

CRWR Online Report 99-2

Building the Foundation for Environmental Risk Assessment at the  
Marcus Hook Refinery Using Geographic Information Systems

by

Andrew Paul Romanek, M.S.E  
Graduate Research Assistant

and

Lesley Hay Wilson, M.S.E  
Graduate Research Assistant  
David R. Maidment, PhD  
Principal Investigator

May 1999

CENTER FOR RESEARCH IN WATER RESOURCES

Bureau of Engineering Research • The University of Texas at Austin  
J.J. Pickle Research Campus • Austin, TX 78712-4497

This document is available online via World Wide Web at  
<http://www.crwr.utexas.edu/online.html>

Copyright

by

Andrew Paul Romanek

1999

## ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. David Maidment, and Lesley Hay Wilson for their continued guidance and interest in this project. I have learned more from them in my two years here than I ever could have expected.

I would also like to thank all of the other Marcus Hook project team members. I have particularly enjoyed working with Jim Rocco and Sue Sharp of British Petroleum Exploration & Oil, Inc. (my research sponsor). I am also indebted to the help of Dr. Robert Gilbert and Lizan Koerner. Special thanks go to Rick Toth, whose wealth of information about the facility was invaluable to my research. It has been a great pleasure to work with all of you.

I sincerely appreciate the help of all the students, staff, and faculty in the EWRE Department. The Maidment Research Team, Ty Lehman, and Francisco Olivera have all helped me overcome many obstacles.

I am also grateful for the encouragement I have received from my friends here in Austin and throughout the world. I will cherish the experiences I have shared with them for the rest of my life.

Finally, I would like to thank my family, specifically my parents, Joe and Mollie, and my sisters, Jennifer and Ashley. There is no question that I would not be the person I am today had it not been for their love and support. Words cannot express what you have meant to me.

May 7, 1999

Building the Foundation for Environmental Risk Assessment at the  
Marcus Hook Refinery Using Geographic Information Systems

by

Andrew Paul Romanek, M.S.E.  
The University of Texas at Austin, 1999  
SUPERVISOR: David R. Maidment

**ABSTRACT**

This research presents a digitally-based methodology for approaching environmental risk assessments at large and complex industrial facilities, using the Marcus Hook Refinery in Pennsylvania as a case study site. The goal of this study is to demonstrate the development of a "digital facility description" and its use as an effective environmental risk assessment tool. The digital facility description is the collection of physical, chemical, geological, and hydrogeological information that has been spatially referenced in a geographic information system (GIS). It provides the mechanism to analyze sources and potential receptors in a spatial framework and to evaluate exposure pathways with models. The digital facility description has two components: (1) a spatial database of regional and facility features, and (2) a relational tabular database of environmental measurements. These two databases are dynamically linked providing a means of evaluating both spatial and temporal relationships. The digital facility description is used to support environmental risk assessment activities, such as map-based modeling and exposure analysis. Specifically, a surface water runoff model and a groundwater model were developed for Marcus Hook. Additionally, source area concentrations and the probabilities that these concentrations are above target levels were analyzed using "risk maps."

## TABLE OF CONTENTS

Section	Page No.
ACKNOWLEDGMENTS .....	ii
ABSTRACT.....	iii
TABLE OF CONTENTS .....	iv
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
1 INTRODUCTION.....	1
1.1 Environmental Risk Assessment .....	1
1.2 Environmental Risk Assessment with GIS .....	11
1.3 Project Background.....	16
1.4 Objective .....	20
1.5 Organization of Thesis .....	21
2 LITERATURE REVIEW .....	22
2.1 Introduction .....	22
2.2 GIS and Cartography .....	23
2.3 GIS and Modeling.....	24
2.4 GIS and Risk Assessment .....	27
3 DIGITAL FACILITY DESCRIPTION .....	29
3.1 Introduction .....	29
3.2 Spatial Database .....	31
3.3 Tabular Database .....	50
3.4 Data Development .....	57
3.4.1 Map Projections .....	57
3.4.2 Determining Data Needs .....	65
3.4.3 Photogrammetry and Spatial Data Creation.....	67
3.4.4 Constructing and Maintaining a Tabular Database.....	80
3.5 Connecting Spatial and Tabular Data.....	85
4 MAP-BASED MODELING .....	95
4.1 Surface Water Models.....	95
4.1.1 Methodology.....	95
4.1.2 Regional Model .....	100
4.1.3 Facility Model.....	109
4.2 Groundwater Model.....	122
4.2.1 Introduction .....	122
4.2.2 Setup and Results.....	124
5 RISK MAPPING .....	133
5.1 Introduction and Methods .....	133
5.1.1 Upper Confidence Limit (95%).....	134

## TABLE OF CONTENTS (cont.)

<b>Section</b>	<b>Page No.</b>
5.1.2 Binomial Distribution .....	139
5.1.3 Bayesian Approach .....	143
5.2 Results .....	147
5.2.1 Derivation of Target Levels.....	147
5.2.2 UCL and Binomial Results.....	150
5.2.3 Bayesian Results .....	158
5.2.4 Conclusions .....	165
6 CONCLUSIONS .....	168
APPENDICES.....	173
A Useful Internet Resources .....	173
B Spatial Data Dictionary.....	176
C Tabular Data Dictionary.....	206
D Mapping Environmental Data Exercise .....	233
E Data Conversions.....	253
F COC Transport Extension.....	258
G Avenue Scripts.....	277
H Risk Mapping Data Sheets.....	290
REFERENCES.....	334
VITA.....	339

## LIST OF TABLES

<b>Table</b>		<b>Page No.</b>
1.1	Environmental Risk Assessment Terms.....	4
3.1	Marcus Hook Regional Coverages.....	35
3.2	Marcus Hook Facility Coverages.....	43
3.3	Marcus Hook Tabular Database Tables .....	53
3.4	Pennsylvania State Plane Coordinate System Parameters.....	60
3.5	Monitoring Well Coordinate Conversions .....	63
4.1	Flow Calculations.....	118
4.2	Expected Mean Concentrations .....	120
4.3	Scatter Point Data File Format.....	129
5.1	Summary of Target Levels .....	149
5.2	Chosen Target Levels.....	149
5.3	Benzene Results Summary .....	152
5.4	Ethyl Benzene Results Summary.....	153
5.5	Naphthalene Results Summary .....	154
5.6	Toluene Results Summary.....	155
5.7	Xylenes Results Summary.....	156

## LIST OF FIGURES

Figure		Page No.
1.1	Example Environmental Risk Assessment Scenario.....	3
1.2	Environmental Risk Assessment Process Diagram.....	5
1.3	Dose-Response Curve .....	6
1.4	Risk Calculation Diagram.....	8
1.5	Corrective Action Approaches.....	10
1.6	Mapping Concentrations in Space and Time.....	12
1.7	GIS Data Models.....	13
1.8	GIS Application in PC-Based Environment .....	15
1.9	Spatial Environmental Risk Assessment Process .....	16
1.10	Case Study Area.....	17
1.11	Risk Communication Diagram .....	20
3.1	Features on a Map Related to Records in a Table.....	30
3.2	Relationship between Regional and Facility Features .....	33
3.3	Physiographic Regions in Pennsylvania.....	36
3.4	Census Tracts in Pennsylvania Displayed by Total Population within Each Tract .....	37
3.5	Streams and Surface Geology in Delaware County, PA.....	38
3.6	USGS Land Use Classifications for Southeast Pennsylvania.....	39
3.7	Digital Elevation Model for the City of Marcus Hook, PA .....	40
3.8	Annual Mean Precipitation Ranges for Southeast Pennsylvania .....	41
3.9	Marcus Hook Refinery Basemap .....	44
3.10	Marcus Hook Refinery Surface Cover .....	45
3.11	Marcus Hook Refinery Structures and Surrounding Properties ..	46
3.12	Marcus Hook Refinery Historical Features .....	47
3.13	Marcus Hook Refinery Oily Sewer Line.....	48
3.14	Marcus Hook Refinery Digital Terrain Model .....	49
3.15	Access Database Functions.....	52
3.16	Marcus Hook Tabular Database Relationships.....	54
3.17	Transformation of Locations on the Earth's Surface to Locations in a Two-Dimensional Coordinate System.....	58
3.18	Transformation of Points on a Globe to a Surface.....	59
3.19	Conversion of Coordinates from a Local System (e.g., u and v) to a Global System (e.g., x and y).....	62
3.20	Comparison of Converted Coordinates to Survey Measurements .....	64
3.21	Photographic Scale Relationships.....	69



## LIST OF FIGURES (cont.)

Figure		Page No.
3.22	(a) Typical Flight Pattern over a Site, (b) Endlap, and (c) Sidelap .....	71
3.23	Ground Target as Seen on Marcus Hook Orthophotograph.....	72
3.24	Photo Control Points Required for 4 Consecutive Photographs..	73
3.25	Image Displacement between Photographs as a Result of Terrain Relief.....	75
3.26	Marcus Hook Refinery Orthophotograph.....	76
3.27	Lube Plant Area CAD Drawings in ArcView .....	79
3.28	Tabular Database Data Levels .....	82
3.29	Entry Spreadsheets for Sample_Collection and Results Tables ..	84
3.30	TransferSpreadsheet Macro in Access .....	84
3.31	Marcus Hook Refinery Sampling Locations .....	89
3.32	Marcus Hook Refinery Groundwater Elevations on 8/17/98 .....	90
3.33	Marcus Hook Refinery Benzene Concentrations in Groundwater (1996).....	91
3.34	Marcus Hook Refinery Benzene Concentrations in Soil 2-8' Below Ground Surface (1996).....	92
3.35	Marcus Hook Refinery Highest Detected Concentrations in Groundwater .....	93
3.36	LNAPL Thickness Measurements in Lube Plant Area .....	94
4.1	<i>COC Transport</i> Methodology.....	98
4.2	<i>COC Transport</i> Program Steps Example .....	99
4.3	Regional Surface Water Runoff Model Results.....	103
4.4	Annual Runoff in the United States (1951-80).....	104
4.5	Interpol Script Methodology.....	105
4.6	Annual Runoff Grid for the United States.....	106
4.7	Annual Runoff Grid for Southeast Pennsylvania.....	107
4.8	Marcus Hook Facility Digital Elevation Model.....	111
4.9	Revised Marcus Hook Facility Digital Elevation Model.....	114
4.10	Facility Surface Water Runoff Model Results.....	115
4.11	Delineated Drainage Areas .....	116
4.12	Flow Calculation Points .....	118
4.13	Predicted Concentration Grid Using Point and Area Sources.....	121
4.14	Lube Plant Area Coverages in GMS .....	125
4.15	Site Conceptual Model in GMS .....	127
4.16	1-Layer Steady State Groundwater Model for the Lube Plant Area .....	130

## LIST OF FIGURES (cont.)

Figure		Page No.
4.17	Comparison of Computed Groundwater Elevations with Observed Groundwater Elevations in the Lube Plant Area.....	131
5.1	95% Upper Confidence Limit for a Data Set with 20 Samples...	135
5.2	UCL Sample Results Produced with Excel .....	138
5.3	Possible Solutions of the Binomial Distribution Method with a Sample Size of Twelve Trials.....	142
5.4	Beta Distribution for a Fraction of 0.85 and a Variance of 0.20 .	145
5.5	Uniform Distribution.....	145
5.6	Prior and Updated Beta Distributions for a Storage Tank Dike Area with Two Hits in Twelve Samples.....	147
5.7	Source Areas Chosen for UCL and Binomial Analyses.....	151
5.8	Source Areas Defined for Bayesian Analysis.....	159
5.9	Prior Grid of the Probability of No Hits in Twenty Samples .....	162
5.10	Updated Grid of the Probability of No Hits in Twenty Samples.....	163
5.11	Change in the Probability of No Hits in Twenty Samples for Each Source Area.....	164

## **CHAPTER ONE**

### **INTRODUCTION**

In today's environmental world, many different approaches exist for determining corrective action activities at a site with a history of chemical releases. These approaches typically include some form of risk assessment in which remediation decisions are made based on potential health risks to receptors. Risk assessments can effectively identify the threats posed by a particular facility, but the risk assessment process is typically very complex, costly, and time consuming. There is a need to develop a more unified corrective action process that efficiently utilizes time and resources while still establishing remediation alternatives that are protective of human health and the environment. Such a process would be particularly useful for large and complex sites that have many different chemical sources and transport pathways. Research is currently underway at The University of Texas at Austin (UT) to develop this type of process by using the Marcus Hook Refinery in Pennsylvania as a case study. The objective of this research is to develop an approach utilizing Geographic Information Systems (GIS) and a decision analysis framework as the basis for making risk-based decisions for corrective action. A complete description of the Marcus Hook project is beyond the scope of this document, and the focus will instead be on the use of GIS at this site and why it can be an effective tool. A GIS framework can serve as the foundation for all steps of the risk assessment process, including risk-based decision making.

#### **1.1 Environmental Risk Assessment**

Environmental risk assessment is the process of evaluating the risks associated with the presence and fate and transport of chemicals in the environment. More specifically, an environmental risk assessment is an analysis

of the potential for adverse effects caused by a chemical(s) of concern from a site to determine the need for remedial action or to develop target levels where remedial action is required (ASTM, 1998). It involves analyzing the sources of a release, the mechanisms of chemical transport, and the potential health risks to receptors. Figure 1.1 shows an example environmental scenario of a release from a storage tank, and Table 1.1 defines some of the key environmental risk assessment terms. From the figure, the storage tank is the source of chemical release, and the soil in the near vicinity of the tank is the source area. One of the transport mechanisms is leaching from soil to groundwater and then movement of the chemical with the groundwater. If a neighboring property owner has a drinking water well downgradient of the storage tank, then this landowner could be exposed to chemicals that have been released from the tank. If the chemical concentrations are high enough, the landowner is susceptible to various health risks. This example is just one possible scenario. Other scenarios could exist based on different combinations of sources, transport mechanisms, and receptors. Example sources at a chemical facility include process units, waste dumps, pipelines, and accidental spills. Some additional transport mechanisms include volatilization from soil, volatilization from groundwater, atmospheric deposition, and surface water runoff. Receptors can be either human or ecological, or both. Receptors are further broken down by potential exposure and toxicity risk. Some examples are children, adult workers, and neighborhood residents.

There are three main components of the environmental risk assessment process: exposure assessment, dose-response assessment, and risk characterization (Figure 1.2). Exposure assessment involves determining the concentration of a chemical of concern (COC) at a receptor based on a source concentration and then determining the actual exposure (i.e., dose) of the receptor to the chemical. An exposure assessment includes source characterization, transport assessment, and exposure analysis. Source characterization studies the

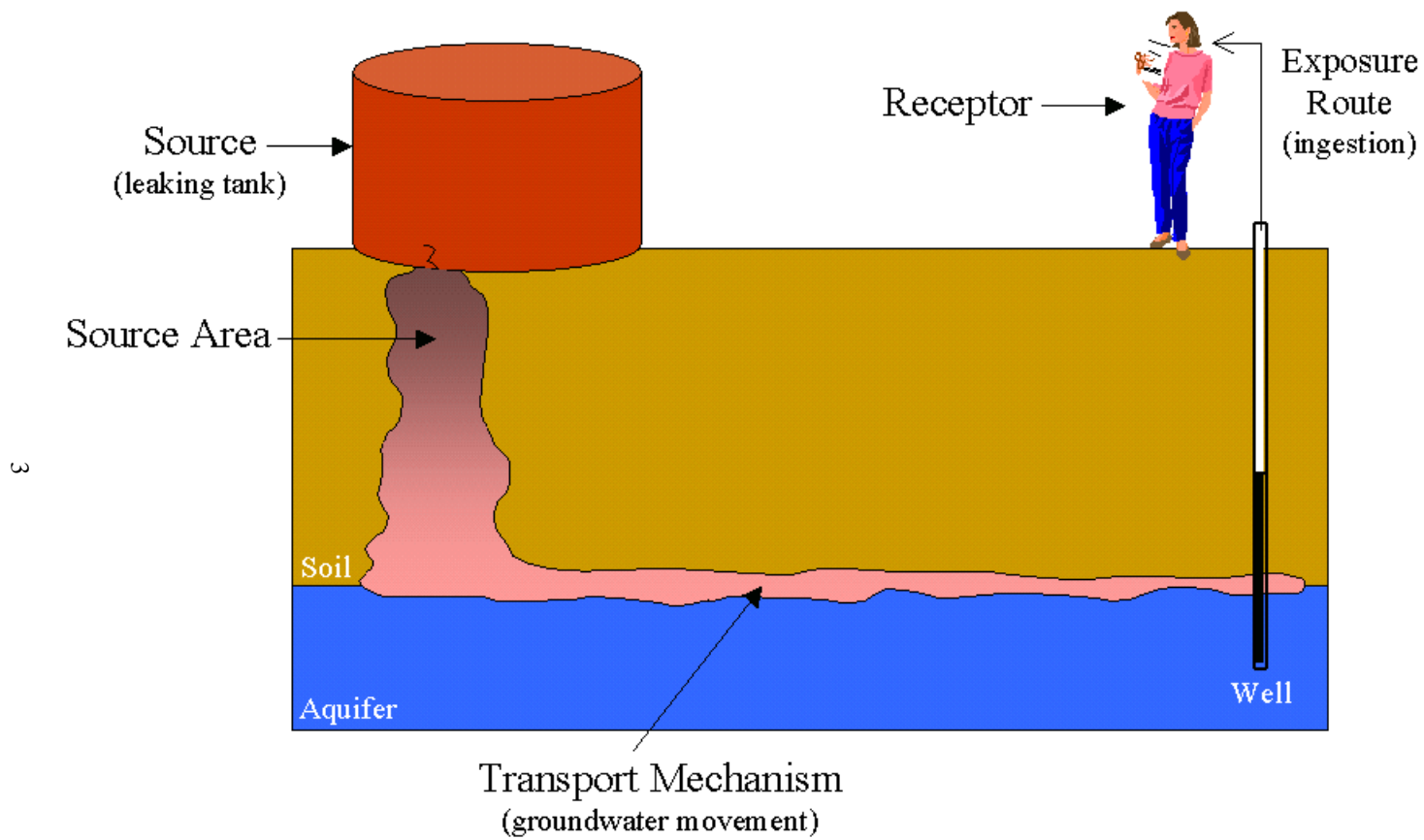


Figure 1.1: Example Environmental Risk Assessment Scenario.

**Table 1.1: Environmental Risk Assessment Terms**

Chemical(s) of Concern (COCs) <sup>1</sup>	Specific constituents and their breakdown products that are identified for evaluation in the risk assessment process.
Chemical Release <sup>1</sup>	Any spill or leak or detection of concentrations of chemical(s) of concern in environmental media.
Exposure <sup>2</sup>	Contact of an organism with chemical(s) of concern at the exchange boundaries (e.g., skin, lungs, liver) and the availability of the chemical of concern for absorption.
Exposure Route <sup>1</sup>	The manner in which a chemical(s) of concern comes in contact with an organism (e.g., ingestion, inhalation, dermal contact).
Natural Attenuation <sup>1</sup>	The reduction in the concentration(s) of chemicals of concern in environmental media due to a combination of one or more naturally occurring physical, chemical, and biological processes (e.g., diffusion, dispersion, absorption, and biodegradation).
Point(s) of Exposure <sup>1</sup>	The point(s) at which an individual or population may come in contact with a chemical(s) of concern originating from a site.
Receptors <sup>1</sup>	Persons that are or may be affected by a release. In the case of ecological receptors, this could include individual organisms, populations, communities, and ecosystems.
Site Conceptual Model <sup>1</sup>	The integrated representation of the physical and environmental context, the complete and potentially complete exposure pathways, and the likely distribution of chemical(s) of concern at a site.
Source	The physical structure or operating unit that may have caused or currently causes a release.
Source Area(s) <sup>1</sup>	The location of non-aqueous phase liquid (NAPL) chemical, the location of highest soil and groundwater concentrations of the chemical(s) of concern, or the location where the chemical(s) of concern was released.
Transport Mechanism	The combination of physical, chemical, and biological processes that move a chemical from the source to the point of exposure.

<sup>1</sup> From *Standard Provisional Guide for Risk-Based Corrective Action* (ASTM, 1998)

<sup>2</sup> From *Guidance Document for the Risk-Based Decision Process for Corrective Action Sites* (BP, 1997)

amount of a release, the rate of release, and the distribution in environmental media. Transport assessments analyze specific pathways, transport rates, and attenuation rates. Upon completion of these two steps, an exposure analysis is performed to determine the exposures for all affected or potentially affected receptors. These exposures are calculated based on the concentration of the COC in the media (e.g., air, water, or soil) and the receptor's amount of exposure to the COC (i.e., duration and frequency). Exposure assessment is the heart of environmental risk assessment. It identifies where the chemicals are, how they are moving, and in what amounts they are reaching specific receptors.

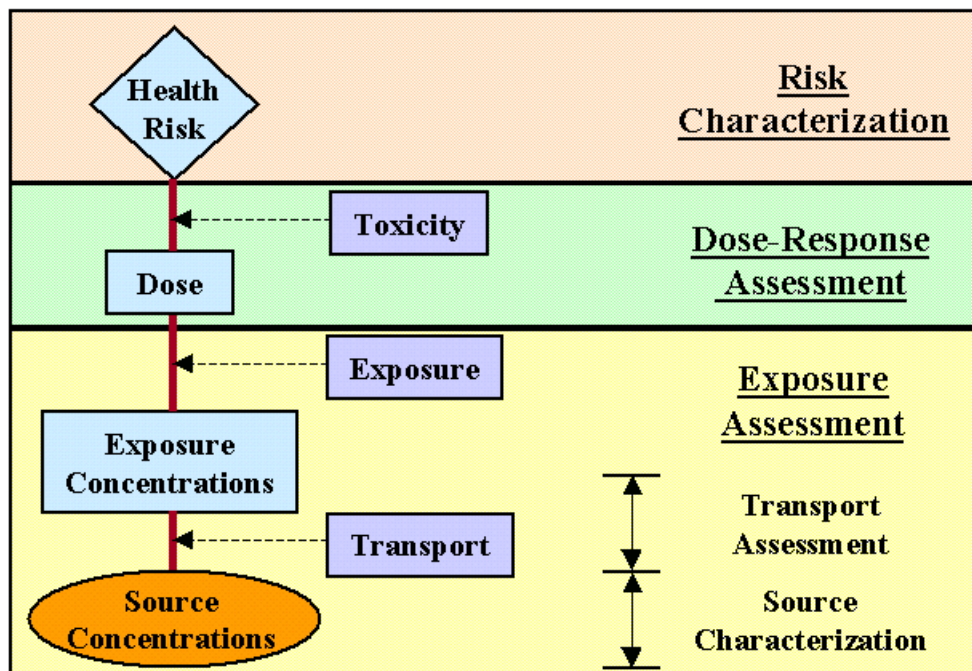
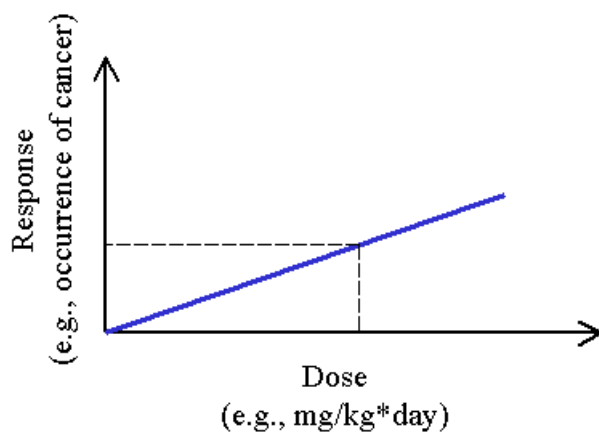


Figure 1.2: Environmental Risk Assessment Process Diagram (Maidment, 1998).

The dose-response assessment involves analyzing receptor responses to a COC based on different magnitudes of dose. This type of assessment usually incorporates dose-response curves, which measure the toxicity of a specific COC

(Figure 1.3). Significant uncertainty exists in dose-response calculations since there are a limited amount of data available on how different COCs affect the human body. As a result of this limited data availability, many dose-response curves are extrapolated from the results of animal studies. Besides an obvious difference between humans and experiment animals such as mice, these studies expose animals to doses that are much higher than the doses typically experienced by humans exposed to a COC in environmental media. Thus, a significant amount of uncertainty exists when extrapolating responses from small dose levels (Hay Wilson, 1998). A dose-response assessment usually does not vary from site to site. Therefore, the Environmental Protection Agency (EPA) and many state agencies have collected the results from dose-response studies to establish toxicity factors for different COCs. These agencies have also developed equations to calculate dose based on specific receptor characteristics.



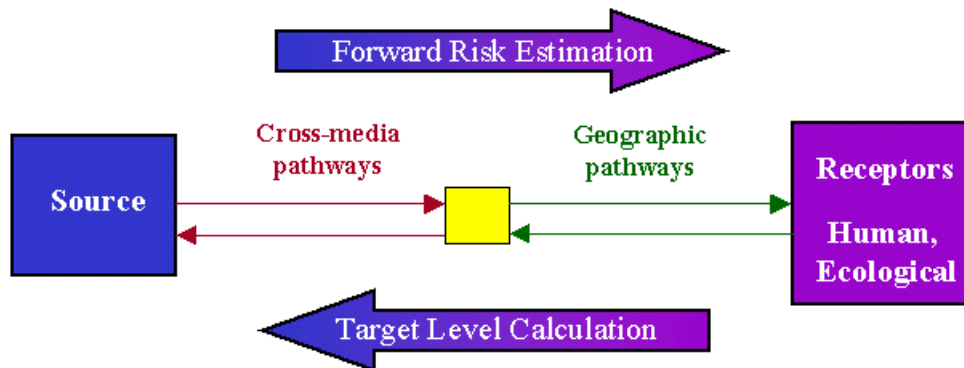
**Figure 1.3: Dose-Response Curve (Maidment, 1998).**

Once the COC concentration at a receptor location is known, and the toxicity criteria have been established, the risk to the receptor can be calculated based on dose and toxicity ( $\text{Risk} = \text{Dose} * \text{Toxicity}$ ). Risk is defined for carcinogenic



chemicals as the potential for incremental human carcinogenic effects, over background cancer occurrence levels, due to exposure to the chemical(s) of concern (ASTM, 1998). For example, if the background cancer risk is 0.25 (i.e., one out of four people are expected to be afflicted with cancer) and the excess risk level is set at  $10^{-6}$  (dimensionless), then the total acceptable risk is 0.250001. Allowable risk levels are set by the EPA and state agencies. There is still considerable debate over which risk level to use for carcinogenic chemicals:  $10^{-4}$ ,  $10^{-5}$ , or  $10^{-6}$ . A risk level of  $10^{-6}$  means that only 1 person out of 1 million in excess of the base line should become afflicted with cancer due to the specific dose of the COC. For non-carcinogenic chemicals, risk is evaluated using a hazard quotient, which is the ratio of the level of exposure of a chemical of concern over a specified time period to a reference dose for that chemical of concern derived for a similar exposure period (ASTM, 1998). The risk is acceptable as long as the hazard quotient is less than one.

There are two types of risk calculations: (1) forward risk estimation and (2) target level calculation (Figure 1.4). Forward risk estimation involves calculating an exposure concentration from a source concentration (exposure assessment) and then calculating the associated health risk for the receptor. Target level calculations involve determining an allowable source concentration based on an acceptable risk to the receptor. If the results of a forward risk estimation indicate that an area is out of compliance (i.e., the calculated risk is greater than the allowable risk), then a target level calculation is used to determine the amount of reduction required in the source area concentration.



**Figure 1.4: Risk Calculation Diagram (Hay Wilson, 1998).**

Environmental risk assessment is a fairly new process and is continually evolving. Some of its roots can be traced back to *Risk Assessment in the Federal Government: Managing the Process* (NAS, 1983). A significant advance then came when the Science Advisory Board (SAB) of the EPA (1992) released *Reducing Risk: Setting Priorities and Strategies for Environmental Protection*. This report created a new line of thinking of looking at things on the basis of their relative risk. Regulatory acts, such as the Resource Conservation and Recovery Act (RCRA), require eliminating generated waste by the best-demonstrated available technology and cleaning contaminated areas to the condition that existed prior to any chemical dumping or releases. While it is extremely difficult and costly to achieve such cleanup standards, it has also been realized that many contaminated areas never impact a receptor. At many sites, plumes of chemicals do not move far enough or fast enough to ever reach a receptor. Additionally, natural attenuation processes such as adsorption, dispersion, and biodegradation may significantly reduce concentrations from a source to a receptor. Therefore, remediation activities now focus on minimizing the potential risks to receptors. While the amount of cleanup and the cost are reduced through the use of risk assessment practices, the bottom line is still the protection of human health and the environment.

Environmental risk assessment is itself part of the larger corrective action process. The traditional corrective action approach as outlined in both RCRA and the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) is to perform some sort of remedial investigation and feasibility study, compare the results of the site investigation to generic standards, determine an appropriate remediation alternative, and then implement this remediation alternative (Figure 1.5a). The remedial investigation and feasibility study involve an assessment of the site and what remedial processes would be effective at the site. Any form of risk assessment usually follows the comparison to generic standards. The difficulty with this approach is that risk management decisions are not incorporated until everything about a site is known, thus resulting in a long protracted corrective action process that is very rigid and enforcement driven (Rocco, 1998). In this traditional approach, a risk assessment is often performed only to change the corrective action requirements determined by the remedial investigation and feasibility study. For instance, a risk assessment might show that an area does not need to be cleaned up to the generic standards because that area will continue to be used for industrial purposes.

In order to address this issue, the American Society for Testing and Materials (ASTM) has released the *Standard Provisional Guide for Risk-Based Corrective Action* (1998), which describes the Risk-Based Corrective Action (RBCA) process. RBCA focuses on addressing the questions that need to be answered by the investigation up front (i.e., what exposures, and therefore, what risks need to be characterized). RBCA incorporates a tiered risk assessment approach in the first phase of the corrective action process (Figure 1.5b). There are three tiers to the process, and each tier requires a more sophisticated analysis. Tier 1 compares exposures at a source area to conservative screening levels, which are set by either the EPA or state agency or based on calculations and policy decisions of the regulatory agency. If the exposure concentrations in the source area are below

the screening levels, then no further action is required. Tier 2 requires more site data as it incorporates exposure pathways using simplistic fate and transport models. Tier 3 requires the highest costs and greatest amount of site data, but it provides the least conservative cleanup alternatives since the uncertainty has been reduced through the site data collection and more sophisticated analyses. The benefits of a tiered approach include an effective utilization of resources and applicability to many sites. The process is also generic enough so that it can fit into the structures of varying regulatory programs.

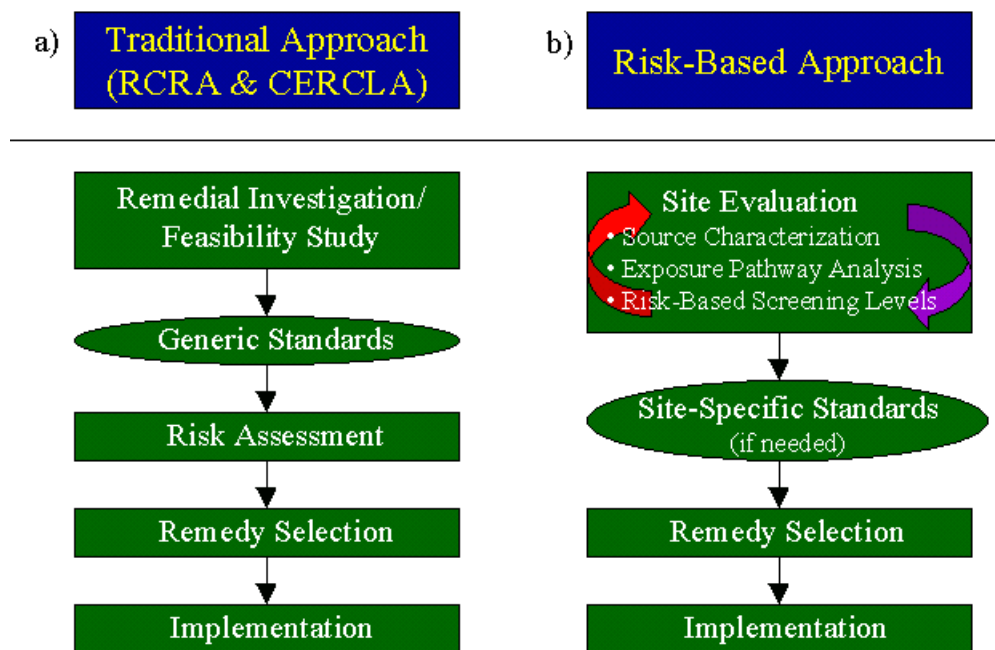
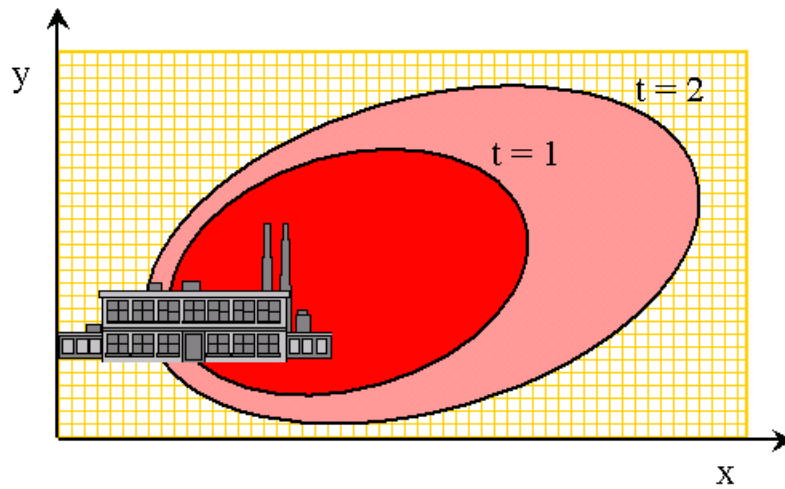


Figure 1.5: Corrective Action Approaches (Rocco, 1998).

## **1.2 Environmental Risk Assessment Using GIS**

Environmental risk assessment using GIS, or spatial environmental risk assessment, is identical to the environmental risk assessment process described above except that it utilizes a spatially-referenced framework to enhance the analysis and risk-based decision-making methods. Spatial risk assessment is the process of 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-referenced, integrated modeling environment (Hay-Wilson, 1998). The methodology is based on physical and hydrogeological properties of the study area and surrounding area, related to each other and to environmental measurements. The spatial and temporal relationships between features and measurements can be particularly useful in constructing a site conceptual model. For example, a map of chemical concentrations and process units might indicate the possible sources of release. Combining the source analysis with groundwater level measurements could indicate potential downgradient receptors. Temporal data attributes are also included in the spatial risk assessment methodology so that, for instance, concentrations can be mapped in both space and time (Figure 1.6). This methodology has an obvious benefit at large and complex sites that have many sources and receptors because the focus of a spatial risk assessment is on the location of features and measurements. The idea is not to focus only on individual areas of a site, but to incorporate the relationships between many areas and view the site in a holistic or site-wide context.

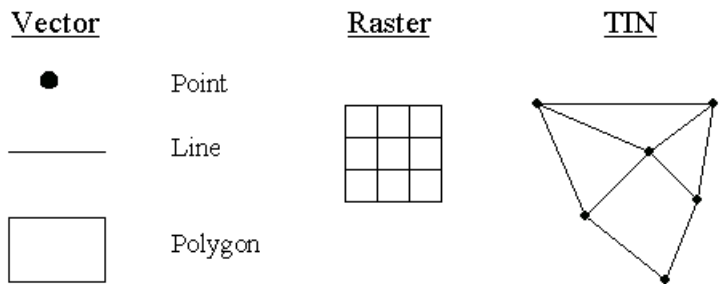


**Figure 1.6: Mapping Concentrations in Space and Time (Maidment, 1998).**

An effective way to construct this spatially-referenced framework is through the use of GIS, which uses a computer system to geographically reference the features of a particular area. The area could be an individual facility, a city, a county, a state, a country, a continent, or even the world. Features could include roads, rivers, buildings, land use, geologic properties, or whatever is of importance for analyzing the study area. GIS is a mapping environment where individual coverages, or themes, can be displayed or removed depending on the intended application. Each coverage is linked to an attribute table so that information is available on the individual features, or records, of the theme. For instance, an applicable coverage for a petroleum refinery is the on-site storage tanks. For each tank, there might be data available on what products are currently stored there and what products have been stored there in the past. Besides this obvious data-mapping advantage, GIS also has the ability to query specific information about a theme. So in the example above, all the tanks that store gasoline could be identified. Geographic Information Systems are predominantly

used for mapping and visual display, but analysis and modeling functions are increasingly being explored.

Three separate data models are supported by GIS: (1) vector data, (2) raster data, and (3) triangulated irregular networks (TINs) (Figure 1.7). Vector data includes feature representation with points, lines, or polygons. For example, the monitoring wells for a site could be mapped as a point data source. Example line features include rivers, roads, and boundaries. Some polygon feature examples are buildings, lakes, and watersheds. While vector data are the most common format, other data sets are better represented with grids, where each cell in the grid has a particular value. This type of format is referred to as raster data and is effective for representing elevations and concentrations. Triangulated irregular networks are the final type of data model and are particularly useful for surface representation and three-dimensional mapping. TINs are constructed by connecting a group of points, such as surveyed elevations. The lines that connect these points form triangles, and since each point in the TIN has an associated value, each triangle in the model (i.e., continuous surface of planar triangles) is sloped. This allows for powerful visualization capabilities with a three-dimensional viewer. The most common method of connecting points to form a TIN model is Delauney triangulation, which maximizes the minimum interior angles of the triangles formed, thereby avoiding long and thin triangles (Jones, *et al.*, 1990).

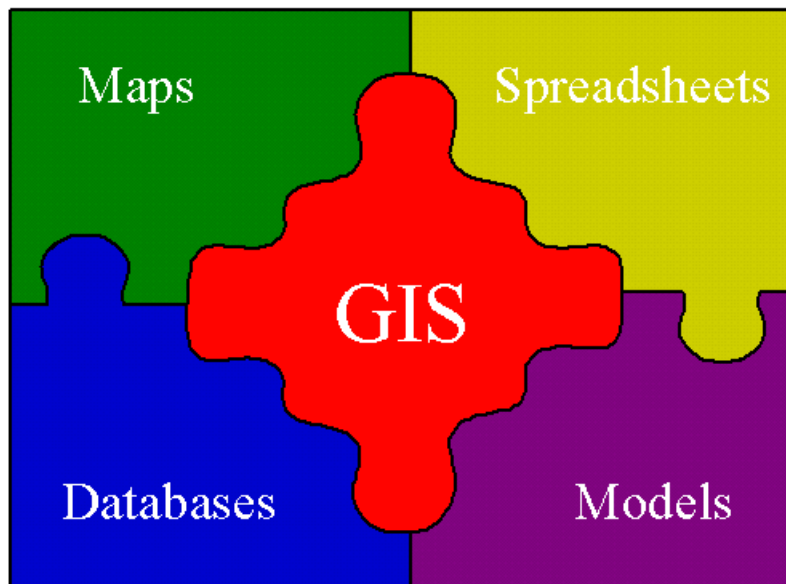


**Figure 1.7: GIS Data Models.**

Any one of the commercially available GIS software packages can be used in the spatial risk assessment process. However, this research has utilized the Environmental Systems Research Institute, Inc. (ESRI) products: ArcView GIS Version 3.1 and ARC/INFO Version 7.2.1. These software packages have been chosen based on their ease of use and worldwide availability. ArcView's graphical interface allows a user to display spatial data, build maps, query data sets, create charts, and perform calculations. ARC/INFO is a more powerful set of programs that is useful for data editing and analysis. ARC/INFO has more functionality than ArcView but is also more expensive (7 or 8 times as much). For purposes of a spatial environmental risk assessment, ArcView is the most effective software tool. Its analytical capabilities have improved significantly over the past few years, and unless otherwise noted, all the methods discussed in this document can be performed with ArcView.

One of the advantages of a GIS such as ArcView is its ability to connect with many different applications in a PC-based environment (Figure 1.8). Each component of the puzzle in the figure is a tool for information processing. Maps can be used to convey geographic features and relationships along with the results of a data analysis. Databases can be used to store physical and chemical data along with geologic and hydrogeologic information. Example spreadsheet uses include concentration data analysis and simple transport algorithms while models utilize sophisticated transport simulations in order to characterize chemical attenuation and migration in the environment. These information-processing tools can be either internal or external to the GIS application. External programs are often dynamically linked to the GIS application so that when information is updated in one location, it is updated throughout the system. The ability to utilize this type of approach has only recently become available with the release of more cost-effective personal computers and more integrated software applications.





**Figure 1.8: GIS Application in PC-Based Environment (Hay Wilson, 1998).**

Spatial environmental risk assessment is a three-step process as shown in Figure 1.9, where each level of the pyramid builds on the previous one. The process begins with the development of spatial and tabular databases. Both on-site and regional data are collected in order to acquire an accurate description of the important site and surrounding area characteristics. This digital facility description is an integral element to the construction of a site conceptual model. The next step is to spatially characterize chemical migration and attenuation in the environment. This process is referred to as map-based modeling, and it further enhances the understanding of the site characteristics. Exposures can then be analyzed by combining map-based models with the digital facility description. Sources and receptors are established with the aid of the digital facility description, and the change in concentration from the source to the receptor is predicted with a map-based model. An exposure assessment incorporating the exposure concentration and the receptor intake then determines whether any

corrective action will be required and/or whether site-specific standards may be necessary.

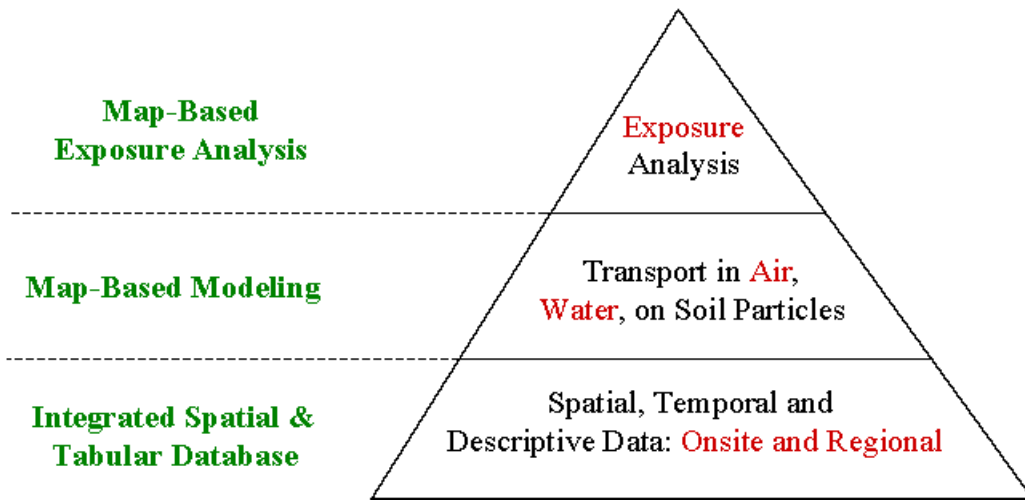


Figure 1.9: Spatial Environmental Risk Assessment Process (Hay Wilson, 1998).

### 1.3 Project Background

In order to develop the methodology and to determine its effectiveness, the spatial environmental risk assessment process is being applied to a real world site that requires corrective actions. The Marcus Hook Refinery is a facility located in Marcus Hook, Pennsylvania, adjacent to the Delaware River (Figure 1.10). The facility has operated as a crude oil refinery and petroleum products terminal since 1900, and it was owned by British Petroleum Exploration & Oil, Inc. (BP) from the 1960's until 1996 when BP sold the facility. BP is responsible for the environmental conditions that existed at the time of sale and the remedial action required based on those conditions. Marcus Hook has many characteristics that make it an excellent case study site. The facility is very big (approximately 350 acres) and old, meaning potentially there have been many releases of many different chemicals. There are, therefore, a lot of potential sources, both past and

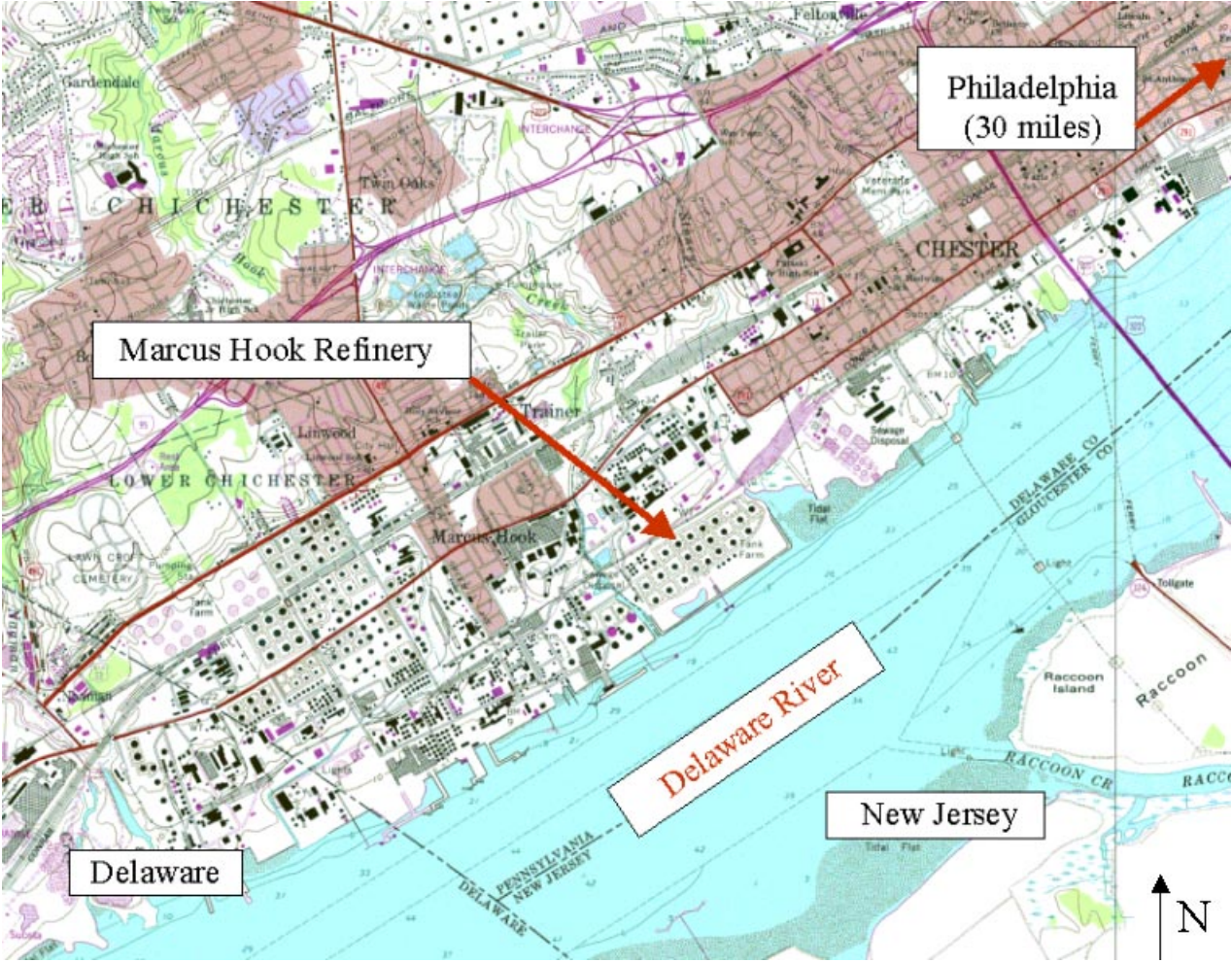


Figure 1.10: Case Study Area.

present, and the location of many of these sources is unknown. There are also many potential receptors. Across the northern boundary of the facility lies a residential neighborhood with a school and baseball field. On the western boundary of the site lie a church and a park. Ecological receptors are also an important consideration at the facility because of on-site wetlands, the Delaware River, and two creeks, Marcus Hook Creek and Stony Run Creek, that wind through the facility. Many different transport mechanisms exist as well including leaching to groundwater, groundwater movement, and discharges to surface water. These characteristics provide an interesting set of challenges to the environmental risk assessment process.

As previously mentioned, the objective of the project is to develop an approach utilizing GIS and a decision analysis framework as the basis for making risk-based decisions for corrective action. The GIS application provides a common spatial framework for the assembly of descriptive data about the facility and the surrounding area. This assembly of descriptive data is referred to as the digital facility description, which has two primary components: (1) a tabular database and (2) a spatial database. The tabular database contains information about environmental measurements at the site, such as concentration data, groundwater levels, and soil profiles. The spatial database operates within the GIS application and describes the facility and regional features along with attributes of these features. Facility features include structures, storage tanks, pipelines, surface cover, and surface hydrology. Regional features include land use, census tracts, soils data, and climate data. The digital facility description is the foundation of the spatial environmental risk assessment process as it provides the information required to evaluate all the following elements of a risk-based approach:

- spatial and temporal characterization of source areas,
- chemical fate and transport analysis,

- receptor identification,
- site conceptual model formulation and exposure pathway analysis,
- risk characterization,
- risk communication, and
- decision analysis framework implementation.

The decision analysis framework provides a rational and defensible basis for the major technical decisions to be made in the corrective action process. It tracks the potential alternatives and outcomes along with the relative importance assigned to these alternatives and outcomes. The decision analysis framework can also be used to determine the value of collecting additional information so that the highest benefit to cost ratio is achieved. Before any additional data are collected at the site, the data from previous studies are collected and organized to obtain as much information as possible about existing site conditions. An evaluation of the data already collected helps identify the value of collecting additional data, including what type of data to collect and where to collect it.

Another important aspect of this project is risk communication (Figure 1.11). Many parties have an interest in the project development and results including BP, the Pennsylvania Department of Environmental Protection (PADEP), EPA Region III, the current property owner, and the Marcus Hook community. To accomplish this task, the visualization capabilities of the GIS application are utilized so that both the technical and non-technical audiences can better understand the results of the data analysis. Some of the communication activities include documenting the new methods in exercises, providing training sessions for the regulatory agencies and eventually for the public, and submitting CD-ROM's of progress and data. The CD-ROM's are submitted in replacement of paper reports. Instead of having to flip through a report and read about someone else's interpretation of the data, the data are presented in a digital format so that a user can perform an analysis. The CD-ROM's include the documents, data displays, animations, and maps that are developed as the risk assessment proceeds

so that information is presented in a clear format. However, the user can create different maps and run different queries to study the material that is of interest to him or her. The goal is to easily allow anyone the opportunity to become involved in assessing conditions at the refinery.

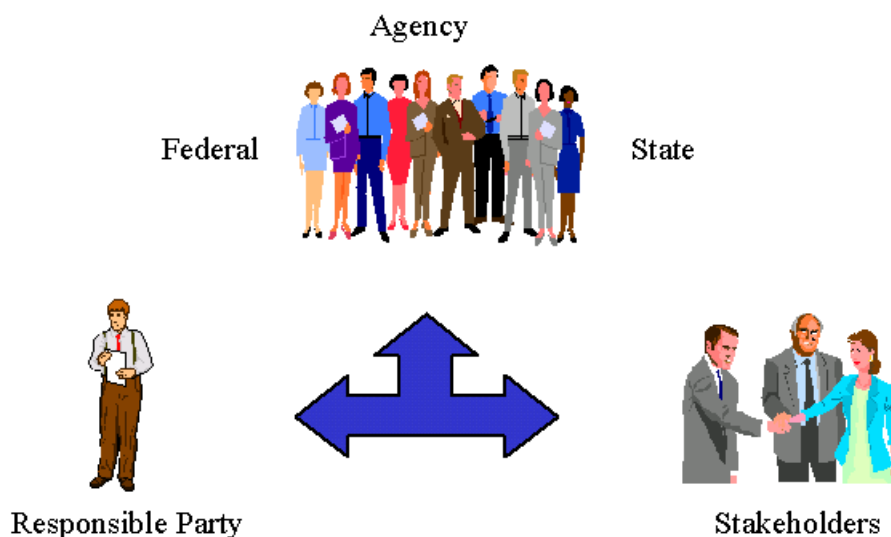


Figure 1.11: Risk Communication Diagram (Maidment, 1998).

#### 1.4 Objective

The goal of this research is to demonstrate the development of a digital facility description and its use as an effective environmental risk assessment tool that can be applied at any site. As a case study example, the Marcus Hook Refinery was used for this development and application. Particular emphasis was placed on all the necessary considerations that are incorporated into the construction of a digital facility description and the specific components which are the most important to the process. This overall goal is achieved by the following specific objectives:

1. Construct a spatial database of important facility and regional features.

2. Construct a tabular database of environmental measurements at the facility.
3. Establish a dynamic link between the spatial and tabular databases in the GIS application.
4. Connect the digital facility description with models so as to analyze transport mechanisms.
5. Analyze chemical concentrations in source areas for the development of "risk maps" using the digital facility description.

### **1.5 Organization of Thesis**

The objectives just described outline the major components of this thesis, which is organized into six chapters. An introduction to the project and environmental risk assessment process is provided in Chapter 1. The thesis also includes a literature review (Chapter 2) to demonstrate the uniqueness of the GIS aspect of this project. Information on how the digital facility description was constructed for the Marcus Hook site and how a digital facility description can be constructed for any site is discussed in Chapter 3. Some example applications of how this description provides the foundation for the spatial environmental risk assessment process are outlined in the following two chapters. These example applications include map-based transport models (Chapter 4), which have been developed for surface water runoff and groundwater flow, and some source area exposure assessments using risk maps (Chapter 5). Finally, the conclusions of this research are presented in Chapter 6. Information is available on the other project aspects such as the site conceptual model and decision analysis framework from the References section. Because all of the work towards the development of a spatial risk assessment methodology has been applied to the Marcus Hook facility, results for this facility are shown in this thesis. However, the overall goal of the research is to develop a process that can be applied to any site with a history of chemical releases.

## **CHAPTER TWO**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

The purpose of this chapter is to discuss some of the previous uses of GIS in the area of environmental risk assessment and to demonstrate the uniqueness of this study. Very few concrete examples have been published about the use of GIS for environmental risk assessment, and of the existing examples, most are limited in scope. There are, however, a wide variety of studies that address the individual aspects of the spatial environmental risk assessment process. These aspects include database development, hydrologic modeling, and statistical analysis of data measurements. Thus, the methods and concepts discussed in chapters 3-5 are not limited to this study, but the specific application of integrating all components for the purposes of a risk assessment at a large facility is unique. None of the reviewed studies demonstrated the use of GIS as the focal point of a framework that is technically defensible and effective for making risk-based decisions in all aspects of an industrial corrective action project. Similarly, no studies were found on the development of a digital facility description.

The previous studies of using GIS in activities related to environmental risk assessment can be broken down into two categories: (1) for cartographic purposes, and (2) for modeling purposes. The first category involves constructing a database of appropriate coverages and then overlaying these coverages for decision-making applications, such as planning and management. The second category extends the cartography application by utilizing data in the compiled database for constructing models. These models can be either internal or external to the GIS application, but the focus is on using a spatial database and GIS for processing model input and output. Both of the above categories have important roles in the spatial risk assessment hierarchy. The cartography (i.e., maps of a



study area) displays important characteristics of an area and how these characteristics might be related. These characteristics play an important role in describing information about sources and receptors. The modeling then provides the mechanism to analyze how chemicals migrate and attenuate from a source to a receptor. The following sections provide a brief summary and outline of the reviewed studies in these two areas and in the specific area of risk assessment.

## **2.2 GIS and Cartography**

Because of GIS's spatial storage and display capabilities, numerous examples exist of using maps for making environmental decisions. *Managing Natural Resources with GIS* (Lang, 1998) provides a good overview of some environmental mapping case studies. These studies involve using GIS in the all of the following areas: oil and gas exploration, agriculture, deforestation, air pollution, water quality, brownfields reclamation, coastal protection, and disaster planning. For instance, Chevron is using GIS to determine oil extraction points that are protective of existing resources. The New Jersey Department of Environmental Protection (NJDEP) is using GIS to overlay sampling locations with points of known source pollution and land use for assessing water quality in the state's major river basins. All of the studies that are described involve building a database of relevant data and then using built-in GIS functions such as overlaying and querying for further analysis and decision-making. As Lang (1998) describes, "By putting their spatial data in an integrated system where it can be organized, analyzed, and mapped, they (the users) find patterns that were previously unrecognized."

Other mapping studies have been more focussed on environmental impact assessments. For instance, a few projects have involved using GIS for selecting sites for the land application of waste (Hendrix and Buckley, 1992; Lober, 1995; Kao, *et al.*, 1996). Hendrix and Buckley (1992) used grid-based calculations, also

referred to as map algebra, incorporating soil, topographic, and land use factors with information about biological, chemical, and physical properties of waste to select appropriate disposal sites. Lober (1995) used a similar methodology, but focussed more of the attention on social factors, such as public opposition. Another project (Yi, 1994) using GIS involved building a database of wetlands in Ohio to aid in natural resources policy development, land use planning, and wetlands management. Similarly Aspinall and Pearson (1996) built a database of ecological species in Scotland for analyzing distribution patterns. Other studies include mapping Lyme Disease risks areas (Glass, *et al.*, 1995), identifying the susceptibility of real estate properties to impacts from others (Hill, 1996), and cataloguing hazardous materials handling sites for hazard assessments in flood prone areas (Harris, 1997). The study by Harris (1997) is interesting because it incorporates exposure pathways based on surface flow and COC profiles at each site. However, the exposure pathways in the associated public health assessment (PHA) are determined simply from the mapped data. All of these examples essentially involve data pattern characterization and do not include any modeling of environmental processes.

### **2.3 GIS and Modeling**

There are a large number of studies that have focused on linking GIS with environmental models. This link between GIS and models is not surprising because, as David Maidment (1996) describes, “Such simulations require data about the environment within which the processes occur, and simulation results provide additional data to enrich environmental description.” The capabilities of GIS allow for an intrinsic connection between preparing data for a model and viewing the model results. Watkins, *et al.* (1996) nicely reviews GIS model setup and some previous examples utilizing GIS for modeling. The article focuses specifically on groundwater connections, but it contains concepts related to all

types of models. For instance, three basic model designs are described: (1) linked GIS models, (2) integrated GIS models, and (3) models embedded in GIS. The first category refers to models that transfer data between programs. Some examples include ARCMOD and the Groundwater Modeling System (GMS), both of which are external to the GIS application. The second category refers to models that are linked through a common interface and share a single database. One example is MODFLOWARC, in which MODFLOW is integrated with ARC/INFO. The final category refers to models that are embedded within the GIS application and utilize built-in capabilities, such as map algebra. The modeling for the Marcus Hook facility, as described in Chapter 4, utilized an embedded model for surface water runoff and a linked model for groundwater flow.

The previous GIS modeling studies vary considerably in scope and objective. Two projects have used a GIS-model system for delineating wellhead protection areas (WHPAs) (Rifai, *et al.*, 1993; Baker *et al.*, 1993). These models utilize the hydrologic parameters assigned or averaged for each well point in the delineation process. They do not incorporate an entire subsurface groundwater model. Another project (Camp and Brown, 1993) utilized GIS for developing a three-dimensional subsurface profile. Arc Macro Language (AML) programs were used to convert well-log information into a profile, to view cross-section diagrams, and to prepare input files for MODFLOW. While fairly interesting, this study does not provide any information about the AMLs or MODFLOW connection. Another interesting project (Ross and Tara, 1993) tied an integrated surface and groundwater hydrologic model to GIS. Coverages of land use, slope, and soil type are used for deriving data such as infiltration. An interface is then utilized to run a set of models (e.g., MODFLOW, the EPA Hydrologic Simulation Program-Fortran (HSPF)). Some of the output examples include water table hydrographs, storm runoff hydrographs, cumulative flows, discharge peaks, and

runoff rates. This link between GIS and hydrologic models is commonly reported in the literature. Shea, *et al.* (1993) used such a model to develop a surface water management plan for a county in Florida. Many research projects at The University of Texas at Austin have also focused on using GIS for analyzing surface water hydraulics (Olivera, *et al.*, 1996; Hellweger and Maidment, 1997). An ArcView extension, *CRWR-PREPRO*, has already been developed that links GIS with the Hydrologic Engineering Center (HEC) models, and the first few steps of this program were used for delineating surface water drainage patterns at the Marcus Hook facility.

Other GIS modeling examples include:

- delineating groundwater flow (McKinney and Tsai, 1996; El-Kadi, 1994),
- managing stream-aquifer interactions with a decision support system (Fredericks, *et al.*, 1998),
- determining groundwater vulnerability to contamination (Sunday, 1996; Evans and Maidment, 1995),
- predicting water quality (Benaman, *et al.*, 1996; DePinto, *et al.*, 1996),
- estimating fugitive emissions from sewer networks (How, *et al.*, 1998), and
- assessing non-point source pollution impacts (Corwin, *et al.*, 1998; Poiani, 1995; Jankowski and Haddock, 1996; Saunders and Maidment, 1996; Quenzer and Maidment, 1998).

Most of these examples could be utilized for purposes of a spatial environmental risk assessment to model hydraulics and/or chemical migration from a source to a receptor. For example, the *COC Transport* extension described in Chapter 4 is based on the work of Ann Quenzer (1998), whose work was preceded by Bill Saunders (1996). However, these models are only one component of the spatial risk assessment process, and the focus of this thesis is on how to build the foundation for spatial environmental risk assessment using an integrated framework. It is not the intention of the author to advocate any particular model for incorporation into this framework. The intention is instead,

to outline some of the numerous possibilities. This open architecture format allows for an incorporation of any model to fit site-specific needs and an incorporation of any more developed model, as it becomes available.

## **2.4 GIS and Risk Assessment**

Studies related to the use of GIS for risk assessment are not limited to cartographic and modeling purposes, as a few studies have focused on using GIS as a primary component of the process. Chen, *et al.* (1998) describe linking GIS with a groundwater model and decision support system for the purposes of making decisions about waste management, pollution control, site remediation, and impact assessment. The authors used fuzzy set theory as the basis for analyzing risks. No discussion is provided on the development of a site conceptual model or the evaluation of any exposure pathways other than groundwater ingestion. There is also no discussion on analyzing sources and receptors. Another project (Miller, *et al.*, 1996) conducted at Oak Ridge Laboratory focused on using GIS to calculate human health risks at a large military facility. This study was oriented towards demonstrating how risks can be effectively displayed on a map. The authors do not consider the geographic relationships between sources and receptors, and they eliminate any discussion of exposure pathways. A discussion is provided on extrapolating a risk surface for groundwater concentrations, but this surface seems to have been created as if the groundwater flow and transport regime was a statistical function instead of a physical process governed by the advection-dispersion equation (Hay Wilson, 1998). Another project involved assessing environmental impact and natural resources damage to aquatic environments from oil and chemical spills (French and Reed, 1996). This project was geared specifically towards satisfying a CERCLA Type A natural resource damage assessment using a computer model.

However, the role of the GIS application and actual model setup are not discussed.

Thus, the research described in this document is unique because it focuses on using GIS as the central component of a detailed spatial environmental risk assessment at a large industrial facility. While the site conceptual model was still being developed at the time this thesis was published, all of the work presented here is aimed at the construction of this model in a spatial framework so that geographic relationships can be established between multiple sources and receptors. Utilizing these geographic relationships, exposure pathways can be evaluated with fate and transport models. This representation of an integrated environmental system within GIS for the purposes of making risk-based decisions has not been developed before.

## **CHAPTER THREE**

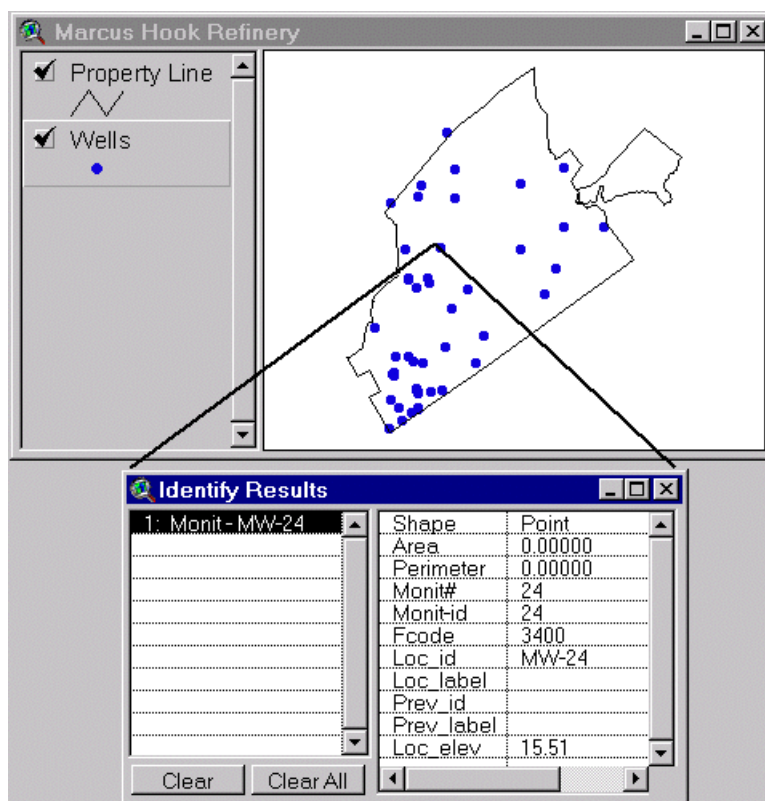
### **DIGITAL FACILITY DESCRIPTION**

#### **3.1 Introduction**

The digital facility description refers to a collection of information pertaining to the characteristics of a facility and its surrounding area that has been compiled within a computer environment. More specifically, the digital facility description is the collection of information describing the physical, geological, hydrologic, and chemical characteristics of a facility and region that has been spatially referenced in a Geographic Information System. Some of the features that could comprise a digital facility description include topography, geology, hydrogeology, environmental sampling locations, current facility features, historical facility features, and off-property features such as land use. The features that comprise a digital facility description are chosen based on the intended application of the description, which, in this case, is to assess the environmental risks posed by a particular facility.

The two major components of the digital facility description are the spatial database and the tabular database. The spatial database contains the facility-specific and regional features along with the attributes of these features. The tabular database is a relational database that contains the environmental measurements that have been collected at the facility. These databases are separate entities and can be analyzed as such, but through the use of built-in PC functionality, they can be dynamically connected within the GIS application. More specifically, data can be imported from a database program into a GIS such as ArcView (Section 3.5). The question then arises: why not combine the two databases into a single database so that there is never a need to import data? The need for two databases results from the required level of detail with environmental measurements and the limitations of the GIS software. In the GIS

environment, each sampling location is a single point on the map that has an associated record in the attribute table as shown in Figure 3.1. However, there could be a vast amount of data associated with that point. For example, if the point is a monitoring well, several different groundwater level measurements and concentration measurements could have been obtained there. Because of the many different types of data and because of temporal variations in each type of data, the most effective form of information management is with a relational database, which minimizes the repetition of records. The current capabilities of ArcView do not provide an efficient way to link multiple tables since each feature can have only one associated record. The solution here is to use a database software package and import selected sets of data into the GIS application.



**Figure 3.1: Features on a Map Related to Records in a Table.**



This chapter documents the first level of analysis for a spatial risk assessment, that is, the development of a digital facility description. The next two chapters demonstrate some of the uses of this description in the next two levels of the spatial risk assessment process: map-based modeling (Chapter 4) and exposure analysis (Chapter 5).

### **3.2 Spatial Database**

The spatial database is the description of the regional and facility-specific features. These features are topologically referenced so that the real world layout and distance between features is preserved. The coverages that comprise the spatial database include topography, geology, hydrogeology, surface waters, historical facility features, current facility features, and off-property features such as surrounding land use. A spatial database incorporates both surface and subsurface representation of the site. The required features depend on the data needs of the environmental risk assessment process (Section 3.4.2). The spatial database must therefore, provide the mechanism to analyze the data that are assembled in the tabular database and to support and communicate the results of the risk-based decision process.

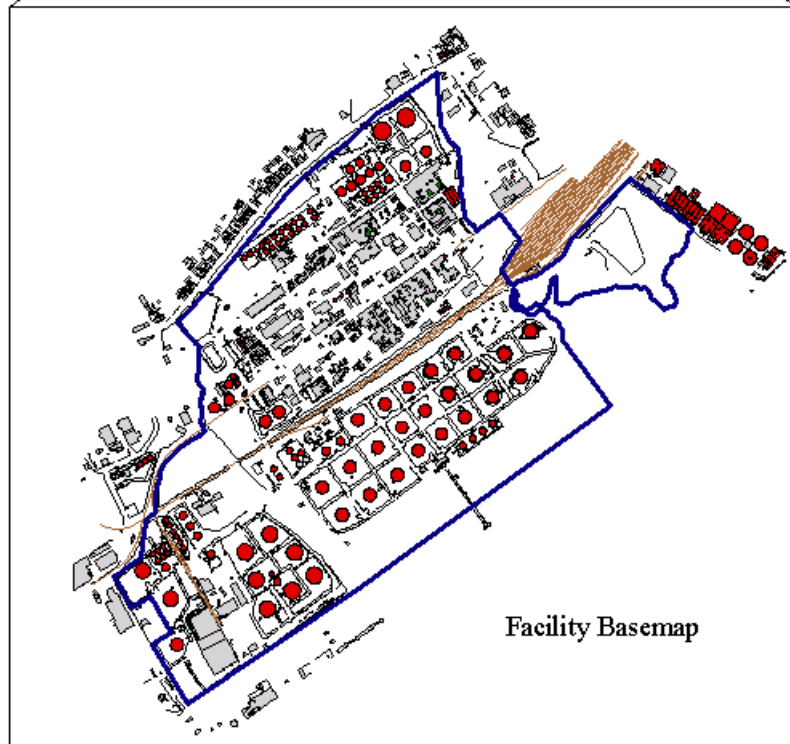
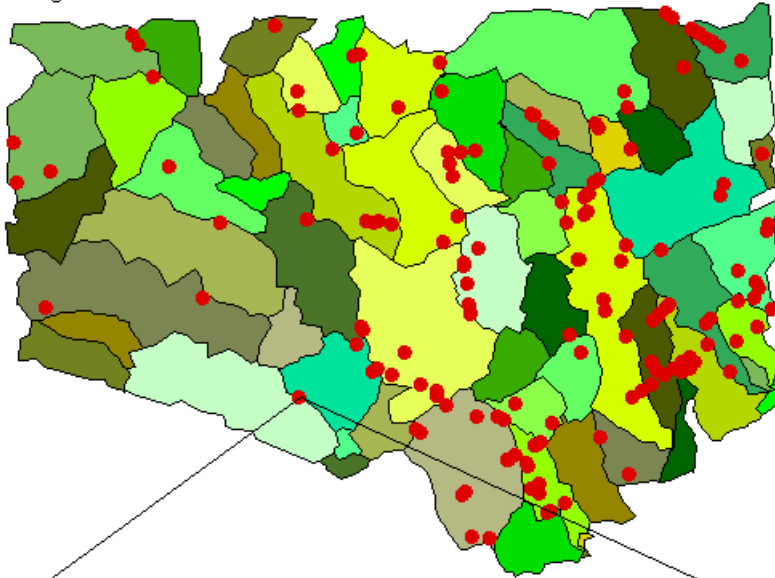
The spatial database has two components: (1) the regional coverages and (2) the facility-specific coverages. Regional features indicate the important characteristics of the area surrounding the facility while facility-specific coverages describe the actual characteristics of the site. Both types of features can be used in the analysis process and in the development of a spatial site conceptual model. All of the spatial coverages, regional and facility-specific, are referenced to the same coordinate system so that these coverages can be overlaid together on a single map (Section 3.4.1). However, a level of scale separates the regional and facility-specific coverages. In the regional model, the facility is just a single point, but on the facility scale, this point can be described by hundreds of

other points, lines, and polygons (Figure 3.2). The regional and facility-specific coverages are defined here as two separate entities, but they can be analyzed together in GIS. Switching from one scale to another in ArcView is accomplished with simple zoom in and zoom out tools.

The first step in the construction of a spatial database is to acquire the regional features. The easiest method of compiling a regional database is to download coverages from the Internet. The expansion of the Internet has led to a large increase in the amount of free spatial data available. The United States Geological Survey (USGS) and EPA have led much of the effort to make ARC/INFO formatted coverages available to anyone. Additionally, state agencies, universities, and private organizations have published data from their projects on the web. The only downfall of the Internet is that it takes a considerable amount of time and effort to search for the vast amounts of spatial data. In order to simplify this process, Appendix A contains a list of useful websites for purposes of regional data collection. It contains addresses for all the sites used for the Marcus Hook regional data collection along with some other generic sites, such as where to find spatial water resources data in any state. Most of these sites contain ARC/INFO coverages in export format (\*.e00). An ARC/INFO export file is an interchange file for use between systems that avoids transferring multiple files for one coverage. An ARC/INFO export file can be imported for use in ArcView or ARC/INFO with the **import** command in ARC/INFO or the **import71** command in ArcView. Some websites contain data sets in a format unrecognized by ARC/INFO. However, these data sets can usually be converted into the desired format, and Appendix E discusses some specific conversion steps.

Thirty regional coverages were compiled for the Marcus Hook project, and these coverages are listed in Table 3.1. This table also identifies the coverages that are the most critical to the spatial environmental risk assessment process.

Regional Watersheds



Facility Basemap

Figure 3.2: Relationship between Regional and Facility Features.

Figures 3.3 - 3.8 display some of these coverages in the ArcView mapping environment. Each figure provides insight into what characteristics can be expected at the facility. For instance, the physiographic regions of Pennsylvania are shown in Figure 3.3. Marcus Hook is located in the coastal plain region, meaning that the geologic conditions of the site are strongly dependent on the features of the Delaware River basin. Figure 3.4 displays the census tracts for Pennsylvania broken down by number of people. This coverage is useful for organizing receptor characteristics because it contains information on the population within each tract, such as number of males, number of females, and number between specified age ranges. The Delaware County streams and surface geology are shown in Figure 3.5. Besides the powerful visual effect of how streams line up with geologic characteristics, this figure provides further insight into the site geological conditions. Other land characteristics such as land use and elevation are also relevant to understanding the site in a regional context. Figure 3.6 shows the USGS land use classifications for southeast Pennsylvania, and Figure 3.7 shows the topographic relief for the same region. The regional climate characteristics can be read from the Vegetation/Ecosystem Modeling and Analysis Project (VEMAP) coverage shown in Figure 3.8. VEMAP is an ongoing project studying the response of biogeography and biogeochemistry to spatial and temporal variability in climate. The compiled data sets are available over the Internet and contain information such as precipitation, temperature, radiation, and wind speed.













**Table 3.1: Marcus Hook Regional Coverages** (refer to Spatial Data Dictionary in Appendix B for more information).

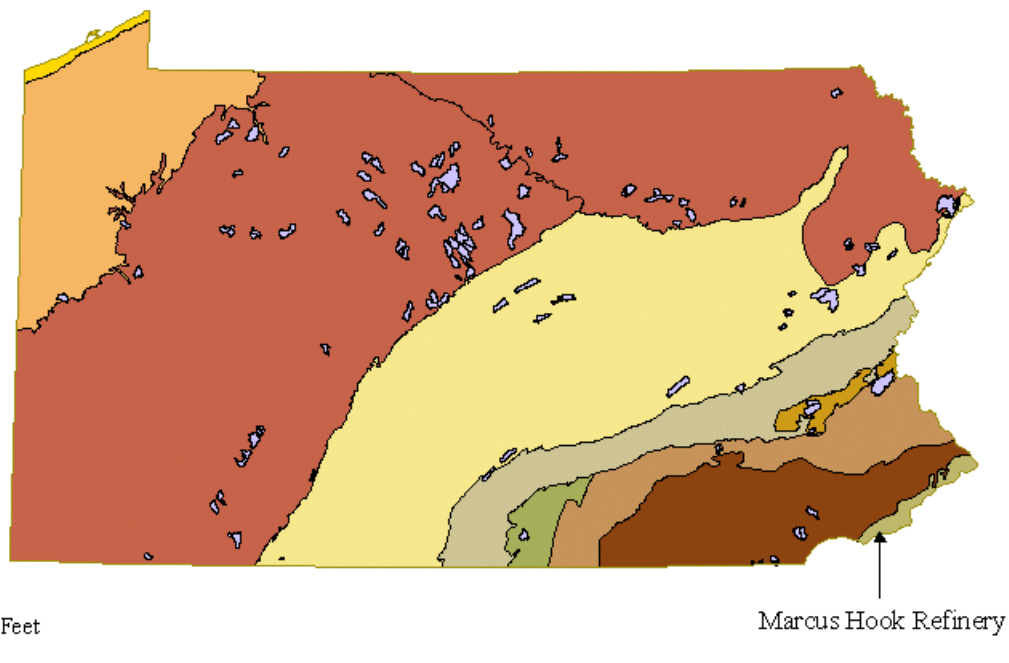
Aquatic Ecoregions	Major Roads
Average Annual Runoff*	Major Watersheds
Census Tracts*	Physiographic Regions
Climate Divisions	Public Water Supplies
County Boundaries	Soil Geographic (STATSGO) Database*
County Parks	State Boundary
County Streams*	State Parks
Digital Elevation Model (DEM)*	State Quadrangle Boundaries
Digital Line Graph Coverages	Stream Flow Basin Characteristics*
EPA Regulated Facilities*	Surface Geology*
EPA River Reach File 1	Surface Water Intakes
Exceptional Value Watersheds	USGS Gage Stations*
Hydrologic Units	USGS Topographic Map
Land Use*	VEMAP Climatological Data*
Major Rivers	Watershed Boundary for Delaware River

\* - Most critical coverages for the Marcus Hook project.

Most coverages that are obtained over the Internet include a metadata file, which describes the information about the coverage, such as the scale, method of creation, and attribute fields. Simply put, metadata is data about data. The scales on the regional coverages for Marcus Hook vary from 1:250,000 to 1:24,000. A scale of 1:24,000 is read as 1 unit (inch, foot, etc.) on the map equals 24,000 of the same units in the actual system. The scale documents the level of detail, and thus accuracy, in the applicable coverage. Therefore, a 1:24,000 scale is more detailed than a 1:250,000 scale. However, in ArcView, all coverages are viewed at the same scale, which is determined by the extent of the viewing window. The features of a coverage will be accurate as long as the viewing window is at a smaller scale than that used to create the coverage. At larger scales of viewing, features are too generalized to appear realistic. The complete spatial data dictionary for Marcus Hook, which includes the associated features for each coverage, is provided in Appendix B.

### Physiographic Regions

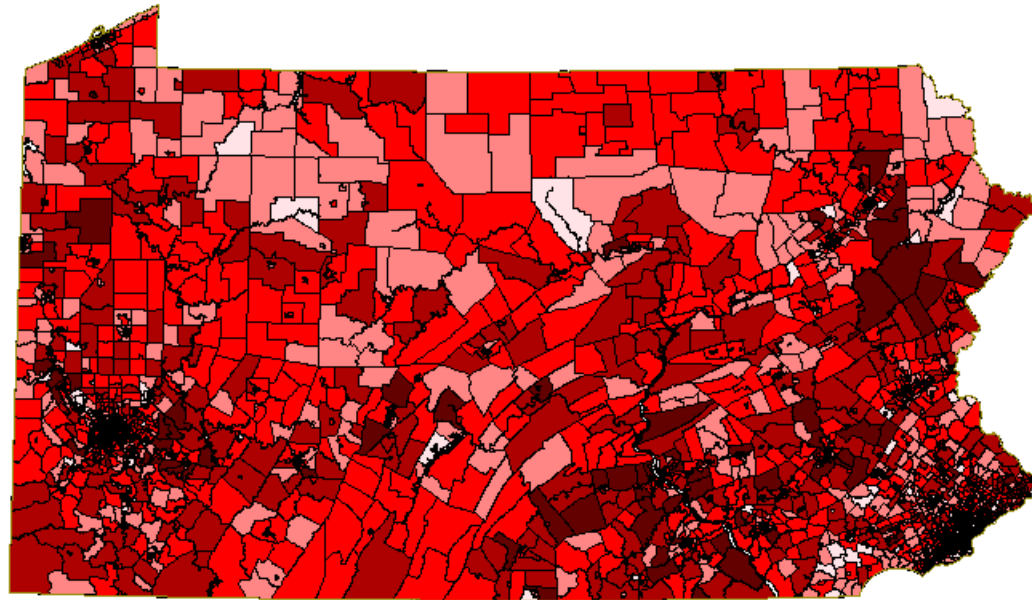
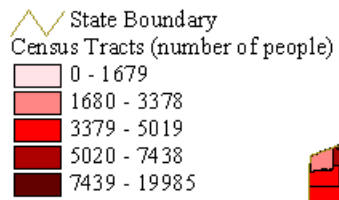
-  State Boundary
-  Exceptional Watersheds
- Physiographic Regions
-  Appalachian Plateau
-  Central Lowlands
-  Coastal Plain
-  Glaciated Appalachian Plateau
-  Great Valley
-  Reading Prong
-  Ridge and Valley
-  South Mountain
-  Triassic Lowlands
-  Virginia Piedmont



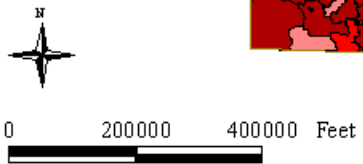
36

Figure 3.3: Physiographic Regions in Pennsylvania.

**Census Tracts (by total number of people)**







37



**Figure 3.4: Census Tracts in Pennsylvania Displayed by Total Population within Each Tract.**

### Surface Geology in Delaware County

-  Streams
- Surface Geology
  -  Anorthosite
  -  Bryn Mawr Fm
  -  Felsic Gneiss, Hornblende bearing
  -  Felsic Gneiss, Pyroxene bearing
  -  Granite Gneiss and Granite
  -  Mafic Gneiss, Hornblende bearing
  -  Mafic Gneiss, Pyroxene bearing
  -  Pegmatite
  -  Pensauken and Bridgeton Fms Undiv
  -  Serpentine
  -  Wissahickon Fm (Albite-Chlorite Schist)
  -  Wissahickon Fm (Oligoclase Mica Schist)

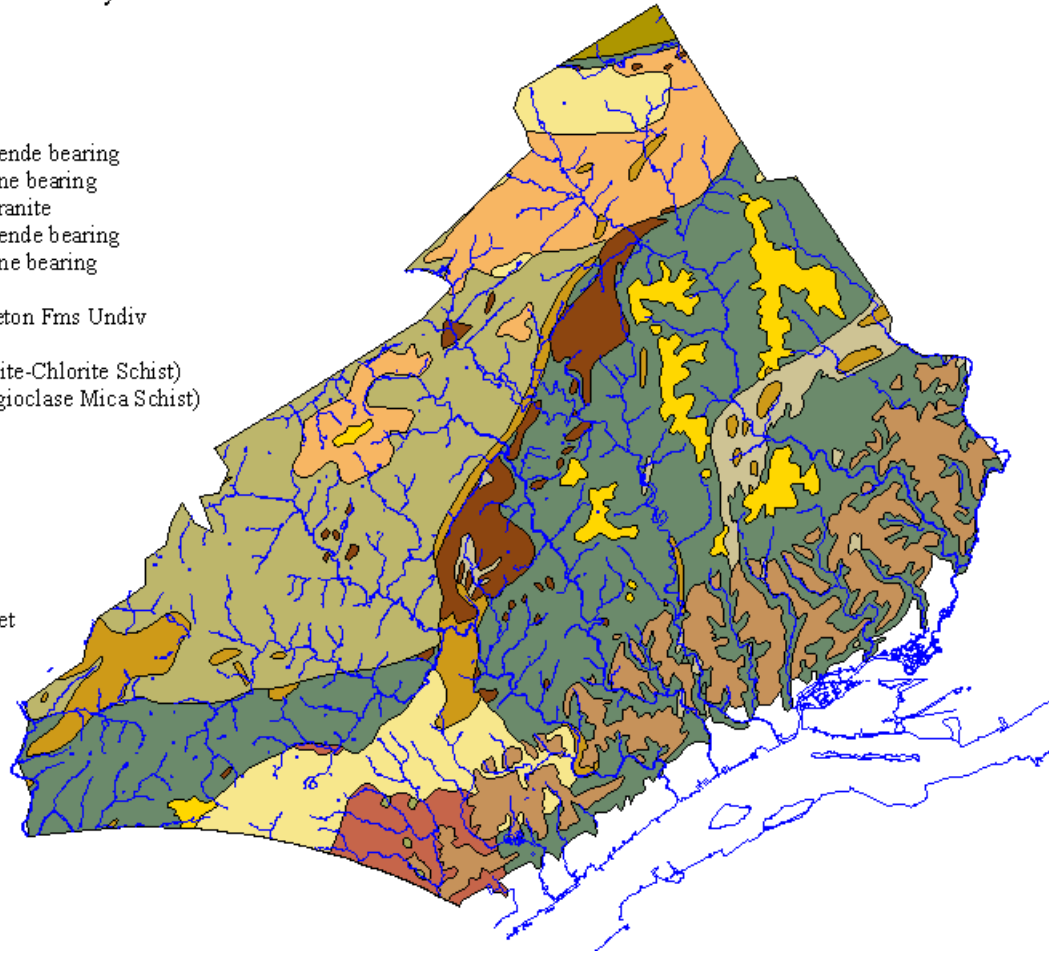
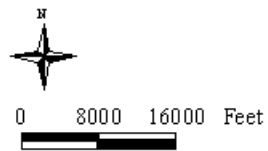
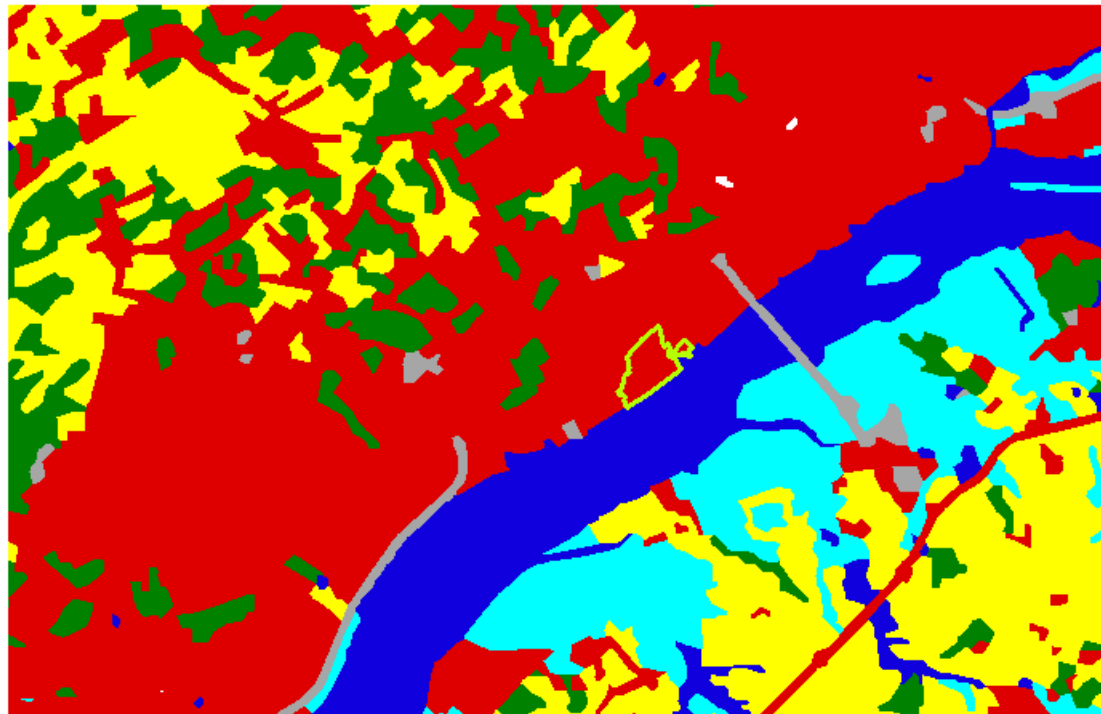


Figure 3.5: Streams and Surface Geology in Delaware County, PA.



