Geophysical Data Fusion for Subsurface Imaging

RECEIVED
MAY 20 1997
OSTI

**Final Report
October 1995**

MASTER

Work Performed Under Contract No.: DE-AC21-92MC29106

For **DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED** ^{HH}

U.S. Department of Energy
Office of Environmental Management
Office of Technology Development
1000 Independence Avenue
Washington, DC 20585

U.S. Department of Energy
Office of Fossil Energy
Federal Energy Technology Center
Morgantown Site
P.O. Box 880
Morgantown, West Virginia 26507-0880

By
Coleman Research Corporation
9302 Lee Highway
8th Floor, Suite 800
Fairfax, Virginia 22031

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

ACKNOWLEDGEMENT

This work was performed under contract number DE-AC21-92MC29106 for the Department of Energy, Morgantown Energy Technology Center. We appreciate the interest and guidance provided by the Contracting Officer's Technical Representatives, Mr. Vijendra P. Kothari of Morgantown, WV and Mr. J. Keith Westhusing of Laramie, WY. We would also wish to thank Supervisory Geologist Dr. Carol Eddy-Dilek of the Environmental Sciences Section, Savannah River Site (Savannah River Technology Center), Aiken, SC for her most detailed and continuous professional assistance throughout this project. Further, we are grateful for Dr. Caroline Purdy's coordination of technical support at Hq DOE and avid interest in seeing this emerging technology transferred to the full Federal community and eventually the private sector.

The principle authors of this report were Mark Blohm; William Hatch; Pieter Hoekstra, PhD PG; David Porter, PhD; Todd Stong, PhD, PE; and James Vandergraft, PhD.

TABLE OF CONTENTS

DISCLAIMER	i
ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	vii
LIST OF TABLES	xi
1.0 EXECUTIVE SUMMARY	1-1
2.0 INTRODUCTION	2-1
2.1 Technical Approach For Data Fusion Methodology Development	2-1
2.2 Geophysical Methods	2-4
2.3 Technical Approach to Research	2-5
2.4 Test Site Characteristics	2-5
2.5 Outline of Report	2-7
2.6 Satisfaction of Phase 3 Objectives	2-7
3.0 DATA FUSION RESEARCH	3-1
3.1 Data Fusion Theory	3-1
3.1.1 Background	3-1
3.1.2 Mathematical Problem	3-11
3.1.3 Data Fusion Approach	3-13
3.1.3.1 Markov Random Field (MRF)	3-13
3.1.3.2 Information Array	3-15
3.1.3.3 Square Root Information Smoother (SRIS)	3-18
3.1.3.4 Generalized Expectation Maximization (GEM) Method	3-20
3.2 Data Fusion System (DFS) Architecture	3-21
3.3 Level 0 Architecture	3-24
3.4 Level 1 Architecture	3-26
3.5 Level 2 Architecture - Analysis Engines	3-29
4.0 NEAR-SURFACE GEOPHYSICAL RESEARCH AND DEVELOPMENT ACHIEVEMENTS	4-1
4.1 Introduction	4-1
4.2 Time Domain Electromagnetic (TDEM)	4-2
4.3 Seismic	4-12

4.3.1	Resolution of Thin Layers	4-13
4.3.2	Vibratory Sources	4-15
4.3.3	Shear Wave Surveys	4-16
4.4	Integrated Interpretation	4-19
4.5	Summary of Advances in Geophysical Technologies	4-22
5.0	NEAR-SURFACE GEOPHYSICAL FIELD DATA COLLECTION AND PROCESSING	5-1
5.1	General	5-1
5.2	Data Acquisition	5-4
5.3	Data Processing	5-8
6.0	GEOPHYSICAL DATA INTERPRETATION	6-1
6.1	General Geology	6-1
6.1.1	Geology, A&M Area	6-3
6.1.2	Geophysical Well Log Data	6-4
6.2	TDEM Results	6-11
6.2.1	General	6-11
6.2.2	Repeatability in TDEM Measurements	6-13
6.2.4	Detectability of Clay Layers with TDEM	6-13
6.2.5	Geoelectric Cross Sections (Lines 1 and 2)	6-16
6.2.6	Conductance Contour Map	6-20
6.3	Reflection Seismic Survey	6-20
6.3.1	Shear Wave Seismic Survey	6-22
6.3.2	Checkshot Surveys	6-22
6.3.3	Results and Interpretation	6-26
6.4	Integrated Interpretation	6-37
6.4.1	Resolution of Geophysical Methods	6-37
6.4.2	Faulting	6-39
6.4.3	TDEM and Seismic	6-51
6.4.4	Stratigraphic Interpretation	6-53
7.0	DATA FUSION APPLICATIONS	7-1
7.1	Fusion Processing Summary	7-1
7.2	Savannah River Site A&M Area	7-2
7.2.1	A&M Line 1	7-6
7.2.1.1	Fusion Overview	7-6
7.2.1.2	Top Of Ellenton Well Picks	7-7
7.2.1.3	Data Description	7-8
7.2.1.4	Seismic Reflection Forward Model	7-9
7.2.1.5	Line 1 Fusion Results	7-11

7.2.1.6	Comparison of Fusion Results with TDEM Ridge Regression	7-13
7.2.1.6.1	Comparison Using Line 1 Out-of-Loop Vertical Field Data	7-13
7.2.1.6.2	Comparison Using Simulated Central Loop TDEM Data	7-16
7.2.1.7	Effect of User Input Spatial Correlation Coefficients	7-19
7.2.1.8	Detection of Geologic Changes	7-21
7.2.2	A&M Line 2	7-24
7.2.2.1	Fusion Overview	7-24
7.2.2.2	Top of Ellenton Well Picks	7-24
7.2.2.3	Data Description	7-24
7.2.2.4	Seismic Reflection Forward Model	7-27
7.2.2.5	Line 2 Fusion Results	7-29
7.3	Fernald South Field Site Characterization	7-35
7.3.1	Three Dimensional Site Lithology	7-36
7.3.2	Contaminant Distribution Superimposed on Lithology	7-36
7.3.3	Greater Miami Aquifer Groundwater Elevation Superimposed on Lithology	7-36
7.4	Marshalltown Iowa Manufactured Gas Plant Bedrock Characterization	7-40
7.5	Hanford 200 West Area Characterization	7-45
7.6	Groundwater Model Calibration at Pantex Using Data Fusion Modeling	7-48
7.6.1	Introduction	7-48
7.6.2	Pantex Model Calibration	7-48
7.6.3	Conclusions	7-56
8.0	CONCLUSIONS	8-1
8.1	Data Fusion Conclusions	8-1
8.2	Geophysical Conclusions	8-3
9.0	RECOMMENDATIONS FOR FUTURE DEVELOPMENTS	9-1
9.1	Data Fusion	9-1
9.2	Near-Surface Geophysical Exploration	9-2
9.2.1	Time Domain Electromagnetic Induction (TDEM)	9-3
9.2.1.1	Noise Rejection	9-3
9.2.1.2	2-D Modeling	9-4
9.2.1.3	Physical Properties from TDEM Data	9-4
9.2.2	Seismic	9-4
9.2.2.1	Polarization Analysis of Multi-Component Seismic Data	9-4
9.2.2.2	Rotation of Shear Wave Data to Natural Coordinates	9-10
10.0	REFERENCES	10-1

11.0 GLOSSARY	11-1
----------------------------	-------------

APPENDICES

Appendix A	Manufacturers Specifications	A-1
	Geonics TEM 47 TDEM Transmitter	A-2
	Geonics Protem TDEM Receiver	A-2
Appendix B	Manufacturers Specifications	B-1
	Geonics EM39 Borehole Conductivity Probe	B2
Appendix C	Manufacturers Specifications	C1
	Mount Sopris Natural Gamma/SP/SPR Probe	C2
	Mount Sopris Digital Logger	C3
Appendix D	Manufacturers Specifications	D1
	iVi "Minivib"	D2
Appendix E	Manufacturers Specifications	E1
	OYO DHS-1 Data Acquisition System	E2
	OYO GS-40D Rotating Coil Geophone	E6
Appendix F	Nonlinear Least Squares Estimates	F-1

LIST OF FIGURES

Figure 2.1	Data Fusion Role in Decision Analysis	2-3
Figure 3.1	Data Fusion Role in Decision Analysis	3-3
Figure 3.2	Remedial Decision Process Supported By Data Fusion	3-4
Figure 3.3	Data Fusion Uses Bayesian Statistics	3-9
Figure 3.4	Spatial Stochastic Process V on Regular Grid	3-14
Figure 3.5	Context Diagram (Level 0)	3-25
Figure 3.6	Level 1 Dataflow Diagram	3-27
Figure 3.7	Analysis Engines: Structure Chart	3-30
Figure 3.8	Analysis Engine: Inversion	3-32
Figure 3.9	Analysis Engine: Geo-Statistics	3-33
Figure 3.10	Analysis Engine: Fusion	3-35
Figure 3.11	Analysis Engine: Hypothesis Testing and Tuning	3-36
Figure 3.12	Analysis Engine: Data Worth Simulation	3-37
Figure 3.13	Analysis Engine: Remediation Alternative Simulation	3-39
Figure 4.1	Well MSB26A Resistivity Log and Lithologic Interpretation	4-3
Figure 4.2	System Waveforms In TDEM	4-4
Figure 4.3	Relation Between Onset of Late Stage and Electrical Resistivity and Transmitter Loop Size	4-6
Figure 4.4	Behavior Of emf_z and emf_x On A Profile Through The Center Of A Square Transmitter Loop	4-7
Figure 4.5	Comparison Of Amplitude Of emf_z and emf_x As A Function Of Time . .	4-9
Figure 4.6	TDEM Apparent Resistivity as Time Curves Derived From Horizontal and Vertical Magnetic Fields Measurement at the Center and Outside the Same Horizontal Transmitter Loop. Line 1, A&M Area, SRS.	4-11
Figure 4.7	Effects Of Wavelet Frequency On Seismic Resolution	4-14
Figure 4.8	Comparison Of P-Wave Amplitude Spectra	4-17
Figure 4.9	Comparison Of P-Wave Source Types	4-18
Figure 4.10	Schematic Presentation Of Particle Motion Of Compressional and Shear Waves	4-20
Figure 4.11	Detectable Thickness Of A Layer (Wavelength/4) As A Function Of Seismic Signal Frequency	4-21
Figure 5.1	Layout of Geophysical Survey Lines - A&M Area, SRS	5-2
Figure 5.2	TDEM Transmitter Loop Layout, Line 1	5-5
Figure 5.3	Schematic Illustration of TDEM Measurements Made From Each Transmitter Loop	5-6
Figure 6.1	Correlation Between Chronostratigraphic, Lithostratigraphic, Hydrostratigraphic and Facies-Based Units	6-2

Figure 6.19	Line 3. Shear wave reflection seismic section. Top: uninterpreted section. Bottom: interpreted section. Well control indicated.	6-32
Figure 6.20	P-wave and S-wave checkshots in Wells MSB21TA and MSB76C: Note depth to water table and increase in P-wave velocity to 6,000 ft/sec below it. S-wave velocity is unchanged.	6-40
Figure 6.21	Basement Fault Map: Seismic lines and faults that offset the top of basement reflection are shown. Seismic lines SRS-7 and 12 are emphasized (modified from Domoracki, 1994).	6-41
Figure 6.22	Portion of Seismic Line SRS-7, showing red, blue, and yellow markers; and the signature of the Pen Branch Fault (PBF) (from Domoracki, 1994).	6-42
Figure 6.23	Velocity Profile Borehole DRB-10: Interval velocity, average velocity, and RMS velocity (from Domoracki, 1994).	6-43
Figure 6.24	Portion of P-wave seismic line SRS-12 showing interpreted faults F1 and F3, and red horizons (Domoracki, 1994), Fault 2 (SRS Hydrogeologic Cross Section C-C', 10/18/92), well locations, and BGD S-wave seismic Line 2.	6-44
Figure 6.25	Seismic Model of "Basement" Fault with 15 Foot Offset at 800 Ft. Depth. Geologic model, top, seismic model, bottom	6-46
Figure 6.26	Seismic Model of "Basement" Fault with 30 Foot Offset at 800 Ft. Depth. Geologic model, top, seismic model, bottom	6-47
Figure 6.27	A&M Area, Hydrogeologic Cross-Section C-C' Showing Fault 2	6-49
Figure 6.28	Top panel: Portion of shear wave reflection seismic Line 2; Well MSB29A, and Horizon "H", the top of the Ellenton clay are indicated. Bottom panel: Seismic model of a normal fault with a 30 foot offset at a depth of 220 feet. Two way time offset is 45 msec. Horizontal and vertical scales are the same of both panels.	6-50
Figure 6.29	Line 1. Integrated Lithologic and Stratigraphic Interpretation from well, TDEM, and shear wave reflection seismic data. Bottom panel shows interpretation with resistivity logs.	6-54
Figure 6.30	Line 2. Integrated Lithologic and Stratigraphic Interpretation from well, TDEM, and shear wave reflection seismic data. Bottom panel shows interpretation with resistivity logs.	6-55
Figure 6.31	Line 3. Integrated Lithologic and Stratigraphic Interpretation from well and shear wave reflection seismic data. Bottom panel shows interpretation with resistivity logs.	6-56
Figure 7.1	Savannah River Site A&M Area Map	7-4
Figure 7.2	Line 1 Elevation	7-5
Figure 7.3	Line 2 Elevation	7-5
Figure 7.4	Line 1 Seismic Checkshot Results	7-10
Figure 7.5	Line 1 Fusion Estimates of Ellenton Depths and Resistivities	7-12
Figure 7.6	Comparison of Fusion and Ridge Regression Estimates from TDEM Data	7-15

Figure 7.7	Fuse3d and Ridge Regression Comparison with Simulated Flat Clay Layer	7-17
Figure 7.8	Fuse3d and Ridge Regression Comparison with Simulated Clay Discontinuity	7-18
Figure 7.9	Effect of User Selection of Spatial Correlation Coefficients	7-20
Figure 7.10	Line 1 Residual Plots Showing Effect of Channel	7-23
Figure 7.11	Time Domain Data Coverage and Data Values	7-26
Figure 7.12	Line 2 Seismic Checkshot Results	7-28
Figure 7.13	Fusion Results Seismic and TDEM Limited to Seismic Line	7-30
Figure 7.14	Fusion Results Time Domain Only	7-32
Figure 7.15	Fusion Results Seismic and TDEM Extended Beyond Seismic Line ...	7-34
Figure 7.16	Fernald: Chair Cut of Fill Showing Lithology	7-37
Figure 7.17	Fernald: Distribution of Uranium in Fill	7-38
Figure 7.18	Fernald: Lower Ground Water Elevations WRT Bottom of Clay	7-39
Figure 7.19	Bedrock Surface from Minimum Tension Gridding of 12 Well Picks ..	7-42
Figure 7.20	Detailed Bedrock Estimate From Fusion Of All Sensor Data	7-43
Figure 7.21	Detailed Bedrock Cross Section By Single Data Type and By Fusion .	7-44
Figure 7.22	Hanford 200 West Site Caliche Topology	7-46
Figure 7.23	Hanford 200 West Z1&Z2 Tile Field Plutonium Plume	7-47
Figure 7.24	Pantex Plant Well Location Map	7-49
Figure 7.25	Fused Fine Grain Zone Elevation (ft)	7-51
Figure 7.26	Head Contours (ft) And Pathlines	7-52
Figure 7.27	Hydraulic Conductivity Contours (ft/yr)	7-53
Figure 9.1	Walkaway	9-6
Figure 9.2	Seismic Raypath Geometry	9-7
Figure 9.3	Particle Motions from Rayleigh and Shear Waves	9-8
Figure 9.4	Shear Wave Propagation	9-12

LIST OF TABLES

Table 3.1	Data Fusion System (DFS) Definitions	3-22
Table 3.2	Data Fusion Definitions	3-23
Table 3.3	Data Fusion - Decision Support	3-28
Table 3.4	Data Fusion - Visualization	3-28
Table 3.5	Data Fusion - Format, Edit and Preprocessing	3-29
Table 3.6	Data Fusion - Analysis Engines	3-29
Table 5.1	Log of Field Activities	5-3
Table 5.2	Seismic Field Parameters	5-8
Table 5.3	Processing Sequence, Savannah River Site	5-9
Table 7.1	Line 1 Well Picks From SRS	78
Table 7.2	Line 1 Survey Data	78
Table 7.3	Line 2 Well Picks From SRS	7-24
Table 7.4	Line 2 Survey Data	7-24

1.0 EXECUTIVE SUMMARY

This report contains the results of a three year, three-phase project, conducted by Coleman Research Corporation (CRC) under U.S. DOE Contract No. DE-AC21-92MC29106. The long-range goal of the project has been to create a means for the more detailed and accurate definition of the near-surface (0-300') geology beneath a site that had been subjected to environmental pollution. The two major areas of research and development have been:

- Improved geophysical field data acquisition techniques;
- Analytical tools for providing the total integration (fusion) of all site data.

The long-range goal of this project has been to mathematically, integrate ("to fuse") the geophysical data that could be derived from multiple sensors with site geologic information and any other type of available site data, to provide a detailed characterization of thin clay layers and geological discontinuities at hazardous waste sites.

In Phase I (reported in August 1993, ref. DOE/MC/29106-3531; DE9400005), improved methods for processing Time Domain Electromagnetic (TDEM) data were developed, high-frequency seismic vibrators were evaluated, and Data Fusion algorithms for processing the data were created. Also in Phase I, the Savannah River Site (SRS) near Aiken, SC was designated by HQ DOE for field experimental work. A NEPA statement was filed.

In Phase II (reported in June 1994), TDEM and shear wave reflection seismic data were collected in the A&M area of the DOE Savannah River Site (SRS), to test the geophysical enhancements and the Data Fusion methodology.

The technical acceptance criteria were:

- Show advances in seismic technology to improve subsurface images;
- Identify thin clay layers and geologic discontinuities;
- Recommend locations for monitoring wells.

Data Fusion was used to process 266 central loop, vertical field out-of-loop, and horizontal field out-of-loop TDEM soundings to produce estimates of the Ellenton clay depth at about 70 meters, along a 1200 meter line (Line 1). This processing is described in Section 7.0. The Data Fusion results were consistent with the Ellenton picks at four wells along the survey line and with the geophysical interpretation presented in Section 6. The TDEM results from both Data Fusion and the geophysical interpretation showed a significant change in the estimated Ellenton resistivity along Line 1. This suggests a

possible change in the clay composition. The resistivity change occurs between existing wells and identifies locations for one or more wells to verify the surface geophysical results.

The seismic survey detected reflecting events above and in the vicinity of the Ellenton. Completion of the Line 1 seismic survey and interpretation of the seismic data was deferred to Phase 3.

Two additional sites within the SRS A&M area were selected for Phase III field work and Phase III Field Operations Plan was prepared.

In Phase III, TDEM and seismic surveys were conducted along three lines in the A&M area of SRS to expand the subsurface imaging of the general area, to subject the newly developed geophysical techniques and the Data Fusion to greater interpretive challenges, and, in the case of Line 3, to provide further information at a complex site that was under active study for potential remediation. The Data Fusion algorithms were further tuned to maximize their treatment of geophysical data and to streamline the end-to-end process of starting with mixed types of data on paper and ending with an interactive 3-D computer graphics model.

The Phase III seismic interpretation, as reported in Section 6.0, identified eleven seismic events. The events corresponding to the top and bottom of the Ellenton, at about 70 meters depth and 90 to 100 meters depth, are identified along all three survey lines. An assessment of the variability of the Ellenton composition along each survey line is also included in Section 6.0. Fusion of the Line 1 seismic and TDEM data, as described in Section 7.0, provides an excellent example of how Data Fusion can be used to combine different types of geophysical data to obtain a single, unified estimate of the Ellenton top and bottom that is consistent with all data types. The well data along Line 1 is included in this combined fusion because the interpreted seismic events are constrained to agree with the well data at each well location .

Due to electromagnetic interference and the apparent disappearance of the resistivity contrast between the Ellenton and its adjacent layers, the TDEM survey data from Lines 2 and 3 did not provide reliable information on the Ellenton depth or composition. Fusion was applied to the Line 2 TDEM and seismic data to demonstrate what happens when Data Fusion is presented with multiple, conflicting data types. The results, as presented in Section 7.0, show, as expected, that the TDEM and seismic data are not consistently interpreted. Data Fusion was presented with some conflicts among the input data types and the fusion results did not satisfy statistical criteria for satisfactory convergence. This indicates the earth model is incomplete or data are corrupted and it is already known there are data difficulties.

The SRS A&M area results presented in Section 7.0 are two dimensional characterizations along individual survey lines. The three A&M area survey lines were too far apart to support a 3-D characterization. If the survey spatial coverage was sufficient for

3-D characterization, then Data Fusion would first be used to analyze the data along each line and produce results that are consistent with those achieved with current practice geophysical methods. Next Data Fusion would be used to obtain a 3-D characterization using all of the data that is judged to be reliable. Some examples of Data Fusion 3-D site characterization on other contracts at Fernald, Hanford, Pantex, and a former Manufactured Gas Plant are summarized in Section 7.0.

2.0 INTRODUCTION

An overview of Data Fusion and geophysical methods is presented with a description of the test site characteristics.

2.1 Technical Approach For Data Fusion Methodology Development

Engineering projects involving hydrogeology are often driven by uncertainties. For activities such as environmental remediation or location of waste management facilities, cost effective solutions often depend on the confidence with which the hydrogeology is known. For example, to create a plume capture zone, answers are needed to questions about the number, depth, location and pumping schedules for purge wells. Computer simulation based on hydrogeological models provides answers. In theory, simulation can provide real-time monitoring of remediation so a plume can be "seen" as it is being cleaned up. But simulation is only as good as its geological and parametric inputs. The earth is very heterogeneous, and typical data sets are fragmented and disparate so there are substantial uncertainties.

Currently, environmental engineers do not have adequate tools to quantify uncertainty so they often rely solely on their judgement to build in sufficient safety margins. This tends to lead to overly conservative decisions that are often inordinately expensive. Data Fusion as developed in this program has value added as an engineering decision tool that quantifies uncertainty. The data fusion quantification of uncertainty is derived from the solution covariance matrix that is a by-product of the constrained, nonlinear least squares estimate of the lithological parameters. The one sigma values extracted from the covariance matrix provide good relative measures of uncertainty within the spatial domain covered by measurements.

Using Data Fusion and modeling, decision makers have a quantitative basis for action so the following benefits can be realized:

- Derives the most out of existing data sets to avoid cost of unnecessary acquisition.
- Establishes data worth in real-time, before expending funds for field data acquisition, to determine if reduction in uncertainty pays for the cost of acquisition.
- Enables remedial simulation to optimize cleanup. Remediation solutions can be exercised in the computer to match quantified safety margins before done in the earth.
- Enables real-time monitoring during remediation. Contaminant plumes can be continuously monitored while they are being cleaned up.

- Provides quantitative basis for cost reduction/avoidance.

Data Fusion and modeling have a solid foundation in the hydrogeological community. Freeze, Massman, and Smith published a framework for hydrogeological decision analysis in Freeze, et.al. (1990). They describe a pragmatic engineering approach to decision making that balances benefits, costs, and uncertainties. We have adopted their viewpoint and approach in our Data Fusion as shown in Figure 2.1. Engineers face uncertainty in parameters (such as hydraulic conductivity) and in the geometry of the problem through the geology. Data Fusion quantifies geological and parametric uncertainty. As shown in Figure 2.1, hydrogeological simulation is performed (e.g., to see plume movement) using geological and parametric inputs. Fusion propagates geological and parametric uncertainties through the simulation so the confidence in plume movement is quantified. Engineering reliability uncertainties in the engineered components of remediation also enter into decisions, but the hydrogeological uncertainties usually dominate.

Freeze, et.al. (1990) explain a limitation that there is no practical approach to incorporate indirect data by working backwards through models (inverse modeling) to quantify uncertainty. In their judgement, there is considerable data worth being lost. Data assimilation methods recently reviewed by McLaughlin (1994) provide a theoretical solution to inverse modeling. However, McLaughlin explains that data assimilation applied to hydrogeology has limitations because it is too numerically demanding. By contrast, data assimilation has become well established over several decades in numerical weather prediction and physical oceanography, but it is not ready for hydrogeology.

Data Fusion as developed in this program is a new computational technology that resolves the limitation explained in Freeze,et.al. by building on the data assimilation approach.

Figure 2.1 Data Fusion Role in Decision Analysis

2.2 Geophysical Methods

All geophysical methods react to differences in physical properties between the strata to be mapped and the surrounding soils and rocks. Three physical properties often allow the detection of clays are:

- Natural gamma ray emissivity. Clays often have a higher content of Thorium (Th) and Potassium (K) than sands and gravels. However, because of rapid attenuation, natural gamma ray emissivity measurements cannot be exploited remotely from the surface. It can be used to map clays and the boundaries between beds only in downhole geophysical logging.
- Electrical resistivity (or the inverse, conductivity). Electrical resistivity of clays is much lower than other soils and rocks and can be mapped from the surface.
- Seismic compression and shear wave velocities and impedances. This method can also be implemented from the surface.

The geophysical methods employed in this program to map clay layers were Time Domain Electromagnetics (TDEM) and Seismic Reflection (using both P- and S-waves). The reasons for selecting these methods is detailed in the Phase I report. In general terms, the main reasons are that both methods:

- Measure parameters (resistivity in TDEM, velocity in Seismic Reflection) that are sensitive to physical property changes between sand and clays;
- Have exploration depth capabilities that can meet or exceed the 300 ft. exploration depth requirement;
- Are capable of attaining the vertical and lateral resolution (mapping accuracy) necessary to distinguish the geologic layers which are significant for environmental restoration activities, and;
- Are non-intrusive.

The detail in the lateral and vertical distribution of clays required for effective design and costing of remediation is far greater than information commonly collected in conventional geotechnology. The underlying concepts of the research and development of this research project are that one geophysical method alone cannot produce a breakthrough in the detection of clay lenses, but that the use of multiple sensors, and integration (fusion) of multiple sensor data, intrusive as well as non-intrusive, can produce significant improvements. The essence of our approach, therefore, is to acquire multiple

sets of geophysical data over the same area and integrate (fuse) the results to produce a consistent image of the subsurface.

A detailed description of the fundamentals of the TDEM and Seismic Reflection is given in the Phase I Topical Report.

2.3 Technical Approach to Research

The lateral and vertical distribution of clays can be determined by intrusive (e.g. borehole) and non-intrusive (geophysical) methods. Intrusive methods have the advantage that they are direct and give unambiguous answers. Their limitations are that they are point measurements, are expensive, pose the risk of cross contaminating aquifers, and can expose personnel to hazardous materials. Surface geophysical methods have the advantage that they are low-cost, have no risk of cross contaminating aquifers and present minimum exposure to personnel. They have the disadvantage of being an indirect measurements that requires calibration and verification.

The trade off between intrusive and non-intrusive methods, therefore, consists of extensive low-cost geophysical surveys, verified and calibrated at optimum locations by intrusive tests. This approach offers continuity in subsurface information at a high level of confidence.

The technical approach used in this project to improve the detection of clay lenses proceeded along three paths:

- 1) Improving the sensitivity of Time Domain Electromagnetic (TDEM) measurements by acquisition of more parameters of the electromagnetic field than is common in present practice, and by new processing techniques.
- 2) Acquiring compressional wave and shear wave seismic reflection data by using state of the art high frequency vibrator technology over the same section as TDEM data and combining the interpretation of both data sets.
- 3) Developing sensor Data Fusion for simultaneous inversion of TDEM and seismic data sets, while incorporating a geological model of the subsurface.

2.4 Test Site Characteristics

The Savannah River Site (SRS) near Aiken, SC was selected as the primary location for testing this project's development of geophysical methods and Data Fusion. The site

lies to the East of the Piedmont where there is considerable inter-layering of fluvial/marine deposits to depths of several hundred feet before bedrock is encountered.

The A&M area of the Savannah River Site was specifically selected for the data acquisition. A detailed view of the geophysical survey lines showing nearby boreholes and landmarks is given in Figure 5.1.

The main site criteria used to select the A&M area were:

- Site has known clay layers and it has geologic (borehole) control.
- The distribution of clays is known to control ground water flow/contaminant migration.
- The site's geologic/hydrologic data base could be made available for correlation with the surface geophysics.

The geophysical lines were selected to avoid (as much as possible) cultural features which could affect the geophysical data. Cultural features defined in this manner would include power lines, metallic fences, buried metallic objects, etc., for TDEM surveys; and vibrating noise sources (such as drill rigs, compressors etc.) for the seismic surveys. Also, near surface soil disturbances (e.g., filled areas) were of concern since they may often cause a deterioration in the quality of the seismic data.

Two hydrogeologic cross-sections (B-B' and E-E'; Figure 6.24 and Figure 6.27 were created from borehole information in the A&M area (U.S. EPA Publication 9355.4-05A., February 1992). These sections show the concentration of the contaminant 1,1,2-trichloroethylene (TCE) in parts per billion (ppb) in ground water and the position of the Ellenton clay at a depth of about 180 to 250 feet. Clean up goals in this area are to reduce TCE levels to 5 ppb. In the A&M area, the Ellenton formation is reported to vary in thickness from 32 to 95 feet and is composed of two major clay layers separated by a poorly sorted sand. The Ellenton clay is understood to be one of the dominant barriers to downward contaminant flow in the A&M area. Section B-B' shows that the Ellenton clay is an effective barrier to TCE migration into deeper aquifers. Section E-E' shows that the TCE contamination has migrated below the Ellenton clay (between Wells MSB37 and MSB8) to the deeper aquifers. Concentration of TCE below the Ellenton clay is 1980 ppb in Well MSB37.

One scenario which can be used to explain this occurrence is that the Ellenton clay may be discontinuous between the two wells, which allows contaminants to migrate downward. Even though there are a substantial number of boreholes in the area, this scenario cannot be confirmed from the intrusive borehole testing. This hydrogeologic section clearly illustrates the role surface geophysical surveys can play in mapping clay

layers and extrapolating continuity (or discontinuity) of clays and other geologic units between boreholes.

2.5 Outline of Report

The form of this report is to progress from a discussion of methodology development to field data collection to data analysis, and finally to a determination of the improvement these efforts have provided beyond conventional practice. Conclusions and recommendations are then presented.

The research efforts presentation will be along two parallel paths (geophysical methods and Data Fusion) which merge at the point where data analysis commences.

The field geophysical methods activities have been divided between Time Domain Electromagnetic Induction measurements and the high frequency, vibrator induced, seismic measurements. Similarly the data analysis for the geophysical activities separately followed these two data gathering techniques. In each case reference control for formation identification was provided by the SRS site geologist.

After separate geophysical interpretations had been developed for these two types of measurements, a combined geophysical interpretation was made. This interpretation challenged the existing wireline borehole data where differences appeared, and to its merit prompted a detailed determination of lithological picks from the existing boreholes. In some cases, actual core samples from wells were analyzed when the data from well logs was questioned. It was determined that electrical and gamma well logs did not have the resolution to identify where clay formation tops and bottoms were along the borehole.

The Data Fusion processing began with the individual examination of TDEM and seismic data, with updates as the geophysicists regularly upgraded their interpretations based on additional and more refined well log data. Once the combined geophysical interpretation was available from the geophysicists, the Data Fusion process was applied to the total data from both sets of geophysical measurements, with well log data used for control. The geophysicists provided further boundary conditions for the fusion process by providing error bounds on data gathering, physical property ranges of site materials, and the measurement methods used in obtaining geophysical measurements.

2.6 Satisfaction of Phase 3 Objectives

The detailed Phase 3 objectives and the report sections describing how these objectives were met are as follows:

1. *Perform field testing in hazardous waste environment and use Data Fusion methodology on specific problems.* The field testing is described in Section 5 and the Data Fusion application in Section 7.

2. *Identify thin clay layers (1 to 3 feet thick) and geologic discontinuities in stratified sand and clay layers to depths of at least 300 feet.* At the selected test areas, within the Savannah River A&M area, the interpreted shear wave seismic data detected a thin clay layer (green clay) at depths of about 170 feet (50 meters) as described in Section 6. Sections 6 and 7 discuss the detection of an apparent channel in the Ellenton clay at a depth of about 230 feet (70 meters). The useful depth of shear wave seismic interpretation extended more than 350 feet as discussed in Section 6.
3. *Perform fusion technology enhancements with TDEM forward modeling for a full 3D capability of Data Fusion. Automated procedures will be developed for users.* The Data Fusion TDEM capabilities were extended to handle horizontal field out-of-loop, and vertical field out-of-loop soundings in addition to vertical field central loop soundings. Detailed results for Line 1 Data Fusion TDEM processing for these data types were reported in the Phase II final report. The Data Fusion TDEM results reported in Section 7 are based only on vertical field out-of-loop data. The automated procedures for the users were developed and used but are not described in this report.
4. *Test and evaluate geophysical technology to include modification of seismic vibrator sources to optimize shallow, high resolution reflection. Equipment shall be modified to include shear wave generation. Identify locations for testing.* Section 4 describes the geophysical technology evaluation. The equipment vendor modified the seismic vibrator source in response to CRC's suggestions. The Savannah River A&M area was selected for testing.
5. *Test and interpret geophysical data using methodologies developed to find better resolution of geology. Tests to be conducted at the same location as Phase II and at another location with hazardous waste using TDEM and seismic data. Interpretations to be compared with all available data on site to access accuracy.* Section 5 describes the test data acquisition. Sections 6 and 7 describe the data interpretation and the identification of possible geologic features. The Phase III survey data was collected at the Phase II site (Line 1) and at 2 other locations (Lines 2 and 3) within the SRS A&M area. The TDEM interpretation results are compared with well picks of the Ellenton clay depth in Sections 6 and 7.
6. *Perform software enhancements to structure software package based on results obtained from testing data to include geophysical data, fusion methodology and sensor forward models.* As noted above, the Data Fusion TDEM capabilities were extended to handle horizontal field out-of-loop, and vertical field out-of-loop soundings in addition to vertical field central loop soundings. The Data Fusion methodology was modified to efficiently handle a survey line as a special 2D case of the more general 3D capability. This

report presents in Section 7 results achieved with the enhanced software but does not document any of the software implementation details.

7. *Prepare topical and final reports as required.* This final report has been submitted to DOE.

3.0 DATA FUSION RESEARCH

The research for this project takes advantage of the results of work on other projects performed in parallel. The special focus for this project is the following:

- Incorporate data and models for advanced time-domain electromagnetic (TDEM) techniques. Including out-of-loop vertical field and horizontal field soundings.
- Incorporate data and models for advanced shallow seismic refraction and reflection methods.
- Add flexibility to incorporate the various practical physical features that are desired. For example, the ability to specify that different TDEM and seismic data sets will only have some geological horizons in common and the ability to specify discontinuities such as pinchouts.
- Optimize the fusion computational performance through improvements in the information filter processing and nonlinear iteration techniques.

In order to understand the role of fusion, the full fusion capability is presented. The next section explains the fundamental theory and notes what parts have been implemented and what parts remain to be completed. This is followed with the Data Fusion System (DFS) software architecture with the portions that remain to be completed noted. Appendix F describes Data Fusion from a user's perspective. The section provides guidance on practical aspects of how to initialize, execute, and validate the fusion results.

3.1 Data Fusion Theory

The key theoretical elements of fusion are presented and motivated by connections with the disciplines from which the fusion ideas have been drawn. Relationships with other methods from the hydrogeological and geophysical communities are described to provide an understanding of the distinguishing features of Data Fusion. A background is presented first to provide a setting for fusion. Then the mathematical problem is presented where it is seen that data assimilation and Data Fusion are different solutions to the same problem. This is followed by a description of the fusion methods.

3.1.1 Background

The historical setting for fusion is presented, beginning with the hydrogeological foundation and a geophysical perspective. Data assimilation is described as a starting point for fusion. Then the following fusion methods are described: Markov Random Field (MRF) model, Square Root Information Smoother (SRIS), and Generalized Expectation Maximization (GEM) method.

Hydrogeological Foundation

Engineering projects involving hydrogeology are often driven by geological and parametric uncertainties. A hydrogeological decision analysis methodology is presented by Freeze, et. al. that balances benefits, costs and uncertainties to optimize economic objectives (Freeze, et. al., 1990; Massman, et. al., 1991; Sperling, et. al., 1992; Freeze, et. al., 1992). The Freeze, et. al. methodology is displayed in Figure 3.1 where the role that Data Fusion plays is also displayed. The advantage of using Data Fusion is that it resolves some limitations described by Freeze, et. al. as explained later in this section.

In Figure 3.1, field data are used to update and calibrate geological and parametric models with quantified uncertainty. Then the models are used in a predictive hydrogeological simulation that provides input with quantified uncertainty to decision analysis. By quantifying uncertainty, regulatory requirements can be met within a quantified safety margin. For example, a purge well system can be designed in the computer by varying the number, depth, location, and pumping schedules for the wells to create a plume capture zone with adequate safety margin.

Freeze, et. al. explain that a rational basis for decision making is provided by optimizing an economic objective as shown in Figure 3.1. The decision is selected that maximizes the benefits less the usual costs less the cost of system failure weighted by the probability of failure. The engineering reliability of the engineered elements enters into the failure probability, but the dominant effect is usually the hydrogeological uncertainty. The role of Data Fusion is to quantify the contribution to the failure probability of the hydrogeological uncertainty.

The goal of the approach is to produce a model to be used for hydrogeological simulation as shown in Figure 3.2. For example, remedial simulation of flow and transport can be performed for alternative remedial actions. Contaminant concentrations can be predicted within uncertainty at regulatory compliance points until performance is acceptable. The inputs for the simulation are provided by fusion as estimates for quantities such as hydraulic conductivity, boundary conditions, and geology within their uncertainties.

The methodology presented by Freeze, et. al. views hydrogeology as a predictive science. Consequently, the fundamental heterogeneity of the subsurface must be accommodated. For example, it is stated by Freeze, et. al. that the hydraulic conductivity must be treated as an autocorrelated spatial stochastic process. In this manner, the variability of the conductivity is modeled, and the spatial continuity of the conductivity is represented as statistical autocorrelation.

Figure 3.1 Data Fusion Role in Decision Analysis

Figure 3.2 Remedial Decision Process Supported By Data Fusion

Freeze, et. al. favor a Bayesian rather than a classical statistical approach while acknowledging the value and importance of the classical approach. Both approaches use statistical spatial continuity models combined with measurements to estimate hydrogeological states. Where they differ is in the use of prior knowledge and in the use of residual fit errors for model tuning.

The classical approach has the following features:

- The user must wait for the first phase of measurements to know anything about the spatial continuity of measured quantities.
- Subsequent phases of measurements are statistically combined with spatial continuity to produce hydrogeological state estimates.
- The use of residual data/model mismatches to tune spatial continuity is considered a dubious procedure.

By contrast, the Bayesian approach has the following features:

- The user starts with prior knowledge, including spatial continuity, based on experience in related situations.
- Measurements are immediately statistically combined with spatial continuity to produce hydrogeological state estimates.
- Residual data/model mismatches are used to tune spatial continuity to reduce the dependence on prior knowledge.

The arguments used by Freeze, et. al. for the Bayesian viewpoint are the following:

- In a hydrogeological context, data are often sparse so classical methods suggest that not much has been learned. But much really has been learned if prior knowledge is considered.
- In an engineering context, the Bayesian viewpoint fits best with the usual engineering sequential design process where the engineer iterates between analysis and measurements.
- The practical reality is that decision making uses prior knowledge anyway, and the Bayesian viewpoint uses prior knowledge in an open and systematic way.

We also take the Bayesian viewpoint with Data Fusion, but recognize the value of the classical approach. In particular, the contributions made by Poeter, et. al. using the classical method are important (Poeter, et. al., 1994). Their approach is to combine hard data (data

with negligible uncertainty such as direct measurements) and soft data (data with non-negligible uncertainty such as geophysical surveys or expert opinion) with spatial continuity to describe the subsurface.

Geostatistical simulation is used with inverse flow modeling in a post processing mode to reduce uncertainty in predicted contaminant concentrations from flow and transport modeling. Geostatistical simulation is an important technique used in such fields as oil reservoir management to produce subsurface realizations that honor the data and spatial continuity. Inverse flow modeling works backwards from hydraulic head data through flow models to estimate hydraulic parameters. The main differences with Data Fusion are the use of classical statistics and the incorporation of inverse modeling in a post processing mode rather than in one processing step.

Freeze, et. al. point out that a limitation of the Bayesian approach is the application of inverse modeling. Some current methods do not incorporate inverse modeling at all. Some methods incorporate it in an approximate manner as Poeter, et. al. do in the above approach. Others such as data assimilation methods incorporate inverse modeling in a fundamental way but find that it is too numerically demanding for practical applications (McLaughlin, 1994). Consequently, it is not practical to combine all the important data sets to reduce and quantify the uncertainty in the subsurface. For example, Freeze, et. al. feel that considerable data worth is being lost by the inability to combine hydraulic head data with hydraulic conductivity data through inverse modeling. Data Fusion resolves this limitation by building on data assimilation to produce a full inverse modeling approach that is numerically practical.

Geophysical Perspective

Current geophysical methods do not provide the same kind of decision analysis foundation for Data Fusion that hydrogeology does. The philosophy of approach is very different in geophysics. For the most part, data are interpreted one data type at a time. Even for one data type, the data are processed one portion at a time. For example, a survey line of electromagnetic soundings will be processed one sounding at a time. A survey line of seismic data will be processed one seismic gather at a time.

When models are used, it is primarily to interpret individual portions of data. The role of the model is to help produce high quality interpreted data. For example, a current approach for processing electromagnetic data is the ridge regression method (Inman, 1975). The data are processed sounding by sounding using a layered earth model to produce interpreted data in the form of resistivity and thickness for each layer. Another current method for electromagnetic data is Occam's inversion (Constable, et. al., 1987). Occam's inversion interprets the smoothest resistivity profile versus depth that fits the data within a prespecified tolerance. By using judgement to adjust the tolerance, the user can produce resistivity profiles that match the data with an acceptable tolerance and are adequately smooth.

