass works well to categorize a node as vulnerable or non-vulnerable, for the purpose of this study.

3.4 Nodal Mass Calculation

For any contaminant intrusion scenario, EPANET toolkit features were used along with C++ codes, to calculate the total nodal mass consumption at each network node. The total chemical contaminant mass Ω_n [M] delivered to a node n at the end of the simulation period was calculated using the Equation 1:

$$\Omega_n = \sum_{i=1}^{4320} C_n(i) Q_n(i) \Delta t \qquad (1)$$

where i is the particular minute during the 72-hour simulation period, $C_n(i)$ is the contaminant concentration [M/L^3] entering the node n during i^{th} minute, $Q_n(i)$ is the corresponding nodal demand [L^3/T], and Δt [T] is the reporting time step (1 minute in this case). The C++ codes used for this calculation are presented in **Appendix II** under the heading of **Program I**. The default simulation parameters used in EPANET are summarized in Table 2.

Table 2: Default EPANET input parameters for simulation.

Features	Details
Total simulation period	72 hours
Simulation start time	9 AM i.e. 00:00 hr
Hydraulic time step	1 hour
Reporting time step	1 minute
Quality time step	1 minute
Minimum water level in tank	56 ft
Initial water level in tank	56.7 ft
Initial pump status	Open
Chemical properties	Conservative
Total chemical mass injected	7200 gram
Injection source type	Mass booster

3.5 System Mass Balance

A preliminary study was done in the network to ensure the system mass balance and veracity of the water quality model. Three storage mixing options in EPANET; (a) Completely Mixed Tank Reactor (CSTR), (b) Last In First Out (LIFO) Plug Flow, and (c) First In First Out (FIFO) Plug Flow were examined to assess the effect of tank mixing on the fate of chemical propagation in CH/BP network. A deterministic simulation was performed for each tank mixing option to simulate the Base-Case injection at node 50 lying in the trunk line of the network (see Figure 1). Evaluation of mass distribution among the pipe links, the demand nodes, and in the storage tank at the end of 72 hour simulation period verified the system mass balance for the network. Figures 3a, 3b, and 3c shows the system mass distribution plot for the CSTR, LIFO, and FIFO respectively when the Base-Case attack was made at node 50.

The EPANET input file used for simulating the network with CSTR tank mixing is presented in **APPENDIX I**.

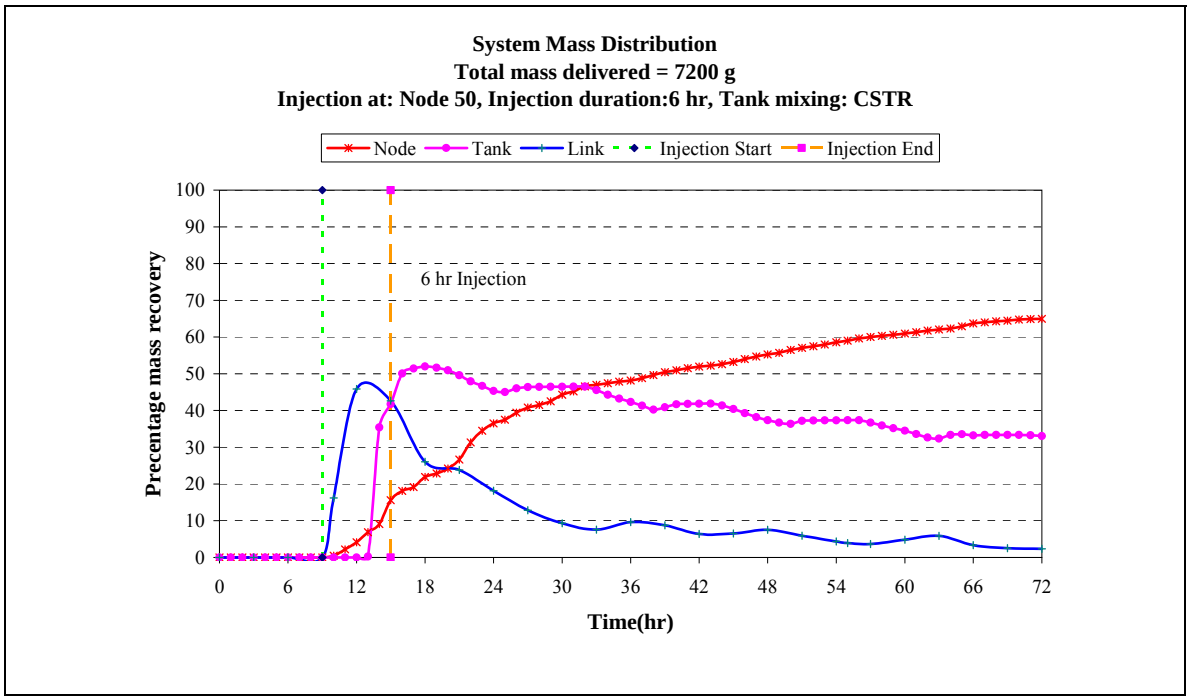

Figure 3a: System mass distribution for network with CSTR tank mixing.

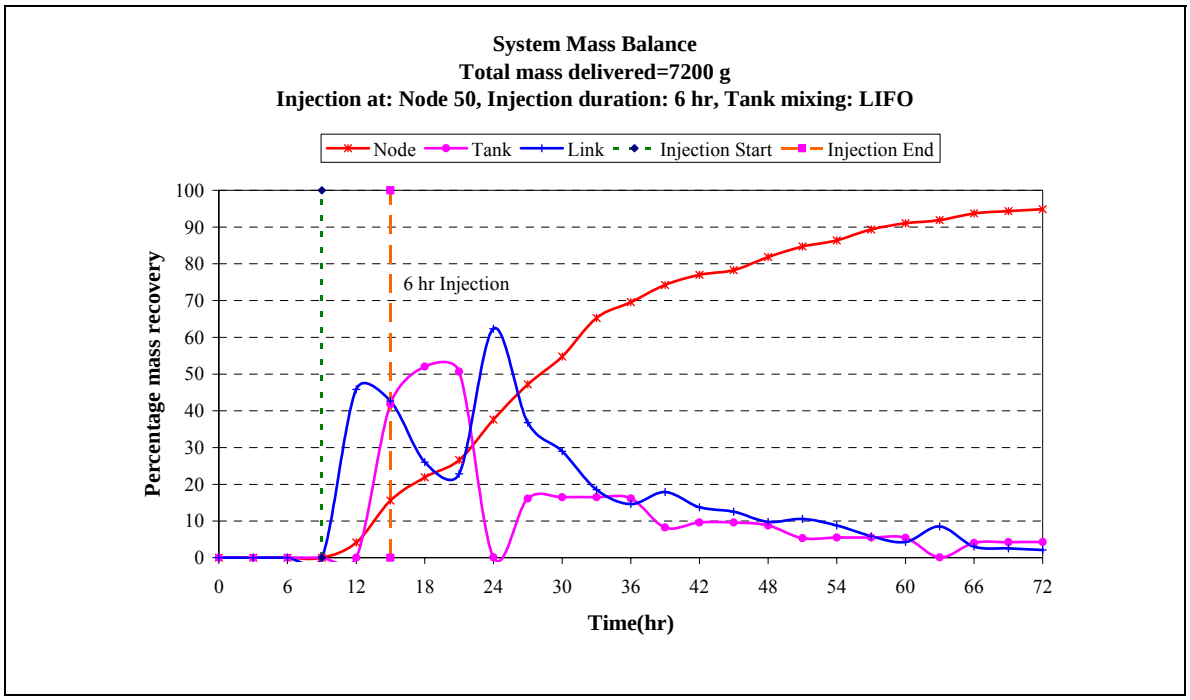

Figure 3b: System mass distribution for network with LIFO tank mixing.

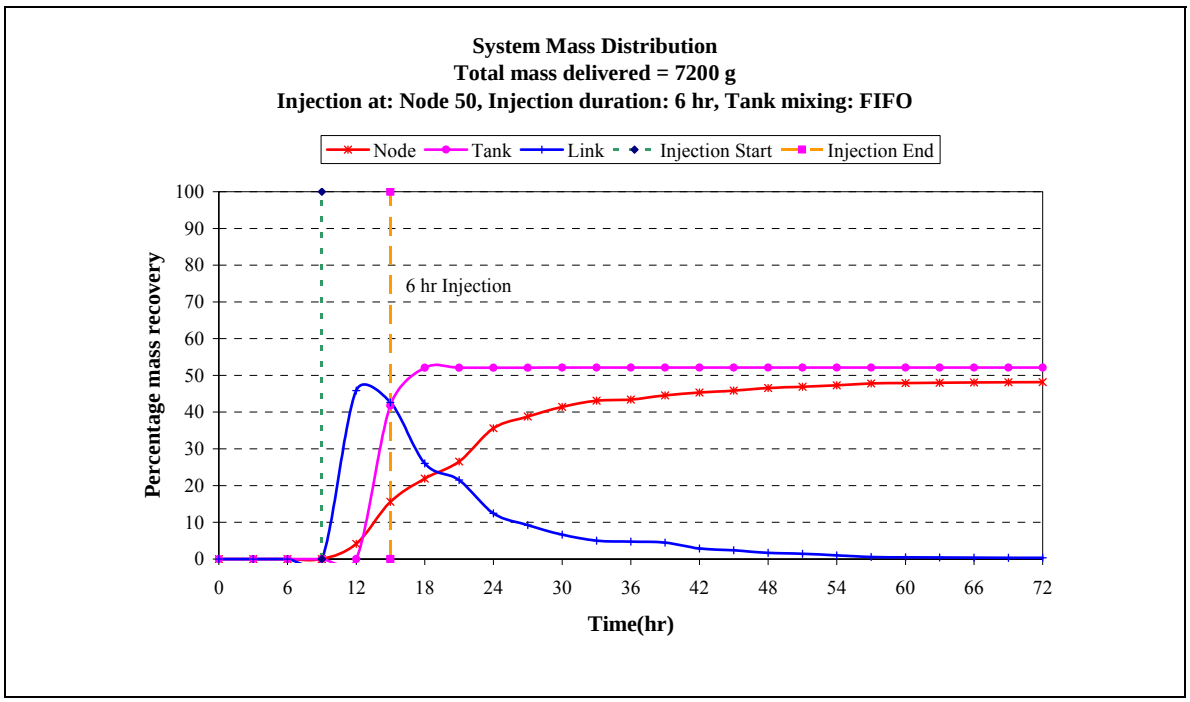

Figure 3c: System mass distribution for network with FIFO tank mixing.

Table 3 summarizes the system mass distribution for three tank mixing conditions considered.

Table 3: System mass distribution for Base-Case injection after 72-hr simulation.

Tank Mixing Option	Mass in links (%)	Mass in storage tank (%)	Mass extracted at nodes (%)	Time for 95% removal (hr)
CSTR	2	33	65	276
LIFO Plug Flow	2	4	94	80
FIFO Plug Flow	1	51	48	160

For FIFO tank mixing, tank played a vital role of retaining more than 51% of mass during the simulation period. The residence time in tank was 89 hours. Any chemical mass that entered the tank with FIFO mixing has to spend this time inside the tank. For LIFO tank mixing almost all chemical mass was consumed at nodes, as tank provided least role in retaining the mass injected. For the CSTR tank mixing, 33% of mass was retained at tank and population was exposed to only 65% of the injected mass.

Clark *et al.* (1996) studied mixing in three different tanks using the compartment model. Model results compared to observed effluent from the CH/BP tank showed that the one-compartment assumption (CSTR) was a good fit for the CH/BP tank. Even though the tank acted as a major component to retain chemical mass for FIFO mixing, simulation continued for extended period showed that it takes only 160 hours for 95% mass removal whereas the same mass was consumed only after 276 hours for CSTR tank.

3.6 Spatial Variability in Network Intrusion

In order to incorporate spatial variability in the attack rather than just one intrusion node, multiple nodes were considered for intrusions. 45 out of 89 CH/BP network nodes (51% of the total) were selected for injection (see cyan colored nodes in Figure 4). These potential injection locations include the storage tank, and all other consumer nodes, except redundant nodes on looping links and some dead-end nodes. The network consists of five zero demand nodes, nodes 1, 28, 32, 38 and 67.

Node 32 lies in main trunk line and it has zero base demand associated with it. Even the nodes have no demand associated with them; four of the zero demand nodes (nodes 1, 32, 38, and 67) were included as injection nodes. Injection at zero-demand nodes allows studying the influence of these zero demand nodes in network vulnerability and examining how the contaminant propagation is likely from any network nodes.

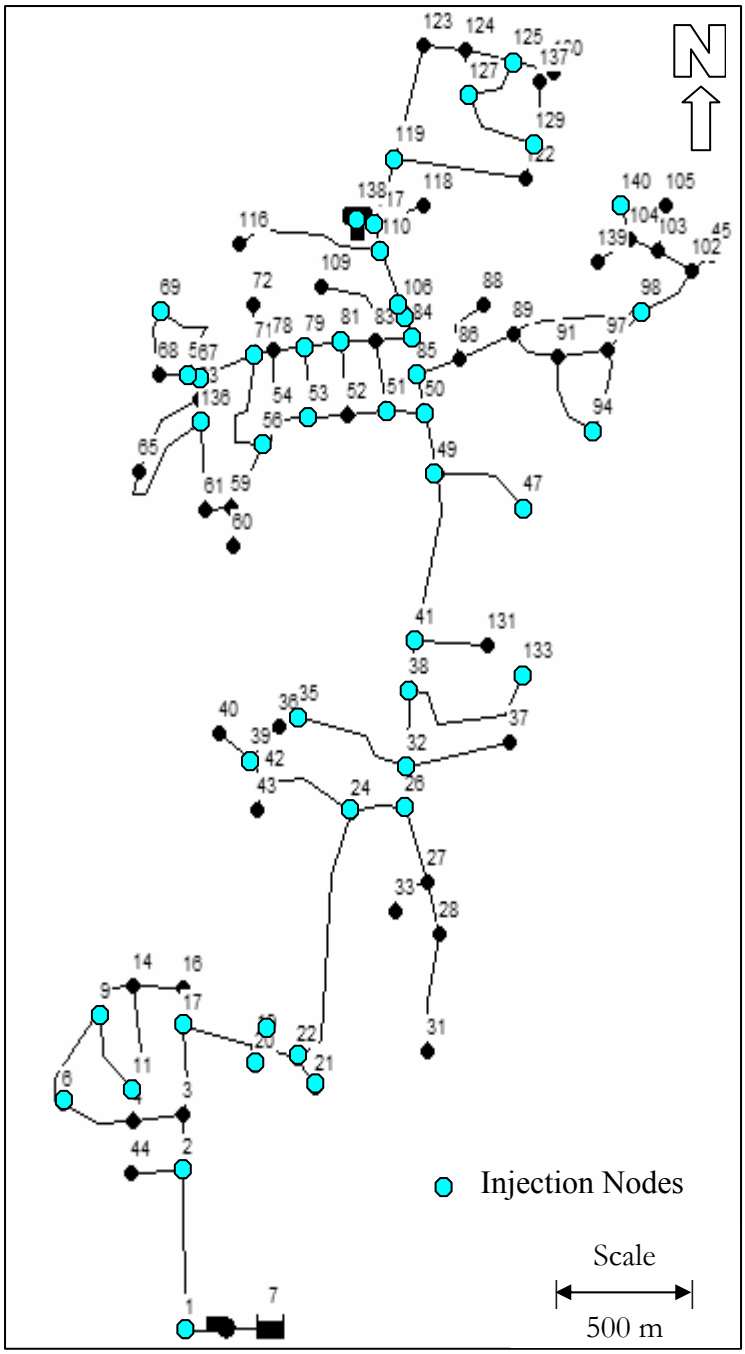

Figure 4: 45 network nodes selected for injection in CH/BP network.

3.7 Vulnerability Study and Zone of Influence

During a simulated network assault, some nodes may receive a high fraction of chemical, others may receive a nominal fraction and some may receive less than a trace of chemical during the 72 hour period. Similarly, an injection made at some nodes can reach higher fraction of population, injection at other nodes may reach to nominal fraction of population and injection at some nodes may have very less impact on to the population. A network can be divided into different zones based upon the injection node's ability to reach the consumers.

A vulnerability study of the CH/BP distribution network was performed by simulating individual Base-Case chemical intrusions at all 45 selected injection nodes in the network to identify the zones where intrusions would be most effective. The tank was assumed to behave as a CSTR and calculation for nodal mass loading at each network node was performed. A threshold nodal mass value of 0.5% of total mass injected i.e., 36 gram, was selected arbitrarily with an assumption that any node receiving chemical mass above that threshold will be deemed vulnerable i.e., it will have consumer exposure associated with it. The percentage population associated with each node is shown in Table 4.

For an injection at particular node, nodes receiving chemical mass above the threshold value were recognized. Then the percentage population associated with those nodes was summed to calculate a total percentage population exposure value for that injection node. In this manner, the percentage population exposure value was calculated for each of the 45 injection nodes.

Table 4: Percentage population associated with each network node

Node	Base Demand Lpm	% Population	Node	Base Demand Lpm	% Population
1	0.00	0.00	63	21.04	1.71
2	18.81	1.53	65	37.85	3.08
3	21.65	1.76	67	0.00	0.00
4	12.79	1.04	68	16.81	1.37
5	20.59	1.68	69	33.65	2.74
6	15.63	1.27	71	33.65	2.74
9	19.87	1.62	72	5.68	0.46
11	14.19	1.16	78	5.68	0.46
14	14.19	1.16	79	8.52	0.69
16	5.68	0.46	81	9.95	0.81
17	7.12	0.58	83	8.52	0.69
19	4.28	0.35	84	2.84	0.23
20	8.52	0.69	85	5.68	0.46
21	8.52	0.69	86	5.68	0.46
22	9.95	0.81	88	11.36	0.93
24	15.63	1.27	89	22.71	1.85
26	17.41	1.42	91	18.47	1.51
27	13.13	1.07	94	15.63	1.27
28	0.00	0.00	97	11.36	0.93
31	26.27	2.14	98	18.47	1.51
32	0.00	0.00	102	8.52	0.69
33	26.27	2.14	103	7.12	0.58
35	25.55	2.08	104	5.68	0.46
36	2.84	0.23	105	11.36	0.93
37	14.19	1.16	106	2.84	0.23
38	0.00	0.00	109	42.05	3.43
39	1.44	0.12	110	1.44	0.12
40	2.84	0.23	116	42.05	3.43
41	11.36	0.93	117	2.84	0.23
42	5.68	0.46	118	14.19	1.16
43	2.84	0.23	119	17.03	1.39
44	2.84	0.23	122	12.79	1.04
45	5.68	0.46	123	8.52	0.69
47	23.13	1.88	124	11.36	0.93
49	11.36	0.93	125	8.52	0.69
50	5.68	0.46	127	17.03	1.39
51	10.64	0.87	129	12.79	1.04
52	9.24	0.75	130	5.68	0.46
53	7.12	0.58	131	46.25	3.77
54	5.68	0.46	133	80.96	6.60
56	50.45	4.11	136	25.25	2.06
59	25.25	2.06	137	5.68	0.46
60	8.40	0.68	139	2.84	0.23
61	16.81	1.37	140	11.36	0.93

A contour plot of population exposure for the network was developed using the 45 population exposure values at 45 injection nodes. Network nodes were then divided into three zones, Red, Yellow, and Green based upon their ability to reach the consumers with chemical mass above the 36 g threshold value. The non-injection nodes also fall in one of the three zones in the contour plot; these nodes shared the same percentage population value represented by the zone they lie on. The red zone contains all injection nodes that delivered at least 36 g of chemical mass to at least 60% (high fraction) of consumers in the CH/BP network. Nodes in Yellow zone will have less impact on consumer (nominal fraction) than nodes in Red zone but more consumer impact than nodes in Green zone. Each network node will lie on one of these three zones, thus defining a Zone of Influence map for the network. Table 5 summarizes the range for each zone in Zone of Influence map for CH/BP.

Table 5: Exposure range in each zone for CH/BP Zone of Influence map.

Zone	**% population exposed above 0.5% threshold**
Red	> 60
Yellow	30-60
Green	< 30

3.8 Measure of Network Vulnerability

For the Base-Case injection made at 45 nodes, the information on percentage population exposed above the 36 g threshold value was used to develop an Exposure CDF. The Exposure CDF is a plot with percentage population exposed greater than the threshold on the X-axis and the CDF of percentage population exposed on Y-axis. This type of plot was introduced by Propato and Uber (2004) in their study of network vulnerability to pathogen intrusion.

The behavior of CDF trajectory provides information on network vulnerability. Utilities would like to have no effect on consumers from any intrusion in the network. In this case, mean population exposed is 0% and the CDF trajectory will follow a vertical path along left end of Y-axis i.e. a vertical line at 0% value of X-axis. But a determined assailant would like to have maximum effect on consumer for every intrusion made; the CDF trajectory in this case will follow a vertical path along right end of Y-axis i.e. a vertical line at 100% of X-axis and mean population exposed in 100%. A real network will fall somewhere between these two extremes; hence the Exposure CDF will follow a trajectory starting from the lower left corner and ending in the upper right corner of the plot.

The CDF trajectory defines two regions: (a) the region below the curve whose area represents the average population percentage that is "safe" (i.e., receiving chemical mass that is below the threshold value) and (b) the region above the curve whose area gives the

mean percentage of the total network population that is exposed to chemical mass above the threshold value.

The **Exposure Index (EI)** is the value of mean percentage population exposed above the threshold. A utility would like to ensure an Exposure Index of 0 whereas an assailant would target for Exposure Index of 100. Thus, the Exposure Index provides a simple measure of network vulnerability. Figure 5 shows an illustrative example of Exposure CDF desired by utilities and assailants, and their comparison with the real network Exposure CDF along with respective Exposure Indices.

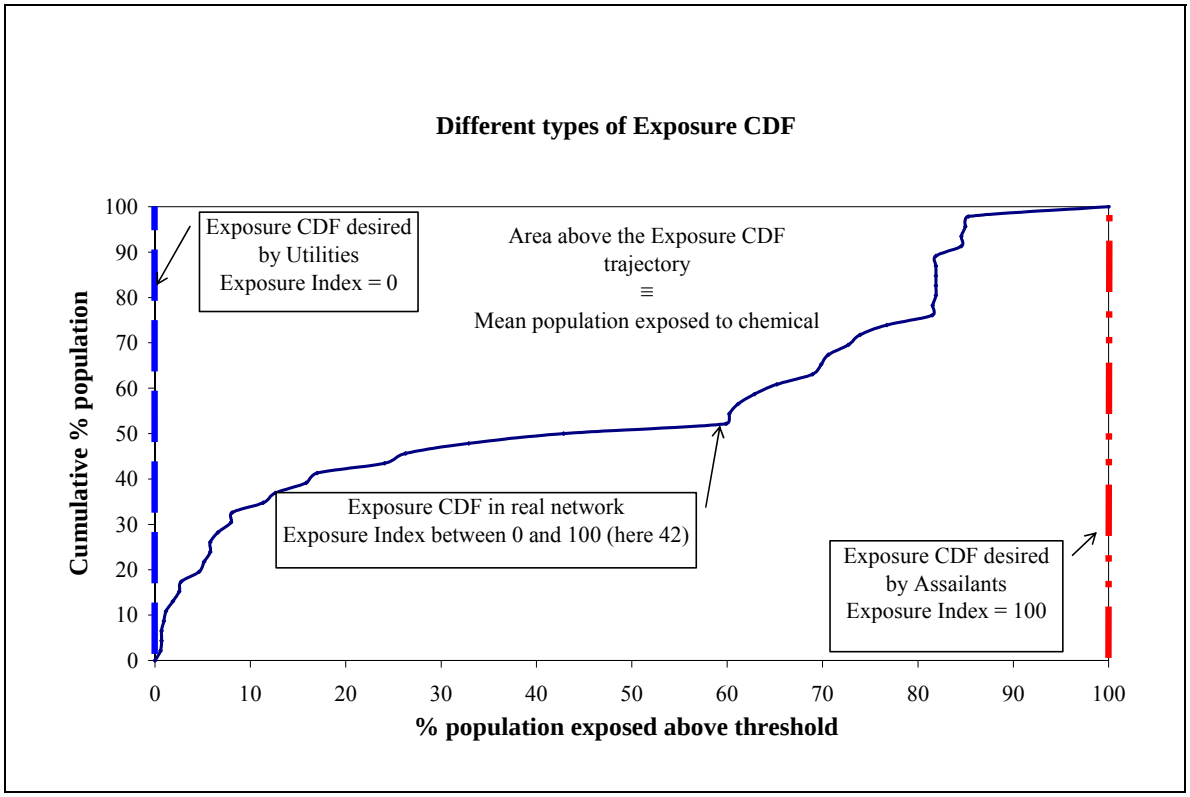

Figure 5: An illustrative example of different types of Exposure CDFs.

3.9 Influences on Exposure Index

The variation of Exposure Index under the influence of various static and dynamic variables was studied. Deterministic simulations were carried out to find the influence of injection timing, injection duration, and tank mixing on the Exposure Index for CH/BP. Injections were made for 3 hr, 6 hr, and 12 hr duration, starting at 09:00 hr and 30:00 hr. This set of injections was performed for three different tank mixing options CSTR, LIFO Plug Flow, and FIFO Plug Flow, thus generating 18 deterministic simulation runs.

Monte Carlo simulation technique was used to simulate the stochastic nature of network demand and study its effect on Exposure Index. The water demand generator program, **PRPsym**, was used to generate stochastic demands for CH/BP. The system base demand was kept constant, with the average system base demand being 1227 Lpm. The coefficient of variation which varies the demand arrival for each node was taken to be 48.9%, and the arrival variation was governed by normal distribution. Following the normal distribution, approximately 95% of the total system demand fell between $(\mu \pm 2\sigma)$, where μ is the mean total system demand (88832 Lpm) and σ is the standard deviation in total system demand (158.86 Lpm) governed by the coefficient of variation in **PRPsym**. The overall system demand was varied between 88514 Lpm and 89150 Lpm, i.e. nearly 1% variation.

100 realizations of Monte Carlo runs were simulated for the Base-Case injection made at each of the 45 selected injection nodes to find the variability of Exposure Index under the influence of varying network demand. The **PRPsym** input file and new EPANET input

file generated for a single run of Base-Case attack made on node 50 is presented in **APPENXIX I**. The C++ codes used for this calculation are presented in **Appendix II** under the heading of **Program II**.

Only the random variation in demand was considered for Monte Carlo simulation so far, other dynamic variables like injection duration, injection timing were treated deterministically. In order to have detailed information about system vulnerability all the dynamic variables have to be treated probabilistically. Before including all dynamic variables for Monte Carlo simulation, an objective way should be sought out to identify significant dynamic variables that can influence network vulnerability, such that effective Monte Carlo simulation can be carried out. Generalized Sensitivity Analysis (GSA) provides way to identify and rank the important dynamic variables based upon their significance to the simulation output.

4. GENERALIZED SENSITIVITY ANALYSIS

4.1 Introduction

GSA is a technique developed by Spear and Hornberger (1980), which identifies the input parameters that lead the simulation output to exceed a specific threshold. As a first step, the input parameters with varying degrees of uncertainty were characterized with a uniform probability density function. Then independent vectors of input parameters were formed using the Latin Hypercube Sampling (LHS) method. Each vector of model inputs was then classified behaviorally into two sample sets: those that created simulation outputs above the threshold ("fail") and those that created outputs below the threshold ("pass") (Makino et al. 2001). GSA was then used to generate a quantitative measure of the difference between "pass" and "fail" sample sets using a non-parametric statistical test like the Kolmogorov-Smirnov (K-S) two-sample test.

Four dynamic variables i) Base Demand, ii) Chemical Mass Loading, iii) Tank Head Fluctuation, and iv) Injection Duration were selected as random input parameters from dynamic variables mentioned in Table 1. Tank head fluctuation refers to tank operation where variable head is added to the initial tank head, and Base Demand is the total system base demand. Table 6 shows the data range used in simulation for four selected dynamic variables, each assumed to be uniformly distributed between their upper and lower limit for LHS. The step followed to perform GSA with the K-S *d* statistic within EPANET toolkit interface is shown in the flowchart (Figure 6).

Table 6: Four input parameters selected from dynamic variable group

Input parameter	Unit	Range	
		Minimum	Maximum
Base Demand	Lpm	380	7600
Injection duration	hr	0.25	72
Tank head fluctuation	m	0.3	30.5
Chemical mass injected	gram	2000	9000

Note: PDF is assumed to be uniform over the given range.

4.2 Assumptions

An average chemical concentration (ratio of total nodal mass consumed to total nodal demand) threshold value was selected, which was assumed to determine the node vulnerability based upon the average concentration received at that node. Any node receiving average chemical concentration above that threshold after the end of simulation was deemed vulnerable i.e. it will have consumer exposure associated with it. Similarly, another parameter, chemical mass threshold (see section 3.3), was also used to determine node vulnerability for the GSA study.

The number of nodes receiving the average chemical concentration (or chemical mass) above the threshold value was counted and the total count of vulnerable nodes for a single simulation run was defined as the Node Count (NC). NC acts as a simulation output and was assumed to provide a binary classification of the input vectors in "pass" and "fail" sample sets. A simulation with NC=0, categorizes the input vector in "pass" sample set, which means none of the nodes were exposed to chemical above the threshold concentration (or mass). Conversely, NC>0, categorizes the input vector in "fail" sample

set, which means at least one node was exposed to chemical above the threshold concentration (or mass).

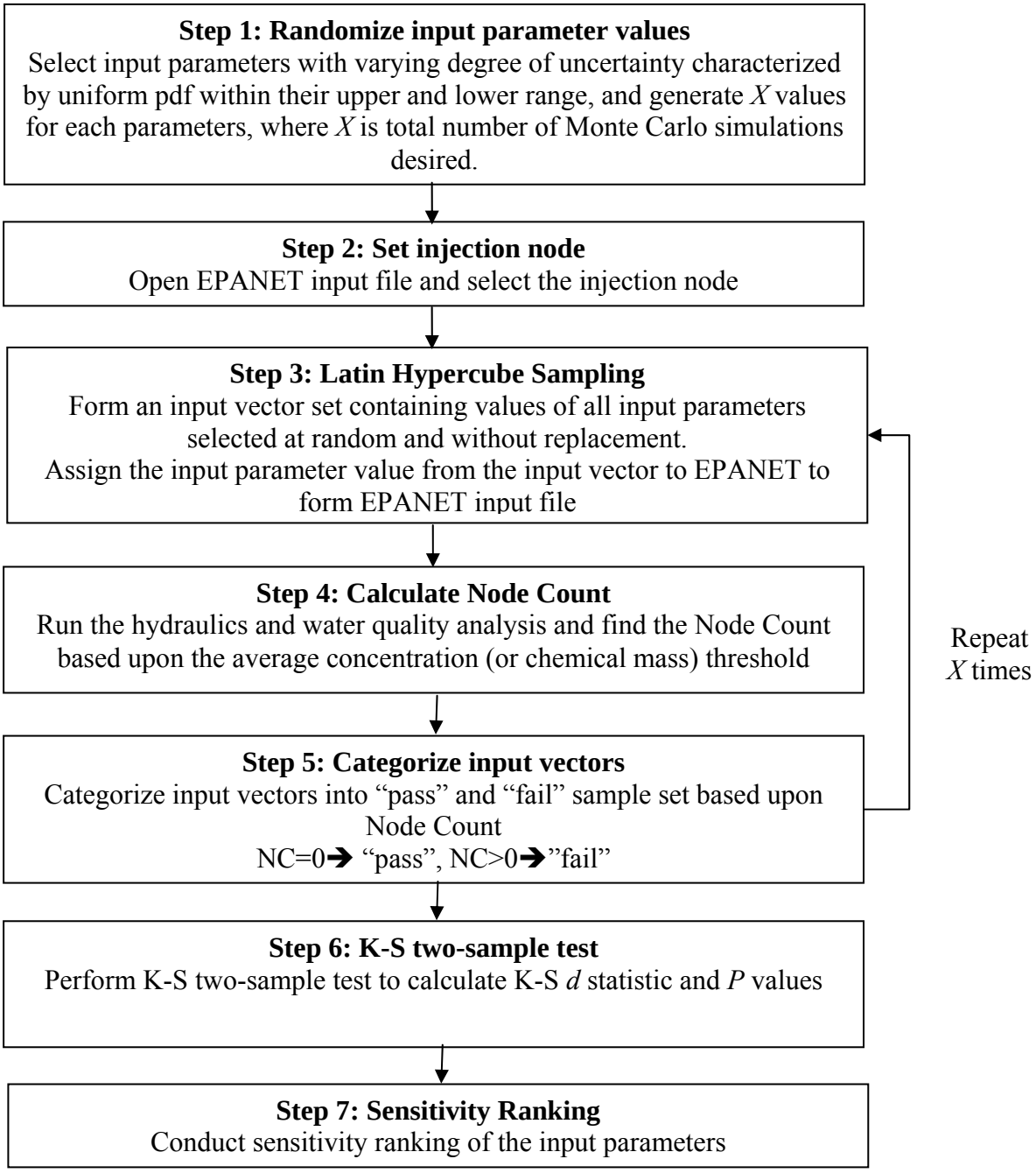

Figure 6: Flow Chart for GSA with K-S *d* statistic.