### Land Use for Southeast Pennsylvania












-  Facility Boundary
- Land Use
  -  Urban
  -  Agriculture
  -  Range Land
  -  Forest
  -  Water
  -  Wetlands
  -  Barren

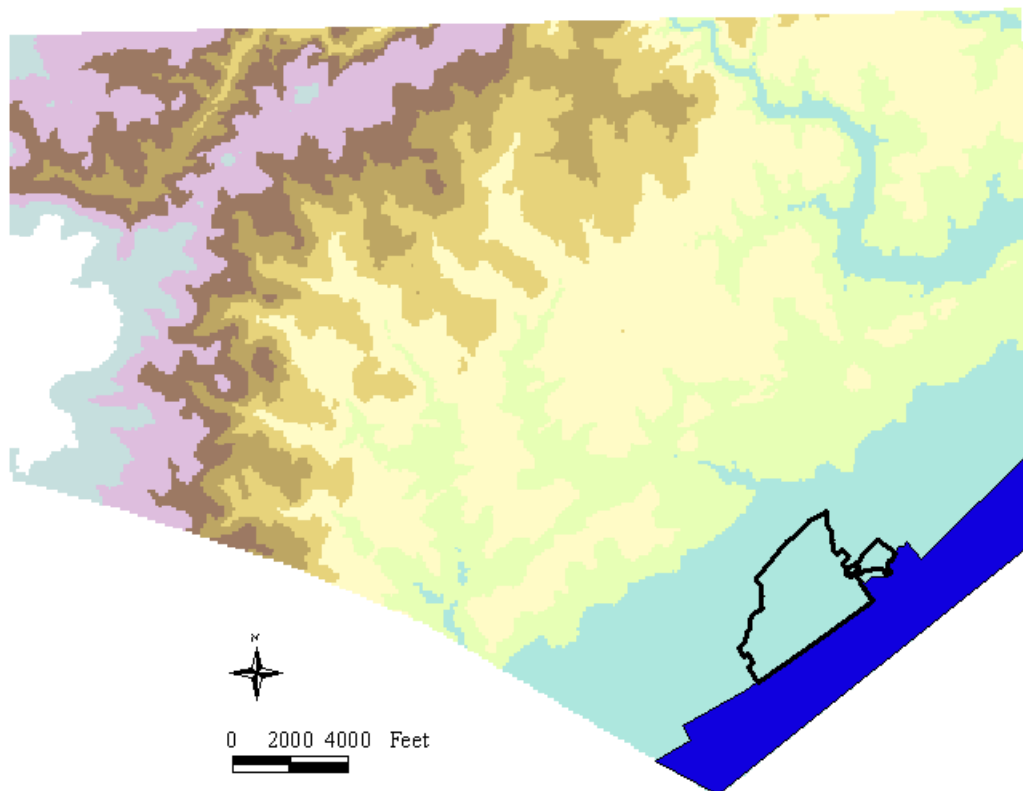


39

Figure 3.6: USGS Land Use Classifications for Southeast Pennsylvania.

**Marcus Hook DEM**

-  Marcus Hook Refinery
-  Delaware River
- Digital Elevation Model (ft)
-  0 - 47
-  47 - 95
-  95 - 142
-  142 - 190
-  190 - 237
-  237 - 285
-  285 - 332
-  332 - 380
-  380 - 427
- No Data



40

**Figure 3.7: Digital Elevation Model for the City of Marcus Hook, PA.**

### VEMAP Climate Data (by precipitation) for Southeast Pennsylvania

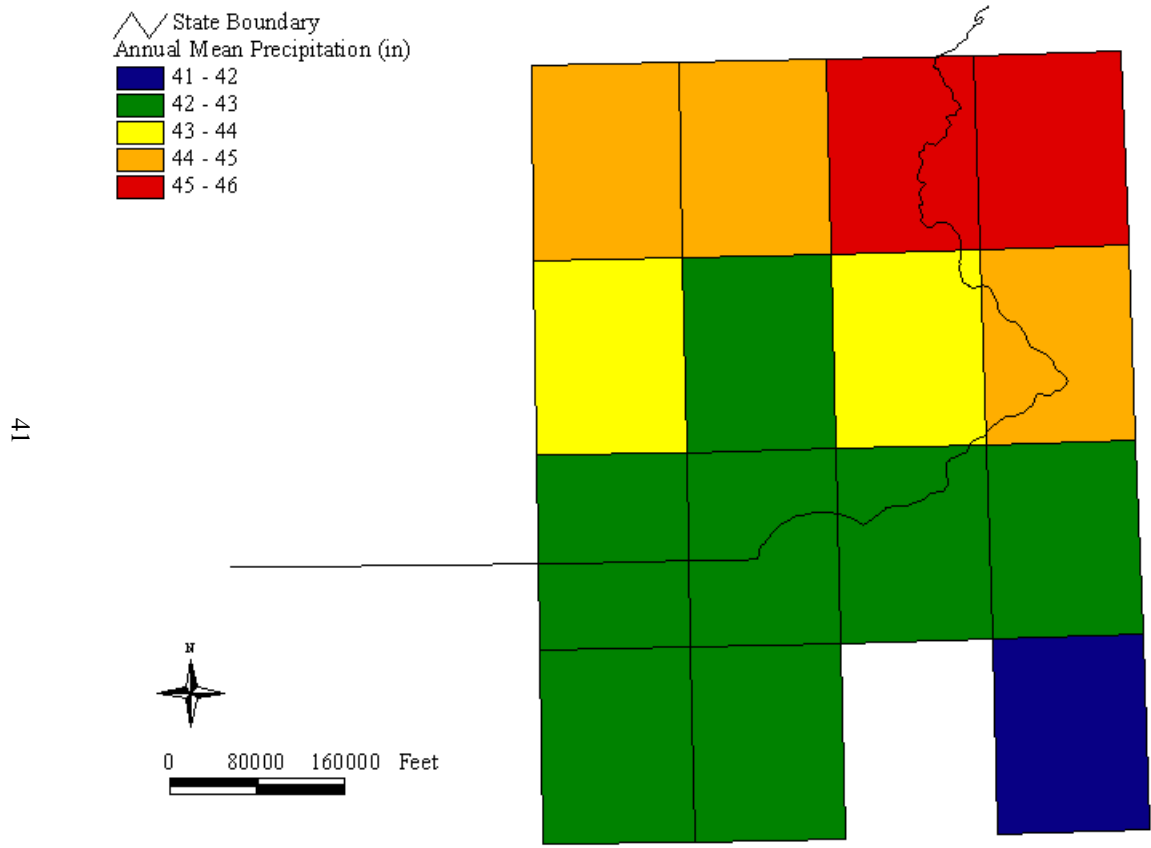


Figure 3.8: Annual Mean Precipitation Ranges for Southeast Pennsylvania.

The second step in the construction of a spatial database is to compile the facility-specific features. Unfortunately, facility-specific features require scales that are much too large to be supported by data on the Internet, but there are other options available for building a facility database. One method is to contract an aerial mapping company to over fly the site, produce photographs, and digitize features from the photographs. BP chose this option for the facility scale mapping at Marcus Hook. This method can be expensive, but it provides a very accurate depiction of facility features. Another method is for the user to digitize the facility features using existing maps or photographs. A user also has the option to use CAD drawings, which can be easily imported and georeferenced in ArcView. All three of these procedures are described and evaluated further in Section 3.4.3.

Table 3.2 lists the 34 coverages developed on a facility scale for the Marcus Hook project, and Figures 3.9 - 3.14 show some of the facility coverages in the ArcView mapping environment. The collection of fundamental features of the facility, as illustrated in Figure 3.9, is referred to as the basemap. The basemap is most useful simply as a map, which describes how physical features of the facility are spatially related to each other and to land characteristics. The surface cover theme, as shown in Figure 3.10, further enhances the understanding of land characteristics, but this theme can also be used in the analysis process. For instance, the expected mean concentration of benzene from runoff from each surface cover type was used in the development of a surface water model for Marcus Hook (Chapter 4). Other coverages are also important to the spatial environmental risk assessment process, such as in the construction of a site conceptual model. Figure 3.11 displays the Marcus Hook facility structures and surrounding properties, and Figure 3.12 displays the historical features of the facility. Each of these coverages can be used to identify sources and receptors. For industrial corrective action projects, the underground sewer and product lines are also features that should be investigated as sources. The Marcus Hook oily-











sewer line is shown in Figure 3.13, and besides identifying a potential source, this coverage was used to delineate sinks in the surface water runoff model. Figure 3.14 shows the digital terrain model as seen with the *3-D Analyst* extension for ArcView. This digital terrain model was utilized to create a facility digital elevation model (2-D) for use in the surface water runoff model. The complete data dictionary provided in Appendix B contains more information on data types and attributes.

**Table 3.2: Marcus Hook Facility Coverages** (refer to Spatial Data Dictionary in Appendix B for more information).

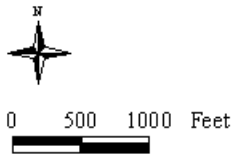
Active Non-Process Areas	Parcels*
Areas of Concern*	Pedestrian Trails
Bridges	Pipelines*
Control Points	Primary Transportation
Cultural	Process Area Vessels
Digital Terrain Model*	Property Boundary*
Electric Utilities	Rail Lines
Fences and Walls	RCRA Units
Former RCRA Units	Steam Utilities
Gas Utilities	Storage Tanks*
Historical Features*	Structures*
Marine Navigation	Sumps
Miscellaneous Features	Surface Cover*
Miscellaneous Utilities	Surface Hydrology*
Monitoring Wells*	Topography
NPDES Sampling Points*	Vegetation
Oily Water Sewer Line*	Water Utilities

\* - Most critical coverages for the Marcus Hook project.

**Marcus Hook Refinery Basemap**

-  Facility Boundary
-  Marine Navigation
-  Fences and Walls
-  Pipelines
-  Railroads
-  Bridges
-  Process Area Vessels
-  Storage Tanks
-  Structures
-  Surface Hydrology

44



**Figure 3.9: Marcus Hook Refinery Basemap.**

### Marcus Hook Refinery Surface Cover

- Surface Cover
- Streams
  - Wetlands
  - River
  - Ponds
  - Tidal Flat
  - Gravel
  - Grass/Vegetation
  - Concrete Pavement
  - Asphalt Pavement
  - Soil (no vegetation)
  - Structures



45

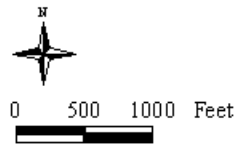


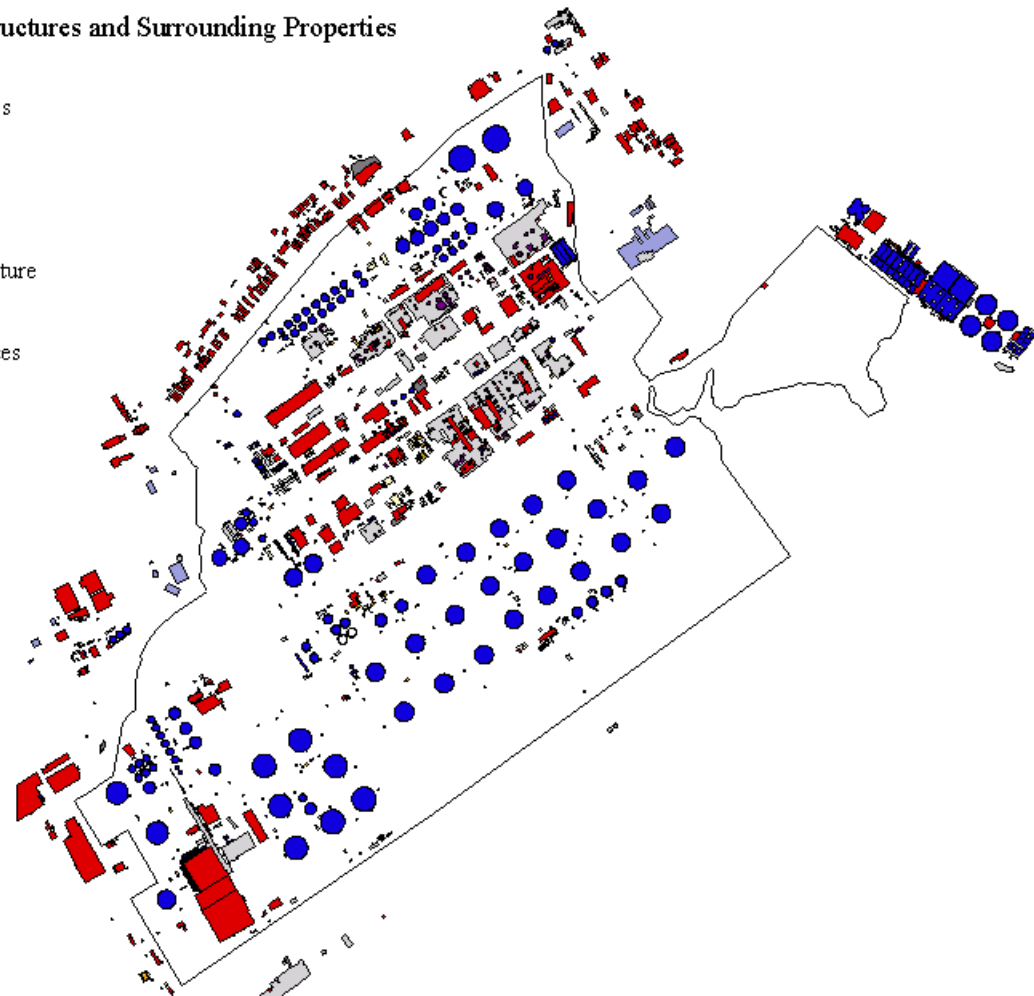


Figure 3.10: Marcus Hook Refinery Surface Cover.

### Marcus Hook Refinery Structures and Surrounding Properties

-  Boundary Line
-  Process Area Vessels
-  Storage Tanks
- Structures
-  Slabs (Asphalt)
-  Slabs (Concrete)
-  Building
-  Miscellaneous Structure
-  Overhead Roofline
-  Ruin/Foundation
-  Trailer Homes/Offices
-  Steps
-  Wooden Decks



46

Figure 3.11: Marcus Hook Refinery Structures and Surrounding Properties.



### Marcus Hook Refinery Historical Features

- ∟ Facility Boundary
- Historical Features
  - Ponds
  - Pump Station/Silo
  - Rubble Piles
  - Storage Tanks
  - Process Areas
  - Basins
  - Impoundments
  - Storage Areas
  - Loading/Unloading Areas

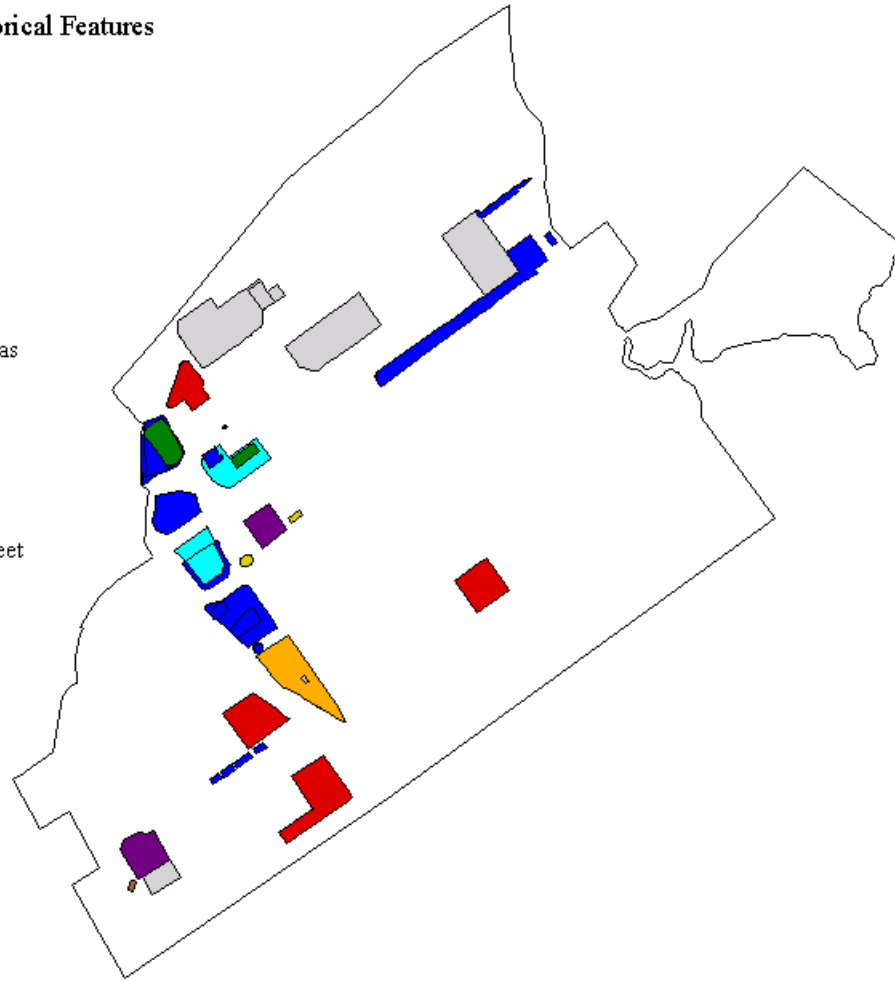
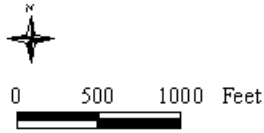


Figure 3.12: Marcus Hook Refinery Historical Features.

### Marcus Hook Refinery Oily Sewer Line

- Facility Boundary
- Oily Sewer
- Drains/Manholes
- Lines
- Surface Hydrology

48

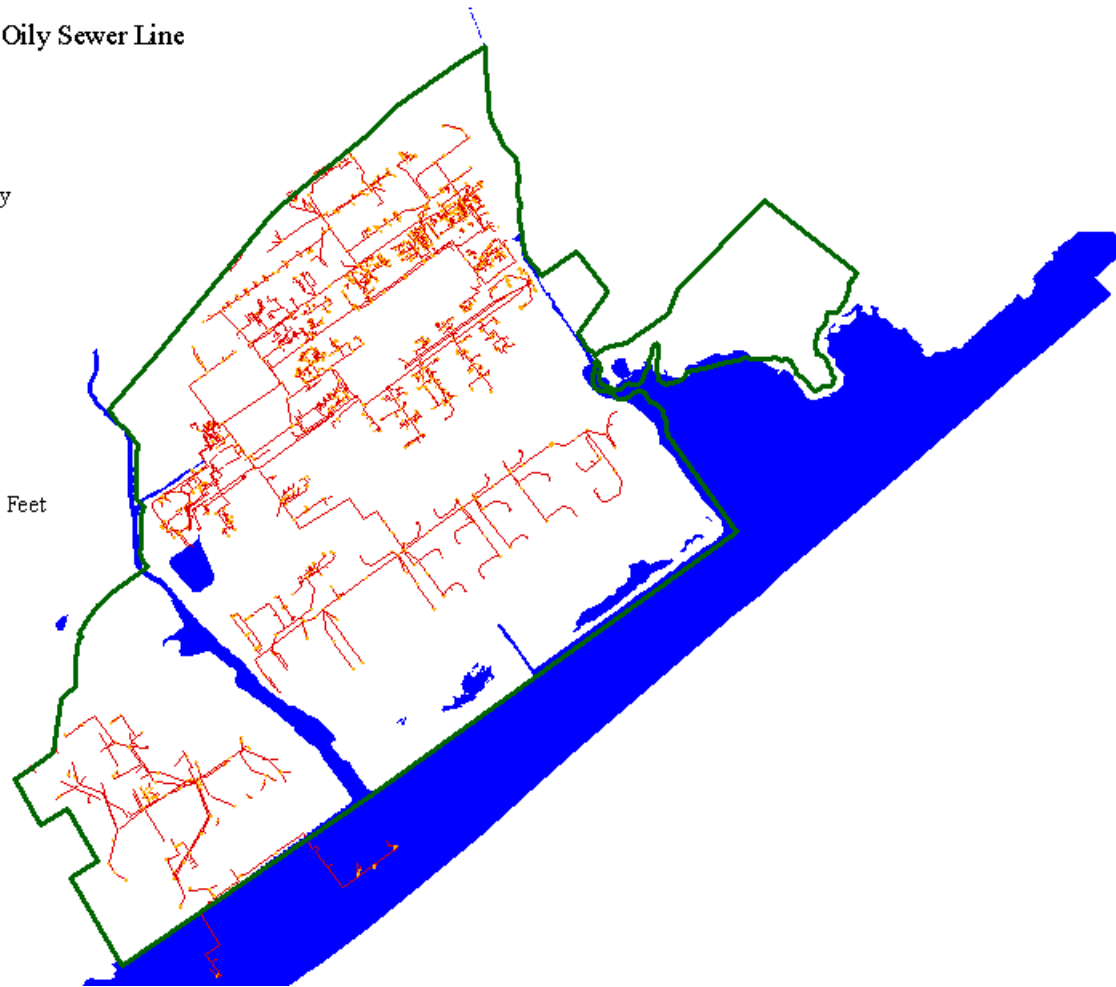
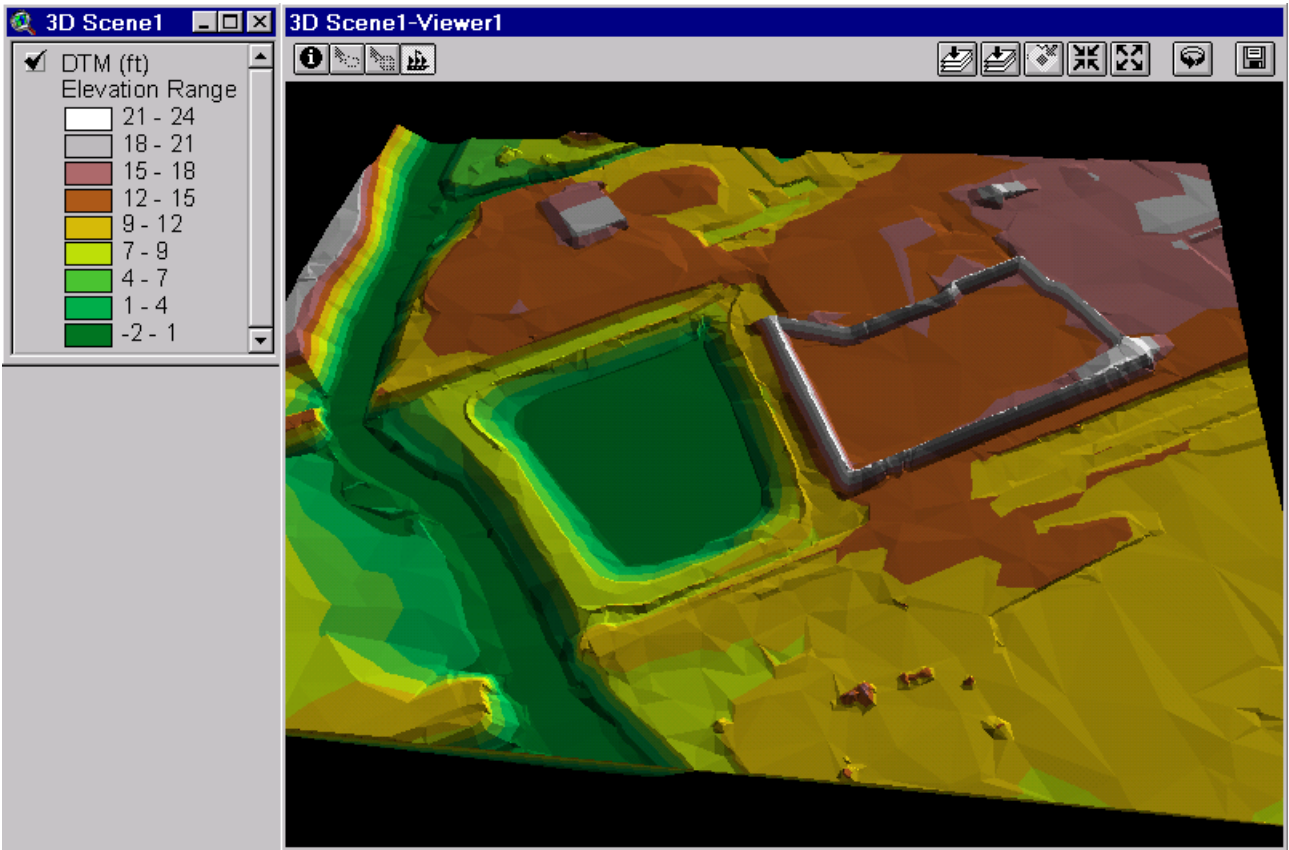


Figure 3.13: Marcus Hook Refinery Oily Sewer Line.



**Figure 3.14: Marcus Hook Refinery Digital Terrain Model.** This figure shows Marcus Hook Creek and the former surface impoundment. These features are located in the west-central portion of the facility.

The compilation of the spatial database for Marcus Hook was conservative on both the regional and facility scales, meaning that as much information as possible was collected. Of the 64 coverages, only about 25 have proved useful to the spatial environmental risk assessment process, and these critical coverages are identified in Tables 3.1 and 3.2. Even within some of the useful coverages is unnecessary information. For instance, the facility surface cover theme contains separate classifications for asphalt pavement and concrete pavement. Thus, a careful planning effort of what coverages are really needed and how each coverage will be used should precede the actual compilation process.

### **3.3 Tabular Database**

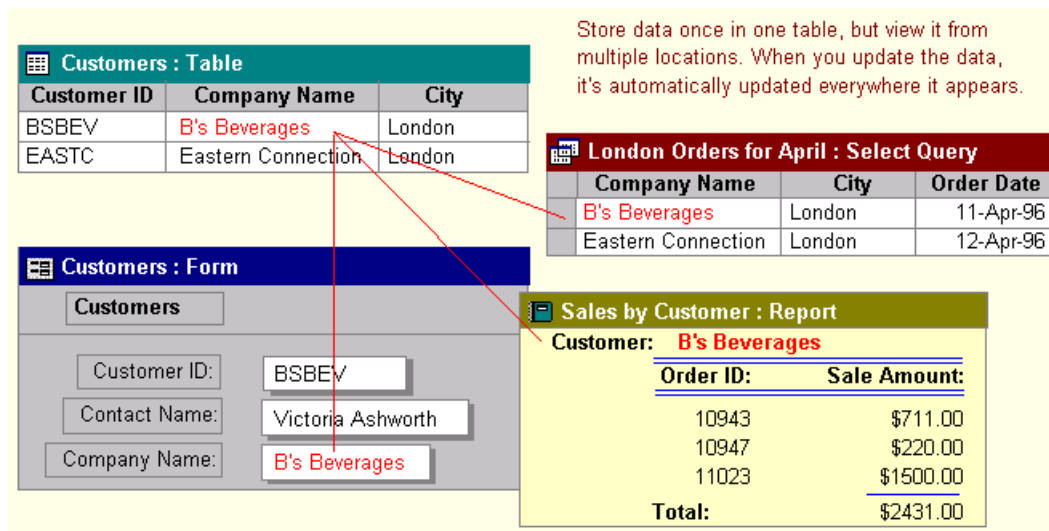
The tabular database consists of all the information pertaining to environmental measurements at the facility. It can contain physical measurements, such as groundwater levels. It can contain chemical measurements, such as concentration sampling results. It can also contain biological measurements, such as degradation rates. Anything that is specific to environmental conditions at a site can be included in the tabular database. The database described below for the Marcus Hook project consists primarily of sampling locations and the measurements taken there. However, a tabular database is flexible and can incorporate all types of site-specific data. The Marcus Hook database is only one example of an effective data management system.

An efficient form of organizing data is through the use of a relational database. A relational database is simply a collection of information that relates to a particular subject or purpose. The database is stored in one file, but this file can contain multiple tables, all of which relate to the particular subject but contain different data related to that subject. For instance, a database for a retail company could contain a table for customer information, another table for product

information, and a final table for shipping orders. All three tables relate to the operations of the company, but they contain different data sets that are easier to manage and understand through the use of separate tables. Tables are the fundamental elements of a database, and all database operations are performed through actions on tables. Tables are essentially storage containers, and database programs such as Microsoft Access can perform various operations on these containers. Data in tables can be managed in the following ways with Access:

1. View, add, and update table data using online **forms**.
2. Find and retrieve just the data you want using **queries**.
3. Analyze or print data in a specific layout using **reports**.

In order to combine data from separate tables into one form, query, or report, relationships or links must be set between each table. In the example described above, a common field such as a customer id could exist in both the customer information and shipping orders tables. By joining this field between tables, data can be retrieved from each table and combined into another table (i.e., a query or report). For example, one could determine all the shipping orders in March for a specific customer. Queries and reports can also perform calculations on data. To extend the above example even further, one could create a report that would calculate the money spent by each customer for orders in March. To accomplish this task, Access (or an equivalent database utility) determines all the orders and products bought for each customer in March from the shipping orders table and then calculates a sum of \$'s spent using prices in the product information table. Figure 3.15 shows some of the database functions that can be performed with Access.



**Figure 3.15: Access Database Functions (Microsoft Corporation, 1996).**

The tabular database for Marcus Hook functions similarly to the database described above except that it contains information specific to environmental data at the facility, such as sampling point locations, concentration measurements, and chemical properties. Table 3.3 lists the 13 tables in the Marcus Hook database and the type of data in each table. A complete tabular data dictionary is provided in Appendix C. The Location table is the key table in the database as all other tables are directly or indirectly linked to this table. It contains information about each sampling location, such as its location identifier, its spatial coordinates, its location type (e.g., well, soil boring), and its surface elevation. Information about particular types of measurements at these locations is separated into other tables to avoid repetition of data. For instance, any locations from which groundwater level measurements were taken are placed into the Groundwater\_Levels table.

**Table 3.3: Marcus Hook Tabular Database Tables** (refer to Tabular Data Dictionary in Appendix C for more information).

<b>Table</b>	<b>Contents</b>
Action_Levels	Target concentration levels set by agencies for each chemical and matrix type.
CAS_List	Name, abstract code, and class for each COC.
Chemical_Properties	Properties for each COC such as vapor pressure and diffusion coefficient.
Grainsize	Sieve analysis results for soil samples.
Groundwater_Levels	Depth to water and immiscible thickness measurements.
Hydraulic	Hydraulic conductivity measurements.
Location	Sampling point coordinates, elevation, type, and coordinate sources.
Results	Laboratory analysis results including concentration and detection limit.
Sample_Analysis	Supplementary analysis information for aqueous samples.
Sample_Collection	Sample collection data such as identification number, matrix, and depth.
Stratigraphy	Stratigraphic units for each soil boring.
Study	Listing of reports from which data in the database originate.
Well	Well construction data including elevations, screen length, and casing type.

These tables are linked through the common **LOC\_ID** field. This established link tells the database program which records in the Groundwater\_Levels table are related to corresponding records in the Location table. If tables are linked correctly, then information from each table can be combined in a query. For example, all of the groundwater measurements for a specific well can be queried. Databases work in both directions, so the locations for all of the wells with a groundwater level measurement above ten feet could also be queried. Figure 3.16 shows all of the database relationships. A link exists wherever there is a small black dot (e.g., the tables along the top row are all linked to Location and not to each other). This figure is a powerful one since it provides an effective means of understanding environmental database design.

The tabular database design for Marcus Hook was originally developed by Parsons Engineering Science using Oracle software. This design was then reconstructed in Access by the author. The same tables originally created by Parsons are still present in the database with the exception of the

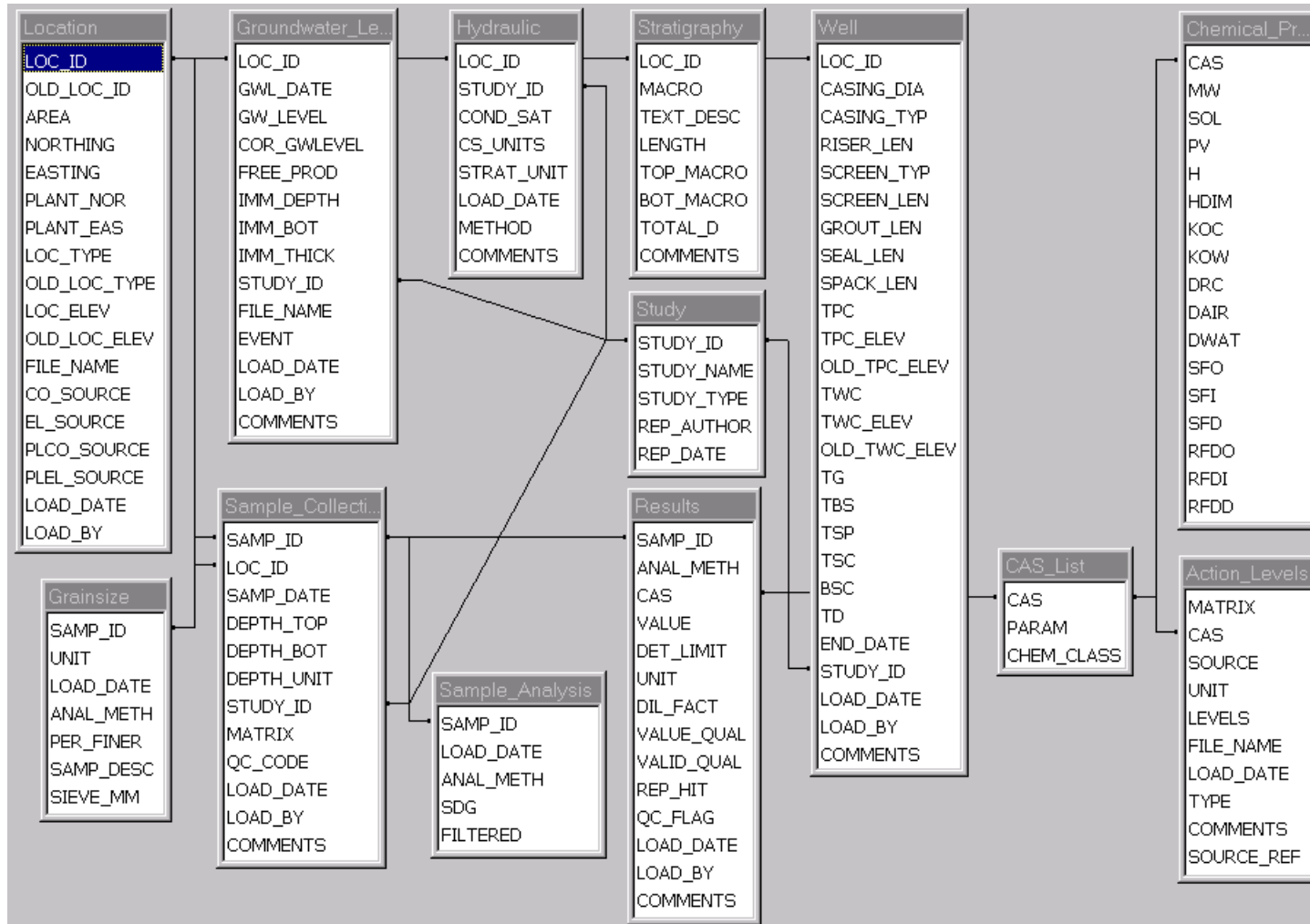


Figure 3.16: Marcus Hook Tabular Database Relationships.



Chemical\_Properties table, which was added later in the development process. There are 13 tables in the current database so that all the data about environmental conditions at the Marcus Hook facility are included in a simplified format that avoids repetition. In this form of management, the database is organized according to the varying levels of detail in the data. For instance, any information related to measurements or observations at each location is directly linked to the collection of sampling points in the Location table. This information includes groundwater levels, hydraulic conductivity measurements, soil profiles, and collected samples. A third level of detail exists only in the current database through the Sample\_Collection table. Any measurements, such as grain sizes and concentrations, are directly linked to this table through the **SAMP\_ID** field. Two more levels of data are included as ancillary information about chemical names, chemical properties, and target concentration levels set by regulatory agencies. These data are linked through the **CAS** field, which is a chemical code number that allows chemicals to be tracked easier.

This database fits the Marcus Hook project data very well. However, modifications to the structure might be appropriate for other sites. For instance, if a site has a precipitation gage sampling location, another table containing precipitation measurements could be created and linked to the Location table. Also of note is that the database structure shown in Figure 3.16 continues to be modified for purposes of the Marcus Hook project. Some of the proposed changes being evaluated at the time this thesis was published included:

- eliminating the Sample\_Analysis table and incorporating the data into the Sample\_Collection table,
- moving the Chemical\_Properties table from this database into the site conceptual model database, and
- creating look up tables for fields with values taken from a set of criteria (e.g., ANAL\_METH, MACRO, and UNIT).

Another tabular database design consideration is creating a more sophisticated structure in which multiple fields from a table are linked to fields in other tables. However, this approach requires an experienced database developer who particularly understands how multiple relationships affect queries. The structure presented in Figure 3.16 is simple, but it eliminates the repetition of data and provides an easy mechanism to extract data from multiple tables. More information on the design and construction is provided in section 3.4.4.

Developing a design is just the first step towards building a tabular database for use in the risk assessment process. The next step is to actually acquire the data and populate the database. The database is first populated with data that have already been collected from previous studies. These data include but are not limited to geotechnical reports, characterization studies, and monitoring programs. Existing information helps build a current picture of the site and determine what additional data need to be collected. In the case of Marcus Hook, dozens of reports and studies already existed at the beginning of the spatial risk assessment project. Organizing and compiling these data in digital form can be very time consuming. For instance, the Results table alone in the Marcus Hook Database has over 30,000 records from previous studies. However, the required effort definitely justifies the end product since all the data can be found in and selected from one file. All the data are linked within the database to the original report from which the data were generated through the Study table. Therefore, the original data source can be consulted when questions arise with particular sets of data. Many projects might have minimal or no existing data. In these cases, the database information will come from the future measurements obtained at the site, and the database design should reflect what type of data will be collected. No matter what type of project data exists, the tabular database is an evolving component of the digital facility description. It is continually updated as new information becomes available.

As opposed to the GIS software arena where software choices are limited, there are many database utilities. Most of these programs are compatible with the development of an environmental risk assessment database. However, Microsoft Access 97 has been chosen as the database program for the Marcus Hook project. Similar to ArcView, Access is easy to use and is a component of the very popular Microsoft Office Professional package. Access is also very compatible for sharing data with other applications, such as ArcView and Excel. Importing data into ArcView is not limited to the Access program, but the procedure with Access is very straightforward (Section 3.5).

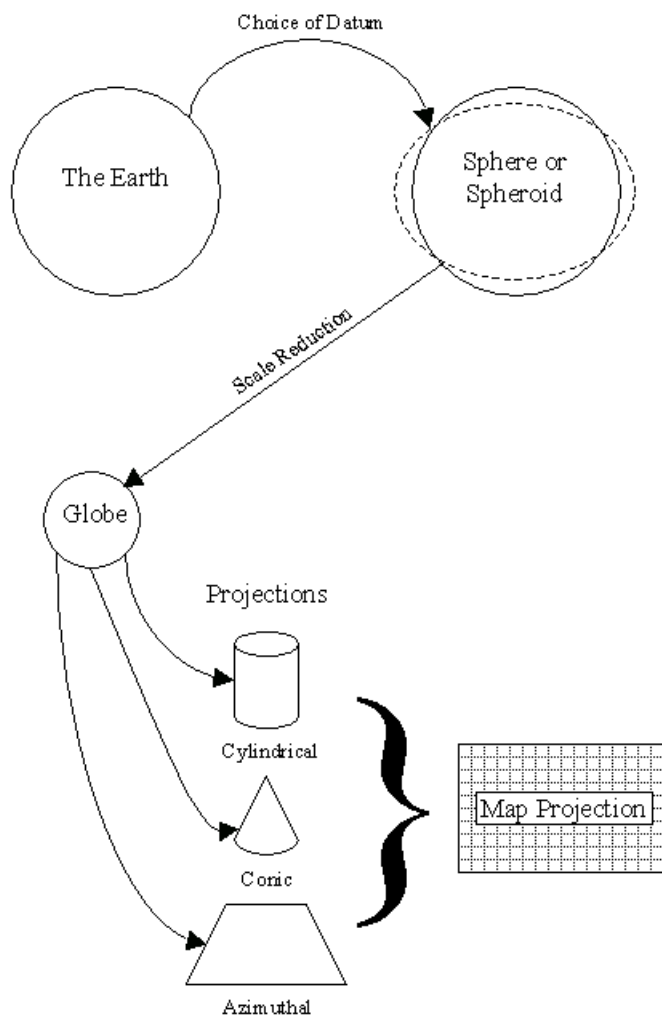
### **3.4 Data Development**

Building and maintaining spatial and tabular databases requires a considerable amount of thought and effort. This section describes some of the considerations that should be incorporated in compiling both databases. These considerations include maintaining a consistent coordinate system, determining what data are actually needed, and organizing the compiled data. This section also includes a description of the aerial mapping process and what other alternatives exist for creating spatial data of facility-specific features. The data development process is the mechanism for creating a powerful digital facility description that can be used for the purposes of risk assessment.

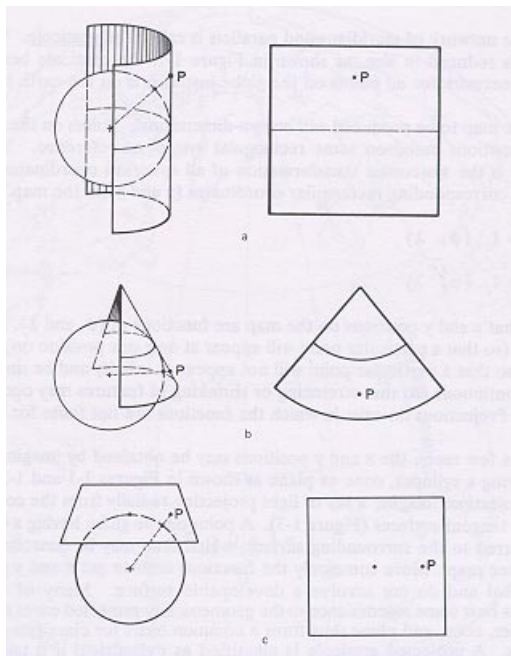
#### **3.4.1 Map Projections**

In order to maintain accurate topological relationships between spatial features, all coverages need to be projected in the same mapping coordinate system. Because features in reality exist on a three-dimensional earth, a difficulty arises when trying to display the same features on a two-dimensional map. There are two steps required to transform spherical coordinates into rectangular coordinates (Figure 3.17). First, a scale reduction is needed to fit the large area of

the earth into the size of a sheet of paper (McDonnell, 1991). Then, a systematic way must be developed of deforming the rounded surface to make it flat (McDonnell, 1991). The simplest way to transform features is to fit a surface, such as a cylinder, cone, or plane, to the globe representing the earth. Each point on the globe is transferred to the surrounding surface as shown in Figure 3.18, and the surface is then laid flat for mapping purposes.



**Figure 3.17: Transformation of Locations on the Earth’s Surface to Locations in a Two-Dimensional Coordinate System (McDonnell, 1991).**



**Figure 3.18: Transformation of Points on a Globe to a Surface (McDonnell, 1991).**

There are a variety of different ways to represent the earth's surface and to transform coordinates from this represented earth to a map. The selection of a map projection depends on the desired scale and intended purpose. There is no projection that possesses a uniform scale everywhere (i.e., scale factor = distance on the map/distance on the globe = 1.0), so a trade-off will always be incorporated in the selection of a projection (McDonnell, 1991). For instance, some projections preserve the area of a feature while others preserve the shape of a feature. For the facility scale projects that are typical of an environmental risk assessment, the differences between projections will not be very evident since the roundness of the earth does not play a significant factor for these small areas. However, the selection of a projection can be important for regional features and for land surveyors. For those working in the United States, a state plane projection is usually the best alternative since, for small sites, it provides a minimal amount of distortion. State plane projections vary from state to state, and

more information on specific parameters can be found in *Map Projections - A Working Manual* (Snyder, 1987).

For the Marcus Hook project, the Pennsylvania State Plane System (South Zone) using the North American Datum of 1983 (NAD83) and North American Vertical Datum of 1988 (NAVD88) was selected as the coordinate system for all coverages. A datum is the name given to a smooth mathematical surface that closely fits the mean sea-level surface throughout the area of interest (Snyder, 1987). The Pennsylvania State Plane System projection parameters are listed in Table 3.4. The projection is Lambert Conformal Conic, which is a standard projection for representing areas whose East-West extent is large compared with their North-South extent. This projection is "conformal" in the sense that lines of latitude and longitude, which are perpendicular to one another on the earth's surface, are also perpendicular to one another in the projected domain (Maidment, 1998). A conformal projection preserves shape since the scale factor at any one point is the same in all directions (McDonnell, 1991). Overall, it provides a good combination of minimum angular and scale distortion (McDonnell, 1991).

**Table 3.4: Pennsylvania State Plane Coordinate System Parameters (Snyder, 1987)**

Projection:	Lambert Conformal Conic
Map Units:	Feet
Datum:	NAD83
1 <sup>st</sup> Standard Parallel:	39 56 00
2 <sup>nd</sup> Standard Parallel:	40 58 00
Central Meridian:	-77 45 00
Latitude:	39 20 00
False Easting:	600000
False Northing:	0

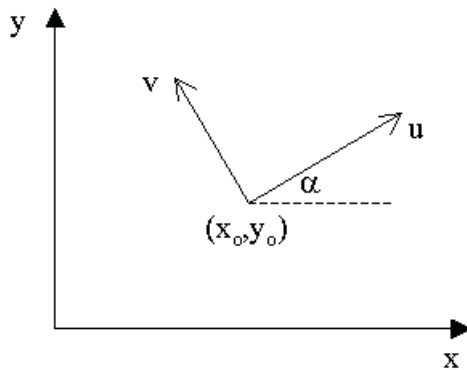
Many times, a data coverage will need to be converted from one projection to another to satisfy the project requirements. This occurrence is very typical for coverages that are downloaded from the Internet. However, both ARC/INFO and

ArcView can perform projection conversions. In ARC/INFO, the **project** command can be used to transform a coverage from one coordinate system to another. In ArcView, an extension called *Projector* can be used for spatial data conversion. Previously, this extension could only be used to view the coverage in a different projection. However, in ArcView Version 3.1, it can be used to perform an actual data conversion. Specific information on performing projection conversions is provided in Appendix E.

Maintaining a consistent coordinate system applies not only to the spatial coverages, but also to the tabular data in the environmental database. Each sampling location or other relevant environmental feature must have spatial coordinates referenced to the correct projection. Otherwise, any data imported from the tabular database into the GIS application will not overlay properly. Typically, measurement locations are surveyed during a study. If the measurements are surveyed using a different coordinate system, they can always be converted into the correct format. Conversions can also be performed with measurements that are marked on a map according to a facility grid. As long as the location of one point in both systems is known along with an angle of rotation, an appropriate data transformation can be accomplished (Figure 3.19). The conversion of coordinates from a local system to a global system can be performed with the following formulae:

- $x = x_o + u \cos \alpha - v \sin \alpha$  [3.1]

- $y = y_o + u \sin \alpha + v \cos \alpha$  [3.2]



**Figure 3.19: Conversion of Coordinates from a Local System (e.g., u and v) to a Global System (e.g., x and y).**

This type of conversion was required for the original data assembled in the Marcus Hook tabular database, which had been referenced to a previously used plant grid coordinate system. This plant grid system consisted of 100-ft. grid cells with an origin in the southwest corner of the facility. Sampling point locations were originally measured in this system with either a map or with a CAD drawing. In order to convert these coordinates into the Pennsylvania State Plane Coordinate System, the following parameters were used:  $x_o = 2626199.942$ ,  $y_o = 183509.695$ ,  $\alpha = 35.6914347$ , and a scale = 1.00038871. The variables  $u$  and  $v$  are the existing plant grid coordinates multiplied by the scale, and  $x$  and  $y$  are the resulting state plane coordinates. Table 3.5 shows the conversions for the monitoring wells at Marcus Hook and compares these conversions with the surveyed measurements obtained as part of the aerial mapping project. This comparison is also presented in Figure 3.20. The conversion errors are minimal for the most part, but some obvious discrepancies exist (e.g., MW-2). Because conversions of this type may result in some errors, field verification of locations may be necessary. In the case of Marcus Hook, the data conversion was a one-time event, as all future measurements are being surveyed using the Pennsylvania State Plane Coordinate System.

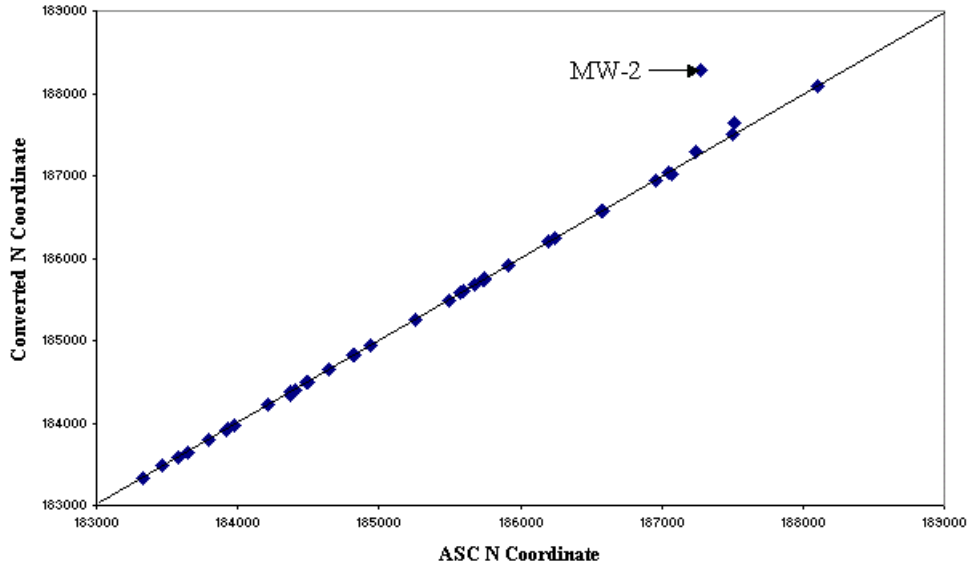


**Table 3.5: Monitoring Well Coordinate Conversions**

Well #	ASC N	ASC E	Grid N	Grid E	Converted N	Diff N	Converted E	Diff E
MW-1	187502	2626930	2813	2922	187501	-0.7	2626932	1.9
MW-2	187281	2627981	2835	4225	188279	998.4	2627978	-2.9
MW-3	187046	2626918	2450	2647	187045	-0.5	2626920	2.3
MW-5	186203	2627987	1144	3022	186203	-0.1	2627988	1.3
MW-6	185913	2628552	578	3312	185912	-0.9	2628554	1.9
MW-7	184646	2626780	583	1132	184645	-1.4	2626780	-0.1
MW-8	184401	2626242	699	554	184401	-0.3	2626243	1.1
MW-9	183972	2626290	322	344	183972	-0.3	2626291	1.1
MW-10	183794	2625885	415	-90	183794	0.3	2625885	-0.1
MW-11	183460	2626079	51	-111	183486	25.8	2626080	1.1
MW-12	184370	2627263	35	1361	184332	-37.9	2627285	21.8
MW-13D	184819	2627389	410	1676	184821	2.3	2627322	-67.0
MW-13S	184822	2627397	405	1695	184828	6.1	2627341	-56.0
MW-14	185495	2628367	325	2948	185495	-0.8	2628405	38.6
MW-16	185574	2627133	1147	1950	185580	6.0	2627115	-17.4
MW-17	186963	2625895	2970	1756	186947	-15.6	2625893	-2.3
MW-18	188100	2626782	3362	3166	188089	-10.8	2626811	28.6
MW-19	187517	2628669	1954	4355	187639	121.4	2628598	-70.5
MW-20	185675	2626518	1570	1521	185673	-2.0	2626519	1.5
MW-21	185597	2626295	1637	1292	185594	-3.0	2626294	-0.8
MW-22D	185733	2626184	1813	1284	185732	-1.2	2626185	1.3
MW-22S	185748	2626170	1832	1281	185745	-2.8	2626171	1.4
MW-23	186204	2626110	2237	1497	186201	-2.8	2626111	0.2
MW-24	186239	2626696	1925	1994	186238	-1.2	2626696	0.2
MW-25	183582	2626234	37	70	183580	-1.8	2626235	0.7
MW-26	183329	2625870	43	-373	183327	-2.2	2625872	1.1
MW-27	183650	2626015	220	-69	183649	-0.9	2626016	0.2
MW-28	183921	2626530	139	508	183919	-1.8	2626531	1.2
MW-29	183931	2626705	46	656	183929	-1.9	2626706	1.1
MW-30	184492	2626161	818	541	184490	-1.6	2626162	1.0
MW-31	184371	2626408	576	671	184369	-1.5	2626409	0.9
MW-32	184489	2625975	925	389	184488	-1.0	2625976	0.9
MW-33	184216	2625947	720	207	184216	-0.7	2625947	0.4
MW-34	184944	2625642	1489	384	184943	-1.1	2625643	0.7
MW-35	186575	2628676	1047	3797	186576	1.2	2628674	-2.2
MW-36	186582	2629330	669	4333	186582	0.4	2629330	-0.2
MW-37	185750	2626494	1645	1545	185748	-1.8	2626495	1.0
MW-38	185256	2626867	1027	1561	185255	-0.8	2626869	1.7
MW-39	187068	2626337	2777	2136	187012	-56.0	2626315	-21.8
MW-40	187246	2626386	2974	2330	187286	40.4	2626357	-28.4

Notes: ASC N + E = state plane coordinates provided by Air Survey Corporation.  
Grid N + E = plant grid coordinates  
Converted N + E = state plane coordinates converted from plant grid coordinates with equations 3.1 and 3.2.  
Diff N + E = ASC - Converted

### Comparison of Northing Coordinates



### Comparison of Easting Coordinates

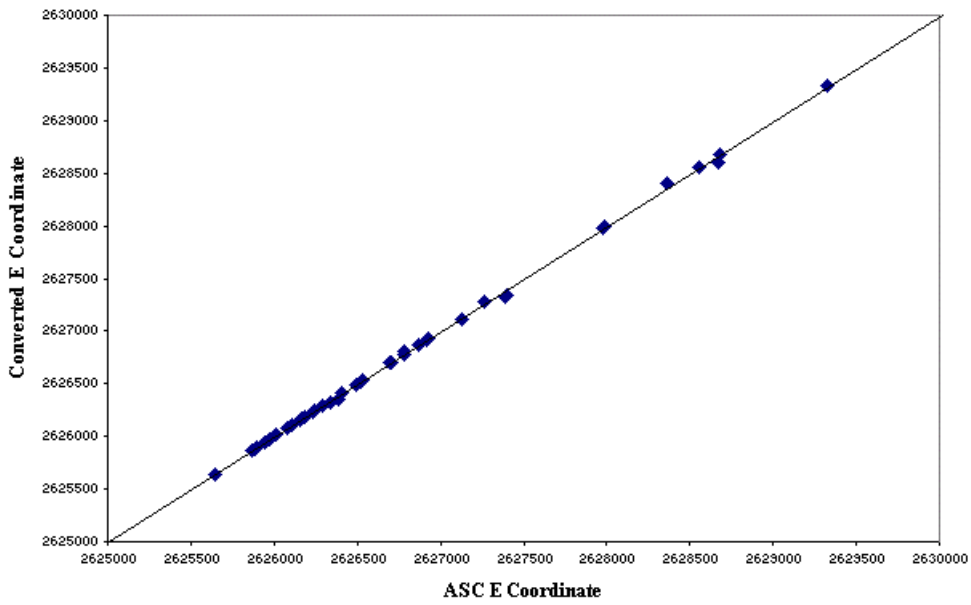


Figure 3.20: Comparison of Converted Coordinates to Survey Measurements.

### 3.4.2 Determining Data Needs

The data that are to be included in a digital facility description are chosen at the beginning of the project based on the questions that need to be answered. In the environmental risk assessment process, all of the following questions need to be addressed:

- What are the sources?
- What are the receptors (both human and ecological)?
- What are the ecological habitats?
- What are the transport pathways?
- What are the risks to the identified receptors?
- What are the remedial action implications based on the calculated risks?

The collection of data for both the spatial and tabular databases focuses on how the collected data can provide answers to these questions. The objective is to obtain any information that could prove useful for understanding the characteristics of the study area, comprising the facility and surrounding properties. The *Risk-Based Decision Process Guidance Manual* (BP, 1997) specifies a default regional area of 2000 feet from the property. At the Marcus Hook facility, this 2000-ft. boundary includes residential properties, industrial properties, and two Superfund sites. However, a study area boundary judgment is not based solely on this detailed value and incorporates any regional features that could provide insight into what conditions to expect at the facility. At Marcus Hook, for instance, it is important to analyze the upstream watersheds for the two creeks that wind through the facility so that potential impacts to the creeks from other facilities can be determined. Other important regional features include topography, land use, surface hydrology, geology, hydrogeology, groundwater use, and climatology. Some of the example applications of these coverages include developing geologic cross-sections, understanding water drainage patterns, and determining drinking water sources.

A spatial database, on a facility scale, includes coverages that define the physical characteristics of the study area and coverages that are useful for analysis purposes. These categories encompass features that identify potential sources and receptors. At Marcus Hook, source features include storage tanks, process units, pipelines, and impoundments while receptor features include houses, schools, wetlands, and creeks. The source and receptor feature compilation also incorporates, to the extent possible, any historical features such as old storage tanks or waste areas. Overlaying historical coverages with COC concentration contours can provide insight into possible sources and source areas (Kim, 1998). The spatial database also includes coverages that can be used for analyzing transport mechanisms. For instance, a surface cover theme can be used to determine runoff and to delineate different recharge areas in a groundwater model (Chapter 4). A final consideration for compiling coverages is to include any features that could be useful simply for display. In the case of the Marcus Hook project, the marine piers might not be part of the site conceptual model, but they provide information about where the ships dock along the Delaware River and pump the crude oil into the facility. A complete display of features at a site can be a very effective presentation tool in meetings with agencies or public officials.

The tabular database incorporates information related to the environmental conditions at the facility. For chemical release facilities, this information encompasses sampling locations, groundwater levels, concentration measurements, soils data, and chemical properties. These data sets provide insight into where certain chemicals are located and how they are migrating across the facility. The difficulty with a tabular database is not necessarily determining what data are needed but where and when to obtain these data. Unlike the spatial database, in which features are not changing very often over time, the tabular database contains measures of conditions that have considerable temporal

variability. For instance, the groundwater concentration for a specific well continually varies based on changing natural groundwater flow processes, such as dispersion and degradation, and source release rates.

Determining when and where to collect data is based on one's knowledge of the facility. For example, knowing which chemicals need to be analyzed in a particular sample depends on an understanding of process characteristics and product storage areas, and knowing where more data would be useful depends on an understanding of existing site data (Chapter 5). Many risk assessments involve subdividing a site on a uniform grid basis and then obtaining samples from each grid cell. This process provides an extensive amount of data, but it requires high costs and often produces unnecessary data. Data are only obtained if these data will provide useful information to the risk assessment.

The digital facility description is an evolving element of the environmental risk assessment process. It is continually updated as new data become available, and the collection of data focuses on how it will address the questions listed above. The spatial and tabular database dictionaries are provided in Appendices B and C so that the reader can review what data are relevant for the Marcus Hook project. These appendices serve only as a guide, and the reader should use his or her own discretion when determining what data will be useful. The data layers in a digital facility description vary from site to site.

### **3.4.3 Photogrammetry and Spatial Data Creation**

Photogrammetry is the process of creating maps from photographic images (Falkner, 1995)\*. It involves interpreting features on a photograph and relating them to ground equivalents. The most common method of obtaining photographs of an area is to take pictures with a camera mounted on a fixed-wing aircraft.

The photographs can then be georeferenced by relating features on the photographs to field surveyed locations. After an image has been referenced, individual features can be digitized to create coverages of the study area. The aerial mapping process is typically very expensive because of the equipment requirements, but it is a powerful tool for developing a facility-specific spatial database.

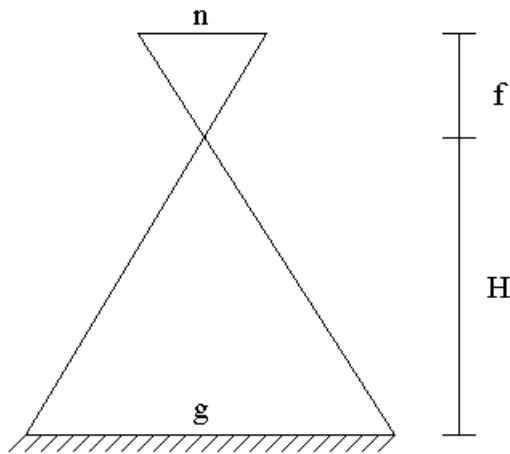
The first step of the aerial mapping process is to prepare a flight plan based on the desired photography scale, the area of study, and the intended application of the photographs. This phase is probably the most important of the process because all of the decisions made in this step will be reflected in the results of the mapping efforts. The most important consideration is the photography scale, which determines the level of feature detail and photo quality. Photo scale is defined as the number of measurement units on the ground that represent one unit of measurement on the photograph. Using the relationship between similar triangles, it can be calculated from the height of the airplane and the focal length of the camera (Figure 3.21):

$$\blacksquare \quad s_p = n / g = f / H \quad [3.3]$$

where  $s_p$  = photo scale,  $n$  = film negative width,  $g$  = ground distance covered by the negative,  $f$  = focal length of the camera, and  $H$  = altitude of aircraft above ground level.

The focal length is the distance from the rear nodal point within the camera lens to the focal plane, and most photomapping cameras have a 6-inch focal length. Therefore, if a photo scale of 1:2,400 (i.e., 1 unit of measurement on the photo = 2,400 equal units of measurement on the ground) is desired, then the aircraft needs to fly at a height of 1,200 feet ( $H = 6 / (1/2400) * 1 \text{ ft} / 12 \text{ in} = 1,200 \text{ ft}$ ). The lower the flight height, the larger the scale will be (i.e., the greater the detail). Larger scale photos provide more accurate feature representation on maps, but they also require higher costs, since more exposures are needed to cover the study

area. The Federal Aviation Agency (FAA) requires that planes fly high enough to glide clear of populated areas should engine power be lost, and they have set a generic height of 1,000 feet above ground level. This minimum flight height corresponds to a maximum photo scale of 1:2,000 for cameras with a focal length of 6 inches.



**Figure 3.21: Photographic Scale Relationships (Falkner, 1995).**

For the Marcus Hook project, four different flying heights were utilized so as to provide flexibility for the mapping applications. The flight heights of 1,000 feet above mean terrain (AMT), 1,500 AMT, 2,400 AMT, and 5,700 AMT yielded photo scales of 1":166', 1":250', 1":400', and 1":1,000', respectively. The 1":166' scale photography was used for the 1":20' topographic mapping. As previously mentioned, this is the largest scale photography available for the facility because of the 1,000 AMT minimum flight height. The 1":250' scale photography was used to produce 1":20' scale enlargements in color and black & white. The 1":400' scale photography was used for creation of digital orthophotographs and facility-specific coverages. Finally, the 1":1,000' scale photography was used for additional mapping functions, such as delineating

features within the detailed study area. The 1":400' scale was chosen for map creation because this scale provided an adequate balance of feature detail and photo quality (ASC, 1997). An evaluation of necessary map accuracy is discussed later in the section.

Incorporating the desired photographic scale and the actual area of the site, the required number of exposures can be determined. Aircraft typically fly over a site in the pattern shown in Figure 3.22a. Photo exposures must overlap in two directions to ensure stereoscopic coverage, which is a three-dimensional optical illusion created by overlapping a pair of two-dimensional photos. This phenomenon of observing a feature from different positions is referred to as the parallax effect, and the overlapping portion of two photos is known as a stereomodel. The common image area on consecutive photographs along a flight strip is called endlap while the overlapping areas of photographs between adjacent flight lines is called sidelap (Figure 3.22b & c). endlap will vary depending on the project, but it typically ranges from 55 - 65% of the length of the photo and is sometimes 80% for projects requiring more accurate coordinates. Sidelap ranges between 20% and 40% of the width of the photo. For the Marcus Hook mapping, the endlap was 60% and the sidelap was 30%. In order to layout an appropriate flight plan, one must know the distance between the centers of consecutive photographs. These distances are termed endlap gain and sidelap gain, and they can be calculated with the following formulae:

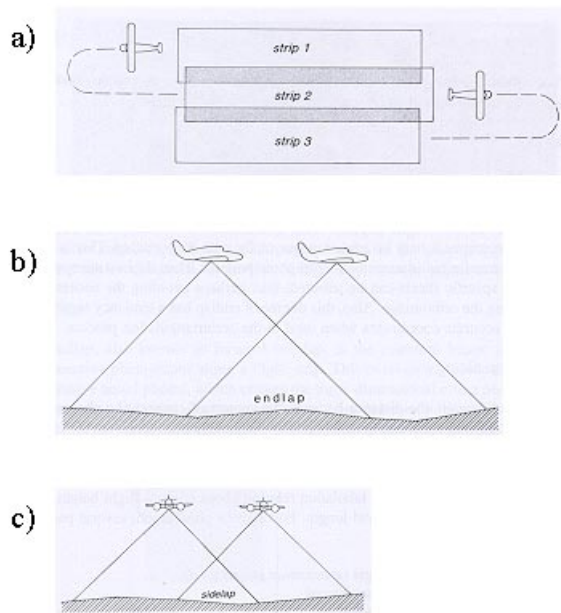
$$\blacksquare \quad g_{end} = s_p \times w \times \left[ \frac{(100 - o_{end})}{100} \right] \quad [3.4]$$

$$\blacksquare \quad g_{side} = s_p \times w \times \left[ \frac{(100 - o_{side})}{100} \right] \quad [3.5]$$

where  $g_{end}$  = endlap gain,  $g_{side}$  = sidelap gain,  $s_p$  = photo scale,  $w$  = width of exposure frame (typically 9 inches),  $o_{end}$  = endlap percent, and  $o_{side}$  = sidelap percent.



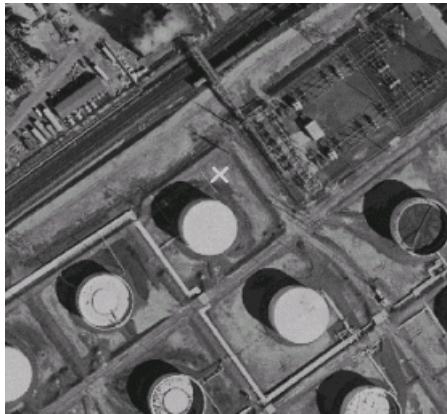
After the flight plan has been prepared, there is a unique set of site conditions that must exist in order to begin the photography. Conditions such as haze, humidity, season, and time of day have effects on the photographic quality and mapping accuracy.



**Figure 3.22: (a) Typical Flight Pattern over a Site, (b) Endlap, and (c) Sidelap (Falkner, 1995).**

Once the aerial photographs from a flight mission have been developed, the next step is orient the photographs to the correct terrain coordinates. This orientation procedure is referred to as georeferencing. Georeferencing is performed using a stereoplotter and the information obtained from a field control survey. A minimum of three horizontal and four vertical control points are required to georeference each stereomodel. The field control survey information can be obtained in one of two ways. One method is to select points from the photograph and then survey these points in the field. Alternatively, ground targets can be laid out and surveyed before the aerial photography begins. Ground

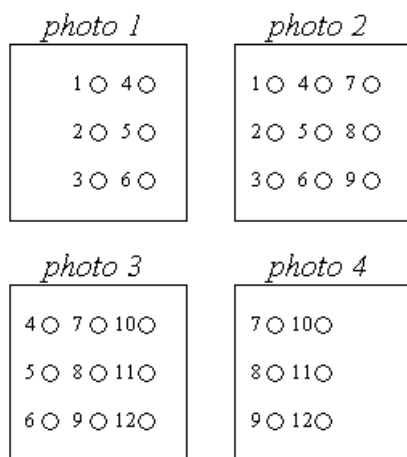
targets must be very visible so that they will be easy to spot on the resulting photographs (Figure 3.23). Surveying can be done using traditional theodolite equipment, or it can be done using Global Positioning System (GPS) devices, which measure the distances between the ground point and a minimum of four satellites. The resection of the divergent rays from the satellites establishes the geographic coordinates of the observing station. A combination of GPS receivers and traditional surveying techniques were used for the field surveying of 39 ground control points at Marcus Hook.



**Figure 3.23: Ground Target as Seen on Marcus Hook Orthophotograph.**

For mapping projects, an additional photo control procedure known as aerotriangulation is required. Also referred to as control triangulation and control bridging, aerotriangulation is an analytical procedure that allows a mapper to utilize a skeletal pattern of field survey control to analytically generate sufficient photo control points to map a project. The procedure begins by choosing six photo control points per stereomodel (Figure 3.24). These points are marked on diapositives, which are the clear-film copies of the aerial photographs that enable a compiler to project light through them and view them in stereoplottor. These artificially created points are referred to as pug points. After point pugging, the

stereomodel information is imported into a computer and processed through an aerotriangulation software package. Typically, this software package utilizes a Least Square Method to adjust the model coordinates to ground control information. Independent stereomodels are joined together using a series of equations that perform a three-dimensional transformation using common pass points. The final solution of the aerotriangulation process is a set of XYZ coordinates for each photo control point, which is imported into the computer to provide control data with which to georeference the stereomodels to the ground. By referencing the ground control points, (x,y,z) coordinates can then be determined for any feature within the model.

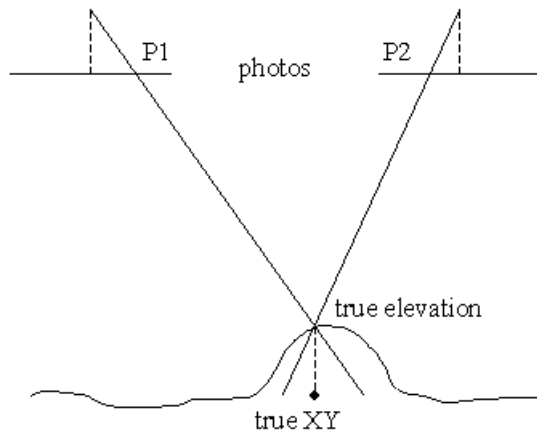


**Figure 3.24: Photo Control Points Required for 4 Consecutive Photographs (Falkner, 1995).**

Once the georeferencing is complete, the mapping phase and actual spatial data creation is ready to begin. Digital data are compiled using a mapping instrument, or digitizer, which exports data to a data collector within a computer. Features are digitized using points, lines, and arcs. After the digitizing procedure, the data are edited, layered into separate coverages, and then translated into the desired output (e.g., ARC/INFO, AutoCAD). Various software packages can be

utilized to aid in performing these processes. For the Marcus Hook mapping project, Kork software was used for modifying data and for generating ARC/INFO coverages (ASC, 1997). The mapping phase also involves determining topographic contours. Accurate contour data strings can be captured from aerial photographs using a stereoplotter, but they can also be captured using a digital terrain modeling (DTM) procedure. This procedure requires inputting a database of spot elevations into the computer, which then interpolates between points using a TIN matrix. Contours can be generated from the resulting TIN, and the resulting contour accuracy depends on how adequately the pattern of collected data points depicts the true shape of landforms (Falkner 158). As previously shown in Figure 3.14, this DTM procedure was used during the Marcus Hook mapping efforts.

Another common component of an aerial mapping project is image rectification. The photography scale calculated using equation 3.3 is actually only an average because of changes in the aircraft height and terrain relief between exposures. These changes in terrain affect not only the photo scale but also the horizontal accuracy of the placement of planimetric and topographic features on a photograph. Figure 3.25 shows the image displacement between two photos as a result of terrain relief. This figure also shows that a feature will be positioned at its true horizontal location if it is sited at its true elevation. In order to correct for image displacement, an orthophoto generation procedure is used to segment a raw image and reconstruct an orthogonal image rendition that fully rectifies image detail. All features in an orthophotograph are accurately related horizontally, and to an observer, it appears as if one is looking directly perpendicular to each feature (hence the name ortho). Figure 3.26 displays the orthophotograph for Marcus Hook.



**Figure 3.25: Image Displacement between Photographs as a Result of Terrain Relief (Falkner, 1995).**

Orthophotographs can be generated as a two-dimensional continuous tone negative or as three-dimensional raster data. Two-dimensional generation involves passing a variable-width slit device across the image while the operator maintains the reference mark on the surface of the spatial terrain. As the slit progresses across the photo, it exposes a negative on an underlying film base. This orthonegative can then be processed into a positive image. Three-dimensional generation, on the other hand, requires scanning the image and then relating these scanned data (i.e., raster data) to a digital terrain model. The coordinate triplet (x,y,z) of each image pixel is calculated in a similar manner to a DTM procedure, and the resulting image matrix has pixels with XYZ coordinates and a brightness value. For projects with multiple exposures, each exposure is rectified and georeferenced separately. These exposures are then organized in an orthophoto workstation, and mosaic lines are established for each exposure. Finally, the images are cut, pasted, and "feathered" together along the mosaic lines using a software program. This three-dimensional method was used for the Marcus Hook mapping, and the resulting orthophoto has a pixel resolution of 1 foot. Orthophotography is not required for all projects, but it is particularly useful

Marcus Hook Refinery Orthophoto (Black & White)

76

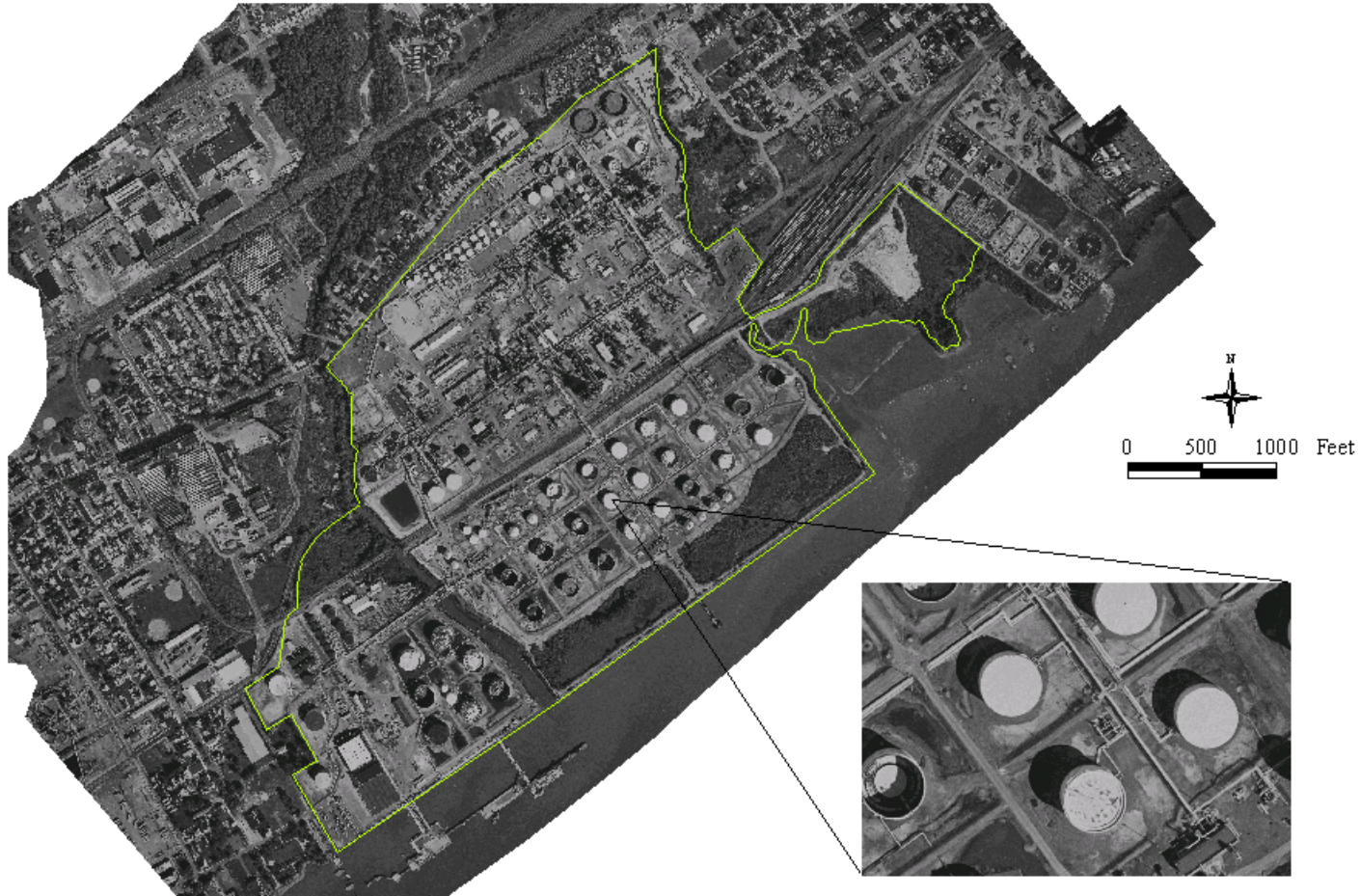


Figure 3.26: Marcus Hook Refinery Orthophotograph (1-ft. pixel resolution).

if the photograph is to serve as a base map, and accuracy is a prime consideration.

The effort involved in the aerial mapping process is quite extensive. Because of all the high-tech equipment and amount of effort required to map a site, the cost for a mapping project can be greater than \$100,000 in 1999 dollars. This option is particularly good for large and complex sites, which have a lot of features that need to be represented. For any particular project, the cost of the mapping effort could be compared to the expected cost of the overall corrective action so as to determine a relative cost to benefit ratio. However, aerial mapping is not the only method of facility scale coverage creation. If a user has access to a digitizing tablet, the process of digitizing features, assigning attributes, and layering coverages can be accomplished directly with ArcView using the *Digitizer* extension in connection with a digitizing tablet. ARC/INFO can also be utilized for digitizing purposes (e.g., with ARCCOGO and ARCPLOT). Since the digitizing is labor-intensive, significant savings may be realized by using these ArcView and ARC/INFO functions, but photographs or maps from which features can be digitized must already exist.

Another alternative is to create a spatial database using CAD drawings. CAD drawings are a popular method of mapping features at a facility, and these drawings can be viewed in ArcView utilizing the *Cad Reader* extension. Both the *Digitizer* and *Cad Reader* extensions are included in the standard ArcView installation package. CAD drawings are added to a View in the same way a theme is added. Whatever coordinate system has been set for the drawing within the AutoCAD program will be the coordinate system within ArcView. However, a drawing can be converted into another projection using a procedure similar to that described in the Map Projections section. For Marcus Hook, many of the CAD drawings that were referenced according to the plant grid coordinate system were added to ArcView projects and then transformed into the Pennsylvania State Plane Coordinate System. This transformation involved creating a World

Transformation File (\*.wld) in Notepad and then assigning this file to the drawing by modifying the theme properties in ArcView. The transformation file includes two coordinate pairs as follows:

u1,v1 x1,y1  
u2,v2 x2,y2

where (u1,v1) and (u2,v2) are points in the plant grid coordinate system and (x1,y1) and (x2,y2) are the same points in the state plane coordinate system.

The u and v coordinates were read off of the drawing theme in ArcView before transformation, and the x and y coordinates were read off of the facility basemap coverages. For accuracy purposes, a different transformation file was used for each CAD drawing. Figure 3.27 shows a CAD drawing from Marcus Hook that has been imported into ArcView, transformed into the correct projection, and separated by layers.

Determining the best method of constructing a spatial database depends on the desired level of accuracy, the site characteristics, any existing information, the development time, the availability of qualified personnel, and the available financial resources. As previously mentioned, the cost for an aerial mapping project runs very high but provides very accurate and easily manageable data. If experienced personnel are available to complete the feature digitizing, an aerial survey company could be contracted just to produce the photography. The photographic component of an aerial mapping project may be only 20% of the total cost (ASC, 1997). The resulting photographs could then be utilized by the user for digitizing features. The accuracy will not be as high because of limited equipment, but costs will be significantly reduced. More detailed information on aerial mapping projects, such as accuracy and scale effects on costs, should be obtained from an actual aerial survey company. Another alternative discussed for creating a spatial database is using CAD drawings. Many times, these drawing



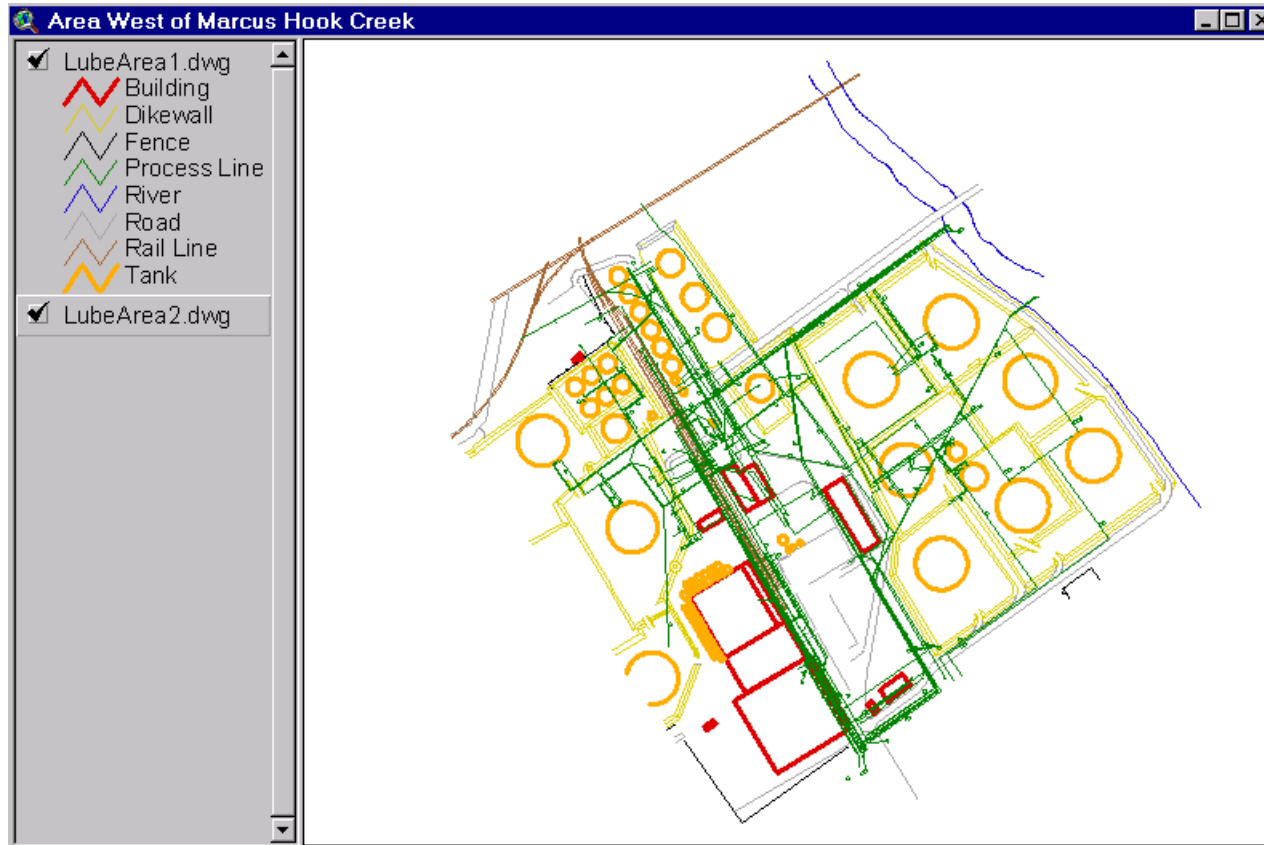


Figure 3.27: Lube Plant Area CAD Drawings in ArcView.

