A Spatial Environmental Risk Assessment Methodology for Risk-Based Decision Making at Large, Complex Facilities

by

Lesley Isabel Hay Wilson, B.S., M.S., P.E.

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin
December 2000

A Spatial Environmental Risk Assessment Methodology for Risk-Based Decision Making at Large, Complex Facilities

Approved by
Dissertation Committee:

Randall J. Charbeneau, Supervisor

David R. Maidment, Co-Supervisor

Robert B. Gilbert

Sheldon Landsberger

Raymond C. Loehr

Dedication

This Dissertation is dedicated to Margaret and James Hay who taught me to strive for excellence, know no limits and set my sights on great things.

Acknowledgements

I wish to express my sincere gratitude to Randall Charbeneau, David Maidment, Robert Gilbert, Raymond Loehr and Sheldon Landsberger for their learned guidance and insight throughout this effort. Encouragement to pursue this work and belief in its value were sustaining motivations provided by James Rocco, John Till and Stephen Pause, for which I am deeply indebted. This research was accomplished with support from the National Science Foundation and with data and information from BP Amoco and their consultant, Langan Engineering. For these, I am very thankful. Special thanks go to Susan Sharp and Ed Merrick of BP Amoco for their support of my work. I wish to thank the other University of Texas students who worked on the risk-based decision support research program, Andrew Romanek, Lizan Mayoral (Koerner) and Julie Kim for all their hard work. I also thank the members of the CRWR research team, especially Dr. Francisco Olivera, for their friendship and GIS guidance. I praise God for His love and the many blessings of my life. Among these are the unwavering love and support of Thomas Wilson, Margaret and James Hay and all the rest of my family, without which I could not have accomplished this work.

A Spatial Environmental Risk Assessment Methodology for Risk-Based Decision Making at Large, Complex Facilities

Publication No._____________

Lesley Isabel Hay Wilson, Ph.D.
The University of Texas at Austin, 2000

Supervisor: Randall J. Charbeneau
Co-Supervisor: David R. Maidment

Spatial Environmental Risk Assessment (SERA) methodology is the integration of geographic information systems (GIS), (i.e., spatial databases), relational databases and environmental models to practically implement environmental risk assessment. The SERA methodology was developed to facilitate evaluation of multiple source areas and exposure pathways that are typical at large, complex facilities. The methodology includes spatial and tabular databases for the site conceptual model, integrated exposure pathway analysis through linkages between the site conceptual model database and fate and transport models. The methodology includes both deterministic and probabilistic calculations. First order uncertainty analysis is used for determining variance in exposure pathway ending concentrations using confidence intervals for modeling input parameters developed based on site data, reference information and judgments about parameter variability.

The impacts of multiple sources at one receptor location are analyzed through a process of segmentation of the exposure pathways, so that parallel and serial exposure pathway segments can be analyzed. The relationships between the exposure pathway segments and the source areas, the model scenario data and the results are managed in the site conceptual model database. Spreadsheets linked to the site conceptual model database are used to calculate attenuation factors using fate and transport models. For the deterministic calculations, the exposure pathway completeness is determined based on comparison of ending concentrations to target levels. For the probabilistic calculations, the exposure pathway completeness is determined based on the probability that the ending concentration is below the target level and the coefficient of variation in the ending concentration is used to determine the value of additional information.

The SERA methodology was implemented using PC-based software and examples from a case study are presented. The case study example results are used to illustrate the visualization benefits of the GIS implementation and the use of the uncertainty results to support decision making. The case study examples illustrated the applicability of the SERA methodology for groups of exposure pathways impacting single receptor locations and the data management benefits were demonstrated for the site conceptual model.

Table of Contents

List of Tables……………………………………………………………..…...…xiii

List of Figures………………………………………………………...……..xv

Nomenclature………………………………………………………....…..xvii

Chapter 1 : Introduction..1
1.1 Background...1
1.2 Problem Statement..1
1.2.1 Critical Issues 3
1.2.1.1 Site Conceptual Model for Site-Wide Risk Assessment 4
1.2.1.2 Uncertainty Analysis 5
1.2.1.3 Data Management...5
1.2.1.4 Solutions Offered by this Research 6
1.3 Research Objectives and Hypothesis....................................7
1.3.1 Spatial Site Conceptual Model 10
1.3.2 Integrating Existing Environmental Modeling into the Spatial Environmental Risk Assessment 11
1.3.3 Uncertainty Analysis 11
1.3.4 Data Management..12
1.4 Organizational Statement 12

Chapter 2 : Literature Review 13
2.1 Risk Assessment...13
2.1.1 National Academy of Sciences Paradigm 16
2.1.1.1 Hazard Identification 17
2.1.1.2 Dose-Response Assessment 17
2.1.1.3 Exposure Assessment 18
2.1.1.4 Risk Characterization 19

    2.1.2 Risk Assessment for Radioactive Materials 19
    2.1.3 Site Conceptual Models...21
    2.1.4 Risk-Based Decisions 22
2.2 Environmental Modeling and Management using GIS 24
    2.2.1 Concepts and terminology 24
    2.2.2 Environmental Modeling and Management using GIS 27
    2.2.3 Evaluation of the GIS Literature Reviewed 35
2.3 Relational Databases..36
2.4 Uncertainty in Pathway Completeness Evaluation.............................37
2.5 Summary..39

Chapter 3 : Methodology 41
3.1 Overview of the Spatial Environmental Risk Assessment Methodology...41
3.2 Data Model 44
    3.2.1 Exposure Pathways...44
        3.2.1.1 Source Area 45
        3.2.1.2 Exposure Pathway Segments......................................46
        3.2.1.3 Transition Points...47
        3.2.1.4 Points of Demonstration 48
    3.2.2 Site Conceptual Model 49
    3.2.3 Temporal Considerations..51
3.3 Calculation Methods...53
    3.3.1 Deterministic Case...59
    3.3.2 Probabilistic Case 60
3.4 Results Evaluation 66
3.5 Summary..71

Chapter 4 : Application Procedure 72
4.1 Software Tools...72
4.2 Representing the Site Conceptual Model Elements using GIS 73

4.2.1 Sources and Source Areas 74
4.2.2 Transition Points..79
4.2.3 Points of Demonstration 81
4.2.4 Lateral Transport Mechanisms 83
4.2.5 Vertical Transport Mechanisms..87
4.2.6 Spatial Site Conceptual Model 87
4.3 Building the Site Conceptual Model Database....................................92
4.3.1 Relationship Keys and Unique Identifiers.................................94
4.3.2 Types of Tables in the Database 100
4.3.2.1 Input Tables 100
4.3.2.2 GIS Attribute Tables...101
4.3.2.3 Spreadsheet Results Tables 102
4.3.2.4 Query Tables...102
4.3.2.5 Reference Tables 102
4.3.3 Primary Tables..103
4.3.3.1 PrelimPathway Table..103
4.3.3.2 Pathway Table 109
4.3.3.3 Pathway_Segments Table...111
4.3.3.4 Chemical of Concern ID Table.................................115
4.3.3.5 Source Variables Table...117
4.3.3.6 Subsurface Variables Table 120
4.3.3.7 Subsurface Correlations Table..................................122
4.3.3.8 Protective Exposure Concentrations Table................123
4.4 Modeling Fate and Transport along an Exposure Pathway using Analytical Models Developed in Spreadsheets 127
4.4.1 Deterministic Case...132
4.4.2 Probabilistic Case 133
4.4.3 Input Parameter Values 136
4.5 Summary..141

Chapter 5 : Results and Discussion 143
5.1 Spatial and Tabular Database Model for Risk Assessment..............144
5.2 Presentation of Examples from the Case Study Facility 147
5.2.1 Example Preliminary Exposure Pathway Evaluation..............148
5.2.2 Example Exposure Pathways...153
5.2.3 Initial Transport Segment Calculation Results.........................159
5.2.4 Discussion of Initial Transport Segment Results 165
5.2.5 Additional Transport Segment Calculation Results 166
5.2.6 Probabilistic Calculations...175
5.2.7 Discussion of Probabilistic Calculation Results.......................183
5.2.8 Additional Calculations for Variance Estimation.....................189
5.2.9 Comparison of Predicted Concentrations to Field Data 193
5.3 Visualization Benefits...200
5.4 Spatial Environmental Risk Assessment Method Evaluation 210
5.4.1 Software Reliability...210
5.4.2 Method Usability 212
5.4.3 Meta Data Management...213
5.4.4 Resources...215
5.5 Application of the Spatial Environmental Risk Assessment Methodology to other Facilities...218
5.5.1 Data Requirements 218
5.5.2 Customizable Features...221
5.5.3 Use of Additional Models..222
5.6 Summary...223
Chapter 6 : Conclusions and Recommendations 224
6.1 Conclusions 224
6.1.1 SERA Data Model 226
6.1.2 Calculation Methods..227
6.1.3 Spatial Site Conceptual Models...228

6.1.4 Uncertainty Analysis 229
6.1.5 Visualization Benefits... 230
6.2 Limitations of the SERA Methodology.. 232
6.3 Recommendations for Future Work 233

Appendices………………………………………………….…………….237

References……………………………………………………...………….418

Vita……………………………………………………………...……….426

List of Tables

Table 1.1 - Improved Methods for Large Facility Risk Assessment......................9
Table 3.1 - Fate and Transport Equations Used in the Example Calculations in this Research..53
Table 4.1- Site Conceptual Model GIS Data Layers..73
Table 4.2 - Example Source Area Attribute Table 78
Table 4.3 - Source Areas and PODs Matching Table for Groundwater Transport84
Table 4.4 - Key Fields for the Pathway and Pathway_Segments Tables 95
Table 4.5 - Preliminary Pathway Table Design...106
Table 4.6 - Soil Depth Range Values Look Up Table..106
Table 4.7 - Screening Values used for the Case Study Facility 108
Table 4.8 - Medium Specific Concentration Table Design................................108
Table 4.9 - Pathway Table Design...110
Table 4.10 - Pathway Segment Table Design..113
Table 4.11 - COC_ID Table Design..116
Table 4.12 - Source Variables Table Design..118
Table 4.13 - Subsurface Input Variables Table Design......................................121
Table 4.14 - Subsurface Input Variable Correlation Coefficients Table Design.124
Table 4.15 - Protective Exposure Concentrations Table Design........................124
Table 4.16 - Example Exposure Pathway Completeness Comparison...............126
Table 4.17 - Example Query Design for the Groundwater Transport Segment..129
Table 4.18 - Results Table Generated in the Spreadsheet for the Probabilistic Case for the Groundwater to Surface Water Exposure Pathway.........................131
Table 4.19 - Variable Description Table Design..138
Table 4.20 - Example Random Variables..139
Table 4.21 - Example Parameters Treated as Constants 141
Table 5.1- Results of the Comparison of Representative Soil Concentrations for Preliminary Source Areas to the Medium Specific Concentrations...........152
Table 5.2 - Example Exposure Pathways..154
Table 5.3 - Case Study Exposure Pathway Records..158
Table 5.4 - Case Study Pathway Segment Records..158
Table 5.5 - Predicted Concentrations at the Points of Demonstration................160
Table 5.6 - Protective Exposure Concentrations 161
Table 5.7 - Exposure Pathway Groups 167
Table 5.8 - Predicted Concentrations at the Points of Demonstration................171
Table 5.9 - Results from Probabilistic Calculations...176
Table 5.10 - Variance in Surface Water and Indoor Air Concentrations 178
Table 5.11 - Probabilistic Comparison of Ending Concentration to Target Level 179
Table 5.12 - Predicted Mean Groundwater Contributions of Benzene to Surface Water for Group G1...185

Table 5.13 - Contributions to Indoor Air Concentrations of Benzene from the Group G6 Source Areas....185
Table 5.14 - Comparison of Variance Results for the Inclusion of the Correlation of Parameters Across Exposure Pathways....192
Table 5.15 - Comparison of Calculated and Measured Concentrations of Benzene in Groundwater....197

List of Figures

Figure 2.1 - Simplified Process View of Risk Assessment 15
Figure 3.1 - Overview of the Spatial Environmental Risk Assessment (SERA) Methodology..41
Figure 3.2 - Exposure Pathway Elements (After Koerner *et al*., 1998) 45
Figure 3.3 - Exposure Pathway Segments and Multiple Contributing Source Areas (After Koerner, *et al*., 1998)..51
Figure 3.4 - Single Exposure Pathway Segments and Exposure Pathways Described by the Attenuation Factors 55
Figure 3.5 - Calculation of Groundwater Transport to Surface Water in the Deterministic Case...62
Figure 3.6 - Calculation of Groundwater Transport to Surface Water in the Probabilistic Case 63
Figure 3.7 - Probabilistic Comparison of Ending Concentration Distribution to Target Level..69
Figure 4.1 - Example Source Areas for the Case Study Facility...........................76
Figure 4.2 - Example Transition Points for the Case Study Facility.....................80
Figure 4.3 - Example Exposure Pathways Illustrating POD Locations 82
Figure 4.4 - Example Exposure Pathways Illustrating the Groundwater Transport Segments...86
Figure 4.5 - Example Site Conceptual Model Map 89
Figure 4.6 - Example Site Conceptual Model Map including the Media Mixing Boxes 91
Figure 4.7 - Site Conceptual Model Database Table Relationships......................99
Figure 4.8 - Site Conceptual Model Database Table Relationships for the Preliminary Exposure Pathway Evaluation 100
Figure 5.1 - Data Model for the Spatial Environmental Risk Assessment Methodology..144
Figure 5.2 - Preliminary Exposure Pathway Evaluation 150
Figure 5.3 - Example Exposure Pathways and Points of Demonstration............156
Figure 5.4 - Complete Groundwater Transport Exposure Pathways...................163
Figure 5.5 - Complete Indoor and Outdoor Air Exposure Pathways 164
Figure 5.6 - Exposure Pathway Groups..168
Figure 5.7 - Exposure Pathway Completeness for Grouped Pathways 174
Figure 5.8 - Complete Exposure Pathways Based on Probabilistic Calculations181
Figure 5.9 - Cumulative Distribution Functions for Predicted Indoor Air Concentrations...182
Figure 5.10 - Contributions to the Variance in the Predicted Surface Water Concentrations for Benzene...186
Figure 5.11 - Variance Contributions for the Volatilization to Indoor Air Exposure Pathway 188

Figure 5.12 - Groundwater Concentrations of Benzene and Subsequent NAPL Thickness Data 195
Figure 5.13 - Map of NAPL Detection Data Used for Source Area Identification 202
Figure 5.14 - NAPL Specific Gravity Data Used for Source Area Identification 203
Figure 5.15 - Example Site Conceptual Model Map 205
Figure 5.16 - Uncertainty Associated with Exposure Pathway Completeness Evaluation 208

Nomenclature

Variable	**Description and Units**
A_F	attenuation factor
AF	areal fraction of cracks in the foundation (dimensionless)
A_M	attenuation mechanism
α_x	longitudinal dispersivity (cm)
α_y	lateral dispersivity (cm)
α_z	vertical dispersivity (cm)
b	confidence level expressed as a percent
CA	groundwater concentration, NAPL partitioning (mg/L)
CB	groundwater concentration, groundwater transport (mg/L)
CC	surface water concentration, surface water mixing (mg/L)
Ccrk	background concentration of a chemical of concern in the surface water (mg/L)
CD	soil vapor concentration, NAPL volatilization (mg/m^3)
CE	vapor concentration, indoor air (mg/m^3)
C_E	segment, or exposure pathway ending concentration
$C_{E,overall}$	overall ending concentration for an exposure pathway group
CF	vapor concentration, outdoor air (mg/m^3)
CG	vapor concentration, indoor air, grouped exposure pathway (mg/m^3)
Cgw	groundwater concentration (mg/L)
CH	vapor concentration, outdoor air, grouped exposure pathway (mg/m^3)
c_i	concentration of compound i in a NAPL mixture (g/L)
COCN	number of chemicals of concern
c.o.v.	coefficient of variation
C_S	source concentration
C_{SAT}	saturated vapor or soil concentration
C_{SOL}	aqueous solubility (mg/L)
C_t	concentration at a transition point
Ctarget	target level concentration
$C_{t,combined}$	combined concentration from multiple source areas at one transition point
Cvap	soil gas concentration (mg/m^3)

D(X)	deterministic value of the variable X
d_{10}	grain size diameter where 10% of the grains are finer than this value (mm)
Dair	diffusion coefficient for the chemical in air (cm^2/sec)
δ_{CE}	coefficient of variation
Degrad	first order degradation constant (day^{-1})
Depth	source zone depth (cm)
Dwat	diffusion coefficient for the chemical in water (cm^2/sec)
E(CE)	expected value of the ending concentration
ER	indoor air exchange rate (sec^{-1})
EyGW	additive model error variable, groundwater (mg/L)
EyI	additive model error variable, indoor air (mg/m^3)
EyO	additive model error variable, outdoor air (mg/m^3)
EySW	additive model error variable, surface water (mg/L)
F-Cgw	groundwater concentration flag
F-Cvap	soil gas concentration flag
Fdepth	depth from ground surface to the bottom of the building (cm)
F-NMF	NAPL mole fraction flag
GWGrad	groundwater gradient (cm/cm)
H	groundwater plume mixing depth at the surface water (cm)
H'	Henry's law partitioning coefficient (atm-m^3/mol)
Henrys	Henry's law partitioning coefficient (dimensionless)
h_{POD}	groundwater elevation at the point of demonstration (cm)
h_{source}	groundwater elevation at the source area (cm)
HydCond	saturated zone hydraulic conductivity (cm/sec)
K	hydraulic conductivity (cm/sec)
Koc	organic carbon partitioning coefficient (L/kg)
LB	Indoor enclosed space volume to area ratio (cm)
LN	number of land use scenarios
Ls	depth to subsurface vapor sources indoor air(cm)
Lsout	depth to subsurface vapor source outdoor air(cm)
μ_{CE}	mean ending concentration
MCL	maximum contaminant level (mg/L)
MixHt	height above ground surface for outdoor air exposure (cm)

MN	number of modeling scenarios
MSC	medium specific concentration
MW_i	molecular weight of compound i (g/mol)
NMF	non aqueous phase liquid (NAPL) mole fraction of chemical of concern in the oil-phase (dimensionless)
NyGW	multiplicative model error variable, groundwater (dimensionless)
NyI	multiplicative model error variable, indoor air (dimensionless)
NyO	multiplicative model error variable, outdoor air (dimensionless)
NySW	multiplicative model error variable, surface water (dimensionless)
Owidth	source area width perpendicular to wind direction (cm)
p	probability
ParDen	soil particle density (g/cm^3)
PN	number of exposure pathway records
PODN	number of points of demonstration
PorVad	vadose zone total porosity (dimensionless)
PorWall	porosity of the soils in the foundation cracks (dimensionless)
PSN	number of exposure pathway segment records
PTN	number of exposure pathway types
P_v	vapor pressure (atm)
Qcrk	representative flow in the surface water (L/sec)
Qcrk,in	representative flow in the surface water, before the groundwater mixing box (L/sec)
Qcrk,out	representative flow in the surface water, after the groundwater mixing box (L/sec)
Qgw	representative groundwater flow (L/sec)
q_x	groundwater flux (cm/sec)
R	universal gas constant (8.2×10^{-5} atm-m^3/(mol-K))
R_c	retardation coefficient (dimensionless)
ρ_{ij}	correlation coefficient between variables i and j (dimensionless)
S	molar solubility (mol/m^3)
SatFoc	fraction organic carbon (g oc/g soil)
σ	standard deviation of the groundwater concentration distribution along the boundary with the surface water (cm)
σ_{CE}	standard deviation of the ending concentration
sDist	groundwater travel distance (cm)

SegN	number of segments in an exposure pathway type
SN	number of source areas
Sol	aqueous solubility for a chemical of concern (mg/L)
T	temperature (K)
TPor	saturated zone porosity (dimensionless)
Var(C_E)	variance in the ending concentration
Var(Cor)	variance contribution due to correlation between variables
Var(I)	independent variance due to individual variables
Vcoc	chemical of concern velocity in groundwater(cm/sec)
WallT	foundation thickness for a building (cm)
WCVad	vadose zone water content (dimensionless)
WCVad	vadose zone water content (dimensionless)
WCWall	water content of soils in the cracks (dimensionless)
Width	source area width (cm)
Wspeed	ambient wind speed (cm/sec)
x_{max}	maximum value for the variable x
x_{min}	minimum value for the variable x

Chapter 1 : Introduction

1.1 BACKGROUND

Risk assessment is a process that supports decision making in environmental management through the estimation of risks to human health and ecological receptors associated with exposure to chemicals of concern present in the environment. It is accomplished by collecting information to construct exposure hypotheses and evaluating those hypotheses to determine the potential for human or ecological exposures to chemicals in the environment at levels that may cause adverse effects. The incorporation of risk assessment as an integral component of an overall risk management decision making process provides the scientific basis for:

- choosing one environmental protection strategy over another,
- deciding which sources are contributing the greatest risks and allocating resources accordingly, or
- determining if corrective measures or abatement activities are necessary.

This risk-based decision process provides a mechanism to determine the necessary and cost-efficient strategy for protection of human health and the environment.

1.2 PROBLEM STATEMENT

In recent years the use of risk assessment and the risk-based decision process in environmental management has gained increasing attention. However, there are a number of significant challenges in the practical application of risk-based decision making at large, complex industrial, energy or defense facilities. For these types of facilities, the implementation of a risk-based approach is difficult and a practical methodology has not been demonstrated (Hay Wilson *et al.*, 1998). The most significant challenges arise from the multiple potential sources, multiple chemicals of concern and many potential receptors. In these large facilities there can literally be hundreds of potential exposure pathways to analyze. In practice, at complex facilities, sources are evaluated individually or in small groups. For example, the corrective action program under the Resource Conservation and Recovery Act (RCRA) encourages this piecemeal evaluation through its focus on individual solid waste management units (SWMU). In general, the implementation of program-specific (e.g., air, water, waste) regulations by the Environmental Protection Agency (EPA) has also perpetuated this non-holistic approach. The same is true for the regulation of Department of Defense and Department of Energy facilities. Individual areas of a facility are studied using a straightforward process to analyze exposure pathways for each source-receptor pair. However, there is typically no attempt to understand the interaction of all of the sources and exposure pathways on facility-wide risks or the affects of these multiple sources and exposure pathways on the environmental management decisions. It is only in recent years that industrial and defense facilities have been asked, or have begun, to think about facilities from a risk assessment perspective in a holistic way. Because of the need for these changes, new tools and methods are needed for the implementation of risk assessment.

It is also the case that in the past, for example, the goals for corrective action projects were based on very low to non-detect concentration values, so the need to understand all of the potential exposures was not as great. In addition, often many individuals are involved, over a number of years, at a significant cost, in the calculations of the risks for each of the different areas for a facility. Many of these facilities are also regulated under different regulatory programs, with different regulatory agencies and no one investigator examines all of the results.

Using a risk-based decision process for environmental management requires the understanding of all of the sources and potential exposure mechanisms. The methodology to comprehensively analyze all of the potential current and future exposure pathways and manage all of the resulting information up until now has not been available. The availability of cost-effective computing power and information systems applications can provide the foundation for the development of computer-based systems to manage information and perform calculations for large numbers of exposure pathways. In particular, the information processing capabilities of Geographic Information Systems (GIS), relational databases, and spreadsheets can provide a comprehensive mechanism to construct a methodology to implement risk assessment at large, complex facilities.

1.2.1 Critical Issues

To develop a methodology to practically implement risk assessment, three critical issues must be resolved. First, better methods for developing and representing the *site conceptual model* and integrating the model into the risk evaluation process are needed. Second, methods that provide a greater level of understanding to the engineers and scientists of the uncertainties involved in the risk assessment calculations are needed and in the process provide the link between data and decisions. Third, procedures for maintaining records of the derivation and quality of attribute data are needed.

1.2.1.1 Site Conceptual Model for Site-Wide Risk Assessment

The site conceptual model is a critical component of the risk assessment process that provides the working hypothesis of all of the potential exposure pathways associated with chemicals of concern at the many potential sources, their movement in the environment and their relationship to potential receptors. The site conceptual model is a synthesis of spatial and observational data. Typically, simplified site conceptual models are used to represent the relationships between the sources and receptors at facilities. However, "using a simplified site conceptual model can potentially lead to an inaccurate understanding about the effects on receptors and expected results from implementing a remedial action alternative" (Koerner *et al.*, 1998). This is particularly true when assessing the environmental risk to the many potential receptors existing on and off of a facility involving multiple sources and receptors, as is the case at most large industrial facilities. It is also often the case that the site conceptual model is developed at the beginning of the risk assessment project and then rarely, if ever, used again. It is usually just a static display of the exposure pathways thought to be of importance to the investigation at an arbitrary time in the process. The typical representation is a series of flow charts (ASTM, 1995, BP, 1997). For a facility with multiple sources of many chemicals of concern and various potential receptors, the presentation of flow charts can be overwhelming, and not very informative.

1.2.1.2 Uncertainty Analysis

Methods that provide a greater level of understanding to the engineers and scientists of the uncertainties involved in the risk assessment calculations are needed. While it is true that uncertainty analysis is very important to the risk assessment process, in all but the most sophisticated and costly risk assessments it is conducted at the end of the evaluation and often only in a qualitative way. In general, investigators acknowledge the importance of the uncertainty analysis, but often the uncertainties are not specifically quantified in favor of a qualitative statement that the conservatism in the calculation process counter-balances the uncertainty in the data. This is likely sufficient for screening calculations for simple facilities. However, the incorporation of uncertainty analysis throughout the calculation process becomes very important to the understanding of a complex facility when a facility-wide focus is adopted.

1.2.1.3 Data Management

Procedures for maintaining records of the derivation and quality of attribute data are needed. Understanding the source and confidence in the data that are being used in an evaluation process is important to understanding the confidence of the results. The EPA has for years focused on the quality of laboratory analytical data and has volumes of guidance about data quality for chemical concentration measurements (e.g., EPA, 1994). The quality of other data and information used in the risk assessment process has not typically been identified. In parallel reasoning to the uncertainty analysis, it is important when applying the risk-based process to complex facilities to understand the data quality issues for all of the data that are used in the calculations. It is particularly important when using computer-based and graphical tools to understand the data quality. For example, in GIS data development, the quality of the spatial location measurements and the information about the development of the spatial feature files are recorded in *meta data* files. Meta data is information about data.

1.2.1.4 Solutions Offered by this Research

This research is focused on developing a methodology to practically implement risk assessment, accounting for the many multiple exposure pathways, at large, complex facilities that will support environmental management decisions. The Spatial Environmental Risk Assessment (SERA) developed by this research will provide a process for identifying and quantifying the potential for adverse effects to human or ecological receptors from chemicals or radioactive materials released to the natural environment within a spatially- and temporally-referenced, integrated modeling environment. The methodology will include solutions designed to address the critical issues identified here. The methodology will also provide a system to holistically evaluate risks at large facilities and provide the information needed to compare the contributions to risks from the various sources. This will in-turn provide the information necessary to implement a rational, practical approach for allocating scare resources for risk-based decisions in environmental management.

1.3 RESEARCH OBJECTIVES AND HYPOTHESIS

Risk assessment is a model description of the real world that is dependent on the spatial and temporal relationships between the sources of chemicals of concern and the potential receptors. In the case of a large, complex facility where there are many sources and potential receptors, this is a very difficult task, which in current practice is infrequently accomplished (Koerner *et al.*, 1998). To address this difficulty, the spatial relationships between multiple sources and receptors can be very effectively represented using GIS. Further, since the GIS can provide the functionality to maintain relationships between descriptions of physical features of a study area and their real locations in space, or spatial features, and the information that defines these features, or attribute data, an effective spatial environmental risk assessment methodology can capitalize on the capabilities of GIS. While the Spatial Environmental Risk Assessment (SERA) methodology could be applied to facilities with any number of sources and potential receptors, the significant benefit of the SERA methodology, as compared to the standard risk assessment practice of individual calculations for a subset of exposure pathways, can be realized at large, complex facilities with many exposure pathways.

The contribution of this research is a new spatial environmental risk assessment methodology using GIS, relational databases and spreadsheets. This research is intended to develop better systems for risk assessment at large, complex facilities that account for the complexities of many sources and receptors. This research is also intended to advance the implementation of innovations in site conceptual models, managing and representing uncertainty and meta data management to support facility-wide risk assessment for environmental management at large, complex facilities. This research utilizes the information processing and data management capabilities of GIS, relational databases and spreadsheets to construct a methodology to implement spatial environmental risk assessment at large, complex facilities.

The SERA methodology uses the GIS and the relational database to provide a common framework to spatially and relationally organize the physical and hydrogeological properties of a facility and the surrounding area, the chemical properties of materials released and the temporally-defined attribute data related to concentrations of chemicals of concern in environmental media and hydrogeological properties. The GIS provides the mechanism for the development of a spatial site conceptual model to define potential exposure pathways and the visualization of the outcomes to communicate results. The relational database and spreadsheets provides the mechanism to conduct exposure assessment modeling, to evaluate the fate and transport of chemicals of concern in the environment. The SERA methodology provides a system for the engineers and scientists conducting risk assessments to manage data, perform fate and transport calculations, and visualize the available information and the uncertainties associated with that information. Another benefit is the presentation methods available from the SERA methodology may allow other interested parties to better understand the risk assessment calculations and the results.

The SERA provides an information-processing concept that more clearly ties the data and information for a study area to the risk-based decisions that are made for the environmental management project. In this manner, the process of evaluation and calculation can be more clearly understood and presented. Table 1.1 summarizes the improvements in risk assessment that will result from this research.

Table 1.1 - Improved Methods for Large Facility Risk Assessment

Risk Assessment Component	Current Methods	SERA Methodology
Project Scope	• Individual regulated units analyzed separately • No integration or comparison across all sources and potential receptors	• Capable of analyzing multiple sources and many potential receptors • Provide the information necessary to compare risks for various sources and allocate resources to address those dominating the risks
Site Conceptual Models	• Static Site Conceptual Model • Typically developed on an individual source basis • Not connected to the calculations	• Spatial Site Conceptual Models • Inclusive of multiple sources, connected to the facility features and calculations

Risk Assessment Component	**Current Methods**	**SERA Methodology**
Integrated Environmental Modeling	• Existing environmental modeling not directly linked to site conceptual model evaluation	• Environmental modeling integrated with spatial and tabular databases • Linked to the site conceptual model and exposure pathway evaluation and analyses
Uncertainty Analysis	• Uncertainty analysis conducted at the end of the process • Not generally addressing facility-wide issues	• Data needed for uncertainty analysis tracked throughout the calculation process • Visualization techniques used to enhance understanding
Data Management	• Data quality descriptions given generally in text of report • Quantitative values for data quality generally for laboratory data only	• Explicit data management to track data quality and calculation assumptions through out the risk assessment process • Easily identified.

1.3.1 Spatial Site Conceptual Model

A spatial site conceptual model procedure has been created using the functionality of the GIS and relational databases for identification of potential exposure pathways and tracking of the exposure pathway completeness. The spatial site conceptual model is linked with environmental modeling for the calculations necessary to conduct the exposure pathway analyses.

1.3.2 Integrating Existing Environmental Modeling into the Spatial Environmental Risk Assessment

The design of models of the natural environment is not the focus of this research. Existing models are used to determine the chemical fate and transport in environmental media for select exposure pathways to calculate the concentrations of chemicals of concern at the points of exposure. The models that are used are analytical solutions for the estimation of chemical concentrations in environmental media. These models are appropriate for initial site-specific risk assessments. The results of the fate and transport and exposure calculations are stored in the relational database and are available to be analyzed using the GIS.

1.3.3 Uncertainty Analysis

Data necessary to understand the uncertainties (e.g., uncertainty in input parameter values, concentration data and models to describe environmental processes) in the modeling results are tracked. The uncertainties in the estimated concentrations at points of exposure are estimated. Examples of data visualization options are presented in the results to enhance the understanding of uncertainties related to the exposure assessment calculations. The availability of uncertainty estimates, even based on screening calculations in the initial exposure pathway evaluations, is important to the evaluation of alternatives for additional actions (Hay Wilson, *et al.*,1998).

1.3.4 Data Management

Procedures have been developed for the documentation of the assumptions and variability of the attribute data that are generated throughout the process. The model input parameters are managed in the relational database and are identified based on unique scenario identifiers so that model runs can be replicated at any point in the process and the results of different scenarios can be easily compared.

1.4 ORGANIZATIONAL STATEMENT

Chapter 2 provides background information on various topics that are important in the development of the Spatial Environmental Risk Assessment (SERA) methodology. Chapter 3 presents the development of the data model and the calculation methods. Chapter 4 presents the application procedure for the SERA methodology using specific PC-based software application tools. Chapter 5 presents the results and discussion of the examples developed for the case study site, meta data procedures, an evaluation of the methodology, and application insights for other facilities. Chapter 6 presents the conclusions, limitations and recommendations for further research. In the appendices, supporting and detailed information are provided. Appendix A includes the fate and transport algorithms used in the implementation of the SERA methodology. Appendix B includes the parameter estimation methods and values developed for the case study facility. Appendix C provides the computer procedures to implement the deterministic case. Appendix D provides the computer procedures to implement the probabilistic case. Appendix E presents the data dictionary and electronic file records.

Chapter 2 : Literature Review

This chapter provides background information on various topics that are important in the development of the Spatial Environmental Risk Assessment (SERA) methodology. The sections are intended to be brief overviews of the topics with references to additional reading with both current literature cited and personal, professional experience included.

2.1 RISK ASSESSMENT

This section provides background information on the current state of knowledge in risk assessment based on personal experience and literature review activities.

The role of risk assessment in the protection of the environment is to provide an understanding of the potential dangers posed to humans and ecological receptors by chemicals or radioactive materials in the environment, and provide quantitative measures of those dangers to decision-makers so that protection alternatives may be evaluated. *Risk* is defined as the potential for adverse consequences. In terms of environmental exposures, risk can be quantified using the toxicity, the measure of response to a dose, and the exposure, based on the absorbed, effective dose (NAS, 1983). For low-level exposures to carcinogenic chemicals or radiation, risk is a measure of the increase in the probability of cancer effects over background incidence levels. The humans, plants and animals that might be affected by a particular environmental condition are called *receptors,* or *potential receptors*. *Toxicity*, in evaluation of chemicals, is a measure of specific adverse effects to the receptor at different dose levels. In radiological risk assessment, toxicity is defined by the *dosimetry* for the radionuclides being evaluated and is a measure of specific adverse effects to the receptor at different dose levels (Turner, 1995). The *source* of chemicals or radiation is the physical structure or process equipment from which a release occurs. It is not usually possible to identify all of the specific sources and releases, so the source areas related to the sources are identified through environmental sampling and other methods. The *source area* is the location of highest concentrations in environmental media. *Exposure assessment* quantifies the concentration of chemicals or radionuclides that reach locations where receptors may come in contact with the chemicals or radionuclides. The process of exposure assessment provides predictions of the fate and transport of chemicals or radionuclides in the environment and the identification of the exposure pathways from the source of the chemicals or radionuclides to the receptors. Exposure pathways can be thought of as either *cross media exposure pathways*, where chemicals or radionuclides are transferred from one media to another (e.g., a chemical in soil leaching into the groundwater), or *single media exposure pathways*, where the chemicals or radionuclides are being transported within a single environmental media (e.g., the chemical dissolved in groundwater moves with the groundwater flow). The *fate* of a compound describes the chemical and physical reactions and changes that the compound undergoes in environmental media, including the radioactive decay of radionuclides. It is convenient to use the term *chemicals of concern* to characterize the group of chemical substances and radionuclides to be evaluated for a particular risk assessment.

The simplified process view of risk assessment used here is illustrated in Figure 2.1.

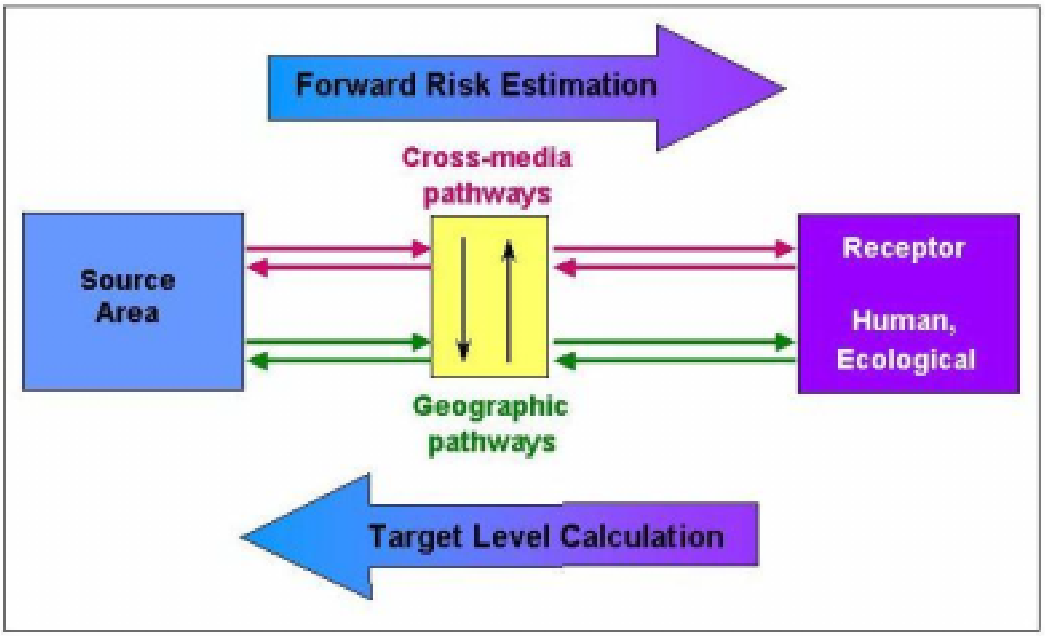


Figure 2.1 - Simplified Process View of Risk Assessment

Figure 2.1 provides a simplified view of the basic elements, or *segments*, of the risk assessment calculation. Each exposure pathway is unique. Some are very simple, composed of only one segment (e.g., direct contact by the receptor with chemicals of concern sorbed on soils). Other exposure pathways are more complex (e.g., benzene as a component of a released petroleum product sorbed on to soils, leached from the soils by infiltration to groundwater and transported with the groundwater flow to a drinking water well). As described in detail by Koerner (1998), it is important to track the individual pathway segments (e.g., benzene transferred from an oil phase to the soil) that combine to constitute a complete exposure pathway in order to account for and analyze the many multiple exposure pathways at a large, complex facility. The segments may occur in series or in parallel. In addition, in the SERA methodology, the physical location of each exposure pathway segment must be identified. For the pathway segments where the start and end of the segment are not physically in the same location, their locations must be identified and the spatial relationship defined by this segment must be incorporated in the calculation process. The process view specifically separates the cross-media and geographic pathway segments, however, the cross-media transfers can occur in parallel with, or at the same time as, the geographic transport or lateral transport (e.g., movement of a chemical of concern with the groundwater flow will at the same time result in sorption of the chemical on the soils through which the groundwater passes). The process view provides the conceptual framework to segment the exposure pathways into analysis elements and to manage the elements in the spatial and tabular databases.

The process of forward risk estimation is discussed in sections 2.1.1 and 2.1.2 and the back calculation of target levels is reviewed in section 2.1.4.

2.1.1 National Academy of Sciences Paradigm

The National Academy of Sciences (NAS, 1983) defines four steps to the risk assessment process: *Hazard Identification, Dose Response Assessment, Exposure Assessment* and *Risk Characterization*.

2.1.1.1 Hazard Identification

Hazard identification is the process of evaluating toxicity data for the materials handled, processed or stored at a facility to determine the potential for illness or injury to humans. An understanding of the exposure scenarios that may result in the effects should be developed (Environ, 1988). The effects are generally classified as *carcinogenic* (cancer-causing), *mutagenic* (causing mutations), *teratogenic* (causing mutations in offspring), and *non-carcinogenic*. These effects are primarily the long-term, or *chronic effects*. It is through this process of hazard identification that the list of chemicals of concern is developed for a facility. The materials review includes those historically used at the property, those associated with current operations and the materials potentially to be used by future activities. The process of historical materials review is complicated, and it may only be possible to rely on the detected concentrations in environmental media to determine what the effects of historical operations have been. Often chemicals of concern are selected to be indicative of other chemicals that are also hazardous.

2.1.1.2 Dose-Response Assessment

Dose-response assessment quantifies the relationship between the amount of a chemical of concern taken in by the body and the resulting degree of illness or injury. The data are often values derived from animal studies and from human epidemiological studies. The process includes a large degree of uncertainty because it involves extrapolating the probability of effects at relatively high exposure concentrations (in the laboratory animal studies) to estimated effects on humans at lower exposure concentrations. Health effects based on lower exposure concentrations are not generally measurable, however the lower concentrations are more often encountered (Till and Meyer, 1990). In order to address the high level of uncertainty that exists in the relationship between environmental dose and human effects, factors of safety are used. The EPA and organizations like the International Commission for Radiological Protection (ICRP) develop the cancer slope factors (for carcinogenic chemicals), reference doses (for non-carcinogenic chemicals), and dose conversion factors (for radionuclides) that are used in relating the environmental concentration to the human doses, with factors of safety included in the values.

2.1.1.3 Exposure Assessment

Exposure assessment quantifies the concentrations of chemicals of concern at locations where the receptors may come into contact with chemicals of concern. The process identifies the elements of each exposure pathway: what is the chemical of concern?, what is the transport mechanism? and who is the receptor? The process of quantifying the exposure includes identifying the concentration in the impacted environmental media, the route of exposure (e.g., ingestion, inhalation, dermal contact), the number of exposure events (i.e., the frequency), and the duration of each exposure event. An individual pathway is made up of a source area of a particular chemical of concern, all of the transfer and transport mechanisms that define the chemical's interaction and movement in environmental media and the receptor, including the route by which a receptor comes in contact with the chemicals of concern. For a large, complex facility with multiple chemicals of concern, multiple sources, and many potential receptors, there can be literally hundreds of exposure pathways to analyze. In addition, the exposure assessment must consider not only the current exposure pathways, but also the potential future exposure pathways as well. The future exposure pathways that are different from the current ones can be, for example, the result of changes in property use on the facility or surrounding properties, changes in the groundwater use in the area, or from delayed releases from a source area.

2.1.1.4 Risk Characterization

Risk characterization is the integration of the analyses of the first three steps (i.e., hazard identification, dose response assessment and exposure assessment) to determine the probabilities that the receptors will experience any of the effects associated with the chemicals of concern under the identified exposure scenarios. Often, the exposure values are derived for typical circumstances and for upper-bound situations so that the values determined in the risk characterization are ranges of risks, or conservative estimates of them (EPA, 1989).

2.1.2 Risk Assessment for Radioactive Materials

In a parallel evaluation, risk can be quantified for radiological assessments by the following relationship:

$$\text{Risk} = f\left(S, T, U, D, R\right)$$ **Equation 2.1**

(Till, 1996).

The variables in Equation 2.1 are defined here. The source term (S) is the identification of the material released to the environment, the manner in which the release occurs (e.g., a continuous release from a process stack or a puff release from an accident) and an associated quantification, if possible, of the volume of material released. This is the parallel to the identification of the chemicals of concern in the hazard identification step and the identification of the magnitude of the sources in the exposure assessment. The environmental transport (T) is the mechanism that moves the material from the source to the receptor. This concept is encompassed in the exposure assessment identified above. The usage factors (U) are all of the parameters that identify the conditions under which an exposure will occur, (e.g., for how long and at what frequency). In the terms of the National Academy of Sciences risk assessment these are known as the exposure factors and they describe the receptors and the activities of the receptor during the exposure. The dose conversion factors (D) are used to translate the exposure to radiation to the actual dose to the affected organs or tissues. The chemical carcinogenic slope factors are a similar quantity and are determined in the dose-response assessment. For radionuclides and chemicals that cause cancer, it is assumed that there is a zero threshold exposure. This means that any exposure adds to the cancer risk. For non-cancer effects there is a lower bounding value, below which no effects are manifested and so no effect is assumed below this value. The conversion to risk (R) is accomplished through risk factors that relate the estimates of cancer to the exposure doses. This is encompassed in the risk characterization.

As Till (1996) presents in his analysis of the risk relationship, uncertainties are involved in each step of the risk assessment process. The description or characterization of risk should include uncertainty analysis. *Monte Carlo analysis* is a fairly complex method, but one that is commonly used in risk assessment to quantify uncertainties. Monte Carlo analysis is the process of incorporating a random sampling of the possible values for the variables and determining the variability of the results. Uncertainty analysis is discussed further in section 2.4.

Till (1996) emphasizes the importance of validating the results of a risk assessment with field data and historical information. Till (1995) also advocates communicating the results of the investigation and actively seeking public participation in the process.

2.1.3 Site Conceptual Models

The *site conceptual model*, which is an "integrated representation of the physical and environmental context, the complete and potentially complete exposure pathways and the likely distribution of chemicals of concerns at a facility, has traditionally been simplified so calculations are straightforward" (Koerner *et al*., 1998). At large, complex facilities there are potentially large numbers of source areas and receptor locations, so the development of a comprehensive site conceptual model that includes all of the potential exposure pathways in a meaningful manner is not often accomplished in current risk assessment practice. However, "using a simplified site conceptual model can potentially lead to an inaccurate understanding about the effects on receptors and expected results from implementing a remedial action alternative" (Koerner *et al.*, 1998). It is often the case that the site conceptual model is developed at the beginning of the risk assessment project and then rarely, if ever, used again. It is often just a static display of the exposure pathways thought to be of importance to the investigation at an arbitrary time in the process. The typical representation is a series of flow charts (ASTM, 1995, BP, 1997). For a facility with multiple sources of many chemicals of concern and various potential receptors the presentation of flow charts can be overwhelming, and not very informative.

If the site conceptual model were truly the working hypothesis of the potential exposure pathways for a facility and were used to communicate among the engineers, decision-makers and interested parties what was being evaluated (and as importantly, what was not being evaluated), it would likely help clarify the process. The site conceptual model could also be a means for different parties to identify the exposure scenarios for which they have the greatest concerns and quickly identify the scenarios that have already been analyzed. The site conceptual model should also be linked to the calculation process that is, it should not just be an abstraction of the project, unrelated to the calculations of risk or target levels. These are goals that have been incorporated in the data model as presented in Chapter 3.

2.1.4 Risk-Based Decisions

Environmental protection decisions are more frequently being based on an evaluation of the risks posed by the industrial activity to be engaged in or the condition of the property being evaluated (Rocco and Hay Wilson, 1998, Washburn and Edelmann, 1998). Risk-based decision making is the process of incorporating the risk assessment tasks and the site evaluation tasks in an iterative calculation process. An understanding of the potential for exposures and the related risks is developed. Here the risks that are being addressed are the risks of adverse human health or ecological effects based on exposure to chemicals of concern in the environment. The understanding of the exposures and risks is used to direct data collection and problem solving. Data are judged to be valuable to the process if they provide information about the potential exposures, and if the decisions about the potential exposures would change as a result of the information. In the process, target levels for concentrations of chemicals of concern in the environment are developed based on "acceptable risk" levels. This process is developed in detail for corrective action projects in the ASTM Standard Guides for Risk-Based Corrective Action (ASTM, 1995a, ASTM, 2000).

The risk-based decision process is an iterative evaluation. Beginning with simple evaluations and progressing to more complex evaluations based on the data and decisions made at each earlier evaluation. The ASTM standard guides are designed using a three-tiered process (ASTM 1995a, ASTM, 2000). The ASTM tier 1 risk evaluation is a *screening level risk assessment* that utilizes source area concentrations compared to statewide standards to evaluate exposure pathways. The ASTM tier 2 risk evaluation is an *initial site-specific risk assessment* that utilizes relatively simple, conservative fate and transport models, non-site specific exposure factors, along with site data and site-specific points of demonstration to evaluate exposure pathways. The ASTM tier 3 risk evaluation is an *advanced site-specific risk assessment* that utilizes sophisticated fate and transport models, and site-specific data, potentially including site-specific exposure factors and toxicology data, to develop target levels.

The risk-based decision process uses the NAS paradigm for relating exposure concentrations to risks, but rather than calculating expected risks, acceptable concentrations at points of exposure are calculated and the relationship to concentrations at the source area is derived from the fate and transport characteristics of the chemical of concern in the environmental media. This is known as the "back calculation" of target levels. Site conditions are compared to target levels and corrective action decisions are made whether remedial action, no additional action or additional risk-based evaluation is appropriate based on the comparison to target levels.

2.2 ENVIRONMENTAL MODELING AND MANAGEMENT USING GIS

2.2.1 Concepts and terminology

Spatial analyses of data are automated using a GIS. Data are defined by spatial objects and tabular information that describes the properties of these objects. The system operates with a one-to-one connection between every spatial object and each record in the tables. The GIS is a database of spatially referenced information, that is, the maps are not stored (ESRI, 1996). Using this data management system, specific data are selected and maps are generated. The user defines the spatial domain of the application, the geographic projection, the map resolution required for the project and the time horizon. In many GIS systems the data are stored in a relational database.

The data structure can be further defined. The data are stored in files called *coverages* or *shapefiles*. A coverage contains only one data type and only one type of spatial feature (e.g., county boundaries, soil sampling locations). A coverage is actually a set of files that define the map features, the coordinates and the spatial relationships (ESRI, 1996). A spatial feature is defined as one of three object types, a *point*, a *line* or a *polygon* (i.e., an area). A line is defined by a series of points. The information stored about a line includes the beginning point on the line, the *from node*, and the ending point of a line, the *to node*. In this way all of the lines that meet at any node are know based on their from nodes or to nodes. A polygon is a series of (x, y) coordinates that are connected. These coordinates define an area. Some GIS systems store points and others store the information about the lines (ESRI, 1996). The information stored also includes topological information that describes the spatial relationships between one feature and another (e.g., what features are contiguous, what features are to the left or right of a given feature).

Coverages of points, lines and polygons are known as *vector* data files. A second important data file structure is *raster* data files. Raster data can either be a picture file, such as a bitmap file that a computer uses to display a photograph, or it can be a *gridded* data file. A gridded data file is a file of square grid cells, all the same size, with values for a single attribute (e.g., water table elevations) (ESRI, 1996). The grids can be employed in mathematical operations. The mathematical expression is applied to each grid cell and a grid file is produced with the operational result stored for each grid cell. This is known as map algebra and the calculation process is integral to the GIS software.

Vector data files may include multiple attributes (e.g., monitoring well data files may include top of casing elevations, water table measurements, thickness of any immiscible phase measurements). Grid data usually includes only one attribute. This would be a specific value for a spatially variable quantity (e.g., ground surface elevations are typically stored as digital elevation model (DEM) grid files).

The raster files and vector files are geographically located using tics, which are points for which the actual locations are known and to which all of the points in the data file are referenced (ESRI, 1996). All of the objects are referenced to longitude and latitude coordinates, not arbitrary x and y locations. The geographic coordinates of the earth's surface can be projected onto a flat map plane using one of many mathematical transformation systems (e.g., Albers Equal Area, Universal Transverse Mercator) (Maidment, 1998).

In this work the ESRI software ArcView 3.1 (ESRI, 1998) is used as the GIS component of the spatial environmental risk assessment methodology because it is available at the University of Texas research computer laboratories, has been used extensively in the research conducted to date and is widely used throughout the world. As discussed in Chapter 3, the methodology is generalized so that other GIS software could be used to implement the same functionality.

2.2.2 Environmental Modeling and Management using GIS

This section provides background information on the current state of knowledge of GIS implemented for environmental modeling and environmental management based on literature review activities and discussions with experts in the GIS field (e.g., Aurora Partnership Meeting). The GIS field has grown rapidly in recent years. There are an abundance of GIS projects for regional environmental management data analysis reported in the literature. These projects typically evaluate spatially varying quantities over watershed or county level scales. The analyses that are performed for these environmental management projects are typically data pattern characterizations. Examples include a project conducted at a county scale for patterns in Lyme disease risk factors (Glass *et al*., 1995). Many applications are described in the literature where the relative locations of properties and the activities on these properties are important. These typically collect geographically referenced features such as census data, locations of EPA permitted facilities, surficial soil types and use buffer distances to determine the susceptibility of properties to impacts by others. One such system is used in a real estate transaction context (Hill, 1996). The system would be particularly useful in real estate transaction audits to identify potential liabilities. Other applications used the geographically referenced features to determine a reasonably optimal location for waste disposal facilities (Hendrix and Buckley, 1992, Lober, 1995, Kao *et al.*, 1996). Another project used a GIS application for hazard assessment in flood prone areas of South Georgia, cataloging hazardous materials handling sites (Harris, 1997). Another system is described for wetlands inventory across the state of Ohio (Yi, 1994). In general, these applications do not include modeling of environmental processes. Other projects have focused on well head protection area mapping for groundwater resource protection (Rafai *et al*., 1993, Baker *et al*., 1993).

Systems that are focused on modeling environmental processes include those developed for surface water hydrology applications, groundwater modeling and surface water modeling. These applications for surface water hydrology, terrain representation and pollutant loading calculation are generally calculations that are completely integrated inside the GIS application and are implemented using both vector and raster data operations (Saunders and Maidment, 1996, Quenzer, 1998, Mizgalewicz and Maidment, 1996, Olivera *et al*., 1996).

There are many examples of connecting GIS software with existing computer codes for groundwater fate and transport modeling (Camp and Brown, 1993, El-Kadi, 1994, McKinney, 1996). The applications use groundwater models such as MODFLOW developed by the US Geological Survey (USGS). In general, all of the linkages include some pre-processing data step in GIS for the development of the model input, running the model outside of the GIS and a post processing of the model output to bring the results back into the GIS. A similar connection between an existing EPA water quality model, WASP5, and ArcView GIS was developed by Benaman *et al*. (1996). The formulation used for the model connection by El-Kadi (1994) for groundwater is conceptually the same as that of Benaman *et al*. (1996) for surface water.

In comparison, a map-based modeling system for regional surface and groundwater flow was developed in ArcView and the object-oriented programming language Avenue, used in ArcView, by Ye *et al*., (1996). Specifically, map-based modeling allows the typical modeling activities to be "activated directly from the model maps" (Ye *et al*., 1996). An interesting feature of the model developed by Ye *et al*., (1996) is the management of spatially referenced time-series data. The data for different times at the same locations are stored in a relational database in the application developed by Ye *et al*., (1996). These are described as one-to-many data types. In their research, simulation results are similarly stored in a tabular database, just as environmental measurement data are stored.

French and Reed (1996) describe a system that incorporates environmental modeling and risk assessment for Natural Resource Damage Assessments as prescribed under the Comprehensive Environmental Response Compensation and Liability Act (CERCLA). This system is focused on short-term impacts to waterways and wetlands habitats. The assessments that are to be implemented with the described system are simplified analyses with "limited field data" (French and Reed, 1996). However, the models that are used in the system are described as multi-dimensional models. It would seem that given limited field data that sufficient information would not be available to run these kinds of sophisticated models. The architecture of the integration of the models with the GIS is not described in the paper. The paper also does not describe the output that are generated or the decision criteria that are used.

A system for estimating risks from numerous facilities across the country is being constructed by EPA and Pacific Northwest National Laboratory (Whelan and Laniak, 1998). The system is being designed to support the EPA rule-making on hazardous waste identification. The system includes environmental transport models for multiple media and many exposure routes. Facilities are necessarily identified as points and not areas because of the scale of the assessment. The use of GIS in the project appears to be as an add-on at the end to locate the facilities and identify their regional context and not as an integral part of the calculation method. They advocate open computer architecture so that new models can replace old models within the system. They also describe an object-oriented window to construct an exposure pathway that links the calculation elements together. The paper does not discuss how multiple exposure pathways are tracked so that a site conceptual model can be developed from the exposure pathway construction window.

As noted by How (1998), most facility level GIS applications are used to provide maps of facility features. These are generally known as automated mapping and facilities management (AM/FM) applications. These are also widely used for municipal utilities management. How (1998) implemented a model integration between the facility level industrial sewer volatile emissions model, naUTilus, and ArcView GIS. The spatial naUTilus model could be implemented as part of the source characterization component for a particular facility in the application of the SERA methodology developed in this research.

An interesting example of simple model incorporation into GIS is provided in a groundwater vulnerability to pesticides application (Sunday, 1996). Screening level models are used for leaching potential, attenuation in the environment and ranking pesticide chemicals based on toxicity, biodegradability and bioaccumulation. The paper uses a case study of an area in southern Iowa to apply the model. Sunday (1996) emphasizes an open modeling architecture rather than "hard-coded" algorithms so that other models can be incorporated in the future. The application uses Arc/INFO and arc macro language (AML) routines to calculate the model output. This is another example where all of the modeling is contained inside of the GIS application. The author also makes a point of mentioning that uncertainty exists when using screening level models and mentions future work on sources of error, but does not discuss any methods employed to analyze the uncertainties.

Another project, focused on a regional scale, is a study of non-point source pollution in the vadose zone (Corwin *et al*., 1998). The authors make the important point about understanding uncertainty when they say that the "sophisticated visualizations created from GIS should never disguise the legitimacy of the rendered results" (Corwin *et al*., 1998). There is a brief discussion in this paper of the sources of uncertainty in environmental modeling but no guidance for handling or visualizing uncertainty is provided.

Additional projects using GIS in environmental management were discussed at the organizing meeting, held September 29-30, 1998, for a group tentatively named the Aurora Partnership. Mark Shaeffer from the Department of the Interior and Jack Dangermond from ESRI generated the idea for the group. The draft stated goal of the group is "to stimulate the development and application of the next generation of information tools and systems for natural resources and environmental decision making" (Aurora Partnership, 1998). The meeting consisted of a series of presentations of work using GIS in environmental management by different individuals from the various participating organizations (e.g., EPA, USGS, The Department of Interior, ESRI).

At the meeting there was extensive discussion of the need for new methods that convey information and engage a broader range of users in environmental management and natural resource protection. One important point that was made is that “technology helps people understand more complex problems than they could before” (Hay Wilson, 1998). This is true not only for lay people but also for engineers and scientists who may be accustomed to looking at a problem in one way based on their analysis methods. New computer-based analyses may help us to identify new insights (Hay Wilson, 1998). One of the presentations at the meeting was of a GIS application developed for a lumber company in California to look at future forest management strategies and incorporated optimization functions for forest utilization and wildlife habitat management (Angelides, 1998).

One project conducted at Oak Ridge National Laboratory focused on human health risks at a large military facility and used GIS to convert the environmental media concentrations to risks (Hargrove *et al.*, 1996). The approach taken in this project is not relevant to the work proposed here. Hargrove *et al*. (1996) did not consider the exposure pathways or the locations of receptors. In fact, they totally eliminated the geographic relationships between sources and receptors, assumed a specific exposure route (i.e., ingestion) and calculated risk-based on human consumption at the groundwater sampling point. This does not seem to clarify the risk, but rather seems to take the reality of the exposure scenario out of the calculations altogether. They include a discussion of interpolating a grid or surface of risk posed by the groundwater concentrations but do not include information about the site geology or groundwater flow direction and discuss kriging the data as if the groundwater flow and transport regime were a statistical function instead of a physical process governed by the advection-dispersion equation.

Another project using GIS in environmental risk assessment is described by Chen *et al*. (1998). They assembled a system that includes modeling of groundwater flow and transport and calculation of risk to receptors. The system connects the models to a GIS. They illustrate the use of the system for a facility that has had a petroleum release. Based on the information presented in the paper, the site conceptual model and exposure pathways, other than groundwater ingestion, are not included in the analyses. Multiple sources and receptors are not discussed.

Three master's theses have been published at The University of Texas at Austin as a result of the risk-based decision support research program. These results are used and cited extensively in this research. The case study facility for this University of Texas research is the former BP Oil Marcus Hook Refinery located along the Delaware River in Marcus Hook, Pennsylvania. An overview of the Marcus Hook Refinery is presented in Appendix B. Romanek *et al*., (1999) developed the *digital facility description* for the case study facility. The digital facility description is the compilation of spatial and tabular data representing the physical features and environmental sampling results at a facility (Romanek *et al*., 1999). Kim *et al*., (2000) developed environmental models and data analyses methods. A procedure to link the GIS data to a numerical groundwater flow model MODFLOW for a portion of the facility known as the Former Lube Plant was also developed. Koerner (1998) developed the first version of a site conceptual model tabular database.

A number of papers were reviewed that discussed the importance of the difference between the GIS data structures and environmental simulation model data structures (Ye *et al*., 1996, Fedra, 1996, Burrough *et al*., 1996, Benaman *et al*., 1996, Maidment, 1996). These data structures impact the level of integration between the simulation model and the GIS. The simple systems are where the model and GIS are separate and data must be selected from the GIS to construct the model input files and run the models and then the results returned to the GIS. A more complex system provides automation for data transfers between the GIS and outside models. Systems with complete integration are those where the calculation models have been completely incorporated inside the GIS. The level of integration is not nearly as important as (i) the consistency of the data structures, (ii) the functionality of the information processing system to share data back and forth with limited manual transfers of the data and (iii) the accuracy with which the natural environment is represented. Computationally, for some environmental processes, outside programs will be more efficient than models run completely in the GIS application language. It may also be a consideration of regulatory acceptability of a particular model that will drive the selection of the particular algorithm or program rather than computational efficiency. This argues for a flexible, open architecture that will accommodate various models and changes to the models in the future.

2.2.3 Evaluation of the GIS Literature Reviewed

Of all the papers reviewed, none discussed a systematic way of incorporating uncertainty or visualizing the uncertainty of the results, although several authors acknowledge the importance of uncertainty analysis. One study of data quality and spatial analysis indicates additional research is needed for uncertainty analysis in environmental modeling in GIS (Aspinall and Pearson, 1996). No specific references to meta data for modeling and risk assessment were included in the papers reviewed. Of the few papers located that discussed risks posed by chemicals in the environment, none deal with the complexities of large facilities with multiple sources, exposure pathways and receptors, and beyond the publications by the risk-based decision support program at the University of Texas at Austin, none of the papers discussed the importance of site conceptual models. None of the papers reviewed discussed developing site conceptual models in a spatial framework.

2.3 RELATIONAL DATABASES

This section briefly reviews background information on relational databases.

Relational databases are also referred to as *tabular databases*. A relational database is composed of a group of tables that are used to manage data for a specific subject. The data are organized into records. Each row in a table is a unique record (e.g., one groundwater monitoring well's information in an environmental measurements database). The columns in a database table are different fields of information (e.g., monitoring well name, x- and y-coordinates, top of casing elevation). Each record contains values in each of the fields. Each of the tables in the database is associated, or *related*, to other tables in the database by common fields.

Each record is unique because each table is designed with one or more *key fields*. The values in the group of key fields for every record in the table must be unique. Therefore, the number of key fields for each table will be determined by how many attributes are needed to uniquely describe each record.

Relationships in a database are evaluated between two tables and are described as *one-to-one*, *one-to-many* and *many-to-many*. The most common relationship included in a database is the one-to-many. It is the relationship between a central table and a detail or supplemental information table. A one-to-one relationship is where each record in one table is related to one record in a second table. This can be used if there are too many fields to be contained in one table. The many-to-many relationship is represented in a database as a number of one-to-many relationships. In this situation, there are multiple rows in each of the two tables that are related to one another (Litwin, *et al*., 1997).

The design of a database takes into account all of the potential information to be stored, the logical groupings for the data and the objectives to be met with the database. Based on this information, the *database schema* can be created. The database schema defines the tables and the fields within each table to be included in the database (Litwin *et al*., 1997). The concept is to store the data efficiently. Data should not be repeated from one table to another in the database. In order to accommodate multiple relationships several tables should be constructed rather than storing lists of information in one field.

Queries are used to collect data from several tables at one time into a new table, called a *query table*. The query tables are live links to the database tables. This means that when data are updated or added to the database the next time the query is executed the new data will appear in the query table. Queries can be used to generate reports or results tables. In most database software, standard mathematical and statistical functions can be performed. However, scientific computing cannot easily be performed in a database field.

2.4 UNCERTAINTY IN PATHWAY COMPLETENESS EVALUATION

This section discusses the methods for estimating uncertainties in modeling calculations for pathway completeness.

In general, in initial site-specific risk assessments the uncertainties associated with modeled or estimated concentrations of chemicals of concern at points of exposure are accounted for by using sufficiently conservative input parameters and representative values (e.g., maximum concentrations, 95-percent upper confidence limits on the mean) that the predicted concentration is expected to be higher than the actual concentrations. In addition, as discussed in section 2.2, in the GIS modeling projects that were reviewed, uncertainty in model output was identified as an important feature. However, of the projects reviewed there were no quantitative methods discussed.

A method for estimating the amount of uncertainty in the predicted concentrations at the points of exposure, or the ending concentrations was sought. The uncertainty analysis necessary is also called forward modeling or error propagation (Gilbert, 1997). Given the distributions for the input variables, what is needed is the distribution of the output variable.

Several methods were reviewed in order to identify an appropriate methodology to implement the uncertainty analysis. Since the fate and transport models being used are spreadsheet-based analytical models, an uncertainty quantification method that can be implemented analytically using spreadsheets was sought. As discussed in section 2.1.2, one of the more common, but fairly complex methods used in probabilistic risk assessments to quantify uncertainties is Monte Carlo analysis. However, Monte Carlo analysis has not had widespread acceptance by the environmental regulatory agencies or responsible parties for risk assessments at facilities with chemical releases.

Prior to implementing a Monte Carlo analysis, or another complex, uncertainty analysis method, it is advisable to implement a simple analytical variance estimation method (Gilbert, 1997). Therefore, using a first order Taylor

Series Approximation for each of the fate and transport models was identified as an appropriate method to estimate the variance in the predicted ending concentrations (Ang and Tang, 1975). The method is exact for linear algorithms, but may also be used as an estimate for non-linear algorithms. A second method that was identified was using commercially available Monte Carlo simulation software that is available as an add-on to the common spreadsheet programs, to perform the uncertainty calculations. This option seemed to be more complex, had the potential difficulty with acceptance, and still required the first order analytical estimate to be calculated. A third option identified was to conduct sensitivity analyses. This is where each input variable is adjusted through its range of possible values and the resulting changes in the ending concentration values are tabulated. This process has the benefit of being simple to implement, however, the overall variance in the ending concentration values are not easily determined. Based on the options reviewed, the first order linear analytical tools were selected because they:

- are exact for linear fate and transport equation,
- may be a reasonable estimate for non-linear equations,
- are a good starting point when using more sophisticated uncertainty evaluations, and
- can be implemented in spreadsheets.

2.5 SUMMARY

This chapter presents the background information on risk assessment and GIS in environmental management. A brief overview of the basic principles of chemical and radiological risk assessment is presented. The current literature on the use of GIS in environmental management with a particular focus on facility-scale environmental modeling is included. Summary information on data management and uncertainty in exposure pathway completeness evaluation is discussed. Chapter 3 presents the methodology for the spatial environmental risk assessment.

Chapter 3 : Methodology

This chapter presents the development of the Spatial Environmental Risk Assessment (SERA) methodology. A brief overview of the methodology and the application procedure is presented first to provide a road map of the overall process. Then, the data model and the calculation methods are presented. The application procedure is discussed in Chapter 4. Detailed algorithms and computer procedures are included in the appendices.

3.1 OVERVIEW OF THE SPATIAL ENVIRONMENTAL RISK ASSESSMENT METHODOLOGY

Figure 3.1 is a flow diagram that illustrates the Spatial Environmental Risk Assessment (SERA) methodology.

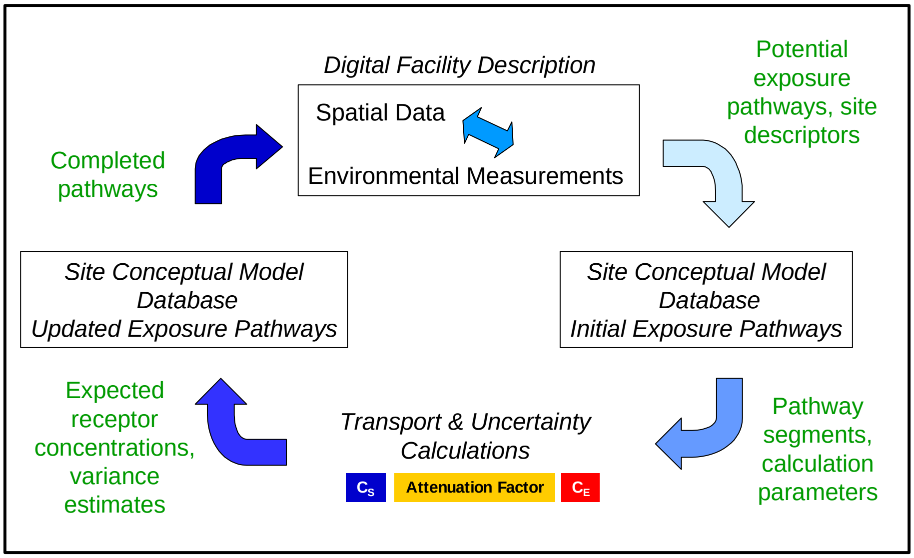


Figure 3.1 - Overview of the Spatial Environmental Risk Assessment (SERA) Methodology

The SERA methodology is an iterative process through which the risk evaluation is refined as additional data are gathered for a facility. The foundation is the *digital facility description*, which is the compilation of spatial and tabular data representing the physical features and environmental sampling results at the facility (Romanek *et al*., 1999). These data are focused on characterizing the potential for risks to human health and the environment under current and future land use scenarios. The *spatial database* is the organized collection of geographic information system (GIS) files that describe the physical features at the facility and the surrounding regional characteristics. The *environmental measurements database* contains the information about environmental sampling and analysis for chemicals of concern in environmental media, site stratigraphic, hydrogeologic and hydrologic information and other relevant data for the characterization of the facility. See Romanek *et al.,* (1999) for a complete discussion of the development of a digital facility description for the former BP Oil Marcus Hook refinery.

Based on the digital facility description the *site conceptual model* is defined. The site conceptual model is an "integrated representation of the physical and environmental context, the complete and potentially complete exposure pathways and the likely distribution of chemicals of concerns at a facility" (Koerner *et al*., 1998). The potential *exposure pathways* and *source area* concentrations are identified in order to build the site conceptual model. An individual exposure pathway is made-up of three major elements: (1) a source of a particular chemical of concern, (2) all of the transfer and transport mechanisms that define the chemical's interaction and movement in environmental media, and

(3) the receptor, including the route by which a receptor comes in contact with the chemicals of concern. The *exposure pathway elements* are the source area, the transfer and transport segments, and the receptor locations. As described by Koerner (1998) the transfer and transport segments represent the fate and transport of chemicals of concern and are connected at *transition points*. The receptor locations are called the *points of demonstration*. The exposure pathway elements are defined in GIS files and the exposure pathways are attributed with the important physical characteristics. The site conceptual model tabular database is populated with the records describing all of the elements of the identified exposure pathways, including the specific fate and transport exposure pathway segments and the required input parameters for the fate and transport models.

The fate and transport models are implemented using algorithms to predict the potential exposure concentrations that can be compared to the *target levels*. The target levels are the protective exposure concentrations or regulatory standards. The results can be displayed using a combination of maps, tables and graphs, comparing the predicted concentrations with the target levels and the displaying the uncertainty in the results.

As additional data are gathered and additional analyses are conducted at the facility, the digital facility description and the site conceptual model are updated and the modeling runs re-executed. The results of these new model simulations can easily be compared to the prior results. Increasingly sophisticated and data-intensive models can also be implemented and the input and results documented in the site conceptual model database.

The next section describes the data model for the spatial environmental risk assessment. The data model is based on a detailed analysis of the information requirements and calculation procedures for implementing risk assessment.

3.2 DATA MODEL

This section describes the data model for the spatial environmental risk assessment (SERA). The data model is based on a detailed analysis of the information requirements and calculation procedures for implementing risk assessment. The data model is the conceptual framework on which all of the details are developed. In general, the objective of a data model is to describe all of the individual steps and elements of a highly complex evaluation process in mathematical and information processing terms. The data model is used as the template for application of the SERA methodology. The data model is also developed to be software and platform independent. Chapter 4 discusses the application of the SERA methodology using PC-based, Microsoft® and ESRI® software applications. There are a number of alternative computer software tools (e.g., Oracle database software) and hardware platforms (e.g., UNIX) that could be used to implement the data model. This section includes discussions of the exposure pathways, the site conceptual model, and a brief discussion of temporal considerations.

3.2.1 Exposure Pathways

As discussed in Chapter 2, an important aspect in the implementation of a risk assessment is defining all of the potential exposures, or exposure pathways, for human and ecological receptors. It is important to recognize that the exposure pathways must be identified individually for each chemical of concern. In the SERA methodology, exposure pathways are defined by the elements that compose an exposure pathway. Figure 3.2 presents these conceptual *exposure pathway elements*.

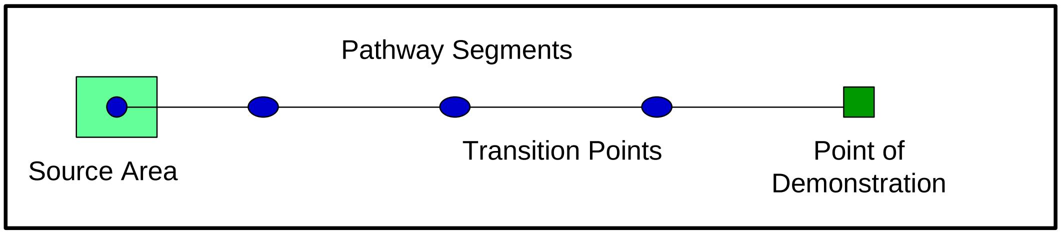


Figure 3.2 - Exposure Pathway Elements (After Koerner *et al*., 1998)

3.2.1.1 Source Area

The start of each exposure pathway is described by a *source area*, or a *point source*. A source area is defined by a physical area where the highest concentrations of chemicals of concern have been detected in an environmental medium (e.g., concentrations of benzene in soils). For the transport calculations, the location of the source area is represented by a point located at the centroid of the area. A point source is an emissions stack, or the like. It is characterized by its point location.

There are several ways to characterize the strength of the source area (e.g., total mass, a constant concentration, a time varying concentration). The model or algorithm selected for the transport calculation will dictate the mathematical form of this characteristic of the source area. As discussed in section 3.3, in the SERA methodology, long-term, steady-state calculations are made to predict the potential exposure point concentrations. This is appropriate because the risks to be defined are those for chronic exposures, exposures of 30 or more years' duration. Therefore, the source is described by a constant concentration that is representative of the source concentration, and conservative, over the entire exposure duration. If transient models were to be used in the transport calculations, then the sources would be described as concentrations or mass inputs as functions of time and an accounting of the time pattern of the exposures, would have to be made.

3.2.1.2 Exposure Pathway Segments

From each source area or point source, each exposure pathway is divided into *exposure pathway segments.* The exposure pathway segments are identified for each of the environmental media through which transport of chemicals of concern away from the source can occur. The exposure pathway segments can represent movement of chemicals of concern as:

- geographic translation in space (e.g., migration with groundwater flow),
- transfer of chemicals of concern from one phase or environmental medium to another at the same geographic location (e.g., partitioning of chemical from a non-aqueous liquid into soil vapor), or
- vertical migration of chemicals of concern through one environmental medium at the same geographic location (e.g., migration of chemicals with soil pore water).

As a result, the exposure pathway segments represent chemical migration functions in all three coordinate directions (i.e., x, y, z). An example exposure pathway segment is the partitioning of a chemical of concern from the sorbed phase on soil in the vadose zone into the soil pore water. The next segment in the exposure pathway is movement of the chemical in the soil pore water through the soil to the groundwater table. The exposure pathway is described as soil leaching to groundwater and consists of two exposure pathway segments. There is no limit to the number of segments in an exposure pathway. For each exposure pathway segment a corresponding transfer or transport calculation algorithm is defined.

3.2.1.3 Transition Points

The exposure pathway segments are connected at *transition points*. A transition point is a location where:

- the chemical of concern is moved from one environmental medium to another,
- there is a change in the direction of movement, within a single environmental medium, or
- there is a change the properties that define the movement, within a single environmental medium.

There are two transition points for the leaching example given in section 3.2.1.2, (1) the point at which the sorbed chemical concentration becomes the starting soil pore water concentration and (2) the point at which the soil pore water concentration becomes the starting groundwater concentration. To look at the situation where a transition point is defined by a change in the environmental properties that define the chemical movement, more detailed knowledge of the vadose zone soils is assumed. For example, if the vadose zone soil were characterized by layering of sand over silt, the movement of soil pore water might be modeled as two exposure pathway segments, one through the sand and one through the silt. This is because each of these soil types exhibits different properties that determine the rate of soil pore water movement. In this case, there would be a transition point at the interface between the sand and the silt layers.