4.3 Sampling and Simulation

An injection node was selected where the variable injection will be made at each run of Monte Carlo simulation. Latin Hypercube Sampling (LHS) scheme was used to generate 1152 input vectors for 1152 Monte Carlo simulation runs.

Since the minimum pattern time step used in EPANET input file for this study was only 15 minutes, the shortest injection duration cannot be less than 15 minutes. Thus for 72 hour simulation duration, only 288 injections could be made each lasting 15 minutes more than the previous injection. The range of other three input parameters, Base demand, Tank head fluctuation, and Chemical mass loading were divided in 288 intervals of equal probability, following a uniform probability distribution. Similarly, for 24 hour simulation, 96 injections were made, thus dividing the range of three input parameters in 96 intervals of equal probability, following uniform probability distribution.

A set of input vectors was formed by combination of each sampled value from these three input parameters, and the same procedure at random and without replacement for 288 sampled values made 288 set of input vectors. A fourth member, injection duration, was then added to the input vector set in ascending order, such that the first input vector has an injection duration of 15 minutes and the last input vector has an injection duration of 72 hours, thus 288 set of input vectors were generated for 288-realization Monte Carlo simulation run. Similar steps were followed to form 96 set of input vectors for 96-realization Monte Carlo simulation run, when the simulation was carried out for only 24 hours.

After a single chain of 288 simulation runs, the seed number used to generate the random number was changed, such that range of three input parameters, Base demand, Tank head fluctuation, and Chemical mass loading were again divided into 288 intervals of equal probability, following uniform probability distribution, but this time generating different random numbers between the range. The same procedure mentioned above was followed to add injection duration as fourth member to form new 288 input vector sets. The seed number was changed for 4 times, such that the simulation runs totaled to 1152 (288x4). Seed numbers 100, 101, 102, and 103 were used for this study.

The above steps were carried out for 72 hour simulation. Similar steps were followed for 24 hour simulation also, but since only 96 simulation was possible for single chain of run, the seed number used to generate the random number was changed from 100 to 111 consecutively (in total 12 times), such that total simulation runs was equal to 1152 (96x12).

The LHS scheme generated 1152 input vectors for 1152-realization Monte Carlo simulation runs. In other words, 1152 EPANET input files were generated with different input vectors within the execution interface of the program written in C++ codes. CSTR tank mixing was assumed in these input files. For all these runs EPANET was able to execute hydraulic and water quality analysis without crashing.

4.4 Node Count Calculation

C++ codes were used to execute EPANET toolkit to simulate the hydraulics and water quality for each input file generated by LHS, thus completing 1152-Monte Carlo simulations. The average concentration $\overline{C_n}$ [M/L^3] delivered to a node *n* at the end of the 72 hr simulation period was calculated using Equation 2:

$$\overline{C_n} = \sum_{i=1}^{4320} \Omega_n(i) / \sum_{i=1}^{4320} Q_n(i) \qquad (2)$$

where *i* is the particular minute during the 72-hour simulation period, $\Omega_n(i)$ is the contaminant mass [M] entering the node *n* during i^{th} minute, $Q_n(i)$ is the corresponding nodal demand [L^3/T]. Equation 1 was used to calculate the nodal mass consumption at the end of simulation period. For 24 hour simulation *i* will range from 1 to 1440 only, for both Equation 1 and 2.

For each run of Monte Carlo simulation, number of nodes receiving average chemical concentration (or chemical mass) above an assumed average concentration (or chemical mass) threshold value was recognized and the total number of these vulnerable nodes was calculated; this total number of vulnerable nodes was defined as Node Count (NC). Thus for 1152-realization Monte Carlo runs, 1152 NC values were generated as simulation output.

4.5 GSA with Kolmogorov-Smirnov (K-S) *d* statistic

Subsequent to Monte Carlo simulation runs, the input parameter vectors are separated into two sample sets, "pass" and "fail" with respect to output above or equal to the specified NC threshold. A NC threshold of 0 was assumed for this binary classification in this calculation.

This classification of "pass" and "fail" sample set is subjective, problem dependent, and may be concerned only with extreme events, but a rule is generally specified for classification based upon the patterns in the output data (Cox and Whitehead 2005). The rule followed for binary classification of input vector set in this study is as follows: if a simulation fails to infect any nodes above the threshold average concentration value i.e. $NC = 0$ then that input vector falls in "pass" sample set, and if even a single node is infected to the chemical above threshold average concentration i.e. $NC > 0$, then that input vector falls in "fail" sample set.

The limitation of K-S test i.e. N_1 and N_2 should be ≥ 20**,** where N_1 and N_2 are number of data in "pass" and "fail" sample sets respectively, was taken in consideration for statistical calculation (Auslander *et al.* 1982). Thus the result of each Monte Carlo run consists of the input vector itself and the NC which determines whether the particular input vector gave output that fell into "pass" or "fail" category.

Sensitivity ranking of the input parameters are then performed using K-S *d* statistic, which indicates the relative importance of input parameter uncertainty in simulation

output (James *et al.* 1996). The K-S two-sample test tests the null hypothesis that the two sample sets are drawn from the same distribution, based on the *d* statistic defined in Equation 3:

$$d = \max_{-\infty<x<\infty} \left| F_P(x) - F_F(x) \right| \qquad (3)$$

where *x* is random input parameter, $F_P(x)$ and $F_F(x)$ are CDF of data set that are above NC ("fail") and data set that are equal to NC ("pass") respectively, *d* is the Kolmogorov-Smirnov two-sample statistic. Note that *d* is actually the maximum vertical distance between $F_P(x)$ and $F_F(x)$. Figure 7 illustrates how *d* is calculated using the K-S two sample test.

If $F_P(x) = F_F(x)$ i.e. null hypothesis is satisfied then it can be inferred that the variation in values of parameter *x* taken alone has no effect on the model output and the parameter appears to be insensitive for set of input parameters defined by a priori distribution. The K-S test is invariant under re-parameterization of *x*, so for the ease of calculation *x* is transformed into values ranging from 0 to 1 using Equation 4:

$$x_{transformed} = (x - x_{\min imum}) / (x_{\max imum} - x_{\min imum}) \qquad (4)$$

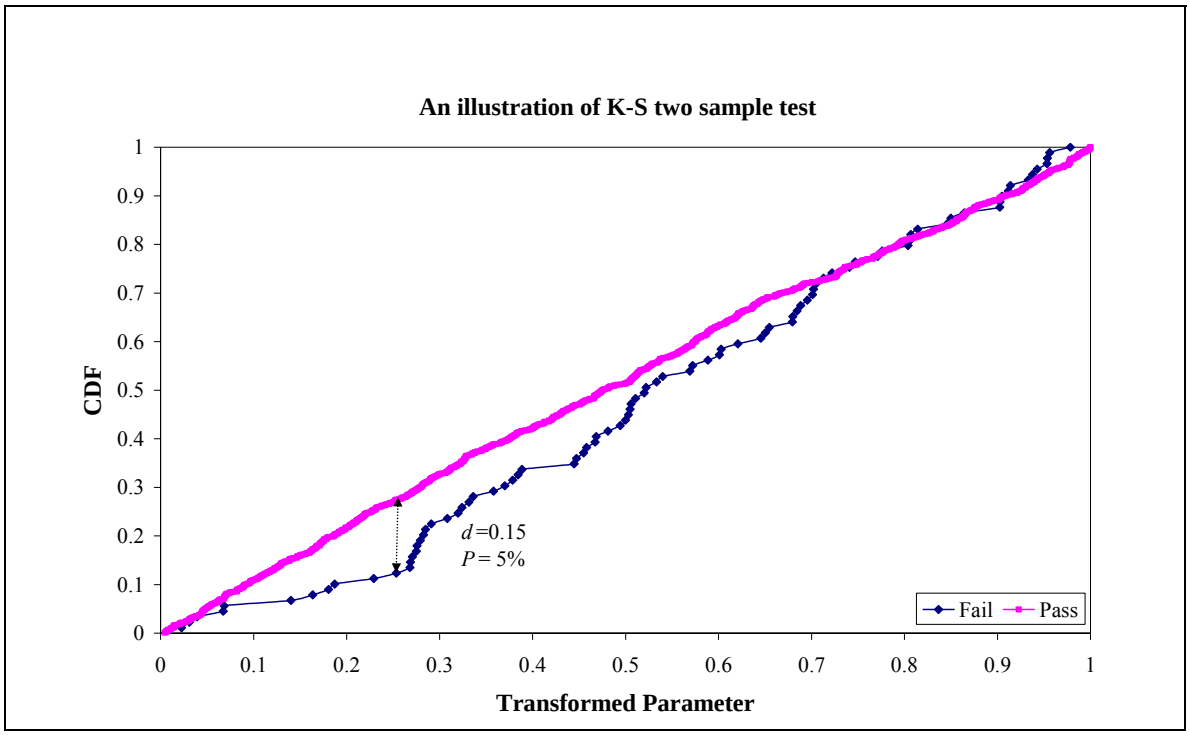

Figure 7: An illustration of K-S two sample test.

To accentuate the difference in importance between different parameters, Equations 5 and 6 as used by James *et al.* (1996) will be used to calculate the level of confidence with which *d* indicates that the realizations in a category "pass" and "fail" are truly different:

$$P = \text{Probability}\,(D > d) = 2\sum_{j=1}^{\infty}(-1)^{j-1} e^{-2j^2\lambda^2} \qquad (5)$$

$$\lambda = d\sqrt{N_e} \qquad (6)$$

where, $N_e = N_1 N_2 /(N_1 + N_2)$

Probability, *P*, is the level of significance; N_1 and N_2 are number of data in "pass" and "fail" sample sets respectively. *D* is the K-S statistic for two samples drawn from the

same population, which is the random maximum difference between CDF of two samples from the same population.

As long as N_1 and N_2 are ≥ 20, the number of realizations needed to achieve a given level of confidence in the conclusions is independent of the number of uncertain parameters because K-S two-sample test is only a function of N_1 and N_2 (Auslander et al. 1982). Lower the P value, greater is the chance of rejection of null hypothesis, which in turn accentuates the sensitivity of the simulation output to that parameter. According to Saltelli et al. (2004), the input parameters can be grouped into three sensitivity classes, based upon the significance level for rejecting the null hypothesis: a) Critical/Significant ($P < 1\%$), b) Important ($1\% \leq P \leq 10\%$), and c) Insignificant ($P > 10\%$).

Thus, K-S d statistic was used as measure of dynamic variable's significance to Monte Carlo simulation. Functions provided in Press *et al.* (2002) for K-S d statistic and P value were used to program the K-S test for this analysis.

The C++ codes written to perform LHS and toolkit execution are presented in **Appendix II** under heading of **Program III**, and the codes used to execute the K-S two-sample test are under the heading **Program IV** in the same appendix. The detailed algorithm followed to perform the GSA for the 72 hour simulation duration and first 288 input vectors from LHS are presented in **Appendix III**.

5. RESULTS

5.1 Likelihood of node selection for intrusion

As a convenient expedient, this study invoked "nodal parity"; that is, each node was assumed to have an equal probability of being selected for injection. However, it might be reasonable to assume that the likelihood a node was selected for injection is proportional to the population (or demand see Table 4) at the node or the relative accessibility of that node to an assailant. In order to examine the effect of the nodal parity assumption on the assessment of network vulnerability, seven separate experiments (E1-E7) were conducted.

For the first six experiments (E1-E6), injection nodes were selected based upon the proportion of population, such that, a node with higher population had more chance of being attacked than a node with less population. In these experiments, the number of attack nodes was less than 45 but in total 45 Base-Case injections were made for each experiment, as repeated attacks were made at nodes with higher population proportion.

E1-E6 differ in terms of number of times a node will be attacked which was determined by cumulative frequency analysis. Based upon the population at a node, a probability of a node being attack was calculated. A cumulative probability value was then assigned to each node. For each experiment (E1-E6), different set of uniformly distributed random numbers (between 1 and 45) were generated and arranged in ascending order. Cumulative frequency analysis was then performed by comparing these random numbers with the range determined from cumulative probability of nodes to calculate a value for

each node. This value determined the number of time a node will be attacked. The last experiment (E7) assumed nodal parity. In this case, 45 different nodes received a Base-Case injection, with each node having equal probability of being selected for injection.

The Exposure CDF for the percentage population receiving chemical mass above the 0.5% threshold was plotted for all seven experiments and the Exposure Index was calculated. As illustrated in Figure 8, the EI is maximized when nodal parity is assumed (case E7). In other words, the vulnerability of network was greatest when it was assumed that all nodes are equally likely to be attacked. Therefore, to guarantee a worst-case attack scenario and to maximum network vulnerability, the nodal parity assumption was adopted in all subsequent network simulations.

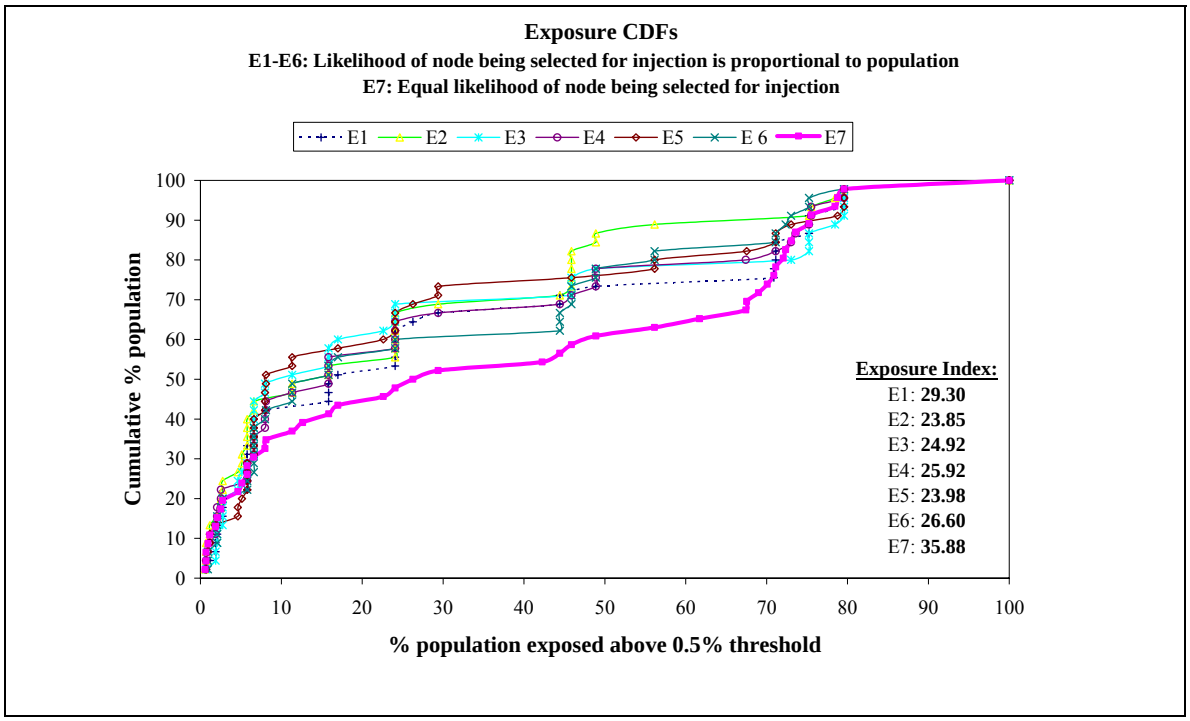

Figure 8: Likelihood of a node being selected for intrusion

5.2 Zone of Influence

The Base-Case intrusion was made at 45 injection nodes in CH/BP water distribution network. The nodal mass values calculated from simulation were compiled in a result matrix, where attack nodes are arranged in columns and receptor nodes are arranged in rows. Each matrix element represents the percentage nodal mass (i.e. percentage of total injected mass consumed at a node) received at the end of the 72-hour simulation. Table 7 shows a small portion of the result matrix for Base-Case intrusions made on CH/BP network with CSTR tank mixing. Nodes receiving chemical mass above 0.5% threshold are shaded in this result matrix.

Table 7: Result matrix for chemical mass consumed at nodes.

Percentage Nodal Mass Loadings						
Receptor Node	**Attack Nodes**					
	1	2	3	...	44	45
1	0	0	0	…	0	0
2	0.313	51.04	0	…	0	0
3	0.384	0.395	2.75	…	0	0
4	0.211	0.217	0.117		0	0
…	…	…	…	…	…	…
87	0.044	0.049	0.146	…	0.553	0.665
88	0.188	0.209	0.610	…	13.50	36.40

A complete result matrix for Base-Case attack on CH/BP network with CSTR tank mixing is presented in **Appendix IV**.

The corresponding percentage population associated with the nodes that received the chemical mass loading above the 0.5% mass threshold was summed to get the total

percentage population exposed for a particular injection node. Table 8 shows the data on percentage population exposed above 0.5% and 0.0% mass thresholds for Base-Case attack on CH/BP.

The data for the percentage population exposed above the 0.5% threshold were used to draw a preliminary contour plot from which population exposure value for non-injection nodes were interpolated. Based upon the ability of the contaminant to reach the consumers from a given injection node, CH/BP network nodes were grouped in three zones Red, Yellow, and Green.

Figure 9 shows the Zone of Influence map for CH/BP network. If an injection is made at a node lying in the Red zone, more than 60% population will receive chemical mass exceeding the 0.5% threshold by the end of the 72-hour simulation period. Similarly, injection at nodes lying in the Yellow zone and Green zone will ensure effects on 30-60% and <30% of population above the threshold, respectively.

Table 8: Percentage population exposed to chemical above mass thresholds.

Rank	% Probability (CDF)	0.5% mass threshold		0.0% mass threshold	
		Injection Node	% population exposed	Injection Node	% population exposed
1	97.83	22	79.58	1	100.00
2	95.65	19	78.77	2	100.00
3	93.48	17	78.45	17	98.24
4	91.30	50	75.52	6	94.27
5	89.13	24	75.24	11	94.16
6	86.96	85	73.59	9	93.00
7	84.78	49	73.04	19	90.22
8	82.61	84	72.35	22	89.76
9	80.43	32	72.06	85	86.63
10	78.26	26	71.14	26	86.63
11	76.09	41	70.90	41	86.63
12	73.91	38	70.09	38	86.63
13	71.74	106	68.99	49	86.63
14	69.57	110	67.55	106	86.63
15	67.39	117	67.43	32	86.63
16	65.22	138	61.70	24	86.63
17	63.04	11	56.15	117	86.63
18	60.87	9	48.89	84	86.05
19	58.70	2	45.88	110	86.05
20	56.52	6	44.43	138	85.82
21	54.35	1	42.26	51	85.82
22	52.17	51	29.40	81	85.13
23	50.00	53	26.27	53	83.74
24	47.83	56	24.07	50	75.52
25	45.65	81	22.62	56	24.07
26	43.48	79	17.01	79	17.01
27	41.30	71	15.85	71	15.85
28	39.13	67	12.65	67	12.65
29	36.96	89	11.34	89	11.34
30	34.78	119	8.10	119	8.10
31	32.61	94	7.99	94	7.99
32	30.43	133	6.60	133	6.60
33	28.26	5	5.79	5	5.79
34	26.09	98	5.79	98	5.79
35	23.91	136	5.14	136	5.14
36	21.74	122	4.63	122	4.63
37	19.57	69	2.74	69	2.74
38	17.39	127	2.54	127	2.54
39	15.22	35	2.08	35	2.08
40	13.04	47	1.88	47	1.88
41	10.87	125	1.16	125	1.16
42	8.70	140	0.93	140	0.93
43	6.52	20	0.69	20	0.69
44	4.35	21	0.69	21	0.69
45	2.17	39	0.58	39	0.58

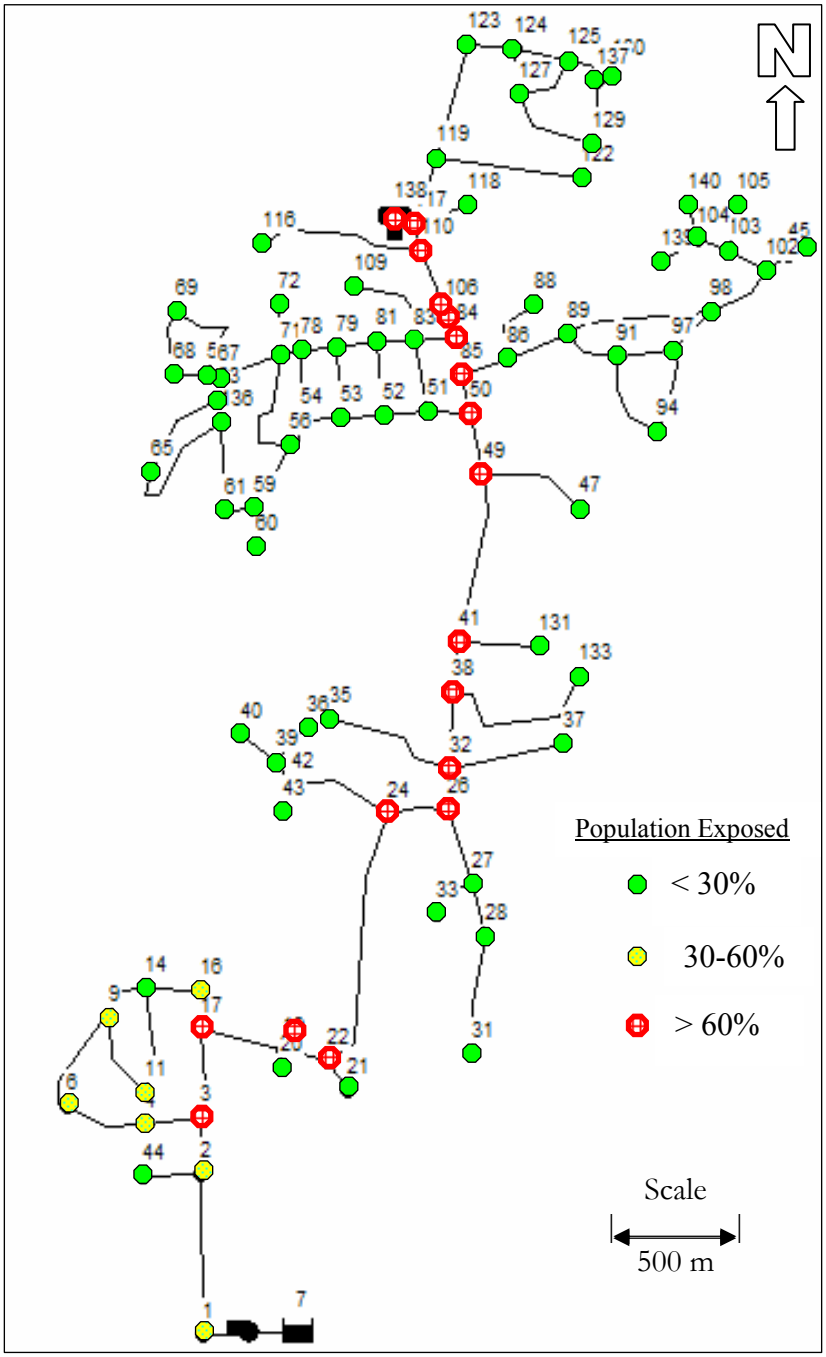

Figure 9: Zone of Influence map for CH/BP water distribution network.

All 18 nodes lying in the Red zone were located along the 12" main trunk of the network. The nodes lying in upper east and west sides of the main trunk fell into Green zone. The Yellow zone was comprised of a cluster of 7 nodes towards the south-western side of the network. In total, 64 nodes fell into Green zone. Clearly, attacks made along the back bone of the network (including the storage tank) reach the highest percentage of consumers in the network. The main trunk line contributes greatly to the vulnerability of the CH/BP water distribution system.

Nodes 56, 109, 116, 131, and 133 are nodes with base demand higher than 37.85 Lpm, (i.e. proportionally higher percentages of population are associated with these nodes), but they lie in the Green zone. In contrast, almost all the nodes lying in Red zone have base demands less than 18 Lpm, but mass injections made at these nodes reached the highest percentage of consumers.

5.3 Exposure CDF and Exposure Index

The CDF for the percentage population exposed above the threshold for Base-Case intrusion in 45 injection nodes is summarized in Table 8. The information was plotted in the Exposure CDF plot with percentage population exposed above threshold data on X-axis and CDF of percentage population exposed on Y-axis. When the threshold value is set to 0.0%, then all nodes receiving chemical mass are vulnerable. Figure 10 shows an Exposure CDF for Base-Case intrusions on CH/BP network when both 0.0% threshold and 0.5% threshold value are considered.

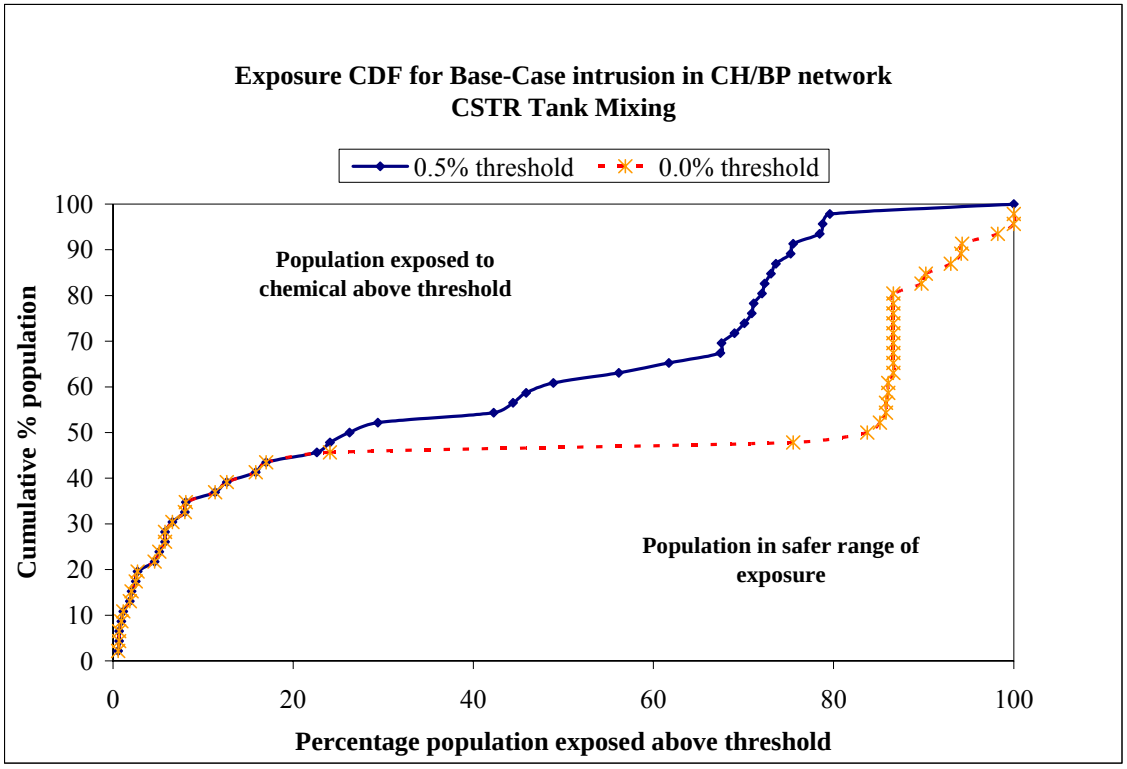

Figure 10: Exposure CDF for Base-Case intrusion in CH/BP.

The area of the region above the Exposure CDF trajectory for 0.5% threshold value on Figure 9 is 35.9 and this agrees with the mean population exposed above the 0.5% threshold for 45 injection nodes. Similarly, the Exposure Index (EI) for 0.0% threshold is calculated to be 50.34. As the mass threshold drops, the Exposure Index increases. Here, Exposure Index will be taken as the area of region above the Exposure CDF trajectory. Table 9 summarizes the statistics for percentage population exposed for Base-Case attack on 45 injection nodes.

The 0.5% threshold Exposure CDF in Figure 9 was related to Zone of Influence map (Figure 9). It showed that the right most part of CDF curve (% exposed > 60) was

formed by the injections made at nodes lying in Red zone of influence, similarly, injection at nodes in less dense Yellow zone formed the center portion of the curve (% exposed between 30-60). The Green zone contains large number of nodes and formed the left portion of the CDF curve that have percentage exposure value <30.

Nodes in Red zone are along the back bone of the network and are connected by 12" mains. Any injection in these nodes will ensure rapid propagation of the contaminant throughout the network and, hence, lead to high population exposures. Yellow zone nodes lie in the south-western portion of a network within a loop connected by 8" pipes. There are only 7 nodes lying in this zone, and injection at any of these node provide nearly same consumer exposure risk. The flatness in the Exposure CDF (see Figure 10) at the center is due to the fact that all injection in that region was made from Yellow zone and their exposure values were also very close.

Most of the nodes in Green zone are towards the upper portion of the network in the looping links and dead end stems away from the main trunk line. Propagation of chemical from nodes in Green zone is limited among themselves or within their vicinity only, so they pose less exposure risk to consumers. More than half of 45 injection nodes lie in the Green zone, and they affected only 0.5% to 29%, as shown by the sharp and dense curve on the left portion of Exposure CDF (see Figure 10).

Table 9: Summary statistics for percentage population exposed.

(Base-Case attack on 45 nodes)

	0.5% mass threshold	**0.0% mass threshold**
Mean (i.e. Exposure Index)	35.9	50.4
Standard Deviation	30.7	41.8
Coefficient of Variation (%)	85.6	83.1

5.4 Results for deterministic variation in network variables

Deterministic simulations were carried out to find the influence of injection timing, injection duration, and tank mixing on the Exposure Index for CH/BP. Injections were made for 3 hr, 6 hr, and 12 hr duration, starting at 09:00 hr and 30:00 hr. This set of injections was performed for three different tank mixing options CSTR, LIFO Plug Flow, and FIFO Plug Flow, thus generating 18 deterministic simulation runs. The total chemical mass injected into the system was kept constant to 7200 g by changing the injection rate depending upon the injection period.

The Exposure Index calculated from 18 deterministic simulations for different injections considered is summarized in Table 10.

Table 10: Exposure Index for different deterministic injections scenarios.