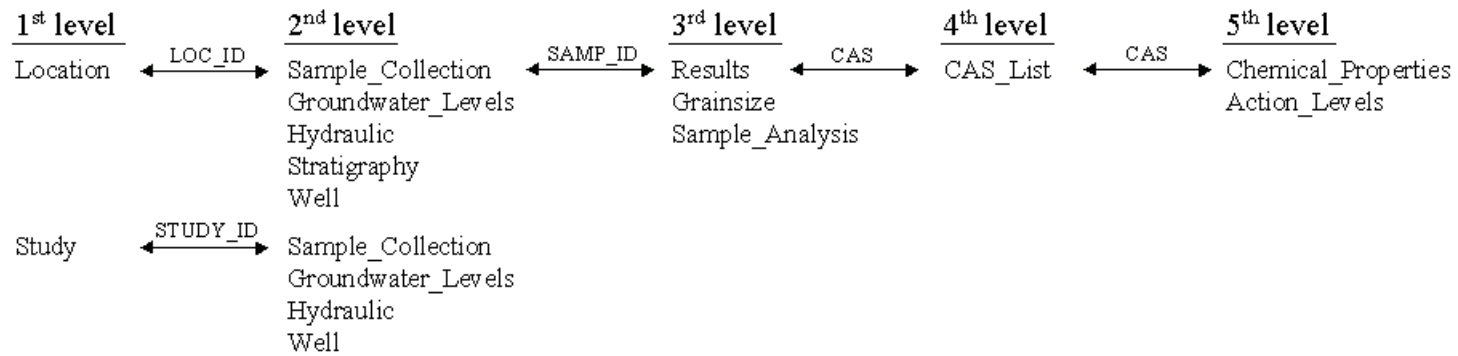
already exist, and the process of incorporating them into a spatial database is very inexpensive. The accuracy is dependent on the original construction of the drawings and any projection conversion errors, which would be expected with a conversion using only two points. An additional difficulty with this method is that the data are not as easy to use. Coverages and shapefiles are much easier to use than CAD drawings, and while CAD drawings can be converted into usable coverages, the process is not very straightforward and requires an experienced GIS user to execute it.

#### **3.4.4 Constructing and Maintaining a Tabular Database**

In order to understand the considerations in the construction of a database, the creation of the Marcus Hook tabular database is described here. As previously mentioned, this database design was originally developed by Parsons Engineering Science using Oracle software. Parsons also populated the database using information that was available from various reports and studies. For instance, the Well table contains information that was found in the well drilling logs such as the casing types, screen lengths, and well casing elevations. Similarly, the Results table contains information that was reported on laboratory data sheets such as the concentrations, detection limits, and dilution factors. After Parsons assembled the data, they exported each individual table from Oracle into dBase files (\*.dbf). These tables were then imported and re-linked in Access. The database originally developed by Parsons was designed to include data from multiple sites. Therefore, a number of fields required for this structure, such as facility name, were not needed for the purposes of this project and were deleted from the database. Any fields that did not have any data associated with them and were not expected to be populated in the future were also eliminated (over 75 fields). The most significant structure modification from the original Oracle design was with respect to relational linkages. The original design contained multiple links

between tables. For instance, the Sample\_Collection, Sample\_Analysis, and Results tables all contained the QC\_CODE, FILTERED, and SAMP\_ID fields, and these fields were linked through all three tables. This multiple linkage and repetition of data were not needed, so these fields now exist in only one table, and each table is linked to the other tables through only one field. As previously mentioned, when the database was constructed in Access, the data were layered according to differing levels of detail. The data levels and the tables associated with each level are illustrated in Figure 3.28. The information contained in the 2<sup>nd</sup> level supports the information contained in the 1<sup>st</sup> level, and so forth. For instance, the Sample\_Collection table provides information on the samples collected at particular locations, and the Results table provides information on the recorded measurements for each sample.

Besides structure modifications, many challenges were also encountered with the data sets. For instance, some of the results data contained in the collection of reports were excluded from the database. This information included data quality levels, dilution factors, and detection limits, all of which are particularly important for understanding uncertainty in the results. To overcome this challenge, a query was created using the Study, Sample\_Collection, and Results tables to determine from which report each record in the results originated. The desired information was then searched for in the original report, and a new results table was created to replace the existing one. Some data input errors were also encountered which resulted in geoprobe sampling locations appearing off-site. Again, the original report was consulted and the spatial coordinates were corrected. In this case, the easting and northing values (according to the plant grid coordinate system) had been switched when the data were input into the database.



**Figure 3.28: Tabular Database Data Levels.** The arrows represent links between tables through the common field listed above the arrow.

After the tabular database is constructed, it must be maintained so that the digital facility description continues to function properly. One of the most important considerations is how to input new data into the database. New data must be organized into a format that is consistent with existing site data. This organization eases the transition of new data into the database. For instance, the LOC\_TYPE field in the Location table of the Marcus Hook database delineates the sampling location type, such as well, soil boring, or geoprobe. All of the soil borings in this table are labeled as "SOIL BORING." If new records are added to the database with a location type of "BORING," then a query of benzene concentrations from borings would not incorporate all of the data. In order to address this issue for the Marcus Hook project, a set of input protocols was developed so that anyone who added new data knew the proper format. The data collector compiled any new information using spreadsheets, which had the same columns as the tables in the database. For example, if concentrations were analyzed from existing groundwater wells, then the data collector filled out a blank spreadsheet of the Sample\_Collection and Results tables (Figure 3.29). The tabular database dictionary was designed to guide the user in adding these data in the correct format (Appendix C). After these spreadsheets were completed, they were added to the database using the TransferSpreadsheet macro in Access (Figure 3.30). This macro is a built-in component of Access that a user can create simply by filling out the Action Arguments shown in the figure. This input system has since been replaced with data forms that reside in the database. These forms are utilized directly in the data collection and analysis processes.

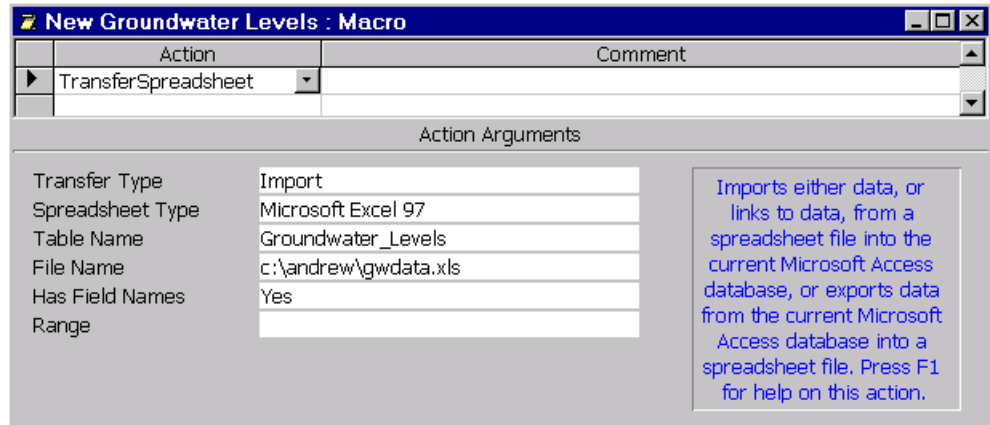
The construction and maintenance of a tabular database could vary considerably between projects. Some projects like Marcus Hook require an extensive effort to compile and then reformat different data sets. Other projects might already have a digital database of information that only requires minimal restructuring. The goal is to develop a product that is easy to use and can support

SAMP_ID	LOC_ID	SAMP_DATE	DEPTH_TOP	DEPTH_BOT	DEPTH_UNIT	STUDY_ID	MATRIX	QC_CODE	LOAD_DATE	LOAD_BY	COMMENTS

SAMP_ID	ANAL_METH	CAS	VALUE	DET_LIMIT	UNIT	DIL_FACT	VALUE_QUAL	VALID_QUAL	REP_HIT	QC_FLAG	LOAD_DATE	LOAD_BY	COMMENTS

**Figure 3.29: Entry Spreadsheets for Sample\_Collection and Results Tables.**

84



**Figure 3.30: TransferSpreadsheet Macro in Access.**

the risk-based decision-making process. The database design outlined in Figure 3.16 has been very effective for the Marcus Hook project, particularly with respect to setting up queries and viewing these queries in ArcView. However, the reader is encouraged to explore other options for data management.

### **3.5 Connecting Spatial and Tabular Data**

An effective risk assessment requires an analysis of both physical and chemical data. Thus, some sort of tool is needed to connect the spatial database with the tabular database. Concentration measurement locations can be queried in Access, but it is difficult to visualize these locations from a table of data. Similarly, a map of facility features is useful for display, but this map needs to be connected with environmental data in order to evaluate exposure pathways. This is where the power of GIS is demonstrated. The basic setup of GIS, in which features on a map are linked with records in a table, provides the necessary framework to accomplish such a task. Systems such as ArcView allow the user to import selected sets of data from another database utility into the GIS. The imported data can then be joined to an existing coverage or used to create a new coverage.

Data can be imported from a database utility into ArcView using an Open Database Connectivity (ODBC) driver. ODBC is a programming interface that enables applications such as ArcView to access data in data management systems that use Structured Query Language (SQL) as a data access standard (Microsoft Corporation, 1996). ODBC capabilities are included in the basic installation of Microsoft Office Professional. They can be accessed on a PC through the Control Panel under 32-bit ODBC. Besides Access, ODBC drivers can be used to connect to dBase files (\*.dbf), FoxPro files (\*.dbf), and Excel files (\*.xls). These applications are all Microsoft related, but an ODBC driver can be set up for most database utilities. If a tabular database program other than Access is being

utilized, information about ODBC should be obtained from the specific software provider.

The ODBC process is simple and it establishes a dynamic link with ArcView, meaning that any time data are updated in the tabular database, these data are also updated in ArcView environment. The process begins by setting a data source using the ODBC Data Source Administrator, which provides the link for ArcView to the database. Data can then be imported from the tabular database into ArcView using the **SQL Connect** command in ArcView. A dialog box appears that allows the user to select data from either tables or queries in the Access database. The selected data form a new table in ArcView, which can be connected to an existing coverage provided there is a common field in each table. However, an even more effective method is to include the spatial coordinates of the measurement location in the selected data set so that a new coverage can be created. New coverages can be created with Avenue scripts or with the **Add Event Theme** command in ArcView, which is the most flexible option. The event theme is specific to the active ArcView project, so each time the ArcView project file is opened, the database is queried and the event theme is updated. The event theme can be converted into a shapefile for transfer between projects, but this shapefile will not be updated as information changes in the tabular database. A potential difficulty arises if the data source name or its file path change. The connection between the ArcView project and the Access database is then lost. This issue is resolved by respecifying the path in the ODBC Data Source Administrator.

Appendix D contains an exercise created by the author for the GIS in Water Resources class entitled *Mapping Environmental Data Stored in Microsoft Access*. This exercise documents the entire process of connecting tabular and spatial data, including how to create queries in Access, how to setup the ODBC driver, how to import data into ArcView, and how to use these data in ArcView



for analysis. Some of the example analysis applications include contouring measurements and solving probabilistic simulations.

The link between the tabular and spatial databases is really the key component of the spatial environmental risk assessment methodology and digital facility description. Some of the maps that have been created using this link for the Marcus Hook project are shown in Figures 3.31 - 3.36. Figure 3.31 displays the sampling locations for Marcus Hook. This figure is simply a map of the Location table in the tabular database and shows where data have been collected. Figures 3.32 and 3.33 show the groundwater elevations and benzene concentrations, respectively, at a specified time. The *Spatial Analyst* extension for ArcView was used to interpolate contours between the point measurements. For each figure, the contours were created using an inverse distance weighting (IDW) interpolation scheme, which interpolates a value for each point in the study area using the inverse distance squared between the location of the current point and the measurement location as a weighting parameter. These figures are useful because one shows the patterns of groundwater flow and the other shows the areas with high benzene concentrations. However, they are not necessarily representative of the environmental conditions at the site. A more detailed analysis would focus on individual areas of the site, incorporate barriers (e.g., surface waters and fault lines), and study the changes between time steps.

There are various means of cartographic representation of point data, and each method is focused on using the most effective means of understanding the data set. The benzene soil concentrations presented in Figure 3.34 are broken down by different colors for each magnitude range. Of particular importance are the no-detect values, which are separated with an outline around the point. In Figure 3.35, the highest recorded concentration at any time for every detected chemical are displayed according to different symbols for each chemical class. This representation not only allows multiple results for a single point to be displayed

together, but also allows for a data pattern characterization that could not be achieved looking at individual chemicals. For instance, the highest detected concentrations of semi-volatile organics appear in north-central area of the facility. Point measurements can also be displayed with physical features as in Figure 3.36. This figure shows LNAPL thickness measurements in the Lube Plant and how these measurements from a specified time are spatially related to structures and surface hydrology. The representation of temporal variations in the tabular database with spatial variations in the GIS is a powerful tool for environmental risk assessment. Future uses of the tabular and spatial link will include more development of the spatial site conceptual model and visualization of uncertainty in the results. Another Access database has been designed for the site conceptual model, and this database will eventually be connected to both the tabular and spatial databases (Koerner, 1998).

Marcus Hook Refinery - Sampling Locations

- ∟ Facility Boundary
- ▲ Soil Vapors
- Soil Borings
- Geoprobes
- ★ Sediments
- Piezometers
- Wells
- × Sumps
- + Sludge Basins
- × Test Pits/Trenches

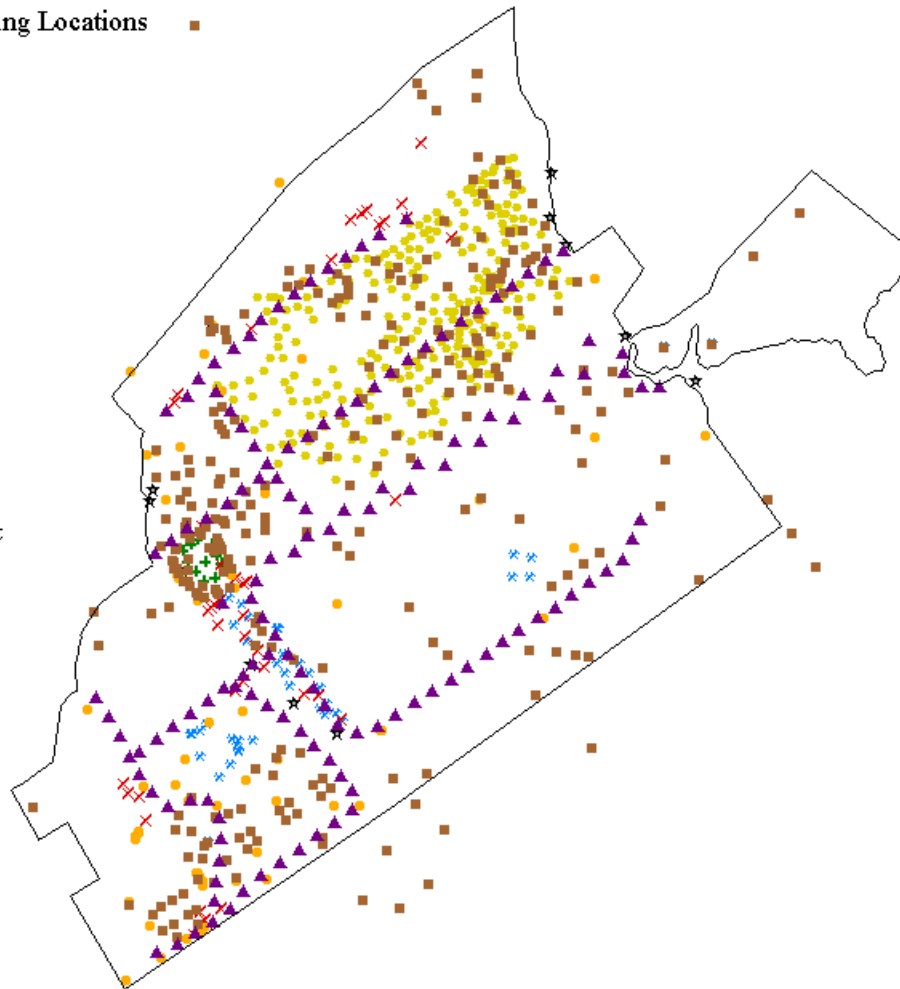
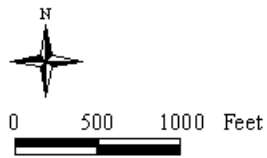


Figure 3.31: Marcus Hook Refinery Sampling Locations.

Marcus Hook Refinery - Groundwater Elevations on 8/17/98

- Facility Boundary
- Groundwater Elevations (ft - NAVD88)
- Groundwater Contours (1 ft intervals)
- Surface Hydrology

06

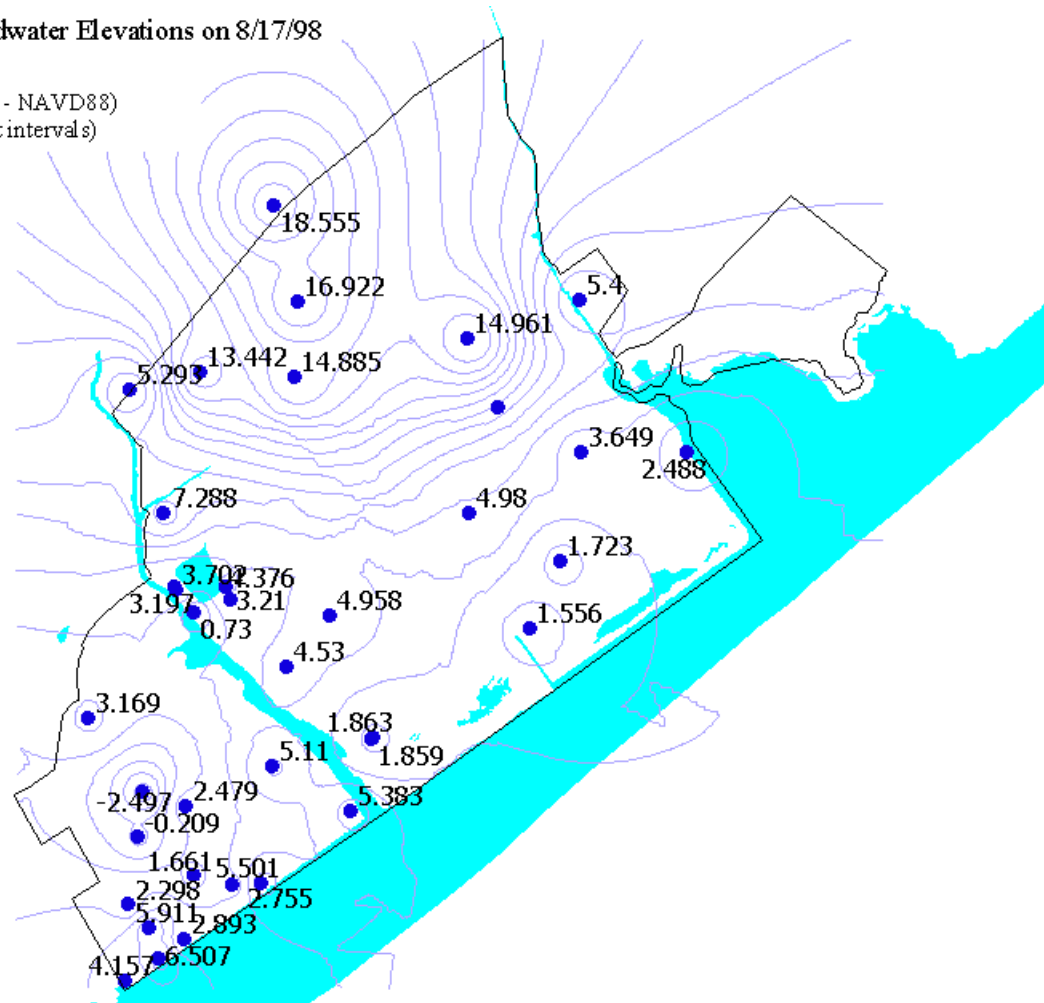
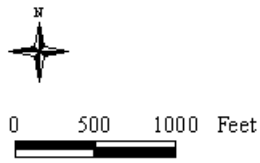


Figure 3.32: Marcus Hook Refinery Groundwater Elevations on 8/17/98.

Marcus Hook Refinery - Benzene Concentrations in Groundwater (1996)

- Facility Boundary
- Benzene Groundwater Concentrations (ug/L)
- Benzene Contours (1000 ug/L intervals)

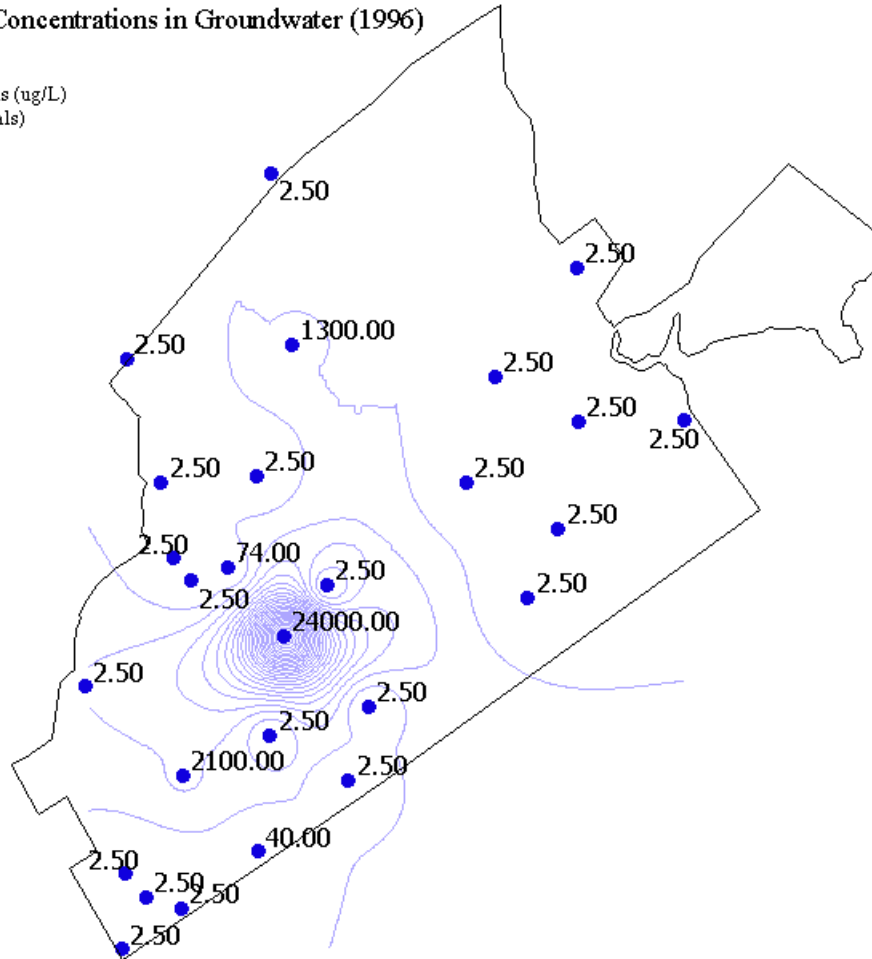
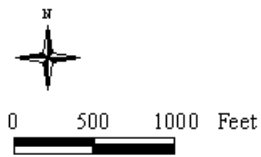


Figure 3.33: Marcus Hook Refinery Benzene Concentrations in Groundwater (1996). The 2.5 values are no-detects with detection limits of 5  $\mu\text{g/L}$ . Half of this value (i.e., 2.5) was used for contouring purposes.

Marcus Hook Refinery - Benzene Concentrations in Soil 2-8' Below Ground Surface (1996)

- Facility Boundary
- Benzene Soil Concentrations (mg/kg)
  - 360 - 725
  - 135 - 360
  - 76.8 - 135
  - 12.5 - 76.8
  - 1 - 12.5
  - No-Detect

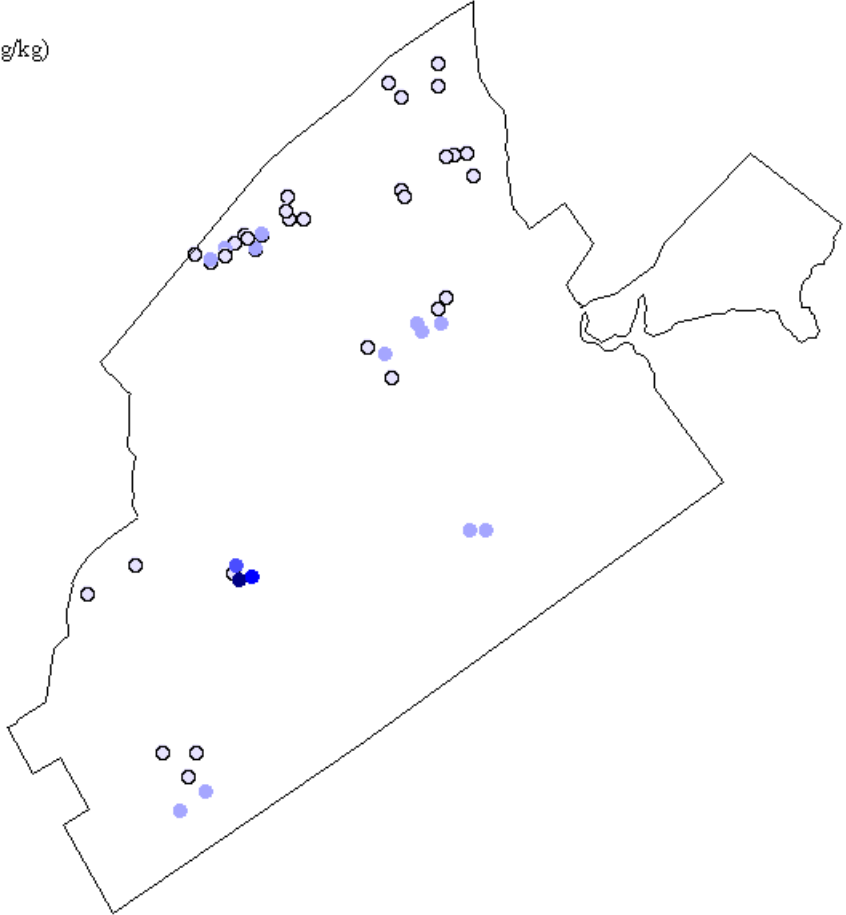
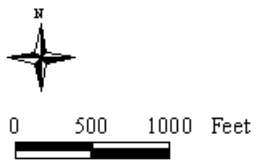


Figure 3.34: Marcus Hook Refinery Benzene Concentrations in Soil 2-8' Below Ground Surface (1996).

### Marcus Hook Refinery - Highest Detected Concentrations in Groundwater

- Facility Boundary
- Maximum Concentrations in GW
- Anions
- Metals
- Other Organics
- Semivolatile Organics
- Volatile Organics

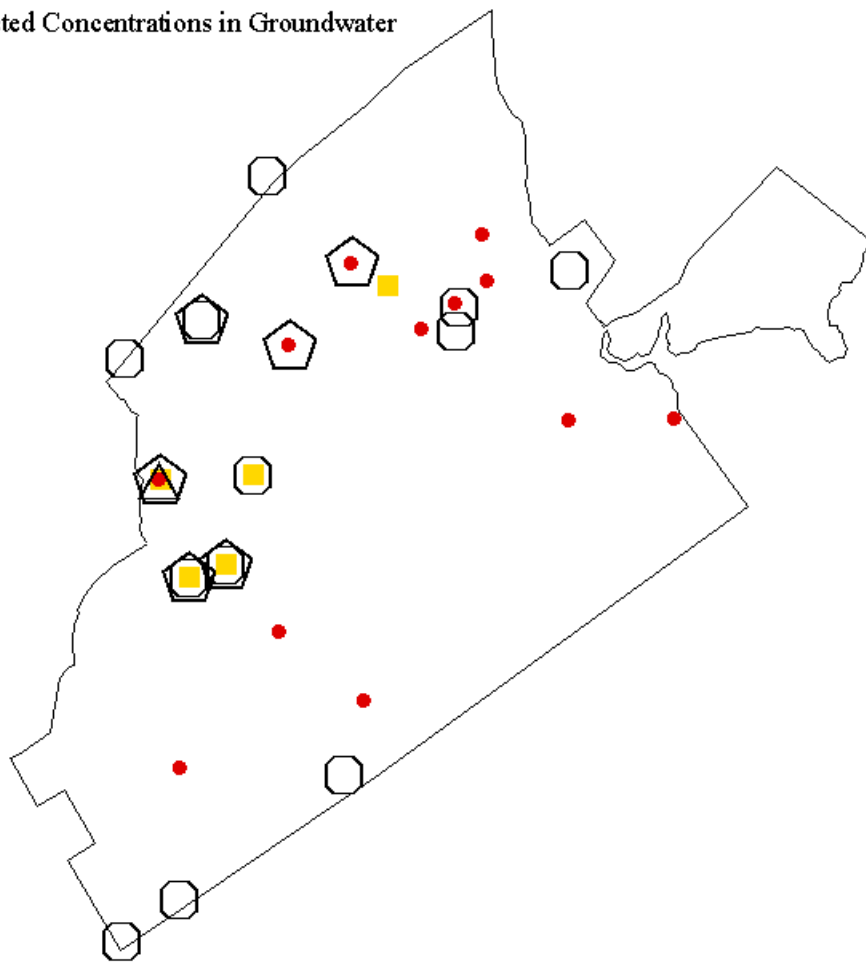
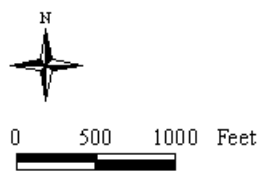


Figure 3.35: Marcus Hook Refinery Highest Detected Concentrations in Groundwater.

### Marcus Hook Refinery - LNAPL Thicknesses in Lube Plant Area

- △ Area West of MH Creek
- LNAPL Thickness (ft)
  - 0
  - 0 - 0.5
  - 0.5 - 2.5
  - 2.5 - 4.5
  - 4.5 - 6.5
  - > 6.5
- Buildings
  - 
  - ▨ Storage Tanks
  - Surface Hydrology

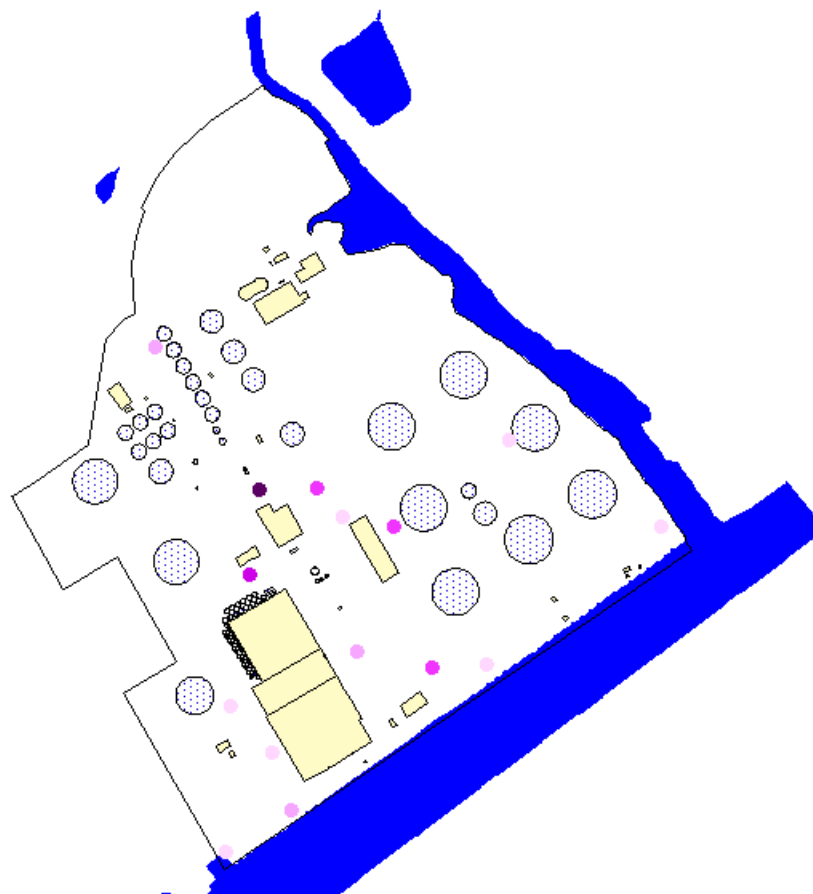
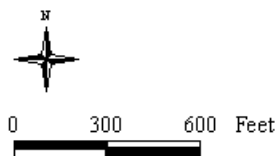


Figure 3.36: LNAPL Thickness Measurements in Lube Plant Area.



## **CHAPTER FOUR**

### **MAP-BASED MODELING**

#### **4.1 Surface Water Models**

This section documents the development of surface water runoff models for the Marcus Hook Refinery and surrounding area. These models provide insight into how COCs move from a surficial source area to a receptor. Because there are three surface water bodies into which runoff drains from the facility, this transport mechanism may be a particularly important exposure pathway in the risk assessment analysis. The process of developing these models involves delineating water drainage patterns using an elevation grid and calculating downstream concentrations based on user-defined point or area sources. These steady-state (i.e., time-averaged) models produce conservative results that are typical of Tier 1 and Tier 2 RBCA analyses (ASTM, 1998). All of the surface water modeling steps can be performed completely within the ArcView system.

##### **4.1.1 Methodology**

There are two major components to developing surface water models for chemical release sites: (1) determining the flow patterns and amounts of surface water runoff for the study area and (2) determining the actual chemical migration based on these runoff characteristics. The first component involves delineating water flow patterns using a digital elevation model (DEM). Each grid cell in a DEM has an associated elevation, and since water flows from high elevation heads to low elevation heads, flow between cells is determined by the path of steepest descent. Once the flow between cells is known, an accumulation scheme can be run to determine how many cells are upstream of any given cell in the DEM. Then, using the flow direction and flow accumulation results, the modeling software can be used to delineate streams and watersheds. All of these

steps can be performed completely within ArcView, and for the Marcus Hook modeling, the *Watershed Delineator* extension was utilized. This extension was developed by ESRI in conjunction with the Texas Natural Resources Conservation Commission (TNRCC). The applicable steps of the *Watershed Delineator* are similar to those found in the similar *CRWR-PrePro* and *Hydrologic Modeling* extensions. More information on all three of these extensions is available from Appendix A. Because all the calculations are grid-based, the ArcView *Spatial Analyst* extension is also required.

The development of surface water runoff models with the *Watershed Delineator* proceeds through six program steps: (1) fill sinks, (2) flow direction, (3) flow accumulation, (4) stream definition, (5) stream segmentation, and (6) watershed delineation. During the fill sinks step, the program searches for all the elevation depressions in the DEM where water could accumulate. The program then fills these depressions, ensuring that all water which falls on the terrain drains to an outlet at the edge of the grid. In the flow direction step, the program determines how water flows from one cell to another. Water can drain from a given cell into any one (but only one) of the surrounding eight cells. In the flow accumulation step, the program determines the number of cells upstream of the given cell based on the flow direction model. Streams can then be defined based on a threshold number of upstream cells using the flow accumulation grid. Once the streams are defined, they can be broken down into separate segments or links. In the final step, the links grid and flow direction grid are used to delineate watersheds. Both the regional and facility-specific runoff characteristics for the Marcus Hook facility were developed using this procedure, and these models are further discussed in the next two sections.

The second major component of surface water modeling is predicting concentrations in the study area based on known inputs to the system. In order to accomplish this aspect of the project, the author has developed a new ArcView

extension called *COC Transport*. *COC Transport* enhances the above modeling process to create a predicted concentration grid based on user-defined system inputs. The conservation of mass principle requires that the sum of inputs to a system must equal the total output plus or minus any changes within the system. Assuming the losses in the system are minimal, the sum of the input loads equals the total output load:

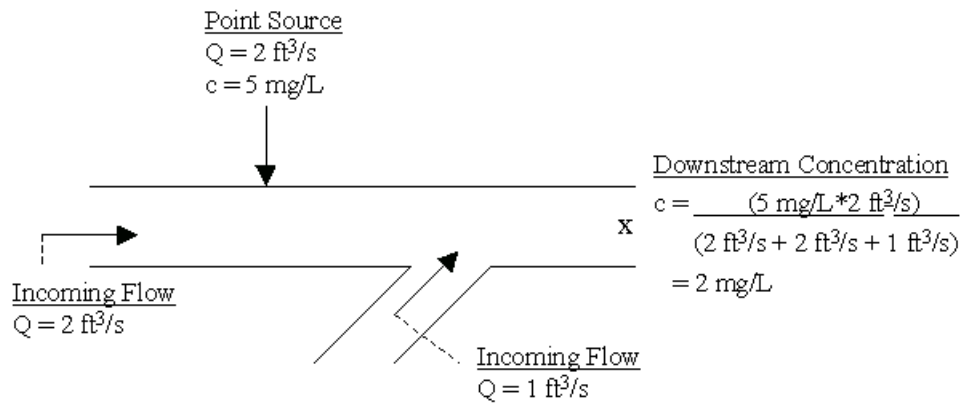
- $W = Q * c * 28.31682$  [4.1]

- $\sum(Q_{in} * c_{in}) = Q_{out} * c_{out}$  [4.2]

- $c_{out} = \frac{\sum(W_{in})}{(Q_{out} * 28.31682)}$  [4.3]

where  $W$  = load [mg/d],  $Q$  = flow [ft<sup>3</sup>/d],  $c$  = concentration [mg/L], and 28.31682 is a conversion factor.

Load is the product of flow and concentration, and it represents the amount of mass input to a system on a time basis. Thus, the concentration at any point in the model can be calculated by summing the inputs to that point and dividing by the output flow. The resulting concentration is a conservative assumption because it does not account for any chemical decay or other losses, such as sorption. The concentration between two points can only decrease because of additional flow input (i.e., dilution). Figure 4.1 shows a simple example that illustrates the *COC Transport* methodology.



**Figure 4.1: COC Transport Methodology.**

*COC Transport* can incorporate inputs from both point and non-point sources. Point source examples include treatment plant outfalls and pipe discharges while non-point source examples include runoff from farmland and urban stormwater runoff. The user has the option to create point sources, area sources (i.e., non-point sources), or both. A series of dialog boxes appear asking the user to define an input flow and either a concentration input or a load input. If no input flow is selected, then the runoff flow at that point is used instead. Once the system inputs are set, the program leads the user through a series of steps to arrive at a predicted concentration grid (Figure 4.2). The first step is to create an input concentration grid based on the defined inputs. Next, a flow grid is calculated using flows from the input coverage and a runoff grid. This flow grid is then accumulated, and a load grid is computed by multiplying the concentration grid by either the flow grid or accumulated flow grid, depending on the type of input coverage. For a point source input coverage, the concentration grid is multiplied by the accumulated flow grid, and for an area source input coverage, the concentration grid is multiplied by the flow grid. The assumption is that for a point source, the flow should be equal to any input flow plus any upstream flow while for an area source the concentration is assumed to be the same value throughout the area, so

Flow Direction

↓	↓	↙
↙	→	↓
→	→	↘

a) Point Sources: Concentrations

0	0	0
2	0	3
0	0	0

\* Accumulated Flow

0	0	0
1	2	3
0	3	8

= Load

0	0	0
2	0	9
0	0	0

→ Accumulated Load

0	0	0
0	0	0
0	2	11

/ Accumulated Flow

0	0	0
1	2	3
0	3	8

= Predicted Concentrations

0	0	0
0	0	0
0	0.7	1.4

66

b) Area Sources: Concentrations

3	1	1
3	0	0
2	2	0

\* Flow

1	1	1
1	1	1
1	1	1

= Load

3	1	1
3	0	0
2	2	0

→ Accumulated Load

0	0	0
3	2	2
0	8	12

/ Accumulated Flow

0	0	0
1	2	3
0	3	8

= Predicted Concentrations

0	0	0
3	1	0.7
0	2.7	1.5

Figure 4.2: COC Transport Program Steps Example.

the flow used in the calculation of load for each grid cell should simply be the flow for that grid cell. After the load grid is calculated, it is accumulated and then divided by the accumulated flow grid to create a predicted concentration grid. Essentially, for each grid cell, the sum of the loads to that point is divided by the sum of the flows to that point.

The *COC Transport* extension is still in a developmental phase, and it has only been tested for a couple of different scenarios. The original concept was to create an extension that could be easily applied and repeated as the modeling dynamics change. This goal has been accomplished, and the extension is generic enough for anyone to use. For instance, the user has the option to create input and runoff coverages or to use preexisting coverages. The only requirements are the *Spatial Analyst* extension and a flow direction theme. The extension can be downloaded at <http://www.ce.utexas.edu/prof/maidment/grad/romanek/research/data.html>. Additional documentation on the process steps and scripts is provided in Appendix F.

#### **4.1.2 Regional Model**

The development of a regional terrain model is important for several reasons. First, it provides an estimate of how much upstream water drains into the two creeks that run through the facility, Marcus Hook Creek and Stony Run Creek, along with the Delaware River. It also identifies runoff from other facilities that could affect either of these creeks. Utilizing the capabilities of ArcView and free spatial data available on the Internet, one can delineate watersheds and then determine which facilities fall in each watershed. This area of Pennsylvania is highly industrialized, and the Marcus Hook Refinery is not the only facility contributing discharges and drainage to the two creeks. Therefore, a screening analysis of upstream impacts is used to estimate background concentration levels in the two creeks. An understanding of the changes in these background levels

due to the Marcus Hook Refinery can then be gained. If this screening assessment indicates the potential for impacts from the refinery to the creeks, then empirical data collection may be necessary.

In order to characterize runoff in the region, the *Watershed Delineator* extension was used in conjunction with some of the coverages included in the spatial database. Specifically, the Marcus Hook city DEM and the coverages of streams in Delaware County, annual runoff, and EPA regulated facilities proved to be particularly useful. The DEM has a grid cell size of 30 meters, meaning that the resulting patterns of water flow are based on drainage from each 900 square meter area. This DEM, which has been cut to the Pennsylvania state boundary line, is shown in Figure 3.7. This DEM was clipped because it was deemed that the elevations on the other side of the Delaware River in New Jersey and southwest of the facility in Delaware were not needed for the analysis.

Before proceeding through the *Watershed Delineator* program steps, the county stream coverage was "burned" into the DEM so as to ensure proper flow accumulation when the model was run. The process of burning in streams requires that the elevations of each stream in the DEM remain constant while the rest of the DEM is raised by some arbitrary value, such as 2000 feet. This arbitrary value must be greater than the highest point in the DEM. Even though the elevations in the burned DEM are not accurate, the differences in elevation between each cell remain constant except at the stream edges. Therefore, the modeled flow direction grid will only be different from the grid that would have resulted with the original DEM around the streams where water should be draining anyway. The burning process simply takes advantage of what is already known about water drainage in the study area. This process can be accomplished using the Map Calculator available with *Spatial Analyst*. The following steps are performed to burn in streams to a DEM: (1) convert the stream coverage to a grid, (2) divide the stream grid by itself, (3) multiply the new stream grid by the DEM

(demstream), (4) add an appropriate offset value to the DEM (demplus), and then (5) merge demstream and demplus. More information on the process can be accessed through the link in Appendix A, and Appendix G contains an Avenue script (burner.ave) that can be used to accomplish the task.

After burning the stream coverage into the DEM, the *Watershed Delineator* was used to model the regional terrain. Figure 4.3 shows the results. Each green area is a watershed, and the red dots represent EPA regulated facilities. There are 39 facilities that lie in the upstream watersheds of Marcus Hook Creek and 24 facilities that lie in the upstream watersheds of Stony Run Creek. The streams in this model have been delineated using a threshold value of 500 upstream cells, meaning that a cell was only identified as a stream if there were at least 500 upstream cells. This threshold provided the closest match of the delineated streams to the existing stream coverage.

The analysis described above was further extended to examine flow magnitudes based on actual measurements of runoff. A runoff value of 20 in/yr for the Marcus Hook facility and surrounding area was obtained using the annual runoff coverage downloaded from the USGS website (Appendix B). This coverage was constructed using available data for the United States from 1951-1980 and is a set of contour lines (Figure 4.4). The difficulty with using these data directly is that none of the contour lines run through the facility. Even though the coverage is fairly detailed, this result is not surprising because the size of the facility is so small compared to the scale seen here. One potential alternative to solve this problem is to interpolate a surface (i.e., a grid) from this set of contour lines. However, the features of *Spatial Analyst* only allow for a surface to be interpolated from a point coverage.



### Regional Surface Water Runoff Results

- Facility Boundary
- EPA Regulated Facilities
- Streams (500)
- 1
- No Data

103

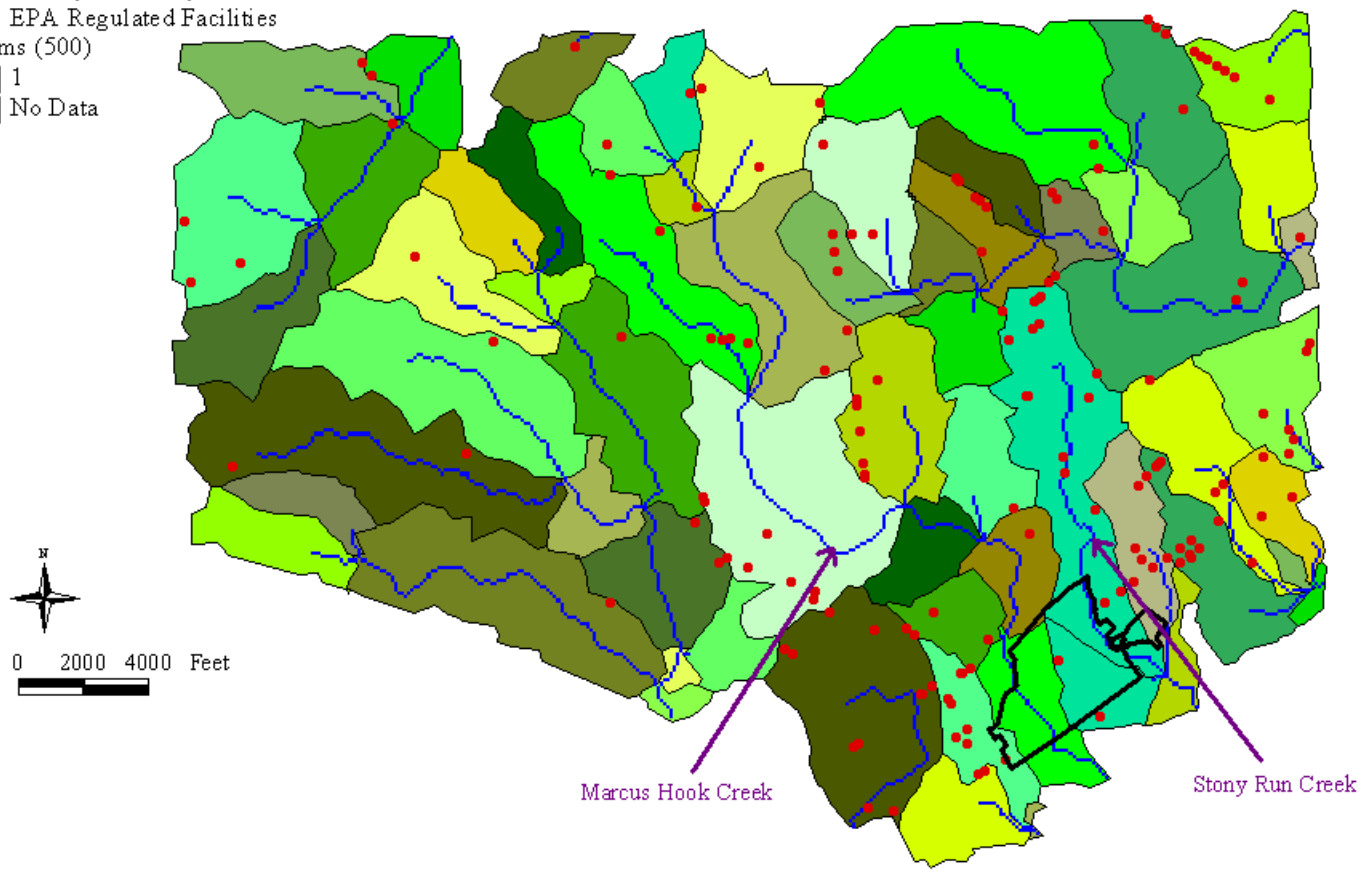
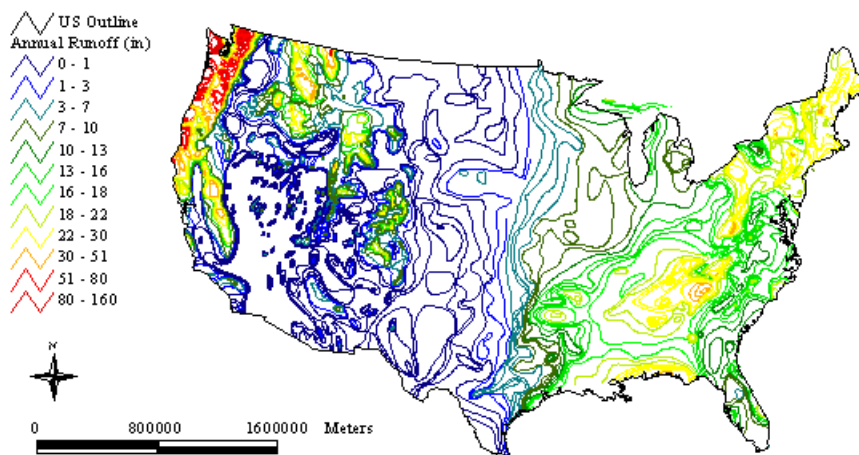


Figure 4.3: Regional Surface Water Runoff Model Results.

### Annual Runoff in United States (Inches)



**Figure 4.4: Annual Runoff in the United States (1951-80).**

Other alternatives are available for creating a runoff surface. One method involves creating a grid and then intersecting this grid with the runoff contours. The resulting coverage is also a line coverage, but there is a node at every location where the grid intersected the contour lines. The nodes can then be converted to a point coverage where each point has a runoff value equal to the contour line from which it is generated. A runoff surface can then be interpolated from the point coverage. While this method was not pursued due to its complexity, these tasks can be performed using ARC/INFO. An alternative method is simply to create a point coverage from the contour lines. The method is not as elegant because a point will be created wherever there is a bend or change in the contour line. However, the procedure is simple, effective, and can be performed completely within ArcView. Appendix G contains a script (interpol.ave) created by the author to interpolate a surface from a set of contour lines (Figure 4.5). This script was used to generate a runoff grid for the entire United States with a cell size of 1 km and a runoff grid for Pennsylvania with a cell size of 500 m as shown in Figures 4.6 and 4.7, respectively.

### Script Methodology

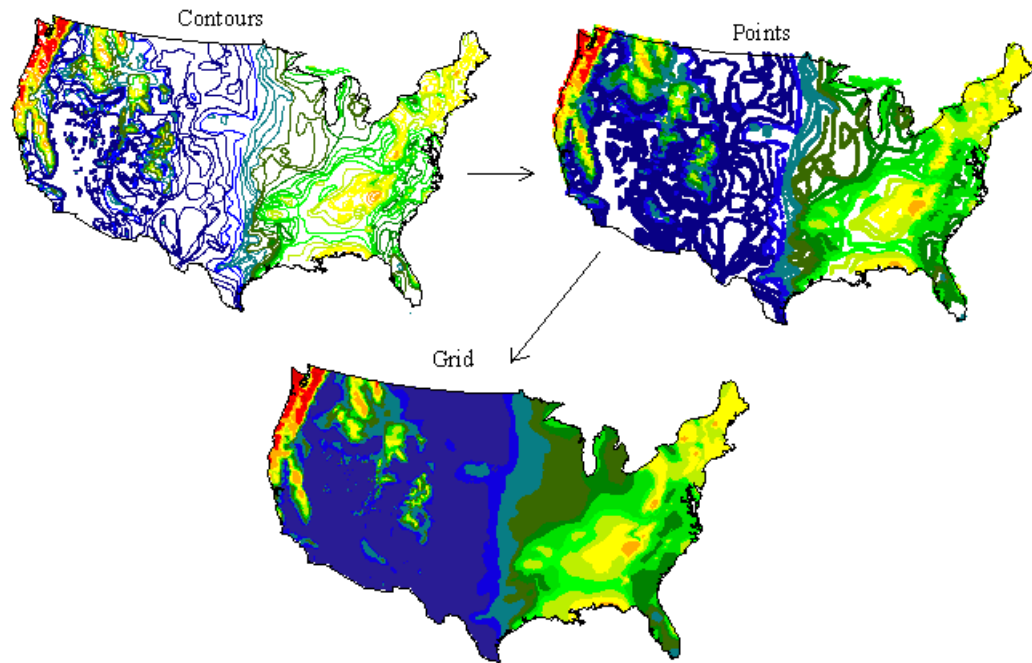
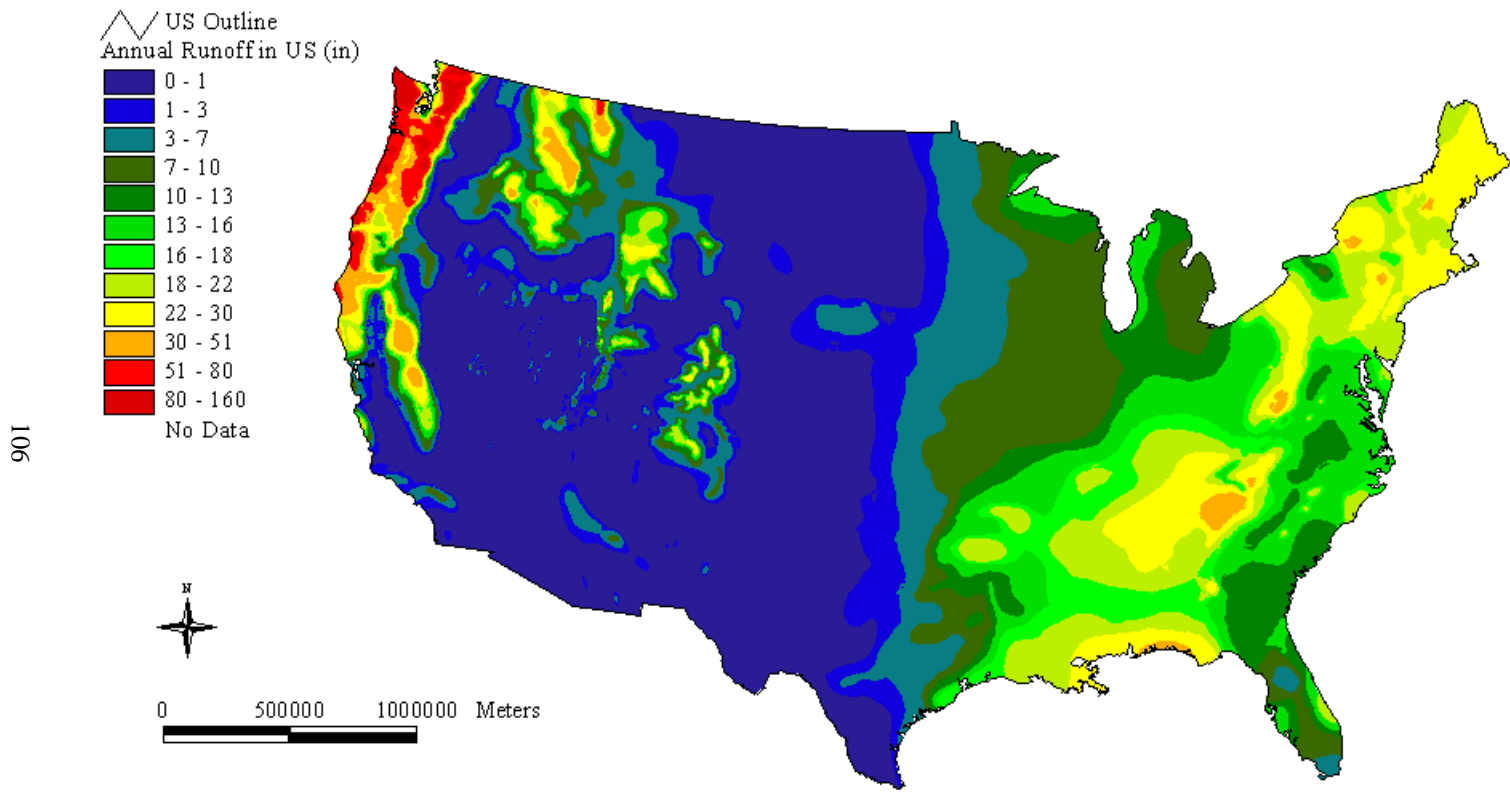


Figure 4.5: Interpol Script Methodology.

**Annual Runoff in United States (Inches)**



**Figure 4.6: Annual Runoff Grid for the United States.**

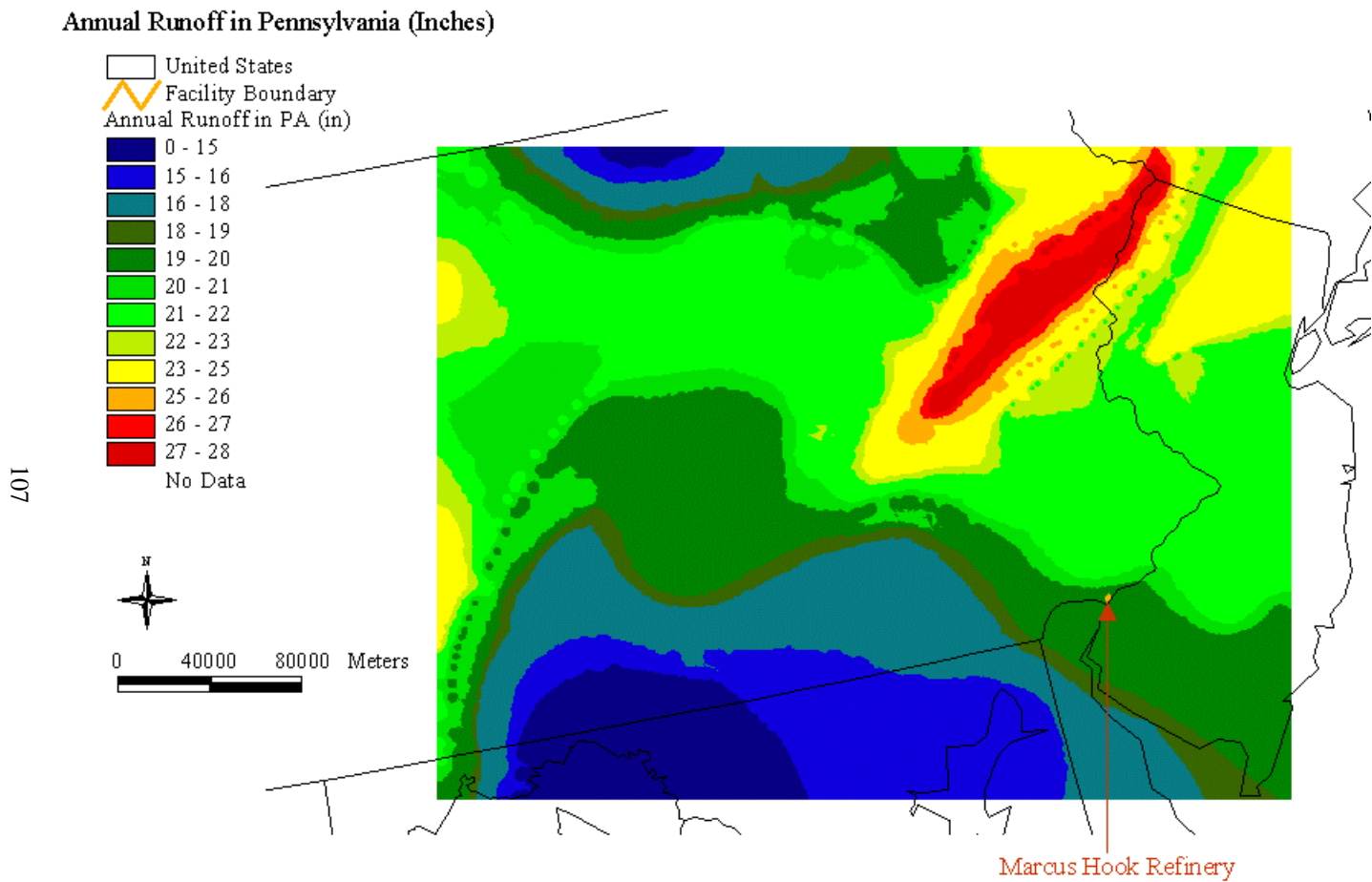


Figure 4.7: Annual Runoff Grid for Southeast Pennsylvania.

From Figure 4.7, the orange dot in the southeast corner of Pennsylvania is the Marcus Hook facility. Utilizing the identify tool in ArcView, an average annual runoff value of 20 inches was determined for the facility. Only one value was obtained because the detail of the contour line coverage does not allow for a more precise interpretation. Creating a much smaller grid would be reading too much from the data, and even with a smaller grid, the changes in runoff values across the site would be very small (probably less than a few tenths of an inch). A more detailed approach is often used to create a runoff grid that incorporates precipitation, infiltration, and evaporation. However, the precipitation data that are currently available over the Internet have cell sizes that are bigger than the entire area of the Marcus Hook Refinery. Thus, the use of only one value is appropriate. This value can be used for calculating flow magnitudes, which is discussed in the next section.

The results shown in Figure 4.3 above can also be used for the second component of surface water modeling for risk assessment purposes, that is determining chemical transport characteristics. The coverage of EPA Regulated facilities provides information about each facility, such as name and id numbers. Further information on these facilities can be obtained with the id numbers from the EPA's Query Mapper webpage (refer to Appendix A for address). Query Mapper is particularly useful for determining if a facility has a permitted discharge to a surface water body, and if so, what parameters are monitored. Actual National Pollutant Discharge Elimination System (NPDES) permits can even be viewed. Utilizing NPDES permit limits or actual monitoring data, point sources can be determined and then incorporated into the surface water model with the *COC Transport* extension. The Query Mapper was used to obtain information on each facility in the watersheds upstream of the Marcus Hook facility. However, of the 53 upstream facilities identified, only two have discharges to either Marcus Hook Creek or Stony Run Creek, and these two

facilities only monitor for pH and temperature. Thus, at this point in the research, a background concentration level has not been estimated for either creek upstream of the facility. If the screening analysis indicates that the surface water runoff pathway may be complete and empirical data collection is required, then background concentration levels could be estimated from these collection and analysis efforts.

#### **4.1.3 Facility Model**

The facility model is important because it characterizes actual water drainage and collection at the facility. Surface water runoff can carry a COC from a source area with high concentrations to a receptor where this COC could affect humans or ecological receptors. For instance, during construction activities, soil is dredged and placed in piles. These piles can have concentrations of COCs, and any precipitation runoff from these piles could carry these COCs across the facility. This runoff could come in contact with a construction worker or drain into one of the creeks where it could affect fish and other organisms. Thus, by knowing how water moves across the surface, one can make better decisions about potential exposures and exposure pathways.

The process used to model surface water runoff for the Marcus Hook facility is identical to that used in the regional model. Before beginning the modeling process, the digital terrain model (DTM) provided by Air Survey Corporation (ASC) had to be converted to a DEM because the watershed modeling tools cannot be used with a TIN. The DTM for the facility has a total of 421,472 triangles, which comprise a surface area of 22,954,375 square feet. Dividing the surface area by the number of triangles yields an average triangle area of approximately 54 square feet. This area would correspond to a grid cell size of about 7.5 feet. However, when trying to maintain accuracy in the conversion of a TIN to a grid, the average area of a triangle should not be used because this area

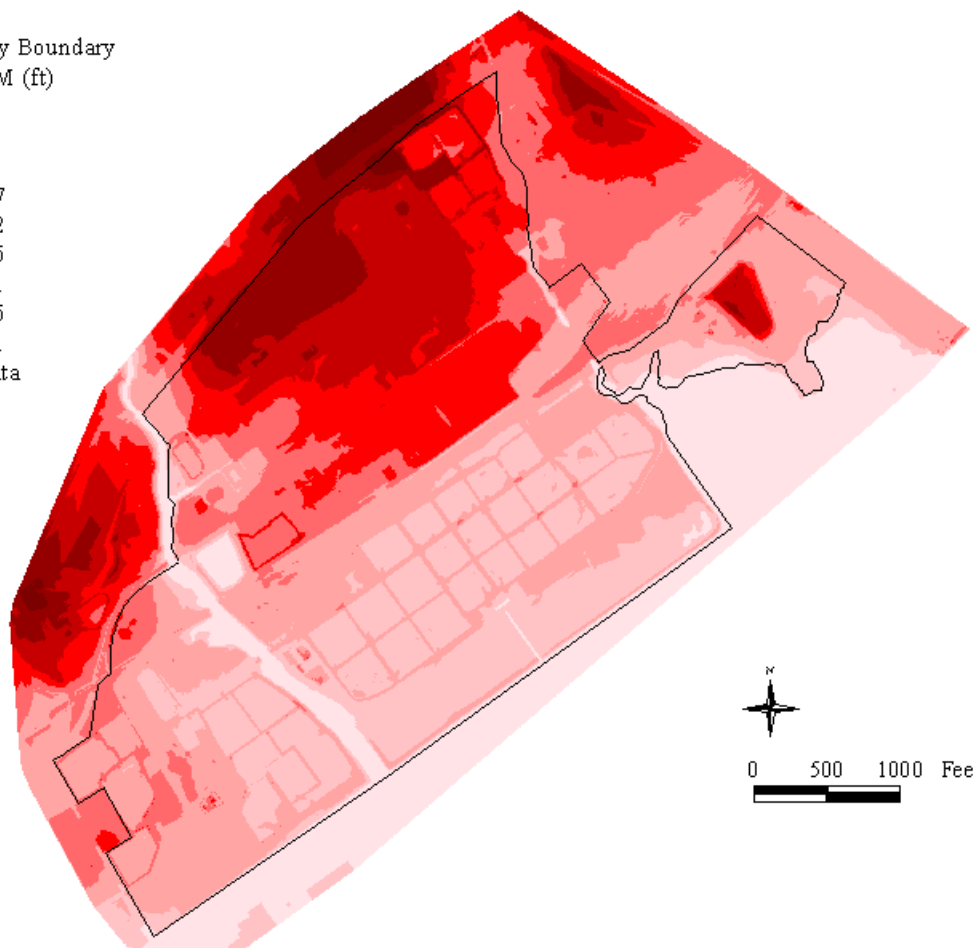
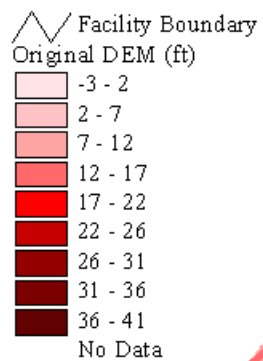
includes some large triangles. The area of the smallest triangle should not be used either because only a limited amount of detail can be obtained from the data. Choosing a grid cell size equal to the smallest area leads to wasted disk space and energy. Instead, some middle ground value should be used. For this project, a few different grid cell sizes were tested, and a 3-foot cell size was chosen as the final value. The *3-D Analyst* extension was utilized to convert the DTM into a DEM, but ARC/INFO could also have been used. The converted DEM is shown in Figure 4.8.

After converting the DTM, the *Watershed Delineator* tools were used with the converted DEM to model runoff. However, a few difficulties were encountered in the initial development attempts. Typically, the *Watershed Delineator* and similar extensions are applied for very large areas with at least 30-meter DEMs, such as in the regional model. However, this facility analysis involved calculating water drainage from each 3-foot grid cell of the study area. This large-scale application introduced many factors that would typically be ignored in a regional model such as buildings, ponds, and sewer drains. In order to produce accurate results, these factors could not be overlooked for the facility scale model. The errors in the standard procedures were due in large part to the processing methods of the filling sinks step. This step modifies a DEM to ensure that only one outlet will exist on the grid. However, there are a number of sinks at the facility that need to be accounted for in order to accurately depict surface water runoff. Some of these sinks include drains, impoundments, and basins.

One of the unique issues with the original facility model development was that the dike areas, which exist around each tank, were being filled to a level terrain in order to prevent flow from accumulating in them. Even though the tank foundations are incorporated in the DEM, the dike areas were being filled to the top of the berm, which is usually a few feet higher than the foundation of a storage tank. In the actual system, each dike area contains a drain to the oily



Marcus Hook Refinery - Digital Elevation Model (ft - NAVD88)



111

Figure 4.8: Marcus Hook Facility Digital Elevation Model.

sewer. This sewer, which covers most of the facility, is designed to collect water that could have come in contact with any products and route that water to the wastewater treatment plant. In addition to the dike area difficulties, other issues were encountered such as flow through buildings and across ponds. The building foundations were not high enough to prevent flow from running across them, especially after the filling-sinks process. Similar to the dike area issue, all the ponds and impoundments on site were being filled to prevent flow from accumulating in them.

Some unique methods were used to correct for these problems. The ArcView fill sinks program setup allows water to drain into no data cells, which are grid cells in the DEM that do not have an associated elevation. Having a no data cell, or sink, in the DEM prevents the program from filling an area during the filling sinks step. Therefore, for the facility model, a no data cell was created in the DEM at all sewer drain locations and at the center of each pond. To accomplish this task, a point coverage of sinks was created in ArcView using the oily sewer coverage and theme editing features (refer to Appendix E for more information). This coverage was then merged with the original DEM, and a no data cell was created in the grid wherever there was a drain. There are approximately 950 drains, each representing an actual capture point in the system. To solve the building issue, the facility coverages of structures and storage tanks were burned into the DEM. The procedure is similar to burning in the streams, except this time the original DEM elevation is held constant, and anywhere there is a building or tank, the elevation is raised 50 feet, which is an arbitrary value set to be above the highest point in the original DEM. Additionally, the facility stream coverage was burned into the DEM in a similar fashion as in the regional model to ensure proper flow accumulation in the creeks and Delaware River. Therefore, the original DEM was modified to include sewer drains, ponds, buildings, tanks, and

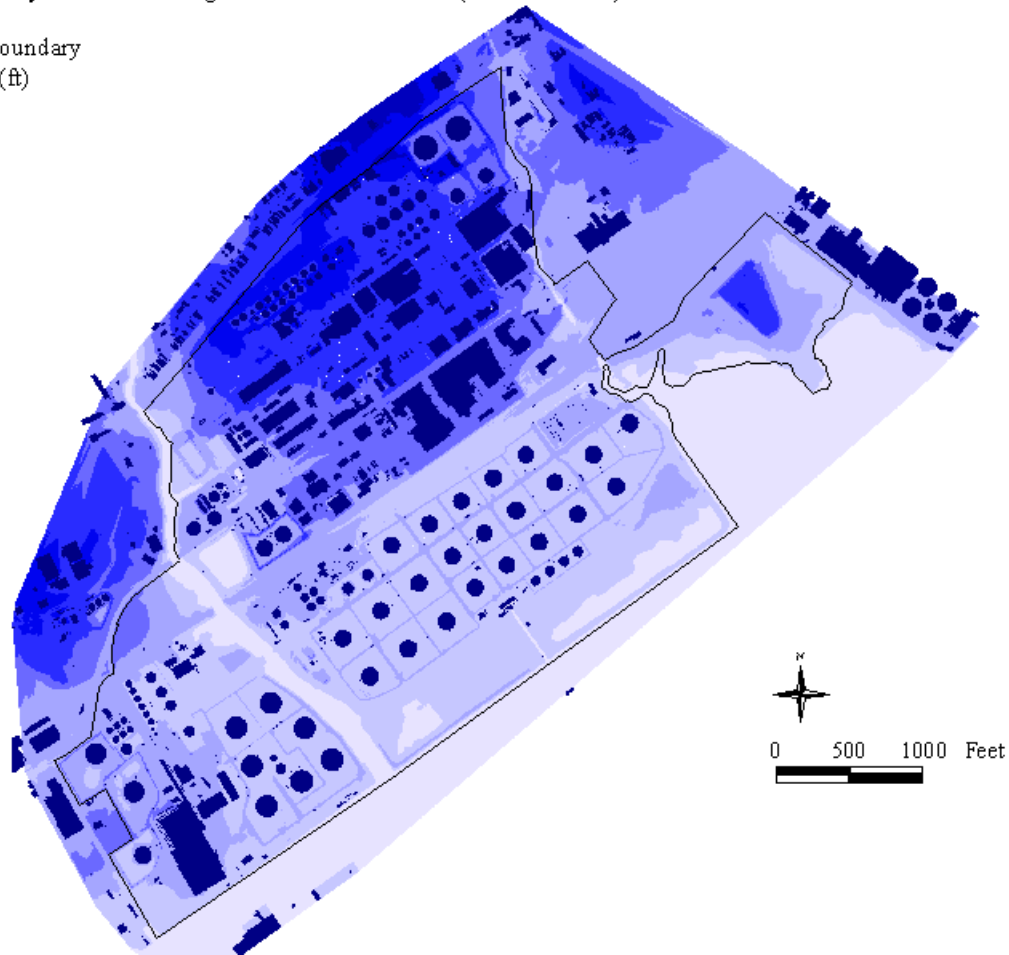
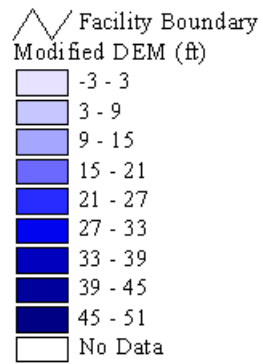
streams before proceeding with the watershed calculations. The modified DEM before any stream burning is shown in Figure 4.9.

Once the DEM manipulation was complete, developing the surface water model was fairly simple. Figure 4.10 shows the facility model results for the southwest corner of the facility. There is no flow through any of the buildings or tanks, and most of the water drains to Marcus Hook Creek and then eventually to the Delaware River. The small streams that appear to start and stop show water accumulation patterns before draining into the oily sewer.

In order to evaluate the strength of the model, the results can be compared with actual runoff observations at the facility. There is no stormwater management plan for the Marcus Hook facility as all runoff is either designed to be captured or is designed to flow through one of the NPDES sampling locations. There are ten NPDES sampling locations at the facility, of which five are used for monitoring stormwater runoff. One of these locations is shown in Figure 4.9, and a delineated runoff stream to the creek passes near this location. In fact, all of the stormwater NPDES locations are located on or very near a runoff stream in this model. Figure 4.11 shows the delineated drainage areas (i.e., watersheds) for the facility. Notice that the drainage areas for the creeks are relatively small, and that when they are larger, they usually contain an NPDES location (these areas are marked in yellow). While there are some additional delineated streams that produce runoff into the creeks, the overall surface drainage patterns appear correct. Also notice that the drainage areas mimic actual expected drainage areas, such as the dike areas in the south-central portion of the facility and the former surface impoundment in the west-central portion.

In order to obtain a more complete model, some preliminary calculations of flow magnitudes were developed using the annual runoff value obtained in the regional analysis. Flow magnitudes can be computed using a weighted flow accumulation process, in which a value such as runoff is assigned for each grid

Marcus Hook Refinery - Modified Digital Elevation Model (ft - NAVD88)



114

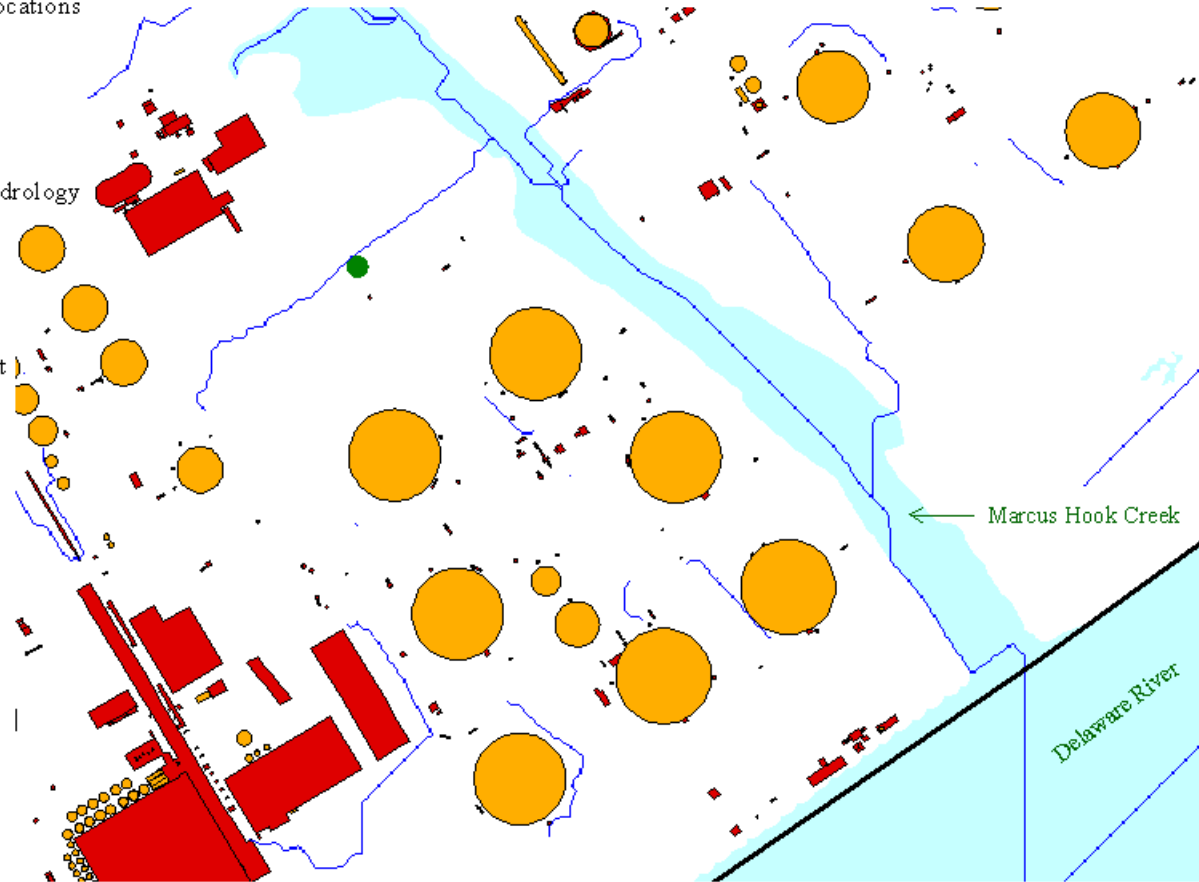
Figure 4.9: Revised Marcus Hook Facility Digital Elevation Model. Some of the no data cells are visible as white dots.

### Facility Surface Water Runoff Results

- Facility Boundary
- NPDES Locations
- Streams 5000
- 1
- No Data
- Tanks
- Buildings
- Surface Hydrology



0 90 180 Feet



115

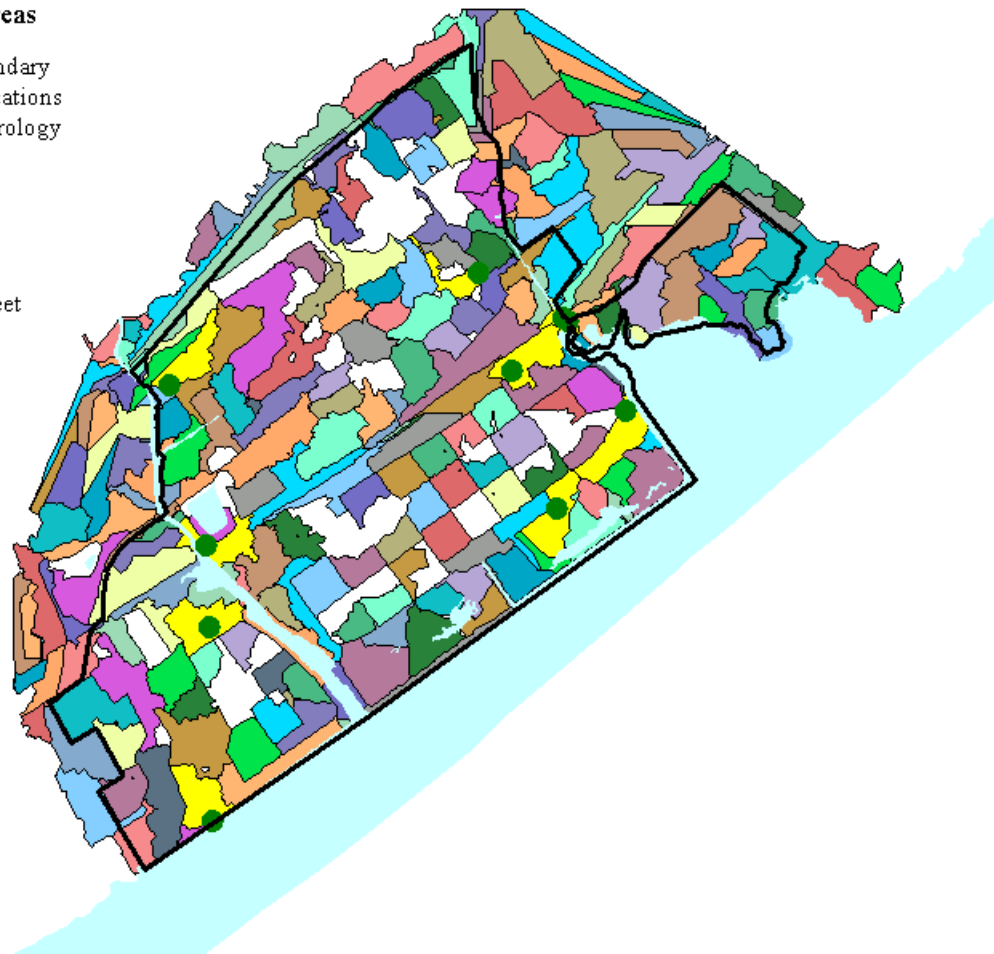
Figure 4.10: Facility Surface Water Runoff Model Results.