The transition points are used to define and to track the start and end of each transfer or transport calculation, except where the exposure pathway segment starts at the source area or ends at the point of demonstration.

3.2.1.4 Points of Demonstration

The end of an exposure pathway occurs at a *point of demonstration*. A point of demonstration is a location where the potential exposure concentration must be known and compared to a target level. In the SERA methodology, a point of demonstration can be:

- a regulatory point of compliance;
- an actual or assumed point of exposure;
- a potential point of exposure;
- a location at which control over land use is no longer certain (e.g., the property boundary); or,
- another location defined arbitrarily.

If an area is to be evaluated, and as a result, a representative potential exposure concentration is needed for an area rather than for one point, several alternatives for the point of demonstration are available. The point of demonstration can be defined as:

- the centroid of the potential exposure area;

- the point closest to the source area given the transport pathway; or,
- a point at the downstream extent of the source area contributions (e.g., a location in a surface water downstream of the discharge of several groundwater sources).

As illustrated in Figure 3.2, each exposure pathway consists of a source area, or point source, one or more exposure pathway segments, the associated transition points and the point of demonstration.

3.2.2 Site Conceptual Model

The site conceptual model for a facility is constructed by identifying all of the source areas and point sources, all of the chemicals of concern, all of the exposure pathways and all of the points of demonstration. For a large facility this may mean defining tens or hundreds of source areas. For each of the source areas there may be many potential exposure pathways and many chemicals of concern. The potential for exposure pathways from different source areas to interact, or to impact a single point of demonstration is high. As discussed in Chapter 1, evaluating and managing the information for many multiple exposure pathways and their interactions has been a complicating factor in current risk assessment practice, and was an important design consideration in the development of the SERA methodology.

For each source element, the chemicals of concern that are associated with the source are identified. Potential exposure pathways and receptor locations are identified and developed for both the current land use case (i.e., the description of the types of potential receptors and their activities) and the future land use case.

The future land use description can be developed based on known plans for the facility, the local planning and zoning designations for the facility, or an evaluation of the potential future land use of the facility and the surrounding area. Identification of the future land use case can be a difficult process, but in order to properly characterize the potential exposure pathways, a potential future land use must be defined.

An important part of the site conceptual model development is defining the spatial features for the site conceptual model. Source areas, exposure pathway segments, transition points and points of demonstration all have spatial characteristics and are defined as spatial database elements. In addition, the site conceptual model elements have many attributes that describe them and define their function from a fate and transport modeling perspective. Within the SERA methodology, these attributes are gathered and managed in the tabular database.

The site conceptual model is constructed segment-by-segment by identifying all of the segments for each exposure pathway, for each chemical of concern, for each source area. As described by Koerner (1998), this is important in order to account for and analyze all of the exposure pathways at a large facility. The segmentation of the exposure pathways therefore is the key to the calculation of ending concentrations at points of demonstration that have multiple source areas contributing to the concentrations. Figure 3.3 illustrates a scenario where two source areas are contributing to the concentrations at one point of demonstration.

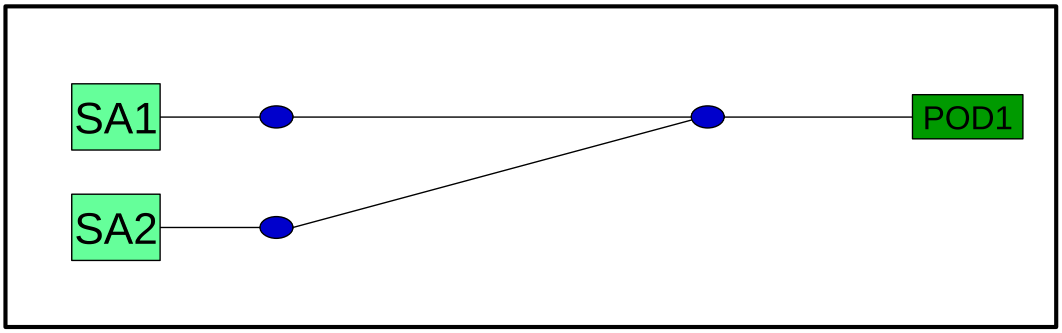


Figure 3.3 - Exposure Pathway Segments and Multiple Contributing Source Areas (After Koerner, *et al*., 1998)

While, Figure 3.3 illustrates two source areas, in reality any number of source areas might act together requiring that many source areas be identified for any one point of demonstration.

The site conceptual model is used to identify all of the exposure pathway segment calculations, the associated input parameter data for each exposure pathway, and to identify all of the source areas that contribute to concentrations at each point of demonstration. The transport calculation results are managed as part of the site conceptual model.

3.2.3 Temporal Considerations

As described previously, the transport calculation models may be identified as steady-state models or as transient models. In addition, some of the exposure pathways identified for evaluation are specifically related to a potential future land use scenario, or may depend not on a land use change but on operational changes at the facility (e.g., exposures to volatile chemical emissions in outdoor air from impacted soils, which occur once a process unit is removed).

These temporal factors are incorporated into the data model in the following ways.

Each exposure pathway is identified as being either a *current* or *potential future exposure pathway*. In this way an evaluation is made separately for the current exposure pathways and for the future exposure pathways. It is possible that at a facility there may be more than two discrete points in time that govern the formulation of exposure pathways. This can be handled by assigning multiple characterizations, rather than just the current or future designations. In this way the accounting for multiple effects can be managed (i.e., when multiple source areas affect one point of demonstration, the exposure pathways that are relevant for a specific time period or receptor activity can be accumulated separately from exposure pathways that are relevant for other time periods or activities). In the examples presented in this research there are exposure pathways defined for the current exposure scenario and there are exposure pathways that are defined for a future exposure scenario.

For the transport modeling, transient modeling may be implemented. These are not specifically included in the application procedures or in the examples presented in this research. However, the framework for gathering the input information, associating it with the exposure pathways through the site conceptual model and evaluating the results, again using the site conceptual model (see section 3.4), is also applicable for transient transport modeling. For any facility, it is recommended that the steady-state evaluations be implemented first and evaluated before more complex and data-intensive transient evaluations are implemented.

3.3 CALCULATION METHODS

This section describes the calculation methods used for the spatial environmental risk assessment. Table 3.1 summarizes the fate and transport equations used in the examples in this research. The equations, the listing of input parameters, and the model assumptions are presented in Appendix A.

Table 3.1 - Fate and Transport Equations Used in the Example Calculations in this Research

<table>
<tr><th>Exposure Pathway Segment</th><th>Equation</th></tr>
<tr><td>NAPL partitioning to groundwater</td><td>$$\mathrm{CA}\left(\frac{\mathrm{mg}}{\mathrm{L}}\right) = \mathrm{NMF} \times \mathrm{Sol}$$</td></tr>
<tr><td colspan="2">Equilibrium partitioning using Raoult's Law (Charbeneau, 2000)</td></tr>
<tr><td>NAPL partitioning to soil vapor</td><td>$$\mathrm{CD}\ (\frac{\mathrm{mg}}{\mathrm{m}^3}) = \mathrm{NMF} \times \mathrm{Henrys} \times \mathrm{Sol} \times \frac{1000\mathrm{L}}{1\mathrm{m}^3}$$</td></tr>
<tr><td colspan="2">Equilibrium partitioning using Raoult's Law and Ideal Gas Law (Charbeneau, 2000)</td></tr>
<tr><td>Groundwater transport</td><td>$$\mathrm{CB}\left(\frac{\mathrm{mg}}{\mathrm{L}}\right) = \mathrm{CA} \times \exp\left[\frac{\mathrm{sDist}}{2 \times \alpha_x}\left(1 - \sqrt{1 + 4\frac{\mathrm{Degrad} \times \alpha_x}{V_{\mathrm{COC}}}}\right)\right] \times \mathrm{erf}\left(\frac{\mathrm{Width}}{4\sqrt{\alpha_y \times \mathrm{sDist}}}\right)\mathrm{erf}\left(\frac{\mathrm{Depth}}{4\sqrt{\alpha_z \times \mathrm{sDist}}}\right)$$</td></tr>
<tr><td colspan="2">Domenico solution to the advection dispersion equation (Domenico, 1987)</td></tr>
<tr><td>Groundwater discharging to surface water and mixing</td><td>$$\mathrm{CC}\left(\frac{\mathrm{mg}}{\mathrm{L}}\right) = \frac{\left(\begin{array}{l} C_{\mathrm{crk}} \times Q_{\mathrm{crk}} + \\ 6 \times 10^{-4} \times \left(\begin{array}{l} \mathrm{HydCond} \times \mathrm{GWGrad} \times \\ \mathrm{H} \times \mathrm{CB} \times \mathrm{Width} \end{array}\right) \end{array}\right)}{Q_{\mathrm{crk}}}$$</td></tr>
<tr><td colspan="2">Simple mass balance model (Charbeneau, 2000)</td></tr>
</table>

Soil vapor transport to outdoor air	$\mathrm{CF}\left(\frac{\mathrm{mg}}{\mathrm{m}^3}\right)=\frac{\mathrm{CD}}{\left(1+\frac{\mathrm{WSpeed}\times\mathrm{MixHt}\times\mathrm{Lsout}}{\mathrm{OWidth}\times\mathrm{Deff}}\right)}$
Diffusion through the vadose zone and an atmospheric box model (Johnson and Ettinger, 1991, ASTM, 1995a)	
Soil vapor transport to indoor air	$\mathrm{CE}=\frac{\mathrm{CD}\times\mathrm{TE1}}{(1+\mathrm{TE1}+\mathrm{TE2})}$
Johnson and Ettinger model (Johnson and Ettinger, 1991, ASTM 1995a)	

The models presented in Table 3.1 are examples of the types of models and the complexity of the models envisioned to be used in the initial evaluations within the SERA methodology. Within the SERA methodology, alternative fate and transport algorithms can be used. The models used in this research were chosen because they are widely accepted and familiar analytical solutions. However, there are many different ways to represent the physical, chemical and hydrogeological processes that are represented by the exposure pathway segments. In addition, other exposure pathways would likely be evaluated in a full application of the SERA methodology to a large facility, therefore additional algorithms beyond those presented in Appendix A would need to be identified.

The important feature of the fate and transport calculations is that they are used to develop a relationship between the starting concentration for an exposure pathway segment and the ending concentration for that exposure pathway segment. They are derived in an *attenuation factor* format. The attenuation factor (A_F) which is a value less than or equal to 1, describes the relationship between the starting concentration (C_S) of a segment and the ending concentration

(C_E) of a segment. Equation 3.1 is the basis for all of the pathway segment calculations.

$$C_E = C_S A_F$$ **Equation 3.1**

Figure 3.4 illustrates this relationship and the parallel relationship for the overall exposure pathway. The attenuation factors are also described in the literature as *transfer factors, natural attenuation factors* or as *dilution attenuation factors* (Bedient, *et al*., 1999, Charbeneau, 2000).

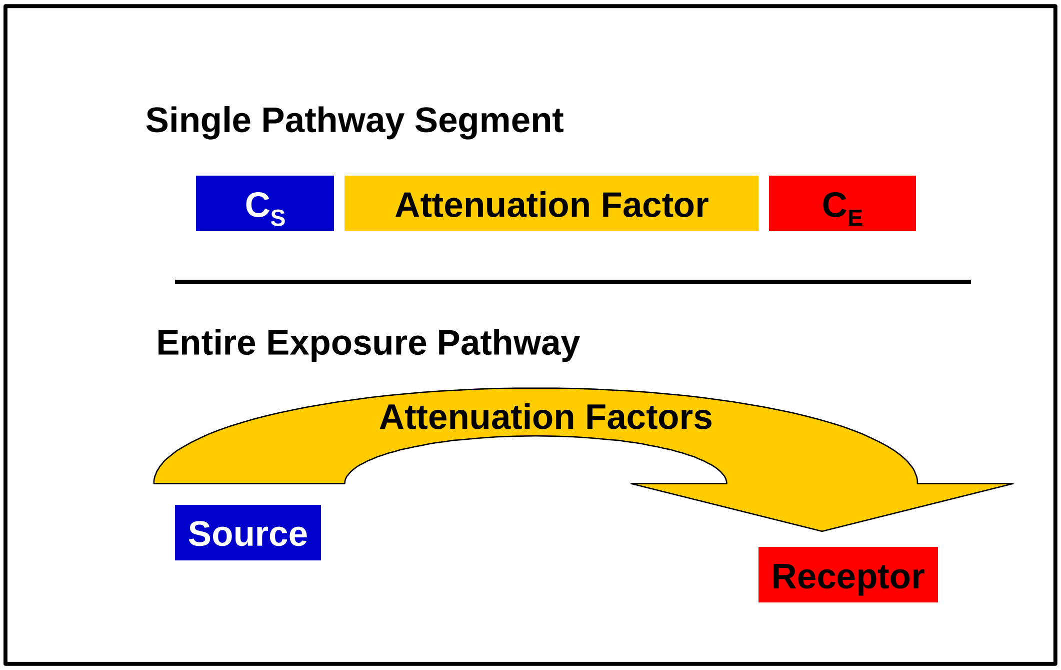


Figure 3.4 - Single Exposure Pathway Segments and Exposure Pathways Described by the Attenuation Factors

The attenuation factors are functions of a number of variables as noted in Equation 3.2.

$$A_F = f\begin{pmatrix}\text{site properties, chemical properties} \\ \text{transport or transfer relationships}\end{pmatrix}$$ **Equation 3.2**

So the attenuation factor can be written as a function of the individual input parameters:

$$A_F = f\left(X_j\right)$$ **Equation 3.3**

where
X_j - individual input parameters.

The ending concentration from one segment is the beginning concentration for the next segment. When the segment ends at a point of demonstration, the ending concentration is the potential exposure concentration at the point of demonstration that is to be compared to the target level.

For the case where there is one source area and one point of demonstration, suppose there are N segments in the exposure pathway and the attenuation factors for each segment are given by $A_F(i)$ where i = 1,2, …, N. Then, if the source concentration is C_S, the ending concentration for segment 1, at the transition point, is defined as:

$$C_t(1) = C_S A_F(1)$$ **Equation 3.4**

where
$C_t(1)$ - concentration at the first transition point

for segment 2:

$$
\begin{aligned}
C_t(2) &= C_t(1)A_F(2) \\
&= C_S A_F(1) A_F(2)
\end{aligned}
$$ **Equation 3.5**

and by extension, for segment N, the ending concentration is:

$$C_E(N) = C_t(N) = C_S \prod_{i=1}^{N} A_F(i)$$ **Equation 3.6**

Therefore, the quantity $\left[\prod_{i=1}^{N} A_F(i)\right]$ is the *pathway attenuation factor* for one source area and one point of demonstration.

The relationship becomes more complicated when there are multiple source areas with multiple exposure pathway segments impacting one transition point or one point of demonstration, as illustrated in Figure 3.3. In this case, the overall ending concentration ($C_{E,overall}$) is a function of all of the contributing pathway segments before the point of demonstration. For the individual source areas, the pathway segments that are independent are calculated up to the transition point where the exposure pathways are combined, using Equation 3.6. At the transition point, where the exposure pathways are combined, the concentrations are summed and compared to a limiting concentration, aqueous solubility (C_{SOL}), vapor saturation (C_{SAT}), or soil saturation ($C_{SAT,SOIL}$), as applicable, to ensure that the combined concentration ($C_{t,combined}$) is physically possible, or if the concentration value at that point is limited by aqueous solubility or saturation.

As an example, for the vapor concentration case, for P source areas:

$$\text{If } \sum_{i=1}^{P} C_t(i) \le C_{SAT} \text{ THEN } C_{t,\text{combined}} = \sum_{k=1}^{P} C_t(i)$$

$$\text{ELSE } C_{t,\text{combined}} = \text{NMF} \times \text{Henrys} \times \text{Sol} \times \frac{1000\text{L}}{1\text{m}^3}$$

Equation 3.7

where
$C_{t,combined}$ - concentration for combined effects of multiple sources at one transition point (mg/m^3)
Henrys - Henry's Law coefficient (dimensionless)
NMF - non-aqueous phase liquid (NAPL) mole fraction (dimensionless)
Sol - aqueous solubility (mg/L).

The aqueous phase and soil cases are similarly stated.

Then the ending concentration for the combined pathway segments is calculated using Equation 3.6 and the $C_{t,combined}$.

For multiple source areas impacting one point of demonstration, the overall pathway attenuation factor is not separable from the starting concentrations, since there are several contributing starting concentrations.

Two cases are important in the estimation of environmental risks, the *deterministic case* and the *probabilistic case*. The deterministic case is used as a screening step to identify whether exposure pathways are potentially complete and will require more detailed study. In the deterministic case, point values are selected for the model input parameters that are "conservative." This means that within the range of likely values for the facility for the parameter, a value is chosen that maximizes the ending concentration. In the probabilistic case, expected values are chosen for each of the input parameters. In addition, estimates of the variance in the inputs are made, so that an estimate of the variance in the ending concentration can be determined. The estimates of the variance in the input parameters are chosen to account for spatial variability in the parameter, measurement uncertainty and other sources of variability in the input values. The probabilistic case requires more information about the facility than the deterministic case in order to develop the variance values for all of the input parameters. As a result, typically the deterministic case will be executed first and then a data-gathering and analysis task will be implemented to develop the values for the probabilistic case. Following which the probabilistic case would be implemented.

3.3.1 Deterministic Case

With one source area and one point of demonstration, the ending concentration (C_E) in the deterministic case is:

$$C_E = C_S \prod_{i=1}^{N} A_{F,i}\left(D\left(X_j\right)\right) \qquad \textbf{Equation 3.8}$$

where
$D(X_j)$ - conservative input value for each parameter $X_{j.}$

In the deterministic case the exposure pathway segments are calculated individually. This means that any number of exposure pathway segments can be combined at any transition point or point of demonstration. The limitation on combinations of exposure pathway segments is, of course, the physical reality of the chemical movement in environmental media. For example, a soil vapor exposure pathway segment from one source area cannot be combined at a transition point with a groundwater concentration exposure pathway segment from another source area.

Each exposure pathway segment is given an attenuation mechanism (A_M) identifier in the site conceptual model database that corresponds to a particular attenuation factor calculation algorithm. In this way, each fate and transport segment is independent of the overall exposure pathway. All of the groundwater transport segments for example can be calculated together, without regard for where they exist on the facility. The independent calculation allows the accumulation of effects from multiple sources on individual points of demonstration. In addition, each exposure pathway segment is related to specific model input parameters, so each of the segments is calculated using the values appropriate to its location on the facility. The group of model input parameters for each calculation is identified as a *model scenario*. A model scenario can apply to one or more exposure pathways.

3.3.2 Probabilistic Case

In the probabilistic case the source areas and the associated exposure pathways that are to be considered together are identified as a group. This is because the variance calculation depends on the particular exposure pathway segments and variables that precede the calculation of the final ending concentration. Unlike the deterministic case, the combination of exposure pathway segments and multiple source areas in the probabilistic case is based on a fixed structure so that the variance calculation can be implemented.

All exposure pathway groups in the probabilistic case are combined at the common point of demonstration. Stated another way, in the probabilistic case the exposure pathways from multiple source areas can only be combined at the point of demonstration. This corresponds to a point of demonstration for an area, as described in section 3.2.1.4. The calculation structure selected for the probabilistic case includes independent exposure pathways until the mixing attenuation factor, or environmental medium *mixing box,* at the end of the exposure pathway. The attenuation factor models in Table 3.1 include the surface water mixing attenuation mechanism and the two vapor exposure pathway models themselves incorporate mixing in the indoor environment or in the air breathing zone box. Further discussion of the attenuation factor models is presented in Appendix A. The surface water mixing attenuation factor is available to be used in the deterministic case; the difference in the probabilistic case is that it is integral to the calculation of the ending concentration and the variance in the ending concentration for the groundwater to surface water probabilistic case. Figure 3.5 illustrates the mixing box concept and the flexibility in point of demonstration location for the deterministic case.

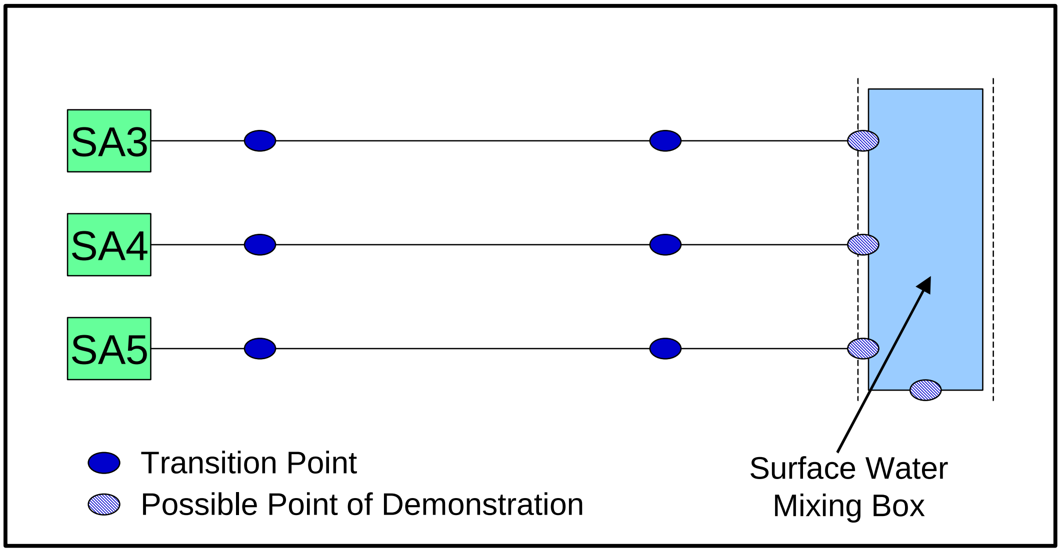


Figure 3.5 - Calculation of Groundwater Transport to Surface Water in the Deterministic Case

Figure 3.6 illustrates the same exposure pathways as Figure 3.5, (i.e., groundwater transport of chemicals of concern to surface water), but for the probabilistic case.

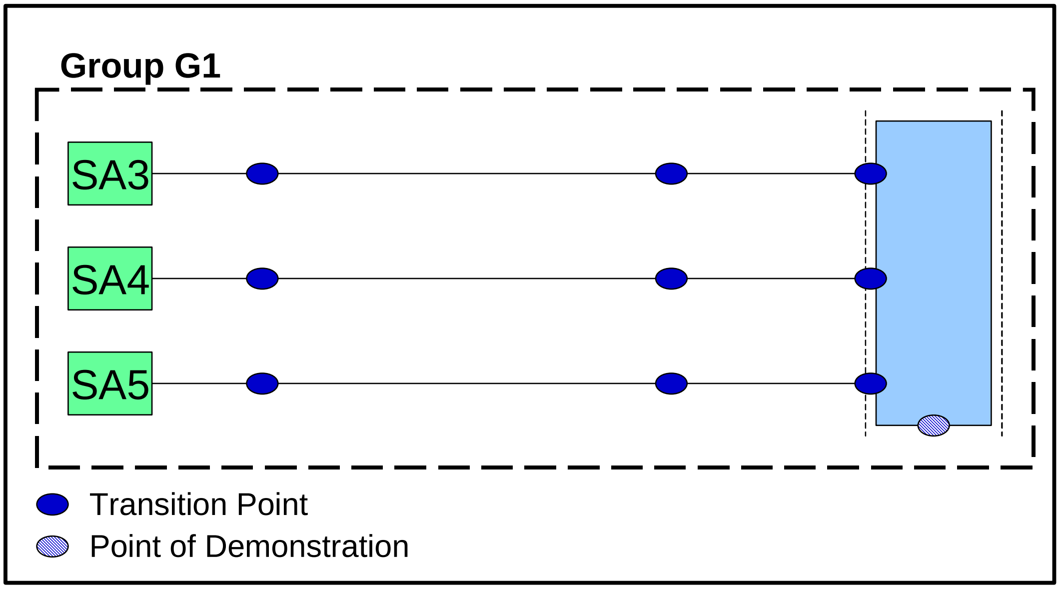


Figure 3.6 - Calculation of Groundwater Transport to Surface Water in the Probabilistic Case

In order to identify all of the variables that contribute to the calculation of the ending concentration, and therefore all of the input variance values that contribute to the variance in the ending concentration, in the probabilistic case, the exposure pathways that contribute to the concentrations at a single point of demonstration are identified in the beginning as part of a group. At the point of demonstration, the combined concentration is calculated and the variance is calculated based on all of the variable parameters for all of the contributing exposure pathways. As in the deterministic case, the ending concentration is checked against aqueous solubility, vapor saturation, or soil saturation, as applicable, to ensure the combined concentration is physically possible.

The expected ending concentration ($E(C_E)$), which is also the mean ending concentration, for each segment is calculated by using the expected value of the starting concentration and the expected values of all of the input parameters ($E(X_i)$). Equation 3.9 is the base equation for the expected concentration calculations.

$$E(C_E) = E(C_S) A_F \left[E(X_i) \right] \qquad \textbf{Equation 3.9}$$

where $A_F[E(X_i)]$ represents any of the attenuation factors.

In order to calculate the variance in the ending concentration, $Var(C_E)$, Equation 3.9 must be transformed into a linear model and the partial derivatives calculated for the ending concentration with respect to all of the input variables. Equation 3.9 can be transformed using a first order Taylor Series approximation (Ang and Tang, 1975). From a linear model the equation for the variance in the ending concentration, $Var(C_E)$, can be written. Equation 3.10 is the general equation for the first order Taylor Series approximation.

$$Y \cong g(\mu_{xi}) + \sum_{j=1}^{m} (X_j - \mu_{xj}) \frac{\partial g(X)}{\partial X_j} \qquad \textbf{Equation 3.10}$$

where there are m input parameters and the partial derivatives are evaluated at the mean values for all of the input parameters.

The first order approximation is exact for linear equations. As the non-linearity of the attenuation factor equation increases, the amount of error in the result increases. This is because the first derivatives are not constant over the entire range of function values when the base equation is non-linear.

For C_E, the first order equation is written in terms of the input variables X_j, and Equation 3.11 is evaluated to estimate $Var(C_E)$.

$$\mathrm{Var}(C_E) = \sum_{j=1}^{n}\left(\frac{\partial C_E}{\partial X_j}\right)^2 \mathrm{Var}(X_j) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left(\frac{\partial C_E}{\partial X_i}\right)\left(\frac{\partial C_E}{\partial X_j}\right)\rho_{ij}\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}$$

Equation 3.11

where

ρ_{ij} - correlation coefficient between variables i and j
$Var(X_i)$ - estimated variance in input parameter X_i
$Var(X_j)$ - estimated variance in input parameter X_j
$\left(\frac{\partial C_E}{\partial X_j}\right)$ - the partial derivative of C_E with respect to the input variable X_j.

It is important to note that the derivatives for $C_E(N)$ with respect to all of the input variables are needed, including the variables that are only input parameters to earlier exposure pathway segments. As an example, if X_1 is an input parameter only for $C_E(1)$, then the derivative with respect to $C_E(N)$ is determined based on the product rule:

$$\frac{\partial C_E(N)}{\partial X_1} = \frac{\partial C_E(1)}{\partial X_1} \times \frac{\partial C_E(N)}{\partial C_E(1)}$$

Equation 3.12

The partial derivatives are calculated for each variable for each attenuation factor model. If a variable is an input to more than one exposure pathway segment the expression is more complex, but dependent on the specific attenuation factors involved. All of the partial derivative equations are included in Appendix A. Additional discussion of the implementation of the probabilistic case is presented in Chapter 4.

This analytical method was chosen for the calculation of the variance values because, in terms of complexity, it is comparable to the attenuation factor algorithms that are used in the SERA methodology. The first order approximation is also a good first calculation to make before more complex uncertainty analyses are implemented. While the method is approximate for non-linear equations, and for the most part the attenuation factor equations are non-linear, the method may provide a sufficient approximation of the variance, even for non-linear equations. In addition, for initial site-specific risk assessment calculations, responsible parties and regulatory agencies may be more comfortable with an analytical solution for the variance calculation than a numerical solution or other more complex solutions such as those based on Monte Carlo simulation.

Appendix A includes the equations for the partial derivatives for each attenuation factor. Appendix D includes the computer procedures used to implement the variance calculations.

3.4 RESULTS EVALUATION

The results for the exposure pathways include the expected concentrations at the points of demonstration and, in the probabilistic case, the variance estimates for those concentrations. The exposure pathway results are compared to the protective exposure point concentrations (i.e., target levels) selected for each type of exposure. In the SERA methodology the protective exposure point concentrations can be from several different sources:

- regulatory standards (e.g., drinking water maximum contaminant levels (MCLs), surface water quality criteria);
- state-defined risk-based values;
- calculated health-protective concentrations based on dose-response relationships; or
- other values.

The protective exposure concentrations are defined by chemical of concern, by environmental medium and by an *exposure mechanism* identifier. The exposure mechanism is used to differentiate between different target levels for a particular chemical and environmental medium (e.g., benzene groundwater ingestion, benzene statewide groundwater protection value, benzene ecological protection value applied to groundwater).

If the concentrations are below the protective exposure point concentrations then the exposure pathway is considered to be *incomplete*. In this sense, a *complete exposure pathway* is one where all of the exposure pathway elements exist (i.e., source, transport mechanism, if necessary, exposure route and receptor) and the predicted chemical of concern concentration at the point of demonstration is above the target level. If any of the exposure pathway elements do not exist, or the concentration of the chemical of concern at the point of demonstration is below the target level, then the exposure pathway is *incomplete* and is eliminated from further consideration.

During the site conceptual model development, potential exposure pathways may be identified, however they may be determined to be incomplete before any fate and transport modeling is conducted. For example, regulatory provisions may preclude the use of the groundwater under a facility from being used as a source of drinking water for on-site workers. The groundwater ingestion exposure pathway would be included in the site conceptual model as a potential exposure pathway, but would be assigned the property *incomplete*. In this way all of the evaluations conducted during the risk assessment are documented and the justification for exposure pathway elimination is maintained in the database records.

If the probabilistic case is implemented for an exposure pathway a more sophisticated comparison is performed. The probability that the ending concentration will be below the protective exposure point concentration is calculated for these exposure pathways and compared to a target probability (e.g., $p = 0.90$). Based on the probability comparison the exposure pathway is considered to be complete or incomplete. Figure 3.7 shows the plot of the predicted ending concentration for a particular exposure pathway and the applicable target level.

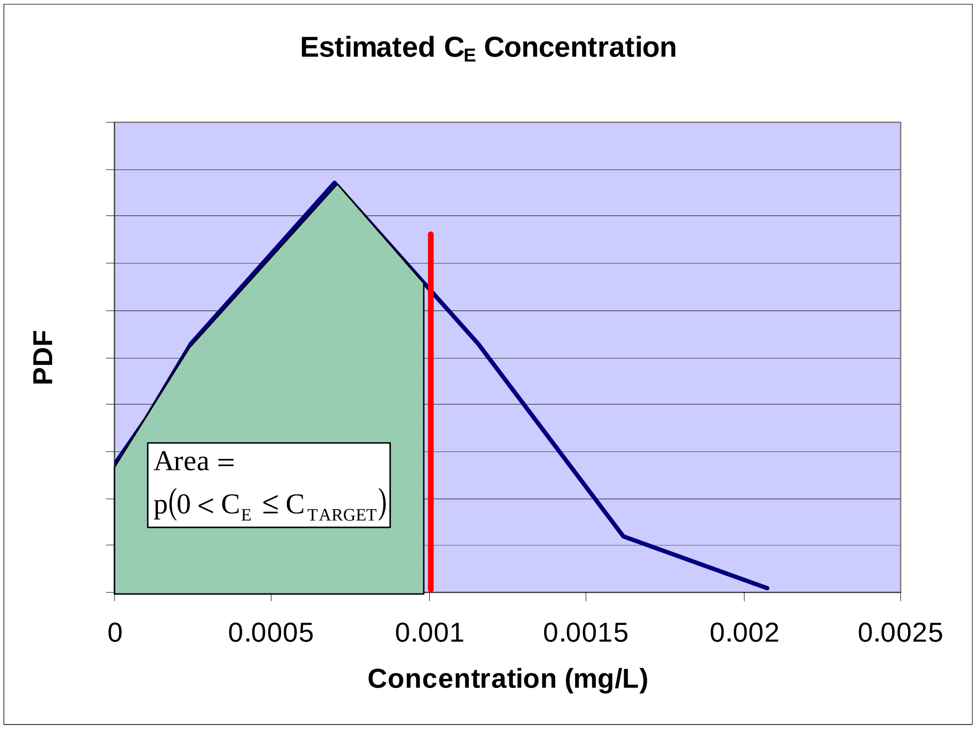


Figure 3.7 - Probabilistic Comparison of Ending Concentration Distribution to Target Level

For the example shown in Figure 3.7, suppose that the target probability of (C_E < Ctarget) is p = 0.9. This means that we require 90-percent certainty that the observed value will be below the target level ($1x10^{-3}$ mg/l). The distribution is truncated at a concentration of zero, since negative concentrations are not physically possible.

For the given mean ($7x10^{-4}$ mg/l) and variance ($2.1x10^{-7}mg^2/l^2$), and standard deviation ($4.6x10^{-4}$ mg/l) of the ending concentration, assuming the ending concentration is normally distributed, the probability that (C_E < Ctarget) is calculated as:

$$p\left(C_E \le C_{TARGET}\right) = \Phi\left(\frac{C_{TARGET} - \mu_{CE}}{\sigma_{CE}}\right) - \Phi(-\infty)$$

$$= \Phi\left(\frac{1\times10^{-3} - 7\times10^{-4}}{4.6\times10^{-4}}\right) - 0 = \Phi(0.655) \qquad \textbf{Equation 3.13}$$

$$p = 0.745$$

where
$\Phi(X)$ - standard normal distribution function
μ_{CE} - mean ending concentration
σ_{CE} - standard deviation of the ending concentration.

Therefore, in this case, we would say that the exposure pathway may in fact be complete (even though the mean value is less than the target level) and that additional evaluations are warranted.

In addition, the coefficient of variation (c.o.v.) can be calculated based on the mean and standard deviation of the ending concentration.

$$\text{c.o.v.} = \delta_{CE} = \frac{\sigma_{CE}}{\mu_{CE}} \qquad \textbf{Equation 3.14}$$

The coefficient of variation (c.o.v.) is an estimate of the variability with respect to the mean value. A large c.o.v. would indicate that there may be value in collecting additional data to refine the variance estimates for the input parameters, while a small c.o.v. would indicate that there may not be value in refining the input parameter variance values.

For the example, the c.o.v. value is 0.65, a relatively small c.o.v., which would indicate that there may not be an opportunity to refine the input parameters.

Based on the probability comparisons, and the c.o.v. values, additional evaluations would be considered. These would include: (1) collecting additional data to refine input parameters and reduce the uncertainty in the results, (2) completing additional modeling, or (3) implementing remedial alternatives. All of these alternatives can be evaluated using a decision analysis framework (e.g., implementing decision trees using utility functions to value alternatives) (Hay Wilson *et al*., 1998).

3.5 SUMMARY

This chapter presents the development of the Spatial Environmental Risk Assessment (SERA) methodology. A brief overview is included to provide a road map of the overall process. Then, the data model and the calculation methods are presented. Finally, the evaluation of the exposure pathway results is discussed. Detailed fate and transport algorithms, case study input values and computer procedures are presented in Appendices A through D. The next chapter presents the application procedure for the SERA methodology using PC-based software tools.

Chapter 4 : Application Procedure

This chapter presents the application procedure for the Spatial Environmental Risk Assessment (SERA) methodology using specific PC-based software tools. The discussion in this chapter builds on the framework presented in the data model and the calculation methods discussed in Chapter 3. The chapter begins with a brief discussion of the software tools used in this research. Then the application is discussed in detail based on its fundamental components, the spatial site conceptual model, the site conceptual model database and the fate and transport algorithms implemented in spreadsheets. Detailed algorithms and computer procedures are presented in the appendices.

4.1 SOFTWARE TOOLS

The application of the Spatial Environmental Risk Assessment (SERA) methodology requires the selection of software tools to perform the many data handling and analysis functions. For this research, the ESRI software ArcView 3.1 (ESRI, 1998) is used as the geographic information system (GIS) component, Microsoft Access 2000 (Microsoft, 1999) is used as the database component and Microsoft Excel 2000 (Microsoft, 1999a) is used as the spreadsheet component. These software applications were selected because they are available at the University of Texas research computer laboratories, have been used extensively in the research conducted to date, can be linked to one another using built-in functions, and are widely used throughout the world.

The remaining sections in this chapter describe how the Spatial Environmental Risk Assessment (SERA) methodology was implemented using the PC-based software tools. This chapter includes some basic examples of the implementation. Additional examples from the case study facility are presented in Chapter 5.

4.2 REPRESENTING THE SITE CONCEPTUAL MODEL ELEMENTS USING GIS

The spatial site conceptual model includes a spatial representation of the exposure pathway elements for multiple sources, exposure pathways and potential receptor locations. The components of the spatial site conceptual model (e.g., sources, transition points, points of demonstration) are represented as individual themes or data layers in a vector GIS database. Table 4.1 includes the data layers used in the case study examples.

Table 4.1- Site Conceptual Model GIS Data Layers

Data Layer	**Type**	**Data Description**
Source Areas	Polygon	Area of NAPL or highest concentration
Source Area Center Points	Point	Centroid of source area
Soil Vapor Transition Points	Point	Location of end of NAPL partitioning to vapor exposure pathway segment
Groundwater Transition Points	Point	Location of end of NAPL partitioning to groundwater exposure pathway segment
Surface Water Transition Points	Point	Location of end of groundwater transport exposure pathway segment
Groundwater Transport Segments	Line	Calculation length for groundwater attenuation mechanism
Surface Water Mixing Boxes	Polygon	Area of application of the target level for surface water

Air Mixing Boxes	Polygon	Area of application of the target level for air
Points of Demonstration	Point	Location of comparison of estimated concentration to target level

These data layers identify the source areas, transport mechanisms and potential receptor locations within the GIS application. The data layers are developed based on the information collected in the digital facility description (e.g., historical information, equipment locations, environmental measurements). As an example, the locations of measured thicknesses of non-aqueous phase liquids (NAPL) floating on the groundwater table represent source areas for chemicals of concern that are part of that NAPL; source area polygons are constructed based on these data.

4.2.1 Sources and Source Areas

Sources can be defined as point locations or as areas. Sources defined as point coverages include individual points for releases (e.g., the location of an emissions stack). If the sources are defined as areas of environmental media containing chemicals of concern (e.g., concentrations of a chemical in soil, or an area of non-aqueous phase liquid (NAPL) on the groundwater table) then they are defined as polygon coverages. In the examples presented in Chapter 5 and Appendix B, the development of the specific source area features is discussed. Figure 4.1 illustrates six example source areas identified for the case study facility. These source areas are defined by measurements of non-aqueous phase liquids (NAPL) on the groundwater table. The potential exposure pathways that are used as examples in this section to describe the GIS element development are:

groundwater transport to surface water; volatilization to the surface and worker exposure in a future indoor environment; and worker exposure to vapors in the outdoor environment.

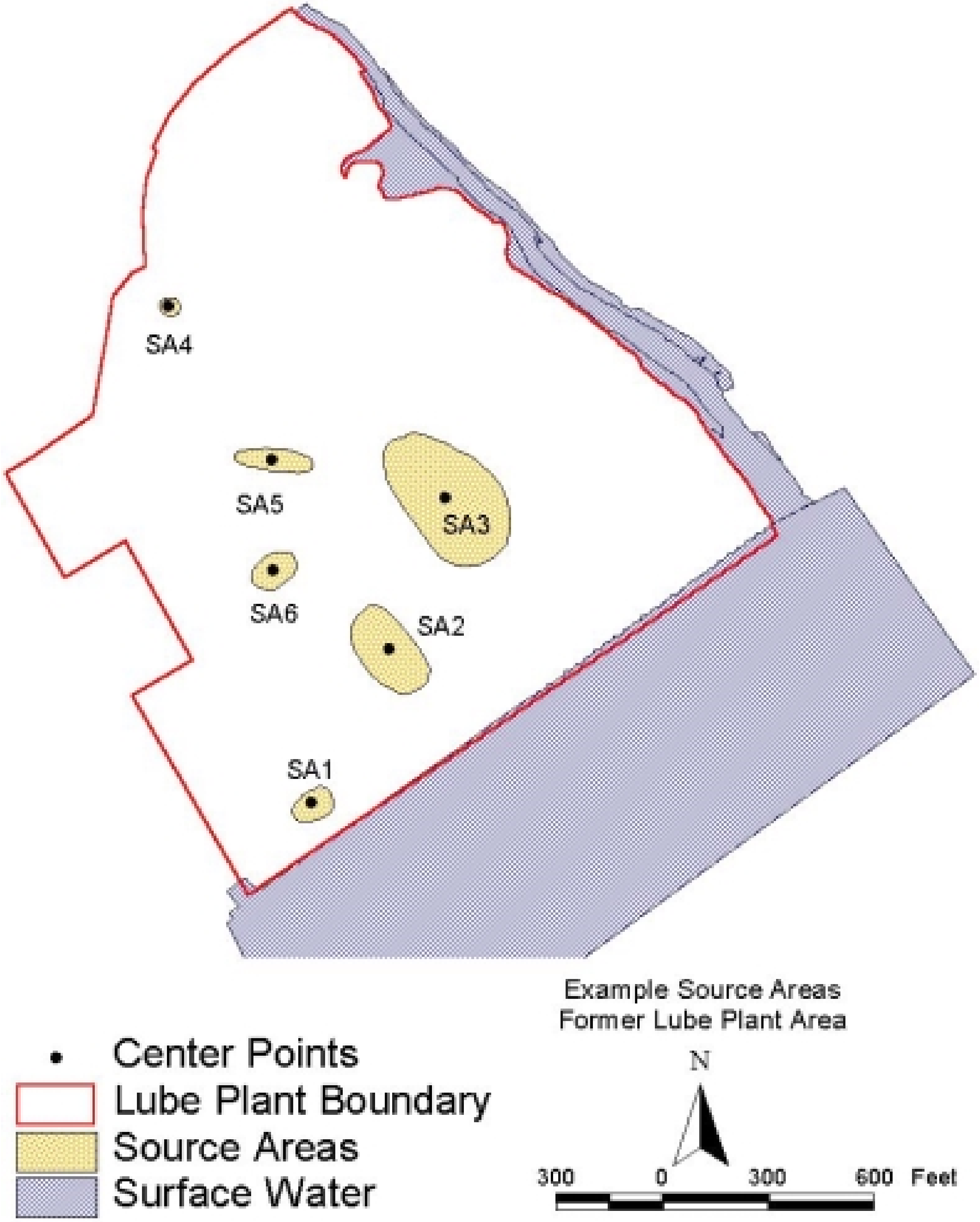


Figure 4.1 - Example Source Areas for the Case Study Facility

Once the source areas are located, the individual spatial objects are attributed with a LOC_ID (e.g., SA1), a source description, a center point (calculated as the centroid of the polygon), the x- and y-coordinates of the center point and the polygon perimeter and area. The LOC_ID and the Location_ID are the same values; in ArcView the length of the field name is limited so in the GIS tables LOC_ID is used, elsewhere Location_ID is used. The calculation of center point location, perimeter, area and assigning x-, and y- coordinates are accomplished using an ArcView extension program called **CRWRVector** developed by Dr. Francisco Olivera at The University of Texas at Austin (Olivera, 1999). Additional information on using the extension is included in Appendix C. In addition, the width of a source area perpendicular to the groundwater flow direction (Gwsource) is measured for any groundwater source since this is a parameter that will be used in the attenuation calculations. The width is measured using the Measuring Tool available in ArcView. Table 4.2 includes the attribute table for the source area shapefile. Each shapefile has an associated attribute table in which data for the particular spatial objects are stored.

Table 4.2 - Example Source Area Attribute Table

Shape	LOC_ID	Description	Area	Perimeter	Gwsource	X_coord	Y_coord
Polygon	SA4	NAPL	2417.468	177.883	54	2625639.23889	184942.92250
Polygon	SA5	NAPL	11288.791	494.264	75	2625931.47830	184507.29001
Polygon	SA6	NAPL	10120.193	368.890	130	2625932.98406	184195.60267
Polygon	SA1	NAPL	8292.789	339.169	120	2626041.52226	183534.11976
Polygon	SA2	NAPL	35138.575	713.364	150	2626260.21263	183971.77386
Polygon	SA3	NAPL	89441.032	1135.852	430	2626418.38263	184397.84589

4.2.2 Transition Points

The transition points represent the transfer of chemicals of concern from one environmental medium to another by physical or chemical processes (e.g., NAPL partitioning to groundwater). The transition points are located based on the location of the source area, the locations of the points of demonstration (see section 4.2.3) and the types of transport mechanisms included in the exposure pathway (see section 4.2.4). The transition points are grouped in themes based on environmental medium (e.g., soil vapor, groundwater, surface water). The points are attributed with a LOC_ID (e.g., TP1), the source area and exposure pathway they are associated with, environmental medium, and x- and y- coordinates, again using the **CRWR Vector** extension. Figure 4.2 includes the transition points for soil vapor, as yellow triangles, surface water as blue circles, and groundwater as purple squares. The location labels are included for all the points. The soil vapor locations are labeled using Italic lettering.

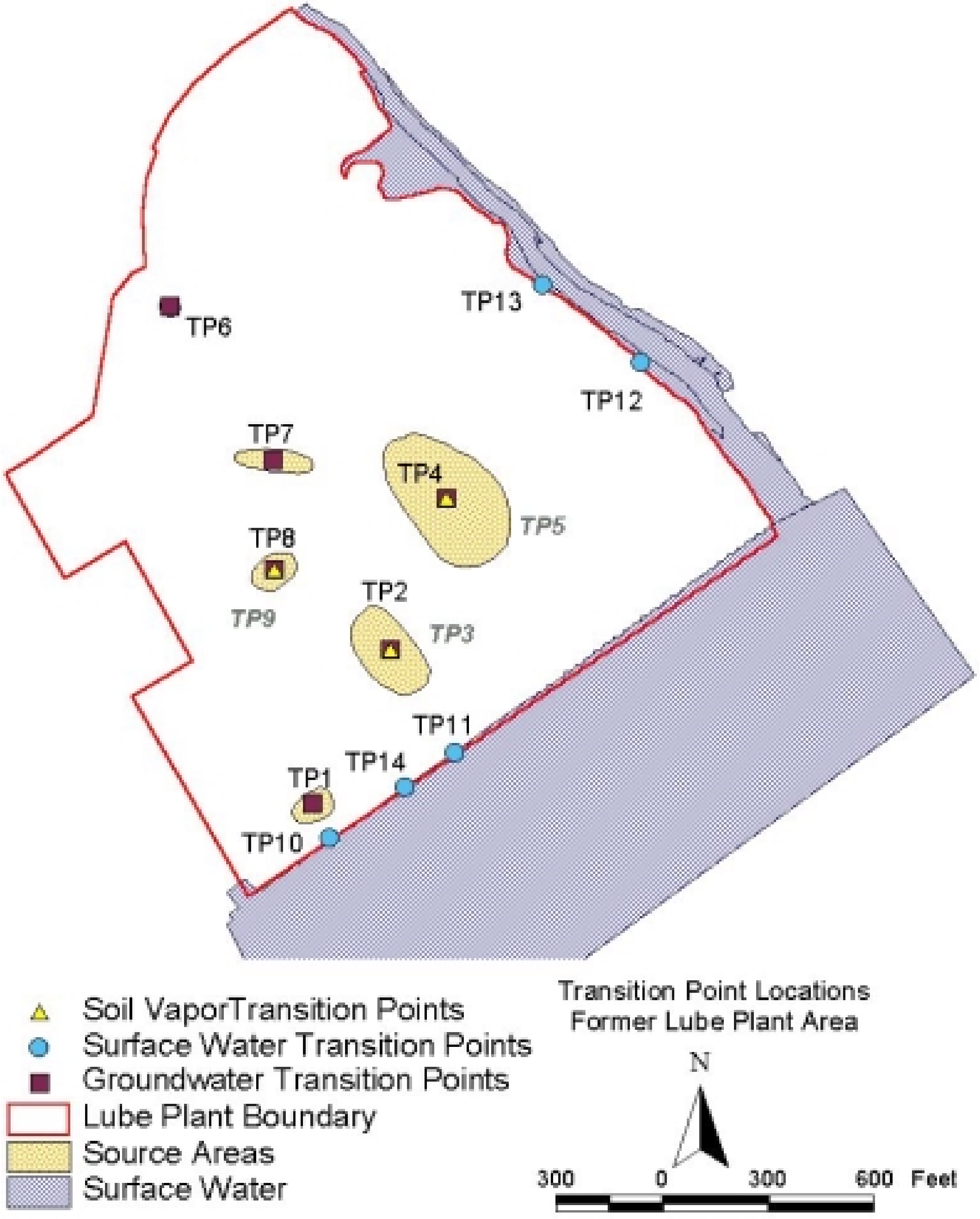


Figure 4.2 - Example Transition Points for the Case Study Facility

4.2.3 Points of Demonstration

Receptor locations are identified based on the spatial data and the current and potential future activities on the facility and surrounding properties. The areas for which receptor identification is needed can be identified based on the digital facility description (e.g., land use data, census data or digital ecological habitat data). In the Spatial Environmental Risk Assessment (SERA), these points at which environmental medium concentrations are calculated are called points of demonstration, to take into account that they can actually be different kinds of points (e.g., points of exposure, points of compliance) depending on the specific application of the SERA. Section 3.2.1.4 discusses the different bases for selecting points of demonstration. The points of demonstration are attributed with POD_IDs, the environmental medium represented at that point, and the x- and y-coordinates, again using the **CRWR Vector** extension. Figure 4.3 includes the points of demonstration (PODs) for the example exposure pathways. The green, circle PODs are for surface water, the magenta, triangle PODs are for indoor air and the orange, square POD is for outdoor air.

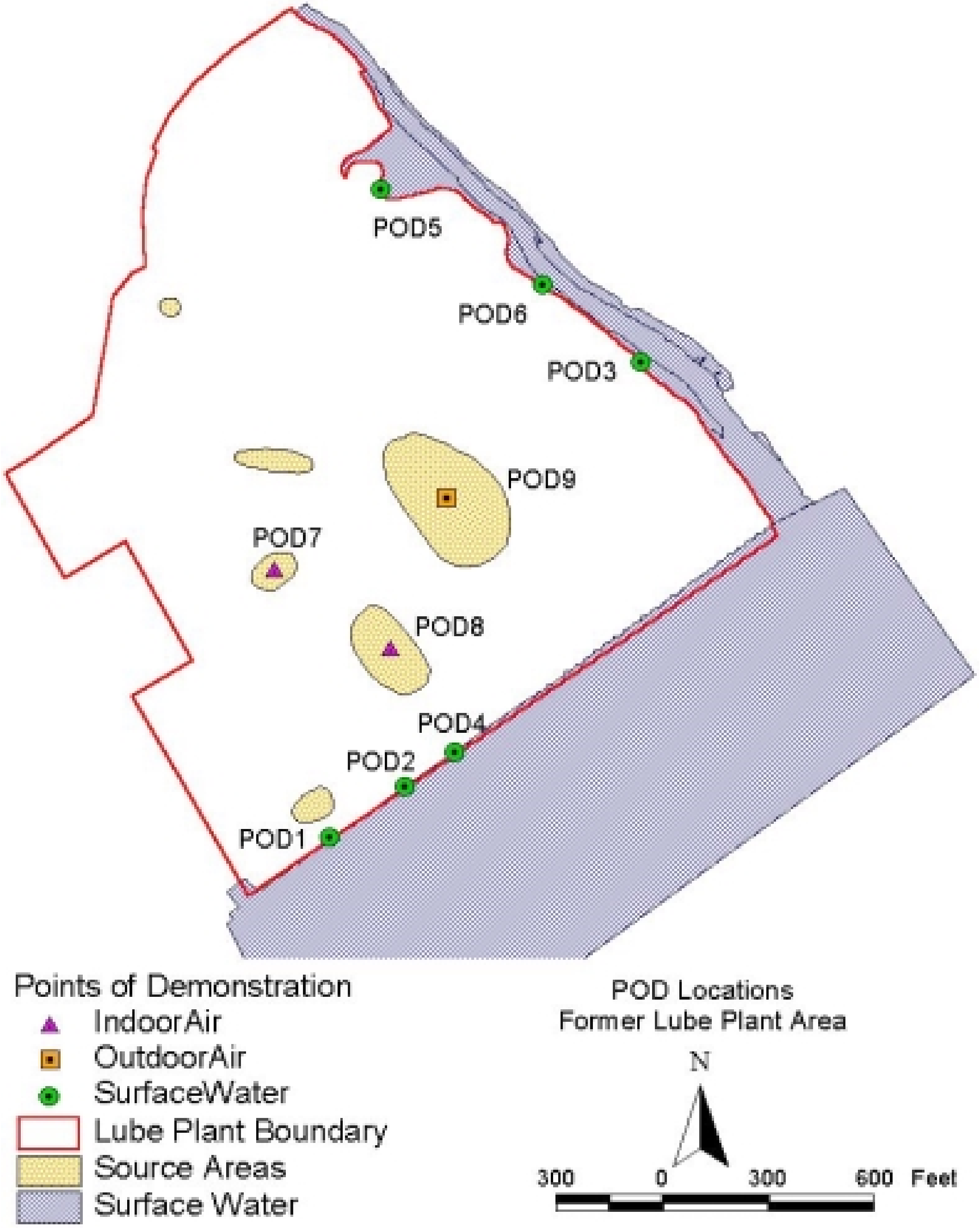


Figure 4.3 - Example Exposure Pathways Illustrating POD Locations

4.2.4 Lateral Transport Mechanisms

The transport mechanisms that link the sources to the points of demonstration are grouped in themes based on environmental medium (e.g., outdoor air, groundwater). Lines define the lateral transport mechanisms such as groundwater flow. The lateral transport segments are constructed in the GIS software using a script (**gwtrans.ave**). The script that takes the beginning point locations (e.g., source area center points, transition points), the ending point locations (e.g., transition points, points of demonstration) and a table that matches the starting locations (e.g., LOC_IDs) with the applicable ending locations (e.g., POD_IDs), and generates a line shapefile of transport segments. Table 4.3 includes the table of source areas matched with their points of demonstration for the examples.

Table 4.3 - Source Areas and PODs Matching Table for Groundwater Transport

LOC_ID	POD_ID	X1	Y1	X2	Y2	ID
SA1	POD1	2626041.5223	183534.1198	2626092.6585	183437.6445	SA1
SA2	POD4	2626260.2126	183971.7739	2626442.9430	183677.1708	SA2
SA3	POD3	2626418.3826	184397.8459	2626968.9893	184781.3776	SA3
SA4	POD5	2625639.2389	184942.9225	2626237.6856	185273.6896	SA4
SA5	POD6	2625931.4783	184507.2901	2626694.0140	185001.0745	SA5
SA6	POD2	2625932.9841	184195.6027	2626305.9474	183581.1985	SA6

The CRWR Vector extension is then used to assign a length value to the attribute table for the new transport mechanism file. The length value is used in the attenuation factor calculation. Additional information on the gwtrans.ave script is included in Appendix C. Figure 4.4 illustrates the groundwater transport mechanism shapefile for the examples. The groundwater transport segments are identified by their concatenated LOC_ID, Pathway_ID and Segment_ID (e.g., SA4P1S2).

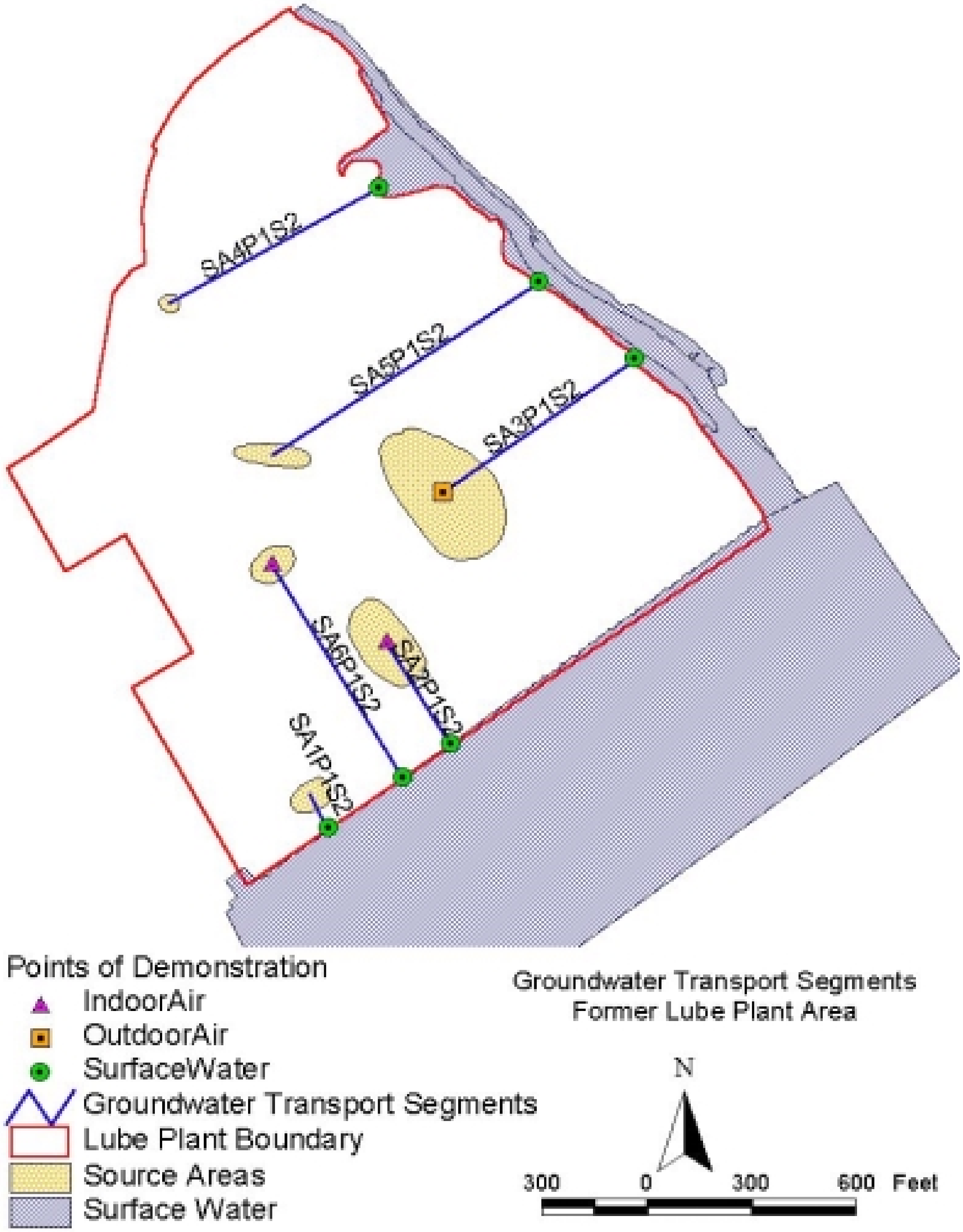


Figure 4.4 - Example Exposure Pathways Illustrating the Groundwater Transport Segments

For other transport mechanisms, or for more sophisticated transport algorithms, other physical characteristics of the sources, source areas, transport segments, and receptor locations, may be needed for the ending concentration calculations (e.g., wind erosion from a surface soil area source may require a cross-wind width for the source area). In these situations, built-in functions of the GIS software, or customized extension programs or scripts, may be used to derive the needed parameter values.

4.2.5 Vertical Transport Mechanisms

In the case where chemicals of concern are moving in the vadose zone, (e.g., with infiltrating water, as vapors rising in the vadose zone), the transport mechanism is a vertical one. The distances between the z-coordinates for the land surface elevation, soil source areas and the groundwater table elevation can be used to determine the transport distances for these vertical transport pathway segments. In the SERA application, the z-distance is estimated based on data from the digital facility description and the value is stored in the Source Variables table in the tabular database (see section 4.3.3.5).

4.2.6 Spatial Site Conceptual Model

In general, the attributes that are assigned to the spatial objects are identifiers that allow them to be matched-up with the corresponding records in the site conceptual model database, values that can be used to classify them within their group (e.g., environmental media identifiers can be used to separate the points of demonstration in maps), and physical values that can be used in the fate and transport calculations. An example map of the site conceptual model based on the case study facility is shown in Figure 4.5. The figure includes six source areas, nine points of demonstration (including six surface water, two indoor air and one outdoor air point of demonstration), and six lateral transport mechanisms.

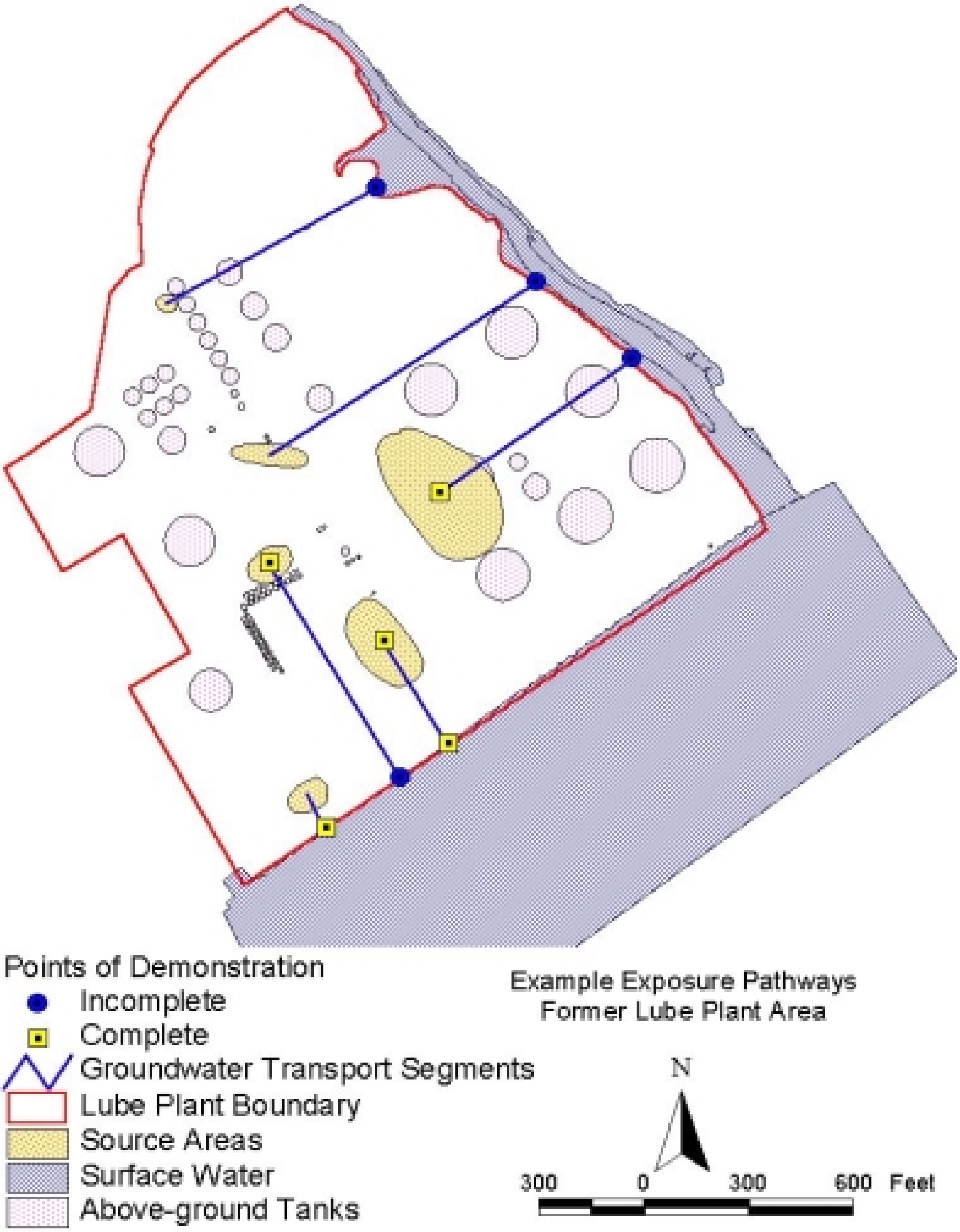


Figure 4.5 - Example Site Conceptual Model Map

Figure 4.5 is an illustration of the results of screening calculations where the points of demonstration are yellow squares for complete exposure pathways (i.e., based on the comparison to the protective exposure concentrations the predicted concentrations are above the target levels) and are blue circles for incomplete exposure pathways. The exposure pathway completeness evaluations are discussed in section 3.4.

As described in Section 3.3.2, for the probabilistic case, the ending concentration is calculated for a single point of demonstration in a mixing box. Figure 4.6 illustrates the same exposure pathways included in Figure 4.5, with the addition of two surface water mixing boxes, an indoor air mixing box (representing a future commercial building constructed over two source areas) and an outdoor air mixing box representing a work area for an onsite worker.

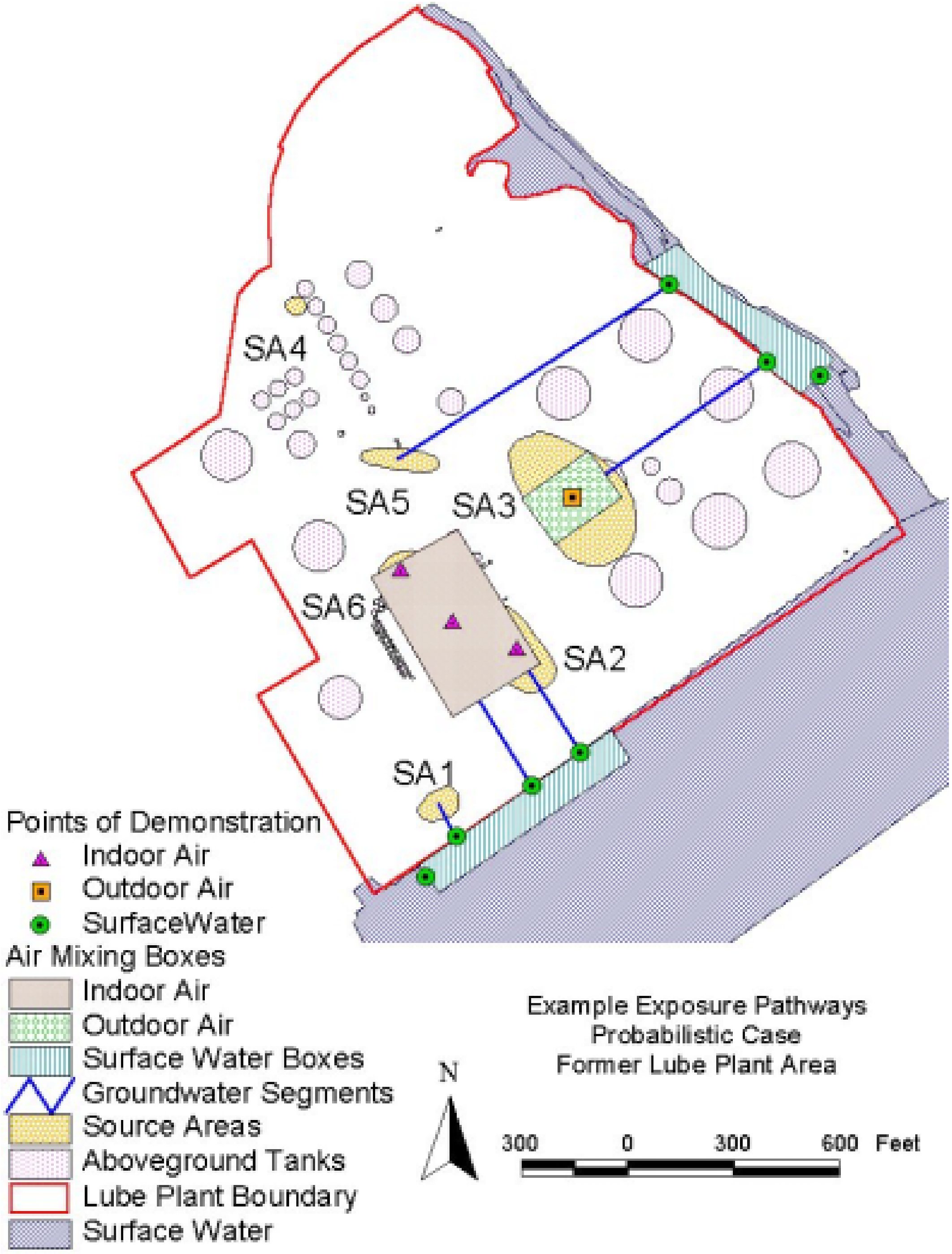


Figure 4.6 - Example Site Conceptual Model Map including the Media Mixing Boxes

Three points of demonstration have been added to the site conceptual model to represent the ending locations for these combined exposure pathways (i.e., the downstream edge of the mixing boxes for the surface waters and the center of the future building mixing box).

The next section describes the development of the site conceptual model database.

4.3 BUILDING THE SITE CONCEPTUAL MODEL DATABASE

As discussed in Chapter 1, the site conceptual model is a critical component of the risk assessment because it provides a description of all the potential exposure pathways associated with chemicals of concern at the many potential source areas, their movement in the environment and their relationship to potential receptors. The site conceptual model involves a synthesis of spatial and observational data. The concept of using a tabular database to build a site conceptual model was developed as part of The University of Texas at Austin's risk-based decision support research program and documented in the master's thesis by Koerner (1998). This conceptual foundation was used in the current research as the starting point of the site conceptual database used in the SERA implementation. Changes and additions were made to almost all of the tables. The most important enhancements to the database were to support the connection of the site conceptual model database to the fate and transport calculations.

The site conceptual model database has six goals:

1. Tracking the components (i.e., source area, transport mechanisms, transition points, point of demonstration, exposure route, receptor) of each exposure pathway;
2. The capacity to evaluate multiple exposure pathways affecting one point of demonstration;
3. Documenting the modeling algorithms and parameters used to analyze the fate and transport of the chemical of concern from source to receptor;
4. Identifying the protective exposure point concentrations that are used to determine if an exposure pathway is complete;
5. Flexibility to include both current and future exposure scenarios and an unlimited number of exposure pathways associated with any source area; and
6. Tracking results of the modeling for concentrations at points of demonstration, including an estimate of the variance in the ending concentration.

An additional goal of the site conceptual model database was specific to the implementation of risk-based corrective action in Pennsylvania for the case study facility. The first evaluation under the Pennsylvania Land Recycling Act (Act 2) is to compare concentrations of chemicals of concern in environmental media (i.e., soil and groundwater) to Medium Specific Concentrations. If the concentrations of chemicals of concern are at or below the appropriate Medium Specific Concentrations values, then no further actions are necessary for that chemical of concern in that medium (Commonwealth of Pennsylvania, 1997). Since this is considered a "preliminary exposure pathway analysis" it is part of the site conceptual model. The results of these comparisons are stored in the site conceptual model database, to document the results and to support the definition of the exposure pathways that will be considered further in the exposure pathway analysis.

The following sections document the design of the site conceptual model database.

4.3.1 Relationship Keys and Unique Identifiers

In the design of a tabular database, the first task is to identify the individual types of records to be stored and the fields associated with each record. The fields of data can be divided based on different types of information, which are the basis for separating tables within the database. As discussed in Section 2.3, in a tabular database the *relationships* refer to the connections between data in one table and the related data in another table. These tables are related to one another based on *key fields*. Key fields contain unique identifiers that make each record in a table unique and different from all of the other records in that table.

Typically there are one or more tables in a database that are central to the information stored in the database. All of the other tables supplement and provide detail to the records in the central tables. For the site conceptual model the central tables are the Pathway table and the Pathway_Segments table. The key fields for the Pathway and Pathway_Segments Tables are given in Table 4.4.

Table 4.4 - Key Fields for the Pathway and Pathway_Segments Tables

Key Field	Value	Table	Description
Location_ID	Alpha numeric value, e.g., SA1	Pathway, Pathway_ Segments	Describes the source area for the exposure pathway
COC_ID	Chemical Abstract Number for the chemical	Pathway, Pathway_ Segments	Identifies the chemical of concern for the exposure pathway
Pathway_ID	Alpha numeric value, e.g., P1	Pathway, Pathway_ Segments	Unique identifier for each exposure pathway originating from a source area
Segment_ID	Alpha numeric value assigned sequentially for each exposure pathway, e.g., S1	Pathway_ Segments	Unique identifier for each segment of an exposure pathway
AboveGrd	Alpha numeric value, e.g., AG1	Pathway, Pathway_ Segments	Modeling scenario identifier for above-ground parameters
ChemVar	Alpha numeric value, e.g., C1	Pathway, Pathway_ Segments	Modeling scenario identifier for chemical properties
Constants	Alpha numeric value, e.g., CN1	Pathway, Pathway_ Segments	Modeling scenario identifier for the physical constants
MdlErrVar	Alpha numeric value, e.g., M1	Pathway, Pathway_ Segments	Modeling scenario identifier for the model error parameters used in the probabilistic case
SubSurVar	Alpha numeric value, e.g., SS1	Pathway, Pathway_ Segments	Modeling scenario identifier for subsurface parameters
SWVar	Alpha numeric value, e.g., SW1	Pathway, Pathway_ Segments	Modeling scenario identifier for surface water parameters

The Location_ID is assigned for an exposure pathway as the source area location (e.g., SA1). The values can be any *alpha numeric string*, although shorter strings are advised (i.e., strings of two or three characters). The location identified corresponds to a physical location within the site that is mapped in the GIS. The COC_ID is the chemical abstracts registry number for the particular chemical. The Pathway_ID is assigned to each exposure pathway from a source area. The convention adopted here is to begin with P1 at each source area and number sequentially all the potential exposure pathways for a particular source area (e.g., P1, P2, etc.). The modeling input parameter value identifiers are as described in Table 4.4. If any of the modeling input parameters do not apply to the particular exposure pathway (e.g., a soil vapor to indoor air exposure pathway would not require the surface water mixing model parameters), then the identifier field is given a value of zero. Each exposure pathway record is related to only one source area or point source record.

The Pathway_Segments table is used to identify all of the calculation segments into which an individual exposure pathway is divided. The Pathway_Segments table therefore requires an additional unique identifier, or key field, to differentiate between the segments that compose a exposure pathway. The additional identifier is the Segment_ID. The Segment_ID convention adopted here is to begin each exposure pathway with S1 and number sequentially all the segments for a particular exposure pathway (e.g., S1, S2, etc). Each segment record is related to only one exposure pathway record, except if the segment begins at a transition point where multiple exposure pathways are combined. These are described as combined segments in section 3.3.

In the implementation of the calculations a unique serial number is needed for each segment record. This unique serial number is created by concatenating the ten key fields listed in Table 4.4 resulting in a unique string. For example, the unique serial number for the outdoor air exposure pathway identified with the example key field values in Table 4.4 is: for the chemical of concern benzene (i.e., chemical abstracts number 71-43-2); from source area SA1; exposure pathway P1; segment S1; above-ground model parameter values AG1; chemical variable values C1; constant values CN1; model error variable values M1; subsurface parameter values SS1. The surface water variables are not used in the outdoor air algorithms, so the surface water scenario value is zero. The unique serial number for this exposure pathway segment is:

SA171-43-2P1S1AG1C1CN1M1SS10.

The first order uncertainty approximation method includes the flexibility to consider correlation among the input values (see section 3.3.2). This is accomplished by segregating the modeling parameter input values into independent sets based on physical attributes. As an example, the above ground parameters like wind speed are physically independent of the subsurface variables like groundwater gradient. Within each model parameter set the variables are potentially correlated. In the database, this translates into separate tables for each variable set and separate variable identifiers.

The connections between the tables in the database through the key fields are shown in Figure 4.7. Only representative tables are show for space. Figure 4.8 shows the relationships for the preliminary exposure pathway evaluation data tables.