Tank mixing	Injection duration hr	Injection period	Injection g/min	Exposure Index	
				Threshold 0.5% (36 g)	Threshold 0.0% (0 g)
CSTR	3	09:00-12:00	40	30.0	44.3
	6	09:00-15:00	20	35.9	50.3
	12	09:00-21:00	10	37.0	52.4
	3	30:00-33:00	40	34.4	48.5
	6	30:00-36:00	20	31.9	48.5
	12	30:00-42:00	10	33.1	52.2
FIFO Plug Flow	3	09:00-12:00	40	26.9	43.2
	6	09:00-15:00	20	30.1	49.6
	12	09:00-21:00	10	31.5	52.8
	3	30:00-33:00	40	32.1	48.5
	6	30:00-36:00	20	30.2	48.5
	12	30:00-42:00	10	30.5	51.4
LIFO Plug Flow	3	09:00-12:00	40	33.9	45.2
	6	09:00-15:00	20	42.1	52.6
	12	09:00-21:00	10	42.2	52.8
	3	30:00-33:00	40	41.2	52.1
	6	30:00-36:00	20	38.8	52.1
	12	30:00-42:00	10	41.7	52.2

EI results in Table 10 shows that when the 0.0% threshold value was considered, the EI for variation in injection duration and timing, for all tank mixing options were very close to each other (43-52%). But the variation in static variable (tank mixing) had significant effect in consumer exposure for the 0.5% threshold value (26.9-41.7%). The LIFO plug flow tank mixing with EI of 42.1 had more impact on consumer than FIFO plug flow

tank mixing with EI of 30.1and CSTR tank mixing with EI of 35.9, when the Base-Case intrusions were simulated in the 45 nodes.

Shorter injection towards the middle of 72-hour simulation period (30:00 hr) was able to affect larger populations than the same injection made during early hours (09:00 hr) but the longer injections made at early simulation hours proved to be more effective than that injected at intermediate simulation hours. This shows significant influence of injection duration on the population exposure and network vulnerability. The Exposure CDFs of 18 deterministic scenarios discussed above are presented in **APPENDIX IV**.

Further calculations were carried in order to check the effect of total simulation period on Exposure Index. In addition to 72-hour simulation of Base-Case attack on 45 injection nodes, two more simulations were conducted where in one case the simulation period was reduced to only 24 hours and in another case the simulation period was extended to 120 hours. Base-Case attacks on 45 nodes were made for both of these simulations.

Figure 11 shows the Exposure CDFs for three different simulation periods. The Exposure CDFs for these three simulations varied with change in simulation period. The Exposure Index increased as the simulation period was increased from 24-hour to 120-hour. For the injection made at node 50 lying in Red zone of influence, the mass consumption at nodes for 24-hour injection was 36%; similarly, 65% and 77% of mass was consumed at nodes for 72-hour and 120-hour simulation. As the simulation duration increased more mass was consumed at the nodes, as the mass retained at the tank starts to

spread into the system. Since, the tank in the system assumes CSTR tank mixing, from Table 3 it is known that the time for 95% mass removal is 276 hours.

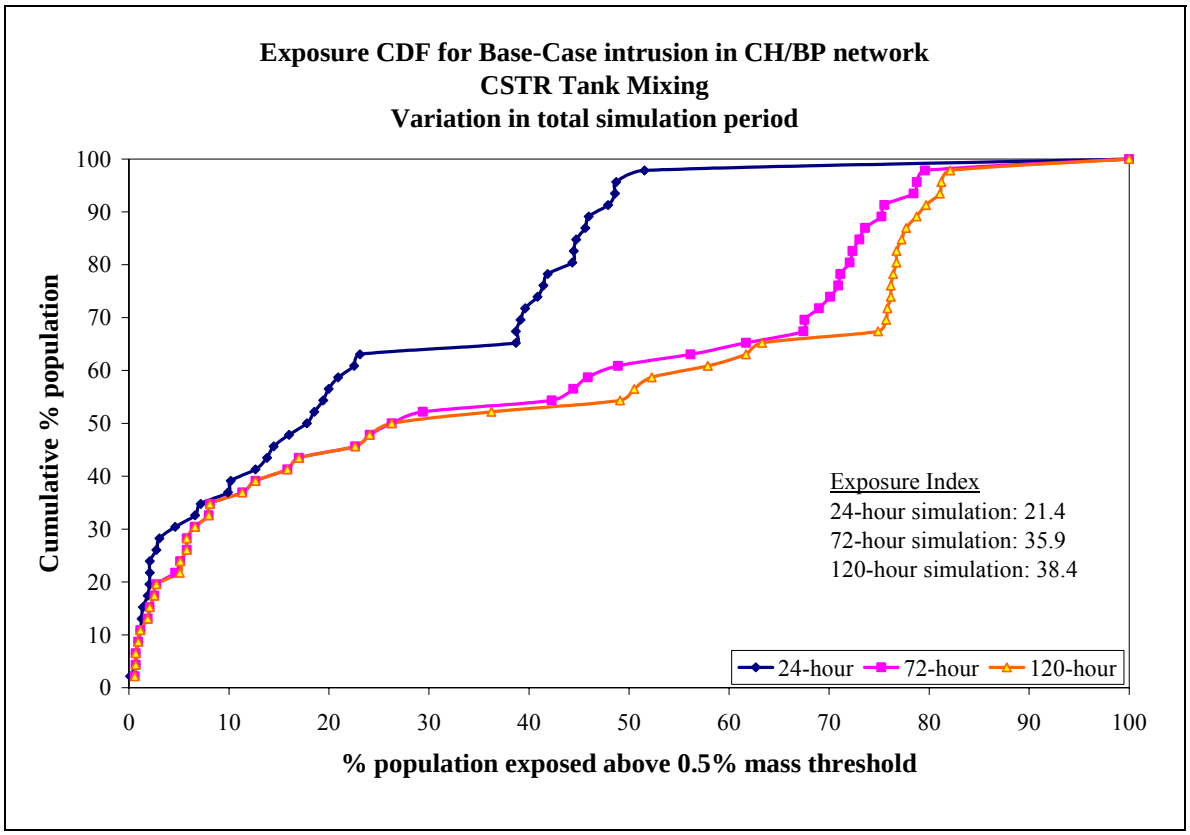

Figure 11: Effect of total simulation period on the Exposure CDF for Base-Case attack made on CH/BP.

The Exposure Index for 24-hour simulation period was only 21.4 where as the Exposure Index increased up to 38.9 for 120-hour simulation period, this is due to the increase in nodal mass consumption as the simulation period was increased. Thus, for CH/BP with CSTR tank mixing and under multiple Base-Case attack scenarios, a complete picture of Exposure CDF can be found if the simulation period is increased further than 276 hours.

5.5 Results for stochastic variation in network demand

Stochastic nature of network demand was introduced in simulation by using demand generator **PRPsym**. The average system base demand of 1227 Lpm was kept constant while the demand multiplier pattern was varied following the normal distribution. The coefficient of variation which varies the demand arrival for each node was taken to be 48.9% in PRPsym. This allowed the overall system demand to fluctuate by 1%. Detailed information on **PRPsym** parameter is discussed in section 3.9.

100 realizations of Monte Carlo simulation of demand were carried out using **PRPsym** and EPANET to calculate the Exposure Index for Base-Case intrusion made at each of the 45 injection nodes. Figure 12 shows the Exposure CDF for 100 Monte Carlo realizations where the curve on the upper side and curve on the lowest side represents the curves with minimum and maximum Exposure Index respectively.

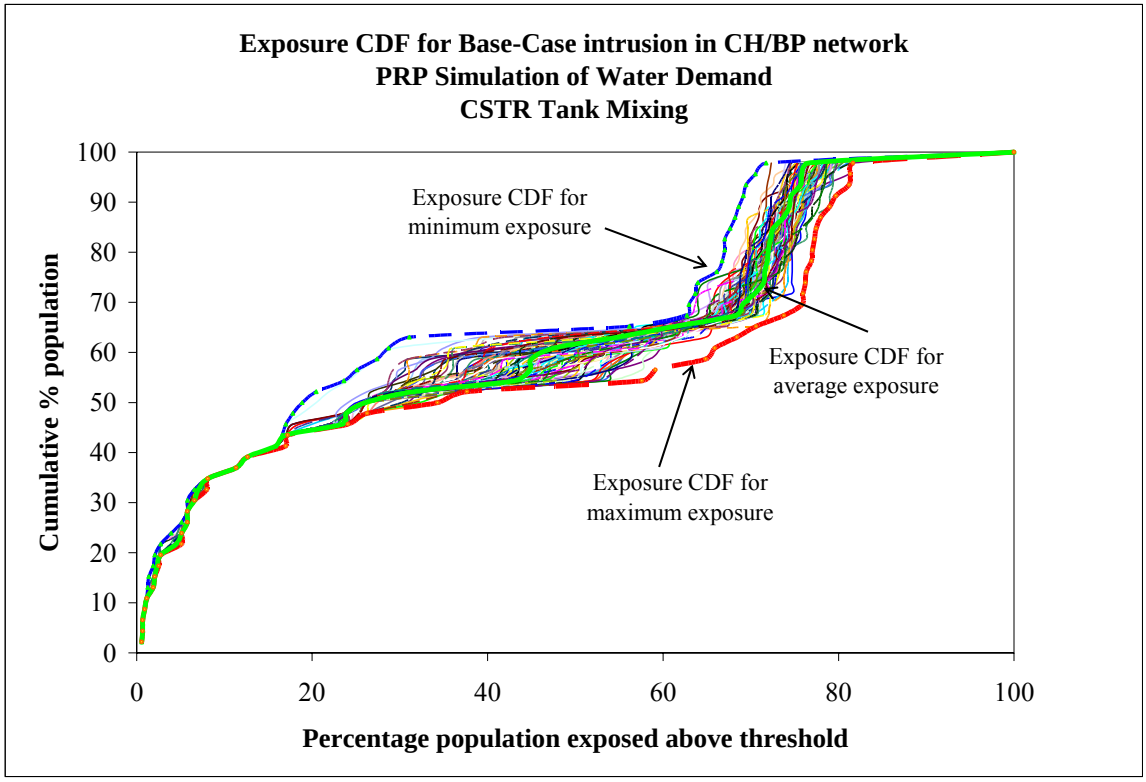

Figure 12: Exposure CDF for Base-Case intrusion in CH/BP network with stochastic water demand.

In Figure 12, the Exposure CDF running through middle of the band provides information on the average Exposure Index for 100 simulation runs. The maximum EI was 37.1, minimum EI was 33.7, and the average EI was 35.7 for this 100 Monte Carlo simulation. Table 11 summarizes the statistics for 100 Monte Carlo simulation of water demand.

Table 11: Exposure Index for 100 Monte Carlo simulation of network demand.

Statistics	Exposure Index
Minimum	33.67
Maximum	37.14
Average	35.72
Standard Deviation	0.71
Coefficient of Variation (%)	2

For about 1% of variation in total system demand there was 10% variation in network vulnerability, which signifies that fluctuation in network demand has large effect on network vulnerability.

5.6 Generalized Sensitivity Analysis of intrusions in CH/BP

Generalized sensitivity analysis using the K-S *d* statistic was performed to rank the four input network variables i) Base Demand, ii) Chemical Mass Loading, iii) Tank Head Fluctuation, and iv) Injection Duration, based upon their significance to network vulnerability. The network solver EPANET was used as a model to simulate hydraulics and water quality in the CH/BP distribution network for 72-hour simulation period. For all intrusion cases, tank mixing was considered as CSTR.

The range of each of the four input variables i) Base Demand, ii) Chemical Mass Loading, iii) Tank Head Fluctuation, and iv) Injection Duration, was divided into 1152 intervals of equal probability using uniform probability distribution between their range specified in Table 5, following the procedure mentioned in section 4.3 and illustrated in Figure 6. As an illustration, the CDF plot for one of the four random input parameters,

Base Demand, is shown in Figure 13, which follows the uniform distribution between the range of 380 Lpm and 7600 Lpm.

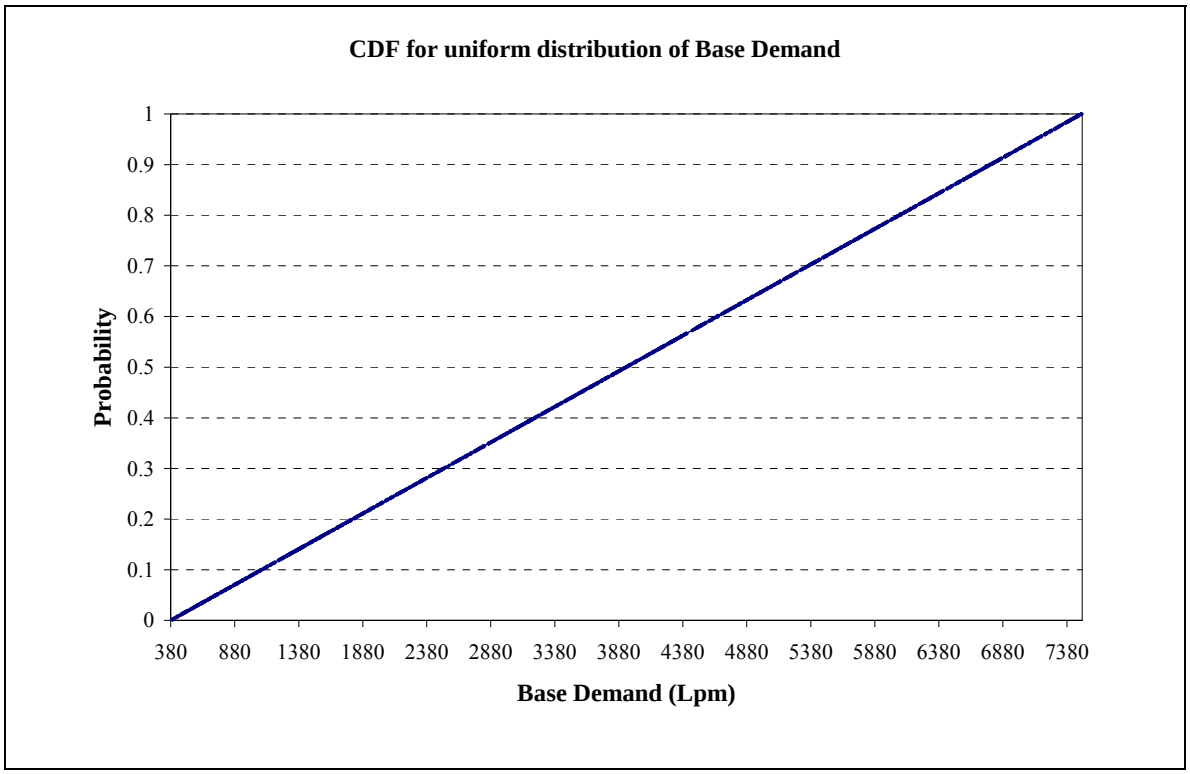

Figure 13: CDF for uniform distribution of Base Demand

For a 72-hour simulation, node 17 (Red zone), node 6 (Yellow zone), and node 81 (Green zone) in the CH/BP network zone of influence map were selected as injection nodes. Three sets of GSA were performed, one each on the selected injection nodes. The only difference between these three GSA experiments was the injection location; otherwise they share same set of 1152 input vectors. Following the Latin Hypercube Sampling, a Node Count was calculated for each Monte Carlo simulation.

Binary classification of input vectors into "pass" and "fail" sample sets was done based upon threshold Node Count of zero. The K-S two-sample test was used to find the maximum difference, *d* statistic, between the CDFs of data in the "pass" and the "fail" sample sets.

5.6.1 Average Concentration Threshold

For the above three GSA study, average chemical concentration threshold was selected as a parameter to define a node vulnerability. If a node received average concentration above the threshold, then Node Count was increased by 1. Instead of focusing on only one threshold value for average concentration, K-S statistics were calculated for a range of average concentration threshold values. Table 12a, 12b, and 12c summarizes the K-S test statistics for injection made at nodes lying in Red, Yellow, and Green zones of influence, when the 72-hour simulation was conducted. The range of threshold value was limited for each GSA experiment because of limitation of K-S test (data in "pass" and "fail" sample sets should be more than 20) and/or consistency of K-S *d* statistic results for the consecutive threshold values.

Table 12a: K-S test statistics for injection made at node 17 (Red Zone)

Average Concentration Threshold	Input Parameter								Total number of samples $N_1 + N_2 = 1152$	
	Base Demand [380-7600] Lpm		Injection Duration [15-4320] min		Tank Head Fluctuation [0.3-30.5] m		Chemical Mass Loading [2000-9000] g			
(mg/l)	*d*	*P*	*d*	*P*	*d*	*P*	*d*	*P*	N_1(Pass) Node Count = 0	N_2(Fail) Node Count > 0
0.2	0.4877	0.00	0.0939	0.12	0.0759	0.31	0.6843	0.00	194	958
0.4	0.6085	0.00	0.0576	0.30	0.0533	0.39	0.3833	0.00	641	511
0.6	0.7657	0.00	0.0463	0.76	0.1004	0.03	0.3186	0.00	874	278
0.8	0.7778	0.00	0.0737	0.39	0.2388	0.00	0.2830	0.00	974	178
1.0	0.8399	0.00	0.1488	0.01	0.3691	0.00	0.1347	0.04	1028	124
1.2	0.8254	0.00	0.1707	0.01	0.4535	0.00	0.1297	0.07	1043	109
1.4	0.8190	0.00	0.1960	0.00	0.4863	0.00	0.1261	0.11	1051	101
1.6	0.8268	0.00	0.2032	0.00	0.4810	0.00	0.1315	0.09	1054	98
1.8	0.8514	0.00	0.2214	0.00	0.4655	0.00	0.1399	0.08	1062	90
2.0	0.8505	0.00	0.2179	0.00	0.4635	0.00	0.1502	0.05	1063	89
2.2	0.8610	0.00	0.2140	0.00	0.4593	0.00	0.1583	0.04	1065	87
2.4	0.8586	0.00	0.2236	0.00	0.4659	0.00	0.1791	0.01	1068	84
2.6	0.8562	0.00	0.2216	0.00	0.4738	0.00	0.1748	0.02	1071	81
2.8	0.8538	0.00	0.2332	0.00	0.4827	0.00	0.1840	0.01	1074	78
3.0	0.8614	0.00	0.2419	0.00	0.4811	0.00	0.1826	0.02	1075	77

Table 12b: K-S test statistics for injection made at node 6 (Yellow Zone)

Average Concentration Threshold	Input Parameter								Total number of samples $N_1 + N_2 = 1152$	
	Base Demand [380-7600] Lpm		Injection Duration [15-4320] min		Tank Head Fluctuation [0.3-30.5] m		Chemical Mass Loading [2000-9000] g			
(mg/l)	*d*	*P*	*d*	*P*	*d*	*P*	*d*	*P*	N_1(Pass) Node Count = 0	N_2(Fail) Node Count > 0
1.4	0.6490	0.00	0.1108	0.83	0.1021	0.89	0.8445	0.00	33	1119
1.6	0.6062	0.00	0.1060	0.57	0.0832	0.84	0.8237	0.00	58	1094
1.8	0.5365	0.00	0.0877	0.49	0.0771	0.66	0.7334	0.00	98	1054
2.0	0.5238	0.00	0.0714	0.54	0.0575	0.80	0.7024	0.00	144	1008
2.2	0.4901	0.00	0.0526	0.78	0.0795	0.27	0.7036	0.00	188	964
2.4	0.5053	0.00	0.0519	0.71	0.0815	0.18	0.6746	0.00	226	926
2.6	0.5171	0.00	0.0780	0.17	0.0837	0.12	0.6203	0.00	259	893
2.8	0.5212	0.00	0.0671	0.27	0.0847	0.08	0.5802	0.00	296	856
3.0	0.4983	0.00	0.0662	0.24	0.1014	0.01	0.5575	0.00	343	809
3.2	0.5000	0.00	0.0610	0.29	0.0621	0.27	0.5159	0.00	393	759
3.4	0.5079	0.00	0.0509	0.49	0.0523	0.45	0.4921	0.00	432	720
3.6	0.5104	0.00	0.0429	0.68	0.0479	0.55	0.4750	0.00	469	683
3.8	0.5488	0.00	0.0385	0.79	0.0509	0.45	0.4257	0.00	523	629
4.0	0.5746	0.00	0.0300	0.96	0.0509	0.45	0.4020	0.00	558	594

Table 12c: K-S test statistics for injection made at node 81 (Green Zone)

Average Concentration Threshold	Input Parameter								Total number of samples $N_1 + N_2 = 1152$	
	Base Demand [380-7600] Lpm		Injection Duration [15-4320] min		Tank Head Fluctuation [0.3-30.5] m		Chemical Mass Loading [2000-9000] g			
(mg/l)	*d*	*P*	*d*	*P*	*d*	*P*	*d*	*P*	N_1(Pass) Node Count = 0	N_2(Fail) Node Count > 0
0.8	0.6486	0.00	0.1054	0.65	0.0954	0.77	0.8365	0.00	51	1101
1.0	0.5738	0.00	0.0646	0.74	0.0755	0.55	0.7025	0.00	126	1026
1.2	0.5117	0.00	0.0589	0.61	0.0977	0.08	0.6856	0.00	203	949
1.4	0.5393	0.00	0.0600	0.45	0.0870	0.09	0.5997	0.00	268	884
1.6	0.5568	0.00	0.0483	0.63	0.0994	0.02	0.5159	0.00	337	815
1.8	0.5737	0.00	0.0454	0.65	0.0646	0.22	0.4475	0.00	410	742
2.0	0.6102	0.00	0.0327	0.92	0.0520	0.43	0.3817	0.00	498	654
2.2	0.6483	0.00	0.0330	0.91	0.0540	0.37	0.3374	0.00	555	597
2.4	0.6613	0.00	0.0317	0.93	0.0496	0.48	0.3259	0.00	593	559
2.6	0.6966	0.00	0.0323	0.93	0.0493	0.49	0.2960	0.00	643	509
2.8	0.7142	0.00	0.0397	0.77	0.0455	0.61	0.2853	0.00	677	475
3.0	0.7507	0.00	0.0397	0.79	0.0429	0.70	0.2632	0.00	715	437

Figures 14a, 14b, and 14c shows the plot of K-S test *d* statistics with respect to different average concentration thresholds considered in calculation for injections made at nodes lying in Red, Yellow, and Green zone of influence. The dashed red line passing horizontally across the plot represents the 10% level of significance. Any input parameters lying below that 10% level of significance (i.e. $P > 10\%$) will have little influence on the simulation output for the given threshold values. Conversely, the input parameters lying above this 10% line will exert significant influence on the simulation output.

For the injection made at node 17 (Red zone), the K-S d statistic value of 0.1 nearly represents the 10% level of significance line. Similarly, for injection in node 6 (Yellow zone) and node 81 (Green zone) the nearest K-S d statistic value that represents 10% level of significance are 0.075 and 0.085 respectively.

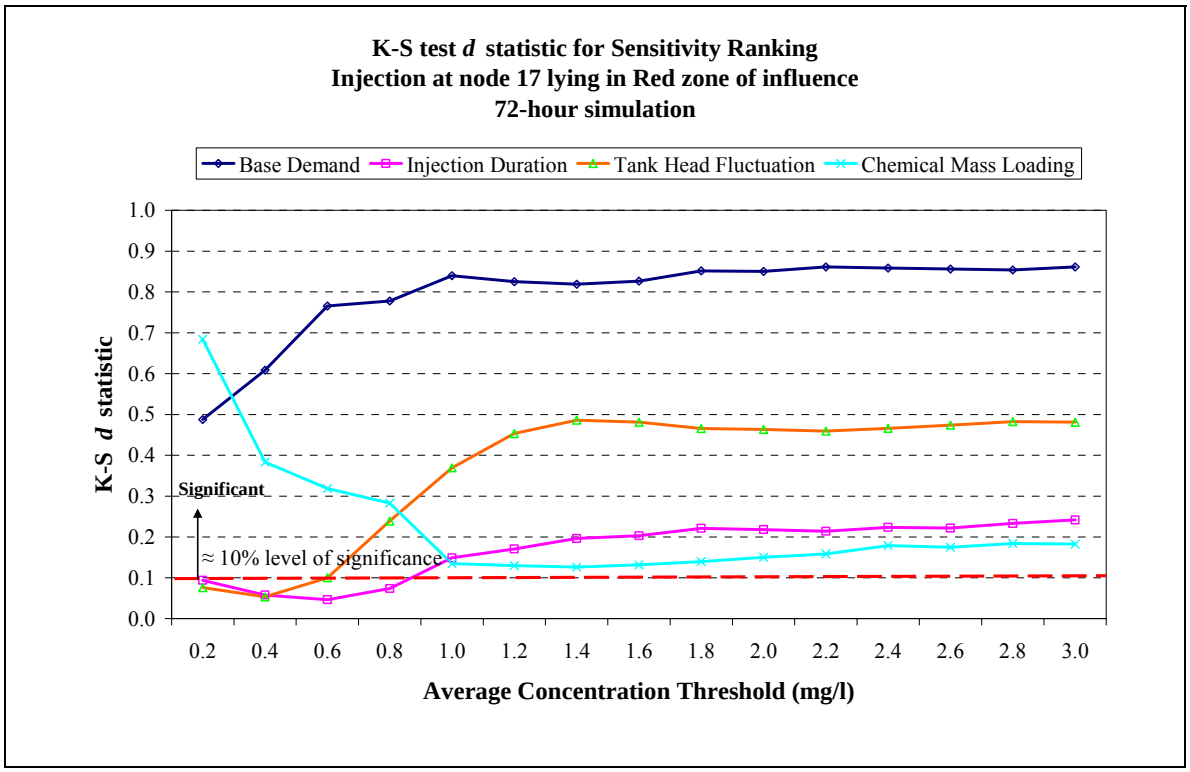

Figure 14a: Variation in Sensitivity Ranking with respect to average concentration threshold value when injection is made at node 17 (Red zone).

For injections made at node 17 (Red zone), and when average concentration threshold was taken as a parameter to define network vulnerability, average nodal concentration values above 1 mg/L are more sensitive to the dynamic variables governed by consumers and utilities (i.e. base demand and tank head fluctuation) than the variables that are controlled by the assailants (injection duration and mass loading). Since the Red zone is mainly comprised of nodes lying along the main trunk line, it appears that consumer behavior and utility practice exert the strongest influences over network vulnerability.

Table 3 shows that for injection made at node 50 (another Red zone node) with CSTR tank at least 33% of mass was retained inside the tank during the simulation period. Any increase or decrease in tank capacity means corresponding fluctuation in percentage of mass retained at tank which directly relates to mass that is available for nodal consumption. Thus the variability in tank head when injection is made at node 17 has strong influence in network vulnerability. GSA results also correctly identified this strong significance of tank head variation and have ranked the parameter higher.

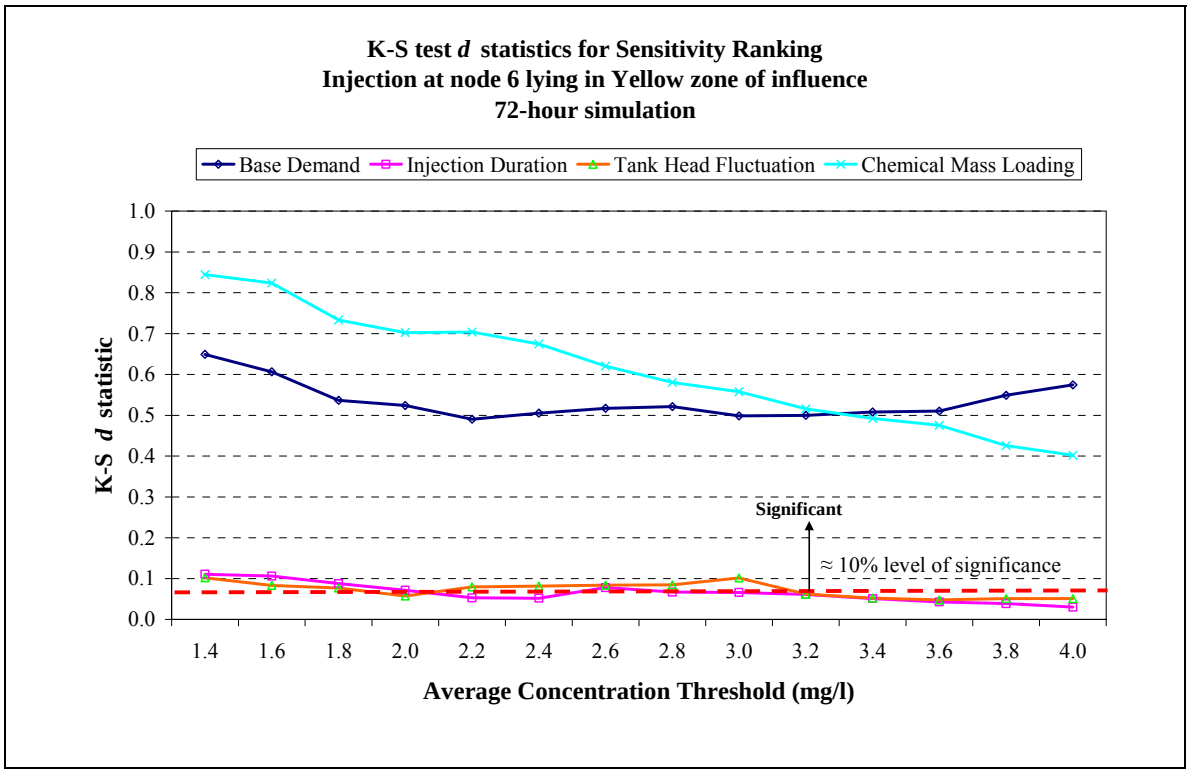

Figure 14b: Variation in Sensitivity Ranking with respect to average concentration threshold value when injection is made at node 6 (Yellow zone).

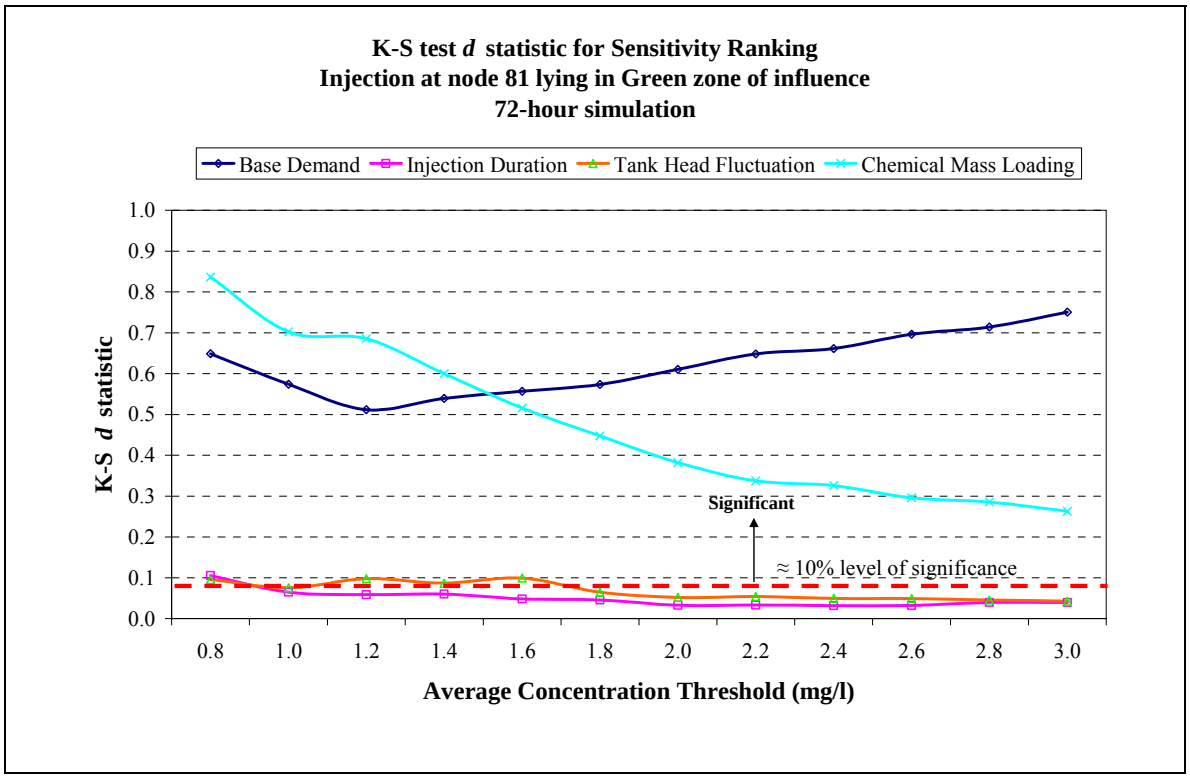

Figure 14c: Variation in Sensitivity Ranking with respect to average concentration threshold value when injection is made at node 81 lying in (Green zone).