**Facility Drainage Areas**

- Facility Boundary
- NPDES Locations
- Surface Hydrology



0 500 1000 Feet



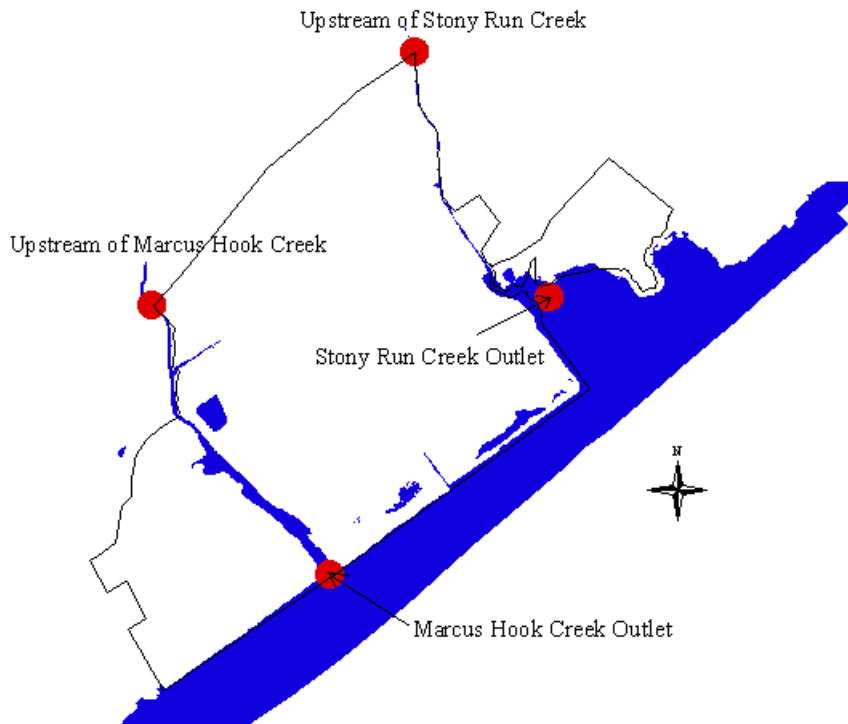
**Figure 4.11: Delineated Drainage Areas (each colored polygon refers to a separate drainage area). The areas with an NPDES sampling location are marked in yellow.**

cell, and the values upstream of a given cell are summed. Instead of accumulating a number of cells, a total amount of runoff is accumulated. However, in this project, the runoff value is the same in every cell: 20 inches. Thus, this value can be multiplied by the accumulated number of cells at the point of interest and by the grid cell area to determine a volume of runoff. The results for both the regional and facility models are shown in Table 4.1 and Figure 4.12. The accumulation was determined by using the identify tool in ArcView with the flow accumulation grids. Of particular interest for this project are the amounts of runoff upstream of the facility and at the facility. Therefore, the number of upstream cells was identified where the two creeks cross the northern boundary of the site and where the creeks outlet to the Delaware River. The facility grid has a much smaller mesh than the regional grid, so the accumulation of cells is larger with the facility grid, but the cell area is also proportionally smaller. As shown in the table, Marcus Hook Creek receives significantly more runoff (7.74 cfs) than Stony Run Creek (1.10 cfs) upstream of the facility, and Stony Run Creek receives a slightly higher amount of additional flow (0.24 cfs compared to 0.21 cfs) as it passes through the facility. This means that, relatively speaking, the surface runoff from the facility potentially has a greater impact on Stony Run Creek, neglecting point discharges from the treatment and cooling water plants. These results were expected based on known creek characteristics. However, actual flow measurements have not yet been obtained to compare against these calculations.

The final step in the analysis performed for the facility surface water model was to determine point and non-point sources and then predict downstream concentrations in the creeks and river. This phase of the model is still in development, but an initial run has been made to predict possible benzene concentrations. The first step involved obtaining the Marcus Hook NPDES permit information. Some of the actual measurements made for monitoring the

**Table 4.1: Flow Calculations**

	Accumulation *	Runoff (ft/yr) *	Cell Area (ft <sup>2</sup> ) =	Flow (ft <sup>3</sup> /s)
<b>Regional</b>				
Upstream of Marcus Hook Creek	14,220	1.667	10,298	7.74
Upstream of Stony Run Creek	2,028	1.667	10,298	1.10
<b>Facility</b>				
Marcus Hook Creek Outlet	441,843	1.667	9	0.21
Stony Run Creek Outlet	496,379	1.667	9	0.24



**Figure 4.12: Flow Calculation Points.**



parameters in the permit were also obtained. Unfortunately, specific parameters such as benzene are not monitored for at the facility.

Instead, the permit contains limitations for a generic oil and grease parameter. In order to determine a benzene concentration for each point source in the model, benzene was assumed to be 2% by volume of the oil and grease concentration limit. This assumption is based on the percentage of benzene in gasoline, which is typically less than 2% (Bond, *et al.*, 1997). This assumption is not likely to be valid because it ignores the differences between oil and gasoline and ignores the volatilization of benzene from the aqueous phase. This assumption was only used as a test case for the modeling algorithm.

Point sources were created with the *COC Transport* extension at each NPDES stormwater runoff location, the wastewater treatment plant outfall, and the creek entry points to the facility. Benzene concentrations of 0.18 mg/L and 0.30 mg/L were used for the treatment plant outfall and NPDES sampling locations, respectively. Runoff flows were used at the NPDES sampling locations, but a flow of 4.46 cfs was assigned at the treatment plant outfall based on known measurements. The flow values from Table 4.1 were used at the creek entry points, but no concentration was assigned for these points because of a lack of upstream data, as discussed in Section 4.1.2. After creating point sources, the remaining program steps were run. The calculated benzene concentrations for Marcus Hook Creek range from 0 µg/L to 180 µg/L, which is above the PADEP limit of 1 µg/L. These calculated concentrations in Marcus Hook Creek are a direct result of the treatment plant input. The concentrations in Stony Run Creek and the Delaware River are 0 µg/L as the runoff inputs have little effect on these systems. These results are not expected to represent actual concentrations in the surface water bodies. Many potential concentration sources such as the guard basins, creek sediments, and groundwater recharges to the surface water bodies

are not included in the model. The model also does not incorporate losses such as sorption and volatilization. The results are only included here to demonstrate the application of the *COC Transport* extension.

This extension was also used to predict benzene concentrations based on area sources. The facility surface cover theme was modified using a script created by the author called *areasources.ave* (provided in Appendix G) to include expected mean concentrations (EMCs) for each surface cover type. Table 4.2 shows the assumed values for this initial benzene analysis. These values were also determined from oil and grease concentrations along with knowledge of the activities associated with each surface cover type. The resulting concentration grid created using both point and area sources is shown in Figure 4.13. Because either gravel or structures cover most of the interior portion of the facility, the predicted runoff concentration in these areas is 200 µg/L. The predicted concentrations in the creeks and river range from 0 µg/L to 200 µg/L, but they are predominantly zero. Again, these results are not expected to represent actual conditions. They have been included in this thesis for illustration purposes only.

**Table 4.2: Expected Mean Concentrations**

Surface Cover Type	EMC (mg/L)
Asphalt Pavement	0.2
Concrete Pavement	0.2
Grass/Vegetation	0
Gravel	0.2
Ponds	0
River	0
Soil (no vegetation)	0
Streams	0
Structures	0.2
Tidal Flat	0
Wetlands	0

**Predicted Concentrations**

- Point Sources
  - ▭ Surface Hydrology
- Predicted Concentrations (mg/L)
- |               |
|---------------|
| 0 - 0.022     |
| 0.022 - 0.044 |
| 0.044 - 0.067 |
| 0.067 - 0.089 |
| 0.089 - 0.111 |
| 0.111 - 0.133 |
| 0.133 - 0.156 |
| 0.156 - 0.178 |
| 0.178 - 0.2   |
| No Data       |



**Figure 4.13: Predicted Concentration Grid Using Point and Area Sources.**

Future use of *COC Transport* could include predicting concentrations based on actual measurements made at runoff and discharge locations. This analysis would likely provide more accurate results, and it would eliminate the very conservative assumptions of using NPDES maximum limits and using high fractions of chemicals in the monitored parameters. Eventually the *COC Transport* extension will be combined with the groundwater model (Section 4.2) to look at the total discharge of chemicals to surface water bodies. This total discharge would include diffuse groundwater transport and surface water runoff.

## **4.2 Groundwater Model**

### **4.2.1 Introduction**

Another important transport mechanism is groundwater flow. Chemicals that are released to the soil can leach into the groundwater and then be carried with the water flow. Eventually, this water could reach a receptor (e.g., a drinking water well) or discharge to a surface water stream. Chemical leaching is a very significant mechanism because it is a continuous source of input to the groundwater that is fueled by the desorption of chemicals from soil. The Marcus Hook Refinery is no exception to this phenomenon. For example, the former lubricant packaging plant area in the southwest corner of the facility has several areas with a known presence of light non-aqueous phase liquid (LNAPL). Thus, it is very important to characterize subsurface fluid movement at the facility. To accomplish this characterization, a groundwater transport model is currently being developed using the Groundwater Modeling System (GMS) software. The goal is to connect the groundwater model with the surface water model so as to develop site-specific target levels for soils and groundwater that are protective of the surface waters. The groundwater model will also be used to investigate the potential for hydraulic connection between the upper saturated zone and the lower

regional aquifer, which is used as a drinking water source across the Delaware River in New Jersey. These groundwater simulations may be important to a determination of the groundwater classification of the surficial aquifer. The groundwater classification system in Pennsylvania depends in part on whether an aquifer is being used as a drinking water source or is connected to an aquifer that is a drinking water source.

GMS is a software utility developed by Brigham Young University (BYU) (1994) for use in groundwater simulations. GMS consists of a graphical interface that links with analysis codes such as MODFLOW, which is a three-dimensional grid-based groundwater modeling program developed by USGS. Other modeling codes to which GMS connects include MT3D, MODPATH, SEEP2D, and FEMWATER. For groundwater modeling purposes, a site conceptual model is first built in the graphical environment. In GMS, a site conceptual model refers to a high-level description of the site that is defined with GIS objects and includes sources/sinks, the boundary of the domain to be modeled, recharge and evapotranspiration zones, and material zones within each of the layers (BYU, 1998). Once the conceptual model is complete, a grid is constructed to fit the study area, and the MODFLOW data are converted from the conceptual model to the cells of the grid. GMS can then run MODFLOW and display the results in its graphical user interface.

One of the advantages of GMS is that it can import and export GIS spatial data sets using a map module. Coverages, shapefiles, and grids can be imported into the GMS environment for use in the analysis process, and the modeling results can be exported back to ArcView for incorporation into the digital facility description. Thus, although GMS is external to the ArcView system, data can be interchanged between them. This data exchange capability and the interface with widely accepted analysis codes like MODFLOW are the primary reasons for the selection of GMS for the groundwater modeling at the Marcus Hook facility.

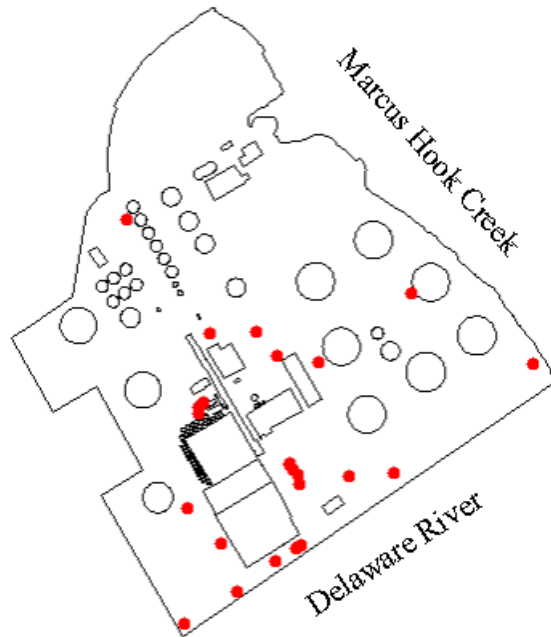
Despite its enormous capabilities, GMS is also very user friendly. However, it is an expensive product, and lower-cost alternatives exist such as MODFLOWARC and Argus. Links to more information on these products and to the actual GMS website are provided in Appendix A. All GMS documentation and a demonstration version of the program can be downloaded from the GMS website. The demo version has the same functionality as the licensed version except that it cannot save or print.

The groundwater model for Marcus Hook is first being developed for the former lubricant packaging plant area, referred to as the lube plant area. The idea is to start at a simple level and demonstrate the modeling capabilities with an example case. Once the model is operational, the impacts of more complex representations of the processes will be evaluated in terms of their impact on decisions. A steady-state solution has already been developed for the lube plant area. This model, its simplifying assumptions, and possible further enhancements are now discussed.

#### **4.2.2 Setup and Results**

This section documents the construction of a steady-state model for the lube plant area. The first step in the process is to build a site conceptual model in the GMS graphical environment. The most effective way to build such a model is to use already existing facility coverages, which can be added to GMS using the **import** command. Although simple in concept, GMS cannot directly read ARC/INFO binary coverages. Coverages must first be converted into ARC/INFO generate files using the **ungenerate** command in ARC/INFO before they can be imported into GMS. A generate file is a text file that contains the x and y coordinates of the points and lines in the coverage. Alternatively, GMS can read ArcView shapefiles for users who do not have access to ARC/INFO. For the lube plant area, coverages of the boundary line, buildings, storage tanks, and wells

were imported into GMS (Figure 4.14). These coverages were clipped in ArcView from the entire facility to the lube plant area before their use in GMS.



**Figure 4.14: Lube Plant Area Coverages in GMS. The red dots are monitoring wells.**

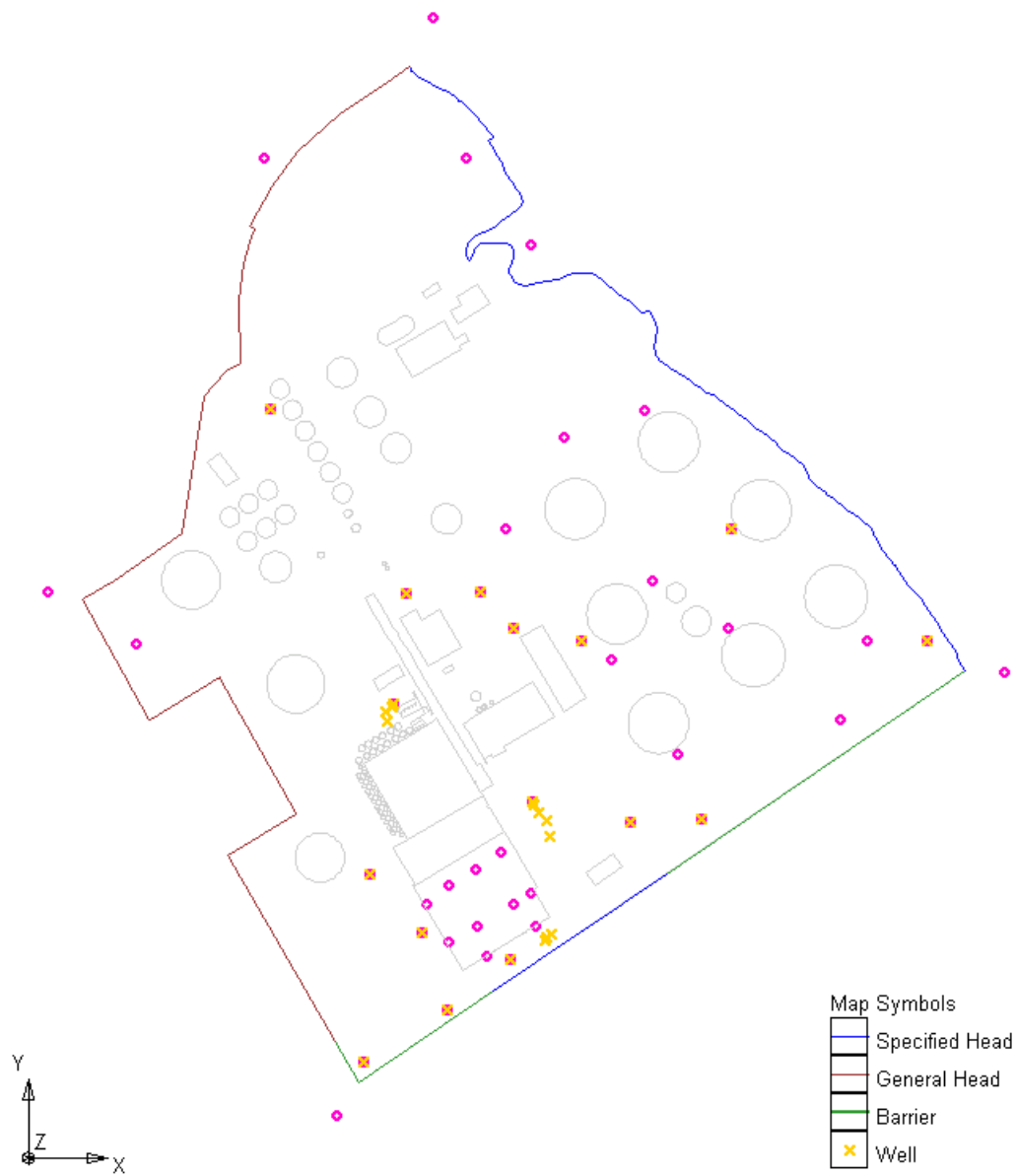
The map created from the coverages of the lube area can be used to build an effective groundwater model. The process is described in detail in the GMS Tutorials manual (BYU, 1998), but some of the unique features for the Marcus Hook model are identified here. The eastern boundary of the model was defined as a specified head arc because of Marcus Hook Creek. The southern boundary of the model was defined as a no-flow boundary except in the bulkhead area where there is a connection between the groundwater system and Delaware River. Heads of -1.7 ft. and -2.5 ft. were assigned at the northern and southern ends of the eastern boundary, respectively, and heads of -2.3 ft. and -2.6 ft. were assigned at the eastern and western ends of the bulkhead area, respectively. These specified heads were assigned based on the free water surface elevations, which

were queried using the digital terrain model. These heads represent values for a particular moment in time (i.e., when these elevations were surveyed during the mapping project). The remaining heads along each arc were linearly interpolated by GMS.

The other boundary arc was identified as general head arc so as to incorporate incoming groundwater flow northwest of the lube plant area. General head cells are often used to simulate lakes, in which the flow in or out of a cell is proportional to the difference in head and the constant of proportionality is conductance. Conductance for an arc is calculated as the hydraulic conductivity of the material along the arc, divided by the thickness of the material, multiplied by the width of the material (BYU, 1998). A conductance of 0.1 ft/d was used for this model, and a range of heads from 3 ft. to 15 ft. was assigned based on an estimated water table elevation along the boundary. Inside the boundary, the building and storage tank coverages were used to delineate recharge areas. Anywhere a building or tank existed, the recharge was set to zero, but elsewhere the recharge rate was set at 25 in/yr, which is the average annual precipitation minus the average annual runoff. This recharge rate is a high estimate since it does not account for evaporation losses. The wells coverage was only used for display, but pumping rates could be set based on some of the existing LNAPL extraction activities. Figure 4.15 shows the created site map in GMS before creating a grid for MODFLOW simulations.

As in any groundwater modeling exercise, the most difficult step in the construction of a groundwater conceptual model is building the subsurface description. The Marcus Hook facility provides a unique set of challenges to this process. Many different soil types are present across the site, and the top layer on the southern half of the facility is predominantly various types of fill since the Delaware River originally covered this portion of the site. To complicate the process even further, the database contains boring data from different studies and





**Figure 4.15: Site Conceptual Model in GMS. The circular points represent stratigraphy data locations.**

different data collectors. Additionally, the data cannot be directly imported from the database into GMS.

Two different options were evaluated to incorporate the subsurface data. The first option was to use the EQuIS Geology software package. EQuIS Geology is a subsurface database management program developed by EarthSoft, Inc. that creates input files for other utilities, such as GMS. The difficulty with using EQuIS Geology in these applications is that it is another data management tool, and it cannot dynamically link to the existing tabular database. Therefore, information would have to be updated in two places instead of one. EQuIS Geology cannot dynamically link with GMS either, and the borehole file that EQuIS Geology creates for GMS requires further modification before it can be incorporated into a groundwater model.

The best solution found was to use Excel to create a tab delimited text file from the Stratigraphy table in the Access database. This text file was then read into GMS as a multiple two-dimensional data set (i.e., each point [x,y] had values associated with the elevations of each layer). However, the file format required some simplifying assumptions to be made of the original data (Table 4.3). Each stratigraphy data location must have the same soil layers. A layer's thickness can be set to zero by using the same elevation for the top of two layers, but layers that only exist in some borings, or layers that have different locations in borings, cannot be incorporated. The soil data for the lube plant were first simplified for the model to a three-layer system of fill, clay, and sand and gravel. The clay layer is not continuous in the lube plant area, and these discontinuities were included in the subsurface description by assigning the same top of elevation to the clay and sand and gravel layers. After creating a three-layer text file and importing it into GMS, the interpolation features of GMS were utilized to determine the elevations of each layer at every point in the model. A linear interpolation scheme was used to determine layer elevations for each cell in the grid created for MODFLOW.

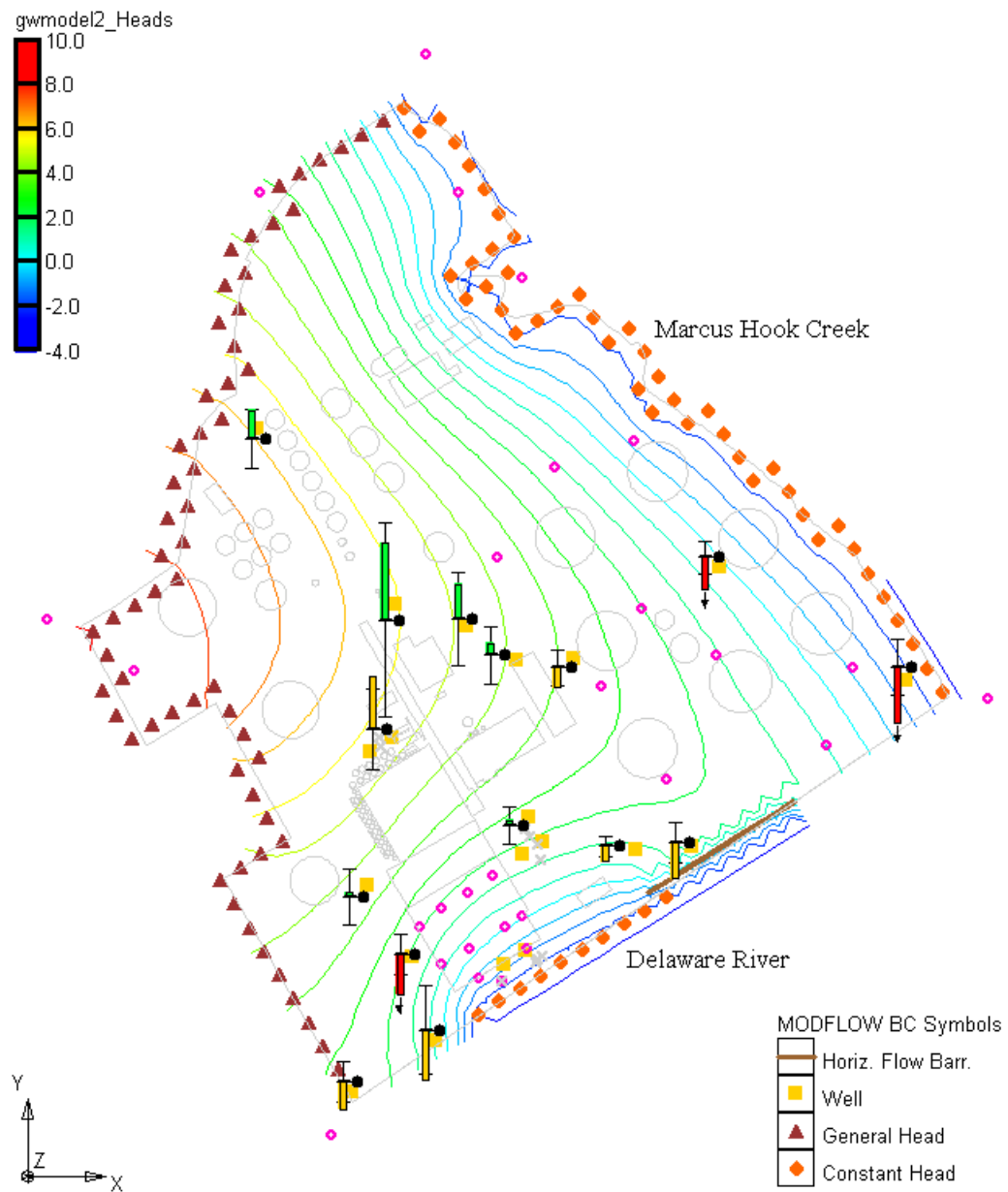
Thus, each grid cell had an associated top of fill elevation, top of clay elevation, top of sand and gravel elevation, and a bottom elevation (i.e., top of bedrock). "Phantom" borings were created at the model edges so that the program could interpolate everywhere within the modeled area beyond the observed boring data, and the stratigraphy for these phantom borings were based on the stratigraphy of the nearest data point. Besides providing elevations, the interpolated data sets were used to define other layer attributes, such as transmissivity.

**Table 4.3: Scatter Point Data File Format**

id	x	y	topA	topB	topC	botC
7	2626780	184646	6.2	-0.8	-3.8	-25.0
8	2626242	184401	9.8	4.8	2.8	-25.0
9	2626290	183972	10.6	-0.4	-3.4	-29.4
10	2625885	183794	10.7	3.7	3.7	-25.0

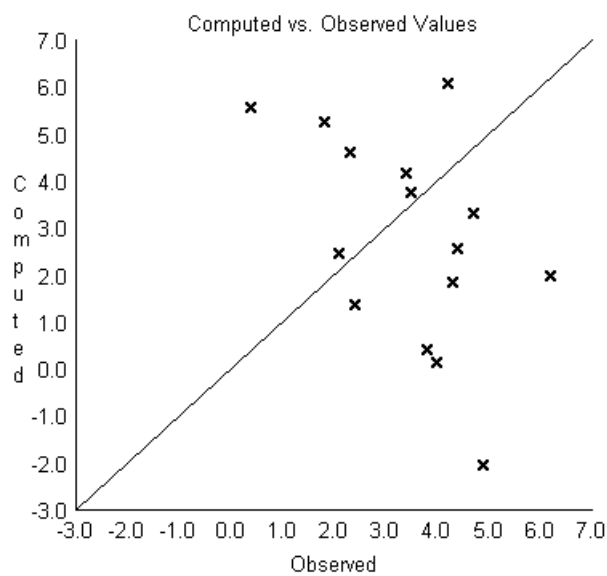
Unfortunately, the attempts to develop a groundwater model using this three-layer system were unsuccessful. Many difficulties were encountered with convergence, dry cells, and comparisons to existing measurements. The results shown in Figure 4.16 are for a one-layer steady-state model. The bottom elevations for this layer were determined based on the estimated bottom elevations of the sand and gravel layer from the Access database. A constant hydraulic conductivity of 10 ft/d was used, and this value was chosen because it provided the best solution to the model (i.e., the error between the computed and observed measurements was minimized). Although this hydraulic conductivity is less than some of the actual hydraulic conductivity measurements listed in the Access database, it is still a very high flow velocity.

The results illustrated in Figure 4.16 are consistent with the expected groundwater contours. Groundwater enters the lube plant from the northwest side and either outlets to Marcus Hook Creek or to the Delaware River along the



**Figure 4.16: 1-Layer Steady State Groundwater Model for the Lube Plant Area.**

bulkhead. The figure also shows a coverage of observation points with the residual error between the observed value at each point and the predicted value at that point. This coverage was created by averaging the groundwater elevations at each measurement location and computing a standard deviation for this average. A green error bar indicates that the computed value is within two standard deviations of the observed value. A yellow bar indicates that the computed value is not within two standard deviations but that the error is less than 200%. A red bar indicates that the error is greater than 200%. Figure 4.17 shows a plot of the computed values versus the observed values.



**Figure 4.17: Comparison of Computed Groundwater Elevations with Observed Groundwater Elevations in the Lube Plant Area.**

The results demonstrate the ability of using GMS and ARC/INFO coverages for the development of a groundwater model. However, these results are just the first step towards an integrated fate and transport model. The hydraulics of the model could be improved to incorporate transient simulations based on varying precipitation, and thus, recharge. An initial attempt has already been made by the

author to predict transient infiltration rates in the lube plant area using the Green & Ampt equations and daily precipitation data (refer to Appendix A). A better subsurface description could also be developed by reinterpreting the soil boring logs and obtaining new soil measurements. Once these hydraulic properties are improved, the model results can be enhanced to include predicting chemical concentrations with MT3D. As already mentioned, the eventual goal is to develop a facility groundwater/surface water interaction model within ArcView and GMS that can be used to develop site-specific target levels that will be protective of the surface water bodies. Any additional fate and transport modeling requirements beyond this site-wide interaction model will be guided by the development of the site conceptual model and specific exposure pathways. The ongoing enhancement of this model is being researched by a new graduate student, Julie Kim.

## **CHAPTER FIVE**

### **RISK MAPPING**

#### **5.1 Introduction and Methods**

This chapter describes the development of maps for the Marcus Hook Refinery that display either the expected concentration for a particular source area or the probability that any sample obtained from a source area will be above a threshold value, such as an agency action level. Maps such as these may be called risk maps. Risk mapping is an exposure analysis activity that involves determining whether the concentrations in a source area could pose risks to receptors. Risk maps are particularly useful tools in evaluating which areas may require remediation, which areas are likely free of concern, and which areas require more data before an appropriate decision can be made. They provide a means to evaluate the data that already exist and to determine future data collection needs. Risk mapping is a Tier 2 Analysis that compares concentrations in a source area with conservative screening values. Although risk mapping does not have to be limited to a source area analysis, no fate and transport modeling is incorporated in the results presented here. Three specific methods of risk mapping are studied: (1) an upper confidence limit, (2) a binomial distribution, and (3) a Bayesian approach. The Upper Confidence Limit method calculates a representative concentration of a COC for each source area while the other two methods calculate the probability that the concentration of a COC in that area will be above a target level.

The development of risk maps is supported by the application of the digital facility description. Concentration data are first extracted from the tabular database, displayed with the spatial database, and finally queried by source area for use in statistical calculations. Once the calculations are complete, the results can be joined to the attributes of the source area coverage so that areas can be

displayed in the ArcView environment according to their relative risk. Note that in this case, risk refers only to the evaluation of source area concentrations against specified target levels. Dose and toxicity assessments for particular receptors are not incorporated. For this project, risk maps were developed for benzene, ethyl benzene, toluene, xylenes, and naphthalene. These COCs were analyzed in soil, groundwater, and soil vapor source areas, which were delineated from the digital facility description based on the portions of the site that contained the most measurements. A total of eight source areas were analyzed using the first two methods while the Bayesian Approach studied benzene soil concentrations for the entire site.

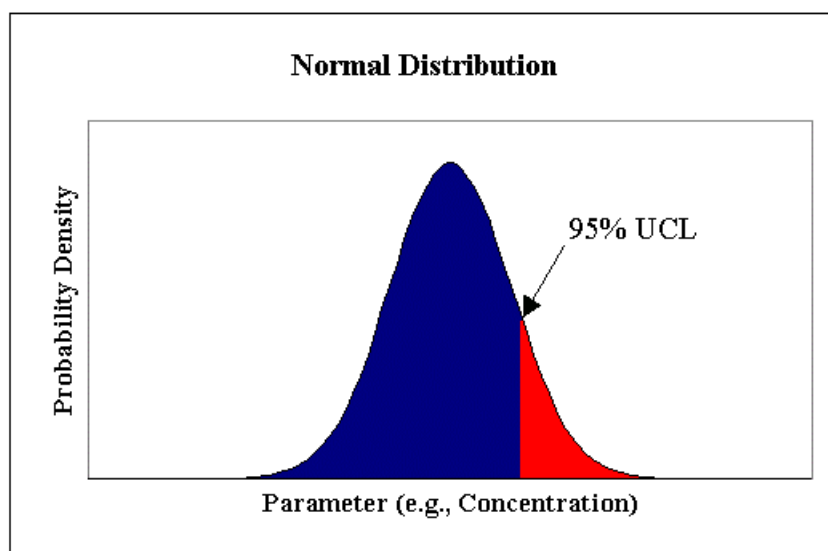
The following sections introduce each risk mapping method. Section 5.2 then discusses the application of these methods with the Marcus Hook facility data and the actual results. For these applications, ArcView was used to analyze historical qualitative data, identify source areas, plot the quantitative data, query the data, and then display the results. Excel was used for all the statistical analysis calculations, and in some cases, as discussed later in the chapter, a dynamic connection was used between Excel and ArcView.

### **5.1.1 Upper Confidence Limit (95%)**

One method of evaluating source area concentrations is to calculate a 95% upper confidence limit (UCL). The 95% UCL is the value that will equal or exceed the true mean of a random data set 95% of the time (Figure 5.1) (BP, 1997). It is a 95% confidence level for the mean concentration, not the data set. This procedure is documented in both the *Guidance Document for the Risk-Based Decision Process for Corrective Action Sites* (BP, 1997) and the *Proposed Texas Risk Reduction Program Rule* (TNRCC, 1998). In the TNRCC manual, the UCL is referred to as a Protective Concentration Level (PCL). The calculation requires previously acquired data within the source area, and the result can be compared



with a target level or risk-based screening level (RBSL), which is a non-site-specific human health risk-based corrective action target level for a COC (ASTM, 1998). The calculated UCL is typically a conservative estimate of the actual concentration, but this method is easily applied and is an effective way to make an initial evaluation for a predefined area. For example, if a UCL is significantly lower than an RBSL, one can expect that no corrective action is required in that area for that COC. Alternatively, if the UCL is close to or greater than the RBSL, one can expect that some sort of corrective action may be necessary or that a more detailed analysis is appropriate.



**Figure 5.1: 95% Upper Confidence Limit for a Data Set with 20 Samples. The UCL is an upper confidence limit of the mean and not the data set.**

The calculation of a UCL is based on the assumption that the supporting data are normally distributed. For concentration measurements, in which values can vary by many orders of magnitude, the distribution is rarely normal. However, the assumption also applies for log transformed data, which often approach normality with concentration data. There are a few different ways to test for

normality in a data set. The first test is simply to plot the cumulative frequency versus the concentration values. For normally distributed data, this plot should be a straight line. To perform this test, the data (in log transformed format) should be listed and then ranked in descending order. The largest value is assigned a rank of one, and measurements with identical values are given the same rank. The cumulative frequency can then be estimated as (BP, 1997):

$$\blacksquare \quad CF = \left[ \frac{i}{(n+1)} \right] * 100 \quad [5.1]$$

where  $CF$  = cumulative frequency,  $i$  = measurement rank, and  $n$  = sample size.

A second test for normality is to calculate the sample coefficient of skewness,  $CS$  or  $\gamma$ , which indicates how "skewed" the data are with respect to the mean. Normally distributed data have a skewness coefficient of zero, but the assumption of normality is valid as long as the skewness coefficient is less than  $|1.0|$  (BP, 1997). Thus, if  $|\gamma| > 1.0$ , the test is considered a "FAIL." The coefficient of skewness can be calculated with the following formula (BP, 1997):

$$\blacksquare \quad \gamma = \frac{\frac{1}{n} * \sum_{i=1}^n (x_i - \mu)^3}{\left( \frac{n-1}{n} \right)^{3/2} * s^3} \quad [5.2]$$

where  $n$  = sample size,  $x_i = i^{\text{th}}$  concentration value,  $\mu$  = data set mean, and  $s$  = standard deviation.

A final test for normality is to calculate the coefficient of variation,  $CV$ , which is the standard deviation of the log transformed data divided by the mean (BP, 1997):

$$\blacksquare \quad CV = \frac{s_y}{\mu} \quad [5.3]$$

If the *CV* is less than 1.0, the test is considered a "PASS." However, the *CV* is not as accurate a measure of skewness as the *CS* is, so the coefficient of skewness test should take precedence over this one (BP, 1997). All three of these tests can be performed using built-in functions of Excel. Summary statistics such as the sample mean, standard deviation, and coefficient of skewness for any data set can be determined using the **Data Analysis/Descriptive Statistics** command (Analysis Tool-Pak required). Figure 5.2 shows some example results produced with Excel.

If the data set passes the tests for normality, then the UCL can be calculated as (BP, 1997):

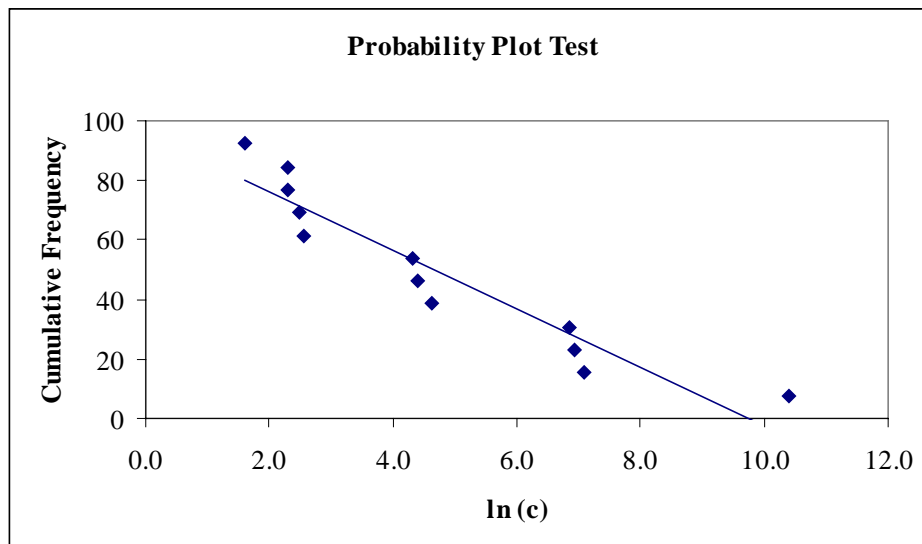
$$\blacksquare \quad 95\% \text{ UCL} = \mu + t\{0.05, n - 1\} * \left( \frac{s_y}{n^{1/2}} \right) \quad [5.4]$$

where  $\mu$  = data set mean,  $n$  = sample size,  $s_y$  = sample standard deviation, and  $t$  = student t-statistic, which can be obtained from tables. Note that when using log transformed data, the inverse natural log ( $e^x$ ) of the result will return the UCL in the correct units.

While this method is simple and straightforward, it should only be used for initial evaluations. One difficulty with this method is that the data set must exhibit behavior similar to a normal distribution in either its original or log transformed format in order to utilize the parametric method of evaluation (BP, 1997). This method also does not account for "no-detects", which are measurements with values below the practical quantification limit of a particular laboratory method. Another difficulty with this method is that the number of data points can have a significant impact on the results. The fewer data points, the more uncertainty there is in the mean value, and thus, the higher the calculated UCL will be. The TNRCC Risk Reduction Manual (1998) requires a data set with a minimum of 10 discrete samples. A final difficulty with this method is that it calculates a single concentration value. The next section examines a more

C ( $\mu\text{g/L}$ )	ln (C) ( $\mu\text{g/L}$ )	Rank	CF	<i>Statistics</i>	
32822.0	10.4	1	7.7	Mean	4.66
1200.0	7.1	2	15.4	Standard Error	0.77
1022.0	6.9	3	23.1	Median	4.36
951.0	6.9	4	30.8	Mode	2.30
102.0	4.6	5	38.5	Standard Deviation	2.67
82.0	4.4	6	46.2	Sample Variance	7.14
75.0	4.3	7	53.8	Kurtosis	0.17
13.0	2.6	8	61.5	Skewness	0.87
12.0	2.5	9	69.2	Range	8.79
10.0	2.3	10	76.9	Minimum	1.61
10.0	2.3	11	84.6	Maximum	10.40
5.0	1.6	12	92.3	Sum	55.89
				Count	12

CV =	0.6
CS  =	0.8
ln (UCL) =	6.0
UCL =	421.1



**Figure 5.2: UCL Sample Results Produced with Excel.**

detailed analysis in which a probability instead of a concentration is determined.

### **5.1.2 Binomial Distribution**

Another method of quantifying potential exposure is to calculate the probability that any sample taken from a source area will be above a target level. Assuming that the concentration of a chemical compound in groundwater or soil is a random function of space and time, a prediction of concentration can only be properly described as a probability function (Evans and Maidment, 1995). If there are only two possible outcomes of an event, this probability function may be modeled by a Bernoulli sequence, which is based on the following assumptions (Ang and Tang, 1975):

1. Each trial has only two possible outcomes: the occurrence or nonoccurrence of an event.
2. The probability of occurrence of the event in each trial ( $p$ ) is constant.
3. The trials are statistically independent.

The validity of the first assumption is easily checked based on the intended application. In the case of risk mapping at the Marcus Hook facility, the only two possible outcomes are that the measured concentration in a sample will be above a target level or that the concentration will not be above a target level. The final two assumptions require closer examination. For example, if one is looking at an entire site, the probability of occurrence in one corner of the site is highly unlikely to be the same as the probability of occurrence in another corner of the site. However, if the study area is small enough and defined properly, this second assumption has justification. As an example, it could be expected that the probability would be constant in the subsurface of a dike area surrounding a storage tank. The final assumption states simply that the occurrence or nonoccurrence of one event does not affect the probability of another event.

In a Bernoulli sequence, the ratio of the number of samples above a target level to the total number of samples can be described by the binomial distribution.

The probability of  $s$  successes (above a target level) in  $n$  trials can be computed as (Evans and Maidment, 1995):

$$\blacksquare e(n, s, P_r) = \binom{n}{s} (P_r)^s (1 - P_r)^{n-s} \quad [5.5]$$

where  $\binom{n}{s}$  is the number of combinations of  $n$  trials that contain  $s$  successes.

The cumulative probability of  $s$  or more successes in  $n$  trials is given by (Evans and Maidment, 1995):

$$\blacksquare E(n, s, P_r) = \sum_{m=s}^n e(n, m, P_r) \quad [5.6]$$

A comparable analogy to this situation is to envision an urn filled with blue and red balls. If balls are drawn from an urn containing an infinite number of balls, or drawn from a finite supply and replaced, the ratio of red balls drawn to total balls drawn can also be governed by the binomial distribution (Evans and Maidment, 1995). The problem is that this ratio is unknown, and it is the value of interest. One way to estimate this ratio is to repeatedly draw balls from the urn or samples from the medium of interest and keep track of the results. An estimate of the probability of success for a single trial is simply the number of successes divided by the number of trials:

$$\blacksquare \hat{P} = \frac{s}{n} \quad [5.7]$$

However, this value is only an estimate, and it would be useful to express this value with some sort of confidence range. For instance, one might determine that there is a 5% chance the probability of success in a single trial will lie outside of the range of 30% to 70%. This concept can be somewhat confusing as a probability range is being set for the calculated fraction.

As an applicable screening range, it is desirable to determine the 90% confidence limits on the exceedance probability, or fraction. There will be a 90% chance that the actual fraction,  $\hat{P}$ , will fall between the lower and upper confidence limits. For example, if an upper bound of 5% was calculated, then one might feel pretty good about the concentrations in that area because it is unlikely that these concentrations would be above that target level. The lower bound,  $P_l$ , can be calculated using the following method (Evans and Maidment, 1995):

- For  $s = 0$  (i.e., no successes),

$$P_l(0) = 0 \quad [5.8]$$

- For  $s = n$  (i.e., all successes),

$$P_l(n) = \sqrt[n]{\alpha} \quad [5.9]$$

- For  $s = 1, 2, \dots, n-1$ , find the value of  $P_l(s)$  such that

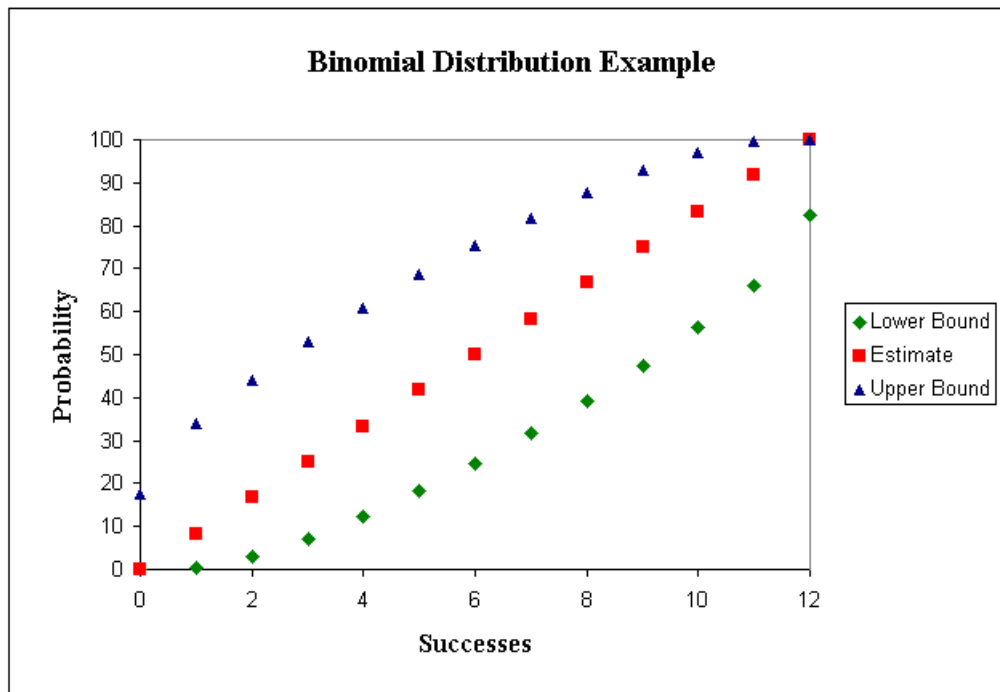
$$1 - E(n, s, P_l(s)) = 1 - \frac{\alpha}{2} \quad [5.10]$$

where  $E(n, s, P)$  is the cumulative binomial probability function (eqn. 5.6) and  $1 - \alpha =$  confidence level (two-sided).

The upper bound,  $P_u(s)$ , can be solved for through symmetry by using the relation (Evans and Maidment, 1995):

- $P_u(s) = 1 - P_l(n - s) \quad [5.11]$

Figure 5.3 shows the estimated fractions and confidence limits for a data set of twelve trials.



**Figure 5.3: Possible Solutions of the Binomial Distribution Method with a Sample Size of Twelve Trials (Evans and Maidment, 1995).**

Equations 5.10 and 5.11 require an iterative solution (inversion of the binomial distribution), and the author has developed an Avenue script to perform this task using ArcView and Excel. As opposed to the Access and ArcView connection discussed in Chapter 3, an ODBC driver is not needed to run this script. The first step is to select the data in the area of interest in ArcView. The script is then run, and it prompts the user for the threshold concentration (target level) and the confidence level. The script determines the number of measurements and successes and sends these data to Excel, which runs through a trial and error solution until each of the above equations are satisfied. This script, binomial.ave, is provided in Appendix G.

The Binomial Distribution method is an effective tool because it provides a probability and range of confidence in this probability that an area might be above a target level instead of simply returning a concentration. It also accounts for "no-



detects”, as these measurements are included in the total number of trials used. Thus, this method incorporates all the available data in an area. However, this method has limitations. The range in concentrations is not included in the calculation. For example, a value of 125 mg/L is not differentiated from a value of 10,000 mg/L when compared with a target level of 100 mg/L. Also, similar to the UCL method, the magnitudes of the confidence limits are highly dependent on the number of data measurements in a specific area. This dependence is shown in the results section.

### **5.1.3 Bayesian Approach**

The Bayesian Approach takes the binomial method one step further by including prior information about historical oil production activities in each source area. In the classical statistical approach, the parameters of the distribution are assumed to be constant, and the sample statistics are used as estimators of these parameters (Ang and Tang, 1975). This classical approach does not contain a provision for combining judgmental information with observational data when estimating the parameters. With a Bayesian Approach, the unknown parameters of a distribution are assumed to be random variables. Prior knowledge of activities in an area can help one make estimates of the distribution parameters. These estimates can then be updated using available data in that area. For example, say it was known that a large amount of gasoline was released from a storage tank. One would expect that the benzene soil concentrations in the surrounding dike area are more likely to be above the applicable target level as compared to another area where no spills have occurred. This conclusion could be further strengthened (or weakened) by analyzing soil samples from that area.

This method also calculates a probability that a source area will have a concentration above a target level. An alternative way to look at it is that this probability represents the fraction of soil in the area with a concentration above

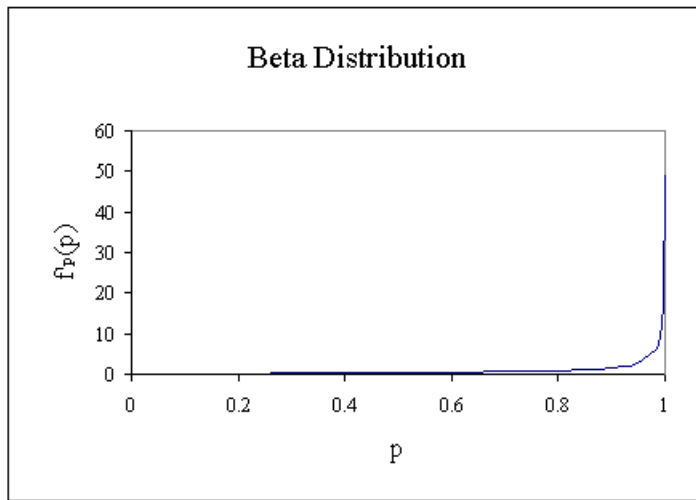
the target level. The method used for this project is a specific application of the Bayesian Approach. The basic random variable (fraction of soil above an action level) is described by a binomial distribution, and the prior and posterior distributions of the fraction are modeled using a beta distribution. A beta distribution can be used to model probability distributions of a random variable that is bounded by finite limits. In this case, the finite limits are 0 and 1. The assumptions of statistically independent samples and a constant fraction in the area are again made because the number of occurrences is assumed to be a binomial variable.

For each area of analysis, a mean fraction and a variance on that fraction are assigned based on prior knowledge of activities in that area. For example, if it is fairly definite that an area of soil will be above a target level based on known releases of product, one might assign a mean fraction of 0.85 with a variance of 0.20. The beta distribution for this fraction and variance is illustrated in Figure 5.4. If no prior knowledge is known for a particular area, a uniform distribution is used. This distribution has a mean of 0.5 with a variance of 0.2887, and the resulting probability density function for  $p$  is shown in Figure 5.5, where  $p$  is the fraction of soil above the action level. After the mean and variance are defined, the parameters of the beta distribution,  $q'$  and  $r'$ , are solved from these equations (Ang and Tang, 1975):

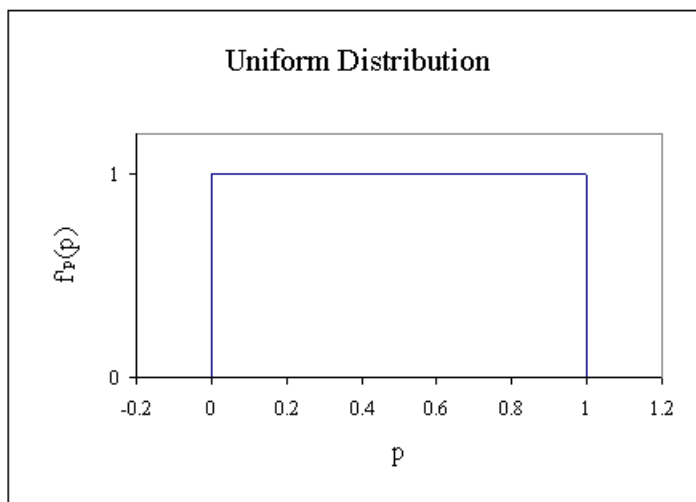
$$\mu'_p = q' / (q' + r') \quad [5.12]$$

$$(\sigma'_p)^2 = (\mu'_p)^2 \frac{r'}{q'(q' + r' + 1)} \quad [5.13]$$

where  $q'$  and  $r'$  are the parameters of the prior distribution,  $\mu'_p$  = assigned mean fraction, and  $\sigma'_p$  = assigned fraction variance. With no prior information,  $q' = r' = 1$  (uniform distribution).



**Figure 5.4: Beta Distribution for a Fraction of 0.85 and a Variance of 0.20.**



**Figure 5.5: Uniform Distribution (i.e., Beta Distribution for a Fraction of 0.5 and a Variance of 0.2887).**

Once the parameters  $q'$  and  $r'$  are calculated, they can be updated using available data in that source area with the following formulae (Ang and Tang 353):

- $q'' = q' + s$  [5.14]
- $r'' = r' + n - s$  [5.15]

where  $q''$  and  $r''$  are the parameters of the posterior distribution,  $s$  = number of samples above a target level, and  $n$  = total number of samples.

The updated mean fraction and variance can then be computed as (Ang and Tang, 1975):

$$\mu_p'' = q'' / (q'' + r'') \quad [5.16]$$

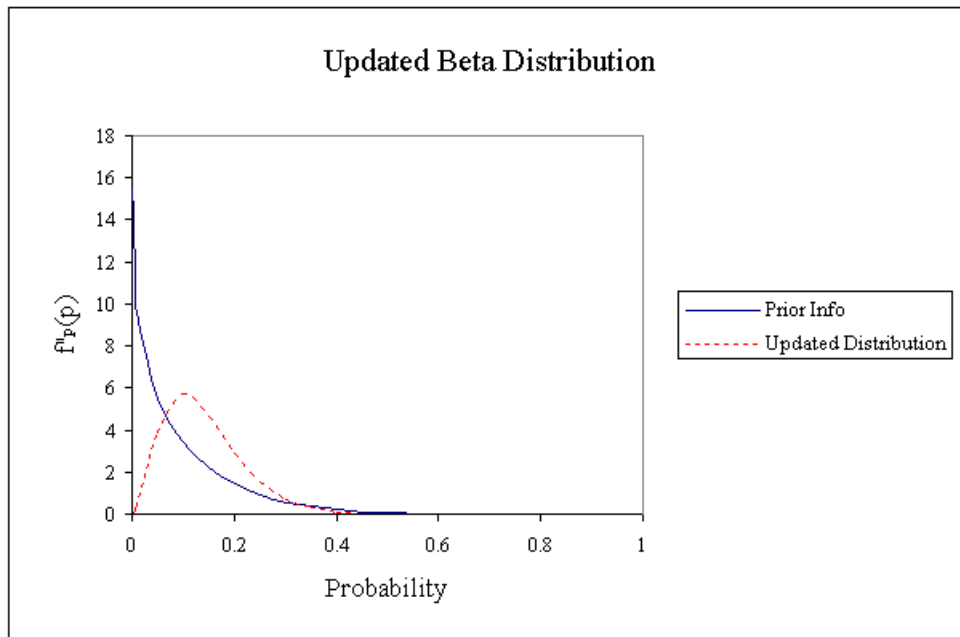
$$(\sigma_p'')^2 = (\mu_p'')^2 \frac{r''}{q''(q'' + r'' + 1)} \quad [5.17]$$

The density function for this distribution is calculated as (Ang and Tang, 1975):

$$f_p''(p) = \frac{\Gamma(q'' + r'')}{\Gamma(q'')\Gamma(r'')} p^{q''-1} (1-p)^{r''-1} \quad [5.18]$$

$$\text{where } \Gamma = \int_0^{\infty} x^{k-1} e^{-x} dx \quad [5.19] \text{ (Ang and Tang, 1975).}$$

The resulting beta distributions can be any shape as shown in Figure 5.6. The prior distribution in this example has a mean and a variance of 0.1 (i.e., 1 violation out of 10 with variance of 0.1 is expected based on prior information). If there are 2 hits (e.g., concentration measurements above the target level) out of 12 measurements, this distribution shifts to the right as shown in the figure. The updated mean fraction and variance are 0.14 and 0.075, respectively. This example demonstrates the capabilities of this approach. It can confirm areas in which the soil concentrations were expected to be above the target level. It can confirm areas in which the soil concentrations were expected to be below the target level. And it can show areas in which the soil concentrations were not expected to be above the target level based on historical information, but the quantitative data suggest otherwise. Additionally, this method can be used to identify areas where more information would be needed before a decision could be made.



**Figure 5.6: Prior and Updated Beta Distributions for a Data Set with Two Hits in Twelve Samples (Original Fraction = 0.1, Original Variance = 0.1).**

## 5.2 Results

This section demonstrates the use of the risk mapping methods described above with the Marcus Hook facility data. The first subsection describes the derivation of target levels, which are used in the statistical calculations. The second and third subsections contain the actual method results for different source areas and COCs. The UCL and Binomial Distribution results are grouped into one section because they are applied to the same source areas and data sets. The final subsection contains an overall discussion of the results and an evaluation of each risk mapping method.

### 5.2.1 Derivation of Target Levels

Before any decisions can be made about which areas require remediation, one must know which target levels are applicable in the environmental medium of interest (e.g., soil and groundwater). Target levels vary depending on the

chemical of concern, the medium of interest, the current use of land, the potential future use of land, and on the exposure route. For instance, the target level would be much more stringent if the property was to be used as a park in the future instead of a refinery. For this risk mapping application, it was assumed that the property will continue to be used for industrial purposes and that the point of exposure is at the source area. Applicable target levels were obtained from three sources: the BP Guidance Document (1997), EPA Region III (1998), and PADEP (1997). The standards taken from the BP Guidance Document are residential RBSL values. The EPA Region III standards are risk-based concentrations developed by the Hazardous Site Cleanup Division in accordance with CERCLA and RCRA. The PADEP standards come from Act 2 of Pennsylvania's Land Recycling Program. This program was adopted in 1995 to make properties impacted by a COC safe and to maintain the productive use of these properties. Act 2 supports three different cleanup options: restoring the site to its background condition before any releases occurred, complying with statewide health-based target levels, and site-specific cleanup based on a detailed risk assessment. Table 5.1 summarizes the target levels obtained from these three sources.

The PADEP values in Table 5.1 are based on a non-used aquifer in a non-residential area with a soil to groundwater value (i.e., COC leaching from soil and moving into groundwater) applying to the entire soil column. The refinery continues to operate, and the underlying aquifer is not used, so these choices are appropriate. The development of the PADEP target levels is well documented, and these values were set as the target levels in the creation of risk maps for the Marcus Hook facility. Because PADEP only lists values for soil and groundwater, the EPA Region III standards were used for air. Table 5.2 shows a breakdown of the target levels used in the calculations. The following conversion was employed for modifying the air target levels:  $\text{ppm} (@ 25^{\circ}\text{C}) = \mu\text{g}/\text{m}^3 * 1 \text{ g}/10^6 \mu\text{g} * 1 \text{ mol}/x \text{ g} * 1 \text{ m}^3/1000 \text{ L} * 22.4 \text{ L}/1 \text{ mol} * 298 \text{ K}/273 \text{ K} * 10^6 \text{ mol vapor}$ .

**Table 5.1: Summary of Target Levels**

COC	BP Guidance Document			EPA Region III			PADEP	
	GW <sup>1</sup> (µg/L)	Soil <sup>2</sup> (mg/kg)	Air <sup>3</sup> (µg/m <sup>3</sup> )	GW <sup>2</sup> (µg/L)	Soil <sup>5</sup> (mg/kg)	Air <sup>6</sup> (µg/m <sup>3</sup> )	GW <sup>3</sup> (µg/L)	Soil <sup>8</sup> (mg/kg)
Benzene	5	6.14	0.678	0.36	200	0.22	500	50
Toluene	1000	> SAT	1780	750	410000	420	100000	10000
Ethyl Benzene	700	> SAT	4470	1300	200000	1000	70000	7000
Xylenes	10000	> SAT	3130	12000	1000000	7300	180000	10000
Naphthalene	20	> SAT	> VP	1500	82000	150	20000	5000

Notes:

<sup>1</sup> Ingestion

<sup>2</sup> Direct Contact with Surficial Soils

<sup>3</sup> Outdoor Inhalation

<sup>4</sup> Tap Water

<sup>5</sup> Ingestion (Industrial)

<sup>6</sup> Ambient

<sup>7</sup> Non-use aquifer, non-residential

<sup>8</sup> Non-use aquifer, non-residential (100\*GW MSC or generic value; whichever is greater)

> VP = Greater than Vapor Pressure

> SAT = Greater than Saturation

**Table 5.2: Chosen Target Levels**

COC	GW (µg/L)	Soil (µg/kg)	Air (ppm)
Benzene	500	50000	6.89E-05
Toluene	100000	10000000	0.11
Ethyl Benzene	70000	7000000	0.23
Xylenes	180000	10000000	1.68
Naphthalene	20000	5000000	0.03

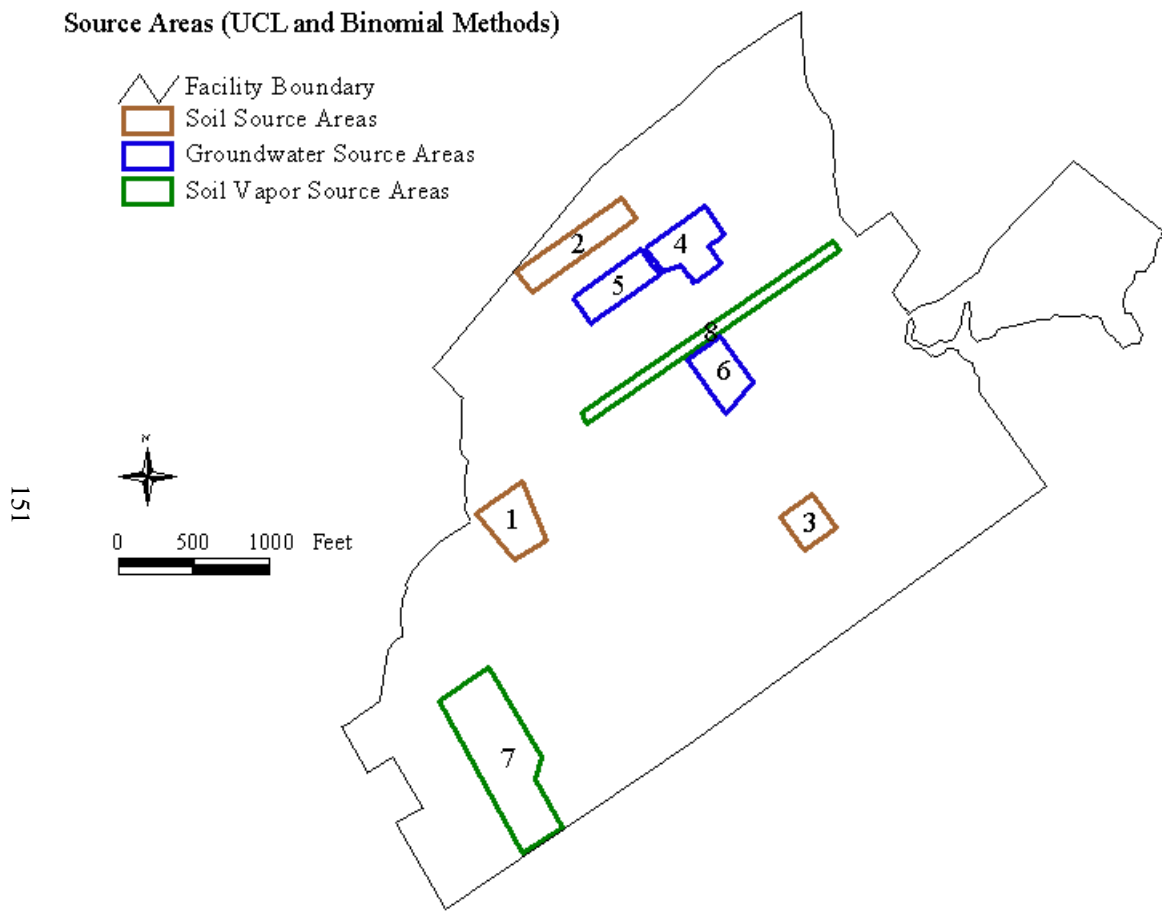
The groundwater and soil target levels are from PADEP (1997) and the air target levels are from EPA Region III (1998).

### 5.2.2 UCL and Binomial Results

For the application of the first two methods with the Marcus Hook facility data, five chemicals were analyzed in eight different source areas. The chosen source areas included a storage tank dike area, a former surface impoundment, a tank farm, three process areas, a pipeline, and the subsurface outside a former lubricant packaging plant (Figure 5.7). Three of these areas (1-3) were analyzed for soil concentrations, another three (4-6) were analyzed for groundwater concentrations, and two (7-8) were analyzed for soil vapor concentrations. The choice of these areas was based primarily on the amount of available data, but some of the data were limited even within these areas, so not all of the chemicals were analyzed in each source area. For instance, no soil vapor data were available for naphthalene, and thus, no calculations were performed. Additionally, the no-detect samples were not included in the UCL analysis. If less than 15% of the samples in a data set are no-detects, the results of the statistical tests will not be substantially affected, and a no-detect can be replaced by a value of 1/2 of the practical quantification limit (BP, 1997). However, in these data sets, it is common to find that more than 15% of the samples are no-detects. Therefore, these no-detects were left out of the calculations.

Tables 5.3 - 5.7 summarize the results of the application of the first two methods for each chemical. For the UCL method, the number of data points represents the number of measurements in that source area excluding no-detects. Test 1 shows whether the data set passed the Coefficient of Skewness test, and Test 2 shows whether the data set passed the Coefficient of Variance Test. The final two columns of the first method report the calculated UCL and whether this UCL is less than the target level. A "NO" means that the representative concentration for the area is above the target level for the particular COC for exposures at the source area. The columns for the Binomial Distribution method display the number of measurements for the area including no-detects, the best





**Figure 5.7: Source Areas Chosen for UCL and Binomial Analyses. Area 1 = surface impoundment, Area 2 = tank farm, Area 3 = storage tank dike area, Area 4 = catalytic cracker process area, Area 5 = alkylation process area, Area 6 = platformers and naphthalene process area, Area 7 = subsurface in vicinity of lubricant packaging plant, and Area 8 = drainage ditch.**

**Table 5.3: Benzene Results Summary**

Area	Medium	Target Level	Method 1 - 95% Upper Confidence Limit					Method 2 - Binomial Distribution				Tier 1 Analysis	
			# of Points	Test 1	Test 2	95% UCL of mean	Below C <sub>target</sub>	# of Points	Estimate	Lower Limit	Upper Limit	Highest Concentration	Below C <sub>target</sub>
1	Soil	50000 µg/kg	14	FAIL	PASS	54 µg/kg	YES	46	0.0	0.0	4.9	4400 µg/kg	YES
2	Soil	50000 µg/kg	14	FAIL	PASS	1918 µg/kg	YES	49	2.0	0.2	9.4	360000 µg/kg	NO
3	Soil	50000 µg/kg	7	PASS	PASS	20644 µg/kg	YES	12	16.7	3.1	43.9	135000 µg/kg	NO
4	Groundwater	500 µg/L	13	PASS	PASS	469 µg/L	YES	22	13.6	3.9	31.6	25829 µg/L	NO
5	Groundwater	500 µg/L	16	PASS	PASS	1391 µg/L	NO	17	47.1	26.1	69.0	45184 µg/L	NO
6	Groundwater	500 µg/L	6	PASS	PASS	641 µg/L	NO	9	22.2	4.2	55.0	1345 µg/L	NO
7	Soil Vapor	6.89E-05 ppm	13	PASS	PASS	20 ppm	NO	15	86.7	63.7	97.6	229 ppm	NO
8	Soil Vapor	6.89E-05 ppm	2	-	FAIL	15538 ppm	NO	16	12.5	2.3	34.4	14 ppm	NO

152

**Notes:**

Test 1 - Coefficient of skewness test for normality (CS < 1.0)

Test 2 - Coefficient of variance test for normality (CV < 1.0)

95% UCL =  $x + t\{0.05, n-1\} * \sigma/n^{0.5}$

**Table 5.4: Ethyl Benzene Results Summary**

Area	Medium	Target Level	Method 1 - 95% Upper Confidence Limit					Method 2 - Binomial Distribution				Tier 1 Analysis	
			# of Points	Test 1	Test 2	95% UCL of mean	Below C <sub>target</sub>	# of Points	Estimate	Lower Limit	Upper Limit	Highest Concentration	Below C <sub>target</sub>
1	Soil	7000000 µg/kg	16	PASS	PASS	307 µg/kg	YES	46	0.0	0.0	4.9	41800 µg/kg	YES
2	Soil	7000000 µg/kg	23	PASS	PASS	5209 µg/kg	YES	49	0.0	0.0	4.6	49000 µg/kg	YES
3	Soil	7000000 µg/kg	8	PASS	PASS	19191 µg/kg	YES	12	0.0	0.0	17.5	346000 µg/kg	YES
4	Groundwater	70000 µg/L	12	PASS	PASS	421 µg/L	YES	22	0.0	0.0	9.9	32822 µg/L	YES
5	Groundwater	70000 µg/L	13	PASS	PASS	449 µg/L	YES	17	0.0	0.0	12.7	5975 µg/L	YES
6	Groundwater	70000 µg/L	5	PASS	PASS	462 µg/L	YES	9	0.0	0.0	22.6	542 µg/L	YES
7	Soil Vapor	0.23 ppm	11	PASS	FAIL	5.5 ppm	NO	15	73.3	49.0	90.4	22 ppm	NO
8	Soil Vapor	0.23 ppm	2	-	PASS	477 ppm	NO	16	12.5	2.3	34.4	12 ppm	NO

153

**Notes:**

Test 1 - Coefficient of skewness test for normality (CS < 1.0)

Test 2 - Coefficient of variance test for normality (CV < 1.0)

$$95\% \text{ UCL} = \bar{x} + t\{0.05, n-1\} * \sigma/n^{0.5}$$

**Table 5.5: Naphthalene Results Summary**

Area	Medium	Target Level	Method 1 - 95% Upper Confidence Limit				Method 2 - Binomial Distribution				Tier 1 Analysis		
			# of Points	Test 1	Test 2	95% UCL of mean	Below C <sub>target</sub>	# of Points	Estimate	Lower Limit	Upper Limit	Highest Concentration	Below C <sub>target</sub>
1	Soil	5000000 µg/kg	22	PASS	PASS	3358 µg/kg	YES	60	0.0	0.0	3.8	12200 µg/kg	YES
2	Soil	5000000 µg/kg	23	FAIL	PASS	8172 µg/kg	YES	41	0.0	0.0	5.5	37000 µg/kg	YES
3	Soil	5000000 µg/kg	5	PASS	PASS	149586 µg/kg	YES	10	0.0	0.0	20.6	262000 µg/kg	YES
4	Groundwater	20000 µg/L	11	FAIL	PASS	105 µg/L	YES	16	0.0	0.0	13.4	17460 µg/L	YES
5	Groundwater	20000 µg/L	12	FAIL	PASS	96 µg/L	YES	15	0.0	0.0	14.2	5182 µg/L	YES
6	Groundwater	20000 µg/L	3	PASS	PASS	16 µg/L	YES	8	0.0	0.0	25.0	10 µg/L	YES

**Notes:**

Test 1 - Coefficient of skewness test for normality (CS < 1.0)

Test 2 - Coefficient of variance test for normality (CV < 1.0)

$$95\% \text{ UCL} = x + t\{0.05, n-1\} * \sigma/n^{0.5}$$

**Table 5.6: Toluene Results Summary**

Area	Medium	Target Level	Method 1 - 95% Upper Confidence Limit					Method 2 - Binomial Distribution				Tier 1 Analysis	
			# of Points	Test 1	Test 2	95% UCL of mean	Below C <sub>target</sub>	# of Points	Estimate	Lower Limit	Upper Limit	Highest Concentration	Below C <sub>target</sub>
1	Soil	10000000 µg/kg	29	FAIL	FAIL	28 µg/kg	YES	46	0.0	0.0	4.9	43100 µg/kg	YES
2	Soil	10000000 µg/kg	6	PASS	PASS	1308 µg/kg	YES	49	0.0	0.0	4.6	3500 µg/kg	YES
3	Soil	10000000 µg/kg	9	FAIL	PASS	11923 µg/kg	YES	12	0.0	0.0	17.5	808000 µg/kg	YES
4	Groundwater	100000 µg/L	11	FAIL	PASS	530 µg/L	YES	22	0.0	0.0	9.9	85420 µg/L	YES
5	Groundwater	100000 µg/L	14	PASS	PASS	697 µg/L	YES	17	0.0	0.0	12.7	16554 µg/L	YES
6	Groundwater	100000 µg/L	6	PASS	PASS	66 µg/L	YES	9	0.0	0.0	22.6	238 µg/L	YES
7	Soil Vapor	0.11 ppm	12	PASS	PASS	21 ppm	NO	15	80.0	56.1	94.4	162 ppm	NO
8	Soil Vapor	0.11 ppm	2	-	FAIL	4539 ppm	NO	16	12.5	2.3	34.4	10 ppm	NO

155

**Notes:**

Test 1 - Coefficient of skewness test for normality (CS < 1.0)

Test 2 - Coefficient of variance test for normality (CV < 1.0)

$$95\% \text{ UCL} = x + t\{0.05, n-1\} * \sigma/n^{0.5}$$

**Table 5.7: Xylenes Results Summary**

Area	Medium	Target Level	Method 1 - 95% Upper Confidence Limit					Method 2 - Binomial Distribution				Tier 1 Analysis	
			# of Points	Test 1	Test 2	95% UCL of mean	Below C <sub>target</sub>	# of Points	Estimate	Lower Limit	Upper Limit	Highest Concentration	Below C <sub>target</sub>
1	Soil	10000000 µg/kg	10	PASS	PASS	51 µg/kg	YES	12	0.0	0.0	17.5	156 µg/kg	YES
2	Soil	10000000 µg/kg	31	PASS	PASS	5062 µg/kg	YES	49	0.0	0.0	4.6	200000 µg/kg	YES
3	Soil	10000000 µg/kg	9	PASS	PASS	54302 µg/kg	YES	12	0.0	0.0	17.5	2060000 µg/kg	YES
4	Groundwater	180000 µg/L	2	-	PASS	8.9 x 10 <sup>8</sup> µg/L	NO	7	0.0	0.0	28.0	2000 µg/L	YES
7	Soil Vapor	1.68 ppm	9	PASS	PASS	12 ppm	NO	15	46.7	24.4	70.1	35 ppm	NO
8	Soil Vapor	1.68 ppm	2	-	PASS	4381 ppm	NO	16	12.5	2.3	34.4	22 ppm	NO

**Notes:**

Test 1 - Coefficient of skewness test for normality (CS < 1.0)

Test 2 - Coefficient of variance test for normality (CV < 1.0)

95% UCL =  $x + t\{0.05, n-1\} * \sigma/n^{0.5}$

estimate probability ( $s/n$ ), the 90% lower exceedance probability, and the 90% upper exceedance probability, respectively. Each table also contains two columns that compare the highest detected concentration in the source area to the target level. This type of comparison is typical of a RBCA Tier 1 Analysis.

There are some interesting things to point out in these results. The UCLs for benzene are below the soil target levels in all three areas analyzed but are above the target levels for all the groundwater and soil vapor source areas. In general, though, the UCL concentrations for the other four chemicals are less than the target level in all source areas except those for soil vapor. In fact, there are no measurements with values above the target level in any soil or groundwater source area for all other COCs except for benzene. This result should be expected as benzene has the lowest target levels of all the COCs studied. These methods are probably not as accurate for the soil vapor data measurements since the detection limits are generally higher than the target levels.

Also notice how the amount of data affects both methods. For example, in Groundwater Source Area 4, there are seven xylene measurements. None of these measurements are above the target level, and five of them are no-detects. However, using only two data points for the UCL method leads to a very high predicted concentration value, which, in this example, is above the target level. Additionally, the binomial calculations indicate with 90% confidence that the exceedance probability is less than 28%. If 50 measurements had been obtained from the area, and no results yielded a value above the target level, the upper exceedance probability would drop to 4.5%.

Besides quantity of measurements, the use of no-detects also plays a prominent role in the method results. For instance consider the benzene measurements in both soil vapor source areas. Source area 7 has 15 benzene measurements, of which 13 are above the target level. Source area 8 has 16 measurements, of which 14 are no-detects. Looking simply at these numbers, one

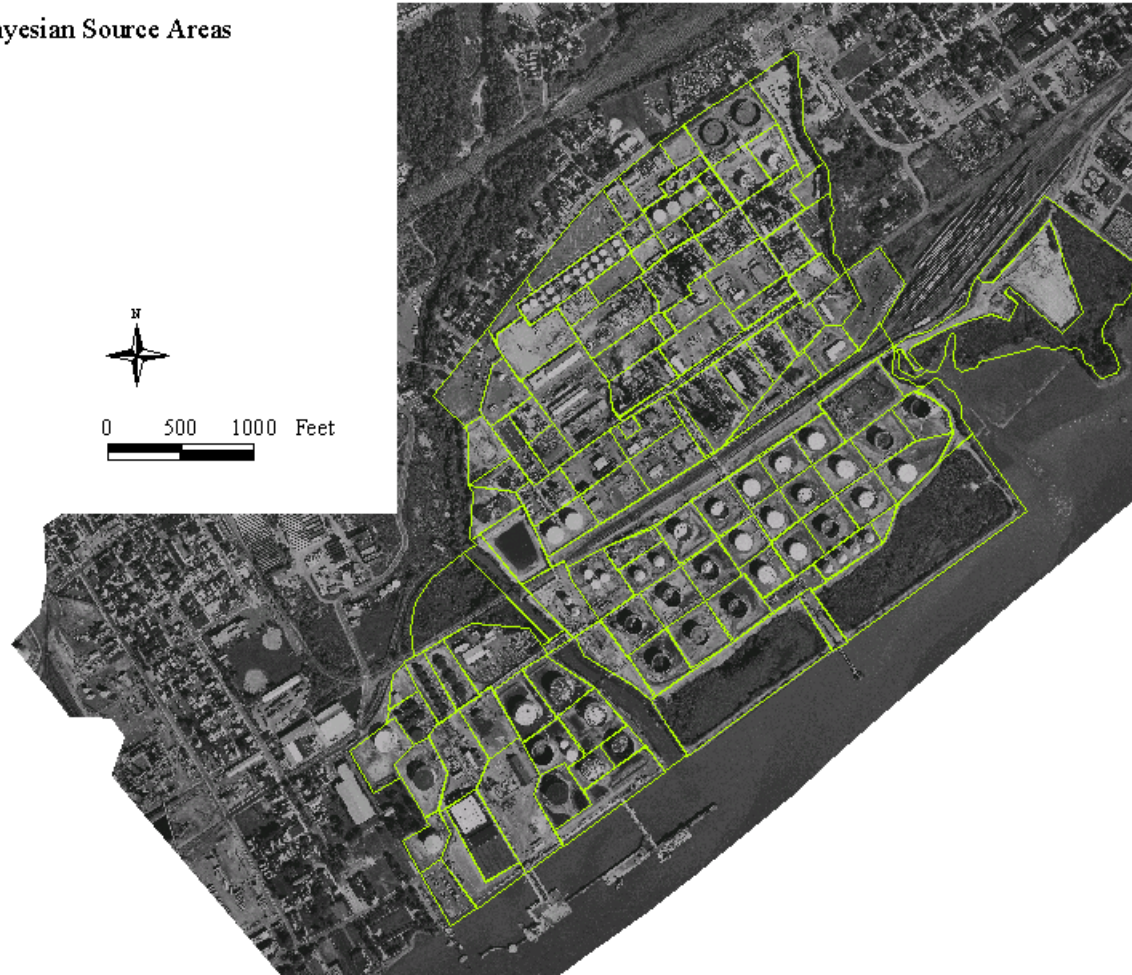
would expect that it is much more likely that source area 7 has a higher fraction of soil with concentrations above the target level than source area 8. This expectation is confirmed with the Binomial Distribution method as the upper confidence levels are 97.6% and 34.4% for areas 7 and 8, respectively. However, the UCL results show that the expected concentration is much greater for source area 8. This result is again a product of having only two data points, and the use of this type of analysis for only two data points is not recommended. Another important thing to notice is that even when using the log transformed data, many of the data sets do not pass either of the normality tests. The Coefficient of Variance Test has a higher PASS ratio, but as mentioned before, this method is not as good a test as the Coefficient of Skewness Test. Appendix H contains all of the Excel data sheets for the application of the UCL method.

### **5.2.3 Bayesian Results**

As an example application, the Bayesian Approach was applied for benzene soil concentrations across the entire site. The first step in the application of this method was to define representative soil source areas across the site. These areas have a consistent soil texture and more importantly, the fractions of soil above a target level in these areas are expected to be constant because the areas were defined based on reasonably consistent historical operations and product storage. Some example areas include dike areas around storage tanks, retention basins, process areas, and historical features such as a former service station. The source areas were defined using the digital orthophotographs and a historical database of operation activities, product storage, and previous releases (Figure 5.8). This database was compiled by the current facility owner based on operation records, characterization studies, and employee knowledge. Using this database, 127 representative areas were defined, and a mean fraction ( $\mu_p$ ) of soil above a target level for benzene along with a variance ( $\sigma_p$ ) on this fraction were assigned for



Bayesian Source Areas



159

Figure 5.8: Source Areas Defined for Bayesian Analysis.

each area. The following values were assigned:

- $(\mu_p = 0.75, \sigma_p = 0.25)$  - Areas storing or working with products high in benzene concentration such as gasoline. Releases of unknown amounts have likely occurred in these areas. Examples include gasoline storage tanks, refining process areas, and a former service station.
- $(\mu_p = 0.25, \sigma_p = 0.25)$  - Areas storing or working with crude which contains benzene but not in concentrations as high as gasoline. Releases could also have occurred in these areas. Examples include crude storage tanks, crude stills, and Marcus Hook Creek sediments.
- $(\mu_p = 0.10, \sigma_p = 0.10)$  - Areas with heavy products that contain low concentrations of benzene. Examples include storage tanks of fuel oil, cutter stock, slop, and jet fuel.
- $(\mu_p = 0.05, \sigma_p = 0.05)$  - Areas expected to be low in benzene concentration but which receive different types of product and waste. Examples include the former surface impoundment and areas around the wastewater treatment plant.
- $(\mu_p = 0.005, \sigma_p = 0.005)$  - Areas expected to have no benzene concentrations above the target level. Examples include wetlands, parking lots, and buildings.
- $(\mu_p = 0.50, \sigma_p = 0.289)$  - Uniform distribution representing areas in which not enough information is known to be able to make an effective judgment. Examples include the railroad tracks where specific loading and unloading practices are unknown and tanks with unknown previous storage.

Using these values, a probability that no remedial action would be required was assigned to each area. This probability assumes that remedial action will be necessary if one sample measurement out of twenty from the area is above a target level. Thus, this value is the probability of no hits in twenty samples, which can be calculated as (refer to Appendix E for derivation):

$$\bullet \quad P(x = 0) = \left[ \frac{\Gamma(q+r) * \Gamma(n+r)}{\Gamma(r) * \Gamma(q+r+n)} \right] \quad [5.20]$$

where  $n = 20$  samples.

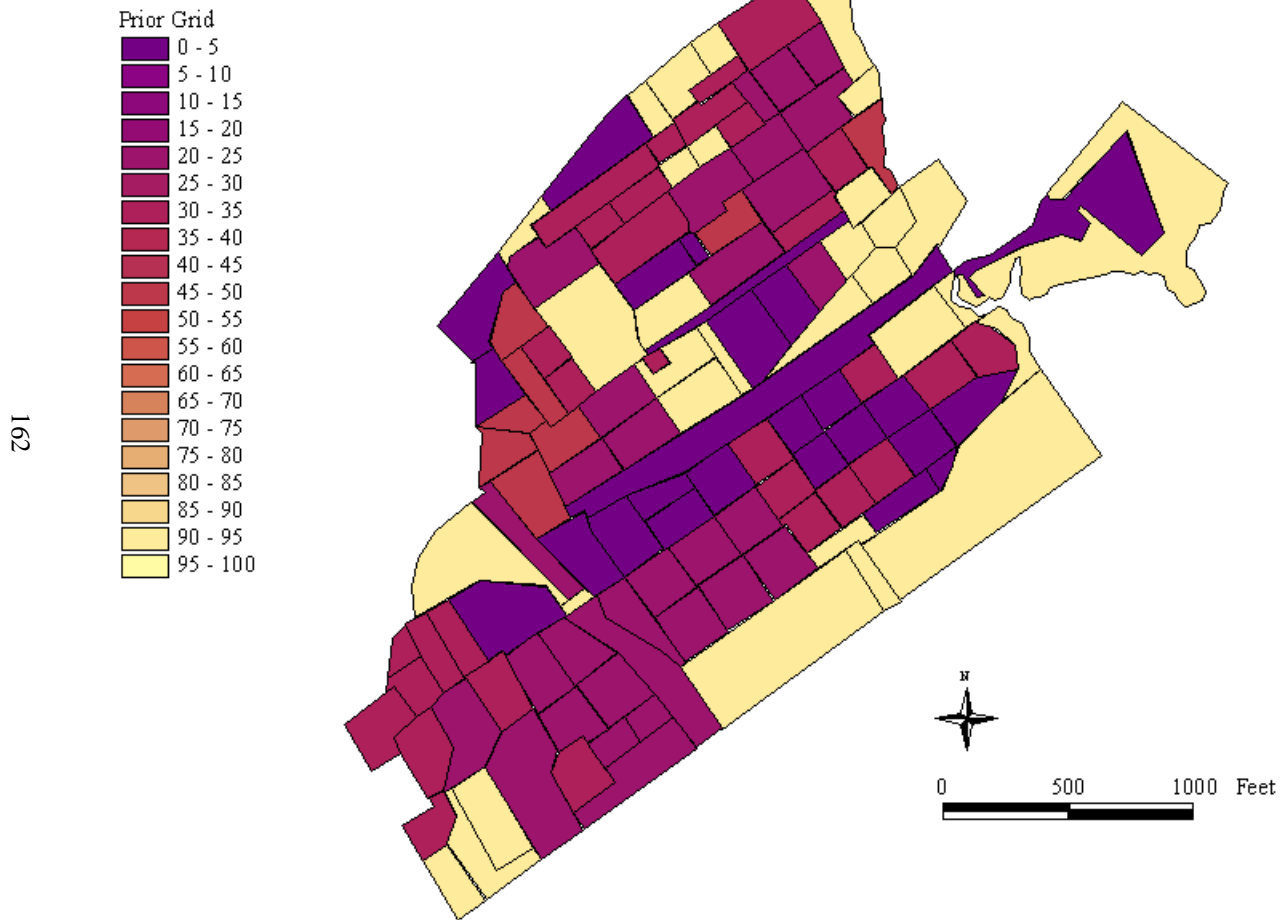
The next step was to update the prior probabilities using any existing benzene soil data in each area. The number of measurements and successes (i.e.,

measurements above a target level) for each area were determined using query tools in ArcView. These data were then input into an Excel spreadsheet that calculated new fractions, variances, and probabilities utilizing Equations 5.12 - 5.20. The updated information was then imported back into ArcView and joined with the coverage of representative areas. Of the 127 areas defined in ArcView, soil data were available for 37 of these areas. Thus, the probability based on the previously assigned fraction and variance does not change for 90 areas. Figures 5.9 and 5.10 show the areas delineated by prior and updated probabilities, respectively. Figure 5.11 shows the probability changes within each area.