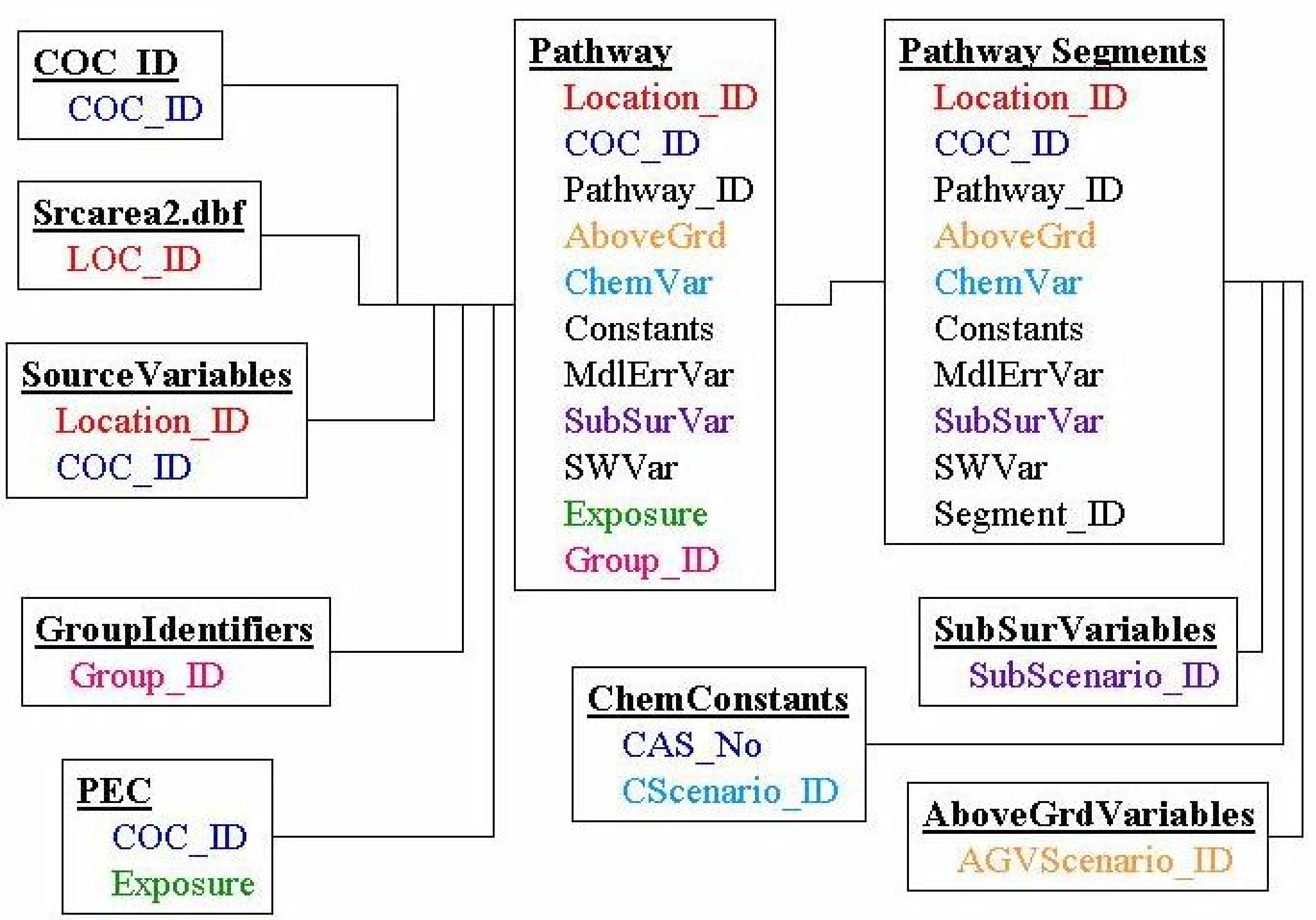


Figure 4.7 - Site Conceptual Model Database Table Relationships

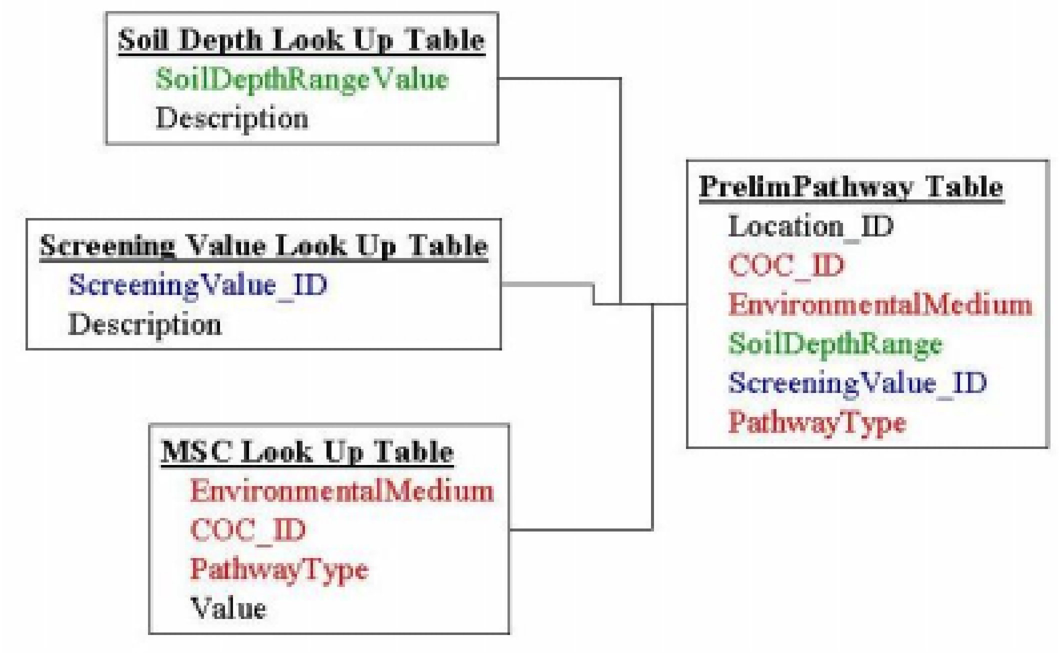


Figure 4.8 - Site Conceptual Model Database Table Relationships for the Preliminary Exposure Pathway Evaluation

4.3.2 Types of Tables in the Database

In the tabular database there are five distinct types of tables. The tables are: input tables; GIS attribute tables; spreadsheet results tables; query tables; and, reference tables. The following sections describe each of the table types.

4.3.2.1 Input Tables

The first group of tables, representing the majority of the tables in the database, is the input tables. Examples of these tables are the Pathway and Pathway_Segments tables. These tables are designed and created in the database software to manage the attribute data for the site conceptual model. These tables contain the primary descriptions of the exposure pathway characteristics. Section

4.3.3.2 and 4.3.3.3 describe the Pathway table and the Pathway_Segments table, respectively.

4.3.2.2 GIS Attribute Tables

The second group of tables is the linked GIS file attribute tables containing the results of the spatial descriptions of the exposure pathway features. These tables are developed in the GIS software as part of the spatial description of the exposure pathway objects and are in a database file format. Discussion of their development is contained in Section 4.2. It is important to note that in this implementation of the Spatial Environmental Risk Assessment (SERA) methodology the data tables that are linked from one software application to another are *live links* (i.e., when data values change in one application they change in all applications where the links exist). If this were not the case, any time additional exposure pathways are evaluated or modeling input values are changed, the table linkages would have to be re-established manually.

Since the GIS attribute tables are easily linked to the database, and more importantly the database tables can be accessed in the GIS software, a *data partitioning rule* has been developed in the SERA methodology, which is to store only the minimum required data in the attribute tables in the GIS files and manage the remainder of the data in the tabular database. This takes best advantage of the strengths of the two software applications. The GIS attribute tables need only the identifiers (e.g., LOC_ID, Pathway_ID, Segment_ID) that make them unique records when read in the tabular database and the sorting parameters (e.g., environmental media identifiers that can be used to separate the points of demonstration in maps), to enable the mapping. The remaining attributes in the GIS files are values that are calculated or measured in the GIS that cannot otherwise be derived in the SERA process.

4.3.2.3 Spreadsheet Results Tables

The third group of tables is the linked spreadsheet results tables. These tables contain the ending concentrations (and in the case of the probabilistic calculations, the variance in the ending concentrations) for the specific attenuation calculation segments. Discussion of the spreadsheet calculations and the results is contained in section 4.4.

4.3.2.4 Query Tables

The fourth group of tables is the query tables developed in the database software. The query tables are critical to the connection between the database records describing exposure pathways and the model calculations for the fate and transport analyses. The queries are used to gather the input data for each exposure pathway segment and organize the data into the correct format and column order to link to the spreadsheet. Specific discussion of the queries used in the implementation is included in the computer procedures in Appendix C and Appendix D.

4.3.2.5 Reference Tables

The fifth group of tables is the reference tables. These are called *look up tables* in the database. These reference tables provide the available values for fields in other tables. An example look up table is the Units_Lookup_Values table that specifies the unit system used for the variables. The table only contains two fields for each record, the Units field and a Description field. In the description field all of the variables using the particular units are listed. In the input tables, for example, the SubSurVariables table that contains all of the model input parameters that describe the soil and groundwater system at the facility, the units fields for each parameter are designed with a look up data source and the field is connected to the Units field in the Units_Lookup_Values table. This connection is implemented during the table design step in the database construction.

4.3.3 Primary Tables

This section presents individual descriptions of the design of the primary tables in the site conceptual model database.

4.3.3.1 PrelimPathway Table

The design of the preliminary exposure pathway evaluation table is shown in Table 4.5. The figure includes the field names, the data type for each field and a brief description of the data stored in that field. This is the table design view in the database software.

The table is designed to document the comparison of the concentrations of chemicals of concern in site environmental media to statewide standards (i.e., for the case study facility, the Pennsylvania Medium Specific Concentrations). The key fields for the table are Location_ID, COC_ID and EnvironmentalMedium. The location description field provides an opportunity to describe the sampling locations or physical location represented by this comparison. At this point in the analyses, the Location_ID may or may not represent a defined source area for the site conceptual model. The COC_ID and the Screening Value_ID are used to select the appropriate Medium Specific Concentration for the comparison to the representative concentrations for the location. The representative concentrations of chemicals of concern that are stored in this table are derived values from the site sampling data. These are the result of queries and analysis of the data stored in the environmental measurements database. The value may be a maximum value for a particular medium in one region of the facility. In this case the Location_ID would be the sampling location identifier for the sample. If the concentration were above the Medium Specific Concentration, further data evaluation would be needed to develop a source area representation for the site conceptual model.

If the concentrations of chemicals of concern in the environmental medium for a Location_ID are above the statewide standards then the applicable exposure pathways (including source areas, transport mechanisms and receptor locations) are identified and new records are generated in the Pathway table.

The PrelimPathway table is designed to track when specific locations, chemicals of concern and environmental media are identified for inclusion in the preliminary exposure pathway analysis, the reasoning for their inclusion, and also when a location, chemical of concern and medium has been eliminated from further consideration. A preliminary exposure pathway would have a Pathway_Status of *PotentialMSC* until the comparisons to the Medium Specific Concentrations have been completed. The results of the comparisons are recorded in the PrelimPathway table as *CompleteMSC*, requiring further evaluation, *IncompleteMSC*, elimination from further consideration.

The information in the table provides the documentation of the initial evaluation for a report, for example to the regulatory agency. Database report tables can be generated for any group of fields and records by constructing a database query. The resulting table can be a part of a paper report to the regulatory agency or the other stakeholders.

In the PrelimPathway table there is a field for the Soil Depth Range. In the Pennsylvania Medium Specific Concentrations values there are different soil depth range values used depending on a characterization of the land use and the types of receptors (i.e., residential or non-residential) for soil standards. The representative concentrations must be identified for the specified soil depth ranges and the corresponding Medium Specific Concentrations values used for the comparison. A look up table is used for the soil depth ranges, Table 4.6.

Table 4.5 - Preliminary Pathway Table Design

Field Name	Field Description
Location_ID	Describes the source location or the physical area of the environmental samples
COC_ID	Defines the COC for this pathway
EnvironmentalMedium	Defines the environmental medium for this comparison
LocationDescription	Description of the location/sampling area
SoilDepthRange	Descriptor for the soil samples included from look-up table
RepresentativeConcentration	Value used for the screening comparison - derived from environmental measurements database
RepConcentrationReference	Reference to query or data table with representative data
ScreeningValue_ID	Numeric value chosen for the comparison
Pathway_Status	Describes the status for the screening comparison
Basis_In	Description field for inclusion of the preliminary pathway
Basis_In_Date	Date that the preliminary pathway was identified for evaluation
Basis_Out	If the pathways are not screened out, they are added to the Pathway table
Basis_Out_Date	Date that the preliminary pathway was screened out
Verified_By	QC field that documents review of all pathway analyses

Notes: Shaded fields are key fields

Table 4.6 - Soil Depth Range Values Look Up Table

Soil Depth Range Value	Description
>15 feet	Residential or Non-Residential
0-15 feet	Residential
0-2 feet	Non-Residential
2-15 feet	Non-Residential
NA	Used for GW MSC

In addition, the Pennsylvania Medium Specific Concentrations depend on the use of the groundwater, the total dissolved solids concentration in the groundwater, the land use, and the equivalency and buffering modifying factors that take into account the attenuation of chemicals of concern in the vadose zone. The Medium Specific Concentration scenarios selected for use at the case study facility are identified in the Screening Value Lookup table shown in Table 4.7 (Walker, 2000).

The Medium Specific Concentration values are stored in the MediumSpecificConc(MSC)Values table. Table 4.8 shows the design of this table.

Table 4.7 - Screening Values used for the Case Study Facility

ScreeningValue_ID	Description
GWScenario1	GW, MSC Value used aquifer, TDS<=2500, residential
GWScenario2	GW, MSC Value used aquifer, TDS>2500, residential
GWScenario3	GW, MSC Value used aquifer, TDS<=2500, non-residential
GWScenario4	GW, MSC Value used aquifer, TDS>2500, non-residential
GWScenario5	GW, MSC Value non-used aquifer, residential
GWScenario6	GW, MSC Value non-used aquifer, non-residential
SoilScenario1	Soil, MSC Value used aquifer, TDS<=2500, residential
SoilScenario2	Soil, MSC Value used aquifer, TDS>2500, residential
SoilScenario3	Soil, MSC Value used aquifer, TDS<=2500, non-residential
SoilScenario4	Soil, MSC Value used aquifer, TDS>2500, non-residential
SoilScenario5	Soil, MSC Value non-used aquifer, residential
SoilScenario6	Soil, MSC Value non-used aquifer, non-residential

Note: all values no soil buffer, no equivalency
(Walker, 2000)

Table 4.8 - Medium Specific Concentration Table Design

Field Name	Field Description
MSCScenario_ID	Unique identifier for the selection of the MSC
EnvironmentalMedium	Designation for the environmental medium for which the MSC applies
COC_ID	Chemical Abstracts registry number
PathwayType	Identifier for MSC values (e.g., soil to groundwater, soil ingestion)
SoilDepthRange	Descriptor for the soil samples included from look-up table
Value	Medium specific concentration value
Units	Look up table for concentration units

Note: Shaded fields are key fields

The specific fields necessary for the PrelimPathway table and the associated support tables (i.e., MediumSpecificConc table, Screening Value Lookup table, Soil Depth Lookup table) will depend on the regulatory structure that is applicable to the facility for which the SERA is implemented. The customization of the SERA methodology is discussed in Section 5.5.

4.3.3.2 Pathway Table

The Pathway table is the focal point of the site conceptual model database. The design of the Pathway table is included in Table 4.9. The key fields for the table are Location_ID, COC_ID, Pathway_ID, and the model input identifiers. There are six model input variable identifiers to identify the parameters needed for the exposure pathway calculations, which are AboveGrd, ChemVar, Constants, MdlErrVar, SubSurVar, and SWVar.

Table 4.9 - Pathway Table Design

Field Name	**Field Description**
Location_ID	Describes the source location
COC_ID	Defines the COC for this pathway
Pathway_ID	Unique identifier for the exposure pathway
AboveGrd	Modeling scenario identifier for aboveground parameters
ChemVar	Modeling scenario identifier for chemical properties
Constants	Modeling scenario identifier for constants
MdlErrVar	Modeling scenario identifier for model error parameters
SubSurVar	Modeling scenario identifier for subsurface parameters
SWVar	Modeling scenario identifier for aboveground parameters
Group_ID	Key field for combined pathways; used in probabilistic calculations
GroupOrderNo	Identifier that places the pathway in the group calculations
Exposure_Mechanism	Selection parameter for the protective exposure concentration
Pathway_Type	Sorting value for the possible pathways
End_Location_ID	Defines the point of exposure or POD
ExpTimingType	Sorting value to identify current or future pathways
Pathway_Status	Describes the status for the pathway
Basis_In	Description field for inclusion of the pathway
Basis_In_Date	Date that the pathway was identified for evaluation
Basis_Out	Description of the information used to exclude the pathway from further evaluation
Basis_Out_Date	Date that the pathway was removed from further evaluation
Verified_By	QC field that documents review of all pathway analyses

Notes: Shaded fields are key fields

The Pathway table defines the type of exposure pathway, including whether the exposure pathway is considered to be an exposure pathway under current land use or future land use, the starting and ending locations, and the basis for identifying the exposure pathway. The Pathway table is used to track the status of the exposure pathway analysis. In addition, the Group_ID is included in the Pathway table to organize the collection of exposure pathways that contribute to the concentration of a chemical of concern at a point of demonstration when the calculations are being completed for the probabilistic case (see section 3.3.2).

As new potential exposure pathways are identified, new records are generated in the Pathway table. As discussed for the PrelimPathway table, the table is designed to track when exposure pathways are identified for inclusion in the exposure analysis and the reasoning for their inclusion and also when an exposure pathway has been eliminated from further consideration. An exposure pathway would have a Pathway_Status of *PotentialPathway* until the evaluations have been completed. The results of the evaluations are recorded in the Pathway table as *CompletePathway*, requiring further actions, *IncompletePathway*, eliminated from further consideration. In addition, any of the fields can be collected into a query and a report table produced to document the exposure pathways considered for the facility.

4.3.3.3 Pathway_Segments Table

The exposure pathway segment table supports the calculation of predicted environmental concentrations at the points of demonstration by dividing the individual exposure pathways into segments that correspond to the fate and transport algorithms. The design of the Pathway_Segments table is included in Table 4.10. The key fields for the table are Location_ID, COC_ID, Pathway_ID, the model input identifiers and the Segment_ID.

Table 4.10 - Pathway Segment Table Design

Field Name	**Field Description**
Location_ID	Describes the source location
COC_ID	Defines the COC for this pathway
Pathway_ID	Unique identifier for the exposure pathway
AboveGrd	Modeling scenario identifier for aboveground parameters
ChemVar	Modeling scenario identifier for chemical properties
Constants	Modeling scenario identifier for constants
MdlErrVar	Modeling scenario identifier for model error parameters
SubSurVar	Modeling scenario identifier for subsurface parameters
SWVar	Modeling scenario identifier for aboveground parameters
Segment_ID	Unique ID for each segment in a pathway
Start_Location_ID	Identifier for the segment starting location
End_Location_ID	Identifier for the segment ending location
Media	Environmental medium for given COC and pathway
Attenuation_Mechanism	Identifies which calculation segment this corresponds to
Segment_Before	Defines the connection between the segments to make-up a pathway
Multiple_Segments_Before	Identifies if the segment has more than one segment before it in the pathway
Segment_After	Defines the connection between the segments to make-up a pathway
Contributions_to_Cstart	Documentation of multiple contributions
Reference	Discussion of the inclusion of the pathway segment and reference to any reports

Note: Shaded fields are key fields

The Attenuation_Mechanism (A_M) field is the identifier that matches the segment record with the applicable spreadsheet calculation to determine the segment ending concentration. The Segment_Before field is given one of two values. If the segment starts at the source area, then the value is *Start*. If the segment is an intermediate segment (i.e., it begins at a transition point), then the unique identifier of the segment before the current segment is entered in the Segement_Before field. If there are multiple segments before this segment, then the Multi_Segment_Before field is used to place a *Yes* flag in that field. The Segment_After field is used to identify if a segment ends at a point of demonstration, then the value is *End*. Otherwise, if a segment is an intermediate segment, then the unique identifier of the segment after is entered into the Segment_After field. The Segment_After field is used in the collection of multiple segments contributing to the concentrations at a single transition point or point of demonstration.

When the Pathway and Pathway_Segments tables are populated it is based on best professional judgment about the exposure pathways that should be evaluated separately and those that act together and must be evaluated at a common point of demonstration. The SERA methodology is not designed to be an expert system to regulate what exposure pathways can or should be evaluated together. As discussed in Section 4.4, the segment divisions are defined by the calculation elements. Additional segments may be needed to describe the particular exposure pathways associated with an application of the SERA methodology. The procedures included here, and in the Appendices, provide the guidance to develop these additional calculations within the same structure. In addition, owing to the group evaluation structure for the probabilistic case, the exposure pathways are evaluated in a more rigid format, therefore the segments represent the entire exposure pathway, (i.e., the points at which exposure pathways are combined are pre-defined in the calculations). For the deterministic case, exposure pathways may be combined at any transition point that is appropriate and the calculation system will support the evaluation of the exposure pathways in that manner. For additional discussions of the calculations see section 4.4.

4.3.3.4 Chemical of Concern ID Table

The COC_ID table is designed to record the list of chemicals of concern that have been considered for the facility. The key field for the table is COC_ID. The design of the COC_ID table is included in Table 4.11.

Table 4.11 - COC_ID Table Design

Field Name	**Field Description**
COC_ID	Chemical abstracts number for the chemical
Chemical_Name	IUPAC name
COC_Status	Identifier that applies to the chemical across the facility, yes/no/potential
Basis_In	Description of reason for inclusion of the chemical
Basis_In_Date	Date that the chemical was added to the COC list
Basis_Out	Description of the reason for eliminating the chemical from further consideration
Basis_Out_Date	Date that the chemical was eliminated from further evaluation
Verified_By	QC field that documents review of all the COC's

Note: Shaded field is key field

The table is designed to track when chemicals are identified for inclusion in the risk-based evaluation and the reasoning for their inclusion and also when a chemical has been eliminated from further consideration at the facility. A chemical would continue to have a COC_Status of *Yes* or *Potential* until it has been fully evaluated across the facility. Once the COC_Status has changed to *No*, then the chemical is no longer a consideration anywhere on the facility.

4.3.3.5 Source Variables Table

The source variables table is used to characterize the source area for both the deterministic and the probabilistic calculations. The design of the Source Variables table is included in Table 4.12.

Table 4.12 - Source Variables Table Design

Field Name	Field Description
Location_ID	Unique location identifier
COC_ID	Chemical at source (multiple COC's for each source and matrix)
Source_Area_Description	Description and location of source at the facility
E(Depth)	Groundwater source area depth perpendicular to GW flow direction
Var(Depth)	Variance of the groundwater source area depth
D(Depth)	High-end value of the source area depth, perpendicular to gw flow (deterministic calculations)
E(Ls)	Depth to top of subsurface sources
Var(Ls)	Variance in the depth to subsurface sources
D(Ls)	High-end value of the depth to subsurface sources (deterministic calculations)
E(Lsout)	Depth to subsurface sources - outdoor air calculations
Var(Lsout)	Variance in the depth to subsurface sources - outdoor air calculations
D(Lsout)	High-end value for the depth to subsurface sources - outdoor air (deterministic calculations)
Source_area_dim_Units	Source area dimension units (all are lengths)
E(NMF)	For NAPL sources, the estimated mole fraction in the NAPL of the given COC
Var(NMF)	Variance for NMF
D(NMF)	High-end value of the NMF (deterministic calculations)
F-NMF	Flag value for NAPL-defined sources --- value is either 0 or 1
E(Cgw)	Estimated groundwater concentration representative of the source
Var(Cgw)	Variance in the groundwater concentration
D(Cgw)	High-end value for the groundwater concentration (deterministic calculations)
F-Cgw	Flag value for groundwater -defined sources --- value is either 0 or 1
E(Cvap)	Estimated vapor concentration representative of the source
Var(Cvap)	Variance in the vapor concentration
D(Cgw)	High-end value for the vapor concentration (deterministic calculations)
F-Cvap	Flag value for vapor concentration-defined sources --- value is either 0 or 1
Var(Owidth)	Variance of the source area width, perpendicular to wind direction
D(Owidth)	High-end value of the source area width, perpendicular to wind (deterministic calculations)
Var(sDist)	Variance of the groundwater travel distance, expected value from GIS
D(sDist)	High-end value of the groundwater travel distance (deterministic calculations)
Var(Width)	Variance in the source area width, perpendicular to gw flow
D(Width)	High-end value of the source area width, perpendicular to gw flow (deterministic calculations)
DataDerivation	Hyperlink to spreadsheet for calculations
Reference	Document source of information included in the source area characterization

The source areas are defined by a number of descriptive parameters. Most of these are used in the segment calculations. The table includes a Description field for the source area identified by the Location_ID. The Location_ID and the COC_ID are the key field fields for this table. They relate the information in this table to the Pathway table and also to the GIS representations of the source areas.

The table includes a reference memo field that can be used to describe the derivation of the source areas. A data description field is also included that provides a hyperlink to a spreadsheet file that includes the calculations used to derive the input values.

The descriptive parameters are defined by three input values. The first is the expected value, or mean value, for the parameter (e.g., E(Depth), the Depth parameter is used to describe the source for the groundwater transport calculation). This is the value used in the probabilistic calculations to calculate the expected concentration value at the point of demonstration. The second value is the variance for the parameter (e.g., Var(Depth)). This value is used in the probabilistic calculations to determine the variance in the concentration at the point of demonstration. The third value is the conservative estimate (i.e., a conservative parameter value that will provide a high estimate of the ending concentration) of the input value (e.g., D(Depth)). This value is used in the deterministic calculation of the concentration at the point of demonstration. The input variables are described along with the algorithms in Appendix A. The derivations of the input values for the case study facility are presented in Appendix B.

4.3.3.6 Subsurface Variables Table

The subsurface variables table is one of the nine model input parameter tables. The values included in this table are used to describe the groundwater and soil at a facility for the fate and transport models. The key field for this table is the SubScenario_ID. It is related to the SubSurVar field in the Pathway table. The design of the Subsurface Variables table is included in Table 4.13.

Table 4.13 - Subsurface Input Variables Table Design

Field Name	**Field Description**
SubScenario_ID	Identifier for the specific group of input values
Description	Description of the group of input values
E(GWGrad)	Expected value for the groundwater gradient
Var(GWGrad)	Variance for the groundwater gradient
D(GWGrad)	High-end value for the groundwater gradient (deterministic calculations)
E(HydCond)	Expected value for hydraulic conductivity
Var(HydCond	Variance in the hydraulic conductivity
D(HydCond)	High-end value of the hydraulic conductivity (deterministic calculations)
Conductivity _Units	Units for hydraulic conductivity
E(TPor)	Expected value for total porosity in saturated zone
Var(TPor)	Variance in the total porosity in the saturated zone
D(TPor)	High-end value for the total porosity in the saturated zone (deterministic calculations)
E(PorVad)	Expected value for the total porosity in vadose zone
Var(PorVad)	Variance in the total porosity in vadose zone
D(PorVad)	High-end value of porosity in vadose zone (deterministic calculations)
E(WCVad)	Expected value for volumetric water content in vadose zone
Var(WCVad)	Variance in the volumetric water content in vadose zone
D(WCVad)	High-end value for the water content in vadose zone (deterministic calculations)
E(SatFoc)	Expected value of the fraction organic carbon in saturated zone
Var(SatFoc)	Variance in the fraction organic carbon in saturated zone
D(SatFoc)	High-end value for the fraction organic carbon in sat zone (deterministic calculations)
DataDerivation	Hyperlink to spreadsheet for calculations

Note: Shaded field is key field

For any particular area of a facility it is expected that there is a set of parameters that describe the subsurface conditions. It is also possible that a generalized set of input parameters might be developed and applied to the entire facility, and as the investigation process continues and more data are collected these values would be refined. Having the input parameters defined with scenario identifiers (e.g., SubScenario_ID) allows for multiple modeling scenarios to be implemented across the facility or even for any individual source area for comparison of results. In addition, the identity of the model values is stored throughout the evaluation process so that the history of the parameter values may be reconstructed at any point.

4.3.3.7 Subsurface Correlations Table

The subsurface correlation table is one of five tables of correlation coefficients. The key field for this table is the SubScenario_ID. It relates the fields in this table to the Subsurface Variables table. The design of the Subsurface Correlations table is included in Table 4.14.

The correlation values are used in the probabilistic calculations. The model input variables are separated into tables that contain only variables that are physically related (e.g., total porosity and groundwater gradient), the correlation coefficients tables correspond to these specific groups of variables. There are no correlations between variables in the separate input tables. The groups that are stored in the separate tables (e.g., SubSurVariables, SWVariables) constitute independent variable sets. In addition, within the potentially correlated parameter sets, correlation coefficients are included only for those parameters that are part of the same group calculations. As an example the vadose zone water-filled porosity is not used along with the groundwater gradient in any of the group calculations, so there is no correlation coefficient value for these two subsurface variables. Therefore, if the attenuation mechanism algorithms are modified, or other groups of pathway segments are constructed, the correlation tables must be altered, as appropriate. Additional discussion of the probabilistic calculation case is included in section 4.4.2. The procedures used to determine the correlation coefficients for the case study facility example calculations are included in Appendix B.

4.3.3.8 Protective Exposure Concentrations Table

The protective exposure concentration table is used to compare calculated concentrations at the points of demonstration (i.e., the ending concentrations from the exposure pathway calculations) to the appropriate target levels. The key fields for this table are the COC_ID and the Exposure_Mechanism. The design of this table is included in Table 4.15.

Table 4.14 - Subsurface Input Variable Correlation Coefficients Table Design

Field Name	**Field Description**
SubScenario_ID	Identifier for the specific group of input values
R-GWGrad/HydCond	Correlation coefficient for gw gradient and sat zone hydraulic conductivity
R-GWGrad/TPor	Correlation coefficient for gw gradient and sat zone total porosity
R-GWGrad/SatFoc	Correlation coefficient for gw gradient and sat zone frac organic carbon
R-HydCond/TPor	Correlation coefficient for hydraulic conductivity and total porosity
R-HydCond/SatFoc	Correlation coefficient for hydraulic conductivity and sat zone frac organic carbon
R-TPor/SatFoc	Correlation coefficient for total porosity and sat zone frac organic carbon
R-PorVad/WCVad	Correlation coefficient for vadose zone total porosity and water content

Note: Shaded field is key field

Table 4.15 - Protective Exposure Concentrations Table Design

Field Name	**Field Description**
COC_ID	Chemical abstracts registry number
Exposure_Mechanism	Identifiers used to differentiate between values (e.g., groundwater ingestion, ecological value)
Media	Environmental medium for which the value is applicable
Values	Concentration target level to which predicted concentrations are compared
Value_Units	Look up table value for concentration units
Comments	Memo field to include whether the value is a site-specific risk-based value, etc.
Reference	Memo field for calculation or regulatory section reference

Note: Shaded fields are the key fields

The target levels may be calculated on a site-specific basis from dose-response equations, they may be state-defined risk-based values or they may be regulatory values (see section 3.4). Using the key fields, the correct target levels are matched-up with each exposure pathway record to determine if the ending concentration is above or below a protective concentration. If the concentrations are below the protective exposure point concentrations then the exposure pathway is considered to be incomplete. If the concentrations are at or above the protective exposure concentrations then the exposure pathway is considered to be complete under the modeling assumptions used in the specific scenario. All of the exposure pathway completeness evaluations are stored in the site conceptual model database. Table 4.16 is an example query results table in the database that compares the groundwater transport segment concentrations to a protective concentration value. The result of the comparison is the Complete? field that compares the E(Cb) concentration (i.e., the surface water expected concentration) to the Value (i.e., protective exposure concentration) field.

Table 4.16 - Example Exposure Pathway Completeness Comparison

UniqueID	Exposure_Mechanism	COC_ID	E(Cb)	Values	Value_Units	Complete?
SA171-43-2P20C1CN1M1SS1SW1S1	Surface Water Value	71-43-2	3.10E-12	0.001	mg/L	N
SA271-43-2P20C1CN1M1SS1SW1S1	Surface Water Value	71-43-2	7.15E-24	0.001	mg/L	N
SA371-43-2P20C1CN1M1SS1SW1S1	Surface Water Value	71-43-2	2.92E-34	0.001	mg/L	N
SA571-43-2P20C1CN1M1SS1SW1S1	Surface Water Value	71-43-2	2.33E-40	0.001	mg/L	N
SA671-43-2P20C1CN1M1SS1SW1S1	Surface Water Value	71-43-2	3.40E-35	0.001	mg/L	N

The next section describes the implementation of the attenuation mechanism calculations based on the data developed both in the spatial database and the tabular database.

4.4 MODELING FATE AND TRANSPORT ALONG AN EXPOSURE PATHWAY USING ANALYTICAL MODELS DEVELOPED IN SPREADSHEETS

The Spatial Environmental Risk Assessment (SERA) methodology implementation uses spreadsheets that have been developed for each attenuation mechanism. There are six mechanisms available in the current versions of the spreadsheets for the deterministic calculations. Table 3.1 includes the list of mechanisms. The details of the models are included in Appendix A. The calculations are based on SI units and the spreadsheets are not designed to handle other unit systems, so the values from spatial database and the values entered in the site conceptual model tabular database must conform to the SI units specified. The length measurements from the GIS files are in English units. The conversion to SI units is handled in the model variable queries before the data are passed to the spreadsheets. The Units_LookUp_Table in the site conceptual model database is provided to help ensure that the proper units are used for each input parameter.

In the probabilistic calculation there are three types of exposure pathways included in the spreadsheet calculations. These are composites of the attenuation mechanisms included in the deterministic case and they are:

1. NAPL partitioning to groundwater with groundwater transport to surface water and surface water mixing for a final ending concentration.

2. NAPL partitioning to soil vapor, vapor transport in the vadose zone and outdoor air mixing for a final ending concentration.
3. NAPL partitioning to soil vapor, vapor transport in the vadose zone and through a building foundation to indoor air mixing for a final ending concentration.

Each spreadsheet is developed with row-wise entries that are iterations of a single calculation. This mimics the database architecture for data records and allows for direct connection between the database and the spreadsheet. The serial number and the input data are accommodated in the initial columns of the spreadsheet. These fields are populated with data records through the **Get External Data** commands in Excel. An Access database driver is established that allows records from the database to be read into the spreadsheet. The database records are developed for each calculation segment through a series of queries that collect all of the records for a specific attenuation mechanism and the specific input data in the columnar order expected in the spreadsheet. An example query is included in Table 4.17.

Table 4.17 - Example Query Design for the Groundwater Transport Segment

<table>
<tr><th colspan="2">Tables included in the Groundwater Transport Query</th></tr>
<tr><td colspan="2">ChemConstants</td></tr>
<tr><td colspan="2">ChemVariables</td></tr>
<tr><td colspan="2">COC_ID</td></tr>
<tr><td colspan="2">Gwtran1.dbf</td></tr>
<tr><td colspan="2">MdlErrorVariables</td></tr>
<tr><td colspan="2">Pathway</td></tr>
<tr><td colspan="2">Pathway Segments</td></tr>
<tr><td colspan="2">Physical Constants</td></tr>
<tr><td colspan="2">Segment Begin Union</td></tr>
<tr><td colspan="2">Source Variables</td></tr>
<tr><td colspan="2">Srcarea2.dbf</td></tr>
<tr><td colspan="2">SubSurVariables</td></tr>
<tr><td colspan="2">UniqueIDBuilder</td></tr>
<tr><th colspan="2">Variable Selections</th></tr>
<tr><th>Variable</th><th>From Table:</th></tr>
<tr><td>UniqueID</td><td>UniqueIDBuilder</td></tr>
<tr><td>D(SatFoc)</td><td>SubSurVariables</td></tr>
<tr><td>D(GWGrad)</td><td>SubSurVariables</td></tr>
<tr><td>D(HydCond)</td><td>SubSurVariables</td></tr>
<tr><td>ParDen</td><td>Physical Constants</td></tr>
<tr><td>D(TPor)</td><td>SubSurVariables</td></tr>
<tr><td>D(Koc)</td><td>ChemVariables</td></tr>
<tr><td>D(Degrad)</td><td>ChemVariables</td></tr>
<tr><td>D(Width)</td><td>Srcarea2.dbf</td></tr>
<tr><td>D(Depth)</td><td>Source Variables</td></tr>
<tr><td>EyGW</td><td>MdlErrorVariables</td></tr>
<tr><td>NyGW</td><td>MdlErrorVariables</td></tr>
<tr><td>CA</td><td>Segment Begin Union</td></tr>
<tr><td>sDist</td><td>Gwtran1.dbf</td></tr>
</table>

The number of calculation records does not need to be known in advance because the formulae to the right of the data columns are automatically copied down when the data are linked into the spreadsheet. Following the columns for the input parameters are the columns containing the formulae for the calculated values. The formulae used in the SERA are written as **User-Defined Functions** in an Excel **Add-In Visual Basic Application**. The segment ending concentration is generally the last column in the sheet. A summary results sheet accompanies each calculation sheet. On the results sheet the serial number (i.e., pathway segment unique identifier) and the ending concentrations appear. In addition, for the probabilistic case, the derivative values for each of the input variables are included in the results sheet. Table 4.18 includes an example results table from the spreadsheets.

Table 4.18 - Results Table Generated in the Spreadsheet for the Probabilistic Case for the Groundwater to Surface Water Exposure Pathway

UniqueID	E(Ca)	E(Cb)	E(CCgXI)	DCCgXI/dCb	DCCgXI/dNMF
SA171-43-2P20C1CN1M1SS1SW1S1	1.75E+01	3.10E-12	3.67E-13	1.19E-01	3.67E-11
SA271-43-2P20C1CN1M1SS1SW1S1	3.5E+01	7.15E-24	1.06E-24	1.48E-01	5.3E-23
SA371-43-2P20C1CN1M1SS1SW1S1	3.5E+01	2.92E-34	1.24E-34	4.25E-01	6.19E-33
SA571-43-2P20C1CN1M1SS1SW1S1	7.0E+01	2.33E-40	1.73E-41	7.41E-02	4.32E-40
SA671-43-2P20C1CN1M1SS1SW1S1	7.0E+01	3.40E-35	4.37E-36	1.28E-01	1.09E-34

Using the **External Data** link from Access to Excel, the ending concentration can be linked into the database and made available for exposure pathway completeness evaluations or as beginning concentrations for subsequent segments.

Construction of the spreadsheets, the queries to collect the calculation input values, the Visual Basic Add-In, and the database-spreadsheet connections are detailed in Appendix C and Appendix D.

4.4.1 Deterministic Case

In the deterministic case the segments are calculated individually. Each segment type is calculated in a separate sheet that contains the attenuation factor algorithm. The ending concentrations are passed back to the database through the results sheet. In the database, the ending concentration from one segment is matched-up with the record for the subsequent segment and becomes the beginning concentration for the next segment.

When the Segment_After value is the same for two, or more ending concentrations, the concentration values are added, checked against a solubility or saturation value, and then passed on to the next segment as the segment starting concentration.

The process of matching-up the segment ending concentration with the correct segment start concentration is accomplished using standard database data handling operations. Queries are written in the database that assign the ending concentration of one segment to the beginning concentration of the next segment based on the Segment_Before value. The Segment_Before and the

Segment_After values are the concatenated, unique serial numbers for the segments. As discussed in Section 4.3, all of the segment connections are identified when the Pathway_Segments table is populated.

The specific procedures used in the deterministic case including the queries used in the matching process are included in Appendix C.

4.4.2 Probabilistic Case

The probabilistic case is formulated in a more rigid structure than the deterministic case. This limitation was imposed because the ending concentration variance calculation depends on the calculation elements and variables that precede the calculation of the final concentration. In the fixed structure chosen for this application, the exposure pathways are considered to be independent until they reach the point of demonstration. At the point of demonstration the combined mean concentration is calculated and checked against saturated vapor or aqueous solubility and the overall variance is calculated. The three probabilistic calculation groups that were implemented for this research are:

- NAPL partitioning to groundwater, groundwater transport and surface water mixing;
- NAPL partitioning to soil vapor and volatilization to indoor air; and
- NAPL partitioning to soil vapor and volatilization to outdoor air.

As discussed in Appendix B and Chapter 5, these were selected based on their applicability at the case study facility.

The variance is calculated for each exposure pathway in the group based on the independent variance contributions (Var(I)) of all of the variable parameters for all of the contributing pathway segments and the variance contribution due to correlation between variables (Var(Cor)) for all of the variables in the pathway segments. The relationship for the independent variance is:

$$\mathrm{Var(I)} = \sum_{j=1}^{n}\left(\frac{\partial C_E}{\partial X_j}\right)^2 \mathrm{Var}(X_j)$$ **Equation 4.1**

And the relationship for the correlated variance is:

$$\mathrm{Var(Cor)} = 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left(\frac{\partial C_E}{\partial X_i}\right)\left(\frac{\partial C_E}{\partial X_j}\right)\rho_{ij}\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}$$ **Equation 4.2**

These equations are the two parts of the variance calculation presented as Equation 3.11. They are separated and named so that the significance of including correlations (rather than making the assumption that all the variables are independent) can be tabulated and assessed.

The variables for two different exposure pathways in the same group are not explicitly considered to be correlated. This may be a limitation in the uncertainty analysis and is discussed further in Chapter 5. The discussion in Chapter 5 includes a method used to develop preliminary results and assess the magnitude of the variance contribution from correlation across exposure pathways in a group.

For the groundwater to surface water mixing group calculation, there are three variables (i.e., hydraulic conductivity, groundwater gradient, source area width) that are part of the groundwater transport and of the surface water mixing attenuation factor models. They are each assumed to be the same variable that is in the two different segment calculations. If, for example, a different hydraulic conductivity were appropriate at the point of discharge to surface water than the value for the groundwater transport segment, then they would be treated as separate variables and correlation would need to be addressed. In the calculations presented here, the three variables that appear in two different segments in the groundwater to surface water group are assumed to be the same variables, or stated another way, they are assumed to be perfectly correlated. It is important to consider variables within the same parameter set that are used within each group calculation and determine if there are correlations that exist.

For the volatilization to outdoor air and the volatilization to indoor air groups there are no variables that appear in more than one attenuation factor model and no variables from the same parameter set that appear in two different attenuation factor models.

In the probabilistic calculation spreadsheet, it is assumed that there will be up to six exposure pathways that combine to contribute to the ending concentration. This value was chosen to allow a reasonable number of exposure pathways to be combined, while making the spreadsheet development manageable. Additional exposure pathways can be added, though it would require amending the spreadsheets to consider more than six contributions. In addition, the source areas are identified as any combination of up to three source areas defined by a NAPL mole fraction (NMF) and up to three source areas defined by a groundwater concentration (Cgw) or a soil gas concentration (Cvap). These also are artificial constraints in order to implement the calculations. These requirements are easily modified with amendments to the spreadsheets. It should be noted that whenever the spreadsheets are changed, if this produces a change in the required input parameters, then the parameter value tables must be changed and the database queries for parameter gathering must be altered appropriately.

In order to implement the group calculations, a Group Order Number is defined in the Pathway table for each exposure pathway in a group. By convention, the first three exposure pathways are those defined by NAPL source areas, the second three exposure pathways are those defined by environmental medium concentrations. For each exposure pathway there are two *Flag* values that must also be completed. The flag F-NMF (i.e., Flag-NAPL mole fraction) is equal to one if a NAPL source area defines the exposure pathway; otherwise the flag is set to zero. The flag F-Cmedium (e.g., F-Cgw) is equal to one if a concentration-based source area defines the exposure pathway. The flags are used in the spreadsheets to include only the applicable calculation elements for the concentration and variance estimates.

4.4.3 Input Parameter Values

In the site conceptual model database the Variable Description table provides information about all of the input parameters used in the application of the SERA methodology. Table 4.19 shows the table design for the Variable Description table.

Table 4.19 - Variable Description Table Design

Field Name	**Field Description**
Variable	Variable name
Description	Description of the variable
Type	Identifier that indicates to which parameter group the variable belongs
Segments	Pathway segment in which the variable is used
Possible Constant?	Yes/No field used in application development for variables treated as constants
Discussion	Additional discussion field for each variable

For the calculations, most of the input parameters are considered to be random variables. Examples of the random variables are included in Table 4.20. For the random variables, estimates of the mean, variance and deterministic or "conservative" estimate for each variable are needed. As used here, the uncertainty described by the variance value is the total uncertainty in the parameter value. The uncertainty may be due to spatial variability, testing variability, the lack of site-specific data and other sources of variability. The complete listing of the input parameters, the method used to identify the input values and the value estimations for the examples are included in Appendix B.

Table 4.20 - Example Random Variables

Variable	**Description**	**Type**	**Segments**	**Discussion**
AF	Areal fraction of cracks in foundation	AbvGrd	Indoor	Not easily measured; may change with time
Degrad	First order degradation constant	Chem	GW	Not easily measured, site properties affect the value, variability in site properties make it variable across a site
Depth	Source zone depth	Source	GW	Measurement is difficult; changes with time
Ls	Depth to subsurface vapor sources	SubSur	Indoor	Estimated value based on site measurements; can change with time

The correlation coefficients are developed based on whether or not there is a relationship between two variables, whether the relationship is expected to be strong or weak, and whether the relationship is positive or negative. The values assigned are from a discrete value function. The discussion of parameter correlation is included in Appendix B.

Some of the input parameters are treated as point values only. The values must be specified, but for the variance calculations they are considered to be constants. Examples of the parameters that are treated as constants are included in Table 4.21.

As a practical matter, the parameters that are treated as constants are such that they are not expected to exhibit much variability based on conditions at the facility. For these parameters, reference values were identified and used in the calculations. Because classifying some of the input parameters as constants simplified the calculation process, and by doing so the results of the calculations were not expected to be significantly affected, the assumption that some of the input values could be treated as constants rather than as random variables was made. By reducing the number of variables it is not expected that the overall variance is reduced rather, that the small variances associated with the parameters treated as constants are accounted for in the variance of the other variables, including the model error variables.

It would be possible, if the variation of a parameter were important in a particular facility application of the spatial environmental risk assessment (SERA) methodology, to convert a value from a constant to a variable within the calculation procedures. The parameter would be moved into the appropriate variable group table in the database, values for the mean and variance would be developed, and the appropriate partial derivative functions developed and added to the variance calculation spreadsheet. In addition, correlation values and relationships would be identified for the parameter within its variable group.

The complete listing of the constants and the value estimations for the examples are included in Appendix B.

Table 4.21 - Example Parameters Treated as Constants

Constant	Description	Type	Segments	Discussion
ER	Indoor air exchange rate	AbvGrd	Indoor	Design parameter for an existing building or a hypothetical building
Dliq	Diffusion coefficient for the chemical in water	Chem	Indoor, Outdoor	Molecular behavior property. Variability is mostly due to laboratory measurement error
ParDen	Soil particle density	SubSur	GW	Constant value for soil grains - by convention
MixHt	Mixing height above ground surface for outdoor air exposure	AbvGrd	Outdoor	Value is estimated based on the simplified box model for the receptor breathing zone height

4.5 SUMMARY

This chapter presents the application procedure for the Spatial Environmental Risk Assessment methodology. A brief overview of the software tools used in this research is included. Then, the application procedure is presented in detail based on its fundamental components: the spatial database; the tabular database; and, the fate and transport algorithms implemented in spreadsheets. Detailed algorithms, case study input values and computer procedures are presented in the Appendices A through D. The next chapter presents the Results and Discussion based on the implementation of the SERA methodology for the examples at the case study facility.

Chapter 5 : Results and Discussion

This chapter presents the discussion and evaluation for the development of the Spatial Environmental Risk Assessment (SERA) methodology, the results of the examples developed for the case study facility, and the benefits and application of the SERA methodology to other facilities.

Risk assessment involves the evaluation of the spatial and temporal relationships between the sources of chemicals of concern and the potential receptors. Where there are many sources and potential receptors, as is the case for many large, complex facilities, this evaluation presents a very difficult and time-consuming task. To address this difficulty, the SERA methodology has been developed to represent, catalog and manage the spatial relationships between multiple sources and receptors and to facilitate the evaluation of exposure pathway completeness. While the spatial risk assessment methodology could be applied to, and benefit facilities with, any number of sources and potential receptors, the significant benefit of this method, as compared to the standard risk assessment practice of individual calculations for a subset of exposure pathways, will be realized at large, complex facilities with many exposure pathways.

This research endeavored to develop better systems for risk assessment that account for the complexities of many sources and receptors. As is presented in this chapter, the spatial environmental risk assessment methodology using GIS as the central integrating element developed by this research accomplishes this objective. The research has realized the improvements identified in Table 1.1, through the implementation of innovations in site conceptual models, representing uncertainty and meta data management to support risk-based decisions in environmental management. The types of analyses presented in these results are typical of an initial site-specific risk assessment. An initial site-specific risk assessment is also known as a *tier 2 risk assessment* in the ASTM risk-based corrective action guides (ASTM, 1995a, 2000). Certainly, these are only the beginning steps and discussion of the some of the additional research that could be undertaken is presented in Chapter 6.

5.1 SPATIAL AND TABULAR DATABASE MODEL FOR RISK ASSESSMENT

This section discusses the data model developed in this research for using spatial and tabular databases in risk assessment. The data model, as presented in Figure 5.1, is used as the template for application of the SERA methodology.

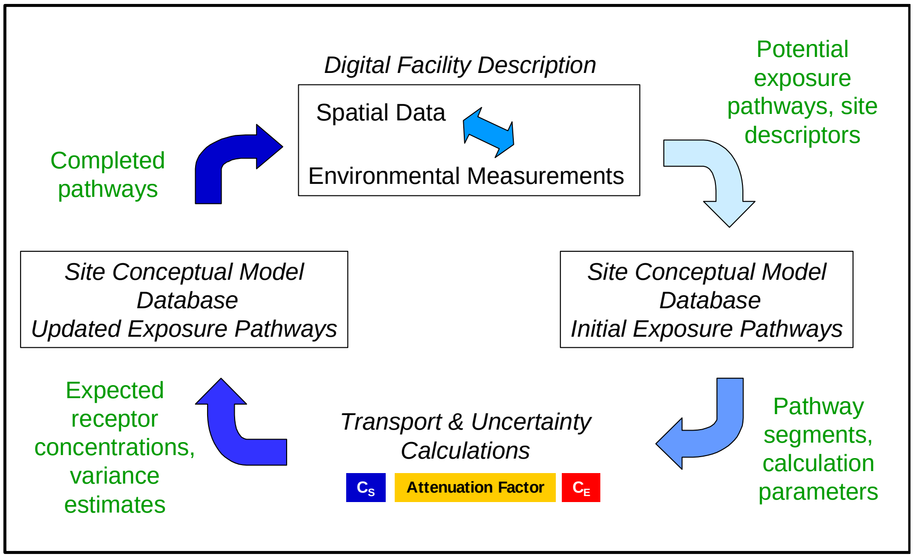


Figure 5.1 - Data Model for the Spatial Environmental Risk Assessment Methodology

The spatial environmental risk assessment methodology uses the spatial and tabular databases to provide a common framework to spatially and relationally organize the physical and hydrogeological properties of a facility and the surrounding area, the chemical properties of materials released and the temporally-defined data related to concentrations of chemicals of concern in environmental media and hydrogeological properties. A fundamental component, and a significant improvement in the SERA methodology over current methods for risk assessment, is the spatial site conceptual model. The spatial site conceptual model with its unique identifiers can be used to manage the exposure pathways related to many sources and many land use (e.g., current, future) configurations. An important feature of the spatial site conceptual model is the connection of the exposure pathway identifiers and elements to the fate and transport calculations for the exposure pathway evaluation process. In addition, the flexibility of the procedure to analyze large numbers of exposure pathways supports facility-wide risk assessment, as compared to the current methods, which focus on separately evaluating individual regulated units.

In the database model for risk assessment, as discussed in Chapter 3, all of the elements of the exposure pathways are included as spatial objects in the GIS shapefiles and are attributed with their important characteristics. The characteristics of the source areas, the receptor locations and the potential transport mechanisms are managed as attribute data in the tabular database. In addition, the modeling input data for the fate and transport calculations associated with each exposure pathway are identified as attributes of that exposure pathway.

The system of unique identifiers for each exposure pathway supports the generation of a tabular database record for each exposure pathway. The unique identifiers for the source areas and the points of demonstration allow the generation of spatial records for each of these entities.

Cataloging all of the exposure pathways and all of the source areas evaluated for a facility provides the mechanism by which the thoroughness of the risk assessment can be identified. Capturing the modeling data in model scenario tables provides the mechanism by which the repeatability of the risk assessment can be tested and the reliability proven. Inclusion of simple uncertainty analysis procedures facilitates the evaluation of the reliability of the results and alternatives analysis for further data collection, risk evaluation or remedial action. These are features that are unique to the SERA methodology and to the underlying spatial and tabular database foundation of the method.

The methodology provides a mechanism for the engineers and scientists conducting risk assessments to visualize the available information through the GIS representation of the exposure assessment data and information and calculation results, and the uncertainties associated with that information. It also provides a mechanism for the other interested parties to better understand the risk assessment calculations and the results. The visual representations of the site conceptual model were presented in Chapter 4. Additional discussions of the visualization benefits of the methodology are presented in section 5.3.

The next section presents a number of examples from the case study facility to demonstrate the implementation of the methodology and to present examples of the pathway completeness evaluations and the types of results that are to be expected from implementing the SERA methodology.

5.2 PRESENTATION OF EXAMPLES FROM THE CASE STUDY FACILITY

The application procedure presented in Chapter 4 was used to evaluate a number of example exposure pathways at the Former Lube Plant of the Marcus Hook Refinery. The details of the exposure pathway development are included in Appendix B. In addition, the input parameter estimation methodology and values are included in Appendix B.

For the case study, the Pennsylvania Act 2 process was used as the basis for the development of the example exposure pathways. The discussion in section 5.5 presents alternatives for the more general application to other facilities. Medium Specific Concentrations for chemicals of concern are defined in Act 2 as statewide standards applicable to all facilities. These Medium Specific Concentrations are related to general exposure pathways (e.g., direct contact with soils, ingestion of groundwater, protection of surface water). As a group, the Medium Specific Concentrations can be thought of as the target levels defined for a "state-defined" or generic site conceptual model. As the screening level risk assessment, the measured concentrations of chemicals of concern in environmental media at a facility are compared to the Medium Specific Concentrations. This comparison is defined in the SERA methodology as the Preliminary Exposure Pathway Evaluation. Under Act 2, if all of the concentrations of the chemicals of concern in the environmental media are not below the Medium Specific Concentrations, then the responsible party has the opportunity to conduct a risk assessment and develop site-specific standards, or the responsible party may choose to use the Medium Specific Concentrations as remedial action target levels (Commonwealth of Pennsylvania, 1997, PaDEP, 1997). If a risk assessment is to be conducted (and for the examples developed here this is assumed to be the case), then a site-specific, site conceptual model is developed. The outcome of the risk assessment is to determine whether any of the exposure pathways are complete and to determine what the appropriate target levels would be for the facility. In the examples presented here, the preliminary exposure pathway evaluation is discussed, the evaluation of the site-specific exposure pathways is demonstrated and several iterations of the calculations are presented to determine which exposure pathways are complete. In addition, a discussion of the options for the complete exposure pathways is presented.

5.2.1 Example Preliminary Exposure Pathway Evaluation

The preliminary exposure pathway evaluation was conducted for several areas of soil concentrations using the Pennsylvania Medium Specific Concentrations. The soil analytical results for a number of example chemicals of concern were compiled in the environmental measurements database using queries. The results were compared to the Pennsylvania Medium Specific Concentration values for direct contact in a non-residential scenario. All of the concentrations were below the direct contact values for benzene, toluene, ethylbenzene, xylenes, anthracene, benzo(a)pyrene, and chrysene. The soil analytical results were then compared to the soil leaching to groundwater Medium Specific Concentrations. The values used in this example are based on a used aquifer, total dissolved solids less than 2500 ppm and non-residential land use. In the site conceptual model database this is identified as Soil Scenario 3. The values are based on 100 times the groundwater Medium Specific Concentration.

The results were linked to ArcView and event themes were constructed based on the sample result locations. The event themes were converted to shapefiles. Romanek *et al*., (1999) provides detailed instructions for constructing event themes. The data were classified based on the individual chemical Medium Specific Concentrations. Figure 5.2 shows the results for benzene (71-43-2), anthracene (120-12-7), benzo(a)pyrene (50-32-8) and chrysene (218-01-9). All of the yellow symbols (i.e., lighter shading) represent sample values that are below the respective Medium Specific Concentration. The respective Medium Specific Concentrations are given in the legend for Figure 5.2. All of the red symbols (i.e., darker shading) represent sample values that are above the respective Medium Specific Concentrations. For this example, all of the values that were equal to the analytical detection limits were taken at their detection limits because in a number of cases the detection limit was above the applicable Medium Specific Concentration. The chrysene results are given as triangles, the anthracene results are given as pentagons, the benzene results are given as circles, and the benzo(a)pyrene results are given as squares.

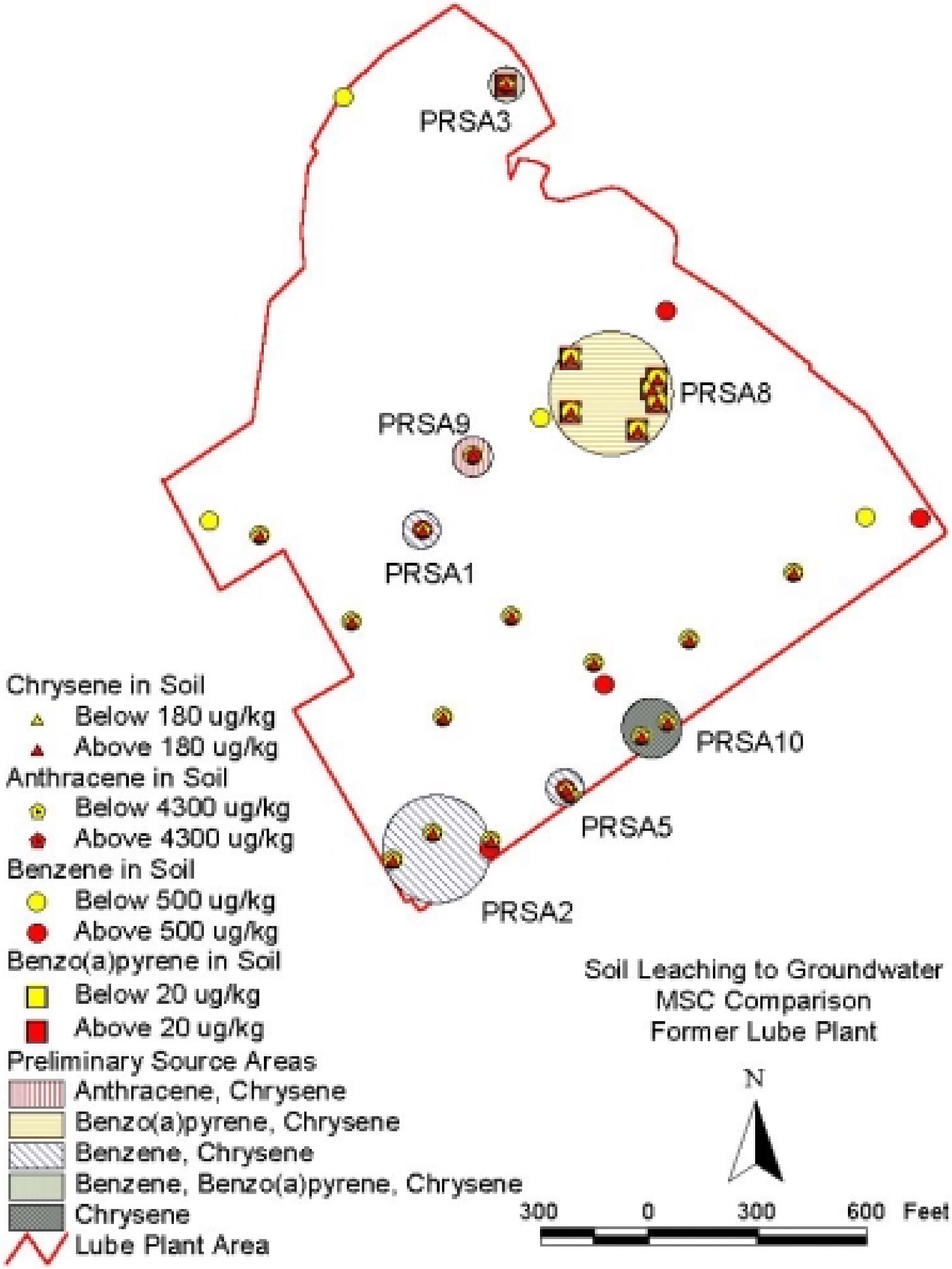


Figure 5.2 - Preliminary Exposure Pathway Evaluation

These data were reviewed and preliminary source areas identified, based on the locations of concentrations of the chemicals that are above their respective Medium Specific Concentrations. A polygon shapefile was constructed with seven example preliminary source areas. These are shown in Figure 5.2. The examples chosen for display are those where there are a number of soil samples together for one chemical of concern that are above the Medium Specific Concentration or there were a number of chemicals of concern at one location that had concentrations above their respective Medium Specific Concentrations. The preliminary source areas are shaded with different colors and patterns, indicating the chemicals of concern that are associated with each area.

Data records were generated in the site conceptual model database in the PrelimPathway table. The representative concentration, in this case the maximum concentration value for each preliminary source area, was recorded in the Preliminary exposure pathway record. The PrelimPathway table includes a reference field for the concentration data that were used to generate the representative concentrations. In this case, the shapefile tables generated from the environmental measurements database are referenced. Using the MSC Lookup Table and the PrelimPathway table, a query table was generated to document the comparison of the representative concentrations to the Medium Specific Concentrations. Table 5.1 includes these results.

Table 5.1- Results of the Comparison of Representative Soil Concentrations for Preliminary Source Areas to the Medium Specific Concentrations

LOC_ID	COC_ID	Chemical	Representative Concentration	MSC Value	Units
PRSA1	218-01-9	Chrysene	0.43	0.18	mg/kg
PRSA1	71-43-2	Benzene	0.64	0.50	mg/kg
PRSA10	218-01-9	Chrysene	2.00	0.18	mg/kg
PRSA2	218-01-9	Chrysene	1.20	0.18	mg/kg
PRSA2	71-43-2	Benzene	1.00	0.50	mg/kg
PRSA3	218-01-9	Chrysene	0.50	0.18	mg/kg
PRSA3	50-32-8	Benzo(a) pyrene	0.50	0.02	mg/kg
PRSA3	71-43-2	Benzene	1.17	0.50	mg/kg
PRSA5	218-01-9	Chrysene	4.20	0.18	mg/kg
PRSA5	71-43-2	Benzene	8.60	0.50	mg/kg
PRSA8	218-01-9	Chrysene	0.75	0.18	mg/kg
PRSA8	50-32-8	Benzo(a) pyrene	0.55	0.02	mg/kg
PRSA9	120-12-7	Anthracene	8.90	4.30	mg/kg
PRSA9	218-01-9	Chrysene	3.90	0.18	mg/kg

Notes: All preliminary source areas are for subsurface soils
All shaded concentration values are above the MSC values.
MSC - Pennsylvania Medium Specific Concentration values
Soil leaching to groundwater values

These soil source areas would be carried forward to the Pathway table for further analyses of the soil leaching to groundwater exposure pathways. In addition, records could be generated in the PrelimPathway table to document the comparisons based on concentrations that were below the Medium Specific Concentrations. The preliminary exposure pathway evaluation is a method to track the comparisons to the statewide standards at the beginning of the risk evaluation.

The additional discussions within section 5.2 are related to another group of source areas. These source areas were identified based on the detection of non-aqueous phase liquids (NAPL) on the water table.

5.2.2 Example Exposure Pathways

Based on the data that were available, a number of non-aqueous phase liquid (NAPL) source areas were identified for the example evaluations. Not all of the source areas were identified, only a representative group in order to demonstrate the application of the SERA methodology. Once the source areas were defined, the possible chemicals of concern were determined from the available data. It must be emphasized that there were limited historical, and no recent composition data available for the NAPL detected at the facility. As discussed in Appendix B, reasonable assumptions were made about the mole fractions of the chemicals of concern in the NAPL in each source area. Additional data collection would be needed to refine these conservative, first assumptions about the source concentrations. However, the estimates that were made likely over-predict the concentrations of the chemicals of concern for the source areas considered. Using these conservative assumptions, it is possible to begin the exposure pathway evaluation process and use the results of the calculations to tailor the additional data collection activities. In addition, the conservative assumptions were useful in the research because they ensured that a number of the ending concentrations for different exposure pathways were non-zero, and were close to the Protective Exposure Concentrations. These were important attributes for the research in order to investigate the various data evaluations and presentations discussed here. Section 5.2.9 includes a discussion of the comparison of the field measured concentration data to the model parameters and discusses areas for further data collection.

In order to define the exposure pathways, the possible transport mechanisms were identified, and then the points of demonstration were determined based on the locations of potential receptors. This information was used to develop the site conceptual model. To be clear, not all of the possible exposure pathways for the Former Lube Plant were identified, only a representative group in order to demonstrate the application of the SERA methodology. The NAPL source areas were chosen because they represent obvious areas that should be addressed in the risk assessment. Other areas of soil and groundwater concentrations should be considered to be source areas, however, these NAPL areas are likely among the most important areas to be addressed.

The exposure pathways evaluated in these examples include seven groundwater transport to surface water exposure pathways, three volatilization to indoor air exposure pathways and three volatilization to outdoor air exposure pathways. Table 5.2 summarizes the exposure pathways included in these examples.

Table 5.2 - Example Exposure Pathways

Source Area	Exposure Pathways
SA4	Groundwater transport to Marcus Hook Creek
SA6	Groundwater transport to Delaware River Volatilization to indoor air
SA7	Groundwater transport to Delaware River Volatilization to outdoor air

SA8	Groundwater transport to Marcus Hook Creek Volatilization to outdoor air
SA9	Groundwater transport to Delaware River Volatilization to indoor air
SA10	Groundwater transport to Delaware River Volatilization to indoor air
SA11	Groundwater transport to Delaware River Volatilization to outdoor air

The chemicals of concern identified for the NAPL source areas in the Former Lube Plant are benzene, ethylbenzene, naphthalene, toluene and total xylenes. These chemicals are components of gasoline, diesel fuel and other petroleum products handled in the Former Lube Plant area. Appendix B provides further discussion of the exposure pathways and the chemicals of concern selected for each source area.

The spatial representations of the exposure pathway elements were developed based on the information in the digital facility description. These data include the environmental measurements database that was updated in May 2000 with field sampling and analytical data collected from September 1999 to March 2000 in the Former Lube Plant. Figure 5.3 shows the locations of the source areas, the groundwater transport segments and the points of demonstration used in the initial deterministic calculations. The site conceptual model database was populated with the exposure pathway data and the parameter values are discussed in Appendix B.

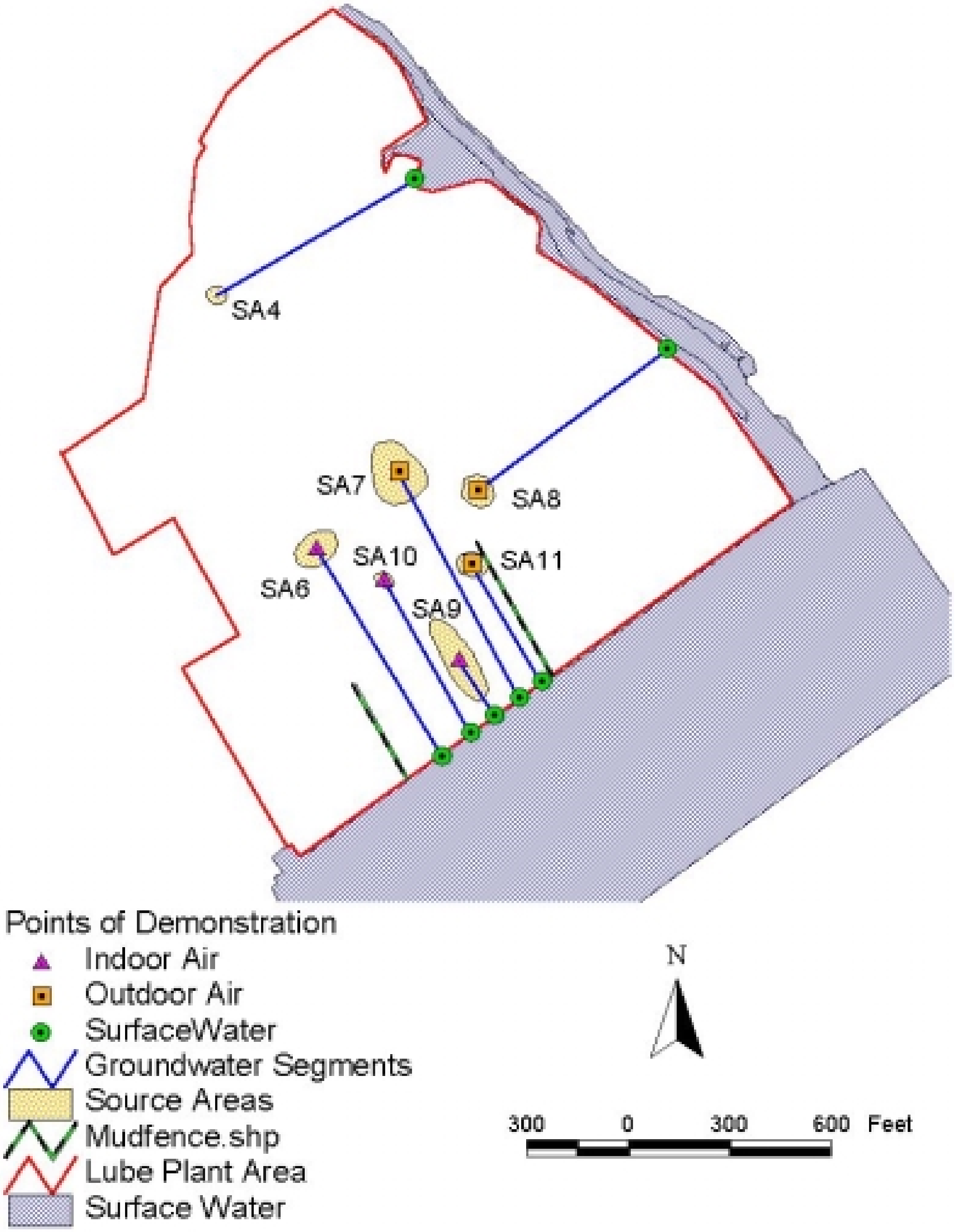


Figure 5.3 - Example Exposure Pathways and Points of Demonstration

In order to verify that all of the needed records were developed for the exposure pathways, a procedure was developed to calculate the number of exposure pathway records and pathway segment records that correspond to the number of source areas and chemicals of concern. It was determined that as the site conceptual model gets larger this procedure would be necessary as a quality control check on the exposure pathway completeness evaluation.

The number of exposure pathway records (PN) can be calculated using the relationship in Equation 5.1, based on the number of:

- source areas (SN),
- chemicals of concern (COCN),
- modeling scenarios (MN),
- exposure pathway types (e.g., volatilization to indoor air, groundwater transport to surface water) (PTN),
- points of demonstration for each pathway type (e.g., location at the creek bank, location in the surface water downstream of the contributing source areas) (PODN), and
- land use scenarios (LN).

$$\mathrm{PN} = \sum_{i=1}^{\mathrm{SN}} \left(\mathrm{LN}_i \times \sum_{j=1}^{\mathrm{PTN}} \left(\mathrm{COCN}_j \times \mathrm{MN}_j \times \mathrm{PODN}_j \right) \right)$$

Equation 5.1

Table 5.3 includes the variable values for each of the example source areas.

Table 5.3 - Case Study Exposure Pathway Records

Source Area	SA4	SA6	SA7	SA8	SA9	SA10	SA11
LN	1	1	1	1	1	1	1
PTN	1	2	2	2	2	2	2
COCN	1	4	4	5	5	5	5
MN	1	1	1	1	1	1	1
PODN	1	1	1	1	1	1	1
Totals	1	8	8	10	10	10	10

For the case study the initial number of pathway records is 57.

The number of exposure pathway segment records (PSN) can be calculated using the relationship in Equation 5.2, based on the number of exposure pathways (PN) and the number of segments for each pathway type (SegN):

$$\mathrm{PSN} = \sum_{i=1}^{\mathrm{SN}} \left(\mathrm{LN}_i \times \sum_{j=1}^{\mathrm{PTN}} \left(\mathrm{COCN}_j \times \mathrm{MN}_j \times \mathrm{PODN}_j \times \mathrm{SegN}_j \right) \right) \qquad \textbf{Equation 5.2}$$

Table 5.4 includes the variable values for each of the example source areas and pathway types.

Table 5.4 - Case Study Pathway Segment Records

Source Area	SA4	SA6	SA7	SA8	SA9	SA10	SA11
LN	1	1	1	1	1	1	1
Number of Segments for each Pathway Type							
GWSW	2	2	2	2	2	2	2
Indoor	0	2	0	0	2	2	0

Air							
Outdoor Air	0	0	2	2	0	0	2
Sum	2	4	4	4	4	4	4
COCN	1	4	4	5	5	5	5
MN	1	1	1	1	1	1	1
PODN	1	1	1	1	1	1	1
Totals	2	16	16	20	20	20	20

For the case study, the initial number of pathway segment records is 114.

5.2.3 Initial Transport Segment Calculation Results

The transport segment calculations were implemented for all of the pathway segments and a predicted concentration at each point of demonstration was calculated for each chemical of concern. The predicted concentrations at the points of demonstration are given in Table 5.5.

Table 5.5 - Predicted Concentrations at the Points of Demonstration

Source Area - POD	**Benzene**	**Ethylbenzene**	**Naphthalene**	**Toluene**	**Total Xylenes**
Exposure Pathway: Groundwater Transport					
Groundwater Concentrations in mg/L	71-43-2	100-41-4	91-20-3	108-88-3	1330-20-7
Protective Exposure Concentrations	**0.001**	**0.58**	**0.01**	**0.33**	**0.21**
SA4 - POD5	NA	NA	3.29E-04	NA	NA
SA6 - POD2	0.06	4.82E-03	NA	0.09	0.03
SA7 - POD12	0.12	9.25E-03	NA	0.16	0.06
SA8 - POD3	0.07	5.22E-03	6.84E-04	0.09	0.03
SA9 - POD14	0.69	5.19E-02	6.80E-03	0.92	0.32
SA10 - POD13	0.03	2.29E-03	3.00E-04	0.04	0.01
SA11 - POD16	0.15	1.16E-02	1.52E-03	0.21	0.07
Exposure Pathway: Volatilization to Indoor Air					
Indoor Air Concentrations in mg/m^3	71-43-2	100-41-4	91-20-3	108-88-3	1330-20-7
Protective Exposure Concentrations	**0.0002**	**1.49**	**0.085**	**0.594**	**1.04**
SA6 - POD20	0.92	0.07	NA	1.39	0.37
SA9 - POD21	1.79	0.15	0.010	2.72	0.72
SA10 - POD22	1.47	0.12	0.008	2.23	0.60
Exposure Pathway: Volatilization to Outdoor Air					
Outdoor Air Concentrations in mg/m^3	71-43-2	100-41-4	91-20-3	108-88-3	1330-20-7
Protective Exposure Concentrations	**0.0007**	**4.47**	**0.085**	**1.78**	**3.13**
SA8 - POD19	0.033	0.003	1.81E-04	0.050	0.013
SA7 - POD17	0.034	0.003	NA	0.052	0.014
SA11 - POD18	0.034	0.003	1.87E-04	0.051	0.014

Note: NA indicates that the chemical was not a COC for that source area
Shaded values indicate concentrations above the protective exposure concentrations.

In general, the predicted concentrations at the points of demonstration are compared to the protective exposure concentrations using queries in the database. The values in Table 5.5 were compared to the protective exposure concentrations identified for each exposure pathway type. Table 5.6 provides information on the protective exposure concentrations, including the references and the values for each chemical of concern.