Similarly, when injections were made at node 6 (Yellow zone) and node 81 (Green zone) only 10% and 5% of chemical mass was retained in the tank after the 72-hour simulation period. So, the fluctuation in tank operations should not have larger influence on the simulation results when injections were made at these nodes. This result was verified by the GSA performed for injection made at node 6 and node 81. In both cases, the variation in Tank Head Fluctuation have insignificant effect on the simulation results, and is mostly classified in insignificant sensitivity class (P>10%).

Node 6 (Yellow zone) is located in the lower and looping part of network away from the mains. Injection at node 6 will highly affect nodes lying in its vicinity while other network nodes receive less chemical concentrations. Change in chemical loading and base demand will affect the simulation output, Node Count. GSA results also showed that variation in base demand and variation in contaminant loading has maximum influence in the simulation. Node 81 (Green zone) lies in bigger loop at the upper west side of main trunk line. Propagation of chemical in that bigger loop is largely driven by demand of nodes lying in that sector of network. A node in that sector receives higher chemical concentration than other nodes in the network. Since the numbers of nodes that are affected by injection made at node 81 are limited, any fluctuation in base demand and chemical loading will affect the concentration received at nodes.

A common average concentration threshold value of 2.0 mg/l was selected for all three GSA experiments to find the influence of the four dynamic variables in the network vulnerability. Based upon the K-S test *d* statistics and *P* values, the four dynamics variables can be ranked depending upon their sensitivity as shown in Table 13:

Table 13: Sensitivity ranking of dynamic variables based upon K-S *d* statistic for average chemical concentration threshold of 2.0 mg/l.

Input Parameters	**Injection Node**		
	Node 17	**Node 6**	**Node 81**
	Rank		
Base Demand	1	2	1
Injection Duration	3	3	4
Tank Head Fluctuation	2	4	3
Chemical Mass Loading	4	1	2

Figure 15a, 15b, and 15c illustrates how K-S *d* statistic was calculated from the difference between the CDFs of input variables grouped in "pass" and "fail" sample sets. The input parameters were grouped into three sensitivity classes a) Critical/Significant (P<1%), b) Important (1% ≤ P ≤ 10%), and c) Insignificant (P > 10%), based upon the significance level for rejecting the null hypothesis. Only the input variables with highest and lowest sensitivity rank are presented in Figures 15a, 15b, and 15c.

For the injection made at node 17 lying in Red zone of influence (see Figure 15a), the highest ranked dynamic variable was Base demand with K-S *d* statistic value of 0.84, and fell in 'Critical' sensitivity class. Chemical mass loading with K-S *d* statistic value of 0.15 was ranked lowest and fell in 'Important' sensitivity class. Similarly, the dynamic variables Tank Head Fluctuation and Injection Duration both were grouped in 'Critical' sensitivity classes. The data range for Base Demand that fell in "fail" category shows that for the variation in lower base demands values, the injection was able to infect more than one node above the threshold concentration. This signifies the strong influence of Base Demand variation on to the simulation output.

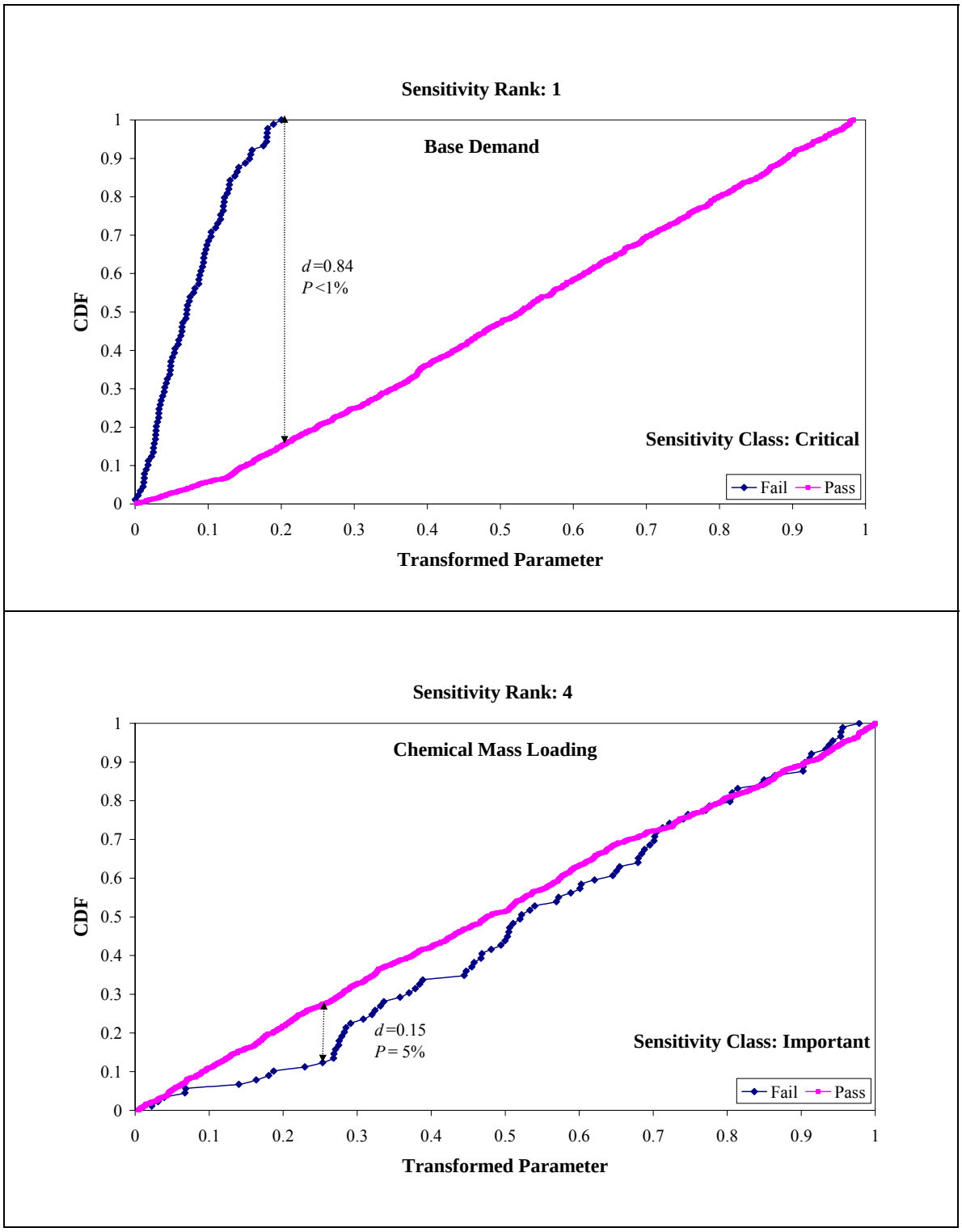

Figure 15a: Highest and Lowest ranked dynamic variables, for injection made at node 17 lying in Red zone and threshold of 2.0 mg/l.

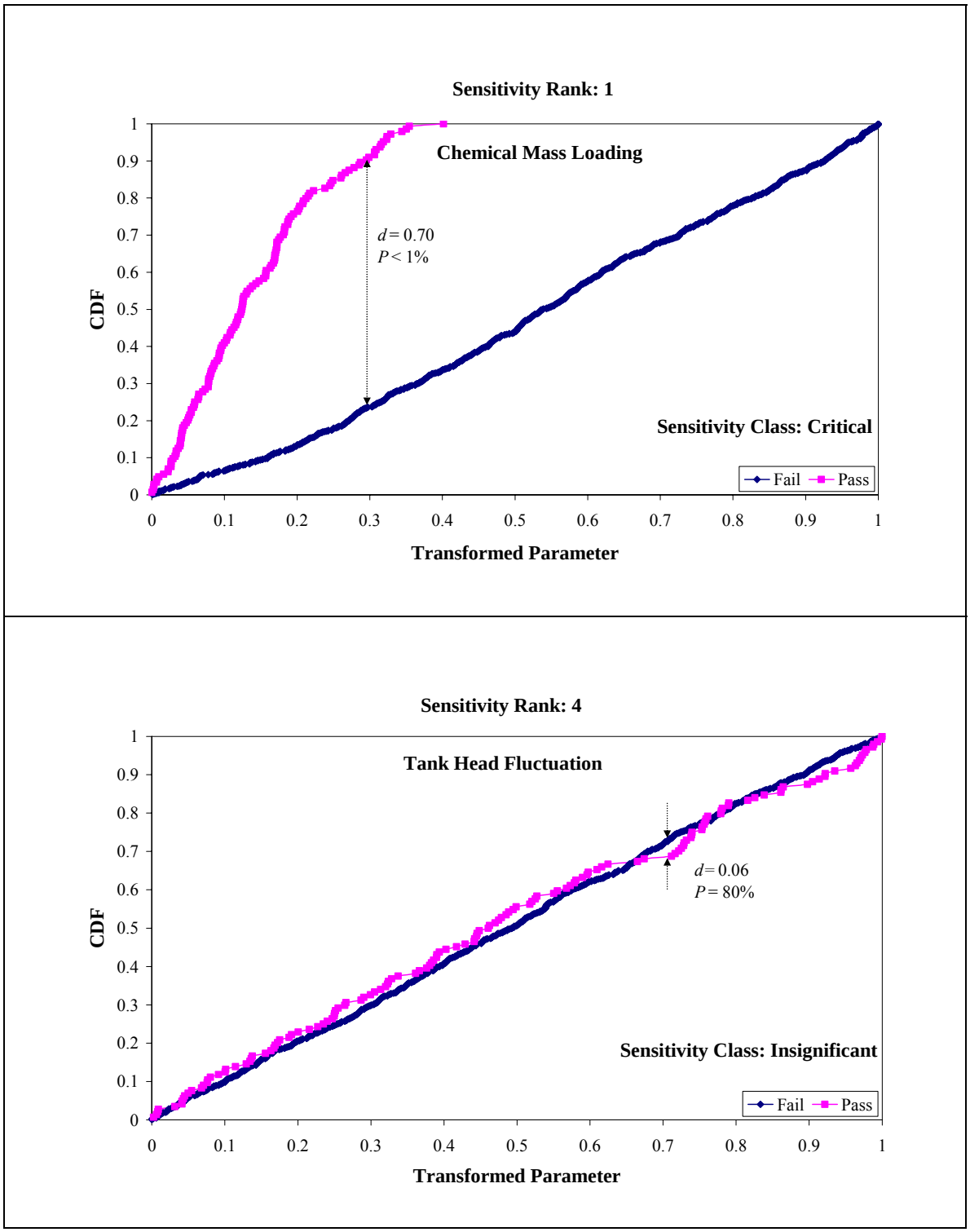

Figure 15b: Highest and Lowest ranked dynamic variables, for injection made at node 6 lying in Yellow zone and threshold of 2.0 mg/l.

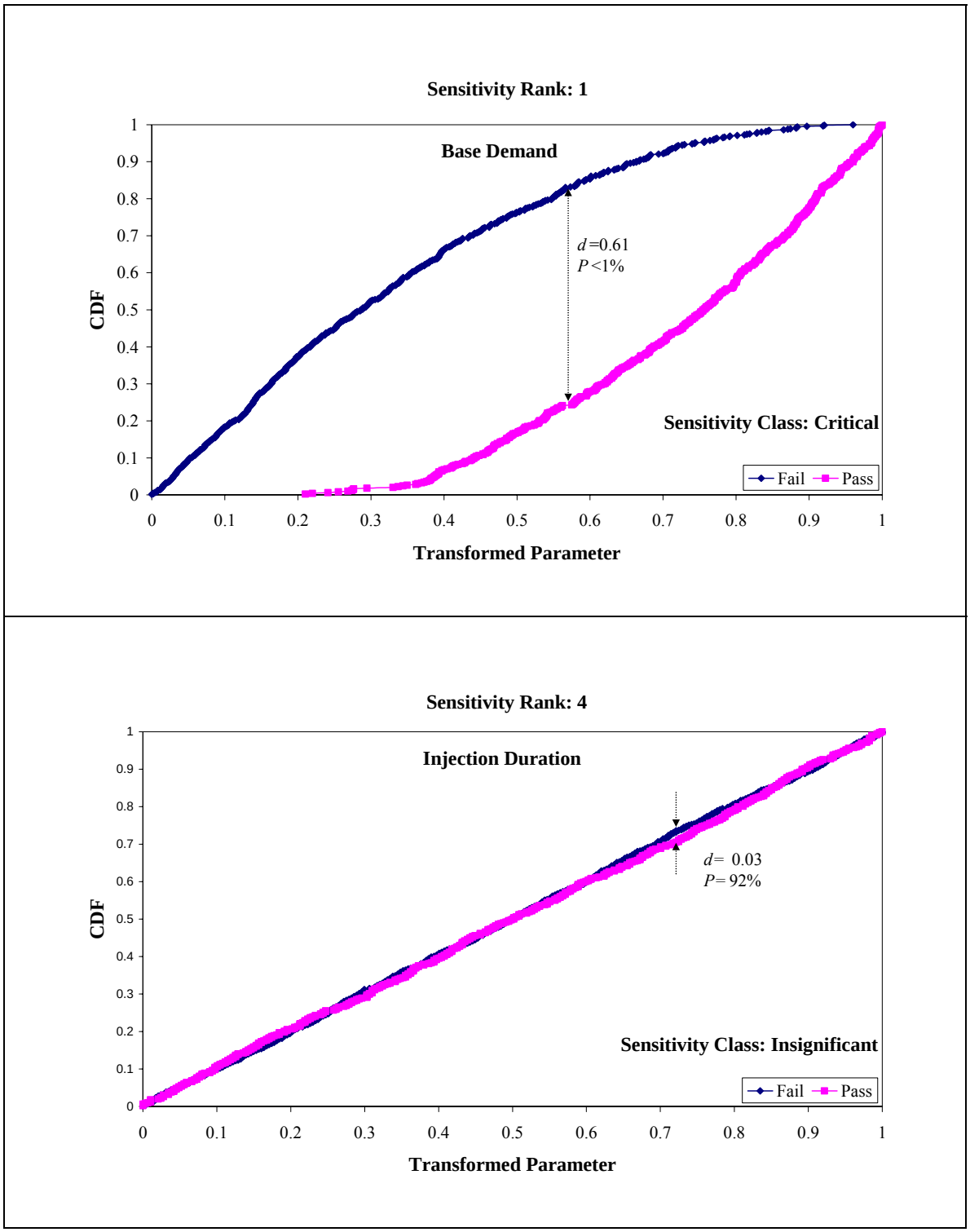

Figure 15c: Highest and Lowest ranked dynamic variables, for injection made at node 81 lying in Green zone and threshold of 2.0 mg/l.

When injection was made at node 6 lying in Yellow zone of influence (see Figure 15b), Chemical mass loading was ranked highest with K-S *d* statistic value of 0.70, and fell in 'Critical' sensitivity class. Tank Head Fluctuation with K-S *d* statistic value of 0.06 was ranked lowest and fell in 'Insignificant' sensitivity class. Similarly, the dynamic variables Base demand and Injection Duration were grouped in 'Critical' and 'Insignificant' sensitivity classes respectively.

Sensitivity ranking for injection made at node 81 lying in Green zone of influence assigned highest and lowest ranking to Base Demand (d=0.61) and Injection Duration (d=0.03) respectively. Base Demand and Chemical mass loading were grouped in 'Critical' sensitivity class while Tank Head Fluctuation and Injection Duration fell in 'Insignificant' sensitivity class.

The sensitivity ranking consistently shows Base Demand as the most significant input parameter for injection made at all three locations. For effective Monte Carlo simulation of network intrusion, Base Demand cannot be omitted as a significant dynamic variable to be simulated. Similarly, Chemical Mass Loading was also identified as another significant variable, and tank head operation proved to be significant when injection was made in the Red zone node.

In order to have complete assessment of vulnerability to network intrusions, the variation of behavior in these consumers and utilities guided variables should be included in simulation along with the variation in attack logistics. For example, TEVA (Murray *et*

al., 2004) considers an ensemble of threat scenarios by incorporating variability in the assailant's behavior only. A more representative simulation of network vulnerability can be achieved if dynamic variables like Base Demand and Tank Head Fluctuation are also included in the threat ensemble.

5.6.2 Nodal Mass Threshold

In the three GSA experiment presented above, average concentration threshold was used as a parameter to define vulnerability of a node. Instead of average concentration, other parameters like nodal mass consumption, number of population effected at a node, etc. can be used to define the node vulnerability. A new set of GSA was performed for the injection made at node 17 (Red zone) but this time with nodal mass consumed as a parameter to define the nodal vulnerability. The same four input variables and their range as mentioned in Table 6 were used for the calculation.

A nodal mass consumed threshold was considered, such that any node receiving nodal mass above the threshold value at the end of simulation will be considered vulnerable and the value of the Node Count will be increased by 1. Taking into account the consistency of K-S statistics for consecutive threshold values and the limitation of K-S two-sample test, a range of threshold value was considered. Figure 16 shows the plot of K-S *d* statistic versus different nodal mass consumed thresholds considered for the GSA.

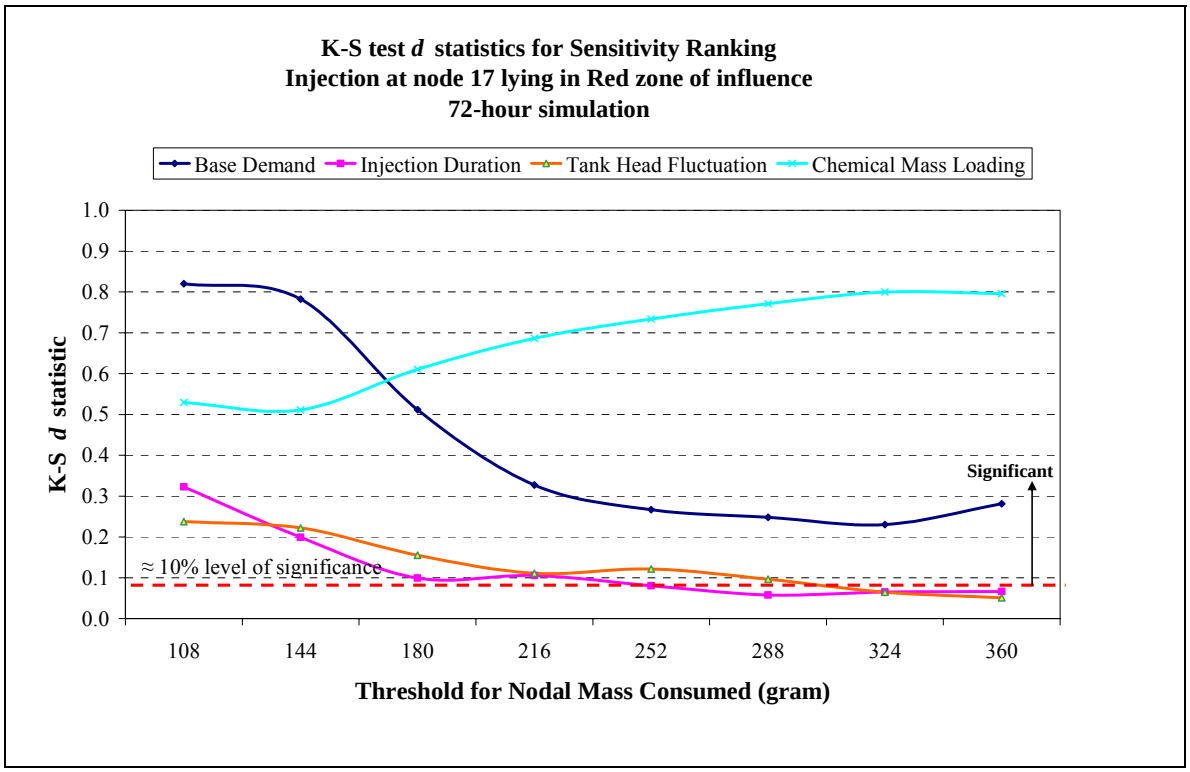

Figure 16: Variation in Sensitivity Ranking with respect to nodal mass consumed threshold value when injection is made at node 17 (Red zone).

Table 14 summarizes the K-S test statistics for injection made at node 17 (Red zone) with nodal mass consumed in grams taken as a threshold to define vulnerability of a node.

This K-S statistics results in Table 14 and K-S *d* statistic plotted in Figure 16 shows that for a range of nodal mass consumed threshold, input variables Base Demand and Chemical Mass Loading falls in "significant" sensitivity class. When average concentration threshold was considered as a parameter to define node vulnerability, for injection made at node 17 (see Figure 14a), Base Demand and Tank Head fluctuation were two most significant input variables.

Table 14: K-S test statistics for injection made at node 17 (Red Zone) with nodal mass consumed as a threshold; 72-hour simulation.

Nodal Mass Consumed Threshold	Input Parameter								Total number of samples $N_1 + N_2 = 1152$	
	Base Demand [380-7600] Lpm		Injection Duration [15-4320] min		Tank Head Fluctuation [0.3-30.5] m		Chemical Mass Loading [2000-9000] g			
(gram)	*d*	*P*	*d*	*P*	*d*	*P*	*d*	*P*	N_1(Pass) Node Count = 0	N_2(Fail) Node Count > 0
108	0.8201	0.00	0.3226	0.00	0.2376	0.04	0.5292	0.00	36	1116
144	0.7824	0.00	0.1992	0.01	0.2219	0.00	0.5113	0.00	72	1082
180	0.5116	0.00	0.0991	0.17	0.1552	0.00	0.6102	0.00	145	1007
216	0.3274	0.00	0.1055	0.03	0.1109	0.02	0.6863	0.00	233	919
252	0.2668	0.00	0.0805	0.10	0.1217	0.00	0.7338	0.00	324	828
288	0.2480	0.00	0.0576	0.34	0.0967	0.01	0.7711	0.00	411	741
324	0.2304	0.00	0.0653	0.18	0.0648	0.18	0.7998	0.00	504	648
360	0.2813	0.00	0.0664	0.16	0.0510	0.44	0.7957	0.00	590	562

But when nodal mass consumed threshold was used to define the node vulnerability, the significance of Tank Head fluctuation decreased in terms of Chemical Mass loading. In both GSA studies Base Demand was recognized as one of the significant input variable.

Figure 17 shows the highest and lowest ranked dynamic variables, for nodal mass threshold of 216 gram. Base Demand was ranked higher with K-S *d* statistic value of 0.69 and fell into "Significant" sensitivity class. Injection Duration was ranked lowest with *d* statistic value of 0.10; it was categorized in "Important" sensitivity class along with Tank Head Fluctuation. Chemical Mass Loading was another variable in "Significant" sensitivity class.

For all of the GSA study presented above, Injection duration proved to be insignificant variable but this result was totally dependent upon the parameter used to define node vulnerability. Since the parameters used to define node vulnerability were average concentration and nodal mass consumed thresholds, which will be calculated at the end of simulation period only, the influence of Injection duration was not properly taken into account. For example, if a node receives higher chemical concentration or higher mass loading at certain point during the simulation period, the effect from that higher mass (or concentration) would not be correctly reflected when the chemical mass or average concentration was calculated at the end of simulation period.

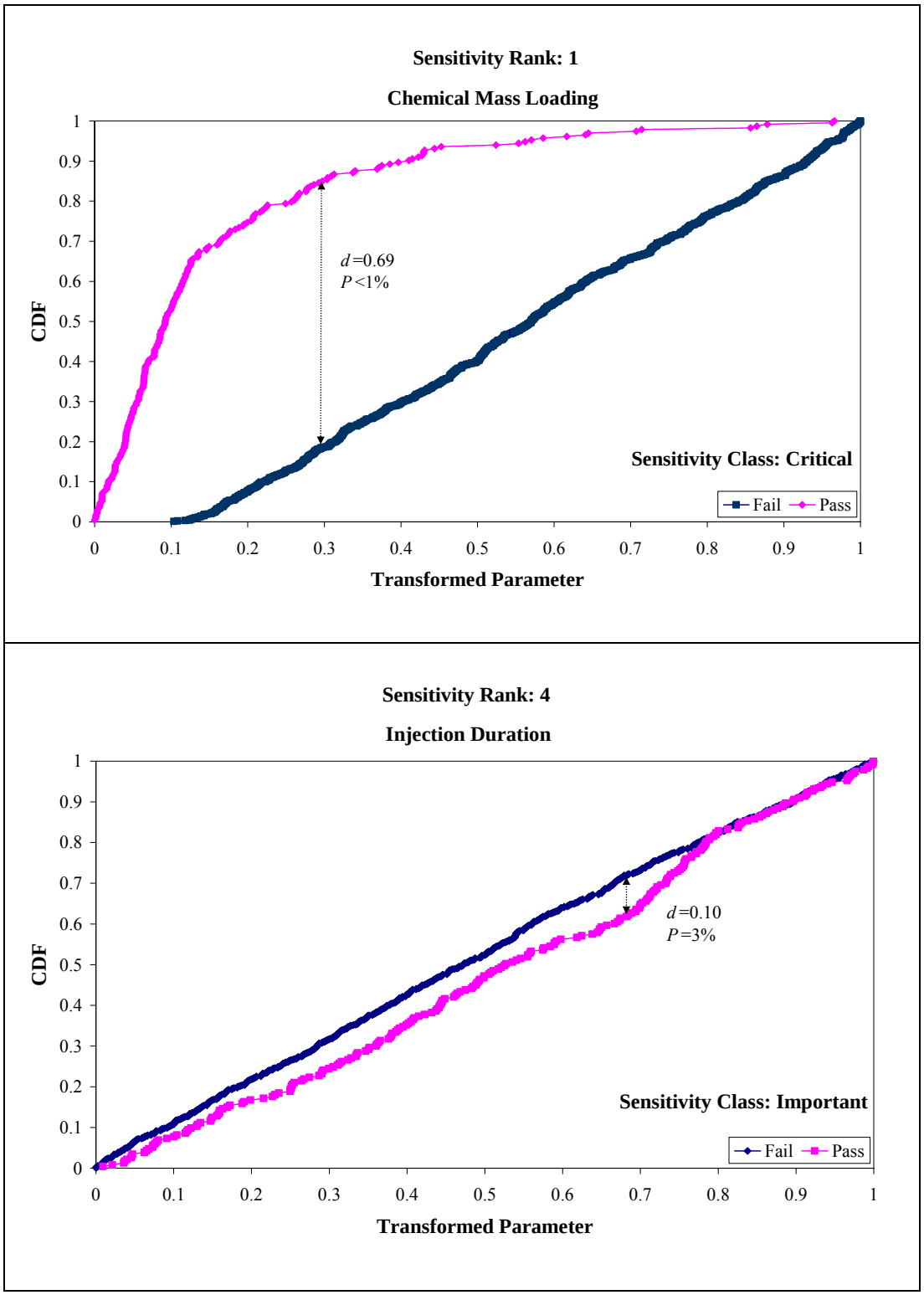

Figure 17: Highest and Lowest ranked dynamic variables, for nodal mass threshold of 216 gram, injection made at node 17 (Red zone).

Additional GSA runs were performed with both nodal mass and average concentration thresholds, but this time the total simulation period was limited to only 24 hours. For the injection made at node 17 (Red zone), and total simulation period of 24-hour, two more GSA study was performed, 1) taking nodal mass consumed as a threshold and 2) taking average concentration as a threshold to define node vulnerability.

Table 15 and 16 summarizes the K-S *d* and *P* statistics when nodal mass consumed and average chemical concentration respectively were taken as thresholds to define node vulnerability. Figure 18 shows the plot of K-S *d* statistics versus the average concentration and nodal mass consumed threshold, when the simulation was carried out for 24 hour.

When a Base Case injection made at node 17 only 33% of mass was consumed at nodes during the 24 hour simulation period, which tells that tank retains larger percentage of injected chemical mass. But for the GSA study, the injection duration was varied from initial 15 minutes period to total 24 hour period.

For the short duration injections made during initial simulation hours, it is possible that more percentage of mass is consumed at nodes, in this case the consumption largely depends upon the demand at the nodes. For longer duration injections, the rate of chemical mass being introduced into the system will be slower, and so does the mass consumption.

Table 15: K-S test statistics for injection made at node 17 (Red Zone) with nodal mass consumed as a threshold; 24-hour simulation.

Nodal Mass Consumed Threshold	Input Parameter								Total number of samples $N_1 + N_2 = 1152$	
	Base Demand [380-7600] Lpm		Injection Duration [15-4320] min		Tank Head Fluctuation [0.3-30.5] m		Chemical Mass Loading [2000-9000] g			
(gram)	*d*	*P*	*d*	*P*	*d*	*P*	*d*	*P*	N_1(Pass) Node Count = 0	N_2(Fail) Node Count > 0
72	0.9150	0.00	0.1492	0.44	0.2887	0.01	0.5278	0.00	35	1117
108	0.8447	0.00	0.1671	0.04	0.2173	0.00	0.3919	0.00	73	1079
144	0.8225	0.00	0.1198	0.12	0.1697	0.01	0.3893	0.00	109	1043
180	0.5500	0.00	0.1125	0.03	0.1052	0.06	0.5167	0.00	192	960
216	0.4341	0.00	0.1170	0.00	0.1195	0.00	0.5942	0.00	294	858
252	0.3989	0.00	0.0804	0.07	0.1133	0.00	0.6501	0.00	390	762
288	0.3712	0.00	0.0599	0.27	0.1047	0.00	0.7020	0.00	462	690
324	0.3545	0.00	0.0406	0.73	0.0727	0.10	0.7333	0.00	544	608
360	0.3602	0.00	0.0475	0.54	0.0852	0.03	0.7792	0.00	612	540

Table 16: K-S test statistics for injection made at node 17 (Red Zone) with average concentration as a threshold; 24-hour simulation.