The results of this method identify areas of concern, areas with concentrations above or below the previously anticipated level, and areas where more information is needed. For example, the expected fraction of soil above a target level was set at 0.05 for the former surface impoundment based on prior information. There are 46 benzene soil measurements from this impoundment and none of them are above the target level (50,000  $\mu\text{g}/\text{kg}$ ). As a result, the mean fraction drops from 0.05 to 0.014 while the probability that the area will not require remedial action jumps from 49.7% to 78.0%. As previously described, this value is the probability of no hits in twenty samples. The results for this area are also consistent with those obtained using the UCL and Binomial methods. The calculated UCL (54  $\mu\text{g}/\text{kg}$ ) is significantly less than the target level while the binomial method's best estimate and upper exceedance probabilities are 0% and 4.9%, respectively.

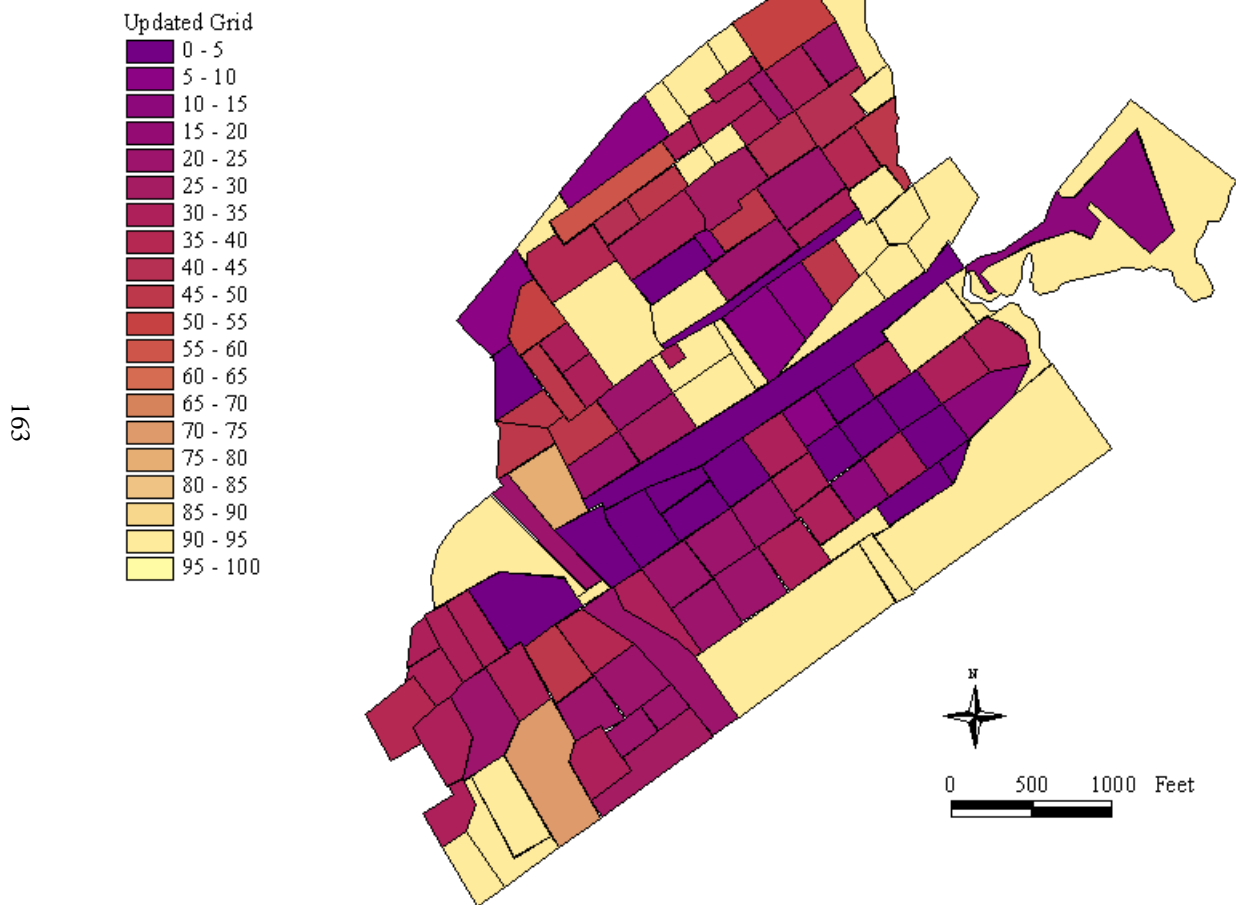
Some areas show higher fractions than what was originally expected. Benzene concentrations were not expected to be above the target level in the dike area surrounding a fuel oil storage tank. However, two samples out of twelve in this area yielded concentrations above the target level. In this example the mean fraction rose from the prior value of 0.1 to 0.14 while the probability of attainment dropped from 33.9% to 12.4%. Other areas returned results that were

### Bayesian Areas by Prior Probability



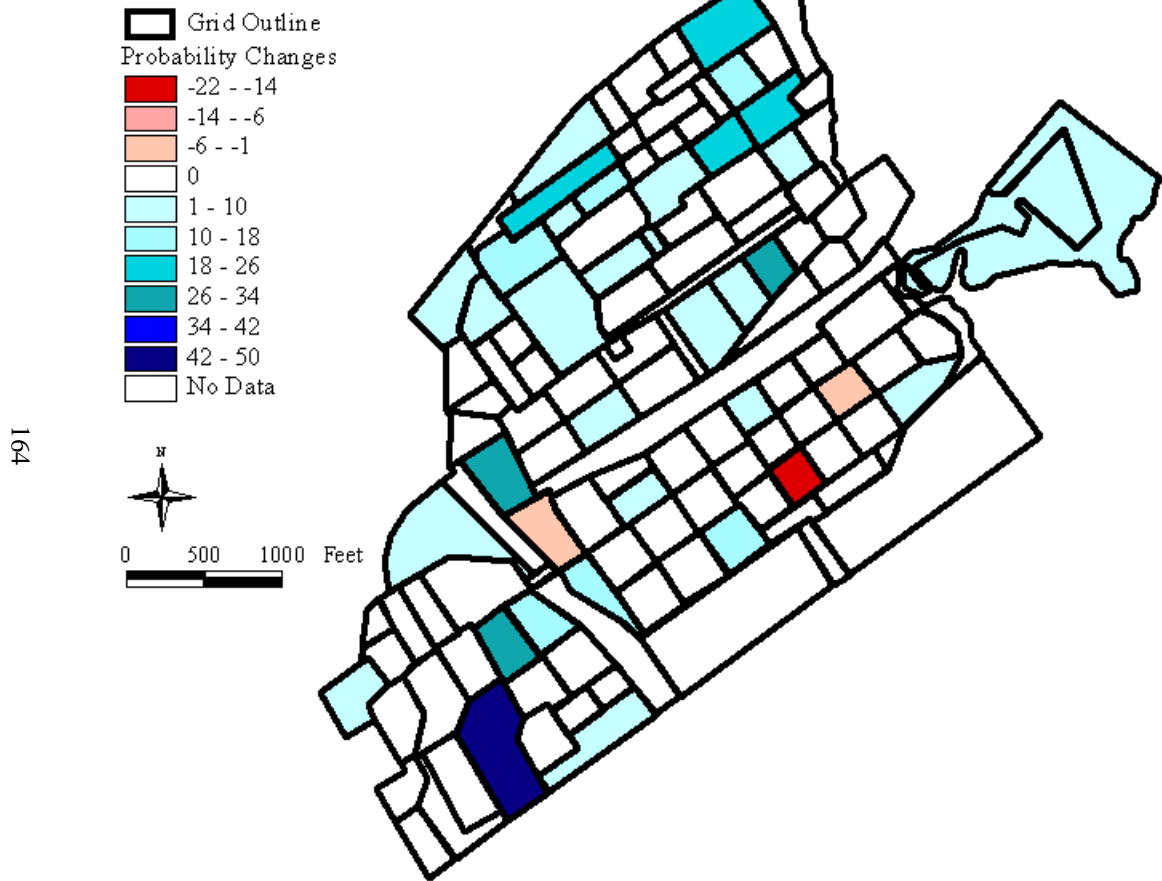
**Figure 5.9: Prior Grid of the Probability of No Hits (i.e., Measurements Above the Target Level) in Twenty Samples.**

### Bayesian Areas by Updated Probability



**Figure 5.10: Updated Grid of the Probability of No Hits (i.e., Measurements Above the Target Level) in Twenty Samples.**

### Bayesian Areas by Change in Probability



164

**Figure 5.11: Change in the Probability of No Hits in Twenty Samples for Each Source Area. The red areas show decreases (i.e., probability of attainment decreases) and the blue areas show increases.**

consistent with the prior information. In one wetland area, the probability of attainment increased slightly from 90.9% to 91.0% based on available data in that area. In areas where little information was known and the uniform distribution was assigned, the prior probability of measuring a soil concentration below the target level is only 4.8%. This probability is so low simply because there is not enough information with which to make an accurate judgement. Therefore, more data would be particularly useful in these areas. Also, because  $q'$  and  $r'$  are 1 for the uniform distribution, any data obtained have a significant impact on the result. Two areas (46 and 122, see Appendix H), one of which contains a storage tank and the other contains bare soil, showed a jump in probability from 4.8% to 13.0% simply from two sample measurements. The complete listing of measurements and probability changes in each area is available in Appendix H.

#### **5.2.4 Conclusions**

The three methods of risk mapping studied in this project all have their own advantages and disadvantages. The Upper Confidence Limit method provides a simple and easy way to calculate a respective concentration for a particular area. It provides a first look at the data and allows one to get a feel for what types of action will be required in future analysis. The disadvantages of this method are that it assumes the data set is normally distributed, it does not account for no-detect measurements, and the results are often highly dependent on the number of sample measurements. The assumption of normality is sometimes a stretch with concentration data, even after a log-normal transformation, because the concentration values can cover such a wide range of magnitudes. While the no-detect measurements were not included in this analysis, more advanced methods exist for incorporating these data into the UCL calculation.

The Binomial Distribution method is more calculation-intensive than the UCL method, but it accounts for the no-detects and quantifies the chance that the

concentrations in an area will be above a target level. It provides both a best estimate probability along with upper and lower confidence limits on this estimate. The upper exceedance probability can be used as a conservative screening level. For instance, if the upper limit probability is less than 10%, one could be fairly certain that the area is below the target level. The results are easy to understand, and they could be used to justify decisions made in developing a remedial action program. The disadvantages of this method are that it does not take into account the magnitude of each measurement, the result is heavily weighed by the amount of data in each area, and it does not incorporate prior information about activities in an area.

The Bayesian Approach, as it is presented here, also does not take the magnitude of each measurement into account, but the results for each area are based on both prior information and available data. The power of this method is obvious. It incorporates all information sources and provides a look at the comparison of exposure concentrations with a target level. This method requires an understanding of previous activities at the site, but these activities can usually be researched. Researching helps the user to obtain a better understanding of the problem and what to expect. Another advantage of this method is that it can be used to identify areas where more data would be useful towards making decisions, and it can be continually updated as more information becomes available. Two of the objectives of the Marcus Hook project are to effectively use the existing site data and to make good decisions about where additional data should be obtained. For instance, it would not be as beneficial to obtain twenty new samples from an area that is already expected to be above a target level than it would to obtain twenty samples from area with very little information. Determining where more data would be useful is referred to as a "value of information" calculation (Koerner, *et al.*, 1998).



In summary, all three of these methods can be used to quantify exposures at a particular source area, and thus provide a solid foundation for making risk-based decisions. All the calculations can be performed using Excel's built-in functions, and the capabilities of ArcView allow the results to be displayed visually.

## **CHAPTER SIX**

### **CONCLUSIONS**

This project presents a digitally-based methodology for approaching risk assessments at large and complex industrial facilities, using the Marcus Hook Refinery in Pennsylvania as a case study site. The objective of this research was to demonstrate the development of a digital facility description and its use as an effective environmental risk assessment tool that can be applied to any site. The digital facility description is the collection of physical, chemical, geological, and hydrogeological information that has been spatially referenced in a geographic information system (GIS). The digital facility description has two components: (1) the spatial database, and (2) the tabular database. Two separate databases were used because the current capabilities of ArcView do not provide an efficient way to relationally connect tables.

The key component towards creating a powerful digital facility description is data development, particularly with respect to determining what data need to be compiled. For the spatial database, the following coverages were the most critical data layers for the environmental risk assessment activities at Marcus Hook:

<b>Regional Spatial Data</b>	<b>Facility Spatial Data</b>
Average Annual Runoff	Areas of Concern
Census Tracts	Digital Terrain Model
County Streams	Historical Features
Digital Elevation Model	Monitoring Wells
EPA Regulated Facilities	NPDES Sampling Points
Land Use	Oily Water Sewer Line
Soil Properties	Parcels
Stream Flow Basin Characteristics	Pipelines
Surface Geology	Property Boundary
USGS Gage Stations	Storage Tanks
Climatological Data	Structures
	Surface Cover
	Surface Hydrology

These coverages provide not only a layout and description of the actual system but also a means for evaluating the source of a chemical release and the migration and attenuation of any chemicals to a receptor. Constructing a spatial database for Marcus Hook required obtaining regional coverages through the Internet and creating facility coverages using aerial photogrammetry. To fit the needs of the project, the regional coverages were reprojected into the Pennsylvania State Plane Coordinate System, which provides a minimal distortion between features for this site. The mapping efforts on the facility scale for Marcus Hook were conservative in the sense that a large number of coverages were created with very accurate feature identification. The detailed accuracy was particularly useful for the orthophotographs and digital terrain model, but it was not as critical for the facility features. Such an extensive effort would not be necessary to utilize this approach again, and other alternatives of coverage creation could be pursued such as digitizing from the orthophotographs and using CAD drawings.

Similar to the spatial database, the tabular database was designed and constructed based on what data were required to support the environmental risk assessment activities. These data included sampling point locations, groundwater levels, soil profiles, type of samples collected, and concentration analyses. The data were compiled from reports and characterization studies and then organized in a relational structure consisting of different data levels, in which the information in each subsequent level further describes the data in the previous level. This structure eliminates the repetition of data and provides a simple format for querying specified information.

The tabular database is a separate entity from the spatial database, but the two can be combined in the GIS application using a dynamic link (ODBC) between the database program and the GIS software. The coordinate systems in each database must be consistent so that features line up properly in the geographic

environment. For Marcus Hook, a transformation of coordinates in the tabular database from a local plant grid coordinate system to the state plane coordinate system was performed using trigonometric relationships. The dynamic link between the two databases at Marcus Hook demonstrates the ability to combine information about environmental conditions that change in time with information about the spatial features of the site. Each time information is updated in either database, it is updated throughout the GIS-tabular database system.

The effectiveness of the digital facility description as an environmental risk assessment tool is shown in the map-based modeling and exposure analysis activities for Marcus Hook. Specifically, models were developed for both surface water runoff and groundwater flow. The surface water runoff delineation tools utilized are normally applied to regional areas with a 30-meter or larger digital elevation model (DEM). However, this thesis effectively illustrates the application of these tools for the Marcus Hook project using a 3-foot cell DEM. Another product of this research is the *COC Transport* extension, which can be used to look at concentration changes between two points. *COC Transport* is a simple grid-based mixing model, and the results can be used to support Risk-Based Corrective Action (RBCA) Tier 1 and Tier 2 analysis activities. Similarly, the groundwater model supports the activities of classifying aquifer status and establishing site-specific target levels. As opposed to the surface water model, which was developed within the GIS application, the groundwater modeling was performed using an external program, the Groundwater Modeling System (GMS). However, this program was linked to the GIS framework through the common use of spatial data from the digital facility description.

The exposure analysis activities focussed on evaluating concentrations in a source area and computing the fraction of each area that could be above a target level. Three different methods of "risk mapping" were studied: (1) a 95% Upper Confidence Limit (UCL), (2) the Binomial Distribution, and (3) a Bayesian

Approach. The UCL method is the simplest of the three studied, and for many projects, it is sufficient enough to demonstrate the attainment of target levels in a source area. The Binomial Distribution method is more sophisticated, but it provides a confidence range on the exceedance probability (i.e., fraction of area above the target level) for each source area. For purposes of the Marcus Hook project, the Bayesian Approach proved to be the most useful as it incorporated historical information about oil production activities in each source area and identified which areas might require corrective action, which areas might not require corrective action, and which areas require more data measurements to make an appropriate evaluation. As opposed to the first two methods, which depend on a large number of measurements for applicable results, the Bayesian Approach helps quantify the value of collecting additional information.

The research at Marcus Hook was in its second year of existence at the time this document was published. Some of the project activities still to be completed included connecting the surface water and groundwater models, specifying the details of the site conceptual model database for an area of the facility, and constructing the spatial site conceptual model. This spatial site conceptual model represents the ultimate goal of the project since it combines each aspect of the spatial risk assessment process: (1) the integrated databases, (2) the map-based modeling, and (3) the exposure analysis. The user will eventually be able to visualize in the GIS application each source, each receptor, the concentration changes from a source to a receptor, and the associated health risks. An evaluator can then make more informed risk-based decisions on what corrective action activities, if any, are appropriate.

Besides these areas of continued research, other enhancements can and will effect the application of a digital facility description in environmental risk assessment. One area of suggested study is constructing a digital facility description at other industrial sites. This research would not only validate the use

of digital facility description for environmental risk assessment purposes, but also enhance the understanding of what data are the most critical and how these data are best organized. One organization improvement could be revising the construction of the tabular database design presented here to be generic enough to fit the purposes of any site. This improvement, perhaps linked to an Intranet application, would provide a common framework for evaluating multiple sites. Another suggestion is further researching the digital facility description as a risk communication tool. Is a digital facility description methodology accepted by the regulatory agencies and public community as an effective means of evaluating the risks to human health and the environment posed by the facility? Other enhancements to the process will come with technological advancements, such as faster computers, improved software utilities, and more sophisticated transport models. Eventually, a spatial environmental risk assessment tool might exist that combines all the components of the process into a single program with no external linkages. While this goal may still be a ways off, the foundation has already been set with the material presented in this thesis on how to represent in GIS the spatial facility features and the environmental sampling data recorded through time at the facility.

## APPENDIX A

### USEFUL INTERNET RESOURCES

This section provides some useful Internet links to spatial data sites, models, presentations, and papers, all of which are relevant for environmental risk assessment. These links were accurate as of May 7, 1999, and most of them can be accessed from the author's homepage:

[www.ce.utexas.edu/prof/maidment/grad/romanek/home.html](http://www.ce.utexas.edu/prof/maidment/grad/romanek/home.html).

#### **Spatial Data Sites**

- USGS Spatial Data: [water.usgs.gov/lookup/getgislist](http://water.usgs.gov/lookup/getgislist)
- Pennsylvania Spatial Data: [www.pasda.psu.edu](http://www.pasda.psu.edu)
- Raster CD's of the United States: [www.horizons.com](http://www.horizons.com)
- Digital Line Graph Data: [edcwww.cr.usgs.gov/pub/data/DLG](http://edcwww.cr.usgs.gov/pub/data/DLG)
- Land Use Data: [ftp.epa.gov/pub/EPAGIRAS](ftp://ftp.epa.gov/pub/EPAGIRAS)
- Digital Elevation Models: [edcwww.cr.usgs.gov/nsdi/gendem.htm](http://edcwww.cr.usgs.gov/nsdi/gendem.htm)
- Hydrologic Unit Maps: [water.usgs.gov/public/GIS/huc.html](http://water.usgs.gov/public/GIS/huc.html)
- SSURGO (County Soil Data): [www.ftw.nrcs.usda.gov/ssur\\_data.html](http://www.ftw.nrcs.usda.gov/ssur_data.html)
- VEMAP Climate Data: [www.cgd.ucar.edu:80/vemap](http://www.cgd.ucar.edu:80/vemap)
- EPA Point Coverages of Regulated Facilities: [ftp.epa.gov/pub/spdata/ef](ftp://ftp.epa.gov/pub/spdata/ef)
- Surf your Watershed: [www.epa.gov/surf](http://www.epa.gov/surf)
- Find Water Resources Data in your State: [water.usgs.gov/public/wrd002.html](http://water.usgs.gov/public/wrd002.html)
- Water Resources Information: [water.usgs.gov](http://water.usgs.gov)
- Useful Internet Sites for Water Resources and GIS:  
[www.ce.utexas.edu/prof/maidment/gishydro/docs/websites/othr\\_web.htm](http://www.ce.utexas.edu/prof/maidment/gishydro/docs/websites/othr_web.htm)

#### **Models and Programs**

- Groundwater Modeling System (GMS): [www.ems-i.com/software\\_gms\\_index.html](http://www.ems-i.com/software_gms_index.html)
- Watershed Delineator Tutorial:  
[www.crrw.utexas.edu/gis/gisenv98/envrep/av3delin/webfiles/av3delin.htm](http://www.crrw.utexas.edu/gis/gisenv98/envrep/av3delin/webfiles/av3delin.htm)
- COC Transport Extension:  
[www.ce.utexas.edu/prof/maidment/grad/romanek/research/data.html](http://www.ce.utexas.edu/prof/maidment/grad/romanek/research/data.html)
- ESRI Scripts Page: [andes.esri.com/arcs-scripts/scripts.cfm](http://andes.esri.com/arcs-scripts/scripts.cfm)
- EPA Subsurface Models: [www.epa.gov/ada/csmos.html](http://www.epa.gov/ada/csmos.html)

- MODular three-dimensional finite-difference ground-water FLOW model (MODFLOW): [water.usgs.gov/software/modflow-96.html](http://water.usgs.gov/software/modflow-96.html)
- MODFLOWARC: [www.oregon.wr.usgs.gov/projs\\_dir/modflowarc/modflowarc.html](http://www.oregon.wr.usgs.gov/projs_dir/modflowarc/modflowarc.html)
- EPA Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) Model: [www.epa.gov/OST/BASINS/](http://www.epa.gov/OST/BASINS/)
- Hydrologic Engineering Center (HEC) Models: [www.wrc-hec.usace.army.mil/](http://www.wrc-hec.usace.army.mil/)
- Argus ONE: [www.argusint.com](http://www.argusint.com)
- EQuIS Geology: [www.earthsoft.com](http://www.earthsoft.com)
- The Hydrologic Modeling Extension can be downloaded via anonymous ftp to <ftp://crwr.utexas.edu/pub/gisclass/urubamba>.

### **Papers**

- "Maximizing the Value of Information in Risk-Based Decision-Making: Challenges and Solutions":  
[www.ce.utexas.edu/prof/maidment/grad/romanek/research/anspaper.pdf](http://www.ce.utexas.edu/prof/maidment/grad/romanek/research/anspaper.pdf)
- CRWR Online Reports: [www.ce.utexas.edu/org/crwr/reports/online.html](http://www.ce.utexas.edu/org/crwr/reports/online.html)

### **Presentations**

- EWRE Seminar - Environmental Risk Assessment with GIS:  
[www.ce.utexas.edu/prof/maidment/grad/romanek/seminar/sld001.htm](http://www.ce.utexas.edu/prof/maidment/grad/romanek/seminar/sld001.htm)
- GIS Enviro '98 - Implementing GIS in Environmental Risk Based Decisions:  
[www.ce.utexas.edu/prof/maidment/grad/romanek/gisenviro98/maidment/sld001.htm](http://www.ce.utexas.edu/prof/maidment/grad/romanek/gisenviro98/maidment/sld001.htm)
- Other presentations available from the GIS Enviro '98 CD-ROM:  
[www.crwr.utexas.edu/gis/gisenv98/present/present.html](http://www.crwr.utexas.edu/gis/gisenv98/present/present.html)

### **Class Projects**

- Estimating Daily Recharge from Infiltration and Evapotranspiration:  
[www.ce.utexas.edu/prof/maidment/grad/romanek/hydproject/report.html](http://www.ce.utexas.edu/prof/maidment/grad/romanek/hydproject/report.html)
- Spatial Analysis of the Marcus Hook Refinery:  
[www.ce.utexas.edu/stu/kimj/proposal.html](http://www.ce.utexas.edu/stu/kimj/proposal.html)
- Other projects available from the GIS Enviro '98 CD-ROM:  
[www.crwr.utexas.edu/gis/gisenv98/class/class.html](http://www.crwr.utexas.edu/gis/gisenv98/class/class.html)



## UT Resources

- GIS Enviro '98: [www.crrw.utexas.edu/gis/gisenv98/GisEnv98.html](http://www.crrw.utexas.edu/gis/gisenv98/GisEnv98.html)
- GIS Hydro '98: [internetcity.crrw.utexas.edu/gis/gishyd98/GisHyd98.htm](http://internetcity.crrw.utexas.edu/gis/gishyd98/GisHyd98.htm)
- Dr. David Maidment's homepage: [www.ce.utexas.edu/prof/maidment](http://www.ce.utexas.edu/prof/maidment)
- CRWR PrePro: [civil.ce.utexas.edu/prof/olivera/prepro/prepro.htm](http://civil.ce.utexas.edu/prof/olivera/prepro/prepro.htm)
- Center for Research in Water Resources: [www.ce.utexas.edu/org/crrw/home.html](http://www.ce.utexas.edu/org/crrw/home.html)

## Other

- Environmental Risk Assessment Class Page:  
[www.ce.utexas.edu/prof/maidment/risk/risksyl.html](http://www.ce.utexas.edu/prof/maidment/risk/risksyl.html)
- GIS in Water Resources Class Page:  
[www.ce.utexas.edu/prof/maidment/giswr98/giswrsyl.html](http://www.ce.utexas.edu/prof/maidment/giswr98/giswrsyl.html)
- EPA Query Mapper: [www.epa.gov/enviro/html/multisystem.html](http://www.epa.gov/enviro/html/multisystem.html)
- National Center for Geographic Information and Analysis: [www.ncgia.ucsb.edu/](http://www.ncgia.ucsb.edu/)
- Pennsylvania's Land Recycling Program:  
[www.dep.state.pa.us/dep/deputate/airwaste/wm/landrecy/MANUAL/Manual.htm](http://www.dep.state.pa.us/dep/deputate/airwaste/wm/landrecy/MANUAL/Manual.htm)
- Proposed Texas Risk Reduction Management Rule:  
[www.tnrcc.state.tx.us/waste/riskrul3.htm](http://www.tnrcc.state.tx.us/waste/riskrul3.htm)
- EPA Media Specific Tools: [www.epa.gov/epahome/dmedia.htm](http://www.epa.gov/epahome/dmedia.htm)
- EPA Region III Risk-Based Concentration Tables:  
[www.epa.gov/reg3hwmd/risk/riskmenu.htm](http://www.epa.gov/reg3hwmd/risk/riskmenu.htm)
- Information on burning streams into a dam can be obtained from the exercise at:  
[www.ce.utexas.edu/prof/maidment/CE397/urubamba/peru.htm](http://www.ce.utexas.edu/prof/maidment/CE397/urubamba/peru.htm).

**APPENDIX B**  
**SPATIAL DATA DICTIONARY**

### **Marcus Hook ArcView Project Files**

<b>Project File</b>	<b>Description</b>
basemap.apr	Facility specific coverages
3dview.apr	Three-dimensional surface representation of refinery
colorphotos.apr	Color orthophotos (separated from basemap because of size)
cadfiles.apr	Facility CAD drawings
facility.apr	Access generated coverages
regional.apr	Regional specific coverages
riskmaps.apr	Risk mapping coverages and data
runoff.apr	Surface water runoff modeling

**FACILITY DATABASE**

All **facility** vector coverages have the attributes listed below. Each individual coverage and its associated feature codes are listed in the next section.

**All Coverages Types**

Attribute	Attribute Description
Shape	Point, Polyline, or Polygon
CoverageName#	Internal feature number
CoverageName-id	User-assigned feature number
Fcode	Feature code
Loc_id	Location identifier
Loc_label	Location label
Prev_id	Previous location identifier
Prev_label	Previous location label
Loc_elev	Location elevation

**Line Coverages** (in addition to the fields listed above)

Attribute	Attribute Description
Fnode#	Internal number of from-node
Tnode#	Internal number of to-node
Lpoly#	Internal number of polygon to left of arc
Rpoly#	Internal number of polygon to right of arc
Length	Length of arc in coverage units

**Polygon Coverages** (in addition to the fields listed above)

Attribute	Attribute Description
Area	Area of polygon feature
Perimeter	Perimeter of polygon feature

**Facility Coverages - Descriptions and Attributes**

<b>Coverage Number</b>	<b>Coverage Name</b>	<b>Coverage Description</b>	<b>Coverage Type</b>	<b>Feature Code</b>	<b>Feature Description</b>
1	SURFHVD	Surface Water	Polygon, Line	1005	Hidden stretches of Stony Run Creek
				1010	Streams
				1020	Wetlands
				1030	River
				1040	Ponds
				1060	Tidal Flat
2	OILSEW	Oily Water Sewer Line	Line	0	Drains/Manholes
				1100	Lines
4	MARNAV	Marine Navigation	Line	1300	Docks/Ramps/Piers
5	BOUND	Boundary	Polygon, Line	1630	Property Boundary
6	CONTROL	Control Points	Point	0	Control Point
7	SURFCOV	Surface Cover	Polygon, Line	1005	Hidden stretches of Stony Run Creek
				1010	Streams
				1020	Wetlands
				1030	River
				1040	Ponds
				1060	Tidal Flat
				7100	Gravel
				7110	Grass/Vegetation
				7120	Concrete Pavement
				7130	Asphalt Pavement
				7140	Soil (no vegetation)
9999	Structures				
8	VEG	Vegetation	Line, Point	1900	Bush (Single)
				1910	Brushline

Coverage Number	Coverage Name	Coverage Description	Coverage Type	Feature Code	Feature Description
10	TRANS	Primary Transportation	Polygon, Line	1940	Tree (Single)
				1950	Treeline
				2100	Centerline
				2110	Curb
				2120	Road (Paved)
				2140	Road (Unpaved)
				2160	Drive (Paved)
				2170	Drive (Unpaved)
				2180	Parking (Paved)
				2190	Parking (Unpaved)
				6700	Parking (Curbed)
11	PED	Pedestrian	Line	2210	Foot Bridge
				2220	Sidewalk
12	BRIDGE	Bridges	Line	2320	Bridge
13	RAIL	Railroad Features	Line, Point	2400	Miscellaneous Railroad Feature
				2410	Railroad
				2430	Railroad Signal
				2440	Railroad Switch Box
14	BLDG	Structures	Polygon, Line	1850	Slabs (Asphalt)
				1860	Slabs (Concrete)
				2500	Building
				2520	Miscellaneous Structure
				2530	Overhead Roofline
				2540	Ruin/Foundation
				2570	Trailer Homes/Offices
				2580	Steps
				2590	Wooden Decks

<b>Coverage Number</b>	<b>Coverage Name</b>	<b>Coverage Description</b>	<b>Coverage Type</b>	<b>Feature Code</b>	<b>Feature Description</b>
15	TANK	Storage Tanks	Polygon, Line	2600	Above Ground Storage Tanks
16	PROARVES	Process Area Vessels	Polygon	2700 2710	Major Vessels and Tanks Stacks
18	FENCE	Fences and Walls	Line	0	Fences and Walls
19	CULT	Cultural	Polygon, Line	3010 3030 3050	Cemetery Outline Rock Outline Pile Outline
21	MISCFEAT	Miscellaneous Features	Line, Point	3240 3250 3260 3270 3280 3290 8910 8920 8930 8950	Miscellaneous Line Miscellaneous Poles Unidentifiable Feature Antenna Air Conditioner Flag Pole Guy Wire Pole Guy Wire Post Mailbox Post
22	MONIT	Monitoring Wells	Point	3400 3410 3420	Monitoring Wells RUST Monitoring Points EIC Temporary Piezometers
23	PIPE	Pipelines	Line, Point	3500 3520 3530	Above Ground Pipelines Columns Supports
24	MISCUT	Miscellaneous Utilities	Line, Point	3600 3610 3620 3630	Cable Marker Lamp Post Manholes Metal Cover



<b>Coverage Number</b>	<b>Coverage Name</b>	<b>Coverage Description</b>	<b>Coverage Type</b>	<b>Feature Code</b>	<b>Feature Description</b>
				3670	Valve
25	ELEC	Electric Utilities	Line, Point	3810	Electric Box
				3830	Meter
				3850	Sub-Station
				3860	Transformers
				3870	Transmission Lines
				3880	Transmission Poles
				3920	Utility Poles
				3940	Street Light
26	GAS	Gas Utilities	Point	4010	Marker
				4060	Valve
28	STEAM	Steam Utilities	Point	4240	Valve
				4250	Vent
30	WATUT	Water Utilities	Point	4400	Fire Hydrants
				4410	Valves
				4420	Meters
				4430	Manholes
31	TOPO	Topography	Line, Point	4500	Index
				4510	Intermediate
				4520	Obscured Index
				4530	Obscured Intermediate
				4540	Depression Index
				4550	Depression Intermediate
				4580	Spot Elevation
35	SAMPPTS	NPDES Sampling Points	Point	6300	NPDES Sampling Locations
40	ANONPROC	Active Non-Process Areas	Polygon, Line	8700	Bulkhead
				8730	Dumpsters

Coverage Number	Coverage Name	Coverage Description	Coverage Type	Feature Code	Feature Description
				8740	Fill Areas
				8750	Ponds
				8760	Pump Station/Silo
				8770	Rubble Piles
				8790	Sumps
				9100	Tanks
				9200	Other Active Non-Process Areas
				9999	Miscellaneous
45	SUMP	Sumps	Point	9400	Sumps
46	OLDRCRA	Former RCRA Units	Polygon, Line	6110	Solid Waste Management Units
				6400	EPA Areas of Concern
47	PARCEL	Parcels	Polygon, Line	9200	Parcels
48	HISTOR	Historical Features	Polygon	8750	Ponds
				8760	Pump Station/Silo
				8770	Rubble Piles
				9100	Tanks
				9201	Process Areas
				9202	Basins
				9203	Impoundments
				9204	Storage Areas
				9205	Loading/Unloading Areas
-	DTM	Digital Terrain Model	Tin	-	-
-	SHEET#.TIF	B & W Orthophoto Sections	Image	-	-
-	N#.TIF	Color Orthophoto Sections	Image	-	-
-	LUBEBLDG.SHP	Structures in Lube Area	Polygon	-	-
-	LUBEHYD.SHP	Surface Hydrology in Lube Area	Polygon	-	-
-	LUBETANK.SHP	Storage Tanks in Lube Area	Polygon	-	-

## **REGIONAL DATABASE**

All **regional** vector coverages have the attributes listed below. Each individual coverage and its additional attributes are listed in the next two sections.

### **All Coverages**

<u>Attribute</u>	<u>Attribute Description</u>
Shape	Point, Polyline, or Polygon
CoverageName#	Internal feature number
CoverageName-id	User-assigned feature number

### **Line Coverages** (in addition to the fields listed above)

<u>Attribute</u>	<u>Attribute Description</u>
Fnode#	Internal number of from-node
Tnode#	Internal number of to-node
Lpoly#	Internal number of polygon to left of arc
Rpoly#	Internal number of polygon to right of arc
Length	Length of arc in coverage units

### **Polygon Coverages** (in addition to the fields listed above)

<u>Attribute</u>	<u>Attribute Description</u>
Area	Area of polygon feature
Perimeter	Perimeter of polygon feature

**Regional Coverages - Descriptions**

<b>Coverage Name</b>	<b>Coverage Description</b>	<b>Coverage Type</b>
BOUNDARY	Pennsylvania State Boundary	Polygon, Line
BPOIL.TIF	USGS Topographic Map	Image
CENTRACTS	Pennsylvania Census Tracts	Polygon
CLIM_DIV	NCDC Climate Divisions for United States (Clipped)	Polygon, Line
CLIMATE	VEMAP Climatological Data for United States (Clipped)	Polygon
COUNTIES	Pennsylvania County Boundaries	Polygon
CPARKS	County Parks in Pennsylvania	Point
DELBASIN	Watershed Boundary for Delaware River	Polygon, Line
DLGBOUND	Marcus Hook Digital Line Graph Coverage of Boundaries	Polygon, Line
DLGHYDRO	Marcus Hook Digital Line Graph Coverage of Hydrography	Polygon, Line
DLGHYPSO	Marcus Hook Digital Line Graph Coverage of Hypsography	Line
DLGPIPE	Marcus Hook Digital Line Graph Coverage of Pipe and Transit Lines	Line
DLGRAIL	Marcus Hook Digital Line Graph Coverage of Railroads	Line
DLGROADS	Marcus Hook Digital Line Graph Coverage of Roads	Line
ECOREGIONS	Aquatic Ecoregions of the Conterminous United States (Clipped)	Polygon
EPA	U.S. EPA Regulated Facilities for Southeast United States (Clipped to Delaware Co.)	Point
EXWATSHD	Boundaries of Exceptional Value Watersheds	Polygon
GEOLOGY	Pennsylvania Surface Geology	Polygon
HYDROLOGIC	1:250,000 Scale Hydrologic Unit of Mid-Atlantic Region (Clipped)	Polygon
LANDUSE#	Four Quadrangles of Land Use in Regions Surrounding the Facility	Polygon
MARCUS_HOOK	Marcus Hook Digital Elevation Model - DEM (ft - NAVD 1929)	Grid, Image
PARKS	State Parks in Pennsylvania	Polygon
PARUNOFF	Runoff grid for Pennsylvania generated using RUNOFF and Spatial Analyst (in inches)	Grid, Image
PHYSREG	Physiographic Regions of Pennsylvania	Polygon, Line
PUBWTSUP	Public Water Supplies	Point

<b>Coverage Name</b>	<b>Coverage Description</b>	<b>Coverage Type</b>
QUADBND	Pennsylvania Quadrangle Boundaries	Polygon
RF1	USEPA River Reach File 1 for Mid-Atlantic Region (Clipped)	Line
RIVERS	Major Rivers in Pennsylvania	Line
ROADS	Major Roads in Pennsylvania	Line
RUNOFF	Average Annual Runoff in the United States, 1951-80 (in)	Line
RV-CHESTER	Streams in Chester County	Line
RV-DELAWARE	Streams in Delaware County	Line
SG-CHESTER	Surface Geology for Chester County	Polygon, Line
SG-DELAWARE	Surface Geology for Delaware County	Polygon, Line
SOILS.SHP	Soil Geographic (STATSGO) Database for Mid-Atlantic Region (Clipped)	Polygon
STRBASIN	WATSTORE Stream Flow Basin Characteristics (Clipped)	Point
STREAMNET	Stations in the USGS's National Stream Quality Accounting Network - NASQAN (Clipped)	Point
SURWTINT	Surface Water Intake	Point
USRUNOFF	Runoff grid for the United States generated using RUNOFF and Spatial Analyst (in inches)	Grid, Image
WATERSHEDS	Major Watersheds in Pennsylvania	Polygon

## Regional Coverages - Attributes

Attributes that are unknown have been left blank. Many of the PASDA coverages, in particular, had metadata files without any attribute information. Many descriptions have also been guessed.

Coverage Name	Attribute	Attribute Description
BOUNDARY	-	-
BPOIL.TIF	-	-
CENTRACTS	Tract	Tract identification number
	Totpers	Total number of persons
	Totfamily	Total number of families
	Tothouseh	Total number of households
	Male	Total number of males
	Female	Total number of females
	Tot_und1	Total number of persons under 1 year of age
	Tot_1and2	Total number of persons 1 to 2 years old
	Tot_3and4	Total number of persons 3 to 4 years old
	Tot_5	Total number of persons 5 years old
	Tot_6	Total number of persons 6 years old
	Tot_7to9	Total number of persons 7 to 9 years old
	Tot_10a11	Total number of persons 10 to 11 years old
	Tot_12a13	Total number of persons 12 to 13 years old
	Tot_14	Total number of persons 14 years old
	Tot_15	Total number of persons 15 years old
	Tot_16	Total number of persons 16 years old
	Tot_17	Total number of persons 17 years old
	Tot_18	Total number of persons 18 years old
	Tot_19	Total number of persons 19 years old
	Tot_20	Total number of persons 20 years old
	Tot_21	Total number of persons 21 years old
	Tot_22t24	Total number of persons 22 to 24 years old
	Tot_25t29	Total number of persons 25 to 29 years old
	Tot_30t34	Total number of persons 30 to 34 years old
	Tot_35t39	Total number of persons 35 to 39 years old
	Tot_40t44	Total number of persons 40 to 44 years old
	Tot_45t49	Total number of persons 45 to 49 years old
	Tot_50t54	Total number of persons 50 to 54 years old
	Tot_55t59	Total number of persons 55 to 59 years old
	Tot_60a61	Total number of persons 60 to 61 years old
	Tot_62t64	Total number of persons 62 to 64 years old
	Tot_65t69	Total number of persons 65 to 69 years old

<b>Coverage Name</b>	<b>Attribute</b>	<b>Attribute Description</b>
	Tot_70t74	Total number of persons 70 to 74 years old
	Tot_75t79	Total number of persons 75 to 79 years old
	Tot_80t84	Total number of persons 80 to 84 years old
	Tot_85ovr	Total number of persons at least 85 years old
	Hh_1pers	Number of households with 1 person
	Hh_2pers	Number of households with 2 persons
	Hh_3pers	Number of households with 3 persons
	Hh_4pers	Number of households with 4 persons
	Hh_5pers	Number of households with 5 persons
	Hh_6pers	Number of households with 6 persons
	Hh_7moper	Number of households with 7 or more persons
	Housunits	Number of house units
	Hu_urb_in	Number of house units in urban areas
	Hu_urb_out	Number of house units out of urban areas
	Hu_r_farm	Number of rural farm house units
	Hu_r_notf	Number of rural house units that are not farms
	Wtr_pub	Number of house units with public water supply
	Wtr_wdril	Number of house units with drilled water supply
	Wtr_wdug	Number of house units with dug water supply
	Wtr_othr	Number of house units with other forms of water supply
	Sew_pub	Number of house units connected to the public sewer system
	Sew_sept	Number of house units with septic tank sewer systems
	Sew_othr	Number of house units with other sewer systems
	Built1989	Number of house units built in 1989 or after
	Built1985	Number of house units built between 1985 and 1988
	Built1980	Number of house units built between 1980 and 1984
	Built1970	Number of house units built between 1970 and 1979
	Built1960	Number of house units built between 1960 and 1969
	Built1950	Number of house units built between 1950 and 1959
	Built1940	Number of house units built between 1940 and 1949
	Built1939	Number of house units built in 1939 or before
	Plumb	Number of house units with plumbing
	Noplumb	Number of house units without plumbing
CLIM_DIV	Cdiv#	Climate division number
	Name	Name of climate division
	St	State abbreviation
	Seq#	Arbitrary sequence number
	St#	State code - not FIPS
	Div#	Climate division within state
	Pre_mean	Mean annual precipitation, 1951-80, inches



<b>Coverage Name</b>	<b>Attribute</b>	<b>Attribute Description</b>
	Pre_std	Standard deviation of annual precipitation
	Tmp_mean	Mean annual temperature, °F
	Tmp_std	Standard deviation of annual temperature
	Pct	Working variable
	Stdiv	Working variable
	Color	Working variable
	Stdiv#	Climate division code
CLIMATE	Fsr	Fractional potential solar radiation - annual average (0-1)
	Irr	Mean daily irradiance - annual average (W/m <sup>2</sup> )
	P	Accumulated precipitation - annual total (mm/year)
	Psr	Potential irradiance - total annual (kJ/m <sup>2</sup> /year)
	Rh	Derived relative humidity - annual average (%)
	Sr	Total incident solar radiation - annual total (kJ/m <sup>2</sup> /year)
	Tm	Mean temperature - annual average (°C)
	Vp	Derived vapor pressure - annual average (mb)
	W	Surface windspeed - annual average (m/s)
	C_atmax	Absolute maximum temperature for the year (°C)
	C_atmin	Absolute minimum temperature for the year (°C)
	C_mtmax	Month of occurrence of maximum temperature (1-12)
	C_mtmin	Month of occurrence of minimum temperature (1-12)
COUNTIES	Mcdcounty	County name (6 letters or less)
	Num	County number identifier
	Name	County name
	State_fips	Federal Information Processing state number (2-digit)
	Cnty_fips	FIPS county number
	Fips	FIPS code (State_fips + Cnty_fips)
	Pop1990	County population in 1990
	Pop90_sqmi	Population in 1990 per square mile
	Households	Number of households
	Males	Number of males
	Females	Number of females
	White	Number of Caucasians
	Black	Number of African-Americans
	Ameri_es	-
	Asian_pi	Number of Asians
	Other	Number of those belonging to another race
	Hispanic	Number of Hispanics
	Age_under5	Number of people under age 5
	Age_5_17	Number of people ages 5 to 17

<b>Coverage Name</b>	<b>Attribute</b>	<b>Attribute Description</b>
	Age_18_64	Number of people ages 18 to 64
	Age_65_up	Number of people at least age 65
	Nevermarry	Number of people that have never been married
	Married	Number of people that are married
	Separated	Number of people that are separated
	Widowed	Number of people that are widowed
	Divorced	Number of people that are divorced
	Hsehd_1_m	Number of households with 1 male
	Hsehd_1_f	Number of households with 1 female
	Marhh_chd	Number of married households with children
	Marhh_no_c	Number of married households with no children
	Marhh_child	Number of married households with a single child
	Fhh_child	-
	Hse_units	Number of house units
	Vacant	Number of vacant house units
	Owner_occ	Number of owner occupied house units
	Renter_occ	Number of renter occupied house units
	Median_val	Median value of house units in \$
	Medianrent	Median monthly rent for rented house units in \$
	Units_1det	-
	Units_1att	-
	Units2	-
	Units3_9	-
	Units10_49	-
	Units50_up	-
	Mobilehome	Number of mobile homes
	No_farms87	Number of farms in 1987
	Avg_size87	Average farm size in 1987 (acres)
	Crop_acr87	Number of acres with crop in 1987
	Avg_sale87	Average farm sale in 1987 (\$)
CPARKS	Unique-id	Unique 10-digit identifier
	Name17	County park name
	Type17	-
DELBASIN	Sqm	Polygon area in square miles
	Wrds#	PA Dept. of Natural Resources stream code number
	Huc	USGS Hydrologic Unit Code (HUC) number
DLGBOUND	-	-
DLGHYDRO	-	-
DLGHYPSO	-	-
DLGPIPE	-	-
DLGRAIL	-	-

<b>Coverage Name</b>	<b>Attribute</b>	<b>Attribute Description</b>
DLGROADS	-	-
ECOREGIONS	Eco	Full ecoregion code
	Pbmin1	State code item used by EPA
	Lwcode	Land/water code (= L for land, = W for water, = ZZ for missing)
	Color	Working item for plotting
	Ecoregion	Ecoregion code (1-76)
	Typical	Indicates strength of association in ecoregion (= 0 for most typical, = 1 for generally typical)
	Fipsst	State code
EPA	Ldip_code	Source of record
	Id	Unique id from respective program system
	Mad_id	Assigned sequential reference number
	Loc_ref_id	Assigned sequential reference number
	Fac_id	EPA Facility Index System (FINDS) identifier
	Facility_n	Name of the facility or site
	Latitude	Latitude of the facility, site, or operable unit
	Longitude	Longitude of the facility, site, or operable unit
	Bnd_flag	Boundary flag (Y indicates an NPL site exists for this id)
	Npl_stat_i	-
	Y_coord	The National Albers meters Y coordinate based on NAD 83 datum
	X_coord	The National Albers meters X coordinate based on NAD 83 datum
	Albers_src	Source for Albers coordinate
	Bvflag	Indicator of most accurate location for an EPA facility as defined by FINDS
	System_id	Unique id from respective program system
	Cer_fac_id	Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS) unique ID name
	Handler_id	Resource Conservation and Recovery Information System (RCRIS) unique ID name
	Tri_facili	Toxic Release Inventory System (TRIS) unique ID name
	Ef_afs_id	Aerometric Information Retrieval System (AIRS) Facility Subsystem (AFS) unique ID name
	Npdes	Permit Compliance System (PCS) unique ID name
	Ern_id	-
	Ncdb_id	-
	Ffis_id	-
Pads_id	-	

<b>Coverage Name</b>	<b>Attribute</b>	<b>Attribute Description</b>
	Site_id	Unique identifier for Envirofacts point coverage
EXWATSHD	Unique-id	Unique 10-digit identifier
	Name02	Exceptional watershed name
GEOLOGY	Fm	Abbreviation for geologic unit
	Name	Name of the geologic unit
	Aqmed	-
	Vadose	-
	Hycond	-
HYDROLOGIC	Huc	USGS Hydrologic Unit Code (HUC) number
	Region	-
	Subregion	-
	Acctunit	-
	Hydrounit	-
LANDUSE#	Lucode	Anderson Land Use Code number
	Descrip	Anderson Land Use category description
MARCUS_HOOK	Value	Elevation in feet above mean sea level (NAVD 88)
PARKS	Unique-id	Unique 10-digit identifier
	Name05	State park name
	Type05	-
	Cal-acre	-
	Legal-acre	-
PARUNOFF	Value	Average annual runoff (inches)
PHYSREG	Regname	Physiographic region name
PUBWTSUP	Name32	Public water supply name
	Type32	-
QUADBND	Name	Quadrangle boundary name
RF1	Rr	Reach file ID - unique identifier for each river reach
	Huc8	Hydrologic cataloging unit code
	Huc6	Hydrologic accounting unit code
	Huc4	Hydrologic subregion code
	Huc2	Hydrologic region code
	Cu	Hydrologic cataloging unit code
	Seq	Reach segment number within CU
	Mi	Mile point within SEG
	Cuseq	Combination of CU and SEG items of RR
RIVERS	-	-
ROADS	-	-
RUNOFF	Inches	Average annual runoff (inches)
RV-CHESTER	Igds-layer	-
	Igds-type	-
	Igds-level	-

<b>Coverage Name</b>	<b>Attribute</b>	<b>Attribute Description</b>
	Igds-ggno	-
	Igds-class	-
	Igds-props	-
	Igds-color	-
	Igds-style	-
	Igds-weight	-
	Igds-text	-
	Igds-font	-
	Igds-entity1	-
	Igds-occur1	-
	Igds-entity2	-
	Igds-occur2	-
	Igds-cpxid	-
	Igds-cpxtype	-
	Igds-offset	-
RV-DELAWARE	Same as RV-CHESTER	
SG-CHESTER	Same as GEOLOGY	
SG-DELAWARE	Same as GEOLOGY	
SOILS.SHP	Muid	Mapunit identification code - used to reference observations
	State	2-character state abbreviation
	Awc	Available water capacity (inches per inch)
	Clay	Percent clay in soil (percent of material less than 2 mm in size)
	Kffact	Actual k factor used in water erosion component of universal soil loss equation
	Om	Organic material in soil (in percent by weight)
	Perm	Permeability of soil (in inches per hour)
	Thick	Cumulative thickness of all soil layers (in inches)
	Hygrp	Hydrologic characteristics of soil ( 1 = high infiltration, 2 = mod infil, 3 = slow infil, 4 = very slow infil)
	Drain	Soil drainage ( 1 = excess, 2 = mod excess, 3 = well, 4 = mod well, 5 = mod poor, 6 = poor, 7 = v poor)
	Slope	Surface slope (in percent)
	Ll	Liquid limit of soil (in percent moisture by weight)
	Ifhydic	Share of map unit components with hydic soils
	Afldfreq	Annual flood frequency ( 1 = greater than 50 pct., 2 = 5 to 50 pct., 3 = 0 to 5 pct., 4 = none)
STRBASIN	Station_no	8-digit station identification number
	Station_na	Station name
	Fipsst	State code

<b>Coverage</b>	<b>Attribute</b>	<b>Attribute</b>
<b>Name</b>	<b>Attribute</b>	<b>Description</b>
	Lat_dd_rep	Reported latitude (decimal degrees)
	Lon_dd_rep	Reported longitude (decimal degrees)
	Drainage_a	Drainage area (square miles)
	Elevation	Elevation of station in feet above mean sea level
	Forest_are	Forested area of basin
	Precip	Mean annual precipitation over basin (inches)
	Flow_peak2	2-year peak flow (cfs)
	Years_peak	Years of record for Flow_peak2
	Flow_avg	Average flow (cfs)
	Years_avg	Years of record for Flow_avg
	Flow_m7_10	7-day, 10-year low flow (cfs)
	Years_m7_1	Years of record for Flow_m7_10
<b>STREAMNET</b>	Station	3-digit sequence number
	Station_id	NASQAN station identifier number
	Name	NASQAN station name
	Huc6	Water resources accounting unit
	Drain	Drainage area (square miles)
	Lat	Latitude
	Long	Longitude
	Status	Status of station ( A = active, I = inactive, D = discontinued)
	Sfr	Streamflow records ( G = includes streamflow records)
	Freq	Frequency of sampling ( B = bimonthly, Q = quarterly, D = bimonthly but drop 1)
	Bdate	Date station became active
	Edate	Date station was discontinued
	Huc2	Water resources region
<b>SURWTINT</b>	Type33	-
	Name33	Surface water intake name
<b>WATERSHEDS</b>	Split	-
	Wshed	Watershed number
	Name	Watershed name

## **DERIVED DATABASE**

**Facility-Specific Access Generated Coverages**

All these coverages were created by importing data into an ArcView table and using the Add Event Theme function. All coverages were then converted to shapefiles. When the facility.apr file is opened, some coverages are created directly from information in the tabular database while some coverages already exist as shapefiles (e.g., the contours). The shapefile coverages do not update when the project file is opened.

<b>Coverage Name</b>	<b>Feature Description</b>	<b>Coverage Type</b>
BASINS	Sludge basin sampling locations	Point
BENGW8010	1996 benzene concentrations in groundwater (8010/8020 analysis method)	Point
BENGW8240	1996 benzene concentrations in groundwater (8240 analysis method)	Point
BENSOIL	1996 benzene concentrations in soil	Point
BENSOIL0-2	1996 benzene concentrations in soil 0 - 2 ft. bgs	Point
BENSOIL2-8	1996 benzene concentrations in soil 2 - 8 ft. bgs	Point
BENSOIL8UP	1996 benzene concentrations in soil > 8 ft. bgs	Point
CTOUR1	Contours of BENGW8240 (1000 ug/L interval)	Line
CTOUR2	Contours of BENGW8010 (10000 ug/L interval)	Line
CTOUR4	Contours of ETHGW8240 (100 ug/L interval)	Line
CTOUR5	Contours of ETHGW8010 (1000 ug/L interval)	Line
CTOUR6	Contours of MTBEGW8020A (10 ug/L interval)	Line
CTOUR8	Contours of NAPGW8270 (10 ug/L interval)	Line
CTOUR9	Contours of NAPGW8010 (1000 ug/L interval)	Line
CTOUR10	Contours of TCEGW8240 (10 ug/L interval)	Line
CTOUR11	Contours of TOLGW8240 (100 ug/L interval)	Line
CTOUR12	Contours of TOLGW8010 (1000 ug/L interval)	Line
CTOUR13	Contours of XYL8240 (1000 ug/L interval)	Line
CTOUR14	Contours of XYL8240B (100 ug/L interval)	Line
CTOUR15	Contours of GWELEVQ1 (1 ft interval)	Line

198



<b>Coverage Name</b>	<b>Feature Description</b>	<b>Coverage Type</b>
CTOUR16	Contours of GWELEVQ2 (1 ft interval)	Line
CTOUR17	Contours of GWELEVQ3 (1 ft interval)	Line
CTOUR18	Contours of GWELEVQ4 (1 ft interval)	Line
CTOUR19	Contours of LUBEGWQ1 (0.5 ft interval)	Line
CTOUR20	Contours of LUBEGWQ2 (0.5 ft interval)	Line
CTOUR21	Contours of LUBEGWQ3 (0.5 ft interval)	Line
CTOUR22	Contours of LUBEGWQ4 (0.5 ft interval)	Line
ETHGW8010	1996 ethylbenzene concentrations in groundwater (8010/8020 analysis method)	Point
ETHGW8240	1996 ethylbenzene concentrations in groundwater (8240 analysis method)	Point
ETHSOIL	1996 ethylbenzene concentrations in soil	Point
ETHSOIL0-2	1996 ethylbenzene concentrations in soil 0 - 2 ft. bgs	Point
ETHSOIL2-8	1996 ethylbenzene concentrations in soil 2 - 8 ft. bgs	Point
ETHSOIL8UP	1996 ethylbenzene concentrations in soil > 8 ft. bgs	Point
GEOPROBES	Geoprobe sampling locations	Point
GWELEVQ1	Site wide groundwater elevation and lnapl thickness data for 10/13/97 in feet - NAVD 1988	Point
GWELEVQ2	Site wide groundwater elevation and lnapl thickness data for 1/5/98 in feet - NAVD 1988	Point
GWELEVQ3	Site wide groundwater elevation and lnapl thickness data for 4/27/98 in feet - NAVD 1988	Point
GWELEVQ4	Site wide groundwater elevation and lnapl thickness data for 8/17/98 in feet - NAVD 1988	Point
GWMCL	All groundwater measurements above an MCL <sup>1</sup>	Point
LNAPLQ1	Lube area groundwater elevation and lnapl thickness data for 10/13/97 in feet - NAVD 1988	Point
LNAPLQ2	Lube area groundwater elevation and lnapl thickness data for 1/5/98 in feet - NAVD 1988	Point
LNAPLQ3	Lube area groundwater elevation and lnapl thickness data for 4/27/98 in feet - NAVD 1988	Point
LNAPLQ4	Lube area groundwater elevation and lnapl thickness data for 8/17/98 in feet - NAVD 1988	Point
LOCATIONS	All sampling locations	Point
LUBEGWQ1	Lube area groundwater elevation and lnapl thickness data for 10/13/97 in feet - NAVD 1988	Point
LUBEGWQ2	Lube area groundwater elevation and lnapl thickness data for 1/5/98 in feet - NAVD 1988	Point
LUBEGWQ3	Lube area groundwater elevation and lnapl thickness data for 4/27/98 in feet - NAVD 1988	Point

<b>Coverage Name</b>	<b>Feature Description</b>	<b>Coverage Type</b>
LUBEGWQ4	Lube area groundwater elevation and lnapl thickness data for 8/17/98 in feet - NAVD 1988	Point
MAXCOCGW	Maximum groundwater concentration for each detected chemical	Point
MAXCOCSOIL	Maximum soil concentration for each detected chemical	Point
MTBEGW8020A	1996 mtbe concentrations in groundwater (8020A analysis method)	Point
NAPGW8010	1996 naphthalene concentrations in groundwater (8010/8020 analysis method)	Point
NAPGW8270	1996 naphthalene concentrations in groundwater (8270 analysis method)	Point
NAPSOIL	1996 naphthalene concentrations in soil	Point
NAPSOIL0-2	1996 naphthalene concentrations in soil 0 - 2 ft. bgs	Point
NAPSOIL2-8	1996 naphthalene concentrations in soil 2 - 8 ft. bgs	Point
NAPSOIL8UP	1996 naphthalene concentrations in soil > 8 ft. bgs	Point
PIEZOMETERS	Piezometer sampling locations	Point
SEDIMENTS	Sediment sampling locations	Point
SOILBORINGS	Soil boring sampling locations	Point
SOILVAPORS	Soil vapor sampling locations	Point
SUMPS	Sump sampling locations	Point
TCEGW8240	1996 tce concentrations in groundwater (8240 analysis method)	Point
TESTPITS	Test pit/trench sampling locations	Point
TOLGW8010	1996 toluene concentrations in groundwater (8010/8020 analysis method)	Point
TOLGW8240	1996 toluene concentrations in groundwater (8240 analysis method)	Point
TOLSOIL	1996 toluene concentrations in soil	Point
TOLSOIL0-2	1996 toluene concentrations in soil 0 - 2 ft. bgs	Point
TOLSOIL2-8	1996 toluene concentrations in soil 2 - 8 ft. bgs	Point
TOLSOIL8UP	1996 toluene concentrations in soil > 8 ft. bgs	Point
WELLS	Well sampling locations	Point
XYLGW8240	1996 total xylenes concentrations in groundwater (8240 analysis method)	Point
XYLGW8240B	1996 total xylenes concentrations in groundwater (8240B analysis method)	Point
XYLSOIL	1996 total xylenes concentrations in soil	Point

<b>Coverage Name</b>	<b>Feature Description</b>	<b>Coverage Type</b>
XYLSOIL0-2	1996 total xylenes concentrations in soil 0 - 2 ft. bgs	Point
XYLSOIL2-8	1996 total xylenes concentrations in soil 2 - 8 ft. bgs	Point
XYLSOIL8UP	1996 total xylenes concentrations in soil > 8 ft. bgs	Point

<sup>1</sup> PaDEP Act II msc, non-use aquifer, non-residential

Note: All concentration values are in units of µg/L for water and µg/kg for soil unless otherwise noted.

## Generated Grids

These grids were generated using either Spatial Analyst and the associated shapefile listed or by using the Watershed Delineator extension.

<b>Coverage Name</b>	<b>Feature Description</b>	<b>Associated Legend</b>
BENSURF1	Surface from BENGW8240	Bengw.avl
BENSURF2	Surface from BENGW8010	Bengw.avl
BUILDGRID	Structures Grid	-
BURNDEM	NEWDEM2 with Surface Hydrology Burned In	Burndem1.avl
BURNDEM2	Marcus Hook DEM with Streams and Rivers Burned In	Burndem2.avl
DELGRID	Delaware River Grid	-
DEMFAC	Facility Flow Accumulation	Flowacc1.avl
DEMFAC1	Regional Flow Accumulation	Flowacc2.avl
DEMFDR	Facility Flow Direction	Flowdir1.avl
DEMFDR1	Regional Flow Direction	Flowdir2.avl
DEMFIL	Filled DEM	Filldem1.avl
DEMFIL1	Filled DEM	Filldem2.avl
DRAINAGE	Facility Drainage Areas	Drainage.avl
ETHSURF1	Surface from ETHGW8240	Ethgw.avl
ETHSURF2	Surface from ETHGW8010	Ethgw.avl
GWELEVQ1	Surface from GWELEVQ1	Gwelev.avl
GWELEVQ2	Surface from GWELEVQ2	Gwelev.avl
GWELEVQ3	Surface from GWELEVQ3	Gwelev.avl
GWELEVQ4	Surface from GWELEVQ4	Gwelev.avl
LUBEGWQ1	Surface from LUBEGWQ1	Lubegwelev.avl
LUBEGWQ2	Surface from LUBEGWQ2	Lubegwelev.avl
LUBEGWQ3	Surface from LUBEGWQ3	Lubegwelev.avl
LUBEGWQ4	Surface from LUBEGWQ4	Lubegwelev.avl
MTBESURF	Surface from MTBEGW8020A	Mtbegw.avl
NAPSURF1	Surface from NAPGW8270	Napgw.avl
NAPSURF2	Surface from NAPGW8010	Napgw.avl
NEWDEM2	Revised DEM with Tanks and Structures Burned In	-
ORIGDEM	DEM Converted from DTM	Origdem.avl
REGDEM	Marcus Hook DEM	-
REGDEM1	Marcus Hook DEM Clipped to PA State Boundary Line	Regdem.avl
REVDDEM	Revised DEM to Include Oily-Sewer Drains	Revdem.avl
STR1000	Facility Delineated Streams with a Threshold of 1000	Str1000f.avl
STR1000	Regional Delineated Streams with a Threshold of 1000	Str1000r.avl
STR10000	Facility Delineated Streams with a Threshold of 10000	Str10000.avl
STR300	Regional Delineated Streams with a Threshold of 300	Str300.avl
STR500	Regional Delineated Streams with a Threshold of 500	Str500.avl

<b>Coverage Name</b>	<b>Feature Description</b>	<b>Associated Legend</b>
STR5000	Facility Delineated Streams with a Threshold of 5000	Str5000.avl
STREAMS	Regional Stream Grid	-
STRLILN500	Stream Link Grid for Streams with a Threshold of 500	Strlin500.avl
TANKGRID	Storage Tank Grid	-
TCESURF	Surface from TCEGW8240	Tcegw.avl
TOLSURF1	Surface from TOLGW8240	Tolgw.avl
TOLSURF2	Surface from TOLGW8010	Tolgw.avl
WATERGRID	Surface Hydrology Grid	-
WATERSH1	Regional Watershed Delineation	Watersh.avl
XYLSURF1	Surface from XYLGW8240	Xylgw.avl
XYLSURF2	Surface from XYLGW8240B	Xylgw.avl

Note: All concentration grids are in units of µg/L and all elevation grids have units of feet - NAVD 88

## Associated Legends

<b>Legend Name</b>	<b>Description</b>	<b>Associated Coverage(s)</b>
Bengw.avl	Benzene groundwater concentrations	BENSURF#
Bensoil.avl	Benzene soil concentrations	BENSOIL
Burndem1.avl	Facility burned dem elevations	BURNDEM
Burndem2.avl	Regional burned dem elevations	BURNDEM2
Drainage.avl	Facility drainage areas	DRAINAGE
Ethgw.avl	Ethylbenzene groundwater concentrations	ETHSURF#
Ethsoil.avl	Ethylbenzene soil concentrations	ETHSOIL
Filldem1.avl	Facility filled dem elevations	DEMFIL
Filldem2.avl	Regional filled dem elevations	DEMFIL1
Flowacc1.avl	Facility flow accumulation	DEMFACTOR
Flowacc2.avl	Regional flow accumulation	DEMFACTOR1
Flowdir1.avl	Facility flow direction	DEMFDIR
Flowdir2.avl	Regional flow direction	DEMFDIR1
Gwelev.avl	Site wide groundwater elevations	GWELEVQ#
Historical.avl	Historical feature types	HISTOR
Landuse.avl	Land use descriptions	LANDUSE#
Lnapl.avl	LNAPL thickness ranges	LNAPLQ#
Lubegwelev.avl	Lube area groundwater elevations	LUBEGWQ#
Maxconc.avl	Maximum concentrations by chemical class	MAXCOCGW, MAXCOCSOIL
Mtbgw.avl	MTBE groundwater concentrations	MTBESURF
Napgw.avl	Naphthalene groundwater concentrations	NAPSURF#
Napsoil.avl	Naphthalene soil concentrations	NAPSOIL
Origdem.avl	Facility dem elevations	ORIGDEM
Parunoff.avl	Annual runoff values in inches	PARUNOFF
Regdem.avl	Regional dem elevations	REGDEM1
Revdem.avl	Facility revised dem elevations	REVDDEM
Str10000.avl	Facility streams with a threshold of 10000 cells	STR10000
Str1000f.avl	Facility streams with a threshold of 1000 cells	STR1000
Str1000r.avl	Regional streams with a threshold of 1000 cells	STR1000
Str300.avl	Regional streams with a threshold of 300 cells	STR300
Str500.avl	Regional streams with a threshold of 500 cells	STR500
Str5000.avl	Facility streams with a threshold of 5000 cells	STR5000
Strlin500.avl	Regional stream links	STRLINL500
Structure.avl	Structure types	BLDG
Surfcover.avl	Surface cover types	SURFCOV
Tcegw.avl	TCE groundwater concentrations	TCESURF
Tolgw.avl	Toluene groundwater concentrations	TOLSURF#
Tolsoil.avl	Toluene soil concentrations	TOLSOIL

<b>Legend Name</b>	<b>Description</b>	<b>Associated Coverage(s)</b>
Topo.avl	Site wide elevations	TOPO
Urunoff.avl	Annual runoff values in inches	USRUNOFF
Watersh.avl	Regional watersheds	WATERSHI
Xylgw.avl	Xylene groundwater concentrations	XYLSURF#
Xylsoil.avl	Xylene soil concentrations	XYLSOIL

Facility scale coverages

Regional scale coverages

Access generated coverages

Spatial Analyst generated grids

Watershed Delineator generated grids

**APPENDIX C**  
**TABULAR DATABASE DICTIONARY**



**Tables:**

<b>Table Name</b>	<b>Field Name</b>	<b>Description</b>
<b>Action_Levels</b>	CAS	Chemical abstract code for the parameter analyzed
	COMMENTS	Comments
	FILE_NAME	Name of the original data file (if applicable)
	LEVELS	Chemical concentration level
	LOAD_DATE	Date this record was loaded or entered into the table
	<b>MATRIX</b>	<b>Sample matrix</b>
	<b>SOURCE</b>	<b>Source of information</b>
	SOURCE_REF	Specific literature reference for identifying source
	<b>TYPE</b>	<b>Broad category of type of activity being monitored</b>
	<b>UNIT</b>	<b>Unit of measurement the final result is reported in</b>
<b>CAS_List</b>	CAS	Chemical abstract code for the parameter analyzed
	<b>CHEM_CLASS</b>	<b>Class given to analyte based on method</b>
	PARAM	Chemical name, physical test performed, or classification for a record
<b>Chem_Properties</b>	CAS	Chemical abstract code for the parameter analyzed
	DAIR	Diffusion Coefficient in air (cm <sup>2</sup> /s)
	DRC	Decay Rate Constant (half-life), day <sup>-1</sup> (days)
	DWAT	Diffusion Coefficient in water (cm <sup>2</sup> /s)
	H	Henry's Law Constant (atm-m <sup>3</sup> /mol)
	HDIM	Dimensionless Henry's Constant
	KOC	Soil-water partition coefficient (L/kg)
	KOW	Octanol-water partitioning coefficient (L/kg)
	MW	Molecular Weight (g/mol)
	PV	Vapor Pressure (mmHg sat)
	RFDD	Dermal Reference Dose (mg/kg-day)

<b>Table Name</b>	<b>Field Name</b>	<b>Description</b>
	RFDI	Inhalation Reference Dose (mg/kg-day)
	RFDO	Oral Reference Dose (mg/kg-day)
	SFD	Dermal Slope Factor (kg-day/mg)
	SFI	Inhalation Slope Factor (kg-day/mg)
	SFO	Oral Slope Factor (kg-day/mg)
	SOL	Solubility @ 20° - 25° in mg/L
<b>Grainsize</b>	<b>ANAL_METH</b>	<b>Sample analysis method</b>
	LOAD_DATE	Date this record was loaded or entered into the table
	PER_FINER	Percent of sample finer than this mesh
	SAMP_ID	Unique sample indentifier
	<b>SAMP_DESC</b>	<b>Description of sample composition</b>
	SIEVE_MM	Standard sieve mesh size in mm
	<b>UNIT</b>	<b>Unit of measure for the sample weights</b>
<b>Groundwater_Levels</b>	COMMENTS	Comments
	COR_GWLEVEL	Corrected groundwater level for the presence of a NAPL
	EVENT	Month year and qualifier
	FILE_NAME	Name of the original data file (if applicable)
	FREE_PROD	Free product present in well on date of measurement?
	GW_LEVEL	Depth to groundwater on this date from top of well casing in feet
	GWL_DATE	Date the depth measurement was taken
	IMM_BOT	Depth to bottom of immiscible layer in feet below ground surface (bgs)
	IMM_DEPTH	Depth to immiscible liquid in feet bgs
	IMM_THICK	Thickness of immiscible layer in feet
	LOAD_BY	Initials of person who loaded this record into the database
	LOAD_DATE	Date this record was loaded or entered into the database
	LOC_ID	Location identifier which is unique for this site

<b>Table Name</b>	<b>Field Name</b>	<b>Description</b>
	STUDY_ID	Identification code given to an investigation
<b>Hydraulic</b>	COMMENTS	Comments
	COND_SAT	Hydraulic conductivity of unit when saturated
	CS_UNITS	Saturated conductivity units
	LOAD_DATE	Date this record was loaded or entered into the table
	LOC_ID	Location identifier which is unique for this site
	MACRO	Macro code associated with the top and bottom macro depths (i.e., stratigraphic uni
	METHOD	Testing method used
	STUDY_ID	Identification code given to an investigation
<b>Location</b>	AREA	Area of the site in which the location resides
	CO_SOURCE	Source for easting and northing coordinates
	EASTING	Easting in Pennsylvania State Plane Projection (South Zone)
	EL_SOURCE	Source for location elevation
	FILE_NAME	Name of the original data file (if applicable)
	LOAD_BY	Initials of person who loaded this record into the table originally
	LOAD_DATE	Date this record was loaded or entered into the table
	LOC_ELEV	Location elevation above sea level in feet (NAVD 1988 datum)
	LOC_ID	Location identifier which is unique to this site
	LOC_TYPE	Location type
	NORTHING	Northing in Pennsylvania State Plane Projection (South Zone)
	OLD_LOC_ELEV	Location elevation above sea level in feet (plant grid datum)
	OLD_LOC_ID	Original loc ids assigned during past studies
	OLD_LOC_TYPE	Original location type assigned during past studies
	PLANT_EAS	Easting based on plant grid reference
	PLANT_NOR	Northing based on plant grid reference
PLCO_SOURCE	Source for plant grid coordinates	

<b>Table Name</b>	<b>Field Name</b>	<b>Description</b>
	<b>PLEL_SOURCE</b>	<b>Source for plant grid elevation</b>
<b>Results</b>	<b>ANAL_METH</b>	<b>Laboratory analysis method number</b>
	CAS	Chemical abstract code for the parameter analyzed
	COMMENTS	Comments
	DET_LIMIT	Detection limit for this record's method and parameter
	DIL_FACT	Dilution factor
	LOAD_BY	Initials of person who loaded this record into the table
	LOAD_DATE	Date this record was loaded or entered into the table
	QC_FLAG	Y for qc samples
	REP_HIT	Report criteria flag - set by reprint scripts, y if record qualifies as a hit
	SAMP_ID	Unique sample identifier
	<b>UNIT</b>	<b>Unit of measurement the final result is reported in</b>
	<b>VALID_QUAL</b>	<b>Validation codes assigned to the sample from data validation process</b>
	VALUE	Final value reported from analysis
	<b>VALUE_QUAL</b>	<b>Qualifier assigned by the laboratory to the value for this record</b>
<b>Sample_Analysis</b>	<b>ANAL_METH</b>	<b>Laboratory analysis method number</b>
	FILTERED	Was the sample filtered (y/n)? (aqueous samples)
	LOAD_DATE	Date this record was loaded or entered into the table
	SAMP_ID	Unique sample identifier
	SDG	Sample delivery group (i.e. laboratory batch)
<b>Sample_Collection</b>	COMMENTS	Comments
	DEPTH_BOT	Depth of sample - bottom of interval
	DEPTH_TOP	Depth of sample - top of interval
	DEPTH_UNIT	Unit of measure for depth of sample
	LOAD_BY	Initials of person who loaded this record into the table
	LOAD_DATE	Date this record was loaded or entered into the table

<b>Table Name</b>	<b>Field Name</b>	<b>Description</b>
	LOC_ID	Location identifier which is unique for this site
	MATRIX	Sample matrix
	QC_CODE	Laboratory designation for qa/qc samples
	SAMP_DATE	Date sample collected
	SAMP_ID	Unique sample identifier
	STUDY_ID	Identification code given to an investigation
<b>Stratigraphy</b>	BOT_MACRO	Bottom of the macro unit in feet below ground surface (bgs)
	COMMENTS	Comments
	LENGTH	Thickness of this particular stratigraphic unit
	LOC_ID	Location identifier which is unique for this site
	MACRO	Macro code associated with the top and bottom macro depths
	TEXT_DESC	Text description corresponding to the stratigraphic unit defined by the micro code
	TOP_MACRO	Top of the macro unit in feet bgs
	TOTAL_D	Total depth of the boring/well in feet
<b>Study</b>	AREA	Area of the site to which the report applies
	KEYWORDS	Key words to search by that originate from report topics
	ORIGINAL_DATA	Is there original data in the report (Y or N)?
	QUAL	Quality rating, or usability scale, for the report
	REP_AUTHOR	Author of report
	REP_DATE	Date of report
	STUDY_ID	Identification code given to an investigation
	STUDY_NAME	Report name
	STUDY_TYPE	Type of study
	SUBMIT_AGENCY	Agency to which the report was submitted
	SUBMIT_DATE	Submittal date
<b>Well</b>	BSC	Depth to the bottom of the screen, in feet below ground surface (bgs)

<b>Table Name</b>	<b>Field Name</b>	<b>Description</b>
	CASING_DIA	Diameter of the protective casing, in inches
	<b>CASING_TYP</b>	<b>Type of protective casing</b>
	COMMENTS	Comments
	END_DATE	Date that the boring and/or well drilling ended
	GROUT_LEN	Grout length, in feet
	LOAD_BY	Initials of person who loaded this record into the table
	LOAD_DATE	Date this record was loaded or entered into the table
	LOC_ID	Location identifier which is unique for this site
	OLD_TPC_ELEV	Top of protective casing elevation, in feet (plant grid datum)
	OLD_TWC_ELEV	Top of the well riser (a.k.a. TR) elevation, in feet (plant grid datum)
	RISER_LEN	Length of the riser, in feet
	SCREEN_LEN	Length of the screen, in feet
	<b>SCREEN_TYP</b>	<b>Type of screen slot</b>
	SEAL_LEN	Seal length, in feet
	SPACK_LEN	Sandpack length, in feet
	STUDY_ID	Identification code given to an investigation
	TBS	Depth to the top of the bentonite seal, in feet bgs
	TD	Total depth of the well or boring, in feet
	TG	Depth to the top of the grout, in feet bgs
	TPC	Height of the protective casing, in feet above the ground surface
	TPC_ELEV	Top of protective casing elevation, in feet (NAVD 1998 datum)
	TSC	Depth to the top of the screen, in feet bgs
	TSP	Depth to the top of the sandpack, in feet bgs
	TWC	Height of the well riser, in feet above the ground surface
	TWC_ELEV	Top of the well riser (a.k.a. TR) elevation, in feet (NAVD 1988 datum)

Note: fields in **red** have a specific set of criteria that can be viewed in the Fields spreadsheet.

**Fields:**

<b>Field</b>	<b>Table(s)</b>	<b>Entries</b>
ANAL_METH	Action_Levels	1 Solids Method
	Grainsize	10 Gross Analysis
	Results	1010 1310 1311 1311/6010 1311/6010A 1311/7041 1311/7060 1311/7420 1311/7470 1311/7470A 1311/7471 1311/7740 1311/8080A 1311/8150B 1311/8240 1311/824060A 1311/8260A 1311/8270 1311/8270B 1311/9045C 1312/6010A 150.1 160.3 17 Bacteria 17 Bacteria Method 17 BOD 17 BOD Method 17 BTEX 17 BTEX Method 17 Carbonate 17 Carbonate Method

Field	Table(s)	Entries
		17 CO2
		17 CO2 Method
		17 COD
		17 COD Method
		17 Diss Metals
		17 Dissolved Metals Method
		17 DO
		17 DO Method
		17 Ethane
		17 Ethane Method
		17 Field pH Method
		17 Field Specific Conductance Method
		17 Fid pH
		17 Fid Spec Cond
		17 Herbicide
		17 Herbicide Method
		17 Metals
		17 Metals Method
		17 Methane
		17 Methane Method
		17 Nitrite
		17 Nitrite Method
		17 Oil & Grease
		17 Oil & Grease Method
		17 Pesticide
		17 Pesticide Method
		17 Sulfide
		17 Sulfide Method
		17 SYOCs
		17 SYOCs Method
		17 TPH Method
		17 VOCs
		17 VOCs Method
		2 BTEX
		2 BTEX Method



Field	Table(s)	Entries
		2 Spec Gravity
		2 Specific Gravity Method
		2 SVOCs
		2 SVOCs Method
		2 Tension
		2 Tension Method
		2 viscosity
		2 Viscosity Method
		218.2
		236.1
		239.2
		24 pH
		24 pH Method
		243.1
		2540G
		2930/2988
		3 Arsenic
		3 Arsenic Method
		3 Base/Neutral Organics Method
		3 BN Orgs
		3 BTEX
		3 BTEX Method
		3 Cadmium
		3 Cadmium Method
		3 Chromium
		3 Chromium Method
		3 Conductivity Method
		3 Corrosivity
		3 Corrosivity Method
		3 Cr(VI)
		3 Cr(VI) Method
		3 EP TOX Metals
		3 EP Tox Metals Method
		3 Ignitability
		3 Ignitability Method

Field	Table(s)	Entries
		3 Lead
		3 Lead Method
		3 Mercury
		3 Mercury Method
		3 Metals
		3 Metals Method
		3 Nitrogen
		3 Nitrogen Method
		3 Oil & Grease
		3 Oil & Grease Method
		3 Oxygen Method
		3 PCB
		3 PCB Method
		3 pH
		3 pH Method
		3 Phenol Method
		3 Phosphorus
		3 Phosphorus Method
		3 Reactivity
		3 Reactivity Method
		3 Solids
		3 Solids Method
		3 TCLP Metals
		3 TCLP Metals Method
		3 TCLP SVOAs Acid Extr
		3 TCLP SVOAs Acid Extractable Method
		3 TCLP SVOAs Base/Neutral Extractable Method
		3 TCLP SVOAs BN Extr
		3 TCLP SVOAs
		3 TCLP SVOAs Method
		3 Temperature Method
		3 TPH
		3 TPH Method
		3 Turbidity Method
		3 SVOAs

Field	Table(s)	Entries
		3 VOAs Method
		3 Water
		3 Water Method
		3005, 3010, 3020/6010
		3005, 3010, 3020/7060
		3005, 3010, 3020/7470
		3005, 3010, 3020/7740
		3005/3010/3020/6010
		3005/3010/3020/7060
		3005/3010/3020/7470
		3005/3010/3020/7740
		3050/6010
		3050/7520
		325, 3
		335, 2
		340, 2
		350.2 Mod.
		350.2M
		351.3 Mod.
		351.3M
		353, 3
		353.3 Mod.
		353.3M
		3550/8100
		3550/8100 Mod.
		3550/8100M
		3550A / 5030A / 8015 Mod.
		3550A / 5030A / 8015M
		360, 1
		365, 2
		365.2 Mod.
		365.2M
		365.3M
		375, 4
		375.4M

Field	Table(s)	Entries
		376.2
		376.2M
		405.1
		410.1
		410.4
		415.1
		418.1
		4500 B M
		4500 B&C M
		4500-CO2
		502 D
		5030/8015
		5030/8015 Mod.
		5030/8015M
		5030A/3550A/8015A Mod.
		5030A/3550A/8015AM
		524.2
		6 Diesel
		6 Diesel Method
		6 Flouride Method
		6 Specific Conductance Method
		600/4-79-020
		601
		6010
		6010A
		602
		624
		625
		7020
		7040
		7041
		7060
		7060A
		7060B
		7080A

Field	Table(s)	Entries
		7090
		7131
		7140
		7190
		7380
		7420
		7420/3050
		7421
		7450
		7470
		7470 A
		7471
		7471A
		7520
		7610
		7740
		7740A
		7760A
		7761
		7770
		8000
		8010
		8010/8020 Mod.
		8010/8020M
		8015 Mod.
		8015A Mod.
		8015AM
		8015B/CA LUFT
		8015M
		8020
		8020A
		8080
		8080A
		8081
		8100 Mod.