Hydrogeological decision analysis takes a predictive science viewpoint. Geophysical data interpretation is oriented more towards a descriptive science viewpoint. The connection is in the promise for geophysical interpretations to aid in quantifying geological uncertainty and parametric uncertainty in the boxes shown in Figure 3.1.

Some efforts along these lines have begun. Rubin, et. al. present a method to map permeability using hydrologic and seismic data (Rubin, et. al., 1992). The work of Poeter, et. al. described above incorporates geophysical interpretations as soft data in geostatistical simulation. Data assimilation methods are beginning to be established in the earth sciences. In the context of data assimilation, McLaughlin suggests that some of the old distinctions between modeling and field work may start to break down (McLaughlin, 1994). Modelers may be able to extract more from field data and field professional personnel may be able to learn more from data produced by model-based assimilation methods. Data Fusion fills the role of solving the problem that data assimilation methods are too numerically demanding.

Data Assimilation

Data assimilation methods are well established in numerical weather prediction, have moved into physical oceanography and are being established in hydrogeology (McLaughlin, 1994; Ghil, et. al., 1989; and Courtier, et. al., 1993). The methods take many forms from adjoint to variational to Kalman filtering as described in the extensive literature cited by Courtier dating back to 1955 with the first introduction of Kalman filtering in 1965.

McLaughlin points out that many of the techniques can be viewed as special cases of the Kalman filter. The Kalman filter provides Bayesian inverse modeling viewing heterogeneities as spatial stochastic processes as needed by the hydrogeological decision analysis methodology. The Kalman filter also includes a model error noise to account for approximation errors in the first principles models used for inverse modeling. Consequently, the Kalman filter produces a solution that honors the data within measurement error noise, the first principles models within model error noise, and spatial continuity within the spatial autocorrelation.

The inclusion of model noise is important to achieve predictive modeling with quantified uncertainty. Applications in a variety of communities have shown the necessity of recognizing that models have noise just like data. For example, satellite orbit determination filters required ad hoc fix up procedures before model noise was used to account for solar wind and gravity disturbances. Sensor mixing for inertial navigation uses model noise for inertial instrument drift and environmental disturbances like gravity errors and, in marine applications, ocean current disturbances.

The inverse problem can be formulated in a geostatistical or indirect iterative manner (Carrera, et. al., 1991). Once statistical correlations are established between measurements and variables to be estimated, the geostatistical approach computes the best linear estimate in one step. The indirect iterative approach iteratively minimizes a least squares penalty function that

penalizes data/model mismatches, excessive model error and, excessive variability in heterogeneous variables. Carrera, et. al. show that the geostatistical approach is mathematically equivalent to one step of the iterative approach. By iterating to convergence, the iterative approach can provide better estimates and better quantification of uncertainties because it is a nonlinear estimator not subject to the linear restriction of the geostatistical approach.

The difficulty with current data assimilation methods is that they are too numerically demanding for practical application. We have formulated the Data Fusion to be numerically practical and to retain the desirable features of the data assimilation methods. In fact, Data Fusion can be viewed as a data assimilation method that is numerically practical in today's UNIX workstations or Pentium class PCs.

Data Fusion

Our Data Fusion approach provides Bayesian inverse modeling as shown in Figure 3.3. It begins with prior knowledge about the state variables to be estimated in the form of first principles models, spatial continuity models as spatial stochastic processes, and uncertain initial conditions. Fusion performs Bayesian updates using measured data and data models as the data is acquired. Posterior state knowledge is produced in the form of state estimates with quantified uncertainties. Residual model fit errors are used as diagnostics to detect discontinuities, perform data validation, and to tune prior statistics such as spatial correlation distances and standard deviations.

Our methods are mathematically equivalent to the Kalman filter, using the Square Root Information Smoother (SRIS) to produce a numerically practical solution. Model noise is incorporated in order to provide a complete predictive modeling capability. Data Fusion is formulated as an indirect iterative method to achieve the best state estimates and quantification of uncertainty.

The key to achieving a numerically practical approach is to return to the basics of spatial stochastic processes. The Kalman filter has difficulty with spatial processing because it uses a causality property to break large processing problems down into a sequence of smaller problems. Causality means that there is a past causing a future. Causality is a powerful property for processes that evolve in time where there is a past and a future. But causality does not work in space.

The Kalman filter was formulated as a generalization of the Wiener filter to incorporate physical models and to be in a form more suitable for computer implementation. But the Wiener filter does not require causality so causality is not an inherent restriction. Wiener's original work and the field of statistical physics are closely tied together. It is through statistical physics that we find the replacement for causality in the concept of a Markov Random Field (MRF) to make computations practical.

Figure 3.3 Data Fusion Uses Bayesian Statistics

Markov Random Field (MRF)

The connection between Bayesian estimation and statistical physics MRF ideas was made in the computer vision community (Geman and Geman, 1984; Clark, et. al., 1990; and Chellappa, et. al., 1991). The MRF provides a way to model large scale statistical structure using only local computations. This means that large unmanageable spatial processing problems can be broken down into smaller local problems that are practical to compute.

MRF models are used in statistical physics for such problems as chemical annealing to determine lowest energy states. This has a direct analogy to Bayesian inverse modeling in determining the minimum value of a least squares penalty function for the indirect iterative method. In computer vision, MRF methods are used as the basis for Data Fusion solutions for stereo vision, shape determination from shading data, tracking object motion, and tomographic image reconstruction. However, computer vision uses the technique of stochastic relaxation to do the actual computing, but we use the SRIS technique.

Square Root Information Smoother (SRIS)

We start with the representation of an MRF as a spatial autoregression (Whittle, 1954). The autoregression has a local computational form that expresses the MRF as an interpolation of nearby values plus an interpolation error that is uncorrelated over space.

The autoregression puts the MRF in a form compatible with the data equation idea used by Bierman for the SRIS (Bierman, 1977). Bierman used the data equation to express prior knowledge on first principles models and statistical correlations as pseudo-data as if it were just more data for doing Bayesian estimation. Since the pseudo-data and data models are all local in space, the processing can be broken down sequentially in space so it becomes practical.

The SRIS has many possible forms depending on the specific details of how data and pseudo-data are represented. In fact, the SRIS is actually a family of algorithms that can be designed according to a set of information principles (Porter, 1991). The nonlinear iteration to produce the indirect iterative solution is provided by the Trust Region method for numerical optimization (Vandergraft, 1985) with the option of a Gauss-Newton step halving method.

Generalized Expectation Maximization (GEM) Method

The GEM method provides a practical means to tune statistical parameters in order to adjust prior knowledge based on the data themselves as shown in Figure 3.3. GEM is mathematically equivalent to the likelihood approach for estimating statistical parameters described by Carrera (Carrera, et. al., 1991). The approach used to calibrate system noise for groundwater simulations by Van Geer, et. al. (Van Geer, et. al., 1991) is similar in concept to GEM. But GEM is much more flexible and provides statistically optimal maximum likelihood estimates for statistical parameters such as spatial correlation distance and standard deviation.

GEM views the measured data as incomplete data for the purpose of estimating the statistical parameters (Dempster, et. al., 1977). A complete data set is specified from which the parameters could easily have been estimated if the complete data had been measured. Then a sufficient statistic of the complete data is estimated in an expectation step based on starting values for the parameters being tuned. The parameters are updated based on the sufficient statistic in a likelihood maximization step and the process is iterated. GEM has desirable statistical convergence properties, has been used in a host of agricultural, economic and scientific applications (Dempster, et. al., 1977), and used for military, scientific and image processing applications (Levy, et. al., 1992).

3.1.2 Mathematical Problem

The Data Fusion problem is presented in a form similar to that presented for data assimilation by McLaughlin (McLaughlin, 1994) in order to see how Data Fusion builds on data assimilation. To keep the ideas specific, a flow and transport example is used. We point out how this readily extends to the full range of subsurface problems.

Equations for steady state ground water flow with adjective dispersive transport in finite difference form linearized around the current solution and accompanied by measured head data and contaminant data take the form of:

Equation (3.1)

$$\begin{array}{rcl}
 A_1 h + A_2 f = q_1 - v_1 & - & \text{Flow Model} \\
 M_1 h = z_1 - v_2 & - & \text{Head Data} \\
 A_{3,t} c_{t+1} + A_{4,t} c_t + A_{5,t} h + A_{6,t} f = q_{2,t} - v_{3,t} & - & \text{Transport Model} \\
 M_{2,t} c_t = z_{2,t} - v_{4,t} & - & \text{Contaminant Data}
 \end{array}$$

- A, M - Banded matrices for physical models and data models.
- f - log hydraulic conductivity modeled as spatial stochastic process with mean f and covariance P_f
- h - Hydraulic head
- q - Sources and sinks
- z - Measured data
- c_t - Contaminant concentrations at time t .
- v - Random measurement noise and model errors normalized to zero mean and identity covariance.

Linearization around the current solution is done to prepare for the next iteration of the indirect iterated approach. The equations are written in a similar format whether they are for measured data or first principles models in order to be compatible with later representations. Notice that the log hydraulic conductivity is a spatial stochastic process. All

the terms on the far right are either measurement noise or model noise and are uncorrelated with each other. The equations have all been divided by the appropriate scaling so the noises are normalized to a standard deviation of one.

Bayesian inverse modeling is accomplished by finding estimates for h , f , and c that minimize the least squares penalty function J

Equation (3.2)

$$J = \|q_1 - A_1 h - A_2 f\|^2 + \|z_1 - M_1 h\|^2 + \sum_t (\|q_{2,t} - A_{3,t+1} c_{i+1} - A_{4,t} c_t - A_{5,t} h - A_{6,t} f\|^2 + \|z_{2,t} - M_{2,t} c_t\|^2)$$

+

$$(f - \bar{f})^T P_f^{-1} (f - \bar{f})$$

Penalizes excessive variability in spatial stochastic process

The terms in the penalty function penalize data/model mismatches, model errors, and excessive variability in log hydraulic conductivity.

The above equations can accommodate any data that can be expressed as a nonlinear function of the states; any model in finite difference, finite element, or any gridded nonlinear form; and any spatial stochastic process with a specified covariance matrix. This includes the

standard models used for geophysical data interpretation. Furthermore, it includes advanced models such as those for electromagnetic transient methods (Hoekstra, et. al., 1992) and for earth penetrating radar (Steinway, 1993).

Discontinuities can be accommodated as hypotheses subject to statistical hypothesis testing. Buried objects could be modeled as primitive building blocks with centroids and spatial dimensions to be estimated. Complex objects could be engaged as combination of simpler building blocks.

A computational problem is posed by the stochastic process term in Equation (3.2). For flexibility in modeling large scale spatial structure, it must be possible for the covariance to extend over a number of grid cells. But this means that the stochastic process term in Equation (3.2) gets out of hand. This is where Data Fusion is needed.

3.1.3 Data Fusion Approach

Data Fusion uses the MRF to solve the computational problem with data assimilation. First the MRF model is described. Then the representation of data and pseudo-data in an information array is described preparatory to presentation of the SRIS. We finish with a description of GEM.

3.1.3.1 Markov Random Field (MRF)

The spatial autoregressive representation of an MRF for the spatial stochastic process V is used

Equation (3.3)

$$V_{i,j} = \sum_{k, l \in S} a_{k,l} V_{i+k, j+l} + w_{i,j}$$

- $w_{i,j}$ - Spatially white (uncorrelated) process noise with variance q
- S - Finite spatial region of support

where the a 's are interpolation coefficients over neighboring cells in the region S . Equation (3.3) is for a 2-D MRF defined on the regular grid shown in Figure 3.4. For a stationary random field with sufficiently high order interpolation, the interpolation error w 's are uncorrelated over space. Equation (3.3) involves only local computations, but models large scale spatial correlation structure. The MRF is readily extended to three dimensions. Furthermore, the correlation structure can be anisotropic with different correlation distances in different directions. Multiple MRF's can be used to model heterogeneities in multiple layers.

		$V_{i+1,j}$		
	$V_{i,j-1}$	$V_{i,j}$	$V_{i,j+1}$	
		$V_{i-1,j}$		

Figure 3.4 Spatial Stochastic Process V on Regular Grid

We have extended the MRF model to a nonuniform slightly deformed grid to be compatible with standard finite difference and finite element grids. Spatially varying coefficients are computed so the underlying correlation structure stays fixed as the grid varies. This is accomplished by using a stochastic partial differential equation representation for the MRF and computing the coefficients so the underlying structure stays fixed.

The MRF can be further extended to explicitly estimate surfaces of spatial discontinuity (Geman and Geman, 1984). Using a binary line Markov process, an additional term is included in the penalty function that allows continuity to be disrupted if a better match to the data can be achieved. This could be useful for such applications as imaging buried objects or for seismic tomography.

3.1.3.2 Information Array

Using a general representation of Equation (3.1), first principles and data models b for region m in 3-D space linearized about current estimates at iteration i for stochastic variables V and remaining variables R with data and model input z and measurement and model noise v are given by:

Equation (3.4)

$$\frac{\phi h_m}{\phi R, V} \begin{bmatrix} R_{m-b} \\ V_{m-b} \\ \vdots \\ R_m \\ V_m \\ \vdots \\ R_{m+b} \\ V_{m+b} \end{bmatrix} = z_m - h_m(\hat{R}_i, \hat{V}_i) - v_m$$

$$L_m \begin{bmatrix} V_{m-b} \\ \vdots \\ V_m \\ \vdots \\ V_{m+b} \end{bmatrix} = 0 - L_m \hat{V}_{m,i} - w_m$$

where the MRF model has interpolation coefficients collected in L and interpolation error w . Since the data models, first principles models, and MRF models are all local in space, they are arranged so they only involve elements of R and V in region m with bandwidth b .

Equation (3.4) is in the data equation form used by Bierman to formulate the SRIS. Prior knowledge is treated as pseudo-data as if it were just additional data. Quantities in Equation (3.4) are assembled in the following information array in the computer for subsequent processing

[illegible]

Equation (3.6)

$$\begin{bmatrix} H_{11} & H_{12} & H_{13} \\ & H_{22} & H_{23} \\ 0 & & H_{33} \\ \hline & & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} z'_1 \\ z'_2 \\ z'_3 \\ \hline e \end{bmatrix} - \begin{bmatrix} v'_1 \\ v'_2 \\ v'_3 \\ \hline r \end{bmatrix} ; \quad E v' v'^T = I \quad \begin{aligned} J &= \|z - H_y\|^2 \\ J &= \|z' - H'_y\|^2 + \|c\|^2 \\ J &= \|e\|^2 \end{aligned}$$

$$\hat{y}_3 = \frac{z'_3}{H_{33}}, \quad \hat{y}_2 = \frac{z'_2 - H'_{23}\hat{y}_3}{H_{22}}, \quad \hat{y}_1 = \frac{z'_1 - H'_{12}\hat{y}_2 - H'_{13}\hat{y}_3}{H'_{11}}$$

The problem is rotated and reflected until the solution can be written by inspection. TR's are applied to make the matrix on the left upper triangular. The corresponding penalty function is shown on the right. Once the matrix is triangular, simple algebra reveals the solution as backward substitution up the diagonal of the matrix. TR can be implemented using the numerically stable work horses of linear algebra, Householder and Givens transformations.

The uncertainty in the estimate error is quantified by:

Equation (3.7)

$$\begin{aligned} H'y &= z' - v' \\ H'\hat{y} &= z' \end{aligned} \Rightarrow H'(y - \hat{y}) = -v' \Rightarrow E(y - \hat{y})(y - \hat{y})^T = H^{-1}H^{-T}$$

$$\text{Information} = (\text{variance of error})^{-1} = H'^T H'$$

$$H' = \text{Square Root Information}$$

The information is defined as the inverse of the variance of the estimate error. Consequently, by Equation (3.7), the matrix square root of the information is the upper triangular matrix in Equation (3.6). This is where the Square Root Information Smoother gets its name.

The SRIS is actually a whole family of algorithms depending on the various details of how the equations are arranged in a particular application (Porter, 1991). Also, there are different ways to manipulate the information array such as the approaches motivated by the conjugate gradient method presented by Golub and Van Loan (Golub and Van Loan, 1989).

3.1.3.3 Square Root Information Smoother (SRIS)

A SRIS processing sequence using TR's is:

Equation (3.8)

**Triangularized information
array from recursion $m - 1$**

**Data and models for
region m including
 R_{m-b} , V_{m-b} for first time**

Started array left behind since R_{m-b}, V_{m-b} never seen again

Triangularized information array to be carried forward from recursion m

Compressed Data

Because the information array is banded, processing proceeds sequentially as a recursion through space down the band of the array to triangularize the matrix. New variables are included only as they are encountered moving down the band. Old variables that are not going to be seen any more moving down the band are left behind. Back substitution can be performed as the last step moving back up the diagonal of the triangularized matrix to form the state estimates

The indirect iterative method can be implemented by iterating the SRIS using the Trust Region numerical optimization method (Vandergraft, 1985). The Trust Region method reduces the penalty function at each iteration as a quadratic minimization within a region of the current solution where the penalty function can be trusted to be approximately quadratic. The Trust Region method also imposes physical reasonableness constraints such as layer widths being greater than zero.

3.1.3.4 Generalized Expectation Maximization (GEM) Method

GEM provides the means to adjust prior knowledge by tuning statistical parameters as shown in Figure 3.3. What is desired is to estimate statistical parameters for effects that enter the measured data directly and indirectly. For example, it may be desired to estimate the correlation distance and standard deviation parameters for the log hydraulic conductivity. Measurements might include non-uniformly sampled hydraulic conductivity and hydraulic head.

GEM views the actual measurements as incomplete data for the purpose of providing a simple estimate of the statistical parameters. A complete data set is defined that would have provided a simple estimate if it had been measured. If the incomplete data is a deterministic function of the complete data independent of the parameters being estimated, then GEM provides a simple iterative estimation procedure.

The GEM answers are usually intuitive and simple. For example, it may be desired to tune the process noise variance q for one of the MRF's

Equation (3.9)

$$V_{ij} = \sum_{k,l} a_{kl} V_{i+k,j+l} + w_{i,j}; \quad Ew^2_{i,j} = q$$

If V were directly measured at all grid locations, then Equation (3.9) could be solved for the interpolation error w at all locations and q would be estimated as the sample variance of the interpolation errors

Equation (3.10)

$$\hat{q} = \frac{1}{N} \sum_{i,j} (V_{ij} - \sum_{k,l} a_{kl} V_{i+k,j+l})^2$$

The complete data is defined as all values of V augmented by the actual measured data. The incomplete measured data is a deterministic function of the complete data independent of the parameter being estimated because the incomplete data is actually part of the complete data. Therefore, a GEM iteration can be formulated.

It might seem reasonable to use the current value for q to estimate V , use the estimate for V in Equation (3.9) to estimate w

Equation (3.11)

$$\hat{w}_{ij} = \hat{v}_{ij} - \sum_{k,l} a_{k,l} \hat{v}_{i+k,j+l}$$

and then estimate q as the sample variance of the estimate of w . This is almost the answer that GEM provides, but it would be biased downward due to SRIS estimate error in w . GEM removes the bias by adding in the estimate error variance in w computed by the SRIS to give the estimate for q

Equation (3.12)

$$\hat{q} = \frac{1}{N} \sum_{ij} (\hat{w}_{ij}^2 + P_{w_{ij}})$$

The above procedure is iterated to convergence.

GEM converges to the statistically optimal maximum likelihood estimates for the parameters. This is mathematically equivalent to the parameter estimation described by Carrera (Carrera, 1991) but much simpler to implement. The actual value for the likelihood function can be computed from SRIS quantities using relationships from Bierman, et. al. (Bierman, et. al., 1990)

Equation (3.13)

$$Q(O, \hat{O}) = E(\ln p(V, R, Z/O) / Z, \hat{O})$$

3.2 Data Fusion System (DFS) Architecture

The Data Fusion System (DFS) supports a wide range of applications by providing a generic and user specific site description. Since no two sites are the same, the DFS must be able to achieve a level of abstraction that supplies the necessary computational functionality and robustness but allows the user to specify the grid, targets under investigation (stratigraphy, plumes, and buried waste), and invasive, and non-invasive data which will be used.

The DFS has an architecture which can be modified to fuse any sensor at any site which has defined models relating measurements to properties of interest. The sensors currently in use at the sites and under development within the environmental community are numerous. Furthermore, the same sensor used at a different site may require different models due to differing geology and hydrogeology. Consequently, a selected list of sensors used at

selected types of sites is implemented and an open architecture is provided which allows additional sensors at additional sites to be added in the future.

The DFS architectural design was created with three goals in mind:

- Maximize the ease of accepting new sensor data, new site information and new model definitions.
- Must be a useable tool accessible to a variety of users: geologists, hydrogeologists, chemists, geophysicists, hydrologists, biologists, computer scientists, managers and technical personnel.
- Must be able to support a single UNIX workstation or a network of workstations.

With the advances in PC hardware and software, a goal for the future is to implement a PC version of Data Fusion. Pentium class machines are more than adequate as a fusion platform.

The following paragraphs describe the DFS architecture in 3 levels. Level 0, the highest level, depicts the relationship between the user, the site and the DFS. Level 1 depicts the relationships within the DFS between the inputs, the most basic processes, and the output. Level 2 breaks down the analytical process into components or subprocesses.

The three levels of the DFS architecture are presented along with working definitions in Tables 3.1 and 3.2 of special terminology. The purpose of the definitions is to provide a common language for the DFS and to provide a fusion math to explain what fusion is doing. Each definition includes a symbolic representation where appropriate (in parentheses), an explanation, and a practical example.

Table 3.1 Data Fusion System (DFS) Definitions

Data Stores (boxes)	A data store is typically represented with a box and contains information in the form of text (including numbers) and graphics
Data Flows (lines w/ arrows)	Shows the movement of data into and out of stores and processes.
Data Processes (circles)	Depict actions being taken on data flows

SITE	All those things that typically constitute the actual physical site and records about the site
USER	All those people who will be interacting with the fusion system. Examples: Scientists, Engineers, Technicians, Managers and Regulators.
Data Fusion System (DFS)	The methodology and software to support DOE sites.
MMI	The Man Machine Interface: all of the processes developed to assist the user's in doing their jobs efficiently and accurately.

Properties (x)	Variables of interest to users to be determined from measurements. <u>Examples:</u> Depth, resistivity, seismic velocity, hydraulic conductivity, and contaminant concentration.
Measurements (z)	Variables acquired in field that have information about properties. <u>Examples:</u> Seismic first-time-of-arrivals, vertical and horizontal loop voltages, lithological picks, hydraulic heads.
Statistical Parameters (p)	Noise and spatial continuity parameters. <u>Examples:</u> Measurement noise standard deviation, and spatial correlation distance.
Model (H)	A mathematical expression that describes the relationship between measurements (z) and the properties of interest (x). <u>Example:</u> $z = H(p) \cdot x$ where the model H depends on parameters p.
Continuity Models (H)	Parameter based models that allow users to describe the variability in properties over space and time. <u>Example:</u> Spatial auto-correlations, Markov Random Field (MRF).
Dynamic Models (H)	Parameter based models that allow users to describe the variation of properties over time and space according to first principles relationships. <u>Example:</u> Groundwater flow partial differential equations.
Forward Models (H)	Parameter based data models that propagate properties forward to measurements. <u>Example:</u> Surface loop voltage measurements related to subsurface geological resistivity.

Table 3.2 Data Fusion Definitions

Sensitivities	Quantities that describe how measurements change when properties are varied. <u>Example:</u> Partials that show how loop voltage changes with resistivity variations.
Interval Halving and Trust Region	Processes that update property (x) estimates to get final property estimates.
Residual Analysis	Shows the user if the measurements (z) are matching what the model H is predicting. <u>Examples:</u> Measurement outlier detection and model diagnostics.

Inversion	The process by which a property (x) is determined from measurements (z) by working backwards through a forward model relating x to z. <u>Example:</u> The estimation of layered geology from single Time Domain EM sounding using ridge regression.
Geostatistics	The process of estimating properties (x) with quantified uncertainty from measurements (z) based on statistical correlations. <u>Example:</u> Interpolation of lithology picks or contaminant concentrations using kriging.

Data Fusion	The process by which prior knowledge of properties (x) is updated using measurements (z) from diverse data types to produce property estimates with quantified uncertainty. Prior knowledge takes the form of first principles and spatial continuity models. <u>Examples:</u> Simultaneous processing of EM and seismic measurements to determine geologic properties. Estimation of hydrological properties from hydraulic head and hydraulic conductivity data.
Hypothesis Testing and Tuning	Ways for users to determine parameters (p) and model structure (H(p)) from actual measurements. <u>Example:</u> Generalized Expectation Maximization (GEM).
Data Worth Simulation	A process used to quantify the improvement in property (x) estimates gained from additional measurements before they are acquired. <u>Example:</u> Prediction of the reduction in the geological property standard deviations from further geophysical data acquisition.
Remediation Alternative Simulation	A process used to evaluate remediation alternatives. <u>Example:</u> Determination of contaminant concentration levels $\pm$ one standard deviation at a compliance surface for a given remediation scenario.

3.3 Level 0 Architecture

The DFS architecture consists of processes and data stores that interact with two fundamental entities: the SITE and the USER, as depicted in Figure 3.5. From the SITE, field measurements, remediation process measurements, and site conditions that enter data stores. These data stores are processed under the USER's control through the MMI to provide increased site understanding through fusion. This increased site understanding is represented through text and graphics. Two of the primary outputs of Data Fusion are Visualization and Decision Support. Visualization gives the user the ability to see in one consistent picture, what is going on, and how to interact with it. Decision Support provides information needed to make decisions about how to perform fusion and how to accomplish the following:

1. *Expedited Site Characterization (ESC):* Geology/hydrogeology and data worth assessment;
2. *Wide Area Characterization:* Remote sensor surveillance and data registration;
3. *In-Situ Remediation:* Predict plumes and zones of remediation during remediation;
4. *Remedial Construction:* 'Look-ahead' for buried waste during construction.

Figure 3.5 Context Diagram (Level 0)

The capability to support decisions is most fully developed for site characterization. Future efforts could expand capabilities to include more support for remediation and wide area characterization. Part of the future efforts would be to provide more software tools such as the following:

- HyperText: A capability allowing access to all information in the current on-line help system, user guides, and training material augmented with both system and functional overviews and detailed explanations.
- On-line Tutorial: A capability directed at the novice user providing demonstrating capabilities of Data Fusion.
- Object Oriented MMI: A capability directed at the novice user that will augment the current tool box MMI providing an information flow architecture.
- Analysis: Capabilities supporting all users in the interpretation of results or input of required input parameters.

3.4 Level 1 Architecture

The DFS architecture can be broken down into four data stores and four processes as shown in Figure 3.6. Non-standardized raw measurements and site information are formatted, edited, and preprocessed to provide sets of well defined data stores for site measurements, time-varying site conditions (e.g., source terms and remedial action), and site property and parameter initial conditions. Various analysis engines are used to refine and update the parameters and properties that are used to describe the site. The MMI allows the USER to visualize and interact with site measurements, site conditions, parameters, and properties, as well as the raw measurements and site information. The MMI provides control of analysis engines covering the following six categories: Inversion, Geostatistics, Fusion, Hypothesis Testing and Tuning (automated version yet to be implemented), Data Worth Simulation, and Remediation Alternative Simulation (yet to be implemented). The processes in Figure 3.6 are summarized in Tables 3.3 through 3.6.

Figure 3.6: Level 1 Dataflow Diagram

Table 3.3 Data Fusion - Decision Support

Inputs	Processing	Outputs
<ul style="list-style-type: none"> • Site <ul style="list-style-type: none"> - Properties and Parameters • Types of Site measurements • User Situation • User Inputs • External Databases • Visualization Data Outputs 	<p>Process According to</p> <ul style="list-style-type: none"> - Situation - What are available measurements - What the user wants to do: <ul style="list-style-type: none"> • Fill in a Parameter • Decide what to do next • Give user help to understand situation 	<ul style="list-style-type: none"> • Updated Site Parameters • Text to describe situation • List of options that user can choose

Table 3.4 Data Fusion - Visualization

Inputs	Processing	Outputs
<ul style="list-style-type: none"> • Data Sets(s) from Site Simulated measurements • Site Properties and parameters • Visualization Method (2-D, 3-D, ...) • Visualization Options and Control 	<ul style="list-style-type: none"> • Quick View: Data displayed quickly, operator can interact • Properties processed into 3-D objects 3-D objects can be viewed • Property and parameter initialization • Measurement editing 	<ul style="list-style-type: none"> • Pictures and data points picked for user guidance • Pictures for interpretation • Properties and Parameters • Measurements sets

Table 3.5 Data Fusion - Format, Edit and Preprocessing

Inputs	Processing	Outputs
<ul style="list-style-type: none"> • Various types of raw measurements • Constant and time varying site conditions • Site parameters 	<ul style="list-style-type: none"> • Smart Data Formatting <ul style="list-style-type: none"> - See the whole file at once: - Define fields by highlight and pull down lists of existing data types - Define records and sub-records through highlight • Data Compression <ul style="list-style-type: none"> - Subsample, average, and combine data to reduce the size but not the information • Data Editing <ul style="list-style-type: none"> - Removing outliers based on statistical analysis 	<ul style="list-style-type: none"> • Site measurements normalized, formatted, compressed, edited • Site properties and parameters initial estimates • Site constant and time-varying conditions standardized

Table 3.6 Data Fusion - Analysis Engines

Inputs	Processing	Outputs
<ul style="list-style-type: none"> • Site Conditions <ul style="list-style-type: none"> - time varying - constant • Properties initial estimates • Parameters initial values • Measurements <ul style="list-style-type: none"> - formatted - normalized 	<ul style="list-style-type: none"> • Measurement prediction <ul style="list-style-type: none"> - Interval halving and trust region • Inversion • Residual analysis • Hypothesis testing and tuning • Pathways to receptors 	<ul style="list-style-type: none"> • Property <ul style="list-style-type: none"> - estimate - confidence - exposure level - predictions • Residual <ul style="list-style-type: none"> - outputs - Model structure - outlier detection • Parameters estimates

3.5 Level 2 Architecture - Analysis Engines

The DFS is designed so the processes within the Analysis Engines stay fixed across the range of DFS applications except for the model process. The DFS is a model-based approach that requires models of appropriate fidelity to function. As new data types are identified that contain needed information about properties for different applications, models must either exist or be developed. However, the DFS provides a generic model structure that will facilitate incorporation of new models. The Analysis Engines contains 6 processes as depicted in Figure 3.7. Each of those subprocesses are described below.

Figure 3.7 Analysis Engines: Structure Chart

In Figure 3.8 (Inversion), time-varying site conditions, current parameters, and current properties are used as input to forward models. These models produce sensitivities and predicted measurements that are compared with formatted actual measurements to update property estimates. Residual analysis provides a measure of how well the updated properties fit the measurements.

The updating is accomplished iteratively by first computing quantities needed for a linear inversion based on sensitivities at the current property estimates. Then the interval halving/trust region process computes a change to the property estimate that reduces a nonlinear least squares cost function. The iteration is constrained to produce properties that are physically possible.

In Figure 3.9 (Geostatistics), parameters are fed to forward and continuity models to produce sensitivities and predict measurements, including continuity pseudo-measurements. The predicted measurements are compared with formatted actual measurements of properties and used with sensitivities to compute final property estimates. Residual analysis provides a measure of how well the final properties fit the measurements.

Figure 3.8 Analysis Engine: Inversion

Figure 3.9 Analysis Engine: Geo-Statistics

The forward models are very simple for Geostatistics since properties must be directly measured. The authors have chosen the implementation for Geostatistics that is in the form of a linear inversion for compatibility with the other analysis engines. Consequently, the estimation is a one step process rather than being iterative.

In Figure 3.10 (Fusion), time-varying site conditions, current parameters, and current properties are used as input to continuity, dynamic, and forward models. These models produce sensitivities and predicted measurements and pseudo-measurements of different types that are compared with formatted actual measurements to update property estimates. The updating is accomplished iteratively using the Square Root Information Smoother (SRIS) and the interval halving/trust region process.

The updating is a Bayesian inverse modeling procedure where the properties are constrained to physically possible values. The SRIS provides the confidence in property estimates as a $\pm$ standard deviation. Note that core fusion is defined by the dashed box in the Figure 3.10 and is used as a subprocess in Figures 3.11 and 3.12. Residual analysis provides a measure of how well the final properties fit the measurements.

In Figure 3.11 (Hypothesis Testing and Tuning), time-varying site conditions, current parameters, and model structure are used in core fusion with formatted actual measurements of different types to determine properties and associated confidences. Final properties and confidence are used in conjunction with residual analysis to tune the final parameters and determine the best structure for the model.

In Figure 3.12 (Data Worth Simulation), simulated time-varying site conditions, parameters, and properties are used as input to continuity, dynamic, and forward models to produce simulated measurements. All of the above quantities are used by core fusion to predict the reduction in property uncertainty that would be obtained if different types of measurements were used in different combinations.

If data has already been fused at the site, then it is unnecessary to simulate measurements. One iteration of fusion is made starting from the fused estimate and telling fusion that it has the new data even though it does not. Fusion will compute the estimate uncertainty as if it actually had the data.

Figure 3.10 Analysis Engine: Fusion

Figure 3.11 Analysis Engine: Hypothesis Testing and Tuning

Figure 3.12 Analysis Engine: Data Worth Simulation

In Figure 3.13 (Remediation Alternative Simulation), time-varying site conditions, parameters, and properties (and their confidence) are used with continuity and dynamic models in a remedial simulation. The simulation is used to determine predicted compliance quantities based on selected remediation alternatives. Further, based on the property confidence from the fusion results, the confidence in compliance quantities is determined. Remedial simulation provides a way to optimize remediation with a quantified safety margin.

Figure 3.13 Analysis Engine: Remediation Alternative Simulation

4.0 NEAR-SURFACE GEOPHYSICAL RESEARCH AND DEVELOPMENT ACHIEVEMENTS

4.1 Introduction

The objectives of the geophysical research have been to increase the sensitivity of traditional electromagnetic and seismic exploration methods. The desire has been to provide finer resolution of site geology in the first 300-500 feet beneath a location which has been subjected to a pollutant release. Interest has been particularly acute for being able to identify clay formations lying between sand-rich unconsolidated sediments, the clay locations, their thickness, extent and if possible their quality as an aquitard. The A&M Area at DOE's Savannah River Site (SRS) near Aiken, SC was selected as the field site for these geophysical developments. The unconsolidated sediments at SRS have a thickness varying from 800 ft to over 1,200 ft. The geophysical research and development (R&D) for mapping the stratigraphy was generally focused in the upper 500 ft. Two surface geophysical methods were applied to differentiating clays and sands, i.e., measuring the geoelectric section and determining seismic reflection events.

Improvements were made in both technologies, but perhaps the greatest improvements in developing an understanding of the lithology and stratigraphy resulted from carefully integrating the information derived from both surface geophysical sensors with the geophysical and stratigraphic logs made available by Westinghouse SRS. Integration allowed interpolation of stratigraphic units, and determining changes in lithology within these units.

For geophysical R&D to be successful requires a number of factors to come together for optimum performance, such as equipment, new theoretical concepts, and computer algorithms making use of new concepts in interpretation. Particularly in the seismic reflection studies at SRS, the R&D was performed at an opportune time when significantly improved seismographs, three-component geophones, and high frequency vibrators became available from commercial manufacturers. Extensive use in field data acquisition was made of these advances in commercial instrumentation. Also, the R&D coincided with a period of intensive research by the oil and gas exploration industry in shear wave reflection and propagation and the influence of anisotropy in acoustic properties of the Earth on seismic wave propagation and reflection. The knowledge resulting from the investigations by the oil and gas industry were of significant benefit to the R&D performed under this contract.

In the following Sections, the improvements made in each component of the geophysical work conducted at SRS are detailed. In Section 4.1, the significant advances in determining the geoelectric section by time domain electromagnetic is discussed, and in Section 4.2, the advances in R&D in seismic reflection exploration are reviewed. The additional information derived by integrating the results of TDEM seismic investigations with geophysical and lithologic logs are discussed in Section 4.3.

4.2 Time Domain Electromagnetic (TDEM)

TDEM is a surface geophysical method that determines the geoelectric section (lateral and vertical resistivity) stratification of the subsurface. From the geoelectric section in turn information about the continuity, thickness, and clay content of clay-rich strata with the sand-rich strata of the unconsolidated sediments underlying SRS can be inferred. Figure 4.1 shows an electrical resistivity log of the A&M area and on this figure soil types have been assigned characteristic ranges of electrical resistivity. Particularly, the clays of the Crouch Branch Confining Unit (CBCU) are characterized by low values of electrical resistivity, and it is precisely this relationship between low electrical resistivity and clay content of soils that was the objective of measuring the geoelectric section.

There are a number of surface geophysical methodologies for determining the geoelectric section. These methods differ in the components of the electromagnetic field measured, and the manner in which electrical current flow is generated in the subsurface. They are far from equivalent in sensitivity to geologic noise, and lateral and vertical resolution. The TDEM method was selected for R&D and use at SRS, because the objectives were within the exploration depth range of available equipment, and the method has a low sensitivity to geologic noise and a better lateral and vertical resolution in determining the geoelectric section.

The most common receivers employed are multi-turn air coils and the effective areas of the receivers vary from $100m^2$ to 10^5m^2 (Hoekstra and Blohm, 1990). The receivers with large effective area are often seismic cables laid out over the ground surface, but coils with effective area of $1000m^2$ and less are wound on a rigid mandrel. With these latter coils, both vertical and horizontal components of the field can be measured.

A common system waveform of TDEM signals is shown in Figure 4.2. The current driven through the transmitter loop is a half-duty cycle waveform and a steady current is terminated by a linear ramp. The ramp time depends on peak current and the dimensions of transmitter loops, and is about 2 to 50 microseconds for exploration depths to 300 feet.

MSB26A

RESISTIVITY (ohm-m)

400
350
300
250
200
150
100
50
0.0

0.0 1500 3000

Figure 4.1 Well MSB26A Resistivity Log and Lithologic Interpretation

Figure 4.2 System Waveforms In TDEM

It has been shown (Kaufman and Keller, 1983) that the sensitivity of the electromotive forces (emf's) to the geoelectric section change with time. At time corresponding to "early stage", emf_z is proportional to ground conductivity, σ , whereas, at late stage emf_z is proportional to $\sigma^{3/2}$. It is therefore advantageous to have the range of measurements mainly in "late stage" because of the high sensitivity to the geoelectric section. The onset of "late stage" is defined by :

$$\text{Equation (4.1)} \quad \frac{\mathfrak{S}}{r} > 16$$

$$\text{where } \mathfrak{S} = \sqrt{2\pi\rho t 10^{-7}}$$

and

r = transmitter-receiver separation

ρ = resistivity

t = time of measurement

Figure 4.3 shows the relation between onset of "late stage" and transmitter loop size. The concept of "late stage" is used in subsequent Sections to discuss improvements in TDEM under this program.

Figure 4.4 shows the behavior of the electromotive force due to the horizontal magnetic field emf_z and the vertical magnetic field emf_x on a line through the center of a 100 meter by 100 meter transmitter loop for ground with a uniform resistivity of 100 ohm-m. For transmitter loops of different dimensions and for horizontally stratified ground or ground of different resistivity, the trends are similar.

Characteristics of the vertical field are:

- emf_z is a maximum in the center of the transmitter loop.
- With increasing distance from the loop, emf_z decays and eventually crosses through zero.

Characteristics of the horizontal field are:

- emf_x is zero in the center of the loop.
- With increasing distance from the loop, emf_x increases to a maximum then decreases and crosses through zero.

132150Z

4-180

Figure 4.3 Relation Between Onset of Late Stage and Electrical Resistivity and Transmitter Loop Size

Figure 4.4 Behavior Of emf_z and emf_x On A Profile Through The Center Of A Square Transmitter Loop

For the same time gates, the maximum amplitude of emf_x is less than the maximum amplitude of emf_z . At the present time the common mode of TDEM surveying consists of making measurements of the vertical field, emf_z , in the center of a non-grounded loop. There are several reasons for this preferred procedure, such as:

- The plotting point of the measurement is well defined, because the induced current is concentric about the center.
- The distance between the transmitter and receiver is minimized, helping lateral resolution.
- Ambient electrical noise is less for the vertical component, and emf_z is a maximum in the center. The signal-to-noise ratio is therefore optimal for emf_z in the center.
- Survey requirements to locate the center are minimal, because the field is relatively flat in the center.

There is, however, one important incentive for investigating measurements of emf_x , and that is its higher sensitivity to the geoelectric section over a certain time range. It has been shown (Kaufman and Keller, 1983) that in late stage, the vertical field, emf_z , is related to conductivity of the geoelectric section and time of measurement by:

$$\text{Equation (4.2)} \quad emf_z \propto K \frac{\sigma^{3/4}}{t^{2.5}}$$

and for the horizontal field, emf_x , we have:

$$\text{Equation (4.3)} \quad emf_x \propto K \frac{\sigma^4}{t^3}$$

where σ is ground conductivity and t is time after current turn-off. It was this increased sensitivity of emf_x to the geoelectric section; σ^2 versus $\sigma^{3/2}$ that was the impetus for developing horizontal field measurements for this investigation. The disadvantages of horizontal field measurements, are:

- emf_x is of lower amplitude and decays faster with time. Figure 4.5 compares the amplitudes of emf_z and emf_x as a function of time. At early time (10^{-2} milliseconds) emf_x is about a factor 3 less than emf_z , but because emf_x decays faster with time, it is an order of magnitude less than emf_z at later times. However, lower amplitudes for emf_x are only significant if noise exceeds signal.