Table 5.6 - Protective Exposure Concentrations

COC	COC_ID	Exposure_Mechanism	Values	Units	Reference
Ethylbenzene	100-41-4	Surface Water Value	0.58	mg/L	1
	100-41-4	Vapor Inhalation Indoor	1.49	mg/m^3	2
	100-41-4	Vapor Inhalation Outdoor	4.47	mg/m^3	2
Toluene	108-88-3	Surface Water Value	0.33	mg/L	1
	108-88-3	Vapor Inhalation Indoor	0.594	mg/m^3	2
	108-88-3	Vapor Inhalation Outdoor	1.78	mg/m^3	2
Total Xylenes	1330-20-7	Surface Water Value	0.21	mg/L	1
	1330-20-7	Vapor Inhalation Indoor	1.04	mg/m^3	2
	1330-20-7	Vapor Inhalation Outdoor	3.13	mg/m^3	2
Benzene	71-43-2	Surface Water Value	0.001	mg/L	1
	71-43-2	Vapor Inhalation Indoor	0.0002	mg/m^3	2
	71-43-2	Vapor Inhalation Outdoor	0.0007	mg/m^3	2
Naphthalene	91-20-3	Surface Water Value	0.01	mg/L	1
	91-20-3	Vapor Inhalation Indoor	0.085	mg/m^3	3
	91-20-3	Vapor Inhalation Outdoor	0.085	mg/m^3	3

Notes:
References
1 - PASWQS, Chpt 16:Toxics
2 - Risk-based Value, BP 1997
3 - Vapor Pressure, BP, 1997
The vapor pressure value is used for naphthalene (91-20-3) in air because the risk-based values are calculated to be higher than the vapor pressure.

Based on the initial comparison of the predicted concentrations to the protective exposure concentrations, there are 8 complete groundwater (i.e., the 6 exposure pathways for benzene, one for toluene and one for xylenes, 6 complete indoor air (i.e., the 3 exposure pathways for each benzene and toluene), and 3 complete outdoor air (i.e., the 3 exposure pathways for benzene) exposure pathways. As discussed in section 3.4, if the predicted concentrations are below the protective exposure concentrations, then the exposure pathway is considered to be incomplete. Complete exposure pathways require further consideration.

Figure 5.4 shows the pathway completeness results for the groundwater transport pathways and Figure 5.5 shows the pathway completeness results for the indoor and outdoor air pathways. These are the results given in Table 5.5. An important result of the initial calculations is the reduction from 57 potentially complete exposure pathways to 17 complete exposure pathways for further evaluation, based on the screening calculations. This allows the investigator to refine input estimates or modeling procedures for a smaller number of exposure pathways, while those exposure pathways that are eliminated are eliminated based on modeled concentrations that are expected to over-predict the actual concentrations reaching the points of demonstration, based on the conservative models and input values chosen.

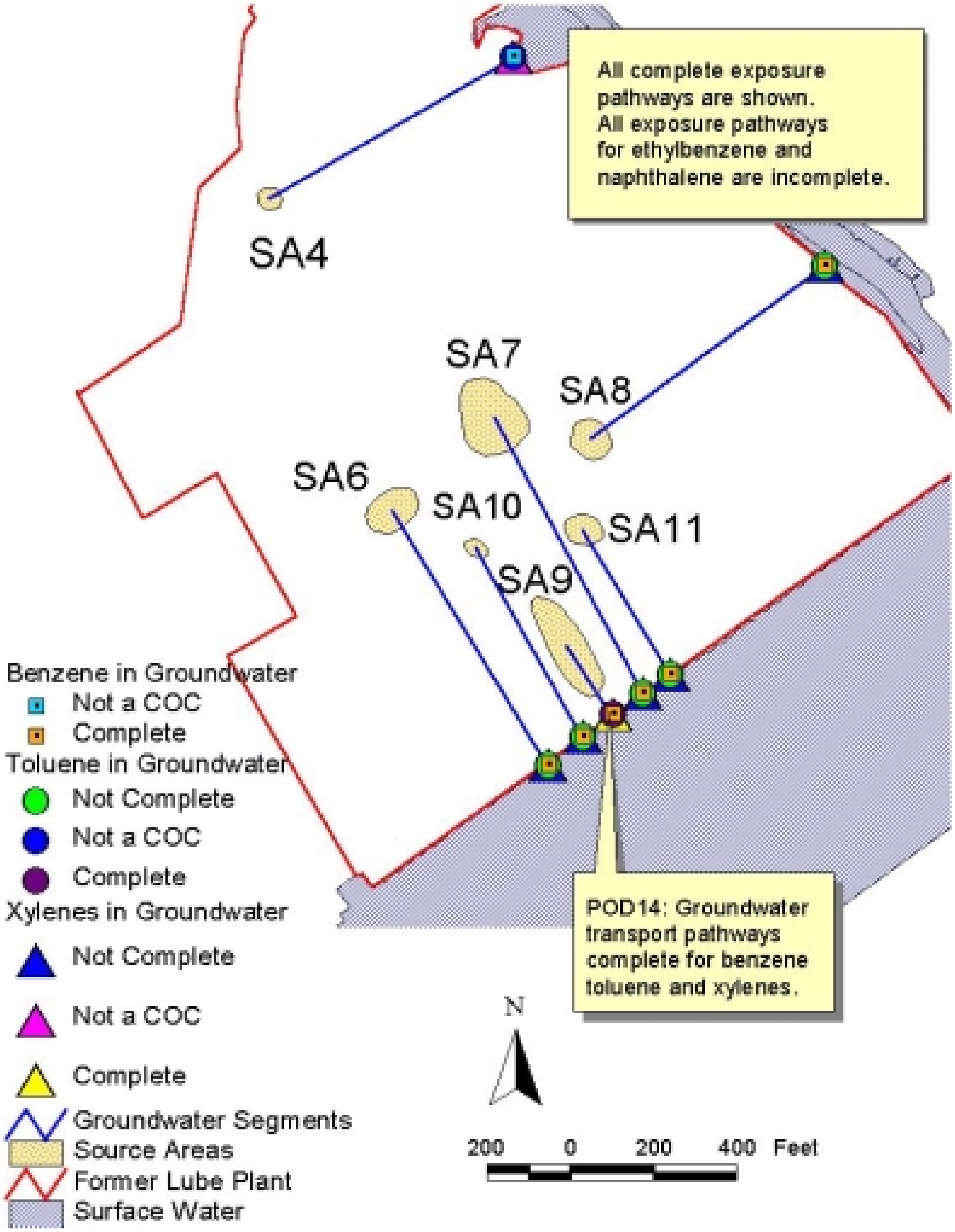


Figure 5.4 - Complete Groundwater Transport Exposure Pathways

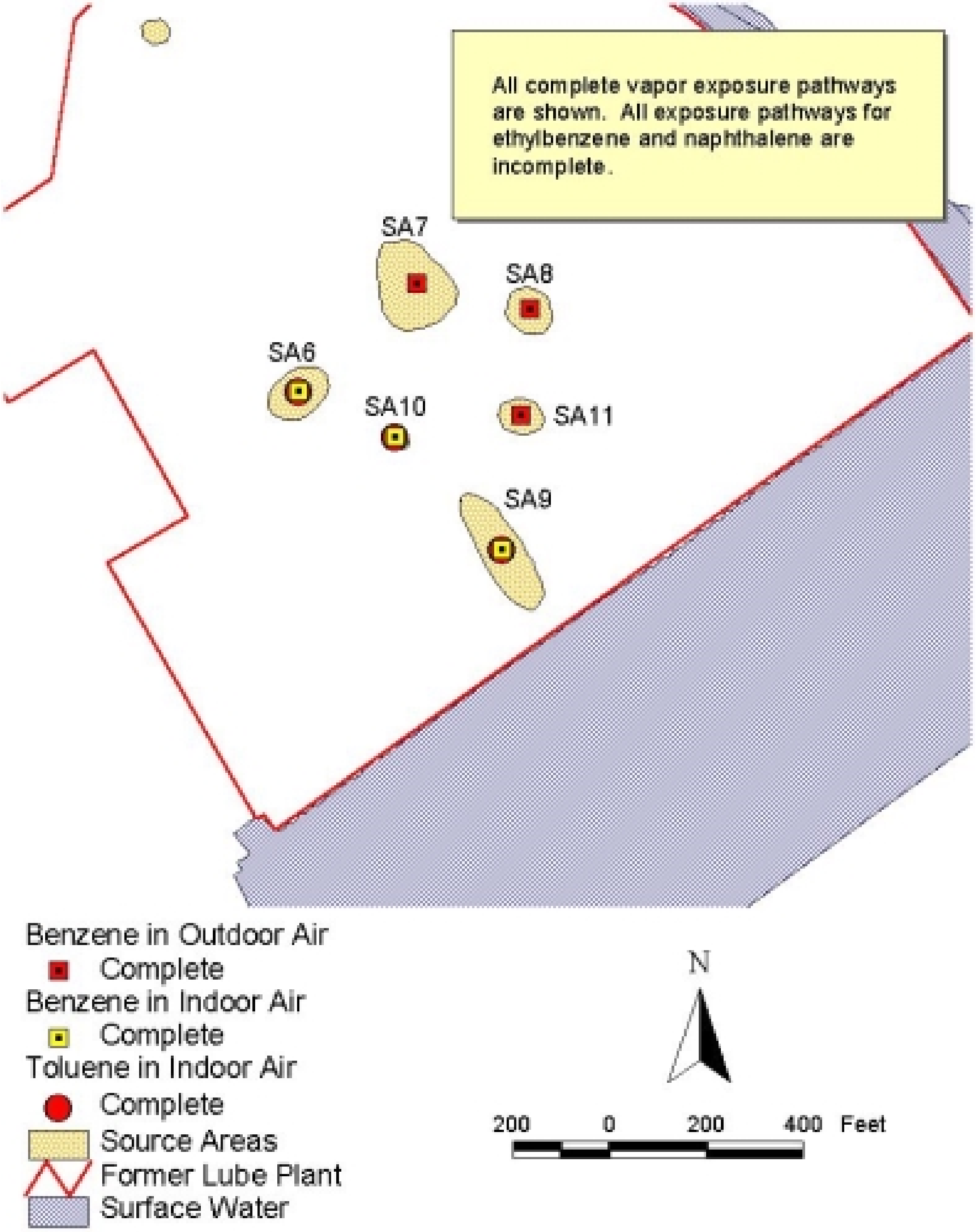


Figure 5.5 - Complete Indoor and Outdoor Air Exposure Pathways

5.2.4 Discussion of Initial Transport Segment Results

For the individual groundwater transport exposure pathways the calculations were completed to the point of demonstration just at the location where the groundwater discharges to surface water. This point of exposure might be important in an ecological exposure evaluation or if the regulatory requirements included a demonstration of attainment of target levels at the property boundary or at a surface water. It was used here to demonstrate the range of calculations that might be required and the versatility of the method application. In the following section the surface water mixing model is used to evaluate the combined effects of the several source areas affecting the same surface water body. In the surface water mixing calculations, the points of demonstration are selected within the Delaware River and Marcus Hook Creek just down-stream of the groundwater discharge locations. Within the Pennsylvania DEP requirements, these are the more likely locations where attainment would be demonstrated.

It should also be noted that for the deterministic groundwater calculations the first order decay rates for each chemical were set to zero in order to demonstrate concentrations greater than zero reaching the groundwater - surface water interface. In the probabilistic calculations, presented in section 5.2.6, the first order decay rates have non-zero mean and variance values.

The calculation algorithms for the indoor air and the outdoor air are based on the conservative (i.e., often over-predicting the concentration of the chemical of concern in the vapor phase within a building or in outdoor air) Johnson and

Ettinger (1991) diffusion model (API, 1998). The model assumes there is no biodegradation of the vapors in the vadose zone. The complete exposure pathways might require additional evaluation before a remedial action decision would be made based on the screening calculation results.

For the benzene and toluene exposure pathways, additional field work might include a determination of the soil properties, including soil moisture content, at the various horizons in the subsurface and a layered implementation of the same attenuation factor model could be made.

In addition, installation of vapor monitoring probes to measure soil gas concentrations of benzene and toluene above the various source areas could be implemented. These measurements would be made at several horizons in the vadose zone to determine the vertical profile of the chemical concentrations. In this way, a site-specific attenuation factor can be developed. Further, API (1998) recommends evaluating the oxygen and carbon dioxide profiles in the vadose zone to develop a more site-specific understanding of the chemical attenuation.

The additional field data would also facilitate an evaluation of the applicability of the vapor transport models as discussed in section 5.2.9.

5.2.5 Additional Transport Segment Calculation Results

In order to evaluate the combined effects of multiple source areas on single points of demonstration, exposure pathway groups were identified. Each exposure pathway group consists of source areas with the same exposure pathway that are all expected to impact the same point of demonstration. Table 5.7 includes the definition of the exposure pathway groups and Figure 5.6 shows these exposure pathway groups. The groundwater to surface water exposure pathway for SA8 for benzene was considered alone, since it was the only source area with benzene to Marcus Hook Creek, with the surface water mixing attenuation factor. The predicted surface water concentration (1.4 E-04 mg/L) is below the protective exposure concentration (1 E-03 mg/L), so the exposure pathway is considered to be incomplete and no additional evaluations were conducted.

Table 5.7 - Exposure Pathway Groups

Group Number: COC_ID	Source Areas	Exposure Pathway	Point of Demonstration
G1:71-43-2 G2:100-41-4 G3:108-88-3 G4:1330-20-7 G5:91-20-3	SA6, SA7, SA9, SA10, SA11	Groundwater to surface water mixing	POD25
G6:71-43-2 G7:100-41-4 G8:108-88-3 G9:1330-20-7 G10:91-20-3	SA6, SA9, SA10	Volatilization to Indoor Air	POD22
G11:71-43-2 G12:100-41-4 G13:108-88-3 G14:1330-20-7 G15:91-20-3	SA7, SA8, SA11	Volatilization to Outdoor Air	POD19
G16	SA4, SA8 - Naphthalene only	Groundwater to surface water mixing	POD28

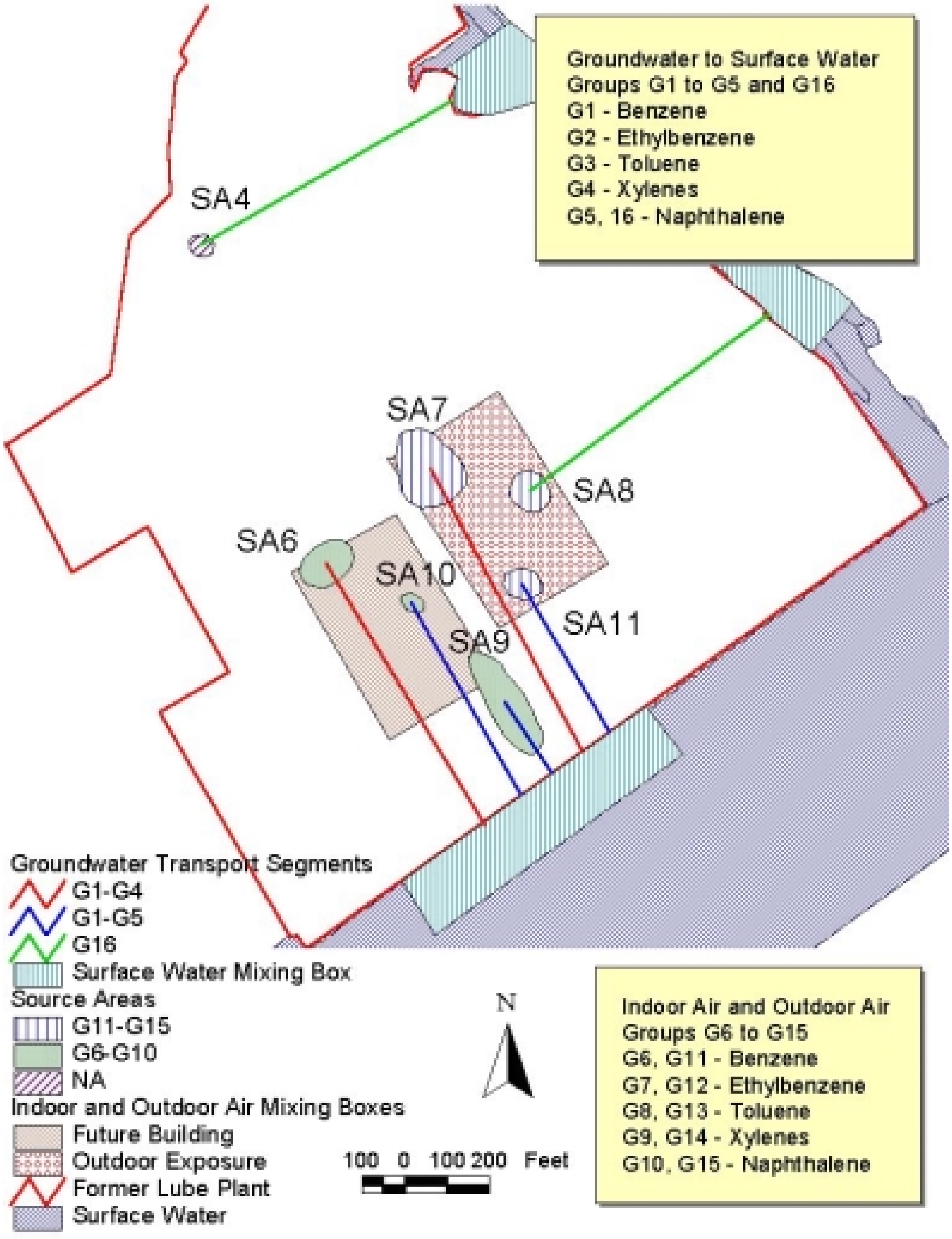


Figure 5.6 - Exposure Pathway Groups

As illustrated in Figure 3.6 and Figure 5.6, the NAPL partitioning and the groundwater transport for the individual source areas are calculated separately. The surface water mixing model is implemented for the combined concentration from the groundwater transport segments and the concentration in surface water is predicted for each chemical of concern.

The group exposure pathway records were generated by entering data in the Pathway and Pathway Segments tables in the site conceptual model database and the Segment_After field was used to identify the pathway segments that belong to each group. These groups were also used in the example probabilistic calculations discussed in section 5.2.6. The resulting number of pathway records for the case study is 110, since two additional exposure pathway records (i.e., one for groundwater and one for soil vapor) were entered for each source area, (except SA4 which only has groundwater exposure pathways and SA8 only the exposure pathway for naphthalene to the creek was included in the groups). In order to calculate the number of pathway segments in the situation where there are groups of source areas affecting one point of demonstration, Equation 5.2 is modified. The number of additional segments includes the number of group segments (GN) for each chemical of concern. Equation 5.3 is the more general form of the equation for the number of pathway segment records.

$$\begin{aligned} \mathrm{PSN} = \sum_{i=1}^{\mathrm{SN}} \left[(\mathrm{LN}_i) \times \sum_{j=1}^{\mathrm{PTN}} \left(\mathrm{COCN}_j \times \mathrm{MN}_j \times \mathrm{PODN}_j \times \mathrm{SegN}_j \right) \right] \\ + \sum_{k=1}^{\mathrm{GN}} \mathrm{COCN}_k \end{aligned}$$

Equation 5.3

For the case study, including the grouped exposure pathways, the number of pathway segment records is 236.

Table 5.8 summarizes the predicted concentration results for the exposure pathway groups.

Table 5.8 - Predicted Concentrations at the Points of Demonstration

Exposure Pathway Group	**Benzene**	**Ethylbenzene**	**Naphthalene**	**Toluene**	**Total Xylenes**
Exposure Pathway: Groundwater Transport and Surface Water Mixing					
Surface Water Concentrations in mg/L	71-43-2	100-41-4	91-20-3	108-88-3	1330-20-7
Protective Exposure Concentrations	**0.001**	**0.58**	**0.01**	**0.33**	**0.21**
POD25(G1-G5)	7.7E-05	5.8E-06	2.3E-05	1.0E-04	1.5E-05
POD28 (G16)	1.4E-04[1]	NA	2.0E-06	NA	NA
Exposure Pathway: Volatilization to Indoor Air					
Indoor Air Concentrations in mg/m^3	71-43-2	100-41-4	91-20-3	108-88-3	1330-20-7
Protective Exposure Concentrations	**0.0002**	**1.49**	**0.085**	**0.594**	**1.04**
POD22 (G6-G10)	4.17	0.34	0.02	6.34	1.69
Exposure Pathway: Volatilization to Outdoor Air					
Outdoor Air Concentrations in mg/m^3	71-43-2	100-41-4	91-20-3	108-88-3	1330-20-7
Protective Exposure Concentrations	**0.0007**	**4.47**	**0.085**	**1.78**	**3.13**
POD19 (G11-G15)	0.10	0.01	3.7E-04	0.15	0.04

Notes: Shaded values indicate concentrations above the Protective Exposure Concentrations

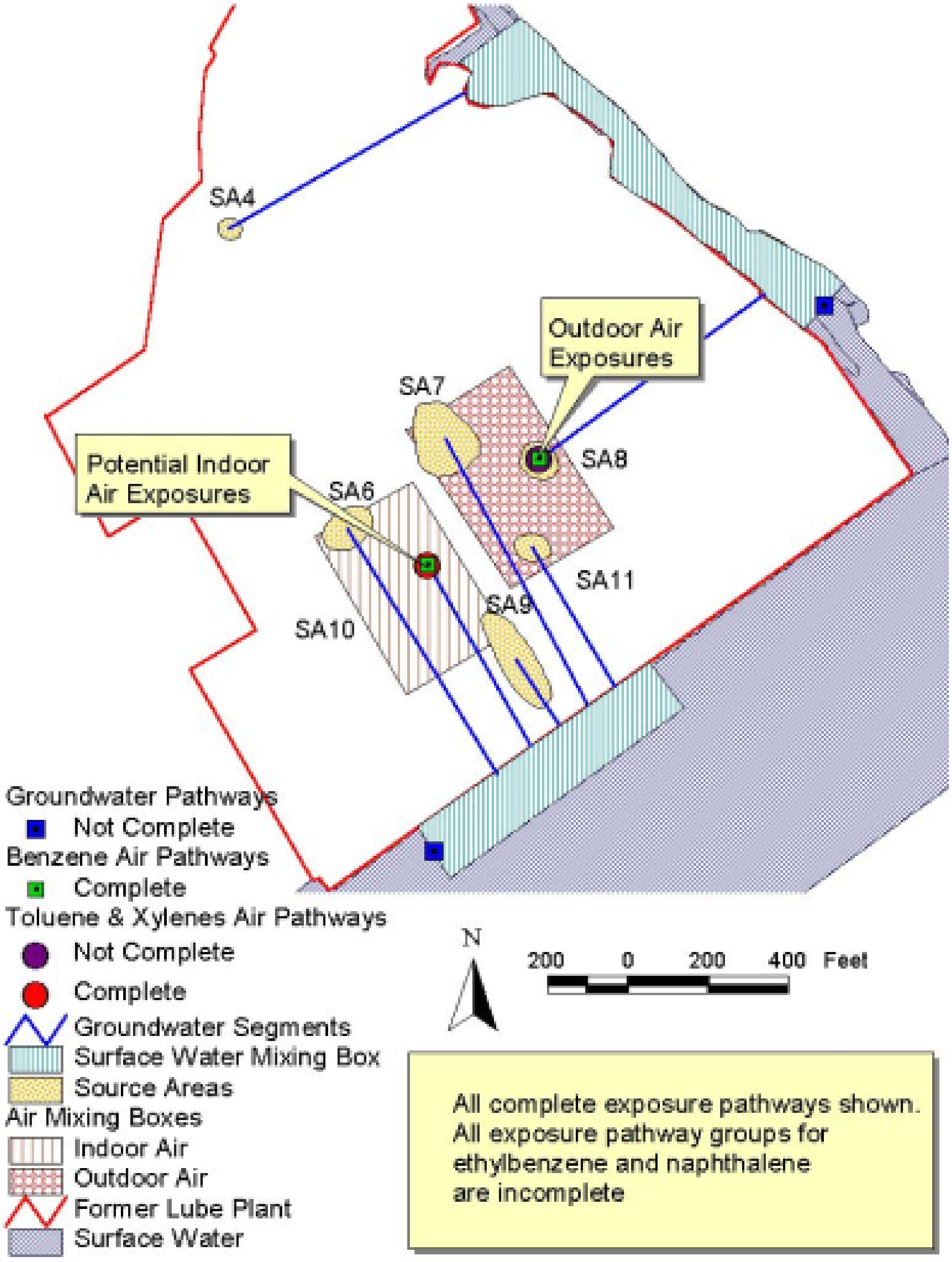
1 - Benzene at POD 28 considered from SA8 individually for surface water mixing

The protective exposure concentrations were again used to evaluate the predicted concentrations. Based on the grouped exposure pathway comparison of the predicted concentrations to the protective exposure concentrations, the groundwater to surface water exposure pathways are not complete. The benzene toluene and xylenes to indoor air and the benzene outdoor air exposure pathways are complete. All of the other vapor exposure pathways are not complete. The results of the additional screening calculations indicate that there are four groups of exposure pathways that require further evaluation. These involve 12 of the original 57 exposure pathways that were considered. Figure 5.7 shows the results of the exposure pathway comparisons. These are also presented in Table 5.8.

As indicated in section 5.2.4, the first order decay rate, describing biodegradation in the groundwater transport pathway was set to zero in these calculations. Particularly in situations where multiple source areas are being evaluated, as is the case in these groups, including biodegradation without field measurements or pilot testing to determine if there is a maximum groundwater concentration above which the biodegradation no longer occurs, or occurs under a process that can no longer be described by a first order process, can involve making assumptions about the site-specific behavior of a very complex process. If there are concerns about the application of the simple first order kinetics, there are several options that can be considered:

- Evaluate the groundwater transport with no degradation (as implemented here for the deterministic calculations).

- Implement the groundwater transport calculations looking at intermediate distances (i.e., insert a transition point between the sources and the point of demonstration and assume impact at that transition point by as many source areas as the circumstances warrant) between the source areas and the point of demonstration to determine if an unacceptable maximum concentration is being predicted. The analysis would be accomplished iteratively, and if an unacceptable maximum were reached, then different decay rate constants could be applied to the segments before and after the transition point.
- Implement a more sophisticated groundwater model that includes a kinetic model that more closes matches the biological activity at the facility, if data are available to support the model.



All complete exposure pathways shown. All exposure pathway groups for ethylbenzene and naphthalene are incomplete

Figure 5.7 - Exposure Pathway Completeness for Grouped Pathways

5.2.6 Probabilistic Calculations

The results of the two phases of calculations performed with the deterministic input values indicate that the indoor air (for benzene, toluene and xylenes) and outdoor air (for benzene) exposure pathways are the only ones that require further consideration. The predicted groundwater concentrations reaching the Delaware River and Marcus Hook Creek are very low compared to the protective exposure concentrations. On this basis, it may be appropriate to eliminate the groundwater exposure pathways from further consideration. For the sake of illustration, and without additional information about the uncertainty in the results all of the grouped exposure pathways and chemicals of concern were used to evaluate the mean and variance data for the predicted concentrations at the points of demonstration.

Table 5.9 includes the results from the probabilistic calculations. The data presented include the expected values, or mean values, the variance values, the standard deviations and the coefficient of variation (c.o.v.) values for the environmental concentrations at the points of demonstration for the same exposure pathway groups discussed in section 5.2.5.

Table 5.9 - Results from Probabilistic Calculations

Exposure Pathway Group	Mean	Variance	S.D.	c.o.v.
Groundwater Transport and Surface Water Mixing				
Surface Water Concentrations in mg/L				
POD25 (G1) Benzene (71-43-2)	1.4E-06	1.4E-10	1.2E-05	8.3
POD25 (G2) Ethylbenzene (100-41-4)	5.3E-16	7.7E-29	8.8E-15	16.6
POD25(G3) Toluene (108-88-3)	1.6E-08	2.8E-14	1.7E-07	10.4
POD25(G4) Xylenes (1330-20-7)	2.6E-10	9.2E-18	3.0E-09	11.8
POD25 (G5) Naphthalene (91-20-3)	2.0E-48	9.7E-93	9.8E-47	50.5
POD28 (G16) Naphthalene (91-20-3)	1.8E-85	1.3E-166	1.1E-83	63.7
Volatilization to Indoor Air				
Indoor Air Concentrations in mg/m^3				
POD22 (G6) Benzene (71-43-2)	1.6E-01	1.8E-01	4.3E-01	2.7
POD22 (G7) Ethylbenzene (100-41-4)	1.3E-02	1.2E-03	3.5E-02	2.7
POD22 (G8) Toluene (108-88-3)	3.1E-01	7.0E-01	8.3E-01	2.7
POD22 (G9) Xylenes (1330-20-7)	6.5E-02	3.0E-02	1.7E-01	2.7
POD22 (G10) Naphthalene (91-20-3)	6.1E-04	4.3E-06	2.1E-03	484.5
Volatilization to Outdoor Air				
Outdoor Air Concentrations in mg/m^3				
POD19 (G11) Benzene (71-43-2)	4.9E-03	2.4E-05	4.9E-03	1.0
POD19 (G12) Ethylbenzene (100-41-4)	4.0E-04	3.2E-06	1.8E-03	4.5
POD19 (G13) Toluene (108-88-3)	9.5E-03	7.7E-05	8.8E-03	0.9
POD19 (G14) Xylenes (1330-20-7)	2.0E-03	6.0E-06	2.4E-03	1.2
POD19 (G15) Naphthalene (91-20-3)	1.8E-05	1.3E-06	1.1E-03	63.2

As indicated by the deterministic calculations, the surface water concentrations resulting from the groundwater discharges are very low concentrations. It is interesting to note that the coefficient of variation values, which are an indication of the whether the dispersion of the values around the mean are significant relative to the mean value, are fairly large for all of the exposure pathways, except for the volatilization to outdoor air exposure pathways. This would indicate, based on the method implemented here, that there is a significant amount of uncertainty in the concentration at the point of demonstration. In addition, when the mean value is very small, as is the case for a number of the exposure pathway groups, the coefficient of variation is often large. An important modifier in the analyses of these results is the approximate nature of the first order uncertainty analysis implemented. Since the transport equations utilized in the exposure pathway calculations are non-linear equations, the first order analysis for the variance is only an approximation, but may provide a reasonable first approximation of the variance in the predicted concentrations.

Often, variance estimates are made based on the assumption that all of the variables are independent. This is not an accurate representation of the relationships between the input parameters for these fate and transport models (Gilbert, 1997). It is interesting to note that the inclusion of the correlation between variables can be significant to the calculations. For the surface water concentration variances, the contribution to the variance due to parameter correlations is a reduction in the variance, which is reported as a negative Var(Cor). For the indoor air concentration variances, the contribution to the variance due to parameter correlations is an increase in the variance, which is reported as a positive Var(Cor). These results indicate that including correlation may provide some reduction in the variance, but may also increase the total variance. The values are dependent on the parameter values and their relationships in the attenuation factor models. Table 5.10 includes the independent variance and the correlated variance values for four of the chemicals of concern in surface water and indoor air.

Table 5.10 - Variance in Surface Water and Indoor Air Concentrations

Group ID	COC	Var(I)	Var(Cor)	Total Variance
Surface Water Concentration Variances (mg/L)^2				
G1	Benzene	1.5E-10	-6.4E-12	1.4E-10
G2	Ethylbenzene	1.5E-28	-7.0E-29	7.7E-29
G3	Toluene	3.9E-14	-1.1E-14	2.8E-14
G4	Xylenes	1.4E-17	-5.0E-18	9.2E-18
Indoor Air Concentration Variances (mg/m^3)^2				
G6	Benzene	1.3E-01	5.7E-02	1.8E-01
G7	Ethylbenzene	8.3E-04	3.8E-04	1.2E-03
G8	Toluene	4.8E-01	2.2E-01	7.0E-01
G9	Xylenes	2.1E-02	9.3E-03	3.0E-02

The results are compared to the protective exposure concentrations and the probability of the ending concentration being below the protective exposure concentration is calculated for each exposure pathway group. The calculation of the probabilities is based on the method presented in Chapter 3. As noted in Chapter 3, a probability of $p = 0.9$ is used as the target probability. This means that we require 90-percent certainty that the predicted value will be below the target level. Because the variables were modeled as normal variates, the ending concentration distributions were also modeled as normal variates. Table 5.11 includes the values for each of the exposure pathway groups.

Table 5.11 - Probabilistic Comparison of Ending Concentration to Target Level

Exposure Pathway Group	Mean	S.D.	PEC [1]	Prob [2]
Groundwater Transport and Surface Water Mixing				
Surface Water Concentrations in mg/L				
POD25 (G1) Benzene (71-43-2)	1.4E-06	1.2E-05	1.0E-03	1.00
POD25 (G2) Ethylbenzene (100-41-4)	5.3E-16	8.8E-15	3.3E-01	1.00
POD25(G3) Toluene (108-88-3)	1.6E-08	1.7E-07	5.8E-01	1.00
POD25(G4) Xylenes (1330-20-7)	2.6E-10	3.0E-09	2.1E-01	1.00
POD25 (G5) Naphthalene (91-20-3)	2.0E-48	9.8E-47	1.0E-02	1.00
POD28 (G16) Naphthalene (91-20-3)	1.8E-85	1.1E-83	1.0E-02	1.00
Volatilization to Indoor Air				
Indoor Air Concentrations in mg/m^3				
POD22 (G6) Benzene (71-43-2)	1.6E-01	4.3E-01	2.0E-04	0.35
POD22 (G7) Ethylbenzene (100-41-4)	1.3E-02	3.5E-02	5.9E-01	1.00
POD22 (G8) Toluene (108-88-3)	3.1E-01	8.3E-01	1.5E+00	0.92
POD22 (G9) Xylenes (1330-20-7)	6.5E-02	1.7E-01	1.0E+00	1.00
POD22 (G10) Naphthalene (91-20-3)	6.1E-04	2.1E-03	8.5E-02	1.00
Volatilization to Outdoor Air				
Outdoor Air Concentrations in mg/m^3				
POD19 (G11) Benzene (71-43-2)	4.9E-03	4.9E-03	7.0E-04	0.20
POD19 (G12) Ethylbenzene (100-41-4)	4.0E-04	1.8E-03	1.8E+00	1.00
POD19 (G13) Toluene (108-88-3)	9.5E-03	8.8E-03	4.5E+00	1.00
POD19 (G14) Xylenes (1330-20-7)	2.0E-03	2.4E-03	3.1E+00	1.00
POD19 (G15) Naphthalene (91-20-3)	1.8E-05	1.1E-03	8.5E-02	1.00

Notes:
1 - Protective Exposure Concentration, or target concentration
2 - Probability that the ending concentration is less than the target concentration
Shaded values indicate that the mean concentration is above the Protective Exposure Concentration, or the probability is below the target probability.

Based on the results of the probabilistic calculations, there are only two complete exposure pathways (i.e., benzene in indoor air and benzene in outdoor air). For the examples, the results are the same using the target probability of being below the protective exposure concentration and comparing the expected concentration to the protective exposure concentration. Figure 5.8 is a map that includes the complete exposure pathways. These are the results that are included in Table 5.11. Comparisons of the predicted concentration distributions to the protective exposure concentrations for benzene, and toluene in indoor air are included as graphs in Figure 5.9.

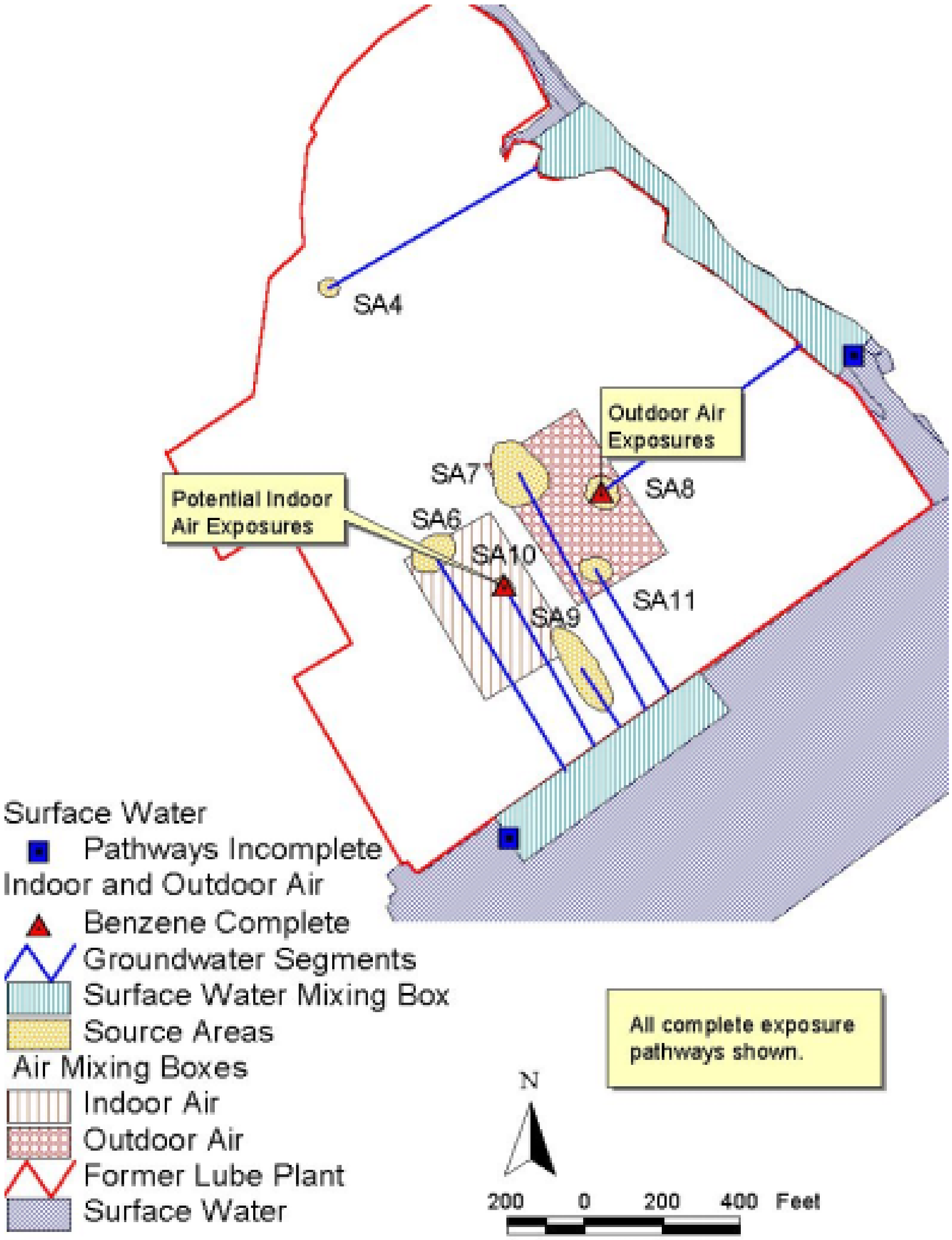
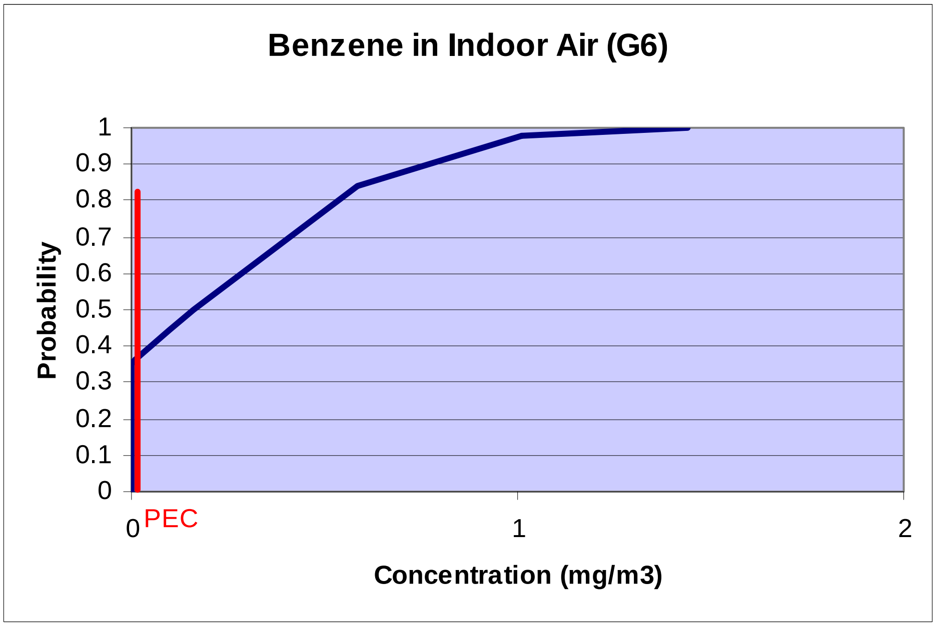


Figure 5.8 - Complete Exposure Pathways Based on Probabilistic Calculations



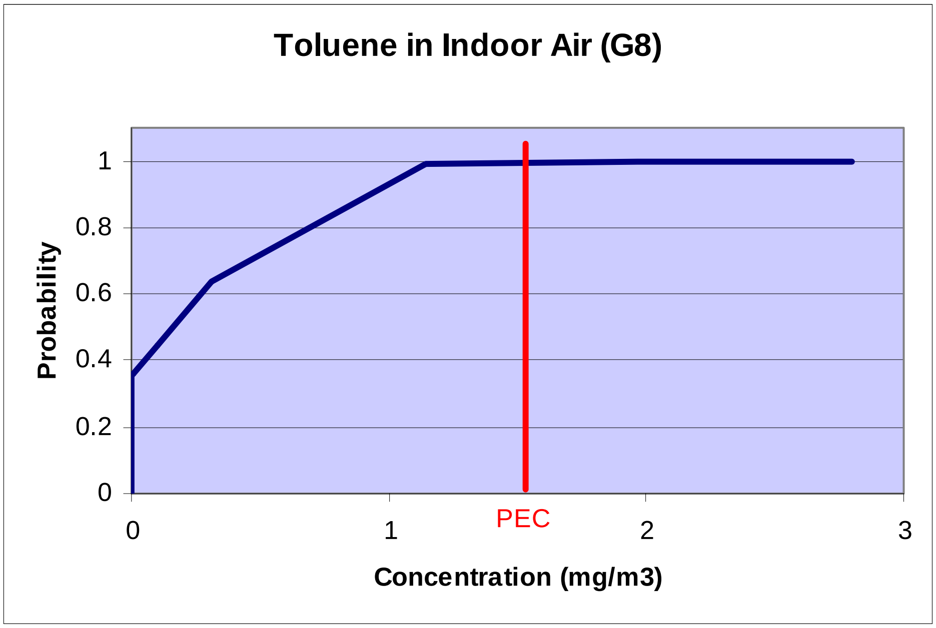


Figure 5.9 - Cumulative Distribution Functions for Predicted Indoor Air Concentrations

The cumulative distribution functions were modeled as normal distributions using the mean and standard deviation data for the predicted indoor air concentrations. Because negative concentrations are physically impossible, the distribution curves were truncated at concentrations of zero. The protective exposure concentrations were added to the plots. The probability of being below any given concentration is given on the y-axis. As indicated in Table 5.9 the coefficient of variation for the indoor air exposure pathways for benzene, toluene, ethylbenzene and xylenes are relatively small, and for naphthalene it is relatively large, based on the small mean value. The probability of being below the target level however seems to be driven by the fact that the benzene protective exposure concentration is several orders of magnitude lower than those of the other chemicals. Also contributing to this spread of values is the chemical volatility, benzene being the highest and naphthalene being the lowest.

5.2.7 Discussion of Probabilistic Calculation Results

As indicated in section 5.2.5, the groundwater to surface water mixing exposure pathways were determined not to be complete based on the deterministic calculations, and for all of the chemicals of concern this determination is easily supported by the probabilistic calculations, even in-light of the significant variability in the results.

For the volatilization to indoor and outdoor air exposure pathways the benzene exposure pathways are determined to be complete based on the deterministic calculations and based on the probabilistic calculations. If the probability of the ending concentration being below the protective exposure concentration, target of p = 0.9 is appropriate, then these two exposure pathways require additional actions. In the case of the toluene and xylenes volatilization to indoor air exposure pathways, by the deterministic calculation the exposure pathways are complete and require further action, however by the probabilistic determinations the exposure pathways are incomplete (i.e., the mean concentrations are below the target levels and the probabilities are above p = 0.9, at 0.92 for toluene and 1.0 for xylenes).

Several analyses can be completed based on these results to help identify further actions that should be taken. In order to provide illustrative examples, the benzene transport to surface water group (i.e., group G1) along with the benzene volatilization to indoor air exposure pathway, are used for the example evaluations that follow.

An initial evaluation is to identify the contributions of the individual source areas to the ending concentration for the exposure pathway group. Table 5.12 includes the predicted mean groundwater concentrations of benzene for the individual source areas in Group G1. The dominant source area is SA9, which would be expected based on the distances from the bulkhead. SA9 has a mean transport distance of 59 m and the next closest source area, SA11, has a mean transport distance of 123 m. This means that the greatest effect on the surface water concentration comes from the groundwater concentrations related to SA9. If further actions were necessary, addressing SA9 would likely reduce the resulting concentrations in the surface water sufficiently without reductions in the other source areas. This is a simple example of the importance of evaluating multiple source areas impacting one point of demonstration. The concept can easily be implemented in the analysis of multiple groups of exposure pathways impacting many points of demonstration.

Table 5.12 - Predicted Mean Groundwater Contributions of Benzene to Surface Water for Group G1

Source Area	Groundwater Concentration mg/L	Percent Contribution
SA6	6.5E-06	0.03
SA7	9.3E-06	0.05
SA9	1.9E-02	98.54
SA10	1.5E-05	0.08
SA11	2.6E-04	1.30
Total	2.0E-02	

In contrast, Table 5.13 includes the predicted mean benzene in indoor air contributions for Group G6. The contributions from the three contributing source areas are nearly equal, and the total concentration is significantly above the protective exposure concentration, so actions may be required for all three source areas.

Table 5.13 - Contributions to Indoor Air Concentrations of Benzene from the Group G6 Source Areas

Source Area	Indoor Air Concentration mg/m^3	Percent Contribution
SA6	5.0E-02	31.13
SA9	5.6E-02	34.90
SA10	5.4E-02	33.97
Total	1.6E-01	

Another evaluation that can be implemented is illustrated in Figure 5.10. The contributions to the variance in the predicted surface water concentration of benzene from Group G1 are graphed in this figure.

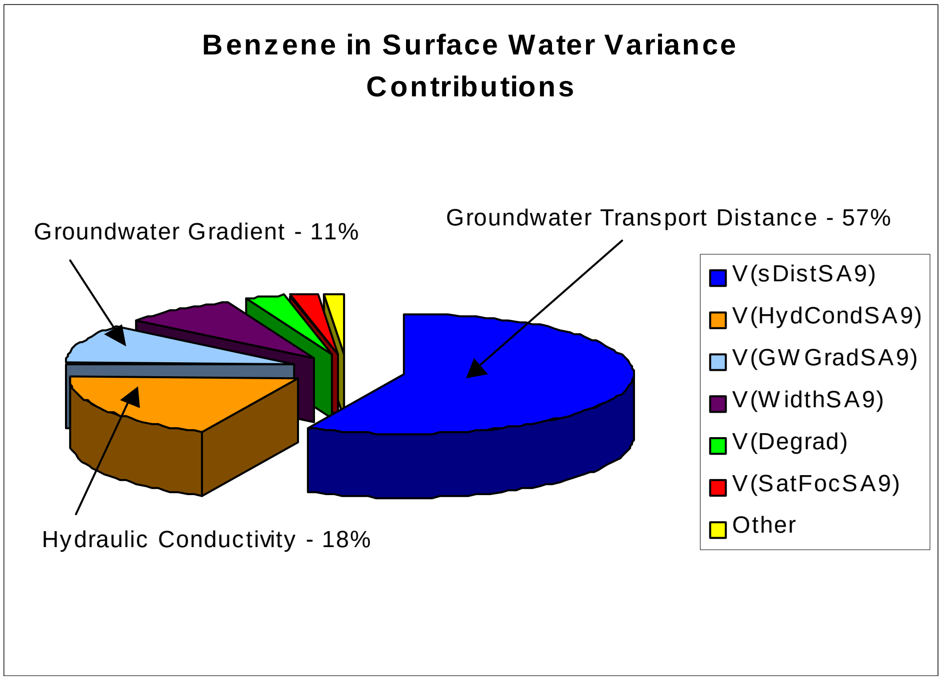


Figure 5.10 - Contributions to the Variance in the Predicted Surface Water Concentrations for Benzene

The variance contribution for each parameter depends on the magnitude of the estimated variance for that parameter and upon the functional relationship (i.e., the models used for the exposure pathway calculations) of the parameter and the ending concentration. The dominance of the variables related to source area SA9 is to be expected, as discussed above, based on its relative distance from the bulkhead. The proportions given in Figure 5.10 are based on the independent variable variance contribution only. The information included in Figure 5.10 is important to the decisions about further data collection, again if further actions were necessary for this exposure pathway group. The groundwater transport distance is an expected dominant variable since the dispersion and biodegradation attenuation mechanisms depend on the downgradient distance from the source. As discussed in Appendix B, the variability in this distance depends on the specificity of the definition of the extent of the source area. Additional data collection might be valuable to refine that definition. In addition, taking into account the functional importance of the hydraulic conductivity in groundwater transport, reducing the variance in the hydraulic conductivity measurements, particularly in the vicinity of SA9, would reduce the variance in the predicted concentrations and therefore would likely be valuable to the facility decision making about further actions.

For the benzene to indoor air the variables contributing the greatest to the ending concentration variance are the areal fraction of cracks in the building foundation, the water content in the porous media in those cracks and the vadose zone water content. These contributions are illustrated in Figure 5.11.

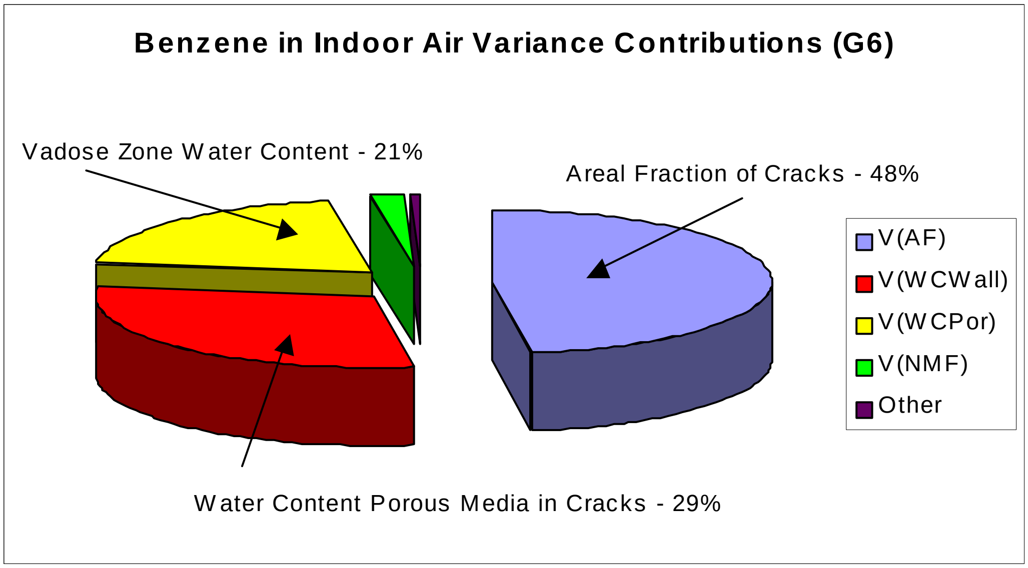


Figure 5.11 - Variance Contributions for the Volatilization to Indoor Air Exposure Pathway

For a future exposure pathway, such as is the case for these examples, it will not be possible to collect measurements for two of the three variables (i.e., areal fraction of cracks in the building foundation and water content of soils in those cracks). Further, the coefficient of variation for this exposure pathway group is relatively small (i.e., c.o.v. = 2.7), so even if data collection were feasible, it may not be valuable because the variance reduction may be insufficient to change the outcome for this exposure pathway evaluation. Finally, as noted in section 5.2.4, the volatilization to indoor air model does not include biodegradation in the vadose zone. It may be most appropriate to collect vadose zone vapor samples for benzene and for biodegradation indicator parameters and derive an empirical attenuation factor rather than continue to use the conservative model.

The development of the probability of being below the protective exposure concentration, the estimates of the variance in the ending concentrations, and those variables contributing the greatest to the overall variance, are important in the evaluation of further actions for the facility. The magnitude of the coefficient of variation (c.o.v.) is an indicator of the value of additional data collection. For small c.o.v. values, additional data collection is not likely to be valuable. Conversely, for large c.o.v. values, additional data collection may be valuable.

The uncertainty in the ending concentrations and the potential for further site sampling, additional modeling or development of remedial actions requires the evaluation of these different alternatives and the potential for additional information to support or change the preferred alternative. The analysis of the different alternatives using decision trees and an estimate of the improvement in the probability for success of an alternative utilizes the uncertainty in the calculated ending concentrations as input values (Hay Wilson, *et al.*, 1998).

The uncertainty data are also used to illustrate the modeling results and the confidence in the modeling results. In section 5.3, a discussion of the benefits of visualizing the modeling results and the use of the probabilistic calculation results are presented.

5.2.8 Additional Calculations for Variance Estimation

In the calculations presented in section 5.2.6, a relatively simple model of variable correlation was used. The variables from specific parameter sets (e.g., subsurface variables, chemical variables) were considered to be potentially correlated and then an evaluation was made to determine the appropriate correlation coefficient. The variables in two different parameter sets however, were considered to be independent. The procedures for estimating the variable relationships and the correlation coefficients are presented in Appendix B.

The example probabilistic calculations were implemented for groups of similar exposure pathways affecting the same point of demonstration. In these calculations the variance in the ending concentration depends on all of the input parameter values for each of the exposure pathways in the group. This means, for instance, in the groundwater to surface water groups, that there is a contribution to the variance from the correlation between the hydraulic conductivity values for each pair of exposure pathways in the group. This cross pathway correlation was examined to determine its effect on the magnitude of the overall variance.

If the situation is considered for two groundwater to surface water exposure pathways affecting one point of demonstration, and the mean ending concentration E(CT) is:

$$\mathrm{E(CT)} = \mathrm{E(C1)} + \mathrm{E(C2)}$$ **Equation 5.4**

Then the variance can be calculated by considering the statistically independent parameter sets. Instead of evaluating the potential correlations for each variable in a specific parameter set across the up to 6 exposure pathways in each group, a lumped correlation coefficient was developed. The lumped correlation coefficient relates all of the variables in one parameter set between two exposure pathways. It is possible to reduce the number of calculations in this way because the variance in any one ending concentration (C1) can be thought of as the sum of the variances due to each of the parameter sets:

$$\begin{aligned}\mathrm{Var(C1)} &= \mathrm{Var(C1)}_{\mathrm{source}} + \mathrm{Var(C1)}_{\mathrm{subsurface}} \\ &+ \mathrm{Var(C1)}_{\mathrm{chemical}} + \mathrm{Var(C1)}_{\mathrm{surfacewater}} + \mathrm{Var(C1)}_{\mathrm{model}}\end{aligned}$$

Equation 5.5

The same expression can be written for a second ending concentration (C2):

$$\begin{aligned}\mathrm{Var(C2)} &= \mathrm{Var(C2)}_{\mathrm{source}} + \mathrm{Var(C2)}_{\mathrm{subsurface}} \\ &+ \mathrm{Var(C2)}_{\mathrm{chemical}} + \mathrm{Var(C2)}_{\mathrm{surfacewater}} + \mathrm{Var(C2)}_{\mathrm{model}}\end{aligned}$$

Equation 5.6

So the additional variance due to the correlation between the parameters in the two exposure pathways (Var(P)) can be written as:

$$\begin{aligned}\mathrm{Var(P)} &= 2\times\rho_{1,2\mathrm{source}}\sqrt{\mathrm{Var(C1)}_{\mathrm{source}}\times\mathrm{Var(C2)}_{\mathrm{source}}} \\ &+ 2\times\rho_{1,2\mathrm{subsurface}}\sqrt{\mathrm{Var(C1)}_{\mathrm{subsurface}}\times\mathrm{Var(C2)}_{\mathrm{subsurface}}} \\ &+ 2\times\rho_{1,2\mathrm{chemical}}\sqrt{\mathrm{Var(C1)}_{\mathrm{chemical}}\times\mathrm{Var(C2)}_{\mathrm{chemical}}} \\ &+ 2\times\rho_{1,2\mathrm{surfacewater}}\sqrt{\mathrm{Var(C1)}_{\mathrm{surfacewater}}\times\mathrm{Var(C2)}_{\mathrm{surfacewater}}} \\ &+ 2\times\rho_{1,2\,\mathrm{model}}\sqrt{\mathrm{Var(C1)}_{\mathrm{model}}\times\mathrm{Var(C2)}_{\mathrm{model}}}\end{aligned}$$

Equation 5.7

The overall variance in the ending concentration (CT) is:

$$\mathrm{Var(CT)} = \mathrm{Var(I)} + \mathrm{Var(Cor)} + \mathrm{Var(P)}$$

Equation 5.8

Where Var(I) and Var(Cor) are the independent and correlated variances presented in section 5.2.6.

For the volatilization to outdoor air and the groundwater to surface water mixing groups, the Var(P) values were calculated and compared to the results presented in section 5.2.6. As a first estimate, all of the parameter sets for all of the exposure pathway pairs were assumed to be strongly and positively correlated, which means that all of the correlation coefficients were assigned a value of 0.9, using the same correlation coefficient value function presented in Appendix B. Table 5.14 includes the results of the calculations.

Table 5.14 - Comparison of Variance Results for the Inclusion of the Correlation of Parameters Across Exposure Pathways

Group ID	COC	Var(I)	Var(Cor)	Var(P)	Total Variance
Surface Water Concentration Variances $(mg/L)^2$					
G1	Benzene	1.5E-10	-6.4E-12	3.5E-12	1.4E-10
G2	Ethylbenzene	1.5E-28	-7.0E-29	5.2E-34	7.7E-29
G3	Toluene	3.9E-14	-1.1E-14	5.6E-17	2.8E-14
G4	Xylenes	1.4E-17	-5.0E-18	5.2E-21	9.1E-18
Outdoor Air Concentration Variances $(mg/m^3)^2$					
G11	Benzene	4.1E-05	-1.7E-05	7.4E-05	9.8E-05
G12	Ethylbenzene	3.3E-06	-1.1E-07	6.0E-06	9.2E-06
G13	Toluene	1.4E-04	-6.4E-05	2.6E-04	3.4E-04
G14	Xylenes	8.8E-06	-2.8E-06	1.6E-05	2.2E-05

Based on these preliminary results, the correlations across exposure pathways are not significant for the estimation of the surface water concentration variances. For the outdoor air concentration variances, the total variance is increased by a factor of four for benzene, toluene and xylenes and by a factor of three for ethylbenzene. For the examples, the exposure pathway completeness evaluations do not change with the inclusion of the additional variance. The inclusion of correlation across exposure pathways in a group may be significant if the concentrations of chemicals of concern were close to the protective exposure concentrations. In these cases, the additional uncertainty in the ending concentration would change the calculated probability of being below the target concentration and potentially change the exposure pathway completeness comparison. It also must be considered that the correlation coefficients may not appropriately be assigned the same values, even for all of the exposure pathways in the groundwater to surface water transport groups (G1-G5) since some of the adjacent exposure pathways are 20 m apart, while the furthest are more than 100 m apart. This might suggest different correlation coefficients based on the distances.

5.2.9 Comparison of Predicted Concentrations to Field Data

The models used in the pathway segment calculations include model error variables that can be used to account for variability for which account is not otherwise taken and to adjust the model output to more closely match the concentrations measured in the field. Originally it was envisioned that there would be sufficient field monitoring data in space and time to evaluate the model error parameters for each model, for the source areas and chemicals of concern. Upon evaluation of the available data, it was determined that a rigorous evaluation of the model error parameters and model calibration would not be possible. For the vapor migration exposure pathways, there are no field measurement data for soil vapor concentrations, so the volatilization models could not be calibrated. In addition, there are no surface water monitoring data, and the predicted concentrations may be below detectable levels, so the surface water mixing model error parameters could not be evaluated.

For the groundwater transport model, a limited comparison between the predicted groundwater concentrations and the measured benzene concentrations was completed. Figure 5.12 shows the available benzene concentration data, from 1996 to 1999 and the NAPL thickness data from February 2000. Circles represent the benzene concentration data, with lighter shading for lower concentrations and darker shading for higher concentrations (i.e., yellow to red to dark blue). The NAPL thickness data are represented by squares of increasing size for increasing thickness. Concentration values that were reported as below the detection limit were assigned the detection limit values for calculation purposes, including the comparisons shown in Table 5.15. For some of the monitoring locations there were several measurements, and for others, installed in 1999, there was only one measured concentration (e.g., MW47, MW51, MW53). As illustrated in Figure 5.12, following the 1999 sampling event, some of the monitoring wells contained NAPL (e.g., MW8 and MW53). In the evaluation of the source area concentrations there were historical data with some NAPL composition results, but no current composition measurements. The source area NAPL mole fraction (NMF) values that were used in the example calculations were based on reference information. Without knowing the NAPL composition, from February 2000 or later, it is not possible to identify whether the prior groundwater concentrations (1999 and earlier) are representative of the groundwater concentration in the presence of the NAPL.

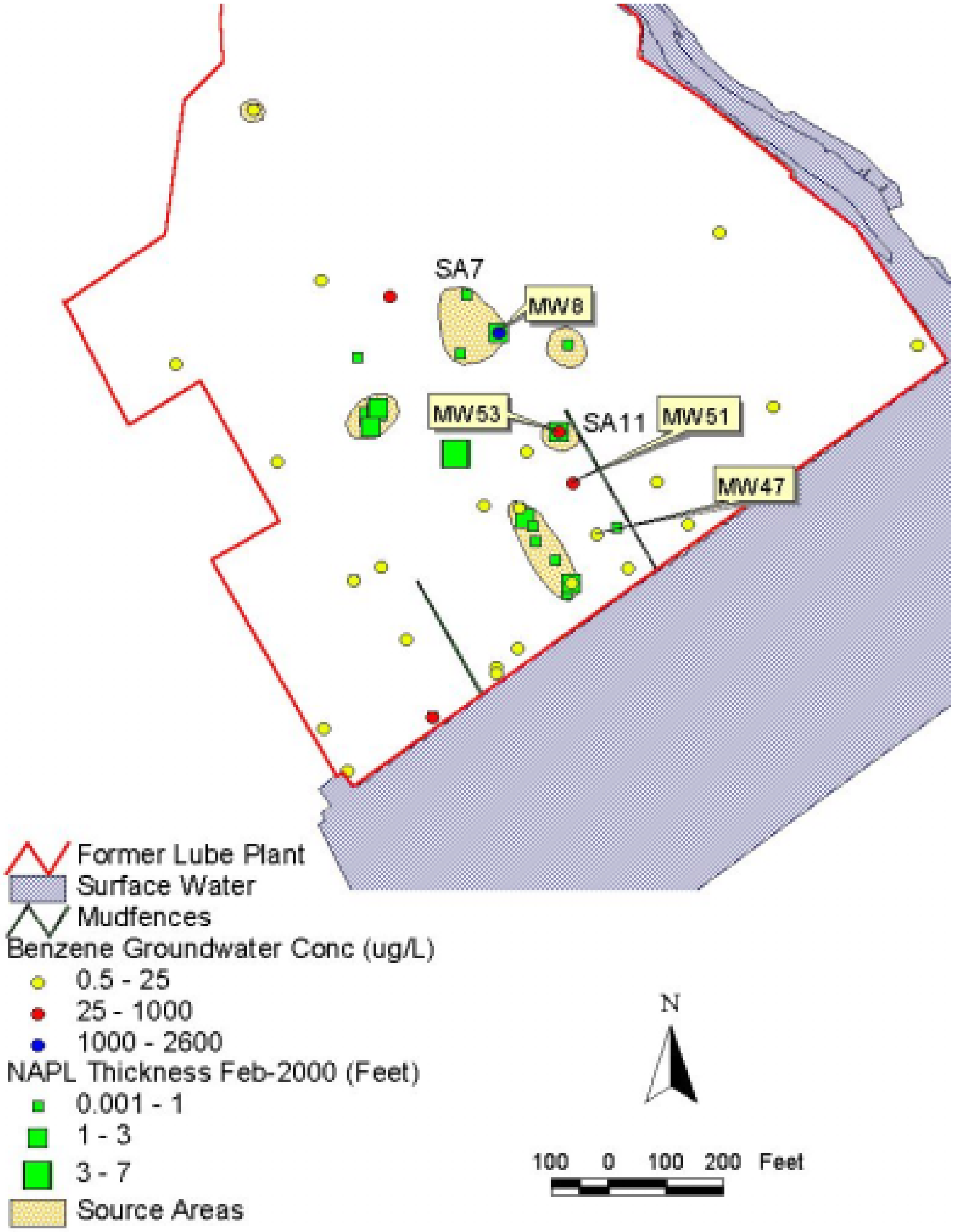


Figure 5.12 - Groundwater Concentrations of Benzene and Subsequent NAPL Thickness Data

In addition, there were only benzene concentration data associated with monitoring wells in SA7 and SA11. Typically, groundwater concentrations are only measured in wells that do not exhibit NAPL thicknesses. This may account for the lack of groundwater concentration data in the other identified source areas.

The benzene data for SA7 and SA11 were evaluated and compared to the concentrations predicted using the groundwater model calculations. The input parameter values used in the calculations presented in section 5.2.6 were used in the calculations presented here, except that the source area concentration (CA) was replaced with the field-measured concentration as noted in Table 5.15. For SA7 the concentration measured in MW8 was used as the source area concentration. For SA11 the concentration measured in MW53 was used as the source area concentration. The expected concentrations were calculated at the monitoring well locations downgradient of MW8 and MW53 (i.e., MW53, MW51, and MW47). The results of these calculations are presented in Table 5.15.

Table 5.15 - Comparison of Calculated and Measured Concentrations of Benzene in Groundwater

Location	Measured Concentrations	Calculated Mean Concentrations	Calculated Variance	Calculated S.D.	Prob[1]
Groundwater Concentrations mg/L					
SA11[2]					
MW-53	6.40E-01				
MW51	3.80E-01	1.84E-03	5.43E-05	7.37E-03	1.00
MW47	<5.00E-03	6.32E-05	5.89E-08	2.43E-04	1.00
SA11[3]					
MW-53	6.40E-01				
MW51	3.80E-01	2.35E-02	5.41E-03	1.53E-01	0.99
MW47	<5.00E-03	6.01E-03	2.01E-04	7.75E-02	0.49
SA11[4]					
MW-53	6.40E-01				
MW51	3.80E-01	3.52E-01	1.25E+00	1.12E+00	0.51
MW47	<5.00E-03	8.44E-02	4.10E-02	2.02E-01	0.35
SA7[2]					
MW8	2.60E+00				
MW53	6.40E-01	4.02E-04	9.09E-06	3.01E-03	1.00
SA7[3]					
MW8	2.60E+00				
MW53	6.40E-01	4.69E-02	8.04E-02	2.17E-01	1.00
SA7[4]					
MW8	2.60E+00				
MW53	6.40E-01	6.53E-01	1.58E+01	3.98E+00	0.50

Notes: Field data from sampling events in 1998 and 1999

1 - Probability that the calculated concentration is less than the measured concentration

2 - Calculated value based on example mean and variance data for all input parameters

3 - Calculated values based on example mean and variance data, except degradation constant set to zero.

4 - Calculated values based on example mean and variance data, except degradation constant: E(9.4E-04), Var(5.76E-04); and, NYGW: E(16), Var(9.46E+01)

Using the expected values for all of the input parameters and the distances between the monitoring wells (Note 2 in Table 5.15), as measured in ArcView, the calculated concentrations are lower than the measured concentrations, which is to be expected since one source area is being modeled to impact the measured location, when in fact there are several source areas, including source areas not identified for the examples, that are impacting the location. It also must be noted that the reported concentration for MW47 is the laboratory detection limit, so the actual concentrations is unknown, but less than the reported value.

A second evaluation was implemented by setting the expected value for the degradation constant to zero (Note 3 in Table 5.15). The calculated concentrations are in reasonable agreement with the field-measured concentrations for SA11 and MW47. For SA7 and MW53 and for the shorter distance between SA11 and MW51, the calculated concentrations are still lower than the measured concentrations. An indicator of the success of the prediction is based on the cumulative distribution for the predicted concentration. Calculating the probability that the predicted concentration is below the measured concentration gives an assessment of whether the measured concentration is within the variation of the predicted concentration. For the zero degradation case (Note 3 in Table 5.15), the probability for SA11and MW47 is close to the median value ($p = 0.49$), which would indicate that the concentrations are in reasonable agreement. For SA7 and MW53, and for SA11 and MW51, the results are not better than the initial estimate.

A third scenario was calculated by using the multiplicative model error variable (Note 4 in Table 5.15). The expected degradation constant was set to its minimum reference value and a multiplicative error value was selected so that the expected values would be in general agreement with the measured values. In this scenario, the results are better. For SA11 and MW51 and SA11 and MW47 the probability values are closer to the median value. This is also true for SA7 and MW53.

There are several issues to consider in this limited evaluation. First, these calculations are based on one source area generating the concentrations at each monitoring well, which is not actually the case. There are number of source areas that would affect the concentrations at the examined monitoring wells, including both the source areas used as examples in this research and potentially other source areas. Only representative source areas were identified for implementation of the examples. In order to properly evaluate the contributions, all of the source areas would need to be identified and concentrations of benzene in the other source areas would have to be measured, or NAPL composition data would have to be generated.

Second, the field-measured data are based only on one point in time. They do not represent long-term concentration values. It is the long-term concentration values that are predicted by the groundwater transport equation in this example. Additional monitoring data to develop expected values and a better idea of the variance in these values, particularly with the intermittent nature of the NAPL presence, would be needed.

Third, other variables impact the ending concentration, including the hydraulic conductivity and the groundwater gradient that were not varied in this comparison. The available data for the facility were used to develop the expected values and variance values for the hydraulic conductivity and the groundwater gradient. Without more information, in the vicinity of the source areas being modeled, it was not judged to be possible to adjust these values.

In this section a number of example calculations have been presented for the case study facility using the available information. Examples of data presentation and evaluations are included.

5.3 VISUALIZATION BENEFITS

Risk assessment requires the assimilation of large amounts of data and information in order to understand the potential human health and environmental risks associated with operations, and in many cases former operations, at an industrial or defense facility. There are a number of ways that maps of data and physical facility features can help to improve understanding of the available data, the potential for exposures and the need for more information. The availability of these figures and maps for the risk assessment is a direct result of the spatial site conceptual model and the data interconnections developed in the spatial environmental risk assessment methodology.

In this section, examples of the use of maps for visualization of the risk assessment process are presented. Some of these figures have been presented elsewhere in the text to support the discussion of the spatial environmental risk assessment methodology.

In developing maps and figures of data, the audience and the message are important considerations. Whether the audience is a group of engineers working together on the risk-based evaluation or whether the audience is a stakeholder group interested in what will happen at a facility in their neighborhood will impact the types of maps and figures that will be helpful to each group. In addition, if the data are to be presented on paper maps, for example, as posters at a public meeting or in a written report, a selection process will be necessary since it is not likely that all of the data can be presented at the same time. If however, the data are to be presented in electronic form and the interested parties have the available software to query the results and view them in any combination, the possibilities are not as limited. It should be the case that implementing the spatial environmental risk assessment methodology for a full-scale risk-based evaluation would provide an analysis tool to the participants that serves to inform the decision process by promoting a better understanding of the data that are available and particularly where there are data gaps.

An example of using the maps for data analysis is presented in Appendix B for the development of the source area representations. Figure 5.13 and Figure 5.14 show the non-aqueous phase liquid (NAPL) detections and specific gravity measurements. As an example, as shown in Figure 5.14 source area SA7 was defined separately from SA8 because the SA7 specific gravity measurements are representative of lighter products than the measurements from SA8.

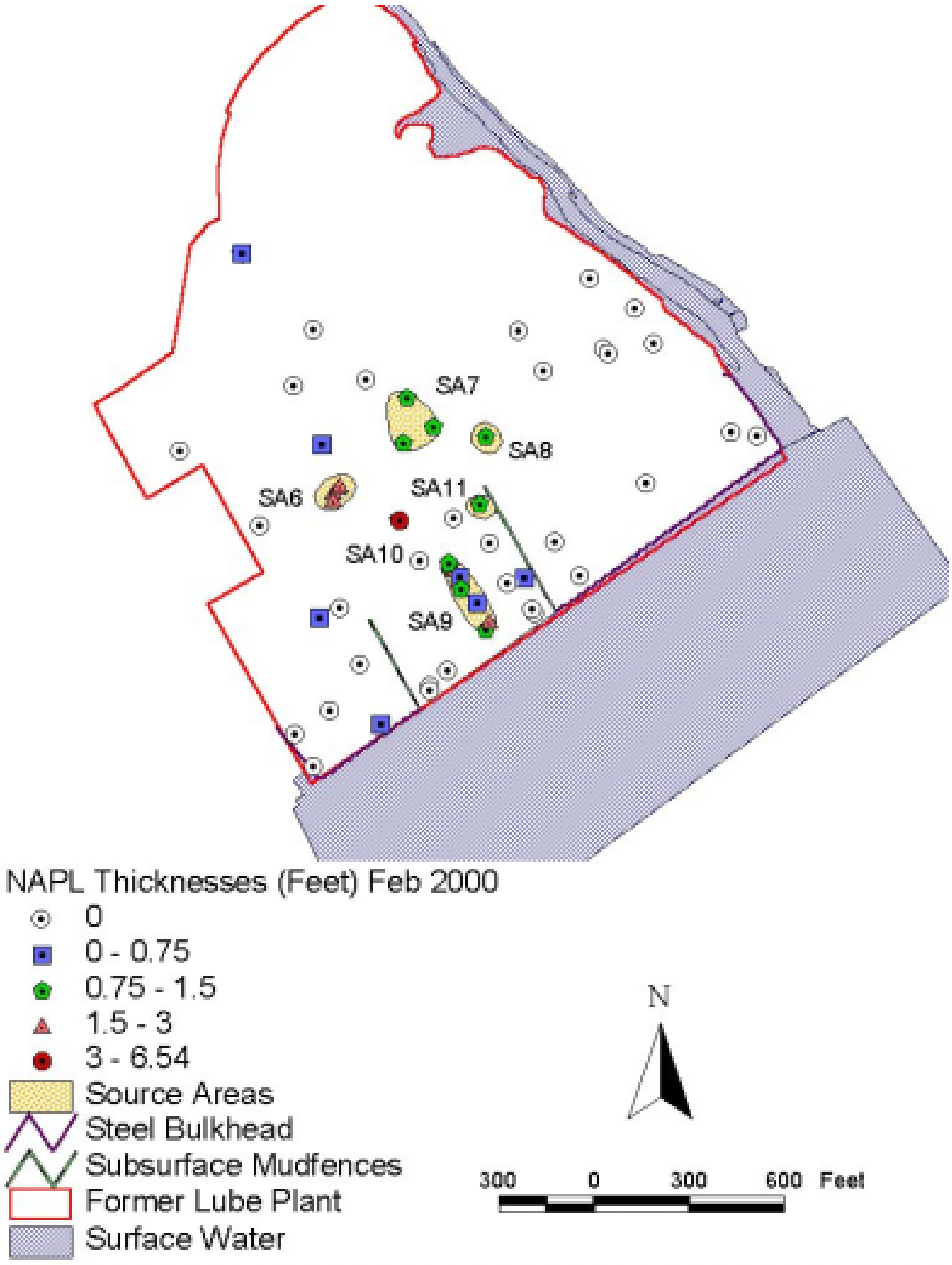


Figure 5.13 - Map of NAPL Detection Data Used for Source Area Identification

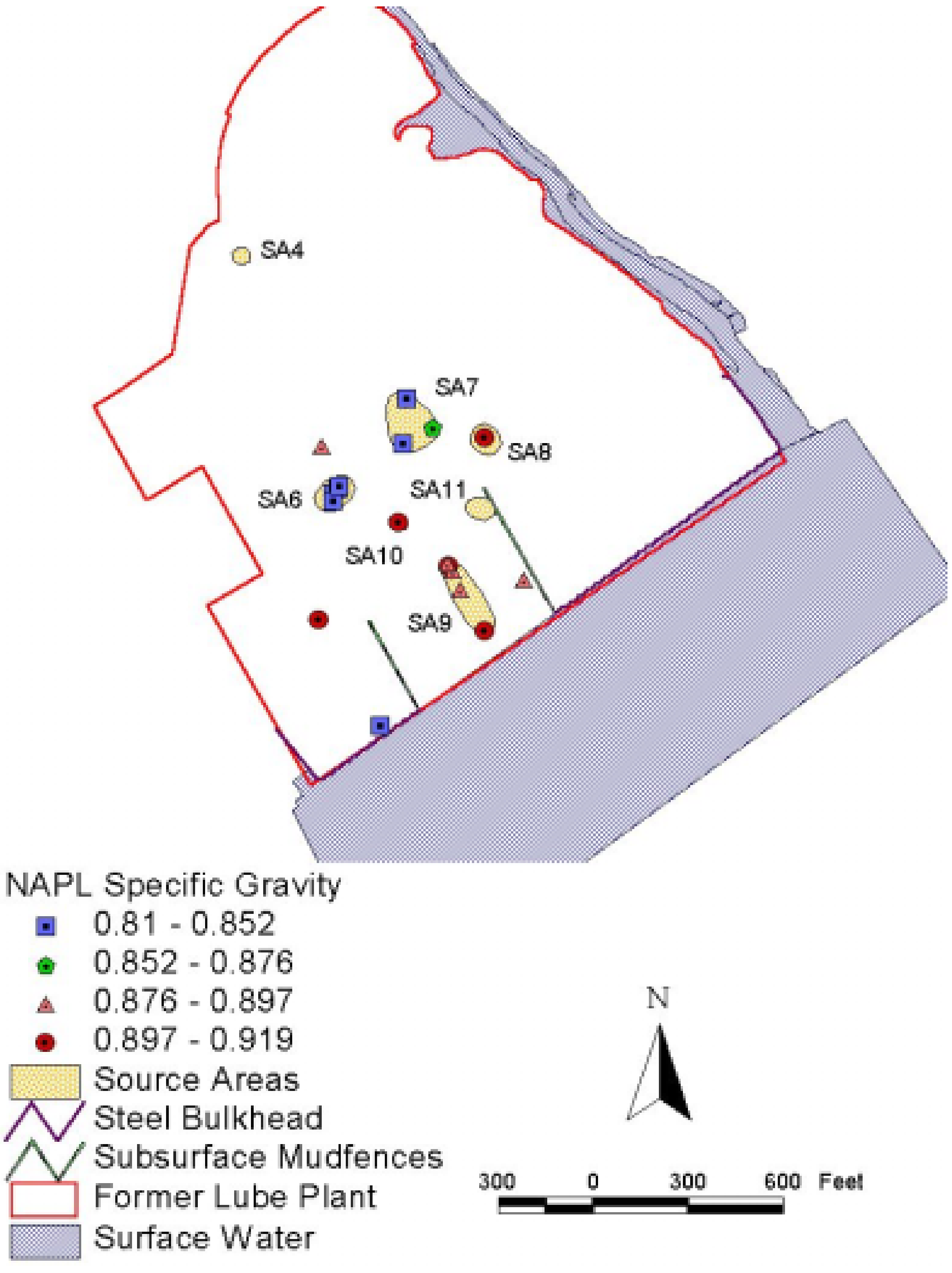


Figure 5.14 - NAPL Specific Gravity Data Used for Source Area Identification

The map of the spatial site conceptual model, included as Figure 5.15, could be used to explain to a stakeholder audience which exposure pathways are being evaluated in the risk assessment. The map includes the source areas, the locations where receptors may come in contact with chemicals of concern and the physical description of the environmental transport of chemicals of concern. These may be the only conceptualizations a layperson has seen describing exposure pathways and the process for risk assessment. These maps may also be useful to the engineers and scientists conducting a risk assessment as a type of "visual accounting" system to ensure that all of the necessary calculations have been performed.