Average Concentration Threshold	Input Parameter								Total number of samples $N_1 + N_2 = 1152$	
	Base Demand [380-7600] Lpm		Injection Duration [15-4320] min		Tank Head Fluctuation [0.3-30.5] m		Chemical Mass Loading [2000-9000] g			
(mg/l)	*d*	*P*	*d*	*P*	*d*	*P*	*d*	*P*	N_1(Pass) Node Count = 0	N_2(Fail) Node Count > 0
0.4	0.6201	0.00	0.1385	0.16	0.0813	0.78	0.8226	0.00	70	1082
0.8	0.5050	0.00	0.0535	0.48	0.0644	0.26	0.5250	0.00	353	799
1.2	0.5883	0.00	0.0545	0.37	0.0849	0.03	0.4212	0.00	665	487
1.6	0.6553	0.00	0.0561	0.47	0.1140	0.01	0.4583	0.00	841	311
2	0.7194	0.00	0.0503	0.82	0.1667	0.00	0.4606	0.00	964	188
2.4	0.7042	0.00	0.1274	0.06	0.2873	0.00	0.4980	0.00	1031	121
2.8	0.7799	0.00	0.1055	0.49	0.2854	0.00	0.4884	0.00	1086	66

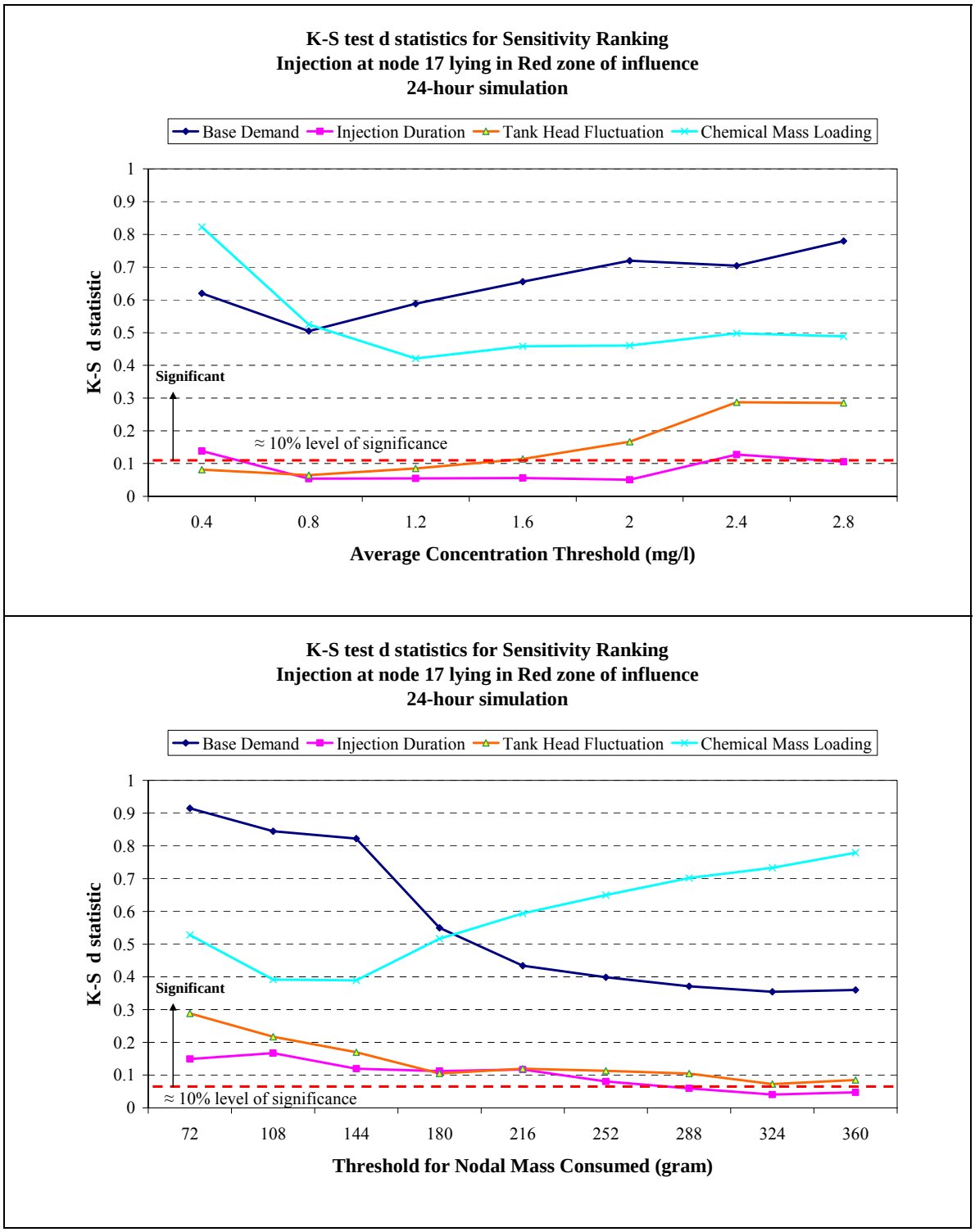

Figure 18: Variation in Sensitivity Ranking with respect to average concentration and nodal mass consumed threshold for 24-hour simulation (injection node 17)

If less amount of chemical mass is injected during longer injection duration, then it will have less effect on consumer as opposed to larger chemical mass injected for longer duration. For any scenario, the amount of mass received at a node is dependent upon the chemical mass injected and the demand associated with the node. For the 24 hour simulation also Base Demand and Chemical Mass Loading were found to be most significant dynamic variables.

6. SUMMARY AND CONCLUSIONS

Network intrusions at multiple locations provide a broader perspective of network vulnerability than an attack made on a single location. Attack on multiple network nodes and an assumption on proportion of population located at nodes provided a means to calculate Exposure Index for a network. The Exposure Index was demonstrated as a measure of network vulnerability which is a simple way of interpreting the simulation results to the utility.

The Zone of Influence plot for a network helps utilities to identify the most vulnerable zones of their network and bolster the nodes lying in those areas against possible contaminant threats. Identification of vulnerable areas in a network helps planning for pre and post contamination events like installation of early warning systems, providing alternative route for drinking water in such emergency situations, planning effective risk communication strategy, and isolating a portion of network in order to prevent further contaminant propagation.

Sensitivity of network vulnerability changed significantly under the influence of static and dynamic network variables. The deterministic and stochastic simulation of intrusions on the network assessed the influence of injection location, injection timing, tank mixing, total simulation period, and system demand on the network vulnerability. While the variation in static variables are network specific and will not change randomly, the stochastic variation in dynamic variables had significant effect on network vulnerability.

This study demonstrated the application of Generalized Sensitivity Analysis using Kolmogorov-Smirnov *d* statistic to rank the dynamic variables that are significant for Monte Carlo simulation of network intrusions. GSA provided a basis to identify and the uncertain network variables, variability of which has to be simulated in order to have broad perspective on network vulnerability and consumer exposure.

The sensitivity ranking of the dynamic variables changed as the threshold values and injection location were changed, but Base Demand and Chemical Mass Loading were consistently identified as significant variable that have to be included in Monte Carlo simulation. Fluctuation in tank head was also significantly affecting the simulation results when injection was made at node lying in Red zone and average concentration threshold value was used to node vulnerability.

The parameter used to define node vulnerability can be problem specific but from the GSA study conducted with different parameters, the results obtained were quite similar. When the parameter to define node vulnerability was changed from average concentration to nodal mass consumed, the effect of variation in Chemical Mass Loading and Base Demand were still more significant than other variables. GSA performed for shorter simulation duration also identified the same dynamic variables as the most significant ones.

Thus, it is important to include all possible combination of assailant's behavior i.e. contaminant scenarios, while simulating a network for its vulnerability to an attack, but the variation in behavior of consumer and utility also have significant effect on determining the network vulnerability and should not be ignored. Variation in demand is always an important variable that have to be included in simulation to assess network vulnerability, while the effect of tank head variation can be significant or insignificant based upon the injection location and network properties.

Further applicability on GSA can be improved by increasing the sample size, increasing the number of input variables, and using different parameters to define network vulnerability. The observations from this study are hypothetical and simulation was done for worst-case scenario taking conservative chemical, but if representative data on threshold value or dosage for specific chemicals are known, GSA can be more effective in order to identify the sensitiveness of these dynamic variables to network vulnerability.

REFERENCES

Auslander, D.M., Spear, R.C., and Young, G.E. (1982). "A simulation-based approach to the design of control systems with uncertain parameters." *Trans. ASME*, 104(1), 20-26.

Bioterrorism Act 2002: *Public Health Security and Bioterrorism Preparedness and Response Act* (Bioterrorism Act), PL 107-188, United States Congress, June 2002. http://www.fda.gov/oc/bioterrorism/PL107-188.html#title4 [accessed 05/24/2005].

Buchberger, S. G., and Wu, L. (1995). "Model for instantaneous residential water demands." *Journal of Hydraulic Engineering*, 121(3), 232-246.

Buchberger, S.G., and Wells, G.J. (1996). "Intensity, duration, and frequency of residential water demands." *Journal of Water Resources Planning and Management*, 122(1), 11-19.

Buchberger, S.G., Carter, J.T., Lee, Y.H., and Schade, T.G. (2003) "Random demands, travel times, and water quality in deadends." *American Water Works Association Research Foundation*, Report 90963F, DeNCer, Colorado, 470 pages.

Clark, R.M., Abdesaken, F., Boulos, P.F., and Mau, R.E. (1996) "Mixing in distribution system storage tanks: Its effect on water quality." *Journal of Environmental Engineering*, 122(9), 814-821

Clark, R.M. and Buchberger, S.G. (2004). "Responding to a contamination threat in a drinking water network: the potential for modeling and monitoring." *Water Supply System Security*, edited by Larry W. Mays, McGraw-Hill, N.Y., 9.1-9.26.

Cox, B.A., and Whitehead, P.G. (2005). "Parameter sensitivity and predictive uncertainty in a new water quality model, Q^2." *Journal of Environmental Engineering*, 131(1), 147-157.

Garcia, V.J., Garcia-Bartual, R., Cabrera, E., Arregui, F., and Garcia-Serra, J. (2004) "Stochastic model to evaluate residential water demands." *Journal of Water Resources Planning and Management*, 130(5), 386-394.

Gargano, R. and Pianese, D. (2000). "Reliability as tool for hydraulic network planning." *Journal of Hydraulic Engineering*, 126(5), 354-364.

Guercio, R., Magini, R. and Pallavicini, I. (2001). "Instantaneous residential water demand as stochastic point process." *Proceedings First International Conference on Water Resources Management*, WIT Press, Southhampton, UK, 129-138.

James, R. Bruce, Gwo, Jin-Ping, and Toran, Laura. (1996). "Risk-cost decision framework for aquifer remediation design." *Journal of Water Resources Planning and Management*, 122(6), 414-420.

Makino, H, McKenna, S.A, and Wakasugi, K. (2001) "Sensitivity analysis of Monte Carlo results using the Kolomogorov-Smirnov *d* statistic." *International Association of Mathematical Geology, Annual Meeting,* Cancun, Mexico. 1-16.

Murray, R., Janke, R., and Uber, J. (2004). "The Threat Ensemble Vulnerability Assessment (TEVA) program for drinking water distribution system security." *Proceedings of EWRI Congress*, Salt Lake City. UT, June, 2004.

Nilsson, K., Buchberger, S.G., and Clark, R.M. (2005). "Simulating exposures to deliberate intrusions into water distribution systems." *Journal of Water Resources Planning and Management*, 131(3), 228-236.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flammery, B.P. (2002). "Statistical Description of Data*." Numerical recipes in C++: the art of scientific computing*, 2nd ed., Cambridge University Press, New York, 625-633.

Propato, M. and Uber, J.G. (2004). "Vulnerability of water distribution systems to pathogen intrusion: how effective is a disinfectant residual?" *Environmental Science and Technology*, 38(13), 3713-3722.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis In Practice: A Guide To Assessing Scientific Models.* John Wiley & Sons Ltd. England.

Shortencarier, M.J. and Helton, J.C. (1999). "A FORTRAN 77 program and user's guide for the statistical analysis of scatterplots to identify important factors in large scale-simulations." Sandia National Laboratories, New Mexico.

Spear, R.C., and Hornberger, G.M. (1980). "Eutrophication in Peel Inlet-II. Identification of critical uncertainties via generalized sensitivity analysis." *Water Research*, 14(1), 43-49.

Walski, T.M., Chase, D.V., Savic, D.A., Grayman, W.M., Bickwith, S., and Koelle, E. (2003). *Advanced Water Distribution Modeling and Management*. Haestad Press, CT.

APPENDICES

APPENDIX I

1. EPANET input file for Base-Case Attack at node 50

2. PRPsym input file for Base-Case Attack at node 50

3. EPANET input file generated by PRPsym for Base-Case Attack at node 50

1. EPANET input file for Base-Case Attack at node 50

CSTR.inp

[TITLE]
"Base-Case"
Injection Loading = 20 mg/min

[JUNCTIONS]

;ID	Elev	Demand	Pattern	
1	50	0.00		;
2	90	4.97		;
3	100	5.72		;
4	75	3.38		;
5	100	5.44		;
6	75	4.13		;
9	60	5.25		;
11	75	3.75		;
14	75	3.75		;
16	100	1.50		;
17	100	1.88		;
19	115	1.13		;
20	115	2.25		;
21	125	2.25		;
22	125	2.63		;
24	150	4.13		;
26	160	4.60		;
27	125	3.47		;
28	110	0.00		;
31	130	6.94		;
32	180	0.00		;
33	150	6.94		;
35	180	6.75		;
36	140	0.75		;
37	180	3.75		;
38	185	0.00		;
39	140	0.38		;
40	140	0.75		;
41	190	3.00		;
42	140	1.50		;
43	120	0.75		;
44	100	0.75		;
45	190	1.50		;
47	200	6.11		;
49	210	3.00		;
50	210	1.50		;Injection Node
51	195	2.81		;
52	180	2.44		;
53	165	1.88		;

54	150	1.50	;
56	150	13.33	;
59	150	6.67	;
60	150	2.22	;
61	150	4.44	;
63	110	5.56	;
65	135	10.00	;
67	100	0.00	;
68	100	4.44	;
69	100	8.89	;
71	160	8.89	;
72	160	1.50	;
78	180	1.50	;
79	183	2.25	;
81	186	2.63	;
83	188	2.25	;
84	190	0.75	;
85	200	1.50	;
86	190	1.50	;
88	190	3.00	;
89	170	6.00	;
91	162	4.88	;
94	150	4.13	;
97	190	3.00	;
98	200	4.88	;
102	185	2.25	;
103	180	1.88	;
104	182	1.50	;
105	190	3.00	;
106	190	0.75	;
109	190	11.11	;
110	230	0.38	;
116	230	11.11	;
117	230	0.75	;
118	230	3.75	;
119	190	4.50	;
122	145	3.38	;
123	110	2.25	;
124	110	3.00	;
125	110	2.25	;
127	110	4.50	;
129	130	3.38	;
130	110	1.50	;
131	190	12.22	;
133	185	21.39	;
136	118	6.67	;
137	120	1.50	;
139	190	0.75	;
140	190	3.00	;

[RESERVOIRS]

;ID Head Pattern
7 50 ;

[TANKS]
;ID Elevation InitLevel MinLevel MaxLevel Diameter MinVol VolCurve
138 235 56.7 56 66 50 0 ;

[PIPES]
;ID Node1 Node2 Length Diameter Roughness MinorLoss Status
2 2 3 550 12 100 0 Open ;
3 3 4 526 8 100 0 Open ;
11 11 4 288 8 100 0 Open ;
16 9 14 576 8 100 0 Open ;
17 14 16 530 12 100 0 Open ;
18 16 17 583 8 100 0 Open ;
21 17 19 815 12 100 0 Open ;
22 19 22 431 12 100 0 Open ;
23 19 20 88 8 100 0 Open ;
25 21 22 288 8 100 0 Open ;
30 26 27 767 12 100 0 Open ;
31 27 28 312 12 100 0 Open ;
35 26 32 383 12 100 0 Open ;
38 32 35 1438 8 100 0 Open ;
41 32 38 743 12 100 0 Open ;
44 38 41 479 12 100 0 Open ;
54 49 50 600 12 100 0 Open ;
55 50 51 420 8 100 0 Open ;
56 51 52 420 8 100 0 Open ;
57 52 53 420 8 100 0 Open ;
58 53 54 420 8 100 0 Open ;
64 59 60 360 8 100 0 Open ;
65 59 61 364 8 100 0 Open ;
67 61 136 838 8 100 0 Open ;
71 63 67 240 12 100 0 Open ;
72 5 68 300 12 100 0 Open ;
73 68 69 623 12 100 0 Open ;
75 67 71 647 12 100 0 Open ;
76 71 72 479 12 100 0 Open ;
83 71 78 216 12 100 0 Open ;
84 78 79 375 8 100 0 Open ;
87 79 81 375 8 100 0 Open ;
90 81 83 375 8 100 0 Open ;
91 83 84 375 8 100 0 Open ;
92 84 85 300 12 100 0 Open ;
93 50 85 400 12 100 0 Open ;
94 85 86 240 12 100 0 Open ;
97 86 89 623 12 100 0 Open ;
101 91 97 527 8 100 0 Open ;
107 97 98 527 8 100 0 Open ;
113 102 103 407 8 100 0 Open ;
114 103 104 312 8 100 0 Open ;

115	103	105	455	8	100	0	Open	;
116	84	106	264	12	100	0	Open	;
120	106	110	647	12	100	0	Open	;
127	110	117	240	12	100	0	Open	;
128	117	118	551	8	100	0	Open	;
129	117	119	671	12	100	0	Open	;
133	119	123	1150	8	100	0	Open	;
134	123	124	455	8	100	0	Open	;
135	124	125	551	8	100	0	Open	;
138	127	124	407	8	100	0	Open	;
139	127	129	1007	8	100	0	Open	;
141	125	130	431	8	100	0	Open	;
143	129	122	335	8	100	0	Open	;
145	51	83	700	6	100	0	Open	;
146	54	78	700	8	100	0	Open	;
156	136	63	216	8	100	0	Open	;
1	1	2	1850	12	100	0	Open	;
158	138	117	200	12	100	0	Open	;
159	137	129	743	8	100	0	Open	;
160	104	139	300	8	100	0	Open	;
161	140	104	400	12	100	0	Open	;
48	127	125	790	8	100	0	Open	;
49	119	122	1413	8	100	0	Open	;
50	110	116	1703	12	100	0	Open	;
122	106	109	887	12	100	0	Open	;
123	86	88	766	8	100	0	Open	;
124	89	98	1200	8	100	0	Open	;
125	89	91	528	8	100	0	Open	;
126	49	47	1079	8	100	0	Open	;
130	91	94	887	8	100	0	Open	;
131	94	97	768	8	100	0	Open	;
132	102	45	250	8	100	0	Open	;
136	98	102	647	8	100	0	Open	;
137	41	131	700	8	100	0	Open	;
142	38	133	1821	8	100	0	Open	;
144	32	37	1102	12	100	0	Open	;
148	81	52	700	6	100	0	Open	;
149	79	53	700	6	100	0	Open	;
150	54	56	384	12	100	0	Open	;
151	136	65	1533	12	100	0	Open	;
152	65	63	839	8	100	0	Open	;
154	71	56	1104	8	100	0	Open	;
155	56	59	958	8	100	0	Open	;
157	49	41	1796	12	100	0	Open	;
36	5	69	1006	12	100	0	Open	;
37	27	33	500	8	100	0	Open	;
40	28	31	1127	8	100	0	Open	;
42	26	24	576	12	100	0	Open	;
43	24	42	1100	8	100	0	Open	;
45	42	43	300	8	100	0	Open	;
46	42	39	300	8	100	0	Open	;

47	39	40	400	8	100	0	Open	;
51	39	36	400	8	100	0	Open	;
52	9	11	910	6	100	0	Open	;
53	14	11	1072	8	100	0	Open	;
59	17	3	838	12	100	0	Open	;
60	9	6	1000	8	100	0	Open	;
61	2	44	400	8	100	0	Open	;
62	24	22	2500	12	100	0	Open	;
4	5	67	83	12	100	0	Open	;
5	6	4	870	8	100	0	Open	;

[PUMPS]

;ID	Node1	Node2	Parameters	
0	7	1	HEAD 1	;

[VALVES]

;ID	Node1	Node2	Diameter	Type	Setting	MinorLoss

[TAGS]

[DEMANDS]

;Junction	Demand	Pattern	Category

[STATUS]

;ID	Status/Setting

[PATTERNS]

;ID	Multipliers					
;Demand Pattern						
1	1.26	1.04	.97	.97	.89	1.19
1	1.28	.67	.67	1.34	2.46	.97
1	.92	.68	1.43	.61	.31	.78
1	.37	.67	1.26	1.56	1.19	1.26
1	.6	1.1	1.03	.73	.88	1.06
1	.99	1.72	1.12	1.34	1.12	.97
1	1.04	1.15	.91	.61	.68	.46
1	.51	.74	1.12	1.34	1.26	.97
1	.82	1.37	1.03	.81	.88	.81
1	.81					
;Injection for 6 hrs- peak						
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	1	1	1
Inj6p	1	1	1	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0

```
 Inj6p          0       0       0       0       0       0
 Inj6p          0       0       0       0       0       0
 Inj6p          0       0       0       0       0       0
 Inj6p          0       0       0       0       0       0

[CURVES]
;ID             X-Value         Y-Value
;PUMP:           Pump Curve
 1              666.62          259.88

[CONTROLS]
Link 0 Closed if Node 138 Above 65
Link 0 Open if Node 138 Below 56

[RULES]

[ENERGY]
 Global Efficiency      75
 Global Price           0
 Demand Charge          0

[EMITTERS]
;Junction       Coefficient

[QUALITY]
;Node           InitQual

[SOURCES]
;Node           Type            Quality         Pattern
 50             MASS            20000           Inj6p

[REACTIONS]
;Type           Pipe/Tank       Coefficient

[REACTIONS]
 Order Bulk             1
 Order Tank             1
 Order Wall             1
 Global Bulk            0
 Global Wall            0
 Limiting Potential     0
 Roughness Correlation  0

[MIXING]
;Tank           Model

[TIMES]
 Duration               72:00
 Hydraulic Timestep     1:00
 Quality Timestep       0:01
 Pattern Timestep       1:00
```

Pattern Start		0:00
Report Timestep		0:01
Report Start		0:00
Start ClockTime		9 am
Statistic	NONE	

[REPORT]

Status	No	
Summary		No
Page	0	

[OPTIONS]

Units	GPM	
Headloss		H-W
Specific Gravity		1
Viscosity	1	
Trials	40	
Accuracy		0.001
Unbalanced		Continue 10
Pattern	1	
Demand Multiplier		1.0
Emitter Exponent		0.5
Quality	Chemical mg/L	
Diffusivity		1
Tolerance		0.01

[COORDINATES]

;Node	X-Coord	Y-Coord
1	1373.55	225.29
2	1359.01	1395.35
3	1359.01	1809.59
4	937.50	1758.72
5	1410.23	7226.40
6	365.89	1890.10
9	675.87	2529.07
11	937.50	1984.01
14	944.77	2754.36
16	1359.01	2739.83
17	1359.01	2449.13
19	1969.48	2311.05
20	1955.30	2177.98
21	2441.86	2005.81
22	2289.24	2216.57
24	2739.83	4018.90
26	3190.41	4055.23
27	3372.09	3502.91
28	3473.84	3125.00
31	3364.83	2267.44
32	3204.94	4353.20
33	3108.61	3304.11
35	2318.31	4716.57

36	2140.78	4656.27
37	4040.70	4527.62
38	3219.48	4912.79
39	1921.97	4403.79
40	1649.85	4594.55
41	3263.08	5276.16
42	2028.57	4235.47
43	1966.85	4033.49
44	918.90	1379.91
45	5713.87	8099.24
47	4164.24	6257.27
49	3452.03	6497.09
50	3350.29	6947.67
51	3037.79	6954.94
52	2710.76	6925.87
53	2398.26	6911.34
54	2093.02	6889.53
56	2009.45	6711.48
59	1744.19	6250.00
60	1773.26	5973.84
61	1533.43	6242.73
63	1489.83	7042.15
65	984.79	6517.55
67	1482.56	7209.30
68	1162.79	7223.84
69	1177.33	7681.69
71	1947.67	7369.19
72	1925.87	7739.83
78	2100.29	7412.79
79	2354.65	7420.06
81	2651.28	7470.48
83	2943.31	7470.93
84	3252.71	7506.56
85	3292.15	7231.10
86	3626.45	7347.38
88	3837.21	7739.83
89	4077.03	7521.80
91	4447.67	7361.92
94	4738.37	6809.59
97	4854.65	7405.52
98	5130.81	7681.69
102	5545.06	7979.65
103	5268.90	8139.53
104	5036.34	8212.21
105	5334.30	8459.30
106	3168.60	7688.95
109	2492.73	7863.37
110	2994.19	8132.27
116	1820.90	8179.12
117	2936.05	8328.49
118	3335.76	8466.57

119	3103.20	8815.41
122	4171.51	8655.52
123	3328.49	9643.90
124	3677.33	9600.29
125	4077.03	9527.62
127	3720.93	9273.26
129	4258.72	8909.88
130	4404.07	9440.41
131	3866.28	5239.83
133	4164.24	5029.07
136	1504.36	6875.00
137	4291.95	9366.72
139	4777.56	8044.76
140	4970.06	8470.13
7	2065.61	234.06
138	2798.67	8340.33

[VERTICES]

;Link	X-Coord	Y-Coord
16	746.81	2733.57
38	2936.97	4429.15
38	2851.56	4589.30
73	1105.95	7562.71
139	3809.92	9045.06
159	4290.09	9224.63
48	3983.07	9329.44
50	2634.05	8159.18
50	2509.78	8251.59
50	1987.17	8280.27
122	2923.29	7710.47
122	2848.73	7795.68
123	3594.31	7593.31
124	4204.81	7613.42
125	4201.42	7396.26
126	3791.35	6501.57
126	3929.82	6480.27
130	4435.74	7129.99
130	4531.60	6932.94
131	4819.18	6991.52
131	4893.74	7316.38
136	5426.29	7827.63
142	3370.63	4893.26
142	3450.52	4664.26
142	4009.70	4728.17
150	2076.53	6751.87
151	1213.79	6693.29
151	1038.04	6352.46
151	947.51	6347.13
152	1171.18	6890.34
154	1881.41	6964.20
154	1775.53	6927.69

154	1782.84	6730.53
155	1877.76	6442.09
155	1764.58	6376.37
157	3476.93	6212.08
36	1431.58	7418.58
36	1551.67	7565.38
36	1359.52	7578.73
37	3091.78	3458.40
40	3361.09	2642.06
42	2973.96	4069.96
43	2339.96	4277.55
51	2059.43	4597.36
52	692.21	2234.01
53	1002.89	2274.75
59	1395.06	2025.19
60	609.66	2498.98
60	295.78	2072.42
60	301.14	1906.09
62	2569.75	3561.38
62	2485.73	2351.45
5	641.00	1741.96

[LABELS]

;X-Coord	Y-Coord	Label & Anchor Node

[BACKDROP]

DIMENSIONS	20.35	-253.27	5808.14	10115.19
UNITS	Feet			
FILE				
OFFSET	0.00	0.00		

[END]

2. PRPsym input file for Base-Case Attack at node 50

PRPbaseCase.txt

PRPsym_V2.1_INPUT_DATA_FILE
Latest_Revision_Date:_February_22_2004

[PART_1:_THE_NETWORK]

Column_1--Sum_of_base_demand_for_all_node_(excluding_leaks,_gpm);
Column_2--Variance_total_demand_(gpm^2);
Column_3--Control_of_arrival_variation(0-no_variation;1-normal_dist;2-Poisson_dist);
Column_4--Random_number_generator_seed_for_indoor_demands_(0<integer<30,000);
Column_5--Random_number_generator_seed_for_outdoor_demands_(0<integer<30,000);
Column_6--Random_number_generator_seed_for_fixture_demands_(0<integer<30,000);
Column_7--Mass_balance_control_(0_within_pattern_segment_time,1_w/o_balance)

[1]	[2]	[3]	[4]	[5]	[6]	[7]
324.21	25237	1	10	10	10	1

[PART_2A:_JUNCTIONS]

Column_1--Node_ID;
Column_2--Number_of_homes_at_node;
Column_3--Base_demand_at_node_(gpm);
Column_4--Ratio_indoor_to_total_nodal_demand;
Column_5--Indoor_pulse_template_ID;
Column_6--Indoor_hourly_pattern_ID;
Column_7--Outdoor_pulse_template_ID;
Column_8--Outdoor_hourly_pattern_ID;
Column_9--Deterministic_use_pattern_ID

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
1	0	0.00	0.67	1	1	1	1	0
2	6	4.97	0.67	1	1	1	1	0
3	9	5.72	0.67	1	1	1	1	0
4	7	3.38	0.67	1	1	1	1	0
5	40	5.44	0.67	1	1	1	1	0
6	11	4.13	0.67	1	1	1	1	0
9	17	5.25	0.67	1	1	1	1	0
11	9	3.75	0.67	1	1	1	1	0
14	10	3.75	0.67	1	1	1	1	0
16	3	1.50	0.67	1	1	1	1	0
17	3	1.88	0.67	1	1	1	1	0
19	4	1.13	0.67	1	1	1	1	0
20	4	2.25	0.67	1	1	1	1	0
21	4	2.25	0.67	1	1	1	1	0
22	8	2.63	0.67	1	1	1	1	0
24	12	4.13	0.67	1	1	1	1	0

26	6	4.60	0.67	1	1	1	1	0
27	1	3.47	0.67	1	1	1	1	0
28	0	0.00	0.67	1	1	1	1	0
31	1	6.94	0.67	1	1	1	1	0
32	0	0.00	0.67	1	1	1	1	0
33	1	6.94	0.67	1	1	1	1	0
35	18	6.75	0.67	1	1	1	1	0
36	2	0.75	0.67	1	1	1	1	0
37	10	3.75	0.67	1	1	1	1	0
38	0	0.00	0.67	1	1	1	1	0
39	1	0.38	0.67	1	1	1	1	0
40	2	0.75	0.67	1	1	1	1	0
41	28	3.00	0.67	1	1	1	1	0
42	3	1.50	0.67	1	1	1	1	0
43	1	0.75	0.67	1	1	1	1	0
44	1	0.75	0.67	1	1	1	1	0
45	2	1.50	0.67	1	1	1	1	0
47	20	6.11	0.67	1	1	1	1	0
49	28	3.00	0.67	1	1	1	1	0
50	3	1.50	0.67	1	1	1	1	0
51	10	2.81	0.67	1	1	1	1	0
52	5	2.44	0.67	1	1	1	1	0
53	4	1.88	0.67	1	1	1	1	0
54	4	1.50	0.67	1	1	1	1	0
56	120	13.33	0.67	1	1	1	1	0
59	70	6.67	0.67	1	1	1	1	0
60	10	2.22	0.67	1	1	1	1	0
61	40	4.44	0.67	1	1	1	1	0
63	50	5.56	0.67	1	1	1	1	0
65	90	10.00	0.67	1	1	1	1	0
67	0	0.00	0.67	1	1	1	1	0
68	40	4.44	0.67	1	1	1	1	0
69	80	8.89	0.67	1	1	1	1	0
71	81	8.89	0.67	1	1	1	1	0
72	3	1.50	0.67	1	1	1	1	0
78	4	1.50	0.67	1	1	1	1	0
79	6	2.25	0.67	1	1	1	1	0
81	10	2.63	0.67	1	1	1	1	0
83	4	2.25	0.67	1	1	1	1	0
84	2	0.75	0.67	1	1	1	1	0
85	4	1.50	0.67	1	1	1	1	0
86	5	1.50	0.67	1	1	1	1	0
88	6	3.00	0.67	1	1	1	1	0
89	20	6.00	0.67	1	1	1	1	0
91	14	4.88	0.67	1	1	1	1	0
94	11	4.13	0.67	1	1	1	1	0
97	6	3.00	0.67	1	1	1	1	0
98	13	4.88	0.67	1	1	1	1	0
102	7	2.25	0.67	1	1	1	1	0
103	8	1.88	0.67	1	1	1	1	0
104	5	1.50	0.67	1	1	1	1	0