Figure 4.5 Comparison Of Amplitude Of emf_z and emf_x As A Function Of Time

- emf_x is zero in the center of the transmitter loop and measurements must, therefore, be made outside the loops. This will require careful surveying to locate receiver positions.

Although differential processing is a highly sensitive procedure for recognizing a conductive layer, its application may be limited. The Russian geophysicists applied differential processing for mapping the contact between brine saturation and hydrocarbon saturation in producing formations on the Siberian Plateau. On the Siberian Plateau petroleum production is in resistive Paleozoic rocks, and the resistivity contrast between brine saturation and hydrocarbon saturation is two orders of magnitude. Differential processing was tested on data acquired at SRS and was not effective. The reasons for that are:

1. The stratigraphy at SRS is complex, consisting of a large number of interbedded sand and clay lenses, so that there is not one conductive layer but many occurring throughout the section.
2. The geologic noise (variation in resistivity not related to the exploration objective) is relatively high. For example, there is significant variation in resistivity of layers above the clays of the Crouch Branch Confining Unit.

Present day practices of TDEM interpretations consist of inversions by one-dimensional (1-D) algorithms. At each receiver station, a 1-D geoelectric profile is derived, and typically 2-D geoelectric cross-sections are constructed from a series of 1-D profiles. Past research (Wilt, 1991) has shown 1-D inversions to be remarkably robust even in the presence of dip in stratigraphy and 2-D discontinuities.

The geoelectric section can be derived from different measurements from the same transmitter loop, such as measurements of the time derivatives of:

- the vertical magnetic field in the center of the transmitter loop,
- the vertical magnetic field outside the transmitter loop, and
- the horizontal magnetic field outside the transmitter loop.

If 1-D inversions are employed, the results need to yield near identical geoelectric sections at least within the range of equivalence. Figure 4.6 compares geoelectric profiles derived from 1-D inversions derived from three sets of measurements. There is agreement between the three data sets providing the evidence that the inversion algorithms developed under this contract for inversions of horizontal field measurements are correct.

RESULTS OF 3 TDEM DATA SETS AT STATION 120 LINE 1, A&M AREA

GEOELECTRIC SECTION DERIVED FROM 1-D INVERSIONS

Figure 4.6 TDEM Apparent Resistivity as Time Curves Derived From Horizontal and Vertical Magnetic Fields Measurement at the Center and Outside the Same Horizontal Transmitter Loop. Line 1, A&M Area, SRS.

Because the geology of the sedimentary section underlying SRS is complex, consisting of many interbedded sand and clay lenses, TDEM is not effective in differentiating individual lenses. The TDEM data must be integrated with seismic and well log information to gain an understanding of the individual clay units. To obtain a measurement of aggregate clay distribution, cumulative conductance is expected to be a meaningful physical quantity. Cumulative conductance is defined as:

Equation (4.4)
$$S = \sum_{i=1}^n \frac{h_i}{\rho_i}$$

where:

S is conductance in Siemens,
 h_i is thickness of i^{th} strata, and
 ρ_i is the resistivity of i^{th} strata.

It was particularly conductance that proved the most useful quantity in arriving at integrated stratigraphic and lithologic interpretations.

4.3 Seismic

Seismic reflection methods image the subsurface by generating seismic waves at the surface and recording the energy after it has been reflected from geologic horizons in the subsurface. This method has been used extensively in the hydrocarbon exploration industry since the 1950's for objectives up to depths of 30,000 ft., however, its application for shallow environmental and engineering objectives has only become prevalent in the last 10 years. In general the same techniques that have been used in exploration for deeper objectives have been scaled down for use in shallow exploration, but there are also requirements which are unique to shallow reflections which requires modification from standard practices. One of the unique requirements of high resolution reflection surveys is the need for non-explosive high frequency sources to improve the resolution of the method.

Therefore, our seismic R&D efforts conducted under this program focused on two areas:

- The use of high frequency vibratory sources and;
- Acquisition and analysis of multi-component seismic reflection surveys

The following Sections will describe the requirements necessary for seismic imaging of thin clay layers and the efforts made under this project to meet those requirements.

4.3.1 Resolution of Thin Layers

The main focus of the project was the detection of thin (less than 3 feet thick) clay layers. The rule of thumb used in seismic exploration is that the layer of interest must be at least as thick as one-quarter of the seismic wavelength. The mathematical expression for wavelength is:

$$\text{Equation (4.5)} \quad \lambda = V/f$$

where:

λ = Wavelength

V = velocity

f = frequency

To detect thin layers, the wavelength of the seismic signal must be shortened as much as possible. As the velocity in the subsurface is fixed, Equation (4.5) shows that the only way to make the seismic wavelength shorter is to increase the frequency of the seismic signal. Figure 4.7 shows a simplified subsurface scenario typical of unconsolidated sediments. Shown on the left of the Figure (4.7a) is the typical subsurface velocity structure for layers 1,2,3,4. To the right of the velocity model, its corresponding reflectivity is shown. The reflection coefficient is given by:

$$\text{Equation (4.6)} \quad R = \frac{(\rho_2 V_2 - \rho_1 V_1)}{(\rho_2 V_2 + \rho_1 V_1)}$$

Where:

R_{12} = reflection coefficient

ρ_1 = density, layer 1, etc.

V_1 = seismic velocity, layer 1, etc.

or ρ_i = density i^{th} layer
 V_i = velocity i^{th} layer

THIN LAYER SUBSURFACE SEISMIC SCENARIO
WITH REFLECTIVITY AND
1. TYPICAL SEISMIC WAVELET
2. HIGH FREQUENCY SEISMIC WAVELET

SUBSURFACE SCENARIO WITH CORRESPONDING
REFLECTIVITY AND SEISMIC TRACE

Figure 4.7 Effects Of Wavelet Frequency On Seismic Resolution

For this case, we have assumed subsurface densities to be constant. At each layer interface, the change in velocity results in a spike on the reflectivity series. To the right of the reflectivity series is shown a typical seismic wavelet. When the reflectivity series is convolved with the seismic wavelet, the product is the seismic trace shown. It can be seen that the large velocity contrast between layer 1 and layer 2 produces a seismic event that is larger in amplitude than the seismic event produced by the smaller velocity contrast between layers 2 and 3. Also, a decrease in velocity (such as that between layers 3 and 4) will produce a seismic event which has the opposite polarity of that produced by a velocity increase (between layers 3 and 4).

Figure 4.7b shows the same four layer model as Figure 4.7a, except layer 3 has been made much thinner. The reflectivity for this model is shown to its right. When the typical seismic wavelet is convolved with the reflectivity series, the reflectivity coefficients from the top and bottom of layer 3 are too close to be resolved. The resultant seismic event is a complex summation of two separate, overlapping wavelets.

In order to resolve the top and bottom of layer 3, a shorter wavelength, or higher frequency event is needed. On the right side of Figure 4.7b, the reflectivity is shown convolved with a high frequency seismic wavelet. This seismic trace clearly shows separate events at the top and bottom of layer 3.

4.3.2 Vibratory Sources

Vibratory sources have been used in the oil and gas industry since the 1960's. These vibrators typically weigh thirty tons, and are capable of vibrating at frequencies of about 4 to 96 hertz. These vibrators were designed for exploration objectives thousands of feet deep, and are not very well suited for mapping targets in the upper 300 feet. A new generation of minivibrators have recently been developed for shallow exploration, weigh from 150 to 7000 pounds, and can sweep up to 1000 hz. The frequencies above 100 hz are required in order to effectively map shallow objectives.

The major advantage that vibrators offer over impact and explosive sources is that the frequency of the vibratory signal can be controlled to a great degree. The iVi Minivib used at SRS in Phase III had the capability to vibrate between 10 and 550 hz.

Traditional impact sources operate by striking a steel plate with a sledgehammer. This leaves very little room for control of the frequency of the seismic signal. Impact sources often contain a high frequency component, but generate a great amount of unusable low-frequency signal as well. With the iVi Minivib, the frequency range and energy in the different frequency bands which are input into the ground can be specified by the operator and therefore frequencies which are more likely to generate ground roll and other noisetrains not desired in reflections seismics can be eliminated or decreased in amplitude relative to those frequencies which generate reflections.

Another benefit of vibratory sources is the ability to obtain useful data in areas of high ambient noise. This is because the correlation of the input signal with the recorded data, tends to decrease the amplitude of uncorrelated noise, such as compressor and vehicle noise.

The ability of vibratory sources to control the frequencies input into the ground is shown in Figure 4.8 where the amplitude vs. frequency plots of data acquired with the iVi minivib and a hammer are compared. The vibrator spectra shows significantly more energy in the frequency range from 100 hz. and above. The benefits of this high frequency content are shown in Figure 4.9, where a reflector is clearly imaged on the shot record obtained using a vibrator and is not visible on the record obtained using a hammer at the same location.

In addition to the higher frequency obtainable from the iVi minivib over larger vibrators there is the increased portability of the units. Vibrators for hydrocarbon exploration are large and heavy, and require special trailers for over the road transport. The minivib comes on several different platforms including a standard pick-up truck, a small trailer or on a modified Ford F850 truck. All of these units are capable of regular highway driving, and this significantly reduces the cost of mobilizing the equipment to the site. Furthermore, the smaller size of the units allows access to areas where it would be difficult for the larger vibrators to go.

Under this project, vibrator sources were extensively tested and found to produce results that were better than those obtained using impact sources. These vibrators were used in compressional and shear waves modes. Furthermore processing of the acquired data was accomplished entirely on PC's, as opposed to UNIX workstations. This ability significantly reduces the cost of processing shallow reflection surveys.

4.3.3 Shear Wave Surveys

Another effort at improving the resolution of high resolution seismic surveys is the use of shear wave reflection methods. Traditional reflection surveys have used compression(P) waves. There are several reasons for this including the ease of generating of compressional waves and less expensive geophones, decreased field effort and better understanding of the propagation of compressional waves in the subsurface. However, recent efforts in the hydrocarbon exploration industry have increased the understanding of the propagation of shear waves in the presence of anisotropy (variation of physical property depending upon direction in which it is measured) and have reduced the cost of three component geophones. In addition, the iVi Minivibs can be configured for generation of horizontally polarized shear waves, which simplifies the field operations.

COMPARISON OF P-WAVE AMPLITUDE SPECTRA

Figure 4.8 Comparison Of P-Wave Amplitude Spectra

COMPARISON OF P-WAVE SOURCE TYPES

Figure 4.9 Comparison Of P-Wave Source Types

Shear waves differ from compressional waves in the particle motion relative to the direction of wave propagation. Compressional wave particle motion is always parallel to the direction of wave propagation. This results in alternating areas where matter is compressed and dilated. This is the method by which sound waves are propagated. This is illustrated in Figure 4.10. In contrast shear wave particle motion is always perpendicular to wave propagation as shown in Figure 4.11, resulting in a shear deformation of the Earth as the wave propagates through the subsurface. Thus shear waves cannot pass through liquid or gaseous media making them insensitive to water saturation.

One of the benefits of shear waves that is of interest in shallow exploration is the possibility of higher resolution because of the shear waves propagate at a speed of less than half that of compressional waves, therefore according to Equation (4.5) the minimum detectable thickness of a geologic layer would be half the size for a shear wave survey than a compressional wave survey given the same frequencies. This is illustrated in Figure 4.11, which shows the minimum detectable thickness of beds vs. frequency for typical velocities at SRS.

Under this R&D program, multi-component surveys were conducted using horizontally polarized shear vibratory sources and three component geophones. The use of multi-component phones allowed the investigation of anisotropy in the sediments underlying the SRS. Field procedures were developed to allow efficient acquisition of multi-component data for high resolution surveys. Vibratory sources were shown to be effective in the generation of high frequency shear waves, and furthermore a processing sequence was developed to produce shear sections.

4.4 Integrated Interpretation

The geoelectric section derived from the TDEM survey provides a highly simplified display of the complex lithology and stratigraphy underlying SRS. From the TDEM survey, changes in accumulative conductance along the measurement profiles are derived. These changes are indicative of changes in lithology, but it is not possible to identify with sufficient accuracy at what depth these changes in lithology occur.

The seismic section shows detail in stratigraphy of different units, and from the seismic sections structure are derived. It is not feasible, however, to derive from the seismic sections changes in lithology, or the characteristics within a given layer. The relation between seismic reflectivity and lithology depends on many factors, and soil type is but one of them.

The resistivity logs along the geophysical survey profiles and the seismic checkshot surveys performed as part of the contract field work provide the information necessary to correlate at well locations the reflectors in the seismic section to stratigraphic units. Also, the resistivity and lithologic logs of the wells allow correlation of the major clay units in the section.

Figure 4.10 Schematic Presentation Of Particle Motion Of Compressional and Shear Waves

Figure 4.11 Detectable Thickness Of A Layer (Wavelength/4)
As A Function Of Seismic Signal Frequency

Integration of the three data sets permits construction of detailed lithologic and stratigraphic sections along the survey lines. Integration results in a product with a value far greater than the sum of its parts. In the integrated products given in Section 6, the seismic sections are used to provide structure and continuity of the stratigraphic units, and the TDEM section yields interpolation of changes in lithology within the stratigraphic units. At the borehole locations the interpretation of the lithology and stratigraphy inferred from TDEM and seismic is closely correlated.

Thus, although advances were made in geophysical acquisition methods, and processing of both best quality geoelectric section and seismic sections, the largest pay-off in R&D came in careful integration of three data sets. As shall be seen in succeeding sections of this report, the complimentary R&D of mathematical Data Fusion provides yet another tool for extracting additional information from the various site data sets. It will be shown that the automation and capacity for simultaneously handling very large amounts of site data by Data Fusion also creates a whole which is considerably larger than the sum of the parts.

4.5 Summary of Advances in Geophysical Technologies

Advances in Time Domain Electromagnetic (TDEM) Acquisition and Processing

The advances in TDEM acquisition and processing made under this contract are:

- Forward model and inversion algorithms were developed, debugged, and tested for 1-D interpretation of recordings of the time derivative of horizontal field measurements.
- Results proved that the 1-D inversions derived from three different data sets, horizontal and vertical magnetic field measurements made outside transmitter loops, and center-loop measurements of the vertical magnetic field agree.
- Experimental data showed that data quality is highest for the geoelectric section at SRS for measurements of the vertical magnetic field outside the transmitter loop.

Advances in Seismic Reflection Data Acquisition and Processing

The advances made in seismic reflection data acquisition and processing are:

- Several high frequency vibrators for generating compression and shear waves were tested, and for the conditions at SRS the iVi Minivibrator with about 7000 lbs hold-down weight proved most effective.

- Shear wave reflection surveys provided the best quality seismic sections. The seismic shear wave reflection survey allowed recognition of thinner layers. The main reason for the better vertical resolution of shear waves is the lower velocity of shear waves compared to compression waves.

Advances in Integration of Different Data Sets

Major unique improvements were made in deciphering the complex lithology and stratigraphy by integrating the information contained in TDEM, seismic, and borehole geophysical logs. The information about lithology and stratigraphy inferred from TDEM, seismic section and borehole resistivity logs is far greater than the sum of the individual components.

5.0 NEAR-SURFACE GEOPHYSICAL FIELD DATA COLLECTION AND PROCESSING

5.1 General

Geophysical data were collected in the A&M area of SRS in Phases II and III of this project. Figure 5.1 shows the extent of data collection for both phases.

In Phase II, both TDEM and seismic data were collected along Line 1. TDEM data were collected along the entire length of the line (Stations 0 to 4000) and seismic data were collected from Stations 0 to 2100. The survey was laid out along an existing dirt road, and measurement locations were initially positioned by compass and chain. Subsequently, SRS personnel resurveyed the measurement locations to obtain the SRS coordinates and elevations for the measurement stations.

In Phase III, TDEM data were collected along the entire length of Line 2. Seismic data were taken on the remainder of Line 1 and across a portion of Line 2.

In Phase II, borehole conductivity and natural gamma ray logs were run in boreholes along Line 1. Seismic checkshots were run in boreholes along both lines in Phases II and III.

A daily log of field activities for Phase II and Phase III is given in Table 5-1.

Figure 5.1 Layout of Geophysical Survey Lines - A&M Area, SRS

Table 5.1 Log of Field Activities

Phase II - Data Acquisition	
Date (1994)	Line/Stations/Boreholes
March 23	Mobilize to Aiken, SC.
March 24	General Employee Training at SRS Training Facility.
March 25	Drug Screening in A.M. Testing of Seismic Vibrator in P.M.
March 26	Continue testing of vibrator, borehole gamma (γ), and conductivity (σ) logging of Well MSB21A. TDEM test data near MSB21A.
March 27	Continue seismic testing. Borehole logging (γ and σ) in Borehole MSB20A. TDEM test near MSB20A, 768 and 780.
March 28	P-wave seismic checkshots in boreholes MSB-21A, MSB-76B and MSB-74C. Borehole logging (γ and σ) in holes MSB-33A, MSB-74B and LWR-2A. TDEM tests near MSB-33A, 78D, 74B and LWR-2.
March 29	Borehole logging MSB26A (γ and σ). Start Line 1 measurements from transmitters 0, 120, 240, 360. Survey seismic line and begin layout of P-wave acquisition spread.
March 30	TDEM measurements, Line 1, transmitters 360, 480, 600, 720, 840, 960. Begin P-wave acquisition production.
March 31	TDEM measurements, Line 1, transmitters 3840, 3720, 3600, 3480, 3360. Continue P-wave acquisition production.
April 1	TDEM measurements, Line 1, transmitters 3240, 3120, 3000, 2880, 2760. Continue P-wave acquisition production.
April 2	TDEM measurements, Line 1, transmitters 2640, 2520, 2400, 2280, repeat 2760, 2880. Continue P-wave acquisition production.
April 3	TDEM measurements, Line 1, transmitters (repeats) 2880, 3000, 3120, 3240, new loops 1080, 1200. Continue P-wave acquisition production.
April 4	TDEM measurements, Line 1, transmitters 1320, 1440, 1560, 1680, 1800, 1920, 2040, 2160. End of TDEM measurements on Line 1. Continue P-wave acquisition production.
April 5	TDEM measurements, Line 2 (Site R), transmitters 60, 180, 300, 420, 540, 660, 780, 900, 1020. Finish Line 2. Continue P-wave acquisition production.
April 6	TDEM measurements near boreholes MSB-12, MSB-30. Change over vibrator mass from P-wave to S-wave mode.
April 7	Testing of shear wave vibrator and shear wave sweep design.
April 8	Begin shear wave acquisition production.
April 9	Continue shear wave acquisition production.
April 10	Continue shear wave acquisition production.
April 11	Continue shear wave acquisition production.
April 12	Continue shear wave acquisition production.
April 13	P- and S-wave checkshots in MSB-21TA, MSB-20A and MSB-76.
April 14	Demobilization from Savannah River Site.
Phase III - Data Acquisition	
October 26	Mobilization to Aiken, SC.
October 27 - November 5	TDEM, Line 2 from Station -2800 to 2200. TDEM, Line 3 from Station 100 to 1500.
October 27 - November 15	Seismic, Line 1. Seismic, Line 2. Borehole seismic checkshots; Boreholes MSB21A, 21TA, 26, 69B, 74C, 78C, 82A, 85B, 85TA.

In addition to the geophysical data collected in this project, other existing data in the A&M area (borehole logs, geologic reports, previous seismic survey data) were supplied by SRS personnel for interpretation (ref. Section 6.0).

5.2 Data Acquisition

Time Domain Electromagnetic (TDEM)

The TDEM data were collected with the Geonics EM 47 TDEM system. Manufacturers specifications for this instrument are given in Appendix A. All data were recorded in the field on a solid state data logger. The data were transferred each evening from the data logger to a PC for preliminary interpretation and data quality analysis.

Figure 5.2 shows the transmitter loop layout along Line 1 for Phase II. This figure also shows the location of nearby boreholes and landmarks. Non-grounded transmitter loops with dimensions of 120 ft. by 120 ft. were used as sources. Loops were positioned end to end for a total of 33 transmitter loops. Five measurement locations were occupied from each transmitter loop as schematically illustrated in Figure 5.3. At the center of the loop the vertical field was measured, and at the 120' and 180' offsets (from each side of the loop) both the vertical (z) and the horizontal (x) fields were measured. Thus, a total of nine separate measurements or soundings were obtained from each transmitter loop resulting in a total of 165 vertical and 132 horizontal field measurements for the entire line.

During Phase III measurements were made from 33 transmitter loops (25 along Line 2, and 8 along Line 3). A transmitter loop size of 200 ft by 200 ft was used for the measurements. From each transmitter loop, measurements were made in the center of the loop and at offsets of 200 feet on each side of the transmitter loop. Both the horizontal (x) and vertical (z) components were measured at each offset station. Only the vertical field was measured at the center-loop station. Thus, a total of 165 measurements were made from the 33 transmitter loops in Phase III.

During TDEM data acquisition in Phase II and III, cultural features (e.g., buried metal culverts, power lines, fences, roads, etc.) were encountered which cause deterioration in data quality. Along Line 1 (Phase II) buried culverts at Stations 1740 and 3240 caused distortion of nearby measurements which were omitted from interpretations. Also, towards the Northeast end of the line (Stations 3480 to 3840), data quality becomes progressively poorer due to noise/coupling effects from the power line and fence near this end of the line. Along Line 2, a buried power line and nearby weather antenna caused distortion of the data between Stations -100 and -900.

Figure 5.2 TDEM Transmitter Loop Layout, Line 1

SCHEMATIC ILLUSTRATION OF TDEM MEASUREMENTS MADE FROM EACH TRANSMITTER LOOP.

Figure 5.3 Schematic Illustration of TDEM Measurements Made From Each Transmitter Loop

Borehole Electromagnetic and Natural Gamma Measurements

In the vicinity of Line 1, boreholes MSB21A, MSB20A, MSB76C, MSB78D and MSB26A were logged with the Geonics EM 39 Conductivity and Mt. Sopris Natural Gamma tools. Manufacturers specifications are given in Appendices B and C. The location of these boreholes with reference to Line 1 are shown on Figure 5.1. The purpose of these measurements was to use the geophysical borehole responses (particularly conductivity) to calibrate the surface TDEM results.

Seismic

The seismic survey was carried out along three lines as shown in Figure 5.1. The Eastern half of Line 1 was a continuation of the survey carried out as part of the contract Phase II. For interpretation, these two records were spliced together. On Lines 1 and 2, the seismic lines were coincident with all or part of the TDEM survey lines; Line 3 has no coincident TDEM survey.

The source used was an iVi Minivibrator; manufacturers specifications are given in Appendix D. This source applies a vibratory signal to a rigid steel plate which was pressed into the ground. The vibrator is capable of generating swept frequencies ranging from 10 Hz to 550 Hz. The frequencies used for this survey were 20 to 120 Hz with a 5db/octave boost. The baseplate can be oriented so that the signal is shaken either along the line of the survey (In-line) or at 90 degrees to the line of the survey (X-line). The source locations (shot points) were located at 5 foot intervals in order to resolve reflectors as close to the surface of the earth as possible.

The signals were recorded using single OYO 3-component geophones with spacing of 5 feet. Only the two horizontal components were recorded as P wave testing showed that no energy was received from below the water table at a depth of about 120 feet. The recording parameters are shown in Table 5.2. The center frequency of the signal was 60 Hz with a bandwidth of 21 octaves. The data were recorded by an OYO DAS-1 seismic system. Manufacturers specifications are given in Appendix E.

Table 5.2 Seismic Field Parameters

Recorded By: DRP
Source: iVi Minivib
Source Interval: 5 feet
*Spread: Off-end, 245.0 - 10.0 - 0.0
Group Interval: 5 feet
Sweep: Non-linear, 20 - 120 Hz, 5 dB/Octave boost
Recorder: OYO DAS-1
Geophones: OYO, 3 component
Stations: 48
Channels: 96
Sample Interval: 2.00 msec
Samples per Trace: 400

- * 245 ft. = distance vibrator to far geophone
 10 ft. = distance vibrator to near geophone
 0 ft. = vibrator distance

Seismic Checkshots

The seismic checkshots at wells MSB21A and MSB78C were acquired with a downhole hydrophone string, and therefore only P-wave velocities were measured. The checkshot surveys at Wells MSB21TA, 26A, 69B, 82A, 85B and 85TA were acquired using a downhole clamping geophone. In these wells, both P-wave and S-wave velocities were recorded.

5.3 Data Processing

TDEM

In Section 6.2, the interpretation of TDEM data using standard geophysical techniques (1-D inversion, 1-D imaging) is given. A description of the fundamentals of TDEM soundings was given in the Phase II report. TDEM data taken during Phase II and Phase III of this project, and data taken in previous surveys (1988) are also incorporated in the interpretation. Two-dimensional geoelectric cross sections were constructed by linking the results of individual one-dimensional (1-D) inversions and images. For 1-D inversions and imaging, the program TEMIXH (Interpex Limited) was used. This program uses a ridge regression process to derive the geoelectric parameters (Inman, 1975). For the 1-D inversions, the minimum number of layers necessary to adequately fit the data was used (i.e., a significant improvement in error between model and data was not achieved by adding additional layers). All of the parameters of the 1-D geoelectric section (resistivities and thicknesses) were allowed to vary freely in the inversions.

For the 1-D imaging, a "smooth" model at each sounding location is estimated by Occam's Inversion (Constable, et al., 1987). Occam's Inversion seeks the smoothest geoelectric model which fits the data to within an expected tolerance. The advantage of this method is that the inversion is not "biased" by initial starting parameters and the results may be more appropriate for lithologies which vary gradually in space (e.g., facies changes).

Seismic

Processing of the seismic data was carried out by Blackhawk Geosciences Division personnel using the program SEISTRIX produced by Interpex Limited. The conventional processing sequence shown in Table 5.3 was chosen. For a complete description of seismic data processing see Yilmaz (1987). In petroleum exploration, because of complex geological structures, migration of the data to position seismic events in the correct space and time relationship is routinely carried out. Black, et al. (1994) showed that migration may be unnecessary for shallow-penetration data. Examination of the SRS data shows that 95% of the horizons dip is at less than 10 msec/75 ft (15 shotpoints). Using the "migrator's equation" (Black, et al., 1994), this representative dip would be migrated less than one sample interval in time (2 msec) and less than four shot points (20 feet) laterally. In addition, f-k migration was tested on the SRS data but the process significantly degraded the data due to smearing by the migration operator, thus masking any improvement in the spatial positioning of the stacked data. Black, et al. (1994), independently came to the same conclusion using other data.

Table 5.3 Processing Sequence, Savannah River Site

Sequence No.	Description
1	Cross-correlation with recorded filtered ground force signal.
2	Input geometry.
3	Resample data from 1 ms to 2 ms.
4	Truncate data at 800 ms.
5	Trace edit.
6	Gain application.
7	Spatial filter in the shot domain.
8	Surface consistent deconvolution.
9	Spatial filter in the receiver domain.
10	Bandpass filtering.
11	Mute of first breaks.
12	Velocity analysis using velocity spectra.
13	Surface consistent statics.
14	Velocity analysis using constant velocity stacks.
15	Normal moveout.
16	Residual statics.
17	Stack.
18	Display.

6.0 GEOPHYSICAL DATA INTERPRETATION

6.1 General Geology

The subsurface geology of the Savannah River Site (SRS) is known mainly from core descriptions supplemented by the following geophysical logs: natural gamma; 16" Normal, 64" Normal, and single point resistivity logs. Multiple geologic classifications of stratigraphy (e.g., Aadland et. al., 1992; Fallaw and Price, 1992) have been prepared for the SRS (Figure 1). In this report, a stratigraphic facies classification based on core descriptions from a site-wide distribution of wells is used for the general geological description, because it links stratigraphic packages to: (a) depositional environment; and (b) units likely to be recognized by surface geophysics. In this report, the core description reported by Everest Geotech Company (Houston, TX) was used for the description of the general geology (of SRS). Geological cross sections along Lines 1 and 2 of the geophysical surveys, based on core descriptions and geophysical well logs, were also supplied by SRS plant personnel for the A&M area. The core descriptions were contained in cross-sections, such as the one shown in Figure 6.2; while the geophysical logs were supplied in digital form. Verification of formation top data, and correlation between geophysical well logs and core data was not supplied to Coleman Research. These two studies show that the geology of the SRS is highly complex with rapid lateral and vertical variations in lithology. Figure 6.1 shows the correlation between chronostratigraphic, lithostratigraphic, and hydrostratigraphic units, and facies-based units (modified from WSRC, 1994 and Everest Geotech, undated).

Everest Geotech carried out a site-wide multi-disciplinary study including core descriptions and geophysical well logs of the sedimentary sequence underlying the SRS. The brief description which follows is taken mainly from their work. The SRS is underlain by crystalline rocks of different ages and compositions which have been identified by drill holes and aero-magnetic data (Petty et. al., 1965; Cumbest et. al., 1992). The Dunbarton graben developed in the crystalline basement in the Triassic, and filled with sediments of late-Triassic age which are now indurated. Subsequent sedimentation continued through the late-Eocene, with final deposition in the Miocene. The attribution of ages to the formations penetrated at SRS have been subject to much revision (Snipes et. al., 1993, and Fallow and Price, 1992). Everest Geotech identified six lithostratigraphic units in their work, unit I is the oldest, unit VI, the youngest; the following description is taken directly from their report.

"The sedimentary succession is sand-rich, generally unconsolidated, and varies in thickness from 800 to over 1,200 feet. Three major facies associations are present: coarse sand, fine silt and clay, and calcareous and glauconitic facies. The coarse sand facies are the predominant facies in the succession and are channel-bar and channel-fill deposits of sandy fluvial systems. Fine silt and clay facies are less common and are part of fining-upward channel sequences, floodplain deposits, and abandoned channel-fill deposits. The calcareous and glauconitic facies are characterized by abrupt lateral and vertical changes in distribution and thickness, and indicate riverine estuarine systems."

Figure 6.1 Correlation Between Chronostratigraphic, Lithostratigraphic, Hydrostratigraphic and Facies-Based Units

"Deposition of these facies occurred from the pre-Santonian through the Recent in multiple sandy fluvial systems that alternated between high and low sinuosity systems. The low sinuosity fluvial systems are multi-story, sheet-like sand deposits with little or no fine-grain sediments and include lithostratigraphic Units II, IV and VI. These sands were deposited during sea level low stands. High sinuosity fluvial systems have a greater proportion of fine silt and clay facies members, are characterized by unique mineralogies and vertically accreted interchannel deposits, and include Units I, III and V. These sediments were deposited during sea level high stands. Riverine estuarine conditions are restricted to the post-Cretaceous portion of the sedimentary section (Unit VI) and have highly variable vertical and lateral distributions. Deposition of these calcareous and glauconitic facies under these conditions occurred during sea level high stands."

The major aquifers, lithostratigraphic Units II, IV and VI, are composed of highly interconnected sheet sands. The major aquitards, Units I, III and V, are composed of more fine-grain sediments and clays with vertically interconnected channel sands. Unit VI contains multiple aquifer and aquitard units. Duncutting highly porous channel sands associated with the major Oligocene sea level lowstand have breached the localized aquitards in Unit VI as well as the more laterally and vertically extensive Unit V aquitard in the vicinity of P18TA, P28TA, P29TA, and P30TA. Well P30TA is shown in Figure 5.1, the others are located to the South of the area of the survey. The breaching of the aquitards is defined from core analysis.

6.1.1 Geology, A&M Area

Note: Well designations

The MSB series of wells were frequently drilled in clusters, separated laterally by only a few feet, in order to test contamination at different depths. In general, the "A" designation is a deep well to test the Crouch Branch Aquifer, the "B" designation is a well to test the Lost Lake Aquifer, and the "C" designation is a well to test the sandy unit immediately below the "Green Clay." The double letter designations, e.g., BB, is a subsequent well drilled between wells B and C. Wells with TA, or TB designations are wells which penetrate the Cretaceous sections. Only the well designations for which the core data was supplied are shown on Figure 5.1.

In order to provide background for the interpretation of the geophysical data, SRS personnel prepared two cross sections along Lines 1 and 2 showing petrographic data, natural gamma ray logs on Line 1, and core petrographic data, natural gamma ray and resistivity logs on Line 2. Core recovery was good with the exceptions of MSB20 and MSB26 on Line 1. The deepest wells penetrated only to the bottom of the Ellenton clay at a depth of approximately 320 feet (40 ft asl). The Ellenton clay is the lower unit of the Black Mingo group of Van Nieuwenhuise and Colquhoun (1982). The Ellenton clay corresponds to the Crouch Branch Confining Unit (CBCU) of Lewis and Aadland (1992). The closest well described by Everest Geotech is P30 (Figure 5.1), which penetrated crystalline basement at a depth of approximately 800 feet (-416 ft bsl). This well lies approximately 900 feet Southeast of Well MSB26. Unit

depth of approximately 800 feet (-416 ft bsl). This well lies approximately 900 feet Southeast of Well MSB26. Unit V of Everest Geotech appears to correspond to the Ellenton clay, and Unit VI appears to correspond to the Tertiary formations younger than the Ellenton clay (Figure 6.1). These two cross sections show that correlating the different lithologic units from well-to-well is very difficult. The lithologies seen in the cores are similar to those described by Everest Geotech and consist of sand, clayey sand, silt-clay and sandy clay deposited in beds which vary greatly in thickness from less than one foot to greater than 20 feet. It was shown from the core of Well P-30 that a sand unit, originating in the upper part of Unit VI (Upland Formation?), cut through the lower part of Unit VI, and through Unit V, the next lower unit (correlated with the CBCU) into the unit below. Such downcutting, on a greater and lesser scale, is probably typical of the depositional architecture of the sedimentary section under consideration. Although lithostratigraphic correlations can be made based on the resistivity and gamma logs, lithologic continuity of the layers is so variable that correlating individual clay layers from well to well is difficult.

6.1.2 Geophysical Well Log Data

The geophysical well logs were analyzed to determine changes in physical properties which may also represent changes in acoustic impedance which would affect the reflection seismic data. In general, increasing clay content correlates with an increase in the amplitude of the gamma ray log, although this correlation is not true of shallow clays at SRS (SRS personnel communication). The gamma ray log, which is a single point measurement, allows the discrimination of very thin beds, much thinner than the resolving capability of surface TDEM and seismic methods. Gamma ray response is low in sand layers. However, granitic gravels at SRS typically have high gamma ray response because they contain Potassium (K) feldspar. The resistivity logs show high resistivities over the sandy layers and lower resistivities over the silt and clay layers.

Natural Gamma Ray Logs

The natural gamma ray logs from the wells along Line 1 are shown in Figure 6.2. This line includes, from West to East, Wells MSB76C, 78D, 20A, 21A and 26A. The gamma ray logs from these wells have been stacked together with the correct vertical location. Correlation of individual gamma responses from log-to-log is problematic, but groupings of responses can be clearly seen to correlate across the sections. It is obvious that the Ellenton ("H" marker) in Well MSB26A has a high gamma ray response below an elevation of 130 feet. The upper surface of the upper clay can be discerned on Wells MSB21A and MSB76C where the response increases rapidly. Above this section there is a sandier interval which extends up to the region of 210 feet elevation. The area of high gamma ray response ("D" marker) between 190 and 200 feet on MSB26A has been designated as the Green Clay (SRS cross section). The response of this unit is similar on Well 76C but is less pronounced on the other wells. Above an elevation of 200 feet, the gamma ray response is subdued indicating a generally sandy section interrupted by more clayey layers. There appears to be greater clay content in wells MSB21A and MSB26A.

LINE 1 GAMMA LOGS

Figure 6.2 Line 1. Natural Gamma Ray Logs. Distances between wells in feet. Letters show position of interpreted seismic horizons. See Figure 6.17.

The gamma ray logs for Line 2 are shown in Figure 6.3. As on Line 1, the Ellenton clay ("H" marker), where it has been logged, is well defined on the gamma ray log at elevations of 148 ft to 160 ft. The "D" marker indicates the "Green Clay" at an elevation of 220 to 225 ft. This unit is succeeded upward by a sandier unit which persists to an elevation of approximately 270 feet. Above this unit, the gamma ray response becomes greater and shows significant clay layers up to 10 feet thick on all five logs.

The gamma ray logs for Line 3 are shown in Figure 6.4. The gamma ray response of the upper Ellenton clay ("H" marker) is quite distinct and shows that the clay is about 30 ft thick. However, the amplitude of the response is only about half of the response on the gamma logs of MSB26A or MSB69A which could be due to log calibration, fast rate of logging leading to poor counting statistics, change in lithology, or all three. The resistivities are higher which indicates that lithology may have changed. The response of the "Green Clay" ("D" marker) at an elevation of about 220 ft is similar to the responses seen on Lines 1 and 2. The responses above 280 feet are very similar to those on Line 2, but less so to those on Line 1.

Resistivity Logs

The resistivity logs are the 16"Normal and 64"Normal. The longer the electrode separations in the logging tool, the greater the penetration of the current. When the 16"Normal log shows lower resistivities than the 64"Normal log, it indicates invasion of the rock close to the wellbore by drilling fluids. However, the 16"Normal is more sensitive to thin beds than the 64"Normal, and was used to describe stratigraphy when its resistivity was not greatly different from the 64"Normal. The resistivities recorded on both logs (Figures. 6.5, 6.6 and 6.7) are very high and indicate that the pore water must be very fresh, as the resistivity of water is greatly diminished by the addition of dissolved salts, particularly sodium chloride. Low dissolved salt content is typical of SRS ground water.

On Line 1, the logs can be divided into several major divisions (Figure 6.5). At an elevation of 130 feet, the transition from the upper Ellenton clay to the next upper unit is marked by a sharp increase in resistivity on all logs. This change is followed by another at an elevation of about 145 feet. This moderately conductive unit is succeeded upwards by a highly resistive unit to an elevation of about 220 feet. A 25 foot thick conductive unit succeeds this resistor upwards on Well MSB26A but is not visible on Wells MSB21A or MSB76C. The upper 100 feet of the logs of MSB26A and 78D are dominated by a massive resistive zone probably reflecting a sand-dominated unit.

The situation is fairly similar on Line 2 (Figure 6.6) except that the upper Ellenton clay is more resistive on wells MSB69A, MSB85C and MSB48A. This relatively conductive zone is succeeded upwards on all four logs by more resistive zones which, in the case of MSB48A, is approximately twice as resistive as on the other wells. Again, a massive resistive unit is shown on MSB48A but which becomes thinner to the West. The upper part of the logs of MSB85A and MSB48A show alternating conductive and resistive units.

Figure 6.3 Line 2. Natural Gamma Ray Logs
Letters show position of interpreted seismic horizons

MMA LOGS

Distances between wells in feet.
 horizons. See Figure 6.18.

LINE 3 GAMMA LOGS

Figure 6.4 Line 3. Natural Gamma Ray Logs. Distances between wells in feet. Letters show position of interpreted seismic horizons. See Figure 6.19.

LINE 1 RESISTIVITY LOGS

Figure 6.5 Line 1. Resistivity Logs. Distance between wells in feet. Letters show position of interpreted seismic horizons. See Figure 6.17.

LINE 2 RESISTIVITY

Figure 6.6 Line 2. Resistivity Logs. Distance between position of interpreted seismic horizons.

LOGS

SB69A $\Leftarrow$ 706' $\Rightarrow$ MSB85A $\Leftarrow$ 983' $\Rightarrow$ MSB48A

wells in feet. Letters show
see Figure 6.20.

The 64" Normal resistivity logs of MSB67B and 68B are shown in Figure 6.7. The Ellenton clay is the conductive unit between 145-163 feet on MSB67B and 135-160 feet on MSB68B. The resistivities are higher than on Lines 1 and 2 and upper part of the Ellenton clay on MSB68B is noticeably more resistive than the lower part probably because it is sandier. There is great similarity between this log and MSB69A. The lower Ellenton is a resistive unit. The upper part of both logs is dominated by a thick resistive body whose upper surface is at an elevation of 280 ft and appears to be correlative to the thick resistive bodies of wells MSB85A and MSB48A (Line 2). This resistive body is succeeded upwards by a 30 foot conductive section at MSB68B. This conductive section is much thinner and more resistive at well MSB67B, a distance of only 250 feet.

Conclusions

The geophysical logs show that seismic reflectors should exist at the interfaces between well-developed clay zones, and well-developed sand zones. The gamma ray logs which tend to more uniquely identify clay beds, will be used to define reflector elevations in the boreholes, and, in turn, interpret the seismic reflectors in terms of geology. The geophysical logs clearly reflect rapid lateral and vertical lithological changes, particularly within the upper Ellenton clay. The clay facies is almost non-existent on the East side of Line 2 (wells MSB82A, 69A, 85A and 48A) and Line 3 (wells MSB67B and 68B).

The authors of the Everest Geotech report classified the depositional environment as a series of stacked, upward-fining units based on the petrographic descriptions. These units had a basal gravel, or coarse sand which became progressively finer-grained towards the top of the section, culminating in well defined clay layers. This depositional environment is also found in the A&M area and is best shown by the behavior of the gamma ray logs from Well MSB85A. This core-derived stratigraphic information corroborates the geophysical well logs in predicting that good seismic reflectors will be found in this section.

6.2 TDEM Results

6.2.1 General

In the following sections, the interpretation of TDEM data using standard geophysical techniques (1-D inversion, 1-D imaging) is given. A description of the fundamentals of TDEM soundings was given in the Phase II report. TDEM data taken during Phase II of this project, and data taken in previous surveys (1988) are also incorporated in the interpretation. Two-dimensional geoelectric cross sections were constructed by linking the results of individual one-dimensional (1-D) inversions and images. For 1-D inversions and imaging the program TEMIXH (Interpex Limited) was used. This program uses a ridge regression process to derive the geoelectric parameters (Inman, 1975). For the 1-D inversions, the minimum number of layers necessary to adequately fit the data was used (i.e., a significant improvement in error between model and data was not achieved by adding additional layers). All of the parameters

LINE 3 RESISTIVITY LOGS

Figure 6.7 Line 3. Resistivity Logs. Distance between wells in feet. Letters show position of interpreted seismic horizons. See Figure 6.21.

of the 1-D geoelectric section (resistivities and thicknesses) were allowed to vary freely in the inversions.