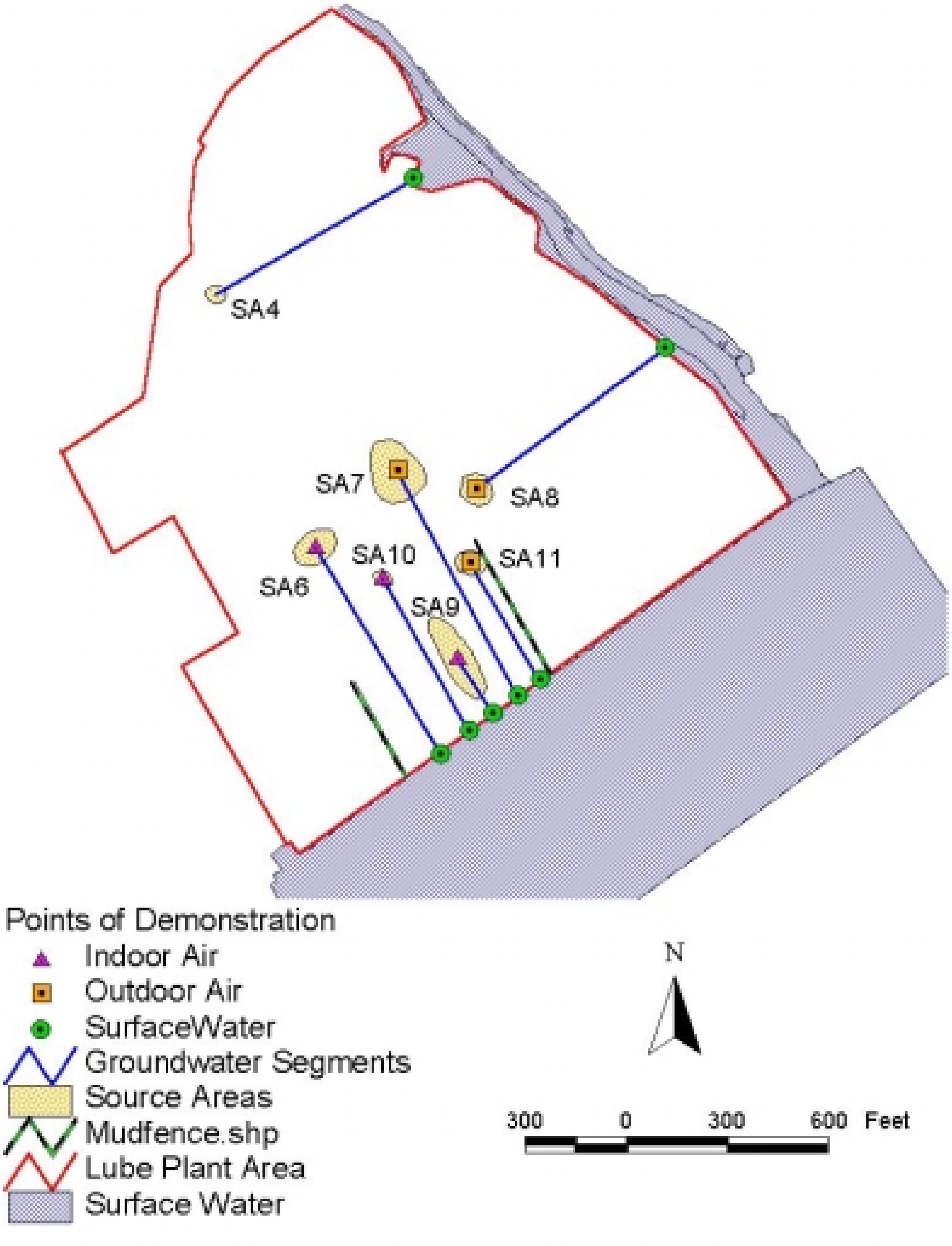


Figure 5.15 - Example Site Conceptual Model Map

The results of the exposure pathway completeness evaluations can be illustrated using the maps. The exposure pathways for which additional evaluations must be performed can be differentiated from those exposure pathways for which no additional actions are necessary. An important consideration in building the maps of the results is deciding which data are to be shown. In the examples presented in section 5.2 there are five chemicals of concern, seven source areas and four types of exposure pathways. In addition, calculations were implemented for individual exposure pathways and for groups of exposure pathways. For the exposure pathway completeness figures selected results are presented here. In a situation where the interested parties can query the electronic maps, this selectivity is not necessary. All of the results are available in the GIS and any combination of results can readily be displayed. Figure 5.4, Figure 5.5, Figure 5.7, and Figure 5.8 are all examples of the display of exposure pathway completeness evaluations in maps.

In Figure 5.4 and Figure 5.5, all of the complete exposure pathways are illustrated. A combination of color and shape of markers is used to differentiate between the exposure pathway completeness (by color) and the chemicals of concern (by shape). The completeness results are illustrated by the markers at the points of demonstration. In order to view markers that are stacked on top of one another (i.e., results for multiple chemicals of concern at individual points of demonstration) the markers are scaled (i.e., the ones shown on top are smaller than the ones shown on the bottom). Text boxes are used to highlight and clarify the results that are presented. In this figure, the size of the marker does not correspond to a concentration or probability value.

In Figure 5.7, all of the complete group exposure pathways are illustrated. In this example there are fewer chemicals of concern for which the exposure pathways are complete. Again the shape of the markers corresponds to the chemical of concern and the color corresponds to exposure pathway completeness. Text boxes are used to clarify the exposure pathway types that are represented by the markers.

In Figure 5.8, the complete exposure pathways are illustrated for the probabilistic calculations. The only complete exposure pathways are the benzene to indoor air and benzene to outdoor air.

As a result of the estimates of the variance in the predicted ending concentrations, information is available about the certainty of the pathway completeness evaluations. In Figure 5.9, the cumulative distribution functions for the predicted indoor air concentrations are presented along with the protective exposure concentrations. The point at which the protective exposure concentration line crosses the cumulative distribution plot represents the probability that the concentration will be below the protective exposure concentration. These data can be combined with the exposure pathway maps. An example is presented as Figure 5.16.

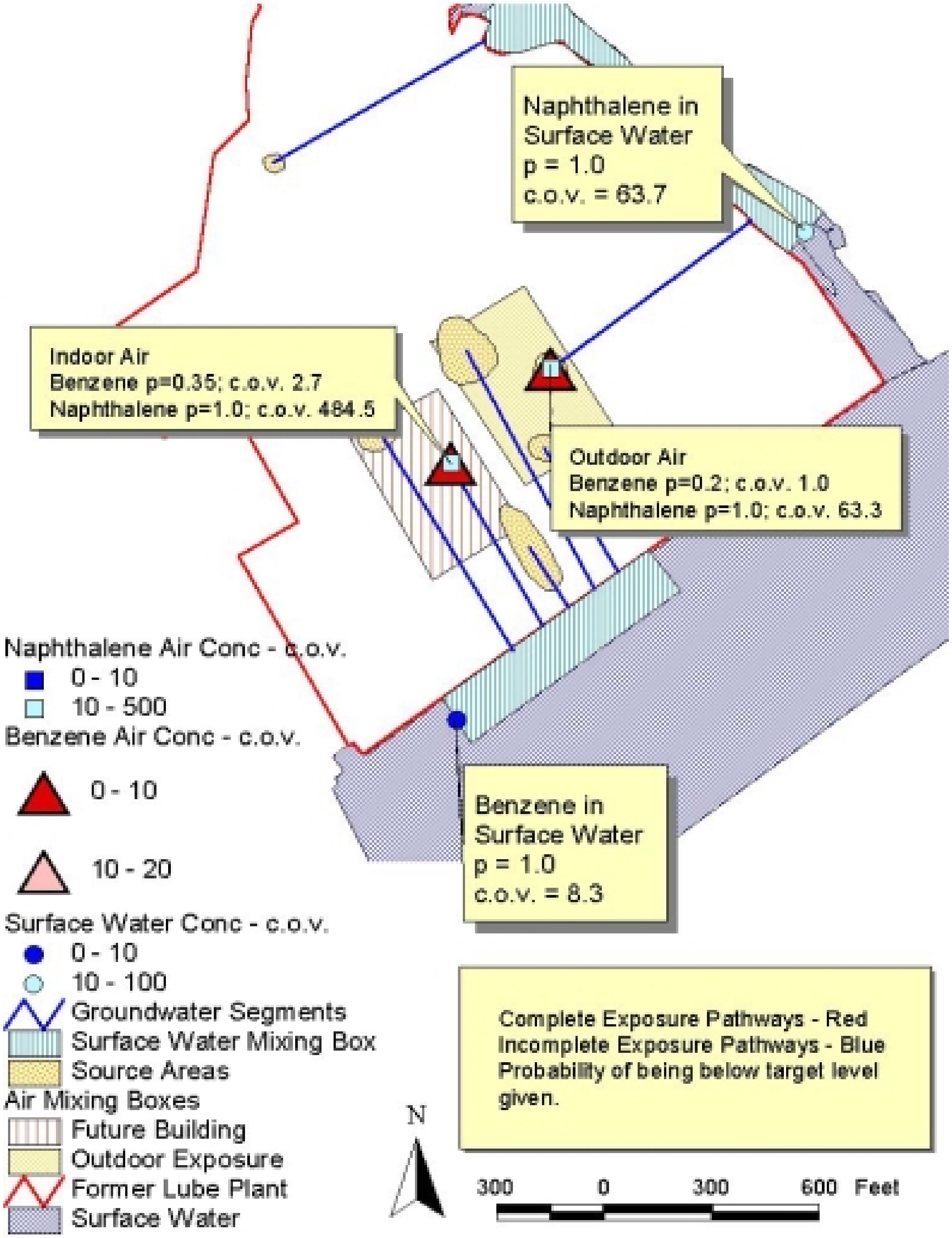


Figure 5.16 - Uncertainty Associated with Exposure Pathway Completeness Evaluation

In Figure 5.16, the uncertainties in the exposure pathway evaluations for benzene and naphthalene are presented. These data are included in Table 5.9 and Table 5.11. For each ending concentration the probability of the concentration being below the protective exposure concentration is given. The coefficient of variation (c.o.v.) for the concentration is also presented. The markers are scaled based on the probability of being below the target level, the lower the probability the larger the marker (i.e., larger the marker the greater the concern). The colors are shaded from pale for high c.o.v. values, more uncertainty, to dark for lower c.o.v. values. In addition, complete exposure pathways are given red symbols and incomplete exposure pathways are given blue symbols. Further, different shapes of symbols represent different media, or different chemicals in the same medium at the same point of demonstration. Benzene in air are triangles, naphthalene in air are squares and both benzene and naphthalene in surface water are circles. In this way, even for the incomplete exposure pathways the certainty in the result and the predicted variability in the ending concentrations are presented.

Using a system such as this provides more information than if all of the exposure pathway completeness data are only presented as yes or no values. An example of which is shown in Figure 4.5. The use of both types of representations may be necessary. For a technical and regulatory audience, Figure 5.16 would be appropriate. For a general audience the results could be reduced to the yes or no maps for all of the exposure pathways. It may be possible to present both maps and discuss that the Figure 4.5-type map is derived from the Figure 5.16-type map.

5.4 SPATIAL ENVIRONMENTAL RISK ASSESSMENT METHOD EVALUATION

In this section the strengths and weaknesses of the SERA methodology are discussed. Comparisons to personal experience with conventional risk assessment procedures are used to provide a basis for valuing the methodology.

5.4.1 Software Reliability

The methodology relies on several different software systems to implement all of the components. The GIS is required for visualization and for managing spatial data, the relational database software is needed for relatively complex data manipulation and storage of the various data that describe the exposure pathways and the input parameters for the calculations, and the spreadsheet is required for implementing the pathway segment calculations. As the number of calculations and the complexity of the queries increased, difficulties were encountered with the links between Excel and Access and maintaining these links. It is not clear that there was a causal relationship between the number of calculations and the difficulties with the links or if they were merely coincident occurrences. Initially, the applications were developed in Windows98® and Office97®, and all of the connections ran smoothly. Errors began occurring in the ODBC connection inside of Excel and in the query generation in Access. A fatal error (e.g., a page fault error in the kernel32.dll file) began occurring in Excel when previously linked data were updated in Excel. In addition, similar illegal operation errors were generated in Access during query completion. After a number of trials, including virus scans and disk de-fragmentation operations, it was noted that when the database operations and the spreadsheet calculations, including the application links, were performed on a computer system with Office 2000® and Windows98®, and on a computer system with Office 2000® and WindowsNT®, these errors did not occur. The Microsoft on-line support was consulted and no specific cause for these errors were found.

The primary system was then upgraded to Windows98® Second Edition, including the SR2b updates. This corrected the page fault error. The final, working version of the kernel32.dll file, which is included in the SR2b update, is dated April 23, 1999. This however did not fix the errors on linking the Access queries to Excel. The system was updated to Office 2000®, including the SR1a update. Once this was completed the linking operations were successful. The final, working version of the ODBCad32.exe file, which is included in the SR1a update, is dated August 8, 1999 (the file version in the File Properties is 3.520.4403.2). It is important that these latest versions of the system files be available in order for the Access and Excel application links to run reliably.

No difficulties were encountered with the links between the ArcView 3.1® or ArcView 3.2 ® software and Access, either with the existing system files or the updated files. It is not clear that the errors were related to the files sizes either, since the number of connections and the number of records in each connection, in the initial examples, used for the examples in Chapter 4, and the final examples, used in Chapter 5, were not large compared to typical databases (i.e., 100 or so records compared to 30,000 records in the environmental measurements database, representative of a typical database). It would seem then that the errors encountered were Microsoft application bugs that happened to be encountered fairly late in the implementation process, but they were likely there all along.

5.4.2 Method Usability

The implementation of the spatial environmental risk assessment to implement (1) a comparison to statewide standards (i.e., screening level risk assessment), (2) a site conceptual model and (3) relatively simple fate and transport models for comparison to risk-based target levels (i.e., initial site-specific risk assessment) is logical and systematic. The method does required knowledge of several software applications and particularly their linkages through the Microsoft ODBC.

The process of defining the source areas, the exposure pathways and modeling variable input values for each of the pathway segments is straightforward. Each source area definition, for example is linked to a visual representation of the source area, which can be overlain in the GIS with other environmental measurements data in a map. Figure 5.13 shows the example source areas for the Former Lube Plant and the NAPL thickness measurements that were used to define the NAPL source areas. The differentiation between SA 9 and SA 11 was made because of the sampling locations between these two areas where NAPL was not detected. The source area definition is also linked to the records in the site conceptual model database that describe each source area, including the chemicals of concern, the modeling variable parameter values such as source depth and distance from a subsurface source to the ground surface.

It can be the case in a conventional risk assessment that the modeling input values that are used in the evaluation are not easily assembled or identified. The calculations may be implemented in various spreadsheets and modeling codes, but the data tables that are generated for the final report and documentation are transcriptions of the input data. The chance for error in translation or omission of data is fairly high in this operating environment. It is also relatively time-consuming to move data by hand from one application to the next. The SERA methodology has the advantage of storing and managing all of the calculation values in the shapefile data tables and in the Access database.

5.4.3 Meta Data Management

Data management is important in the application of environmental models to evaluate complex natural systems. The *meta data* is the information about the quality, derivation and basis for the data being used. In particular, in assessing risks posed by chemicals of concern at a large facility, the task of meta data management is critical to the success of the project. An important aspect of data management is tracking the source of all of the input values, the methods used to derive the values, the evaluations that have been conducted, the results and the decisions made based on the results. This information is the basis for the repeatability of the calculations and the demonstration that all of the exposure pathways that need to be analyzed have been analyzed. All of this information is managed in the spatial and tabular databases integral to the SERA methodology.

An important feature of the SERA methodology is the systematic organization of the data and information used in the risk assessment process. The database structure for the information provides the flexibility to add data fields for a specific facility application as the circumstances warrant. Each aspect of the site conceptual model development is documented in the spatial and tabular databases and in linked spreadsheet files. As an example, in each input parameter table (e.g., Source Variables, SubSurVariables and AboveGrd Variables) there is a Data Derivation field that is a hyperlink reference to the spreadsheet, or query from the environmental measurements database, from which the parameter values were derived. These linked files were used extensively in the development of Appendix B. In addition, the site conceptual model database includes a table called Variable Descriptions that documents all of the input parameters used in the modeling and provides a description, the segments in which the parameter is used and provides a discussion of each.

The shapefiles for the geographic representation of the site conceptual model components provide the visual documentation of all of the source areas and exposure pathways that were considered in the risk assessment process. The identity of these files and their contents are collected in supporting tables like the ones included in Appendix E. During the course of the risk assessment the understanding of the sources and exposure pathways will evolve as more data are collected and the exposure pathway evaluations are conducted. Each of these representations of the source areas, transport mechanisms and points of demonstration are saved in separate versions of the shapefiles. This is so that a record of the evolution of the site conceptual model can be maintained.

5.4.4 Resources

For any risk assessment, development of the site conceptual model involves the data evaluation and parameter estimation presented in Appendix B. In addition, the model identification for the various exposure pathways, as presented in Appendix A is performed for any risk assessment. Therefore the resources to complete these tasks are not considered to be incremental to the SERA methodology. In fact, based on personal experience, the task of identifying parameter values and identifying source areas and exposure pathways was more efficient and easier to track using the spatial and tabular databases than prior experiences with paper and spreadsheet notes and tables. Two additional tasks, that are not typically included in an initial site-specific risk assessment, were determining the amount of variability in each of the parameter values and the correlation relationship between the variables. It is estimated that the time involved in developing the confidence intervals, the mean and variance values and the correlation values for all of the variables (i.e., 95 variables for the case study example), including the chemical-specific parameters, is approximately 20 to 50 minutes per variable. This was based on the fact that the digital facility description, and most importantly, the environmental measurements database had been constructed by others prior to the work conducted here. Developing the digital facility description is detailed in Romanek, *et al*., (1999).

The development of the database records for all of the exposure pathways and pathway segments amounted to approximately 2 to 5 minutes per record. This was once the database schema had been developed and the database tables designed. This task therefore was not considered to be time-intensive. In addition, the development of a database schema and table design would be conducted once and then may be applied at many facilities.

The development of the spatial representations of the site conceptual model elements involves some tasks that are independent of the number of source areas involved (e.g., generating a source area shapefile, a point of demonstration shapefile). However, these tasks are relatively simple. Based on the spatial representations for the case study, the development of the all of the files for the spatial site conceptual model involves approximately 45 to 75 minutes per source area.

The creation of the model spreadsheets for the deterministic calculations was relatively straightforward, once the linking process between the database and the spreadsheet was identified. Programming the attenuation mechanism functions in visual basic for the Excel Add-In was also easily implemented once the programming steps had been identified. Based on the calculations available to date in the SERA application example, the implementation of a calculation spreadsheet involves approximately 30 to 60 minutes per exposure pathway.

The creation of the model spreadsheets for the probabilistic calculations was considerably more time-intensive than the deterministic calculations. This is due to the number of derivative equations that are needed for the calculation of the variance. A derivative equation is needed for each ending concentration with respect to each variable involved in the calculation of that ending concentration. Each of the derivative equations was developed by-hand. In the examples used for the case study, there are 3 exposure pathways with 7 segment calculations. For the groundwater transport to surface water mixing there are 20 variables, for the volatilization to indoor air there are 11 variables and for the volatilization to outdoor air there are 10 variables. This results in a minimum of 41 derivative equations. Because the original equations are not basic functions, intermediate derivatives were calculated in order to implement the product rule and the quotient rule in the differentiation (Salas and Hille, 1982). As a result there are 53 derivative equations in the SERA Excel Add-In.

Development of the derivative equations for additional exposure pathway models would depend on the complexity of the equations and the number of variables involved in the equations. It is estimated based on the development time for the equations included in the SERA examples that the probabilistic equation development and quality control checking involves, at a minimum, 4-8 hours per exposure pathway.

The savings in evaluation and reporting time were not quantified because a typical risk assessment report was not generated for the case study. Based on professional experience, it is expected that the documentation of the site conceptual model and the pathway results could be as much as 50-percent faster than using typical spreadsheet, flowchart and table development since all of the data reside in the spatial and tabular databases so the results can be exported efficiently.

In terms of computing resources, the files for the case study facility (including the digital facility description files and the intermediate files) account for approximately 270 MB of file space for 2000 files. The primary computer system running the SERA application is a Pentium II 366 MHz, with 128 MB RAM and a 6 GB hard drive. The applications run quickly on this computer, so it would seem that it is more than adequate. The software requirements are described in section 5.4.1.

5.5 APPLICATION OF THE SPATIAL ENVIRONMENTAL RISK ASSESSMENT METHODOLOGY TO OTHER FACILITIES

This section discussed the data requirements for use of the SERA methodology, the customizable features and the use of additional environmental models in the implementation of the SERA methodology.

5.5.1 Data Requirements

In order to implement the spatial environmental risk assessment (SERA) methodology at a facility there are a number of data requirements that must be addressed. Assembling existing information and organizing it into a format that is compatible with the database-driven approach for the SERA will likely be the most time-consuming step in the process. The digital facility description is an important element in the data evaluation process. The digital facility description should include the physical features that characterize sources, could impact environmental transport or identify receptor locations. The types of features that should be included at a minimum are:

- facility boundaries,
- adjacent land use,
- surface waters, on the facility and in the near vicinity,

- surface cover,
- current and historical structures,
- current and historical process features,
- current and historical waste handling areas,
- groundwater monitoring wells,
- soil borings, and
- other environmental sampling locations.

The digital facility description must also include a database of the environmental measurements. This can be a relational database as described in Romanek *et al*., (1999) or it can be a flat file table, such as a spreadsheet table. The construction of the database will depend on the number of sampling locations, the number of chemicals of concern and the expected sampling period. As stated earlier, for the entire Marcus Hook refinery the environmental measurements database contains more than 30,000 records. In the case of a facility of similar complexity, it is recommended that a relational database be used.

Based on the information in the digital facility description, the site conceptual model spatial and tabular databases can be developed. The extent of the existing information will determine the completeness of the spatial site conceptual model. It may be the case that very little information is available at the beginning of the project and data collection, both historical file reviews and field sampling, would need to be conducted to assemble the information necessary to construct the site conceptual model. It should be emphasized that if additional data collection is conducted after the site conceptual model is developed, the site conceptual model should be updated and revised based on the new information. The site conceptual model is intended to be an evolving representation of the current understanding of the potential for exposures.

Once the exposure pathways have been identified, the applicable screening level fate and transport models are identified. The models used in the SERA application should be seen as a guide to the complexity of the models to be used in the screening process, but these are not the only models that could or should be used to evaluate even the same exposure pathways that were included in the case study examples. The models must be available in the attenuation factor format, or they must be re-derived to be in the attenuation factor format. The input data that are required for the transport pathway calculations depends directly on the equations selected. As described in Appendix B, a combination of field-collected data that are site-specific and reference values was used in the calculations for the case study examples. The same will likely be the case in most other applications. Typically, it is not practical or resource-efficient to develop site-specific values for all of the model input parameters. The equations to be used and the input data required will dictate any additional data collection.

Based on experience, and the iterative nature of the risk-based decision making process included in the SERA methodology, it is advisable to begin the calculation process early in the evaluation and use the results of the early calculations to determine what refinements to the site conceptual model may be needed, what additional data are needed and what additional site investigation would prove beneficial. The need for field sampling data to verify modeling results argues for implementing the calculations and then determining where the valuable verification sampling should be conducted. In addition, as the data from the environmental measurements database are studied to determine the source areas and the potential exposure pathways, data gaps will be identified through mapping of the data and evaluation of the data to determine potential variability and uncertainty.

5.5.2 Customizable Features

The SERA methodology is very flexible in its implementation elements. There are a number of customizable features. Certainly, the site conceptual model database itself can be customized to accommodate particular exposure pathway types. Additional look-up tables can be developed to support the database population efforts and to provide consistency in the database records.

The existing fate and transport equations used in the application can be replaced with other equations or the equations may be revised based on the facility specifics. For example, if the site geology is well known and the vadose zone is known to consist of heterogeneous layers, the vadose zone diffusion equation used here could be re-written as a combination of resistances in series. As noted in section 5.5.1 new models, either for the included exposure pathways or for alternative exposure pathways, can be used as long as they are formulated in the attenuation factor format.

The tables included in the database for the preliminary pathway evaluation should be customized to apply to the applicable regulatory agency's default standards. The changes that must be made will depend on the manner in which the default standards are identified for comparison to site environmental concentrations.

5.5.3 Use of Additional Models

As indicated in sections 5.5.1 and 5.5.2, additional models may be used for the pathway evaluations. If the models are not developed in a spreadsheet then the connection to the site conceptual model database will have to be worked out separately. In the SERA methodology, the model input values are identified as records in a query results table. This table would have to be converted to the format required by the modeling code. Typically, models are developed in FORTRAN to run one simulation at a time and return the simulation results. Additional programming would be required to use the model directly. A file conversion from a table record to an input array would need to be performed and then the model code executed. Additional file handling would need to be developed to accommodate the multiple calculation records included in the input data query. The model would need to be run for each input record and an output record written to an output data file for each model run. The output data file would then need to be written to a database table format.

In general, simulation models would be used in an advanced site-specific risk assessment, for specific exposure pathways and chemicals of concern based on the results of the initial site-specific risk assessment. The advanced site-specific risk assessment is also known as a *tier 3 risk assessment* in the ASTM risk-based corrective action guides (ASTM, 1995a, 2000).

5.6 SUMMARY

This chapter presents the results of the Spatial Environmental Risk Assessment methodology development. The results of the examples developed for the case study are presented and discussed. An evaluation of the costs and benefits of the SERA methodology is presented. A discussion of the application of the SERA methodology to other facilities is also included. The next chapter presents the Conclusions, Limitations and Recommendations that have been developed as a result of this research.

Chapter 6 : Conclusions and Recommendations

This chapter presents the conclusions, limitations and recommendations based on the development of the Spatial Environmental Risk Assessment (SERA) methodology.

6.1 CONCLUSIONS

The research conducted to develop the data model and calculation framework for the SERA methodology provided the opportunity to investigate a number of difficult issues in the implementation of risk-based decision making at large, complex facilities. Through this work:

- A practical methodology for conducting environmental risk assessment in a spatial and tabular database format with the development of the SERA data model has been accomplished.
- The implementation of the SERA data model and calculation methods for the evaluation of multiple source areas affecting the same point of demonstration using analytical models to describe chemical fate and transport in the environment has been developed and implemented.
- A spatial and tabular database for managing the site conceptual model information, including the meta data for modeling, throughout the risk evaluation has been developed and implemented.
- A probabilistic calculation method and tracking of the necessary information to identify uncertainties throughout the risk evaluation and to quantify the uncertainties in the exposure pathway evaluation has

been developed and implemented. The uncertainty analysis is an analytical method that is commensurate with the complexity of the fate and transport models and incorporates data collected for the facility, reference information for parameter values and professional judgment on the variability of parameter values.

- The visualization benefits of the SERA methodology using a case study facility to develop example applications of the spatial environmental risk assessment methodology and exposure pathway analyses are presented.

Overall, the SERA methodology provides significant advantages over current risk assessment practice. These advantages include:

- Effective data management and accessibility,
- Visualization of the site conceptual model,
- Visualization of the modeling results and the uncertainty associated with the results,
- Capacity to analyze many multiple exposure pathways at one time,
- Procedures to analyze multiple source areas impacting one point of demonstration and comparison of the multiple contributions,
- Links to the decision analysis framework to support risk management decisions,
- Ease of implementation for multiple modeling scenarios,
- Support for an iterative calculation process,
- Documentation and repeatability of the model calculations, and

- Scalability for more sophisticated modeling, if needed.

In the deterministic case, the data management support alone is a significant advantage. In the probabilistic case, while the time commitment may be quite large, the advantages are significant. The SERA methodology expands the initial site-specific risk assessment to include estimating uncertainty, which is not typically conducted quantitatively as part of an initial site-specific risk assessment. Understanding the uncertainty in the results is important for focusing on those exposure pathways that contribute the greatest risk and require the most resource-intensive actions.

The following sections present the conclusions for each of the major objective areas: SERA data model, calculation methods, spatial site conceptual models, uncertainty analysis and tracking, and visualization benefits.

6.1.1 SERA Data Model

The SERA data model provides the conceptual framework for the implementation of environmental risk assessment using spatial and tabular databases. The data model is independent of the software applications used in the implementation and in the examples presented in this research. The SERA data model depends, in general, on the quantity and quality of data, as demonstrated in the examples, that are typically available for a facility during a corrective action project. The methodology is logical and systematic. From this perspective, it is considered to be a practical approach.

6.1.2 Calculation Methods

The exposure pathway evaluations are implemented in an attenuation factor format, as described in Chapter 3, using the spatial site conceptual model and the calculation spreadsheets. The benefits include:

- Segmenting the exposure pathways into their calculation components to provide the mechanism by-which multiple source areas affecting one transition point or point of demonstration can be analyzed.
- The independence of the SERA methodology from the fate and transport equations used to describe the chemical movement in the environment. This means that the fate and transport models used in the application presented here are not the only models that can be used. In fact, the data model describes the fate and transport models as the relationship between the starting and ending concentrations for each exposure pathway segment. In a particular facility application there may be sufficient empirical data for a specific exposure pathway to allow the development of a site-specific attenuation factor. This attenuation factor could then be used in the same manner as the model equations used here. The independence of the method from the models allows the flexibility in application to any facility for the SERA methodology to include the fate and transport models that are appropriate to the facility and the data that are available.
- The iterative nature of the risk-based decision making process that is fundamental to the SERA methodology, allows the exposure pathway

analysis calculations to begin early and focus the additional data collection and evaluation activities. In addition, beginning with conservative assumptions for the input parameters allows refinements of input estimates or modeling procedures to be made for a smaller number of exposure pathways, while those exposure pathways that are eliminated, are eliminated based on conservative models and input values that are expected to over-predict the actual concentrations reaching the points of demonstration.

6.1.3 Spatial Site Conceptual Models

The spatial site conceptual model represents a significant improvement over current methods for risk assessment. The innovations include:

- flexibility to analyze large numbers of exposure pathways,
- a connection between the exposure pathway identifiers and the model input parameters,
- an efficient and easy to track method for identifying parameter values and identifying source areas and exposure pathways,
- an advantage over conventional risk assessment, in storing and managing all of the calculation values in the shapefile data tables and the tabular database,
- meta data management through the systematic storage of all of the information required in the risk assessment,
- direct linkages between the software applications, including explicit connection between model values and their derivation through data file

linkages, and tracking of modeling scenarios through a system of unique identifiers, and

- improved reliability over conventional risk assessment methods because it is a transparent modeling environment where each exposure pathway that is analyzed is saved in the database, including all of the input parameters and the decisions about each exposure pathway to facilitate an inventory of all of the evaluated exposure pathways and all of the decisions associated with the exposure pathways.

6.1.4 Uncertainty Analysis

While the uncertainty analysis implemented here is approximate for non-linear equations, the process is an improvement over what is typically completed for an initial site-specific risk assessment, therefore, provides value to the methodology. It may be seen as somewhat time-intensive and potentially unsatisfactory. Ideas for further work on the uncertainty analysis methods are discussed in section 6.3.

The variance estimation method includes not only measured data and reference information, but also incorporates judgments about the expected variability through the estimation of confidence intervals. This procedure, reported by Muchard (1997), provides an opportunity to include the potential sources of variability and to use limited data to begin to understand the uncertainty in the results. In addition, inclusion of the model error variables in the fate and transport models provides the opportunity to account for other sources of variance that are not easily quantified, or for which account is not otherwise taken. The variance evaluation implemented in this research defines a straightforward procedure to assign correlation values for pairs of modeling variables. Based on the case study examples, the inclusion of the correlation between variables can be significant to the results.

The results of the uncertainty analyses for the case study provided the opportunity to evaluate two additional exposure pathway completeness evaluation measures. The first measure, used in this research, is based on the probability of an ending concentration being above the target level. By setting this target probability the results of the probabilistic calculations can be analyzed based on the comparison of the mean value to the target level, but more importantly, the results can be evaluated based on the potential to be below the target level. If the probability that the ending concentration will be below the target level does not meet the target probability then additional actions should be taken. The second measure, used in this research, is the coefficient of variation (c.o.v.) for the ending concentration. It provides a measure of the value of collecting additional information about the input parameters to reduce the uncertainty. If the c.o.v. is low then additional data collection is likely not to be valuable. If the c.o.v. is high, then the opportunity for valuable additional investigation exists.

6.1.5 Visualization Benefits

As with any complex process, the more illustrated the process, discussion and results, the better understood the process is likely to be. Greater understanding is likely for lay people and for engineers and scientists. Authors such as Tufte (1997) have prepared extensive volumes on the value of visualizing data and the methods for using color, shading, line weight and format. The critical first step to the visualization process however, is making all of the data available in a consistent and appropriate format and providing the tools to easily generate different maps, pictures and graphs so that visualizations of data can be experimented with and evaluated.

The spatial environmental risk assessment (SERA) methodology uses an information processing method that makes the data accessible within a software environment that allows experimentation with visualizations. The SERA methodology provides a number of important opportunities to use maps, tables and graphs to bring better understanding to the process of risk assessment. The case study examples were used to present a number of different visualizations of the information being generated and the understanding of the exposure pathways involved. The work completed here included initial evaluations of visualization for risk assessment and presenting uncertainty data. Additional investigations are needed in this area and are discussed in section 6.3.

The GIS is designed to link the site conceptual model and data tables to maps, so locations of predicted concentrations can be easily presented. The implementation of a spatial site conceptual model provides an opportunity to visualize all of the exposure pathways that are included in the evaluations and the results of the evaluation using status identifiers (e.g., complete, not complete or to be evaluated) to code the various exposure pathways in an area of a facility.

The parameter values for the fate and transport models were plotted on graphs and associated with maps. The maps of environmental data were used to interpret locations of source areas and select parameter values. The uncertainty information (e.g., the c.o.v. and the probability of being below the target level) for the parameter values and for the ending concentrations were plotted on maps and provided in tables. These proved to be important to evaluate the veracity of the exposure pathway completeness evaluations and to illuminate the areas where additional investigation may yield refined parameter values and higher confidence in the exposure pathway results.

6.2 LIMITATIONS OF THE SERA METHODOLOGY

The SERA methodology was developed with the goal of providing better methods for implementing environmental risk assessment at large, complex facilities. The method achieves that goal in a number of dimensions, as discussed in section 6.1. However, it is important to recognize that there are limitations to these methods.

In the GIS application, the development of the source area objects and the receptor areas, (e.g., the surface water mixing and outdoor air mixing boxes) are based on the free-hand drawing with the irregular polygon tool; these polygons maybe more approximate than might be needed. Additional methods to automate the development of these polygons based on available data could be beneficial.

The probabilistic case was formulated in a rigid structure to accommodate the development of the necessary partial derivatives for the first order uncertainty analysis. This makes the implementation of the methodology somewhat cumbersome. Before the exposure pathway evaluations can be implemented, the group structures must be defined and the derivative equations worked-out. In addition, for the groundwater to surface water exposure pathway group there were three pathway segments and more than 10 variables involved in the calculation. This made the spreadsheet development more challenging than the others and the development of the derivative equations time-consuming. In addition, the structure is limited to six source areas, so if during the risk assessment additional contributing source areas are identified, and the number for a group is larger than six, the calculation spreadsheets and the data collection queries would have to be updated.

The implementation of the first order analysis requires the estimation of variance data for all of the parameters and consideration of correlations. The structure for including correlation here is a simplified approach and may not be flexible enough for other applications. In order to considered correlation between variables in multiple segments (e.g., two different hydraulic conductivities for two segments) in one exposure pathway group, and among the variables for different exposure pathways in the groups, modifications to the existing spreadsheets would be required. The necessary modifications were made in order to implement the additional calculations presented in section 5.2.8, however, the variance calculation spreadsheets are not designed to be universally applicable.

6.3 RECOMMENDATIONS FOR FUTURE WORK

The work completed during this research represents the beginning steps in the implementation of the spatial and tabular database model for environmental risk assessment. The following items have been identified as additional development work that could be implemented in order to enhance the performance and utility of the SERA methodology.

- Develop additional procedures and codes to interface the spatial site conceptual model with non-Excel based transport models.
- Develop additional procedures and codes to incorporate more a flexible calculation framework for the first order uncertainty analysis. The current probabilistic calculation is limited to the group scenarios that have been identified. An expanded implementation of the methodology might include the development of a more flexible probabilistic calculation case. This would likely be implemented as a separate software program in Visual Basic or C++ that would be interfaced with the spatial and tabular databases.
- For the implementation of the first order uncertainty analysis, a method for further linearizing the fate and transport models before implementing the mean and variance calculations may provide a better estimate of the parameters of the ending concentrations. The calculations would be implemented in natural logarithm space instead of real space as was implemented in this research. The natural logarithms of the attenuation factor equations and all of the input variables would be needed. The results would be the natural logarithms of the mean ending concentrations and the variance in the ending concentrations.

- Future enhancements in the GIS software may be useful in streamlining the connections between the applications. The site conceptual model tabular database and the spatial database may ultimately be stored in one *geospatial database*. The new ESRI Arc/INFO 8.0 operating environment, through the geospatial database, more fully integrates the spatial and tabular databases (Zeiler, 1999). In addition, the changes in the GIS software will allow for the implementation of scientific calculations in Visual Basic or C++ that are tied directly to the spatial objects as method of these objects. The spatial objects can be programmed to have behaviors so that they "function" in the modeling environment as they do in reality. For example, a groundwater transport segment could be given a behavior that implements the Domenico solution to the advection dispersion equation so the starting concentration and the ending concentration are properties of the transport segment and are calculated within the GIS. This may also foster a "cleaner" connection between the spatial and the functional features of each component of the site conceptual model.
- It should be the case that implementing the spatial environmental risk assessment methodology for a full-scale risk-based evaluation would provide an analysis tool to the participants that serves to inform the decision process by promoting a better understanding of the data that are available and particularly where there are data gaps. However, this was not rigorously tested in this research. An evaluation of the

effectiveness of the communication with a typical large facility stakeholder group that includes a diverse range of participants could be conducted to determine the value of the visualizations.

Appendices

Apendix A : Algorithms for Transport Segments

This Appendix presents the algorithms used in the application of the SERA methodology presented in this research. In the equations, the ending concentrations in the various environmental media were given different variable names in order to distinguish them. Table A.1 includes the listing of the variable names and their descriptions.

Table A.1 - Ending Concentration Variable Names

Variable	Description
CA	Groundwater concentration, NAPL partitioning
CB	Groundwater concentration, groundwater transport
CC	Surface Water concentration, surface water mixing
CD	Soil Vapor concentration, NAPL volatilization
CE	Vapor concentration, indoor air
CF	Vapor concentration, outdoor air
CG	Vapor concentration, indoor air, grouped exposure pathway
CH	Vapor concentration, outdoor air, grouped exposure pathway

A.1 NON-AQUEOUS PHASE LIQUID (NAPL) PARTITIONING

This equation represents the NAPL dissolving in groundwater, creating concentrations of chemicals of concern in groundwater (**CA**). The equation is based on equilibrium partitioning, which is represented by the chemical's aqueous solubility (**Sol**). Because the chemicals of concern are part of an oil-phase mixture (i.e., the non-aqueous phase liquid (NAPL), the partitioning to groundwater is limited by the mole fraction of the chemical in the oil-phase (**NMF**), (Charbeneau, 2000).

$$\mathrm{CA}\left(\frac{\mathrm{mg}}{\mathrm{L}}\right) = \mathrm{NMF} \times \mathrm{Sol}$$ **Equation A.1**

The derivative needed for the variance calculation is:

$$\frac{\partial \mathrm{CA}}{\partial \mathrm{NMF}} = \mathrm{Sol}$$ **Equation A.2**

Variable List:
NMF - Non aqueous phase liquid (NAPL) mole fraction (dimensionless) of chemical of concern in the oil-phase.

Constant List:
Sol - aqueous solubility for a chemical of concern (mg/L).

A.2 NON-AQUEOUS PHASE LIQUID (NAPL) VOLATILIZATION

This equation represents the NAPL volatilizing into the vadose zone air-filled porosity, creating concentrations of chemicals of concern in soil vapor (**CD**). The equation is based on equilibrium partitioning. Because the chemicals of concern are part of an oil-phase mixture the partitioning is limited by the mole fraction of the chemical in the oil-phase. If the vapor is assumed to be an ideal gas, then the vapor concentration is defined as:

$$CD = NMF \times \frac{P_V \times MW}{R \times T}$$ **Equation A.3**

where
P_v - vapor pressure (atm)
MW - molecular weight (g/mol)
T- temperature (K)
R - universal gas constant (8.2×10^{-5} atm-m^3/(mol-K)).

The Henry's law coefficient (H') is defined as the ratio of vapor pressure to molar solubility (**S**), (Charbeneau, 2000).

$$H'\left(\frac{atm - m^3}{mol}\right) = \frac{P_v}{S}$$ **Equation A.4**

where
S - molar solubility (mol/m^3).

The dimensionless Henry's coefficient (**Henrys**) is

$$Henrys = \frac{H'}{R \times T} = \frac{P_v}{S \times R \times T}$$ **Equation A.5**

The molar solubility (S) is related to the aqueous solubility (Sol):

$$S\left(\frac{mol}{m^3}\right) = \frac{Sol\left(\frac{mg}{L}\right)}{MW\left(\frac{g}{mol}\right)} \times \left(\frac{1g}{1000mg}\right) \times \left(\frac{1000L}{1m^3}\right)$$ **Equation A.6**

The vapor concentration can be written as

$$CD\left(\frac{mg}{m^3}\right) = NMF \times Henrys \times MW \times \left(\frac{Sol\left(\frac{mg}{L}\right)}{MW}\right) \times \left(\frac{1000L}{m^3}\right)$$

$$\mathrm{CD}\ (\frac{\mathrm{mg}}{\mathrm{m}^3}) = \mathrm{NMF} \times \mathrm{Henrys} \times \mathrm{Sol} \times \frac{1000\mathrm{L}}{1\mathrm{m}^3}$$ **Equation A.7**

The derivatives needed for the variance calculation are:

$$\frac{\partial \mathrm{CD}}{\partial \mathrm{NMF}} = \mathrm{Henrys} \times \mathrm{Sol} \times 1000$$ **Equation A.8**

$$\frac{\partial \mathrm{CD}}{\partial \mathrm{Henrys}} = \mathrm{NMF} \times \mathrm{Sol} \times 1000$$ **Equation A.9**

Variable List:
Henrys - Henry's law partitioning coefficient (dimensionless)
NMF - Non aqueous phase liquid (NAPL) mole fraction (dimensionless) of chemical of concern in the oil-phase.

Constant List:
Sol - aqueous solubility for a chemical of concern (mg/L).

A.3 GROUNDWATER TRANSPORT

The lateral transport of chemicals of concern with groundwater flow can be modeled using any of a large number of solutions to the advection dispersion equation. The form of the boundary conditions and the configuration of the source area will define how the solution is written. The solution used in these calculations is by P.A. Domenico (Domenico, 1987). The source is conceptualized as a vertical plane perpendicular to the groundwater flow direction. It is described by a depth of the source zone (**Depth**) and a width of the source zone (**Width**). The groundwater source concentration (**CA**) is constant over the entire plane and constant over time. Figure A.1 illustrates the groundwater model source configuration.

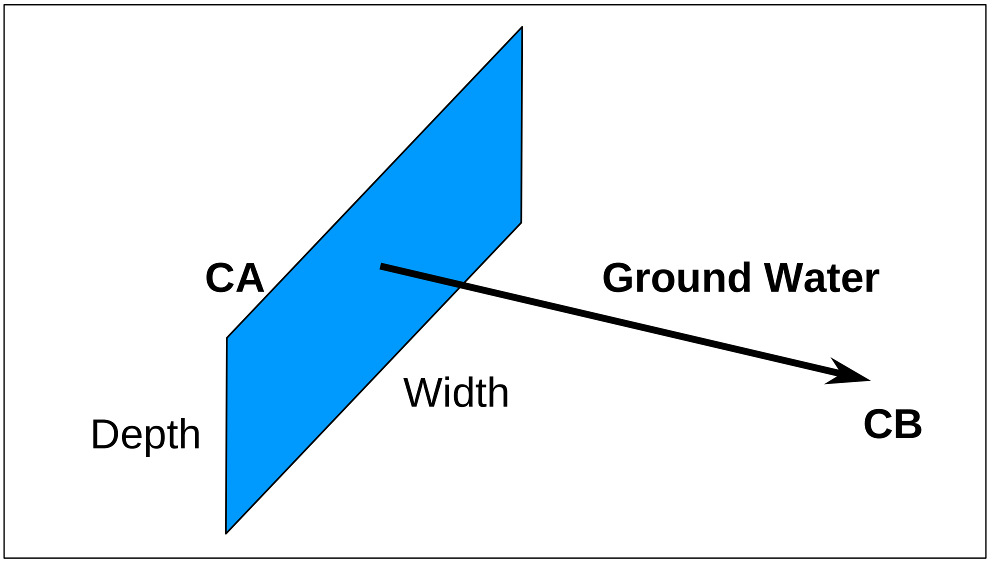


Figure A.1 - Groundwater Model Source Configuration

The equation is a steady-state solution, representing the concentration at the ending point (**CB**) at longer times. First order decay is included. In addition, the equation is solved for the center-line concentration because this will represent the highest concentration at any specified distance downgradient. The ratio of the decay constant (**Degrad**) to the chemical velocity (**Vcoc**), and the distance away from the source (**sDist**), have the greatest effect on the ending concentration. The aquifer is assumed to be isotropic and homogeneous.

$$CB = CA \times \exp\left[\frac{sDist}{2 \times \alpha_x}\left(1 - \sqrt{1 + 4\frac{Degrad \times \alpha_x}{V_{COC}}}\right)\right] \times \operatorname{erf}\left(\frac{Width}{4\sqrt{\alpha_y \times sDist}}\right)\operatorname{erf}\left(\frac{Depth}{4\sqrt{\alpha_z \times sDist}}\right)$$ **Equation A.10**

Dispersion is included in three dimensions. The relationships of the longitudinal (α_x,), lateral (α_y,) and vertical (α_z) dispersivities to the transport distance are based on a U.S. EPA estimation formula (Charbeneau, 2000).

$$\begin{aligned} \alpha_x &= 0.1 \times sDist \\ \alpha_y &= 0.33 \times \alpha_x = 0.033 \times sDist \\ \alpha_z &= 0.05 \times \alpha_x = 0.005 \times sDist \end{aligned}$$ **Equation A.11**

The chemical velocity is defined as the ratio of the groundwater velocity to the chemical-specific retardation coefficient.

$$V_{COC} = \frac{V_{GW}}{Rc} = \frac{HydCond \times GWGrad}{TotalPor \times Rc}$$ **Equation A.12**

The retardation coefficient (**Rc**) is based in the partitioning between the soil in the saturated zone and the groundwater. The chemical's organic carbon partitioning (**Koc**) and the fraction of organic carbon in the soil (**SatFoc**) govern the amount of partitioning between the water and the soil.

$$\mathrm{Rc} = 1 + \left(\frac{\mathrm{ParDen} \times (1 - \mathrm{TotalPor}) \times \mathrm{Koc} \times \mathrm{SatFoc}}{\mathrm{TotalPor}} \right)$$

Equation A.13

Substituting the relationships for the dispersivities, the chemical velocity and the retardation coefficient, the resulting Equation A.14 is the groundwater transport equation used in the calculations. Included in Equation A.14 are two model error variables. These are the mulitplicative error term (**NyGW**) and the additive error term (**EyGW**). Because any model is a simplification of the complex processes in a real groundwater flow system, and because the assumption of an isotropic, homogeneous aquifer typically does not reflect reality, the model error variables are used as a simple method to calibrate the model to fit the site-specific groundwater flow.

$$\mathrm{CB}\left(\frac{\mathrm{mg}}{\mathrm{L}}\right) = \mathrm{CA} \times \mathrm{NYGW} \times \exp\left[5 \times \left(1 - \sqrt{\begin{array}{l} 1 + \dfrac{0.4 \times \mathrm{Degrad} \times \mathrm{sDist}}{\mathrm{HydCond} \times \mathrm{GWGrad} \times 86400} \\ \times \mathrm{TPor} + \mathrm{ParDen} \times (1 - \mathrm{TPor}) \\ \times \mathrm{Koc} \times \mathrm{SatFoc} \end{array}} \right) \right] \times$$
$$\mathrm{erf}\left(\frac{\mathrm{Width}}{0.7303 \times \mathrm{sDist}} \right) \times \mathrm{erf}\left(\frac{\mathrm{Depth}}{0.3001 \times \mathrm{sDist}} \right) + \mathrm{EyGW}$$

Equation A.14

The derivatives are included in the code listing from the Excel Visual Basic Application presented in section A.8.

Variable List:
CA - source area groundwater concentration (mg/L)
Degrad - first order degradation constant (day^{-1})

Depth - source zone depth (cm)
EyGW - additive model error variable (mg/L)
GWGrad - groundwater gradient (cm/cm)
HydCond - saturated zone hydraulic conductivity (cm/sec)
Koc - organic carbon partitioning coefficient (L/kg)
NyGW - multiplicative model error variable (dimensionless)
SatFoc - fraction organic carbon (g oc/g soil)
sDist - groundwater travel distance (cm)
TPor - saturated zone porosity (dimensionless)
Width - source zone width (cm).

Constant List:
ParDen - soil particle density (g/cm^3).

A.4 SOIL VAPOR MIGRATION TO OUTDOOR AIR

The vertical transport of chemicals of concern in the vadose zone soils to outdoor air (**CF**) is modeled using the equations presented by Johnson and Ettinger, (1991) and ASTM, (1995a). The model assumes diffusive, one-dimensional transport of vapors through the porosity of the vadose zone. The outdoor air is modeled as a mixing box. A constant wind velocity (**WSpeed**) is used along with a source area width (**Owidth**) that is perpendicular to the wind direction to define the mixing box. The source vapor concentration (**CD**) is constant in time. The ambient air concentration of the chemical of concern, away from the source area, is assumed to be zero. The equation is a steady state equation. Figure A.2 shows a schematic of the calculation assumptions.

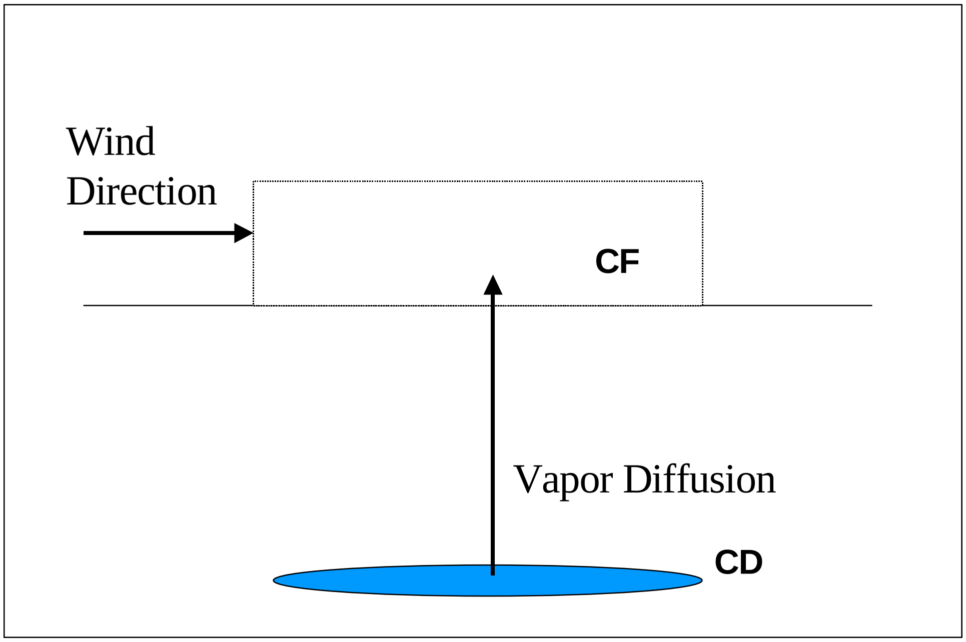


Figure A.2 - Soil Vapor to Outdoor Air Schematic

$$CF\left(\frac{mg}{m^3}\right) = \frac{CD \times NyO}{\left(1 + \frac{WSpeed \times MixHt \times Lsout}{OWidth \times Deff}\right)} + EyO \qquad \textbf{Equation A.15}$$

Equation A.15 includes a multiplicative model error variable (**NyO**) and an additive model error variable (**EyO**). As with the groundwater transport equation, these variables are used to assist in fitting the equation to the site-specific situation. The effective diffusion coefficient (**D_{eff}**) through the vadose zone soil is based on the molecular diffusion in air (**D_{air}**) and in water (**D_{wat}**) for each chemical, the porosity of the soil (**PorVad**), and the water filled porosity

(**WCVad**). Equation A.16 uses the Millington-Quirk expression for diffusion through porous media (Johnson and Ettinger, 1991).

$$\mathrm{Deff} = \mathrm{Dair} \times \frac{(\mathrm{PorVad} - \mathrm{WCVad})^{3.33}}{\mathrm{PorVad}^2} + \frac{\mathrm{Dwat}}{\mathrm{Henrys}} \times \frac{\mathrm{WCVad}^{3.33}}{\mathrm{PorVad}^2} \qquad \textbf{Equation A.16}$$

The total porosity, water-filled porosity and the wind speed (**WSpeed**) have the greatest effect on the ending outdoor air concentration.

The derivatives are included in the code listing from the Excel Visual Basic Application presented in section A.8.

Variable List:
CD - vapor concentration (mg/m^3)
EyO - additive model error variable (mg/m^3)
Henrys - Henry's law partitioning coefficient (dimensionless)
Lsout - depth to subsurface vapor source (cm)
NyO - mulitplicative model error variable (dimensionless)
OWidth - source area width perpendicular to wind direction (cm)
PorVad - vadose zone total porosity (dimensionless)
WCVad - vadose zone water content (dimensionless)
WSpeed - ambient wind speed (cm/sec).

Constant List:
Dair - Diffusion coefficient for the chemical in air (cm^2/sec)
Dwat - Diffusion coefficient for the chemical in water (cm^2/sec)
MixHt - height above ground surface for outdoor air exposure (cm).

A.5 SOIL VAPOR MIGRATION TO INDOOR AIR

The vertical transport of chemicals of concern in the vadose zone soils to an indoor environment is modeled in the same manner as the soil vapor migration to outdoor air with the outdoor air mixing box being replaced by a building. The equations used are based on Johnson and Ettinger, (1991), and, (1995a). In addition to the diffusion through the vadose zone soils, there is diffusion through the building foundation. This is modeled as diffusion through cracks in the foundation that are filled with porous media. Figure A.3 shows a schematic of the calculation assumptions.

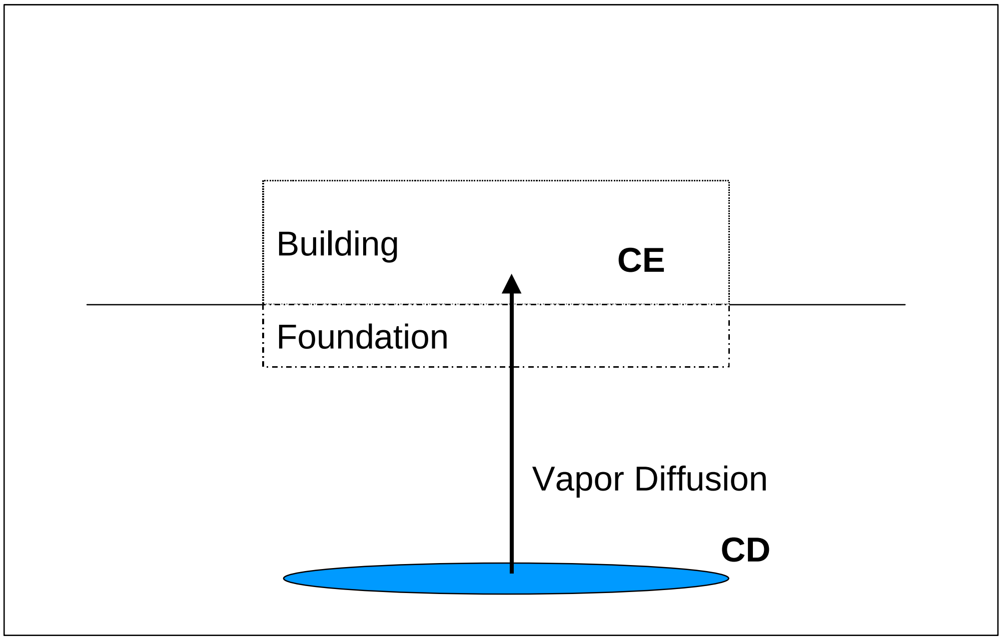


Figure A.3 - Schematic for Soil Vapor Migration to Indoor Air

The indoor air ending concentration (**CE**) is calculated using Equation A.17.

$$\mathrm{CE} = \frac{\mathrm{CD} \times \mathrm{TE1} \times \mathrm{NyI}}{(1 + \mathrm{TE1} + \mathrm{TE2})} + \mathrm{EyI}$$ **Equation A.17**

Equation A.17 includes a mulitplicative error variable (**NyI**) and an additive error variable (**EyI**). The terms **TE1** and **TE2** are defined as follows:

$$\mathrm{TE1} = \frac{\dfrac{\mathrm{Deff,vad}}{\mathrm{Ls}}}{\mathrm{ER} \times \mathrm{LB}}$$ **Equation A.18**

$$\mathrm{TE2} = \frac{\mathrm{Deff,vad} \times \mathrm{WallT}}{\mathrm{Deff,wall} \times \mathrm{Ls} \times \mathrm{AF}}$$ **Equation A.19**

The effective diffusion in the vadose zone (**Deff,vad**) is calculated using Equation A.16. The effective diffusion coefficient through the foundation, or basement wall (**Deff,wall**) is also calculated using Equation A.16. The porosity of the foundation cracks (**PorWall**) is substituted for the vadose zone porosity (**PorVad**) and the water-filled porosity of the foundation cracks (**WCWall**) is substituted for the vadose zone water-filled porosity (**WCWall**).

The areal fraction of cracks in the foundation (**AF**) and the water-filled porosities have the greatest effect on the ending indoor air concentration.

The derivatives are included in the code listing from the Excel Visual Basic Application presented in section A.8.

Variable List:
AF - areal fraction of cracks in the foundation (dimensionless)
CD - source area vapor concentration (mg/m^3)
EyI - additive model error variable (mg/m^3)
Henrys - Henry's law partitioning coefficient (dimensionless)
Ls - depth to subsurface vapor sources (cm)

NyI - mulitplicative model error variable (dimensionless)
PorVad - vadose zone total porosity (dimensionless)
PorWall - porosity of the soils in the foundation cracks (dimensionless)
WCVad - vadose zone water content (dimensionless)
WCWall - water content of soils in the cracks (dimensionless).

Constant List:
ER - Indoor air exchange rate (sec^{-1})
LB - Indoor enclosed space volume to area ratio (cm)
WallT - Foundation thickness for a building (cm).

A.5.1 Air Mixing Models

For both the indoor air and outdoor air exposure pathways, group exposure pathways are evaluated as the sum of the individual ending concentrations since the exposure pathway models themselves incorporate mixing in the indoor environment or in the outdoor air breathing zone box. The resulting indoor air concentration (**CG**) and the resulting outdoor air concentration (**CH**) are compared to the saturated vapor concentration. If the calculated concentration is below the saturated vapor concentration then the calculated concentration is used as the estimate of the exposure point concentration (i.e., the point of demonstration concentration), if the calculated concentration is above saturation, then the saturated vapor concentration is used as the estimate of the exposure point concentration (C_E). Equation A.20 is used for the group exposure pathways for indoor air. Equation A.21 is used for the group exposure pathways for outdoor air.

$$CG = \sum_{i=1}^{N} CE \qquad \textbf{Equation A.20}$$

$$\mathrm{CH} = \sum_{i=1}^{N} \mathrm{CF}$$ **Equation A.21**

where
N - number of exposure pathways in the group.

$$\begin{aligned}&\text{If CG}\left(\text{or CH}\right) \le C_{SAT} \text{ THEN } C_E = \text{CG}\left(\text{or CH}\right)\\&\text{ELSE } C_E = \text{NMF} \times \text{Henrys} \times \text{Sol} \times \frac{1000\text{L}}{1\text{m}^3}\end{aligned}$$ **Equation A.22**

The saturation check is given in Equation A.22. The aqueous solubility check is developed in a parallel manner.

A.6 SURFACE WATER MIXING MODEL

The mixing of groundwater flow with surface water flow for a resulting concentration in the surface water is based on a simple mass balance and the advective flux from the groundwater. Figure A.4 shows a schematic of the calculation assumptions.

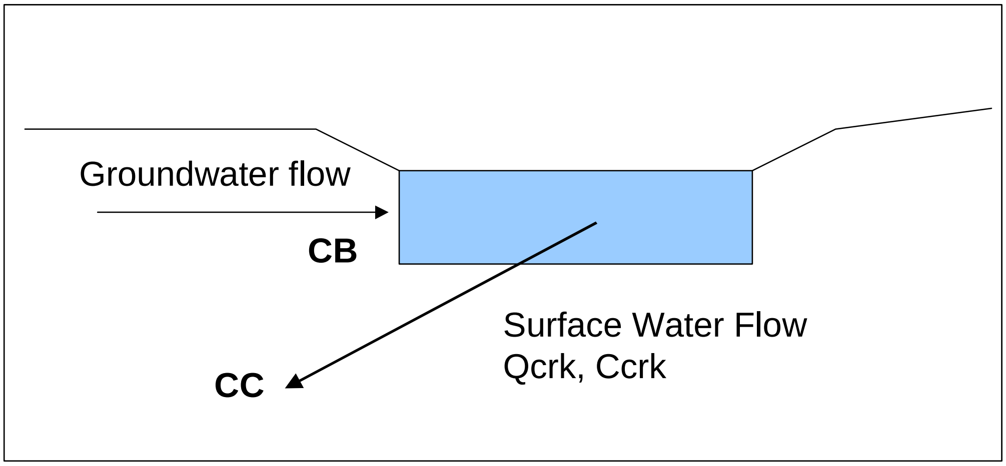


Figure A.4 - Schematic Cross Section of Groundwater Mixing in Surface Water Calculation

The water flow balance for the surface water is given by Equation A.23.

$$Q_{crk,out} = Q_{gw} + Q_{crk,in}$$

Equation A.23

The mass balance for the chemical is given by Equation A.24.

$$Q_{crk,out} \times CC = Q_{gw} \times CB + Q_{crk,in} \times C_{crk}$$

Equation A.24

The groundwater advective flux (Q_{gw}*CB) can be quantified using an equation presented by Charbeneau, (2000):

$$Q_{gw} \times CB = \sqrt{2\pi} \times q_X \times H \times CB \times \sigma$$

Equation A.25

The groundwater flux (**q_X**) can be written in terms of the hydraulic conductivity (**HydCond**) and the groundwater gradient (**GWGrad**) as presented in Equation A.26.

$$q_X = \mathrm{HydCond} \times \mathrm{GWGrad}$$

Equation A.26

The groundwater concentrations along the boundary with the surface water can be assumed to have a Gaussian distribution with a standard deviation (**σ**). The standard deviation can be estimated based on the source area width (**Width**), as presented in Equation A.27.

$$\sigma = 0.25 \times \mathrm{Width}$$

Equation A.27

Solving Equation A.24 for the surface water ending concentration and substituting Equation A.25 through Equation A.27 into Equation A.24, an expression for the surface water concentration (**CC**) is developed. Based on the results presented in Romanek *et al*., (1999) and Kim *et al*., (2000), the surface water flows can be taken to be much larger than the groundwater flows. This simplifies the mass balance equation.

$$\mathrm{CC}\left(\frac{\mathrm{mg}}{\mathrm{L}}\right) = \frac{\left(C_{crk} \times Q_{crk} + \sqrt{2\pi} \times \left(\begin{array}{l}\mathrm{HydCond} \times \mathrm{GWGrad} \times \mathrm{H} \times \\ \mathrm{CB} \times 0.25 \times \mathrm{Width} \times \frac{1\mathrm{L}}{1000\mathrm{cm}^3}\end{array}\right)\right) \times \mathrm{NySW}}{Q_{crk}} + \mathrm{EySW}$$

Equation A.28

Equation A.28 includes a multiplicative model error variable (**NyGW**) and an additive model error variable (**EyGW**).

The derivatives are included in the code listing from the Excel Visual Basic Application presented in section A.8.

Variable List:
CB - groundwater concentration at the surface water (plume center line concentration) (mg/L)
Ccrk - background concentration of a chemical of concern in the surface water (mg/L)
EySW - additive model error variable (mg/L)
GWGrad - groundwater gradient (cm/cm)
H - plume mixing depth at the surface water (cm)
HydCond - saturated zone hydraulic conductivity (cm/sec)
NySW - multiplicative model error variable (dimensionless)
Qcrk - representative flow in the surface water (L/sec)
Width - source area width (cm).

Constant List:
None.

A.7 PROGRAMMING QUALITY CONTROL

A numerical comparison was performed to verify the results of the equations programmed in the Excel Add-In. The results of the derivative equations were compared to the base transport equations. As presented in section A.8, 52 derivative functions were developed in the Excel Add-In to implement the variance calculations. Each derivative was reviewed for accuracy both at the derivation stage and once the function was programmed in Visual Basic to be sure that no transposition errors had occurred. Then, approximately 20 percent of the equations were selected to be numerically tested against the base equations. The derivative functions to be tested were selected from all of the transport segments.

In addition, since the derivatives of one equation for each of the input variables are related, the verification of one variable for an equation was taken to validate the calculus that went into developing the derivatives for all of the variables for that base equation.

Each chosen derivative function was tested at two points around the mean value for the given input variable and the results for the derivative compared to the change in the base equation concentration divided by the change in the input variable. Equation A.29 gives the test condition that was used.

$$\begin{aligned}&\text{IF } \frac{\Delta C}{\Delta X} \approx \frac{\partial C}{\partial X} \text{ THEN } F'(X) \text{ IS CORRECT} \\ &\quad \text{ELSE CHECK } F'(X)\end{aligned}$$

Equation A.29

The results of the numerical checks are presented in Table A.2. As can be seen in the table the calculated values for the derivatives very closely match the results of the derivative functions.

Table A.2 - Numerical Comparison of Derivative Results

Patitioning Equations		**Diffusion Coefficient**	
Example 1		**Example 3**	
Equation	Function Name:	Equation	Function Name:
dCD/dHenrys	dCDH	dDeff/dTPor	dDeffTPor
Calculated	Function	Calculated	Function
21875	21875	0.0173117	0.017330625
21875	21875	0.017303185	0.017313606
Groundwater Transport		**Example 4**	
Example 7B		Equation	Function Name:
Equation	Function Name:	dDeff/dWCPor	dDeffWCPor
dKX/dHydCond	dKXHydC	Calculated	Function
Calculated	Function		
		-0.02276018	-0.022761951
0.000421045	0.000421269	-0.022756636	-0.022754864
0.000421347	0.000421123		
		Example 5	
Example 9		Equation	Function Name:
Equation	Function Name:	dDeff/dH	dDeffH
dKX/dsDist	dKXsDist	Calculated	Function
Calculated	Function		
		-1.00205E-05	-1.00159E-05
-1.83543E-09	-1.83535E-09	-1.00296E-05	-1.00342E-05
-1.83551E-09	-1.83566E-09		
		Indoor Air	
Example 10		**Example 6**	
Equation	Function Name:	Equation	Function Name:
dQX/dWidth	dQXWidth	dCE/dT1	dCET1
Calculated	Function	Calculated	Function
7.53052E-05	7.53052E-05	461.22466	461.22024
7.53052E-05	7.53052E-05	461.22908	461.23792

Table A.2 - Comparison of Numerical Derivative Results (continued)

Surface Water Mixing		**Outdoor Air**	
Example 7A		**Example 2**	
Equation	Function Name:	Equation	Function Name:
dCCgX/dHydCond	dCCgXIHydC	dCF/dCD	dCFCD
Calculated	Function	Calculated	Function
2.413E-04	2.419E-04	8.43961E-07	8.43961E-07
2.413E-04	2.419E-04	8.43961E-07	8.43961E-07
Example 8			
Equation	Function Name:		
dCCgXI/dH	dCCgXIH		
Calculated	Function		
3.84015E-07	3.84015E-07		
3.84015E-07	3.84015E-07		

A.8 VISUAL BASIC APPLICATION CODE LISTING

The fate and transport equations included in this Appendix were coded in an Excel Visual Basic Application as User-Defined functions. These are stored in an Excel Add-In.

For the probabilistic calculations, a derivative equation is needed for each ending concentration with respect to each variable involved in the calculation of that ending concentration. In the examples used for the case study, there are 3 exposure pathways with 6 segment calculations. For the groundwater transport to surface water mixing there are 20 variables, for the volatilization to indoor air there are 11 variables and for the volatilization to outdoor air there are 10 variables. This results in a minimum of 41 derivative equations. Because the original equations are not basic functions, intermediate derivatives were calculated in order to implement the product rule and the quotient rule in the differentiation (Salas and Hille, 1982). In addition, in order to develop the partial derivatives for the groundwater transport equation it was split into four smaller functions that are multiplied together to yield the original equation as shown in Equation A.30.

$$
\begin{aligned}
&\mathrm{HX} = \mathrm{CA} \times \mathrm{NYGW} \\
&\mathrm{KX} = \exp\left[5 \times \left(1 - \sqrt{\begin{array}{l} 1 + \dfrac{0.4 \times \mathrm{Degrad} \times \mathrm{sDist}}{\mathrm{HydCond} \times \mathrm{GWGrad} \times 86400} \\ \times \mathrm{TPor} + \mathrm{ParDen} \times (1 - \mathrm{TPor}) \\ \times \mathrm{Koc} \times \mathrm{SatFoc} \end{array}}\right)\right] \\
&\mathrm{QX} = \mathrm{erf}\left(\frac{\mathrm{Width}}{0.7303 \times \mathrm{sDist}}\right) \\
&\mathrm{RX} = \mathrm{erf}\left(\frac{\mathrm{Depth}}{0.3001 \times \mathrm{sDist}}\right) \\
&\mathrm{CB} = \mathrm{HX} \times \mathrm{KX} \times \mathrm{QX} \times \mathrm{RX} + \mathrm{EYGW}
\end{aligned}
$$

Equation A.30

The error function is an integral, exponential function that cannot be integrated analytically:

$$
\mathrm{erf}(X) = \frac{2}{\sqrt{\Pi}} \int_0^X e^{-\xi^2} \, d\xi
$$

Equation A.31

(Charbeneau, 2000).

A power series with 5 terms is used in place of the exponential function and the integral is evaluated as 5 separate terms (Salas and Hille, 1982). Equation

A.32 presents the power series. This substitution is used to develop the needed partial derivatives for the QX and RX terms in the groundwater transport equation.

$$e^{-x^2} = 1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} + \frac{x^8}{4!}$$ **Equation A.32**

As a result of the intermediate steps, there are 53 derivative equations in the SERA Excel Add-In.

The partial derivatives of all of the segment ending concentrations with respect to each of the variables that contribute to that ending concentration were developed as functions. In the visual basic code, the variables of the partial derivatives are named using a concatenation of the two variable names involved. For example, the derivative function **dCCQcrk** is the partial derivative of the surface water ending concentration (**CC**) with respect to the surface water flow (**Qcrk**). The code is separated into six modules. These are Deriv1, Deriv2, Diffusion, GWFunctions, Transport and Variance. The partial derivatives are generally included in the modules Deriv1 and Deriv2. The diffusion through porous media equation and the associated partial derivatives are included in the module Diffusion. The groundwater fate and transport equations are included in the module GWFunctions. The non-groundwater fate and transport equations are included in the module Transport. The variance calculation equations are included in the module Variance.

The following sections include the code listings for each of the six modules.