105	5	3.00	0.67	1	1	1	1	0
106	52	0.75	0.67	1	1	1	1	0
109	50	11.11	0.67	1	1	1	1	0
110	51	0.38	0.67	1	1	1	1	0
116	50	11.11	0.67	1	1	1	1	0
117	2	0.75	0.67	1	1	1	1	0
118	9	3.75	0.67	1	1	1	1	0
119	15	4.50	0.67	1	1	1	1	0
122	8	3.38	0.67	1	1	1	1	0
123	4	2.25	0.67	1	1	1	1	0
124	9	3.00	0.67	1	1	1	1	0
125	6	2.25	0.67	1	1	1	1	0
127	15	4.50	0.67	1	1	1	1	0
129	9	3.38	0.67	1	1	1	1	0
130	2	1.50	0.67	1	1	1	1	0
131	60	12.22	0.67	1	1	1	1	0
133	140	21.39	0.67	1	1	1	1	0
136	60	6.67	0.67	1	1	1	1	0
137	3	1.50	0.67	1	1	1	1	0
139	2	0.75	0.67	1	1	1	1	0
140	6	3.00	0.67	1	1	1	1	0
-1	0	0	0	0	0	0	0	0

[PART_2B:_FIXTURES]
Column_1--Fixture_ID;
Column_2--Node_ID_from_part_2A;
Column_3--Base_demand_at_fixture_(gpm);
Column_4--Ratio_indoor_to_total_nodal_demand;
Column_5--Indoor_pulse_template_ID;
Column_6--Indoor_hourly_pattern_ID;
Column_7--Outdoor_pulse_template_ID;
Column_8--Outdoor_hourly_pattern_ID;
Column_9--Deterministic_use_pattern_ID

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
-1	0	0	0	0	0	0	0	0

[PART_3:_EXECUTION_CONTROL]

Column_1--Simulation_start_time_(hr);
Column_2--Simulation_duration_(hr)
Column_3--function_switch_for_volume_compare_tool_(1-on;0-off)
Column_4--time_averaging_interval_used_for_volume_compare_(second)
Column_5--Period_for_compare(hour)
Column_6--shifting_interval(minute)
Column_7--Options_for_output_format(0-normal_output;_1-EPANET_input_file_only)
Column_8--Name_of_EPANET_input_file_to_be_read

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
0	72	0	300	6	15	1	CSTR.inp

Provide_specific_node_IDs_to_output
Note:_If_the_first_line_is_set_0,output_all_node
0
-1

Provide_time_averaged_intervals(second)
3600
-1

[PART_4:_PRP_PARAMETERS]

INDOOR_Pulse_Characteristics
Column_1--Template_ID;
Column_2--Average_demand_intensity_(gpm);
Column_3--Variance_demand_intensity_(gpm^2);
Column_4--Probability_distribution_for_intensity_ID;
Column_5--Average_demand_duration_(min);
Column_6--Variance_demand_duration_(min^2);
Column_7--Probability_distribution_for_duration_ID

[1]	[2]	[3]	[4]	[5]	[6]	[7]
1	2.25	1.55	0	1.00	4.00	0
-1	0	0	0	0	0	0

INDOOR_Arrival_Pattern
Column_1--Pattern_ID;
Column_2--Pattern_segment_number;
Column_3--Segment_start_time_(hr);
Column_4--Segment_duration_(hr);
Column_5--Segment_arrival_multiplier

[1]	[2]	[3]	[4]	[5]
1	1	0	1	1.26
1	2	1	1	1.04
1	3	2	1	0.97
1	4	3	1	0.97
1	5	4	1	0.89
1	6	5	1	1.19
1	7	6	1	1.28
1	8	7	1	0.67
1	9	8	1	0.67
1	10	9	1	1.34
1	11	10	1	2.46
1	12	11	1	0.97
1	13	12	1	0.92
1	14	13	1	0.68
1	15	14	1	1.43
1	16	15	1	0.61
1	17	16	1	0.31

1	18	17	1	0.78
1	19	18	1	0.37
1	20	19	1	0.67
1	21	20	1	1.26
1	22	21	1	1.56
1	23	22	1	1.19
1	24	23	1	1.26
1	25	24	1	0.6
1	26	25	1	1.1
1	27	26	1	1.03
1	28	27	1	0.73
1	29	28	1	0.88
1	30	29	1	1.06
1	31	30	1	0.99
1	32	31	1	1.72
1	33	32	1	1.12
1	34	33	1	1.34
1	35	34	1	1.12
1	36	35	1	0.97
1	37	36	1	1.04
1	38	37	1	1.15
1	39	38	1	0.91
1	40	39	1	0.61
1	41	40	1	0.68
1	42	41	1	0.46
1	43	42	1	0.51
1	44	43	1	0.74
1	45	44	1	1.12
1	46	45	1	1.34
1	47	46	1	1.26
1	48	47	1	0.97
1	49	48	1	0.82
1	50	49	1	1.37
1	51	50	1	1.03
1	52	51	1	0.81
1	53	52	1	0.88
1	54	53	1	0.81
1	55	54	1	0.81
1	56	55	1	1.26
1	57	56	1	1.04
1	58	57	1	0.97
1	59	58	1	0.97
1	60	59	1	0.89
1	61	60	1	1.19
1	62	61	1	1.28
1	63	62	1	0.67
1	64	63	1	0.67
1	65	64	1	1.34
1	66	65	1	2.46
1	67	66	1	0.97
1	68	67	1	0.92

1	69	68	1	0.68
1	70	69	1	1.43
1	71	70	1	0.61
1	72	71	1	0.31
-1	0	0	0	0

OUTDOOR_Pulse_Characteristics

[1]	[2]	[3]	[4]	[5]	[6]	[7]
1	4.00	1.00	0	15	2700	0
-1	0	0	0	0	0	0

OUTDOOR_Hourly_Arrival_Pattern

[1]	[2]	[3]	[4]	[5]
1	1	0	1	1.26
1	2	1	1	1.04
1	3	2	1	0.97
1	4	3	1	0.97
1	5	4	1	0.89
1	6	5	1	1.19
1	7	6	1	1.28
1	8	7	1	0.67
1	9	8	1	0.67
1	10	9	1	1.34
1	11	10	1	2.46
1	12	11	1	0.97
1	13	12	1	0.92
1	14	13	1	0.68
1	15	14	1	1.43
1	16	15	1	0.61
1	17	16	1	0.31
1	18	17	1	0.78
1	19	18	1	0.37
1	20	19	1	0.67
1	21	20	1	1.26
1	22	21	1	1.56
1	23	22	1	1.19
1	24	23	1	1.26
1	25	24	1	0.6
1	26	25	1	1.1
1	27	26	1	1.03
1	28	27	1	0.73
1	29	28	1	0.88
1	30	29	1	1.06
1	31	30	1	0.99
1	32	31	1	1.72
1	33	32	1	1.12
1	34	33	1	1.34
1	35	34	1	1.12

1	36	35	1	0.97
1	37	36	1	1.04
1	38	37	1	1.15
1	39	38	1	0.91
1	40	39	1	0.61
1	41	40	1	0.68
1	42	41	1	0.46
1	43	42	1	0.51
1	44	43	1	0.74
1	45	44	1	1.12
1	46	45	1	1.34
1	47	46	1	1.26
1	48	47	1	0.97
1	49	48	1	0.82
1	50	49	1	1.37
1	51	50	1	1.03
1	52	51	1	0.81
1	53	52	1	0.88
1	54	53	1	0.81
1	55	54	1	0.81
1	56	55	1	1.26
1	57	56	1	1.04
1	58	57	1	0.97
1	59	58	1	0.97
1	60	59	1	0.89
1	61	60	1	1.19
1	62	61	1	1.28
1	63	62	1	0.67
1	64	63	1	0.67
1	65	64	1	1.34
1	66	65	1	2.46
1	67	66	1	0.97
1	68	67	1	0.92
1	69	68	1	0.68
1	70	69	1	1.43
1	71	70	1	0.61
1	72	71	1	0.31
-1	0	0	0	0

Note:Probability_distribution_ID:
(0=lognormal_distribution;_1=normal_distribution;_2=exponential_distribution)

[PART_5:DETERMINISTIC_HOURLY_FLOW_PATTERN]
Column_1--Pattern_ID;
Column_2--Pattern_segment_number;
Column_3--Segment_start_time_(hr);
Column_4--Segment_duration_(hr);
Column_5--Segment_demand_value_(gpm)

[1]	[2]	[3]	[4]	[5]
-1	0	0	0	0

[PART_6:_END]

3. EPANET input file generated by PRPsym for Base-Case Attack at node 50

New EPANET input file generated by PRPsym by executing PRPbaseCase.txt

[TITLE]
"Base-Case Attack on node 50"

[JUNCTIONS]

;ID	Elev	Demand	Pattern	
1	50	0.000		;
2	90	4.970	2	;
3	100	5.720	3	;
4	75	3.380	4	;
5	100	5.440	5	;
6	75	4.130	6	;
9	60	5.250	9	;
11	75	3.750	11	;
14	75	3.750	14	;
16	100	1.500	16	;
17	100	1.880	17	;
19	115	1.130	19	;
20	115	2.250	20	;
21	125	2.250	21	;
22	125	2.630	22	;
24	150	4.130	24	;
26	160	4.600	26	;
27	125	3.470	27	;
28	110	0.000		;
31	130	6.940	31	;
32	180	0.000		;
33	150	6.940	33	;
35	180	6.750	35	;
36	140	0.750	36	;
37	180	3.750	37	;
38	185	0.000		;
39	140	0.380	39	;
40	140	0.750	40	;
41	190	3.000	41	;
42	140	1.500	42	;
43	120	0.750	43	;
44	100	0.750	44	;
45	190	1.500	45	;
47	200	6.110	47	;
49	210	3.000	49	;
50	210	1.500	50	;
51	195	2.810	51	;
52	180	2.440	52	;
53	165	1.880	53	;
54	150	1.500	54	;
56	150	13.330	56	;

59	150	6.670	59	;
60	150	2.220	60	;
61	150	4.440	61	;
63	110	5.560	63	;
65	135	10.000	65	;
67	100	0.000		;
68	100	4.440	68	;
69	100	8.890	69	;
71	160	8.890	71	;
72	160	1.500	72	;
78	180	1.500	78	;
79	183	2.250	79	;
81	186	2.630	81	;
83	188	2.250	83	;
84	190	0.750	84	;
85	200	1.500	85	;
86	190	1.500	86	;
88	190	3.000	88	;
89	170	6.000	89	;
91	162	4.880	91	;
94	150	4.130	94	;
97	190	3.000	97	;
98	200	4.880	98	;
102	185	2.250	102	;
103	180	1.880	103	;
104	182	1.500	104	;
105	190	3.000	105	;
106	190	0.750	106	;
109	190	11.110	109	;
110	230	0.380	110	;
116	230	11.110	116	;
117	230	0.750	117	;
118	230	3.750	118	;
119	190	4.500	119	;
122	145	3.380	122	;
123	110	2.250	123	;
124	110	3.000	124	;
125	110	2.250	125	;
127	110	4.500	127	;
129	130	3.380	129	;
130	110	1.500	130	;
131	190	12.220	131	;
133	185	21.390	133	;
136	118	6.670	136	;
137	120	1.500	137	;
139	190	0.750	139	;
140	190	3.000	140	;

[RESERVOIRS]

;ID	Head	Pattern	
7	50		;

[TANKS]

;ID	Elevation	InitLevel	MinLevel	MaxLevel	Diameter	MinVol	VolCurve
138	235	56.7	56	66	50	0	;

[PIPES]

;ID	Node1	Node2	Length	Diameter	Roughness	MinorLoss	Status	
2	2	3	550	12	100	0	Open	;
3	3	4	526	8	100	0	Open	;
11	11	4	288	8	100	0	Open	;
16	9	14	576	8	100	0	Open	;
17	14	16	530	12	100	0	Open	;
18	16	17	583	8	100	0	Open	;
21	17	19	815	12	100	0	Open	;
22	19	22	431	12	100	0	Open	;
23	19	20	88	8	100	0	Open	;
25	21	22	288	8	100	0	Open	;
30	26	27	767	12	100	0	Open	;
31	27	28	312	12	100	0	Open	;
35	26	32	383	12	100	0	Open	;
38	32	35	1438	8	100	0	Open	;
41	32	38	743	12	100	0	Open	;
44	38	41	479	12	100	0	Open	;
54	49	50	600	12	100	0	Open	;
55	50	51	420	8	100	0	Open	;
56	51	52	420	8	100	0	Open	;
57	52	53	420	8	100	0	Open	;
58	53	54	420	8	100	0	Open	;
64	59	60	360	8	100	0	Open	;
65	59	61	364	8	100	0	Open	;
67	61	136	838	8	100	0	Open	;
71	63	67	240	12	100	0	Open	;
72	5	68	300	12	100	0	Open	;
73	68	69	623	12	100	0	Open	;
75	67	71	647	12	100	0	Open	;
76	71	72	479	12	100	0	Open	;
83	71	78	216	12	100	0	Open	;
84	78	79	375	8	100	0	Open	;
87	79	81	375	8	100	0	Open	;
90	81	83	375	8	100	0	Open	;
91	83	84	375	8	100	0	Open	;
92	84	85	300	12	100	0	Open	;
93	50	85	400	12	100	0	Open	;
94	85	86	240	12	100	0	Open	;
97	86	89	623	12	100	0	Open	;
101	91	97	527	8	100	0	Open	;
107	97	98	527	8	100	0	Open	;
113	102	103	407	8	100	0	Open	;
114	103	104	312	8	100	0	Open	;
115	103	105	455	8	100	0	Open	;
116	84	106	264	12	100	0	Open	;

120	106	110	647	12	100	0	Open	;
127	110	117	240	12	100	0	Open	;
128	117	118	551	8	100	0	Open	;
129	117	119	671	12	100	0	Open	;
133	119	123	1150	8	100	0	Open	;
134	123	124	455	8	100	0	Open	;
135	124	125	551	8	100	0	Open	;
138	127	124	407	8	100	0	Open	;
139	127	129	1007	8	100	0	Open	;
141	125	130	431	8	100	0	Open	;
143	129	122	335	8	100	0	Open	;
145	51	83	700	6	100	0	Open	;
146	54	78	700	8	100	0	Open	;
156	136	63	216	8	100	0	Open	;
1	1	2	1850	12	100	0	Open	;
158	138	117	200	12	100	0	Open	;
159	137	129	743	8	100	0	Open	;
160	104	139	300	8	100	0	Open	;
161	140	104	400	12	100	0	Open	;
48	127	125	790	8	100	0	Open	;
49	119	122	1413	8	100	0	Open	;
50	110	116	1703	12	100	0	Open	;
122	106	109	887	12	100	0	Open	;
123	86	88	766	8	100	0	Open	;
124	89	98	1200	8	100	0	Open	;
125	89	91	528	8	100	0	Open	;
126	49	47	1079	8	100	0	Open	;
130	91	94	887	8	100	0	Open	;
131	94	97	768	8	100	0	Open	;
132	102	45	250	8	100	0	Open	;
136	98	102	647	8	100	0	Open	;
137	41	131	700	8	100	0	Open	;
142	38	133	1821	8	100	0	Open	;
144	32	37	1102	12	100	0	Open	;
148	81	52	700	6	100	0	Open	;
149	79	53	700	6	100	0	Open	;
150	54	56	384	12	100	0	Open	;
151	136	65	1533	12	100	0	Open	;
152	65	63	839	8	100	0	Open	;
154	71	56	1104	8	100	0	Open	;
155	56	59	958	8	100	0	Open	;
157	49	41	1796	12	100	0	Open	;
36	5	69	1006	12	100	0	Open	;
37	27	33	500	8	100	0	Open	;
40	28	31	1127	8	100	0	Open	;
42	26	24	576	12	100	0	Open	;
43	24	42	1100	8	100	0	Open	;
45	42	43	300	8	100	0	Open	;
46	42	39	300	8	100	0	Open	;
47	39	40	400	8	100	0	Open	;
51	39	36	400	8	100	0	Open	;

52	9	11	910	6	100	0	Open	;
53	14	11	1072	8	100	0	Open	;
59	17	3	838	12	100	0	Open	;
60	9	6	1000	8	100	0	Open	;
61	2	44	400	8	100	0	Open	;
62	24	22	2500	12	100	0	Open	;
4	5	67	83	12	100	0	Open	;
5	6	4	870	8	100	0	Open	;

[PUMPS]

;ID	Node1	Node2	Parameters	
0	7	1	HEAD 1	;

[VALVES]

;ID	Node1	Node2	Diameter	Type	Setting	MinorLoss

[TAGS]

[DEMANDS]

;Junction	Demand	Pattern	Category

[STATUS]

;ID	Status/Setting

[PATTERNS]

;ID	Multipliers					
2	1.31067	0.66201	0.91984	1.75648	1.39423	1.37276
2	0.66065	0.75646	0.79091	1.45838	2.63337	0.49725
2	1.11756	1.04016	0.75364	0.69897	0.67629	0.48665
2	1.46679	2.13484	1.62666	0.65072	2.09858	2.44404
2	1.21769	1.67070	1.74524	2.02875	1.59638	1.88771
2	0.69848	2.19549	2.00232	3.46780	2.34480	2.62758
2	2.08976	3.27873	2.35779	1.22502	0.74921	1.09424
2	0.11294	0.74381	0.89579	1.21548	1.07595	1.15724
2	0.56305	0.43805	0.86185	0.39435	0.37839	0.46207
2	0.57103	1.08351	0.51375	0.86902	0.58637	0.95051
2	1.82102	1.28059	1.16899	1.09152	0.89682	2.86240
2	1.94533	2.25924	2.02640	2.15978	0.93509	0.10792
3	1.09539	0.70203	0.15848	0.66060	1.01702	1.48458
3	1.49271	0.69557	0.60926	0.72026	1.55625	1.34540
3	0.97718	0.32923	1.12987	0.86162	0.08525	0.58670
3	0.53658	0.47445	1.26360	1.13212	0.51260	1.42812
3	0.10987	0.97041	1.23540	1.04635	1.12713	0.79455
3	1.86151	0.98751	0.49178	1.18316	1.45204	1.19037
3	1.27441	1.24924	0.95155	1.10644	1.18270	0.68475
3	0.70784	0.98492	1.52291	1.56740	0.70968	0.80310
3	1.24802	1.87533	0.50917	0.77293	0.71841	0.97548
3	0.48772	0.67518	0.61129	1.09016	1.61937	1.99401

3	1.67038	0.45144	0.69528	0.22575	2.42158	4.22897
3	1.28432	0.98733	0.71009	1.03888	0.49481	0.18125

This is hourly demand multipliers generated by **PRPsym** at only three nodes (node#2, node#3 and node#140) in the CH/BP all-pipes network. These multipliers are inserted into the **EPANET** input file by the **PRPsym** code. Owing to the random nature of the water demands, each node has a unique 72-hr demand multiplier sequence.

140	0.62553	1.18913	1.31953	0.28647	0.77906	0.87205
140	2.49215	0.55027	0.33993	0.22990	1.40812	0.41396
140	1.06297	0.98488	0.86235	0.49339	0.21094	0.33449
140	0.38786	1.17755	1.71720	1.33485	1.40125	0.95890
140	0.05745	1.69173	1.55627	0.45925	1.22771	0.55119
140	0.86860	2.85842	0.99185	0.64145	1.44547	0.68795
140	1.41176	1.37456	1.00662	0.49328	0.33859	0.08334
140	0.40082	0.27629	1.41747	1.53175	1.22866	0.07902
140	0.45499	1.55161	1.24811	0.89896	0.59199	1.02496
140	1.25724	1.49492	0.24784	1.14258	0.95698	0.37892
140	0.72829	0.77960	0.81805	0.53210	0.95472	1.21129
140	0.67490	0.38362	0.48196	1.00965	0.07279	0.11268

;Demand Pattern

1	1.26	1.04	.97	.97	.89	1.19
1	1.28	.67	.67	1.34	2.46	.97
1	.92	.68	1.43	.61	.31	.78
1	.37	.67	1.26	1.56	1.19	1.26
1	.6	1.1	1.03	.73	.88	1.06
1	.99	1.72	1.12	1.34	1.12	.97
1	1.04	1.15	.91	.61	.68	.46
1	.51	.74	1.12	1.34	1.26	.97
1	.82	1.37	1.03	.81	.88	.81
1	.81	1.26	1.04	0.97	0.97	0.89
1	1.19	1.28	0.67	0.67	1.34	2.46
1	0.97	0.92	0.68	1.43	0.61	0.31

;Injection for 6 hrs- peak

Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	1	1	1
Inj6p	1	1	1	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0
Inj6p	0	0	0	0	0	0

```
 Inj6p           0       0       0       0       0       0
 Inj6p           0       0       0       0       0       0

[CURVES]

[CURVES]
;ID             X-Value         Y-Value
;PUMP:            Pump Curve
 1              666.62          259.88

[CONTROLS]
Link 0 Closed if Node 138 Above 65
Link 0 Open if Node 138 Below 56

[RULES]

[ENERGY]
 Global Efficiency      75
 Global Price           0
 Demand Charge          0

[EMITTERS]
;Junction       Coefficient

[QUALITY]
;Node           InitQual

[SOURCES]
;Node           Type            Quality         Pattern
 50             MASS            20000           Inj6p

[REACTIONS]
;Type           Pipe/Tank       Coefficient

[REACTIONS]
 Order Bulk             1
 Order Tank             1
 Order Wall             1
 Global Bulk            0
 Global Wall            0
 Limiting Potential     0
 Roughness Correlation  0

[MIXING]
;Tank           Model

[TIMES]
 Duration               72:00
 Hydraulic Timestep     1:00
 Quality Timestep       0:01
 Pattern Timestep       1:00
```

Pattern Start		0:00
Report Timestep		0:01
Report Start		0:00
Start ClockTime		9 am
Statistic	NONE	

[REPORT]

Status	No	
Summary		No
Page	0	

[OPTIONS]

Units	GPM	
Headloss		H-W
Specific Gravity		1
Viscosity	1	
Trials	40	
Accuracy		0.001
Unbalanced		Continue 10
Pattern	1	
Demand Multiplier		1.0
Emitter Exponent		0.5
Quality	Chemical mg/L	
Diffusivity		1
Tolerance		0.01

[COORDINATES]

;Node	X-Coord	Y-Coord
1	1373.55	225.29
2	1359.01	1395.35
3	1359.01	1809.59
4	937.50	1758.72
5	1410.23	7226.40
6	365.89	1890.10
9	675.87	2529.07
11	937.50	1984.01
14	944.77	2754.36
16	1359.01	2739.83
17	1359.01	2449.13
19	1969.48	2311.05
20	1955.30	2177.98
21	2441.86	2005.81
22	2289.24	2216.57
24	2739.83	4018.90
26	3190.41	4055.23
27	3372.09	3502.91
28	3473.84	3125.00
31	3364.83	2267.44
32	3204.94	4353.20
33	3108.61	3304.11
35	2318.31	4716.57

36	2140.78	4656.27
37	4040.70	4527.62
38	3219.48	4912.79
39	1921.97	4403.79
40	1649.85	4594.55
41	3263.08	5276.16
42	2028.57	4235.47
43	1966.85	4033.49
44	918.90	1379.91
45	5713.87	8099.24
47	4164.24	6257.27
49	3452.03	6497.09
50	3350.29	6947.67
51	3037.79	6954.94
52	2710.76	6925.87
53	2398.26	6911.34
54	2093.02	6889.53
56	2009.45	6711.48
59	1744.19	6250.00
60	1773.26	5973.84
61	1533.43	6242.73
63	1489.83	7042.15
65	984.79	6517.55
67	1482.56	7209.30
68	1162.79	7223.84
69	1177.33	7681.69
71	1947.67	7369.19
72	1925.87	7739.83
78	2100.29	7412.79
79	2354.65	7420.06
81	2651.28	7470.48
83	2943.31	7470.93
84	3252.71	7506.56
85	3292.15	7231.10
86	3626.45	7347.38
88	3837.21	7739.83
89	4077.03	7521.80
91	4447.67	7361.92
94	4738.37	6809.59
97	4854.65	7405.52
98	5130.81	7681.69
102	5545.06	7979.65
103	5268.90	8139.53
104	5036.34	8212.21
105	5334.30	8459.30
106	3168.60	7688.95
109	2492.73	7863.37
110	2994.19	8132.27
116	1820.90	8179.12
117	2936.05	8328.49
118	3335.76	8466.57

119	3103.20	8815.41
122	4171.51	8655.52
123	3328.49	9643.90
124	3677.33	9600.29
125	4077.03	9527.62
127	3720.93	9273.26
129	4258.72	8909.88
130	4404.07	9440.41
131	3866.28	5239.83
133	4164.24	5029.07
136	1504.36	6875.00
137	4291.95	9366.72
139	4777.56	8044.76
140	4970.06	8470.13
7	2065.61	234.06
138	2798.67	8340.33

[VERTICES]

;Link	X-Coord	Y-Coord
16	746.81	2733.57
38	2936.97	4429.15
38	2851.56	4589.30
73	1105.95	7562.71
139	3809.92	9045.06
159	4290.09	9224.63
48	3983.07	9329.44
50	2634.05	8159.18
50	2509.78	8251.59
50	1987.17	8280.27
122	2923.29	7710.47
122	2848.73	7795.68
123	3594.31	7593.31
124	4204.81	7613.42
125	4201.42	7396.26
126	3791.35	6501.57
126	3929.82	6480.27
130	4435.74	7129.99
130	4531.60	6932.94
131	4819.18	6991.52
131	4893.74	7316.38
136	5426.29	7827.63
142	3370.63	4893.26
142	3450.52	4664.26
142	4009.70	4728.17
150	2076.53	6751.87
151	1213.79	6693.29
151	1038.04	6352.46
151	947.51	6347.13
152	1171.18	6890.34
154	1881.41	6964.20
154	1775.53	6927.69

154	1782.84	6730.53
155	1877.76	6442.09
155	1764.58	6376.37
157	3476.93	6212.08
36	1431.58	7418.58
36	1551.67	7565.38
36	1359.52	7578.73
37	3091.78	3458.40
40	3361.09	2642.06
42	2973.96	4069.96
43	2339.96	4277.55
51	2059.43	4597.36
52	692.21	2234.01
53	1002.89	2274.75
59	1395.06	2025.19
60	609.66	2498.98
60	295.78	2072.42
60	301.14	1906.09
62	2569.75	3561.38
62	2485.73	2351.45
5	641.00	1741.96

[LABELS]

;X-Coord	Y-Coord	Label & Anchor Node

[BACKDROP]

DIMENSIONS	20.35	-253.27	5808.14	10115.19
UNITS	Feet			
FILE				
OFFSET	0.00	0.00		

[END]

APPENDIX II

1. Program I: Percentage nodal mass consumption

2. Program II: PRP simulation of water demands

3. Program III: Latin Hypercube Sampling within EPANET execution interface

4. Program IV: Kolmogorov-Smirnov two-sample test

1. Program I: Percentage nodal mass consumption

```
/*************************************************************************
Program:      SysMassNK.cpp
              Calculates chemical mass percentage present at each node for each
              reporting timestep and calculates the total chemical mass
              consumed over the EPS and tank mass at end of each hydraulic
              time step

Author:       Nabin Khanal
              Civil and Environmental Engineering, University of Cincinnati

Date:         October 11, 2004

Last Updated: November 15, 2004
Libraries linked: epanet2vc.lib for EPANET
*************************************************************************/
// Includes
#include <iostream>
#include "epanet2.h"
#include <fstream>
using namespace std;

// Constants
const double GalToLit = 3.785;  //gallon to liter conversion 1 gallon =3.785 liter
const double CfToLit = 28.32;   //cubic feet to liter conversion 1 cftt = 28.316 liter
const float TankDia = 50;       //tank diameter (ft)
const float TankMin = 235.0;    //minimum tank water level (ft)

// Function Declarations
void NodalMass();       //calculatess total mass consumed at demand nodes
void TankMass();        //calculates total mass in tank

// Global Variables
```

```
int   i,NumNodes, type, errcode,j;
long  t, tstep, r,ht,du,qs; // r=reporting time step, ht=hydraulic time step
float q, d, h, ChemMas, TotalChem, TankMas, TotalQ;
float TotalChemArr[90] = {0};
float DemandArr[90] = {0};
float nk[90]={0};
char  id[15];
char  SystemMassFile[25] = {"SysMassNK.txt"};// Output filename
char  inpEpanetFile[25]; //Variable to store Epanet input filename
char  rptEpanetFile[25]={"1.rpt"}; //Epanet report filename
char nkoutfile[25]= {"nkout.txt"};// Output file name

// Main program body
void main()
{       // Start main
        //Object to manage output stream
        ofstream fout(SystemMassFile);// Writes to file SysMassNK.txt
        fout << "Total System Contaminant Mass Consumed (g)"<<endl;
        fout << "Run Date: " <<__DATE__<<endl;
        ofstream nkout(nkoutfile);// Writes to file nkout.txt
        nkout<<"Time\t";
        for (int cc=1;cc<=90;cc++)
        {// start cc
                nkout<<cc<<"\t";// Writing the nodex index numbers
        }// end cc
        nkout<<endl<<endl;
//      Asks for EPANET input file for analysis
        cout<<"Enter the EPANET input filename(<file.inp>):"<<inpEpanetFile;
        cin>>inpEpanetFile;
        cin.get();

        fout<<"EPANET input file:"<<inpEpanetFile<<endl;
        fout<<"EPANET report file:"<<rptEpanetFile<<endl<<endl<<endl;
        fout << "Hour\t\tNodal(g)\tTank (g)\tDemands(gallons)"<<endl;
```

```
// Opens EPANET toolkit and runs hydraulic analysis
errcode=ENopen(inpEpanetFile,rptEpanetFile,"");
if (errcode>0)
{ENclose();
return;}
ENsolveH();

// Displays EPANET node & time setting information
ENgetcount(EN_NODECOUNT, &NumNodes);
cout << "Number of Nodes: " << NumNodes <<endl<<endl;
ENgettimeparam(0, &du);
cout << "Simulation duration: " << du/3600<<"hrs"<<endl<<endl;
ENgettimeparam(1, &ht);
cout << "Hydraulic Time Step: " << ht<<"sec" <<endl<<endl;
ENgettimeparam(2, &qs);
cout << "Quality Time Step: " << qs<<"sec" <<endl<<endl;
ENgettimeparam(5, &r);
cout << "Reporting Time Step: " << r <<"sec"<<endl<<endl;

// Starts EPANET water quality simulation
ENopenQ();
ENinitQ(0);

do
{// Start do-while loop
        ENrunQ(&t);
        ChemMas = 0.0;
        TotalChem = 0.0;
        TankMas = 0.0;
        TotalQ = 0.0;
        for (i = 1; i<= NumNodes; i++)
        {// start i
                ENgetnodeid(i, id);
                ENgetnodetype(i, &type);
```

```
                        ENgetnodevalue(i, EN_DEMAND, &d);
                        ENgetnodevalue(i, EN_QUALITY, &q);
                        ENgetnodevalue(i, EN_HEAD, &h);
                        if (type == 0)                  //nodes
                                NodalMass();// Calling function NodalMass
                        else if (type == 2)             //tanks
                                TankMass();// Calling function TankMass
                        else if (type == 1)             //reservoirs
                                ChemMas = 0.0;
                }// end i
                fout << t/3600;  //time in hour
                fout << "\t\t" << TotalChem;//total Nodal mass (g)
                fout << "\t\t" << TankMas;//total Tank mass (g)
                fout << "\t\t" << TotalQ;//total demand (gallons)
                fout << endl;
//-------------------------Writes the time series cumulative nodal mass --------
/*      nkout<<t/3600;
                for (int kk=1;kk<=90;kk++)
                {// start kk
                        nkout<<"\t"<<nk[kk];//Nodal Mass
                }//end kk
                        nkout<<endl;*/
//---------------------------------------------------------------------------
                ENnextQ(&tstep);

        }       while (tstep > 0);// Repeats do-while loop
//---------------Writes cumulative nodal mass percentage at end of simulation only----
nkout<<t/3600;
for (int pp=1;pp<=90;pp++)
                {//start pp
                        nkout<<"\t"<<(100*nk[pp])/7200;//Percentage Nodal Mass
                                                        // mass injected is 7200 gram
                }//end pp
                        nkout<<endl;
```

```
//-------------------------------------------------------------------------------
        cout << "Total mass consumed at Nodes: " << TotalChem <<"g"<<endl;
        cout << "Total mass at Tank: " << TankMas <<"g"<<endl;
        cout << "Demands: " << TotalQ <<"gallons"<<endl;
        fout <<endl<<endl;
        fout.close();
        nkout.close();
        ENcloseQ();// End water quality simulation
        ENclose();// Close the EPANET input file

}       // End main

//Functions

// Function to calculate mass consumed at nodes
void NodalMass()
{
        ChemMas = d*(ht/60)*GalToLit*(q/1000);      // Calculates nodal mass in each time
                                                    //step (gram)
        TotalChemArr[i] = TotalChemArr[i] + ChemMas;//Adds chemical mass
                                                    //at each time step
        TotalChem = TotalChem + TotalChemArr[i];// total mass consumed
        j=i;
        nk[j]=TotalChemArr[i];
        DemandArr[i] = DemandArr[i] + d;
        TotalQ = TotalQ + DemandArr[i]*(r/60); // total demand (gallons)
}// End function NodalMass
// Function to calculate mass at tank
void TankMass()
{
        TankMas = (h-TankMin)*(3.14*TankDia*TankDia/4)*CfToLit*(q/1000); //Tank chem
mass calc (g)
}// End function TankMass
// End of SysMassNK.cpp
```