For the 1-D imaging, a "smooth" model at each sounding location is estimated by Occam's Inversion (Constable et. al., 1987). Occam's Inversion seeks the smoothest geoelectric model which fits the data to within an expected tolerance. The advantage of this method is that the inversion is not "biased" by initial starting parameters and the results may be more appropriate for lithologies which vary gradually in space (e.g., facies changes).

The interpretation of the geoelectric parameters into geologic/lithologic units was accomplished by assigning characteristic resistivity ranges to lithologic units using borehole information supplied by SRS personnel.

6.2.2 Repeatability in TDEM Measurements

In Phase III, several repeats of soundings made in Phase II were performed. Figure 6.8 shows the apparent resistivity curve for Sounding 132150Z taken in Phase II and the repeat data (4-180) taken in Phase III. The time period between these measurement sets was approximately seven months. Repeatability between the two apparent resistivity curves is extremely good; difference between data sets is less than 5% for all time gates.

6.2.3 Horizontal and Vertical Components

One objective of this contract is the development of algorithms for the forward modeling and inversion of horizontal field TDEM data. It has been shown in theory that an increase in sensitivity to the geoelectric section can be achieved with measurements of the horizontal field. Figure 6.9 shows the apparent resistivity curves derived from vertical field measurements in the center (Figure 6.9a) and outside (Figure 6.9b) the transmitter loop, and derived from horizontal field measurements outside (Figure 6.9c) the transmitter loop. These data were acquired from transmitter loop 1200 (Line 2) during Phase II. Below the apparent resistivity curves are the geoelectric sections derived from 1-D inversions of the three data sets. The agreement between the three data sets is excellent. All three data sets clearly map a clay layer at a depth of about 75 m (250 ft) below surface.

6.2.4 Detectability of Clay Layers with TDEM

The detectability of clay layers by electromagnetic methods depend on a number of factors, such as:

- A. Electrical conductivity, and conductivity contrast between clay layer and sand layers,
- B. Conductance of clay layer, (i.e., product of thickness and conductivity),

Figure 6.8 TDEM Soundings 132150Z Taken During Phase II and 4-180 Taken During Phase III. Comparison Of The EMFs Show That They Differ By Less Than 15% At All Time Gates.

Figure 6.9 Apparent Resistivity Derived from Horizontal and Vertical Magnetic Fields from a Single Transmitter Loop

- C. Depth of clay layer, and
- D. Complexity of geoelectric section, (e.g., the number of clay layers embedded in sands).

The influence of these factors are summarized in the detectability graph in Figure 6.10. For the construction of this graph, the response of clay layers embedded in a background resistivity of 1000 ohm-m was computed. Clay layers were placed at a depth of 300 ft. (91.4 m). Three layer thicknesses, 1 m, 3 m and 5 m, were chosen, and the emf's calculated for variable resistivity. On the vertical axis, the percentage difference between the electromotive force (emf) induced in a geoelectric section with a clay layer, and the emf of a geoelectric section without a clay layer (i.e., a half space), is plotted. It can be seen that thin layers can be detected if they are conductive and vice versa. Based on experimental evidence, a difference of 10% in emf is detectable. This difference is easily measured at the surface by the EM47 equipment.

Also shown on the graph is the property of the clay layer measured at the Savannah River Site and the observed percent differences in emf. Thus, the measurement at SRS verify the computations. The conclusion from this graph for the geoelectric section prevalent at the A & M site at SRS is that clay layers of relative low resistivity (< 30 ohm-m) and with a thickness of 3 ft (1m), are expected to be detectable. Part of the reason for the good detectability is the high resistivity of the sand matrix.

Factors that will deteriorate detection would be a sand matrix of lower resistivity, and multiple clay layers. In the event of multiple clay layers, the conductance of TDEM measurements likely will indicate the aggregate conductance of clay layers within certain intervals, but likely will not be able to distinguish between individual lenses.

6.2.5 Geoelectric Cross Sections (Lines 1 and 2)

The dominant objective of the TDEM survey is to map the continuity, depth, and thickness of clay layers in the stratigraphic column. Figures 6.11 and 6.12 show the geoelectric sections along Lines 1 and 2. Superimposed on the geoelectric section is the interpretation showing continuity of clay layers. The geoelectric section was derived from 1-D (ridge regression) inversion using discrete layers (3 or 4). Figures 6.11 and 6.12 also show the geoelectric section for Lines 1 and 2 using Occam's Inversion in which smoothly-varying resistivity layering is used. The conclusions derived from these data sets are:

1. Only the Ellenton clays, at a depth of about 250 ft (70 m) are mapped by the TDEM soundings. The variation in resistivity of the Ellenton clay is due to depositional facies change from clay to sand (SP2500E, Line 1), or by down cutting by later sand-filled channels (SP3300E, Line 1). The seismic horizons show that the thickness of the Ellenton clay varies little along the lines of the survey.

Figure 6.10 Response of clay layers of different thickness and resistivity embedded in a background resistivity of 1000 ohm-m. Results show that a clay layer of 60 ohm-m and thickness greater than 1.5m can be detected by EM-47 system.

WEST

LINE
TDEM
2-D GEOELECTRIC
(RIDGE REGR)

Figure 6.12 Line 2 Interpretation TDI conductance, and Occam's inversion shear

2

SEISMIC SECTION (PRESSION)

VERSION

ELEVATIONS OF INTERPRETED SEISMIC HORIZONS

2. Clay layers above the Ellenton (e.g., Green Clay) are not evident in the TDEM data

To better bring out the discontinuities in the geoelectric parameters, the conductance of the section from the surface to a depth of 400 ft was computed; conductance of a section is defined as:

$$\text{Equation (6.1)} \quad \text{Conductance (S)} = \sum_{i=1}^u \frac{h_i}{\rho_i} \text{ Siemens}$$

where h_i is the thickness and ρ_i is the resistivity of the i th layer. Conductance correlates to overall clay content. To illustrate their relationship, conductance is plotted below the 1-D discrete layer geoelectric section (Figures 6.11 and 6.12).

Inasmuch as the TDEM response comes only from the Ellenton clay, shown by the high-amplitude signals at late time gates, i.e., deep penetration, the change in conductance of the section can be directly correlated to change in electrical conductance of the Ellenton clay. To summarize, where the Ellenton is conductive, it has a high clay content, where it is more resistive, it has become sandy either by: (a) change in depositional environment; or (b) downcutting by later river channels.

6.2.6 Conductance Contour Map

In order to illustrate the variation of conductance across the A&M area, all of the TDEM data from all phases of the contract were assembled and contoured (Figure 6.13). Although station density is sparse away from Lines 1 and 2, the results show considerable variation in conductance which is related to lithologic changes in the clay content of the Ellenton Formation. There is a very steep gradient, from high conductance on the West to low conductance on the East, which runs approximately NNW-SSE through Well MSB82C on Line 2. The change in conductance is supported by the resistivity logs which penetrate the Ellenton (Figures 6.5, 6.6 and 6.7). The gradient probably represents a boundary in the Ellenton between clay-dominated facies to the West and sand-dominated facies to the East.

6.3 Reflection Seismic Survey

In reflection seismology, a seismic source at the surface sends a signal into the earth which is reflected back to the surface from interfaces between strata of different acoustic impedances; the reflections from deeper horizons occur at progressively later times. Given adequate well control, the seismic method is the surface geophysical method with the greatest resolution of thin layers, and is the one which can be related directly to stratigraphy. At SRS, the ability to perform checkshots (described below) for velocity control, core descriptions from wells on the seismic lines, and geophysical well logs, provided many of the elements necessary to interpret the seismic data in detail.

TOTAL CONDUCTANCE (SURF)

LEGEND

△ 1988 T

○ 1994 T

CE TO 400 FT.)

Figure 6.13 Total Conductance. Map of A&M Area based on data from 1988 and 1989 contract surveys.

6.3.1 Shear Wave Seismic Survey

Tests by Blackhawk Geosciences Division (BGD) at SRS in March 1994 (contract, Phase II Report, 1994) showed that very little P-wave information was returned to the surface (the P-wave survey could distinguish only the water table). The P-wave is a compressional wave and travels through the earth in a series of compressions and rarefactions as the particles sequentially move forward and back along the direction of travel of the wave (Figure 4.12). Conversely, when a shear wave (S-wave) passes through the earth, the particles vibrate in all directions at 90 degrees to the direction of travel of the wave (Figure 4.12). Shear waves travel only through the rock matrix, not through the water in the pores. Depending on the velocity structure of the earth, and the presence of fractures, certain components of the S-wave may be delayed with respect to other components; this phenomenon, known as birefringence, has been observed in both fractured rock, and in clay layers. In the March 1994 tests at SRS, BGD showed that it was possible to distinguish good reflectors using a shear-wave source. The survey described below was carried out using shear waves generated by a vibrational source and recorded by special geophones which record in orthogonal directions.

6.3.2 Checkshot Surveys

Checkshot surveys are carried out in order to define the change of velocity with depth, and, if several wells are surveyed, the lateral variation in velocity can also be determined. The technique consists of lowering a 3-component geophone down the well, clamping it against the casing at 10 or 20 foot intervals, and recording the seismic signal from the surface. The following wells were surveyed in this way:

Line 1: MSB21TA, MSB26A

Line 2: MSB69B, MSB89A, MSB85B, MSB85TA.

(See note on page 6-3 regarding well designations and locations.)

The resulting curves of two-way-time versus depth are shown in Figures 6.14 and 6.15. Comparison of the two sets of checkshot results show that the velocity field in the area is quite uniform. A polynomial was fitted to the checkshot curves in order to derive a function for depth conversion of the two-way times to the seismic horizons (Figure 6.16). However, the individual checkshots for each well were used to determine their relationship between the geophysical well logs, the core data, and the seismic horizon.

Figure 6.14 Shear Wave Checkshots SRS - Line 1

Figure 6.15 Shear Wave Checkshots SRS - Line 2

Polynomial Used for Conversion to Depth from Two-Way Travelltimes
 $Z=8.504+T(0.518)+T^{**}2(0.000416)$

Figure 6.16 Depth-Time Curve Used to Convert Seismic Two-Way Times To Depth

6.3.3 Results and Interpretation

Because In-line and X-line orientations for both source and receivers (geophones) were used, four data sets were recorded along Lines 1, 2 and 3:

<u>Source</u>	<u>Receiver</u>
X-line	X-line
In-line	In-line
X-line	In-line
In-line	X-line

There are differences between each of the four data sets indicating that the seismic energy may have traveled along the path source-reflector-receiver at different velocities. The X-line/X-line components generated the most continuous reflectors, and, so, were used for the initial interpretation of the data which is presented below. The other three components also contain significant information concerning the nature of the stratigraphy, but this information can be accessed only by further processing of the data.

The final processed sections of the X-line/X-line data are shown on the upper part of Figures 6.17, 6.18 and 6.19; shot point numbers are listed along the top of the section and time is displayed vertically at intervals of 10 milliseconds. Reflectors (time horizons) were picked on the basis of their continuity across the seismic sections, and correlation with the geophysical well logs. Each reflector was given a letter designation, starting with "A" for the earliest (shallowest) discernable and proceeding to "L" for the latest (deepest). The character of the reflectors is described in detail below. Not all reflectors occur on each line which may be due to lateral stratigraphic change and/or low reflectivity between the layers. On each line, reflectors A and C are the two earliest events. On Line 1, they occur at 140 msec and 185-200 msec two-way travel time (TWT), respectively, while on Lines 2 and 3, they occur earlier at 95-140 msec and 110-180 msec (TWT), respectively. Because these two reflectors are the earliest, they were given the same designation on each line and may represent the same geologic horizon. Reflector B varies considerably in time but its earliest occurrence is at the same time, 150 msec at SP475, Line 1, and 150 msec at SP1060, Line 2; it does not occur on Line 3. On both Lines 1 and 2, reflector B has an angular relationship with reflectors A and C and may represent the same event. Reflectors D, F and H are seen on all three lines. Reflectors I and J are seen on Lines 2 and 3; reflectors E and G occur only on Line 1, while reflectors K and L occur on Lines 1 and 3 (Figures 6.17, 6.18 and 6.19).

LINE 1A

Figure 6.17A Line 1. Shear wave r
section. Bottom: interpret
Figure is presented in 3 panels, A

MSB20A

250 265 280 295 310 325 340 355 370 385 400 415 4.

ection seismic section. Top: uninterpreted
section. Well control indicated.
B, and C, from Southwest to Northeast.

MSB20A

250 265 280 295 310 325 340 355 370 385 400 415 4.

LINE 1B

Figure

17B

LINE 1C

LINE 1

D:\SRS2\L1BLU\ALL1.STK

RECORDING PARAMETERS

RECORDED BY:
INSTRUMENT:
SOURCE:
SOURCE INTERVAL: 5.00 ft
CHANNELS: 48
SAMPLES/TRACE: 800
SPREAD: OFF-ON: 0 - 20.0 - 255.0

DATE RECORDED: 2-FEB-0
FIELD FILTERS:
GEOPHONES:
GROUP INTERVAL: 5.00 ft
SAMPLE INTERVAL: 1.00 ms
FOLD: 2400 X

PROCESSING SEQUENCE

1) CROSSCORRELATION
2) CMP SORT
3) TRACE EDIT
4) RESAMPLE
0.500 ms TO 1.000 ms
5) WINDOW DATA
0 ms TO 800 ms
6) GAIN
TRACE BALANCING
RMS AMPLITUDE: 0 - 798 ms
7) FK FILTER
POLYNOMIAL FILTER
8) SURFACE CONSISTENT PREDICTIVE DECONVOLUTION
SHOTS AND RECEIVERS
CMP: 5 ms
OPERATOR LENGTH: 100 ms

9) BANDPASS FILTER
20-40-80-120 Hz
ZERO PHASE
10) MUTE FIRST BREAKS
11) NORMAL MOVEOUT
STRETCH MUTE: 130%
12) SURFACE-CONSISTENT RESIDUAL STATICS
CORRELATION WINDOW: 180 TO 500 ms
MAXIMUM SHIFT: 5 ms
13) CMP-CONSISTENT RESIDUAL STATICS
PILOT 9 CMPS
CORRELATION WINDOW: 180 TO 500 ms
MAXIMUM SHIFT: 5 ms
14) CMP STACK
15) RECORD MERGE

DISPLAY PARAMETERS

HORIZONTAL SCALE: 30.00 ft / in
POLARITY: NORMAL
DATE PROCESSED: 24-JAN-85

VERTICAL SCALE: 7.00 in / sec
GAIN: AGC - WINDOW 250 ms

CEES-Blackhawk Geosciences Division

SOURCE-RECEIVER GEOMETRY LAYOUT

SOURCE RECEIVER

X-LINE X-LINE

LINE 1

D:\SRS2\L1BLU\ALL1.STK

RECORDING PARAMETERS

RECORDED BY:
INSTRUMENT:
SOURCE:
SOURCE INTERVAL: 5.00 ft
CHANNELS: 48
SAMPLES/TRACE: 800
SPREAD: OFF-ON: 0 - 20.0 - 255.0

DATE RECORDED: 2-FEB-0
FIELD FILTERS:
GEOPHONES:
GROUP INTERVAL: 5.00 ft
SAMPLE INTERVAL: 1.00 ms
FOLD: 2400 X

PROCESSING SEQUENCE

1) CROSSCORRELATION
2) CMP SORT
3) TRACE EDIT
4) RESAMPLE
0.500 ms TO 1.000 ms
5) WINDOW DATA
0 ms TO 800 ms
6) GAIN
TRACE BALANCING
RMS AMPLITUDE: 0 - 798 ms
7) FK FILTER
POLYNOMIAL FILTER
8) SURFACE CONSISTENT PREDICTIVE DECONVOLUTION
SHOTS AND RECEIVERS
CMP: 5 ms
OPERATOR LENGTH: 100 ms

9) BANDPASS FILTER
20-40-80-120 Hz
ZERO PHASE
10) MUTE FIRST BREAKS
11) NORMAL MOVEOUT
STRETCH MUTE: 130%
12) SURFACE-CONSISTENT RESIDUAL STATICS
CORRELATION WINDOW: 180 TO 500 ms
MAXIMUM SHIFT: 5 ms
13) CMP-CONSISTENT RESIDUAL STATICS
PILOT 9 CMPS
CORRELATION WINDOW: 180 TO 500 ms
MAXIMUM SHIFT: 5 ms
14) CMP STACK
15) RECORD MERGE

DISPLAY PARAMETERS

HORIZONTAL SCALE: 30.00 ft / in
POLARITY: NORMAL
DATE PROCESSED: 24-JAN-85

VERTICAL SCALE: 7.00 in / sec
GAIN: AGC - WINDOW 250 ms

Figure 6.17C

LINE 2A

Figure 6.18A Line 2. Shear wave section. Bottom: interpreted Figure is presented in 2

Reflection seismic section. Top: uninterpreted
Interpreted section. Well control indicated.
Channels, A and B, from West to East

MSB85A

LINE 2B

MSB85A

LINE 2

D:\SRS2\L2BLU\STAT2.STK

RECORDING PARAMETERS

RECORDED BY: INSTRUMENT: SOURCE: SOURCE INTERVAL: 5.00 ft CHANNELS: 48 SAMPLES/TRACE: 400 SPREAD: OFF-END: 245.0 - 10.0 - 0	DATE RECORDED: 0-JAN-0 FIELD FILTERS: GEOPHONES: GROUP INTERVAL: 5.00 ft SAMPLE INTERVAL: 2.00 ms FOLD: 2400 X
---	---

PROCESSING SEQUENCE

1) IMPORT FIELD RECORDS 2) CMP SORT 3) TRACE EDIT 4) CMP SORT 5) TRACE EDIT 6) RESAMPLE 7) WINDOW DATA 8) GAIN 9) FK FILTER 10) SURFACE CONSISTENT PREDICTIVE DECONVOLUTION	11) BANDPASS FILTER 12) TRACE EDIT 13) FK FILTER 14) MUTE FIRST BREAKS 15) SURFACE-CONSISTENT RESIDUAL STATICS 16) NORMAL MOVEOUT 17) SURFACE-CONSISTENT RESIDUAL STATICS 18) CMP STACK
--	--

DISPLAY PARAMETERS

HORIZONTAL SCALE: 30.00 ft / in POLARITY: NORMAL DATE PROCESSED: 23-JAN-85	VERTICAL SCALE: 7.00 in / sec GAIN: AGC - WINDOW 258 ms
--	--

CEES - Blackhawk Geosciences Div.

LINE 2

D:\SRS2\L2BLU\STAT2.STK

RECORDING PARAMETERS

RECORDED BY: INSTRUMENT: SOURCE: SOURCE INTERVAL: 5.00 ft CHANNELS: 48 SAMPLES/TRACE: 400 SPREAD: OFF-END: 245.0 - 10.0 - 0	DATE RECORDED: 0-JAN-0 FIELD FILTERS: GEOPHONES: GROUP INTERVAL: 5.00 ft SAMPLE INTERVAL: 2.00 ms FOLD: 2400 X
---	---

PROCESSING SEQUENCE

1) IMPORT FIELD RECORDS 2) CMP SORT 3) TRACE EDIT 4) CMP SORT 5) TRACE EDIT 6) RESAMPLE 7) WINDOW DATA 8) GAIN 9) FK FILTER 10) SURFACE CONSISTENT PREDICTIVE DECONVOLUTION	11) BANDPASS FILTER 12) TRACE EDIT 13) FK FILTER 14) MUTE FIRST BREAKS 15) SURFACE-CONSISTENT RESIDUAL STATICS 16) NORMAL MOVEOUT 17) SURFACE-CONSISTENT RESIDUAL STATICS 18) CMP STACK
--	--

DISPLAY PARAMETERS

HORIZONTAL SCALE: 30.00 ft / in POLARITY: NORMAL DATE PROCESSED: 23-JAN-85	VERTICAL SCALE: 7.00 in / sec GAIN: AGC - WINDOW 258 ms
--	--

CEES - Blackhawk Geosciences Div.

Figure 6.18B

LINE 3

D:\SRS2\L3BLU\CVS.STK

RECORDING PARAMETERS	
RECORDED BY:	DATE RECORDED: 0-JAN-0
INSTRUMENT:	FIELD FILTERS:
SOURCE:	GEOPHONES:
SOURCE INTERVAL: 5.00 ft	GROUP INTERVAL: 5.00 ft
CHANNELS: 48	SAMPLE INTERVAL: 2.00 ms
SAMPLES/TRACE: 400	FOLD: 2400 X
SPREAD: OFF-DIC: 245.0 - 10.0 - 0	

PROCESSING SEQUENCE	
1) IMPORT FIELD RECORDS	SHOTS AND RECEIVERS
2) CMP SORT	GAP: 5 ms
3) TRACE EDIT	OPERATOR LENGTH: 100 ms
4) RESAMPLE	9) FK FILTER
1,000 ms TO 2,000 ms	POLYDROMAL FILTER
5) WINDOW DATA	10) BANDPASS FILTER
0 ms TO 800 ms	10-20-60-120 Hz
6) GAIN	ZERO PHASE
ACC = WINDOW: 51 ms	11) MUTE FIRST BREAKS
7) FK FILTER	12) NORMAL MOVEOUT
POLYDROMAL FILTER	STRETCH MUTE: 150%
8) SURFACE CONSISTENT PREDICTIVE DECONVOLUTION	13) CMP STACK

DISPLAY PARAMETERS	
HORIZONTAL SCALE: 30.00 in / in	VERTICAL SCALE: 7.00 in / sec
POLARITY: NORMAL	GAIN: ACC = WINDOW 250 ms
DATE PROCESSED: 07-FEB-03	

CEES-Blackhawk Geosciences Division

LINE 3

D:\SRS2\L3BLU\CVS.STK

RECORDING PARAMETERS	
RECORDED BY:	DATE RECORDED: 0-JAN-0
INSTRUMENT:	FIELD FILTERS:
SOURCE:	GEOPHONES:
SOURCE INTERVAL: 5.00 ft	GROUP INTERVAL: 5.00 ft
CHANNELS: 48	SAMPLE INTERVAL: 2.00 ms
SAMPLES/TRACE: 400	FOLD: 2400 X
SPREAD: OFF-DIC: 245.0 - 10.0 - 0	

PROCESSING SEQUENCE	
1) IMPORT FIELD RECORDS	SHOTS AND RECEIVERS
2) CMP SORT	GAP: 5 ms
3) TRACE EDIT	OPERATOR LENGTH: 100 ms
4) RESAMPLE	9) FK FILTER
1,000 ms TO 2,000 ms	POLYDROMAL FILTER
5) WINDOW DATA	10) BANDPASS FILTER
0 ms TO 800 ms	10-20-60-120 Hz
6) GAIN	ZERO PHASE
ACC = WINDOW: 51 ms	11) MUTE FIRST BREAKS
7) FK FILTER	12) NORMAL MOVEOUT
POLYDROMAL FILTER	STRETCH MUTE: 150%
8) SURFACE CONSISTENT PREDICTIVE DECONVOLUTION	13) CMP STACK

DISPLAY PARAMETERS	
HORIZONTAL SCALE: 30.00 in / in	VERTICAL SCALE: 7.00 in / sec
POLARITY: NORMAL	GAIN: ACC = WINDOW 250 ms
DATE PROCESSED: 07-FEB-03	

CEES-Blackhawk Geosciences Division

Figure 6.19 Line 3. Shear wave reflection seismic section. Top: uninterpreted section. Bottom: interpreted section. Well control indicated.

The geologic interpretation of the seismic sections was carried out by identifying time horizons on the seismic sections, calculating their depths using the checkshot data, and correlating them with the gamma ray and resistivity well log cross sections (Figures 6.2 through 6.7). The seismic reflectors so identified are shown on the gamma ray and resistivity log cross-sections (Figures 6.2 through 6.7), and on the lower panels of Figures 6.17, 6.18 and 6.19. The descriptions given below are keyed to Figures 6.17, 6.18 and 6.19.

Line 1

- Reflector A. Dips from West to East across the section. It is most continuous from SP640 to SP880, and has a pronounced dip between SP730 and SP775. On the gamma ray logs, reflector A occurs between an upper sandier layer and a lower, clay-dominated layer. It is not defined at Well 21A, and appears as a reflector below a clay dominated layer at an elevation of 260 feet at Well MSB26A.
- Reflector B occurs between SP475 and 880 and has a pronounced "channel" appearance between SP730 and 805 (see F, below). It is clearly defined on the resistivity log of Well MSB26A (Figure 6.5) as the base of a thick high-resistivity unit, probably a down-cutting sand.
- Reflector C. Partly discontinuous, and completely missing between SP730 and SP805; almost horizontal. On the gamma ray logs, it maps the transition from an upper sandy layer to a lower clay layer except at MSB26AA where it marks the transition from a clay layer to a sandier layer.
- Reflector D. Continuous across the section, except between SP730 and SP805. On the gamma ray logs, it is identified as the top of the "Green Clay" and marks the transition from a sandier layer above to a layer with higher clay content below. The "Green Clay", with an indicated maximum thickness of approximately 6 feet (SRS cross-section A-A') is too thin to have both top and bottom surfaces imaged as separate reflectors.
- Reflector E. Occurs only between SP112 and 378 at the West end of the line. On Well 76C, it can be seen to mark the transition between an upper sandy layer and a lower clay layer. The upper sandy layer may represent a channel sand which was not deposited East of SP378.
- Reflector F. Occurs between SP550 and 880 at the East end of the line. It is possible to trace it between SP730 and 805. It's defined on Well MSB26A as the transition between a sandy upper unit and a clay lower unit.

- Reflector G. Occurs only between SP715 and 850. Horizontal between SP805 and 880, then increases markedly in time to a maximum of 470 msec then rapidly rises to its original level. At Well MSB26A it is the transition between clay upper unit and a sand lower unit, and is interpreted to be a channel sand which has scoured downwards through the clay layers below. It truncates reflectors H, I and J.
- Reflector H. Undulates continuously across the section except between SP700 and 805. Marks the top of the upper clay of the Ellenton Formation. The transition is particularly well marked on the gamma ray log of Well MSB26A.
- Reflector I. More or less continuous across the section except between SP730 and 805. Marks the top of the lower clay, Ellenton Formation (SRS cross-section A-A'). Clearly defined on the gamma ray log of Well MSB26TA at an elevation of 92 feet.
- Reflector J. Relatively continuous across the section except where cut out by reflector G between SP730 and 805. Marks the bottom of the lower clay, Ellenton (SRS cross-section A-A').
- Reflector K. Quite continuous from SP420 to 880, including between SP730 and 805. There is no well information available at this depth on the line.
- Reflector L. Relatively continuous across the whole section at about 600 msec which represents a depth of approximately 400 feet. There is no well information available at this depth on the line.

Line 2

Few of the reflectors on Line 2 are completely continuous along the line. The reasons for the lack of continuity are primarily due to, in order:

- Lack of geological continuity,
- Reflections from out-of-the-plane of the section,
- Signals not recorded due to near-surface conditions.

A geologic cross-section prepared independently by SRS personnel shows lack of lateral continuity of lithologic units. Nonetheless, the reflectors identified on the section can also all be identified and correlated on the gamma ray logs except those which occur below the logged sections of the wells. A brief description of each reflector and its geological significance follows:

- Reflector A. Clearly defined as a peak-trough-peak doublet at Well MSB85, discontinuous across most of the line. Represents the change from a sand-dominated section to a clay-dominated section at a depth of approximately 60 feet. Dips gently to the East.
- Reflector B. Clearly defined at Wells MSB29A and MSB82A. Dips to the East from the 150 ms (TWT) at Well MSB29A to 210 ms (TWT) at Well MSB82A. Probably represents the bottom of a sandy interval which has downcut into the lower clay sections.
- Reflector C. Clearly defined as a peak-trough-peak doublet at Well MSB69, and is fairly continuous at the East end of the line. Represents the transition from a clay dominated to a sand-dominated section. Dips gently to the East.
- Reflector D. This reflector is continuous, and is a clearly defined doublet which appears at the bottom of a thick sandy unit overlying a thin clay identified as the "Green Clay" (SRS cross-section C-C'). The reflector undulates considerably which probably indicates downcutting by the streams which laid down the overlying sands.
- Reflector E is not found on Line 2.
- Reflector F. Reflector F is defined at MSB82A, MSB69 and MSB85. Towards the West end of the line, it terminates between shotpoints 1165 and 1180. This reflector is at the interface between an overlying sand and an underlying clay. It can be seen on Figures 6.3 and 6.6 that the sand thickness becomes much greater at Wells MSB82A, MSB69 and MSB85, with a maximum thickness of 25 feet at MSB82A. The thickening is probably due to downcutting.
- Reflector G is not found on Line 2.
- Reflector H. Reflector H is a clearly defined peak-through-peak doublet at Wells MSB82A, MSB69A and MSB85A. It can be traced continuously across the whole line. This reflector occurs at the depth/time identified by SRS personnel as the clay at the top of the Ellenton formation, and is shown on the gamma ray logs (Figure 6.3) as the interface between the sandy interval above and a more clayey interval below. It shows little variation in depth across the line.
- Reflector I. Reflector I can be distinguished on the gamma ray logs of Wells MSB29TA and MSB82TA. The corresponding reflector is quite discontinuous which may indicate a change in acoustic impedance between the more clayey layer above and the sandier unit below. This reflector may represent the bottom of the lower clay of the Ellenton (SRS cross-section C-C').

- Reflector J. Reflector J is represented by a peak-through-peak doublet on the seismic section at around the two-way time representing a depth of 345 feet. This reflector may correspond to the bottom of the Ellenton (SRS cross-section C-C').

Line 3

Line 3 is a 500-foot long line situated due South of the Eastern half of Line 2 (Figure 5.1). Wells MSB67B and 68B occur at shot point 2070 and immediately East of shot point 2115, respectively (Figure 6.19). Most of the reflectors recognized on Lines 1 and 2 also appear on Line 3. No geological information was available for these wells. The brief description of each reflector (time horizon) and its geological significance follows:

- Reflector A. Clearly defined as a peak-trough-peak doublet at both wells. Between MSB68 on the East, and MSB67B in the center of the line, a distance of 250 feet, the time at which reflector A occurs changes from 125 MS two-way travel time (TWT) to 100 MS (TWT); this 25 msec difference represents an elevation change of about 16.5 feet, and the effect is quite marked on the seismic section. Occurs at the base of a low resistivity section, probably a clay.
- Reflector B. Does not occur on Line 3.
- Reflector C. Well defined doublet, sharing the lower peak of reflector A. Also decreases in time between Wells MSB68 and 67. Represents the top of a massive resistive unit at both wells.
- Reflector D. Continuous across the section, based on similarities of the geophysical well logs with Lines 1 and 2, is identified as the interface between a sand above and the "Green Clay" below.
- Reflector E. Does not occur on Line 3.
- Reflector F. Does not occur on Line 3.
- Reflector G. Does not occur on Line 3.
- Reflector H. Based on similarities with the logs of Lines 1 and 2, is identified as the upper surface of the upper Ellenton clay. Continuous across the section.
- Reflector I. It continues across the section, occurs about 35 feet below the conductive part of the Ellenton clay and was identified on Lines 1 and 2 as the bottom of the lower clay of the Ellenton.

- Reflector J. Continuous across the section, identified on Lines 1 and 2 as the bottom of the Ellenton Formation.
- Reflector K. Also continuous across the section, no geological identification made as the wells do not go deep enough in the A&M area.
- Reflector L. Continuous across the area, no geological interpretation made.

6.4 Integrated Interpretation

The integrated interpretation results described below show the power of combining surface geophysical measurements which measure different properties of the earth with good geological and geophysical well control to produce an accurate synthesized interpretation of stratigraphy and lithology. Specifically, the shear wave seismic measures the acoustic impedance (product of velocity and density) of the rock matrix (shear waves do not propagate through water), and TDEM measures the electrical conductivity of the rock, which depends on water salinity and clay content. Because the well information (cores and geophysical logs) shows that the majority of clay in the section occurs in the Ellenton, the TDEM interpreted conductances are a measure of the Ellenton's clay content. The TDEM survey shows that there is considerable lateral change in conductance in the Ellenton Formation, which may be caused by either facies changes or faulting. First, the resolution of the geophysical methods employed will be discussed in Section 6.4.1. The possibility that faults may exist in the A&M area will be examined in Section 6.4.2. In Section 6.4.3, the TDEM and seismic results will be presented. The stratigraphic interpretation of those results, synthesized with well log and core descriptions from well-based geological cross-section supplied by SRS, will be described in Section 6.4.4.

6.4.1 Resolution of Geophysical Methods

Time Domain Electromagnetics

At SRS, the TDEM data of this survey show that all of the response comes from the electrically-conductive Ellenton clay. Two parameters are realized from interpretation of the data: (a) depth to the top of the clay layer; and (b) the conductance of the clay layer which is the product of clay thickness and its conductivity.

Depth Resolution

Empirical results over many years show that depth resolution is on the order of 5 to 10% of the depth to the layer. At the A&M area, depth resolution of the top of the Ellenton clay at a depth of 210 feet, is 11-21 ft.

Conductance Resolution

The accuracy of the conductance measurement is greater than 90% (Kaufman and Keller, 1983). Determining its component parts, thickness and conductivity requires independent information about either of these parameters.

Reflection Seismic

The resolving capability of reflection seismic depends on the wavelength of the seismic signal:

$$\text{Wavelength} = \text{Velocity/Frequency}$$

Model studies show that beds whose thickness is one quarter of a wavelength can be resolved. Because reflections are created by energy returning from the upper and lower surfaces of a bed, the depth resolution to the top of a bed will also be one quarter of a wavelength. Offsets in a bed surface due to faulting can be resolved if they are larger than one quarter of a wavelength (Sheriff, 1978).

Shear Waves

In the A&M area, shear wave velocity, measured from time section, is 1,300 ft/sec, and center frequency is 60 Hz giving a wavelength of 21 feet. Thus:

$$\begin{aligned}\text{Depth and bed thickness resolution} &= 5.25 \text{ feet} \\ \text{Fault resolution} &= 5.25 \text{ feet.}\end{aligned}$$

Figure 6.20 shows P-wave and S-wave check shot data from Wells MSB21TA and MSB76C (contract, Phase II Report). Above the water table, the S-wave velocity, 1250 ft/sec, is exactly one-half of the P-wave velocity, 2,500 ft/sec. Below the water table, the P-wave velocity increases to 6,000 ft/sec but the S-wave velocity remains constant. This disparity can be explained by the fact that the sediments in the upper 400 feet at SRS are unconsolidated having never been buried.

Compression (P) Waves

Domoracki (1994) derived a P-wave velocity of 6,600 ft/sec which compares favorably with BGD check shot data; center frequency is 50 Hz, giving a wavelength of 132 feet. Thus:

$$\begin{aligned}\text{Depth and bed thickness resolution} &= 33 \text{ feet} \\ \text{Fault resolution} &= 33 \text{ feet.}\end{aligned}$$

6.4.2 Faulting

Basement faulting within SRS has been defined using core data (Snipes, et. al., 1993), gravity and magnetics supplemented by core data (Cumbest, et. al., 1992), and P-wave reflection seismic (Chapman and DiStefano, 1989). Faulting within the stratigraphic section has been defined using core data (Snipes, et. al., 1993), and in the A&M Area, using geophysical well logs (SRS Hydrogeologic Cross-Section C-C', October 18, 1992, Figure 6.27, this report). These methods can be very accurate when the fault offset is close to vertical as in normal or reverse faults; it is more difficult to accurately define other types of faulting such as strike-slip or thrust.

Within SRS, the fault with the best documentation is the Pen Branch fault which runs Northeast-Southwest approximately through the center of the site (Snipes, et. al., 1993), and appears to mark the Western edge of the Triassic-Jurassic (?) clastic sediments (Cumbest, et. al., 1992; Figure 6.21). The most definitive indicator of the Pen Branch fault is the offset seen on a seismic reflector originally identified with the interface between the Cretaceous and pre-Cretaceous rocks (Chapman and DiStefano, 1989). Figure 6.22 shows the signature of the Pen Branch fault on seismic line SRS-7 (Domoracki, 1994); the signature is a clear offset of the peak-trough-peak doublet which represents the sediment-basement interface. To the Northwest of the Pen Branch fault, the basement is igneous, and metamorphic rock, to the Southeast of the Pen Branch fault, the pre-Cretaceous rocks are Triassic redbeds developed in the Dunbarton basin. These P-wave seismic results, as well as the results from gravity and magnetic modeling, show that faulting in this area involves basement displacement.

Faulting: A&M Area

The borehole velocity profile from Well DRB10 (Figure 6.23) shows that the Triassic basin fill has velocities in the same range as does the crystalline basement at the site, i.e., average of 15,000 fps (4,500 mps) (Domoracki, 1994). The upper surface of the crystalline rock is covered by a saprolite zone up to 30 feet thick; there is a similar weathered zone on top of the Triassic redbeds. The saprolite/weathered zone has a seismic velocity of 7,000 fps (2,143 mps). The sediments above the saprolite zone average about 6,600 fps (2,000 mps). The large velocity contrasts between the Mesozoic sediments and the saprolite/weathered zone and the underlying crystalline basement-Triassic redbeds means that a complex reflected signal will return to the surface.

A portion of seismic line SRS-12, originally recorded and processed by Conoco (Chapman and DiStefano, 1989), and reprocessed by Domoracki (1994) is shown in Figure 6.24. Of the three reflectors interpreted by Domoracki, only one reflector could be unquestionably determined in this study. This reflector was identified to be basement and is a doublet consisting of a peak-trough-peak sequence. The location of two faults identified by Domoracki are also indicated on the section at SP350 and 510 (Stieve, personal communication). The location of wells MSB43A, 29A, 82A, 69A, 85A and 48A are also shown, as is the location of BGD shear wave reflection Line 2.

Figure 6.20 P-wave and S-wave checkshots in Wells MSB21TA and MSB76C: Note depth to water table and increase in P-wave velocity to 6,000 ft/sec below it. S-wave velocity is unchanged.

Figure 6.21 Basement Fault Map: Seismic lines and faults that offset the top of basement reflection are shown. Seismic lines SRS-7 and 12 are emphasized (modified from Domoracki, 1994).

LINE SRS-7

Figure 6.22 Portion of Seismic Line SRS-7, showing red, blue, and yellow markers; and the signature of the Pen Branch Fault (PBF) (from Domoracki, 1994).

Figure 6.23 Velocity Profile Borehole DRB-10: Interval velocity, average velocity, and RMS velocity (from Domoracki, 1994).

Figure 6.24 Portion of P-wave seismic line SRS-12 showing interpreted faults F1 and F3, and red horizons (Domoracki, 1994), Fault 2 (SRS Hydrogeologic Cross Section C-C', 10/18/92), well locations, and BGD S-wave seismic Line 2.

The criteria used by Domoracki to recognize faults is described in his report (1994, pg. 48) and is reproduced here.

"The identification of faults is based on time offset of reflections and is restricted to offsets that demonstrably originate in the basement and penetrate upward into the Coastal Plain sediments. The identification of faults is not always unambiguous; small time offsets and fault planes oblique to the line of profile complicate the issue. Furthermore, some faults of considerable offset are detected on only one or two profiles and thus, the orientation of these structures is subjective. Minor static busts can imitate the appearance of faults and can lead to the misinterpretation of the sense of fault movement (Appendix A). Draped folding (Stearns, 1978) of sediments over a basement fault is not mapped as a fault penetrating to the level of the folding regardless of the radius of curvature of the folding. That is, deformation by faulting (offset) is distinguished from deformation by folding.

It must be emphasized that if a particular time horizon is observed to not be offset by a fault it does not preclude the possibility that the reflector is offset in depth. The statement of whether or not a time horizon is faulted must be qualified as being within the limit of resolution of the data. Generally, the resolution of two events separated by greater than an eighth or quarter wavelength is possible (Sheriff, 1977). Therefore, for the SRS data, given an average frequency of 50 Hz, and an average velocity of the Coastal Plain sediments of 2000 m/s, the vertical resolution limit is 5-10 m. That is, faults of less than 5-10 m throw are at the resolution limit of these data."

Domoracki (1994) indicates that, among others, line SRS-12 had particularly bad statics problems, "for certain problem lines residual statics were recomputed several times for a given pass until a suitable section was obtained". The effect of improper, or incomplete, statics corrections is that time horizons may be displaced, earlier or later, from their proper position. At shotpoints 349 and 335, two velocity analyses were carried out; the lower peak of the "basement doublet" is degraded at each of these points. Between the two velocity analyses, this time horizon is pulled upwards; this behavior is also shown at the velocity analyses carried out at shotpoints 395 and 500, and is clearly mimicked on the blue reflector. At neither fault is there evidence of a distinct discontinuity in the basement reflector, nor is there any other indicator which fits Domoracki's fault criteria.

Seismic modeling was carried out by BGD to determine what the signature of a normal fault with displacements of 15 feet and 30 feet would be at a depth of 800 feet. All seismic parameters were taken from Domoracki's reprocessing of the original Conoco data. It can be seen from Figures 6.25 and 6.26 that even a 15 foot displacement on the basement would result in a time displacement of one-half cycle of the seismic signal. No such displacement is seen at any point on the basement reflector of SRS-12. Therefore, time displacements, and/or changes in character of the doublet representing the basement reflector, must be attributed to data acquisition and processing. If the faulting is basement controlled, then any displacement on time horizons higher in the section, i.e., on the blue and yellow horizons, would probably also be due to data acquisition and/or processing problems.

Figure 6.25 Seismic Model of "Basement" Fault with 15 Foot Offset at 800 Ft. Depth.
Geologic model, top, seismic model, bottom.

Figure 6.26 Seismic Model of "Basement" Fault with 30 Foot Offset at 800 Ft. Depth.
Geologic model, top, seismic model, bottom.