Field	Table(s)	Entries
		8100M
		8150
		8240
		8240B
		8260
		8260A
		8270
		8270B
		9 Oil & Grease Dif Method
		9 Oil & Grease Difference Method
		9 Oil & Grease Extr
		9 Oil & Grease Extraction Method
		9 Solids Method
		9010
		9010A
		9020
		9020B
		9038
		9040B
		9045B
		9045C
		9050
		9060
		9060 Mod.
		9060M
		9065
		9071
		9071 Mod.
		9071M
		9080
		9081
		9081/6010
		9132
		9215B
		9215D

Field	Table(s)	Entries
		9252
		9310
		9320
		Analysis by ICP
		ASTM 9045B
		ASTM 9045C
		ASTM 9060
		ASTM 9080
		ASTM D 1290
		ASTM D 1498
		ASTM D 2161, unit=centipoise
		ASTM D 2161, unit=cSt
		ASTM D 2216
		ASTM D 3278
		ASTM D 3278 Setflash Closed Cup Tester
		ASTM D 4318
		ASTM D 854
		ASTM D 95
		ASTM D1498
		ASTM D2161
		ASTM D422
		Atomic Absorption-Cold Vapor Technique
		CA LUFT Gasoline
		CE 3-72
		Coleman 50-Hg Analyzer
		Coleman 50-Mercury Analyzer System
		D-1298
		D-3237
		D-3606
		D-4294
		D-4817
		EPA CLP Inorganic SOW-ILM03
		EPA Subsection 261.24
		Field
		Fld

Field	Table(s)	Entries
		Gas Chromatograph
		GC
		ICP
		ILM01.0
		Inorg SOW-ILM03
		Method 1312/6010A
		Method 150.1
		Method 7761
		Mod. 4500 B
		Mod. 4500 B & C
		Modified 351.3
		Modified 353.3
		Modified 365.3
		Modified 375.4
		Modified 376.2
		Modified 9060
		RCRA TSR 3015
		Section 7.3.3.2
		Section 7.3.4.1
		SM 450-CO2D
		Sohio Test E13
		Sohio Test E8
		Walkley-Black
		Walkley-Black Agronomy No.9 29-3.5.2
AREA	Location	Currently being determined
	Study	Currently being determined
CASING_TYP	Well	PVC SCH 40
CHEM_CLASS	CAS_List	Anions
		Bacteria
		Bioremediation
		Chemical Characteristics
		Gases
		Herbicides
		Metals
		Nitroaromatics



<b>Field</b>	<b>Table(s)</b>	<b>Entries</b>
<b>CO_SOURCE</b>	Location	Other Organics PCBs/Pesticides Physical Characteristics Radionuclides Semivolatile Organics Volatile Organics Waste Characteristics ASC Control Report Converted from plant grid coordinates Estimated from ArcView Surveyed with state plane datum
<b>CS_UNITS</b>	Hydraulic	cm/sec
<b>EL_SOURCE</b>	Location	ASC Control Report Converted from plant grid elevation Estimated from dtm
<b>KEYWORDS</b>	Study	Currently being determined
<b>LOC_TYPE</b>	Location	API Separator Boring/Test Trench FCC Staging Area Geoprobe Impoundment Pond MH Guard Basin Piezometer Process Vessel Bottoms SC Guard Basin Sediment Sludge Basin Soil Boring Soil Vapor Sump Surge Basin Test Trench/Pit Well
<b>MACRO</b>	Hydraulic Stratigraphy	BR (Bedrock) CG (Clayey Gravel)

Field	Table(s)	Entries
		CL (Clay)
		CO (Concrete)
		FC (Fine to Coarse Sand)
		FL (Fill)
		FM (Fine to Medium Sand)
		GR (Gravel)
		NR (No Recovery)
		PT (Peat)
		RS (River Silt)
		SC (Clayey Silt)
		SG (Sand and Gravel)
		SI (Silt)
		SI/CL (Mud)
		SS (Sandy Silt)
		SSC (Sandy Clay)
		WA (Water)
		WBR (Weathered Bedrock)
<b>MATRIX</b>	Action_Levels	AIR
	Sample_Collection	BASIN SEDIMENT
		BASIN SLUDGE
		BASIN WATER
		CLARIFIER SLUDGE
		CLARIFIER WATER
		FACILITY SLUDGE
		FCC FINES
		FILTER CAKE
		GROUND WATER
		LEACHATE WATER
		LNA PL
		POND SLUDGE
		QA WATER
		SEDIMENT
		SEPARATOR SLUDGE
		SLUDGE BASIN SLUDGE
		SLUDGE BASIN WATER

<b>Field</b>	<b>Table(s)</b>	<b>Entries</b>
		SOIL SOIL VAPOR SURFACE IMPOUNDMENT SLUDGE TISSUE TREATMENT WATER WASTE TREATMENT EFFLUENT WATER
<b>METHOD</b>	Hydraulic	Bower and Rice Cone Penetrometer Falling Head Permeability Triaxial Permeability
<b>PLCO_SOURCE</b>	Location	Digitized Export from CAD drawing Not specified Scaled from drawing
<b>PLEL_SOURCE</b>	Location	Estimated Not specified Well logs
<b>QC_CODE</b>	Sample_Collection	DU (Duplicate) FB (Field Blank) LB (Leachate Blank) LS (Leachate Spike) SA (Sample) TB (Trip Blank)
<b>QUAL</b>	Study	Currently being determined
<b>SAMP_DESC</b>	Grainsize	Brown Clayey Gravel with Sand Brown Organic Clayey Sand Brown Sand wit Silt and Organics Brown Silt Dark Brown Silt Gray Elastic Silt Gray Fat Clay Gray Fat Clay with Organics Gray Lean Clay with Organics Gray Lean Clay with Sand

Field	Table(s)	Entries
		Gray Organic Clay
		Gray Organic Elastic Silt
		Gray Organic Fat Clay
		Gray Silt
		Gray Silty Sand
		Light Brown Silty Clay
		Reddish Brown Organic Silty Sand
		Sand with Gravel, Some Silt
		Sand with Silt
		Sand with Silt and Gravel
		Sandy Silt
		Silt with Sand
		Tan and Brown Gravel with Sand and Organics
		Tan Sand with Silt and Organics
<b>SCREEN_TYP</b>	Well	0.01
		0.02
<b>SOURCE</b>	Action_Levels	57FR30798 (Vol. 57 Federal Register)
		PADEP ACT 2 (Internet Tables)
		RB1Q97 (EPA Region III Risk-Based Concentrations)
<b>STUDY_TYPE</b>	Study	Baseline Study Supplement
		Characterization Study
		Closure Performance
		CPT Study
		Current Condition Summary
		Geotech.
		Groundwater Data
		Hydrogeologic Investig. Identification Program
		IS Closure Plan
		IS GW Monitoring
		Misc. Data
		NJIT Soil Fracturing
		Phase I Impounding Pond
		Preliminary Assessment
		RCRA Assessment

Field	Table(s)	Entries
		RCRA Waste Program
		Recovery Testing
		Remediation Report
		Site Characterization
		Soil Reuse Survey
		TOSCO Baseline Study
		UST Investigation
		Currently being determined
<b>SUBMIT_AGENCY</b>	Study	
<b>TYPE</b>	Action_Levels	
		AIR
		FISH
		INDUSTRIAL
		MSC THRESHOLD, ALL AQUIFERS
		MSC THRESHOLD, NON-Residential, 0-2 feet
		MSC THRESHOLD, NON-Residential, 2-15 feet
		MSC, NON-Residential, 0-2 feet
		MSC, NON-Residential, 2-15 feet
		MSC, NON-Use AQUIFER, NON- Residential
		MSC, NON-Use AQUIFER, RESIDENTIAL
		MSC, RESIDENTIAL, 0-15 feet
		MSC, SOIL/GW, NON-Use AQUIFER, NON-Residential, 100 x GW MSC
		MSC, SOIL/GW, NON-Use AQUIFER, NON-Residential, GENERIC
		MSC, SOIL/GW, NON-Use AQUIFER, RESIDENTIAL, 100 X GW MSC
		MSC, SOIL/GW, NON-Use AQUIFER, RESIDENTIAL, GENERIC
		MSC, SOIL/GW, USED AQUIFER, TDS <= 2500, NON-Residential, 100 x GW MSC
		MSC, SOIL/GW, USED AQUIFER, TDS <= 2500, NON-Residential, GENERIC
		MSC, SOIL/GW, USED AQUIFER, TDS <= 2500, RESIDENTIAL, 100 x GW MSC
		MSC, SOIL/GW, USED AQUIFER, TDS <= 2500, RESIDENTIAL, GENERIC
		MSC, SOIL/GW, USED AQUIFER, TDS > 2500, NON-Residential, 100 x GW MSC
		MSC, SOIL/GW, USED AQUIFER, TDS > 2500, NON-Residential, GENERIC
		MSC, SOIL/GW, USED AQUIFER, TDS > 2500, RESIDENTIAL, 100 x GW MSC
		MSC, SOIL/GW, USED AQUIFER, TDS > 2500, RESIDENTIAL, GENERIC
		MSC, USED AQUIFER, TDS <= 2500, NON-Residential
		MSC, USED AQUIFER, TDS > 2500, NON-Residential
		MSC, USED AQUIFER, TDS <= 2500, RESIDENTIAL
		MSC, USED AQUIFER, TDS > 2500, RESIDENTIAL

Field	Table(s)	Entries
		RESIDENTIAL
		SOIL/GW, THRESHOLD MSC
		SOIL-AIR
		SOIL-GW
		SUBPTS, SOIL
		SUBPTS, WATER
		TAP WATER
		THRESHOLD MSC, RESIDENTIAL_SOIL_0-15 feet
UNIT		
	Action_Levels	%
	Grainsize	% Vol
	Results	% Wt
		% W/W CARBON
		1/100 ML
		1/100ML
		CENTIPOISE
		CFU/G
		CFU/ML
		CS
		CST
		DEG. C
		DEG. F
		Fibers/L
		G/CC
		G/GAL
		MEQ AS NA
		MEQ/100G
		mg/kg
		mg/kg
		mg/l
		mg/L
		mm
		MV
		NA
		PCI/L
		PCT REC

Field	Table(s)	Entries
		ppm
		SU
		ug/kg
		ug/kg
		ug/l
		ug/L
		ug/m3
		umhos/cm
		WEIGHT %
VALID_QUAL	Results	~
		+/- 1
		+/- 10
		+/- 14
		+/- 24
		+/- 5
		+/- 6
		+/- 9
		+/- 1
		+/- 11
		+/- 13
		+/- 15
		+/- 16
		+/- 17
		+/- 19
		+/- 2
		+/- 20
		+/- 21
		+/- 24
		+/- 3
		+/- 4
		+/- 7
		+/- 8
		+/- 9
		«
		B

Field	Table(s)	Entries
		C1
		E
		J
		U
<b>VALUE_QUAL</b>	Results	1 (Initial health and safety screening evaluation) 2 (Laboratory methods adapted for field use) 3 (Approved laboratory methods with complete QA/QC) 4 ("State of the Art" methods for a particular COC)



**Queries (information pulled from multiple tables in Access):**

Query Name	Description
Benzene_GW96_8010_8020	1996 benzene concentrations in groundwater (8010/8020 analysis method)
Benzene_GW96_8240	1996 benzene concentrations in groundwater (8240 analysis method)
Benzene_Soil_96	1996 benzene concentrations in soil
Chemicals_Above_GW_MCLs	All groundwater measurements above an MCL <sup>1</sup>
Chemicals_in_GW_MAX	Maximum groundwater concentration for each detected chemical
Chemicals_in_Soil_MAX	Maximum soil concentration for each detected chemical
Ethylbenzene_GW96_8010_8020	1996 ethylbenzene concentrations in groundwater (8010/8020 analysis method)
Ethylbenzene_GW96_8240	1996 ethylbenzene concentrations in groundwater (8240 analysis method)
Ethylbenzene_Soil_96	1996 ethylbenzene concentrations in soil
Geoprobes	Geoprobe sampling locations
Locations	All sampling locations
Lube_Area_GW_Q1	Lube area groundwater elevation and lnapl thickness data for 10/13/97
Lube_Area_GW_Q2	Lube area groundwater elevation and lnapl thickness data for 1/5/98
Lube_Area_GW_Q3	Lube area groundwater elevation and lnapl thickness data for 4/27/98
Lube_Area_GW_Q4	Lube area groundwater elevation and lnapl thickness data for 8/17/98
MTBE_GW96_8020A	1996 mtbe concentrations in groundwater (8020A analysis method)
Naphthalene_GW96_8010_8020	1996 naphthalene concentrations in groundwater (8010/8020 analysis method)
Naphthalene_GW96_8270	1996 naphthalene concentrations in groundwater (8270 analysis method)
Naphthalene_Soil_96	1996 naphthalene concentrations in soil
Piezometers	Piezometer sampling locations
Sediments	Sediment sampling locations
Site_Wide_GW_Q1	Site wide groundwater elevation and lnapl thickness data for 10/13/97
Site_Wide_GW_Q2	Site wide groundwater elevation and lnapl thickness data for 1/5/98
Site_Wide_GW_Q3	Site wide groundwater elevation and lnapl thickness data for 4/27/98
Site_Wide_GW_Q4	Site wide groundwater elevation and lnapl thickness data for 8/17/98

<b>Query Name</b>	<b>Description</b>
Sludge_Basins	Sludge basin sampling locations
Soil_Borings	Soil boring sampling locations
Soil_Vapors	Soil vapor sampling locations
Sumps	Sump sampling locations
TCE_GW96_8240	1996 tce concentrations in groundwater (8240 analysis method)
TCE_GW96_8240B	1996 tce concentrations in groundwater (8240B analysis method)
Test Pits/Trenches	Test pit/trench sampling locations
Toluene_GW96_8010_8020	1996 toluene concentrations in groundwater (8010/8020 analysis method)
Toluene_GW96_8240	1996 toluene concentrations in groundwater (8240 analysis method)
Toluene_Soil_96	1996 toluene concentrations in soil
TotalXylenes_GW96_8240	1996 total xylenes concentrations in groundwater (8240 analysis method)
TotalXylenes_GW96_8240B	1996 total xylenes concentrations in groundwater (8240B analysis method)
TotalXylenes_Soil_96	1996 total xylenes concentrations in soil
Wells	Well sampling locations

<sup>1</sup> PaDEP Act II msc, non-use aquifer, non-residential

Note: All concentration values are in units of ug/L for water and ug/kg for soil unless otherwise noted. All groundwater elevations are in units of feet - NAVD 1988. The groundwater elevations have not been corrected for the presence of LNAPL.

## APPENDIX D

### ARCVIEW-ACCESS EXERCISE

#### Mapping Environmental Data Stored in Microsoft Access

Andrew Romanek and David Maidment

The University of Texas at Austin

Available on the web at

[www.ce.utexas.edu/prof/maidment/giswr98/riskmap/webfiles/riskmap.html](http://www.ce.utexas.edu/prof/maidment/giswr98/riskmap/webfiles/riskmap.html)

#### Table of Contents

- [Goals of the Exercise](#)
- [Computer and Data Requirements](#)
- [Introduction - What is a Relational Database](#)
- [Procedure](#)
  1. [Acquire the Coverages and Database](#)
  2. [Display the Coverages](#)
  3. [Understanding Relationships](#)
  4. [Creating Queries in Access](#)
  5. [Setting up the ODBC Driver](#)
  6. [Importing Data into ArcView](#)
    1. [Joining Data with an Existing Coverage](#)
    2. [Creating a New Point Coverage](#)
  7. [Creating Contour Grids](#)
    1. [Inverse Distance Weighting \(IDW\) Method](#)
    2. [Spline Method](#)
  8. [Preparing Data for a Risk Map](#)
    1. [Using the Binomial Distribution](#)

#### Goals of the Exercise

This exercise covers the following concepts:

1. Working with the relational database Microsoft Access.
2. Importing data from Access into ArcView.
3. Creating contour plots.
4. Calculating exceedance probabilities of specific chemical levels.

Upon completion of this exercise, you should be able to set up simple queries in Access, import data from a database into ArcView, join these data to an existing coverage, and create different surfaces based on concentration measurements. This exercise is intended to introduce you to the numerous possibilities of using Access and ArcView for display and analysis.

#### Computer and Data Requirements

To complete this exercise, you will need access to a computer with ArcView GIS 3.1 (or 3.0a) and Microsoft Access '97. Your computer must also have an Open DataBase Connectivity

(ODBC) driver for Microsoft Access databases. This driver allows ArcView to locate tables and queries in Access, and thus retrieve any desired information. The existence of this driver can be checked by opening the control panel and double clicking on the ODBC icon. Select the ODBC drivers tab to see if you have an Access driver. More information on ODBC drivers and SQL connections is available from the ArcView help topics menu.

The Creating Contours step requires the Spatial Analyst extension.

### **Introduction - What is a relational database**

A database is simply a collection of information that relates to a particular subject or purpose. The database is stored in one file, but this file can contain multiple tables, all of which relate to the particular subject but contain different data related to that subject. For instance, a database for a retail company could contain a table for customer information, another table for product information, and a final table for shipping orders. All three tables relate to the operations of the company, but they contain different data sets that are easier to manage and understand through the use of separate tables. Tables are the fundamental elements of a database, and all database operations are performed through actions on tables. Tables are essentially storage containers, and database programs such as Microsoft Access can perform various operations on these containers. Data in tables can be managed in the following ways with Access:

1. View, add, and update table data using online **forms**.
2. Find and retrieve just the data you want using **queries**.
3. Analyze or print data in a specific layout using **reports**.

In order to combine data from separate tables into one form, query, or report, relationships or links can be set between each table. In the example described above, a common field such as a customer id could exist in both the customer information table and in the shipping orders table. By joining this field between tables, data can be retrieved from each table and combined into another table (a query or report). For example, one could determine all the shipping orders in March for a specific customer. Queries and reports can also perform calculations on data. To extend the above example even further, one could create a report that would calculate the money spent by each customer for orders in March. To accomplish this task, Access (or an equivalent database utility) would determine all the orders and products bought for each customer in March from the shipping orders table and then calculate a sum of \$'s spent using prices in the product information table. This figure displays the data management operations in Access:

Store data once in one table, but view it from multiple locations. When you update the data, it's automatically updated everywhere it appears.

Customer ID	Company Name	City
BSBEV	B's Beverages	London
EASTC	Eastern Connection	London

Company Name	City	Order Date
B's Beverages	London	11-Apr-96
Eastern Connection	London	12-Apr-96

Customer ID:	BSBEV
Contact Name:	Victoria Ashworth
Company Name:	B's Beverages

Customer: B's Beverages	
Order ID:	Sale Amount:
10943	\$711.00
10947	\$220.00
11023	\$1500.00
<b>Total:</b>	<b>\$2431.00</b>

(From Access help menu)

In this exercise, you will be using a sample environmental database for an oil refinery located in Marcus Hook, PA. This database is set up like any other relational database, but contains information specific to environmental data at this facility. This database contains information such as sampling point locations, groundwater levels, and chemical analytical results. We could determine, for instance, the location of a specific well, information about that well, the groundwater levels that have been observed at the well, and the chemical concentrations that have been measured at the well. Databases work in both directions, so we could also find all the locations at which a specific chemical of concern (COC) was measured. Pretty cool!

With an environmental database, it would be beneficial to not only create reports and queries, but also to view and analyze this information. We can query concentration measurement locations in Access, but it is difficult to visualize these locations from a table of data. That is where ArcView becomes a very useful tool because of both its display and data storage capabilities. Let us see how this works.

## Procedure

### Acquire the Coverages and Database

There are a total of nine files that you will need to complete this exercise. The files can be found on the LRC server at [lrc/class/maidment/giswr/riskmap](http://lrc/class/maidment/giswr/riskmap). The files are also available via anonymous ftp from <ftp://ftp.crwr.utexas.edu/pub/gisclass/riskmap>. [Instructions on how to use anonymous ftp](#). Here is the file breakdown:

- **bound.e00** - Marcus Hook Refinery property line
- **monit.e00** - On-site monitoring wells
- **sourcearea.e00** - Concentration source area
- **usgsmap.tif** - USGS map for Marcus Hook and the surrounding area
- **usgsmap.tfw** - Geographic reference of the USGS map (without this file, the image would not be in the correct projection)
- **database.mdb** - Microsoft Access environmental database
- **cocvalue.ave** - Script to create a point coverage of concentration data
- **gwelev.avl** - Legend for the groundwater maps
- **coc.avl** - Legend for the concentration maps

Some of these files are in export format, so you will need to use the ArcView **Import71** command or the Arc/Info **import** command. These are the commands if you are using Arc/Info:

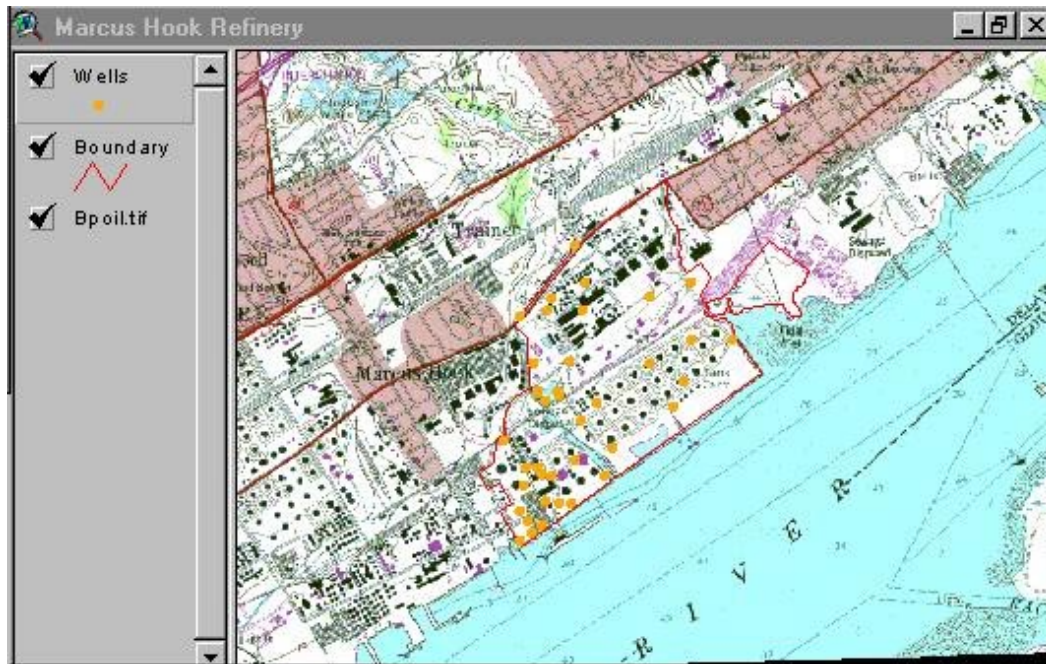
- Arc: **import cover bound.e00 boundary**
- Arc: **import cover monit.e00 wells**
- Arc: **import cover sourcearea.e00 sourcearea**

Note that the file **monit.e00** is imported as a new name, **wells**, and similarly, **bound.e00** becomes **boundary** when imported. **Usgsmap.tif** is an image file. We will examine this further in the next section.

### **Display the Coverages**

Start up ArcView. Open a new view and add the themes **Boundary** and **Wells** from your working directory. You should now see the refinery boundary line and the layout of wells across the site. In order to get a better idea of where this refinery is located and its general layout, let's add the USGS map. Add the theme **usgsmap.tif** to your view. Make sure you have image data source selected. This map has a raster data structure. If you zoom in closely, you will see that the map is a collection of grid cells. Each cell has a different color, and the size of these cells determines the pixel resolution for the image. Zoom out to obtain a clearer view.


This refinery lies along the Delaware River. Marcus Hook is located in Southeast Pennsylvania, about twenty miles South of Philadelphia. The data are given in State Plane Coordinates for the South Region of Pennsylvania. This area is highly industrialized, but there is a residential area across from the northern boundary of the refinery. Your view should look something like this:

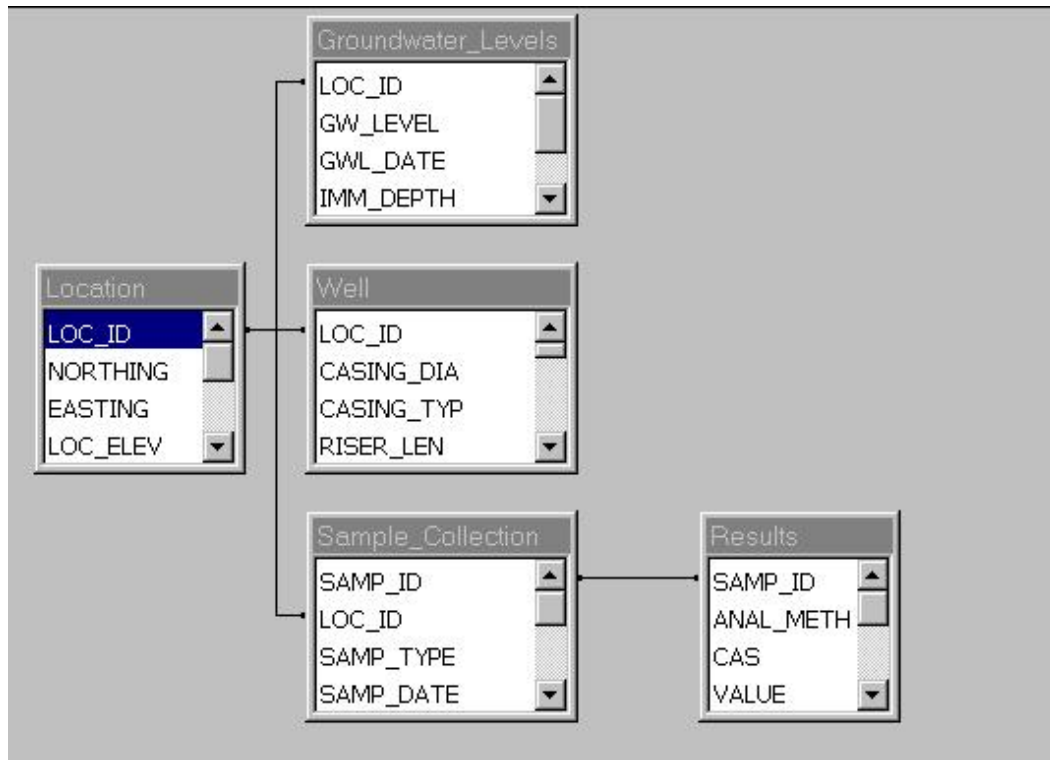


The University of Texas is currently involved with a research project at this facility, and the project goal is to develop a program for corrective action that will protect human and ecological receptors both on and off the property. The primary steps in the risk assessment process are to identify sources, source areas, transport mechanisms, points of exposure, exposure routes, and receptors. Sources are the physical structures that could cause a chemical release such as a tank or pipeline. Source areas are the locations of the highest concentrations in environmental media, such as groundwater or soil. A transport mechanism is the combination of chemical, physical, and biological processes that move a chemical from the source to the point of exposure. The point of exposure is the location at which an individual may come in contact with a chemical. The exposure route is the manner in which a chemical comes in contact with an organism, and receptors are persons that are or may be affected by a release. In this exercise, we will focus on identifying source areas based on concentration measurements taken from the soil on-site. We will also look at groundwater level measurements to identify potential transport pathways.

### **Understanding Relationships**

From the Start menu, choose Programs and Microsoft Access. Select **Open an Existing Database** and choose the database you obtained in the first step. The database window will appear with five different tables listed. Again, these tables contain separate types of data all

related to the environmental conditions at the refinery. Click on the  icon to see the relationships between these tables. This view shows the tables, the fields within each table, and the links between tables:



In order to understand these relationships, let us look at the Sample\_Collection and Results tables. The Sample\_Collection table is linked to the Results table through the common SAMP\_ID field. Sample\_Collection provides information related to each sample that has been collected, including the type of sample and location where it was obtained. The Results table displays all the concentration measurements for each sample. These concentrations are determined using specified analytical methods. View the data in any table by returning to the original "Database" window that you got when you opened the database and then double-clicking on any of the tables to see its contents. Click on each of the tables so that you can see what it contains.

More information on what each field refers to can be obtained from the [Database Dictionary](#).



## Creating Queries in Access

In this step, you will learn how to take information from different tables and combine it into one table using a query. We want to create a query that will determine the groundwater level measurements on a particular date. From the main Database window, click on the **Queries** tab and select **New**.


Access "Wizards" can be useful tools for designing queries, forms, and reports. However, we will use the design view to create our query, so select **Design View** and click OK.

You will be prompted for the tables and/or queries you want to obtain data from for your query. Add the **Location**, **Groundwater\_Levels**, and **Well** tables and then close the window. You will then see a clipboard containing the tables you added and an empty query form. Notice that the relationships among tables are also shown on this clipboard. You can modify these relationships by double clicking on any of the links between tables. When creating queries, you need to show all tables in the relational chain even if you are not obtaining data from all of them. For instance, if you wanted to obtain the groundwater levels for a specific well and also list attributes for that well, you would need to show the Groundwater\_Levels, Well, and Location tables even if you did not want to include any fields from the Location table.


From Groundwater\_Levels, click on the **LOC\_ID** field and drag it into the first empty column on the form. Also select **GWL\_DATE** and **GW\_LEVEL** from this table and add them to the query form. From the Well table, select **TWC\_ELEV**, which is the top of the well casing elevation (also referred to as the top of the well riser elevation). The GW\_LEVEL is the depth to groundwater as measured from the top of the well casing. By subtracting this depth from the elevation of the top of the well casing, the elevation of the water table is found, and from these data a map of water table elevations can be created. The elevations used here are referenced to the North American Vertical Datum of 1988 (NAVD88), which is the most accurate vertical datum presently available.

To calculate the elevations, we will create a new field using an equation. To do this, move the cursor to the next empty column field and type **GW\_ELEV:[TWC\_ELEV]-[GW\_LEVEL]**. This expression creates a new column called GW\_ELEV using the given mathematical expression.

Besides performing calculations, you can also set criteria so that the query returns only the type of data you desire. For this query, we want to limit the groundwater level measurements to a single date, so type **#4/30/94#** in the Criteria field for the GWL\_DATE column. Also type **Is Not Null** in the Criteria field for the GW\_LEVEL column so as to eliminate measurements with no values.

A final step that may not always be necessary but is good to do anyway, is to remove duplicated records from your selected set. To do this, click in the gray area within the Query Window (so that the cursor is not identifying a particular table or field in the query), and then click on the  icon in the main tool bar. Switch **Unique Values** from **No** to **Yes** and then close the window. Your completed form should look like this:



Now click on  to run your query. You should have 36 records. Click on the x and then say Yes to Save Changes? Call your query **GW\_Elevations**. Be sure to include the underscore because ArcView is not able to read spaces, and thus, without the underscore in the table name, you will not be able to import your data into ArcView later in the exercise. If you don't get the correct number of records, close the query window, open a new query, and repeat the instructions carefully. Make sure that you've added all three tables that are needed. The Location table is needed to complete the relational linkage between the Groundwater\_Levels and Wells data even though none of the query fields are drawn from the Location table.

We are also interested in benzene concentrations at the refinery, so we will need to create another query to obtain these data. Again select New and choose Design View. Add the **Location**, **Sample\_Collection**, and **Results** tables. Include the following fields in your query design:

- From Sample Collection: **LOC\_ID**, **SAMP\_ID**, and **MATRIX**
- From Location: **EASTING** and **NORTHING**
- From Results: **CAS**, **VALUE**, **DET\_LIMIT**, and **UNIT**

The order you put these in will not matter, but be sure your query contains these fields because they will be searched for later with an Avenue script. As criteria, type **"SOIL"** in the MATRIX column, **"71-43-2"** in the CAS column, and **"UG/KG"** in the UNIT column. Be sure

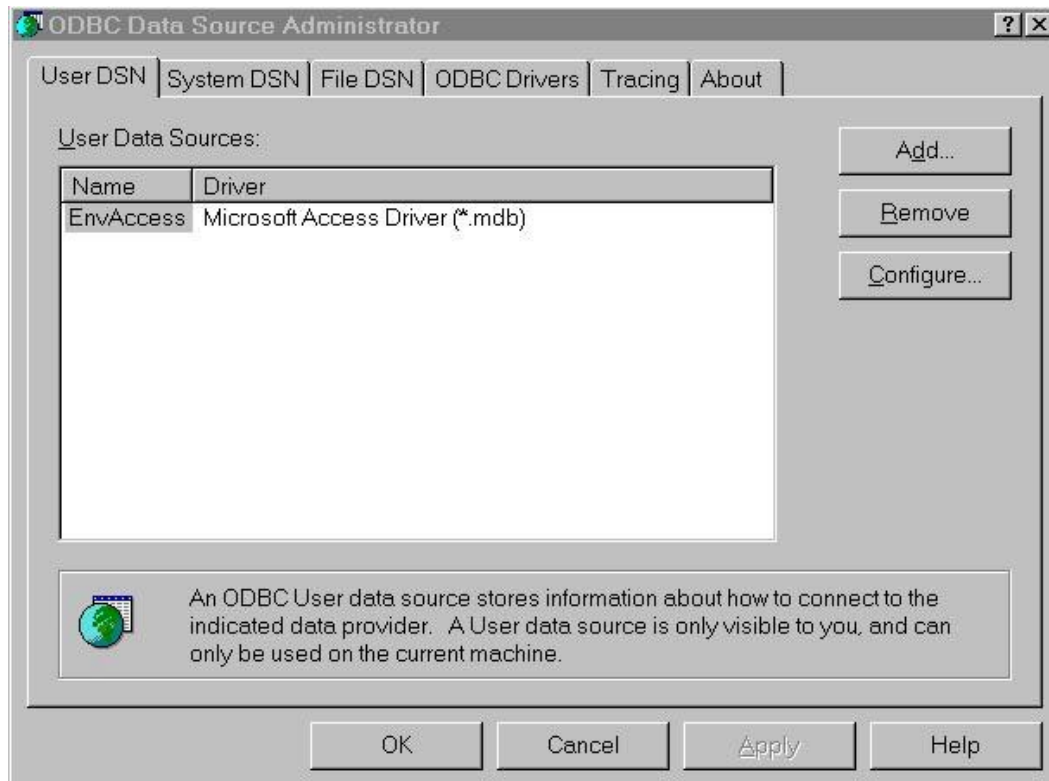
to include the quote marks. Run the query and then save it as **Benzene\_Soil**. You should have 177 records. The CAS field 71-43-2 is the identification number for benzene. Note that there are three tables involved in this query and two different key fields are used to connect records in these tables. The records with no data in the Value field had a measurement that was below the detection limit as specified in the DET\_LIMIT field. The detection limit is the minimum concentration that can be measured in an analysis. There are 52 records with a value greater than the detection limit recorded. Note that detection limits vary depending on the analysis method and the actual constituents in a sample. Some samples with no detected concentration but high detection limits might have a value that is greater than that for a sample with a detected concentration but low detection limit. This problem is a data quality issue that will not be discussed here.

*To be turned in: Write in words what has been done in the query you have just executed. Which fields are used to connect the tables? Which fields are used to select the required records? What is the query that has been executed? Which fields are picked up as ancillary information about the selected records? Choose one record from the resulting Benzene\_Soil table and describe what information this record represents (utilize the database dictionary provided above).*

### **Setting up the ODBC Driver**

Before importing data into ArcView, we need to tell the computer where to find the data. Open your Control Panel and double click on the ODBC icon.

Under the User DSN tab: click **Add**, select the Microsoft Access Driver, and then click **Finish**. For Data Source Name, type **EnvAccess**. Under Database, click **Select** and choose the location path for **database.mdb**. Choose OK and your driver should now appear under User Data Sources and look something like this:



Click OK to leave.

## Importing Data into ArcView

### Joining Data with an Existing Coverage

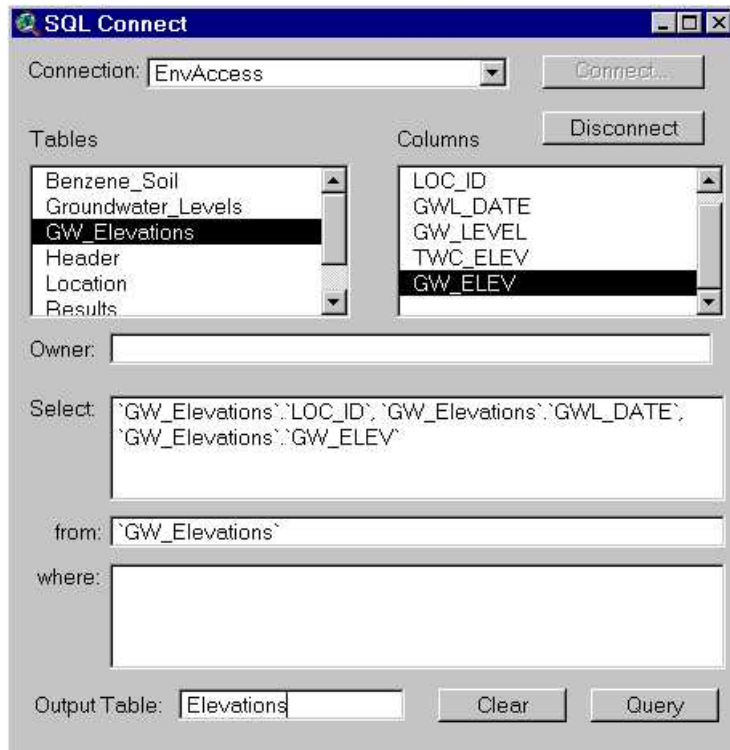
You have now built your database and are ready to import data into ArcView. From your previously opened project, open the attributes table for Wells. Notice the Loc\_Id field has identifiers that are the same as the LOC\_ID field in the database. If you are going to import data from a database and connect it to an existing coverage, you need to have two fields with identical values so that a link can be established.

Make the project window active and select **SQL Connect** under **Project**. In the Connection box, you should be able to scroll down and choose **EnvAccess**. If not, the driver is not set up properly. Click on **Connect** and you will see all the tables and queries that exist in the database (8 in all).

Click once on GW\_Elevations to see the fields in that query. Double click on **LOC\_ID**, **GWL\_DATE**, and **GW\_ELEV**. This process selects the fields we want to import. Now double

click on **GW\_Elevations** to set the input source. Criteria expressions can be added in the Where field, but this task is much easier to accomplish in Access. For Output Table, type **Elevations**.

Your form should look like this:



Click on **Query** and ArcView creates a table of data using the database.

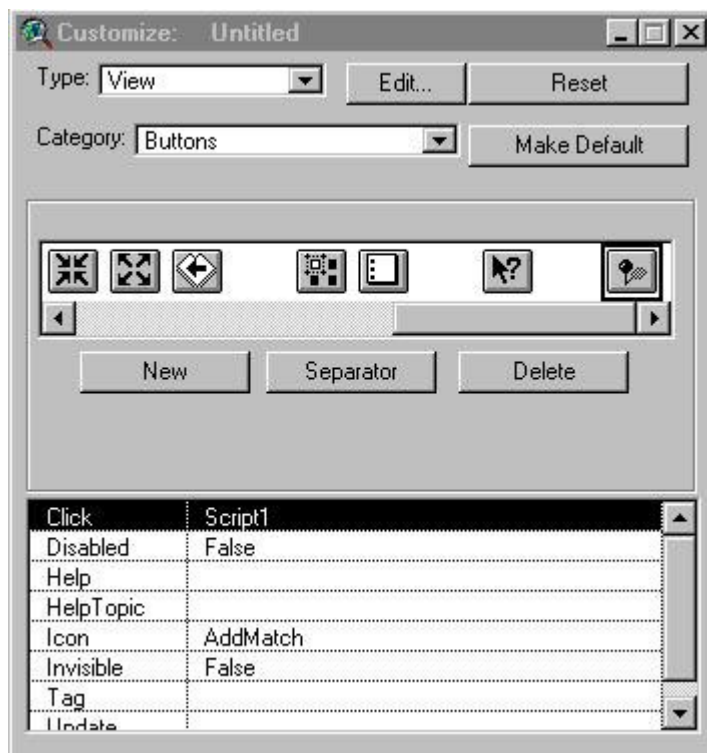
Close the window and open the Attributes of Wells table. We now want to join the data from the query with the well coverage. Make the Elevations table active and click on *LOC\_ID*. Now make the Attributes of Wells table active and click on *Loc\_id*. Use the **Table/Join** command to include the data you imported in the well attributes. If you go back to your view and use the identify tool to click on a well, you will now see the two fields you imported. The water level was not measured at some of the monitoring wells on 4/30/94 so the *GW\_LEVEL* field is blank for them. **NOTE:** Each time you open this project, ArcView will requery the database and rejoin the data. This feature is particularly useful if you make a lot of changes to the database. Neat!

### Creating a New Point Coverage

In the previous step, we already had a coverage to which we could join data. However, this luxury might not always be available. In this step, we will create our own coverage using location

coordinates from the database. We already determined the points and data we are interested in with our benzene concentration query, but we now want to import these data. A script has been created to do this for you. The script is very similar to the one used in the Building a Basemap exercise, but it includes a section to read the coordinates from the database (using SQL Connect) and additional fields in the output table for the data we want.

In the Project window, switch to the Script icon and select **New**. Load the text file **cocvalue.ave** and then compile the script. The script requires that the view to be the active document. You can use the Window menu to switch back and forth between the view and the script, but it is often easier to customize your view properties to eliminate the switching procedure. To do this, open your View and double click on any part of the blank gray area of the toolbar. You should see the **customize** box appear. Change the **Category** to **Buttons**. Scroll all the way to the right and click once on the help icon. Click **Separator** and then **New** to create a new icon, one space away from the help icon. Go to the table in the lower part of the window, and Double click on **Click** and select **Script1** (Script1 has to be open in the project for this to be a valid choice). Then double click on Icon and choose an appropriate button. Your customize box should look like:



Do not hit Reset or Make Default here! Close the customize window by clicking on the x in the upper right corner and click on your new icon to run the script. When prompted to Enter the COC Query Name, type **Benzene\_Soil**. For the output table name, type **Benzene Soil Data**. Finally, save the new shapefile as **benzene** to your working directory. What this script does is create a new point coverage in ArcView using the Easting and Northing values that were contained in the data extracted from MS Access in the Benzene\_Soil query. An alternative method to using a script is to import the data using SQL Connect and then display the data with the Add Event Theme option available from the View menu. This alternative provides more flexibility when adding different types of queries.

Make the new theme visible and you will see that the benzene measurements are scattered across the site. Some areas contain many measurements and others very few. Open the attributes table and take a look at the data. The records with an empty value field are the "no detect" measurements.

*To be turned in: Use the Statistics command for the Value field and report the mean, median, range, maximum, minimum, and standard deviation of the measurements. What do these values tell you about benzene concentrations measurements at the refinery?*

You can find out more about how to statistically analyze the benzene data by looking at the [Statistical Analysis of Environmental Data](#) exercise for the Environmental Risk Assessment Course.

### **Creating Contour Grids**

In a perfect world, we would know the groundwater level and concentration value at any location, and at any time. However, it would be an extremely time-consuming and expensive task to visit every location on the site and determine the groundwater level or concentration at that location. Additionally, this study would only give us values for one moment in time. How then can we track and evaluate information using a limited data set? One method of analysis that will estimate measurements at any location is interpolating a surface through the existing data points. The values of this surface will be equal to the measurements at the measurement locations, and will be estimated for all other locations. There will likely be errors in the estimates, but the surface can still be a useful tool for identifying areas of concern and determining where more measurements are needed. For this exercise, we will examine two different methods of estimation:

the Inverse Distance Weighting (IDW) Method and Spline Method. Both methods can be run using the Spatial Analyst extension in ArcView.

### **Inverse Distance Weighting (IDW) Method**

For the Inverse Distance Weighting Method, a fine mesh grid is laid over the study area and a value is interpolated for each grid cell using the inverse distance squared between the cell location and the measurement location as a weight for each measurement. The interpolator assumes that each point has a local influence that diminishes with distance. It weighs the points closer to the processing cell greater than those farther away. You can specify the number of points to interpolate from, or optionally, you can specify all the points within a radius to interpolate from. The power parameter in the IDW interpolation controls the significance of the surrounding points on the interpolated value. Higher powers result in less influence from more distant points. You can also specify barriers beyond which the surface will not interpolate for the input point. For instance, you could specify a barrier for a cliff in your study area.

Turn on the Spatial Analyst extension from the File menu (Project window active). Open your View and make the Wells theme active. Choose **Interpolate Grid** under **Surface**. For Output Grid Extent, choose **Same as Boundary**. The Output Grid Cell Size, Number of Rows, and Number of Columns will be set according to the extent of the boundary theme. However, you can change these values if you want a finer or larger grid cell size. You might want to choose a 25' cell size, for example. Remember that your grid will only be as accurate as your data measurements (i.e., a fine grid will not be any more accurate than a larger grid if the amount of data points is small). Accept the defaults for your grid parameters.

You will now see the Interpolate Surface dialog box. Choose **GW\_ELEV** for the Z Value Field and accept the default surface parameters. If you don't see GW\_ELEV as an available field at this point, it means that you did not do the database query correctly using the ODBC connection through EnvAccess. You should see a grid appear of groundwater elevation measurements for the refinery. You can use the legend editor to color your grid, and there are even a few elevation color schemes. You can also load the legend, **gwelev.avl**, which has been provided for you. The elevation values for your grid are the elevations above mean sea level, so water will flow from the higher values to the lower ones. If you want to see the input data values for the groundwater elevations, you can use the Theme/Auto-Label command with the Wells theme. Your grid should look something like the one provided below. To save your grid, choose **Save Data Set** under



**Theme.** Call your grid **gwidw**. Use the theme properties to change the theme name to **GW IDW Surface**.

*To be turned in: Which well has the highest recorded groundwater level? Use the measuring tool to determine the distance across the property to the boundary line near the Delaware River. What is the maximum gradient of groundwater flow on this site? All map units, vertical and horizontal are in feet.*

### **Spline Method**

Another method for interpolating surfaces is the Spline Method. The Spline Method uses a polynomial function to fit a surface which passes through all the data values and which is smoothed so that there are not as many peaks and pits as in the IDW Method. Conceptually, it is like bending a sheet of rubber to pass through the points, while minimizing the total curvature of the surface. This method can overshoot estimated values if there are large changes within a short horizontal distance, but the method is well suited for gently varying surfaces. You can specify two different kinds of surfaces. The Regularized method yields a smooth surface while the Tension method tunes the stiffness of the surface according to the character of the modeled phenomenon. The weight parameter for the Regularized method defines the weight of the third derivative of the surface in the curvature minimization expression, while the weight parameter for the Tension method defines the weight of tension. The number of points parameter identifies the number of points per region used for local approximation.

Follow the same procedure above, except this time choose **Spline** for the Interpolate Surface Method. Also choose **Tension** for Type with a weight of **10**. The default value range will be somewhat larger for this grid, and you can use the `gwelev.avl` legend to allow for a better comparison with the IDW grid. Note that the grid is only accurate within the facility boundary area. If you click on a point in the Delaware River, you will notice it has a large elevation value. Its elevation should be zero (assuming the river is at mean sea level). Again check the layout below to compare your answer. Save your grid as **gwspline** and rename it **GW Spline Surface**.

Repeat this procedure for the concentration data (use Value field). For the Spline Method, use **Tension** for Type with a weight of **10000**. Call your two grids **cocidw** and **cocspline** and rename them **Benzene IDW Surface** and **Benzene Spline Surface**. Load the legend `coc.avl` for both grids to obtain a better comparison. Here is what your four grids should look like:

### Groundwater Elevations

IDW Method



Spline Method



### Benzene Concentrations

IDW Method



Spline Method



We could also have generated contour lines instead of grids. The procedure is identical except that you choose Create Contours instead of Interpolate Surface from the Surface menu. Feel free to experiment with this option and with different weight parameters. You will notice that the weights can have a significant influence on the results. You can also overlay the contours created for a particular interpolated surface so as to more easily understand what the interpolated grid represents. If you already have an interpolated grid and wish to create the contours from that, just highlight the grid and then use **Surface/Create Contours**.

*To be turned in: Briefly describe the benzene concentration grids. Which sample contains the highest benzene concentration? Which interpolation method do you think is better for mapping groundwater measurements and which interpolation method is better for mapping benzene concentrations? Present a copy of the map that you feel best presents the groundwater elevation data and another for the benzene concentration data.*

### Preparing Data for a Risk Map

Concentration maps for the entire site can give us a good idea of the range in concentration data. However, many of these data measurements come from different sources and some areas have considerably more measurements than others do. For a more detailed analysis, let us focus our study on one particular area. Add the **sourcearea** coverage to your view. Use the legend editor to make the polygon transparent and select an appropriate color for the outline. This polygon encloses a former surface impoundment that has received excess sewer effluent, sand filter backwash, stormwater runoff, and tank farm wastes.

It turns out that we have a lot of benzene soil measurements from this area. We can query these measurements by making the benzene.shp theme active and selecting **Select by Theme** under **Theme**. Select features of the active theme that **Intersect** the selected features of **Sourcearea** and choose **New Set**. Open the attributes table of benzene.shp and use the promote icon to view these records. You should see 28 selected records. We now want to analyze these data using Microsoft Excel, so use the **Export** command under **File**. Choose dBase for the export format and save the table to your working directory as **records.dbf**. ArcView will only export the highlighted records to your dBase file. Start Excel and open the table (be sure to select dBase Files for Files of type:). You should see something very familiar to the table you had in ArcView.

### Using the Binomial Distribution

For the source area we have selected, we want to determine the probability the any soil sample taken from this area will be above a threshold concentration, such as an EPA action level. There are only two possible outcomes for each soil sample we obtain: a value above the threshold or a value below the threshold. Problems of this type can be modeled by a Bernoulli sequence, which is based on the following assumptions:

1. Each trial has only two possible outcomes: occurrence or non-occurrence of an event.
2. The probability of occurrence of the event in each trial is constant.
3. The trials are statistically independent.

For our small area of analysis, we can assume that the final two assumptions are valid. However, this assumption would not be safe if we were looking at the entire facility. The probability of having a sample value above the threshold concentration in the Southwest corner of the refinery is not likely to be the same as the probability of having a sample value above the threshold concentration in the Northeast corner of the refinery.

If the soil samples are taken from an area containing an infinite number of samples, or are taken from a finite supply and replaced, the ratio of soil samples above a threshold to the total number of samples obtained will be described by a binomial distribution:

The probability of drawing  $s$  soil samples (violations) in  $n$  trials is equal to

$$e(n, s, P_T) = \binom{n}{s} (P_T)^s (1 - P_T)^{n-s} \quad \text{where } \binom{n}{s} \text{ is the number of combinations of } n \text{ trials that contain } s \text{ successes.}$$

The cumulative probability of  $s$  or more successes in  $n$  trials is given by the sum of all  $e(n, m, P_T)$  with  $m$  greater than or equal to  $s$ :

$$E(n, s, P_T) = \sum_{m=s}^n e(n, m, P_T)$$

In our case, we do not know the ratio of measurements exceeding to measurements not exceeding the threshold. This ratio can be estimated by repeatedly making soil measurements and keeping track of the number exceedances and non-exceedances. Thus, the best estimate ratio of exceedances to non-exceedances is simply the ratio of these two outcomes using the obtained measurements. The expected accuracy of this estimate increases as more samples are taken. From our data set, we can obtain the best estimate probability that any soil sample in the area will be in violation of a set threshold by dividing the number of violations by the total number of samples. You can use your Excel spreadsheet to calculate this probability. Highlight your records and then choose **Data/Sort**. Sort by **Value** in **Descending** order. Your first five results should look like this:

	A	B	C	D	E	F	G	H	I
1	SAMP_ID	LOC_ID	EASTING	NORTHING	MATRIX	CAS	VALUE	DET_LIMIT	UNIT
2	#12 (6-8)	Boring 12	2626306.435486	186010.037826	SOIL	71-43-2	4400.0	10.0	ug/kg
3	#11 (4-6)	Boring 11	2626342.137652	186034.686628	SOIL	71-43-2	477.0	10.0	ug/kg
4	#2A (8-10)	Boring 2A	2626461.858781	185796.115014	SOIL	71-43-2	70.0	10.0	ug/kg
5	SI-12	12/94SI-12	2626331.097450	185812.124064	SOIL	71-43-2	19.0	0.5	ug/kg
6	SI-7	12/94SI-7	2626415.386158	185791.993126	SOIL	71-43-2	14.0	0.5	ug/kg

For illustration purposes in this exercise, let us use a threshold concentration of 15  $\mu\text{g}/\text{kg}$  (this is actually much lower than EPA target levels for benzene in soil). The best estimate probability = (# of samples greater than 15  $\mu\text{g}/\text{kg}$ )/(total # of samples). One of the benefits of this method is that it includes the no detect data. However, this method does not take into account the actual value measurements (i.e., for the given threshold, a sample with a value of 50,000  $\mu\text{g}/\text{kg}$  is not distinguished from a sample with a value of 20  $\mu\text{g}/\text{kg}$ ). In this case, the total number of samples is equal to the number of records in the table, namely 28.

*To be turned in: What is the best estimate probability for the source area data?*

We now have a best estimate probability, but it will be more appropriate to express the probability estimate as a range of possible values with a degree of confidence that the true probability falls in that range. A typical statement of this type is "The probability of measuring a concentration above the target level in a single trial lies between 40% and 60% with a confidence level of 95%." As an appropriate screening level for our data, we would like to calculate the upper bound on our probability estimate with a confidence level of 95% (two-sided).

To estimate an upper bound, we need to find the value of  $P_u(s)$  that will satisfy the following equation:

$$1 - E(n, n-s, 1-P_u(s)) = 1 - \alpha/2 \text{ where } \alpha = 1 - \text{confidence limit}$$

This task can be accomplished using Excel with the Solver Add-In. Select **Tools/Add-Ins** and verify that **Analysis ToolPak** and **Solver Add-In** are checked. Solver is installed on the machines in the LRC, but if you are using your own version of Microsoft Office, you might have to add Solver using your setup CD. Move to the right side of your spreadsheet and edit it as follows: in cell J1, type **0**; in cell K1, type **=1-binomdist(n-s,n,1-J1,true)+binomdist(n-s,n,1-J1,false)**; in cell K2, type **0.025** ( $=\alpha/2$ ); and in cell L1, type **=J1\*100**. Put the actual number of successes and trials in place of s and n (s = 4, so n-s = 24, and n = 28), respectively. You don't have to do this task in this location, but make sure you set up the references correctly. Binomdist is a built-in function of Excel that can be used for binomial distribution calculations. The first term is the number of successes. The second term is the number of trials. The third term is the probability, and the fourth terms tells Excel whether it should calculate the probability mass function (false) or the cumulative binomial (true). It should also be noted that Excel defines the cumulative binomial as the sum from x = 0 to x = s as opposed to the sum from x = s to x = n as described above. This is the reason for the strange looking formula you wrote on your spreadsheet.

Now we are ready to solve for the upper bound. Select **Tools/Solver**. We want to change the first cell until it solves the equation, and then place the result in the last cell. Thus, type **\$J\$1** under By Changing Cells, and type **\$L\$1** under Set Target Cell. In the Subject to the Constraints section, click **Add** and type Cell Reference: **\$K\$1 >= \$K\$2**, and then click OK. Make sure Equal To: is set to Max and then click **Solve**. Accept the solution and the value in cell L1 is your upper bound (%). Your completed Solver form should look like this:



*To be turned in: What is the upper bound probability that a sample selected from this source area will contain at least 15  $\mu\text{g}/\text{kg}$  of benzene?*

Now that we have an upper bound, we could add this information to our Sourcearea attributes table in ArcView. If we had other source areas, we could follow a similar procedure and then use the legend editor to color the source areas according to their upper exceedance probability. This map would be a useful tool for displaying possible risks associated with each source area. OK, you are done!

## APPENDIX E

### DATA CONVERSIONS

#### Projections

Projections can be modified using either ARC/INFO or ArcView. The only requirement is that the user know the input and output projection parameters. Here is an example conversion for a coverage in geographic coordinates to the Pennsylvania State Plane System (South Zone) using ARC/INFO:

- Arc: project cover input\_file\_name output\_file\_name
- Project: input
- Project: projection geographic
- Project: units dd
- Project: datum NAD83
- Project: parameters
- Project: output
- Project: projection lambert conformal conic
- Project: units feet
- Project: datum NAD83
- Project: parameters
- 1<sup>st</sup> Standard Parallel: 39 56 0.000
- 2<sup>nd</sup> Standard Parallel: 40 58 0.000
- Central Meridian: -77 45 0.000
- Reference Latitude: 39 20 0.000
- False Easting: 600000
- False Northing: 0.000000
- Project: end

To modify a projection in ArcView, the user must have the *Projector* extension, which can be downloaded from ESRI's scripts page ([andes.esri.com/arcscripsts/scripts.cfm](http://andes.esri.com/arcscripsts/scripts.cfm)). The process is similar to that in ARC/INFO except that dialog boxes are utilized to select the input and output projection parameters. This extension includes many of the standard projections, but also has an option to use a custom projection. The user should always be aware of the projection parameters so that no inconsistencies arise in the output data. *Projector* is a very viable alternative for those users who do not have ARC/INFO.

## **Topographic Maps**

As previously mentioned, most of the regional coverages that are obtained from the internet are already in ARC/INFO format and only need to be imported and reprojected. However, there are a few exceptions, such as the USGS topographic maps. A USGS map can be exported from a Horizons Technology Sure! Maps Raster CD-ROM. Horizons sells a variety of CD-ROMS, including one containing maps of cities in the Northeast at a scale of 1:24,000. The map extent is selected in the Sure! Maps program and then exported as a Tagged Image File (\*.tif) that is referenced in geographic coordinates. This image can be viewed as is in ArcView, but it will not be very useful unless the desired coordinate system is geographic. To convert the image into the state plane system (or other appropriate coordinate system), the image must be converted into a grid, which is reprojected and then converted back to an image. The steps required in ARC/INFO are listed here:

- Arc: imagegrid file.tif gridname colors
- Arc: grid
- Grid: display 9999
- Grid: mape gridname
- Grid: gridpaint gridname
- Grid: setwindow \* (use this command to cut unnecessary portions of the image)
- Grid: newgridname = gridname
- Grid: quit
- Arc: project grid newgridname newimagename
- Arc: gridimage newimagename colors newimagename.tif TIFF

The USGS map was the only regional coverage obtained for Marcus Hook that was not free. An alternative to using a USGS map is to obtain an orthophotograph of the regional area. These orthophotographs are also available on the Internet (usually for free) and typically have a pixel resolution of 1 - 3 meters.

## **VEMAP Climate Data**

The VEMAP climate data also require some special modifications in order to view the output in ArcView. The data are available in SVF format, which consists of two header lines followed by an integer array. Separate files exist for each month and for each parameter (e.g., precipitation, solar radiation). There are 5,520 values in each file, and each value corresponds to a grid cell that covers part of the United States. In order to incorporate this information in a coverage, the file values must be manipulated with Excel and then joined within ArcView to an ARC/INFO created coverage. The first step is to create the coverage in ARC/INFO as documented here:

- Arc: generate coveragename
- Generate: fishnet



- Fishnet Origin Coordinate (X,Y): -124.5,25.0
- Y-Axis Coordinate (X,Y): -124.5,49.0
- Cell Size (Width,Height): 0.5,0.5
- Number of Rows, Columns: 48,115
- Generate: quit
- Arc: build coveragename poly

These parameters originate from the VEMAP website. The resulting coverage is a polygon coverage in the form of a grid that covers the U.S. The grid cells are 0.5° longitude x 0.5° latitude. Because it is created using geographic coordinates, it must be reprojected into the coordinate system chosen for the particular project.

The problem with the created coverage is that the grid cells are labeled beginning with the lower-left corner of the grid while the SVF files are arranged so that the integer array begins at the upper-left corner of the grid. In order to join the values in the SVF files to their appropriate grid cell, the original files must be manipulated to obtain a string of values that is linked to identification numbers. The simplest way to do this is to use Excel. For each file, the headers should be removed and the values should be imported into Excel. Once in Excel, the matrix of values must be transposed (using the Cut and Paste Special commands), and then each column (beginning with the leftmost) must be cut and pasted beneath the next adjacent column. This process will produce one column of values beginning with the first value in the created coverage. This procedure of cutting and pasting can be tedious, especially with a different file for each parameter. To simplify the process, a macro can be recorded so that the procedure only needs to be run once. After creating one column of values for each parameter that is desired in the resulting coverage, a single table should be created that contains a column with labels from 1 to 5520 and a column for each climatic parameter. This table can be saved as a dBase file (\*.dbf), added to an ArcView project, and then joined to the ARC/INFO created polygon coverage. The final process step, which is not necessarily required, is to cut the coverage to the selected region of study instead of for the entire U.S. This step can be accomplished by selecting the desired grid cells with the select tool and then converting the coverage to a shapefile. Only the selected cells will be exported into the shapefile, and any joined data will become a permanent part of the theme.

### **No Data Cells**

The most effective way to incorporate sinks into a surface water runoff model is to revise a DEM by modifying cell values. If a grid cell is assigned a value of NO DATA, the watershed delineator programs will allow water to outlet to that cell. One way to incorporate no data cells is to use ARCTOOLS within ARC/INFO. However, another way is to create a point coverage of

sinks in ArcView, merge this coverage with the DEM, and then assign the no data cells using GRID in ARC/INFO. The procedure is as follows:

1. Create a shapefile of sinks from scratch or from an existing coverage.
2. Add a column to the attribute table of the shapefile and populate it with a number value that is higher than the highest elevation in the original DEM (same number for each sink).
3. Convert the shapefile to a grid using the created column as the field for the grid cell values.
4. Merge the created drain grid with the existing dem (a script, drainer.ave, is provided in Appendix G)
5. Start ARC/INFO and type grid.
6. Enter the following command: `outgrid = setnull (ingrid > x, ingrid)` where ingrid is the merged DEM and x is the some value between the highest elevation point in the original DEM and the arbitrarily assigned value in step 2. Outgrid can then be used as the starting DEM in the surface water modeling process.

Alternatively, the final three steps can be performed completely within ArcView through an Avenue script with the following lines:

- `grid1 = (SinkGrid.isnull).con(1.asgrid, 0.asgrid)`
- `newgrid = DEMGrid/grid1`

where SinkGrid is the grid created from the point coverage of sinks and DEMGrid is the original DEM.

### Probability Equation Derivation

The probability of  $x$  occurrences among  $n$  trials in a Bernoulli sequence is given by the binomial probability mass function as follows (Ang and Tang, 1975):

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

For uncertain  $p$ ,

$$P(X = x) = \int_0^1 P(X = x) f_p(p) dp$$

$$f_p(p) = \frac{1}{B(q, r)} p^{q-1} (1 - p)^{r-1} \quad (\text{Beta Function})$$

$$\therefore P(X = x) = \int_0^1 \binom{n}{x} p^x (1 - p)^{n-x} \frac{1}{B(q, r)} p^{q-1} (1 - p)^{r-1} dp$$

$$P(X = x) = \binom{n}{x} \frac{1}{B(q, r)} \int_0^1 p^{x+q-1} (1 - p)^{n-x+r-1} dp$$

$$P(X = x) = \binom{n}{x} \frac{B(x+q+1, n-x+r-1)}{B(q, r)}$$

Through the relation  $B(q, r) = \frac{\Gamma(q)\Gamma(r)}{\Gamma(q+r)}$ ,

$$P(X = x) = \binom{n}{x} \frac{\Gamma(q+r)}{\Gamma(q)\Gamma(r)} \frac{\Gamma(x+q)\Gamma(n-x+r)}{\Gamma(q+r+n)}$$

For  $x = 0$ ,

$$P(X = 0) = \frac{\Gamma(q+r)\Gamma(n+r)}{\Gamma(r)\Gamma(q+r+n)}$$

## APPENDIX F

### COC TRANSPORT EXTENSION

#### Documentation

COC Transport is an ArcView extension that enhances surface water modeling to include concentration calculations. Most of the existing hydrology applications for ArcView are currently focussed on delineating water drainage patterns using digital elevation models. These applications calculate water flow based on the paths of steepest descent (i.e., water flows from high head to low head). Some example applications include delineating streams and watersheds, and preparing input files for external programs (e.g., HMS). COC Transport provides an additional function for these models of computing downstream concentrations from known inputs to the system. The conservation of mass principle requires that the sum of inputs to a system must equal the total output. Assuming the losses in the system are minimal, the sum of the input loads equals the total output load:

- Load ( $W$ ) = Flow ( $Q$ ) \* Concentration ( $c$ ) [1]
- $\Sigma (Q_{in} * c_{in}) = Q_{out} * c_{out}$  [2]
- $c_{out} = \Sigma (W_{in}) / Q_{out}$  [3]

Therefore, if the inputs to the system are known, then a downstream concentration can be calculated. The resulting concentration at each point downstream is a conservative approximation because it does not account for any chemical decay or other losses, such as sorption.



The extension can incorporate inputs from both point sources, such as a treatment plant outfall, and area sources, such as runoff flow from farmland. The user has the option to create point sources, area sources, or both. The user can also run the extension with a custom input theme. Once the inputs are defined, the program will proceed through the steps necessary to calculate concentrations. The result is a grid of concentrations, where each grid cell is treated as an outlet, and the concentration is computed from equation 3.

The only requirements for the extension are Spatial Analyst, a flow direction grid, and knowledge of the system inputs. The following sections detail the individual process steps. If you have any questions, please contact the author. The actual extension can be downloaded from [www.ce.utexas.edu/prof/maidment/grad/romanek/research/data.html](http://www.ce.utexas.edu/prof/maidment/grad/romanek/research/data.html).

## Process Steps

1. *Set Analysis Extent*: This is a required step that determines the cell size and extent for all the grids that will be created. The user needs to set the Analysis Extent and Analysis Cell Size to **Same as Flow Direction**, where flow direction is the name of your flow direction grid. The flow direction grid is chosen so that the flow and load grids will accumulate properly since the flow direction grid is the foundation for the accumulation process. Once the analysis extent has been set, it will remain the same and does not need to be changed unless a process separate from the extension is run.

Note: Either one of the following two steps or both can be run depending on your system inputs. The methods used to calculate concentrations vary depending on the type of sources you have. Thus, many of the later process steps will ask for which type of sources are being used so that the extension knows what section of code to run. The user can skip the next two steps if he/she has already created an input source theme. However, the input source theme must contain the following three fields: Concentration, Flow, and Load.

2. *Point Sources*: Creates a shapefile of points, where each point has a predefined concentration or load, and flow. The first part of the program simply creates the shapefile. The user must then use the special blue diamond tool  and click on point source locations. The user has the option to use either a concentration based or load based point source. The user must enter a concentration [mg/L] if concentration based is chosen and a load [mg/d] if load based is chosen. Both options also prompt the user for a location description and a flow [ft<sup>3</sup>/s]. If the flow is a result of runoff, simply click on OK to proceed (the runoff flow will be incorporated later). If a flow is given, then the program will calculate via equation 1, either concentration or load, whichever parameter was not defined. Otherwise, the field records are left blank. As many points as necessary can be added. When finished, simply reply No to the question, "Would you like to add another point?"
3. *Area Sources*: Creates a shapefile of areas, where each area has a predefined concentration or load, and flow. The procedure is identical to that for *Point Sources*, but the user must now use the special green diamond tool  to create areas. The green diamond tool is similar to

the polygon tool, so each mouse click will add a vertex to the polygon. Double click on the final vertex to complete the shape.

4. *Concentration Grid*: Creates a grid of concentrations from a previously created point source theme, area source theme, or both a point and area source theme. The resulting grid will contain no data cells everywhere except where there is a point or area. In these locations, grid cell values are determined by the concentration value from the input theme(s). Note that when using both point and area sources, a point source will always have precedent over an area source. For example, if a point source lies within an area, the concentration grid will have a value equal to the concentration of the area everywhere within the area except at the point source, where the value will equal the concentration of the point source. The resulting concentration grid will not include any concentrations that are the result of a load with runoff flow. These loads will be incorporated in the *Load Grid* step.
5. *Flow Grid*: Creates a flow grid using flows from the input source theme(s) and runoff. The flow grid values will be equal to the values from the runoff grid everywhere except where there is a point or area source with a defined flow. In these areas, the flow grid will have values equal to the flows defined in the input theme(s). The user can use a previously created runoff grid (must have the same extent and cell size as the flow direction grid) or he/she has the option to create a runoff grid [in/yr]. This runoff grid will have a user-defined, constant value everywhere. Simpler models might not need to incorporate runoff. In these circumstances, the runoff grid value can be set to 0.
6. *Accumulated Flow Grid*: Creates an accumulated flow grid using a flow direction grid. For each grid cell, the program sums all of the upstream flow. The flow direction grid provides the network of which cells are upstream of others.
7. *Load Grid*: Creates a grid of loads from the concentration grid, flow or accumulated flow grid, and input source theme(s). The program multiplies the concentration grid by a flow grid to compute load. This grid is then merged with the point source theme, area source theme, or both to create a load grid that has the value of the multiplied concentration and flow grids everywhere except where there is a point or area with a predefined load. Thus, this step incorporates those points or areas in which a load was defined with runoff flow. If the input

theme is a point source theme, the program multiplies the concentration grid by the accumulated flow grid. However, if the input theme is an area source theme, the program multiplies the concentration grid by the flow grid. The assumption is that for a point source, the flow should be equal to any input flow plus any upstream flow (e.g., runoff). For example, if a point source is a runoff sampling location, the flow should be equal to the accumulated flow at that point, and the concentration should be equal to the value observed at that point. For area sources, the concentration is assumed to be the same value throughout the area, so the flow used in the calculation of load for each grid cell should simply be the flow for that grid cell. To enhance your understanding of the methodology, envision a drainage area. This drainage area could be modeled as an area source or as a point source at the outlet.

8. *Accumulated Load Grid*: Creates an accumulated load grid using a flow direction grid. The process is identical to that used in the *Accumulated Flow Grid* step except that the calculated load grid is used.
9. *Predicted Concentration Grid*: Creates a predicted concentration grid from the accumulated flow and load grids. The predicted concentration grid is computed by dividing the accumulated load grid by the accumulated flow grid. Essentially, for each grid cell, the sum of the loads to that point is divided by the sum of the flows to that point (equation 3).

## Extension Scripts

### COCTransport.Points

```
theView = av.GetActiveDoc
```

```
If a theme in the view is being edited, Stop Editing it before creating new theme
editThm = theView.GetEditableTheme
if (editThm <> nil) then
  doSave = MsgBox.YesNoCancel("Save edits to "+editThm.GetName+"?", "Stop
Editing", true)
  if (doSave = nil) then
    return nil
  end
  if (editThm.StopEditing(doSave).Not) then
    MsgBox.Info("Unable to Save Edits to "
      + editThm.GetName +
      ", please use the Save Edits As option", "")
    return nil
  else
    theView.SetEditableTheme(NIL)
  end
end
```

```
'Define theme
class = Point
def = av.GetProject.MakeFileName("ptsrc", "shp")
def = FileDialog.Put(def, "*.shp", "New Theme")
```

```
'Create attribute table
if (def <> nil) then
  tbl = FTab.MakeNew(def, class)
  if (tbl.HasError) then
    if (tbl.HasLockError) then
      MsgBox.Error("Unable to acquire Write Lock for file " + def.GetBaseName,
        "")
    else
      MsgBox.Error("Unable to create " + def.GetBaseName, "")
    end
  end
  return nil
end
```

```
fld = Field.Make("ID", #FIELD_DECIMAL, 5, 0)
desc = Field.Make("Description", #FIELD_CHAR, 50, 0)
conc = Field.Make("Concentration", #FIELD_DECIMAL, 10, 2)
flow = Field.Make("Flow", #FIELD_DECIMAL, 18, 6)
load = Field.Make("Load", #FIELD_DECIMAL, 18, 2)
fld.SetVisible(TRUE)
desc.SetVisible(TRUE)
conc.SetVisible(TRUE)
flow.SetVisible(TRUE)
load.SetVisible(TRUE)
tbl.AddFields({ fld, desc, conc, flow, load})
tbl.SetEditable(FALSE)
end
```

```
'Create theme and add to view
theTheme = FTheme.Make(tbl)
theView.AddTheme(theTheme)
theTheme.SetActive(TRUE)
theTheme.SetVisible(TRUE)
theTheme.SetName("Point Sources")
theView.SetEditableTheme(theTheme)
av.GetProject.SetModified(TRUE)
MsgBox.Info("Use the special blue diamond tool to add point
sources.", "Application Info")
```



## COCTransport.PointTool

```
theView = av.GetActiveDoc
```

```
' Create point and populate attribute table
pt = theView.GetDisplay.ReturnUserPoint
theTheme = theView.GetEditableTheme
if (theTheme <> nil) then
  theList = {"Concentration Based", "Load Based"}
  theChoice = MsgBox.ChoiceAsString(theList, "Choose the type of point
coverage:", "Coverage Selection")
  thePrj = theView.GetProjection
  if (thePrj.IsNull.Not) then
    pt = pt.ReturnUnprojected(thePrj)
  end
  theField = theTheme.GetFTab.FindField("Shape")
  theIDField = theTheme.GetFTab.FindField("ID")
  theDescField = theTheme.GetFTab.FindField("Description")
  theConcField = theTheme.GetFTab.FindField("Concentration")
  theFlowField = theTheme.GetFTab.FindField("Flow")
  theLoadField = theTheme.GetFTab.FindField("Load")
  theTheme.GetFTab.BeginTransaction
  theNumber = theTheme.GetFTab.GetNumRecords
  rec = theTheme.GetFTab.AddRecord
  theTheme.GetFTab.SetValue(theField, rec, pt)
  theTheme.GetFTab.SetValue(theIDField, rec, theNumber + 1)
  n = MsgBox.Input("Enter a location description:", "Location Input", "")
  theTheme.GetFTab.SetValue(theDescField, rec, n)
  if (theChoice = "Concentration Based") then
    c = MsgBox.Input("Enter a Concentration Value [mg/L]:", "Concentration
Input", "")
    f = MsgBox.Input("Enter a Flow [ft^3/d] or click OK to use runoff value:", "Flow
Input", "0")
    conc = c.AsNumber
    flow = f.AsNumber
    theTheme.GetFTab.SetValue(theConcField, rec, conc)
    theTheme.GetFTab.SetValue(theFlowField, rec, flow)
    if (flow <> 0) then
      theTheme.GetFTab.SetValue(theLoadField, rec, conc*flow*28.31682)
    else
      theTheme.GetFTab.SetValue(theLoadField, rec, nil)
      theTheme.GetFTab.SetValue(theFlowField, rec, nil)
```

```
end
else
  l = MsgBox.Input("Enter a Load [mg/d]:", "Load Input", "")
  f = MsgBox.Input("Enter a Flow [ft^3/d] or click OK to use runoff value:", "Flow
Input", "0")
  load = l.AsNumber
  flow = f.AsNumber
  theTheme.GetFTab.SetValue(theLoadField, rec, load)
  theTheme.GetFTab.SetValue(theFlowField, rec, flow)
  if (flow <> 0) then
    theTheme.GetFTab.SetValue(theConcField, rec, load/flow/28.31682)
  else
    theTheme.GetFTab.SetValue(theConcField, rec, nil)
    theTheme.GetFTab.SetValue(theFlowField, rec, nil)
  end
end
theTheme.GetFTab.EndTransaction
theTheme.GetFTab.GetSelection.ClearAll
theTheme.GetFTab.GetSelection.Set(rec)
theTheme.GetFTab.UpdateSelection
else
  gp = GraphicShape.Make(pt)
  theView.GetGraphics.UnselectAll
  gp.SetSelected(TRUE)
  theView.GetGraphics.Add(gp)
end

' Continue adding points?
theEnd = MsgBox.YesNo("Would you like to add another point?", "More
Points", TRUE)
if (theEnd = false) then
  theTheme.StopEditing(True)
  theTheme.ClearSelection
  exit
end
av.GetProject.SetModified(true)
```

## COCTransport.Areas

```
theView = av.GetActiveDoc
```

```
If a theme in the view is being edited, Stop Editing it before creating new theme
editThm = theView.GetEditableTheme
if (editThm <> nil) then
  doSave = MsgBox.YesNoCancel("Save edits to "+editThm.GetName+"?", "Stop
Editing", true)
  if (doSave = nil) then
    return nil
  end
  if (editThm.StopEditing(doSave).Not) then
    MsgBox.Info("Unable to Save Edits to "
      + editThm.GetName +
      ", please use the Save Edits As option", "")
    return nil
  end
  theView.SetEditableTheme(NIL)
end
```

```
'Define theme
class = Polygon
def = av.GetProject.MakeFileName("arsrc", "shp")
def = FileDialog.Put(def, "*.shp", "New Theme")
```

```
'Create attribute table
if (def <> nil) then
  tbl = FTab.MakeNew(def, class)
  if (tbl.HasError) then
    if (tbl.HasLockError) then
      MsgBox.Error("Unable to acquire Write Lock for file " + def.GetBaseName,
        "")
    else
      MsgBox.Error("Unable to create " + def.GetBaseName, "")
    end
    return nil
  end
  fld = Field.Make("ID", #FIELD_DECIMAL, 5, 0)
  desc = Field.Make("Description", #FIELD_CHAR, 50, 0)
  conc = Field.Make("Concentration", #FIELD_DECIMAL, 10, 2)
```

```
flow = Field.Make("Flow", #FIELD_DECIMAL, 18, 6)
load = Field.Make("Load", #FIELD_DECIMAL, 18, 2)
fld.SetVisible(TRUE)
desc.SetVisible(TRUE)
conc.SetVisible(TRUE)
flow.SetVisible(TRUE)
load.SetVisible(TRUE)
tbl.AddFields({ fld, desc, conc, flow, load})
tbl.SetEditable(FALSE)
theTheme = FTheme.Make(tbl)
end
```

```
'Create theme and add to view
theView.AddTheme(theTheme)
theTheme.SetActive(TRUE)
theTheme.SetVisible(TRUE)
theTheme.SetName("Area Sources")
theView.SetEditableTheme(theTheme)
av.GetProject.SetModified(TRUE)
MsgBox.Info("Use the special green diamond tool to add area sources as
polygons.", "Application Info")
```

## COCTransport.AreaTool

```
theView = av.GetActiveDoc
```

```
' Create areas and populate attribute table
```

```
p = theView.ReturnUserPolygon
```

```
theTheme = theView.GetEditableTheme
```

```
if (p.IsNotNull) then
```

```
    return nil
```

```
else
```

```
    if (theTheme <> nil) then
```

```
        theList = { "Concentration Based", "Load Based" }
```

```
        theChoice = MsgBox.ChoiceAsString(theList, "Choose the type of area
```

```
coverage:", "Coverage Selection")
```

```
        theTheme.GetFTab.BeginTransaction
```

```
        thePrj = theView.GetProjection
```

```
        if (thePrj.IsNotNull.Not) then
```

```
            p = p.ReturnUnprojected(thePrj)
```

```
        end
```

```
        theField = theTheme.GetFTab.FindField("Shape")
```

```
        theIDField = theTheme.GetFTab.FindField("ID")
```

```
        theDescField = theTheme.GetFTab.FindField("Description")
```

```
        theConcField = theTheme.GetFTab.FindField("Concentration")
```

```
        theFlowField = theTheme.GetFTab.FindField("Flow")
```

```
        theLoadField = theTheme.GetFTab.FindField("Load")
```

```
        theNumber = theTheme.GetFTab.GetNumRecords
```

```
        rec = theTheme.GetFTab.AddRecord
```

```
        theTheme.GetFTab.SetValue(theField,rec,p)
```

```
        n = MsgBox.Input("Enter a location description:", "Location Input", "")
```

```
        theTheme.GetFTab.SetValue(theDescField,rec,n)
```

```
        if (theChoice = "Concentration Based") then
```

```
            c = MsgBox.Input("Enter a Concentration Value [mg/L]:", "Concentration
```

```
Input", "")
```

```
            f = MsgBox.Input("Enter a Flow [ft^3/d] or click OK to use runoff
```

```
value:", "Flow Input", "0")
```

```
            conc = c.AsNumber
```

```
            flow = f.AsNumber
```

```
            theTheme.GetFTab.SetValue(theConcField,rec,conc)
```

```
            theTheme.GetFTab.SetValue(theFlowField,rec,flow)
```

```
            if (flow <> 0) then
```

```
                theTheme.GetFTab.SetValue(theLoadField,rec,conc*flow*28.31682)
```

```
            else
```

```
                theTheme.GetFTab.SetValue(theLoadField,rec,nil)
```

```
                theTheme.GetFTab.SetValue(theFlowField,rec,nil)
```

```
            end
```

```
        else
```

```
            l = MsgBox.Input("Enter a Load [mg/d]:", "Load Input", "")
```

```
            f = MsgBox.Input("Enter a Flow [ft^3/d] or click OK to use runoff
```

```
value:", "Flow Input", "0")
```

```
            load = l.AsNumber
```

```
            flow = f.AsNumber
```

```
            theTheme.GetFTab.SetValue(theLoadField,rec,load)
```

```
            theTheme.GetFTab.SetValue(theFlowField,rec,flow)
```

```
            if (flow <> 0) then
```

```
                theTheme.GetFTab.SetValue(theConcField,rec,load/flow/28.31682)
```

```
            else
```

```
                theTheme.GetFTab.SetValue(theConcField,rec,nil)
```

```
                theTheme.GetFTab.SetValue(theFlowField,rec,nil)
```

```
            end
```

```
        end
```

```
        theTheme.GetFTab.GetSelection.ClearAll
```

```
        theTheme.GetFTab.GetSelection.Set(rec)
```

```
        theTheme.GetFTab.UpdateSelection
```

```
        theTheme.GetFTab.EndTransaction
```

```
    else
```

```
        gp = GraphicShape.Make(p)
```

```
        theView.GetGraphics.UnselectAll
```

```
        gp.SetSelected(TRUE)
```

```
        theView.GetGraphics.Add(gp)
```

```
    end
```

```
end
```

```
' Continue adding areas?
```

```
theEnd = MsgBox.YesNo("Would you like to add another area?", "More
```

```
Areas", TRUE)
```

```
if (theEnd = false) then
```

```
    theTheme.StopEditing(True)
```

```
    theTheme.ClearSelection
```

```
    exit
```

```
end
```

```
av.GetProject.SetModified(true)
```

## COCTransport.ConcentrationGrid

```
' Theme Selection
theView = av.GetActiveDoc
theActThemes = theView.GetThemes
theList = { "Point Sources", "Area Sources", "Both Point and Area Sources" }
theCoverages = MsgBox.ChoiceAsString(theList, "Choose the type of coverages to
use", "Coverage Selection")
if (theCoverages = "Point Sources") then
  theTheme = MsgBox.Choice(theActThemes, "Choose the POINT source
theme", "Theme Selection")
  theFTab = theTheme.GetFTab
elseif (theCoverages = "Area Sources") then
  theTheme = MsgBox.Choice(theActThemes, "Choose the AREA source
theme", "Theme Selection")
  theFTab = theTheme.GetFTab
else
  thePointTheme = MsgBox.Choice(theActThemes, "Choose the POINT source
theme", "Theme Selection")
  theAreaTheme = MsgBox.Choice(theActThemes, "Choose the AREA source
theme", "Theme Selection")
  thePointFTab = thePointTheme.GetFTab
  theAreaFTab = theAreaTheme.GetFTab
end

' File Name
cdef = av.GetProject.MakeFileName("concgnd", "")
cFN = SourceManager.PutDataSet(GRID, "Save concentration grid to:", cdef, TRUE)
if (cFN = NIL) then
  return NIL
end

' Set extent and cell size for conversion if not already set
ae = theView.GetExtension(AnalysisEnvironment)
box = Rect.Make(0@0, 1@1)
cellSize = 1
if ((ae.GetExtent(box) <> #ANALYSEENV_VALUE) or (ae.GetCellSize(cellSize)
<> #ANALYSEENV_VALUE)) then
  ce = AnalysisPropertiesDialog.Show(theView, TRUE, "Conversion Extent:" ++
theTheme.GetName)
  if (ce = NIL) then
    return NIL
```

266

```
end
ce.GetCellSize(cellSize)
ce.GetExtent(box)
end

if ((theCoverages = "Point Sources") or (theCoverages = "Area Sources")) then

' Obtain field and modify theme table (points or areas)
cField = theFTab.FindField("Concentration")
if (cField = NIL) then
  MsgBox.Info("There is no concentration field for this coverage.", "Error")
  exit
end
theFTab.SetEditable(TRUE)
outfield = List.Make
outfield.Add(Field.Make("Dummy", #FIELD_DECIMAL, 18, 2))
outfieldc = outfield.DeepClone
theFTab.AddFields(outfieldc)
dField = theFTab.FindField("Dummy")
for each tabrec in theFTab
  thecNum = theFTab.ReturnValue(cField, tabrec)
  if (thecNum <> nil) then
    theFTab.SetValue(dField, tabrec, thecNum)
  else
    theFTab.SetValue(dField, tabrec, nil)
  end
end
theFTab.SetEditable(FALSE)
ae.GetCellSize(cellSize)

' CONCENTRATION GRID Conversion (points or areas)
thePrj = theView.GetProjection
theCGrid = Grid.MakeFromFTab(theFTab, thePrj, dField, { cellSize, box })
if (theCGrid.HasError) then
  MsgBox.Error(theTheme.GetName ++ "could not be converted to a
concentration grid", "Conversion Error")
  return NIL
end
theCGrid.SaveDataSet(cFN)
concthm = GTheme.Make(theCGrid)
theFTab.SetEditable(TRUE)
theFTab.RemoveFields(outfieldc)
```

```

theFTab.SetEditable(FALSE)
else
theAreaFTab.SetValue(dField2,tabrec,nil)
end
end
theAreaFTab.SetEditable(FALSE)
ae.GetCellSize(cellSize)
' CONCENTRATION GRID Conversion (points and areas)
thePrj = theView.GetProjection
thePointCGrid = Grid.MakeFromFTab(thePointFTab,thePrj,dField1,{ cellSize,
box })
if (thePointCGrid.HasError) then
MsgBox.Error(thePointTheme.GetName ++ "could not be converted to a
grid", "Conversion Error")
return NIL
end
theAreaCGrid = Grid.MakeFromFTab(theAreaFTab,thePrj,dField2,{ cellSize,
box })
if (theAreaCGrid.HasError) then
MsgBox.Error(theAreaTheme.GetName ++ "could not be converted to a
grid", "Conversion Error")
return NIL
end
ListGrid = { theAreaCGrid }
ConcGrid = thePointCGrid.Merge(ListGrid)
ConcGrid.SaveData.Set(cFN)
concthm = GTheme.Make(ConcGrid)
thePointFTab.SetEditable(TRUE)
thePointFTab.RemoveFields(outfieldc1)
thePointFTab.SetEditable(FALSE)
theAreaFTab.SetEditable(TRUE)
theAreaFTab.RemoveFields(outfieldc2)
theAreaFTab.SetEditable(FALSE)
end
' Add Flow Grid to View
theView.AddTheme(concthm)
concthm.SetName("Concentration Grid (mg/L)")
MsgBox.Info("Any loads that were assigned with runoff flow do not appear in this
grid and will be incorporated in the create load grid step.", "NOTE")