2. Program II: PRP simulation of water demands

```
/*********************************************************************
Program:        PrpEPanetNK.cpp

                Reads PRPsym input file and uses PRPsym dll a library to
                demand generator program PRPsym to generate new
                EPANET .inp file with random demands associated to each
                nodes. EPANET toolkit is executed to run hydraulic and
                water quality analysis for newly formed EPANET .inp file,
                and calculates the demand at nodes and nodal mass
                consumption for the entire simulation period.

                Monte Carlo simulation of water demand is performed by
                executing this program for number of times, where for each
                runs, PRPsym generates random demands to network
                nodes.

Author:         Nabin Khanal
                Civil and Environmental Engineering, University of Cincinnati

Date:           December 2, 2004

Last updated:   February 10, 2005

Reference:      Nilsson et al. (2005)

Libraries linked:   PRP_Generator.lib for PRPsym
                    epanet2vc.lib for EPANET
*********************************************************************/
// Includes and headers
#include <iostream>
#include <cstdlib>
#include <fstream>
```

```
#include <string>
#include "epanet2.h" //Header file for EPANET
#include "PRP_Generator.h"//Header file for PRPsym

using namespace std;
using std::string;

// Function Declaration
void RandNum();
void NodalMass();

//Constants
const int  intNodes=90;  // total number of nodes in network
const int Max = 90;        // maximum number of nodes used in calculation
const double GalToLit = 3.785; //gallon to liter conversion 1 gallon = 3.785 liter
const double CfToLit = 28.32; //cubic feet to liter conversion 1 cft = 28.316 liter
const float TankDia = 50; //tank diameter (ft)
const float TankMin = 235.0;      //minimum tank water level (ft)

//Global Variables
int  z;                           //number of simulations
char ch[16];                      //node name array
char MyNodes[Max][16];            //array containing specified nodes
char inpEpanetFile[20];           //EPANET input file
char inpPRPsymFile[20];           //PRPsym input file
char outPRPsymFile[20];           //PRPsym output file

// Main program body
void main()
{//Start main
        cout << "Enter EPANET input file name: ";
        cin >> inpEpanetFile;
        cout << "Enter PRPsym input file name <*.txt>: ";
        cin >> inpPRPsymFile;
```

```
        cout << "Enter PRPsym output file name: ";
        cin >> outPRPsymFile;
        cin.get();
        RandNum();//Calling function RandNum
}//End main

// Functions
// Function that generates random number, so that it can be used to randomize the seed
// no. in PRPsym

void RandNum()
{//Start function RandNum
        // Local variables for function RandNum
        int SeedValue, NumRuns, seed;
        char  RandNumFile[25] = {"RandNum.txt"};
        char s1[20], s2[20];
// Class variable for PRPsym
        PRP_Generator prp;
        // Asks for Seed number and total number of simulations desired
        cout << "Enter Seed Number (0<S<30,000): ";//Seed for rand # generator
        cin >> SeedValue;
        cout << "Enter number of simulation runs: ";//Number of Simulations
        cin >> NumRuns;
        cout << endl<<endl;

        // Writing generated random numbers to RandNum.txt
        ofstream fout(RandNumFile);     //Uses fout to write to RandNum.txt
        fout << "Random Number File"<<endl;
        fout << "Run Date: " << __DATE__ << endl<<endl;
        fout << "EPANET input File: " << inpEpanetFile << endl<<endl;
        fout << "PRPsym input File: " << inpPRPsymFile << endl<<endl;
        fout << "Random number Generator Seed: " << SeedValue << endl;
        fout << "Number of Simulations: " << NumRuns;
        fout <<  endl<<endl;
```

```
        fout << "Run#\t" << "PRPseed#"<<endl<<endl;

        // Generates a set of pseudorandom variables for a seed number
        srand(SeedValue);
        for (z = 1; z <= NumRuns; z++)
                {// start z
                        //Generates random numbers (1 to 29,999) based on seed
                        seed = (1 + rand() % 29999);
                        cout << "Run: " << z << " of " << NumRuns;
                        cout <<"\tPRPseed: " << seed << endl<<endl;
                        fout << z << "\t"; //run number
                        fout << seed << "\n";     //seed number
                        strcpy(s1, outPRPsymFile);       //PRPsym output filet name
                        strcpy(s2, inpPRPsymFile);       //PRPsym input file name
                        prp.controller(s1,s2,seed);          //PRPsym DLL
                        NodalMass();//Calling function NodalMass
                }// end z

        fout <<endl<<endl<<"Finished!!!";
        fout.close();
}       // End function RandNum


//Function that calculates nodal mass consumption at each nodes
void NodalMass()
{// Start function NodalMass
        //Local variables for function NodalMass
        int   i,NumNodes, type, errcode;
        long  t, tstep, r,ht,du,qs;
        float q, d, h;
        float ChemMass, TotalChem, TankMass, TotalQ,NodeDemand;
        float TotalChemArr[90] = {0};
        float DemandArr[90] = {0};
        float chemical[90]={0};
```

```
float demand[90]={0};
char  id[15];
char  massFile[20]= {"NodalMass.txt"};
char  demandFile[20] = {"NodalDemand.txt"};
char tankmassFile[20]={"TankMass.txt"};

ofstream nkout(massFile, ios::app);//Uses nkout to write to NodalMass.txt
ofstream nkdmd(demandFile,ios::app);//Uses nkdmd to write to NodalDemand.txt
ofstream nktank(tankmassFile,ios::app);//Uses nktank to write to TankMass.txt

// Sets 4 decimal precision to outputs
nkout.precision (4);
nkdmd.precision (4);
nktank.precision (4);

// Opens EPANET file with .inp extension
// Epanet input file generated by PRPsym has filename format as follows:
// <outPRPsymFile>_EPANET_dt_<hydraulic time step in PRPsym>.inp
errcode=ENopen("nk_EPANET_dt_3600.inp","1.rpt","");
// Here outPRPsymFile=nk and hydraulic time step=3600
if (errcode>0)
        {
                ENclose();
                return;
        }

// Performs hydraulic simulation using EPANET engine
ENsolveH();

// EPANET node & time setting information
ENgetcount(EN_NODECOUNT, &NumNodes);// Number of network nodes       cout <<
"Number of Nodes: " << NumNodes <<endl<<endl;

ENgettimeparam(0, &du);// Retrieves simulation duration (du) information
```

```
        cout << "Simulation duration: " << du/3600<<"hrs"<<endl<<endl;
        ENgettimeparam(1, &ht);// Retrieves hydraulic time step (ht) used for simulation
        cout << "Hydraulic Time Step: " << ht<<"sec" <<endl<<endl;
        ENgettimeparam(2, &qs);// Retrieves quality time step (qs) used in simulation
        cout << "Quality Time Step: " << qs<<"sec" <<endl<<endl;
        ENgettimeparam(5, &r);// Retrieves reporting time step (r) used in simulation
        cout << "Reporting Time Step: " << r <<"sec"<<endl<<endl;
        // Starts EPANET water quality simulation
        ENopenQ();
        ENinitQ(0);
do
        {//Start do-while loop
        ENrunQ(&t);
        ChemMass = 0.0;
        TotalChem = 0.0;
        TankMass = 0.0;
        TotalQ = 0.0;

        for (i = 1; i<= NumNodes; i++) //Start loop for total no. of nodes

                {// start i
                        ENgetnodeid(i, id);
                        //id=node id assigned by EPANET to ith node
                        ENgetnodetype(i, &type);
                        //type=type of node (0=node,1=reservoir,2=tank)
                        ENgetnodevalue(i, EN_DEMAND, &d);
                        // d=demand at the node
                        ENgetnodevalue(i, EN_QUALITY, &q);
                        //q=chemical concentration
                        ENgetnodevalue(i, EN_HEAD, &h);
                        //h=head at the node

        if (type == 0)    //nodes
        {// start if for nodes
```

```
                // calculating nodal mass in each quality time step
                ChemMass = d*(qs/60)*GalToLit*(q/1000);
                //Adds chem mass each tstep
                TotalChemArr[i] = TotalChemArr[i] + ChemMass;
                TotalChem = TotalChem + TotalChemArr[i];

                NodeDemand=d*(qs/60);
                DemandArr[i] = DemandArr[i] + NodeDemand;//In gallons
                TotalQ = TotalQ + DemandArr[i]; //total demand (gallon)

        } //end of if for nodes

                else if (type == 2)          //tanks

        {//start if for tanks
        //Tank chem mass calc (g)
        TankMass = (h-TankMin)*(3.14*TankDia*TankDia/4)*CfToLit*(q/1000);
        }//end if for tanks

                else if (type == 1){ChemMass = 0.0;}//reservoirs

        }// end i

        ENnextQ(&tstep);// Starts water quality calculation for next time step

} while (tstep > 0);//End do-while loop


//-----------Writes demand and cumulative % nodal mass at end of simulation only----
                for (int pp=1;pp<=90;pp++)
                {// start pp
                        nkout<<"\t"<<TotalChemArr[pp]/72;
                        //Percentage Nodal Mass [100*Chemical/7200]
                        nkdmd<<"\t"<<DemandArr[pp];//Nodal demand
```

```
                }// end pp

                nktank<<TankMass/72<<endl;
                nkout<<endl;
                nkdmd<<endl;

//-------------------------------------------------------------------------------------

        cout << "Total mass consumed at Nodes: " << TotalChem <<"g"<<endl;
        cout << "Demands: " << TotalQ <<"Gallon"<<endl;
        nkout.close();
        nktank.close();
        nkdmd.close();

        ENcloseQ();// Close EPANET water quality simulation
        ENclose();// Close EPANET

}// End function nodal mass

// End of program PRPEpanetNK.cpp
```

3. Program III: Latin Hypercube Sampling within EPANET execution interface

```
/*****************************************************************************
Program:        LHSforGSA.cpp
                Performs Latin Hypercube Sampling (LHS) by generating 288
                input vector comprising of four input parameters a)injection
                duration, b) tank head fluctuation, c) nodal base demand, and d)
                chemical mass loading. These input parameters are randomly
                varied within the program interface and 288 EPANET .inp files are
                created.

                After simulation of hydraulics and water quality, the program will
                finally calculate Node Count i.e. Number of nodes receiving
                chemical concentration above the assumed threshold value


Author:         Nabin Khanal
                Civil and Environmental Engineering, University of Cincinnati

Date:           June 1st 2005

Last Updated:   July 15th 2005
Libraries linked:   epanet2vc.lib for EPANET
*****************************************************************************/
// Includes
# include <iostream>
# include <fstream>
# include <stdio.h>
# include <cstdlib>
# include "epanet2.h"
# include <time.h>
```

```
#define DIM 288// Total number of runs possible for this program

using namespace std;

//Function Declarations
double RandNum(double, double);
float RandNumFloat(float, float);

//Constants
const int  TotSim= 288;
// Total number of multipliers in the injection/demand pattern equivalent to total no. of //runs
const int seedValue=100; // Seed value used for this study are 100, 101, 102, and 103
const float threshold=2.0;
const int injNode=11; // injNode=Node Index  Node 17= Node index 11(Red zone node),
 //Node 6=Node index 6(Yellow zone node), Node 81=Node index 54(Green zone node)

const int  intNodes=90;//number of nodes for analysis and total nodes in network
const int Max = 90;
const double GalToLit = 3.785;//gal to liter conversion 1 gallon= 3.785 liter
const double CfToLit = 28.32;//cubic feet to liter conversion 1 cft = 28.316 liter
const float TankMin = 56.0;       //minimum tank water level (ft)

// Global Variables
long t, tstep, duration,hydtime,qualitystep,rptime,patime,ht,r,qs;
double baseDMD,BASEdmd;
float basedmd,bASEDMD,factor,total1,total2,Stype,Spatt,Tpatt,setting,level;
float d,q,h,total11,total22,ototal,obaseDMD,oSpatt,oStype,Smass;
float timepattern[DIM],injRate[DIM],newTankhead[DIM];
float status,TankMas,NodeDemand,TotalChem,ChemMas,TotalQ;
float multipliers[DIM]={0}; // 15 minutes pattern step for injection for total simulation
                            //period of 72 hrs
int jj,gg,patindex,count,cindex,cctype,numctrls,pindex,ntype;
int lindex=0,success=0,nindex=90;// Link index for pump and Node index for tank
int tindex = 0,nodeValue;
```

```
int errcode,nnodes,type;
float DemandArr[90] = {0};
float aConc[90]={0};
float TotalChemArr[90] = {0};
char injection[4]={"Inj"};//Id label for injection pattern
char pump[5]={"PUMP"};// Id label for pump
char  id[15];

//Variables used for cross-checking the outputs
float tPTT,iRate,nlevel,nsetting,oRate,oPTT,olevel,osetting;
int ncctype,nlindex=0,nnindex=90,olindex=0,onindex=90,ocount,occtype;

// Random Chemical Mass in gram
double cMass[DIM];
const double chemMassFrom= 2000.0;
const double chemMassTo= 9000.0;

// Random Tank Head Fluctuation in ft
double tHead[DIM];
const double theadFrom=1.0;
const double theadTo=100.0;

// Random Nodal Base Demand, GPM
float bsDMD[DIM];
float multFactorFrom=0.308;
float multFactorTo= 6.17;
//***********Main Program Body**********

void main()

{       // Start main

        srand(seedValue);
        char * EpaInpFile = "17.inp", * EpaRptFile = "1.rpt";
```

```
        // Change the input file name to 11.inp for inj. at node 17 (index 11);
        // 6.inp for inj. at node 6(index 6); and 81.inp for node 81 (index 54)

        errcode = ENopen(EpaInpFile, EpaRptFile,"");
        if (errcode>0)
        {
                ENclose();
        return;
        }

// Reading default EPANET values to variables
        errcode= ENgetcount(EN_NODECOUNT, &nnodes);
        errcode= ENgettimeparam(EN_DURATION, &duration);
        errcode= ENgettimeparam(EN_QUALSTEP, &qualitystep);
        errcode= ENgettimeparam(EN_REPORTSTEP, &rptime);
        errcode= ENgettimeparam(EN_PATTERNSTEP, &patime);


// Files generated as program output
ofstream fout1("Randomness.txt");
fout1.precision(3);// Random factor that multiplied the base demand
ofstream fout2("Data.txt");
fout2.precision(5);// Input vector set
ofstream fout3("Bdmd.txt");
fout3.precision(4);//Base demand for each nodes
ofstream fout4("aConc.txt");
fout4.precision(4);// Average concentration at a node
ofstream fout5("Dmd.txt");
fout5.precision(4);// Cumulative water consumption
ofstream fout6("Mass.txt");
fout6.precision(4);// Cumulative nodal mass consumed
fout2<<"InjRate"<<"\t"<<"InjDur"<<"\t"<<"maxTank"<<"\t"<<"Basedmd"<<"NodeValue"<<en
dl;
fout2<<"mg/min"<<"\t"<<"minutes"<<"\t"<<"feet"<<"\t"<<"GPM"<<"      "<<endl;
```

```
//******************Start of the LHS*******************

for (int k=0;k <TotSim;k++)
{          //Start of simulation loop
// Initializing variable with value 0
for (int yy=1;yy<=nnodes;yy++)
{
 TotalChemArr[yy]=0;
 DemandArr[yy]=0;
 aConc[yy]=0;
}

//***************Injection Duration*****************************
//Increase 1 in injection pattern multipliers          for each run
          for(int zz=(k+1);zz<=288;zz++)// 288 is total number of multipliers
                    {//Start kk
                    timepattern[k]=1;
//Assigning first pattern multiplier the value of 1, rest is 0; 1 represents 15 min injection
                    timepattern[zz]=0;
                    }
          count=0;
          count=k+1;
//Set the changing multipliers to the injection pattern(2)
ENsetpattern(2,timepattern,DIM);
// 2 is the pattern id for the injection, at each step 15 min injection period will be added
//*********************************************************

//******************Chemical Mass Loading**************
// Generation of Random Chemical Mass
cMass[k]=RandNum(chemMassFrom,chemMassTo); //Generating random chemical mass to be
injected
// Calculating Injection Rate for each run
injRate[k]=(1000*cMass[k])/(15*count);
```

```
//Based upon injection pattern and chemical mass, injection rate is calculated (mg/min)
// Set the injection rate to injection node
ENsetnodevalue(injNode,EN_SOURCEQUAL,injRate[k]);
//*************************************************

//**************Tank Head Fluctuation*************
tHead[k]= RandNum(theadFrom, theadTo);// Additional fluctuation in tank head from minimum
level
newTankhead[k]=TankMin+tHead[k];// Total fluctuation in tank head

// Set the pump controls according to new tank head
        ENgetcontrol(1, &cctype, &lindex, &setting,&nindex,&level);
        // Changing the simple control to update the head of water in tank

        if (cctype == EN_HILEVEL)
        {
                ENsetcontrol(1, cctype, lindex,setting,nindex, newTankhead[k]);
 //Tank head varies with each loop cycle
        }
// New tank head and pump controls updated
//*********************************************************

//*******************Base Demand************************
bsDMD[k]=RandNumFloat(multFactorFrom,multFactorTo);
// Random factor generation to increase/decrease base demand

//Multiply base demand of each node with a random factor
                total1=0;
                total2=0;
                for (int j = 1; j <=nnodes; j++)

                        {// start j
                        errcode = ENgetnodevalue(j, EN_BASEDEMAND, &basedmd);
                        if (errcode == 0)
```

```
                    {
                        total1+=basedmd;
                        baseDMD=basedmd*(bsDMD[k]);
                        ENsetnodevalue(j, EN_BASEDEMAND, baseDMD);
                        total2+=baseDMD;
                    }
                }//end j
//*****************************************************************
//******************End of LHS*************************************

//**********Solving Hydraulics and Water Quality***************
                errcode=ENsolveH();
                success++;
                cout<<"Run # "<<" "<<success<<" Done!!!"<<endl;
                ENopenQ();
                ENinitQ(0);

        do
        {
                ENrunQ(&t);
                ChemMas = 0.0;
                nodeCount=0;

                for (int i = 1; i<= nnodes; i++)
                {//start i
                        ENgetnodeid(i, id);//Node id
                        ENgetnodetype(i, &type);// Node type
                        ENgetnodevalue(i, EN_DEMAND, &d);//Water demand
                        ENgetnodevalue(i, EN_QUALITY, &q);//Chemical concentration
                        ENgetnodevalue(i, EN_HEAD, &h);// Node head
                        ENgettimeparam(EN_DURATION, &duration);// Simulation time
                        ENgettimeparam(EN_HYDSTEP, &hydtime);//Hydraulic time
```

```
                if (type == 0)    //nodes
                        {
                                ChemMas = d*(hydtime/60)*GalToLit*(q);
                        // calculating nodal mass in each hydraulic time step...mg
                                TotalChemArr[i] = TotalChemArr[i] + ChemMas;
                                NodeDemand=d*(hydtime/60)*GalToLit;
                                DemandArr[i] = DemandArr[i] + NodeDemand;//demand
                                if (DemandArr[i]==0)
                                {
                                        DemandArr[i]=0.00000001;
                                }
                                aConc[i]=TotalChemArr[i]/DemandArr[i];
                                //Average chemical concentration
                                if (aConc[i]>=threshold)
                                {nodeCount+=1;}

                        }//end if
                }// end i
                ENnextQ(&tstep);
        }       while (tstep > 0);// Next quality step
// Writing outputs to txt files
        for (int gg=1;gg<=nnodes;gg++)
                {
                        fout5<<DemandArr[gg]<<"\t";//Nodal Demand
                        fout6<<TotalChemArr[gg]<<"\t";//Chemical mass
                        fout4<<aConc[gg]<<"\t";//Avg. concentration
                }
                        fout4<<endl;
                        fout5<<endl;
                        fout6<<endl;
        ENcloseQ();
//***************** End of hydraulics and water quality simulaiton*********

// Outputs for Input Vector set
```

```
//*******************************
ENgetnodevalue(injNode,EN_SOURCEQUAL,&oRate);
//fout<<"Rate:"<<oRate<<endl;
ocount=0;
for (int hh=1;hh<=DIM;hh++)
        {
                ENgetpatternvalue(2,hh,&oPTT);
                if (oPTT==1){ocount+=1;}
                // fout1<<oPTT;
}
//cout<<"DmdMultCount:"<<ocount<<endl;
ENgetcontrol(1, &occtype, &olindex,&osetting,&onindex,&olevel);
//cout<<"New Tank Level:"<<olevel<<endl;
ototal=0;
for (int ff=1;ff<=nnodes;ff++)
{
ENgetnodevalue(ff, EN_BASEDEMAND, &obaseDMD);
ototal+=obaseDMD;
}
//fout1<<ototal;
fout2<<oRate<<"\t"<<ocount*15<<"\t"<<olevel<<"\t"<<ototal<<"\t"<<nodeValue<<endl;
//*****************************************

// Re-update base demand to original values
                total11=0;
                total22=0;
                for (int cc = 1; cc <=nnodes; cc++)
                        {// start cc
                        errcode = ENgetnodevalue(cc, EN_BASEDEMAND, &bASEDMD);
                        {
                                total11+=bASEDMD;
                                BASEdmd=bASEDMD*(1/(bsDMD[k]));
                                ENsetnodevalue(cc, EN_BASEDEMAND, BASEdmd);
                                total22+=BASEdmd;
```

```
                        }
                }//end cc

}// End of simulation loop

ENclose();// Close EPANET input file
cout<<"!!!! File closed!!! Simulation Ended!!!!"<<endl;
}// End main

 //Functions

// Generates double random number between the specified range
 double RandNum(double start, double end)
 {// start function RandNum
        for (int z=1;z<=TotSim;z++)
        {
                double range=(end-start);;
                double s=start+(double)range*rand()/(RAND_MAX+1);
                return s;
        }
 }// end function RandNum

// Generates float random number between the specified range
 float RandNumFloat(float start,float end)
 {//start
        for (int z=1;z<=TotSim;z++)
        {
                float range=(end-start);;
                float s=start+(float)range*rand()/(RAND_MAX+1);
                return s;
        }
 }// end function RandNumFloat
// End of program LHSforGSA.cpp
```

4. Program IV: Kolmogorov-Smirnov two-sample test

```
/*************************************************************************
Program:        KStest.cpp

                This program reads file N1.txt and N2.txt as inputs and tests
                the routine KSTWO to check if data in files N1.txt and N2.txt have
                same distributions by using the Kolmogorov-Smirnov two-sample test
                and returns K-S statistics.

                N1.txt contains data that falls in "fail" sample set
                N2.txt contains data that falls in "pass" sample set

Author:         Nabin Khanal
                Civil and Environmental Engineering, University of Cincinnati

Date:           May 20, 2005

Last updated:   July 10, 2005

Reference:      Functions reference: Press et al. (2002)
                C++ codes reference: Moreau, J-P.
                http://perso.wanadoo.fr/jean-pierre.moreau/Cplus/tkstwo_cpp.txt
                 [Accessed May 16, 2005]

*************************************************************************/
// Includes
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <iostream>
#include <fstream>
// Define number of data in file N1.txt and N2.txt
#define N1 1075
```

```
#define N2 77
using namespace std;
double DATA1[N1],DATA2[N2];

// Function Qks called by subroutine KSTWO
  double PROBKS(double ALAM) {
  double eps1=0.001, eps2=1.e-8;
  double a2,fac,temp,term,termbf;
  int j;
  a2=-2*ALAM*ALAM;
  fac=2.0;
  temp=0.0; termbf=0.0;
  for (j=1; j<=100; j++) {
    term=fac*exp(a2*j*j);
    temp += term;
    if (fabs(term) < eps1*termbf || fabs(term) < eps2*temp) return temp;
    fac=-fac;
    termbf=fabs(term);
  }
  return 1;
}

// Function that sorts an array RA of length n in ascending
//  order by the Heapsort method
void HPSORT(int n, double *RA)  {
  int i,ir,j,l;
  double rra;
  l=(n/2)+1;
  ir=n;
 e10:
  if (l > 1) {
    l--;
    rra=RA[l];
  }
```

```
  else {
    rra=RA[ir];
    RA[ir]=RA[1];
    ir--;
    if (ir==1) {
      RA[1]=rra;
      return;
    }
  }
  i=l;
  j=l+l;
e20:
  if (j <= ir) {
    if (j < ir)
      if (RA[j] < RA[j+1]) j++;
    if (rra < RA[j]) {
      RA[i]=RA[j];
      i=j; j=j+j;
    }
    else
      j=ir+1;
    goto e20;
  }
  RA[i]=rra;
  goto e10;

}

double MAX(double a, double b) {
  if (a>=b) return a;
  else return b;
}
```

```
/*************************************************
 Given an array DATA1 of N1 values, and an array  DATA2 of N2 values,
this routine returns the K-S  statistic D, and the significance level PROB for
 the null hypothesis that the data sets are drawn  from the same distribution.
 Small values of PROB  show that the cumulative distribution function  of DATA1
 is significantly different from that of  DATA2. The arrays DATA1 and DATA2
are modified by being sorted in ascending order (heapsort).
*************************************************/
void KSTWO(double *DATA1,int n1,double *DATA2,int n2,double *D,double *PROB) {
  int J1,J2;
  double DT,EN1,EN2,FN1,FN2,FO1,FO2;
  HPSORT(n1,DATA1);
  HPSORT(n2,DATA2);
  EN1=n1; EN2=n2;
  J1=1; J2=1;
  FO1=0; FO2=0;
  *D=0;
e1:if (J1 <= n1 && J2 <= n2)  {
            if (DATA1[J1] < DATA2[J2])  {
              FN1=J1/EN1;
              DT=MAX(fabs(FN1-FO2),fabs(FO1-FO2));
              if (DT > *D) *D=DT;
              FO1=FN1;
              J1++;
            }
            else {
              FN2=J2/EN2;
              DT=MAX(fabs(FN2-FO1),fabs(FO2-FO1));
              if (DT > *D) *D=DT;
              FO2=FN2;
              J2++;
     }
     goto e1;
```

```
  }
  *PROB=PROBKS(sqrt(EN1*EN2/(EN1+EN2))*(*D));
}

void main()
        {//start main
        double D,PROB;
        // Read files N1.txt in array DATA1, N2.txt in array DATA2
        ifstream infile1("N1.txt",ios::in);
        ifstream infile2("N2.txt",ios::in);

        for (int i=0;i<N2;i++)
                {
                        infile2>>DATA2[i];
                }

        for (int j=0;j<N1;j++)
                {
                        infile1>>DATA1[j];
                }

        KSTWO(DATA1,N1,DATA2,N2,&D,&PROB);
         cout.precision(4);
         cout<<"K-S test d statistic= "<<D<<endl;
         cout<<"K-S test P value= "<<PROB<<endl;

        }//end main

        // End of program KStest.cpp
```

APPENDIX III

1. GSA Algorithm

2. Input vector sets generated by GSA program

1. GSA Algorithm

Following are the steps coded to perform Generalized Sensitivity Analysis using Kolmogorov-Smirnov *d* statistic to rank the four input parameters based upon their sensitivity to simulation output, Node Count:

1. Identify the dynamic input variables to be included in the simulation; four variables are selected for this analysis.
    - Base Demand
    - Tank Head Fluctuation
    - Chemical Mass Loading
    - Injection Duration

2. Using EPANET toolkit function (**ENopen**), open an existing EPANET input file where demand pattern, injection pattern, injection node, injection source type are already defined.

3. The range for four dynamic variables were defined as follows:
    - Base Demand: Average system base demand was varied from 100 GPM to 2000 GPM, i.e. from 380 Lpm to 7600 Lpm. [Original average system base demand was 321.24 GPM (1227 Lpm)].
    - Tank Head Fluctuation: Highest water level in tank was varied from 1 ft to 100 ft, i.e. from 0.3 to 30.5 m.
    - Chemical Mass Loading: Conservative chemical mass injected into the system was varied from 2000 gram to 9000 gram.
    - Injection Duration: The injection duration was varied from 15 minutes (smallest time pattern in the input file) to total simulation duration (24-hour or 72-hour).

4. Determine the number of simulation runs.
   Since, minimum injection time period is 15 minutes, only 288 injection durations, starting from 15 min, 30 min, 45 min, and so on till 72 hr is possible. Thus, only

288 simulations run were possible in a single code execution for 72-hour simulation and 96 simulations run were possible for in a single code execution for 24-hour simulation.

5. Generate random values for each input variables using Uniform Probability Distribution, except Injection Duration.

6. Perform **Latin Hypercube Sampling (LHS)** to generate 288 input parameter vectors. Following steps were followed to perform LHS with the EPANET toolkit:

   a. Injection Duration: First the injection pattern was varied using **ENsetpattern** function in toolkit. For each run change the multipliers in injection pattern such that the injection duration is changed. For the first run only first demand multiplier will be 1 and remaining 287 multipliers will have 0 values. For second run first two multipliers will be 1, and will be rest zero. Following the same procedure for the last run all multipliers will be 1. For each run count the number of multiplier with the value equivalent to 1, and save it in a variable **Count**.

   b. Chemical Mass Loading: Take the random value generated for the chemical mass loading. Calculate the Injection Rate for the source as follows: Injection Rate = [Chemical Mass Loading/(**Count** x 15 min)] (mg/min)
   Using **ENsetnodevalue** function, assign the newly calculated injection rate to the injection node.

   c. Tank Head Fluctuation: The maximum water level in tank was calculated as Minimum tank head + Additional head. The minimum tank head was 56 ft, and additional head was calculated as a random number between 1 and 100 ft, following a uniform pdf. Using the **ENsetcontrol** function the pump controls were updated to operate up to the newly defined maximum water level.

d. Base Demand: Find out the number of nodes in the network, and for each node get the original base demand value defined in the input file. Select a random factor, generated from range of 0.31 to 6.17 (uniformly distributed) and multiply the original base demand, such that the average base demand lies between the ranges of 99 GPM to 1982 GPM. Using **ENsetnodevalue,** update the base demand for each node. Thus, LHS was performed using EPANET toolkit, and one input parameter vector was created for the first run.

7. Run the hydraulic and water quality simulation for the input vector created; calculate the nodal mass loading, nodal demand, and the average nodal concentration for the network. If the average concentration or chemical mass of a node is greater than the pre-defined average concentration or chemical mass threshold value, node was deemed as vulnerable. Count the number of nodes that are vulnerable and save it in a variable Node Count.