Faulting: Well MSB29

A hydrogeologic cross-section C-C' (dated 10/18/92), based on core descriptions, natural gamma ray and resistivity logs from wells MSB43A, 29A, 82A, 69A, 85A and 48A was supplied by SRS personnel (Figure 6.27) (Lewis and Aadland, 1995). A down-to-the-West fault is shown cutting through Well MSB29 from an elevation of 320 feet to depths below 0 feet elevation. The fault is apparently defined by a separation of some 30 feet on the upper surface of the "lower clay" (upper clay?) confining zone of the Ellenton formation (CBCU). On the East side of MSB29, the elevation of this surface is given as 162 feet, and on the West side of MSB29, the elevation is shown as 128 feet. Using the geological cross sections supplied by SRS personnel, the top was identified on the gamma ray and resistivity cross sections (Figures 6.3 and 6.6). On the gamma ray and resistivity log cross sections (Figures 6.3 and 6.4), the top of the Ellenton formation was defined on Wells MSB29 and MSB43A at elevations of 142 feet and 141 feet, respectively. At Well MSB82A, the elevation is 162 feet. This represents a change in elevation of about 20 feet between Wells MSB82A and MSB29A, which are 790 feet apart (i.e., a gradient of 0.25 feet per 100 feet). A change in elevation of only 20 feet in 790 feet seems not to call for a fault to explain it.

A portion of shear wave seismic Line 2 across Well MSB29A is shown on the upper panel of Figure 6.28, Horizon "H", the top of the Ellenton clay (Figure 6.18) is indicated. The lower panel shows the seismic model of the fault shown on Figure 6.27. The parameter used for the model are the same as those used for the shear wave reflection survey. The time scale on both panels is the same. Although Horizon "H" varies in time from 345 msec (TWT) at SP1045 to 335 msec (TWT) at SP1000 and shows minor displacements of 5 msec (TWT) at SP1062, now here does it have a displacement corresponding to the 45 msec (TWT) that would represent a 30 foot fault with normal throw. None of the other horizons (not indicated) show evidence of displacement.

The results of Chapman and DiStefano (1989), Cumbest, et. al., (1992), and Snipes, et. al., (1993), show that basement is involved in all faulting described at SRS. Inspection of the basement reflector shown in Figure 6.24 reveals no discontinuity in the area of Well MSB29A; the peak-through-peak double picked as representing the top of the basement shows no deflection whatsoever.

Conclusions

The evaluation of the P-wave reflection seismic data of Line 2 in the section shown in Figure 6.24 (Domoracki, 1994), the BGD S-wave seismic reflection data (Figure 6.28), and the geophysical well logs reveal that:

- No fault leading to displacements as defined by Domoracki, or which resemble the model studies, exist on the basement reflector defined by Domoracki (1994). Any departures from the average time horizon are probably due to data acquisition and/or processing problems.

A/M AREA HYDROGEOLOGIC

Figure 6.27 A&M Area, Hydrogeologic C-C' Showing

LINE 2

MSB29A

Figure 6.28 Top panel: Portion of shear wave reflection seismic Line 2; Well MSB29A, and Horizon "H", the top of the Ellenton clay are indicated.
Bottom panel: Seismic model of a normal fault with a 30 foot offset at a depth of 220 feet. Two way time offset is 45 msec.

Horizontal and vertical scales are the same of both panels.

- The fault at Well MSB29A (fault 2, Figure 6.24, Figure 6.27) is substantiated neither by the geophysical well log data, the P-wave seismic data (Figure 6.27), nor by the S-wave reflection seismic survey which shows that reflector "H" is continuous through the "fault" location.

6.4.3 TDEM and Seismic

The TDEM interpretations are shown on Figures 6.11 and 6.12, together with the seismic time horizons converted to depth using the depth-time curve of Figure 6.16.

Line 1

The 1-D geoelectric section, conductance and Occam's Inversion of the TDEM data are shown in Figure 6.11. The main feature of these interpretations is: West of TDEM sounding 1800, high conductivity values occur in a layer ranging from 15-20 feet thick at an elevation of approximately 125 feet asl. Resistivity of this layer increases to the East and maintains approximately the same resistivity until point 3375, whereupon the resistivity decreases to the same level as the layer West of point 1800. This behavior is also described by the curve of conductance (layer thickness divided by resistivity) which is high in the same zones that the resistivity is shown to be low in the upper panel. The results of using Occam's Inversion shows continuous variation in resistivity, laterally, and with depth; the color scheme shows the variation in resistivity ranging from low (orange), to high (dark blue). This display very clearly shows how the two zones of low resistivity are separated by a zone of much higher resistivity, and that the lateral resistivity variation may be gradational. The low resistivity zone at the East end of the line can also be seen between elevations of 45 and 130 feet on the resistivity log of MSB26A (Figure 6.5). Unfortunately, the other wells along Line 1 were not logged deep enough to characterize this layer.

Of particular interest on the seismic interpretation, is the elevation of time/depth horizon H which represents the top of the Ellenton. It occurs at an elevation of approximately 115 feet, the same depth as the upper surface of the conductive layer determined by the TDEM, and shows continuity across much of the line as do time/depth horizons I and J. The continuity of these three horizons are disrupted between shotpoints 720 and 800 by time/depth horizon G, which is interpreted to be the lower surface of a paleochannel. On the Occam's Inversion, this channel region is shown as an area of slightly lower resistivity which indicates that the upper levels of the channel may contain sufficient clays to be a more conductive unit than the lower levels. Horizon G is terminated upwards by horizon F which may represent the upper surface of the silty overbank deposits of the channel to the sides of the channel and the fine-grained sediments over the channel proper. Horizon B mimics the shape of horizon G and may represent a later channel development. The step between SP730 and 745 may be caused by the inability of the seismic technique to image a steep bank. It is possible that the downcutting of "B", due to a drop in sea level occurred at the same point as "G" due to persistent drainage locations. The "channel" area today is occupied by a slight

topographic low which connects to a stream to the Northwest. Horizon D coincides with the "Green Clay".

Line 2

The results of the interpretations of Line 2 are shown in Figure 6.12; the seismic survey covered a shorter section of the line than did the TDEM survey. The 1-D inversions shows that a zone of low resistivity lies between points -1000 and -2500. Its upper surface is interpreted to be at an elevation of 150 feet and rises towards the East to an elevation of 200 feet asl. No data were obtained over a 250 foot stretch centered on point -700 because of electrical interference. The zone of low resistivity increases in resistivity and changes depth between points -400 and -100. The conductance plot shows higher conductances than on Line 1 indicating that either the clays have decreased in resistivity or that they are thicker. Zones of low resistivity can be seen on Well MSB43A at an elevation of 135 feet, this zone persists downwards to an elevation of 65 feet (Figure 6.5). The correspondence of the well log of MSB43A with both the 1-D inversion and the Occam's Inversion is very good. In the region between sounding -1000 and -100, centered on Well MSB29A, there was considerable electrical interference which resulted in no useable data between soundings -860 and -550. In the rest of this region, the data may be questionable. The resistivity log of Well MSB89A shows a conductive zone between the elevations of 145 ft. and 95 ft. which is not seen on the TDEM soundings. The seismic interpretation shows continuous horizons across this region also showing that the TDEM data may be contaminated by noise. It is interesting to note that the less-resistive zone between points -100 and -200 occur at the depth indicated by the well logs of MSB29 and 82A. Unfortunately, the resistivity log of MSB82A reaches only the upper surface of the Ellenton. The section is completely resistive between points -100 and 1300 which is corroborated by Wells 69A and 85A where the Ellenton is seen to be non-conductive. The gamma ray logs show that there is significant response in this zone perhaps indicating clay content, but these clays (?) are not conductive. East of point 1300, a conductive zone is seen close to the surface on the Occam's Inversion. This zone can also be seen on the resistivity logs of MSB85A and 48A.

The seismic horizons, for the most part, are continuous across the line. Horizon H represents the top of the Ellenton and occurs at an elevation of approximately 125 feet, 10 feet higher than on Line 1. Horizon J occurs at an elevation of 25 feet and may represent the bottom of the Ellenton. Between those two horizons is horizon I which may represent the bottom of the middle sand aquifer zone of the Ellenton. Horizon B appears to be the horizon of a down-cutting sand body which merges with horizon D at shot point 1280. Horizon D correlates with the "Green Clay". Horizon F may represent the upper surface of a sandy clay layer which has been cut into by the sand layers above. Horizon A appears to be the bottom of a conductive unit which gives the conductive response seen on the East end of the Occam's Inversion. Horizon C represents the interface between an upper clay and a lower massive sand body. Both of these clays die out to the West although the reflectors continue.

Line 3

No TDEM soundings were made along seismic Line 3.

6.4.4 Stratigraphic Interpretation

To create the stratigraphic cross sections shown in Figures 6.29, 6.30 and 6.31, the following data were used:

- Wells
 - Core description: Lithologies, bed boundaries
 - Gamma Ray Logs: Depth to bed boundaries, differentiation of sand and clay
 - Resistivity Logs: Electrical conductivities, differentiation of resistive and conductive layers.
- TDEM
 - Depth to conductive layers
 - Conductivity (resistivity) of conductive layers
 - Presence or absence of conductive layers.
- Shear Wave Seismic
 - Definition of stratigraphy
 - Lateral continuity of stratigraphy

In order to adequately define the stratigraphy, four lithologies were identified based on well log resistivities and core descriptions: (1) sand; (2) clayey sand; (3) sandy clay; and (4) clay. It was felt that this distribution would adequately characterize the stratigraphic section.

Line 1

The stratigraphic interpretation of the geophysics along Line 1 is shown in Figure 6.29. The principal stratigraphic interval of interest is the Ellenton Formation which shows considerable clay development along the length of the line and is bounded by horizons H (upper) and J (lower). The intermediate horizon, I, is the lower surface of the Ellenton middle aquifer sand zone. This sand zone is seen to be developed in the center of the line on either

Figure 6.29 Line 1. Integrated Litho
TDEM, and shear wave reflection se
with

ologic and Stratigraphic Interpretation from well,
smic data. Bottom panel shows interpretation
resistivity logs.

Figure 6.30 Line 2. Integrated Lithologic and Structural Data, TDEM, and shear wave reflection seismic data with 1

ic and Stratigraphic Interpretation from well,
ic data. Bottom panel shows interpretation
istivity logs.

WEST

INTEGRATED LITHOLOGIC AND STRATIGRAPHIC INTERPRETATION - LINE 3

EAST

MSB67B

MSB68B

SHOTPOINT 2020

2040

2060

2080

2100

2120

ELEVATION(Feet)

LEGEND

CLAY	SAND
SANDY CLAY	CLAYEY SAND

50 0 50
Scale in Feet

C:\GEO\SOFT\SAVAN3\LITHBASE.PLT

Figure 6.31 Line 3. Integrated Lithologic and Stratigraphic

WEST

INTEGRATED LITHOLOGIC AND STRATIGRAPHIC INTERPRETATION - LINE 3

EAST

MSB67B

MSB68B

SHOTPOINT 2020

2040

2060

2080

2100

2120

ELEVATION(Feet)

LEGEND

CLAY	SAND
SANDY CLAY	CLAYEY SAND

50 0 50
Scale in Feet

side of well MSB21A. At shotpoint 720, at the East end of the line, there is a well-developed fluvial channel, whose bottom surface is shown by horizon G, which has cut from above the Ellenton into the units below the Ellenton. This channel is terminated upwards by horizon F. The channel is filled with material whose electrical properties are similar to those of the beds on either side and, so, appears continuous on the TDEM data. There is a slight increase in conductivity coincident with the channel which may be due to clay deposition in the channel filled up. From the geophysical data alone, it is not possible to determine the hydrological importance of this channel. Between horizons F and D at the East half of the line a thick sandy zone is developed which changes to a clayey sand at shot point 370. This clayey sand has been cut into by horizon E which probably represents another down-cutting episode. Horizon E is succeeded upwards by horizon D. The layer between horizons D and C becomes progressively finer upwards. At the East end of the line, sand dominates between horizons C and B and above horizon C; horizon B may represent another down-cutting episode. At well MSB21A, a clayey sand is succeeded upwards by a sand and sand dominates the upper part of the section at the West end of the line. Two other horizons, K and L, are shown on the line. Because there is no well information at this depth, these two horizons were not identified with geology.

Line 2

The stratigraphy derived from the geophysical interpretations of Line 2 is shown in Figure 6.29. The Ellenton at the West end of the line is a well developed clay overlying clayey sand. To the East, both of these units become progressively sandier. Above the Ellenton, on the West third of the line, a clayey sand is succeeded upwards by a thicker sandy clay. On the East end of the line, the Ellenton, here a sandy clay and clayey sand unit is succeeded upwards by a clayey sand and sandy clay unit. In the center third of the line, the lower Ellenton is a sand unit and above the Ellenton, a thick sandy unit persists to the surface. This sandy unit is well defined at this level towards the West end of the line, but is thin towards the East, and is succeeded upwards by a clay unit East of shotpoint 1450. The "G Clay" occurs at the base of the upper sand zone and is represented by horizon D. This horizon is continuous across the line but the clay probably becomes much sandier and thinner, particularly in the middle of the section. Figure 6.30 clearly shows the lateral change in lithology which can occur in a fluvial depositional system, most importantly in the Ellenton. The figure also shows the, probably permeable, system, of sands that permeates vertically through the section in the middle of the line.

Line 3

The stratigraphy of Line 3 is shown in Figure 6.31. The Ellenton is poorly developed, having clay developed only in a 15-foot layer. The rest of the Ellenton is made up of sand, clay and sand. Above the Ellenton, there is a thick sandy clay succeeded upwards by massive sand units.

Conclusion

The stratigraphic sections of Lines 1, 2 and 3 clearly show the rapid horizontal variations in lithology and depositional environments which exist in SRS. The Ellenton is thicker and perhaps has higher clay content under Line 1 than under Line 2. Under Line 1, the Ellenton is breached by a sand unit which cuts completely. On Line 2, the clay facies of the Ellenton is found only in the West half being considerably sandier. On Line 3, the stratigraphy is very similar to the Ellenton of Line 2 having a thin Ellenton. These results point to the fact that there is hydraulic communication between the aquifers above and below the Ellenton due to clay deposition and downcutting by river channels.

7. DATA FUSION APPLICATIONS

Under this contract, the Blackhawk Geophysical Division (BHD) of Coleman Research Corporation (CRC) conducted time domain electromagnetic (TDEM) and seismic surveys within the Savannah River Site (SRS) A&M area. Joint TDEM and seismic surveys were conducted along two lines as per the survey map in Figure 7.1. This survey layout was designed to exercise the new TDEM and seismic methods proposed for this contract. As a result the Data Fusion application was limited to a two dimensional earth model on each individual survey line. The survey lines were too far apart to support an extension to a three dimensional subsurface model without the incorporation of additional well or seismic data. The Data Fusion results from processing indirect data (TDEM and seismic) were consistent with two dimensional sections produced by expert geoscientists, as reported in Section 6 of this report, from the same data.

A logical extension of the work at SRS would be to continue the geophysical survey and Data Fusion application leading to a three dimensional model of the A&M area. This model could be updated rapidly as new data becomes available to assure that remediation decisions are based on a "best current" site geology.

The following sections present details of the Data Fusion processing and results for SRS and summarize the results of similar efforts at other sites. At each of the other sites, the data fusion was applied with a three dimensional earth model. In two of these cases, three dimensional contaminant distributions were estimated from well data. The contaminant distribution were superimposed on the estimated site lithology.

7.1 Fusion Processing Summary

The *fuse3d* earth model is a layered earth defined on a 3D grid. In the single line case, as for the Savannah River A&M Line 1 and Line 2, this reduces to a 2D model. In addition to layer thickness, each layer may have other properties such as resistivity and seismic velocity that also vary over the 3D grid. The thickness of the bottom layer is always assumed to be infinite. Each property (layer thickness, layer resistivity and layer seismic velocity) is modeled as a zero mean, spatially correlated random field added to a 2D polynomial trend plus known or hypothesized features such as a channel or fault. This model attempts to match the local statistical characteristics of the subsurface and to incorporate deterministic effects (large scale trends). The random field x and y spatial

correlation coefficients and spatial sigmas may be independently specified by the user for each property within each layer.

The initial earth model is derived from a site conceptual model and from geologic interpretations of well data and geophysical survey data. The grid spacing is determined by the sensor resolution, the desired resolution of the estimated lithology and computational resource constraints. The initial values of the spatial correlation coefficients are selected to quantify apriori knowledge of the subsurface degree of smoothness. These coefficients quantify the geoscientist's apriori knowledge of the spatial variability of the site geology.

The *fuse3d* program is a constrained nonlinear least squares estimator. The spatial continuity is a soft constraint which is treated as a measurement (called a pseudo measurement). The pseudo measurement residuals at each grid point quantify the degree of departure from spatial smoothness at that grid point relative to the adjacent grid points. The least squares fit produces an estimated geology that honors well picks, seismic data, electromagnetic data and spatial continuity.

7.2 Savannah River Site A&M Area

The geology of the Savannah River Site (SRS) A&M area is described in the preceding sections. The map at Figure 7.1 shows the locations of geophysical survey Lines 1 and 2. The Line 1 location was selected by BHD to provide a "best case" for demonstrating the geophysical nonintrusive survey methods and data fusion. This location is not close to any power lines, which generate TDEM interference, and provides a hard, flat surface that is ideal for the seismic survey. A 1200 meter long survey line was established with 3 wells directly on the line and two additional wells within 100 meters of the line. The seismic and TDEM survey lines are coincident. The site ground surface elevation, as per the plot in Figure 7.2, varies less than four meters along Line 1. These elevations are based on a survey by SRS.

The Line 2 location was selected by SRS and CRC to verify an existing geohydrological section. A power line conduit and a weather antenna created some TDEM interference. A survey line was established with six wells on or near the line. The TDEM line is about 1500 meters long and the seismic line about 850 meters long. The Line 2 origin, TDEM station 0., is located about 880 meters from the west end and 670 meters from the east end. Line coordinates are negative to the west of the origin and positive to the east. From Figure 7.3, the Line 2 elevation varies less than eight meters. These elevations were extracted from an SRS contour map.

The following sections present detailed Data Fusion results for Line 1 and Line 2.

Figure 7.1. Savannah River Site A&M Area Map

Figure 7.2. Line 1 Elevation

Figure 7.3. Line 2 Elevation

7.2.1 A&M Line 1

7.2.1.1 Fusion Overview

The geophysical survey data collected along Line 1 presented CRC with several "first opportunities" for the application of Data Fusion. These included the following:

- Fusion of Out-Of-Loop TDEM data - This was the first Data Fusion processing of vertical field and horizontal field out-of-loop TDEM data. New TDEM forward models from INTERPEX were integrated into the *fuse3d* program for these data types.
- Fusion of Different Types of TDEM Data - During Phase 2 of this contract all 266 TDEM soundings along Line 1 were jointly inverted by *fuse3d* to obtain an estimated lithology for Line 1.
- Fusion of TDEM and Seismic Reflection Data - the TDEM and seismic surveys both produced a response from the top of the Ellenton Clay. The results show that the high resolution, interpreted seismic data determined the horizon depths. The TDEM data then defines the lateral variation in resistivity within a layer and the resistivity contrast between layers.
- Alternate Seismic Reflection Forward Model - Along Line 1, the seismic checkshot data did not show much lateral variation in seismic velocity. Therefore, an alternate seismic reflection forward model that takes advantage of this lateral consistency was implemented. This model uses a regression equation, developed from the checkshot data, to convert one way travel time to depth.
- Modeling of a Channel - The interpreted seismic data indicated that there was a channel present that cut through the Ellenton Clay. A deterministic component was added to the *fuse3d* trend to account for the channel presence.

The processing of the Line 1 data taught several lessons. With seismic reflection data, it may not be productive to fuse preliminary versions of the interpreted seismic data. It was found that the final, definitive version of the interpreted seismic reflection data differed significantly from the preliminary version. The seismic interpretation process is very labor intensive. A coherent seismic reflection "picture" may not be available until the expert seismic interpreter has carefully integrated all of the auxiliary (well and cross section) data with information about the regional geology and the processed seismic data.

The TDEM data, which was available several months before the interpreted seismic data, provided a good quick look at the site geology and identified possible variations of the Ellenton clay composition.

Beyond this quick look the authors probably spent too much time trying to extract a more precise characterization from the TDEM data. The better characterization was possible only after the final seismic interpretation became available. In TDEM fusion applications, the availability of additional horizon depth information from wells, seismic surveys, geoprobe and other sources will greatly improve the accuracy of the characterization.

7.2.1.2 Top Of Ellenton Well Picks

Savannah River Site (SRS) provided Coleman Research (CRC) with a recent (February, 1995) "M Area Geophysical Study Control Cross Section" which shows four of the Line 1 wells, as listed in Table 7.1. Three of these wells (MSB76, MSB21, and MSB26) intersect the top of Ellenton. Wells MSB21 and MSB26 also intersect the bottom of Ellenton. Table 7.1 gives the distance from the Line 1 origin, the ground elevation, the well depth and the depth to the top of the Ellenton for each of three wells as extracted from the cross section. The first Table 7.1 Ellenton top entries in italics were provided by SRS at an earlier date (October, 1994). MSB-20 is not deep enough to intersect the top of Ellenton. The Ellenton depth value provided by SRS is an interpolated value. The second Table 7.1 entry for depth to top of Ellenton in italics was derived by BHD from the analysis of well logs and of soil samples to support the seismic reflection data interpretation. All of the plots in the following sections show the well picks received from SRS in October 1994 in comparisons with Data Fusion estimates for depth to the top of Ellenton. The variability in the Table 7.1 well picks gives some measure of the variability in well lithology picks among different experts who may not all be looking at exactly the same supporting data. The well picks were not directly used in the fusion of the Savannah River A&M data described in the following paragraphs.

TABLE 7.1. Line 1 Well Picks From SRS

Well Id	Line Dis. meters	Elev. meters	Depth meters	Ellenton Top meters
MSB-76	18.5	106.9	74.1	73.2 72.5 73.2
MSB-20	273	107.3	61.9	-- 73.2 68.6
MSB-21	495.	107.5	103.	75.6 73.2 71.6
MSB-26	1196.	109.5	106.7	69.8 70.7 70.1

7.2.1.3 Data Description

Table 7.2 lists the Line 1 geophysical survey data. This includes three types of time domain electromagnetic (TDEM) data and seismic shear wave reflection data. All of these data types were collected along the entire 1200 meter line. During the TDEM quality assurance processing, it was determined that a buried culvert and power lines had contaminated a few soundings. These soundings were omitted from the analyses described in this report. The Table 7.2 entries are the number of TDEM soundings remaining after the questionable data was omitted. Additional seismic data was collected but not analyzed during this contract.

TABLE 7.2. Line 1 Survey Data

Survey Data Type	No. Soundings
TDEM Vertical Field Central Loop	26
TDEM Vertical Field Out-of-loop	124
TDEM Horizontal Field Out-of-loop	116
Seismic Reflection Shear Wave	780

Our initial Data Fusion processing showed that the TDEM horizontal field data contained higher levels of measurement noise than the vertical field data. When the horizontal field data was fused with the vertical field data, it was determined that the horizontal field data contributed very little to the resolution

of the Ellenton depth and conductance. For this reason and to reduce the time consumed by Data Fusion computer runs, only the vertical field out-of-loop data was used in producing the Data Fusion results described in this report.

7.2.1.4 Seismic Reflection Forward Model

The *fuse3d* seismic reflection forward model currently assumes that the measurement is a two way travel time in the vertical direction to a given seismic reflector at a given shotpoint. In computing the expected travel time, it is assumed that within each earth layer the seismic velocity is constant in the vertical direction but may vary in the horizontal direction. In addition other measurement types such as seismic refraction are required for resolution of the depth-velocity ambiguity.

At the SRS A&M site very high quality seismic reflection data was obtained. The seismic checkshot data from MSB21 and MSB26, that are 561 meters apart, indicated that there is not much lateral variation in seismic velocity. The seismic refraction data quality was not as good and Line 1 was not long enough to get refractions off of the Ellenton top and bottom. For these reasons an alternate seismic reflection measurement forward model was used. A second order polynomial, expressing depth as a function of one way travel time was fit to the checkshot data. This polynomial was then used to convert the interpreted seismic reflection measurements to depth values. The partial derivatives of the conversion equation were used to convert the travel time measurement standard deviations to depth standard deviations. Figure 7.4 includes a plot of the checkshot data and lists the regression coefficient values for the second order polynomial fit.

Checkshot Regression Analysis

Fitting 1 way travel time as a function of depth, the travel time in milliseconds is given by:

$$t = B0 + B1*x + B2*x^2$$

where $x = ((d/.3048) - 175)/350$ and d = depth in meters, $B0=132.58$, $B1=240.91$, $B2=-68.40$

Using this regression equation, the velocity at depth d can be expressed as

$$v1(d) = (1000*.3048*350) / (B1 + 2B2*x)$$

where $v1$ is in meters per second, and x is defined above.

Fitting depth as a function of 1 way travel time, the depth in meters is given by:

$$d = (B0 + B1*y + B2*y^2) * .3048$$

where $y = (t-110)/220$, t is the 1 way travel time in milliseconds, and d is the depth in meters. $B0=142.67$, $B1=308.61$, $B2=80.56$

Using this second regression equation, the velocity at a given travel time is given by:

$$v2(t) = (B1 + B2*y) * 1000 * .3048 / 220$$

where $v2$ is in meters per second and y is defined above.

The upper left plot shows the checkshot 1 way travel time in seconds as a function of depth for wells MSB21TA and MSB26B. The lower left plot shows the seismic refraction velocity vs depth as computed using the above equations for $v1(d)$ and $v2(d)$.

SAVANNAH RIVER A&M LINE 1 SEISMIC REFLECTION CHECKSHOT

Coleman Research Computational Engineering Division 2/9/95

Figure 7.4. Line 1 Seismic Checkshot Results

7.2.1.5 Line 1 Fusion Results

The Data Fusion results for Line 1 are summarized in Figure 7.5. The top plot is from the fusion of TDEM and seismic reflection data. The estimated parameters were the layer thicknesses and the within layer resistivities. The resistivity values in ohm-meters are shaded on the black and white illustration to show the lateral variability within each layer and the resistivity contrast between layers. The x coordinate is distance from the line origin (west end) and the y coordinate is depth from the surface. There is a discontinuity in the Ellenton at about 1000 meters from the line origin which has been interpreted as a channel. The resistivity contrast between the Ellenton and the adjacent, higher resistivity, layers disappears between 750 and 950 meters from the line origin. This has been interpreted as a change in the soil composition of the Ellenton which may alter its aquitard properties.

The second and third plots in Figure 7.5 show the Ellenton thickness and resistivity estimates from TDEM data alone. The Ellenton depth is in reasonable agreement with the above combined estimate but the channel at 1000 meters from the line origin is not detected. The Ellenton region of increased resistivity extends from 600 meters through the end of the line. The Ellenton thickness is not well determined from the TDEM data alone. If the depth sigma bands were shown on these plots then the estimates of depth to top of Ellenton and depth to bottom of Ellenton would be indistinguishable. Thus it is necessary to fuse the TDEM data with another data type, such as seismic reflection, in order to determine the Ellenton thickness.

Figure 7.5. Line 1 Fusion Estimates Of Ellenton Depths And Resistivities

7.2.1.6 Comparison Of Fusion Results With TDEM Ridge Regression

The *fuse3d* program uses TDEM forward model code provided by INTERPEX, Ltd. of Golden Colorado. INTERPEX uses this same forward model code in the TEMIX program. TEMIX is a commercial program used by many geophysists to invert single sounding TDEM data to obtain the model parameters for a 1D layered earth at the sounding location. A single TDEM sounding typically consists of voltage measurements at 13 to 18 delay times (time gates). The delay time is the time in milliseconds since the leading edge of the transmit pulse. The TEMIX 1D inversion is a nonlinear least squares estimate which uses ridge regression to solve the linearized subproblem at each iteration. (Inman 1975) TEMIX produces a measure of the uncertainty in the 1D solution by determining individually how much each estimated parameter can be perturbed without increasing the measurement residual RMS beyond a user specified threshold. This perturbation gives an "equivalence band" about each estimated parameter.

Fuse3d is a nonlinear least squares estimator that jointly inverts all of the TDEM data in a given set of soundings under user controlled spatial continuity constraints on the estimated parameters. In the Line 1 examples presented here, *fuse3d* is inverting 124 out-of-loop vertical field soundings or a total of about 2000 individual measurements in a single computer run.

7.2.1.6.1 Comparison Using Line 1 Out-Of-Loop Vertical Field Data

Figure 7.6 compares the *fuse3d* and TEMIX results over three sections of A&M line 1 using exactly the same TDEM measurements. These plots show that the TEMIX and *fuse3d* produce compatible results given the same input data. The TEMIX results are displayed as error bars. The interior dot is the "best" estimated value and the extreme dots show the equivalence bounds computed within TEMIX. In some cases there may be more than one out-of-loop soundings recorded and plotted at the same location causing the error bars to appear to have multiple interior dots.

It has been well documented that the 1D inversion problem that is solved within TEMIX is very poorly conditioned. (Constable 1987) In a poorly conditioned inversion small changes in the input data produce large changes in the solution. Ridge regression is one method used to improve the conditioning and thus obtain a more stable solution (Hoerl 1970). An alternate method of stabilizing the solution is to obtain more data. The Data Fusion program *fuse3d* does this by using spatial continuity constraints as pseudo data and by jointly inverting all of the soundings. The constraint coefficients (spatial τ and spatial sigma) are

selected by the expert geoscientist to be consistent with his or her apriori knowledge of the site geology.

A common mistake among geoscientists has been to interpret the 1D results as the "TDEM data" and to attribute differences between the 1D results and Data Fusion results to the Data Fusion's refusal to honor the "TDEM data". Both the 1D results and Data Fusion results are derived from the same TDEM measurement values via different inversion procedures. Thus, where the 1D inversions show some departure from the Data Fusion results, the differences are due to differences in the inversion procedures where both inversion procedures are attempting to honor the same TDEM measurements.

SAVANNAH RIVER A&M LINE 1 COMPARISON OF FUSE3D AND 1D INVERSIONS
 Coleman Research Computational Engineering Division 11/7/94

Figure 7.6. Comparison Of Fusion And Ridge Regression Estimates From TDEM Data

7.2.1.6.2 Comparison Using Simulated Central Loop TDEM Data

In the first simulated example a three layer earth model with constant resistivities and layer thicknesses with a thin clay layer at 70 meters depth was used. There were 26 simulated central loop soundings distributed along a 1000 meter line. These simulated soundings were at the same locations and had the same time gates as the actual Line 1 central loop soundings. Zero mean normally distributed random noise was added to the measurements. The noise standard deviation was five percent of the noise free measurement value with the minimum allowable standard deviation being one nanovolt per ampere-meter squared. The plots in Figure 7.7 compare the Data Fusion and ridge regression results. The Data Fusion results very closely follow the simulated lithology. The ridge regression results have much greater variability and one peak that might be misinterpreted as a discontinuity in the clay.

Next the simulated lithology was modified to include a step change in the clay layer resistivity from 6 ohm-meters to 15 ohm-meters at a distance of 600 meters from the line origin. The results plotted in Figure 7.8 show that the ridge regression shows a change in resistivity at 600 meters and a change in estimated clay depth from 70 meters to 140 meters. The Data Fusion correctly detects only a change in clay resistivity at 600 meters.

These plots show the fuse3d and 1D inversion results for a line of simulated central loop, vertical field time domain electromagnetic soundings. The earth model used to generate the simulated data includes 3 layers with constant layer width and resistivity.

Top Layer	width	70 meters
	res.	800 ohm-meter
Clay Layer	width	5 meters
	res.	6 ohm-meter
Bottom Layer	width	infinite
	res.	1000 ohm-meter

Current practice is to independently invert each sounding. Data fusion, as implemented in fuse3d, jointly inverts all of the soundings under user specified spatial continuity constraints on layer width and resistivity.

Figure 7.7. Fuse3d And Ridge Regression Comparison With Simulated Flat Clay Layer

Comparison Of Data Fusion And "Current Practice" With The Addition Of A Step Increase In Clay Layer Resistivity From 6 To 15 Ohm-meters At 600 Meters From The Line Origin

These plots show the fuse3d and 1D inversion results for a line of central loop time domain electromagnetic soundings. The earth model used to generate the simulated data includes three layers. The layer widths and the first and third layer resistivities are constant.

The second layer, which is the simulated clay layer, has a step increase in resistivity from 6 to 15 ohm-meters at 600 meters from the line origin. The earth model parameters are as follows:

Top Layer	width	70 meters
	res.	800 ohm-meter
Clay Layer	width	5 meters
	res.	6 ohm-meter
		from 0 to 600 m.
		15 ohm-meter
		from 600 to 1200 m.
Bottom Layer	width	infinite
	res.	1000 ohm-meter

Figure 7.8. Fuse3d And Ridge Regression Comparison With Simulated Clay Discontinuity

7.2.1.7 Effect Of User Input Spatial Correlation Coefficients

Fuse3d requires the user to quantify the amount of spatial continuity present by specifying a spatial correlation distance, τ , for each property to be estimated. Our experience with this type of parameter in other applications leads us to expect that Data Fusion results will not be highly sensitive to the values of τ provided by a trained geoscientist who understands the site geology. This conjecture was verified by processing the line 1 TDEM data with τ varied from 100 to 500 meters (10% to 50% of the length of line 1) for all of the parameters except the Ellenton resistivity. The spatial τ of the Ellenton resistivity was held constant at 100 meters. The estimated parameters are the four within layer resistivities and the three layer widths. The bottom layer is assumed to have infinite width. The resultant estimates of depth to the top of the Ellenton Clay are plotted in Figure 7.9. As would be expected, the smoothness of the estimate increases with increasing τ but the shape of the plot does not change very much. Increasing τ by a factor of 5, from 100 to 500 meters, increased the RMS of the standardized measurement residuals by only ten percent.

The recommended fusion practice is to select initial values for spatial τ from the geoscientist's apriori knowledge of the site geology. All of the initial problems in fusing the available sensor data will involve data discrepancies and selection of an appropriate layered earth model. After a converged estimate with acceptable residual statistics has been obtained, then test the sensitivity of the estimate to the values of spatial τ .

SAVANNAH RIVER A&M LINE 1 EFFECT OF SPATIAL CORRELATION

Computational Engineering Division 12/5/94

Figure 7.9. Effect Of User Selection Of Spatial Correlation Coefficients

7.2.1.8 Detection Of Geologic Changes

In discussions about Data Fusion, earth scientists often ask if Data Fusion is capable of detecting geologic changes that occur over short horizontal distances. The relevant issues are:

1. *Sensor resolution* - any geological interpretation is going to be limited by the lateral resolution of the sensors,
2. *Data fusion usage* - given the lateral resolution of the sensors used in a given survey, how would Data Fusion be applied to detect and model geologic changes.

The ability of TDEM data to resolve geologic changes is limited by the transmitter loop size. In a homogeneous earth the TDEM current pattern spreads with increasing depth outward like a smoke ring from the transmitter loop. Thus the loop size gives the "best possible" resolution at the surface and the resolution degrades with increasing depth.

The resolution of seismic reflection data is determined by the size of the Fresnel zone. This size is a function of signal frequency, depth and seismic velocity. Resolution decreases with increasing depth. A 3D seismic survey yields much better resolution than a 2D survey. It is estimated that the A&M line 1 interpreted channel boundaries are accurate to within ten to twenty meters. This geologic feature has a horizontal extent of about 90 meters and a depth change of about 30 meters.

If an unmodeled geological change is present and it's spatial extent is greater than the resolution and measurement noise limits of the geophysical sensors then the presence of the change will be accompanied by the occurrence of statistically large measurement residuals, pseudo measurement residuals or both in the summary reports and plots produced by *fuse3d*. Additional deterministic inputs to the earth model are then derived to account for the suspected geologic change. For example, in the case of the A&M line 1 a channel through the Ellenton clay at 1000 meters from the line origin as per the results displayed in Figure 7.5 was specified.

The plots in Figure 7.10 show examples of measurement residuals from the A&M line 1 data. These are plots of the *standardized residuals*, that is the differences between the computed and actual measurement values divided by the measurement standard deviation. If the least squares estimator has converged and all of the user's input assumptions about the measurement error statistics and the spatial correlation statistics are valid, then the standardized residuals should approximately follow a normal distribution with mean zero and variance equal to

one.

About 1000 meters from the line origin there appears to be a channel that cuts through the Ellenton clay. In the first case (top row) a deterministic feature at the channel location was not inserted. The seismic residuals clearly show an anomaly at about 1000 meters. The TDEM residuals do not show this feature. In the second case (bottom row) a channel in the Ellenton in the neighborhood of 1000 meters was added. This changed the residual plot slightly but the anomaly is still present. The anomaly was detected, the location was as previously defined in the seismic interpretation, and with more work it was possible to define the appropriate combination of deterministic features that better fit the data in the neighborhood of the anomaly. When we have better modeled the geologic feature, anomaly in the standardized residual plots will decrease in magnitude. If the model is highly successful, the residual anomaly will disappear.

SAVANNAH RIVER A&M LINE 1 FUSION OF SEISMIC REFLECTION & TDEM MEASUREMENT RESIDUALS

Coleman Research Computational Engineering Division 2/24/95

Figure 7.10. Line 1 Residual Plots Showing Effect Of Channel

7.2.2 A&M Line 2

7.2.2.1 Fusion Overview

The TDEM data quality along Line 2 was not as good as in the Line 1 survey. In addition, the seismic survey covered less than half of Line 2. As a result there is less than 200 meters of joint seismic and possibly good TDEM coverage. The interpreted seismic data provides excellent information on the horizon depths (top of Ellenton and bottom of Ellenton). The resistivity estimates from the TDEM are less accurate than those for Line 1.

7.2.2.2 Top Of Ellenton Well Picks

SRS provided CRC with a geologic cross section titled "Integrated Demonstration Site Hydrogeologic Cross Section" which includes four wells. One objective of the Line 2 survey was to verify this cross section. The well picks of the top of Ellenton Clay from this cross section are listed in Table 7.3. The Line 2 seismic and well log interpretation by BHD revised these top of Ellenton picks by less than 1 foot; therefore, only one set of well picks are presented for Line 2.

TABLE 7.3. Line 2 Well Picks From SRS

Well Id	Line Dis. meters	Elev. meters	Depth meters	Ellenton Top meters
MSB-82A	33.5	113.5	77.1	64.
MSB-69	190.	116.6	93.4	67.
MSB-85TA	403.	115.5	90.7	67.
MSB-48C	685.	110.	57.	60.

7.2.2.3 Data Description

The TDEM survey extends over the entire 1500 meter line. The seismic survey covers from -280 meters to 550 meters. The number of TDEM soundings and seismic reflection shot points are given in Table 7.4.

TABLE 7.4. Line 2 Survey Data

Survey Data Type	No. Soundings
TDEM Vertical Field Central Loop	14
TDEM Vertical Field Out-of-loop	46
Seismic Reflection Shear Wave	545

A buried power line and a weather station antenna induced contamination in the TDEM data from -300 meters to -50 meters. In addition there was no TDEM response to any geologic features at the approximate depth of the Ellenton clay between +50 meters and the eastern end of Line 2. These conditions limited the simultaneous Ellenton visibility for TDEM and seismic data to the 100 meter Line 2 segment between -50 meters and +50 meters.

Figure 7.11 shows the TDEM coverage (top panel) and measurement values (bottom panel). The TDEM coverage plot shows the time gates at each sounding location that provided a measurable receiver response. The plot of measurement values shows the degree of spatial continuity at each time gate. The sawtooth effect is produced when two adjacent out-of-loop soundings with the same "plot point" give slightly different measurement values at a time gate. The sounding location or plot point for an out-of-loop sounding is defined as the average of the receiver coil position and the position of the center of the transmitter loop. Note that, on the measurement value plot, the TDEM measurement values are shifted upward in the vicinity of the buried power line and weather antenna. The effect of these interference sources may extend beyond those measurements between -180 meters and -50 meters that have been identified as being corrupted.

Figure 7.11. Time Domain Data Coverage And Data Values

7.2.2.4 Seismic Reflection Forward Model

Line 2 seismic shear wave checkshot data was obtained from four wells, MSB82A, MSB96B, MSB85TA and MSB85B. Wells MSB82A and MSB85TA are 372 meters apart. The checkshot data, plotted in Figure 7.12, is very consistent among these wells. This implies that, as with Line 1, there is not very much lateral variation in seismic velocity. A second order polynomial, expressing one way travel time as a function of depth, was fit to the checkshot data. The polynomial coefficients were used to convert the seismic reflection measurements to depth measurements and the travel time measurement sigmas to depth measurement sigmas.

From a 2d order polynomial fit to the above checkshot data, the depth in meters as a function of the one way travel time in milliseconds is given by the following equation:

$$d = .3048 * (B0 + B1 * x + B2 * x^2)$$

where $x = (t - 110)/220$, t is the one way travel time in milliseconds, $B0 = 135.29$, $B1 = 313.30$, $B2 = 140.54$

SAVANNAH RIVER A&M LINE 2 SEISMIC CHECKSHOT PLOTS

Coleman Research Computational Engineering Division 2/15/95

Figure 7.12. Line 2 Seismic Checkshot Results

7.2.2.5 Line 2 Fusion Results

Our experience in processing the Line 1 seismic and TDEM data showed that the seismic data will almost completely determine the depths to geologic horizons that provide a reflecting boundary. The TDEM data will then provide estimates of the lateral variation of resistivity within each layer and of the resistivity contrast between layers. Figure 7.13 shows the results of fusing the seismic reflection data and that TDEM data which is coincident with the seismic data. The earth model contains four layers, a near surface conductive layer, an upper sandy clay or clayey sand layer, the Ellenton clay, and a lower sandy clay or clayey sand layer. The lateral variation of estimated resistivity within each layer is color coded (intensity shaded in the black and white illustration). Thus the shaded stripes within each layer represent lateral changes in resistivity, not geological features. In Figure 7.13 it is noted where the resistivity estimate is unreliable due to electromagnetic interference and to lack of electromagnetic response from the Ellenton. Within the approximately 100 meter span that may have useful TDEM data, the Ellenton resistivity is estimated to be less than two hundred ohm-meters. The adjacent layers have estimated resistivities greater than 800 ohm-meters.