```

---

```

theFTab.SetEditable(FALSE)
else
' Obtain field and modify theme table (points)
cField1 = thePointFTab.FindField("Concentration")
if (cField1 = NIL) then
MsgBox.Info("There is no concentration field for the point coverage.", "Error")
exit
end
thePointFTab.SetEditable(TRUE)
outfield1 = List.Make
outfield1.Add(Field.Make("Dummy",#FIELD_DECIMAL,18,2))
outfieldc1 = outfield1.DeepClone
thePointFTab.AddFields(outfieldc1)
dField1 = thePointFTab.FindField("Dummy")
for each tabrec in thePointFTab
theNum1 = thePointFTab.ReturnValue(cField1,tabrec)
if (theNum1 <> nil) then
thePointFTab.SetValue(dField1,tabrec,theNum1)
else
thePointFTab.SetValue(dField1,tabrec,nil)
end
end
thePointFTab.SetEditable(FALSE)
' Obtain field and modify theme table (areas)
cField2 = theAreaFTab.FindField("Concentration")
if (cField2 = NIL) then
MsgBox.Info("There is no concentration field for the area coverage.", "Error")
exit
end
theAreaFTab.SetEditable(TRUE)
outfield2 = List.Make
outfield2.Add(Field.Make("Dummy",#FIELD_DECIMAL,18,2))
outfieldc2 = outfield2.DeepClone
theAreaFTab.AddFields(outfieldc2)
dField2 = theAreaFTab.FindField("Dummy")
for each tabrec in theAreaFTab
theNum2 = theAreaFTab.ReturnValue(cField2,tabrec)
if (theNum2 <> 0) then
theAreaFTab.SetValue(dField2,tabrec,theNum2)

```

## COCTransport.FlowGrid

```
' Theme Selection
theView = av.GetActiveDoc
theActThemes = theView.GetThemes
theList = { "Point Sources", "Area Sources", "Both Point and Area Sources" }
theCoverages = MsgBox.ChoiceAsString(theList, "Choose the type of coverages to
use", "Coverage Selection")
if (theCoverages = "Point Sources") then
  theTheme = MsgBox.Choice(theActThemes, "Choose the POINT source
theme", "Theme Selection")
  theFTab = theTheme.GetFTab
elseif (theCoverages = "Area Sources") then
  theTheme = MsgBox.Choice(theActThemes, "Choose the AREA source
theme", "Theme Selection")
  theFTab = theTheme.GetFTab
else
  thePointTheme = MsgBox.Choice(theActThemes, "Choose the POINT source
theme", "Theme Selection")
  theAreaTheme = MsgBox.Choice(theActThemes, "Choose the AREA source
theme", "Theme Selection")
  thePointFTab = thePointTheme.GetFTab
  theAreaFTab = theAreaTheme.GetFTab
end

' Set extent and cell size for conversion if not already set
ae = theView.GetExtension(AnalysisEnvironment)
box = Rect.Make(0@0,1@1)
cellSize = 1
if ((ae.GetExtent(box) <> #ANALYSENV_VALUE) or (ae.GetCellSize(cellSize)
<> #ANALYSENV_VALUE)) then
  ce = AnalysisPropertiesDialog.Show(theView, TRUE, "Conversion Extent:" ++
theTheme.GetName)
  if (ce = NIL) then
    return NIL
  end
  ce.GetCellSize(cellSize)
  ce.GetExtent(box)
end
ae.GetCellSize(cellSize)

' Grid Selection
```

208

```
theQuestion = MsgBox.YesNo("Would you like to create a runoff grid (Choose No
if you have already created a runoff grid [in/yr] that has the same dimensions as
your flow direction grid)?", "Runoff Grid", TRUE)
if (theQuestion = FALSE) then
  RunoffGrid = MsgBox.Choice(theActThemes, "Choose the predefined runoff
grid", "Runoff Selection")
else
  theRunoff = MsgBox.Input("Enter a constant runoff value [in/yr]:", "Runoff
Creation", "20").AsNumber
  RunoffGrid = (cellSize*cellSize*theRunoff*0.0002283).AsGrid
end

' File Names
fdef = av.GetProject.MakeFileName("flowgrd", "")
fFN = SourceManager.PutDataSet(GRID, "Save flow grid to:", fdef, TRUE)
if (fFN = NIL) then
  return NIL
end

if ((theCoverages = "Point Sources") or (theCoverages = "Area Sources")) then

' Obtain field to convert with (points or areas)
fField = theFTab.FindField("Flow")
if (fField = NIL) then
  MsgBox.Info("There is no flow field for this coverage.", "Error")
  exit
end
theFTab.SetEditable(TRUE)
outfield = List.Make
outfield.Add(Field.Make("Dummy", #FIELD_DECIMAL, 18, 2))
outfieldc = outfield.DeepClone
theFTab.AddFields(outfieldc)
dField = theFTab.FindField("Dummy")
for each tabrec in theFTab
  thefNum = theFTab.ReturnValue(fField, tabrec)
  if (thefNum <> nil) then
    theFTab.SetValue(dField, tabrec, thefNum)
  else
    theFTab.SetValue(dField, tabrec, nil)
  end
end
theFTab.SetEditable(FALSE)
```

```

' FLOW GRID Conversion (points or areas)
thePrj = theView.GetProjection
theFGrid = Grid.MakeFromFTab(theFTab,thePrj,dField,{cellSize, box})
if (theFGrid.HasError) then
    MsgBox.Error(theTheme.GetName ++ "could not be converted to a flow
grid", "Conversion Error")
return NIL
end
ListGrid = {RunoffGrid}
FlowGrid = theFGrid.Merge(ListGrid)
FlowGrid.SaveDataSet(FFN)
flowThm = GTheme.Make(FlowGrid)
theFTab.SetEditable(TRUE)
theFTab.RemoveFields(outfieldc)
theFTab.SetEditable(FALSE)
else
end

' Obtain field to convert with (points)
ifField1 = thePointFTab.FindField("Flow")
if (ifField1 = NIL) then
    MsgBox.Info("There is no flow field for the point coverage.", "Error")
exit
end
thePointFTab.SetEditable(TRUE)
outfield1 = List.Make
outfield1.Add(Field.Make("Dummy",#FIELD_DECIMAL,18,2))
thePointFTab.AddFields(outfieldc1)
dField1 = thePointFTab.FindField("Dummy")
for each tabrec in thePointFTab
    thefNum1 = thePointFTab.ReturnValue(ffield1,tabrec)
    if (thefNum1 <> nil) then
        thePointFTab.SetValue(dField1,tabrec,thefNum1)
    else
        thePointFTab.SetValue(dField1,tabrec,nil)
    end
end

' Obtain field to convert with (areas)
ifField2 = theAreaFTab.FindField("Flow")
if (ifField2 = NIL) then
    MsgBox.Info("There is no flow field for the area coverage.", "Error")
exit
end
theAreaFTab.SetEditable(TRUE)
outfield2 = List.Make
outfield2.Add(Field.Make("Dummy",#FIELD_DECIMAL,18,2))
theAreaFTab.AddFields(outfieldc2)
dField2 = theAreaFTab.FindField("Dummy")
for each tabrec in theAreaFTab
    thefNum2 = theAreaFTab.ReturnValue(ffield2,tabrec)
    if (thefNum2 <> nil) then
        theAreaFTab.SetValue(dField2,tabrec,thefNum2)
    else
        theAreaFTab.SetValue(dField2,tabrec,nil)
    end
end
theAreaFTab.SetEditable(FALSE)

' FLOW GRID Conversion (points and areas)
thePrj = theView.GetProjection
thePointFGrid = Grid.MakeFromFTab(thePointFTab,thePrj,dField1,{cellSize,
box})
if (thePointFGrid.HasError) then
    MsgBox.Error(theTheme.GetName ++ "could not be converted to a
grid", "Conversion Error")
return NIL
end
theAreaFGrid = Grid.MakeFromFTab(theAreaFTab,thePrj,dField2,{cellSize,
box})
if (theAreaFGrid.HasError) then
    MsgBox.Error(theTheme.GetName ++ "could not be converted to a
grid", "Conversion Error")
return NIL
end
ListGrid1 = {RunoffGrid}
ListGrid2 = {theAreaFGrid}
FlowGrid1 = thePointFGrid.Merge(ListGrid2)
FlowGrid2 = FlowGrid1.Merge(ListGrid1)
FlowGrid2.SaveDataSet(FFN)

```

```

flowthm = GTheme.Make(FlowGrid2)
thePointFTab.SetEditable(TRUE)
thePointFTab.RemoveFields(outfieldc1)
thePointFTab.SetEditable(FALSE)
theAreaFTab.SetEditable(TRUE)
theAreaFTab.RemoveFields(outfieldc2)
theAreaFTab.SetEditable(FALSE)

end

' Add Flow Grid to View
theView.AddTheme(flowthm)
flowthm.SetName("Flow Grid (ft^3/d)")

```

270

## COCTransport.FlowAccumulation

```

' Get Themes
theView = av.GetActiveDoc
theThemes = theView.GetThemes
if (theThemes.Count = 0) then
  MsgBox.Error("No themes found", "Error")
  exit
end
theGThemes = List.Make
for each themerec in theThemes
  if (themerec.GetClass.GetClassName = "gtheme") then
    theGThemes.Add(themerec)
  end
end
if (theGThemes.Count = 0) then
  MsgBox.Error("No grid themes found", "Error")
  exit
end
fdrtheme = MsgBox.ListAsString(theGThemes, "Choose the FLOW DIRECTION
grid:", "Flow Direction Selection")
if (fdrtheme = nil) then
  exit
end
flowtheme = MsgBox.ListAsString(theGThemes, "Choose the FLOW grid to
accumulate:", "Flow Grid Selection")
if (flowtheme = nil) then
  exit
end

' File Names
flacdef = av.GetProject.MakeFileName("flowacc", "")
flFN = SourceManager.PutDataSet(GRID, "Save accumulated flow grid
to:", flacdef, TRUE)
if (flFN = NIL) then
  return NIL
end

' Calculate Flow Accumulation
fdrgrid = fdrtheme.GetGrid
flowgrid = flowtheme.GetGrid
flacgrid = fdrgrid.FlowAccumulation(flowgrid)

```



```

flaggrid.SaveDataSet(flFN)
flagtheme = Gtheme.Make(flaggrid)
theView.AddTheme(flagtheme)
flagtheme.SetVisible(false)
flagtheme.SetName("Flow Accumulation (ft^3/d)")

```

## COCTransport.LoadGrid

```

' Theme Selection
theView = av.GetActiveDoc
theActThemes = theView.GetThemes
theList = { "Point Sources", "Area Sources", "Both Point and Area Sources" }
theCoverages = MsgBox.ChoiceAsString(theList, "Choose the type of coverages to
use", "Coverage Selection")
theGThemes = List.Make
for each themerec in theActThemes
  if (themerec.GetClass.GetClassName = "gtheme") then
    theGThemes.Add(themerec)
  end
end
if (theGThemes.Count = 0) then
  MsgBox.Error("No grid themes found", "Error")
  exit
end
if (theCoverages = "Point Sources") then
  theTheme = MsgBox.Choice(theActThemes, "Choose the POINT source
theme", "Theme Selection")
  theFTab = theTheme.GetFTab
  conctheme = MsgBox.ListAsString(theGThemes, "Choose the
CONCENTRATION grid:", "Concentration Grid Selection")
  if (conctheme = nil) then
    exit
  end
  flowactheme = MsgBox.ListAsString(theGThemes, "Choose the FLOW
ACCUMULATION grid:", "Flow Accumulation Grid Selection")
  if (flowactheme = nil) then
    exit
  end
elseif (theCoverages = "Area Sources") then
  theTheme = MsgBox.Choice(theActThemes, "Choose the AREA source
theme", "Theme Selection")
  theFTab = theTheme.GetFTab
  conctheme = MsgBox.ListAsString(theGThemes, "Choose the
CONCENTRATION grid:", "Concentration Grid Selection")
  if (conctheme = nil) then
    exit
  end
end

```

```

flowtheme = MsgBox.ListAsString(theGThemes,"Choose the FLOW grid:","Flow
Grid Selection")
if (flowtheme = nil) then
    exit
end
else
    thePointTheme = MsgBox.Choice(theActThemes,"Choose the POINT source
theme","Theme Selection")
    theAreaTheme = MsgBox.Choice(theActThemes,"Choose the AREA source
theme","Theme Selection")
    thePointFTab = thePointTheme.GetFTab
    theAreaFTab = theAreaTheme.GetFTab
    concntheme = MsgBox.ListAsString(theGThemes,"Choose the
CONCENTRATION grid:","Concentration Grid Selection")
    if (concntheme = nil) then
        exit
    end
end
flowtheme = MsgBox.ListAsString(theGThemes,"Choose the FLOW grid:","Flow
Grid Selection")
if (flowtheme = nil) then
    exit
end
flowwactheme = MsgBox.ListAsString(theGThemes,"Choose the FLOW
ACCUMULATION grid:","Flow Accumulation Grid Selection")
if (flowwactheme = nil) then
    exit
end
end
' File Name
ldef = av.GetProject.MakeFileName("loadgrd","")
lFN = SourceManager.PutDataSet(GRID,"Save load grid to: ",ldef,TRUE)
return NIL
end
' Set extent and cell size for conversion if not already set
ae = theView.GetExtension(AnalysisEnvironment)
box = Rect.Make(0@0,1@1)
cellSize = 1
if (ae.GetExtent(box) <> #ANALYSISISENV_VALUE) or (ae.GetCellSize(cellSize)
<> #ANALYSISISENV_VALUE)) then

```

```

ce = AnalysisPropertiesDialog.Show(theView,TRUE,"Conversion Extent: " ++
theTheme.GetName)
if (ce = NIL) then
    return NIL
end
ce.GetCellSize(cellSize)
ce.GetExtent(box)
end
if (theCoverages = "Point Sources") then
' Obtain field and modify theme table (points)
lField = theFTab.FindField("LLoad")
if (lField = NIL) then
    MsgBox.Info("There is no load field for this coverage. ", "Error")
    exit
end
theFTab.SetEditable(TRUE)
outfield = List.Make
outfieldc = outfield.DeepClone
theFTab.AddFields(outfieldc)
dField = theFTab.FindField("Dummy")
for each tabrec in theFTab
    theNum = theFTab.ReturnValue(lField,tabrec)
    if (theNum <> nil) then
        theFTab.SetValue(dField,tabrec,theNum)
    end
end
theFTab.SetEditable(FALSE)
ae.GetCellSize(cellSize)
' LOAD GRID Conversion (points)
thePrj = theView.GetProjection
theLGrid = Grid.MakeFromFTab(theFTab,thePrj.dField,{ cellSize, box })
if (theLGrid.HasError) then
    MsgBox.Error(theTheme.GetName ++ "could not be converted to a load
grid", "Conversion Error")
    return NIL
end
concngrid = concntheme.GetGrid
flowwacgrid = flowwactheme.GetGrid

```

```

LoadGrid1 = flowacgrid*concgriid*28.31682.AsGrid
ListGrid = {LoadGrid1}
LoadGrid2 = theLGrid.Merge(ListGrid)
LoadGrid2.SaveDataSet(IFN)
loadthm = GTheme.Make(LoadGrid2)
theFTab.SetEditable(TRUE)
theFTab.RemoveFields(outfieldc)
theFTab.SetEditable(FALSE)

elseif (theCovages = "Area Sources") then

' Obtain field and modify theme table (areas)
Ifield = theFTab.FindField("Load")
if (Ifield = NIL) then
  MsgBox.Info("There is no load field for this coverage.", "Error")
  exit
end
theFTab.SetEditable(TRUE)
outfield = List.Make
outfield.Add(Field.Make("Dummy", #FIELD_DECIMAL, 18, 2))
outfieldc = outfield.DeepClone
theFTab.AddFields(outfieldc)
dField = theFTab.FindField("Dummy")
for each tabrec in theFTab
  theNum = theFTab.ReturnValue(Ifield, tabrec)
  if (theNum <> nil) then
    theFTab.SetValue(dField, tabrec, theNum)
  end
end
theFTab.SetEditable(FALSE)
ae.GetCellSize(cellSize)

' LOAD GRID Conversion (areas)
thePrj = theView.GetProjection
theLGrid = Grid.MakeFromFTab(theFTab, thePrj.dField, {cellSize, box})
if (theLGrid.HasError) then
  MsgBox.Error(theTheme.GetName ++ "could not be converted to a load
grid", "Conversion Error")
  return NIL
end
concgriid = conctheme.GetGrid
flowgrid = flowtheme.GetGrid

```

---

```

LoadGrid1 = flowacgrid*concgriid*28.31682.AsGrid
ListGrid = {LoadGrid1}
LoadGrid2 = theLGrid.Merge(ListGrid)
LoadGrid2.SaveDataSet(IFN)
loadthm = GTheme.Make(LoadGrid2)
theFTab.SetEditable(TRUE)
theFTab.RemoveFields(outfieldc)
theFTab.SetEditable(FALSE)

else

' Obtain field and modify theme table (points)
Ifield1 = thePointFTab.FindField("Load")
cField1 = thePointFTab.FindField("Concentration")
if (Ifield1 = NIL) then
  MsgBox.Info("There is no load field for the point coverage.", "Error")
  exit
end
end
if (cField1 = NIL) then
  MsgBox.Info("There is no concentration field for the point coverage.", "Error")
end
thePointFTab.SetEditable(TRUE)
outfield1 = List.Make
outfield1.Add(Field.Make("Dummy", #FIELD_DECIMAL, 18, 2))
outfield1.Add(Field.Make("Dummy2", #FIELD_DECIMAL, 18, 2))
outfield1c = outfield1.DeepClone
thePointFTab.AddFields(outfield1c)
dField1 = thePointFTab.FindField("Dummy")
d2Field1 = thePointFTab.FindField("Dummy2")
for each tabrec in thePointFTab
  theNum1 = thePointFTab.ReturnValue(Ifield1, tabrec)
  thecNum1 = thePointFTab.ReturnValue(cField1, tabrec)
  if (theNum1 <> nil) then
    thePointFTab.SetValue(dField1, tabrec, theNum1)
  end
end
if (thecNum1 <> nil) then
  thePointFTab.SetValue(d2Field1, tabrec, thecNum1)
end
end
thePointFTab.SetEditable(FALSE)

' Obtain field and modify theme table (areas)

```

```

IfField2 = theAreaFTab.FindField("Load")
if (IfField2 = NIL) then
  MsgBox.Info("There is no load field for the area coverage.", "Error")
  exit
end
theAreaFTab.SetEditable(TRUE)
outfield2 = List.Make
outfield2.Add(Field.Make("Dummy", #FIELD_DECIMAL, 18, 2))
outfield2 = outfield2.DeepClone
theAreaFTab.AddFields(outfield2)
dField2 = theAreaFTab.FindField("Dummy")
for each tabrec in theAreaFTab
  theNum2 = theAreaFTab.ReturnValue(IfField2, tabrec)
  if (theNum2 <> nil) then
    theAreaFTab.SetValue(dField2, tabrec, theNum2)
  end
end
theAreaFTab.SetEditable(FALSE)
ae.GetCellSize(cellSize)

' LOAD GRID Conversion (points and areas)
thePrj = theView.GetProjection
thePointLGrid = Grid.MakeFromFTab(thePointFTab, thePrj, dField1, {cellSize,
box})
if (thePointLGrid.HasError) then
  MsgBox.Error(thePointTheme.GetName ++ "could not be converted to a
grid", "Conversion Error")
  return NIL
end
thePointCGrid = Grid.MakeFromFTab(thePointFTab, thePrj, d2Field1, {cellSize,
box})
if (thePointCGrid.HasError) then
  MsgBox.Error(thePointTheme.GetName ++ "could not be converted to a
grid", "Conversion Error")
  return NIL
end
theAreaLGrid = Grid.MakeFromFTab(theAreaFTab, thePrj, dField2, {cellSize,
box})
if (theAreaLGrid.HasError) then
  MsgBox.Error(theAreaTheme.GetName ++ "could not be converted to a
grid", "Conversion Error")
  return NIL
end

concegrid = concetheme.GetGrid
flowgrid = flowtheme.GetGrid
flowacgrid = flowachtheme.GetGrid
LoadGrid1 = flowgrid*concegrid*28.31682
LoadGrid2 = flowacgrid*thePointCGrid*28.31682.AsGrid
ListGrid1 = {LoadGrid1}
ListGrid2 = {LoadGrid2}
LoadGrid3 = theAreaLGrid.Merge(ListGrid1)
LoadGrid4 = thePointLGrid.Merge(ListGrid2)
ListGrid3 = {LoadGrid3}
LoadGrid5 = LoadGrid4.Merge(ListGrid3)
LoadGrid5..SaveDataSet(IFN)
loadthm = GTheme.Make(LoadGrid5)
thePointFTab.SetEditable(TRUE)
thePointFTab.RemoveFields(outfieldc1)
thePointFTab.SetEditable(FALSE)
theAreaFTab.SetEditable(TRUE)
theAreaFTab.RemoveFields(outfieldc2)
theAreaFTab.SetEditable(FALSE)

end

' Add Load Grid to View
theView.AddTheme(loadthm)
loadthm.setname("Load Grid (mg/d)")
end

```

## COCTransport.LoadAccumulation

```
275 ' Get Themes
theView = av.GetActiveDoc
theThemes = theView.GetThemes
if (theThemes.Count = 0) then
  MsgBox.Error("No themes found", "Error")
  exit
end
theGThemes = List.Make
for each themerec in theThemes
  if (themerec.GetClass.GetClassName = "gtheme") then
    theGThemes.Add(themerec)
  end
end
if (theGThemes.Count = 0) then
  MsgBox.Error("No grid themes found", "Error")
  exit
end
fdrtheme = MsgBox.ListAsString(theGThemes, "Choose the FLOW DIRECTION
grid:", "Flow Direction Selection")
if (fdrtheme = nil) then
  exit
end
loadtheme = MsgBox.ListAsString(theGThemes, "Choose the LOAD grid to
accumulate:", "Load Grid Selection")
if (loadtheme = nil) then
  exit
end

' File Names
ldacdef = av.GetProject.MakeFileName("loadacc", "")
ldFN = SourceManager.PutDataSet(GRID, "Save accumulated load grid
to:", ldacdef, TRUE)
if (ldFN = NIL) then
  return NIL
end

' Calculate Flow Accumulation
fdrgrid = fdrtheme.GetGrid
loadgrid = loadtheme.GetGrid
ldacgrid = fdrgrid.FlowAccumulation(loadgrid)
```

```
ldacgrid.SaveDataSet(ldFN)
ldacgtheme = Gtheme.Make(ldacgrid)
theView.AddTheme(ldacgtheme)
ldacgtheme.SetVisible(false)
ldacgtheme.SetName("Load Accumulation (mg/d)")
```

## COCTransport.Predicted

```
276 'Get Themes
theView = av.GetActiveDoc
theThemes = theView.GetThemes
if (theThemes.Count = 0) then
  MsgBox.Error("No themes found", "Error")
  exit
end
theGThemes = List.Make
for each themerec in theThemes
  if (themerec.GetClass.GetClassName = "gtheme") then
    theGThemes.Add(themerec)
  end
end
if (theGThemes.Count = 0) then
  MsgBox.Error("No grid themes found", "Error")
  exit
end
flowactheme = MsgBox.ListAsString(theGThemes, "Choose the FLOW
ACCUMULATION grid:", "Flow Accumulation Grid Selection")
if (flowactheme = nil) then
  exit
end
loadactheme = MsgBox.ListAsString(theGThemes, "Choose the LOAD
ACCUMULATION grid:", "Load Accumulation Grid Selection")
if (loadactheme = nil) then
  exit
end
flowtheme = MsgBox.ListAsString(theGThemes, "Choose the FLOW grid:", "Flow
Grid Selection")
if (flowtheme = nil) then
  exit
end
loadtheme = MsgBox.ListAsString(theGThemes, "Choose the LOAD grid:", "Load
Grid Selection")
if (loadtheme = nil) then
  exit
end
```

```
' File Names
concddef = av.GetProject.MakeFileName("predgrd", "")
pFN = SourceManager.PutDataSet(GRID, "Save concentration grid
to:", concddef, TRUE)
if (pFN = NIL) then
  return NIL
end

' Compute Concentration Grid
ldgrid = loadtheme.GetGrid
flgrid = flowtheme.GetGrid
ldacgrid = loadactheme.GetGrid
flacgrid = flowactheme.GetGrid
tempgrid = (ldgrid.IsNull).con(0.AsGrid, ldgrid)
concgriid = (tempgrid+ldacgrid)/(flgrid+flacgrid)/(28.31682.AsGrid)
concgriid.SaveDataSet(pFN)
concgtheme = Gtheme.Make(concgriid)
theView.AddTheme(concgtheme)
concgtheme.SetVisible(false)
concgtheme.SetName("Predicted Concentrations (mg/L)")
```

## **APPENDIX G**

### **AVENUE SCRIPTS**

All of these scripts were created by the author, and most of them can be downloaded from [www.ce.utexas.edu/prof/maidment/grad/romanek/research/data.html](http://www.ce.utexas.edu/prof/maidment/grad/romanek/research/data.html).

## AreaSources.ave

```
'-----  
'--- Creation Information ---  
'-----  
'  
'Name: areasources.ave  
'Version: 1.0  
'Date: April 1, 1999  
'Author: Andrew Romanek  
'      Center for Research in Water Resources  
'      The University of Texas at Austin  
'      romanek@mail.utexas.edu  
'  
'-----  
'--- Purpose/Description ---  
'-----  
'  
278 'Modifies the surface cover theme to include expected mean concentrations for the  
COC  
Transport extension. A view must be the active document.  
  
Theme selection  
theView = av.GetActiveDoc  
theThemeList = theView.GetThemes  
theTheme = MsgBox.Choice(theThemeList,"Choose the Surface Cover  
Theme","Theme Selection")  
  
'Set the table editing properties  
thefab = theTheme.GetFTab  
if (thefab.CanEdit) then  
    thefab.SetEditable(true)  
else  
    MsgBox.Info("Can't edit the theme table","Error")  
    exit  
end  
  
'Add the new field  
outfields=List.Make  
outfields.Add(Field.Make("Concentration",#FIELD_DECIMAL,10,2))  
outfields.Add(Field.Make("Flow",#FIELD_DECIMAL,18,6))
```

```
outfields.Add(Field.Make("Load",#FIELD_DECIMAL,18,2))  
outfields=List.Make  
thefab.addfields(outfields)
```

```
Define fields  
codefield = thefab.findfield("Fcode")  
conclfield = thefab.findfield("Concentration")  
flowfield = thefab.findfield("Flow")  
loadfield = thefab.findfield("Load")
```

```
EMC values  
theLabels = {"Soil", "Gravel", "Grass/Vegetation", "Streams", "River", "Tidal  
Flats", "Wetlands", "Ponds", "Structures", "Pavement"}  
theDefaults = {"0", "0", "0", "0", "0", "0", "0", "0", "0", "0"}  
theParameters = MsgBox.MultiInput("Enter the EMC Values in  
[mg/L]:", "Expected Mean Concentrations", theLabels, theDefaults)  
c1 = theParameters.Get(0).AsNumber  
c2 = theParameters.Get(1).AsNumber  
c3 = theParameters.Get(2).AsNumber  
c4 = theParameters.Get(3).AsNumber  
c5 = theParameters.Get(4).AsNumber  
c6 = theParameters.Get(5).AsNumber  
c7 = theParameters.Get(6).AsNumber  
c8 = theParameters.Get(7).AsNumber  
c9 = theParameters.Get(8).AsNumber  
c10 = theParameters.Get(9).AsNumber
```

```
'Set extent and cell size for conversion if not already set  
ae = theView.GetExtension(AnalysisEnvironment)  
box = Rect.Make(0@0,1@1)  
cellSize = 1  
if ((ae.GetExtent(box) <> #ANALYSENV_VALUE) or (ae.GetCellSize(cellSize)  
<> #ANALYSENV_VALUE)) then  
    ce = AnalysisPropertiesDialog.Show(theView, TRUE, "Conversion Extent:" ++  
theTheme.GetName)  
    if (ce = NIL) then  
        return NIL  
    end  
    ce.GetCellSize(cellSize)  
    ce.GetExtent(box)  
end
```



```

Runoff value
theQuestion = MsgBox.YesNo("Would you like to use the default runoff value of
20 in/yr?", "Runoff Value", TRUE)
if (theQuestion = FALSE) then
theRunoff = MsgBox.Input("Enter a new runoff value [in/yr].", "Runoff
Selection", "").AsNumber
else
theRunoff = 20
end
ae.GetCellSize(cellSize)
theFlow = cellSize*cellSize*theRunoff*0.0002283

'Add EMC, flow, and load values to table
for each rec in thefab
code = thefab.returnValue(codefield,rec)
if (code = 1005) then
thefab.returnValue(concfield,rec,c4)
elseif (code = 1010) then
thefab.returnValue(concfield,rec,c4)
elseif (code = 1020) then
thefab.returnValue(concfield,rec,c7)
elseif (code = 1030) then
thefab.returnValue(concfield,rec,c5)
elseif (code = 1040) then
thefab.returnValue(concfield,rec,c8)
elseif (code = 1060) then
thefab.returnValue(concfield,rec,c6)
elseif (code = 7100) then
thefab.returnValue(concfield,rec,c2)
elseif (code = 7110) then
thefab.returnValue(concfield,rec,c3)
elseif (code = 7120) then
thefab.returnValue(concfield,rec,c10)
elseif (code = 7130) then
thefab.returnValue(concfield,rec,c10)
elseif (code = 7140) then
thefab.returnValue(concfield,rec,c1)
elseif (code = 9999) then
thefab.returnValue(concfield,rec,c9)
else
MsgBox.Info("Something is wrong here.", "Error")
exit
end

theConcentration = thefab.returnValue(concfield,rec)
thefab.setValue(flowfield,rec,theFlow)
thefab.setValue(loadfield,rec,theFlow*theConcentration*28.31682)
end
thefab.SetEditable(false)
end

```

## Binomial.ave

```
'-----  
'--- Creation Information ---  
'-----  
,
```

```
Name: binomial.ave  
Version: 2.0  
Date: April 26, 1999  
Author: Andrew Romanek  
, Center for Research in Water Resources  
, The University of Texas at Austin  
, romanek@mail.utexas.edu  
,
```

```
'-----  
'--- Purpose/Description ---  
'-----  
,
```

```
Calculates the upper and lower confidence limits of the exceedance  
probability of the binomial distribution using a selected set of  
data and Excel. Requirements: Excel must be open with the Analysis  
ToolPak turned on and a blank spreadsheet showing. A view must be  
the active document with one theme selected. The selected theme  
must have a "Value" field.
```

```
Get Active Theme  
theView = av.GetActiveDoc  
theList = theView.GetActiveThemes  
theCount = theList.Count  
if (theCount > 1) then  
  MsgBox.Info("Only one theme can be active", "Error")  
  exit  
end  
theTheme = theList.Get(0)
```

```
Get Number of Trials and Successes  
theFTab = theTheme.GetFTab  
theSelected = theFTab.GetSelection  
theFTab.RememberSelection  
n = theSelected.Count  
thebitmap = theSelected.Count  
if (thebitmap = 0) then
```

```
  MsgBox.Info("You must select some records to use this script", "Error")  
  exit  
end  
theThreshold = MsgBox.Input("Enter the Threshold Concentration:", "Threshold  
Choice", "1000")  
expr = "([Value] >= " + theThreshold + ")"  
step = theFTab.Query(expr, theSelected, #VTAB_SELTYPE_AND)  
theFTab.UpdateSelection  
theTrial = theFTab.GetSelection  
s = theTrial.Count  
MsgBox.Info("The number of trials = "+n.AsString +NL+ "The number of  
violations = "+s.AsString, "Selected Area Parameters")  
lastSelection = theFTab.GetLastSelection  
theFTab.SetSelection(lastSelection.clone)  
theFTab.UpdateSelection
```

```
Get Confidence Limit  
theInterval = MsgBox.Input("Enter the two-sided confidence level:", "Confidence  
Interval", "0.90")  
alpha = 1 - theInterval.AsNumber
```

```
Link to Excel  
sysClient = DDEClient.Make("excel", "system")  
if (sysClient.HasError) then  
  MsgBox.Error(sysClient.GetErrorMsg, "")  
  exit  
end
```

```
Create the Spreadsheet  
selection = sysClient.Request("selection")  
spreadSheet = selection.Left(selection.IndexOf("!"))  
ssClient = DDEClient.Make("excel", spreadSheet)  
ssClient.Poke("R1C1", "n =")  
ssClient.Poke("R2C1", "s =")  
ssClient.Poke("R3C1", "alpha =")  
ssClient.Poke("R5C1", "p-upper")  
ssClient.Poke("R5C3", "p-lower")  
ssClient.Poke("R7C1", "=A6*100")  
ssClient.Poke("R7C3", "=C6*100")  
ssClient.Poke("R1C2", n)  
ssClient.Poke("R2C2", s)  
ssClient.Poke("R3C2", alpha)
```

```

ssClient.Poke("R6C2","=1-binomdist(B1-B2,B1,1-A6,true)+binomdist(B1-
B2,B1,1-A6,false)")
ssClient.Poke("R6C4","=abs(binomdist(B2,B1,C6,false)-
binomdist(B2,B1,C6,true))")

Calculate the Upper Bound
if (s = 0) then
  pu = (1 - (alpha^(1/n))) * 100
else
  alpha1 = alpha/2
  alpha2 = alpha1.AsString
  av.ShowMsg("Calculating Upper Bound")
  for each i in 0 .. 1 by 0.001
    av.SetWorkingStatus
    ssClient.Poke("R6C1",i)
    test = ssClient.Request("R6C2")
    if (test <= alpha2) then
      break
    end
  end
end
av.ClearMsg
av.ClearStatus
pu = ssClient.Request("R7C1")
end

Calculate the Lower Bound
if (s = 0) then
  pl = 0
elseif (s = n) then
  pl = (alpha^(1/n)) * 100
else
  alpha3 = 1 - (alpha/2)
  alpha4 = alpha3.AsString
  av.ShowMsg("Calculating Lower Bound")
  for each i in 0.001 .. 0.999 by 0.001
    av.SetWorkingStatus
    ssClient.Poke("R6C3",i)
    test2 = ssClient.Request("R6C4")
    if (test2 <= alpha4) then
      break
    end
  end
end
end

```

```

av.ClearMsg
av.ClearStatus
pl = ssClient.Request("R7C3")
end

Results
pb = s/n * 100
pb.SetFormat("d,d")
MsgBox.Info("The best estimate probability = "++pb.AsString+NL+"The upper
exceedance probability = "++pu.AsString+NL+"The lower exceedance probability
="++pl.AsString,"Results (%)")

```

## Burner.ave

```
'-----  
'--- Creation Information ---  
'-----  
,
```

```
Name: burner.ave  
'Version: 1.0  
Date: June 19, 1998  
'Author: Andrew Romanek  
,  
      Center for Research in Water Resources  
,  
      The University of Texas at Austin  
,  
      romanek@mail.utexas.edu  
,
```

```
'-----  
'--- Purpose/Description ---  
'-----  
,
```

```
Burns a stream grid into a Digital Elevation Model. Requirements:  
'Spatial Analyst and the view must be active. Saves the dem in the  
'current directory.
```

```
'Get the grids  
theView = av.GetActiveDoc  
theThemeList=theView.GetThemes  
theStreams = MsgBox.Choice(theThemeList,"Choose the Stream GRID","Stream  
Grid Selection")  
StreamGrid = theStreams.GetGrid  
theDEM = MsgBox.Choice(theThemeList,"Choose the DEM","DEM Selection")  
DEMGrid = theDEM.GetGrid
```

```
'Create stream grid  
unitstream = StreamGrid/StreamGrid  
demstream = DEMGrid*unitstream
```

```
'Raise the DEM  
theElevation = MsgBox.Input("Enter the arbitrary elevation rise. This value should  
be greater than the highest point in the DEM.", "Elevation Rise", "5000")  
demplus = DEMGrid+theElevation.AsNumber
```

```
'Merge the grids  
listGrid = { demplus }
```

```
outGrid = demstream.Merge(listGrid)
```

```
'Save and add to view  
outGrid.SaveDataSet("burndem".AsFileName)  
outTheme = GTheme.Make(outGrid)  
theView.AddTheme(outTheme)  
outTheme.setname("Burned DEM")
```

## Cocvalue.ave

```
'-----  
'--- Creation Information ---  
'-----  
,  
Name: cocvalue.ave  
Version: 1.0  
Date: February 23, 1998  
Author: Andrew Romanek  
,  
      Center for Research in Water Resources  
,  
      University of Texas at Austin  
,  
      romanek@mail.utexas.edu  
,  
'-----  
'--- Purpose/Description ---  
'-----  
,  
Creates a point coverage of COC concentrations from a query in  
Microsoft Access. Requirements: ODBC connection titled BPAccess,  
the query must contain the requested fields, and the view must  
be active.  
  
theView=av.GetActiveDoc  
  
theSQL=SQLCon.Find("BPAccess")  
theQuery=MsgBox.Input("Enter the COC Query Name:", "Choose the COC  
Query", "")  
astring="Select * from" ++ theQuery  
theVTab=VTab.MakeSQL(theSQL,astring)  
myTable=Table.Make(theVtab)  
av.GetProject.AddDoc(mytable)  
mytable.GetWin.Open  
thetable=MsgBox.Input("Enter a Name for the Output Table:", "Choose Table  
Name", "")  
mytable.setname(thetable)  
  
theProject=av.GetProject  
theDocs=theProject.GetDocs  
  
'--- IDENTIFY INPUT TABLE  
intablename=thetable
```

```
if (intablename=nil) then  
  exit  
end  
  
intable=theproject.finddoc(intablename)  
invtab=intable.getvtab  
infields=invtab.getfields  
  
'--- IDENTIFY INPUT FIELDS  
xfield=invtab.findfield("Easting")  
yfield=invtab.findfield("Northing")  
idfield=invtab.findfield("Samp_ID")  
locfield=invtab.findfield("Loc_ID")  
matfield=invtab.findfield("Matrix")  
casfield=invtab.findfield("CAS")  
valfield=invtab.findfield("Value")  
detfield=invtab.findfield("Det_Limit")  
unifield=invtab.findfield("Unit")  
  
'--- READ AND PROCESS DATA  
'--- UNITS OF X AND Y ARE FEET  
  
OutFileName=FileDialog.Put("outfile".asfilename,"*.shp", "Output Shape File")  
if(OutFileName=Nil)then  
  exit  
end  
OutFileName.SetExtension("shp")  
OutFTab=FTab.MakeNew(OutFileName,point)  
outTheme=Ftheme.make(outftab)  
  
CREATE FIELDS FOR THE NEW POINT TABLE  
outFields=List.Make  
outfields.Add(Field.Make("Samp_ID",#FIELD_CHAR,20,0))  
outfields.Add(Field.Make("Loc_ID",#FIELD_CHAR,20,0))  
outFields.Add(Field.Make("Easting",#FIELD_DECIMAL,15,6))  
outFields.Add(Field.Make("Northing",#FIELD_DECIMAL,15,6))  
outFields.Add(Field.Make("Matrix",#FIELD_CHAR,15,0))  
outFields.Add(Field.Make("CAS",#FIELD_CHAR,10,0))  
outFields.Add(Field.Make("Value",#FIELD_DECIMAL,10,1))  
outFields.Add(Field.Make("Det_Limit",#FIELD_DECIMAL,5,1))  
outFields.Add(Field.Make("Unit",#FIELD_CHAR,8,0))  
outFieldsc=outFields.DeepClone
```

```

outftab.setvalue(ocasfield,newrec,CAS)
outftab.setvalue(ovalfield,newrec,Concentration)
outftab.setvalue(odetfield,newrec,Detection_Limit)
outftab.setvalue(ounifield,newrec,Unit)
end

outftab.setctable(false)
mytable.GetWin.Close

```

```

outftab.addfields(outfields)
theView.addtheme(outTheme)

if(outFtab.CanEdit)then
outFtab.SetEditable(true)
else
msgbox.info("Can't edit the output theme. ", "Error")
exit
end

```

#### IDENTIFY FIELDS FOR WRITING

```

shpField=outFtab.findfield("shape")
oidfield=outftab.findfield("Samp_ID")
oxfield=outftab.findfield("Easting")
oyfield=outftab.findfield("Northing")
olocfield=outftab.findfield("Loc_ID")
omatfield=outftab.findfield("Matrix")
ocasfield=outftab.findfield("CAS")
ovalfield=outftab.findfield("Value")
odetfield=outftab.findfield("Det_Limit")
ounifield=outftab.findfield("Unit")

```

for each rec in invtab

```

Samp_ID=invtab.returnvalue(idfield,rec)
Easting=invtab.returnvalue(xfield,rec)
Northing=invtab.returnvalue(yfield,rec)
Loc_ID=invtab.returnvalue(locfield,rec)
Matrix=invtab.returnvalue(matfield,rec)
CAS=invtab.returnvalue(casfield,rec)
Concentration=invtab.returnvalue(valfield,rec)
Detection_Limit=invtab.returnvalue(detfield,rec)
Unit=invtab.returnvalue(unifield,rec)
newrec=outFtab.AddRecord
pt=point.make(Easting,Northing)
outFtab.SetValue(shpField,newrec,pt)
outftab.setvalue(oidfield,newrec,Samp_ID)
outftab.setvalue(oxfield,newrec,Easting)
outftab.setvalue(oyfield,newrec,Northing)
outftab.setvalue(olocfield,newrec,Loc_ID)
outftab.setvalue(omatfield,newrec,Matrix)

```

## Drainer.ave

```
'-----  
'--- Creation Information ---  
'-----  
'  
'Name: drainer.ave  
'Version: 1.0  
'Author: Andrew Romanek  
'      Center for Research in Water Resources  
'      The University of Texas at Austin  
'      romanek@mail.utexas.edu  
'-----  
'--- Purpose/Description ---  
'-----  
'  
'Merges a drain grid into a dem. The dem elevation remains  
'constant everywhere except where a drain exists.  
  
'Get the grids  
theView = av.GetActiveDoc  
theThemeList=theView.GetThemes  
theDrains = MsgBox.Choice(theThemeList,"Choose the Drain GRID","Drain Grid  
Selection")  
DrainGrid = theDrains.GetGrid  
theDEM = MsgBox.Choice(theThemeList,"Choose the DEM","DEM Selection")  
DEMGrid = theDEM.GetGrid  
  
'Merge the grids  
listGrid = {DEMGrid}  
outGrid = DrainGrid.Merge(listGrid)  
  
'Save and add to view  
outGrid.SaveDataSet("burndem".AsFileName)  
outTheme = GTheme.Make(outGrid)  
theView.AddTheme(outTheme)  
outTheme.setname("Merged DEM")
```

285

## Interpol.ave

```
'-----  
'--- Creation Information ---  
'-----  
'  
'Name: interpol.ave  
'Version: 1.1  
'Date: September 29, 1998  
'Author: Andrew Romanek  
'      Center for Research in Water Resources  
'      The University of Texas at Austin  
'      romanek@mail.utexas.edu  
'-----  
'--- Purpose/Description ---  
'-----  
'  
'Interpolates a grid from a set of contour lines. Requirements:  
'Spatial Analyst, the extent of the output grid, the view must  
'be active, and one polyline theme must be active.  
  
'Get the contour lines and choose an interpolation field.  
theView = av.GetActiveDoc  
theThemeList = theView.GetActiveThemes  
if (theThemeList.Count = 1) then  
    theTheme = theThemeList.Get(0)  
elseif (theThemeList.Count = 0) then  
    MsgBox.Info("One theme must be active.", "Error")  
    exit  
else  
    MsgBox.Info("Only one theme can be active.", "Error")  
    exit  
end  
if (theTheme.GetFTab.FindField("shape").GetType<>#field_shapeline) then  
    MsgBox.Info("The active theme must be a polyline theme.", "Error")  
    exit  
end  
theFTab = theTheme.getftab  
theFieldList = theFTab.GetFields  
theField = MsgBox.Choice(theFieldList,"Choose the Interpolation Field","Field  
Selection")
```

```

' Create a point coverage.
theNewShapefile = av.GetProject.GetWorkDir.MakeTmp("newpoint", "shp")
newftab = FTAB.MakeNew(theNewShapefile,Point)
valuefield = Field.Make("Value",#FIELD_DECIMAL,16,2)
newftab.SetEditable(true)
newftab.AddField({valuefield})

step = 0
countup = 0
av.SetStatus(0)
av.ShowMsg("Creating Points...")
for each rec in theFTab
    countup = countup + 1
    av.SetStatus(countup/theFTab.GetNumRecords*100)
    theShape = theFTab.ReturnValue(theFTab.FindField("Shape"),rec)
    theValue = theFTab.ReturnValue(theFTab.FindField(theField.AsString),rec)
    pointslst = theShape.AsMultiPoint.AsList
    for each p in pointslst
        step = step + 1
        newrec = newftab.addrecord
        newftab.setvalue(newftab.findfield("Shape"),newrec,p)
        newftab.setvalue(newftab.findfield("Value"),newrec,theValue)
    end
end
newftab.seteditable(false)

av.ClearMsg
newtheme = ftheme.make(newftab)

' Grid Parameters
theLabels = {"Lower Left Easting","Lower Left Northing","Upper Right
Easting","Upper Right Northing","Cell Size"}
theRectangle = theTheme.ReturnExtent
theLeft = theRectangle.GetLeft.AsString
theRight = theRectangle.GetRight.AsString
theTop = theRectangle.GetTop.AsString
theBottom = theRectangle.GetBottom.AsString
theDefaults = {theLeft,theBottom,theRight,theTop,"1000"}
theParameters = MsgBox.MultiInput("Choose the Output Grid
Parameters","Interpolation Grid Parameters",theLabels,theDefaults)
x1 = theParameters.Get(0).AsNumber

y1 = theParameters.Get(1).AsNumber
x2 = theParameters.Get(2).AsNumber
y2 = theParameters.Get(3).AsNumber
therect = Rect.MakeXY(x1,y1,x2,y2)
cellsize = theParameters.Get(4).AsNumber

' Interpolation Method
theinterpList = InterpolationDialog.Show(newtheme,cellsize)
if (theinterpList.Count < 2) then
    return NIL
end
thezField = theinterpList.Get(0)
anInterp = theinterpList.Get(1)
aPrj = theView.GetProjection
theptFTab = newtheme.GetFTab

' Create Grid and Add to View
r = Grid.MakeByInterpolation(theptFTab,aPrj,thezField,anInterp,{cellsize,therect})
if (r.HasError) then
    return NIL
end
gthm = GTheme.Make(r)
theView.addtheme(gthm)
gthm.setname("Surface From" ++ theTheme.AsString)

```



## Newfield.ave

```
'-----  
'--- Creation Information ---  
'-----  
,
```

```
Name: newfield.ave  
Version: 1.0  
Date: August 26, 1998  
Author: Andrew Romanek  
, Center for Research in Water Resources  
, University of Texas at Austin  
, romanek@mail.utexas.edu  
,
```

```
'-----  
'--- Purpose/Description ---  
'-----  
,
```

```
'Adds a new field to an existing theme table that will maintain  
'any detected concentration values and will add a value you choose  
'for the no-detect samples. The value added can be 0, 1/2 the  
'detection limit, or a custom number. The purpose of this new field  
'is to aid in contouring applications.  
Requirements: The view must be active, and there must be a theme  
table with Value and Det_Limit fields.
```

```
Find the table  
theProject = av.GetProject  
theView = av.GetActiveDoc  
theThemeList = theView.GetThemes  
theTheme = MsgBox.Choice(theThemeList,"Choose the Theme Table to  
Modify","Theme Table Selection")
```

```
'Set the editing properties  
thefab = theTheme.GetFTab  
if (thefab.CanEdit) then  
    thefab.SetEditable(true)  
else  
    MsgBox.Info("Can't edit the theme table","Error")  
    exit  
end
```

```
'Add the new field  
outfields=List.Make  
outfields.Add(Field.Make("Newfield",#FIELD_DECIMAL,10,2))  
outfieldsc=outfields.DeepClone  
thefab.addfields(outfieldsc)
```

```
Define fields  
valfield = thefab.findfield("Value")  
detfield = thefab.findfield("Det_Limit")  
newfield = thefab.findfield("Newfield")
```

```
'Choose the modification  
choice1 = "Convert no-detect values to zero"  
choice2 = "Convert no-detect values to 1/2 of detection limit"  
choice3 = "Choose a custom number for no-detect values"  
theList = {choice1,choice2,choice3}  
theChoice = MsgBox.ChoiceAsString(theList,"Choose the  
Modification","Modification Selection")
```

```
Return and modify values  
if (theChoice = choice1) then  
    for each rec in thefab  
        Concentration = thefab.returnvalue(valfield,rec)  
        Detection = thefab.returnvalue(detfield,rec)  
        if (Concentration.IsNull) then  
            thefab.setvalue(newfield,rec,0)  
        else  
            thefab.setvalue(newfield,rec,Concentration)  
        end  
    end  
elseif (theChoice = choice2) then  
    for each rec in thefab  
        Concentration = thefab.returnvalue(valfield,rec)  
        Detection = thefab.returnvalue(detfield,rec)  
        if (Concentration.IsNull) then  
            thefab.setvalue(newfield,rec,Detection/2)  
        else  
            thefab.setvalue(newfield,rec,Concentration)  
        end  
    end  
elseif (theChoice = choice3) then  
    theNumber = MsgBox.Input("Enter a value:","Input Form","0")
```

```

for each rec in theftab
  Concentration = theftab.returnvalue(valfield,rec)
  Detection = theftab.returnvalue(detfield,rec)
  if (Concentration.IsNull) then
    theftab.setvalue(newfield,rec,theNumber.AsNumber)
  else
    theftab.setvalue(newfield,rec,Concentration)
  end
end
end
else
  MsgBox.Info("Something is wrong here","Error")
  exit
end

theftab.SetEditable(false)

```

288

## Startup.ave

```

'-----
'--- Creation Information ---
'-----
'
Name: startup.ave
Version: 1.1
Date: August 27, 1998
Author: Andrew Romanek
'
'       Center for Research in Water Resources
'       The University of Texas at Austin
'       romanek@mail.utexas.edu
'
'-----
'--- Purpose/Description ---
'-----
'
Renames the title bar,d isplays a banner, and maximizes all
'documents to the extent of the computer screen.

av.SetName("Marcus Hook Refinery - ")
av.Maximize
MsgBox.Banner("banner.bmp".AsFileName,4,"")
thewindow = av.getproject.getwin
thewindow.maximize
therectangle = av.getproject.getwin.returnextent
thewindow.restore
thewidth = therectangle.getx
theheight = therectangle.gety
thedocs = av.getproject.getdocs
for each i in thedocs
  i.getwin.resize(thewidth,theheight)
end

```

## Tanksbuild.ave

```
'-----  
'--- Creation Information ---  
'-----  
,
```

```
'Name: tanksbuild.ave  
'Version: 1.0  
'Author: Andrew Romanek  
'      Center for Research in Water Resources  
'      The University of Texas at Austin  
'      romanek@mail.utexas.edu  
,
```

```
'-----  
'--- Purpose/Description ---  
'-----  
,
```

```
'Merges a tank grid and a building grid into a dem. The dem  
'elevation remains constant everywhere except where a tank or  
'building exists. Note: the tank and building grids should have  
'a value that is higher than the highest point in the dem.
```

```
'Get the Grids  
theView = av.GetActiveDoc  
theThemeList = theView.GetThemes  
aGridName = MsgBox.Choice(theThemeList,"Choose the Tanks Grid","Tank Grid  
Selection")  
bGridName = MsgBox.Choice(theThemeList,"Choose the Buildings  
Grid","Building Grid Selection")  
cGridName = MsgBox.Choice(theThemeList,"Choose the DEM","DEM  
Selection")  
aGrid = aGridName.GetGrid  
bGrid = bGridName.GetGrid  
cGrid = cGridName.GetGrid
```

```
'Merge the Tanks  
listGrid = {cGrid}  
outGrid = aGrid.Merge(listGrid)  
outGrid.SaveDataSet("newdem1".AsFileName)  
outTheme = GTheme.Make(outGrid)
```

```
theView.AddTheme(outTheme)
```

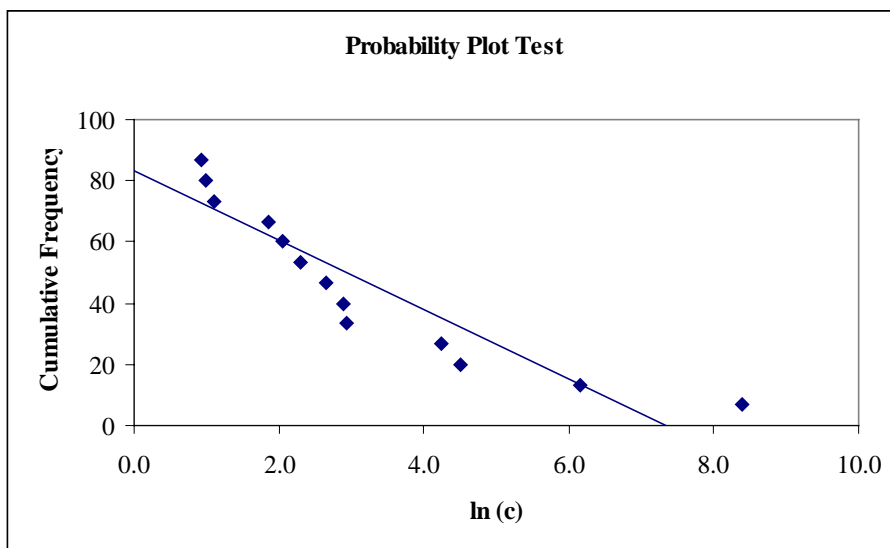
```
'Merge the Buildings  
listGrid = {outGrid}  
outGrid = bGrid.Merge(listGrid)  
outGrid.SaveDataSet("newdem2".AsFileName)  
outTheme = GTheme.Make(outGrid)  
theView.AddTheme(outTheme)
```

**APPENDIX H**  
**RISK MAPPING DATA SHEETS**

**Benzene Data for Source Area 1 (Soil)**

C ( $\mu\text{g}/\text{kg}$ )	ln (C) ( $\mu\text{g}/\text{kg}$ )	Rank	CF	<i>Statistics</i>	
4400.0	8.4	1	6.7		
477.0	6.2	2	13.3	Mean	2.91
90.0	4.5	3	20.0	Standard Error	0.61
70.0	4.2	4	26.7	Median	2.47
19.0	2.9	5	33.3	Mode	#N/A
18.0	2.9	6	40.0	Standard Deviation	2.29
14.0	2.6	7	46.7	Sample Variance	5.23
10.0	2.3	8	53.3	Kurtosis	1.35
7.8	2.1	9	60.0	Skewness	1.13
6.4	1.9	10	66.7	Range	8.61
3.0	1.1	11	73.3	Minimum	-0.22
2.7	1.0	12	80.0	Maximum	8.39
2.5	0.9	13	86.7	Sum	40.78
0.8	-0.2	14	93.3	Count	14

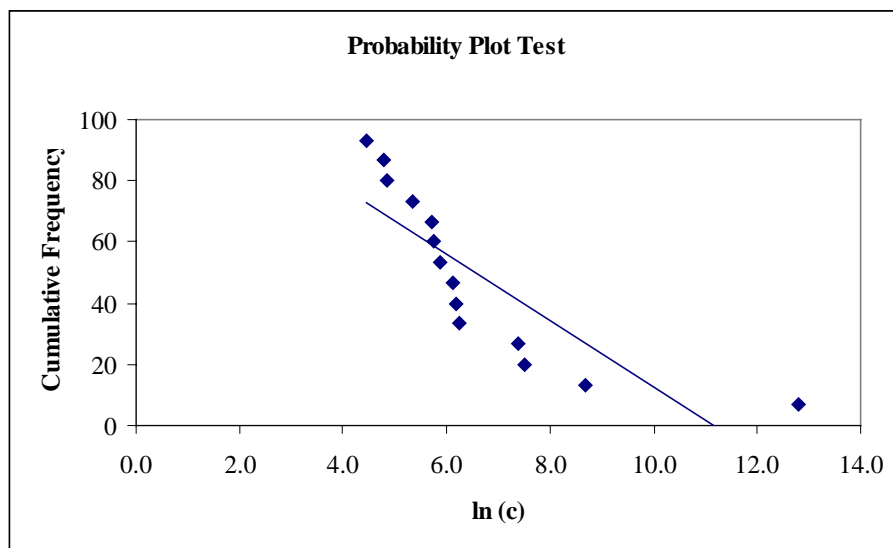
CV =	0.79
CS  =	1.01
ln (UCL) =	4.00
UCL =	54.3



**Benzene Data for Source Area 2 (Soil)**

C ( $\mu\text{g}/\text{kg}$ )	ln (C) ( $\mu\text{g}/\text{kg}$ )	Rank	CF	<i>Statistics</i>	
360000.0	12.8	1	6.7		
6000.0	8.7	2	13.3	Mean	6.55
1800.0	7.5	3	20.0	Standard Error	0.57
1600.0	7.4	4	26.7	Median	5.99
510.0	6.2	5	33.3	Mode	#N/A
490.0	6.2	6	40.0	Standard Deviation	2.13
460.0	6.1	7	46.7	Sample Variance	4.56
350.0	5.9	8	53.3	Kurtosis	5.50
310.0	5.7	9	60.0	Skewness	2.14
300.0	5.7	10	66.7	Range	8.34
210.0	5.3	11	73.3	Minimum	4.45
130.0	4.9	12	80.0	Maximum	12.79
120.0	4.8	13	86.7	Sum	91.68
86.0	4.5	14	93.3	Count	14

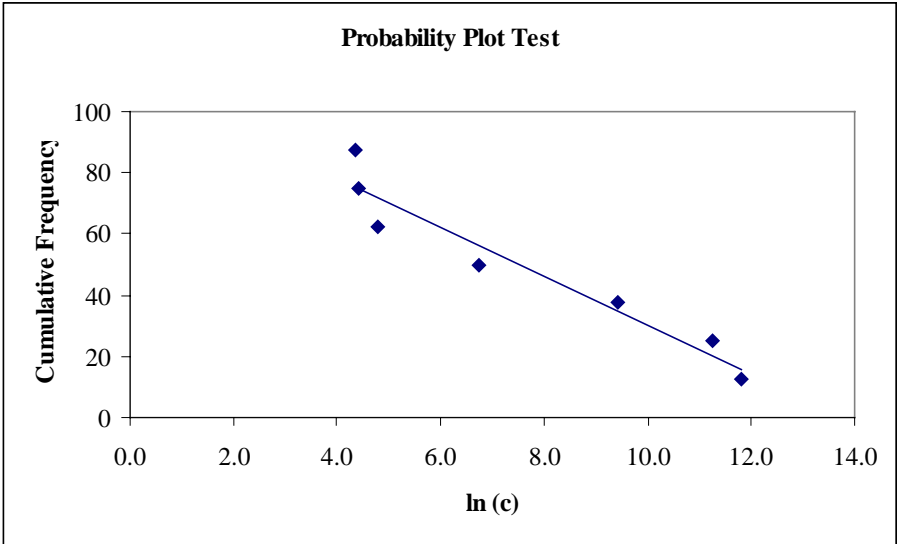
CV =	0.33
CS  =	1.91
ln (UCL) =	7.56
UCL =	1917.7



**Benzene Data for Source Area 3 (Soil)**

C (µg/kg)	ln (C) (µg/kg)	Rank	CF	<i>Statistics</i>	
135000.0	11.8	1	12.5	Mean	7.54
76800.0	11.2	2	25.0	Standard Error	1.23
12500.0	9.4	3	37.5	Median	6.74
842.0	6.7	4	50.0	Mode	#N/A
120.0	4.8	5	62.5	Standard Deviation	3.26
84.0	4.4	6	75.0	Sample Variance	10.60
78.0	4.4	7	87.5	Kurtosis	-2.13
				Skewness	0.35
				Range	7.46
				Minimum	4.36
				Maximum	11.81
				Sum	52.81
				Count	7

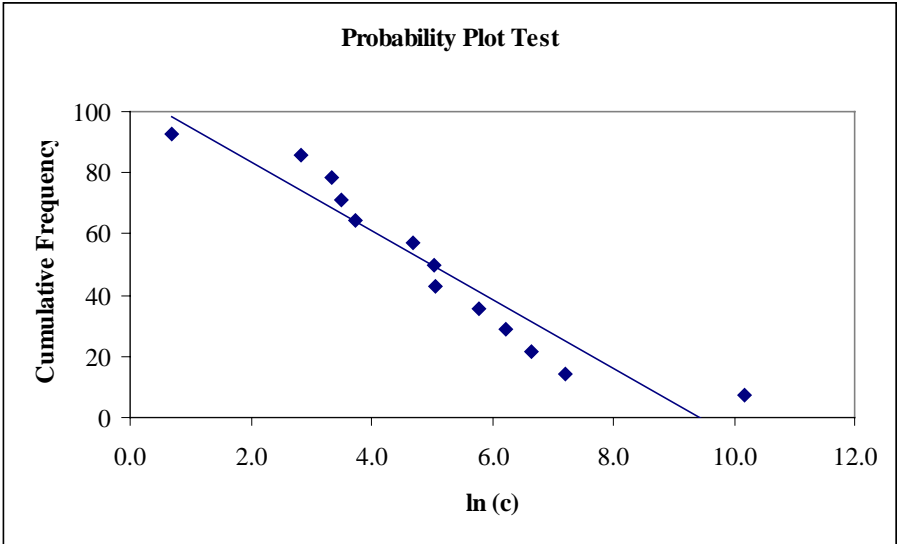
CV =	0.43
CS  =	0.27
ln (UCL) =	9.94
UCL =	20644.1



**Benzene Data for Source Area 4 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
25829.0	10.2	1	7.1		
1338.0	7.2	2	14.3	Mean	4.99
763.0	6.6	3	21.4	Standard Error	0.65
499.0	6.2	4	28.6	Median	5.04
322.0	5.8	5	35.7	Mode	#N/A
159.0	5.1	6	42.9	Standard Deviation	2.35
154.0	5.0	7	50.0	Sample Variance	5.51
110.0	4.7	8	57.1	Kurtosis	1.17
42.0	3.7	9	64.3	Skewness	0.43
33.0	3.5	10	71.4	Range	9.47
28.0	3.3	11	78.6	Minimum	0.69
17.0	2.8	12	85.7	Maximum	10.16
2.0	0.7	13	92.9	Sum	64.88
				Count	13

CV =	0.47
CS  =	0.38
ln (UCL) =	6.15
UCL =	469.4

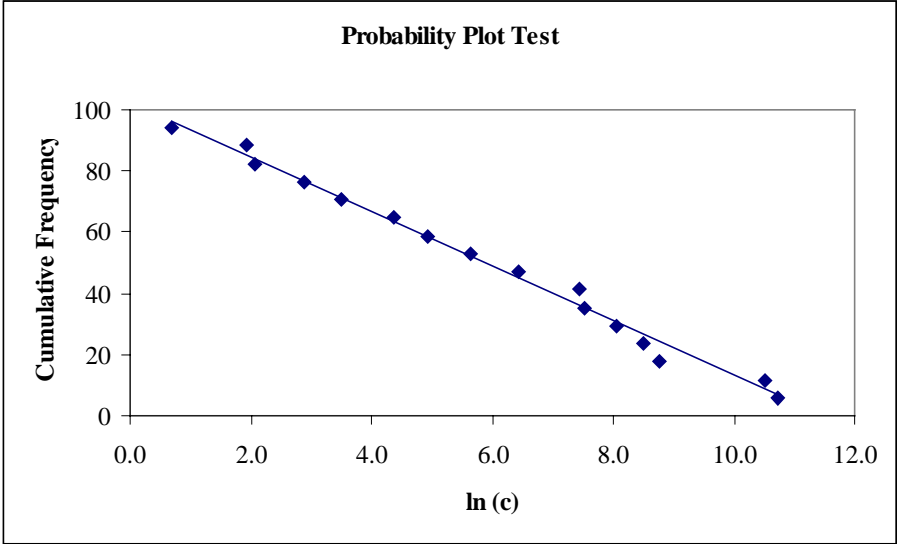




**Benzene Data for Source Area 5 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
45184.0	10.7	1	5.9		
36742.0	10.5	2	11.8	Mean	5.88
6482.0	8.8	3	17.6	Standard Error	0.78
4879.0	8.5	4	23.5	Median	6.03
3157.0	8.1	5	29.4	Mode	#N/A
1855.0	7.5	6	35.3	Standard Deviation	3.11
1709.0	7.4	7	41.2	Sample Variance	9.66
619.0	6.4	8	47.1	Kurtosis	-1.09
281.0	5.6	9	52.9	Skewness	-0.06
139.0	4.9	10	58.8	Range	10.03
80.0	4.4	11	64.7	Minimum	0.69
33.0	3.5	12	70.6	Maximum	10.72
18.0	2.9	13	76.5	Sum	94.01
8.0	2.1	14	82.4	Count	16
7.0	1.9	15	88.2		
2.0	0.7	16	94.1		

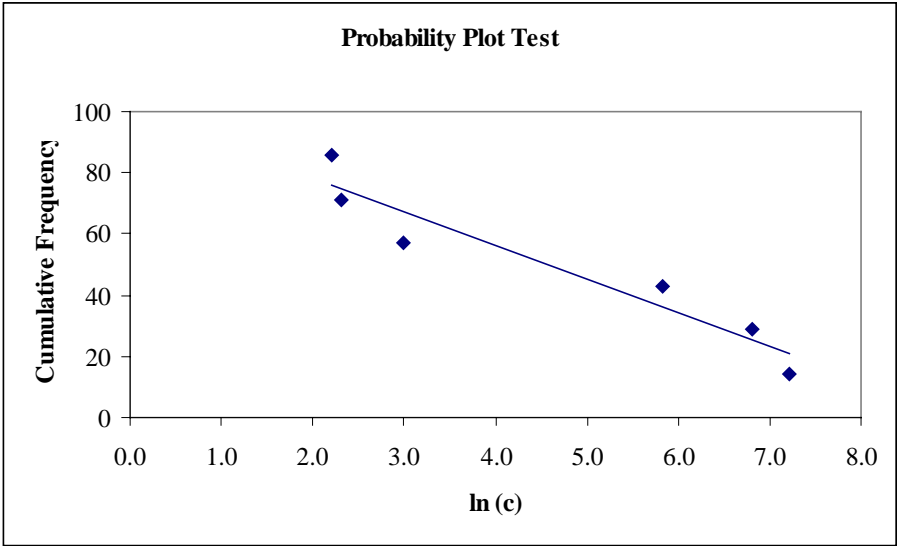
CV =	0.53
CS  =	0.06
ln (UCL) =	7.24
UCL =	1391.4



**Benzene Data for Source Area 6 (Groundwater)**

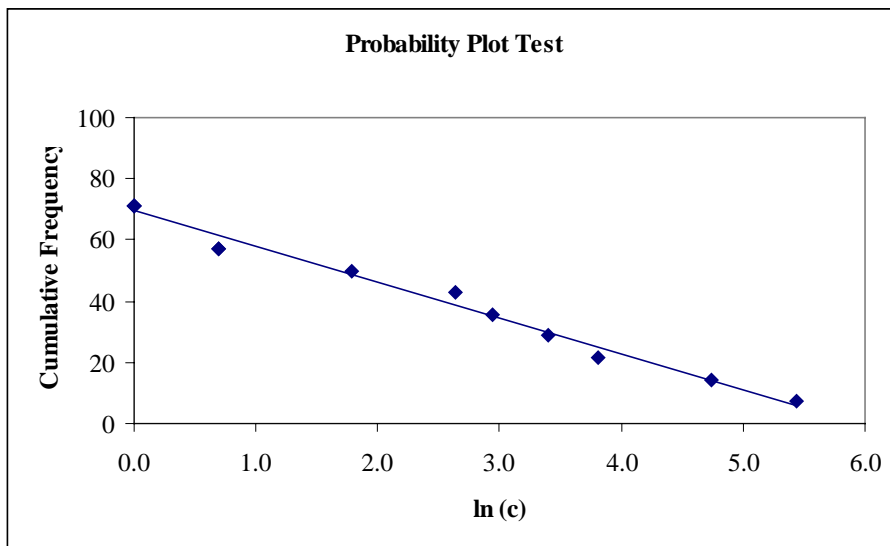
C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
1345.0	7.2	1	14.3	Mean	4.56
910.0	6.8	2	28.6	Standard Error	0.95
342.0	5.8	3	42.9	Median	4.42
20.0	3.0	4	57.1	Mode	#N/A
10.0	2.3	5	71.4	Standard Deviation	2.32
9.0	2.2	6	85.7	Sample Variance	5.36
				Kurtosis	-2.83
				Skewness	0.09
				Range	5.01
				Minimum	2.20
				Maximum	7.20
				Sum	27.35
				Count	6

CV =	0.51
CS  =	0.06
ln (UCL) =	6.46
UCL =	641.1



**Benzene Data for Source Area 7 (Soil Vapor)**

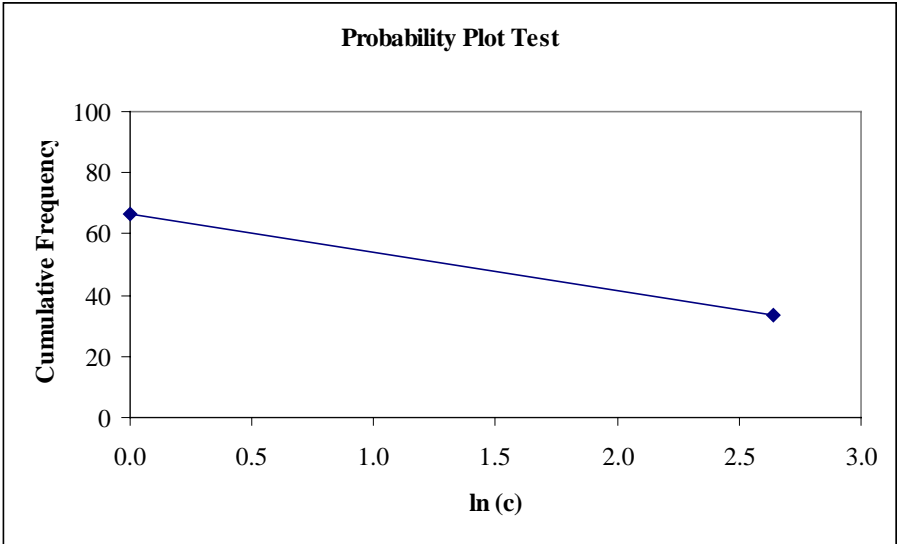
C (ppm)	ln (C) (ppm)	Rank	CF	<i>Statistics</i>		
229.0	5.4	1	7.1			
115.0	4.7	2	14.3	Mean	2.01	CV = 0.97
45.0	3.8	3	21.4	Standard Error	0.54	CS  = 0.39
30.0	3.4	4	28.6	Median	1.79	ln (UCL) = 2.97
19.0	2.9	5	35.7	Mode	0.00	UCL = 19.5
14.0	2.6	6	42.9	Standard Deviation	1.94	
6.0	1.8	7	50.0	Sample Variance	3.78	
2.0	0.7	8	57.1	Kurtosis	-1.22	
2.0	0.7	8	57.1	Skewness	0.44	
1.0	0.0	10	71.4	Range	5.43	
1.0	0.0	10	71.4	Minimum	0.00	
1.0	0.0	10	71.4	Maximum	5.43	
1.0	0.0	10	71.4	Sum	26.15	
				Count	13	



**Benzene Data for Source Area 8 (Soil Vapor)**

C (ppm)	ln (C) (ppm)	Rank	CF	<i>Statistics</i>	
14.0	2.6	1	33.3	Mean	1.32
1.0	0.0	2	66.7	Standard Error	1.32
				Median	1.32
				Mode	#N/A
				Standard Deviation	1.87
				Sample Variance	3.48
				Kurtosis	-
				Skewness	-
				Range	2.64
				Minimum	0.00
				Maximum	2.64
				Sum	2.64
				Count	2

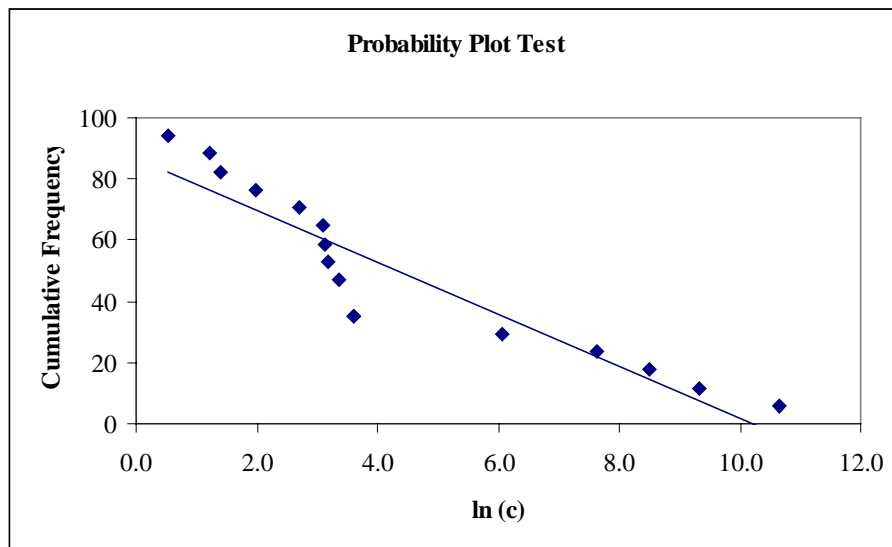
CV =	1.41
CS  =	-
ln (UCL) =	9.65
UCL =	15537.8



**Ethyl Benzene Data for Source Area 1 (Soil)**

C ( $\mu\text{g}/\text{kg}$ )	ln (C) ( $\mu\text{g}/\text{kg}$ )	Rank	CF	<i>Statistics</i>	
41800.0	10.6	1	5.9		
11200.0	9.3	2	11.8	Mean	4.38
4880.0	8.5	3	17.6	Standard Error	0.77
2070.0	7.6	4	23.5	Median	3.27
428.0	6.1	5	29.4	Mode	3.61
37.0	3.6	6	35.3	Standard Deviation	3.09
37.0	3.6	6	35.3	Sample Variance	9.53
29.0	3.4	8	47.1	Kurtosis	-0.43
24.0	3.2	9	52.9	Skewness	0.87
23.0	3.1	10	58.8	Range	10.11
22.0	3.1	11	64.7	Minimum	0.53
15.0	2.7	12	70.6	Maximum	10.64
7.3	2.0	13	76.5	Sum	70.01
4.1	1.4	14	82.4	Count	16
3.4	1.2	15	88.2		
1.7	0.5	16	94.1		

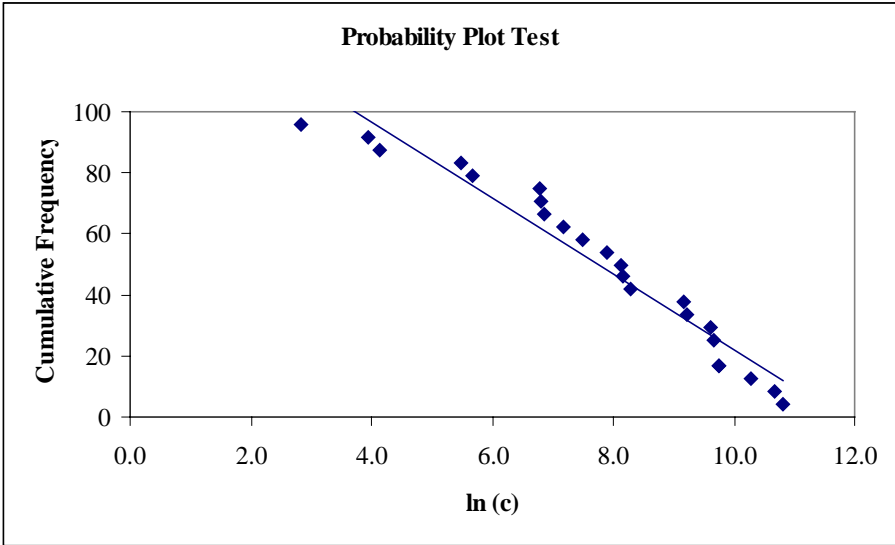
CV =	0.71
CS  =	0.79
ln (UCL) =	5.73
UCL =	307.4



**Ethyl Benzene Data for Source Area 2 (Soil)**

C (µg/kg)	ln (C) (µg/kg)	Rank	CF	<i>Statistics</i>	
49000.0	10.8	1	4.2		
43000.0	10.7	2	8.3	Mean	7.76
29000.0	10.3	3	12.5	Standard Error	0.46
17000.0	9.7	4	16.7	Median	8.13
17000.0	9.7	4	16.7	Mode	9.74
16000.0	9.7	6	25.0	Standard Deviation	2.22
15000.0	9.6	7	29.2	Sample Variance	4.92
10000.0	9.2	8	33.3	Kurtosis	-0.29
9600.0	9.2	9	37.5	Skewness	-0.66
4000.0	8.3	10	41.7	Range	7.97
3500.0	8.2	11	45.8	Minimum	2.83
3400.0	8.1	12	50.0	Maximum	10.80
2700.0	7.9	13	54.2	Sum	178.6
1800.0	7.5	14	58.3	Count	23
1300.0	7.2	15	62.5		
950.0	6.9	16	66.7		
910.0	6.8	17	70.8		
880.0	6.8	18	75.0		
290.0	5.7	19	79.2		
240.0	5.5	20	83.3		
62.0	4.1	21	87.5		
52.0	4.0	22	91.7		
17.0	2.8	23	95.8		

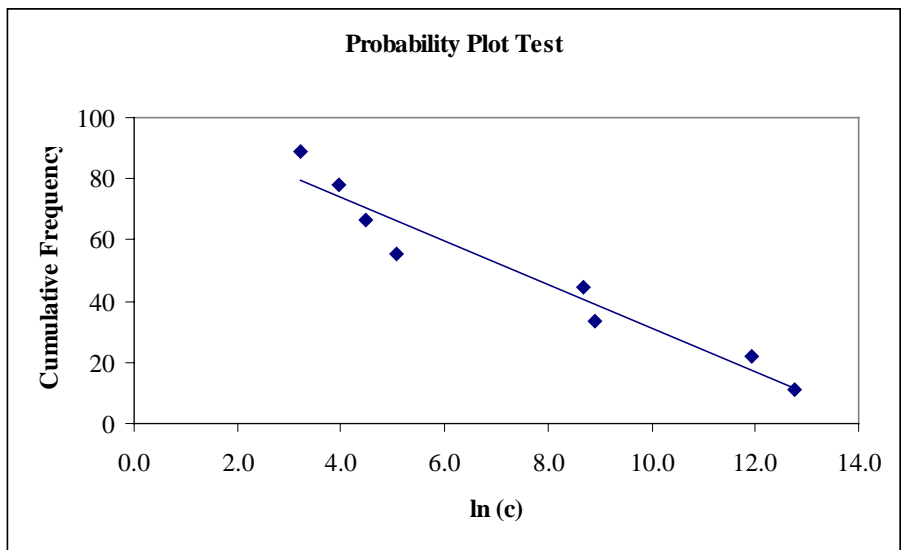
CV =	0.29
CS  =	0.61
ln (UCL) =	8.56
UCL =	5209.4



**Ethyl Benzene Data for Source Area 3 (Soil)**

C ( $\mu\text{g}/\text{kg}$ )	ln (C) ( $\mu\text{g}/\text{kg}$ )	Rank	CF	<i>Statistics</i>	
346000.0	12.8	1	11.1	Mean	7.38
153000.0	11.9	2	22.2	Standard Error	1.31
7420.0	8.9	3	33.3	Median	6.87
5830.0	8.7	4	44.4	Mode	#N/A
159.0	5.1	5	55.6	Standard Deviation	3.71
89.0	4.5	6	66.7	Sample Variance	13.78
52.0	4.0	7	77.8	Kurtosis	-1.63
25.0	3.2	8	88.9	Skewness	0.40
				Range	9.54
				Minimum	3.22
				Maximum	12.75
				Sum	59.00
				Count	8

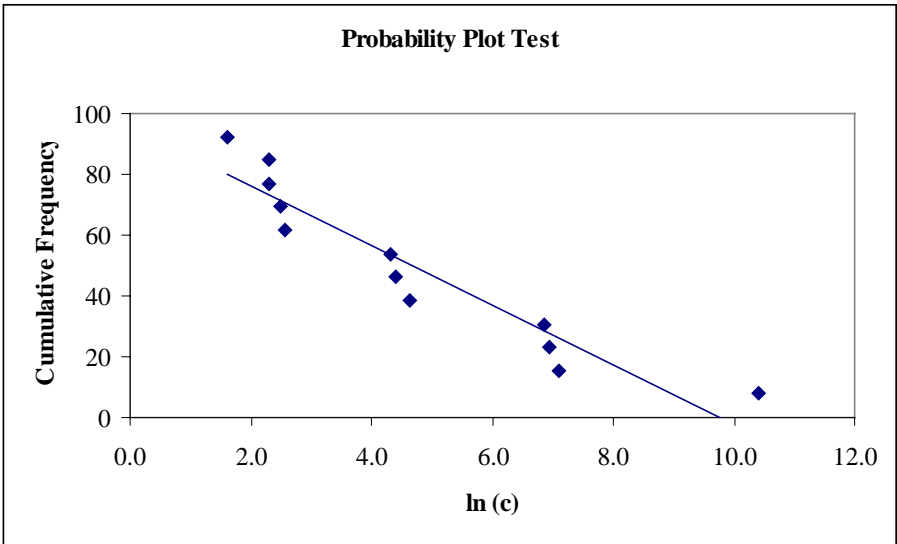
CV =	0.50
CS  =	0.32
ln (UCL) =	9.86
UCL =	19191.4



**Ethyl Benzene Data for Source Area 4 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
32822.0	10.4	1	7.7		
1200.0	7.1	2	15.4	Mean	4.66
1022.0	6.9	3	23.1	Standard Error	0.77
951.0	6.9	4	30.8	Median	4.36
102.0	4.6	5	38.5	Mode	2.30
82.0	4.4	6	46.2	Standard Deviation	2.67
75.0	4.3	7	53.8	Sample Variance	7.14
13.0	2.6	8	61.5	Kurtosis	0.17
12.0	2.5	9	69.2	Skewness	0.87
10.0	2.3	10	76.9	Range	8.79
10.0	2.3	11	84.6	Minimum	1.61
5.0	1.6	12	92.3	Maximum	10.40
				Sum	55.89
				Count	12

CV =	0.57
CS  =	0.76
ln (UCL) =	6.04
UCL =	421.1

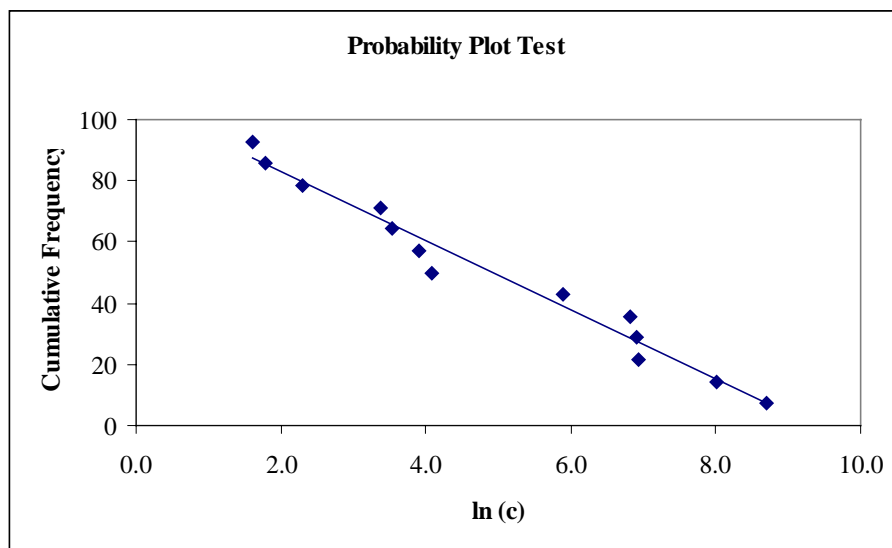




**Ethyl Benzene Data for Source Area 5 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
5975.0	8.7	1	7.1		
3025.0	8.0	2	14.3	Mean	4.91
1022.0	6.9	3	21.4	Standard Error	0.67
1009.0	6.9	4	28.6	Median	4.08
925.0	6.8	5	35.7	Mode	#N/A
361.0	5.9	6	42.9	Standard Deviation	2.42
59.0	4.1	7	50.0	Sample Variance	5.84
50.0	3.9	8	57.1	Kurtosis	-1.44
34.0	3.5	9	64.3	Skewness	0.11
29.0	3.4	10	71.4	Range	7.09
10.0	2.3	11	78.6	Minimum	1.61
6.0	1.8	12	85.7	Maximum	8.70
5.0	1.6	13	92.9	Sum	63.86
				Count	13