A.8.1 Deriv1

```
Public Function dCDNMF(Sol, Henrys)
dCDNMF = Henrys * Sol * 1000
End Function

Public Function dCDH(NMF, Sol)
dCDH = NMF * Sol * 1000
End Function

Public Function dCANMF(Sol)
dCANMF = Sol
End Function

Public Function dCECD(Te1, Te2, NYI)
dCECD = Te1 * NYI / (1 + Te1 + Te2)
End Function

Public Function dCET1(VapConc, Te1, Te2, NYI)
dCET1 = VapConc * NYI * (1 + Te2) / (1 + Te1 + Te2) ^ 2
End Function

Public Function dCET2(VapConc, Te1, Te2, NYI)
dCET2 = -VapConc * Te1 * NYI / (1 + Te1 + Te2) ^ 2
End Function

Public Function dCCfXNYSW(Ccrk)
dCCfXNYSW = Ccrk
End Function

Public Function dCENYI(VapConc, Te1, Te2, NYI)
dCENYI = (VapConc * Te1) / (1 + Te1 + Te2)
End Function

Public Function dCFNYO(VapConc, WSpeed, MixHt, Lsout, OWidth, Deffv)
dCFNYO = VapConc / (1 + WSpeed * MixHt * Lsout / (OWidth * Deffv))
End Function
```

```
Public Function dCFCD(WSpeed, MixHt, Lsout, OWidth, Deffv, NYO)
dCFCD = NYO / (1 + WSpeed * MixHt * Lsout / (OWidth * Deffv))
End Function

Public Function dCFWSpeed(VapConc, WSpeed, MixHt, Lsout, OWidth, Deffv,
NYO)
dCFWSpeed = -(VapConc * NYO * MixHt * Lsout / (OWidth * Deffv)) / (1 +
WSpeed * MixHt * Lsout / (OWidth * Deffv)) ^ 2
End Function

Public Function dCFLsout(VapConc, WSpeed, MixHt, Lsout, OWidth, Deffv,
NYO)
dCFLsout = -(VapConc * NYO * MixHt * WSpeed / (OWidth * Deffv)) / (1 +
WSpeed * MixHt * Lsout / (OWidth * Deffv)) ^ 2
End Function

Public Function dCFOWidth(VapConc, WSpeed, MixHt, Lsout, OWidth, Deffv,
NYO)
dCFOWidth = (VapConc * NYO * MixHt * WSpeed * Lsout / (OWidth ^ 2 *
Deffv)) / (1 + WSpeed * MixHt * Lsout / (OWidth * Deffv)) ^ 2
End Function

Public Function dCFDeffv(VapConc, WSpeed, MixHt, Lsout, OWidth, Deffv,
NYO)
dCFDeffv = (VapConc * NYO * MixHt * WSpeed * Lsout / (OWidth * Deffv ^
2)) / (1 + WSpeed * MixHt * Lsout / (OWidth * Deffv)) ^ 2
End Function

Public Function dCBXI(HX, KX, RX, QX, dHX, dKX, dRX, dQX)
dCBXI = HX * KX * (QX * dRX + RX * dQX) + QX * RX * (HX * dKX + KX
* dHX)
End Function

Public Function dCCfXCcrk(NYSW)
dCCfXCcrk = NYSW
End Function

Public Function dCCgXIHydC(HydC, GWGrad, CB, Width, dCB, H, NYSW,
Qcrk)
dCCgXIHydC = (0.00063 * HydC * GWGrad * Width) * (dCB * H * NYSW /
Qcrk) + (CB * H * NYSW / Qcrk) * (0.00063 * Width * GWGrad)
End Function
```

```
Public Function dCCgXIGWGrad(HydC, GWGrad, CB, Width, dCB, H, NYSW,
Qcrk)
dCCgXIGWGrad = (0.00063 * HydC * GWGrad * Width) * (dCB * H * NYSW /
Qcrk) + (CB * H * NYSW / Qcrk) * (0.00063 * Width * HydC)
End Function

Public Function dCCgXIWidth(HydC, GWGrad, CB, Width, dCB, H, NYSW,
Qcrk)
dCCgXIWidth = (0.00063 * HydC * GWGrad * Width) * (dCB * H * NYSW /
Qcrk) + (CB * H * NYSW / Qcrk) * (0.00063 * HydC * GWGrad)
End Function

Public Function dCCgXICB(HydC, GWGrad, Width, H, NYSW, Qcrk)
dCCgXICB = (0.00063 * HydC * GWGrad * H * Width * NYSW) / Qcrk
End Function

Public Function dCCgXIXI(HydC, GWGrad, dCB, Width, H, NYSW, Qcrk)
dCCgXIXI = (0.00063 * HydC * GWGrad * H * dCB * Width * NYSW) / Qcrk
End Function

Public Function dCFH(VapConc, WSpeed, MixHt, Lsout, OWidth, Deffv, dCD,
dDeffv, NYO)
dCFH = ((1 + WSpeed * MixHt * Lsout / (OWidth * Deffv)) * dCD * NYO +
VapConc * NYO * (WSpeed * MixHt * Lsout * dDeffv / (OWidth * Deffv ^ 2)))
/ (1 + WSpeed * MixHt * Lsout / (OWidth * Deffv)) ^ 2
End Function

Public Function dCEH(VapConc, Te1, Te2, dCD, dTe1, dTe2, NYI)
dCEH = ((1 + Te1 + Te2) * NYI * (VapConc * dTe1 + Te1 * dCD) - (VapConc *
Te1 * NYI) * (dTe1 + dTe2)) / (1 + Te1 + Te2) ^ 2
End Function

Public Function dCEXI(VapConc, Te1, Te2, dTe1, dTe2, NYI)
dCEXI = ((1 + Te1 + Te2) * (VapConc * NYI * dTe1) - (VapConc * NYI * Te1)
* (dTe1 + dTe2)) / (1 + Te1 + Te2) ^ 2
End Function

Public Function dCCgXINYSW(HydC, GWGrad, CB, Width, H, Qcrk)
dCCgXINYSW = (0.00063 * HydC * GWGrad * H * CB * Width) / Qcrk
End Function
```

```
Public Function dCCgXIH(HydC, GWGrad, CB, Width, NYSW, Qcrk)
dCCgXIH = (0.00063 * HydC * GWGrad * CB * Width * NYSW) / Qcrk
End Function

Public Function dCCgXIQcrk(HydC, GWGrad, CB, Width, H, NYSW, Qcrk)
dCCgXIQcrk = (-0.00063 * HydC * GWGrad * H * CB * Width * NYSW) /
Qcrk ^ 2
End Function
```

A.8.2 Deriv2

```
Public Function dT1Ls(Deffv, Ls, ER, LB)
dT1Ls = -Deffv / (ER * LB * Ls ^ 2)
End Function

Public Function dT1Deffv(Ls, ER, LB)
dT1Deffv = 1 / (ER * LB * Ls)
End Function

Public Function dT2Deffv(Deffw, Ls, WallT, AF)
dT2Deffv = WallT / (Deffw * Ls * AF)
End Function

Public Function dT2Deffw(Deffv, Deffw, Ls, WallT, AF)
dT2Deffw = -Deffv * WallT / (Ls * AF * Deffw ^ 2)
End Function

Public Function dT2Ls(Deffv, Deffw, Ls, WallT, AF)
dT2Ls = -Deffv * WallT / (Deffw * AF * Ls ^ 2)
End Function

Public Function dT2AF(Deffv, Deffw, Ls, WallT, AF)
dT2AF = -Deffv * WallT / (Deffw * Ls * AF ^ 2)
End Function

Public Function dfSWHydC(Qcrk, GWGrad, H, NYSW)
dfSWHydC = NYSW * (0.000627 * GWGrad * H / Qcrk)
End Function

Public Function dfSWGWGrad(Qcrk, HydC, H, NYSW)
dfSWGWGrad = NYSW * (0.000627 * HydC * H / Qcrk)
End Function
```

```
Public Function dT1XI(Ls, ER, LB, dDeffv)
dT1XI = dDeffv / (Ls * ER * LB)
End Function

Public Function dT2XI(Ls, AF, WallT, Deffv, Deffw, dDeffv, dDeffw)
dT2XI = (WallT / (AF * Ls)) * (Deffw * dDeffv - Deffv * dDeffw) / (Deffw ^ 2)
End Function
```

A.8.3 Diffusion

```
Public Function Deff(Dair, Dwat, Henrys, TPor, WCPor)
Deff = Dair * (TPor - WCPor) ^ 3.33 / TPor ^ 2 + Dwat / Henrys * WCPor ^ 3.33
/ TPor ^ 2
End Function

Public Function dDeffH(Dwat, Henrys, TPor, WCPor)
dDeffH = -1 / Henrys ^ 2 * Dwat * WCPor ^ 3.33 / TPor ^ 2
End Function

Public Function dDeffTPor(Dair, Dwat, Henrys, TPor, WCPor)
dDeffTPor = Dair * (3.33 *TPor* (TPor - WCPor) ^ 2.33 - 2 * (TPor - WCPor) ^
3.33) / TPor ^ 3 - 2 * Henrys * Dwat * WCPor ^ 3.33 / TPor ^ 3
End Function

Public Function dDeffWCPor(Dair, Dwat, Henrys, TPor, WCPor)
dDeffWCPor = 3.33 * (-Dair * (TPor - WCPor) ^ 2.33 / TPor ^ 2 + Dwat / Henrys
* WCPor ^ 2.33 / TPor ^ 2)
End Function
```

A.8.4 GWFunctions

```
Public Function AlphaX(sDist)
AlphaX = 0.1 * sDist
End Function

Public Function AlphaY(sDist)
AlphaY = 0.0333 * sDist
End Function

Public Function AlphaZ(sDist)
AlphaZ = 0.00563 * sDist
End Function
```

```
Public Function Retard(ParDen, TPor, Koc, SatFoc)
Retard = 1 + (ParDen * (1 - TPor) * Koc * SatFoc / TPor)
End Function

Public Function Vcoc(HydC, GWGrad, TPor, R)
Vcoc = HydC * GWGrad / (TPor * R)
End Function

Public Function dKXsDist(KX, sDist, Degrad, HydC, GWGrad, TPor, ParDen, Koc, SatFoc)
dKXsDist = KX * (-2.5) * 1 / Sqr(1 + (0.4 * Degrad * sDist / (86400 * HydC * GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))) * (0.4 * Degrad / (86400 * HydC * GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))
End Function

Public Function dKXDegrad(KX, sDist, Degrad, HydC, GWGrad, TPor, ParDen, Koc, SatFoc)
dKXDegrad = KX * (-2.5) * 1 / Sqr(1 + (0.4 * Degrad * sDist / (86400 * HydC * GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))) * (0.4 * sDist / (86400 * HydC * GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))
End Function

Public Function dKXHydC(KX, sDist, Degrad, HydC, GWGrad, TPor, ParDen, Koc, SatFoc)
dKXHydC = KX * (2.5) * 1 / Sqr(1 + (0.4 * Degrad * sDist / (86400 * HydC * GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))) * (0.4 * sDist * Degrad / (86400 * HydC ^ 2 * GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))
End Function

Public Function dKXGWGrad(KX, sDist, Degrad, HydC, GWGrad, TPor, ParDen, Koc, SatFoc)
dKXGWGrad = KX * (2.5) * 1 / Sqr(1 + (0.4 * Degrad * sDist / (86400 * HydC * GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))) * (0.4 * sDist * Degrad / (86400 * HydC * GWGrad ^ 2) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))
End Function

Public Function dQXWidth(sDist, Width)
X = Width / (0.7303 * sDist)
Pi = 3.141592654
```

```
dQXWidth = 2 / Sqr(Pi) * 1 / (0.7303 * sDist) * (1 - X ^ 2 + X ^ 4 / 2 - X ^ 6 / 6 +
X ^ 8 / 24)
End Function

Public Function dQXsDist(sDist, Width)
X = Width / (0.7303 * sDist)
Pi = 3.141592654
dQXsDist = 2 / Sqr(Pi) * (-Width / (0.7303 * sDist ^ 2)) * (1 - X ^ 2 + X ^ 4 / 2 -
X ^ 6 / 6 + X ^ 8 / 24)
End Function

Public Function dRXsDist(sDist, Depth)
X = Depth / (0.3001 * sDist)
Pi = 3.141592654
dRXsDist = 2 / Sqr(Pi) * (-Depth / (0.3001 * sDist ^ 2)) * (1 - X ^ 2 + X ^ 4 / 2 -
X ^ 6 / 6 + X ^ 8 / 24)
End Function

Public Function dRXDepth(sDist, Depth)
X = Depth / (0.3001 * sDist)
Pi = 3.141592654
dRXDepth = 2 / Sqr(Pi) * (1 / (0.3001 * sDist)) * (1 - X ^ 2 + X ^ 4 / 2 - X ^ 6 / 6
+ X ^ 8 / 24)
End Function

Public Function dKXTPor(KX, sDist, Degrad, HydC, GWGrad, TPor, ParDen,
Koc, SatFoc)
dKXTPor = KX * (-2.5) * 1 / Sqr(1 + (0.4 * Degrad * sDist / (86400 * HydC *
GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))) * (1 - ParDen * Koc *
SatFoc) * 0.4 * Degrad * sDist / (86400 * HydC * GWGrad)
End Function
Public Function dKXKoc(KX, sDist, Degrad, HydC, GWGrad, TPor, ParDen,
Koc, SatFoc)
dKXKoc = KX * (-2.5) * 1 / Sqr(1 + (0.4 * Degrad * sDist / (86400 * HydC *
GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))) * (ParDen * SatFoc)
* (1 - TPor) * 0.4 * Degrad * sDist / (86400 * HydC * GWGrad)
End Function

Public Function dKXSatFoc(KX, sDist, Degrad, HydC, GWGrad, TPor, ParDen,
Koc, SatFoc)
```

```
dKXSatFoc = KX * (-2.5) * 1 / Sqr(1 + (0.4 * Degrad * sDist / (86400 * HydC * GWGrad) * (TPor + ParDen * (1 - TPor) * Koc * SatFoc))) * (ParDen * Koc) * (1 - TPor) * 0.4 * Degrad * sDist / (86400 * HydC * GWGrad)
End Function

Public Function dHXgwC(NYGW)
dHXgwC = NYGW
End Function

Public Function dHXNYGW(WatConc)
dHXNYGW = WatConc
End Function
```

A.8.5 Transport

```
Public Function CA(NMF, Sol)
CA = NMF * Sol
End Function

Public Function CD(NMF, Sol, Henrys)
CD = NMF * Henrys * Sol * 1000
End Function

Public Function CF(VapConc, WSpeed, MixHt, Lsout, OWidth, Deffv, NYO, EYO)
CF = (VapConc * NYO) / (1 + WSpeed * MixHt * Lsout / (OWidth * Deffv)) + EYO
End Function

Public Function CE(VapConc, Te1, Te2, NYI, EYI)
CE = (VapConc * Te1 * NYI) / (1 + Te1 + Te2) + EYI
End Function

Public Function Term1(Deffv, Ls, ER, LB)
Term1 = (Deffv / Ls) / (ER * LB)
End Function

Public Function Term2(Deffv, Deffw, Ls, WallT, AF)
Term2 = (Deffv * WallT) / (Deffw * Ls * AF)
End Function

Public Function CB(HX, KX, QX, RX, EYGW)
CB = HX * KX * QX * RX + EYGW
```

```
End Function

Public Function CCfX(Ccrk, EYSW, NYSW)
CCfX = Ccrk * NYSW + EYSW
End Function


Public Function KX(sDist, Degrad, HydC, GWGrad, TPor, ParDen, Koc, SatFoc)
KX = Exp(5 * (1 - Sqr(1 + (0.4 * Degrad * sDist / (86400 * HydC * GWGrad) *
(TPor + ParDen * (1 - TPor) * Koc * SatFoc)))))
End Function

Public Function HX(WatConc, NYGW)
HX = WatConc * NYGW
End Function

Public Function CCgXI(HydC, GWGrad, H, CB, Width, NYSW, Qcrk)
CCgXI = 0.00063 * HydC * GWGrad * H * CB * Width * NYSW / Qcrk
End Function

Public Function CBd(WatConc, NYGW, sDist, AlphaX, Degrad, Vcoc)
CBd = WatConc * NYGW * Exp(sDist / (2 * AlphaX) * (1 - Sqr(1 + 4 * Degrad
* AlphaX / Vcoc)))
End Function

Public Function CC(Ccrk, Qcrk, HydC, GWGrad, H, gwConc, Width, NYSW,
EYSW)
CC = ((Ccrk * Qcrk + 2.5066 * (HydC * GWGrad * H * gwConc * 0.00025 *
Width)) * NYSW) / Qcrk + EYSW
End Function
```

A.8.6 Variance

```
Public Function VarI(AI, VarXI)
VarI = AI ^ 2 * VarXI
End Function

Public Function VarIJ(AI, AJ, RHOIJ, VarXI, VarXJ)
VarIJ = AI * AJ * RHOIJ * (VarXI * VarXJ) ^ 0.5
End Function
```

Appendix B : Parameter Estimation

This Appendix provides the rationale and methodology for the development of the input values for the example calculations presented in Chapter 5. The example calculations were prepared for the former BP Oil Marcus Hook Refinery. A brief overview of the Marcus Hook Refinery is also presented in this Appendix.

B.1 MARCUS HOOK REFINERY

In order to provide context and examples for this research the former BP Oil Marcus Hook Refinery was used as a case study. The Marcus Hook Refinery is located in the cities of Marcus Hook and Trainer, Pennsylvania, adjacent to the Delaware River (Figure B.1). The 300-acre facility has been operated as a crude oil refinery and petroleum products storage facility since the early 1900's. BP Exploration & Oil Inc. (BP) owned and operated the facility from 1969 until February 1996 when it was sold to Tosco. As part of this sale, BP has retained responsibilities to address certain environmental conditions that existed at the facility prior to its sale to Tosco.

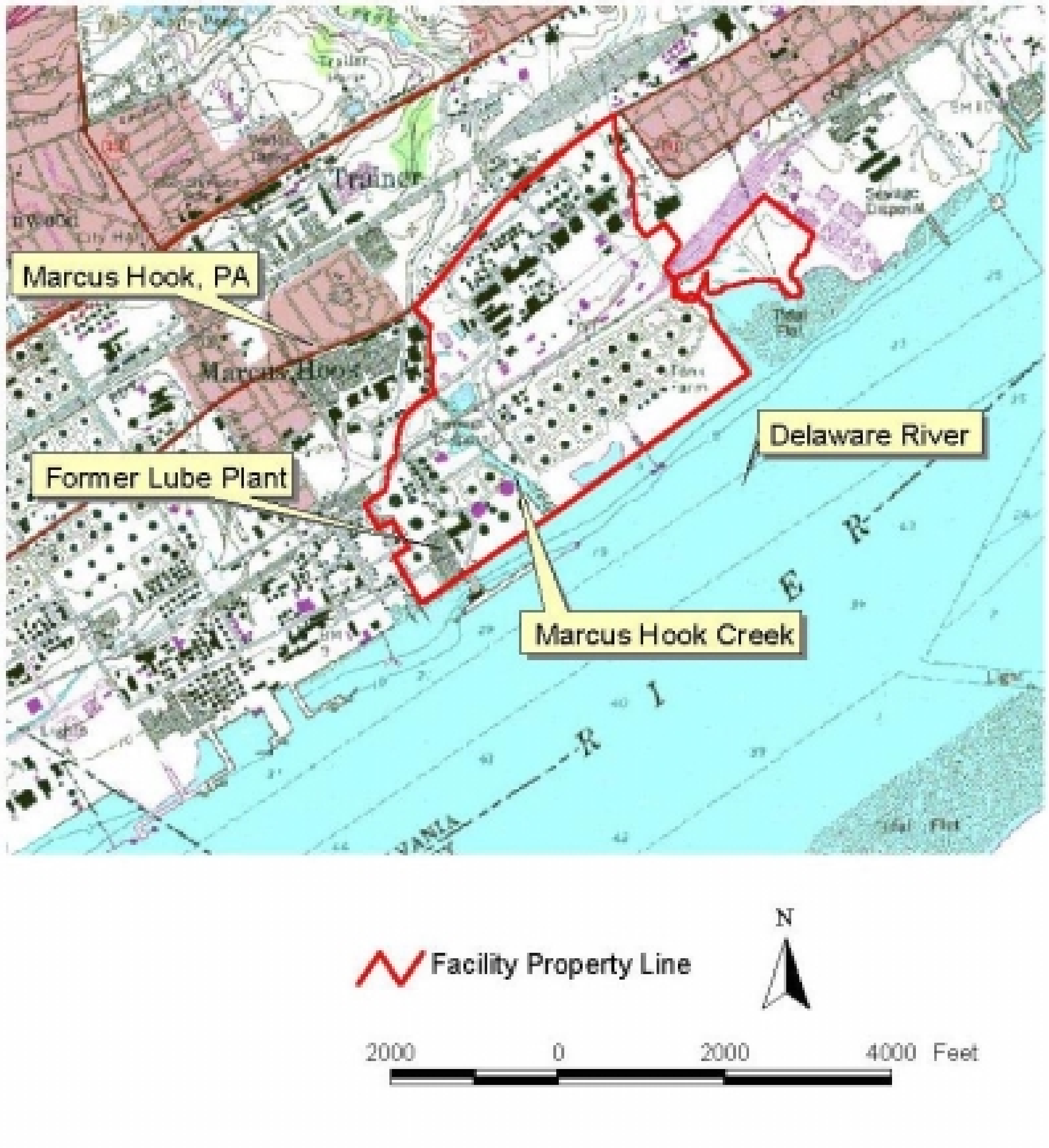


Figure B.1 - Marcus Hook Refinery

An area of approximately 70 acres, known as the Former Lube Plant is the focus of the examples for this research. Figure B.2 is a map of the major structures in the Former Lube Plant. The Former Lube Plant is bounded by the

Delaware River to the southwest, by residential and other industrial properties to the northwest, and by Marcus Hook Creek to the east.

An important feature at the Former Lube Plant is a bulkhead that is located along the Delaware River, shown on the map in Figure B.2 as a purple line. The bulkhead is constructed of steel, except for approximately 500-feet, where a wooden structure exists. The steel portion of the bulkhead provides a hydraulic barrier between the groundwater and the surface water. The wooden structure does not provide a hydraulic barrier. It is located between the two subsurface "mud fences" depicted as green lines running perpendicular to the river. The aboveground storage tanks and the lubricants processing plant buildings are also show in Figure B.2. The shaded areas, designated "Processing Area, Heavy Fuels and West Tank Farm" were developed by Langan (1999) to describe the historical activities in the different sections of the Former Lube Plant.

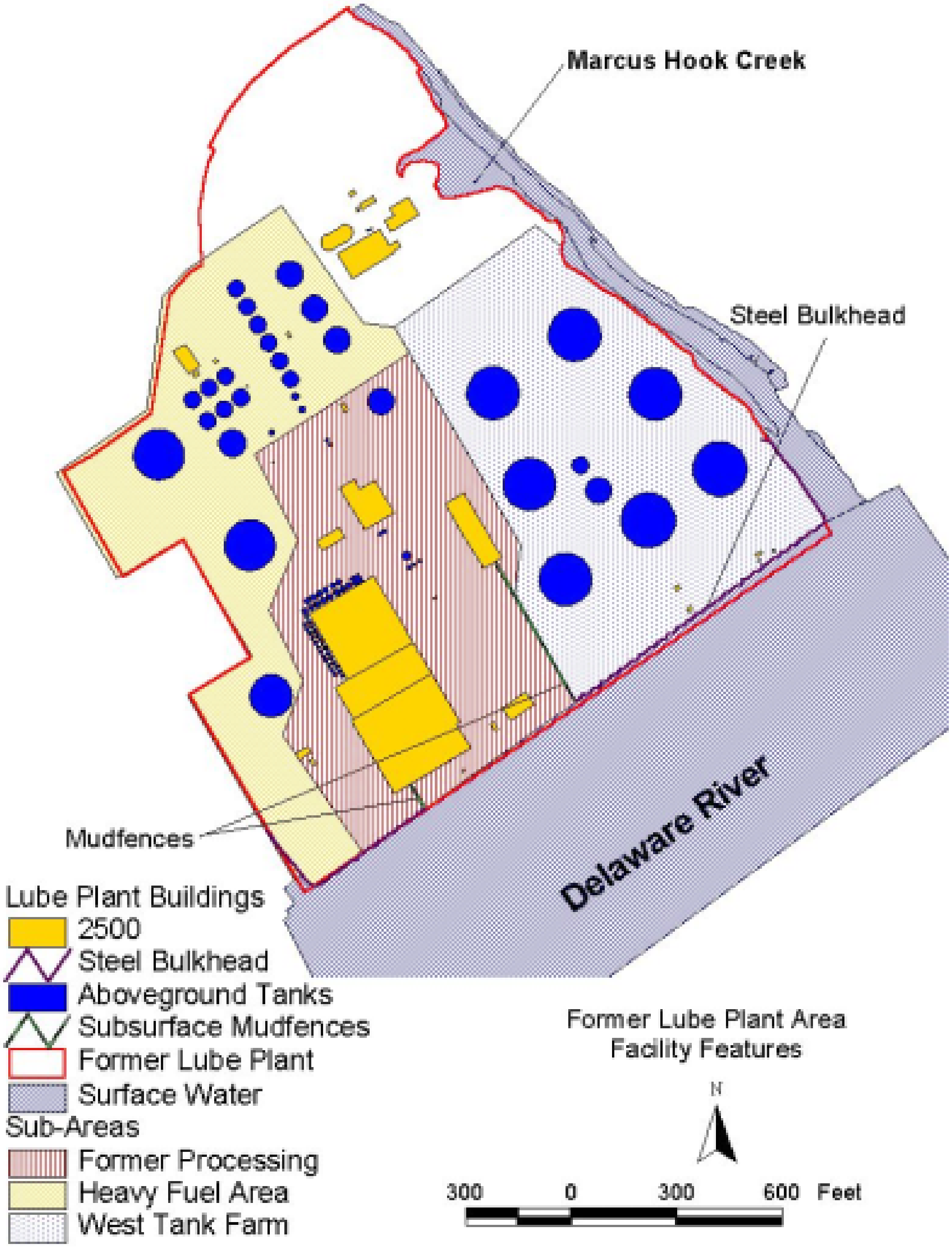


Figure B.2 - Former Lube Plant

The general groundwater flow at the former Lube Plant is from the northwest to the southeast towards the Delaware River and Marcus Hook Creek. Groundwater elevation measurements collected on November 1, 1999 were used to evaluate the groundwater flow directions. Depths to groundwater and NAPL were measured at 47 locations in the Former Lube Plant. These results are similar to previous measurement events (Langan 1999, Kim *et al.*, 2000). The groundwater elevations were corrected for the presence of NAPL using the equations presented in Kim *et al.*, (2000). The resulting corrected elevations were contoured using the **Spatial Analyst** extension in ArcView. The results are shown in Figure B.3.

The inverse distance weighting interpolation method was used for contouring the elevation data. This method gives more weight to the measured values closer to a point than values farther away (Romanek *et al.*, 1999). The mapped data include the calculated elevation values, given in feet above mean sea level, and the contours of equal head elevation. The predicted groundwater flow is perpendicular to the contour lines from areas of higher elevation to areas of lower elevation. The influence of the steel bulkhead and the mud fences in limiting the groundwater flow, and "funneling" groundwater flow to the Delaware River between the mud fences, can be seen in the flow direction results. These data are used in the subsurface parameter estimation presented in section B.3.5.

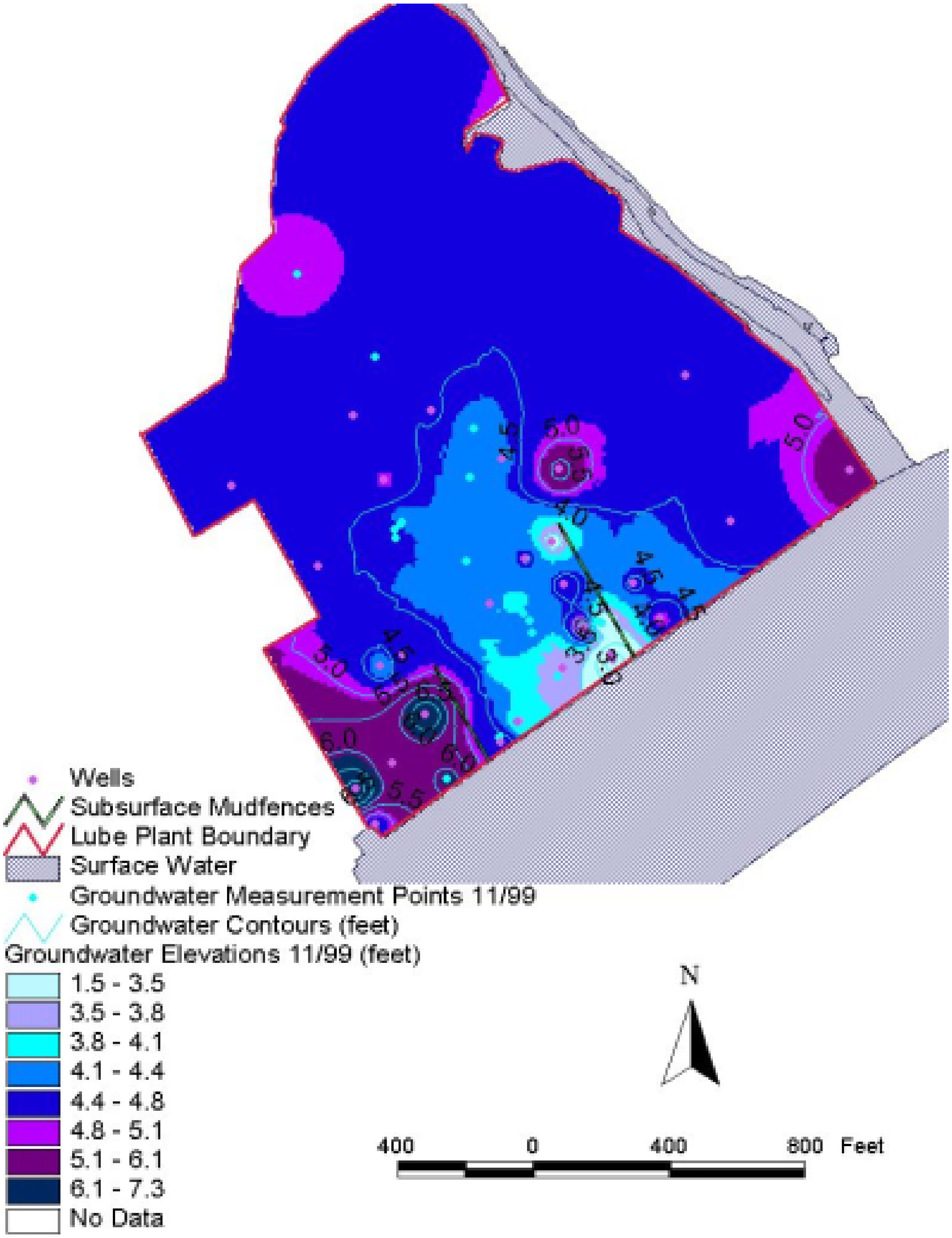


Figure B.3 - Groundwater Elevation Data, November 1, 1999

Site characterization data, foundation boring data and other environmental data collection programs have been conducted at the facility over the past 30 years. These data are included in the digital facility description. Importantly, for the work here, several data collection activities have been conducted associated with, and since, the 1996 property transaction. As a result, the parameter estimation tasks implemented in this research have been based on these sources. The digital facility description by Romanek, *et al*., (1999), the environmental measurements database, the data evaluations prepared by Kim, *et al*., (2000) and the data compilations for the Former Lube Plant prepared by Langan (1999) and (2000) were all used extensively for the parameter estimation. These references are also important sources of additional descriptions of the site conditions. In addition, for parameters that were treated as non-site specific parameters, standard data references were consulted, including ASTM (1995a), ASTM (1999), Charbeneau (2000), Freeze and Cherry (1979), Schwarzenbach *et al*., (1993) and BP, (1997).

B.2 PARAMETER ESTIMATION METHODOLOGY

For the parameters of interest there were varying levels of available information upon which to base the estimates for the example calculations. The parameters were separated into three general classes. The first class consists of those parameters that are treated as site-specific random variables. The second class consists of those parameters that are treated as non-site-specific random variables. The third class consists of those parameters that are treated as constants. These classes are discussed in detail in this section.

An approach for estimating the necessary parameter values was used that incorporates several different methods to develop the mean, variance and deterministic or "conservative" estimate for each input value. As used here, the uncertainty described by the variance value is the total uncertainty in the parameter value. The uncertainty may be due to spatial variability, testing variability, the lack of site-specific data and other sources of variability.

For the parameters that are treated as site-specific variables and site measurements have been made to estimate the parameter value (e.g., hydraulic conductivity by slug tests and grain size analyses), the sample mean and variance, along with judgments about the variability of the parameter, were used to derive the estimated mean and variance to represent the area to be modeled. For the parameters that are treated as site-specific variables and no direct measurements were made, but characteristics of the facility indicate the approximate values for the parameters (e.g., air and water filled porosity in the vadose zone based on the characteristics of the soil type) reference values were used along with judgments about the data distribution and the variability to develop the needed parameter values. For the parameters that were treated as non-site specific parameters (e.g., areal fraction of cracks in a foundation, Henry's law coefficient) several references, including those listed in section B.1, were used along with judgments about the data distribution and the variability to develop the needed parameter values.

As a practical matter, the parameters that are treated as constants are such that they are not expected to exhibit much variability based on conditions at the facility (e.g., chemical aqueous solubility, soil particle density, building air exchange rate) and so their contribution to the overall variance is accounted for in the variance of the other variables, including the model error variables. For these parameters, reference values were identified and used in the examples. This distinction between variables and constants is only applicable to the probabilistic calculations, since in the deterministic case all of the parameters are constants.

It must be emphasized that in reality all of the input values are random variables that are site-specific, and in-fact may have specific data distributions within sub-areas at a facility. In order to develop an approximation of the variability in the predicted exposure pathway ending concentrations within a practical, analytical calculation method, the variables were divided into the three classes described here and the needed parameter values estimated.

In general, if no specific data distribution was indicated by the reference sources or other information, the procedure outlined by Muchard (1997) was used. The variables were estimated to follow a normal distribution. For the estimate, reasonable minimum (x_{min}) and maximum (x_{max}) values were determined for the parameter and the mean (μ_x) value was calculated. The mean value is calculated as the mid-point, or median value, between the minimum and maximum values. Equation B.1 shows the relationship used to calculate the mean values.

$$\mu_x = \frac{x_{min} + x_{max}}{2} \qquad \textbf{Equation B.1}$$

Based on the amount of data and information available, including the site sampling data, a level of confidence was assigned for the minimum and maximum values for each parameter (i.e., the probability that the actual value lies between the minimum value and the maximum value). This level of confidence is used to define the confidence interval in order to calculate the standard deviation (σ_x) and the variance (σ_x^2). The confidence intervals were set at levels that were believed to be applicable to each variable, but that were expected to be low values to include the variability for which account is not otherwise taken. Equation B.2 is based on the equation for the standard normal distribution and gives the relationship for calculating the standard deviation.

$$\sigma_x = \frac{x_{\min} - \mu_x}{\Phi^{-1}\left(\frac{1-\frac{b}{100}}{2}\right)}$$

Equation B.2

where

b = confidence level expressed as a percent

Φ^{-1} = Inverse of the standard normal distribution cumulative density function (Muchard, 1997).

The coefficient of variation (c.o.v.) is also calculated for each parameter. Equation B.3 gives the relationship for the coefficient of variation.

$$\text{c.o.v.} = \frac{\sigma_x}{\mu_x}$$

Equation B.3

The coefficient of variation is an estimate of the variability with respect to the mean value. It provides a perspective about the magnitude of the variance given the magnitude of the mean value. As an example, if the c.o.v. is large, then the variability is large, regardless of the value of the variance. (Ang and Tang, 1975). An important qualifier is if the mean value is very small, then the c.o.v. is likely to be large.

B.3 DISCUSSION OF PARAMETER VALUES

This section describes the example source areas chosen to demonstrate the SERA methodology, the exposure pathways to be modeled and the input parameter values needed to execute the exposure pathway calculations. This section provides additional detail to the discussion presented in Chapter 5.

B.3.1 Source Areas

Seven source areas were identified for the Former Lube Plant. These are not intended to be all of the areas that might or should be evaluated for the Former Lube Plant, but are intended to be a representative group that can be used to demonstrate the SERA methodology. In addition, one or more chemicals of concern were chosen for each source area. More chemicals may be representative of the source areas, however, a smaller number of chemicals of concern were chosen as examples for this demonstration. The source areas that were chosen are all areas where non-aqueous phase liquids (NAPL) have been detected on the groundwater table as of February 2000. Figure B.4 shows the locations of these source areas.

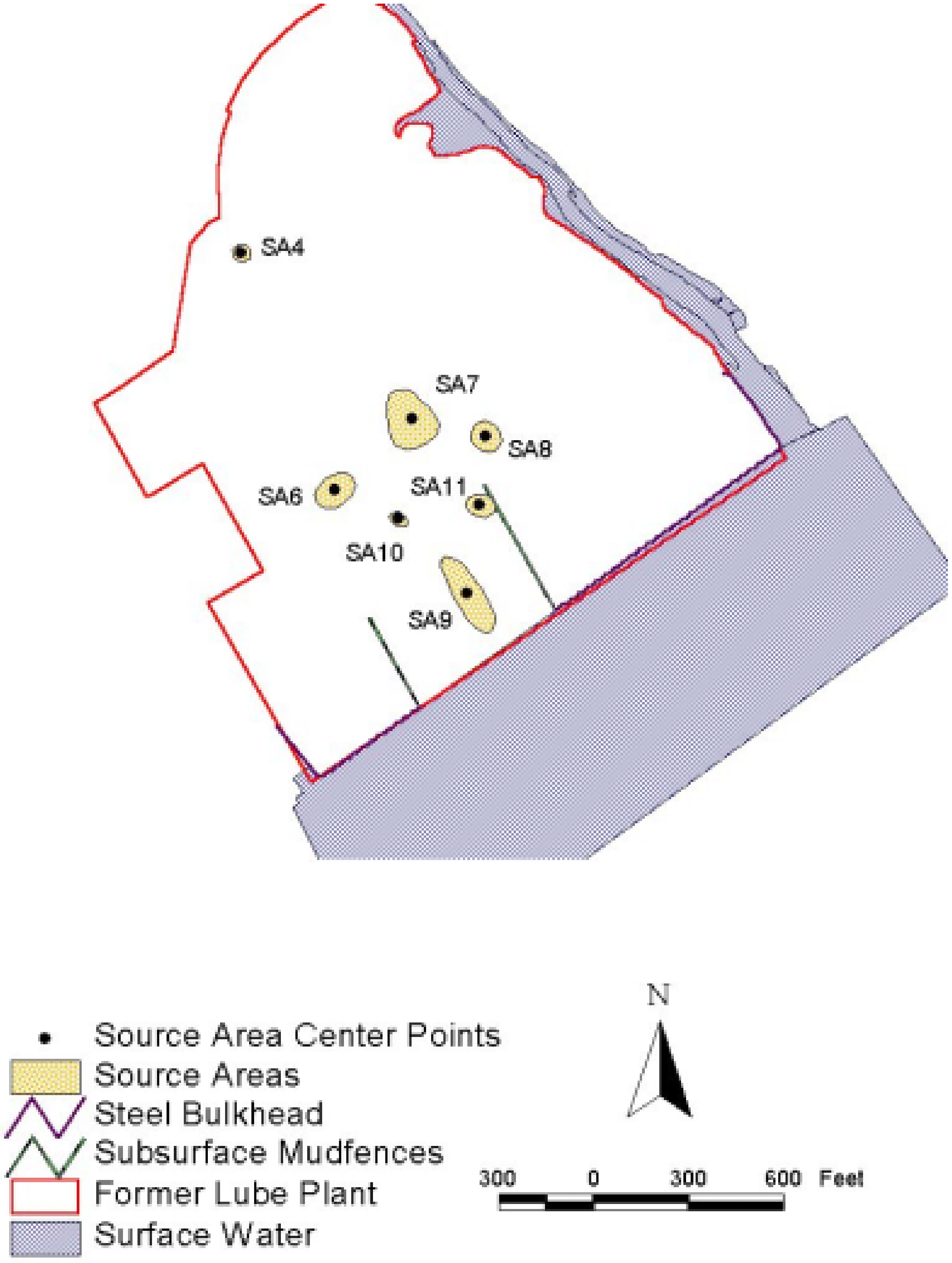


Figure B.4 - Example NAPL Source Areas in Former Lube Plant

B.3.1.1 Environmental Measurement Database

Data from the environmental measurements database (Langan, 2000) were used to identify the source areas. The perimeter of each source area was identified based on the location of sampling points where any NAPL was detected surrounding an area where the greatest thickness was measured. Separating one source area from an adjacent source area was accomplished using the original source areas described by Langan (1999) and updating them with the specific gravity measurements (i.e., fluid density relative to that of water given a value of 1.0) that have been made for NAPL samples from the monitoring wells. Maps of the different parameters included in this section were developed using the environmental measurements database and the ArcView GIS software. Queries were developed in Access to select the desired parameter values and then they were mapped in the GIS using an ODBC driver for the database and the SQL Connect commands in ArcView. Procedures for the ODBC driver set-up, and the SQL Connect are included in Romanek *et al*., (1999).

B.3.1.2 NAPL Thickness

The data for NAPL thickness measurements were selected for the most-recent sampling event, February 16, 2000. Figure B.5 shows the example source areas and the NAPL thickness measurements (in feet). The open circles are groundwater measurement locations (e.g., monitoring wells) where no NAPL was detected. The blue squares are measurements up to 0.75 feet. The green pentagons are from 0.75 to 1.5 feet. The pink triangles are 1.5 to 3.0 feet. The red octagon (one measurement) is above 3.0 feet, at 6.54 feet. This maximum thickness is represented by Source Area SA10, see section B.3.1.5.6 for further discussion of SA10.

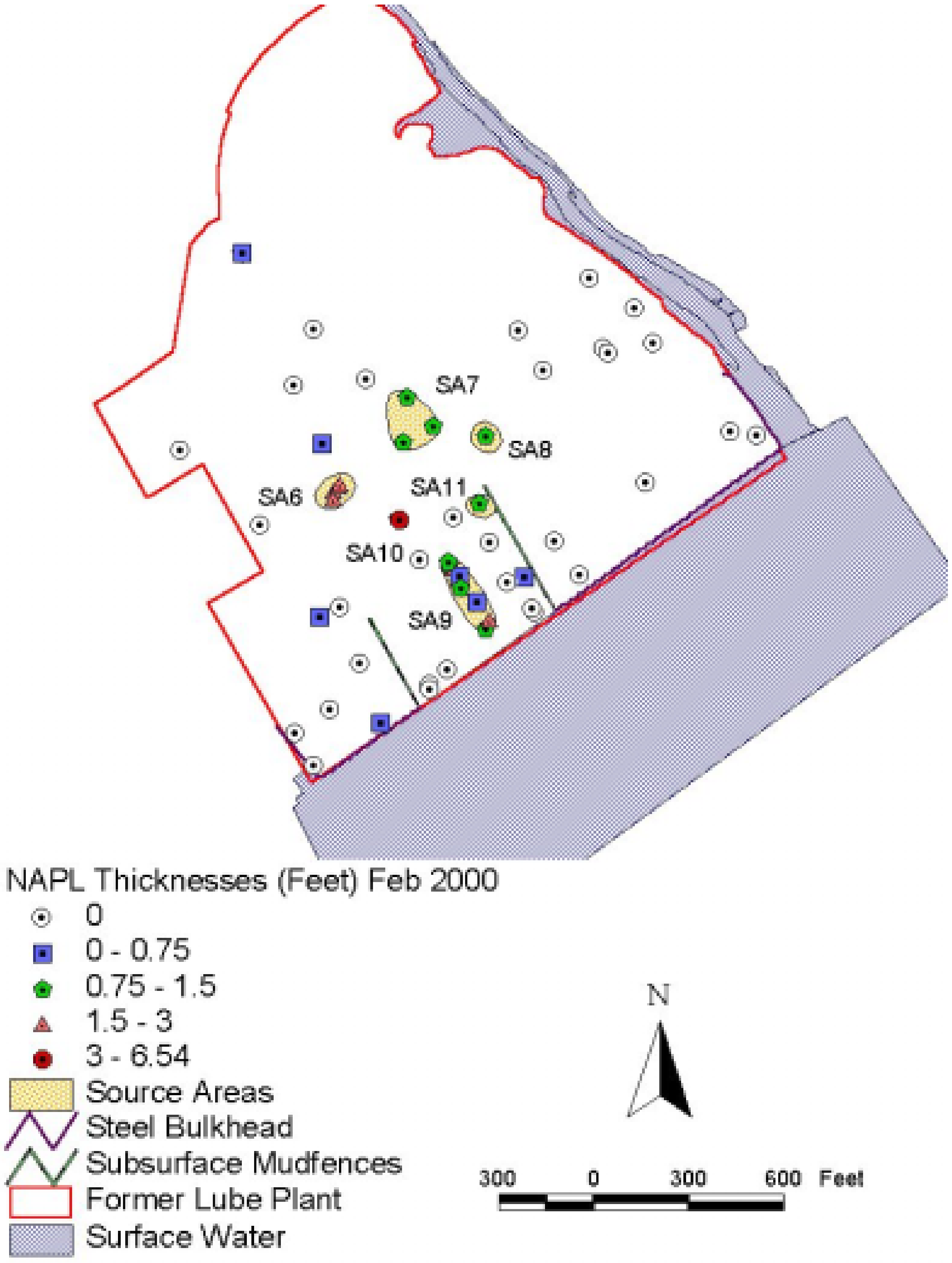


Figure B.5 - NAPL Detection on Groundwater from the February 2000 Monitoring Event

B.3.1.3 Specific Gravity

The data for specific gravity measurements from March 1999 and November 1999, were also mapped. Figure B.6 shows the example source areas and the specific gravity measurements (dimensionless). The blue squares are measurements from 0.81 to 0.852. The green pentagons are from 0.852 to 0.876. The pink triangles are 0.876 to 0.897 feet. The red circles are 0.897 to 0.919. The absolute value of the specific gravity measurement is not as important as the relative differences for measurements near one another. These measurements were used to distinguish the likely extent of one source area as separate from another source area. This is illustrated by SA7 and SA8 where the SA7 samples represent the lighter end of the range of values while the SA8 sample is at the heavier end of the range. Lower molecular weight hydrocarbons will have a lower specific gravity. In addition, for the same petroleum hydrocarbon mixture, more recently released petroleum hydrocarbons will have a lower specific gravity than petroleum hydrocarbons that have been exposed to weathering processes in the subsurface. These relative characteristics and comparisons were used to identify the separate source areas.

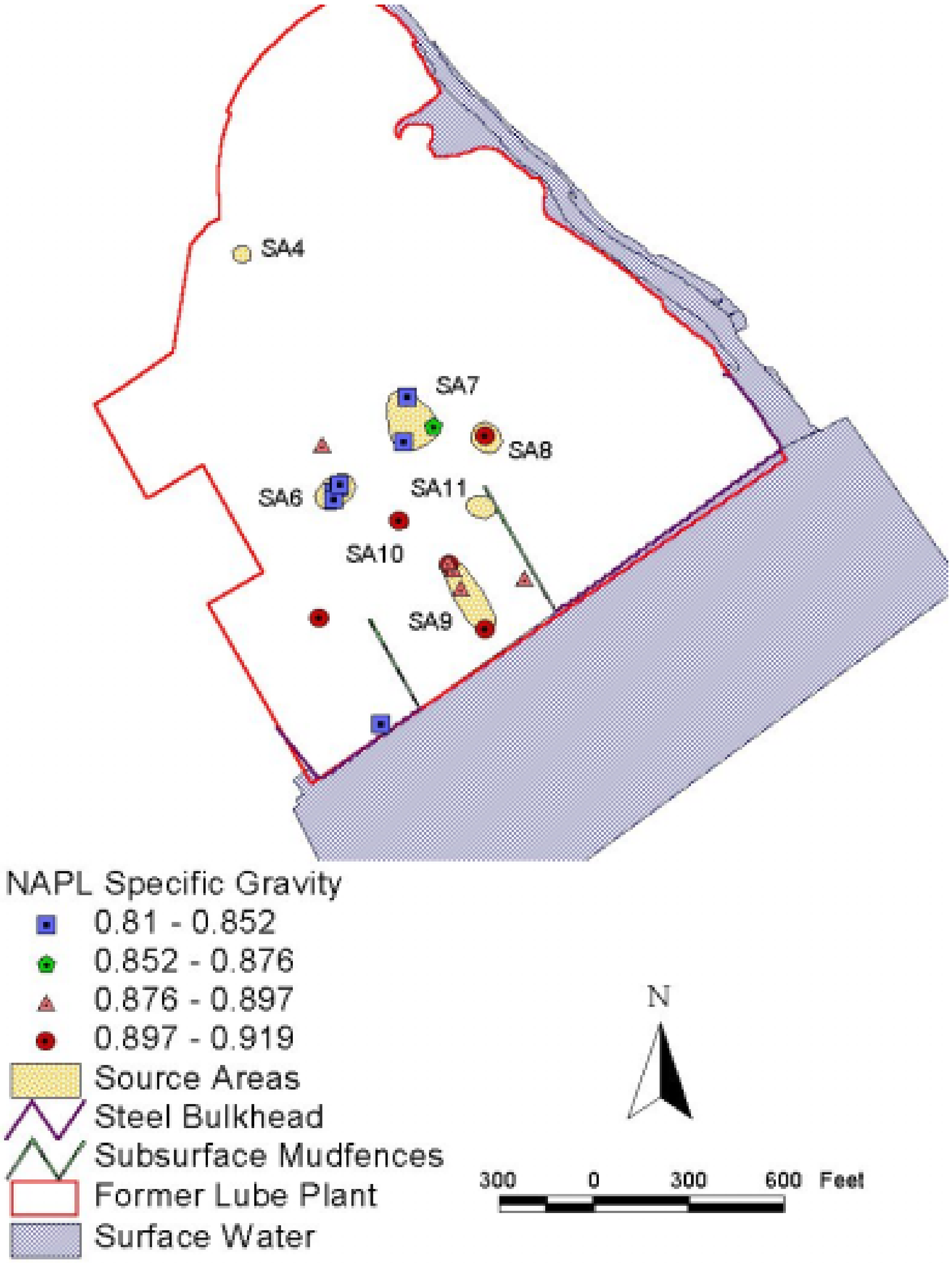


Figure B.6 - NAPL Specific Gravity Measurements from 1999

B.3.1.4 NAPL Mole Fraction

Specific compositional analyses of all of the NAPL encountered in the monitoring wells were not available. These data would have been helpful, not only in distinguishing between source areas, but also in determining specific chemicals of concern for each source area and their respective NAPL mole fractions.

For the calculations, the NAPL mole fraction (**NMF**) is used to estimate the concentrations in groundwater and in soil vapor as presented in sections A.1 and A.2. The mole fraction for any compound *i* in a mixture of N compounds is defined as:

$$NMF_i = \frac{\frac{c_i}{MW_i}}{\sum_{j=1}^{N} \frac{c_j}{MW_j}}$$

Equation B.4

where:
c_i - concentration of compound *i* in the mixture (g/L)
MWi - molecular weight of compound *i* (g/mol).
(Charbeneau, 2000).

Because specific compositional data were not available for the source areas in the examples, an approximate method was needed to describe the NMF's for each chemical of concern for each source area. The estimated NMF's are based on literature values compiled in a petroleum hydrocarbons review document prepared by the American Petroleum Institute (API, 1994). The values reported in the API document are individual chemical weight percent values for different classes of products. These values are used here for the NMF values because the identity of the NAPL mixtures at the Former Lube Plant are not known and, therefore, the molecular weight of the NAPL mixtures at the facility are not known. As an approximation, the concentrations (or weight percent values) are taken as suitable surrogate measures of the mole fractions. In addition, since the NAPL at the Former Lube Plant are more weathered and likely more heterogeneous than the fresh products reported in the API document, the NMF values are more likely to be over-estimates of the true mole fractions, than under-estimates for the individual chemicals of concern at the Former Lube Plant.

B.3.1.5 Source Area and Exposure Pathways Descriptions

The locations at which NAPL were detected, and those where NAPL were not detected along with the specific gravity measurements and the historical information provided by Langan (1999), were used to identify the source areas and to select representative chemicals of concern. As additional characterization data are collected, the source areas and their compositional characteristics may be refined.

For the source areas where it is suspected that gasoline is a component of the NAPL, benzene, toluene ethylbenzene and total xylenes are used as the chemicals of concern. Benzene is typically the primary chemical of concern for gasoline releases because of its toxicity and mobility in the environment. The three other aromatic hydrocarbons are also typically analyzed for a gasoline release. For areas where it is suspected that heavier fuels or petroleum mixtures (e.g., diesel, heating oil) are components of the NAPL, naphthalene is used as the chemical of concern because it is commonly an indicator of heavier fuels (e.g., diesel, heating oil) that is also relatively mobile in the environment.

Descriptions of each source area are included in the following sections.

B.3.1.5.1 Source Area SA4

SA4 is defined as the area around monitoring well MW-34. Langan (1999) describes this area as the "heavy fuels area" based on historical operations. They report impacted soils and NAPL detections in this area in prior investigations. There has been limited investigation near SA4, so its extent is uncertain, but it is assumed to be a relatively small area near MW-34. The chemical of concern identified for this source area is naphthalene. The potential exposure pathway analyzed for the chemicals of concern for SA4 is transport in groundwater to surface water (i.e., Marcus Hook Creek). Groundwater transport to the Delaware River was not chosen as a complete exposure pathway for the calculations because of the distance (more than 1400 feet) from SA4 to the river, so attenuation would make this exposure pathway insignificant. This is demonstrated by using the mean NMF value presented in Table B.1 for naphthalene, which results in a CA concentration of 0.4 mg/L. Using the mean values for the other input parameters given in this Appendix for the groundwater transport equation, the predicted groundwater concentration at the river is 3×10^{-46} mg/L.

The NAPL mole fraction for naphthalene is estimated from literature values of composition of middle distillate fuels (API, 1994). The NAPL mole fraction values are given in Table B.1.

B.3.1.5.2 Source Area SA6

SA6 is defined by monitoring wells MW-33, MP-5 and MP-7. It is in an area of the site characterized by Langan (1999) and Kim, *et al.*, (2000) as a former processing area along with the location of current and out of service underground piping. Given what is known about this area, the NAPL detected here could be a composite of a number of the products handled by the Refinery. Since the specific gravity of the NAPL is relatively low, for the samples collected from the wells, in the range 0.81-0.852, the NAPL is likely to contain at least some proportion of gasoline. In 1996 the benzene fraction of the NAPL was measured at 0.003 (reported as 0.3-percent by weight) in a sample taken from MW-33. Therefore, benzene, toluene, ethylbenzene and xylenes are considered to be the chemicals of concern for this source area. The exposure pathways identified for this source area include ground water transport to surface water (i.e., the Delaware River) and volatilization to indoor air. Groundwater transport to Marcus Hook Creek was not chosen as a complete exposure pathway for the calculations because the funneling effect of the wooden bulkhead imparts a groundwater gradient towards the river from SA6 and creates a gradient away from the creek between SA6 and the creek. This can be seen in Figure B.7. The volatilization to indoor air is considered to be a potentially complete exposure pathway based on a future industrial building in the vicinity of the now-vacant former lubricants warehouse buildings.

The estimated NAPL mole fraction value for benzene, toluene, ethylbenzene and xylenes for this source area are based on the reported concentration values and the literature values for gasoline (API, 1994). The NAPL mole fraction parameters are given in Table B.1.

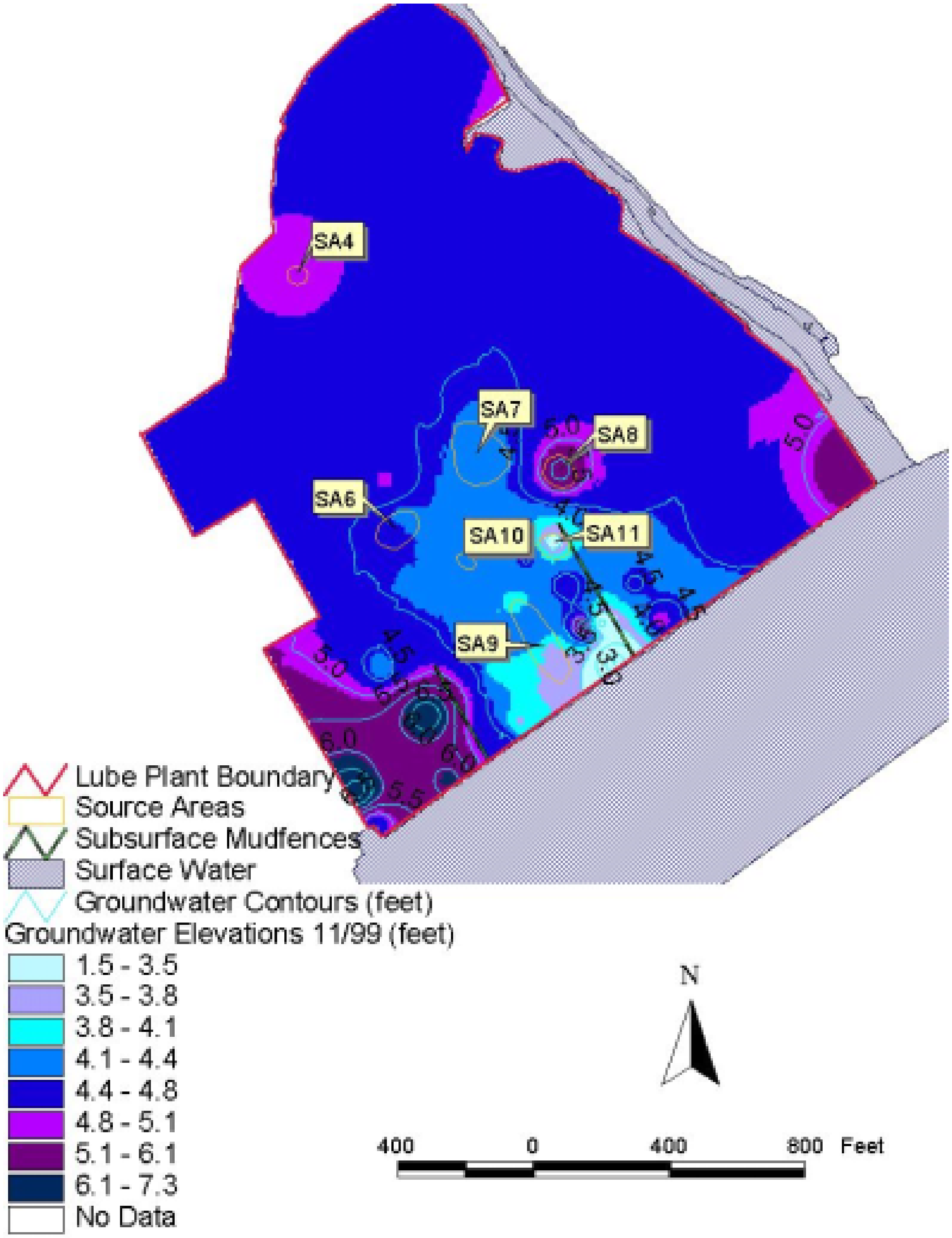


Figure B.7- Groundwater Flow Directions and Locations of Source Areas

B.3.1.5.3 Source Area SA7

SA7 is defined by monitoring wells MW-8, MW-30 and MW-55. It is in an area of the site characterized by Langan (1999) and Kim, *et al*., (2000) as a former processing area along with the location of current and out of service underground piping. Given what is known about this area, the NAPL detected here could be a composite of a number of products handled by the Refinery. The relatively low specific gravity (with an average of three measurements of 0.85) would indicate that, at least in part, the NAPL is composed of gasoline. In 1995 and 1996 the NAPL from MW-30 was analyzed for various parameters, including benzene content, MTBE content and specific gravity. MTBE was reported at the value given as the detection limit, 0.01-percent by volume. Benzene was reported at 0.01-percent by weight in 1995 and 0.29-percent by weight in 1996. The specific gravity was reported at 0.809 in 1995 and 0.831 in 1996.

Based on the available information, the chemicals of concern identified for this source area are benzene, ethylbenzene, toluene and total xylenes, which are characteristic of gasoline. The exposure pathways identified for this source area include ground water transport to surface water (i.e., the Delaware River) and volatilization to outdoor air. Groundwater transport to Marcus Hook Creek was not chosen as a complete exposure pathway for the calculations because it appears from the November 1999 groundwater elevation measurements that there is a shallow groundwater gradient away from the creek, between SA7 and the creek.

This can be seen in Figure B.7. The outdoor air exposure pathway is relevant for an industrial worker in the Former Lube Plant.

The estimated NAPL mole fraction value for benzene for this source area is based on the reported concentration values and the literature values for gasoline. The values for ethylbenzene, toluene and xylenes are based on literature values for gasoline (API, 1994). The NAPL mole fraction parameters are given in Table B.1.

B.3.1.5.4 Source Area SA8

SA8 is defined by monitoring well MW-31. It is in an area of the site characterized by Langan (1999) and Kim, *et al*., (2000) as the West Tank Farm and adjacent to the location of current and out of service underground piping. Given what is known about this area, the NAPL detected here could be a composite of a number of products handled by the Refinery. The specific gravity measurement from MW-31 was 0.91; at the heavier end of the measured range. No compositional data were available for MW-31. The tanks in the West Tank Farm include crude oil storage tanks. Based on the location, near underground piping and the crude storage tanks and based on the higher specific gravity, the chemicals of concern chosen for this source area are benzene, toluene, ethylbenzene, xylenes and naphthalene. The exposure pathways identified for this source area include ground water transport to surface water (i.e., Marcus Hook Creek) and volatilization to outdoor air. Groundwater transport to the Delaware River was not chosen as a complete exposure pathway for the calculations because of the location of SA8 relative to the steel bulkhead and the eastern mud fence. Groundwater flow from the area of SA8 to the river is precluded by these two features (See Figure B.7). The outdoor air exposure pathway is relevant for an industrial worker in the Former Lube Plant.

The estimated NAPL mole fraction values for benzene, toluene, ethylbenzene, xylenes and naphthalene, are based on literature values (API, 1994). The NAPL mole fraction parameters are given in Table B.1.

B.3.1.5.5 Source Area SA9

SA9 is defined by monitoring wells MW-9, MW-45, MP-1, MP-2 and MP-4. It is in an area of the site characterized by Langan (1999) and Kim, *et al.*, (2000) as a former processing area along with the location of current and out of service underground piping. Given what is known about this area, the NAPL detected here could be a composite of a number of products handled by the Refinery. The relatively high specific gravity (average of five measurements of 0.90) would indicate that, at least in part, the NAPL is composed of heavier products, or weathered light products. In 1995 and 1996 the NAPL from MW-9 was analyzed for various parameters, including benzene content, MTBE content and specific gravity. MTBE was reported at the value given as the detection limit, 0.01-percent by volume. Benzene was reported at 0.4-percent by weight in 1996. The specific gravity was reported at 0.92 in 1991, 0.884 in 1995 and 1.003 in 1996.

Based on the location, near underground piping, and based on the higher specific gravity, the chemicals of concern chosen for this source area are benzene, toluene, ethylbenzene xylenes and naphthalene. The exposure pathways identified for this source area include ground water transport to surface water (i.e., the Delaware River) and volatilization to indoor air. Groundwater transport to Marcus Hook Creek was not chosen as a complete exposure pathway for the calculations because of the location of SA9 relative to the eastern mud fence. Groundwater flow from SA9 to the creek is precluded by the mud fence (See Figure B.7). The volatilization to indoor air is considered to be a potentially complete exposure pathway based on a future industrial building in the vicinity of the now-vacant former Lubricants Warehouse buildings.

The estimated NAPL mole fraction value for benzene for this source area is based on the reported concentration values and literature values. The values for toluene, ethylbenzene, xylenes and naphthalene are based on literature values (API, 1994). The NAPL mole fraction parameters are given in Table B.1.

B.3.1.5.6 Source Area SA10

SA10 is defined by monitoring well MW-54. It is in an area of the site characterized by Langan (1999) and Kim, *et al*., (2000) as a former processing area along with the location of current and out of service underground piping. Given what is known about this area, the NAPL detected here could be a composite of a number of products handled by the Refinery. The specific gravity of the NAPL is relatively high for the samples collected from the wells, 0.905, and would indicate that, at least in part, the NAPL is composed of heavier products, or weathered light products. No compositional analyses are available for the NAPL in this area. Therefore, based on the specific gravity data and the location near the underground piping, benzene, toluene, ethylbenzene, xylenes and naphthalene are considered to be the chemicals of concern for this source area. The exposure pathways identified for this source area include ground water transport to surface water (i.e., the Delaware River) and volatilization to indoor air. Groundwater transport to Marcus Hook Creek was not chosen as a complete exposure pathway for the calculations because of the location of SA10 relative to the eastern mud fence. Groundwater flow from SA10 to the creek is precluded by the mud fence (See Figure B.7). The volatilization to indoor air is considered to be a potentially complete exposure pathway based on a future industrial building in the vicinity of the now-vacant former Lubricants Warehouse buildings.

The estimated NAPL mole fraction values for benzene, toluene, ethylbenzene, xylenes and naphthalene are based on literature values (API, 1994). The NAPL mole fraction parameters are given in Table B.1.

B.3.1.5.7 Source Area SA11

SA11 is defined by monitoring well MW-53. It is in an area of the site characterized by Langan (1999) and Kim, *et al*., (2000) as a former processing area along with the location of current and out of service underground piping. Given what is known about this area, the NAPL detected here could be a composite of a number of the products handled by the Refinery. Since there are no compositional or specific gravity measurements for the NAPL in this well and it is possible that both gasoline and heavier products are included in this NAPL, the chemicals of concern are estimated to be benzene, toluene, ethylbenzene, xylenes and naphthalene. It is considered to be a separate source area from SA10 and SA9 because there are a number of wells located between these three source areas that do not exhibit NAPL thicknesses (i.e., MW-50, MW-51, MW-52). The exposure pathways identified for this source area include ground water transport to surface water (i.e., the Delaware River) and volatilization to outdoor air. Groundwater transport to Marcus Hook Creek was not chosen as a complete exposure pathway for the calculations because of the location of SA11 relative to the eastern mud fence. Groundwater flow from SA11 to the creek is precluded by the mud fence (See Figure B.3). The outdoor air exposure pathway is relevant for an industrial worker in the Former Lube Plant.

The estimated NAPL mole fraction values for benzene, toluene, ethylbenzene, xylenes and naphthalene are based on literature values (API, 1994). The NAPL mole fraction parameters are given in Table B.1.

Table B.1 - Source Area NAPL Mole Fraction (NMF) Values

Variable ID	Min	Max	Confidence Interval %	Mean	SD	c.o.v.	Variance	D(X)
NMFSA491-20-3	0.000076	0.025	80	1.25E-02	9.72E-03	7.76E-01	9.46E-05	1.25E-02
NMFSA6108-88-3	0.045	0.21	80	1.28E-01	6.44E-02	5.05E-01	4.14E-03	1.28E-01
NMFSA61330-20-7	0.037	0.145	80	9.10E-02	4.21E-02	4.63E-01	1.78E-03	9.10E-02
NMFSA6100-41-4	0.007	0.028	80	1.75E-02	8.19E-03	4.68E-01	6.71E-05	1.75E-02
NMFSA671-43-2	0.007	0.038	80	2.25E-02	1.21E-02	5.38E-01	1.46E-04	2.25E-02
NMFSA7108-88-3	0.045	0.21	80	1.28E-01	6.44E-02	5.05E-01	4.14E-03	1.28E-01
NMFSA71330-20-7	0.037	0.145	80	9.10E-02	4.21E-02	4.63E-01	1.78E-03	9.10E-02
NMFSA7100-41-4	0.007	0.028	80	1.75E-02	8.19E-03	4.68E-01	6.71E-05	1.75E-02
NMFSA771-43-2	0.007	0.038	80	2.25E-02	1.21E-02	5.38E-01	1.46E-04	2.25E-02
NMFSA8108-88-3	0.045	0.21	80	1.28E-01	6.44E-02	5.05E-01	4.14E-03	1.28E-01
NMFSA81330-20-7	0.037	0.145	80	9.10E-02	4.21E-02	4.63E-01	1.78E-03	9.10E-02
NMFSA8100-41-4	0.007	0.028	80	1.75E-02	8.19E-03	4.68E-01	6.71E-05	1.75E-02
NMFSA871-43-2	0.007	0.038	80	2.25E-02	1.21E-02	5.38E-01	1.46E-04	2.25E-02
NMFSA891-20-3	0.000076	0.025	80	1.25E-02	9.72E-03	7.76E-01	9.46E-05	1.25E-02

Notes: Variable ID is variable name, source area, and COC_ID concatenated
All data distributions are assumed to be normal
Naphthalene data from API (1994) Tables A-13, A-14
Benzene, Toluene, Ethylbenzene, Xylenes data from API (1994) Table A-7
All values are dimensionless

Table B.1 - Source Area NAPL Mole Fraction (NMF) Values (continued)

Variable ID	Min	Max	Confidence Interval %	Mean	SD	c.o.v.	Variance	D(X)
NMFSA9108-88-3	0.045	0.21	80	1.28E-01	6.44E-02	5.05E-01	4.14E-03	1.28E-01
NMFSA91330-20-7	0.037	0.145	80	9.10E-02	4.21E-02	4.63E-01	1.78E-03	9.10E-02
NMFSA9100-41-4	0.007	0.028	80	1.75E-02	8.19E-03	4.68E-01	6.71E-05	1.75E-02
NMFSA971-43-2	0.007	0.038	80	2.25E-02	1.21E-02	5.38E-01	1.46E-04	2.25E-02
NMFSA991-20-3	0.000076	0.025	80	1.25E-02	9.72E-03	7.76E-01	9.46E-05	1.25E-02
NMFSA10108-88-3	0.045	0.21	80	1.28E-01	6.44E-02	5.05E-01	4.14E-03	1.28E-01
NMFSA101330-20-7	0.037	0.145	80	9.10E-02	4.21E-02	4.63E-01	1.78E-03	9.10E-02
NMFSA10100-41-4	0.007	0.028	80	1.75E-02	8.19E-03	4.68E-01	6.71E-05	1.75E-02
NMFSA1071-43-2	0.007	0.038	80	2.25E-02	1.21E-02	5.38E-01	1.46E-04	2.25E-02
NMFSA1091-20-3	0.000076	0.025	80	1.25E-02	9.72E-03	7.76E-01	9.46E-05	1.25E-02
NMFSA11108-88-3	0.045	0.21	80	1.28E-01	6.44E-02	5.05E-01	4.14E-03	1.28E-01
NMFSA111330-20-7	0.037	0.145	80	9.10E-02	4.21E-02	4.63E-01	1.78E-03	9.10E-02
NMFSA11100-41-4	0.007	0.028	80	1.75E-02	8.19E-03	4.68E-01	6.71E-05	1.75E-02
NMFSA1171-43-2	0.007	0.038	80	2.25E-02	1.21E-02	5.38E-01	1.46E-04	2.25E-02
NMFSA1191-20-3	0.000076	0.025	80	1.25E-02	9.72E-03	7.76E-01	9.46E-05	1.25E-02

Notes: Variable ID is variable name, source area, and COC_ID concatenated
All data distributions are assumed to be normal
Naphthalene data from API (1994) Tables A-13, A-14
Benzene, Toluene, Ethylbenzene, Xylenes data from API (1994) Table A-7
All values are dimensionless

B.3.2 Source Area Dimension Variables

This section presents the variables used to describe the source area dimension variables for the exposure pathways chosen and the algorithms used. For each source area specific dimensions are needed depending on the exposure pathways to be evaluated. The dimensions are generally estimated based on the data available for the site and are described in the following sections.

B.3.2.1 Ground Water Transport to Surface Water Exposure Pathway

For the groundwater transport to surface water exposure pathway, the width of the source area perpendicular to the groundwater flow direction (**Width**), the thickness or depth of the concentrations of chemicals of concern in the source area below the water table (**Depth**) and the distance from the source area to the transition point at the surface water are needed (**sDist**). The estimates for the Width and sDist values are developed in ArcView. The Width is estimated using the Measuring Tool. The sDist is developed using the **gwtrans.ave** script. The minimum groundwater travel distance (sDist-Min) is estimated as the distance from the downgradient edge of the source area to the surface water. The maximum distance (sDist-Max) is estimated from the upgradient edge of the source area to the surface water. Figure B.8 shows the source areas, along with the groundwater transport segments for the groundwater to surface water exposure pathways. Table B.2 includes the values for the groundwater transport segment travel distances.

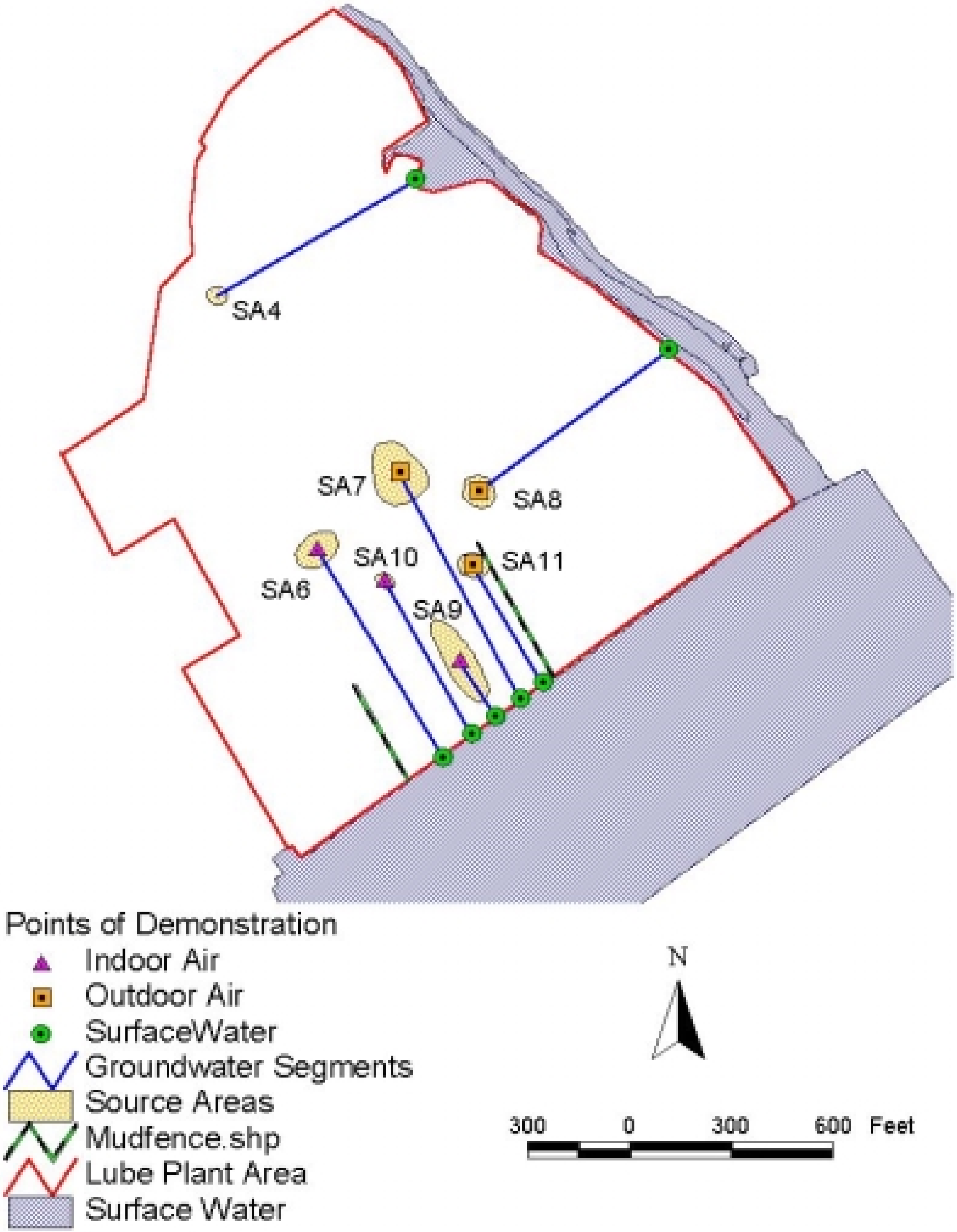


Figure B.8 - Groundwater Transport Segments and Points of Demonstration

Table B.2 - Groundwater Transport Segment Distances

Variable	Source Area	Min	Max	Confidence Interval %	Mean	SD	c.o.v.	Variance	D(X)
sDist	SA4	2.00E+04	2.17E+04	80	2.08E+04	6.35E+02	3.05E-02	4.03E+05	2.08E+04
sDist	SA8	1.99E+04	2.26E+04	80	2.12E+04	1.06E+03	5.01E-02	1.13E+06	2.12E+04
sDist	SA6	2.05E+04	2.33E+04	80	2.19E+04	1.10E+03	5.02E-02	1.21E+06	2.19E+04
sDist	SA10	1.54E+04	1.69E+04	80	1.62E+04	6.17E+02	3.82E-02	3.81E+05	1.62E+04
sDist	SA9	1.78E+03	1.01E+04	90	5.94E+03	2.52E+03	4.25E-01	6.37E+06	5.94E+03
sDist	SA7	2.00E+04	2.57E+04	80	2.29E+04	2.25E+03	9.86E-02	5.08E+06	2.29E+04
sDist	SA11	1.11E+04	1.35E+04	80	1.23E+04	9.39E+02	7.64E-02	8.82E+05	1.23E+04

Notes: All data distributions are assumed to be normal
All values in cm, except variance in cm^2 and c.o.v. is dimensionless

Figure B.9 shows the source areas along with buffer areas, estimated to be 10 percent larger than each source area. With no additional data to describe the source areas, the 10 percent buffer value was chosen as an approximation. These values were used to determine the mean and the variance for the Width variables.