Figure 7.13. Fusion Results Seismic And TDEM Limited To Seismic Line

Since the TDEM survey along Line 2 covered more than twice the area of the seismic survey, the authors attempted to use the TDEM data to extend the fusion results beyond the western (left hand) limits of the seismic survey. Figure 7.14 shows the fusion results using only TDEM data over a segment of Line 2 where there was an electromagnetic response from the Ellenton. The $\pm$ one sigma bands of depth to top of Ellenton and depth to bottom of Ellenton completely overlay the estimate of the Ellenton thickness. The one sigma bands are smaller at the TDEM sounding locations and become larger as distance to the nearest soundings increases. The interference from the buried power line and weather antenna may have biased the Ellenton depth estimate toward the surface. These fusion results did not satisfy the *fuse3d* statistical criteria for satisfactory least squares convergence indicating that either the earth model is incomplete or that some of the data is corrupted. In this case it has already been established that some of the TDEM data is corrupted.

These fusion results involving the Line 2 TDEM data are presented to demonstrate what happens when Data Fusion is used to process bad or questionable data. Figures 7.13 through 7.15 do not present reasonable estimates of the Line 2 geology.

Figure 7.14. Fusion Results Time Domain Only

Next the authors attempted to expand the Line 2 results to the west of the seismic line by fusing the above described extended TDEM data with the seismic data. These results are shown in Figure 7.15 . These fusion results also did not satisfy the *fuse3d* convergence criteria. To the west (left) of the seismic line, the top of Ellenton and bottom of Ellenton estimates make an abrupt move toward the surface. This could be attributed to corrupt TDEM data. Alternately the TDEM may be sensing a resistivity contrast due to a different geologic feature than that at which the seismic reflection occurs.

Figure 7.15. Fusion Results Seismic And TDEM Extended Beyond Seismic Line

7.3 Fernald South Field Characterization

In cooperation with Fernald Environmental Restoration Management Company (FERMCO), CRC applied Data Fusion to map the surface of an old stream bed that had been filled in with waste materials including flyash. The following types of location tagged data were used in this analysis:

- Lithological picks from well core samples,
- Uranium contaminant data from borings,
- Lithological picks from soil borings,
- Uranium contaminant data from monitoring wells,
- Historical ground water elevations from monitoring wells,
- EM31 and EM34 frequency domain electromagnetic data,
- Lithological picks from cone penetrometer data,
- Real time radiological data.

In addition FERMCO provided historical air photos , historical topological maps, current topological maps, plus building, road, trench and well layouts in InterGraph design format.

The Fernald site, just north of Cincinnati, Ohio, has a very complicated and heterogeneous geology. The area of concern is composed of fluvial deposits on top of a glacial moraine. A three layer earth model was determined to be sufficient for this application. These layers were as follows:

- Fill put in place after 1952,
- Clay Layer,
- Sand and Gravel - part of the Greater Miami Aquifer.

The following sections present selected results from the South Field characterization. As a direct result of this success, FERMCO has awarded CRC additional tasking to analyze similar data types to identify locations, within an authorized waste storage area, where the geology offers natural, long term protection against contaminant leakage.

7.3.1 Three Dimensional Site Lithology

From the Data Fusion solution, Figure 7.16 is a chair cut showing the depth to bottom of the fill and the one sigma uncertainty (lighter grey strip) of this depth. Each square on the fill bottom surface represents an individual pick of the fill bottom. The squares above the fill bottom are locations where no "classified fill" was found.

7.3.2 Contaminant Distribution Superimposed On Lithology

The distribution of uranium in the fill is given in Figure 7.17 as contours of isosurfaces of constant uranium concentration. Note that in this particular chair cut, the highest uranium concentration seems to be resting on the estimated bottom of fill/top of clay. Having the top and bottom elevation contours of the fill, it is now possible to estimate the volume of fill that needs to be removed under different proposed remediation scenarios.

7.3.3 Greater Miami Aquifer Groundwater Elevation Superimposed On Lithology

Figure 7.18 is a three dimensional picture of the elevation of the lower aquifer surface as computed from eight ground water elevation measurements taken on October 13 and 14 of 1993. Note that the Greater Miami Aquifer is separated from the contaminated fill by a clay layer that is 10 to 20 feet thick. Accurate knowledge of the clay depth and depth uncertainty is required to guide any remediation procedures.

Figure 7.16. Fernald: Chair Cut Of Fill Showing Lithology

Figure 7.17. Fernald: Distribution Of Uranium In Fill

Figure 7.18. Fernald: Lower Ground Water Elevations WRT Bottom Of Clay

7.4 Marshalltown Iowa Manufactured Gas Plant Bedrock Characterization

The IES utilities site in Marshalltown Iowa has been the location for industrial operations since the 1870's. These included electrical generation and the manufacture of gas. Gas manufacture ceased in the early 1950's when natural gas became available. The site experienced the release of several types of contaminants associated with gas production.

The Marshalltown IES site was selected for a demonstration of the Expedited Site Characterization Method (ESC) by Ames Laboratory. As part of this effort, Ames selected BHD to perform seismic surveys to map the bedrock topology underlying the site. One of the anticipated site remediation procedures includes the removal of contaminants that are resting on the bedrock. The estimate of the bedrock topology would be used to determine the volume of soil to be removed and to direct the removal effort.

Prior to the ESC startup, depth to bedrock picks were available from twelve monitoring wells. During the ESC, in addition to the seismic data, Geoprobe measurements of the lithological features above the bedrock were taken.

The following figures demonstrate the performance of Data Fusion for the three dimensional characterization of bedrock from multiple types of direct (well, Geoprobe) and indirect (seismic refraction) measurements. Figure 7.19 shows the bedrock topology as could be determined prior to the ESC, from a minimum tension gridding of the 12 well picks. Figure 7.20 shows the much more detailed bedrock topology as obtained from Data Fusion of the well picks, seismic refraction measurements, and Geoprobe picks for the layers above the bedrock.

For an east-west cross section, Figure 7.21 compares bedrock topologies from the current practice of interpreting each data type separately (top two panels) with the Data Fusion method of jointly interpreting the same data types (third panel). Note that the fusion result contains a depression, not apparent from the individual interpretations, where contaminants may be collected. The bottom panel shows the Data Fusion results when the Geoprobe data is fused along with the well and seismic data. These pannels demonstrate the value added by data fusion. Using existing wells and a relatively inexpensive seismic survey, The Data Fusion produces a much more detailed bedrock map than either data set by itself. In addition the Data Fusion quantifies the uncertainty of this map. The more expensive (with respect to the seismic survey) Geoprobe survey verifies the seismic and well Data Fusion results but only contributes a small amount of additional detail to the bedrock map.

The Data Fusion results were *cross validated* by using data from only three of the twelve monitoring wells. Other sensor inputs included the seismic refraction data and the Geoprobe pick data. The resultant bedrock map predicted depth to bedrock, within the statistical confidence bounds, at all nine of the "left out" wells.

Former Manufactured Gas Plant: Bedrock Map
From Minimum Tension of 12 Well Bedrock Picks
No Data Fusion – Minimum Detail

Figure 7.19. Bedrock Surface From Minimum Tension Gridding of 12 Well Picks

Former Manufactured Gas Plant Bedrock Map From
Fusion of 12 Well Bedrock Picks, 1013 Seismic Refraction
Measurements And 196 Geoprobe Picks of Clay Above Bedrock
Full Data Fusion - Maximum Detail

Figure 7.20. Detailed Bedrock Estimate From Fusion Of All Sensor Data

Figure 7.21. Detailed Bedrock Cross Section By Single Data Type And By Fusion

7.5 Hanford 200 West Area Characterization

One of the first Data Fusion applications was at the Hanford 200 West Area in eastern Washington state. The objective was to map a thin caliche layer 30 to 40 meters below the surface in the vicinity of the Z1 and Z2 tile fields. Scattered well data with caliche picks was provided by the Hanford site. In order to obtain a more detailed characterization, BHD conducted a geophysical survey to obtain densely spaced non-intrusive data. The survey was conducted in two phases. During phase 1, TDEM, DC resistivity and seismic methods were tested. The TDEM methods failed because of the large amounts of cultural interference from buried construction debris. The DC resistivity soundings yielded some information on the very shallow subsurface, but nothing that was judged to be helpful in mapping the caliche. Attempts by other contractors to conduct seismic surveys at Hanford had not been successful. Possibly due to the winter conditions during the survey phase 1 (frozen ground surface), BHD obtained successful results for both the seismic reflection and refraction tests. Therefore, the phase 2 survey design consisted of four seismic reflection and refraction lines which are shown in Figure 7.22. The availability of Data Fusion allowed BHD to use a new "roll along" method to collect very high density seismic refraction data along lines 1 and 2. The "current practice" GRM seismic layout was used to collect lower density seismic refraction data along lines 3 and 4. Due to the site specific safety practices, it was not possible to obtain any seismic checkshot data.

Figure 7.22 shows the Data Fusion results using seismic measurements and well picks for the Hanford 200 West Area. The top of the caliche layer, a possible barrier to contaminant movement, is shown at the bottom of the chair cut. Top of caliche picks were obtained from 25 wells. The individual interpretations for seismic reflection and seismic refraction data, without checkshot information, provide almost no information on the caliche topology. Using reflection data alone, inadequate seismic velocity information was available to interpret travel times as depths to the caliche. The caliche did not support a refraction so no depth information was available from the refraction measurements. Data Fusion, by jointly processing all of the data uses seismic velocity information from the refraction measurements to interpret the reflection travel times as depths. This demonstrates the value added by the ability of Data Fusion to produce a complete picture from multiple sensors.

Figure 7.23 shows a plutonium plume within a portion of the characterization area. Note that the plume is above the estimated caliche depth with a few isolated points right on the caliche. This figure provides an independent verification of the caliche topology estimated by the Data Fusion.

Hanford 200 West Area: Fusion Results

Figure 7.22. Hanford 200 West Site Caliche Topology

Hanford 200 West Area Tile Field: Plutonium Plume and Lithology

Figure 7.23. Hanford 200 West Z1&Z2 Tile Field Plutonium Plume

7.6 Groundwater Model Calibration at Pantex Using Data Fusion Modeling

7.6.1 Introduction The U.S. Department of Energy (DOE) Pantex plant is located near Amarillo in the panhandle region of Texas. This Southern High Plains region is a well-defined plateau that slopes gently toward the southeast. Surface soils in this region are primarily clays and clay loams. The climate is arid with an average 18 to 19 inches of precipitation per year. The regional land surface is characterized by numerous playa basins that vary in diameter from a few hundred feet to one mile. Water accumulates in these depressions following heavy rains. Most of the playa water evaporates but a portion percolates downward to provide groundwater recharge. The playas may be related to geologic structural depressions caused by faulting in the area.

Regional hydrogeology is characterized by two principal aquifers: an upper zone (commonly referred to as the perched aquifer) at a depth of 200 to 300 feet below the Pantex site, and the deeper, thicker Ogallala aquifer whose base ranges from 650 feet to 900 feet below the site. A layer of fine-grained sediments, termed the Fine-Grained Zone (FGZ) generally occurs between the perched and Ogallala aquifers and acts as an aquitard (flow restriction zone) between the two. The Ogallala aquifer is the main water supply in the Southern High Plains, and pumping from this exceeds recharge.

The Pantex plant has operated as one of the Federal government's key conventional and nuclear weapons facilities since the 1940's. The plant has been used for fabrication of chemical high explosives and components of nuclear weapons. Contaminants from various waste products and processes at the plant have percolated into the perched groundwater zone beneath the plant and have migrated laterally with flowing groundwater in that zone. The perched aquifer is not used for drinking water or agricultural purposes. Also, there is no evidence of contamination in the Ogallala aquifer. In recent years, the DOE has expended considerable effort to characterize the nature and extent of groundwater contamination associated with the site. That effort is still on-going with the ultimate aim of determining and implementing appropriate remedial measures. The goal of the study described here was to use Data Fusion modeling to calibrate a groundwater model near Zone 12 of Pantex, primarily to define the potential pathways to the Ogallala aquifer.

7.6.2 Pantex Model Calibration Figure 7.24, from the U.S. Army Corps of Engineers report (1995), shows the Pantex region and the location of wells. The location of Zone 12 of the plant is also shown in the figure. Zone 12 is of special interest because it is an area where plant activities generated considerable wastes that may have contributed to explosive-related contaminants (such as RDX) in the perched aquifer. Monitoring wells near the plant which reach the perched water but may or may not penetrate the FGZ are also shown. These

wells monitor the contamination in the perched water. Principal playa areas on the site are also shown. For reference purposes, the outlines of Zone 12 and the playas are shown in Figures 7.25 through 7.27 that display fusion results.

We relied heavily on data and information contained in the U.S. Army Corps of Engineers report (1995) and the Argonne National Laboratory report (1994). In addition, approximately 30,000 pages of data from a variety of sources were reviewed for relevant data. This included well logs, geologic and hydrogeologic framework studies, seismic interpretations, water level measurements, slug tests, and geotechnical and analytical chemistry laboratory analyses. These were reviewed to develop a hydrological modeling framework for inverse modeling using Data Fusion.

Because the saturated thickness of the perched zone is relatively thin (varying from 0 to 60 feet), it is necessary to map the top of the FGZ in order to compute the flow within the aquifer. Data available for mapping the FGZ consisted of the lithologic and geophysical logs from about 50 wells around the Pantex area. These data tend to be clustered in the south-eastern and northern sectors of the modeled area (Figure 7.25). The elevation at the top of the FGZ was estimated using the Data Fusion software where the spatial continuity model assumed a statistical variation of 30 feet with a horizontal correlation distance of 2000 feet. Because the quality of the well log data varied significantly, the accuracy of the depth measurement was assumed to vary from 3 to 20 feet. Figure 7.25 shows the resulting top of FGZ estimate. Large variations in the surface elevation are apparent as are the large areas with no wells that have a corresponding large uncertainty in the estimate (as great as 45 feet). A channel in the top of the FGZ running generally southeast is apparent in the vicinity of Zone 12.

Because there are no natural hydrological boundaries around the Pantex area, it was necessary to extend the region of the hydrological model far outside Pantex. This greatly reduces the sensitivity of the flow to the boundary conditions at the perimeter of the model. The extent of the model was approximately 100,000 by 100,000 feet on a variably spaced 32-by-34 grid. There are no wells intersecting the top of the FGZ outside Pantex. Therefore, significant extrapolation was necessary to extend the top of the FGZ shown in Figure 7.25 to the larger area. This was done the following two different ways to determine sensitivity to the extrapolation: 1) assuming that the average FGZ slopes 2 feet per thousand downward to the southwest and 2) assuming that it is flat. The resulting flow estimate within the Pantex area was relatively insensitive to the assumed slope outside Pantex. Results presented here used the flat assumption.

The data available for hydrological modeling consists of 37 perched water level

measurements and several hydraulic conductivity (slug) tests. Because the primary goal was to model the flow of perched water, simplifying assumptions which have little effect on the perched water flow were used. A two layer model was assumed in which the perched water was treated as an unconfined aquifer with the FGZ acting as a low-permeability layer separating the FGZ from the underlying Ogallala aquifer. For convenience in the model, the Ogallala was treated as a confined aquifer. In reality, however, it behaves more as an unconfined aquifer. This simplifying assumption has no significant effect on perched water flow. Recharge was included for playas and inter-playa areas but higher rates were specified for the playa areas. Also, vertical groundwater movement was allowed to be greater beneath the playa areas. Since little information on flow in the larger region was available, the outer boundary of the model was assumed to be a no-flow boundary for the perched aquifer layer. Head dependent flow boundaries were used for the eastern and southwestern Ogallala boundaries. The Amarillo well field was modeled as a flux of 410 million cu-ft/year which is the approximate current pumping rate.

The hydrological fusion uses a steady-state finite-difference flow model which is similar in many respects to the U.S. Geological Survey MODFLOW model and is called MODLIKE. The most significant differences involve the method by which variations in conductivity are handled (MODFLOW assumes discontinuities while fusion assumes smooth variations between nodes) and the method for handling the variable water level in cells of an unconfined aquifer (fusion never lets cells become completely dry). This latter difference is probably the most significant since it allowed fusion to compute a solution for hydraulic conditions for which MODFLOW could not compute a solution. Similar fusion results have been obtained using a variably saturated model called VAM3D from HydroGeoLogic, Inc., validating the MODLIKE assumptions.

Fusion models the log hydraulic conductivity as a spatially correlated stochastic process with a systematic trend as advocated by Freeze et. al. (1990). For this model, the standard deviation in log hydraulic conductivity was assumed to be 0.4 with a 2000 foot correlation distance. The correlation distance is the distance beyond which the hydraulic conductivity is not substantially correlated. The average measured conductivity of the Ogallala is about 10,000 ft/year.

In order to minimize the sensitivity of the computed perched water flow to modeling assumptions, fusion was allowed to estimate critical parameters of the model. In addition to estimation of the head and log hydraulic conductivity at each node, fusion estimated the recharge in playa 1, and the vertical groundwater movement for the entire region, the Pantex region, and the four playas. Playa 1 was treated differently because the wastewater discharge from Pantex operations (originally drawn from the Ogallala aquifer) is channeled to playa 1. Recharge for

the entire region (except Pantex) was set to the average listed in several studies, 2.4 in/year. Recharge within playas 2 through 4 was assumed to be 6.7 in/year based on studies of the area.

Fusion was run with different setups specifying different conceptual models and different variables to be estimated. For example, the vertical groundwater movement for playa 1 was originally the same as for the other playas, but the converged fusion results suggested treating playa 1 differently. Even though the estimation is highly nonlinear (principally because of the variable water elevation in the perched aquifer) fusion produced a converged numerical model in only 8 or 9 iterations (approximately 15 minutes of SGI Indigo II time) for a given conceptual model.

Conventional manual calibration methods require hand tweaking of the numerical model for each conceptual model that is considered. In general, manual calibration can take days or weeks for a given conceptual model and there is no objective way to know if the calibration has converged. If it appears that the conceptual model should be changed, then the whole process needs to be laboriously repeated. The rapid convergence provided by Data Fusion replaces the slow hand tweaking. This frees the hydrogeologist to focus on the scientific and engineering issues of getting a good conceptual model rather than being burdened by tweaking yet another numerical model.

Figures 7.26 and 7.27 show the estimated head contours, flow pathlines, and log hydraulic conductivity. A good fit was obtained between the fusion model results and the field data. The RMS head residual is only 0.4 feet. Figure 7.27 shows the fusion estimates for heterogeneity in hydraulic conductivity. As explained by Freeze et. al. (1990), modeling of hydraulic conductivity heterogeneity is critical for confident groundwater modeling. Flow pathlines shown in Figure 7.26 are determined with confidence within the region of data coverage. Pathlines on the boundaries of the data coverage are sensitive to model boundary conditions and need additional data acquisition to be determined with confidence. Also, the data were not available for this study to separate recharge and vertical groundwater movement since a model can be calibrated with a low value of one as long as the other is low and vice versa. With more information that should be available about recharge, vertical groundwater movement could be separately estimated.

The flow pathlines shown in Figure 7.26 emanate from playa 1 at the top of the figure due to mounding of water under the playa from plant wastewater discharge to the playa. Water flowing south from playa 1 tends to part around Zone 12 with the bulk of the flow to one side or the other of Zone 12 but not directly through Zone 12. Then the flow bends to the east in the direction of the

channel in the top of the FGZ aquitard before it comes to the boundary of the data coverage. As additional data is acquired to the east, it is anticipated that it will be found that the flow continues to the east. Since fusion converges so rapidly, fusion could be used to update flow models in the field to provide real time guidance for site investigation.

7.6.3 Conclusions A groundwater model was calibrated for a perched aquifer under Zone 12 at Pantex using hydraulic head data, slug tests, and recharge information. The MODLIKE numerical model that is similar to the USGS MODFLOW model was validated by comparison with the variably saturated VAM3D model from HydroGeoLogic, Inc. A good model fit was achieved with an RMS head residual of only .4 feet. Hydraulic conductivity heterogeneity was estimated in order to provide flow pathlines with confidence within the region of data coverage.

As the conceptual model was modified to improve results, fusion converged rapidly for each conceptual model. Generally it took 8 or 9 numerical iterations at approximately 15 minutes of Indigo II time. This means that fusion modeling is fast enough to be used for real time field model updating. This provides a capability that has never before been possible with conventional manual model calibration.

As additional data is gathered to the east of Zone 12, Data Fusion modeling could be used to provide guidance of site investigation. Flow models could be updated in real time in the field to support decisions about placement of future borings. Further, acquisition of additional information about recharge would allow fusion to provide understanding about the separation of recharge and vertical groundwater movement effects.

Figure 7.25. PANTEX Fused Fine Grain Zone Elevation

Figure 7.26. PANTEX Head Contours And Pathlines

8.0 CONCLUSIONS

This program has produced a combination of advancements which now make it possible to provide distinctly finer resolution of geologies beneath contaminant sites. Given, good sharp seismic reflectors, it appears possible, with the enhanced geophysics and the Data Fusion System, to identify clay layers at depths as great as 300 feet. These new technologies have been applied successfully at the project's test location, the Savannah River Site at Aiken, SC, and at several other locations with significantly different geologic characteristics. The Data Fusion System technology has been successfully transferred to the geologic environment. Further, its computational power and ability to quantify uncertainty have been extended to hydrologic model calibration.

The program's specific conclusions are presented below, with the Data Fusion System and the geophysical advancements separately grouped.

8.1 Data Fusion Conclusions

A theoretical framework has been provided for engineering decision analysis that quantifies geological and hydrogeological uncertainty. Diverse non-invasive and invasive data sets are fused by relating measurements and targets of interest through first principles and statistical models. Data Fusion has been performed on UNIX workstations and Pentium PC's. A fusion track record of performance has been established in applications at the Savannah River Site and at DOE sites in Washington, Texas, and Ohio, plus a commercial site in Iowa.

Using Data Fusion and modeling, decision makers have a quantitative basis for action so the following benefits can be realized:

- Gets the most out of existing data sets to avoid cost of unnecessary acquisition.
- Estimates data worth in real-time before data acquisition to determine if reduction in uncertainty pays for the cost of acquisition.
- Enables remedial simulation to optimize cleanup. Remediation can be done in the computer to get it right with quantified safety margins before it is done in the earth.
- Enables real-time monitoring during remediation. Contaminant plumes can be seen while they are being remediated.
- Provides quantitative basis for cost reduction/ avoidance.

The Data Fusion implementation does not have the data limitations experienced by other approaches such as geostatistics and data assimilation. In particular, Data Fusion has the following features:

- Measured data related to target variables of interest can be fused through nonlinear forward models for the data.
- Spatial heterogeneity is incorporated using Markov Random Field (MRF) models for spatial continuity.
- First principles models relating the variables of interest are incorporated (such as ground water flow models, seismic reflection, electrical induction).
- Fusion updates prior knowledge using Bayesian inverse modeling to provide estimates of the variables of interest and quantified estimate uncertainty.
- The Data Fusion architecture has data management and visualization pre- and post-processing capabilities to provide an efficient and useable engineering decision tool.
- The Data Fusion has followed high software development standards as an open and extendable product. New data and models are incorporated by linking in software modules specific to the data and models.
- Data Fusion provides the basis for a service business that has been used for jobs at other DOE, DOD and commercial sites.

DATA FUSION APPLICATIONS:

A Data Fusion track record has been developing through applications at a number of sites described in the body of the report.

At Savannah River, Aiken, SC, fusion of electromagnetic data along Line 1 described in the body of the report was performed to provide an interpretation of the top of the Ellenton layer target. Fusion was contrasted with a current method of separately inverting data obtained at each location along the line. A labor intensive manual integration of single sounding inversions along the line with current methods produces results comparable to fusion. However, fusion provides quantification of the uncertainty in the interpreted top of the Ellenton not provided by either of the current methods. The Data Fusion results were validated by comparison with well picks that were not included in the run to test fusion.

Geophysical seismic reflection and electromagnetic data sets were fused along Line 1 to interpret the Ellenton layer. Each data set constrained the other, balancing their relative sensitivities and uncertainties. The seismic data provided information about the top and bottom of the Ellenton that in turn helped the electromagnetic data to provide an interpretation that distinguishes between width and resistivity of the Ellenton. This produced a resistivity map of the Ellenton. There is some evidence that resistivity can serve as a permeability surrogate for flow and transport simulation. Current geophysical methods do not provide this kind of joint interpretation capability.

Fusion of seismic and electromagnetic data along Line 2 at Savannah River showed the data sets to be inconsistent with each other except for a small portion of the line.

8.2 Geophysical Conclusions

The objectives of the geophysical work, as applied to the test site at SRS, has been to provide better information about the continuity, thickness, and clay content of the clay-rich strata within the succession of sand-rich unconsolidated sediments underlying the site. The unconsolidated sediments at SRS have a thickness varying from 800 ft to over 1,200 ft. The geophysical research and development (R&D) for mapping the stratigraphy was focused in the upper 500 ft. of the A&M area of SRS. Two surface geophysical methods were applied to differentiating clays and sands, i.e., measuring the geoelectric section and determining seismic reflection events.

The results of the geophysical surveys indicate that improvements in equipment and data analysis suggested in this R&D can be applied in practice. The measurements of the geoelectric section (by time domain electromagnetic; TDEM) detected a clay layer at a depth of about 70 meters, known as the Ellenton. This clay layer is part of the Crouch Branch Confining Unit (CBCU) which is reported to be one of the dominant barriers to downward contamination in the A&M area. The TDEM data suggest that significant lateral changes in the lithology of the CBCU occur in the A&M area. With the TDEM methods, a high density data set was obtained (approximately 4 to 8 times more data than taken in routine surveys) including measurements of the horizontal field. Software developed in this program allowed for the one-dimensional inversion and interpretation of this data.

For measures of the seismic section at SRS, vibrator sources were extensively tested and found to produce results that were better than those obtained using impact sources. These vibrators were used in compressional and shear wave modes using multi-component geophones. Processing of the data was accomplished entirely on PC's, significantly reducing the cost of processing shallow seismic reflection data. Shear wave reflection data provided the best quality seismic sections in the A&M area. The use of seismic shear wave reflection data allowed recognition of clay at 300 ft. depth. The main reason for the better vertical resolution of shear waves is the lower velocity of shear

waves compared to compression waves. The seismic sections derived from the data, show detail in stratigraphy of different units, including the CBCU.

Improvements were made in both technologies, but perhaps the greatest improvements in developing an understanding of the lithology and stratigraphy resulted from carefully integrating the information derived from both surface geophysical sensors with the geophysical and stratigraphic logs made available by Westinghouse, SRS. Integration allowed interpolation of stratigraphic units, and determining changes in lithology within these units.

For geophysical R&D to be successful requires a number of factors to come together for optimum performance, such as equipment, new theoretical concepts, and computer algorithms making use of new concepts in interpretation. Particularly in the seismic reflection studies at SRS, the R&D was performed at an opportune time when significantly improved seismographs, three-component geophones, and high frequency vibrators became available from commercial manufacturers. Extensive use in field data acquisition was made of these advances in commercial instrumentation. Also, R&D coincided with a period of intensive research by the oil and gas exploration industry in shear wave reflection and propagation, and the influence of anisotropy in acoustic properties of the Earth on seismic wave propagation and reflection. The knowledge resulting from the investigations by the oil and gas industry were of significant benefit to the R&D performed under this contract.

9.0 RECOMMENDATIONS FOR FUTURE DEVELOPMENTS

This multi-year program has produced two types of recommendations. First, having already exercised Data Fusion at several different DOE sites as a part of its development, there is a good understanding of new ways Data Fusion may be applied. Second, the acquisition of data under this contract has produced recommendations for enhanced survey methods and geophysical data processing.

9.1 Data Fusion

New technology development is recommended within the context of fusion capabilities that are needed before, during, and after remediation and waste management activities. The existing Data Fusion system has substantial capabilities as shown by the service work that is being performed at several DOE and commercial sites. However, additional fusion developments are needed as described below:

PC Implementation

Data Fusion is currently implemented on a UNIX workstation, primarily because of pre-and post-processing issues. Although computational algorithms themselves have been run on high end PC's, the data management, display, and 3-D visualization needs have been best met in the past with a UNIX workstation. Given that many environmental engineering firms in the commercial sector resist securing a workstation, when their engineers do all other work on a PC, action is needed to move the full Data Fusion activity to a high-end PC.

Full Transport Fusion (inverse modeling)

For flow and transport, fusion does flow inverse modeling, and work is on-going to provide the capability to propagate uncertainties through the transport calculation. Data Fusion provides the first practical capability to take advantage of the considerable data worth in hydraulic head data through flow inverse modeling. Considerable data worth also exists in contaminant concentration data which Data Fusion could use to refine transport modeling.

Data Fusion applies to transport in theory, but is in practice, limited by excessive computations. A technique has been identified that promises to solve the computational limitation that is a sparse least squares method motivated by the conjugate gradient method. What is needed is to develop, implement, and test the technique.

With completion of transport fusion, one will have the generic capability to statistically update time-dependent spatial maps of a broad class of processes. Where sufficient science has been done to provide predictive models, adequate data is available to estimate what is desired, and the form and scale is similar to flow and transport, the

door is opened for Data Fusion to make contributions. Applications similarly become possible in such areas as airborne contaminant monitoring, oceanography, and monitoring of biodegradation for bioremediation.

Improved Model Fidelity

Models of varying levels of fidelity are used in the geophysical and hydrogeological communities to relate variables of interest to measured data. The authors have developed a Data Fusion capability that incorporates general model structures. Then specific models were implemented for fusion applications. For example, seismic refraction uses an approximate ray-tracing model for a layered earth and time-domain electromagnetic (TDEM) uses a layered earth solution of Maxwell's equations. But there is no inherent limitation of Data Fusion to layered earth models, and some of the applications described require more heterogeneous models. Furthermore, there exist models that relate properties at a more fundamental level such as relating seismic properties to hydraulic properties. Fundamental property models may contain valuable prior knowledge that could further reduce uncertainties.

Additional development is needed to incorporate higher fidelity models. For seismic, wavefront tracing, wave-theoretic (for full waveform inversion), and various pre-stack models should be considered. For TDEM, gridded earth finite element representation of Maxwell's equations should be considered.

Discontinuity Detection And Spatial Continuity Tuning

Currently Data Fusion uses hypothesis testing to locate a discontinuity. Additional discontinuity models are also needed for faults and other abrupt changes in subsurface features. An alternative to hypothesis testing is direct estimation of discontinuities using a technique from "computer vision". Continuous properties with boundaries of discontinuity are estimated as coupled Markov Random Fields (MRF's) where the discontinuous boundary is a binary line process. The coupled MRF appears to be consistent with the basic form of this project's data fusion algorithm.

Currently the spatial continuity correlation and standard deviation parameters are manually tuned based on the least squares measurement residuals. We have an approach that promises to automate the tuning based on the Generalized Expectation Maximization (GEM) method. GEM is widely used in a number of communities.

9.2 Near-Surface Geophysical Exploration

The DOE's fostering of refined geophysical techniques for specific application to environmental challenges is rather unique. Although other Federal agencies have participated with the commercial sector in developing 1) very shallow systems (ground penetrating radar and shallow looking frequency domain electromagnetics) for detecting

buried objects, and 2) more acceptable intrusive methods like the cone penetrometer, none is known to have previously invested in non-intrusive, geophysical methods for the 15 to 300 foot depth at an environmental site. Now that this technological area has been opened, numerous extensions are becoming clearer. Thin layered soil structures, the precise extent of cohesive bodies, and perched water bodies are each of high interest to geoscientists, and potentially open to the new geophysical methods developed under this program.

9.2.1 Time Domain Electromagnetic Induction (TDEM)

This relatively inexpensive, large area coverage, geophysical method offer much promise when combined with other geophysical sensing systems.

9.2.1.1 Noise Rejection

Currently, one of the main limitations with TDEM measurements at SRS (and other built-up sites) is interference from infrastructure or cultural objects (e.g., power lines, fences, buried pipelines, etc.). In general, it is not possible to remove the interference caused by metallic objects, such as fences and buried pipelines, and they must be avoided. Interference caused by power lines may be caused by (i) direct EM interferences caused by the current carried in the power line, or (ii) coupling interaction between the TDEM and the power line. It is standard practice with TDEM to mitigate direct EM interference by stacking (summing) data over a period of time. This method is effective if the power line is well regulated. If the power line is not well regulated in frequency and amplitude, then stacking may not succeed. Existing commercially available shallow TDEM equipment performs analog stacking of received signals during the off-time period of the current waveform. After the stack has been completed then the data are digitized. With digital-full waveform data acquisition, noise sources could be more accurately quantified and removed. Previous experience by Blackhawk with deep digital TDEM systems (EM42) has shown that even partial waveform digital acquisition resulted in signal-to-noise improvements of a factor of 10 to 100. With true, full waveform digital acquisition systems, it may also be possible to acquire multi-channel (multi-receiver) data. This may also decrease measurement time required for each sounding. Three possible noise mitigation methods should be evaluated:

- Digital filtering in the time domain,
- Digital analysis/filtering in the frequency domain, and
- Deconvolution of noise by simultaneous (multi-channel) measurements of horizontal and vertical fields.

Evaluation of these noise reduction methods can be accomplished using synthetic data and then tested on the data already acquired at SRS.

Improvements in signal-to-noise ratio may be possible by simply increasing transmitter power for TDEM measurements. These modifications to the TDEM transmitter system have been made but not as yet tested.

9.2.1.2 2-D Modeling

The TDEM data collected at SRS were evaluated with forward and inverse modeling programs which assume a local 1-dimensional (1-D) geoelectric layering at each TDEM sounding location. The results of the interpretation infer 2-D geologic structure. 2-D forward modeling programs have been developed (Hoekstra, et. al., 1992), however, they require extensive CPU time for forward calculations and thus are inappropriate for the inversion process. An evaluation of the influence of 2-D structures on TDEM interpretation at SRS should be made using the existing 2-D forward modeling code. Further, the computational effectiveness of Data Fusion should be investigated for its potential for reducing the extensive CPU time.

9.2.1.3 Physical Properties from TDEM Data

Very likely the conductance derived from the TDEM strongly correlates to clay content, and therefore, vertical permeability. To increase the utility and reliability of the data the correlation between conductance, clay content, and vertical permeability needs to be established. This would require integration of surface and borehole geophysics, lithologic information, and physical testing.

9.2.2 Seismic

9.2.2.1 Polarization Analysis of Multi-Component Seismic Data

A major barrier to obtaining high quality near surface reflection data is the presence of strong ground roll events which can arrive at the geophones simultaneously with reflecting events. This is shown in Figure 9.1. Several methods have been used in the past to remove the ground roll energy from recorded seismograms including spatial filtering in the frequency-wave number (F-K) domain or by bandpass filtering. The drawback of these techniques is that they require the desired signal (reflecting events) and noise (ground roll) to be sufficiently separated in the frequency or wave number domain so that a filter can be applied which sufficiently removes the noise and still leaves sufficient signal. This generally requires reflections have frequencies above 200 Hz., which is normally not the case except in areas where the water table is within a few feet of the surface and explosive or projectile sources can be used; conditions which are rarely met in environmental geophysical surveys.

Another possible method for removing the ground roll is the use of polarization filters. It relies on the different direction of travel for the different seismic wave trains and the different particle motion (particle motion is the path that an element of earth will move through as the seismic wave travels through the Earth) which occur in the different types of energy. As shown in Figure 9.2, seismic reflections travel down to a geologic layer and then are reflected back to the surface, therefore their direction of propagation is near vertical. Furthermore the particle motion is always in the direction of travel for compressional (P) wave reflections and orthogonal for shear waves (Figure 9.3). This is in contrast to a majority of ground roll energy which is in the form of Rayleigh waves and which travel along the ground surface (Figure 9.2) and has a particle motion which is retrograde elliptical (Figure 9.3).

In order to characterize particle motion at least two, but preferably three orthogonal components must be recorded. This significantly increases the number of channels required to record the data, however, in cases where multi-component data is already being recorded this does not significantly increase the seismograph requirements.

Polarization filters have been used in analysis of seismograms from earthquake events for several years. Several methods have been used to compute the polarization of the data and subsequently filter the results, Vidale (1986) uses a method based on the analytic signal to compute the polarization of the data over a time window. Park and Lingberg (1987) use a method which can compute the polarization of the data as a function of frequency, by computing the polarization of the data in the frequency domain. However, the computation is still done only over a window of time. This basic approach is followed by Shieh and Hermann (1990) and Perelberg and Hornbostel (1994), where several examples are provided which show a significant decrease in the amplitude of ground roll after polarization filter processing. According to personnel at Amoco Production Co., polarization filtering is used extensively in their multi-component processing flow, with very good results. Other examples of polarization filtering can be found in Miles (1988).

Other possible methods of filtering could also be possible, such as matched filtering between the different components to remove (or enhance) seismic wave trains which appear on two or more orthogonal receivers (i.e., Ground Roll). Another possible use for polarization filtering is isolating the ground roll, the converse of what we have been proposing so far. The ground roll can then be analyzed and an inversion performed in order to derive the near surface velocity structure (upper 10-20 ft) this can then be used to perform near surface static corrections on seismic reflection data. Static corrections have proven to be difficult for shear wave reflection data, due to lateral variations in near surface shear wave velocities. Inversion of surface waves to derive subsurface velocities is done routinely on seismological data to determine the velocity structure of the Earth's crust (Patton, 1980).

LEGEND:

- REFRACTION
- GROUND ROLL
- REFLECTIONS
- AIR BLAST

REFLECTORS ① ② ③

CEES BLACKHAWK GEOSCIENCES DIVISION

WALKAWAY

Figure 9.1

RAYLEIGH WAVES

○ NORMAL POSITION OF PARTICLE

- POSITION DURING PASSAGE OF SHEAR PULSE

CEES BLACKHAWK GEOSCIENCES DIVISION

**PARTICAL MOTIONS FROM
RAYLEIGH AND SHEAR
WAVES**

Figure 9.3

The type of work that has been described in this Section has not been performed on high resolution seismic reflection data. If polarization filtering is successful at significantly reducing the ground roll energy in shallow reflection data, then it could significantly improve the quality and reliability of the data.

Efforts that would be required to further investigate polarization analysis and filtering are:

1. *Acquire seismic reflection data with all three components.* previously only one or two components have been acquired simultaneously. The increase in recording capacity could be handled by adding expansion boxes to currently used seismographs, or the addition of another seismograph to record the third component. Three component phones exist that could be used in the field acquisition effort. The recording of a third component is not expected to significantly increase the amount of field effort.
2. *Develop polarization analysis software.* This would be done within the structure of the Seismic UNIX(SU) a public domain seismic processing package which is supported by the Center for Wave Phenomena (CWP) at the Colorado School of Mines. This software is currently installed on our Silicon Graphics workstation and has been used as a framework for developing software for multi-component seismic data acquired under a program for DOE's Fossil Energy program.
3. *Investigate Different Methods of Filtering.* Analysis of different methods of polarization filtering will be done on shallow reflection data collected at SRS to evaluate the different types of filtering on high resolution reflection data. This will be performed on field records using the SU software, methods which appear to provide good results will be applied over records collected over a section of a line and processed to final stack and compared to evaluate the improvement in the final seismic section by using polarization filters.
4. *Investigate Near Surface Velocity Information Contained in Surface Waves.* Polarization filters will be applied to the data to produce records containing primarily surface waves. Inversion processes will be tested to determine the ability of these methods to produce near surface velocity information. The accuracy of the methods will be evaluated by applying the method to seismic data obtained near boreholes where surface to borehole velocity information has been obtained. If the surface wave inversion method provides accurate velocity information, then the static corrections derived from the derived velocity information will be applied to the reflection data, and the improvement in data quality evaluated.

9.2.2.2 Rotation of Shear Wave Data to Natural Coordinates

It has been demonstrated in many cases (Alford, 1986) that the Earth is not a seismically isotropic medium and that it is in most cases anisotropic. Seismic anisotropy has also been shown to be present in shallow unconsolidated sediments (Bates, 1989). This anisotropy can be due to several possible factors including clay particle orientation, regional stress, and fracturing. There are several different models of anisotropy that have been postulated including azimuthal anisotropy, transverse anisotropy and others. They differ based on the number of axes of symmetry. Further information on this subject can be found in Crampin (1981).

One of the effects of many types anisotropy is that a shear wave which is polarized in one direction may split into two polarized shear waves which have different velocities. This effect is shown in Figure 9.4. This can cause significant problems in the final shear wave stacks if this effect is present and is not accounted for. One manifestation of this is that shear energy will be apparent on geophones which are oriented perpendicularly to the source polarization. Another is discontinuous reflectors on shear wave shot records. These effects are present on the shear wave data we have collected at SRS. In order to extract the maximum amount of information from multi-component data, one should try to process the data so that the geophone and source orientations are lined up with the orientation of shear wave splitting, this is known as rotating the data into its natural coordinates.