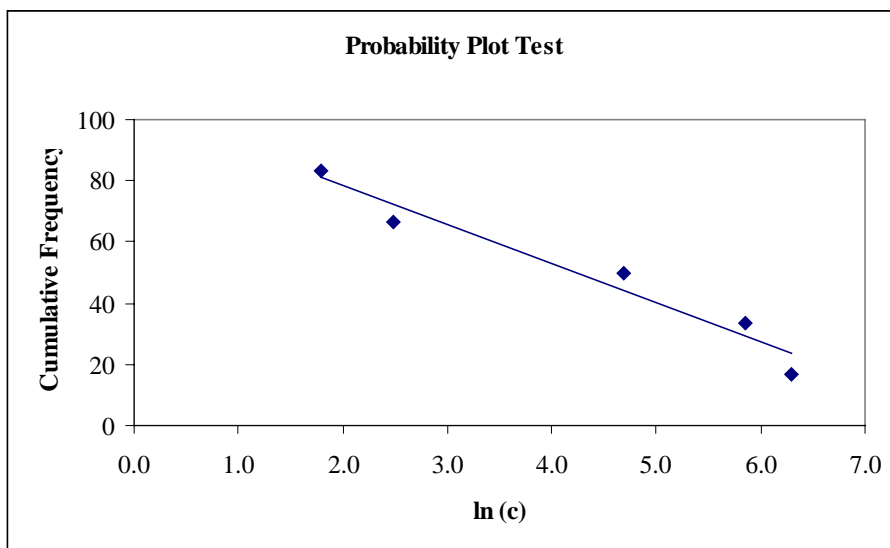
CV =	0.49
CS  =	0.10
ln (UCL) =	6.11
UCL =	449.1



**Ethyl Benzene Data for Source Area 6 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
542.0	6.3	1	16.7	Mean	4.22
350.0	5.9	2	33.3	Standard Error	0.90
108.0	4.7	3	50.0	Median	4.68
12.0	2.5	4	66.7	Mode	#N/A
6.0	1.8	5	83.3	Standard Deviation	2.01
				Sample Variance	4.03
				Kurtosis	-2.63
				Skewness	-0.32
				Range	4.50
				Minimum	1.79
				Maximum	6.30
				Sum	21.11
				Count	5

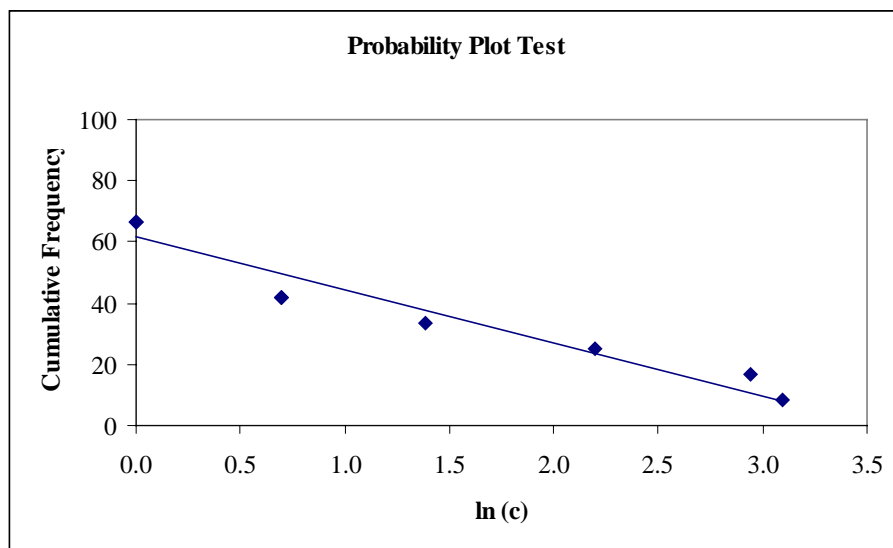
CV =	0.48
CS  =	0.22
ln (UCL) =	6.14
UCL =	462.1



**Ethyl Benzene Data for Source Area 7 (Soil Vapor)**

C (ppm)	ln (C) (ppm)	Rank	CF	<i>Statistics</i>	
22.0	3.1	1	8.3		
19.0	2.9	2	16.7	Mean	1.06
9.0	2.2	3	25.0	Standard Error	0.36
4.0	1.4	4	33.3	Median	0.69
2.0	0.7	5	41.7	Mode	0.00
2.0	0.7	5	41.7	Standard Deviation	1.18
2.0	0.7	5	41.7	Sample Variance	1.40
1.0	0.0	8	66.7	Kurtosis	-0.76
1.0	0.0	8	66.7	Skewness	0.85
1.0	0.0	8	66.7	Range	3.09
1.0	0.0	8	66.7	Minimum	0.00
				Maximum	3.09
				Sum	11.70
				Count	11

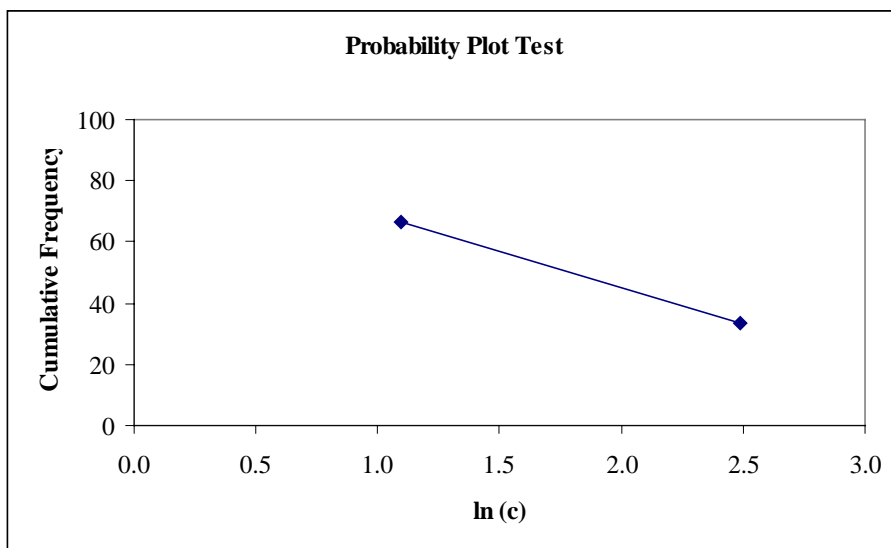
CV =	1.11
CS  =	0.73
ln (UCL) =	1.71
UCL =	5.5



**Ethyl Benzene Data for Source Area 8 (Soil Vapor)**

C (ppm)	ln (C) (ppm)	Rank	CF	<i>Statistics</i>	
12.0	2.5	1	33.3	Mean	1.79
3.0	1.1	2	66.7	Standard Error	0.69
				Median	1.79
				Mode	#N/A
				Standard Deviation	0.98
				Sample Variance	0.96
				Kurtosis	-
				Skewness	-
				Range	1.39
				Minimum	1.10
				Maximum	2.48
				Sum	3.58
				Count	2

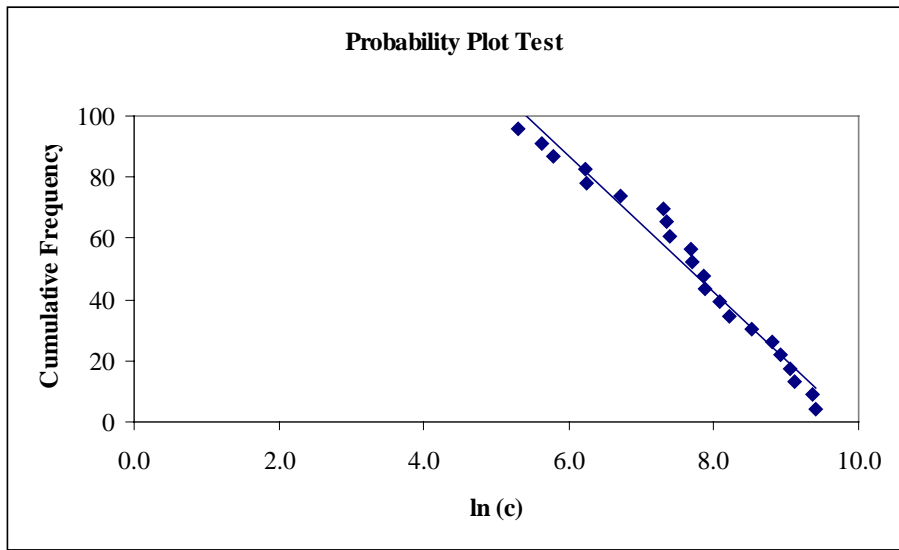
CV =	0.55
CS  =	-
ln (UCL) =	6.17
UCL =	477.4



**Naphthalene Data for Source Area 1 (Soil)**

C (µg/kg)	ln (C) (µg/kg)	Rank	CF	<i>Statistics</i>	
12200.0	9.4	1	4.3	Mean	7.66
11500.0	9.4	2	8.7	Standard Error	0.27
9090.0	9.1	3	13.0	Median	7.78
8500.0	9.0	4	17.4	Mode	#N/A
7460.0	8.9	5	21.7	Standard Deviation	1.25
6730.0	8.8	6	26.1	Sample Variance	1.55
5050.0	8.5	7	30.4	Kurtosis	-0.86
3695.0	8.2	8	34.8	Skewness	-0.39
3241.0	8.1	9	39.1	Range	4.11
2620.0	7.9	10	43.5	Minimum	5.30
2560.0	7.8	11	47.8	Maximum	9.41
2240.0	7.7	12	52.2	Sum	168.6
2180.0	7.7	13	56.5	Count	22
1640.0	7.4	14	60.9		
1561.0	7.4	15	65.2		
1481.0	7.3	16	69.6		
828.0	6.7	17	73.9		
512.0	6.2	18	78.3		
501.0	6.2	19	82.6		
327.0	5.8	20	87.0		
280.0	5.6	21	91.3		
201.0	5.3	22	95.7		

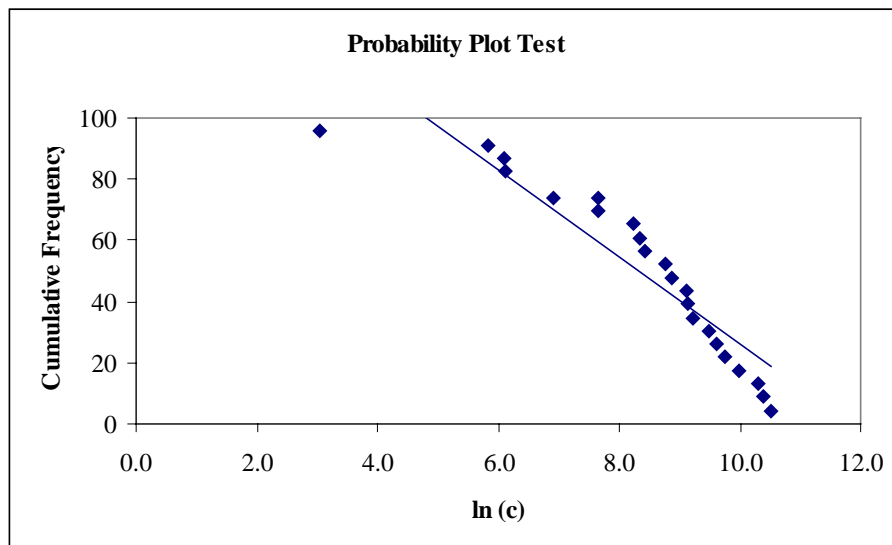
CV =	0.16
CS  =	0.37
ln (UCL) =	8.12
UCL =	3357.8



**Naphthalene Data for Source Area 2 (Soil)**

C ( $\mu\text{g}/\text{kg}$ )	ln (C) ( $\mu\text{g}/\text{kg}$ )	Rank	CF	<i>Statistics</i>	
37000.0	10.5	1	4.3		
32000.0	10.4	2	8.7	Mean	8.34
30000.0	10.3	3	13.0	Standard Error	0.39
22000.0	10.0	4	17.4	Median	8.82
17000.0	9.7	5	21.7	Mode	7.65
15000.0	9.6	6	26.1	Standard Deviation	1.83
13000.0	9.5	7	30.4	Sample Variance	3.34
10000.0	9.2	8	34.8	Kurtosis	1.92
9200.0	9.1	9	39.1	Skewness	-1.29
9100.0	9.1	10	43.5	Range	7.47
7100.0	8.9	11	47.8	Minimum	3.04
6400.0	8.8	12	52.2	Maximum	10.52
4600.0	8.4	13	56.5	Sum	183.4
4200.0	8.3	14	60.9	Count	22
3800.0	8.2	15	65.2		
2100.0	7.6	16	69.6		
2100.0	7.6	17	73.9		
1000.0	6.9	17	73.9		
460.0	6.1	19	82.6		
440.0	6.1	20	87.0		
340.0	5.8	21	91.3		
21.0	3.0	22	95.7		

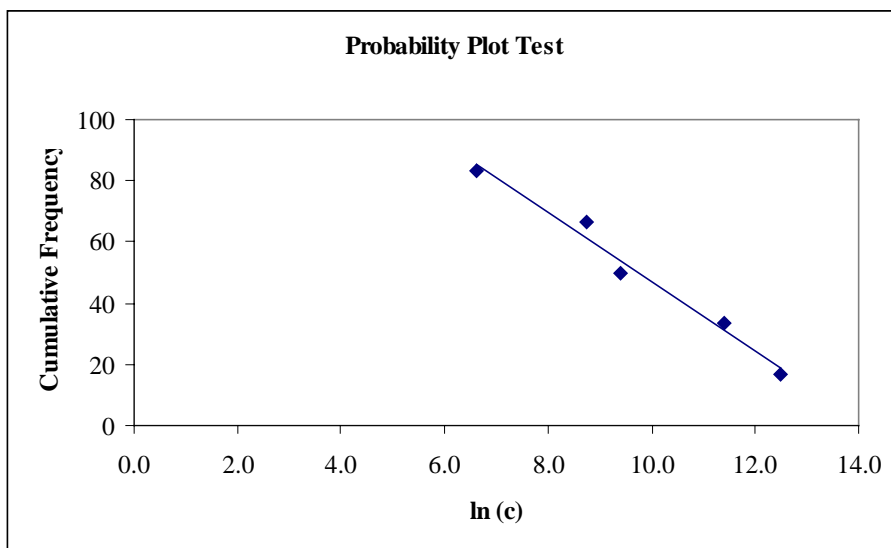
CV =	0.22
CS  =	1.20
ln (UCL) =	9.01
UCL =	8171.8



**Naphthalene Data for Source Area 3 (Soil)**

C ( $\mu\text{g}/\text{kg}$ )	ln (C) ( $\mu\text{g}/\text{kg}$ )	Rank	CF	<i>Statistics</i>	
262000.0	12.5	1	16.7	Mean	9.73
89100.0	11.4	2	33.3	Standard Error	1.03
12200.0	9.4	3	50.0	Median	9.41
6210.0	8.7	4	66.7	Mode	#N/A
756.0	6.6	5	83.3	Standard Deviation	2.29
				Sample Variance	5.26
				Kurtosis	-0.85
				Skewness	-0.19
				Range	5.85
				Minimum	6.63
				Maximum	12.48
				Sum	48.64
				Count	5

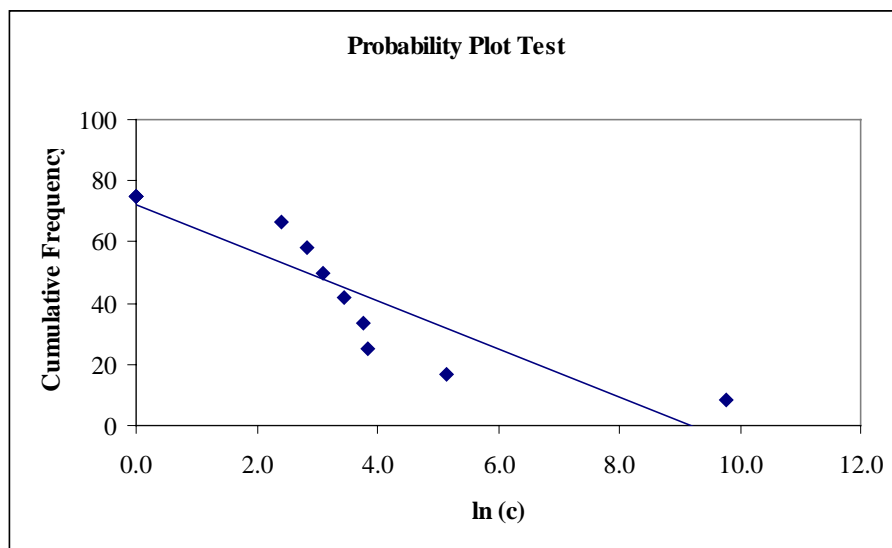
CV =	0.24
CS  =	0.13
ln (UCL) =	11.92
UCL =	149586



**Naphthalene Data for Source Area 4 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
17460.0	9.8	1	8.3		
170.0	5.1	2	16.7	Mean	3.11
46.0	3.8	3	25.0	Standard Error	0.85
43.0	3.8	4	33.3	Median	3.09
31.0	3.4	5	41.7	Mode	0.00
22.0	3.1	6	50.0	Standard Deviation	2.81
17.0	2.8	7	58.3	Sample Variance	7.91
11.0	2.4	8	66.7	Kurtosis	2.46
1.0	0.0	9	75.0	Skewness	1.17
1.0	0.0	9	75.0	Range	9.77
1.0	0.0	9	75.0	Minimum	0.00
				Maximum	9.77
				Sum	34.25
				Count	11

CV =	0.90
CS  =	1.00
ln (UCL) =	4.65
UCL =	104.6

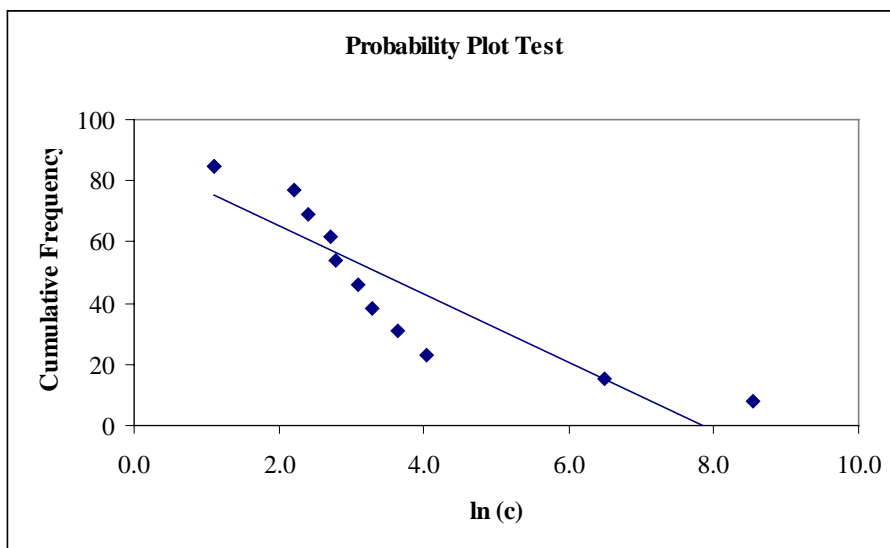




**Naphthalene Data for Source Area 5 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
5182.0	8.6	1	7.7		
662.0	6.5	2	15.4	Mean	3.45
57.0	4.0	3	23.1	Standard Error	0.62
38.0	3.6	4	30.8	Median	2.93
27.0	3.3	5	38.5	Mode	1.10
22.0	3.1	6	46.2	Standard Deviation	2.14
16.0	2.8	7	53.8	Sample Variance	4.60
15.0	2.7	8	61.5	Kurtosis	2.15
11.0	2.4	9	69.2	Skewness	1.46
9.0	2.2	10	76.9	Range	7.45
3.0	1.1	11	84.6	Minimum	1.10
3.0	1.1	11	84.6	Maximum	8.55
				Sum	41.39
				Count	12

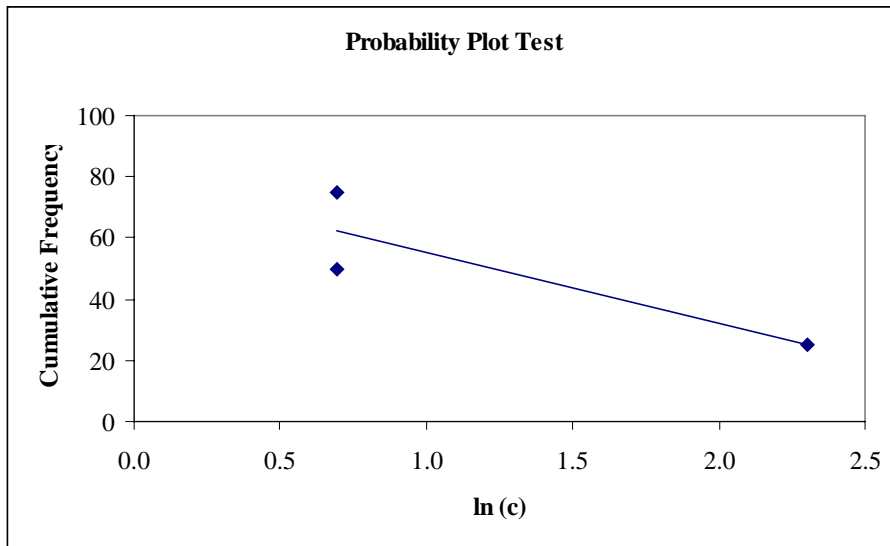
CV =	0.62
CS  =	1.27
ln (UCL) =	4.56
UCL =	95.7



**Naphthalene Data for Source Area 6 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
10.0	2.3	1	25.0	Mean	1.23
2.0	0.7	2	50.0	Standard Error	0.54
2.0	0.7	3	75.0	Median	0.69
				Mode	0.69
				Standard Deviation	0.93
				Sample Variance	0.86
				Kurtosis	-
				Skewness	1.73
				Range	1.61
				Minimum	0.69
				Maximum	2.30
				Sum	3.69
				Count	3

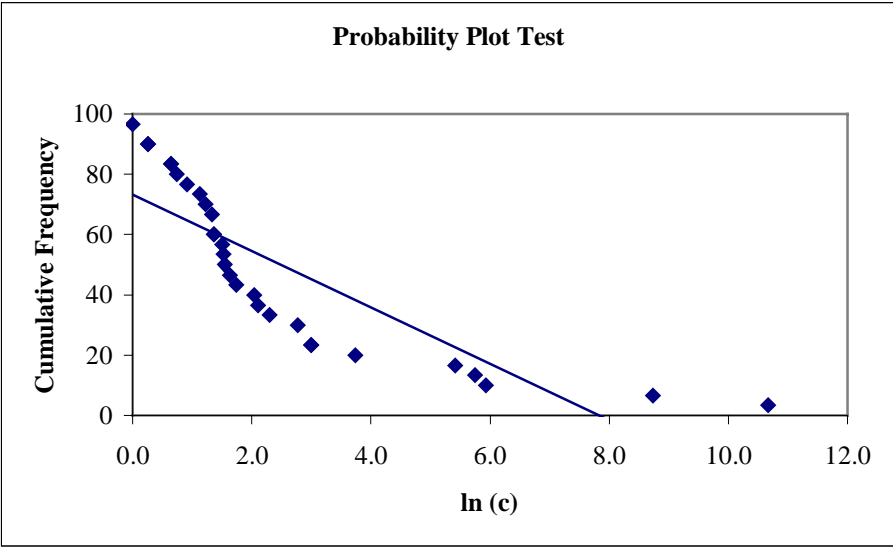
CV =	0.76
CS  =	0.71
ln (UCL) =	2.80
UCL =	16.4



**Toluene Data for Source Area 1 (Soil)**

C (µg/kg)	ln (C) (µg/kg)	Rank	CF	<i>Statistics</i>	
43100.0	10.7	1	3.3		
6180.0	8.7	2	6.7	Mean	2.53
377.0	5.9	3	10.0	Standard Error	0.47
313.0	5.7	4	13.3	Median	1.55
224.0	5.4	5	16.7	Mode	3.00
42.0	3.7	6	20.0	Standard Deviation	2.53
20.0	3.0	7	23.3	Sample Variance	6.40
20.0	3.0	7	23.3	Kurtosis	3.50
16.0	2.8	9	30.0	Skewness	1.90
10.0	2.3	10	33.3	Range	10.67
8.2	2.1	11	36.7	Minimum	0.00
7.7	2.0	12	40.0	Maximum	10.67
5.7	1.7	13	43.3	Sum	73.27
5.1	1.6	14	46.7	Count	29
4.7	1.5	15	50.0		
4.6	1.5	16	53.3		
4.5	1.5	17	56.7		
3.9	1.4	18	60.0		
3.9	1.4	18	60.0		
3.8	1.3	20	66.7		
3.4	1.2	21	70.0		
3.1	1.1	22	73.3		
2.5	0.9	23	76.7		
2.1	0.7	24	80.0		
1.9	0.6	25	83.3		
1.9	0.6	25	83.3		
1.3	0.3	27	90.0		
1.3	0.3	27	90.0		
1.0	0.0	29	96.7		

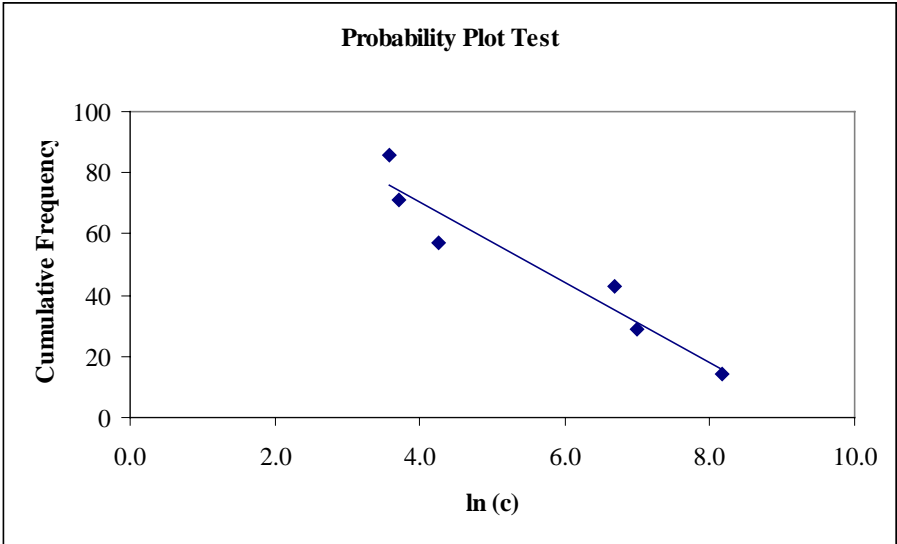
CV =	1.00
CS  =	1.80
ln (UCL) =	3.33
UCL =	27.8



**Toluene Data for Source Area 2 (Soil)**

C (µg/kg)	ln (C) (µg/kg)	Rank	CF	<i>Statistics</i>	
3500.0	8.2	1	14.3		
1100.0	7.0	2	28.6	Mean	5.57
800.0	6.7	3	42.9	Standard Error	0.80
71.0	4.3	4	57.1	Median	5.47
41.0	3.7	5	71.4	Mode	#N/A
36.0	3.6	6	85.7	Standard Deviation	1.95
				Sample Variance	3.82
				Kurtosis	-2.36
				Skewness	0.22
				Range	4.58
				Minimum	3.58
				Maximum	8.16
				Sum	33.41
				Count	6

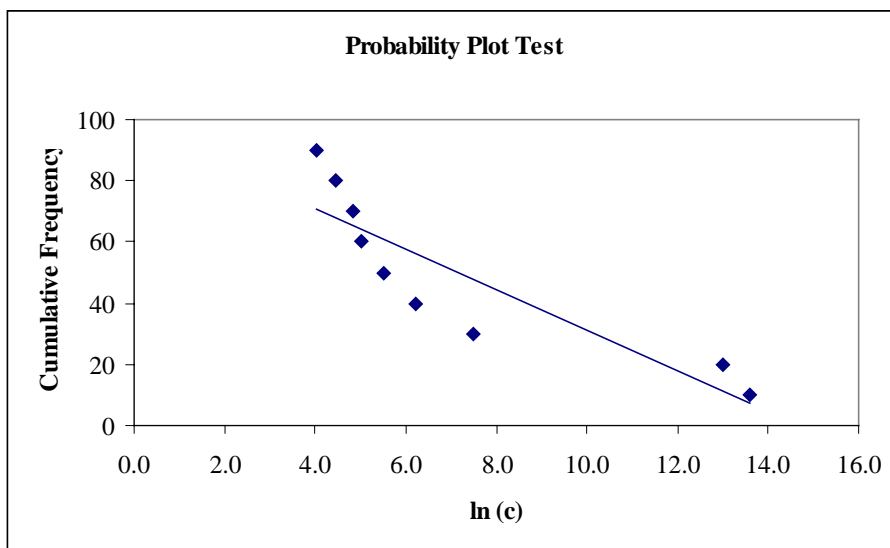
CV =	0.35
CS  =	0.16
ln (UCL) =	7.18
UCL =	1307.9



**Toluene Data for Source Area 3 (Soil)**

C ( $\mu\text{g}/\text{kg}$ )	ln (C) ( $\mu\text{g}/\text{kg}$ )	Rank	CF	<i>Statistics</i>	
808000.0	13.6	1	10.0	Mean	7.13
436000.0	13.0	2	20.0	Standard Error	1.22
1810.0	7.5	3	30.0	Median	5.50
499.0	6.2	4	40.0	Mode	#N/A
244.0	5.5	5	50.0	Standard Deviation	3.65
152.0	5.0	6	60.0	Sample Variance	13.29
125.0	4.8	7	70.0	Kurtosis	0.18
86.0	4.5	8	80.0	Skewness	1.32
56.0	4.0	9	90.0	Range	9.58
				Minimum	4.03
				Maximum	13.60
				Sum	64.13
				Count	9

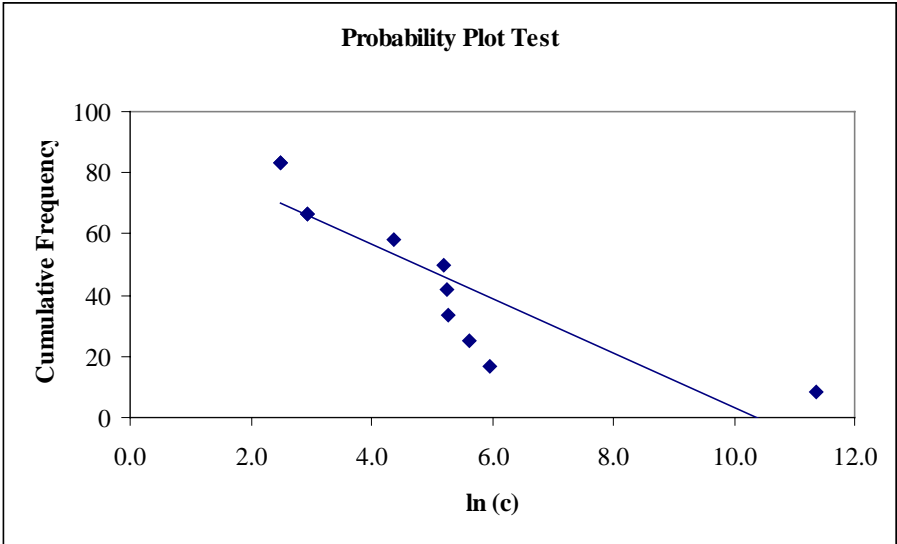
CV =	0.51
CS  =	1.09
ln (UCL) =	9.39
UCL =	11922.7



**Toluene Data for Source Area 4 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
85420.0	11.4	1	8.3		
384.0	6.0	2	16.7	Mean	4.90
274.0	5.6	3	25.0	Standard Error	0.76
197.0	5.3	4	33.3	Median	5.18
192.0	5.3	5	41.7	Mode	2.94
178.0	5.2	6	50.0	Standard Deviation	2.51
80.0	4.4	7	58.3	Sample Variance	6.32
19.0	2.9	8	66.7	Kurtosis	4.26
19.0	2.9	8	66.7	Skewness	1.75
12.0	2.5	10	83.3	Range	8.87
12.0	2.5	10	83.3	Minimum	2.48
				Maximum	11.36
				Sum	53.88
				Count	11

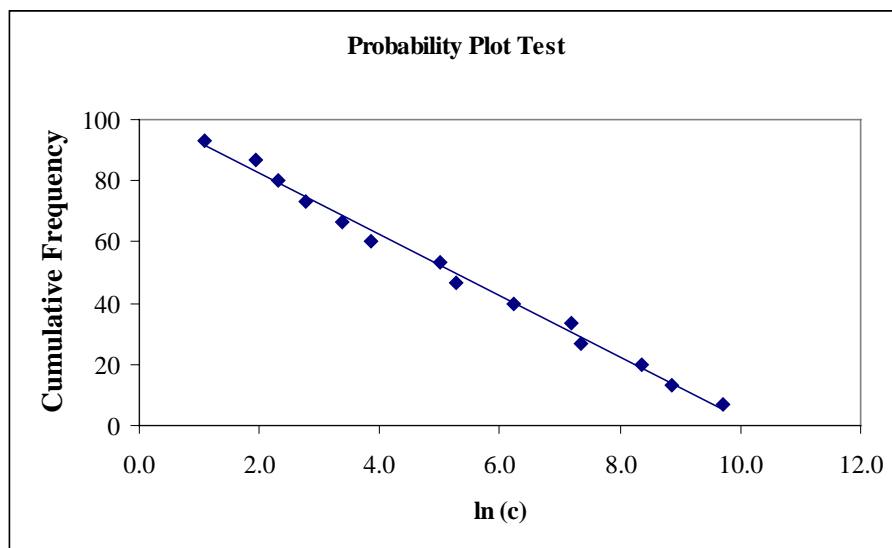
CV =	0.51
CS  =	1.50
ln (UCL) =	6.27
UCL =	529.6



**Toluene Data for Source Area 5 (Groundwater)**

C ( $\mu\text{g/L}$ )	ln (C) ( $\mu\text{g/L}$ )	Rank	CF	<i>Statistics</i>	
16554.0	9.7	1	6.7		
6967.0	8.8	2	13.3	Mean	5.23
4216.0	8.3	3	20.0	Standard Error	0.74
1559.0	7.4	4	26.7	Median	5.13
1302.0	7.2	5	33.3	Mode	#N/A
509.0	6.2	6	40.0	Standard Deviation	2.77
193.0	5.3	7	46.7	Sample Variance	7.69
147.0	5.0	8	53.3	Kurtosis	-1.28
48.0	3.9	9	60.0	Skewness	0.12
29.0	3.4	10	66.7	Range	8.62
16.0	2.8	11	73.3	Minimum	1.10
10.0	2.3	12	80.0	Maximum	9.71
7.0	1.9	13	86.7	Sum	73.28
3.0	1.1	14	93.3	Count	14

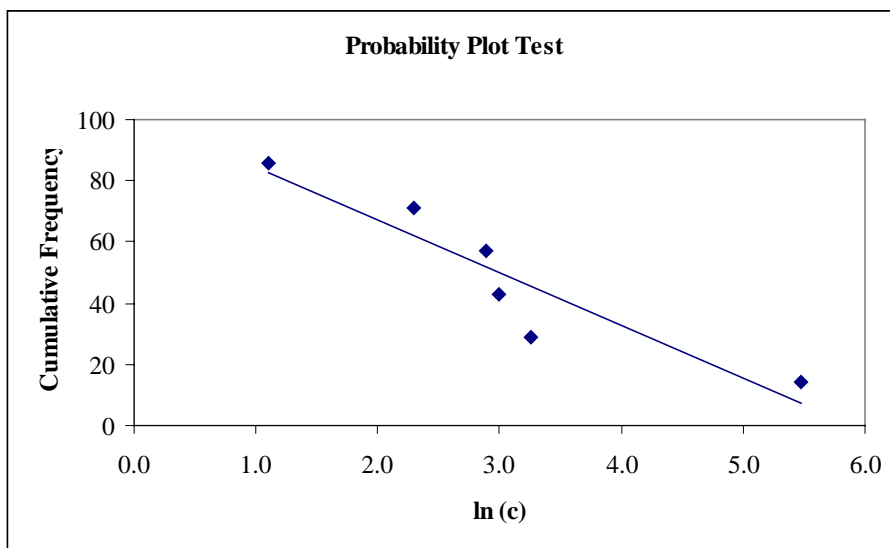
CV =	0.53
CS  =	0.11
ln (UCL) =	6.55
UCL =	697.2





**Toluene Data for Source Area 6 (Groundwater)**

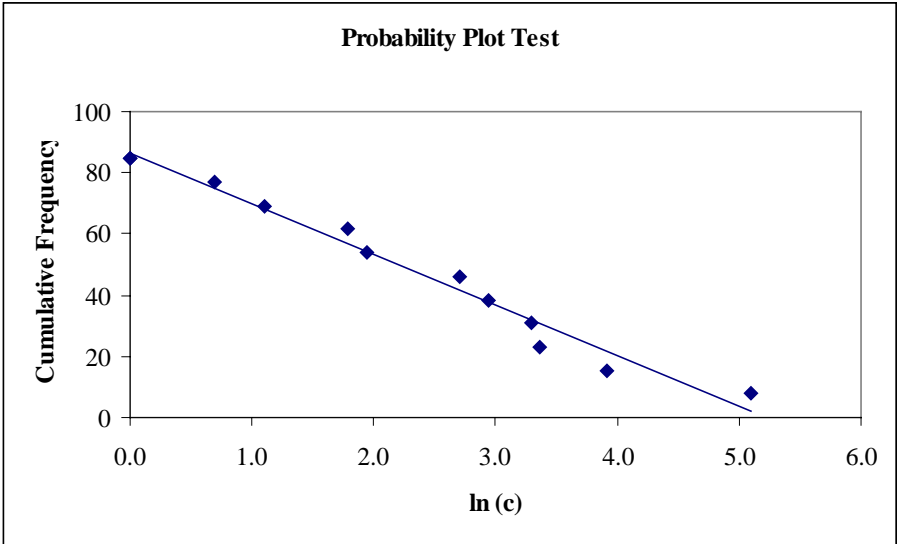
C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>										
238.0	5.5	1	14.3	Mean	3.00	<table border="1"> <tr> <td>CV =</td> <td>0.48</td> </tr> <tr> <td> CS  =</td> <td>0.58</td> </tr> <tr> <td>ln (UCL) =</td> <td>4.18</td> </tr> <tr> <td>UCL =</td> <td>65.6</td> </tr> </table>	CV =	0.48	CS  =	0.58	ln (UCL) =	4.18	UCL =	65.6
CV =	0.48													
CS  =	0.58													
ln (UCL) =	4.18													
UCL =	65.6													
26.0	3.3	2	28.6	Standard Error	0.59									
20.0	3.0	3	42.9	Median	2.94									
18.0	2.9	4	57.1	Mode	#N/A									
10.0	2.3	5	71.4	Standard Deviation	1.43									
3.0	1.1	6	85.7	Sample Variance	2.06									
				Kurtosis	2.11									
				Skewness	0.79									
				Range	4.37									
				Minimum	1.10									
				Maximum	5.47									
				Sum	18.02									
				Count	6									



**Toluene Data for Source Area 7 (Soil Vapor)**

C (ppm)	ln (C) (ppm)	Rank	CF	<i>Statistics</i>	
162.0	5.1	1	7.7		
50.0	3.9	2	15.4	Mean	2.24
29.0	3.4	3	23.1	Standard Error	0.46
27.0	3.3	4	30.8	Median	2.33
19.0	2.9	5	38.5	Mode	0.00
15.0	2.7	6	46.2	Standard Deviation	1.60
7.0	1.9	7	53.8	Sample Variance	2.55
6.0	1.8	8	61.5	Kurtosis	-0.77
3.0	1.1	9	69.2	Skewness	0.09
2.0	0.7	10	76.9	Range	5.09
1.0	0.0	11	84.6	Minimum	0.00
1.0	0.0	11	84.6	Maximum	5.09
				Sum	26.84
				Count	12

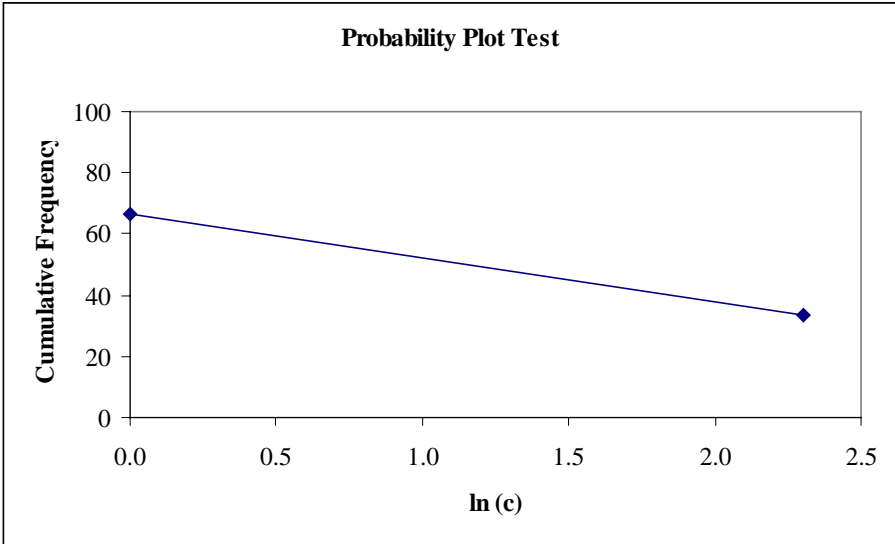
CV =	0.71
CS  =	0.08
ln (UCL) =	3.06
UCL =	21.4



**Toluene Data for Source Area 8 (Soil Vapor)**

C (ppm)	ln (C) (ppm)	Rank	CF	<i>Statistics</i>	
10.0	2.3	1	33.3	Mean	1.15
1.0	0.0	2	66.7	Standard Error	1.15
				Median	1.15
				Mode	#N/A
				Standard Deviation	1.63
				Sample Variance	2.65
				Kurtosis	-
				Skewness	-
				Range	2.30
				Minimum	0.00
				Maximum	2.30
				Sum	2.30
				Count	2

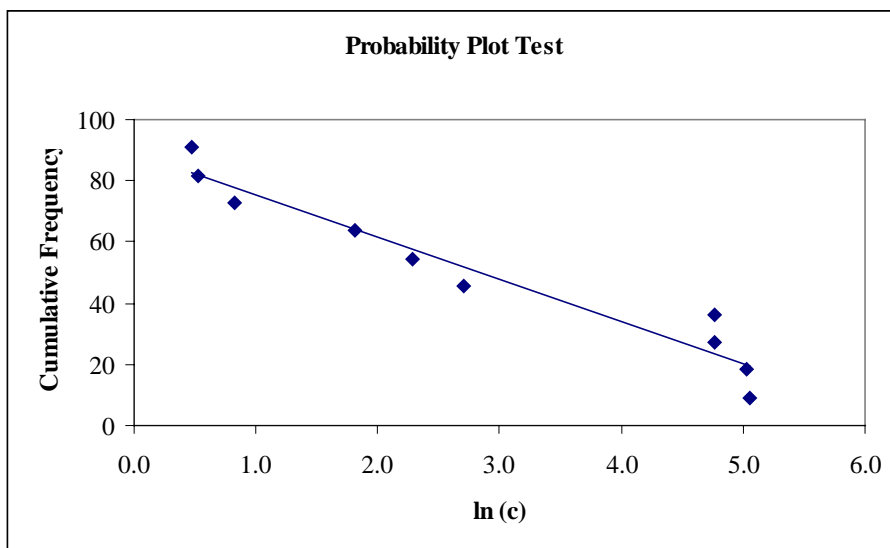
CV =	1.41
CS  =	-
ln (UCL) =	8.42
UCL =	4539.4



**Xylenes Data for Source Area 1 (Soil)**

C ( $\mu\text{g}/\text{kg}$ )	ln (C) ( $\mu\text{g}/\text{kg}$ )	Rank	CF	<i>Statistics</i>	
156.0	5.0	1	9.1		
152.0	5.0	2	18.2	Mean	2.82
118.0	4.8	3	27.3	Standard Error	0.61
117.0	4.8	4	36.4	Median	2.50
15.0	2.7	5	45.5	Mode	#N/A
9.8	2.3	6	54.5	Standard Deviation	1.93
6.1	1.8	7	63.6	Sample Variance	3.72
2.3	0.8	8	72.7	Kurtosis	-1.94
1.7	0.5	9	81.8	Skewness	0.04
1.6	0.5	10	90.9	Range	4.58
				Minimum	0.47
				Maximum	5.05
				Sum	28.24
				Count	10

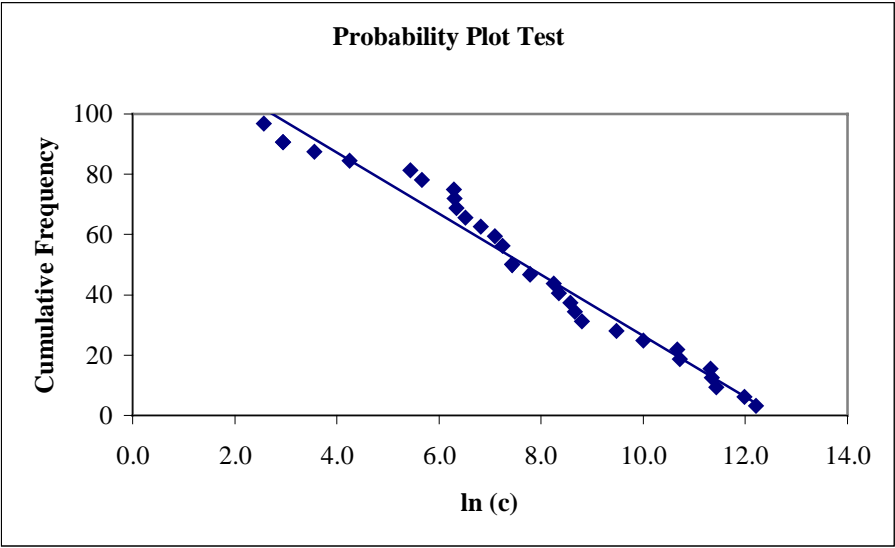
CV =	0.68
CS  =	0.03
ln (UCL) =	3.94
UCL =	51.5



**Xylenes Data for Source Area 2 (Soil)**

C (µg/kg)	ln (C) (µg/kg)	Rank	CF	<i>Statistics</i>	
200000.0	12.2	1	3.1		
160000.0	12.0	2	6.3	Mean	7.69
92000.0	11.4	3	9.4	Standard Error	0.49
84000.0	11.3	4	12.5	Median	7.44
82000.0	11.3	5	15.6	Mode	7.44
45000.0	10.7	6	18.8	Standard Deviation	2.75
43000.0	10.7	7	21.9	Sample Variance	7.58
22000.0	10.0	8	25.0	Kurtosis	-0.70
13000.0	9.5	9	28.1	Skewness	-0.16
6600.0	8.8	10	31.3	Range	9.64
5800.0	8.7	11	34.4	Minimum	2.56
5300.0	8.6	12	37.5	Maximum	12.21
4200.0	8.3	13	40.6	Sum	238.4
3800.0	8.2	14	43.8	Count	31
2400.0	7.8	15	46.9		
1700.0	7.4	16	50.0		
1700.0	7.4	16	50.0		
1400.0	7.2	18	56.3		
1200.0	7.1	19	59.4		
920.0	6.8	20	62.5		
680.0	6.5	21	65.6		
570.0	6.3	22	68.8		
550.0	6.3	23	71.9		
540.0	6.3	24	75.0		
290.0	5.7	25	78.1		
230.0	5.4	26	81.3		
70.0	4.2	27	84.4		
35.0	3.6	28	87.5		
19.0	2.9	29	90.6		
19.0	2.9	29	90.6		
13.0	2.6	31	96.9		

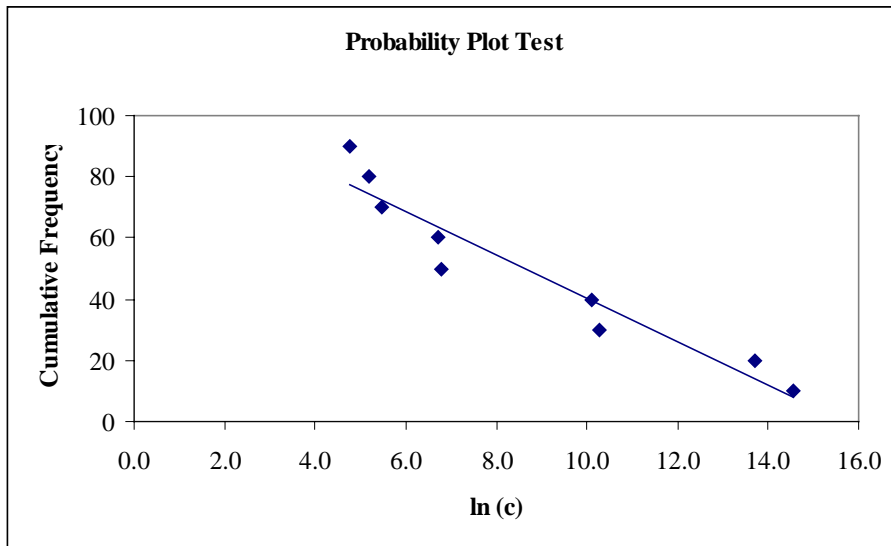
CV =	0.36
CS  =	0.15
ln (UCL) =	8.53
UCL =	5061.5



**Xylenes Data for Source Area 3 (Soil)**

C (µg/kg)	ln (C) (µg/kg)	Rank	CF	<i>Statistics</i>	
2060000	14.5	1	10.0		
887000	13.7	2	20.0	Mean	8.61
28800	10.3	3	30.0	Standard Error	1.23
24000	10.1	4	40.0	Median	6.76
866.0	6.8	5	50.0	Mode	#N/A
829.0	6.7	6	60.0	Standard Deviation	3.69
240.0	5.5	7	70.0	Sample Variance	13.65
178.0	5.2	8	80.0	Kurtosis	-1.13
118.0	4.8	9	90.0	Skewness	0.66
				Range	9.77
				Minimum	4.77
				Maximum	14.54
				Sum	77.50
				Count	9

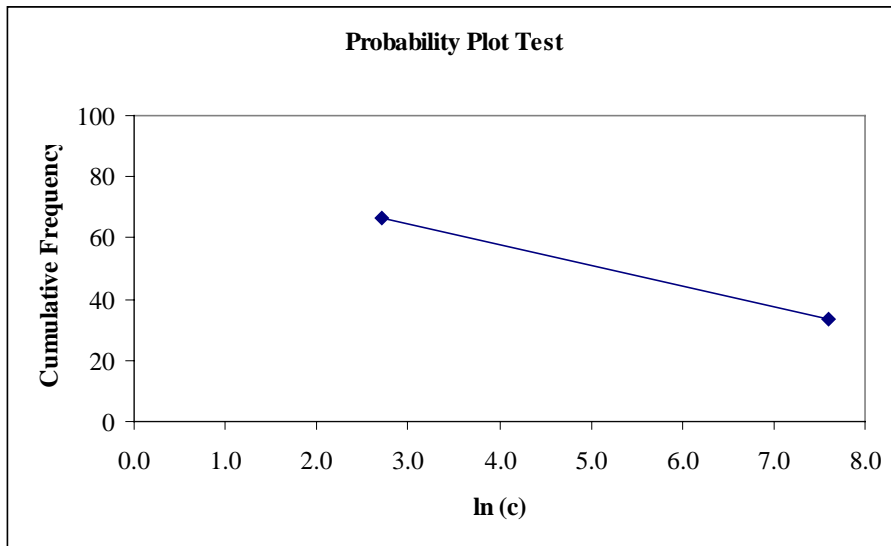
CV =	0.43
CS  =	0.54
ln (UCL) =	10.90
UCL =	54301.5



**Xylenes Data for Source Area 4 (Groundwater)**

C (µg/L)	ln (C) (µg/L)	Rank	CF	<i>Statistics</i>	
2000.0	7.6	1	33.3	Mean	5.15
15.0	2.7	2	66.7	Standard Error	2.45
				Median	5.15
				Mode	#N/A
				Standard Deviation	3.46
				Sample Variance	11.97
				Kurtosis	-
				Skewness	-
				Range	4.89
				Minimum	2.71
				Maximum	7.60
				Sum	10.31
				Count	2

CV = 0.67  
 |CS| = -  
 ln (UCL) = 20.60  
 UCL = 8.9E+08

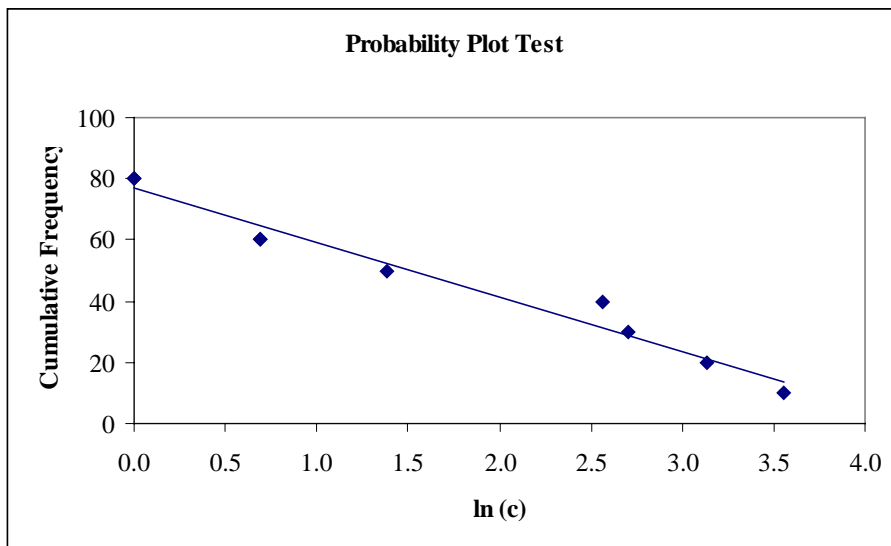




**Xylenes Data for Source Area 7 (Soil Vapor)**

C (ppm)	ln (C) (ppm)	Rank	CF	<i>Statistics</i>	
35.0	3.6	1	10.0		
23.0	3.1	2	20.0	Mean	1.64
15.0	2.7	3	30.0	Standard Error	0.46
13.0	2.6	4	40.0	Median	1.39
4.0	1.4	5	50.0	Mode	0.69
2.0	0.7	6	60.0	Standard Deviation	1.38
2.0	0.7	6	60.0	Sample Variance	1.89
1.0	0.0	8	80.0	Kurtosis	-1.83
1.0	0.0	8	80.0	Skewness	0.12
				Range	3.56
				Minimum	0.00
				Maximum	3.56
				Sum	14.74
				Count	9

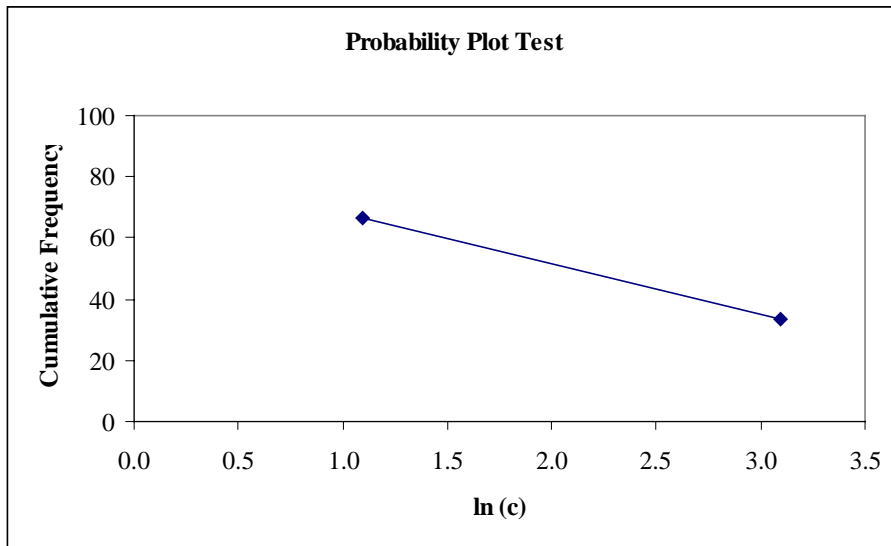
CV =	0.84
CS  =	0.10
ln (UCL) =	2.49
UCL =	12.1



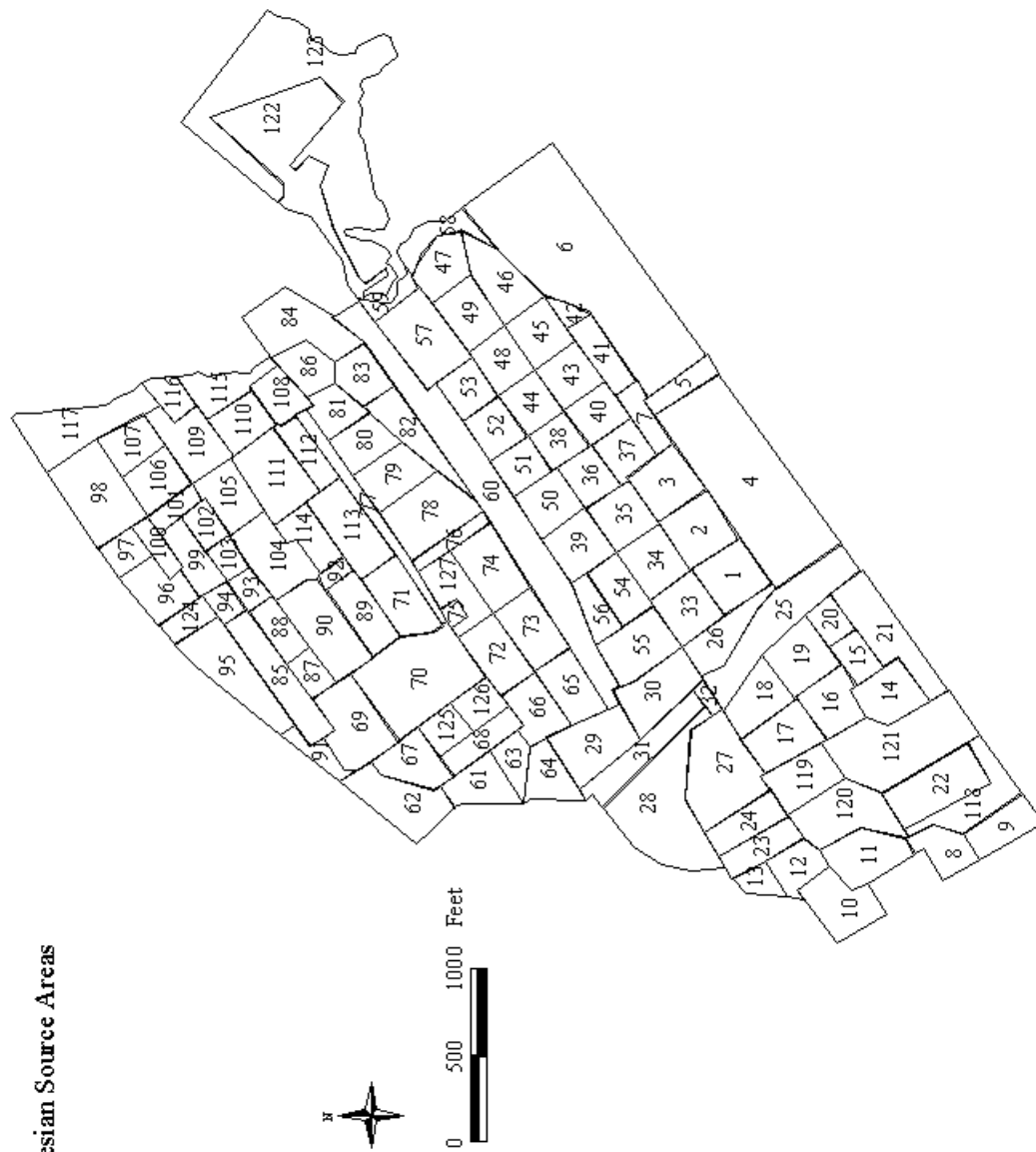
**Xylenes Data for Source Area 8 (Soil Vapor)**

C (ppm)	ln (C) (ppm)	Rank	CF	<i>Statistics</i>	
22.0	3.1	1	33.3	Mean	2.09
3.0	1.1	2	66.7	Standard Error	1.00
				Median	2.09
				Mode	#N/A
				Standard Deviation	1.41
				Sample Variance	1.98
				Kurtosis	-
				Skewness	-
				Range	1.99
				Minimum	1.10
				Maximum	3.09
				Sum	4.19
				Count	2

CV = 0.67  
|CS| = -  
ln (UCL) = 8.38  
UCL = 4380.5



Bayesian Source Areas



Bayesian Results by Area												
ID	n	s	$\mu_p'$	$\sigma_p'$	$q'$	$r'$	$q''$	$r''$	$\mu_p''$	$\sigma_p''$	Pprior	Pupdated
1	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
2	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
3	2	0	0.250	0.250	0.600	1.8	0.600	3.8	0.136	0.148	21.1	32.4
4	2	0	0.005	0.005	0.990	197.0	0.990	199.0	0.005	0.005	90.9	91.0
5	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
6	1	0	0.005	0.005	0.990	197.0	0.990	198.0	0.005	0.005	90.9	90.9
7	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
8	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
9	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
10	1	0	0.100	0.100	0.818	7.4	0.818	8.4	0.089	0.089	33.9	36.6
11	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
12	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
13	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
14	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
15	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
16	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
17	7	0	0.250	0.250	0.600	1.8	0.600	8.8	0.064	0.076	21.1	48.6
18	3	0	0.250	0.250	0.600	1.8	0.600	4.8	0.111	0.124	21.1	36.6
19	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
20	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
21	1	0	0.250	0.250	0.600	1.8	0.600	2.8	0.176	0.182	21.1	27.4
22	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
23	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
24	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
25	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
26	2	0	0.250	0.250	0.600	1.8	0.600	3.8	0.136	0.148	21.1	32.4
27	0	0	0.500	0.289	1.000	1.0	1.000	1.0	0.500	0.289	4.8	4.8
28	14	0	0.005	0.005	0.990	197.0	0.990	211.0	0.005	0.005	90.9	91.4
29	46	0	0.050	0.050	0.905	17.2	0.905	63.2	0.014	0.015	49.7	78.0
30	14	4	0.500	0.289	1.000	1.0	5.000	11.0	0.312	0.112	4.8	0.9
31	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
32	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
33	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
34	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
35	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
36	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
37	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
38	0	0	0.750	0.250	1.286	0.4	1.286	0.4	0.750	0.263	0.9	0.9
39	0	0	0.750	0.250	1.286	0.4	1.286	0.4	0.750	0.263	0.9	0.9
40	12	2	0.100	0.100	0.818	7.4	2.818	17.4	0.140	0.075	33.9	12.4

ID	n	s	$\mu_p'$	$\sigma_p'$	$q'$	$r'$	$q''$	$r''$	$\mu_p''$	$\sigma_p''$	$P_{\text{prior}}$	$P_{\text{updated}}$
41	0	0	0.500	0.289	1.000	1.0	1.000	1.0	0.500	0.289	4.8	4.8
42	0	0	0.500	0.289	1.000	1.0	1.000	1.0	0.500	0.289	4.8	4.8
43	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
44	0	0	0.750	0.250	1.286	0.4	1.286	0.4	0.750	0.263	0.9	0.9
45	0	0	0.500	0.289	1.000	1.0	1.000	1.0	0.500	0.289	4.8	4.8
46	2	0	0.500	0.289	1.000	1.0	1.000	3.0	0.250	0.194	4.8	13.0
47	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
48	2	2	0.500	0.289	1.000	1.0	3.000	1.0	0.750	0.194	4.8	0.1
49	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
50	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
51	2	0	0.750	0.250	1.286	0.4	1.286	2.4	0.346	0.219	0.9	6.1
52	0	0	0.750	0.250	1.286	0.4	1.286	0.4	0.750	0.263	0.9	0.9
53	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
54	1	0	0.750	0.250	1.286	0.4	1.286	1.4	0.474	0.259	0.9	3.4
55	0	0	0.750	0.250	1.286	0.4	1.286	0.4	0.750	0.263	0.9	0.9
56	0	0	0.750	0.250	1.286	0.4	1.286	0.4	0.750	0.263	0.9	0.9
57	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
58	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
59	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
60	0	0	0.500	0.289	1.000	1.0	1.000	1.0	0.500	0.289	4.8	4.8
61	0	0	0.500	0.289	1.000	1.0	1.000	1.0	0.500	0.289	4.8	4.8
62	1	0	0.500	0.289	1.000	1.0	1.000	2.0	0.333	0.236	4.8	9.1
63	0	0	0.050	0.050	0.905	17.2	0.905	17.2	0.050	0.050	49.7	49.7
64	0	0	0.050	0.050	0.905	17.2	0.905	17.2	0.050	0.050	49.7	49.7
65	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
66	0	0	0.050	0.050	0.905	17.2	0.905	17.2	0.050	0.050	49.7	49.7
67	4	0	0.050	0.050	0.905	17.2	0.905	21.2	0.041	0.041	49.7	54.8
68	0	0	0.050	0.050	0.905	17.2	0.905	17.2	0.050	0.050	49.7	49.7
69	3	0	0.250	0.250	0.600	1.8	0.600	4.8	0.111	0.124	21.1	36.6
70	25	0	0.005	0.005	0.990	197.0	0.990	222.0	0.004	0.004	90.9	91.8
71	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
72	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
73	1	0	0.250	0.250	0.600	1.8	0.600	2.8	0.176	0.182	21.1	27.4
74	1	0	0.005	0.005	0.990	197.0	0.990	198.0	0.005	0.005	90.9	90.9
75	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
76	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
77	0	0	0.500	0.289	1.000	1.0	1.000	1.0	0.500	0.289	4.8	4.8
78	3	0	0.750	0.250	1.286	0.4	1.286	3.4	0.273	0.186	0.9	8.8
79	3	0	0.750	0.250	1.286	0.4	1.286	3.4	0.273	0.186	0.9	8.8
80	7	0	0.250	0.250	0.600	1.8	0.600	8.8	0.064	0.076	21.1	48.6
81	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9

ID	n	s	$\mu_p'$	$\sigma_p'$	$q'$	$r'$	$q''$	$r''$	$\mu_p''$	$\sigma_p''$	$P_{\text{prior}}$	$P_{\text{updated}}$
82	4	0	0.005	0.005	0.990	197.0	0.990	201.0	0.005	0.005	90.9	91.0
83	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
84	2	0	0.005	0.005	0.990	197.0	0.990	199.0	0.005	0.005	90.9	91.0
85	48	1	0.100	0.100	0.818	7.4	1.818	54.4	0.032	0.023	33.9	56.8
86	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
87	3	0	0.100	0.100	0.818	7.4	0.818	10.4	0.073	0.075	33.9	41.3
88	5	0	0.100	0.100	0.818	7.4	0.818	12.4	0.062	0.064	33.9	45.3
89	1	0	0.750	0.250	1.286	0.4	1.286	1.4	0.474	0.259	0.9	3.4
90	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
91	3	0	0.005	0.005	0.990	197.0	0.990	200.0	0.005	0.005	90.9	91.0
92	1	0	0.500	0.289	1.000	1.0	1.000	2.0	0.333	0.236	4.8	9.1
93	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
94	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
95	2	0	0.750	0.250	1.286	0.4	1.286	2.4	0.346	0.219	0.9	6.1
96	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
97	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
98	11	0	0.100	0.100	0.818	7.4	0.818	18.4	0.043	0.045	33.9	54.6
99	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
100	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
101	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
102	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
103	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
104	1	0	0.250	0.250	0.600	1.8	0.600	2.8	0.176	0.182	21.1	27.4
105	5	0	0.250	0.250	0.600	1.8	0.600	6.8	0.081	0.094	21.1	43.3
106	2	0	0.250	0.250	0.600	1.8	0.600	3.8	0.136	0.148	21.1	32.4
107	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
108	2	0	0.005	0.005	0.990	197.0	0.990	199.0	0.005	0.005	90.9	91.0
109	5	0	0.250	0.250	0.600	1.8	0.600	6.8	0.081	0.094	21.1	43.3
110	2	0	0.100	0.100	0.818	7.4	0.818	9.4	0.080	0.081	33.9	39.0
111	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
112	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
113	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
114	0	0	0.050	0.050	0.905	17.2	0.905	17.2	0.050	0.050	49.7	49.7
115	0	0	0.050	0.050	0.905	17.2	0.905	17.2	0.050	0.050	49.7	49.7
116	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
117	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
118	2	0	0.005	0.005	0.990	197.0	0.990	199.0	0.005	0.005	90.9	91.0
119	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
120	0	0	0.250	0.250	0.600	1.8	0.600	1.8	0.250	0.235	21.1	21.1
121	24	0	0.250	0.250	0.600	1.8	0.600	25.8	0.023	0.028	21.1	70.7
122	2	0	0.500	0.289	1.000	1.0	1.000	3.0	0.250	0.194	4.8	13.0

ID	n	s	$\mu_p'$	$\sigma_p'$	$q'$	$r'$	$q''$	$r''$	$\mu_p''$	$\sigma_p''$	$p_{\text{prior}}$	$p_{\text{updated}}$
123	25	0	0.005	0.005	0.990	197.0	0.990	222.0	0.004	0.004	90.9	91.8
124	0	0	0.005	0.005	0.990	197.0	0.990	197.0	0.005	0.005	90.9	90.9
125	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
126	0	0	0.100	0.100	0.818	7.4	0.818	7.4	0.100	0.099	33.9	33.9
127	1	0	0.005	0.005	0.990	197.0	0.990	198.0	0.005	0.005	90.9	90.9

## REFERENCES

- Air Survey Corporation (ASC). *Marcus Hook Refinery GIS Basemap Program Proposals*. Sterling, VA. 1997.
- American Society for Testing and Materials (ASTM). *Standard Provisional Guide for Risk-Based Corrective Action*. PS104-98. 1998.
- Ang, A. H-S. and W. H. Tang. *Probability Concepts in Engineering Planning and Design: Volume 1 - Basic Principles*. New York: John Wiley & Sons, 1975.
- Aspinall, R. J. and D. M. Pearson. "Data Quality and Spatial Analysis: Analytical Use of GIS for Ecological Modeling." *GIS and Environmental Modeling: Progress and Research Issues*. Ed. Goodchild *et al.* Fort Collins: GIS World Books, 1996. 35-39.
- Baker, C. P., M. D. Bradley, and S. M. Kaczor Bobiak. "Wellhead Protection Area Delineation: Linking Flow Models with GIS." *Journal of Water Resources Planning and Management*. 119 (1993): 275-87.
- Benaman, J., N. E. Armstrong, and D. R. Maidment. *Modeling of Dissolved Oxygen in the Houston Ship Channel using WASP5 and Geographic Information Systems*. CRWR Online Report The University of Texas at Austin, Department of Civil Engineering. Austin, Texas. 1996.
- Bond, J. A., T. L. Leavens, M. J. Seaton, and M. A. Medinsky. "Predicting the Toxicity of Chemical Mixtures Based on Knowledge of Individual Components." *CIIT Activities*. 17 (1997). Internet address: [www.ciit.org/ACT97/ACTIVITIESDEC97/dec97.html](http://www.ciit.org/ACT97/ACTIVITIESDEC97/dec97.html).
- BP Exploration & Oil Inc. (BP). *Guidance Document for the Risk-Based Decision Process for Corrective Action Sites*. Cleveland: BP Exploration & Oil Inc. 1997.
- Brigham Young University (BYU) - Engineering Computer Graphics Laboratory (ECGL). *Groundwater Modeling System for the PC Version 2.1*. Provo, Utah. 1998.
- Brigham Young University (BYU) - Engineering Computer Graphics Laboratory (ECGL). *Groundwater Modeling System Version 2.1 Reference Manual*. Provo, Utah. 1998.
- Brigham Young University (BYU) - Engineering Computer Graphics Laboratory (ECGL). *Groundwater Modeling System Version 2.1 Tutorials*. Provo, Utah. 1998.
- Camp, C. V. and M. C. Brown. "GIS Procedure for Developing Three-Dimensional Subsurface Profile." *Journal of Computing in Civil Engineering*. 7 (1993): 296-309.
- Chen, Z., G. H. Huang, and A. Chakma. "Integrated Environmental Risk Assessment Through a GIS-Based Decision-Support System." Published On-line in *ESRI International User Conference 1998 Proceedings*. Internet address: [www.esri.com/library/userconf/proc98/PROCEED/TO800/PAP769/P769.HTM](http://www.esri.com/library/userconf/proc98/PROCEED/TO800/PAP769/P769.HTM).



- Corwin, D. L., K. Loague, and T. R. Ellsworth. "GIS-based Modeling of Non-Point Source Pollutants in the Vadose Zone." *Journal of Soil and Water Conservation*. 53 (1998): 34-38.
- DePinto, J. V., J. F. Atkinson, H. W. Calkins, P. J. Densham, W. Guan, H. Lin, F. Xia, P. W. Rodgers, and T. Slaweki. "Development of GEO-WAMS: A Modeling Support System to Integrate GIS with Watershed Analysis Models." *GIS and Environmental Modeling: Progress and Research Issues*. Ed. Goodchild *et al.* Fort Collins: GIS World Books, 1996. 271-76.
- El-Kadi, A. I. "Use of a Geographic Information System in Site-Specific Ground-Water Modeling." *Ground Water*. 32 (1994): 617-25.
- Environmental Protection Agency (EPA). *Reducing Risk: Setting Priorities and Strategies for Environmental Protection: The Report of the Science Advisory Board, Relative Risk Reduction Strategies Committee to William K. Reilly*. Washington: The Science Advisory Board, 1992.
- Environmental Protection Agency Region III (EPA Region III). *Risk-Based Concentration Table*. 1998. Internet address: [www.epa.gov/reg3hwmd/risk/riskmenu.htm](http://www.epa.gov/reg3hwmd/risk/riskmenu.htm).
- Environmental Systems Research Institute (ESRI). *ArcView GIS Version 3.1*. Redlands, CA. 1998.
- Environmental Systems Research Institute (ESRI). *ARC/INFO Version 7.2.1*. Redlands, CA. 1998.
- Evans, T. A. and D. R. Maidment. *A Spatial and Statistical Assessment of the Vulnerability of Texas Groundwater to Nitrate Contamination*. CRWR Online Report 95-4. The University of Texas at Austin, Department of Civil Engineering. Austin, Texas. 1995.
- Falkner, E. *Aerial Mapping: Methods and Applications*. Boca Raton, FL: CRC Press, 1995.
- Fredericks, J. W., J. W. Labadie, and J. M. Altenhofen. "Decision Support System for Conjunctive Stream-Aquifer Management." *Journal of Water Resources Planning and Management*. 124 (1998): 69-78.
- French, D. P. and M. Reed. "Integrated Environmental Impact Model and GIS for Oil and Chemical Spills." *GIS and Environmental Modeling: Progress and Research Issues*. Ed. Goodchild *et al.* Fort Collins: GIS World Books, 1996. 197-98.
- Glass, G. E., B. S. Schwartz, J. M. Morgan, D. T. Johnson, P. M. Noy, and I. Ebenezer. "Environmental Risk Factors for Lyme Disease Identified with Geographic Information Systems." *American Journal Public Health*. 85 (1995): 944-48.
- Harris, J. S. "Evaluating Possible Human Exposure Pathways to Populations Relative to Hazardous Material Sites." Published On-line in *ESRI International User Conference 1997 Proceedings*. Internet address: [www.esri.com/library/userconf/proc97/PROC97/TO450/PAP409/P409.HTM](http://www.esri.com/library/userconf/proc97/PROC97/TO450/PAP409/P409.HTM)

- Hay Wilson, L. "A Spatial Risk Assessment Methodology for Environmental Risk-Based Decision Making at Large, Complex Facilities." Dissertation Proposal. The University of Texas at Austin, Department of Civil Engineering. Austin, Texas. 1998.
- Hellweger, F. and D. R. Maidment. *HEC-PREPRO: A GIS Preprocessor for Lumped Parameter Hydrologic Modeling Programs*. CRWR Online Report 97-8. The University of Texas at Austin, Department of Civil Engineering. Austin, Texas. 1997.
- Hendrix, W. G. and D. A. Buckley. "Use of Geographic Information Systems for Selection of Sites for Land Application of Sewage Waste." *Journal of Soil and Water Conservation*. 47 (1992): 271-75.
- Hill, G. "System Helps Identify Property Risks." *American City and County*. 111.11 (1996): 10.
- How, C. F. H., R. L. Corsi, and D. R. Maidment. *Fugitive Emissions of VOCs from Industrial Sewer Networks: Integration of naUTilus and ArcView*. CRWR Online Report 98-2. The University of Texas at Austin, Department of Civil Engineering. Austin, Texas. 1998.
- Jankowski, P. and G. Haddock. "Integrated Nonpoint Source Pollution Modeling System." *GIS and Environmental Modeling: Progress and Research Issues*. Ed. Goodchild *et al.* Fort Collins: GIS World Books, 1996. 209-11.
- Jones, N. L., S. G. Wright, and D. R. Maidment. "Watershed Delineation with Triangle-Based Terrain Models." *Journal of Hydraulic Engineering*. 116 (1990): 1232-1251.
- Kao, J. J., W. Y. Chen, H. Y. Lin, and S. J. Guo. "Network Expert Geographic Information System for Landfill Siting." *Journal of Computing in Civil Engineering*. 10 (1996): 307-17.
- Kim, J. "Spatial Analysis of the Marcus Hook Refinery." Term Project Report. 1998. Internet address: [www.ce.utexas.edu/stu/kimj/proposal.html](http://www.ce.utexas.edu/stu/kimj/proposal.html).
- Koerner, L. N., L. Hay Wilson, A. P. Romanek, J. R. Rocco, S. L. Sharp, and R. B. Gilbert. "Maximizing the Value of Information in Risk-Based Decision-Making: Challenges and Solutions." American Nuclear Society Conference, Pasco WA., April 3-5, 1998. Internet address: [www.ce.utexas.edu/prof/maidment/grad/romanek/research/anspaper.pdf](http://www.ce.utexas.edu/prof/maidment/grad/romanek/research/anspaper.pdf).
- Koerner, L. N. *Development of a Site Conceptual Model Using a Relational Database*. Masters Thesis. The University of Texas at Austin, Department of Civil Engineering. Austin, Texas. 1998. Internet address: [internetcity.crrw.utexas.edu/gis/gisenv98/library/koerner/abstract.html](http://internetcity.crrw.utexas.edu/gis/gisenv98/library/koerner/abstract.html).
- Lang, L. *Managing Natural Resources with GIS*. Redlands, CA: ESRI Press, 1998.
- Lober, D. J. "Resolving the Siting Impasse: Modeling Social and Environmental Locational Criteria with a Geographic Information System." *Journal of the American Planning Association*. 61 (1995): 482-95.
- Maidment, D. R. "Environmental Modeling within GIS." *GIS and Environmental Modeling: Progress and Research Issues*. Ed. Goodchild *et al.* Fort Collins: GIS World Books, 1996. 315-323.

- Maidment, D. R., L. Hay Wilson, J. E. Till, J. R. Rocco, A. P. Romanek, and J. Benaman. "GIS Enviro 98: Implementing GIS in Environmental Risk Based Decisions." *ESRI International User Conference 1998 Pre-Conference Seminar*. Internet address: [www.ce.utexas.edu/prof/grad/romanek/gisenviro98/maidment/sld001.htm](http://www.ce.utexas.edu/prof/grad/romanek/gisenviro98/maidment/sld001.htm).
- Maidment, D. R. *GIS in Water Resources, Exercise 3: Map Projections*. Internet address: [www.ce.utexas.edu/prof/maidment/giswr98/ex398/mapproj.htm](http://www.ce.utexas.edu/prof/maidment/giswr98/ex398/mapproj.htm). 1998.
- McDonnell, P. W. *Introduction to Map Projections*. 2<sup>nd</sup> ed. Rancho Cordova, CA: Landmark Enterprises, 1991.
- McKinney, D. C. and H. L. Tsai. "Multigrid Methods in GIS Grid-Cell-Based Modeling Environment." *Journal of Computing in Civil Engineering*. 10 (1996): 25-30.
- Microsoft Corporation. *Microsoft Access 97*. Redmond, WA. 1996.
- Miller, M. R., P. R. Coleman, D. L. Pack, and R. C. Durfee. "GIS and Risk Assessment: A Fruitful Combination." Published On-line in *ESRI International User Conference 1996 Proceedings*. Internet address: [www.esri.com/library/userconf/proc96/TO50/PAP028/P28.HTM](http://www.esri.com/library/userconf/proc96/TO50/PAP028/P28.HTM).
- National Academy of Sciences (NAS). *Risk Assessment in the Federal Government: Managing the Process*. Washington: National Academy Press, 1983.
- Olivera, F., R. J. Charbeneau, and D. R. Maidment. *Spatially Distributed Modeling of Storm Runoff and Non-Point Source Pollution Using Geographic Information Systems*. CRWR Online Report 96-4. The University of Texas at Austin, Department of Civil Engineering. Austin, Texas. 1996.
- Pennsylvania Department of Environmental Protection (PADEP). *Pennsylvania's Land Recycling Program*. 1997. Internet address: [www.dep.state.pa.us/dep/deputate/airwaste/wm/landrecy/MANUAL/Manual.htm](http://www.dep.state.pa.us/dep/deputate/airwaste/wm/landrecy/MANUAL/Manual.htm)
- Poiani, K. A. "GIS-Based Nonpoint Source Pollution Modeling: Considerations for Wetlands." *Journal of Soil and Water Conservation*. 50 (1995): 613-19.
- Quenzer, A. M. and D.R. Maidment. *A GIS Assessment of the Total Loads and Water Quality for the Corpus Christi National Estuary Program*. CRWR Online Report 98-1. The University of Texas at Austin, Department of Civil Engineering. Austin, Texas. 1998.
- Rifai, H. S., L. A. Hendricks, K. Kilburn, and P. B. Bedient. "A Geographic Information System User Interface for Delineating Wellhead Protection Areas." *Ground Water*. 31 (1993): 480-88.
- Rocco, J. R. "A Regulatory Perspective for Risk-Based Decisions for Corrective Action." Lecture. 1998. Internet address: [www.crrw.utexas.edu/gis/gisenv98/class/risk/lecture/rocco/sld001.htm](http://www.crrw.utexas.edu/gis/gisenv98/class/risk/lecture/rocco/sld001.htm).
- Ross, M. A. and P. D. Tara. "Integrated Hydrologic Modeling with Geographic Information

- Systems." *Journal of Water Resources Planning and Management*. 119 (1993): 129-141.
- Saunders, W. K. and D. R. Maidment. *A GIS Assessment of Nonpoint Source Pollution in the San Antonio - Nueces Coastal Basin*. CRWR Online Report 96-1. The University of Texas at Austin, Department of Civil Engineering. Austin, Texas. 1996.
- Shea, C., W. Grayman, D. Darden, R. M. Males, and P. Sushinsky. "Integrated GIS and Hydrologic Modeling for Countywide Drainage Study." *Journal of Water Resources Planning and Management*. 119 (1993): 112-28.
- Sunday, T. U. "Interactive Modeling of Ground-Water Vulnerability within a Geographic Information System Environment." *Ground Water*. 34 (1996): 618-27.
- Synder, J. P. *Map Projections - A Working Manual*. U.S. Geological Survey Professional Paper 1395. Washington: U.S. Government Printing Office, 1987.
- Texas Natural Resource Conservation Commission (TNRCC). *Proposed Texas Risk Reduction Program Rule*. 1998. Internet address: [www.tnrcc.state.tx.us/waste/riskrul3.htm](http://www.tnrcc.state.tx.us/waste/riskrul3.htm)
- Watkins, D. W., D. C. McKinney, D. R. Maidment, and M. D. Lin. "Use of Geographic Information Systems in Ground-Water Flow Modeling." *Journal of Water Resources Planning and Management*. 122 (1996): 88-96.
- Yi, G. C. "Development of Ohio's GIS-Based Wetlands Inventory." *Journal of Soil and Water Conservation*. 49 (1994): 23-28.

## VITA

Andrew Paul Romanek was born in Houston, Texas on July 3, 1975, the son of Mollie Elizabeth Romanek and Joseph John Romanek, II. After completing his work at Strake Jesuit College Preparatory, Houston, Texas, in 1993, he entered the University of Notre Dame in Notre Dame, Indiana. He graduated *magna cum laude* from Notre Dame in May 1997, receiving a degree of Bachelor of Science in Civil Engineering. In August 1997, he entered the Graduate School at the University of Texas at Austin.

Permanent Address: 5328 Longmont Dr.  
Houston, Texas 77056

This thesis was typed by the author.