8. For a single simulation run write the inputs vector set and the Node Count, into text files. Change the base demand to the original base demand; divide the updated base demand by the same random factor used in step **6d**.

9. Repeat steps 5-9 for remaining simulation runs. The random seed number was changed and the code was executed for three more times, such that total simulation runs was 1152. Seed numbers 100 to 103 were used for 72-hour simulation and seed numbers 100 to 111 were used for 24-hour simulation.

10. From the results saved in the spreadsheet, categorize the input vector set into two sets “pass” and “fail” based upon the Node Count. If the Node Count is 0, the input vector was grouped in “pass” sample set, and if the Node Count is greater than 0, the input vector was grouped in “fail” sample set. Data in “pass” sample was saved in file N1.txt, and data in “fail” sample set was saved in file N2.txt.

11. Another code was executed to perform the K-S two-sample test. The code reads the txt files N1.txt and N2.txt and provides the K-S test *d* statistic and probability *P* value as an output. Based upon the K-S test *d* statistic value the sensitivity ranking for the four input variables were performed.

12. Steps 2-13 were repeated to calculate the K-S test statistics for different average concentration or chemical mass threshold values considered.

2. Input vector sets generated by GSA program

For injection made at node 17 and average concentration threshold of 3.0 mg/l, following is the data generated by Program III, which performs the LHS and provides Node Count as final simulation output. Seed number 100 was used and a single program execution results with 288 run is shown here:

Run #	Injection Rate mg/min	Injection Duration min	Tank Head feet	System Base Demand GPM	Chemical Mass gram	Node Count
		Input 1	Input 2	Input 3	Input 4	Output
1	138530	15	60.7	410.2	2078.0	0
2	185610	30	131.0	745.9	5568.3	0
3	161690	45	62.1	1433.0	7276.1	0
4	53375	60	129.0	1967.4	3202.5	0
5	82419	75	106.9	761.7	6181.4	0
6	53141	90	68.4	158.2	4782.7	0
7	64273	105	80.7	998.3	6748.7	0
8	29343	120	104.8	470.9	3521.2	0
9	21453	135	135.1	1812.8	2896.2	0
10	29582	150	135.5	466.9	4437.3	0
11	21997	165	108.7	1115.7	3629.5	0
12	13277	180	118.8	1240.9	2389.9	0
13	45946	195	70.5	766.0	8959.5	0
14	28683	210	93.7	1764.7	6023.4	0
15	18881	225	109.5	989.3	4248.2	0
16	9610.6	240	88.0	1112.6	2306.5	0
17	16751	255	96.5	1822.3	4271.5	0
18	21056	270	114.4	406.3	5685.1	0
19	18466	285	61.8	802.8	5262.8	0
20	7804.6	300	152.9	423.0	2341.4	0
21	23481	315	111.8	781.5	7396.5	0
22	12236	330	102.9	1014.1	4037.9	0
23	7028.7	345	69.8	1750.2	2424.9	0
24	10880	360	73.6	1466.8	3916.8	0
25	11754	375	64.0	606.8	4407.8	0
26	17329	390	106.1	485.5	6758.3	0
27	8668.5	405	99.9	1097.1	3510.7	0
28	21425	420	135.3	1761.5	8998.5	0
29	12881	435	105.2	1388.9	5603.2	0
30	5786.9	450	94.6	213.0	2604.1	0
31	14422	465	92.9	1518.5	6706.2	0
32	12715	480	96.3	1133.3	6103.2	0
33	8338.8	495	57.9	1380.0	4127.7	0
34	13753	510	129.7	1958.5	7014.0	0
35	13757	525	66.9	281.8	7222.4	2

Run #	Injection Rate mg/min	Injection Duration min	Tank Head feet	System Base Demand GPM	Chemical Mass gram	Node Count
		Input 1	Input 2	Input 3	Input 4	Output
36	11176	540	85.2	1139.9	6035.0	0
37	10864	555	107.5	1502.8	6029.5	0
38	6933.9	570	96.1	274.8	3952.3	0
39	12695	585	71.7	122.6	7426.6	10
40	6247.9	600	77.4	774.8	3748.7	0
41	6151.7	615	131.1	1036.1	3783.3	0
42	11344	630	110.6	371.3	7146.7	0
43	3512.8	645	101.1	1730.3	2265.8	0
44	7958.2	660	78.6	871.0	5252.4	0
45	10094	675	73.4	1366.1	6813.5	0
46	4897.3	690	94.5	859.8	3379.1	0
47	8892.3	705	123.2	858.6	6269.1	0
48	4101.1	720	118.1	431.1	2952.8	0
49	10940	735	66.5	1737.6	8040.9	0
50	7984.7	750	114.8	416.1	5988.5	0
51	3798.4	765	138.1	831.2	2905.8	0
52	4294.2	780	101.9	1067.3	3349.5	0
53	8807.3	795	129.7	129.3	7001.8	0
54	2759	810	80.3	1692.0	2234.8	0
55	10402	825	122.3	1300.6	8581.7	0
56	5341.2	840	112.7	896.0	4486.6	0
57	4142.9	855	127.5	136.2	3542.2	0
58	9984.6	870	101.6	1827.5	8686.6	0
59	6845.9	885	70.2	402.6	6058.6	0
60	6739.4	900	117.1	1605.0	6065.5	0
61	6175.8	915	118.0	1199.3	5650.9	0
62	4609.7	930	101.8	1285.9	4287.0	0
63	2719.1	945	89.5	1297.9	2569.5	0
64	5860	960	66.4	618.8	5625.6	0
65	7627.2	975	87.0	1315.7	7436.5	0
66	4295	990	83.6	1523.4	4252.1	0
67	2409	1005	66.0	1855.7	2421.0	0
68	7550.2	1020	138.2	1498.9	7701.2	0
69	8498.3	1035	91.2	872.4	8795.7	0
70	2342.4	1050	135.8	618.6	2459.5	0
71	2214.7	1065	145.8	1791.2	2358.7	0
72	5886.2	1080	129.5	1318.9	6357.1	0
73	4378.6	1095	101.9	949.7	4794.6	0
74	3140.3	1110	101.6	690.2	3485.7	0
75	2198.8	1125	60.9	361.1	2473.7	0
76	2642.8	1140	122.8	1337.5	3012.8	0
77	1991.8	1155	79.5	1688.0	2300.5	0
78	4225.6	1170	67.8	443.7	4944.0	0
79	6701	1185	133.4	1979.1	7940.7	0
80	4967.1	1200	148.1	1591.4	5960.5	0
81	4849.2	1215	63.6	1599.0	5891.8	0

Run #	Injection Rate mg/min	Injection Duration min	Tank Head feet	System Base Demand GPM	Chemical Mass gram	Node Count
		Input 1	Input 2	Input 3	Input 4	Output
82	3480.5	1230	113.4	1743.8	4281.0	0
83	5911.3	1245	102.0	968.8	7359.6	0
84	2762.9	1260	73.6	1245.2	3481.3	0
85	2023	1275	116.1	1531.9	2579.3	0
86	2085.6	1290	81.9	1667.0	2690.4	0
87	6120.6	1305	90.2	488.8	7987.4	0
88	3005.3	1320	118.5	1108.7	3967.0	0
89	4770.3	1335	101.0	990.3	6368.4	0
90	2442.3	1350	91.5	818.3	3297.1	0
91	6329.4	1365	61.0	1060.6	8639.6	0
92	6490.5	1380	152.9	1904.8	8956.9	0
93	2290.2	1395	127.4	1810.8	3194.8	0
94	3027.6	1410	72.4	1247.9	4268.9	0
95	3187.7	1425	93.1	547.8	4542.5	0
96	3811.9	1440	60.0	348.4	5489.1	1
97	2931.2	1455	154.6	1250.6	4264.9	0
98	3409.4	1470	134.9	266.1	5011.8	0
99	1965.9	1485	103.5	1290.9	2919.4	0
100	3119.6	1500	115.8	1107.0	4679.4	0
101	4178.6	1515	82.0	793.9	6330.6	0
102	3338.4	1530	58.0	580.2	5107.8	0
103	2322.4	1545	64.2	323.6	3588.1	1
104	1877.9	1560	98.3	1497.6	2929.5	0
105	2496.6	1575	107.5	1424.5	3932.1	0
106	4494.2	1590	97.1	1616.3	7145.8	0
107	4582.8	1605	129.0	1694.3	7355.4	0
108	2884.1	1620	141.7	894.3	4672.2	0
109	3754.6	1635	104.2	373.7	6138.8	0
110	1883.7	1650	113.7	1864.1	3108.1	0
111	3895.4	1665	134.1	450.7	6485.8	0
112	2417.8	1680	65.4	1118.4	4061.9	0
113	4022.1	1695	77.3	1527.4	6817.5	0
114	2610.6	1710	117.8	1595.0	4464.1	0
115	3496.4	1725	66.9	1796.4	6031.3	0
116	4891.6	1740	95.2	814.5	8511.4	0
117	4537.5	1755	152.2	1762.7	7963.3	0
118	2616.3	1770	115.9	1889.7	4630.9	0
119	1285.1	1785	76.7	491.6	2293.9	0
120	4715.8	1800	103.2	940.3	8488.4	0
121	1720.3	1815	77.2	1185.7	3122.3	0
122	2159.8	1830	78.0	623.0	3952.4	0
123	1200.3	1845	134.0	959.1	2214.6	0
124	2403.6	1860	85.1	1768.4	4470.7	0
125	3432.7	1875	57.6	1424.4	6436.3	0
126	3630.2	1890	95.6	181.5	6861.1	9
127	1552.7	1905	93.2	1283.4	2957.9	0

Run #	Injection Rate mg/min	Injection Duration min	Tank Head feet	System Base Demand GPM	Chemical Mass gram	Node Count
		Input 1	Input 2	Input 3	Input 4	Output
128	1287.6	1920	90.3	996.6	2472.2	0
129	3276.9	1935	134.4	393.5	6340.8	0
130	1213.8	1950	57.9	1189.5	2366.9	0
131	4317.2	1965	65.4	845.6	8483.3	0
132	2546.1	1980	145.7	1617.4	5041.3	0
133	3234.1	1995	119.6	424.4	6452.0	0
134	2758.6	2010	107.6	786.5	5544.8	0
135	1173.8	2025	74.0	1870.2	2376.9	0
136	3501.6	2040	71.8	1470.6	7143.3	0
137	2271	2055	75.0	506.3	4666.9	0
138	1526.9	2070	113.2	1753.0	3160.7	0
139	1309.8	2085	90.1	586.6	2730.9	0
140	1345.7	2100	127.8	1386.0	2826.0	0
141	1425	2115	79.6	833.1	3013.9	0
142	3335.6	2130	112.5	170.8	7104.8	0
143	2958.8	2145	110.1	437.4	6346.6	0
144	2352.7	2160	107.3	1952.8	5081.8	0
145	1968.3	2175	63.3	1702.8	4281.1	0
146	1270.9	2190	64.5	1236.7	2783.3	0
147	1834.9	2205	87.8	1599.4	4046.0	0
148	3517.3	2220	98.0	1340.1	7808.4	0
149	1770.4	2235	136.8	479.2	3956.8	0
150	1389.1	2250	69.6	779.5	3125.5	0
151	3255.2	2265	76.5	1131.9	7373.0	0
152	1874.8	2280	131.5	1874.5	4274.5	0
153	1529.5	2295	114.4	681.6	3510.2	0
154	1021.3	2310	91.3	933.9	2359.2	0
155	1059.6	2325	94.2	1341.1	2463.6	0
156	2136.2	2340	155.0	126.6	4998.7	0
157	1930.3	2355	145.4	1124.6	4545.9	0
158	2202.1	2370	140.0	1288.4	5219.0	0
159	992.19	2385	69.9	399.6	2366.4	0
160	3519.4	2400	123.6	236.2	8446.6	0
161	1185.3	2415	63.7	1778.7	2862.5	0
162	1125.5	2430	149.5	1940.8	2735.0	0
163	3516.2	2445	76.2	482.8	8597.1	0
164	3029.1	2460	61.7	739.5	7451.6	0
165	1999.1	2475	60.8	1805.2	4947.8	0
166	2916.7	2490	111.4	1216.7	7262.6	0
167	1679.1	2505	115.9	1810.2	4206.1	0
168	2289.2	2520	91.5	279.1	5768.8	7
169	3526	2535	95.2	1950.9	8938.4	0
170	1532.8	2550	94.9	1881.6	3908.6	0
171	2393.9	2565	66.3	900.2	6140.4	0
172	3357.5	2580	76.7	713.1	8662.4	0
173	2920.9	2595	151.1	1825.8	7579.7	0

Run #	Injection Rate mg/min	Injection Duration min	Tank Head feet	System Base Demand GPM	Chemical Mass gram	Node Count
		Input 1	Input 2	Input 3	Input 4	Output
174	3069.6	2610	121.3	1067.2	8011.7	0
175	1974.4	2625	135.9	982.2	5182.8	0
176	1492.4	2640	68.7	266.5	3939.9	2
177	1587.4	2655	141.2	480.1	4214.5	0
178	3306.4	2670	116.0	346.3	8828.1	0
179	1209.4	2685	61.7	134.3	3247.2	2
180	864.64	2700	115.5	443.3	2334.5	0
181	3068.6	2715	83.1	222.8	8331.2	8
182	1131	2730	105.9	1759.1	3087.6	0
183	1571.7	2745	75.3	879.0	4314.3	0
184	2406.9	2760	122.9	510.0	6643.0	0
185	3137	2775	138.0	1379.3	8705.2	0
186	1278.5	2790	90.4	1737.8	3567.0	0
187	2775.4	2805	83.8	1153.9	7785.0	0
188	2314.4	2820	112.4	1384.2	6526.6	0
189	938.91	2835	148.1	1113.4	2661.8	0
190	821.91	2850	152.8	1472.5	2342.4	0
191	2631	2865	93.2	971.3	7537.8	0
192	2909.3	2880	80.1	272.1	8378.8	5
193	862.19	2895	144.4	965.6	2496.0	0
194	1675.9	2910	140.5	461.3	4876.9	0
195	1184.4	2925	102.6	1636.5	3464.4	0
196	2778.9	2940	97.0	1656.9	8170.0	0
197	2224.7	2955	75.0	444.1	6574.0	0
198	874.08	2970	73.1	1202.8	2596.0	0
199	2604.6	2985	121.4	1702.7	7774.7	0
200	2556.9	3000	152.6	1496.4	7670.7	0
201	1046.5	3015	126.1	155.1	3155.2	0
202	2347.2	3030	114.1	1194.3	7112.0	0
203	1261.6	3045	121.0	434.8	3841.6	0
204	968.58	3060	85.9	367.3	2963.9	0
205	2751.4	3075	83.8	1228.9	8460.6	0
206	2184.3	3090	76.1	191.7	6749.5	7
207	2535.8	3105	78.8	1657.9	7873.7	0
208	1169.5	3120	94.0	530.4	3648.8	0
209	1282	3135	99.4	1924.1	4019.1	0
210	1120.8	3150	66.6	779.5	3530.5	0
211	694.61	3165	123.7	513.0	2198.4	0
212	1460.9	3180	85.2	1219.1	4645.7	0
213	1824	3195	118.8	768.8	5827.7	0
214	1643.1	3210	68.9	1953.1	5274.4	0
215	752.57	3225	148.3	615.3	2427.0	0
216	1236.3	3240	148.2	1744.5	4005.6	0
217	1336.6	3255	138.6	1472.1	4350.6	0
218	1519.7	3270	62.0	1432.0	4969.4	0
219	1916.9	3285	87.1	816.6	6297.0	0

Run #	Injection Rate mg/min	Injection Duration min	Tank Head feet	System Base Demand GPM	Chemical Mass gram	Node Count
		Input 1	Input 2	Input 3	Input 4	Output
220	798.06	3300	105.8	352.1	2633.6	0
221	2380.9	3315	135.3	709.6	7892.7	0
222	1223.4	3330	59.2	757.3	4073.9	0
223	735.53	3345	124.5	952.6	2460.3	0
224	693.02	3360	130.5	307.4	2328.5	0
225	1592.9	3375	147.8	1106.4	5376.0	0
226	2034.5	3390	86.5	637.0	6897.0	0
227	1176.9	3405	151.2	1561.9	4007.3	0
228	2353.4	3420	91.5	1725.9	8048.6	0
229	1180.6	3435	149.4	383.6	4055.4	0
230	2238.8	3450	73.5	1312.6	7723.9	0
231	1443.3	3465	111.4	1753.4	5001.0	0
232	2530.8	3480	104.9	1927.1	8807.2	0
233	2058.6	3495	91.6	875.6	7194.8	0
234	808.01	3510	72.7	1060.7	2836.1	0
235	1932.1	3525	60.6	344.9	6810.7	1
236	1725.2	3540	58.1	1662.9	6107.2	0
237	913.52	3555	153.0	1686.9	3247.6	0
238	1984.3	3570	89.4	1952.9	7084.0	0
239	1539.1	3585	93.8	683.5	5517.7	0
240	2309.8	3600	83.1	446.9	8315.3	2
241	1815.6	3615	88.6	740.4	6563.4	0
242	597.16	3630	134.2	918.9	2167.7	0
243	872.62	3645	152.3	1668.2	3180.7	0
244	1063.1	3660	145.7	935.8	3890.9	0
245	1130.6	3675	90.1	1009.3	4155.0	0
246	1703.6	3690	78.1	1605.6	6286.3	0
247	1144.1	3705	64.6	1752.9	4238.9	0
248	1846	3720	132.9	1823.3	6867.1	0
249	684.87	3735	72.4	1095.7	2558.0	0
250	881.28	3750	119.9	591.8	3304.8	0
251	1018.7	3765	128.8	335.8	3835.4	0
252	1453.2	3780	89.0	1081.7	5493.1	0
253	2190.7	3795	93.2	233.1	8313.7	9
254	1338.7	3810	59.8	402.4	5100.4	1
255	1658	3825	116.2	711.0	6341.9	0
256	1419.2	3840	102.2	310.4	5449.7	5
257	805.91	3855	108.5	1645.1	3106.8	0
258	1613.4	3870	71.2	998.0	6243.9	0
259	995.22	3885	71.4	869.4	3866.4	0
260	1419.1	3900	86.4	1893.4	5534.5	0
261	1554.4	3915	88.1	1016.1	6085.5	0
262	1696.9	3930	72.6	1466.5	6668.8	0
263	512.55	3945	98.0	814.1	2022.0	0
264	1573.4	3960	147.6	1024.5	6230.7	0
265	1107.5	3975	79.5	510.5	4402.3	0

Run #	Injection Rate mg/min	Injection Duration min	Tank Head feet	System Base Demand GPM	Chemical Mass gram	Node Count
		Input 1	Input 2	Input 3	Input 4	Output
266	2127.9	3990	98.2	1261.6	8490.3	0
267	2183.2	4005	127.9	1373.3	8743.7	0
268	2005.4	4020	97.2	1971.3	8061.7	0
269	1363.6	4035	154.6	1194.4	5502.1	0
270	2129.1	4050	153.1	694.2	8622.9	0
271	844	4065	128.5	575.0	3430.9	0
272	524.49	4080	87.7	366.7	2139.9	0
273	1510.7	4095	126.1	1111.4	6186.3	0
274	1304.2	4110	69.9	1325.1	5360.3	0
275	1259.2	4125	66.1	237.0	5194.2	3
276	1887.6	4140	141.8	451.8	7814.7	0
277	1441.9	4155	155.5	837.3	5991.1	0
278	1335.4	4170	132.2	313.4	5568.6	0
279	1313.1	4185	78.8	583.2	5495.3	0
280	1387.1	4200	154.3	1524.2	5825.8	0
281	1638.8	4215	59.8	444.7	6907.5	1
282	2045.8	4230	108.5	1970.4	8653.7	0
283	696.04	4245	75.8	1190.2	2954.7	0
284	908.76	4260	117.9	1797.7	3871.3	0
285	1864.6	4275	122.3	1003.8	7971.2	0
286	688.24	4290	135.8	1019.6	2952.5	0
287	838.13	4305	110.3	1143.6	3608.1	0
288	1386.3	4320	73.0	1014.5	5988.8	0

APPENDIX IV

1. Exposure CDFs for 18 deterministic simulations

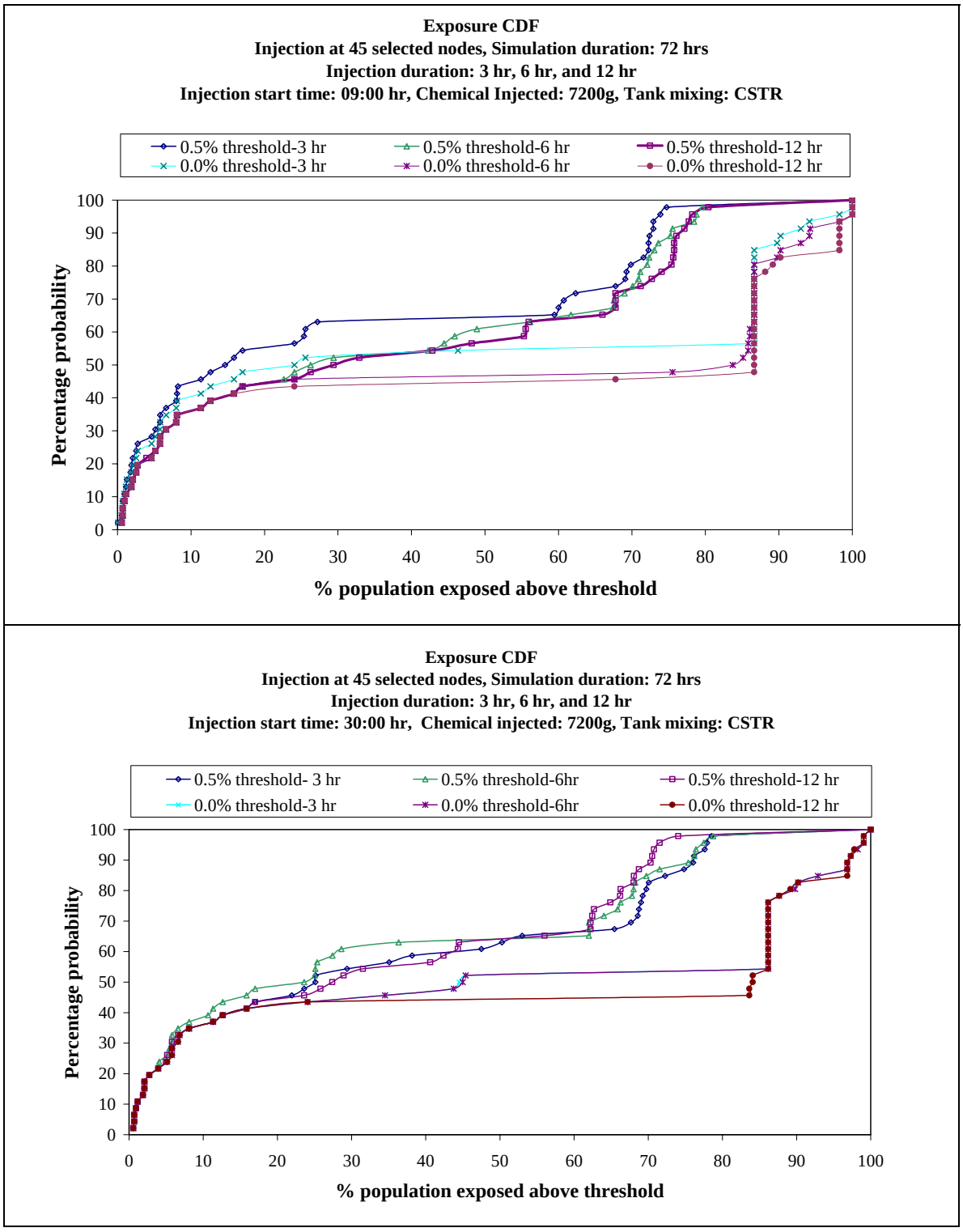

Figure A1.1: Exposure CDFs for system with CSTR tank mixing and varying injection duration and timing.

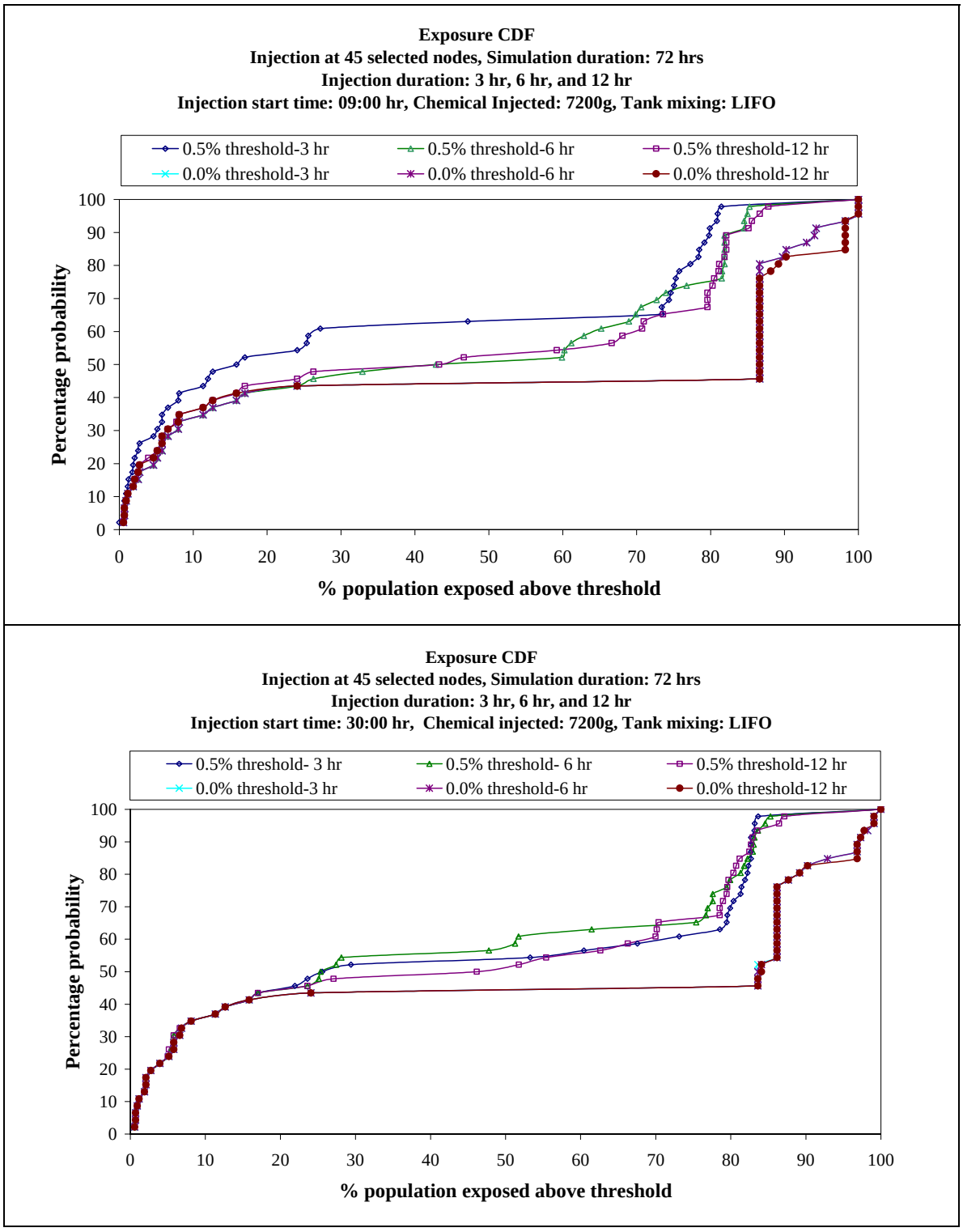

Figure A1.2: Exposure CDFs for system with LIFO tank mixing and varying injection duration and timing.

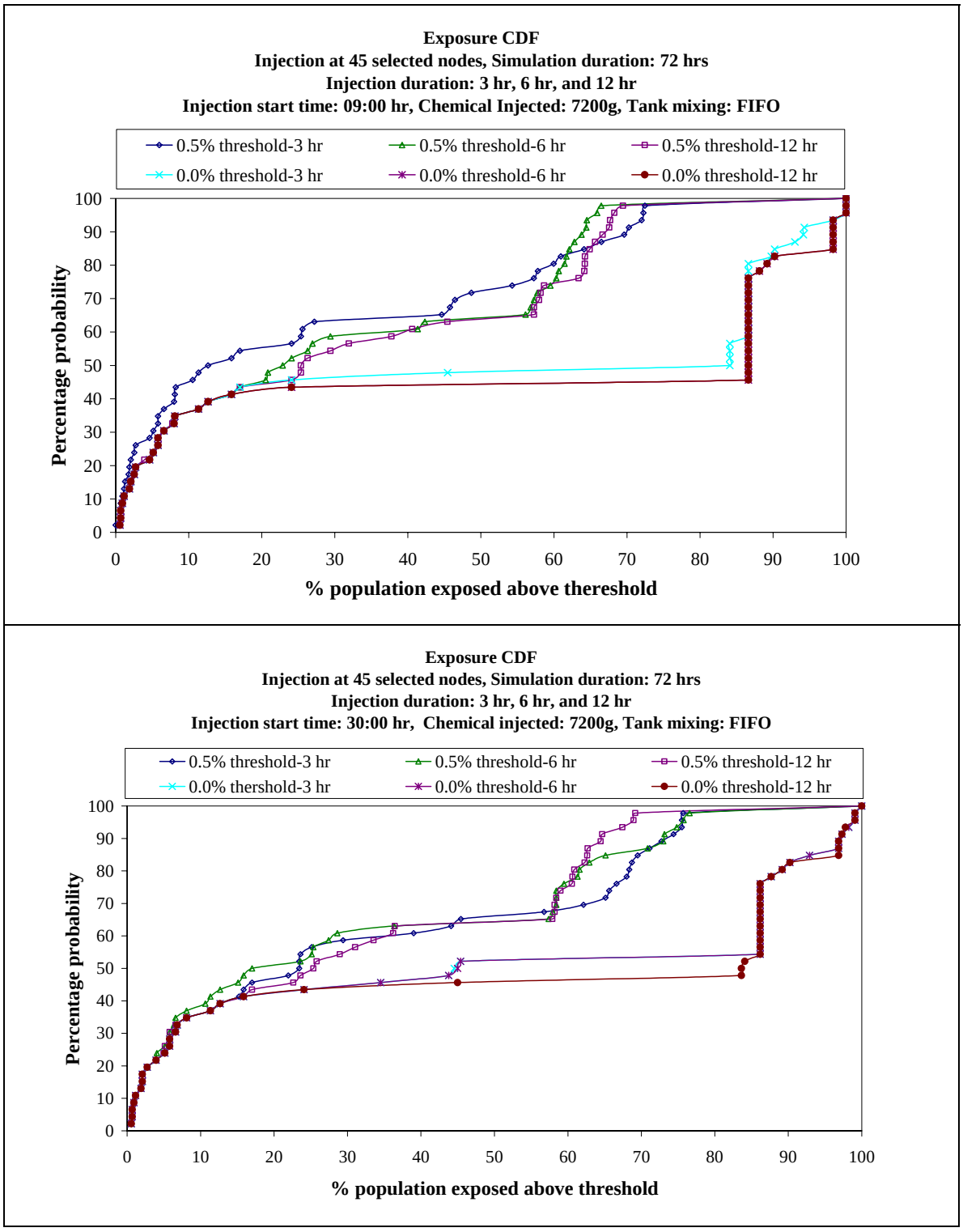

Figure A1.3: Exposure CDFs for system with FIFO tank mixing and varying injection duration and timing.