A method that has been used in the hydrocarbon industry is to mathematically rotate data after acquisition, so that it appears as if the source and receiver axes correspond to the axes of shear wave splitting. However, the orientation of the natural coordinate system is not known, and must be determined from the data. The method proposed in Alford (1986) is to rotate the data in 5 to 10 degree increments and determine which orientation produces the best data on the diagonal components (the orientations at which the shear wave are generated after splitting) and the least amount of energy on the off-diagonal components. Another method for determining the required rotation and then rotating the data is proposed in Li and Crampin (1993) where a linear transform is applied to four (4) component data (two orthogonal receiver and two orthogonal sources). The angle is computed by forming a covariance matrix of the recorded time series and computing the eigenvalues and eigenvectors. The largest eigenvalue will then correspond to the eigenvector which will contain the information on the angles of shear wave splitting. The method proposed by Li and Crampin offers several advantages over other methods which use the rotation scanning to determine the angle. These advantages include:

- No assumptions regarding orthogonal splitting of shear waves,
- Allows pre-stack analysis and rotation of surface reflection data, and
- Significantly reduced computational requirements.

This method of shear wave data rotation has been coded by personnel at CEES-BGD and integrated into the SU seismic data processing system. This algorithm has also been tested on synthetic data provided by the University of Edinburgh.

To our knowledge, this type of rotation has not been applied to shallow seismic data. We propose that data that has been collected at the site for this project be evaluated to determine if shear wave splitting has occurred and if it has, whether or not rotation into natural coordinate system will improve the quality of the data.

Additional work which would be required would include:

- Evaluation of existing data to determine if shear wave splitting had occurred, and
- Rotation of data in raw state and reprocessing to final stack.

BLACKHAWK GEOSCIENCES DIVISION

SHEARWAVE PROPAGATION

Figure 9.4

10.0 REFERENCES

- Alford, R. M., 1986, "Shear Data in The Presence of Azimuthal Anisotropy," *56th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts*, pp. 476-479.
- Anderson, W. L., 1973, "Fortran IV programs for the determination of the transient tangential electric field and vertical M-Layer stratified earth by numerical integration and digital linear filtering," *U.S. Geol. Surv., Rept. USGS-GD-73-017*, 82.
- Bay Area, "CA and Estimation of Horizontal Stresses' Ratio from V_{s1}/V_{s2} ," *Geophysics*, 56, No. 6, pp. 822-832.
- Bierman, G. J., "Factorization Methods for Discrete Sequential Estimation," (New York: Academic Press, 1977).
- Bierman, G. J., Belzer, M. R., Vandergraft, J. S. and Porter, D. W., "Maximum Likelihood Estimation Using Square Root Information Filters," *IEEE, Trans. on Automatic Control*, 35, No. 12, Dec. 1990.
- Black, R. A., Steeples, D. W. and Miller, R. D., 1994, "Migration of shallow seismic reflection data," *Geophysics*, 59, No. 3, pp. 402-410.
- Carrera, J., and Glorioso, L., 1991, "On Geostatistical Formulations of the Ground Water Flow Inverse Problem," *Advances in Water Resources*, 14, No. 5, pp. 273-283.
- Chapman, W. L., and DiStefano, M. P., 1989, "Savannah River Plant Seismic Survey, 1987-1988", Research Report 1809-005-006-1-89: Conoco Inc. 11088.
- Chellappa, R. and Jain, A., *Markov Random Fields: Theory and Application*, (Boston: Academic Press, 1991).
- Clark, J. J. and Yuille, A. L., *Data Fusion for Sensory Information Processing Systems*, (Boston: Kluwer Academic Publishers, 1990).
- Coleman Research Corporation, 1994, "Phase II Topical Report; Geophysical Data Fusion for Subsurface Imaging."
- Coleman Research Corporation, 1993, "Phase I Topical Report; Geophysical Data Fusion for Subsurface Imaging."
- Constable, S. C., Parker, R. L. and Constable, C. G., 1987, "Occam's Inversion; A practical algorithm for generating smooth models from electromagnetic sounding data," *Geophysics*, 52, No. 3, pp. 289-300.

- Courtier, P., Derber, J., Errico, R., Louis, J. F. and Vukicevic, T., 1993, "Important Literature on the use of Adjoint, Variational Methods and the Kalman Filter in Meteorology," *Tellus*, 45A, pp. 342-357.
- Crampin, S., 1981, "A Review of Wave Motion in Anisotropic and Cracked Elastic Media," *Wave Motion*, 3, pp. 343-391.
- Cumbest, R. J., Price, V. and Anderson, E. E., 1992; "Gravity and Magnetics Modeling of the Dunbarton Triassic Basin, South Carolina," *Southeastern Geology*, 33, No. 1, pp. 37-51.
- Daley, R., *Atmospheric Data Analysis*, (New York: Cambridge University Press, 1991).
- Dempster, A. P., Laird, N. M. and Rubin, D. B., "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Society, Ser., B.39*, 1977.
- Domoracki, W. J., 1994, "A Geophysical Investigation of Geologic Structure and Regional Tectonic Setting at the Savannah River Site, South Carolina," *USDOE Report WSRC-TR-94-0317*, pp. 173.
- "Draft Phase I Report and Phase II Work Plan: Expedited Site Characterization for the DOE Pantex Zone 12 Groundwater Investigation," Argonne National Laboratory, September 1994.
- Everest Geotech Company, 1988, "Integrated model of the geologic system at Savannah River Plant", Aiken, South Carolina.
- Fallow, W. C. and Price, V., 1992, "Outline of stratigraphy at the Savannah River Site," in Fallow, W. C. and Price, V., Eds., "Geological Investigations of the Central Savannah River area," South Carolina and Georgia, *Carolina Geological Society Field Trip Guidebook*, Nov. 13-15, Augusta, GA.
- Freeze, R. A., Massmann, J., Smith, L., Sperling, T. and James, B., "Hydrogeological Decision Analysis: 1.A Framework," *Ground Water*, Sep.-Oct. 1990.
- Freeze, R. A., Massmann, J., Smith, L., Sperling, T. and James, B., "Hydrogeological Decision Analysis: 4. The Concept of Data Worth and Its Use in the Development of Site Investigation Strategies," *Ground Water*, Jul.-Aug. 1992.
- Gelb, A., *Applied Optimal Estimation*, (Mass: M.I.T. Press, 1974).
- Geman, S. and Geman, D., 1984, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, pp. 721-741.

- Ghil, M., 1989, "Meterological Data Assimilation for Oceanographers, Part 1, Description and Theoretical Framework," *Dynamics of Atmospheres and Oceans*, 13, pp. 171-218.
- Gibbs, B. P., "Precision Recursive Estimator For Ephemeris Refinement (PREFER)," *Goddard Space Flight Center, Flight Mechanics/Estimation Theory Symposium*, Oct. 1985.
- Gibbs, B. P. and Estes, R. H., "Geomagnetic Field Modeling by Optimal Recursive Filtering," *Fourth Scientific Assembly, International Association of Geomagnetism and Aeronomy*, Apr. 1982.
- Gibbs B. P. and Porter, D.W., "Development and Evaluation of an Adaptive Algorithm for Predicting Tank Motion," *Proceedings of Conference on Decision and Control*, Dec. 1980.
- Gibbs, B. P., "Jump Detection/Estimation Using Stepwise Regression," *IEEE Trans. on Aerospace and Electronic Systems*, 28, No. 4, Oct. 1992.
- Golub G. H. and Van Loan, C. F., *Matrix Computations*, (Baltimore: The Johns Hopkins University Press, 1989).
- Hall, D. L., *Mathematical Techniques in Multisensor Data Fusion*, (Boston: Artech House, 1992).
- Hoekstra, P. and Blohm, M. W., 1990, "Base Histories of Time-Domains Electromagnetic Soundings in Environmental Geophysics: in Ward, S. H., ed., *Geotechnical and Environmental Geophysics*, No. 9, pp. 1-16.
- Hoekstra, P., Stoyer, C. H., James, B. A. and Blohm, M. W., "Characterization of Reservoir Rocks and Fluids by Surface Electromagnetic Transient Methods," Blackhawk Geosciences, Inc., Interpex Limited, and James Consulting Services, *Annual Report U.S. Department of Energy*, Contract No. DE-AC22-90BC14476, Jan. 1992.
- Hoerl, A. E. and Kennard, R. W., "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 12, No. 1, Feb. 1970, pp. 55-67.
- Inman, G. R., 1975, "Resistivity Inversion with Ridge Regression," *Geophysics*, 40, No. 5, pp. 798-817.
- Kalman, R. E., 1960, "A New Approach to Linear Filtering and Prediction Problems," *Trans, ASME, Ser. D: J. Basic Eng.*, 82, pp. 35-45.
- Kaufman, A. A. and Keller, G. V., 1983, "Frequency and Transient Soundings," *Elsevier Science Publishers B.V.*, pp. 685.

- Levy, L. J. and Porter, D. W., "Large-Scale System Performance Prediction with Confidence from Limited Field Testing Using Parameter Identification," *the Johns Hopkins APL Technical Digest*, 13, No. 2, 1992.
- Lewis, S. E. and Aadland, R. K., 1992, "Hydrogeologic Setting of A&M Area: Framework for Groundwater Transport (U)," *USDOE Report WSRC-TR-92-355*, Westinghouse Savannah River Co., Savannah River Site, Aiken, SC 29808, pp. 101.
- Li, X. Y. and Crampin, S., 1993, "Linear Transform Techniques for Processing Shear-Wave Anisotropy in Four-Component Seismic Data," *Geophysics*, 58, No. 2, pp. 240-256.
- Lynn, H. B., 1991, "Field Measurements of Azimuthal Anisotropy: First 60 M San Francisco Bay Area, CA and Estimation of Horizontal Stresses' Ration from V_{s1}/V_{s2} ," *Geophysics*, 56, No. 6, pp. 822-832.
- Massmann, J., Freeze, R. A., Smith, L., Sperling, T. and James, B., "Hydrogeological Decision Analysis: 2. Applications to Ground-Water Contamination," *Ground Water*, Jul.-Aug. 1991.
- McLaughlin, D., "Recent Developments in Hydrologic Data Assimilation," Prepared for the U.S. Report to the IUGG, Jun. 1994.
- Miles, D. R., 1988, "Three Component Seismic Data: Researcher's Toy or Interpreters Tool?," *The Leading Edge*, 7, No. 11, pp. 28-31.
- Mosteller and Tukey, 1977, "Data Analysis and Regression," *Addison-Wesley*, Chapter 10.
- "Pantex Plant Groundwater Sampling Report August 1994 Sampling Round," U.S. Army Corps of Engineers, Tusa District, March 1995.
- Park, J., Vernon, F. L. and Lindberg, C. R., 1987, "Frequency Dependent Polarization Analysis of High Frequency Seismograms," *Journal of Geophysical Research*, 92, No. B12, pp. 12.664-12.674.
- Patton, H., 1980, "Crust and Upper Mantle Structure of the Eurasian Continent From the Phase Velocity and Q of the Surface Waves, Rev.," *Geophys. Space Phys.*, 18, pp. 605-625.
- Perelberg, A. I. and Hornbostel, S. C., 1994, "Applications of Seismic Polarization Analysis," *Geophysics*, 59, No. 1, pp. 119-130.
- Petty, A. J., Petrafeso, F. A. and Moore, F. C., Jr., 1965, "Aeromagnetic map of the Savannah River Plant area, South Carolina and Georgia," *U.S. Geol. Surv. Geophys. Inv. Map*, GP489.

- Porter, D. W., Gibbs, B. P., Vandergraft, J. S., Hatch, W. E., Kelsey, P. W., Hoekstra, P., Hoekstra, B., Blohm, M., Phillips, D. and Bates, R., "Waste Site Characterization Using Data Fusion Workstation at Hanford and Savannah River DOE Sites," presented at ER93 -- Environmental Remediation Conference in Oct. 1993. Coleman Research Corporation, 9891 Broken Land Parkway, Suite 200, Columbia Maryland 21046.
- Porter, D. W. and Englar, T. S., "Multi-object Tracking via a Recursive Generalized Likelihood Approach," *Proceedings of Conference on Decision and Control*, Dec. 1979.
- Porter, D. W., "Quantitative Data Fusion: A Distributed/Parallel Approach to Surveillance, Tracking, and Navigation using Information Filtering," *Fifth Joint Data Fusion Symposium, Johns Hopkins University/Applied Physics Laboratory*, Oct. 1991.
- Poeter, E. P. and McKenna, S. A., "Geostatistical Simulation and Inverse Flow Modeling to Reduce Uncertainty Associated with Flow and Transport Predictions," *Proceedings of the 1994 Ground Water Modeling Conference*, Ft. Collins, CO, International Ground Water Modeling Center Publication 1994.
- Rubin, Y., Mavko, G. and Harris, J., "Mapping Permiability in Heterogeneous Aquifers Using Hydrologic and Seismic Data," *Water Resources Research*, 28, No. 7, 1992.
- Sheriff, R. E., 1978, "A First Course in Geophysical Exploration and Interpretation," (Boston: IWRDC), pp. 313.
- Shieh, C. F. and Herrmann, R. B., 1990, "Ground Roll: Rejection Using Polarization Filters," *Geophysics*, 55, No. 9, pp. 1216-1222.
- Snipes, D. S., Fallaw, W. C., Price, V. and Cumbest, R. J., 1993, "The Pen Branch fault: documentation of late Cretaceous-Tertiary faulting in the Coastal Plain of South Carolina," *Southeastern Geology*, 33, No. 4, pp. 195-218.
- Sperling, T., Freeze, R. A., Massmann, J., Smith, L. and James, B., "Hydrogeological Decision Analysis: 3. Application to Design of a Ground-Water Control Systems at an Open Pit Mine," *Ground Water*, May-Jun. 1992.
- Steinway, W., "Earth Penetration Radar Imaging System (EPRIS)," Coleman Research Corporation, *Final Report for U.S. Army and Department of Energy*, Contract No. DAAA21-90-D-0007, Aug. 1993.
- TEMIXH, "Program for Interpretation of TDEM Data", Interpex Ltd., 715 14th Street, Golden, CO 80401, C. Stoyer, President.
- U.S. Environmental Protection Agency (U.S. EPA), Feb. 1992. "Evaluation of Ground-Water Extraction Remedies: Phase II. Case Studies and Updates," *Publication 9355.4-05A*.

Vandergraft, J. S., "Efficient Optimization Methods for Maximum Likelihood Parameter Estimation," *Proceedings of 24th Conference on Decision and Control*, Ft. Lauderdale, FL Dec. 1985.

VanGeer, F. C., Te Stroet, C. B. M. and Yangxiao, Z., 1991, "Using Kalman Filtering to Improve and Quantify the Uncertainty of Numerical Ground Water Simulations, 1. The role of System Noise and its Calibration," *Water Resource Res.*, 27, No. 8, pp.1987-1994.

Van Nieuwenhuise, D. S., and Colquhoun, D. J., 1982, "The Paleocene-lower Eocene Black Mingo Group of the East-central Coastal Plain of South Carolina," *South Carolina Geology*, 26, No. 2, pp. 47-67.

Vidale, J., 1986, "Complex Polarization Analysis of Particle Motion," *Bulletin of Seismological Society of America*, 76, No. 5, pp. 1393-1405.

Westinghouse Savannah River Company, 1994, "Hydrogeologic Framework for the A&M Area (U)," Savannah River Site, Aiken, SC 29808.

Westinghouse Savannah River Company, 1992a. "Hydrogeologic Setting of the A/M Area: Framework for groundwater transport (4)," *WSRC-TR-92-355*, Prepared for the Westinghouse Savannah River Company, Savannah River Site, Aiken, SC 29808.

Widess, M. B., 1973, "How Thin is a Thin Bed," *Geophysics*, 38, pp. 1176-1180.

Wiener, N., 1930, "Generalized Harmonic Analysis," *Acta Math*, 55, pp. 117-258.

Wilt, M. J., 1991, "Interpretation of Time Domain Electromagnetic Soundings Near Geological Contacts", PhD Thesis, University of California, Earth Sciences Division, Lawrence Berkeley Laboratory, pp. 187.

Yilmaz, O., 1987, "Seismic Data Processing," *Soc. Expl. Geophys. Inv. Geophys.*, 2, pp. 526.

11.0 GLOSSARY

central loop TDEM - a TDEM measurement configuration with the receiver placed in the center of the transmitter loop.

measurement forward model - the equations and logic used to compute a predicted measurement value given the required earth parameter values.

measurement residual - for a least squares estimator, such as fuse3d or TEMIX, the measurement residual is defined as the difference between the predicted and actual measurement value. The predicted measurement value is obtained by substituting the currently estimated earth model parameters, at the measurement location, into the measurement forward model. Measurement residuals are often referred to simply as residuals.

out-of-loop TDEM - a TDEM measurement configuration with the receiver placed outside of the transmitter loop.

plot point - for an out of loop TDEM sounding the receiver is placed outside the transmitter loop. Within the Data Fusion, the sounding position is taken to be the average of the receiver position and the transmitter loop center position. This position is called the plot point.

residual RMS - square root of the sum of squared standardized measurement residuals divided by the number of measurements. If all of the assumed measurement sigmas are correct, the measurement models are correct, the earth model is correct and the nonlinear least squares estimator has converged, then the residual RMS should be close to 1.0.

spatial tau - the distance at which the correlation between elements of a spatial random field has decayed from 1.0 to .37.

standardized measurement residual - the measurement residual divided by the measurement noise standard deviation. All Fuse3d residual reports and plots use the standardized measurement residuals. The standardized measurement residuals are often referred to as the standardized residuals, or in the Data Fusion application, simply as the residuals.

TDEM - time domain electromagnetic

Appendix A

Manufacturers Specifications

Geonics TEM47 TDEM Transmitter
Geonics Protem TDEM Receiver

Geonics Limited
8-1745 Meyerside Drive
Mississauga, Ontario
Canada L5T 1C6

PROTEM RECEIVER

It is well known that there is a trade-off between depth of exploration and target definition in terms of conductivity, extent and orientation. Greatest depth is obtained with large fixed loop Turam-type systems which generate large half space responses and along with current gathering makes target detection difficult. Better spatial resolution is obtained with a moving transmitter configuration with a short intercoil spacing but is limited to a shallower depth of exploration. These variations in survey requirements make system flexibility an important design consideration.

Time Domain systems are also now routinely employed for general geological exploration such as for freshwater aquifers in bedrock fractures, and mapping groundwater contaminant plumes. Mapping to the shallow depths required in these applications requires a very wide bandwidth and many narrow sampling gates.

Recognition of these diverse requirements led Geonics to develop the extremely flexible PROTEM time domain system. The digital, 3 channel receiver is used with any of the 3 TEM transmitters and choice of receiver coil to cover all applications. With its 23 bit resolution, system bandwidth of 500 kHz, microsecond sampling gates and simultaneous XYZ component measurements, the PROTEM receiver provides the ultimate in time domain capability. Used with the Geonics 3-component coil, mineral surveys are greatly speeded up with more data in either the fixed loop or slingram mode. The three component measurement also allows a quick and accurate check on geoelectric sounding data for lateral variations in conductivity which could invalidate a layered-earth interpretation.

Specifications

MEASURED QUANTITY	Rate of decay of induced magnetic field along 3 axes, in nV/m ²
EM SENSOR	Air- and ferrite-cored coils
CHANNELS	3 in parallel or sequential
TIME GATES	20 geometrically spaced, from 6 μ s to 800 ms
DYNAMIC RANGE	23 bits (132 dB)
FREQUENCY	0.3, 0.75, 3, 7.5, 30, 75 and 285 Hz or 0.25, 0.625, 2.5, 6.25, 25, 62.5 and 262.5 Hz
INTEGRATION TIME	2, 4, 8, 15, 30, 60, 120 or 240 s
DISPLAY	240 x 64 dot graphic LCD
DATA HANDLING	Solid-state memory for 3300 data-sets. RS-232 output
SYNCHRONIZATION	Reference cable or, optionally, highly stable quartz crystal
POWER SUPPLY	12 V rechargeable battery for 8 h continuous operation
WEIGHTS	15 kg
DIMENSIONS	34 X 38 X 27 cm

TEM47 TRANSMITTER

Three interchangeable transmitters - TEM47, TEM57 and TEM37 - are used with the PROTEM receiver and the appropriate receiver coil to make up different PROTEM systems for various applications such as mineral exploration, structural mapping, resistivity sounding and contaminant plume mapping.

The TEM47 is the smallest, lightest battery operated transmitter with a very fast turn-off time to enable the near surface response to be measured. The PROTEM 47 (PROTEM receiver and TEM47 transmitter) is most often used for shallow geoelectric sounding looking for conductive contaminant plumes, saline intrusion or general stratigraphy mapping. In this mode single turn transmitter loops from 5 m up to 100 m on a side with turn-off times as short as half a microsecond can be used to give maximum near surface resolution.

The maximum transmitter output of 3 A into a 100 m x 100 m loop gives a good response and resolution to depths down to 150 m making this the ideal instrument for resistivity sounding over a large area.

The TEM47 uses a reference cable to achieve the high synchronization accuracy required for shallow sounding. Regardless of application, the high-frequency receiver coil is used in PROTEM 47 systems. This receiver coil has the bandwidth necessary to capture the earliest portion of the transient decay.

When the TEM47 is used in a PROTEM 47 system for profiling, it supplies 3 A to an 8-turn, 5 x 5 m moving transmitter loop to provide a dipole moment of 600 ampere square metres. With base frequency of 75 Hz and 20 gates from 49 μ s to 2.9 ms this configuration is optimal for Slingram (horizontal loop) surveys for mineral exploration to shallow depths, and for groundwater exploration in bedrock fractures. Electrical sounding is performed simultaneously with the search for fault or dike-like targets.

Specifications

CURRENT WAVEFORM	Bipolar rectangular current with 50 % duty cycle
FREQUENCY	30, 75 or 285 Hz where powerline frequency is 60 Hz. 25, 62.5 or 262.5 Hz where powerline frequency is 50 Hz
TURN-OFF TIME	2.5 μ s at 3 A into 40x40 m loop. Faster into smaller loop
TRANSMITTER LOOP	5 x 5 to 100 x 100 m single turn loop, or 5 x 5 m 8-turn loop
OUTPUT VOLTAGE	0 to 9 V, continuously variable
POWER SUPPLY	Internal 12 V rechargeable battery
BATTERY LIFE	5 h continuous operation at 2 A output
DIMENSIONS	10.5 x 24 x 32 cm
WEIGHT	5.3 kg

Appendix B
Manufacturers Specifications
Geonics EM39 Borehole Conductivity Probe

Geonics Limited
8-1745 Meyerside Drive
Mississauga, Ontario
Canada L5T 1C6

BOREHOLE CONDUCTIVITY PROBE

EM39

The EM39 provides measurement of the electrical conductivity of the soil and rock surrounding a borehole or monitoring well, using the inductive electromagnetic technique. The unit employs coaxial coil geometry with an intercoil spacing of 50 cm to provide a substantial radius of exploration into the formation while maintaining excellent vertical resolution. Measurement is unaffected by a conductive borehole fluid or the presence of plastic casing. The instrument operates to a depth of 500 metres.

The combination of a large conductivity range, high sensitivity and very low noise and drift, allows accurate measurement of subsurface conditions; typical applications include groundwater contamination monitoring, groundwater and mineral exploration, and general geotechnical applications.

The 4-conductor EM39 probe can be used with many available borehole logging systems or with Geonics dedicated winch and console system.

Measurements can be either recorded with the DL720 digital data logger, or viewed in real time using the EM39RT program with field computer.

Specifications

MEASURED QUANTITY	Apparent conductivity in millisiemens per metre (mS/m)
PRIMARY FIELD SOURCE	Self-contained dipole transmitter
SENSOR	Self-contained dipole receiver
INTERCOIL SPACING	50 cm
OPERATING FREQUENCY	39.2 kHz
CONDUCTIVITY RANGES	$\pm$ 100, 1000, 10000 mS/m
DEPTH	200 m (500 m cable optional)
MEASUREMENT PRECISION	$\pm$ 0.1 % of the full scale
MEASUREMENT ACCURACY	$\pm$ 5 % at 30 mS/m
NOISE LEVEL	< 0.5 mS/m
POWER SUPPLY	10 disposable "D" cells, or 12 VDC external power source
DIMENSIONS	Probe: 3.6 cm diameter, 163cm length
WEIGHTS	Probe: 2.2 kg Shipping: 90 kg (2 boxes)

Appendix C

Manufacturers Specifications

**Mount Sopris Natural Gamma/SP/SPR Probe
Mount Sopris Digital Logger**

Mount Sopris Instrument Co., Inc.
17301 West Colfax Ave., Suite 255
Golden, CO 80401

HLP-2375/S COMBINATION NATURAL GAMMA/SP/SPR PROBE

The Mount Sopris HLP-2375/S combination natural gamma/SP/Single Point Resistance probe provides 3 measurements that are invaluable for most logging operations. The natural gamma detector features a high sensitivity scintillation detector based on a large 7/8" (22 mm) diameter by 3" (76 mm) long NaI crystal which provides excellent repeatability and definition, even through steel casing.

In holes filled with conductive fluid, the SP and Single Point Resistance measurements provide important information on fluid salinity contrast, stratigraphy, and qualitative resistivity contrast.

The probe operates on a single conductor, and is designed to log gamma data going down, and SP and SPR coming up. Automatic data merging is provided with the **LOGSHELL** program when the HLP-2375 is logged with the MGX digital logger.

Power 28 VDC @ 85 ma for gamma
SP/SPR passive (depend on surface VM and Current Generator)

Length 120 cm. Diameter 4.1 cm (with insulation) Weight .65 kg.

HLP-2375/S

MOUNT SOPRIS
Instrument Company, Inc.

MGX PORTABLE DIGITAL LOGGER

Now you can log
with your
own PC!

Console/Winch:

Console mounted directly on winch side plate. Controls all probe functions. Outputs data in serial format directly to PC
Small footprint (19" L X 11.4" H X 9.5" W)(485 mm X 290 mm X 242 mm)
Single conductor for ease of maintenance. Motor drive with speed control
110/220 VAC power (12 VDC inverter option)
Cable capacity-1000' (305 m) of 0.10", 660' (200 m) of .0125" diameter.
Total weight 69 lbs (31 kg) including heavy duty polypropylene transit case.

Software:

LOGSHELL menu-driven acquisition and processing software with pull-down HELP. Provides full onscreen *waterfall* display while logging, direct hardcopy output to dot-matrix, ink jet or thermal printer. Includes PRNPLOT, SCRNPLOT, INTERP, and PRNHEAD functions for enhanced custom data presentation. Requires 386SX MSDOS compatible portable PC with hard drive and VGA graphics. (PC\Printer may be supplied by customer or ordered through MSI).

Accessories:

System comes complete with standard Gamma/SP/Single Point Resistance probe, sturdy anodized aluminum tripod with removable pulley for rig operations, cable line wiper, consumables, and rugged probe/tripod transit case.

Post-It™ brand fax transmittal memo 7671 # of pages > 3

To Ann Winda	From Tim Lobco
Co. Blackhawk	Co. MSI
Dept.	Phone # 279-0171
Fax # 278-0289	Fax #

Example of LOGSHELL menu screen ▽

Logging Menu

Probe DataFile Depth Log Printer Exit

CURRENT PARAMETERS:
 Probe: MLP-2375/S
 Measurements: Natural Gamma Ray
 Depth Reference: R&P Electrode 2.5 Ft. from Top
 Starting Depth: 200

Selected Data File: testwell.dat
 Begin Logging.

Press F1 for Help

AVAILABLE MEASUREMENTS

SINGLE FUNCTION PROBES:

Motorized Caliper (three-arm)
 Temperature
 Fluid resistivity
 Flowmeter
 Guard Resistivity
 Neutron
 Gamma-Gamma (4π)

COMBINATION PROBES:

Gamma/SP/Single Point Res.
 Normal Res. (8-16-32-64")
 Temperature/Fluid Res.
 Caliper/Temperature

PRNPL0T EXAMPLE PLOT
 (1/2 SCALE)

MOUNT SOPRIS INSTRUMENT CO., INC.
 17301 WEST COLFAX AVENUE SUITE 255
 GOLDEN CO 80401 USA
 PHONE 303 279-3211 FAX 303 279-2730

Appendix D
Manufacturers Specifications
iVi "Minivib"

Industrial Vehicles International, Inc.
Tulsa, Oklahoma

'minivib'

Trailer Deployment System T-7000

The **T-7000** deployment system is one of several ways in which the 'minivib' can be configured. This trailer system is available in three different models. The most versatile configuration is the **T-7000W**. This system has several water tanks which can be filled to increase the trailer weight to provide the hold-down weight required. With this trailer in its heaviest configuration, the full output of the 'minivib' can be utilized in the 10 to 190 Hz. band. This makes it ideal for surveys with deeper targets. After a survey is completed, the water can then be drained from the tanks and the unit can be driven to its next survey. In addition to the **T-7000W**, there is also the **T-7000L** and the **T-7000S**. These models are described on the back.

T-7000 FEATURES

- ★ The **T-7000** allows the operator to use the full capability of the 'minivib' throughout its frequency and power range.
- ★ Electric brakes are provided on the trailer axles.
- ★ A fifth-wheel trailer hitch allows for convenient hook-up and good maneuverability.
- ★ The weight of the trailer is centered around the vibrator so that the trailer weight can be fully utilized for hold-down.
- ★ The **T-7000W** has a hydraulically powered water pump for filling the water tanks on location.

INDUSTRIAL VEHICLES INTERNATIONAL, INC., TULSA, OKLAHOMA, U.S.A.

'minivib'

Shear Wave Option

The 'minivib' is a controllable vibratory source developed for high resolution seismic surveys. This source has the ability to generate a user determined signal which falls within a frequency range of 10 to 550 hz. Included in the 'minivib' package is a portable IBM PC based controller used for controlling and monitoring vibrator functions.

The Shear Wave Option gives the 'minivib' the additional feature of allowing the vibrator mass to operate horizontally. With the standard 'minivib' the mass operates vertically. The Shear Wave Option includes a mass mounting fixture, a mass rotating mechanism, and a shear wave baseplate attachment. The mass rotating cylinder is designed to reposition the mass for inline and crossline production shear wave operation.

Bottom of the baseplate attachment

'minivib' with the Shear Wave Option

Shear Wave Option Features

1. The mass mounting fixture mounts the 'minivib' mass in a horizontal position on the baseplate. The mass assembly is identical to the standard P-wave mounting.
2. The mass rotating cylinder rotates the mass and the mass rotating fixture 90 degrees relative to the baseplate. This is done through a remote controlled electric switch.
3. The mass locking clamp is located internally in the baseplate. This is a spring actuated and hydraulic released clamp which is controlled remotely.
4. The shear wave baseplate attachment is a structure with multiple wedges which have been designed to maximize shear wave coupling to the earth, and at the same time minimize environmental damage to the earth.

INDUSTRIAL VEHICLES INTERNATIONAL, INC.

6737 East 12th Street, Tulsa, Oklahoma, 74112 U.S.A.

Phone (918)836-6516

TULSA, OKLAHOMA

FAX (918)838-9529 Telex 49-7574

Appendix E

Manufacturers Specifications

OYO DHS-1 Data Acquisition System
OYO GS-40D Rotating Coil Geophone

OYO Geospace
7334 N. Gessnet
Houston, TX 77040

DAS-1

Data Acquisition System

- 24, 48, 96, and 144 channel systems
- 24-bit Sigma Delta A/D technology
- Dynamic resolution of 114 dB (measured).
- System dynamic range of 132 dB (measured)
- 32-bit stacking
- Digital data storage in SEG-1, SEG-2 or SEG-D on floppy disk, hard disk, mini-cartridge tape, high capacity DAT, 9-track tape, or 3480 cartridge
- Large wide-angle EL VGA display
- Internal and/or external thermal plotting
- Post acquisition digital filtering
- Sample rate from 31.25 μ s to 4 ms
- Programmable field layout
- Noise editing signal enhancement
- Playback gain control
- Software roll-along capability
- Vibroseis control software

DAS-1

DATA ACQUISITION SYSTEM

DAS-1 is a multipurpose hi-performance seismic data acquisition system designed for engineering, environmental, and oil & gas exploration surveys. DAS-1 exceeds current seismograph performance standards and is easily adapted for both land and marine operations. Field applications include refraction profiling, high resolution P and S wave reflection, VSP, crosshole tomography, multi-component P and S wave borehole surveys, and acoustic emissions monitoring. System configurations range from low cost, portable 24 channel units to sophisticated 144 channel acquisition and processing systems.

With up to 80,000 samples per trace, high dynamic range, and high dynamic resolution, DAS-1 is ideal for recording signals generated from either impulsive or vibratory energy sources. Data may be recorded onto floppy disks, hard disk or high capacity tape storage systems. Plotting capability includes direct output to an internal 4.4 inch thermal printer or to an external 8.5 inch thermal plotter.

DAS-1 incorporates an 80486 processor as its main system controller. All system operation and acquisition functions such as parameter input, automatic roll-along, data recording, and plotter output are managed by menu-driven software. For quality control of instrument operation and data acquisition parameters, DAS-1 features a full array of system diagnostics and performance tests. As a portable processing system, DAS-1 can run DOS based software for in-field processing and analysis of seismic data. To enhance processing capabilities, the system includes auxiliary ports for connections to an external keyboard, a color VGA monitor, a mouse, and a printer.

STANDARD CONFIGURATIONS

24 Channel System: 2 AUX channels, 4 MB acquisition memory, 4 MB system memory, 1.44 MB 3.5 inch floppy disk drive, 120 MB hard disk, and internal thermal printer

48 Channel System: 2 AUX channels, 16 MB acquisition memory, 16 MB system memory, 1.44 MB 3.5 inch floppy disk drive, 120 MB hard disk, and internal thermal printer

96 Channel System: 4 AUX channels, 16 MB acquisition memory, 16 MB system memory, 1.44 MB 3.5 inch floppy disk drive, 120 MB hard disk, internal thermal printer, and a 48 channel expansion unit

144 Channel System: 6 AUX channels, 16 MB acquisition memory, 16 MB system memory, 1.44 MB 3.5 inch floppy disk drive, 120 MB hard disk, internal thermal printer, and two 48 channel expansion units

STANDARD AND OPTIONAL EQUIPMENT

FEATURES	24 Ch. System	48 Ch. System	96 Ch. System	144 Ch. System
Acquisition Memory	Standard: 4 MB Optional: 16 MB	Standard: 16 MB	Standard: 16 MB	Standard: 16 MB
System Memory	Standard: 4 MB Optional: 16/32 MB	Standard: 16 MB Optional: 32 MB	Standard: 16 MB Optional: 32 MB	Standard: 16 MB Optional: 32 MB
Maximum Samples per Trace	Standard: 30 K Optional: 80 K	Standard: 40 K Optional: 80 K	Standard: 20 K Optional: 40 K	Standard: 13 K Optional: 27 K
Hard Disk Capacity	Standard: 120 MB Optional: 240/480 MB	Standard: 120 MB Standard: 240/480 MB	Standard: 120 MB Optional: 240/480 MB	Standard: 120 MB Optional: 240/480 MB
1.44 MB Floppy Disk	Standard	Standard	Standard	Standard
120 MB Tape Back-up System	Optional	Optional	Optional	Optional
2 Gigabyte DAT	Optional	Optional	Optional	Optional
1600/6250 bpi Tape Drive	Optional	Optional	Optional	Optional
3480 Format Tape Cartridge	Optional	Optional	Optional	Optional
Internal 4.4 inch Thermal Printer	Standard	Standard	Standard	Standard
External 608P 8.5 inch Plotter	Optional	Optional	Optional	Optional
Expansion Unit(12-48 channels)	None	None	One	Two
Connectors	Cannon	Cannon/Bendix	Bendix	Bendix
Vibroseis Interface	Optional	Optional	Optional	Optional

DAS-1 SPECIFICATIONS

DATA ACQUISITION

Number of Channels	24, 48, 96 or 144
Number of Auxiliary Channels	2 - 6 (12 dB Gain)
Sampling Intervals	31.25, 62.5, 125, 250 μ sec 0.5, 1, 2, and 4 ms
Number of Samples per Trace	48 channels - 80 K max 96 channels - 40 K max 144 channels - 27 K max
Acquisition Memory	4 MB (expandable to 16 MB)
System Memory	4 MB (expandable to 32 MB)
Time Standard Accuracy	0.005%
To to To accuracy	$\pm 31.25 \mu$ sec
Delay Time	0 - 99 ms
Pre-Trigger	512 samples
Preamplifier	
Input Protection	Diode/filter
Gain Constants	24, 48 dB
Gain Accuracy (channel/channel)	$\pm 1\%$
Input Impedance - Differential	20K/0.005 μ F
- Common Mode	5K/0.02 μ F
Common Mode Rejection Ratio	80 dB
Maximum Input Signal	22 mVRMS @ 48 dB K-gain 350 mVRMS @ 24 dB K-gain

Equivalent Input Noise

@ 2 ms Sample Interval	0.1 μ VRMS @ 48 dB K-Gain
(500 Ohms Source, 125 Hz)	0.7 μ VRMS @ 24 dB K-Gain
DC Offset	0.1 μ V @ 48 dB K-Gain 0.15 μ V @ 24 dB K-Gain

Frequency Response

Alias Filter (High-Cut)	
Analog	4.7 KHz @ 12 dB/octave
Digital	4K, 2K, 1K, 500, 250, 125 and 62.5 Hz @ 96 dB/octave

Low Cut Filter

Analog (default)	3 Hz @ 6 dB/octave
Digital - frequency selectable	3 Hz or 20-240 Hz in 20 Hz steps Slope: 6 dB/octave
Digital Filtering (Post Acquisition Filtering)	Low-pass, High-pass, Band-pass, Multiple notch

A/D Resolution

@ $f_c = 62.5 - 2000$ Hz	24 bits
@ $f_c = 4000$ Hz	18 bits
Dynamic Resolution (Measured)	108 dB @ 48 dB K-Gain
@ $f_c = 3 - 125$ Hz	114 dB @ 24 dB K-Gain
System Dynamic Range (Measured)	132 dB
Crossfeed Isolation	100 dB
Total Harmonic Distortion	0.005%

DATA DISPLAY & RECORDING

Display (PC-AT compatible)	640 x 480 - EL VGA (Electro-luminescent)
Mass Storage (Basic Systems)	
Floppy Disk (3.5 inch)	1.44 MB
Hard Disk	120 MB
Recording Formats	SEG-1, SEG-2, SEG-0
Mass Storage (Optional)	
Hard Disk Expansion	240/480 MB
Tape Back-up System	120 MB
DAT Tape Storage System	2 Gigabytes
External Tape Drives	
3480 Format Tape Cartridge	
1600/6250 bpi 9-Track Tape Drive	

PERIPHERAL INTERFACE

RS-232C, Centronics, VGA Analog,
Keyboard, SCSI (Optional)

PRINTERS

Internal Thermal Printer	
Resolution	203 dots/in (8 dots/mm)
Plot Width	4.1 in (104 mm)
Paper Width	4.4 in (112 mm)
Plot Speed	1 in/sec (25 mm/sec)
External Thermal Plotter (GS-608P)	
Resolution	203 dots/in (8 dots/mm)
Plot Width	8.3 in (211 mm)
Paper Size	8.5 in (216 mm)
Plot Speed	1 in/sec (25 mm/sec)

SYSTEM TESTS

Power-Up Diagnostic	Functional
System Diagnostic	Board Level
Performance	Equivalent input noise: Preamp gain accuracy: Impulse response Crossfeed isolation: Total harmonic distortion Geophone test

PHYSICAL

Master Control Unit	
Dimensions	17 x 19 x 11.5 inches (43.2 x 48.3 x 29.2 cm)
Weight	55 lbs. (25 Kg)
Expansion Module	
Dimensions	8 x 19 x 9 inches (20.3 x 48.3 x 22.9 cm)
Weight	15 lbs. (6.8 Kg)
Thermal Plotter (GS-608P)	
Dimensions	13.5 x 14 x 9 inches (34.3 x 35.6 x 22.9 cm)
Weight	24 lbs (10.9 Kg)

ENVIRONMENTAL

Temperature	
Operating	0°C to 50°C
Storage	-40°C to 85°C
Humidity	
Operating	5-95% non-condensing
Storage/Transfer	Waterproof

POWER REQUIREMENTS

Master Controller	12 VDC
Standby	6 A type
Operating (with 608P plotter)	17 Amp

Specifications are subject to change without notice

DAS-1 PERFORMANCE

The 24 bit Sigma Delta technology increases system dynamic range by 18 dB and instantaneous dynamic range by 24 dB to give superior digital recording resolution. DAS-1 exceeds the performance of conventional floating point amplifier systems with lower instrument noise, higher instantaneous dynamic range, and lower total harmonic distortion (0.005%). With higher instantaneous dynamic range, the full amplitude spectrum is recorded; thus increasing the resolution of the seismic data. Since the system noise and distortion is lower, important high frequency signals that were previously lost in high amplitude low frequency noise can now be recorded without using analog low-cut filters. Unwanted signals are removed without adding unnecessary distortion by using high precision digital filters after A/D conversion. Compared to conventional IFP systems, DAS-1 produces a true digital representation of the seismic signal.

OYO GEOSPACE

USA: OYO GEOSPACE INSTRUMENTS, INC.
9777 West Gulf Bank #10 • Houston TX 77040 U.S.A
PH 713-937-5800 • 1-800-824-2319
FAX 713-937-1161 • TLX 790433

CAN: OYO GEOSPACE CANADA, INC.
2735-37th Avenue NE • Calgary, AB T1Y-5R8, Canada
PH 403-250-9600 • FAX 403-250-9643
TLX 03-821172

UK: OYO UK LTD.
F3 Bramingham Business Park, Enterprise Way
Luton, Bedfordshire, LU3 4BU, England
PH 0582-573980 • FAX 0582-574945 • TLX 825522

JAPAN: OYO CORPORATION Instruments Division
2-19 Daitakubo 2-chome • Urawa, Saitama 336, Japan
PH • 048-882-5371 • FAX 048-885-6424

DAS-1 24 BIT SYSTEM

PORTABLE IFP SYSTEM

DAS-1 NOISE SPECTRUM

GS-40D

ROTATING COIL GEOPHONE

FREQUENCY RESPONSE CURVES

Specifications are subject to change without notice.