The Depth variables were estimated and set at the same value for all of the source areas since there are no site measurements to evaluate this distance. Because the concentrations in groundwater are a function of the NAPL floating on the groundwater table, the Depth is not expected to be very large. The extent of the mixing of the dissolved chemicals below the water table is represented by the Depth parameter. A minimum of 10 cm and a maximum of 100 cm were chosen to represent all of the source areas.

Table B.3 includes the values for each of the groundwater source areas for the Width and Depth parameters.

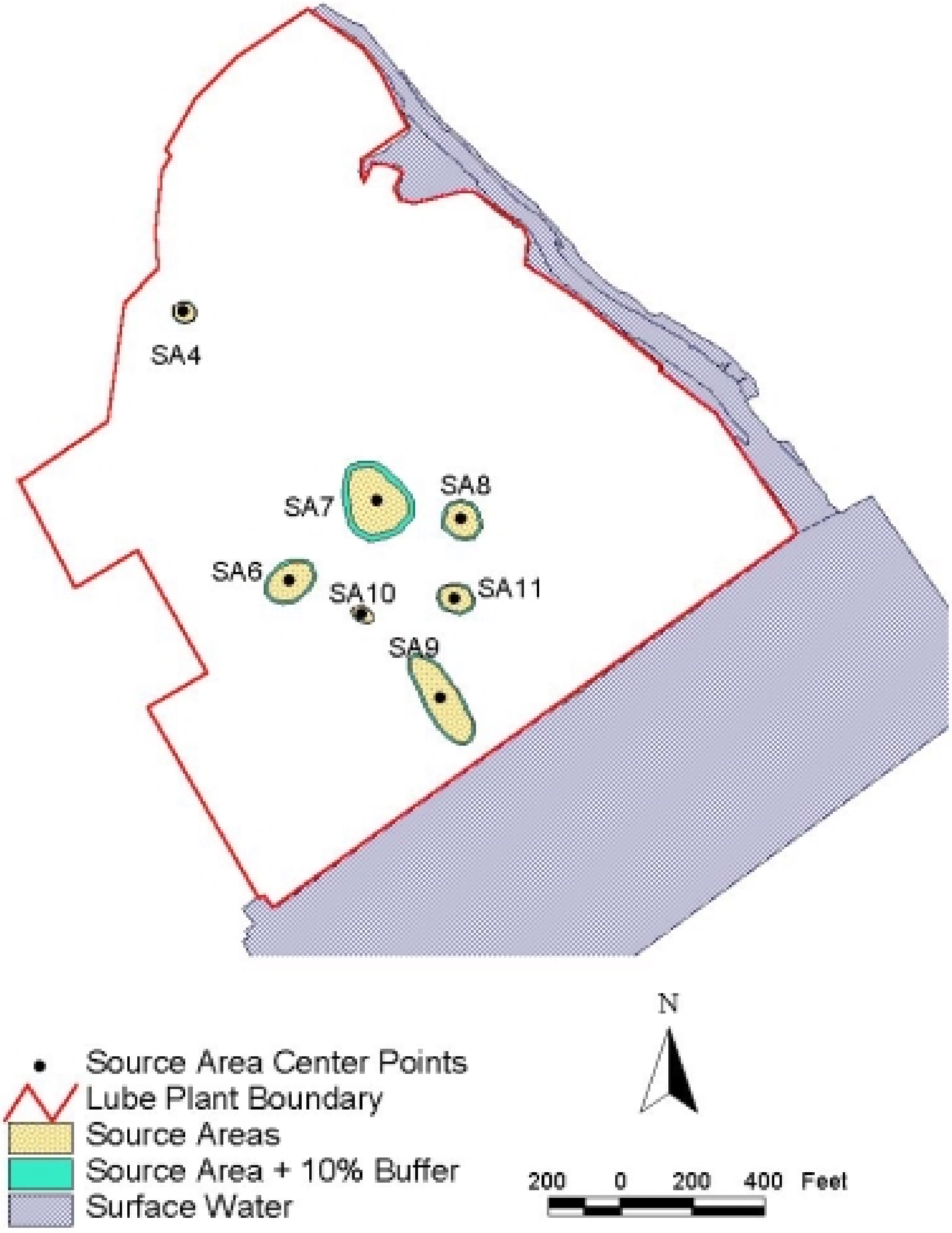


Figure B.9 - Source Areas with 10-Percent Buffer

Table B.3 - Groundwater Exposure Pathway Source Area Width and Depth

Variable	Source Area	Min	Max	Confidence Interval %	Mean	SD	c.o.v.	Variance	D(X)
Depth	All	1.00E+01	1.00E+02	90	5.50E+01	2.74E+01	4.97E-01	7.48E+02	5.50E+01
Width	SA4	1.01E+03	1.24E+03	90	1.13E+03	6.86E+01	6.08E-02	4.70E+03	1.13E+03
Width	SA8	2.26E+03	2.77E+03	90	2.51E+03	1.53E+02	6.08E-02	2.34E+04	2.51E+03
Width	SA6	2.22E+03	2.72E+03	90	2.47E+03	1.50E+02	6.08E-02	2.25E+04	2.47E+03
Width	SA10	5.49E+02	6.71E+02	90	6.10E+02	3.71E+01	6.08E-02	1.37E+03	6.10E+02
Width	SA9	1.92E+03	2.35E+03	90	2.14E+03	1.31E+02	6.12E-02	1.71E+04	2.14E+03
Width	SA7	5.24E+03	6.40E+03	90	5.82E+03	3.54E+02	6.08E-02	1.25E+05	5.82E+03
Width	SA11	1.69E+03	2.07E+03	90	1.88E+03	1.14E+02	6.08E-02	1.31E+04	1.88E+03

Notes: All data distributions are assumed to be normal
All values in cm, except variance in cm^2 and c.o.v. is dimensionless

B.3.2.2 Volatilization to Outdoor Air Exposure Pathway

For the volatilization to outdoor air exposure pathway the needed dimensions are the distance from the soil vapor source to the ground surface (**Lsout**) and the width of the source area perpendicular to the wind direction (**Owidth**).

The vertical dimensions for this volatilization exposure pathway were developed for each source area based on the depth to groundwater measurements and the ground surface elevation data. The ground surface elevation file is the digital elevation model (DEM) from the 1998 aerial survey of the facility (Romanek, *et al*., 1999). In addition, each groundwater monitoring record includes a top of casing elevation. Because of the potential for errors in reading ground surface elevation at each monitoring well location from the DEM, and the vertical control of the DEM, the larger of the DEM elevation value or the top of casing elevation was used to represent the ground elevation. Table B.4 includes the groundwater monitoring well data used in the evaluation for each source area. Because of the frequency of the measurements, groundwater monitoring data were used for all dates later than 1996.

Table B.4 - Monitoring Wells Used for each Source Area for Vadose Zone Thickness Evaluation

Source Area	Monitoring Well
SA4	MW-34
SA6	MW-33
SA7	MW-30
SA8	MW-31
SA9	MW-48
SA10	MW-54
SA11	MW-53

The vadose zone thicknesses, described by the variable **Lsout**, are used to describe the location of the soil vapor concentrations in the subsurface. The descriptive statistics of the corrected groundwater elevation measurement data were calculated in Excel. The sample mean and variance from these data represent a time variation in the vadose zone thickness value. Additional variability exists due to spatial variability across a source area and measurement variation, as well as other sources of variability. The modeling is based on long-term, steady state conditions and one dimensional transport from the center of the source area to the ground surface, or building foundation. Therefore, the time varying measurements were taken as a surrogate for all the variability in these values and a confidence interval applied to the minimum and maximum values for the available data. These values are included in Table B.5. To provide perspective on the variation in the available measurements, the corrected groundwater elevations as a function of time for MW-33 and MW-48 are included as Figure B.10 and Figure B.11, respectively.

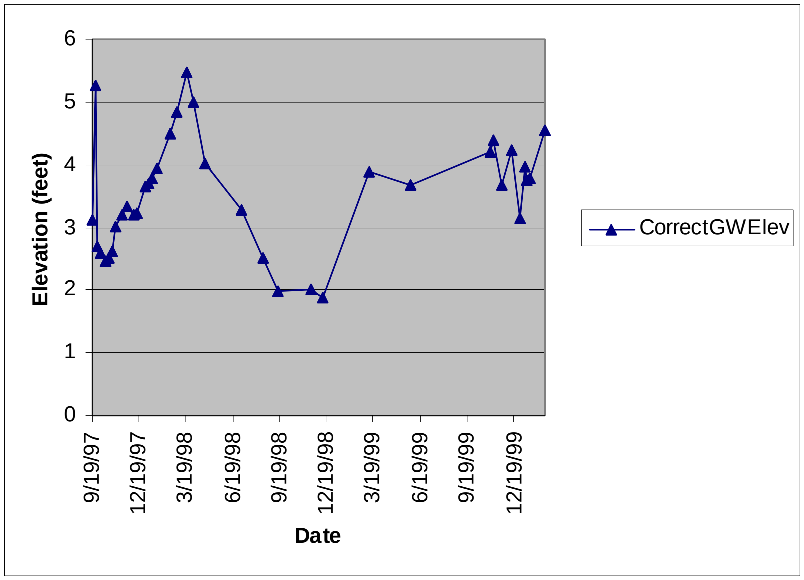


Figure B.10 - Corrected Groundwater Elevations for MW-33 as a Function of Time

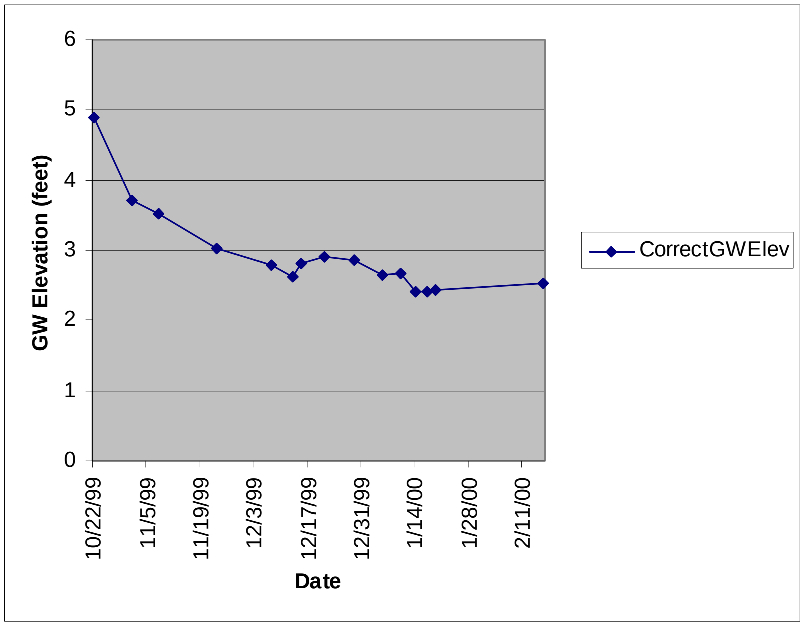


Figure B.11 - Corrected Groundwater Elevations for MW-48 as a Function of Time

The source areas, for which the volatilization to outdoor air exposure pathway was considered, for these calculations, are SA7, SA8, and SA11.

The values for Owidth were developed in ArcView using the measuring tool. It was assumed that the prevailing wind direction at the Former Lube Plant was from the southwest, so the Owidth values were measured from northwest to southeast. The +/-10 percent buffer was used for estimating the minimum width (Owidth-Min) and the maximum width (Owidth-Max).

Table B.5 includes the values for Lsout and Owidth for each of the source areas.

B.3.2.3 Volatilization to Indoor Air Exposure Pathway

For the volatilization to indoor air exposure pathway, the needed dimension is the distance from the soil vapor source to the bottom of the building foundation (**Ls**).

The relationship between the vertical transport distances for the indoor and outdoor air exposure pathways is given as:

$$\mathrm{Ls} = \mathrm{Lsout} - \mathrm{Fdepth}$$ **Equation B.5**

where:
Fdepth (cm) - the depth from ground surface to the bottom of the building foundation
Ls (cm) - the depth from the soil vapor source to the bottom of the building foundation
Lsout (cm) - the depth from the soil vapor source to the ground surface.

A building constructed on a slab set at 1 foot (30.48 cm) below ground surface for the bottom of the slab was assumed for the volatilization to indoor air exposure pathway. The volatilization to indoor air exposure pathway is considered to be a potentially complete exposure pathway for source areas SA6, SA9 and SA10, based on a future industrial building in the vicinity of the now-vacant former Lubricants Warehouse buildings. The values for Ls were calculated in the same manner as Lsout. Table B.4 includes the identity of each monitoring well used for each source area. Table B.5 includes the values for Ls for each source area.

Table B.5 - Source Area Dimensions for the Vapor to Indoor and Vapor to Outdoor Exposure Pathway Segments

Variable	Source Area	Min	Max	Confidence Interval %	Mean	SD	c.o.v.	Variance	D(X)
Ls	SA6	2.78E+02	3.87E+02	90	3.32E+02	3.32E+01	9.99E-02	1.10E+03	3.32E+02
Ls	SA9	1.00E+02	1.76E+02	90	1.38E+02	2.31E+01	1.67E-01	5.33E+02	1.38E+02
Ls	SA10	1.43E+02	2.21E+02	90	1.82E+02	2.38E+01	1.30E-01	5.64E+02	1.82E+02
Lsout	SA7	2.82E+02	3.89E+02	90	3.35E+02	3.27E+01	9.75E-02	1.07E+03	3.35E+02
Lsout	SA8	1.74E+02	2.21E+02	90	1.97E+02	1.43E+01	7.24E-02	2.04E+02	1.97E+02
Lsout	SA11	1.43E+02	1.69E+02	90	1.56E+02	7.91E+00	5.08E-02	6.26E+01	1.56E+02
Owidth	SA7	5.05E+03	6.17E+03	90	5.61E+03	3.41E+02	6.08E-02	1.16E+05	5.61E+03
Owidth	SA8	2.83E+03	3.46E+03	90	3.15E+03	1.91E+02	6.08E-02	3.66E+04	3.15E+03
Owidth	SA11	2.31E+03	2.83E+03	90	2.57E+03	1.56E+02	6.08E-02	2.44E+04	2.57E+03

Notes: All data distributions are assumed to be normal
All values in cm, except variance cm^2 and c.o.v. is dimensionless

B.3.3 Chemical Variables

This section presents the variables used to describe the fate and behavior of the chemicals of concern. The chemical variables used in the equations are the first order degradation (**Degrad**), the organic carbon partitioning coefficient (**Koc**) and the Henry's law equilibrium partitioning coefficient (**Henrys**). The degradation constant is not easily measured, the site properties affect the value, the variability in site properties make it variable across a facility, and the value is non-negative. The Henry's law coefficient represents the relationship between the concentration of a chemical dissolved in water and its vapor concentration. The Henrys values are not easily measured. In addition, site conditions and temperature will affect the value; the natural logarithm of the Henrys value is a function of temperature (Schwarzenbach, *et al.*, 1993) and the values are non-negative. The organic carbon partitioning is not easily measured. It is not a fundamental chemical property. The organic phase used to measure the Koc values is different at the facility from the measurement fluid.

There were a significant number of data sources for these parameters. These reference values are tabulated in Table B.6. The values chosen for the parameter ranges and the calculated mean and variance values are given in Table B.7.

Table B.6 - Chemical Variables

COC_ID	Chemical Name	Variable	Data Sources	ASTM Value	Charbeneau, 2000 (Table 5.4.1)	BP RBDP Values	Other Reference Values
71-43-2	Benzene	Degrad	ASTM 1995a; BP, 1997	0.007, 0.0085, 0.095, 0.007 to 0.024		0.00096, 7.00e-2	
71-43-2	Benzene	Henrys	ASTM1995a; BP, 1997; Schwarzenbach et al, 1993	0.22	0.2288	2.28E-01	T= 22 C Henrys = 0.204; T=28 C, Henrys = 0.258
71-43-2	Benzene	Koc	ASTM1995a; BP, 1997; Schwarzenbach et al, 1993.	38.02	8.30E+01	58.9	17 L/kg (Figure 11.7, Schwarzenbach et al)
100-41-4	Ethyl-benzene	Degrad	ASTM1995a; BP, 1997			0.00304, 0.1	
100-41-4	Ethyl-benzene	Henrys	ASTM1995a; BP, 1997; Schwarzenbach et al, 1993	0.32	0.263	3.23E-01	
100-41-4	Ethyl-benzene	Koc	ASTM1995a; BP, 1997; Schwarzenbach et al, 1993.	1.29E+03	1.10E+03	3.63E+02	

Notes: All values are literature values, not site-specific
Degrad - 1/days
Henrys - dimensionless
Koc - L/kg

Table B.6 - Chemical Variables (continued)

COC_ID	Chemical Name	Variable	Data Sources	ASTM Value	Charbeneau, 2000 (Table 5.4.1)	BP RBDP Values	Other Reference Values
91-20-3	Naphthalene	Degrad	ASTM1995a; BP, 1997			0.00269, 0.7	
91-20-3	Naphthalene	Henrys	ASTM1995a; BP, 1997; Schwarzenbach et al, 1993	0.049	0.047	1.98E-02	T= 22 C Henrys = 0.282; T=28 C, Henrys = 0.369
91-20-3	Naphthalene	Koc	ASTM1739-95; BP, 1997; Schwarzenbach et al, 1993.	1.29E+03	1.30E+03	2.00E+03	
108-88-3	Toluene	Degrad	ASTM1995a; BP, 1997	0.011, 0.067		0.025, 0.1	
108-88-3	Toluene	Henrys	ASTM1995a; BP, 1997; Schwarzenbach et al, 1993	0.26	0.261	2.72E-01	
108-88-3	Toluene	Koc	ASTM1995a; BP, 1997; Schwarzenbach et al, 1993.	134.9	3.00E+02	1.82E+02	

Notes: All values are literature values, not site-specific

Degrad - 1/days

Henrys - dimensionless

Koc - L/kg

Table B.6 - Chemical Variables (continued)

COC_ID	Chemical Name	Variable	Data Sources	ASTM Value	Charbeneau, 2000 (Table 5.4.1)	BP RBDP Values	Other Reference Values
1330-20-7	Xylenes (total)	Degrad	ASTM1995a; BP, 1997	0.004 to 0.014		0.0019, 0.05	
1330-20-7	Xylenes (total)	Henrys	ASTM1995a; BP, 1997; Schwarzenbach et al, 1993; Value from Charbeneau, 2000 is for o-Xylene	0.29	0.209	2.90E-01	
1330-20-7	Xylenes (total)	Koc	ASTM1995a; BP, 1997; Schwarzenbach et al, 1993. Value from Charbeneau, 2000 is for o-Xylene	239.88	8.30E+02	2.40E+02	

Notes: All values are literature values, not site-specific

Degrad - 1/days

Henrys - dimensionless

Koc - L/kg

Table B.7 - Chemical Variable Values

COC_ID	Variable	Min	Max	Confidence Interval %	Mean	SD	c.o.v.	Variance	D(X)
71-43-2	Degrad	9.60E-04	9.50E-02	95	4.80E-02	2.40E-02	5.00E-01	5.76E-04	0.00E+00
71-43-2	Henrys	2.04E-01	2.58E-01	99	2.31E-01	1.05E-02	4.54E-02	1.10E-04	2.31E-01
71-43-2	Koc	1.70E+01	8.30E+01	99	5.00E+01	1.28E+01	2.56E-01	1.64E+02	5.00E+01
100-41-4	Degrad	3.04E-03	1.00E-01	95	5.15E-02	2.47E-02	4.80E-01	6.12E-04	0.00E+00
100-41-4	Henrys	2.63E-01	3.23E-01	95	2.93E-01	1.53E-02	5.22E-02	2.34E-04	2.93E-01
100-41-4	Koc	3.63E+02	1.29E+03	99	8.26E+02	1.80E+02	2.18E-01	3.22E+04	8.26E+02
91-20-3	Degrad	2.69E-03	7.00E-01	95	3.51E-01	1.78E-01	5.06E-01	3.16E-02	0.00E+00
91-20-3	Henrys	1.98E-02	3.69E-01	99	1.94E-01	6.78E-02	3.49E-01	4.59E-03	1.94E-01
91-20-3	Koc	1.29E+03	2.00E+03	99	1.65E+03	1.38E+02	8.38E-02	1.90E+04	1.65E+03
108-88-3	Degrad	1.10E-02	1.00E-01	95	5.55E-02	2.27E-02	4.09E-01	5.15E-04	0.00E+00
108-88-3	Henrys	2.60E-01	2.72E-01	95	2.66E-01	3.06E-03	1.15E-02	9.37E-06	2.66E-01
108-88-3	Koc	1.35E+02	1.82E+02	95	1.58E+02	1.20E+01	7.58E-02	1.44E+02	1.58E+02
1330-20-7	Degrad	1.90E-03	5.00E-02	95	2.60E-02	1.23E-02	4.73E-01	1.51E-04	0.00E+00
1330-20-7	Henrys	2.09E-01	2.90E-01	95	2.50E-01	2.07E-02	8.28E-02	4.27E-04	2.50E-01
1330-20-7	Koc	2.40E+02	8.30E+02	95	5.35E+02	1.51E+02	2.81E-01	2.27E+04	5.35E+02

Notes: All data distributions are assumed to be normal
Degrad - 1/days
Henrys - dimensionless
Koc - L/kg

B.3.4 Above Ground Variables for Volatilization Exposure Pathways

This section describes the variables used in the calculation for volatilization to indoor air exposure pathway for the infiltration and mixing of chemicals of concern in the building and the volatilization to outdoor air for the mixing of chemicals of concern in ambient air. The aboveground variables are the wind speed (**WSpeed**), the areal fraction of foundation cracks (**AF**), the porosity of the soils in the foundation cracks (**PorWall**) and the water-filled porosity of the soils in the foundation cracks (**WCWall**). The areal fraction of cracks is not easily measured and likely changes over time. The porosity and water content of the soils in the foundation cracks are quantities that are not easily measured. They are modeling assumptions for the potentially complex situation of air infiltration from the vadose zone into a building. The assumption is that the particles are a mixture of sand from foundation construction and dust from indoor environment. The wind speed values are based on average reference data included in ASTM (1995a) and ASTM (1999).

The references for the aboveground variable values for the volatilization exposure pathways (i.e., indoor air and outdoor air) are given in Table B.8. The parameter values used for the aboveground variables are given in Table B.9.

Table B.8 - Aboveground Variables

Variable	ASTM Example Value	BP RBDP Value	ASTM, 1999	Other Reference Values
AF	0.01	0.01	0.001, 0.1	0.01 to 0.001
WCWall	0.12	0.19	0.05, 0.4	
TPorWall	0.38	0.44	0.25, 0.5	
WSpeed	225	225	0, 2200	

Notes: All parameters are dimensionless except Wspeed
Wspeed - cm/sec
References:
BP 1997, ASTM 1995a, 1999, 2000, Johnson & Ettinger, 1991.

Table B.9 - Aboveground Variable Values

Variable	Min	Max	Confidence Interval %	Mean	SD	c.o.v.	Variance	D(X)
AF	1.00E-03	1.00E-01	90	5.05E-02	3.01E-02	5.96E-01	9.06E-04	1.00E-01
WCWall	5.00E-02	4.00E-01	90	2.25E-01	1.06E-01	4.73E-01	1.13E-02	1.20E-01
TPorWall	2.50E-01	5.00E-01	90	3.75E-01	7.60E-02	2.03E-01	5.78E-03	4.40E-01
WSpeed	1.00E+01	1.00E+03	99	5.05E+02	1.92E+02	3.81E-01	3.69E+04	1.00E+02

Notes: All parameters are dimensionless except Wspeed - cm/sec
All of the data distributions are assumed to be normal

B.3.5 Subsurface Variables

This section describes the groundwater and soil variables used in the groundwater transport and the vadose zone vapor migration algorithms. The site is generally underlain by fill material (to depths of approximately 10 to 14 feet), silty clay (to depths of approximately 20 to 26 feet) and sand and gravel (to depths of up to 44 feet). The fill material is described as everything from silt, clay and sand to cinders bricks and ashes (Langan, 2000, Kim *et al.*, 2000). The groundwater table is encountered between 6 and 12 feet below grade across the Former Lube Plant, so it is hypothesized that the saturated zone exists in all three generalized soil zones.

Initially, it was expected that separate subsurface modeling scenarios could be developed for the *Former Processing Area*, the *Heavy Fuels Area* and the *West Tank Farm* since it appears that there may be significant amounts of fill material in the Former Processing Area as well as in the West Tank Farm, while the Heavy Fuels Area appears to be mostly native materials (Kim *et al.*, 2000, Langan, 1999 and 2000). However, most of the field data that have been collected are from the central portion of the Former Lube Plant; in the Former Processing Area. In addition, the limited data outside of the Former Processing Area did not appear to be significantly different from the data collected in the Former Processing Area. Therefore, one modeling scenario (SS1) has been developed to apply across the Former Lube Plant using all of the available data.

The groundwater gradient (**GWGrad**) was estimated based on site measurements. The groundwater elevation at each source area ($\mathbf{h_{source}}$) and at the corresponding point of demonstration (**h_{POD}**) were determined from the interpolated groundwater grid (i.e., based on the November 1, 1999 groundwater measurements). Because the groundwater transport model used in these calculations is a one-dimensional flow model, the simple gradient between each source area and point of demonstration was used to estimate the applicable groundwater gradient for the calculations. Equation B.6 was used to calculate the groundwater gradient.

$$\mathrm{GWGrad} = \frac{h_{\mathrm{source}} - h_{\mathrm{POD}}}{\mathrm{sDist}}$$ **Equation B.6**

A GWGrad-Min value was calculated using the sDist-Max value for the individual source area. The GWGrad-Max value was calculated using the sDist-Min value. The results of these calculations were used to determine the overall minimum GWGrad and overall maximum GWGrad for the Former Lube Plant. Because this is a simplified data evaluation for a complex flow system, a relatively low confidence interval of 80 percent was applied to the range of values.

The hydraulic conductivity (**HydCond**) of the saturated zone was calculated based on 34 grain size measurements and 11 slug tests of monitoring wells. The grain size values were converted to hydraulic conductivity using the Hazen power-law equation (Freeze and Cherry 1979).

$$K = Ad_{10}^{2}$$ **Equation B.7**

where
K - hydraulic conductivity (cm/sec)
A - coefficient = 1
d_{10} - grain-size diameter at which 10-percent by weight of the soil particles are finer than this value (mm).

The results of the slug tests are tabulated in the environmental measurements database. The groundwater transport model is based on a homogeneous, isotropic system so only one hydraulic conductivity value is used in the calculation. However, there are at least three geologic zones through which groundwater is moving. Even in a relatively simple geologic system, there can be large variations in hydraulic conductivity, therefore it is a difficult task to estimate a single value that represents the overall, or effective hydraulic conductivity. To account for the higher uncertainty in the effective hydraulic conductivity value, the minimum and maximum values from the data set were used along with a confidence interval, to estimate the mean and variance values to apply across the Former Lube Plant. A comparison using a lognormal distribution and a normal distribution to estimate the hydraulic conductivity was made. Given the data range, the lognormal distribution parameters were calculated and then the mean and variance were calculated. The confidence interval method required a 99.99-percent confidence interval for the calculated mean to be between the minimum and maximum. A better alternative was to use a normal distribution and assign a low confidence interval, 70-percent, so that the form of the distribution was not as much a factor (Davenport, 1970).

All of the hydraulic conductivity data are summarized in Figure B.12.

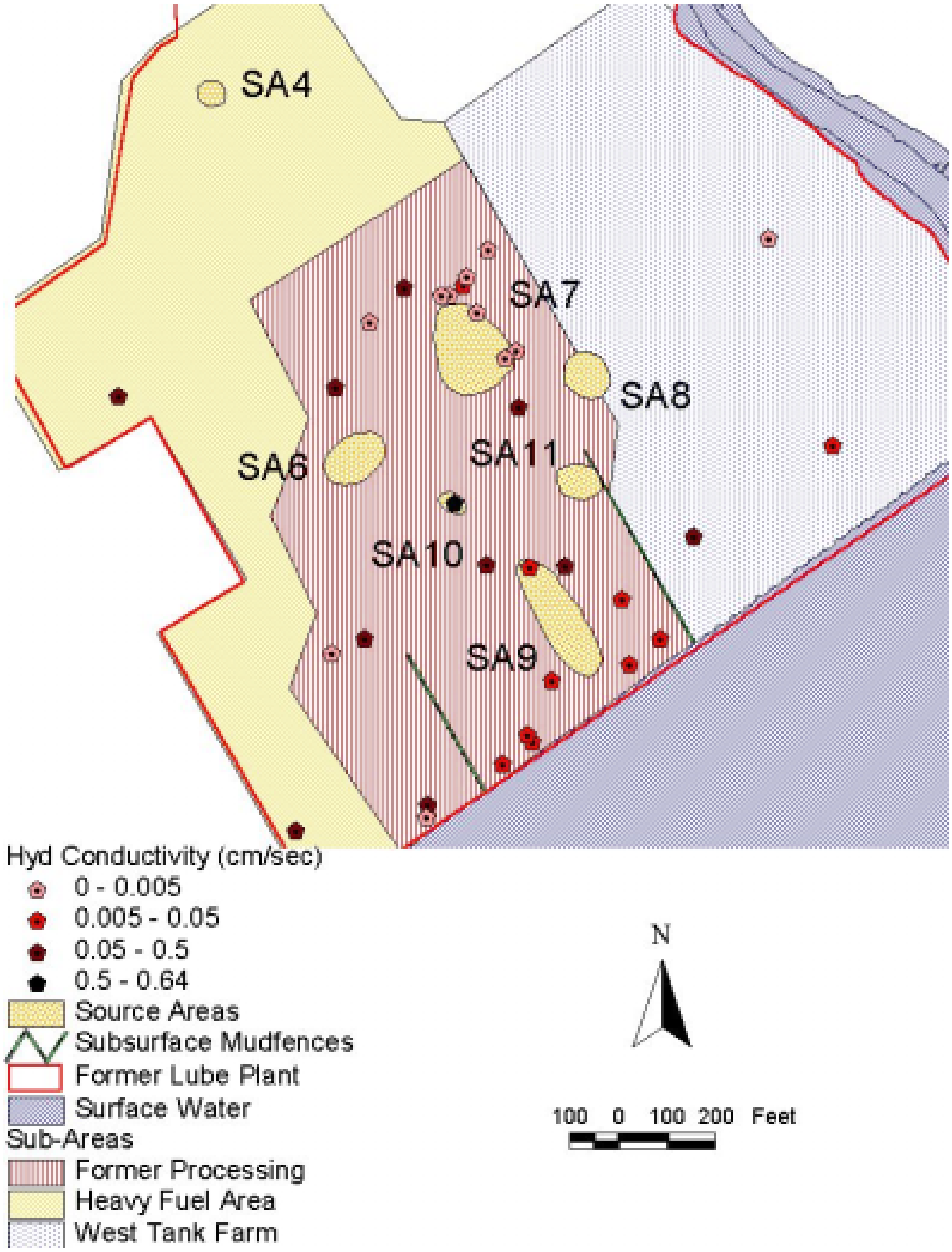


Figure B.12 - Hydraulic Conductivity Data (cm/sec)

There are no site-specific measurements of saturated zone porosity (**TPor**), vadose zone porosity (**PorVad**) or vadose zone water content (**WCVad**). Reference values from BP (1997) and ASTM (1999) were used to estimate these parameters. For the saturated zone, the values for TPor cover the ranges for gravel, sand and silt. For the unsaturated zone, the values for WCVad and PorVad cover the ranges for sand and silt.

For the fraction of organic carbon content of soils in the saturated zone (**SatFoc**) reference values from BP (1997), ASTM (1999) and PaDEP (1997) were evaluated and the PaDEP default value used for the D(X) value for SatFoc. In addition, there were 19 measurements of fraction organic content in the environmental measurements database from September 1999. The site samples were taken during soil boring installations in the Lube Plant. The minimum and maximum values from the sample results for soils in the saturated zone, generally at depths below 10 feet, were used along with a confidence interval, for the mean and variance estimates.

The references for the subsurface variable values are given in Table B.10. The subsurface variable values are given in Table B.11.

Table B.10 - Subsurface Variables

Variable	Data Sources	ASTM Example Value	BP RBDP Value	ASTM, 1999	Other Reference Value
GWGrad	Site Data - 11/1/99 elevation data				
HydCond	Site Data - 45 measurements				
TPor	ASTM 1999, BP 1997	0.38	0.44	0.25,0.5	
PorVad	ASTM 1999, BP 1997	0.38	0.44	0.25,0.5	
WCVad	ASTM 1999, BP 1997	0.12	0.19	0.05, 0.4	
SatFoc	ASTM 1999, BP 1997, PaDEP 1997; 19 saturated zone samples from 9/99	0.01	0.01	0.0001, 0.03	0.005

Notes: All variables a dimensionless, except HydCond - cm/sec

Table B.11 - Subsurface Variable Values

Variable	Min	Max	Confidence Interval %	Mean	SD	c.o.v.	Variance	D(X)
GWGrad	4.06E-04	1.08E-02	80	5.6E-03	4.1E-03	7.2E-01	1.7E-05	5.6E-03
HydCond	1.00E-05	6.40E-01	70	3.2E-01	3.1E-01	9.6E-01	9.5E-02	3.2E-01
TPor	2.50E-01	5.00E-01	90	3.8E-01	7.6E-02	2.0E-01	5.8E-03	4.4E-01
PorVad	2.50E-01	5.00E-01	95	3.8E-01	6.4E-02	1.7E-01	4.1E-03	4.4E-01
WCVad	5.00E-02	4.00E-01	95	2.3E-01	8.9E-02	4.0E-01	8.0E-03	1.9E-01
SatFoc	5.00E-04	5.30E-02	95	2.7E-02	1.3E-02	5.0E-01	1.8E-04	5.0E-03

Notes: All of the data distributions are assumed normal
All variables a dimensionless, except HydCond - cm/sec

B.3.6 Surface Water Mixing Variables

This section describes the variables used in the surface water mixing model. The surface water flow rate (**Qcrk**) for Marcus Hook Creek was estimated by Kim, *et al*., (2000) using a drainage area ratio method and flow data for the Chester Creek available from the USGS. The long-term median flow value for the Delaware River is 2.4 x 10^3 ft^3/sec (USGS, 2000). The mixing in the river is estimated to occur only in a small portion of the river close to the shoreline. A value of 10-percent of the total flow was estimated to be the mean mixing flow for the groundwater discharging to the river. A relatively low confidence interval of 70-percent was assigned to the river mixing flow.

The groundwater plume mixing depth (**H**) at the point where the groundwater discharges to the surface water was estimated based on the source depth parameter.

The surface water mixing variables are given in Table B.12.

Table B.12 - Surface Water Mixing Variables

Variable	SW Scenario ID	Min	Max	Confidence Interval %	Mean	SD	c.o.v.	Variance	D(X)
Qcrk	SW1	2.1E+02	3.4E+02	80	2.8E+02	5.0E+01	1.8E-01	2.5E+03	2.8E+02
H	SW1	1.0E+02	3.0E+02	70	2.0E+02	9.6E+01	4.8E-01	9.3E+03	2.0E+02
Qcrk	SW2	5.7E+03	7.7E+03	70	6.7E+03	9.6E+02	1.4E-01	9.3E+05	6.7E+03
H	SW2	1.0E+02	3.0E+02	70	2.0E+02	9.6E+01	4.8E-01	9.3E+03	2.0E+02

Notes: Qcrk values are in L/sec
H values are in cm
All data distributions are assumed to be normal
The Qcrk for the Delaware River is 10-percent of the long-term median flow
SW_ScenarioID SW1 is Marcus Hook Creek
SW_ScenarioID SW2 is the Delaware River
Flow Data References: Kim *et al*., 2000, USGS 2000

B.3.7 Surface Water Mixing Chemical Variables

The surface water mixing model requires an input chemical concentration for each chemical of concern in the upstream surface water. Because there are no field measurements of these input concentrations and very little information about how to estimate specific chemical concentrations for surface water runoff from urban land, the values for all the chemicals (i.e., mean, variance and D(X)) were set to zero. Romanek, *et al*., (1999) attempted to find good data for benzene concentrations in urban runoff and most of the measurements that were available were total petroleum hydrocarbon (TPH) data and not benzene data. In addition, the chemicals of concern studied in these example calculations are relatively volatile, so concentrations in surface water are not expected to persist.

B.3.8 Chemical Property Values

This section describes the chemical property values that are treated as constants in the calculations. The chemical property values used in the exposure pathway calculations are the molecular diffusion coefficient in air (**Dair**) the molecular diffusion coefficient in water (**Dwat**), and the chemical solubility in water (**Sol**).

The chemical property constants are given in Table B.13.

Table B.13 - Chemical Property Constants

COC_ID	Chemical Name	Constant	Value	Units
71-43-2	Benzene	Dair	8.80E-02	cm2/sec
71-43-2	Benzene	Dliq	9.80E-06	cm2/sec
71-43-2	Benzene	Sol	1.75E+03	mg/L
100-41-4	Ethylbenzene	Dair	7.50E-02	cm2/sec
100-41-4	Ethylbenzene	Dliq	7.80E-06	cm2/sec
100-41-4	Ethylbenzene	Sol	1.69E+02	mg/L
91-20-3	Naphthalene	Dair	5.90E-02	cm2/sec
91-20-3	Naphthalene	Dliq	7.50E-06	cm2/sec
91-20-3	Naphthalene	Sol	3.10E+01	mg/L
108-88-3	Toluene	Dair	8.70E-02	cm2/sec
108-88-3	Toluene	Dliq	8.60E-06	cm2/sec
108-88-3	Toluene	Sol	5.26E+02	mg/L
1330-20-7	Xylenes (total)	Dair	7.20E-02	cm2/sec
1330-20-7	Xylenes (total)	Dliq	8.50E-06	cm2/sec
1330-20-7	Xylenes (total)	Sol	1.98E+02	mg/L

Data Source for all values BP (1997)

B.3.9 Physical Property Values

This section describes the physical property values that are treated as constants in the calculations. The indoor air exchange rate (**ER**) for a hypothetical building is a design parameter and may be specified by a building code. The volume to plan area ratio for a building (**LB**) is essentially the indoor height of the building. This is a design parameter. The foundation slab or wall thickness (**WallT**) is a design parameter for the building. The soil particle density (**ParDen**) is considered to be a constant value. For the outdoor air inhalation scenario a mixing height (**MixHt**) is needed. The value is based on the simplified box model and is the height of the receptor breathing zone.

The physical property values are given in Table B.14.

Table B.14 - Physical Constants

Constant	Data Sources	Value	Units
ER	BP, 1997	2.30E-04	1/sec
LB	BP, 1997	4.88E+02	cm
WallT	BP, 1997	1.50E+01	cm
ParDen	Charbeneau, 2000	2.65E+00	g/cm3
MixHt	BP, 1997	2.00E+02	cm

B.3.10 Model Error Variables

This section describes the variables used to account for the differences between the results of the simplified mathematical algorithms and actual behavior of chemicals of concern in the natural environment. The model error variables are included for the groundwater transport, the vapor to indoor air, the vapor to outdoor air and the surface water mixing models. The model error variables are designed to be used to account for variability that would be difficult to quantify otherwise, or is not quantified elsewhere in the procedure and to provide an opportunity using field data, to calibrate the environmental models. Two types of errors are modeled. The first is a multiplicative error (i.e., **NyGW, NyI, NyO, NySW**); where the actual ending concentration value is some multiple larger or smaller than the predicted ending concentration. The second is an additive error (i.e., **EyGW, EyI, EyO, EySW**) where the actual ending concentration value is larger or smaller than the predicted ending concentration by some discrete amount. The mean values and the deterministic values for the multiplicative error terms are set to one and the additive terms are set to zero in the initial model runs. In the probabilistic case, because there are no data to indicate what other value should be chosen; the variance values are set to zero for the additive terms and 0.25 for the multiplicative terms. The variance in the multiplicative error terms is based on a c.o.v. value of 0.5. If sufficient site chemical concentration data are available, the model error variables can then be used during the model calibration to adjust the model results to agree with site data and to recalculate the variance in the ending concentrations. Additional discussion of the use of the model error variables is included in Chapter 5.

B.4 PARAMETER CORRELATIONS

This section describes the methodology used to assign correlation coefficients for parameters within each random variable set. It must be emphasized that this methodology only applies to variables within a specific parameter set (i.e., random variables that are dependent or potentially dependent upon one another). The separation of variables into random variable sets permits the treatment of each set of variables as statistically independent of one another. The correlations therefore that are developed for each parameter are only for the other variables within the parameter's specific variable set. This is a simplification of the reality of the correlated variables, but it provides a practical and implementable solution for incorporating correlations.

The correlation coefficient (ρ_{ij}) values for which are developed in this section, represents the degree to which two variables are related. This is expressed in Equation B.8.

$$\rho_{ij} = \frac{\mathrm{COV}(i, j)}{\sigma_i \sigma_j}$$ **Equation B.8**

where
COV(i,j) - covariance between i and j
σ_i - standard deviation of i.
(Benjamin and Cornell, 1970)

In addition, the value of the correlation coefficient is bounded as shown in Equation B.9.

$$-1 \leq \rho_{i,j} \leq 1$$ **Equation B.9**

When two random variables are *positively correlated,* larger than mean values of the variable i are expected to occur when larger than mean values of the variable j occur. The same is true if both variables tend to be smaller than the mean. In the case where larger than mean values of i occur with smaller than mean values of j, the random variables are said to be *negatively correlated*. If the two variables are independent, then the correlation coefficient is zero (Gilbert, 1997). However, small values of correlation coefficients do not necessarily mean that the variables are independent. It is important to recognize that a value of +/-1 for the correlation coefficient implies a linear relationship between the variables i and j. Therefore, variables may be stochastically dependent, and in fact be functionally dependent, with some higher order functionality, but if they are not linearly dependent, then the correlation coefficient will have an absolute value less than one (Benjamin and Cornell, 1970, DeVore, 1995).

As presented in section B.2, many of the variable input values are based on reference information and professional judgment. Very few of the variables had sufficient field-measured data values to develop estimates of the population mean and variance from the sample data. In addition, for the case study facility there were no parameter studies to look at the potential for correlation in site parameter values based on field measurements. Therefore, no direct data analyses could be implemented to develop the correlation coefficients to be used in the example calculations. Instead of assuming that all of the input values are statistically independent, which is known not to be the case since there are many related variables used in the model calculations (e.g., total soil porosity and water-filled porosity), a simple procedure was implemented to assign correlation coefficients to variable pairs.

It was decided that a discrete value function would be used to assign correlation values, instead of a continuous function where the correlation coefficients could have any real number value. This approach was used because it makes the determination of correlation coefficients simpler. In addition, since the method is approximate (i.e., no site field studies have been conducted to determine parameter correlations) it uses an appropriate level of complexity to evaluate the variance.

For each parameter pair within a set the following steps were implemented:

1. A determination of the type of correlation possible:
    - No correlation
    - Positive Correlation
    - Negative Correlation
    - Perfect (e.g., the variable with itself, linearly related variables).

2. For parameter pairs with no correlation, the coefficient value assign is zero.
3. For parameter pairs with perfect correlation, the coefficient value assign is one.
4. For the remaining parameter pairs, a determination is made whether the correlation is weak or strong.
5. The correlation values are assigned based on the following function:
    - Weak correlation +/- 0.5
    - Strong correlation +/- 0.9
6. To check the assigned correlation values, the determinant of each matrix of correlation coefficients must be positive and definite.

The central question when evaluating parameter correlation is "when variable I is underestimated (or over estimated) what happens to variable J?" The determination must be made based on the functional relationship between the variables and the site data available. In the cases where limited site data are available, then the determination is made based on the literature sources for the individual parameters. As an example, if the hydraulic conductivity is measured and it is underestimated then it is likely that the soil porosity would be underestimated. This means the variables are positively correlated. In addition, since the hydraulic conductivity is functionally related to the porosity of the soils, but the texture of the soils has a modifying influence on that relationship, see section B.4.3, a weak correlation relationship is assumed. Therefore, the correlation coefficient for hydraulic conductivity **(HydCond)** and porosity **(TPor)** is 0.5.

Each parameter pair was evaluated individually, with respect to the relationship between the parameters, and with respect to other properties (e.g., each of the chemical property variables was evaluated with respect to the other two property variables and with respect to molecular weight).

The correlations are only assigned for variables within a parameter set that are also used in a group calculation. It may be true that there are other correlations that exist, however, the mechanism to account for any affect of correlation on the potential exposure concentrations is limited to the variables within the sequence of segments that make-up an exposure pathway, and between two variables at a time. As a check on the assigned correlation values, the determinant of each matrix of correlation coefficients was calculated. The determinant must be positive and definite for the correlation values to be valid for the group of variables involved. All of the determinants for the values presented here were positive and definite. The Excel function **mdeterm** is used for the calculation.

The following sections provide the values used for the example calculations for the correlation coefficients for each parameter set.

B.4.1 Above Ground Variable Correlations for Volatilization Exposure Pathways

This section describes the correlation coefficients for the variables used in the volatilization to indoor air and the volatilization to outdoor air exposure pathway calculations. Of the variable pairs in this group, only three were non-zero and non-one. These are AF and WCWall, AF and TPorWall, and WCWall and TPorWall. The areal fraction (**AF**) of cracks in the foundation would be related to the porosity (**TPorWall**) and the water content of the soils in the cracks (**WCWall**). With time the areal fraction of cracks would be higher and the porosity of the soils would be lower, since the assumption is that the soil starts out being sand from the foundation construction, but more fine particles would collect in the cracks as the building ages, so the sediments would start out as well-sorted and with time would be more heterogeneous. The water content would decrease with lower total porosity in the cracks.

The porosity and the water content of the soils are related by Equation B.10.

$$\mathrm{TPorWall} = \mathrm{WCWall} + \mathrm{ACWall}$$ **Equation B.10**

where
ACWall - volumetric air content of soils (dim)
WCWall - volumetric water content of soils (dim)
TPorWall - total porosity of soils (dim).

As the porosity decreases, in general the water content would be expected to decrease within the cracks because there is less volume for water. So the relationship between total porosity and water content is not linear; it is expected to be a weak relationship, and a positive one.

Table B.15 includes the parameter pairs in this group, the correlation type, an indication of the functional relationship, the weak or strong determination and the correlation coefficient value. Table B.16 includes the correlation coefficients in a matrix format.

Table B.15 - Aboveground Variable Correlation Coefficients

Parameter Pair	Type	Functional Relationship	Weak/Strong	Value
AF/AF	Perfect			1
AF/WCWall	Negative		Weak	-0.5
AF/TPorWall	Negative		Weak	-0.5
AF/WSpeed	None			0
WCWall/WCWall	Perfect			1
WCWall/TPorWall	Positive	Yes	Weak	0.5
WCWall/Wspeed	None			0
TPorWall/TPorWall	Perfect			1
TPorWall/WSpeed	None			0
Wspeed/Wspeed	Perfect			1

Table B.16 - Aboveground Variables Correlation Matrix

	AF	WCWall	TPorWall	WSpeed
AF	1	-0.5	-0.5	0
WCWall		1	0.5	0
TPorWall			1	0
WSpeed				1

B.4.2 Chemical Variable Correlations

This section presents the correlation coefficients for the variables used to describe the fate and behavior of the chemicals of concern. The three variables in this parameter set are Henry's Law partitioning coefficient (**Henrys**), the first order decay constant (**Degrad**) and the organic carbon partitioning coefficient (**Koc**). All of these variables were considered to be correlated, based on chemical properties.

The Henrys partitioning coefficient is the ratio of vapor pressure to solubility as shown in Equation A.4, Appendix A. In general, for higher molecular weight organic chemicals the Henrys value decreases. Table B.6 Chemical Variable Values indicates this, comparing the mean Henrys value for Benzene ($2.3\text{x}10^{-1}$) and the mean Henrys for Naphthalene ($1.94\text{x}10^{-1}$). The degradation constant, Degrad, in general, is lower for higher molecular weight chemicals. It takes longer to degrade more complex, larger molecules. The organic partitioning constant Koc, is higher for larger molecules. Based on this simplified information, the relationship between Degrad and Henrys is expected to be positive. When the Henrys values decrease the Degrad values decrease, but because there is no direct functional relationship, and the structure and activity of different molecular groups affect both the degradability of a chemical and its solubility and vapor pressure, the relationship is considered to be weak. The Degrad value is considered to be negatively correlated with Koc. When Koc is high, the Degrad value is expected to be low. Again, the relationship is expected to be weak. Finally, when Henrys values increase the Koc values are expected to decrease, so the parameters are negatively correlated, but the relationship is weak. Table B.17 includes all of the values for the chemical variable set. Table B.18 includes the correlation coefficient matrix.

Table B.17 - Chemical Variable Correlation Coefficients

Parameter Pair	Type	Functional Relationship	Weak/Strong	Value
Degrad/Degrad	Perfect			1
Degrad/Henrys	Positive		Weak	0.5
Degrad/Koc	Negative		Weak	-0.5
Henrys/Henrys	Perfect			1
Henrys/Koc	Negative		Weak	-0.5
Koc/Koc	Perfect			1

Table B.18 - Chemical Variable Correlation Matrix

	Degrad	Henrys	Koc
Degrad	1	0.5	-0.5
Henrys		1	-0.5
Koc			1

B.4.3 Model Error Variable Correlations

This section describes the correlation coefficients for the model error variables used in the segment calculations. This variable set includes the groundwater multiplicative and additive error variables (NYGW, EYGW) the surface water multiplicative and additive error variables (NYSW, EYSW), indoor air multiplicative and additive error variables (NYI, EYI), and the outdoor air multiplicative and additive error variables (NYO, EYO). Since only the groundwater and surface water variables are used in the same group calculations, these are the only variable for which correlation coefficients would be needed. Without any information about the applicability of the models to the case study facility, through surface water sampling data for instance, the groundwater and surface water model error variables were considered to be independent.

B.4.4 Subsurface Variable Correlations

This section describes the correlation coefficients for the variables used in the groundwater transport pathway and the vadose zone vapor migration algorithms. The parameters that are expected to be related are total porosity in the saturated zone (**TPor**), hydraulic conductivity (**HydCond**), groundwater gradient (**GWGrad**). The vadose zone parameters of total porosity (**PorVad**) and (**WCVad**) are also expected to be related.

For the saturated zone parameter, the functional relationship is given in Equation B.11.

$$V_{GW} = \frac{\mathrm{HydCond} \times \mathrm{GWGrad}}{\mathrm{TPor}}$$ **Equation B.11**

The relationship between hydraulic conductivity and total porosity is made more complex because the porosity as well as the texture of the soils has an influence on the hydraulic conductivity. Clean sands can have a high porosity and a high hydraulic conductivity, while fine silts and clays can also have a high porosity but a low hydraulic conductivity (Freeze and Cherry, 1979). It is expected that the correlation coefficient is positive, but the relationship is weak.

For groundwater gradient and hydraulic conductivity, the relationship is strong. If the gradient is high, it is likely that the hydraulic conductivity is low, that there is a greater resistance to groundwater flow, but if the hydraulic conductivity is high, in general, the gradient will be low. The correlation is therefore, strong and negative.

For groundwater gradient and porosity the soil texture and heterogeneity factors into the assessment of the porosity. A clean sand generally, has a low groundwater gradient and a high porosity, while a heterogeneous silty, clayey sand may have a high porosity and a high groundwater gradient. The relationship between groundwater gradient and porosity is negative and weak.

For the vadose zone parameters, the functional relationship is parallel to Equation B.10 for the soils in the foundation cracks. Equation B.12 shows the relationship between PorVad and WCVad.

$$\text{PorVad} = \text{WCVad} + \text{ACVad} \qquad \textbf{Equation B.12}$$

where
ACVad - volumetric air content of soils(dim)
WCVad - volumetric water content of soils (dim)
PorVad - total porosity of soils (dim).

The relationship is made more complex because the porosity as well as the texture of the soils determines the water content. Clean sands can have a high porosity but a low water content, while fine silts and clays can also have a high porosity but a high water content. For heterogeneous soils, like those that are found at the Former Lube Plant, it is expected that when the total porosity increases, the water content increases. The correlation coefficient is positive, but the relationship is weak.

Table B.19 includes all of the values for the subsurface variable set. Table B.20 includes the correlation coefficient matrix for the saturated zone variables. Table B.21 includes the correlation coefficient matrix for the vadose zone variables.

Table B.19 - Subsurface Variable Correlation Coefficients

Parameter Pair	Type	Functional Relationship	Weak/Strong	Value
GWGrad/GWGrad	Perfect			1
GWGrad/HydCond	Negative	Yes	Strong	-0.9
GWGrad/TPor	Negative	Yes	Weak	-0.5
GWGrad/SatFoc	None			0
HydCond/HydCond	Perfect			1
HydCond/TPor	Positive	Yes	Weak	0.5
HydCond/SatFoc	None			0
TPor/TPor	Perfect			1
TPor/SatFoc	None			0
SatFoc/SatFoc	Perfect			1
PorVad/PorVad	Perfect			1
PorVad/WCVad	Positive	Yes	Weak	0.5
WCVad/WCVad	Perfect			1

Table B.20 - Saturated Zone Variables Correlation Matrix

	GWGrad	HydCond	TPor	SatFoc
GWGrad	1	-0.9	-0.5	0
HydCond		1	0.5	0
TPor			1	0
SatFoc				1

Table B.21 - Vadose Zone Variables Correlation Matrix

	PorVad	WCVad
PorVad	1	0.5
WCVad		1

B.4.5 Surface Water Mixing Variable Correlations

This section describes the correlation coefficients for the variables used in the surface water mixing algorithm. This variable set includes the surface water flow rate (**Qcrk**) and the groundwater plume depth (**H**) at the surface water.

These variables are not expected to be related. Table B.22 includes all of the values for the surface water variable set.

Table B.22 - Surface Water Variable Correlation Coefficients

Parameter Pair	Type	Functional Relationship	Weak/Strong	Value
Qcrk/Qcrk	Perfect			1
Qcrk/H	None			0
H/H	Perfect			1

B.4.6 Source Area Variable Correlations

This section describes the correlation coefficients for the variables used to describe each of the source areas. Three variables were considered to be related in this parameter set. These are the groundwater source zone depth (**Depth**), the groundwater source zone width (**Width**) and the groundwater travel distance (**sDist**). The Width and the Depth are expected to be positively correlated because they are a function of the same interpretation of the extent of the source area concentrations. The Width and the sDist are expected to be correlated because the width and length of the source area define the location of the source area centroid, and based on the method used here, the length defines the variability in the sDist dimension. The Width and the sDist are expected to be negatively correlated. The sDist and the Depth are negatively correlated based on their relationships to the Width. All three of the correlations are expected to be weak.

This variable set also includes the depth to vapor sources for the outdoor air calculation (**Lsout**) and the source width perpendicular to the wind direction (**Owidth**). These variables are not expected to be related. The depth to vapor sources for the indoor air calculation (**Ls**) is included in this parameter set, but it is the only variable from that segment that describes the source area, so it does not have any potential correlation relationships.

Table B.23 includes all of the values for the source area variable set. Table B.24 includes the correlation coefficient matrix for the groundwater source area variables. Table B.25 includes the correlation coefficient matrix for the outdoor air source area variables.

Table B.23 - Source Area Variable Correlation Coefficients

Parameter Pair	Type	Functional Relationship	Weak/Strong	Value
Depth/Depth	Perfect			1
Depth/Width	Positive		Weak	0.5
Depth/sDist	Negative		Weak	-0.5
Width/Width	Perfect			1
Width/sDist	Negative		Weak	-0.5
sDist/sDist	Perfect			1
Lsout/Lsout	Perfect			1
Owidth/Owidth	Perfect			1
Lsout/Owidth	None			0
Ls/Ls	Perfect			1

Table B.24 - Groundwater Source Area Variable Correlation Matrix

	Depth	Width	sDist
Depth	1	0.5	-0.5
Width		1	-0.5
sDist			1

Table B.25 - Outdoor Air Source Area Variable Correlation Matrix

	Lsout	Owidth
Lsout	1	0
Owidth		1

Appendix C : Computer Implementation and Procedures for Deterministic Calculations

This Appendix presents the computer procedures used in the application of the Spatial Environmental Risk Assessment (SERA) methodology. Specifically the methods for developing the spatial database for the site conceptual model, the site conceptual model tabular database, and the deterministic spreadsheet calculations. The discussion of the User Defined Functions in Excel is also presented. The development of the exposure pathways in the site conceptual model database and the spreadsheet calculations for the probabilistic case are presented in Appendix D.

C.1SPATIAL DATABASE DATA DEVELOPMENT

The digital facility description is used to construct the GIS files for the site conceptual model. Table C.1 includes the listing of the types of files used in the case study examples. Other files may be necessary in a specific application depending on the facility conditions and the surrounding land use and features. Appendix E includes the complete listing, by file name, of the spatial database files.

The environmental monitoring data, information about historical releases, information about waste handling areas and process areas, among other information, are used to identify source areas. Source areas are defined in a polygon shapefile. The shapefile is constructed to identify the approximate extent of all of the source areas. Shapefiles are constructed in ArcView using the **View,**

New Theme commands. A polygon shapefile is selected and the irregular polygon-drawing tool is used. The new theme is saved as a shapefile by using the **Theme Convert To Shapefile** command.

During the course of data collection it is likely that the source areas will be revised and updated. The shapefiles are identified sequentially by date and version (e.g., sourcav1.shp, sourcav2.shp) so that the updates to the identified areas can be tracked during the risk evaluation. The files are tracked in an electronic files data table as illustrated in Appendix E. This documentation is critical to maintaining records of the meta data for the site conceptual model.

Table C.1- Site Conceptual Model GIS Data Layers

Data Layer	Type	Data Description
Source Areas	Polygon	Area of NAPL or highest concentration
Source Area Center Points	Point	Centroid of source area
Soil Vapor Transition Points	Point	Location of end of NAPL partitioning to vapor exposure pathway segment
Groundwater Transition Points	Point	Location of end of NAPL partitioning to groundwater exposure pathway segment
Surface Water Transition Points	Point	Location of end of groundwater transport exposure pathway segment
Groundwater Transport Segments	Line	Calculation length for groundwater attenuation mechanism
Surface Water Mixing Boxes	Polygon	Area of application of the target level for surface water
Air Mixing Boxes	Polygon	Area of application of the target level for air
Points of Demonstration	Point	Location of comparison of estimated concentration to target level
Analysis Boundary	Polygon	Facility area or boundary included in the risk assessment
Surface Water	Polygon	Location of bounding surface water

Once the source area shapefile has been constructed, the file is edited to add and populate fields in the data table. The data table is identified as a file of the same name as the shapefile, except it has a *.dbf file extension (e.g., sourcav1.dbf). The features that are added to the data table include:

- LOC_ID, the unique identifier for the source area that will match the records in the site conceptual model relational database;

- Description, a field with a brief description of the source area (e.g., NAPL, soil concentrations);
- x-,y-coordinates, added using CRWR Vector;
- area, added using CRWR Vector;
- perimeter, added using CRWR Vector;
- Width, measured using the Measuring Tool;
- Owidth, measured using the Measuring Tool; and,
- Any other physical dimensions of the source area needed for the exposure pathway calculations.

The CRWR Vector extension is a group of scripts developed at The University of Texas at Austin, Center for Research in Water Resources, for basic operations on vector data (Olivera, 1999). Table C.2 includes the functions in the extension.

Table C.2 - CRWR Extension Functions

CRWR-Vector
• Project
• Fishnet
• Theissen polygons
• Clip by graphic
• Polygons to polylines
• Poly to poly property transfer
• Update feature geometry
• Add XY to table
• Add record number to table

(Olivera, 1999)

The scripts operate on vector shapefiles. The various operations are available to match the type of shapefile (e.g., area is only available for polygon files, x- and y-coordinates are available for point files and polygon files). The x- and y-coordinates for a polygon shapefile are the centroid coordinates. Update feature geometry adds length to line shapefiles, and area and perimeter to polygon shapefiles.

The exposure pathways for each source area are identified along with the chemicals of concern that apply to each source area. These data are added to the site conceptual model relational database.

The x-, and y-coordinates for the source areas are the centroids of the source area polygons. These can be converted to a source area center points shapefile by selecting the source area shapefile to be the active theme and then using the **Add Event Theme** command. The new theme is converted to a shapefile, as described above. This file is also identified by a version number and date.

The points of demonstration are identified based on the exposure pathways. Using the **Add New Theme** command, a point theme is constructed. The fields added to the data table include:

- POD_ID,
- Environmental Medium, and
- x- and y-coordinates.

The new theme is converted to a shapefile.

The transition points are identified based on the exposure pathways. Using the **Add New Theme** command, a point theme is constructed. The fields added to the data table include:

- LOC_ID,
- Description (e.g., NAPL to Air)
- Elevation
- X- and y-coordinates
- Source area LOC_ID
- Pathway ID, and
- Segment ID.

Since a number of the transition points will necessarily be located at the same x, y location, but have different z values, the transition points should be segregated into different shapefiles by environmental medium (e.g., soil vapor, groundwater, soil leachate). This, along with the elevation values, could enable a three-dimensional representation of the site conceptual model using the ArcView extension **3D-Analyst**. Each point in the transition points shapefiles, as well as the source areas, source area center points and points of demonstration would require a z-coordinate field. This was not implemented in the examples in this research.

The source areas with groundwater transport exposure pathways are identified. The starting and ending locations for each groundwater transport segment are identified. These could be LOC_IDs from source area center points, transition points and points of demonstration. A data table of the matching points is constructed. The table consists of the matching LOC_IDs and POD_IDs.

The matching points table is used in the gwtrans.ave script to build the line shapefile representing the groundwater transport segments. The listing of the gwtrans.ave script is included in section C.6. The CRWR Vector extension is used to add the length to the shapefile data table. The data table for the resulting groundwater transport line shapefile is shown in Table C.3.

Table C.3 - Example Data Table for the Groundwater Transport Shapefile

LOC_ID	SEG_ID	Length	PATH_ID
SA4	S1	685.012	P1
SA5	S1	911.961	P1
SA3	S1	654.879	P1
SA2	S1	176.576	P1

If the points of demonstration are identified for areas (e.g., surface water, indoor air) the indoor air, outdoor air and surface water mixing box shapefiles are constructed. The mixing box files are polygon shapefiles and are constructed in the same manner as the source area shapefile. The fields added to the data table include:

- LOC_ID,
- Environmental Medium, and
- Name.

The next section describes the database operations to create the exposure pathway records and to develop the model calculation queries.

C.2SITE CONCEPTUAL MODEL DATABASE

Along with the development of the spatial data for the site conceptual model, the exposure pathway records are added to the site conceptual model database. The Pathway and Pathway_Segments tables are populated first. Based on the specific records developed for the exposure pathways, the other database tables are populated, as necessary.

The modeling scenarios are identified based on the available facility information. As presented in Appendix B, once all of the exposure pathways are known the individual model input parameter values are developed. These are saved in the site conceptual model database tables.

The exposure pathway records and model input parameters can be entered into the data tables in the spreadsheet view, or input screens can be developed in Access to facilitate the data entry. In the research conducted here all of the records were entered in the spreadsheet views for the tables.

The shapefile data tables are linked into Access. The **File, Get External Data, Link Tables** commands are used to link the source area, points of demonstration and groundwater transport tables to the database. The link option is used, selecting dBase files, instead of import, so that the database records are updated when the GIS files are updated. If the shapefiles are revised and saved as a new revision (i.e., the shapefile name is changed), then the link must be updated. Table C.4 includes the linked data tables from the spatial database.

Table C.4 - Spatial Database Tables Linked to the Access Database

Table Name	**Description**
Gwtran1	Groundwater transport segments
Lubepod4	Points of demonstration
Srcarea2	Source areas

For the calculations for the transport pathways, the Excel spreadsheets are developed with a sheet for the calculation and a related sheet for the summarized results. Section C.4 includes a discussion of the Excel sheet construction. Table C.5 includes the linked tables from Excel.

Table C.5 - Spreadsheet Results Tables Linked to the Access Database

Table Name	**Description**
GWResults	Deterministic groundwater calculation results
NAPLPResults	Deterministic NAPL partitioning to groundwater calculation results
NAPLVResults	Deterministic NAPL volatilizing to soil vapor calculation results
SWMixResults	Deterministic surface water mixing calculation results
VtoIResults	Deterministic vapor to indoor air calculation results
VtoOResults	Deterministic vapor to outdoor air calculation results

Using the **File, Get External Data, Link Tables** commands, the results sheets from Excel are linked into the database. The results sheets must be linked into the database before the model input parameter queries, discussed in section C.3, can be implemented. This requires that one record of information be included in the Excel file when it is constructed. This acts as a placeholder. When the actual data are linked from the Access queries, the placeholder record is over-written. The results sheets are all of the same format. They have two columns (i.e., fields) of data, the UniqueID and the ending concentration.

C.3DATA HANDLING QUERIES

The data handling queries are constructed in Access. The following list summarizes the query construction. The specific queries including the tables used and the field positions are given in Table C.6. The data handling queries are constructed and re-executed in the same order. If the input values change, each of the queries is re-executed to collect the revised input parameters. If the input table names change (e.g., the shapefiles are updated), then the queries must be updated).

The queries must be cycled through the number of times based on the longest exposure pathway. That is, the exposure pathway with the largest number of segments determines how many times the queries are cycled through in order to develop the final results. All of the results sheets will have been over-written and the final results can be analyzed.

The following queries are used in the deterministic calculations:

- **UniqueIDBuilder**: Defines the pathway segment UniqueID based on the Pathway_Segments Table. The 10 key fields are concatenated. The UniqueIDBuilder query table is related to the Pathway_Segments Table by the 10 key fields.
- **ResultsUnion**: Collects all of the segment ending concentrations into one table using a select query.
- **SegmentStartConc**: Identifies any segments that start with a groundwater or vapor concentration instead of a NMF. The expression to define the starting concentration is:
    - Cstart: IIf([SourceVariables]![D(Cgw)]>0 And [Pathway_Segments]![Media]="Groundwater",[SourceVariables]![D(Cgw)],IIf([SourceVariables]![D(Cvap)]>0 And [Pathway_Segments]![Media]="Air",[SourceVariables]![D(Cvap)],Null)).

- **SegmentConc**: Checks for multiple segments before and separates single segment exposure pathways from the branched expoure pathways; SegmentConc chooses only the segments where the Multiple Segments Before field is "No."
- **SegementConcA**: Attributes each segment with Cstart, Medium and COC_ID.
- **SegmentConc2:** Selects segments where the Multiple Segments Before field is "Yes."
- **SegmentConc2A**: Selects all the segments with the same Segment_After ID, including medium, COC_ID and attenuation mechanism, attributes each segment with its ending concentration using the Results Union.
- **SegmentConc2B:** Sums all of the segments with the same Segment_After ID, grouped by medium and COC_ID, SegmentConc2B
- **SegmentConc2C**: Attributes each segment with Cstart, Medium and COC_ID.
- **BeginUnion**: Combine the two sets of starting concentrations through a select query.
- **SegmentBeginConc**: Checks all of the starting concentrations against solubility or saturated vapor concentrations.
  - Need a medium, COC_ID and representative NMF for solubility calculation. A value is chosen that is higher than

all values in use, say 0.1, so that the segment concentrations are used in preference to a simulated solubility-limited concentration. The selection statement is:

    - Cstart = If BeginUnion.Cstart > 0.1* Csol and Media = "Groundwater", then (0.1*Csol), If BeginUnion.Cstart > 0.1*Csatvap and Media = "Air", then (0.1*Csatvap), Else [BeginUnion.Cstart].

- **SegmentBeginUnion**: Combines the calculated starting concentrations with the specified concentrations. The SQL statement is:
  - SELECT * FROM SegmentBeginConc WHERE [Cstart] Is Not Null UNION SELECT * FROM SegmentStartConc WHERE [Cstart] Is Not Null.
- Collect the segment model parameters for each pathway segment calculation (see Table C.7, Table C.8, and Table C.9), using the **SegmentBeginUnion** as the starting source for concentration data. The UniqueID relates the **SegmentBeginUnion** query to the **UniqueIDBuilder** query table.

Table C.6 - Access Queries for Assigning the Starting Concentrations

Order Number	1	2	3	4	5
Query Name	UniqueIDBuilder	ResultsUnion	SegmentStartConc	SegmentConc	SegmentConcA
Tables	Pathway_Segments	GWResults, NAPLVResults, NAPLPResults, VtoIResults,VtoOResults SWMixResults	Pathway, Pathway_Segments, UniqueIDBuilder, SourceVariables	Pathway_Segments, UniqueIDBuilder, ResultsUnion	Pathway_Segments, UniqueIDBuilder, SegmentConc
Variable Position:	UniqueID	UniqueID	UniqueID	UniqueID	UniqueID
	Location_ID	Cend	Cstart	SegmentBeforeID	Cstart
	COC_ID			Cend	Media
	Pathway_ID			Multi Segments Before	COC_ID
	AboveGrd			COC_ID	
	ChemVar				
	Constants				
	MdlErrVar				
	SubSurVar				
	SWVar				
	Segment_ID				

Table C.6 - Access Queries for Assigning Starting Concentrations (continued)

Order Number	6	7	8	9	10
Query Name	SegmentConc2	SegmentConc2A	SegmentConc2B	SegmentConc2C	BeginUnion
Tables	Pathway_Segments, UniqueIDBuilder	Pathway_Segments, UniqueIDBuilder, SegmentConc2, ResultsUnion	SegmentConc2A	SegmentConc2B	SegmentConc2C, SegmentConcA
Variable Position:	UniqueID	UniqueID	Segment_After	UniqueID	UniqueID
	Multi Segments Before	Segment_After	Sum of Cend	Cstart	Cstart
		Media	Count of Cend	Media	Media
		Attenuation_Mech	Media	COC_ID	COC_ID
		COC_ID	COC_ID		
		Cend			

Table C.6 - Access Queries for Assigning Starting Concentrations (continued)

Order Number	11	12	13
Query Name	SegmentBeginConc	SegmentBeginUnion	ChemCheck
Tables	BeginUnion, ChemCheck	SegmentBeginConc, SegmentStartConc	COC_ID, ChemVariables, ChemConstants
Variable Position:	UniqueID	UniqueID	COC_ID
	Cstart	Cstart	Sol
			SatVap

Table C.7 - Access Query Tables for the NAPL Partitioning and Volatilization Segments

Query	**NAPLVolatilization**	**NAPLPartitioning**
Table Name	DNAPLVol	DNAPLPart
Excel Sheet	NAPLVolatilization	NAPLPartitioning
Variable Position:	UniqueID	UniqueID
	Sol	Sol
	Henrys	NMF
	NMF	

Table C.8 - Access Query Tables for the Vapor to Indoor and Outdoor Air Segments

Query	**VaportoIndoor**	**VaportoOutdoor**
Table Name	DVaportoIndoor	DVaportoOutdoor
Excel Sheet	VaportoIndoor	VaportoOutdoor
Variable Position:	UniqueID	UniqueID
	TPorVad	TPorVad
	WCVad	WCVad
	Ls	Lsout
	ER	Wspeed
	LB	Owidth
	WallT	MixHt
	AF	Henrys
	TPorWall	Dair
	WCWall	Dwat
	Henrys	EYO
	DAir	NYO
	Dwat	CD
	EYI	
	NYI	
	CD	

Table C.9 - Access Query Tables for the Groundwater and Surface Water Segments

Query	**GWTransport**	**SWMixing**
Table Name	DGWTransport	DSWMix
Excel Sheet	GWTransport	SWMixing
Variable Position:	UniqueID	UniqueID
	SatFOC	HydCond
	GWGrad	GWGrad
	HydCond	Width
	ParDen	Qcrk
	Tpor	Ccrk
	Koc	H
	Degrad	CB
	Width	EySW
	Depth	NySW
	Eygw	
	Nygw	
	CA	
	sDist	

C.4 EXCEL SHEET CONSTRUCTION

Each pathway segment that is to be calculated has a separate sheet in Excel. This is the sheet to which the query results table with the model parameters is connected. Following each calculation sheet there is a results sheet. On the results sheet there are cell references to the calculation sheet to obtain the values for the segment unique identifier and the ending concentration. This second sheet, the results sheet, is linked back to Access. The Excel sheets used in the deterministic calculations are listed in Table C.10.

Table C.10 - Excel Sheets for the Deterministic Calculations

Name	Description
Dnmfseg.xls	NAPL partitioning and volatilization
Dairseg.xls	Volatilization to indoor and outdoor air
Dgwseg.xls	Groundwater and surface water

Once the queries have been established and the Excel sheets have been created, the query results for each pathway segment calculation are linked to the corresponding Excel sheet for each pathway segment. The links are established in Excel using the **Data, Get External Data, New Database Query** commands. The connection to the Access database is established using an ODBC driver. The ODBC driver for the site conceptual model database is established in **Control Panel, ODBC, User DSN**. The Microsoft Access Driver is selected and the site conceptual model database is selected from the appropriate folder on the computer. Romanek *et. al.*, (1999) includes additional instructions on setting up an ODBC driver.

The query table is selected from the database in the dialog windows in Excel. Excel is using a sub-program called MSQuery to access the table records. The queries are not executed inside of Excel using MSQuery because of the number of initial queries that are needed to build the final parameter value query.

The Excel sheets can be set to refresh data upon opening them. Care should be taken with this function in the developmental stage, if the names of the queries are changed once the connection is made, or the Access file name is changed, additional steps will be needed to re-establish the connections. Also, there is an option in Excel to remove the linked data when the file is closed. This option should not be used, since removing the external data also removes the columns established for the data. This can cause errors in the cell references in the formulae.

In Excel when using the Refresh Data option, be sure that the cell or cells selected are within the external data range to be refreshed.

If a link exists between an Access query and Excel and the query changes, even if it is only the name of a field, the Too Few Parameters error results. This error is avoided by not refreshing the data after the query change. In the **Tools** use the **Edit Query** command and then re-establish the link to the query table. The **Edit Query** command can also be used the change the path or the ODBC driver and update the data range.

If the files are moved to a different computer, or to different folders on the same computer, the locations of the files can be identified in Excel using the **Edit, Links, Change Source** commands. In the dialog box, the database file can be located in the appropriate folder. This will update the path on all of the references. This is also true for the location of the Excel Add-In for the user-defined functions, described in section C.5.

C.5 CREATING USER-DEFINED FUNCTIONS IN EXCEL

In the Spatial Environmental Risk Assessment (SERA) methodology the environmental fate and transport calculations are implemented by row-wise calculations in spreadsheets. There are 53 partial derivatives, 13 transport algorithms and 2 variance equations to be calculated within the spreadsheets. In addition, all of the equations are called more than once in the spreadsheets. A method was needed to program the equations and then to be able to reuse the same equations without needing to re-enter cell references and formulae directly in the cells. An Excel Add-In Visual Basic Application (VBA) was created to address these computational needs. The functions programmed in visual basic are available within any spreadsheet as long as the Add-In file is available in the user's profile on the specific computer. Once the Add-In is available to a spreadsheet the user-defined functions are implemented just as any of the standard Excel functions would be. An added benefit of the VBA program is that error checking in the original formulae is easier than in calculations written directly in cells and the cell calculations are easier to check since the value in the cell uses an equation name.

The following steps are used to create User-Defined Functions.

1. In Excel, without a spreadsheet file open, choose **Tools, Macro**, and then **Visual Basic Editor**. This will open the Visual basic GUI.
2. In the project window select VBAProject (Book 1).
3. Choose **Insert, Module** (i.e., a set of visual basic code).
4. Choose **Insert Procedure** (i.e., a single group of commands).
5. Assign a name for the function (e.g., ExArea) and choose Function with public scope and all local variables not static.
6. In the Module window Public Function ExArea() will be displayed.
7. Write the variable names for all parameters needed for the in the Function ExArea, e.g., ExArea(Length, Width).

8. On the next line under the Function declaration write the mathematical relationship for ExArea, e.g., ExArea = Length*Width.
9. Next choose **Debug, Compile VBAProject**. If there are any errors in the formula an error dialog box will be generated.
10. To save the Function for use in a spreadsheet choose **File, Save As.**
11. Use **Save As Type = Microsoft Excel Add-In** (*.xla).
12. For Microsoft Windows NT machines save the file in Winnt/Profiles/UserName/ApplicationData/Microsoft/AddIns

The VBAProject will now have the saved name. There can be multiple Modules in one project. In the SERAv1.xla there are six modules:

- Deriv1
- Deriv2
- Diffusion
- GWFunctions
- Transport
- Variance

These correspond to the code listing sections given in Appendix A.

The modules are named in the properties window. A name is a property of an object in VBA. Module 1 is replaced with a name that helps to identify the functions in that module.

When finished adding and editing functions, choose **Save** and then **Exit and Return to Excel**.

In Excel choose **Tools, Add-Ins**. A dialog box will be displayed. The new Add-In file will either be listed in the available Add-Ins or use the **Browse** command to locate the file.

The functions are used in Excel just as any of the pre-defined functions. They will be listed in the User-Defined category. In the spreadsheet, the function help box can also be used with the User-Defined functions.

C.6 SCRIPT FOR LATERAL TRANSPORT SEGMENTS

The **gwtrans.ave** script can be used for construction of any lateral transport exposure pathway segment (e.g., one dimensional air transport), but was originally written for the groundwater transport exposure pathway.

```
' Script: gwtrans.ave
' Author: Francisco Olivera, Lesley Hay Wilson
' Center for Research in Water Resources (CRWR)
' University of Texas at Austin
' October 30, 1999

' Purpose: Create a polyline shape-file of lateral transport segments from a table of
user specified
'          starting and ending points.
' Input:  An attribute table for a point shape-file of the starting points
'          (generally this is the shape-file of source area center points)
'         An attribute table for a point shape-file of the ending points
'          (generally this is the shape-file of points of demonstration)
'         A dbf table for the connections between source areas and PODs
'            The starting points attribute table should have:
'          (1) "LOC_ID": starting point identifier
'          (2) "x_coord": x-coordinate of the starting point
'          (3) "y_coord": y-coordinate of the starting point
'
'            The ending points attribute table should have:
'          (1) "POD_ID": point of demonstration identifier
'          (2) "x_coord": x-coordinate of the end point
```

```
'	(3) "y_coord": y-coordinate of the end point
'
'	The connection table should have:
'	(1)  "LOC_ID": starting point identifier
'	(2)  "POD_ID": point of demonstration identifier

' Output: A polyline shape-file of line segments. The table associated with this
shape-file
'	stores the polyline shapes, id's and the XY values for the starting and
ending points.

' ***********************************************************

' (1) INPUT STATEMENTS

LableList = {"Name of the View","Name of the Start Table","Name of the End
Table", "Name of Connection Table", "Name of the output ShapeFile"}
DefaultList = {"View1","Attributes of Centerpts.shp", "Attributes of
Lubepod3.shp","tablept.dbf",""}
InputList = msgbox.multiInput("Enter the input
names:","INPUT",LableList,DefaultList)
ViewName = InputList.Get(0)
TableStartName = InputList.Get(1)
TableEndName = InputList.Get(2)
TableConn = InputList.Get(3)
TranShp = InputList.Get(4)

' (2) INITIAL SETTINGS

TheProject = av.GetProject

TheView = TheProject.FindDoc(ViewName)
If (TheView = nil) then
  msgbox.info(ViewName ++ "not found", "INPUT ERROR !!!")
  exit
End

TheStartTable = TheProject.FindDoc(TableStartName)
If (TheStartTable = nil) then
  msgbox.info(TableStartName ++ "not found", "INPUT ERROR !!!")
  exit
End
```

```
TheEndTable = TheProject.FindDoc(TableEndName)
If (TheEndTable = nil) then
  msgbox.info(TableEndName ++ "not found", "INPUT ERROR !!!")
  exit
End

TheConnTable = TheProject.FindDoc(TableConn)
If (TheConnTable = nil) then
  msgbox.info(TableConn ++ "not found", "INPUT ERROR !!!")
  exit
End

TheSFTab = TheStartTable.GetVTab
TheEFTab = TheEndTable.GetVTab
TheConnVTAb = TheConnTable.GetVTab

' (3) INITIALIZING FIELDS OF THE INPUT TABLES

ConnStartId = TheConnVTab.FindField("LOC_ID")
ConnEndId   = TheConnVTab.FindField("POD_ID")

StartId = TheSFTab.FindField("LOC_ID")
EndId   = TheEFtab.FindField("POD_ID")
xFieldStart = TheSFTab.FindField("x_coord")
yFieldStart = TheSFTab.FindField("y_coord")
xFieldEnd = TheEFTab.FindField("x_coord")
yFieldEnd = TheEFTab.FindField("y_coord")

' (4) ADDING FIELDS TO THE OUTPUT TABLES

TheConnVTab.SetEditable(TRUE)

x1Field = TheConnVTab.FindField("x1")
If(x1Field=NIL) then
   x1Field = Field.Make("x1", #Field_Float, 16, 4)
   TheConnVTab.AddFields({x1Field})
End

y1Field = TheConnVTab.FindField("y1")
If(y1Field=NIL) then
   y1Field = Field.Make("y1", #Field_Float, 16, 4)
```

```
  TheConnVTab.AddFields({y1Field})
End

x2Field = TheConnVTab.FindField("x2")
If(x2Field=NIL) then
  x2Field = Field.Make("x2", #Field_Float, 16, 4)
  TheConnVTab.AddFields({x2Field})
End

y2Field = TheConnVTab.FindField("y2")
If(y2Field=NIL) then
  y2Field = Field.Make("y2", #Field_Float, 16, 4)
  TheConnVTab.AddFields({y2Field})
End

IDField = TheConnVTab.FindField("ID")
If(IDField=NIL) then
  IDField = Field.Make("ID", #FIELD_CHAR, 16, 0)
  TheConnVTab.AddFields({IDField})
End

' (5) CREATING AN EMPTY POLYGON SHAPE-FILE

aFileName = av.GetProject.GetWorkDir.MakeTmp(TranShp, "")
TranFTab = FTab.MakeNew (aFileName, POLYLINE)
TranFTab.SetEditable(TRUE)
TheIDField = Field.Make("ID", #Field_CHAR, 16, 0)
TranFTab.AddFields({TheIDField})
TheShapeField=TranFTab.FindField("Shape")
TheShapeField.SetVisible(TRUE)
TheIDField=TranFTab.FindField("ID")

TranFTab.SetEditable(TRUE)

x1F = TranFTab.FindField("x1")
If(x1F=NIL) then
  x1F = Field.Make("x1", #Field_Float, 16, 4)
  TranFTab.AddFields({x1F})
End

y1F = TranFTab.FindField("y1")
If(y1F=NIL) then
```

```
  y1F = Field.Make("y1", #Field_Float, 16, 4)
  TranFTab.AddFields({y1F})
End

x2F = TranFTab.FindField("x2")
If(x2F=NIL) then
  x2F = Field.Make("x2", #Field_Float, 16, 4)
  TranFTab.AddFields({x2F})
End

y2F = TranFTab.FindField("y2")
If(y2F=NIL) then
  y2F = Field.Make("y2", #Field_Float, 16, 4)
  TranFTab.AddFields({y2F})
End

IDF = TranFTab.FindField("ID")
If(IDF=NIL) then
  IDF = Field.Make("ID", #Field_CHAR, 16, 0)
  TranFTab.AddFields({IDF})
End

'  (6)  MATCHING THE STARTING AND ENDING LOCATIONS USING
THE CONNECTION TABLE

For each ConnRec in TheConnVTab

  StartPoint = TheConnVTab.ReturnValue(ConnStartId,ConnRec)
  EndPoint   = TheConnVTab.ReturnValue(ConnEndId,ConnRec)

  For each SRec in TheSFTab
    xStart = TheSFtab.ReturnValue(xFieldStart,SRec)
    yStart = TheSFTab.ReturnValue(yFieldStart,SRec)
    StartIdSFTab = TheSFTab.ReturnValue(StartID,SRec)

    if(StartIdSFTab = StartPoint) then

      x1 = xStart
      y1 = yStart
      ID = StartIdSFTab
   end
  End
```

```
  For each ERec in TheEFTab
    xEnd = TheEFtab.ReturnValue(xFieldEnd,ERec)
    yEnd = TheEFTab.ReturnValue(yFieldEnd,ERec)
    EndIdEFTab = TheEFTab.ReturnValue(EndID,ERec)

    if(EndIdEFTab = EndPoint) then
      x2 = xEnd
      y2 = yEnd
    end

 End

  TheConnVTab.SetValue(x1Field,ConnRec,x1)
  TheConnVTab.SetValue(y1Field,ConnRec,y1)
  TheConnVTab.SetValue(x2Field,ConnRec,x2)
  TheConnVTab.SetValue(y2Field,ConnRec,y2)
  TheConnVTab.SetValue(IDField,ConnRec,ID)


' (7) CREATING A LINE AND ADDING IT TO THE SHAPE FILE

  Segment = Line.Make(x1@y1, x2@y2)
  TranFTab.AddRecord
  TranFTab.SetValue(TheShapeField,ConnRec, Segment.AsPolyLine)

  TranFTab.SetValue(x1F,ConnRec,x1)
  TranFTab.SetValue(y1F,ConnRec,y1)
  TranFTab.SetValue(x2F,ConnRec,x2)
  TranFTab.SetValue(y2F,ConnRec,y2)
  TranFTab.SetValue(IDF,ConnRec,ID)

End

' (8) CREATING A THEME FOR THE SHAPE-FILE AND DISPLAYING IT

TranThm = FTheme.Make(TranFTab)
TranThm.SetName(TranShp)
TheView.AddTheme(TranThm)
TranThm.SetVisible(True)

' (9) FINAL STATEMENTS
```

```
TheSFTab.SetEditable(FALSE)
TheEFTab.SetEditable(FALSE)
TranFTab.SetEditable(FALSE)
TheConnVTab.SetEditable(FALSE)
```

Appendix D : Computer Implementation and Procedures for Probabilistic Calculations

This Appendix presents the computer procedures used in the application of the Spatial Environmental Risk Assessment (SERA) methodology. Specifically the methods for developing the site conceptual model database records and the spreadsheets for the probabilistic calculations.

D.1 SPATIAL DATABASE DEVELOPMENT

The exposure pathway groups for the probabilistic calculations are based on the same exposure pathway GIS layers that are described in Appendix C. Additional points of demonstration are likely to be needed and can be added to the points of demonstration shapefile. If additional source areas or exposure pathways are to be included in the probabilistic calculations, then they are added to the appropriate shapefiles.

D.2SITE CONCEPTUAL MODEL DATABASE

The exposure pathway records are added to the site conceptual model database for the exposure pathway groups. The Pathway and Pathway_Segments tables are populated first. The groups are identified and the exposure pathways are assigned positions in the groups. The first three positions are for NAPL source areas, the second three are for concentration-based source areas. For the examples, groups G1 to G5 have five NAPL source areas. Based on the NMF data for the last two source areas, mean and variance values were calculated for the groundwater concentration values (Cgw) and recorded in the database. The groups are based on a single chemical of concern, an exposure pathway type (e.g., volatilization to indoor air) and each group is associated with only one point of demonstration. For the probabilistic calculations the model correlation parameter tables must be populated. In addition, for any groups that do not have 6 exposure pathways in the group, the Group Support Tables must be populated.

The Group Identifiers table describes each of the groups. The modeling scenarios for any variables that apply only to the group are identified in the Group Identifiers table. These are variables like the Test NMF for the solubility check and the background concentration in the surface water, which are both part of only the combined group calculations and not part of the individual exposure pathway calculations within the groups.

The Group Support tables are used in the input parameter variable gathering to act as placeholders in the groups that do not have 6 exposure pathways. The tables are:

- GroupSupportIn
- GroupSupportOut
- GroupSupportSW
- GroupSupportSWVar

The key fields are the Group_ID and the Group Order No. The tables have all the same fields as key query tables, but the values are all zero. In section D.3 the use of the Group Support Tables in the queries is presented.

As in the deterministic case, for the calculations, the Excel spreadsheets have sheets with the summarized results that are linked back to database. Table

D.1 includes the spreadsheet results tables that are linked to the site conceptual model database.

Table D.1. - Spreadsheet Results Tables Linked to the Access Database

Table Name	Description
PSWResults	Probabilistic groundwater and surface water calculation results
PSWVarRes	Probabilistic groundwater and surface water variance results
VapInResults	Probabilistic indoor air calculation results
VapOutResults	Probabilistic outdoor air calculation results
Var(CE)Results	Probabilistic indoor air variance results
Var(CF)Results	Probabilistic outdoor air variance results
Var(CG)Results	Probabilistic indoor air variance results for grouped exposure pathways
Var(CH)Results	Probabilistic indoor air variance results for grouped exposure pathways

D.3DATA HANDLING QUERIES

The data handling queries are constructed in Access. The following list summarizes the query construction. There is a set of queries for each of the three types of group calculations. For volatilization to indoor air and volatilization to outdoor air, there are two sets of queries to link two input value data tables to Excel. For groundwater to surface water mixing there are three sets of queries to link three input value data tables to Excel. The data handling queries are constructed and re-executed in the same order. If the input values change, each of the queries is re-executed to collect the revised input parameters.

D.3.1 Queries for Volatilization to Outdoor Air

The following queries are used for the calculation of outdoor air concentration means and variances, the specific tables used and field positions are given in Table D.2:

- **PVapOut:** Collects all of the input data for the individual exposure pathway calculations. The mean values for all of the input parameters are collected, as are the variance values and correlation coefficients.
- **PVOStart:** Brings together the exposure pathways identified by group, flag (i.e., F-NMF, F-Cvap) and group order number.
- **PVOVarQ1:** Assembles the identifiers from PVOStart and the results for the means and the variances from the results sheets.
- **PVOVarUnion:** Joins the results in PVOVarQ1 with the GroupSupportOut table, which provides placeholder records where there are less than 6 exposure pathways in a group.
- **PVOVarQ2:** Selects the mean and variance data for Group Order No. = 1.
- **PVOVarQ3:** Selects the mean and variance data for Group Order No. = 2 and adds that data to the data for the prior Group Order numbers.

- **PVOVarQ4:** Selects the mean and variance data for Group Order No. = 3 and adds that data to the data for the prior Group Order numbers.
- **PVOVarQ5:** Selects the mean and variance data for Group Order No. = 4 and adds that data to the data for the prior Group Order numbers.
- **PVOVarQ6:** Selects the mean and variance data for Group Order No. = 5 and adds that data to the data for the prior Group Order numbers.
- **PVOVarQ7:** Selects the mean and variance data for Group Order No. = 6 and adds that data to the data for the prior Group Order numbers.
- **PVOGroup1:** Selects the TestNMF and the model scenario identifiers for each group for chemical variables and above ground variables.
- **PVOGroup2:** Selects the mean value for the Henrys constant and the solubility to use in the saturated vapor check at the end of the vapor to outdoor air calculation.
- **PVOVarQ8:** Adds the parameters from PVOGroup2 to the results in PVOVarQ7. This query is linked to the Excel sheet for the group mean and variance calculation.

Table D.2 - Access Queries for Collecting Input Data for the Outdoor Air Calculations

OrderNumber	1	2	3	4
Query Name	PVapOut	PVOVarStart	PVOVarQ1	PVOVarUnion
Tables	Pathway, Pathway_Segments, UniqueIDBuilder, SourceVariables, SubSurVariables, MdlErrorVariables, scrarea2, SourceCorrelations, ChemConstants, Physical Constants, Chem Variables, AboveGrd Variables, SubSurCorrelations	Pathway, Pathway_Segments, UniqueIDBuilder, SourceVariables	Var(CF)Results, PVOVarStart, VapOutResults	PVOVarQ1, GroupSupportOut
Variable Position:	UniqueID	UniqueID	UniqueID	UniqueID
	F-NMF	Group_ID	Group_ID	Group_ID
	F-Cvap	GroupOrderNo	GroupOrderNo	GroupOrderNo
	Sol	F-NMF	F-NMF	F-NMF
	DAir	F-Cvap	F-Cvap	F-Cvap
	Dliq	Attenuation_Mechanism	CF	CF
	MixHgt		Var(CF)	Var(CF)
	E(Henrys)			
	E(PorVad)			
	E(WCVad)			
	E(WSpeed)			
	E(Lsout)			
	E(Owidth)			
	EyO			
	NyO			
	E(NMF)			
	E(Cvap)			
	Var(Henrys)			
	Var(PoVad)			
	Var(WCVad)			
	Var(WSpeed)			

OrderNumber	1	2	3	4
	Var(Lsout)			
	Var(OWidth)			
	Var(EyO)			
	Var(NyO)			
	Var(NMF)			
	Var(Cvap)			
	R-PorVad/WCVad			
	R-Lsout/OWidth			

Table D.2 - Access Queries for Collecting Input Data for the Indoor Air Calculations (continued)

OrderNumber	5	6	7	8	9
Query Name	PVOVarQ2	PVOVarQ3	PVOVarQ4	PVOVarQ5	PVOVarQ6
Tables	PVOVarUnion	PVOVarUnion, PVOVarQ2	PVOVarUnion, PVOVarQ3	PVOVarUnion, PVOVarQ4	PVOVarUnion, PVOVarQ5
Variable Position:	Group_ID	Group_ID	Group_ID	Group_ID	Group_ID
	F-NMF1	F-NMF1	F-NMF1	F-NMF1	F-NMF1
	E(CF1)	F-NMF2	F-NMF2	F-NMF2	F-NMF2
	Var(CF1)	E(CF1)	F-NMF3	F-NMF3	F-NMF3
		E(CF2)	E(CF1)	F-Cvap1	F-Cvap1
		Var(CF1)	E(CF2)	E(CF1)	F-Cvap2
		Var(CF2)	E(CF3)	E(CF2)	E(CF1)
			Var(CF1)	E(CF3)	E(CF2)
			Var(CF2)	E(CF4)	E(CF3)
			Var(CF3)	Var(CF1)	E(CF4)
				Var(CF2)	E(CF5)
				Var(CF3)	Var(CF1)
				Var(CF4)	Var(CF2)
					Var(CF3)
					Var(CF4)
					Var(CF5)

Table D.2 - Access Queries for Collecting Input Data for the Indoor Air Calculations (continued)

OrderNumber	10	11	12	13
Query Name	PVOVarQ7	PVOGroup1	PVOGroup2	PVOVarQ8
Tables	PVOVarUnion, PVOVarQ6	GroupIdentifiers	PVOGroup1, ChemConstants, ChemVar	PVOVarQ7, PVOGroup2
Variable Position:	Group_ID	Group_ID	Group_ID	Group_ID
	F-NMF1	ChemVar	COC_ID	F-NMF1
	F-NMF2	COC_ID	TestNMF	F-NMF2
	F-NMF3	AboveGrd	E(Henrys)	F-NMF3
	F-Cvap1	TestNMF	Sol	F-Cvap1
	F-Cvap2			F-Cvap2
	F-Cvap3			F-Cvap3
	E(CF1)			TestNMF
	E(CF2)			E(Henrys)
	E(CF3)			Sol
	E(CF4)			E(CF1)
	E(CF5)			E(CF2)
	E(CF6)			E(CF3)
	Var(CF1)			E(CF4)
	Var(CF2)			E(CF5)
	Var(CF3)			E(CF6)
	Var(CF4)			V(CF1)
	Var(CF5)			V(CF2)
	Var(CF6)			V(CF3)
				V(CF4)
				V(CF5)
				V(CF6)

D.3.2 Queries for Volatilization to Indoor Air

The following queries are used for the calculation of indoor air concentration means and variances, the specific tables used and field positions are given in Table D.3:

- **PVapIn:** Collects all of the input data for the individual exposure pathway calculations. The mean values for all of the input parameters are collected, as are the variance values and correlation coefficients.
- **PVIStart:** Brings together the exposure pathways identified by group, flag (i.e., F-NMF, F-Cvap) and group order number.
- **PVIVarQ1:** Assembles the identifiers from PVIStart and the results for the means and the variances from the results sheets.
- **PVIVarUnion:** Joins the results in PVIVarQ1 with the GroupSupportIn table, which provides placeholder records where there are less than 6 exposure pathways in a group.
- **PVIVarQ2:** Selects the mean and variance data for Group Order No. = 1.
- **PVIVarQ3:** Selects the mean and variance data for Group Order No. = 2 and adds that data to the data for the prior Group Order numbers.
- **PVIVarQ4:** Selects the mean and variance data for Group Order No. = 3 and adds that data to the data for the prior Group Order numbers.
- **PVIVarQ5:** Selects the mean and variance data for Group Order No. = 4 and adds that data to the data for the prior Group Order numbers.

- **PVIVarQ6:** Selects the mean and variance data for Group Order No. = 5 and adds that data to the data for the prior Group Order numbers.
- **PVIVarQ7:** Selects the mean and variance data for Group Order No. = 6 and adds that data to the data for the prior Group Order numbers.
- **PVIGroup1:** Selects the TestNMF and the model scenario identifiers for each group for chemical variables and above ground variables.
- **PVIGroup2:** Selects the mean value for the Henrys constant and the solubility to use in the saturated vapor check at the end of the vapor to indoor air calculation.
- **PVIVarQ8:** Adds the parameters from PVIGroup2 to the results in PVIVarQ7. This query is linked to the Excel sheet for the group mean and variance calculation.

Table D.3 - Access Queries for Collecting Input Data for the Indoor Air Calculations

OrderNumber	1	2	3	4
Query Name	PVapIn	PVIVarStart	PVIVarQ1	PVIVarUnion
Tables	Pathway, Pathway_Segments, UniqueIDBuilder, SourceVariables, SubSurVariables, MdlErrorVariables, scrarea2, ChemConstants, Physical Constants, Chem Variables, AboveGrd Variables, AboveGrd Correlations, SubSurCorrelations	Pathway, Pathway_Segments UniqueIDBuilder, SourceVariables	Var(CE)Results, PVIVarStart, VapInResults	PVIVarQ1, GroupSupportIn
Variable Position:	UniqueID	UniqueID	UniqueID	UniqueID
	F-NMF	Group_ID	Group_ID	Group_ID
	F-Cvap	GroupOrderNo	GroupOrderNo	GroupOrderNo
	Sol	F-NMF	F-NMF	F-NMF
	DAir	F-Cvap	F-Cvap	F-Cvap
	Dliq	Atten_Mechanism	CE	CE
	ER		Var(CE)	Var(CE)
	LB			
	WallT			
	E(Henrys)			
	E(AF)			
	EYI			
	E(Ls)			
	NYI			
	E(PorVad)			
	E(PorWall)			
	E(WCVad)			
	E(WCWall)			
	E(NMF)			
	E(Cvap)			

OrderNumber	1	2	3	4
	Var(Henrys)			
	Var(AF)			
	Var(EYI)			
	Var(Ls)			
	Var(NYI)			
	Var(PorVad)			
	Var(PorWall)			
	Var(WCVad)			
	Var(WCWall)			
	Var(NMF)			
	Var(Cvap)			
	R-PorVad/WCVad			
	R-PorWall/WCWall			
	R-AF/WCWall			
	R-AF/TPorWall			

Table D.3. Access Queries for Collecting Input Data for the Indoor Air Calculations (Continued)

OrderNumber	5	6	7	8	9
Query Name	PVIVarQ2	PVIVarQ3	PVIVarQ4	PVIVarQ5	PVIVarQ6
Tables	PVIVarUnion	PVIVarUnion, PVIVarQ2	PVIVarUnion, PVIVarQ3	PVIVarUnion, PVIVarQ4	PVIVarUnion, PVIVarQ5
Variable Position:	Group_ID	Group_ID	Group_ID	Group_ID	Group_ID
	F-NMF1	F-NMF1	F-NMF1	F-NMF1	F-NMF1
	E(CE1)	F-NMF2	F-NMF2	F-NMF2	F-NMF2
	Var(CE1)	E(CE1)	F-NMF3	F-NMF3	F-NMF3
		E(CE2)	E(CE1)	F-Cvap1	F-Cvap1
		Var(CE1)	E(CE2)	E(CE1)	F-Cvap2
		Var(CE2)	E(CE3)	E(CE2)	E(CE1)
			Var(CE1)	E(CE3)	E(CE2)
			Var(CE2)	E(CE4)	E(CE3)
			Var(CE3)	Var(CE1)	E(CE4)
				Var(CE2)	E(CE5)
				Var(CE3)	Var(CE1)
				Var(CE4)	Var(CE2)
					Var(CE3)
					Var(CE4)

OrderNumber	5	6	7	8	9
					Var(CE5)

Table D.3. Access Queries for Collecting Input Data for the Indoor Air Calculations (Continued)

OrderNumber	10	11	12	13
Query Name	PVIVarQ7	PVIGroup1	PVIGroup2	PVIVarQ8
Tables	PVIVarUnion, PVIVarQ6	GroupIdentifiers	PVIGroup1, ChemConstants, ChemVar	PVIVarQ7, PVIGroup2
Variable Position:	Group_ID	Group_ID	Group_ID	Group_ID
	F-NMF1	ChemVar	COC_ID	F-NMF1
	F-NMF2	COC_ID	TestNMF	F-NMF2
	F-NMF3	AboveGrd	E(Henrys)	F-NMF3
	F-Cvap1	TestNMF	Sol	F-Cvap1
	F-Cvap2			F-Cvap2
	F-Cvap3			F-Cvap3
	E(CE1)			TestNMF
	E(CE2)			E(Henrys)
	E(CE3)			Sol
	E(CE4)			E(CE1)
	E(CE5)			E(CE2)
	E(CE6)			E(CE3)
	Var(CE1)			E(CE4)
	Var(CE2)			E(CE5)
	Var(CE3)			E(CE6)
	Var(CE4)			V(CE1)
	Var(CE5)			V(CE2)
	Var(CE6)			V(CE3)
				V(CE4)
				V(CE5)
				V(CE6)

D.3.3 Queries for Groundwater to Surface Water Mixing

The following queries are used for the calculation of surface water concentration means and variances, the specific tables used and field positions are given in Table D.4:

- **PathTest:** Collects the input data to calculate the groundwater mean concentrations, the surface water mean concentrations (for each exposure pathway independently) and to calculate the partial derivatives.
- **SWMixStart:** Brings together the exposure pathways identified by group, flag (i.e., F-NMF, F-Cvap) and group order number.
- **SWMixStart2:** Collects the variance input data for the groundwater segments.
- **SWMixStart3:** Connects the exposure pathway identifiers from SWMixStart with the variance data from SWMixStart2.
- **SWMixUnion:** Joins the input data in SWMixStart3 with the GroupSupportSW table, which provides placeholder records where there are less than 6 exposure pathways in a group.
- **SWMixQ1:** Selects the variance data for Group Order No. = 1.
- **SWMixQ2:** Selects the variance data for Group Order No. = 2.
- **SWMixQ3:** Selects the variance data for Group Order No. = 3.
- **SWMixQ4:** Selects the variance data for Group Order No. = 4.
- **SWMixQ5:** Selects the variance data for Group Order No. = 5.

- **SWMixQ5A:** Selects the variance data for Group Order No. = 6.
- **SWGroup1:** Selects the TestNMF and the model scenario identifiers for each group for chemical variables and surface water variables.
- **SWGroup2:** Selects the values for the group surface water variables, the correlation coefficients, and the solubility check at the end of the surface water calculation.
- **SWMixQ6:** Assembles input variables for the surface water calculations. This query is connected to the Excel sheet SWVarInput.
- **SWVarStart:** Brings together the exposure pathways identified by group, flag (i.e., F-NMF, F-Cvap) and group order number.
- **SWVarQ1:** Using the groups identified by SWVarStart, the results for the mean surface water concentrations (for individual exposure pathways) and the partial derivatives are connected.
- **SWVarUnion:** Joins the input data in SWVarQ1 with the GroupSupportSWVar table, which provides placeholder records where there are less than 6 exposure pathways in a group.
- **SWVarQ2:** Selects the mean data and the partial derivatives for Group Order No. = 1.
- **SWVarQ3:** Selects the mean data and the partial derivatives for Group Order No. = 2.

- **SWVarQ4:** Selects the mean data and the partial derivatives for Group Order No. = 3.
- **SWVarQ5:** Selects the mean data and the partial derivatives for Group Order No. = 4.
- **SWVarQ6:** Selects the mean data and the partial derivatives for Group Order No. = 5.
- **SWVarQ7:** Selects the mean data and the partial derivatives for Group Order No. = 6. This query table is linked to the Excel sheet SWVarInput2.

Table D.4 - Access Queries for the Groundwater to Surface Water Calculations

OrderNumber	1	2	3	4
Query Name	PathTest	SWMixStart	SWMixStart2	SWMixStart3
Tables	Pathway, Pathway_Segments, UniqueIDBuilder, SourceVariables, SubSurVariables, MdlErrorVariables, ChemConstants, PhysicalConstants, ChemVariables, Srcarea2, gwtran1, SWVariables	Pathway, Pathway_Segments, UniqueIDBuilder	Pathway, Pathway_Segments, UniqueIDBuilder, SourceVariables, SubSurVariables, MdlErrorVariables ChemVariables, ChemConstants, SWVariables	SWMixStart, SWMixStart2
Variable Position:	UniqueID	UniqueID	UniqueID	UniqueID
	F-NMF	Group_ID	F-NMF	Group_ID
	F-Cgw	GroupOrderNo	F-Cgw	GroupOrderNo
	ParDen	Attenuation_Mech	Var(NMF)	F-NMF
	Sol		Var(Cgw)	F-Cgw
	E(Degrad)		Var(Degrad)	Var(NMF)
	E(Depth)		Var(Koc)	Var(Cgw)
	E(GWGrad)		Var(H)	Var(Degrad)
	E(HydCond)		Var(Qcrk)	Var(Koc)
	E(Koc)		Var(NySW)	Var(H)
	E(NMF)		Var(GWGrad)	Var(Qcrk)
	E(SatFoc)		V(HydCond)	Var(NySW)
	E(sDist)		V(SatFoc)	Var(GWGrad)
	E(TPor)		V(TPor)	V(HydCond)
	E(Width)		V(EyGW)	V(SatFoc)
	E(H)		V(NyGW)	V(TPor)
	E(NySW)		V(Width)	V(EyGW)
	E(EySW)		V(Depth)	V(NyGW)
	E(Qcrk)		V(sDist)	V(Width)
	E(EyGW)		Atten_Mech	V(Depth)
	E(NyGW)			V(sDist)
	E(Cgw)			

Table D.4 - Access Queries for the Groundwater to Surface Water Calculations (Continued)

OrderNumber	5	6	7
Query Name	SWMixUnion	SWMixQ1	SWMixQ2
Tables	SWMixStart3, GroupSupportSW	SWMixUnion	SWMixUnion, SWMixQ1
Variable Position:	UniqueID	Group_ID	Group_ID
	Group_ID	F-NMF1	GroupOrderNo
	GroupOrderNo	Var(Degrad1)	F-NMF1
	F-NMF	Var(Koc1)	F-NMF2
	F-Cgw	Var(H1)	Var(Degrad1)
	Var(NMF)	Var(Qcrk1)	Var(Degrad2)
	Var(Cgw)	Var(NySW1)	Var(Koc1)
	Var(Degrad)	Var(NMF1)	Var(Koc2)
	Var(Koc)	V(GWGrad1)	Var(H1)
	Var(H)	V(HydCond1)	Var(H2)
	Var(Qcrk)	V(SatFoc1)	Var(Qcrk1)
	Var(NySW)	V(TPor1)	Var(Qcrk2)
	Var(GWGrad)	V(EyGW1)	Var(NySW1)
	V(HydCond)	V(NyGW1)	Var(NySW2)
	V(SatFoc)	V(Width1)	V(NMF1)
	V(TPor)	V(Depth1)	V(NMF2)
	V(EyGW)	V(sDist1)	V(GWGrad1)
	V(NyGW)		V(GWGrad2)
	V(Width)		V(HydCond1)
	V(Depth)		V(HydCond2)
	V(sDist)		V(SatFoc1)
			V(SatFoc2)
			V(TPor1)
			V(TPor2)
			V(EyGW1)
			V(EyGW2)
			V(NyGW1)
			V(NyGW2)
			V(Width1)
			V(Width2)
			V(Depth1)
			V(Depth2)
			V(sDist1)

OrderNumber	5	6	7
			V(sDist2)

Table D.4 - Access Queries for the Groundwater to Surface Water Calculations (Continued)

OrderNumber	8	9	10
Query Name	SWMixQ3	SWMixQ4	SWMixQ5
Tables	SWMixUnion, SWMixQ2	SWMixUnion, SWMixQ3	SWMixUnion, SWMixQ3
Variable Position:	Group_ID	Group_ID	Group_ID
	GroupOrderNo	GroupOrderNo	GroupOrderNo
	F-NMF1	F-NMF1	F-NMF1
	F-NMF2	F-NMF2	F-NMF2
	F-NMF3	F-NMF3	F-NMF3
	Var(Degrad1)	F-Cgw1	F-Cgw1
	Var(Degrad2)	Var(Degrad1)	F-Cgw2
	Var(Degrad3)	Var(Degrad2)	Var(Degrad1)
	Var(Koc1)	Var(Degrad3)	Var(Degrad2)
	Var(Koc2)	Var(Degrad4)	Var(Degrad3)
	Var(Koc3)	Var(Koc1)	Var(Degrad4)
	Var(H1)	Var(Koc2)	Var(Degrad5)
	Var(H2)	Var(Koc3)	Var(Koc1)
	Var(H3)	Var(Koc4)	Var(Koc2)
	Var(Qcrk1)	Var(H1)	Var(Koc3)
	Var(Qcrk2)	Var(H2)	Var(Koc4)
	Var(Qcrk3)	Var(H3)	Var(Koc5)
	Var(NySW1)	Var(H4)	Var(H1)
	Var(NySW2)	Var(Qcrk1)	Var(H2)
	Var(NySW3)	Var(Qcrk2)	Var(H3)
	V(NMF1)	Var(Qcrk3)	Var(H4)
	V(NMF2)	Var(Qcrk4)	Var(H5)
	V(NMF3)	Var(NySW1)	Var(Qcrk1)
	V(GWGrad1)	Var(NySW2)	Var(Qcrk2)
	V(GWGrad2)	Var(NySW3)	Var(Qcrk3)
	V(GWGrad3)	Var(NySW4)	Var(Qcrk4)
	V(HydCond1)	V(NMF1)	Var(Qcrk5)
	V(HydCond2)	V(NMF2)	Var(NySW1)
	V(HydCond3)	V(NMF3)	Var(NySW2)
	V(SatFoc1)	V(Cgw1)	Var(NySW3)
	V(SatFoc2)	V(GWGrad1)	Var(NySW4)

OrderNumber	8	9	10
	V(SatFoc3)	V(GWGrad2)	Var(NySW5)
	V(TPor1)	V(GWGrad3)	V(NMF1)
	V(TPor2)	V(GWGrad4)	V(NMF2)
	V(TPor3)	V(HydCond1)	V(NMF3)
	V(EyGW1)	V(HydCond2)	V(Cgw1)
	V(EyGW2)	V(HydCond3)	V(Cgw2)
	V(EyGW3)	V(HydCond4)	V(GWGrad1)
	V(NyGW1)	V(SatFoc1)	V(GWGrad2)
	V(NyGW2)	V(SatFoc2)	V(GWGrad3)
	V(NyGW3)	V(SatFoc3)	V(GWGrad4)
	V(Width1)	V(SatFoc4)	V(GWGrad5)
	V(Width2)	V(TPor1)	V(HydCond1)
	V(Width3)	V(TPor2)	V(HydCond2)
	V(Depth1)	V(TPor3)	V(HydCond3)
	V(Depth2)	V(TPor4)	V(HydCond4)
	V(Depth3)	V(EyGW1)	V(HydCond5)
	V(sDist1)	V(EyGW2)	V(SatFoc1)
	V(sDist2)	V(EyGW3)	V(SatFoc2)
	V(sDist3)	V(EyGW4)	V(SatFoc3)
		V(NyGW1)	V(SatFoc4)
		V(NyGW2)	V(SatFoc5)
		V(NyGW3)	V(TPor1)
		V(NyGW4)	V(TPor2)
		V(Width1)	V(TPor3)
		V(Width2)	V(TPor4)
		V(Width3)	V(TPor5)
		V(Width4)	V(EyGW1)
		V(Depth1)	V(EyGW2)
		V(Depth2)	V(EyGW3)
		V(Depth3)	V(EyGW4)
		V(Depth4)	V(EyGW5)
		V(sDist1)	V(NyGW1)
		V(sDist2)	V(NyGW2)
		V(sDist3)	V(NyGW3)
		V(SDist4)	V(NyGW4)
			V(NyGW5)
			V(Width1)
			V(Width2)

OrderNumber	8	9	10
			V(Width3)
			V(Width4)
			V(Width5)
			V(Depth1)
			V(Depth2)
			V(Depth3)
			V(Depth4)
			V(Depth5)
			V(sDist1)
			V(sDist2)
			V(sDist3)
			V(SDist4)
			V(SDist5)

Table D.4 - Access Queries for the Groundwater to Surface Water Calculations (Continued)

OrderNumber	11	12	13	14
Query Name	SWMixQ5A	SWGroupQ1	SWGroupQ2	SWMixQ6
Tables	SWMixUnion, SWMixQ3	GroupIdentifiers	SWGroupQ1, ChemConstants, ChemCorrelations, SubSurCorrelations, SubSurVariables, SWChemVariables, SWVarCorrelations, SWVariables, SourceCorrelations	SWGroup2, SWMix5A
Variable Position:	Group_ID	Group_ID	Group_ID	Group_ID
	GroupOrderNo	ChemVar	E(Ccrk)	F-NMF1
	F-NMF1	COC_ID	E(H)	F-NMF2
	F-NMF2	MdlErrVar	E(Qcrk)	F-NMF3
	F-NMF3	SubSurVar	TestNMF	F-Cgw1
	F-Cgw1	SWVar	Sol	F-Cgw2
	F-Cgw2	SRVar	NySW	F-Cgw3
	F-Cgw3	TestNMF	EySW	E(Ccrk)
	Var(Degrad1)		Var(Ccrk)	E(H)
	Var(Degrad2)		Var(EySW)	E(Qcrk)
	Var(Degrad3)		Var(NySW)	TestNMF

OrderNumber	11	12	13	14
	Var(Degrad4)		R-Degrad/Koc	Sol
	Var(Degrad5)		R-Qcrk/H	NySW
	Var(Degrad6)		R-GWGrad/HydCond	EySW
	Var(Koc1)		R-GWGrad/SatFOC	Var(Ccrk)
	Var(Koc2)		R-GWGrad/TPor	Var(Degrad1)
	Var(Koc3)		R-HydCond/TPor	Var(Degrad2)
	Var(Koc4)		R-HydCond/SatFOC	Var(Degrad3)
	Var(Koc5)		R-TPor/SatFOC	Var(Degrad4)
	Var(Koc6)		R-Depth/Width	Var(Degrad5)
	Var(H1)		R-sDist/Depth	Var(Degrad6)
	Var(H2)		R-sDist/Width	Var(Koc1)
	Var(H3)			Var(Koc2)
	Var(H4)			Var(Koc3)
	Var(H5)			Var(Koc4)
	Var(H6)			Var(Koc5)
	Var(Qcrk1)			Var(Koc6)
	Var(Qcrk2)			Var(H1)
	Var(Qcrk3)			Var(H2)
	Var(Qcrk4)			Var(H3)
	Var(Qcrk5)			Var(H4)
	Var(Qcrk6)			Var(H5)
	Var(NySW1)			Var(H6)
	Var(NySW2)			Var(Qcrk1)
	Var(NySW3)			Var(Qcrk2)
	Var(NySW4)			Var(Qcrk3)
	Var(NySW5)			Var(Qcrk4)
	Var(NySW6)			Var(Qcrk5)
	V(NMF1)			Var(Qcrk6)
	V(NMF2)			Var(EySW)
	V(NMF3)			Var(NySWG)
	V(Cgw1)			Var(NySW1)
	V(Cgw2)			Var(NySW2)
	V(Cgw3)			Var(NySW3)
	V(GWGrad1)			Var(NySW4)
	V(GWGrad2)			Var(NySW5)
	V(GWGrad3)			Var(NySW6)
	V(GWGrad4)			V(NMF1)
	V(GWGrad5)			V(NMF2)

OrderNumber	11	12	13	14
	V(GWGrad6)			V(NMF3)
	V(HydCond1)			V(Cgw1)
	V(HydCond2)			V(Cgw2)
	V(HydCond3)			V(Cgw3)
	V(HydCond4)			V(GWGrad1)
	V(HydCond5)			V(GWGrad2)
	V(HydCond6)			V(GWGrad3)
	V(SatFoc1)			V(GWGrad4)
	V(SatFoc2)			V(GWGrad5)
	V(SatFoc3)			V(GWGrad6)
	V(SatFoc4)			V(HydCond1)
	V(SatFoc5)			V(HydCond2)
	V(SatFoc6)			V(HydCond3)
	V(TPor1)			V(HydCond4)
	V(TPor2)			V(HydCond5)
	V(TPor3)			V(HydCond6)
	V(TPor4)			V(SatFoc1)
	V(TPor5)			V(SatFoc2)
	V(TPor6)			V(SatFoc3)
	V(EyGW1)			V(SatFoc4)
	V(EyGW2)			V(SatFoc5)
	V(EyGW3)			V(SatFoc6)
	V(EyGW4)			V(TPor1)
	V(EyGW5)			V(TPor2)
	V(EyGW6)			V(TPor3)
	V(NyGW1)			V(TPor4)
	V(NyGW2)			V(TPor5)
	V(NyGW3)			V(TPor6)
	V(NyGW4)			V(EyGW1)
	V(NyGW5)			V(EyGW2)
	V(NyGW6)			V(EyGW3)
	V(Width1)			V(EyGW4)
	V(Width2)			V(EyGW5)
	V(Width3)			V(EyGW6)
	V(Width4)			V(NyGW1)
	V(Width5)			V(NyGW2)
	V(Width6)			V(NyGW3)
	V(Depth1)			V(NyGW4)

OrderNumber	11	12	13	14
	V(Depth2)			V(NyGW5)
	V(Depth3)			V(NyGW6)
	V(Depth4)			V(Width1)
	V(Depth5)			V(Width2)
	V(Depth6)			V(Width3)
	V(sDist1)			V(Width4)
	V(sDist2)			V(Width5)
	V(sDist3)			V(Width6)
	V(SDist4)			V(Depth1)
	V(SDist5)			V(Depth2)
	V(SDist6)			V(Depth3)
				V(Depth4)
				V(Depth5)
				V(Depth6)
				V(sDist1)
				V(sDist2)
				V(sDist3)
				V(SDist4)
				V(SDist5)
				V(SDist6)
				R-Degrad/Koc
				R-Qcrk/H
				R-GWGrad/HydCond
				R-GWGrad/SatFoc
				R-GWGrad/TPor
				R-HydCond/TPor
				R-HydCond/SatFoc
				R-TPor/SatFoc
				R-Depth/Width
				R-sDist/Depth
				R-sDist/Width

Table D.4 - Access Queries for the Groundwater to Surface Water Calculations (Continued)

OrderNumber	15	16	17
Query Name	SWVarStart	SWVarQ1	SWVarUnion
Tables	Pathway, Pathway_Segments, UniqueIDBuilder	SWVarStart, PSWResults	SWVarQ1, GroupSupportSWVar
Variable Position:	UniqueID	UniqueID	UniqueID
	Group_ID	Group_ID	Group_ID
	GroupOrderNo	GroupOrderNo	GroupOrderNo
	Attenuation_Mechanism	E(CCgXI)	E(CCgXI)
		dCCgXI/dCb	dCCgXI/dCb
		dCCgXI/dNMF	dCCgXI/dNMF
		dCCgXI/dCgw	dCCgXI/dCgw
		dCCgXI/dNyGW	dCCgXI/dNyGW
		dCCgXI/dEyGW	dCCgXI/dEyGW
		dCCgXI/dDegrad	dCCgXI/dDegrad
		dCCgXI/dKoc	dCCgXI/dKoc
		dCCgXI/dDepth	dCCgXI/dDepth
		dCCgXI/dGWGrad	dCCgXI/dGWGrad
		dCCgXI/dHydCond	dCCgXI/dHydCond
		dCCgXI/dSatFoc	dCCgXI/dSatFoc
		dCCgXI/dsDist	dCCgXI/dsDist
		dCCgXI/dTotalPor	dCCgXI/dTotalPor
		dCCgXI/dWidth	dCCgXI/dWidth
		dCCgXI/dH	dCCgXI/dH
		dCCgXI/dNYSW	dCCgXI/dNYSW
		dCCgXI/dQcrk	dCCgXI/dQcrk

Table D.4 - Access Queries for the Groundwater to Surface Water Calculations (Continued)

OrderNumber	18	19	20
Query Name	SWVarQ2	SWVarQ3	SWVarQ4
Tables	SWVarUnion	SWVarQ2, SWVarUnion	SWVarQ3, SWVarUnion
Variable Position:	Group_ID	Group_ID	Group_ID
	GroupOrderNo	GroupOrderNo	GroupOrderNo
	E(CCgX1)	E(CCgX1)	E(CCgX1)
	dCCgX1/dCb	E(CCgX2)	E(CCgX2)

OrderNumber	18	19	20
	dCCgX1/dNMF	dCCgX1/dCb	E(CCgX3)
	dCCgX1/dCgw	dCCgX2/dCb	dCCgX1/dCb
	dCCgX1/dNyGW	dCCgX1/dNMF1	dCCgX2/dCb
	dCCgX1/dEyGW	dCCgX2/dNMF2	dCCgX3/dCb
	dCCgX1/dDegrad	dCCgX1/dNyGW	dCCgX1/dNMF1
	dCCgX1/dKoc	dCCgX2/dNyGW	dCCgX2/dNMF2
	dCCgX1/dDepth	dCCgX1/dEyGW	dCCgX3/dNMF3
	dCCgX1/dGWGrad	dCCgX2/dEyGW	dCCgX1/dNyGW
	dCCgX1/dHydCond	dCCgX1/dDegrad	dCCgX2/dNyGW
	dCCgX1/dSatFoc	dCCgX2/dDegrad	dCCgX3/dNyGW
	dCCgX1/dsDist	dCCgX1/dKoc	dCCgX1/dEyGW
	dCCgX1/dTotalPor	dCCgX2/dKoc	dCCgX2/dEyGW
	dCCgX1/dWidth	dCCgX1/dDepth	dCCgX3/dEyGW
	dCCgXI1/dH	dCCgX2/dDepth	dCCgX1/dDegrad
	dCCgXI1/dNYSW	dCCgX1/dGWGrad	dCCgX2/dDegrad
	dCCgXI1/dQcrk	dCCgX2/dGWGrad	dCCgX3/dDegrad
		dCCgX1/dHydCond	dCCgX1/dKoc
		dCCgX2/dHydCond	dCCgX2/dKoc
		dCCgX1/dSatFoc	dCCgX3/dKoc
		dCCgX2/dSatFoc	dCCgX1/dDepth
		dCCgX1/dsDist	dCCgX2/dDepth
		dCCgX2/dsDist	dCCgX3/dDepth
		dCCgX1/dTotalPor	dCCgX1/dGWGrad
		dCCgX2/dTotalPor	dCCgX2/dGWGrad
		dCCgX1/dWidth	dCCgX3/dGWGrad
		dCCgX2/dWidth	dCCgX1/dHydCond
		dCCgXI1/dH	dCCgX2/dHydCond
		dCCgXI2/dH	dCCgX3/dHydCond
		dCCgXI1/dNYSW	dCCgX1/dSatFoc
		dCCgXI2/dNYSW	dCCgX2/dSatFoc
		dCCgXI1/dQcrk	dCCgX3/dSatFoc
		dCCgXI2/dQcrk	dCCgX1/dsDist
			dCCgX2/dsDist
			dCCgX3/dsDist
			dCCgX1/dTotalPor
			dCCgX2/dTotalPor
			dCCgX3/dTotalPor
			dCCgX1/dWidth

OrderNumber	18	19	20
			dCCgX2/dWidth
			dCCgX3/dWidth
			dCCgXI1/dH
			dCCgXI2/dH
			dCCgXI3/dH
			dCCgXI1/dNYSW
			dCCgXI2/dNYSW
			dCCgXI3/dNYSW
			dCCgXI1/dQcrk
			dCCgXI2/dQcrk
			dCCgXI3/dQcrk

Table D.4 - Access Queries for the Groundwater to Surface Water Calculations (Continued)

OrderNumber	21	22	23
Query Name	SWVarQ5	SWVarQ6	SWVarQ7
Tables	SWVarQ4, SWVarUnion	SWVarQ5, SWVarUnion	SWVarQ6, SWVarUnion
Variable Position:	Group_ID	Group_ID	Group_ID
	GroupOrderNo	GroupOrderNo	E(CCgX1)
	E(CCgX1)	E(CCgX1)	E(CCgX2)
	E(CCgX2)	E(CCgX2)	E(CCgX3)
	E(CCgX3)	E(CCgX3)	E(CCgX4)
	E(CCgX4)	E(CCgX4)	E(CCgX5)
	dCCgX1/dCb	E(CCgX5)	E(CCgX6)
	dCCgX2/dCb	dCCgX1/dCb	dCCgX1/dCb
	dCCgX3/dCb	dCCgX2/dCb	dCCgX2/dCb
	dCCgX4/dCb	dCCgX3/dCb	dCCgX3/dCb
	dCCgX1/dNMF1	dCCgX4/dCb	dCCgX4/dCb
	dCCgX2/dNMF2	dCCgX5/dCb	dCCgX5/dCb
	dCCgX3/dNMF3	dCCgX1/dNMF1	dCCgX6/dCb
	dCCgX4/dCgw	dCCgX2/dNMF2	dCCgX1/dNMF1
	dCCgX1/dNyGW	dCCgX3/dNMF3	dCCgX2/dNMF2
	dCCgX2/dNyGW	dCCgX4/dCgw	dCCgX3/dNMF3
	dCCgX3/dNyGW	dCCgX5/dCgw	dCCgX4/dCgw
	dCCgX4/dNyGW	dCCgX1/dNyGW	dCCgX5/dCgw
	dCCgX1/dEyGW	dCCgX2/dNyGW	dCCgX6/dCgw
	dCCgX2/dEyGW	dCCgX3/dNyGW	dCCgX1/dNyGW
	dCCgX3/dEyGW	dCCgX4/dNyGW	dCCgX2/dNyGW
	dCCgX4/dEyGW	dCCgX5/dNyGW	dCCgX3/dNyGW
	dCCgX1/dDegrad	dCCgX1/dEyGW	dCCgX4/dNyGW
	dCCgX2/dDegrad	dCCgX2/dEyGW	dCCgX5/dNyGW
	dCCgX3/dDegrad	dCCgX3/dEyGW	dCCgX6/dNyGW
	dCCgX4/dDegrad	dCCgX4/dEyGW	dCCgX1/dEyGW
	dCCgX1/dKoc	dCCgX5/dEyGW	dCCgX2/dEyGW
	dCCgX2/dKoc	dCCgX1/dDegrad	dCCgX3/dEyGW
	dCCgX3/dKoc	dCCgX2/dDegrad	dCCgX4/dEyGW
	dCCgX4/dKoc	dCCgX3/dDegrad	dCCgX5/dEyGW
	dCCgX1/dDepth	dCCgX4/dDegrad	dCCgX6/dEyGW
	dCCgX2/dDepth	dCCgX5/dDegrad	dCCgX1/dDegrad
	dCCgX3/dDepth	dCCgX1/dKoc	dCCgX2/dDegrad

OrderNumber	21	22	23
	dCCgX4/dDepth	dCCgX2/dKoc	dCCgX3/dDegrad
	dCCgX1/dGWGrad	dCCgX3/dKoc	dCCgX4/dDegrad
	dCCgX2/dGWGrad	dCCgX4/dKoc	dCCgX5/dDegrad
	dCCgX3/dGWGrad	dCCgX5/dKoc	dCCgX6/dDegrad
	dCCgX4/dGWGrad	dCCgX1/dDepth	dCCgX1/dKoc
	dCCgX1/dHydCond	dCCgX2/dDepth	dCCgX2/dKoc
	dCCgX2/dHydCond	dCCgX3/dDepth	dCCgX3/dKoc
	dCCgX3/dHydCond	dCCgX4/dDepth	dCCgX4/dKoc
	dCCgX4/dHydCond	dCCgX5/dDepth	dCCgX5/dKoc
	dCCgX1/dSatFoc	dCCgX1/dGWGrad	dCCgX6/dKoc
	dCCgX2/dSatFoc	dCCgX2/dGWGrad	dCCgX1/dDepth
	dCCgX3/dSatFoc	dCCgX3/dGWGrad	dCCgX2/dDepth
	dCCgX4/dSatFoc	dCCgX4/dGWGrad	dCCgX3/dDepth
	dCCgX1/dsDist	dCCgX5/dGWGrad	dCCgX4/dDepth
	dCCgX2/dsDist	dCCgX1/dHydCond	dCCgX5/dDepth
	dCCgX3/dsDist	dCCgX2/dHydCond	dCCgX6/dDepth
	dCCgX4/dsDist	dCCgX3/dHydCond	dCCgX1/dGWGrad
	dCCgX1/dTotalPor	dCCgX4/dHydCond	dCCgX2/dGWGrad
	dCCgX2/dTotalPor	dCCgX5/dHydCond	dCCgX3/dGWGrad
	dCCgX3/dTotalPor	dCCgX1/dSatFoc	dCCgX4/dGWGrad
	dCCgX4/dTotalPor	dCCgX2/dSatFoc	dCCgX5/dGWGrad
	dCCgX1/dWidth	dCCgX3/dSatFoc	dCCgX6/dGWGrad
	dCCgX2/dWidth	dCCgX4/dSatFoc	dCCgX1/dHydCond
	dCCgX3/dWidth	dCCgX5/dSatFoc	dCCgX2/dHydCond
	dCCgX4/dWidth	dCCgX1/dsDist	dCCgX3/dHydCond
	dCCgXI1/dH	dCCgX2/dsDist	dCCgX4/dHydCond
	dCCgXI2/dH	dCCgX3/dsDist	dCCgX5/dHydCond
	dCCgXI3/dH	dCCgX4/dsDist	dCCgX6/dHydCond
	dCCgXI4/dH	dCCgX5/dsDist	dCCgX1/dSatFoc
	dCCgXI1/dNYSW	dCCgX1/dTotalPor	dCCgX2/dSatFoc
	dCCgXI2/dNYSW	dCCgX2/dTotalPor	dCCgX3/dSatFoc
	dCCgXI3/dNYSW	dCCgX3/dTotalPor	dCCgX4/dSatFoc
	dCCgXI4/dNYSW	dCCgX4/dTotalPor	dCCgX5/dSatFoc
	dCCgXI1/dQcrk	dCCgX5/dTotalPor	dCCgX6/dSatFoc
	dCCgXI2/dQcrk	dCCgX1/dWidth	dCCgX1/dsDist
	dCCgXI3/dQcrk	dCCgX2/dWidth	dCCgX2/dsDist
	dCCgXI4/dQcrk	dCCgX3/dWidth	dCCgX3/dsDist
		dCCgX4/dWidth	dCCgX4/dsDist

OrderNumber	21	22	23
		dCCgX5/dWidth	dCCgX5/dsDist
		dCCgXI1/dH	dCCgX6/dsDist
		dCCgXI2/dH	dCCgX1/dTotalPor
		dCCgXI3/dH	dCCgX2/dTotalPor
		dCCgXI4/dH	dCCgX3/dTotalPor
		dCCgXI5/dH	dCCgX4/dTotalPor
		dCCgXI1/dNYSW	dCCgX5/dTotalPor
		dCCgXI2/dNYSW	dCCgX6/dTotalPor
		dCCgXI3/dNYSW	dCCgX1/dWidth
		dCCgXI4/dNYSW	dCCgX2/dWidth
		dCCgXI5/dNYSW	dCCgX3/dWidth
		dCCgXI1/dQcrk	dCCgX4/dWidth
		dCCgXI2/dQcrk	dCCgX5/dWidth
		dCCgXI3/dQcrk	dCCg6/dWidth
		dCCgXI4/dQcrk	dCCgXI1/dH
		dCCgXI5/dQcrk	dCCgXI2/dH
			dCCgXI3/dH
			dCCgXI4/dH
			dCCgXI5/dH
			dCCgXI6/dH
			dCCgXI1/dNYSW
			dCCgXI2/dNYSW
			dCCgXI3/dNYSW
			dCCgXI4/dNYSW
			dCCgXI5/dNYSW
			dCCgXI6/dNYSW
			dCCgXI1/dQcrk
			dCCgXI2/dQcrk
			dCCgXI3/dQcrk
			dCCgXI4/dQcrk
			dCCgXI5/dQcrk
			dCCgXI6/dQcrk

D.4 EXCEL SHEET CONSTRUCTION

Each group exposure pathway calculation has a separate Excel workbook. In each workbook there are several calculation sheets that have links to the

Access queries and there are other calculation sheets that reference the data from the linked sheets. Table D.5 includes the names of each of the sheets in the indoor air workbook and a brief description of each sheet's function. Table D.6 includes the information for the outdoor air workbook and Table D.7 includes the information for the groundwater to surface water workbook. In general, the calculations were split across several Excel sheets because there is a much higher number of fields in these calculations than the deterministic calculations and so the spreadsheets are easier to manage.

Table D.5 - Excel Workbooks for Indoor Air Probabilistic Calculations

Excel Sheet	Linked Access Query Table	Description
Indoor Air: PVapIn.xls		
VaportoIndoor E(C)	PVapIn	Calculation of mean results for individual exposure pathways
VapInResults	None	Cell references to calculated results
VapInVar(CE)	None	Calculation of variance results for individual exposure pathways
Var(CE)Results	None	Cell references to calculated results
GroupVar	PVIVarQ8	Calculation of mean and variance results for the groups
Var(CG)Results	None	Cell references to calculated results

Table D.6 - Excel Workbooks for Outdoor Air Probabilistic Calculations

Excel Sheet	Linked Access Query Table	Description
Outdoor Air: PVapOut.xls		
VapOut E(C)	PVapOut	Calculation of mean results for individual exposure pathways
VapOutResults	None	Cell references to calculated results
VapOutVar(C)	None	Calculation of variance results for individual exposure pathways

Var(CF)Results	None	Cell references to calculated results
GroupVar(2)	PVOVarQ8	Calculation of mean and variance results for the groups
Var(CH)Results	None	Cell references to calculated results

Table D.7 - Excel Workbooks for Groundwater to Surface Water Probabilistic Calculations

Excel Sheet	Linked Access Query Table	Description
Groundwater to Surface Water: PGWSW.xls		
GWTrans(g'(x))	PathTest	Calculation of mean results for individual exposure pathways
SWMix E(C)	None	Calculation of partial derivatives for individual exposure pathways
SWResults	None	Cell references to calculated results
SWVarInput	SWMixQ6	Input sheet for variance data
SWVarInput (2)	SWVarQ7	Input sheet for mean data and partial derivatives
SWVar(C)	None	Calculation of mean and derivative results for the groups
SWVar(2)	None	Calculation of the variance results
SWVarRes	None	Cell references to calculated results

As discussed in Appendix C, once the queries have been established and the Excel sheets have been created, the query results for each calculation are linked to the corresponding Excel sheet. The Excel sheets start out with an example record in each calculation sheet. The results sheet for each calculation is linked to the Access file. The queries in Table D.2, Table D.3 and Table D.4 are executed in order. Each time the results for the query are updated the example records are replaced with the actual records. Once the connections have been established, if new records are added to the site conceptual model, the queries must be executed again in the order given in Table D.2, Table D.3 and Table D.4. The Excel sheets can be set to refresh data upon opening them. Care should be taken with this function in the developmental stage, if the names of the queries are changed once the connection is made, or the Access file name is changed additional steps will be needed to re-establish the connections. Also, there is an option in Excel to remove the linked data when the file is closed. This should not be used since removing the external data also removes the columns established for the data. This can cause errors in the cell references in the formulae.

In Excel when using the Refresh Data option, be sure that the cell or cells selected are within the external data range to be refreshed.

If a link exists between an Access query and Excel and the query changes, even if it is only the name of a field, the Too Few Parameters error results. This error is avoided by not refreshing the data after the query change. In the **Tools** use the **Edit Query** command and then re-establish the link to the query table. The **Edit Query** command can also be used the change the path or the ODBC driver and update the data range.

If the files are moved to a different computer, or to different folders on the same computer, the locations of the files can be identified in Excel using the **Edit, Links, Change Source**, using the dialog box, find the database file in the appropriate folder. This will update the path on all of the references.

Appendix E : Data Dictionary and Archive Records

This Appendix includes the file archive records and the data dictionary tables for the spatial environmental risk assessment.

E.1 SPATIAL DATABASE

The spatial database files for the case study facility include files from the digital facility description, derived data files from the environmental measurements database and site conceptual model data files based on the available information.

All of the spatial database files are in the Pennsylvania State Plane System (South Zone). This projection uses the North American Datum of 1983 (NAD83) and North American Vertical Datum of 1988 (NAVD88). The projection parameters for the Pennsylvania State Plan System (South Zone) are given in Table E.1(Romanek, *et. al*., 1999).

Table E.1 - Projection Parameters for the Pennsylvania State Plane Projection (South Zone)

Projection	**Lambert Conformal Conic**
Map Units	Feet
Datum	NAD83
First Standard Parallel	39 56 00
Second Standard Parallel	40 58 00
Central Meridian	-77 45 00
Latitude	39 20 00
False Easting	600000
False Northing	0

The spatial database files used in this research are listed in Table E.2. The table includes a brief description of the features in each shapefile or coverage, the file type and the location on the CD-ROM produced to document this work.

Table E.2 - Spatial Database Files

File Name	Feature Description	File Type	Location
airbox.shp	locations of indoor and outdoor air mixing attenuation factors; included in the evaluation of grouped exposure pathways	polygon	\scm\shapefiles
airboxA.shp	air mixing boxes for the chapter 4 examples	polygon	\scm\shapefiles
airpodsg.shp	indoor and outdoor air points of demonstration for grouped exposure pathways	point	\scm\shapefiles
anthsoil.shp	soil concentration data for anthracene	point	\DataAnalysis\shapefiles
aoc-border.shp	sub-areas in the lube plant defined by Langan (1999)	polygon	\facility
bapsoil.shp	soil concentration data for benzo (a) pyrene	point	\DataAnalysis\shapefiles
benzgw.shp	benzene detections in groundwater in lube plant from 1996-1999	point	\DataAnalysis\shapefiles
bldgclip.shp	buildings in the lube plant	polygon	\facility
boundary.shp	Marcus Hook Refinery property line	line	\facility
bpoil.tif	USGS quadrangle map	image	\facility
bzgw1199.shp	detections of benzene in groundwater (11/99)	point	\DataAnalysis\shapefiles
bzsoil.shp	soil concentration data for benzene	point	\DataAnalysis\shapefiles
chrysoil.shp	soil concentration data for chrysene	point	\DataAnalysis\shapefiles
ctrpts.shp	source area center points for SERA exercise	point	\scmexercise\serazip
gwc121098.shp	groundwater benzene concentrations (12/98)	point	\scmexercise\serazip
gwcontours4.shp	interpolated contours of groundwater elevation from November 1999	arc	\DataAnalysis\shapefiles
gwelev2k.shp	groundwater elevation data February 2000	point	\DataAnalysis\shapefiles
gwelev3	groundwater elevation grid file	grid	\DataAnalysis\gwelev3
gwelev4	interpolated surface of groundwater elevations (11/99)	grid	\DataAnalysis\gwelev4
gwelevfeb2k.shp	NAPL thickness data, groundwater elevation data from February 2000	point	\DataAnalysis\shapefiles

File Name	Feature Description	File Type	Location
gwpods.shp	individual exposure pathway groundwater points of demonstration	point	\scm\shapefiles
gwpodsg.shp	groundwater to surface water points of demonstration for grouped exposure pathways	point	\scm\shapefiles
gwtran1.shp	groundwater transport segments	arc	\scm\shapefiles
gwtrans3.shp	groundwater transport segments for SERA exercise	arc	\scmexercise\serazip
hydcond3.shp	hydraulic conductivity data (cm/sec) from the lube plant	point	\DataAnalysis\shapefiles
inpods.shp	indoor air points of demonstration	point	\scm\shapefiles
lubearea.shp	lube plant area	polygon	\facility
lubebound.shp	perimeter boundary of the lube plant	arc	\facility
lubepod3.shp	points of demonstration in the lube plant for the SERA exercise	point	\scmexercise\serazip
lubepod4A.shp	PODs for the individual exposure pathway examples	point	\scm\shapefiles
lubepod5.shp	PODs for the chapter 4 examples	point	\scm\shapefiles
lubetank.shp	above ground storage tanks in the lube plant	polygon	\facility
lubpod4.shp	Locations of points of demonstration for the example exposure pathways in the lube plant	point	\scm\shapefiles
matchpt.dbf	corresponding source area center points and points of demonstration for the groundwater transport segments	data table	\scm\shapefiles
mudfence.shp	approximate location of the mud fences in the lube plant	arc	\facility
napl2_2k.shp	NAPL thickness measurements from February 2000	point	\DataAnalysis\shapefiles
napldet.shp	occurrence of NAPL in the lube plant, based on all sampling dates	point	\DataAnalysis\shapefiles
naplsg.shp	NAPL specific gravity data	point	\DataAnalysis\shapefiles
nwscrpts.shp	source area center points	point	\scm\shapefiles
origdem2	ground surface elevation model from aerial survey	grid	\DataAnalysis\origdem2
outpod2.shp	outdoor air points of demonstration	point	\scm\shapefiles
prelim.shp	preliminary source areas	polygon	\DataAnalysis\shapefiles
src2centr.shp	center points of the source areas for the Chapter 5 examples	point	\scm\shapefiles
srcarea.shp	source areas for the SERA exercise	polygon	\scmexercise\serazip

File Name	**Feature Description**	**File Type**	**Location**
srcarea2.shp	approximate locations of NAPL in the lube plant for the example exposure pathways	polygon	\DataAnalysis\shapefiles
srcbuff.shp	estimated source areas plus 10-percent buffer	polygon	\DataAnalysis\shapefiles
steelbulk4.shp	location of the steel bulkhead in the lube plant	arc	\facility
surfwatlube.shp	surface water locations adjacent to the lube plant	polygon	\facility
swbox.shp	surface water mixing box locations	polygon	\scm\shapefiles
swboxA.shp	surface water mixing box locations for the chapter 4 examples	polygon	\scm\shapefiles
tablept	corresponding source area center points and points of demonstration for the groundwater transport segments for the SERA exercise	data table	\scmexercise\serazip
transpts.shp	source area transition points	point	\scm\shapefiles
transptsa.shp	air transition points	point	\scm\shapefiles
transptsw.shp	groundwater to surface water transition points	point	\scm\shapefiles
vadthck2	calculated grid file based on DEM and groundwater elevations	grid	\DataAnalysis\vadthck2
wells.shp	locations of groundwater monitoring wells prior to 1999 field investigation	point	\DataAnalysis\shapefiles
wells1199.shp	locations of groundwater elevation measurements from November 1999	point	\DataAnalysis\shapefiles
wells2K.shp	locations of groundwater monitoring wells as of February 2000	point	\DataAnalysis\shapefiles

Notes: All files are ArcView shapefiles, created in version 3.1
All shapefiles include several files including a *.dbf file with the same name as the *.shp file.
Only *.shp files are listed.

In addition to the data files that were generated, ArcView project files were created. These files are included in Table E.3.

Table E.3 - ArcView Project Files

File Name	Description	Location
scm1.apr	case study site conceptual model	\DataAnalysis
data1.apr	data evaluation for the model input parameters	\DataAnalysis
data2.apr	data evaluation for the preliminary exposure pathway evaluation	\DataAnalysis
data3.apr	data evaluation for the groundwater model calibration	\DataAnalysis
mhrloc.apr	Location map for the Marcus Hook Refinery	\DataAnalysis
scriptest.apr	Initial examples of MHR lube plant site conceptual model	\scmexercise

Notes:
All files are ArcView 3.1 Project Files

The ArcView scripts, legends and extension files used in this research are included in Table E.4.

Table E.4 - ArcView Scripts, Legends and Extension Files

File Name	Description	File Type	Location
airbox.avl	legend for indoor and outdoor air mixing boxes	ArcView legend	\scm
bound.avl	legend for lube plant boundary	ArcView legend	\scm
complete.avl	legend for exposure pathway completeness	ArcView legend	\scm
gwgrid.avl	legend for groundwater elevation grid	ArcView legend	\scm
gwgrid2.avl	legend for groundwater elevation grid	ArcView legend	\scm
gwtrans.ave	transport segment defining script	Avenue script	\scm
gwtrans.avl	legend for groundwater transport segments	ArcView legend	\scm
hydc1.avl	legend for hydraulic conductivity data	ArcView legend	\scm
hydc2.avl	legend for hydraulic conductivity data	ArcView legend	\scm
hydc3.avl	legend for hydraulic conductivity data	ArcView legend	\scm
napl.avl	legend for NAPL detections	ArcView legend	\scm
naplsg.avl	legend for NAPL specific gravity data	ArcView legend	\scm
origdem.avl	legend for digital elevation model	ArcView legend	\scm

File Name	Description	File Type	Location
pods.avl	legend for points of demonstration	ArcView legend	\scm
soilbenz.avl	legend for soil concentration data	ArcView legend	\scm
source.avl	legend for source areas	ArcView legend	\scm
swater.avl	legend for surface water	ArcView legend	\scm
swbox.avl	legend for surface water mixing boxes	ArcView legend	\scm
vector11.avx	vector operations extension	Avenue extension	\scm
vector11.html	vector operations extension help file	html	\scm

E.2 TABULAR DATABASES

The relational database files used in this research include the environmental measurements database files and the site conceptual model database files. These files are included in Table E.5.

Table E.5 - Relational Database Files

File Name	Description	File Type	Location
bp-data2K.mdb	Environmental measurements database, including all queries for Appendix B data evaluation	Access 2000	\DataAnalysis
bp-proto.mdb	Environmental measurements database, including all queries for Appendix B data evaluation	Access97	\DataAnalysis
SCMdata.mdb	Site conceptual model file, contains no links or queries. Base file for the SCM exercise, 11/99	Access97	\scmexercise\serazip
SCMdata2K.mdb	Site conceptual model file, contains no links or queries. Base file for the SCM exercise, 11/99	Access 2000	\scmexercise\serazip
SCMmaster.mdb	Site conceptual model file, contains no links or queries. Includes the Chapter 5 example input values and exposure pathways	Access97	\scm
SCMres2K.mdb	Site conceptual model file, including all queries and links. Answer file for the SCM exercise	Access 2000	\scmexercise\serazip
Uncdatav2.mdb	Site conceptual model file, contains all links and queries. Includes the Chapter 5 example input values and exposure pathways	Access 2000	\Uncert

E.3 CALCULATION FILES

The calculation files used in this research are Excel spreadsheet files. There are data analysis files that were used in the parameter estimation and preparation of the results and there are spreadsheet files that are linked to the site conceptual model database for the fate and transport calculations. In addition several verification files (e.g., gwcalibration.xls and Eqncheckv2.xls) are included in the file records. These were used to check the calculation methods and the results. The calculation files are listed in Table E.6.

Table E.6 - Calculation Files

File Name	Description	File Type	Location
airsegmaster.xls	calculation spreadsheet for the deterministic indoor and outdoor air exposure pathways; contains no links	Excel 2000	\scm
dairseg.xls	calculation spreadsheet for the deterministic indoor and outdoor air exposure pathways; includes external links	Excel 2000	\Uncert
datadictionaryv1.xls	tables with the file records	Excel 2000	\scm
dgwseg.xls	calculation spreadsheet for deterministic calculations for groundwater and for surface water, file includes the external data links	Excel 2000	\Uncert
dnmfseg.xls	calculation spreadsheet for deterministic calculations for NAPL partitioning and volatilization; includes external links	Excel 2000	\Uncert
dsegmaster.xls	calculation spreadsheets for deterministic calculations, file does not include any external data links	Excel 2000	\scm
Eqncheckv2.xls	calculation spreadsheet for numerical check of the derivative functions	Excel 2000	\DataAnalysis
FOCdata.xls	measured data values for fraction organic carbon, calculation of mean and variance	Excel 2000	\DataAnalysis

File Name	**Description**	**File Type**	**Location**
grainsizev3.xls	data evaluation of grain size results, updated from J.Kim and D. Petrecca	Excel 2000	\DataAnalysis
gwcalibration.xls	calculation spreadsheet for comparison of predicted groundwater concentrations to field data	Excel 2000	\DataAnalysis
gwelevations.xls	data evaluation of groundwater elevation measurements	Excel 2000	\DataAnalysis
gwsegmaster.xls	calculation spreadsheet for the deterministic groundwater and surface water exposure pathways; contains no links	Excel 2000	\scm
gwswquery.xls	tables with the Access - Excel connection queries	Excel 2000	\Uncert
LogNormalVar.xls	calculation spreadsheet to check distribution results for log-normal variables	Excel 2000	\DataAnalysis
nmfsegmaster.xls	calculation spreadsheet for deterministic calculations for NAPL partitioning and volatilization; contains no links	Excel 2000	\scm
pgwsw.xls	calculation spreadsheets for probabilistic calculations for groundwater to surface water, file includes the external data links	Excel 2000	\scm
pgwswtest.xls	verification spreadsheet for the probabilistic groundwater to surface water calculation, includes the calculation fields for the correlation of variables between exposure pathways in a group	Excel 2000	\DataAnalysis
PrelimPQ.xls	tables for the preliminary pathway evaluation	Excel 2000	\DataAnalysis
pvapin.xls	calculation spreadsheets for probabilistic calculations for indoor air, file includes the external data links	Excel 2000	\Uncert
pvapout.xls	calculation spreadsheets for probabilistic calculations for outdoor air, file includes the external data links	Excel 2000	\Uncert
querytables.xls	tables with the Access - Excel connection queries	Excel 2000	\Uncert
resultsv3.xls	final results tables and graphs for Chapter 5	Excel 2000	\Uncert
segment.xls	calculation spreadsheet for all segments in the SERA exercise	Excel 97	\scmexercise\serazip

File Name	**Description**	**File Type**	**Location**
serav1.xla	Excel Add-In file for the user defined functions	Excel 2000	\Uncert
variablesv2.xls	parameter estimation calculations	Excel 2000	\Uncert

Vita

Lesley Isabel Hay Wilson was born in Chicago, Illinois on August 3, 1965, the daughter of Margaret Gibson Hastie Hay and James Potts Hay. After completing her studies at Lane Technical High School, Chicago, Illinois, in 1983 she attended Northwestern University where she studied chemical engineering. Ms. Hay Wilson worked in the cooperative engineering program with Standard Oil of Ohio during her undergraduate years. She received a Bachelor of Science from Northwestern University in Chemical Engineering in December 1987. Ms. Hay Wilson held positions in environmental engineering for BP Exploration & Oil Inc. for more than 10 years. In 1996 she began her studies at the Graduate School of The University of Texas at Austin. In 1998 she was awarded a Master of Science from The University of Texas in Civil Engineering, Environmental and Water Resources Engineering. Ms. Hay Wilson has been an instructor for many short courses and lectures in risk assessment, risk-based decision making and corrective action. She has presented several conference papers on the use of GIS in risk-based decision making.

Permanent address: 3117 Honey Tree Lane, Austin, Texas 78746

This dissertation was typed by the author.