OYO GEOSPAC

7334 N. Gessner • Houston, TX 77040 U.S.A.
TEL (713) 939-9700 • TLX 790433 • FAX (713) 937-8262
Bay R, 1350 42nd Ave. S.E. • Calgary, Alberta T2G-4V6, Canada
TEL (403) 287-3050 • TLX 03-821172 • FAX (403) 243-6220

APPENDIX F
DATA FUSION TOOL:
USER'S DESCRIPTION

Contents

1	Nonlinear Least Squares Estimates	1
1.1	Parameters to be Estimated	2
1.2	Grid Orientation and Spacing	4
1.3	Pre-Screening Measurements	5
1.4	Automatic Data Editing	6
1.5	Forward Models	7
1.6	Spatial Continuity	8
1.6.1	Spatial Correlation Distance and Variance of Random Parameters . .	9
1.6.2	Polynomial Coefficients and Order	9
1.7	Pinchouts and Lenses	10
1.8	Execution Algorithms	10
1.8.1	Nonlinear Least Squares and the SRIS Algorithm	10
1.8.2	Householder Transformation	17
1.8.3	Trust Region Algorithm	17
1.8.4	Constrained Step Halving Optimization	22
1.9	Post-Execution Validation	24

1 Nonlinear Least Squares Estimates

The Data Fusion System (DFW) is a tool that allows the user to process geological, geophysical and hydrological data from different types of invasive, non-invasive, and remote sensors to optimally estimate geology and hydrogeological parameters. For example, electromagnetic seismic and borehole data could be combined to optimally estimate a layered earth model which could then be input to flow and transport simulation to predict where contaminants are migrating. The Data Fusion System is not just bringing together the different types of data, but the different types of data are being mathematically weighted according to their accuracy and statistically combined in order to yield the optimal parameter estimates. The parameter estimates are optimal in the sense of being the solution to a nonlinear least squares problem that incorporates data models, first principles models, and spatial continuity models.

The nonlinear least squares problem arises most commonly from data fitting applications where one is attempting to fit measured data z_i , $i = 1, \dots, m$, with a data model $h(s)$ that is nonlinear in s (an n -vector of parameters). The differences between the actual measurements and what the model predicts are called the residuals, i.e., the residuals are

$$r_i(s) = z_i - h_i(s)$$

where $h_i(s)$ is the data model prediction for the specified location and time. For this problem, a useful cost function is the weighted sum of squares of the residuals

$$C(s) = r(s)^T W^{-1} r(s) = \sum_{i=1}^m \left(\frac{r_i(s)}{w_i} \right)^2$$

where the m -vector $r(s)$ has $r_i(s)$ as its i^{th} component, and W is a diagonal weighting matrix with diagonal elements w_i^2 . The weightings w_i are generally chosen to be the standard deviation of the measurement noise. Then the nonlinear least squares problem is

$$\text{minimize } C(s)$$

where the minimization is taken over all n -vectors s . The parameters s that minimize $C(s)$ fit the measurements as closely as possible in the sense that the weighted sum of the squares of the residuals is minimized. This example is similar to the problem we are solving, except that instead of just fitting the measurements, the parameters also fit the pseudomeasurements that account for spatial continuity and first principles models and fit the prior information on initial conditions.

One method for computing the least squares estimates for a nonlinear model is to use a series of linear approximations to the problem. Suppose we have a current state estimate, s_k . Let $r_k^L(s)$ be the linearization of the residual vector $r(s)$ around the point s_k . Then a logical way to use the linearization to solve the nonlinear least squares problem, given s_k , is to choose the next iterate s_{k+1} as the solution to the linear least squares problem

$$\text{minimize } (r_k^L(s))^T W^1 r_k^L(s). \quad (1.1)$$

This approach of solving a nonlinear problem by a series of linear least squares problems is called the Gauss-Newton method. Our implementation of it uses the Square Root Information Smoother (SRIS) algorithm to solve (1.1). The main point to remember here is that the goal of the Data Fusion System is to compute estimates of the desired parameters which minimize a quadratic cost function (i.e., least squares).

1.1 Parameters to be Estimated

When defining the problem to be solved by the Data Fusion System, it is first necessary to determine the parameters to be estimated for a site model. In most cases, selection of parameters to be estimated (states) is a relatively easy task. When processing electromagnetic (EM) data (e.g. time-domain, frequency domain, direct resistivity, etc.), it is necessary to estimate the layer widths and resistivities since these are the parameters required by forward models such as INTERPEX models used in the DFW. Processing of seismic reflection and refraction data requires estimation of layer widths and seismic velocities (either compression or shear). We will eventually use alternate forward EM and seismic models which are gridded instead of layered. Well measurements are either assumed to be direct measurements of a property within a layer or of a continuous rather than layered property. Hydrological flow models use hydraulic head, hydraulic conductivity, sources/sinks and recharge as states, while contaminant transport models will also include contaminant concentrations and sources. Hydrological modeling commonly uses either full 3D gridded models or quasi-3D where quasi-3D assumes a layer cake structure but leakage between layers is included.

For successful fusion of different data types in layered earth models, it is necessary that some layers used in different measurement forward models be common to more than one data type either directly or through spatial continuity. Consider the example shown in Figure 1.1 for seismic refraction and time-domain EM (TDEM). The first seismic velocity, v_1 , is assumed to apply to both layers 1 and 2, while v_2 and v_3 are associated with layers 3 and 4 respectively. The first and third resistivities correspond to layers 1 and 4 respectively, while R_2 corresponds to layers 2 and 3. Thus information about the depth to the bottom of layer 2 is obtained from the seismic data, and the electromagnetic data contains information about the depth to the bottom of layer 1. However both data types will contribute information about the depth to the bottom of layer 3.

Ideally it would be desirable to estimate fundamental properties of the earth in common between seismic velocity, resistivity, hydraulic conductivity and possibly other properties. This will require future work to incorporate fundamental property models.

Another issue related to selection of states is deletion of unobservable parameters. States may be unobservable because the available data is not a function of a particular state, or because the measurement data is a function of only a single combination of states. In Figure 1.1, the seismic data is a function of the layer 2 depth but not layer 1 depth (since no change in velocity occurs between the layers). Thus, if the seismic data was the only

Figure 1.1: Example of Different Data Types Using Common Layers

data type available, layer 1 depth would be unobservable and should be deleted from the estimated states. As an example of the second case, consider Figure 1.2: a thin clay layer is imbedded between two other layers which have much greater resistivity than the clay. If TDEM data were being used to estimate the depth and resistivities of the layers, the emf received at the surface would be a function of the resistance which is the ratio of resistivity to thickness. That is, one can not separate the layer width from the resistivity if the layer is thin. In this case it would be best to set the layer resistivity (or thickness) to a reasonable value and only estimate thickness (respectively, resistivity).

Figure 1.2: Example of a Thin Layer

The user may not know in advance that a particular combination of parameters is unobservable. Excessive nonlinear iterations are a symptom of observability problems. One can examine the state correlation coefficients printed at the end of the fusion run. Correlation co-

efficients greater in absolute value than 0.95 indicate potential problems, and values greater than 0.999 are a big concern. In this case, one of the states should be deleted. However, it must be realized that the data fusion is only estimating the observable linear combination of parameters, and any errors in the assumed value of the fixed parameter will cause similar errors in the estimated value of the associated parameters. Hopefully this approach will result in a good solution for those states which are largely uncorrelated with the problem states.

1.2 Grid Orientation and Spacing

Estimation is accomplished by discretizing space so it is necessary to define the positions at which the parameters will be estimated. This is accomplished by specifying a grid that covers the area under consideration. The parameters (states) will be estimated at the grid intersections (which are treated as the centers of grid cells in hydrological modeling). The fusion workstation allows for the grid to be one-dimensional (a line with subdivisions), two-dimensional (a flat grid or a vertical grid), or three-dimensional (a six-sided polyhedron with subdivisions in all three dimensions). The grid intersections are denoted by (x_i) , (x_i, y_j) , or (x_i, y_j, z_k) in the one, two, and three-dimensional cases respectively, and s_i , s_{ij} , and s_{ijk} denote the state at those positions. To define the grid for a given dimension, it is necessary to specify the grid orientation and grid spacing.

There are two considerations in selecting the orientation of the grid: fusion computational burden and anisotropy. The fusion execution time and storage will be linearly proportional to the number of grid column states which must be carried in memory as will be apparent when the algorithms are described. For local measurements (e.g. TDEM, reflection, wells, etc), the number of states carried is the number of states in all adjacent grid columns. However, for extended measurements such as seismic refraction, the number of grid columns which must be carried in memory is equal to the maximum number of grid columns which are crossed by any individual source-receiver combination, plus adjacent columns. Thus it is desirable to align all the refraction lines parallel to each other and to orient the grid so that the columns are parallel to the refraction lines. Making the rows shorter than the columns will also reduce computations, but this may conflict with other requirements which are more important.

The other factor of interest is anisotropy of the hydrogeology, i.e., changes in spatial continuity with direction. The fusion workstation allows for different correlation distances for spatial continuity in different directions. It thus may be desirable to orient the grid consistent with the anisotropy.

There are at least two conflicting considerations in selecting the grid spacing: computation time and model accuracy. If the grid spacing is halved in one dimension, the number of states will double, and the fusion time and storage will at least quadruple! On the other hand, large grid spacing is undesirable due to errors in modeling. The software interpolates between parameters at the nearest grid locations when computing a forward model at a given measurement location. We are currently using linear interpolation, but other interpolations schemes that use just the adjacent points are possible. Modeling errors can result if the

linear interpolation between grid nodes is not accurate. Further finite difference or finite element representations of first principles models can be inaccurate if grid spacing is too large.

Variable grid spacing and slightly deformed grids are commonly used in hydrological modeling where areas near point sources/ sinks, near discontinuities or around contaminant migration will be modeled with smaller spacing than at the boundaries. However, the ratios between successive grid spacings must not be too large. For example the finite difference equations are only second-order accurate for constant spacings. It has been found that spacing ratios less than 1.5 give acceptable results.

Another consideration is the size of the measurement footprint. For example, TDEM loops can be 30 meters on a side and the effective current loop expands as it propagates downward. Since the measurement is averaging the soil properties over this area, little will be gained when using grid spacings smaller than about 50 meters if TDEM measurements are the primary data type. A good rule-of-thumb is to use grid spacings which keep the total number of states less than 10,000 for a Silicon Graphics R 4000 class of computer, but this number is increasing as our algorithms and computers get faster.

1.3 Pre-Screening Measurements

Now that we know what parameters need to be estimated, and have defined the grid, we next prepare the measurements for processing. Anyone who has had tried to solve a nonlinear estimation problem using data types which do not directly measure the states knows the importance of pre-editing the measurements. Anomalous measurements can cause anomalous estimates and convergence problems. Although the fusion software has automated data editing capabilities, it is better to edit anomalous data before fusion to avoid wasting time looking for anomalies that might have easily been found in the original data. The importance of pre-editing the measurements cannot be over emphasized.

We strongly recommend plotting all the data in ways that will allow anomalous measurements to be identified. For example, when processing sounding data taken along a line, it is very informative to plot several closely-spaced soundings on the same plot. Since the data was taken in nearby locations, the measurement plots should almost overlay. This is particularly useful with TDEM data, and with refraction data. Any individual time gates or geophone times which do not fit the general pattern should be given special attention. Also, there may be entire data sets which appear to be biased or otherwise appear invalid.

This process can be automated somewhat by fitting the data from single data sets using low-order polynomials. Measurement residuals which exceed a user-supplied multiple of measurement noise standard deviation can be flagged for possible editing. Capabilities for plotting and fitting raw measurement data are included in the current fusion software.

Additional insight can be obtained by inverting data at a single location along a survey line, or by inverting all the data along a survey line using a polynomial model for the spatial variation in states along the line. For example, the TDEM measurements can be inverted to compute the depths and resistivities of layers. These estimates can be compared with

well data from nearby wells. If this is satisfactory, this can be done for all the data along a line: first for individual data types and then for all the data if the different types appear consistent.

This staged approach to data fusion may appear to require more effort than simply mixing all the data in one big fusion run, but it will probably save much effort and produce a better result. The single-location and line inversions execute quickly (seconds or minutes) and thus many runs can be made in a short time span. By comparing estimates from individual data types, it may be discovered that the different types are not measuring the same layers. Once the user has confidence in the data, one should have confidence in the final fusion result, or at least know where to start looking if problems are encountered. Furthermore, much of the above processing is current practice for conventional data interpretation.

1.4 Automatic Data Editing

As noted previously, the fusion software has provisions for automatically editing measurements which do not seem to fit the fusion models. However, the user should not depend too heavily on this capability because it is better to remove anomalous data before fusion.

Anomalous data can be detected through isolated excessively large residuals. A residual that is 3 to 5 times larger in magnitude than its standard deviation is statistically unlikely. The measurement that produced the residual should be considered for deletion. Data that are automatically deleted in a fusion run should always be carefully examined to be sure the data really are just outliers and not symptoms of some other problem. Although fusion can compute the residual standard deviation, for simplicity we use a conservative upper bound provided by the measurement noise standard deviation. This is an upper bound because the parameters generally fit some noise so the residual is statistically smaller than the measurement noise.

Sometimes it is desired to consider a group of m measurements together for editing, for example, from a single sounding. The logical extension of the above is to use the root mean square α of the residuals normalized by the measurement noise standard deviation and delete the measurements if α is greater than 3 to 5 where

$$\alpha = \sqrt{\frac{1}{m} \sum_{i=1}^m \frac{(z_i - h_i(s))^2}{\sigma_{zi}^2}} \quad (1.2)$$

However, initial iterations of fusion will have large residuals just because fusion has not converged. But, anomalous data can be damaging to convergence so the editing threshold on α must be inflated. An overall root mean square RMS is computed for all the data and used to scale the threshold.

The fusion editing logic rejects any measurements in iteration ℓ for which $\alpha > \beta \text{ RMS}_{\ell-1}$ where $\text{RMS}_{\ell-1}$ is the residual RMS from the previous iteration and β is a user supplied editing threshold. Values of β from 3 to 5 are typical. Since RMS_0 is not available on the first iteration, the user must input this. We suggest $\text{RMS}_0 = 10$ on the first iteration unless

the run is a restart (use the RMS from the last iteration). This editing logic has been used in many applications, including many years in satellite orbit determination programs.

1.5 Forward Models

Recall from Section 1 that we are trying to minimize a cost function involving the residuals, which are the differences between the actual measurements and what is predicted by the model. That is, the residuals are

$$r_i(s) = z_i - h_i(s),$$

where s is the state vector which includes the properties of interest, $h(s)$ is the data model which predicts the measurements based on the state, and the z_i 's are the actual measurements. As we will see, minimization of the cost function at each iteration requires evaluation of the differences $z_i - h_i(s)$, and the partial derivatives of $h(s)$ with respect to the state components. So for each type of data (e.g., electromagnetic, seismic), we need to have the mathematical model $h(s)$ that relates the properties of interest (the parameters) to the measurements. If we do not have an analytical expression for the model, then we at least need to have the differences $z_i - h_i(s)$, and we need to be able to numerically evaluate the partial derivatives.

The models that map the properties (parameter estimates) to the expected measurements are called the forward models, that is, $z = h(s) + v$, where v is noise. First principles models such as flow and transport have the same form where z includes source/sink terms and v is statistical model error. First principles models are included in fusion just as if they were data models as pseudomeasurements. In the data fusion workstation, the forward models that we use are usually obtained through commercial software products.

Consider for example forward models for the time domain electromagnetic (TDEM) data. TDEM data is collected with a system that consists of a transmitter and a receiver. In environmental applications, the transmitter configuration often used is a square loop of wire laid on the ground surface. A multi-turn air coil receiver is placed in the center of the loop. The current waveform driven through the square loop consists of equal intervals of time-on and time-off. These intervals (and inversely the repetition frequencies) are usually different for each sweep. The current driven through the transmitter loop creates a primary magnetic field which results in electromagnetic induction during turn-off time, which in turn results in an eddy current flow in the subsurface. The intensity of these currents at a certain time and depth depends on the ground resistivity. The receiver measures the emf (electromotive force) due to the secondary magnetic field caused by these ground eddy currents.

When processing TDEM data, the geophysical properties that we want to estimate are the widths of the subsurface layers and the resistivity for each layer. At each iteration of the estimation procedure, we have initial estimates of the widths and resistivities. In order to update these estimates, we need to evaluate the $z_i - h_i(s)$ terms, and the partial derivatives of $h(s)$.

The software used for TDEM data to compute the predicted measurements $h(s)$ for given parameter estimates was obtained from INTERPEX. This software requires input of

the number of layers, the values for the widths of the layers and the resistivities of the layers, and the time gates at which the actual measurements were taken. The forward model maps the layer widths and resistivities to the expected emf values for the times that were input. These are used to evaluate the differences between the measured emf and the predicted emf, i.e., the residual $z_i - h_i(s)$. Also the partial derivatives can be computed numerically by perturbing each input parameter in s . (The user also inputs information about how the TDEM data was gathered such as the size of the square transmitter loop, the elevation if known, station coordinates, preamp gain, additional gain on applicable channels, run-on compensation, turn-on time constant, and for each sweep – the frequency, current, gain setting, receiver coil area, and ramp (turn-off) time.)

1.6 Spatial Continuity

The concept of spatial continuity is motivated by two desires. First, as discussed above, the data fusion algorithms work by superimposing a grid on the site under investigation. Actual measurements may occur at points other than the grid intersections, so we must express a measurement as a combination of states at the surrounding grid locations. Second, in estimating the parameters of interest, we use the prior knowledge that heterogeneous parameters can vary only so much between nearby locations. For example, in estimating the depth of a certain layer, the answer that we get at one point should not be substantially different than the answer we get at a nearby point. We include a method of incorporating spatial continuity, and in doing so, we also minimize the effect of measurement errors.

In the algorithms, spatial continuity is imposed by modeling state variables in the fusion model as the sum of two components: a systematic trend, and a Markov Random Field (MRF) that provides a correlated variation about the trend. The trend is modeled as a spatial polynomial, while the MRF is modeled as the output of a n autoregressive process, e.g., for a two-dimensional model using a first order autoregression and second order polynomial:

$$s_{ij} = p_{ij} + v_{ij} \quad (1.3)$$

$$p_{ij} = c_0 + c_1 x_i + c_2 y_j + c_3 x_i^2 + c_4 y_j^2 + c_5 x_i y_j \quad (1.4)$$

$$v_{ij} = a(v_{i+1,j} + v_{i-1,j}) + b(v_{i,j+1} + v_{i,j-1}) + w_{ij} \quad (1.5)$$

where

- s_{ij} is the state at position x_i, y_j ,
- p_{ij} is the trend evaluated at position x_i, y_j ,
- v_{ij} is the MRF at x_i, y_j ,
- c are the polynomial coefficients
- a, b are the autoregressive coefficients, and
- w_{ij} is white noise input with zero mean and variance σ_2 .

If $a = b$, then this is isotropic; otherwise it is anisotropic.

The MRF (1.5) has the same form as a measurement and is treated as a pseudomeasurement in fusion. The MRF is incorporated in the algorithms by generating residuals $r_{i,j}$ using the pseudomeasurements, e.g. for the two-dimensional case,

$$r_{i,j} = v_{ij} - a(v_{i+1,j} + v_{i-1,j}) - b(v_{i,j+1} + v_{i,j-1})$$

and then processing these pseudomeasurements in the same way as the measurements are processed.

The autoregressive coefficients (a and b) indicate the extent of the surrounding states influence on the value at a given state. As the grid points are taken closer together, the larger the coefficients should be. In the fusion workstation, the autoregressive coefficients are computed internally from other inputs.

1.6.1 Spatial Correlation Distance and Variance of Random Parameters

There are two inputs relating to the spatial continuity model that the user supplies. The spatial continuity model has a spatial autocorrelation function which is exponential in three dimensions, but deviates from exponential in one or two dimensions (see Figure 1.3). In two dimensions, except near zero, we have

$$E[(v(x_1, y_1) - \bar{v}(x_1, y_1))(v(x_2, y_2) - \bar{v}(x_2, y_2))] \approx \sigma^2 \exp \left\{ \frac{-\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}}{\tau} \right\}$$

where the average value $\bar{v}$ is obtained from the polynomial (1.4). The user needs to input τ which is the distance at which the correlation coefficient falls to $1/e=0.37$ (possibly different for different directions), and input σ which is the standard deviation of the output of (1.5) as defined in the preceding section.

Ideally the correlation distance τ and variance σ^2 should be chosen from statistical analysis of data. However, the data may not be initially available or may not be in a form suitable for easy interpretation. Thus the scientist's and engineers prior insight and experience must often be used to select initial τ and σ^2 . As additional data are acquired τ and σ^2 should be updated based on residuals. The statistical likelihood function is computed by fusion. Statistically optimal maximum likelihood parameter estimates are found by selecting τ and σ^2 to maximize the likelihood.

Grid spacing should be selected to accommodate the correlation distance. It is expected that the ratio of correlation distance to grid spacing should be greater than 1.0. Note that the fusion software internally computes the relationship between the output variance, the correlation distances, the autoregressive coefficients, and the white noise input variance.

1.6.2 Polynomial Coefficients and Order

The user must select the order of the polynomial trend, and initialize the polynomial coefficients. Since the purpose of the polynomial is to model the spatial trend in the parameters, it should usually be low order. Generally it will not be necessary to use more than a second-order polynomial (6 coefficients), and a linear model will be adequate in many situations. If the region of interest is very small, or the parameter is known to be relatively constant, a single coefficient (zero order) may be acceptable. It is not desirable to use high order polynomials because of the well-known problems associated with extrapolation of data.

Figure 1.3: Continuous Autocorrelation Functions in One, Two, and Three Dimensions

1.7 Pinchouts and Lenses

The current version of the data fusion software has the capability to constrain parameters at specified locations to fixed values, and to constrain parameter values in different layers to be equal. This capability can be used to hypothesize pinchouts or lenses. As shown in Figure 1.4, layer 2 pinches-out in the region from x_1 to x_2 . This would be done by adding pseudomeasurements of layer 2 width equal to zero at several nodes between x_1 and x_2 . The spacing between these pseudomeasurements should be approximately equal to the spatial correlation distance τ in that direction. It may also help the optimization to set other layer 2 parameters to fixed values.

Geologic lenses are another case in which this capability may be useful (Figure 1.5). If the region of the lense is small compared to the total region modeled, it would be better to specify the region in which the layer does exist rather than the region in which it does not. This can be done by user input.

1.8 Execution Algorithms

1.8.1 Nonlinear Least Squares and the SRIS Algorithm

Recall from Section 1 that the nonlinear least squares problem is to minimize the cost function $C(s)$. Often the cost function is the weighted sum of squares of residuals $r_i(s)$, where the

Figure 1.4: Example of a Pinchout

Figure 1.5: Example of a Lense

residuals are the differences between the measurements z_i and what the data model $h(s)$ predicts:

$$r_i(s) = z_i - h_i(s), \quad i = 1, \dots, m.$$

In our case, we also have a prior estimate of the initial condition s_p , estimate error covariance P , and residuals for the pseudomeasurements. Then the cost function includes the prior information, the weighted sum of squares of the measurement residuals, and the weighted sum of squares of the pseudomeasurement residuals $r_p(s)$ where W_p is the spatial continuity and first principles model error noise covariance matrix:

$$C(s) = (s - s_p)^T P^{-1} (s - s_p) + r(s)^T W^{-1} r(s) + r_p(s)^T W_p^{-1} r_p(s)$$

and the nonlinear least squares problem becomes

$$\text{minimize } C(s) \tag{1.6}$$

where the minimization is taken over all n -vectors s .

The methods used to solve nonlinear least squares problems are iterative procedures, that is, the update of the parameter estimates at each iteration is a function of the current parameter estimates, and the procedure for updating continues until some criterion of convergence is met. The Gauss-Newton method that we use for solving the nonlinear least squares problem is an iterative procedure in which the update at each iteration is the linearized correction to the current state vector, i.e., the problem is linearized at each iteration by using the linearization of the residual function around the current state estimate. Hence at each iteration of the nonlinear least squares problem, we are solving a linear least squares problem.

For simplicity in the following equations, we drop the distinction between the residuals for the measurements and the pseudomeasurements, and we assume that the residuals have been normalized by $W^{-1/2}$ or $W_p^{-1/2}$. If s_k denotes the current estimate of the state at the k^{th} iteration, then the linearization of the residual function around s_k is

$$r_k^L(s) = r(s_k) - J(s_k)(s - s_k) \tag{1.7}$$

where $J(s_k)$ is the Jacobian matrix of $h(s)$ at s_k , an $m \times n$ matrix of partial derivatives in which $J(s)_{ij} = \partial h_i(s) / \partial s_j = -\partial r_i(s) / \partial s_j$. Then substituting this linearization for the residuals in the cost function, we get the linear least squares problem which is solved at each iteration of the nonlinear least squares problem.

$$\text{minimize } [(s - s_p)^T P^{-1} (s - s_p) + (r_k^L(s))^T r_k^L(s)] \tag{1.8}$$

The algorithm that we use to solve the linear least squares problem is the Square Root Information Smoother or SRIS. As the name implies, this algorithm makes use of the square root of the information matrix. The information matrix is the inverse of the estimate error covariance matrix. The prior error covariance P is factored into square root information form

$$P = R^{-1}(R^{-1})^T$$

where the square root information R is upper triangular. Then using this prior and combining (1.7) and (1.8), our goal becomes

$$\text{minimize } [(s - s_p)^T R^T R(s - s_p) + [r(s_k) - J(s_k)(s - s_k)]^T [r(s_k) - J(s_k)(s - s_k)]]$$

or equivalently,

$$\text{minimize } \left\| \begin{bmatrix} R & R(s_p - s_k) \\ J(s_k) & r(s_k) \end{bmatrix} \begin{bmatrix} s - s_k \\ -I \end{bmatrix} \right\|^2 \quad (1.9)$$

The norm of a vector v is $\|v\| = \sqrt{v^T v}$. The norm in (1.9) is unchanged if we multiply the matrices by an orthogonal matrix, H , so we choose H to be the Householder transformation (see Section 1.8.2) that makes the left-hand matrix in (1.9) upper triangular. So now we are minimizing

$$\begin{aligned} \left\| H \begin{bmatrix} R & R(s_p - s_k) \\ J(s_k) & r(s_k) \end{bmatrix} \begin{bmatrix} s - s_k \\ -I \end{bmatrix} \right\|^2 &= \left\| \begin{bmatrix} R_{k+1} & \alpha_{k+1} \\ 0 & e_{k+1} \end{bmatrix} \begin{bmatrix} s - s_k \\ -I \end{bmatrix} \right\|^2 \\ &= \|R_{k+1}(s - s_k) - \alpha_{k+1}\|^2 + \|e_{k+1}\|^2 \end{aligned} \quad (1.10)$$

which is clearly minimized for

$$s - s_k = R_{k+1}^{-1} \alpha_{k+1}. \quad (1.11)$$

This is the solution of the linear least squares problem.

We will explain further below how the measurements are processed, and how (1.11) is actually solved using back substitution (backward sweep) through the information matrix, but for now we briefly describe the rest of the procedure. Once we have a value for $(s - s_k)$ from (1.11), we test if any physical constraints are violated by $s_{k+1} = s_k + (s - s_k)$, and we test if the cost function evaluated at s_{k+1} has not been significantly reduced compared to the cost function evaluated at s_k . (An increase in the cost function could happen if the step is too large since the linearization that we used for the residual becomes less valid as we move away from s_k .) If either of these is the case, then the step is altered by the trust region algorithm or by the step-halving optimization algorithm (both discussed below) until $(s - s_k)$ is found for which $s_{k+1} = s_k + (s - s_k)$ satisfies the above tests. This completes one iteration of the nonlinear least squares problem. To continue, the linearization of the residual function at s_{k+1} is evaluated and the entire procedure is repeated until a convergence criterion is met.

Measurement Processing In order to understand how the measurements are processed, we first need to understand the structure of the state and how the model states are conceptually ordered. We ultimately have to think of stacking the states in a state vector. We give a horizontal two-dimensional example; the one-dimensional and three-dimensional cases are straightforward extensions. The idea is that a two-dimensional grid can be thought of one-dimensionally by stacking the columns of the grid. Then for each grid cell, all the state variables local to that grid cell can be strung out in a vector. Thus if we have a grid with M rows and N columns, the elements of the state vector can be written as

$$s = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix}$$

and each component s_j has the substructure

$$s_j = \begin{bmatrix} s_{1j} \\ s_{2j} \\ \vdots \\ s_{Mj} \end{bmatrix}$$

where each component s_{ij} is the vector of the local variables (e.g., depth, resistivity) corresponding to grid cell location (i, j) . With this structure, not all adjacent grid cells will have all corresponding variables adjacent to each other in the state vector. However, for any cell, the variables for that cell and all adjacent cells will be within three column lengths in the state vector. If all measurements and pseudomeasurements can be processed using only parameter values at adjacent nodes, then only three grid columns of states need to be carried in memory at any one time; this fact is used to reduce the storage of the SRIS algorithm. (For extended measurements such as seismic refraction, the number of grid columns which must be carried in memory is equal to two more than the maximum number of grid columns which are crossed by any individual source-receiver combination.)

To run the SRIS algorithm, the matrix that is transformed to upper triangular in equation (1.9) is

$$\begin{bmatrix} R & R(s_p - s_k) \\ J(s_k) & r(s_k) \end{bmatrix}. \quad (1.12)$$

Because of the large size of this matrix, transforming it to upper triangular is a problem both in terms of computation and storage. However, because of the state structure described above and the limited "extent" of measurements or pseudomeasurements, the amount of storage and computation can be greatly reduced. We will show that only part of $J(s_k)$ needs to be stored at any time while processing measurements.

There are two general types of measurements to be processed. One is the physical measurement. This type of measurement should have a partial derivative matrix with only a few local nonzero elements since a measurement at any location can be modeled as a nonlinear function of an interpolation of the surrounding grid cell states (except for extended

measurements). Thus only these states would be involved in the partial derivative matrix for this measurement.

The other type of measurement is the pseudomeasurement for the purpose of enforcing spatial continuity and incorporating first principles models. Recall from Section 1.6 that all spatially continuous state variables are modeled as the sum of two components: a systematic trend, and a random but correlated variation about the trend. The trend is modeled as a spatial polynomial, while the random variation is modeled as the output of an MRF, e.g., in the two-dimensional case:

$$s_{ij} = p_{ij} + v_{ij} \quad (1.13)$$

$$p_{ij} = c_0 + c_1 x_i + c_2 y_j + c_3 x_i^2 + c_4 y_j^2 + c_5 x_i y_j \quad (1.14)$$

$$v_{ij} = a(v_{i+1,j} + v_{i-1,j}) + b(v_{i,j+1} + v_{i,j-1}) + w_{ij}. \quad (1.15)$$

Since $v_{ij} = s_{ij} - p_{ij}$, the partial derivatives for each pseudomeasurement involve five state variables (s_{ij} , $s_{i+1,j}$, $s_{i-1,j}$, $s_{i,j+1}$, $s_{i,j-1}$, which are within three columns in the grid) and all of the polynomial coefficient variables for that state type.

This localness of the measurements and pseudomeasurements allows us to store and operate on only the blocks in the partial derivative matrix corresponding to the adjacent grid columns, the block for the polynomial coefficients, and the corresponding blocks of the square root information matrix R . The total size of the blocks that need to be stored depends on whether we are solving a one-dimensional, two-dimensional, or three-dimensional problem, and on whether extended measurements such as seismic refraction are included. We refer to the number of states that need to be stored as the "bandwidth" of the measurements.

As an example, in the two-dimensional case without extended measurements, the part of the matrix in (1.12) that needs to be stored for the first set of calculations is

$$\left[\begin{array}{cccc|c} * & 0 & 0 & 0 & \\ 0 & * & 0 & 0 & \\ 0 & 0 & * & 0 & \\ 0 & 0 & 0 & * & \\ \hline J_1 & J_2 & J_3 & J_c & r(s_k) \end{array} \right]$$

where the asterisks represent blocks with nonzero elements and J_i (respectively J_c) is the block of the partial derivative matrix containing all elements relating to the i^{th} grid column (respectively, the polynomial coefficients). To continue the calculations, the SRIS algorithm loops through all the grid columns, and for each grid column, processes all measurements and pseudomeasurements for adjacent grid columns. When the grid column index is $i + 1$, the square root information matrix is stored for grid columns i , $i + 1$, and $i + 2$. Pseudomeasurements for grid column $i + 1$ are processed first, followed by all measurements located between grid columns i and $i + 1$. Then the upper rows of the square root information matrix (corresponding to the states for grid column i) are written to disk since these blocks will remain unchanged by subsequent measurements, the remaining rows are shifted to the top, and the next row block of the information matrix is inserted. This process is repeated until all grid measurements and pseudomeasurements have been processed.

Backward Sweep of SRIS Recall that at the k^{th} iteration, we solve for the increment $(s - s_k)$ from (1.11):

$$R_{k+1}(s - s_k) = \alpha_{k+1}.$$

By using the upper triangular structure of the square root information matrix R_{k+1} , $(s - s_k)$ could be computed by direct back-substitution. We use an approach similar to back-substitution that uses retriangularization with Householder transformations.

Continuing the two-dimensional example without extended measurements, we rewrite the increment $\Delta s = s - s_k$ as

$$\Delta s = \begin{bmatrix} \Delta s_1 \\ \Delta s_2 \\ \vdots \\ \Delta s_N \\ \Delta s_c \end{bmatrix}$$

where Δs_i (respectively Δs_c) is the part of the state associated with the i^{th} grid column (respectively, polynomial coefficients). We similarly subdivide the square root information matrix and the vector α . Then after processing the measurements and pseudomeasurements for the last grid column, the reduced square root information matrix that we have stored gives us

$$\begin{bmatrix} R_{N-2,N-2} & R_{N-2,N-1} & R_{N-2,N} & R_{N-2,c} \\ 0 & R_{N-1,N-1} & R_{N-1,N} & R_{N-1,c} \\ 0 & 0 & R_{N,N} & R_{N,c} \\ 0 & 0 & 0 & R_{c,c} \end{bmatrix} \begin{bmatrix} \Delta s_{N-2} \\ \Delta s_{N-1} \\ \Delta s_N \\ \Delta s_c \end{bmatrix} = \begin{bmatrix} \alpha_{N-2} \\ \alpha_{N-1} \\ \alpha_N \\ \alpha_c \end{bmatrix}. \quad (1.16)$$

We will start with this reduced matrix and retrieve rows from the disk as needed, thereby making a backward sweep through the square root information matrix.

To initialize the process, the bottom two row blocks of (1.16) are solved for Δs_N and Δs_c (by back substitution) and the covariance matrix is computed as the inverse of the information. The next step is to rotate the columns of the left-hand matrix, and reorder the components of the vector, to get the equivalent system

$$\begin{bmatrix} R_{N-2,N} & R_{N-2,N-2} & R_{N-2,N-1} & R_{N-2,c} \\ R_{N-1,N} & 0 & R_{N-1,N-1} & R_{N-1,c} \\ R_{N,N} & 0 & 0 & R_{N,c} \\ 0 & 0 & 0 & R_{c,c} \end{bmatrix} \begin{bmatrix} \Delta s_N \\ \Delta s_{N-2} \\ \Delta s_{N-1} \\ \Delta s_c \end{bmatrix} = \begin{bmatrix} \alpha_{N-2} \\ \alpha_{N-1} \\ \alpha_N \\ \alpha_c \end{bmatrix}.$$

This can be retriangularized using a Householder transformation to obtain

$$\begin{bmatrix} R'_{N,N} & R'_{N,N-2} & R'_{N,N-1} & R'_{N,c} \\ 0 & R'_{N-2,N-2} & R'_{N-2,N-1} & R'_{N-2,c} \\ 0 & 0 & R'_{N-1,N-1} & R'_{N-1,c} \\ 0 & 0 & 0 & R_{c,c} \end{bmatrix} \begin{bmatrix} \Delta s_N \\ \Delta s_{N-2} \\ \Delta s_{N-1} \\ \Delta s_c \end{bmatrix} = \begin{bmatrix} \alpha'_N \\ \alpha'_{N-2} \\ \alpha'_{N-1} \\ \alpha_c \end{bmatrix}.$$

We then use the bottom two row blocks to solve for Δs_{N-1} . Now Δs_N has already been solved for, and no other variables depend on it, so we can eliminate Δs_N from the system by

deleting its row block. In its place, we can import from the disk the row block corresponding to Δs_{N-3} , which gives us a system of the same form as (1.16), so we can perform the same column rotation to produce a system that we can solve for Δs_{N-2} . Then we can eliminate the row block corresponding to Δs_{N-1} and in its place import the row block corresponding to Δs_{N-4} , and so forth until we have completely solved for Δs .

1.8.2 Householder Transformation

As we see above, the SRIS algorithm requires transforming matrices to upper triangular without changing the norms, and the workhorse that is used for this is the well-known and numerically stable Householder transformation. We will briefly explain the idea of the Householder transformation. If v is a non-zero n -vector, then an $n \times n$ matrix H of the form

$$H = I - \left(\frac{2}{v^T v} \right) v v^T$$

is called a Householder transformation, and the vector v is called a Householder vector. When a vector x is multiplied by H , it is reflected in the hyperplane perpendicular to the span of v . It is easy to verify that H is orthogonal ($H^T H = I$), so that premultiplying a matrix by H does not change the norm of the matrix.

Householder transformations can be used to zero out selected components of a vector. In particular, suppose we have any non-zero n -vector x , and we want Hx to be a multiple of $e_1 = [1, 0, 0, \dots, 0]^T$. We have

$$Hx = x - \left(\frac{2v^T x}{v^T v} \right) v$$

and thus, Hx being in the span of e_1 implies v is in the span of $\{x, e_1\}$. Setting $v = x + \alpha e_1$ gives

$$Hx = \left(1 - 2 \frac{x^T x + \alpha x_1}{x^T x + 2\alpha x_1 + \alpha^2} \right) x - \left(2\alpha \frac{v^T x}{v^T v} \right) e_1.$$

In order for the coefficient of x to be zero, we need to set $\alpha = \pm \|x\|_2$. Thus,

$$v = x \pm \|x\|_2 e_1 \text{ implies } Hx = \mp \|x\|_2 e_1.$$

A matrix could be made upper triangular by applying this to the first column of the matrix, then applying it to the second column without changing the first element, then to the third column, etc. This has been implemented in a way that avoids multiplication by zeroes.

1.8.3 Trust Region Algorithm

The trust region algorithm ensures that the step Δs that is added to the parameter estimate vector improves the cost function and ensures that the parameter estimates are reasonable (such as resistivities and layer widths that are positive). Trust region algorithms constrain

the parameter estimates to stay within a region that can be "trusted", i.e., within a region of values that are reasonable and within a region where the minimization is logical because the function being minimized is well approximated by a quadratic. When the function is approximately quadratic in a certain region, the step $(s - s_k)$ determined by the Gauss-Newton method leads to a new estimate which is near the minimum in that region. We are using our own trust region algorithm that was designed to take advantage of the quantities available in square root information filtering and designed to readily include the physical constraints.

Unlike most trust region algorithms which use a circle around the current estimate, we use a (multi-dimensional) rectangular trust region based on the fact that the diagonals of the inverse of the information matrix are a measure of the uncertainties in the parameter estimates. Recall from above that the information matrix is the inverse of the error covariance matrix so the standard deviation (square root of the variance) of the i^{th} component of the current estimate, $(s_k)_i$, is

$$\sigma_i = \sqrt{[R_k^{-1}(s_k)R_k^{-T}(s_k)]_{ii}}$$

where R_k is the square root of the information matrix which we used for the SRIS algorithm. It is reasonable to allow the parameters to vary from their current values by some multiple of their uncertainties. That is, we consider trust regions to consist of all vectors s whose components all satisfy

$$(s_k)_i - r\sigma_i \leq s_i \leq (s_k)_i + r\sigma_i$$

where r is called the radius of the trust region. If r is properly adjusted, then the minimum of the quadratic approximation to the cost function inside the trust region should give a reasonable decrease in the true cost function. If it does not, then r must be too big. On the other hand, if r is too small, then our movement away from s_k toward the minimum of the cost function will be very slow. Hence, the logic of the trust region algorithm involves adjusting r to keep it as large as possible, but small enough so that within the trust region the quadratic cost function approximates the true cost function.

Now consider the effect of adding parameter constraints. That is, suppose it is known that the i^{th} component of the parameter vector must be constrained to lie in some interval

$$l_i \leq s_i \leq u_i.$$

Our use of rectangular trust regions allows for a simple modification to take the physical constraints into account. For this, we use as the trust region the set of vectors s such that

$$\hat{l}_i \leq s_i \leq \hat{u}_i$$

where

$$\hat{l}_i = \text{maximum of } (l_i, (s_k)_i - r\sigma_i),$$

$$\hat{u}_i = \text{minimum of } (u_i, (s_k)_i + r\sigma_i).$$

The trust region algorithm can best be understood by referring to Figure 1.6. Starting with the current state estimate s_k , the SRIS computes the initial quantities $R(s_k)$, α_k , and

the cost function $C(s_k)$. The SRIS then computes the step $\Delta s = (s - s_k)$, and the parameter uncertainties σ_i 's. Also, the first time through the loop, the trust region radius r is initialized. The trust region radius and the uncertainties are then used to compute the boundaries of the trust region.

If the increment Δs violates any constraints, a logic vector keeps track of which constraints are active at the current estimate: the logic vector has a component of +1 or +2 if the corresponding state component is at its upper limit where +1 is used for a trust region constraint and +2 for a physical constraint, similarly -1 or -2 if the state component is at its lower limit, and 0 for an inactive constraint. If no constraints are active, then the increment Δs is used. If there are active constraints, then the constrained minimization loop is entered. To understand constrained minimization, consider the state vector partitioned into the constrained and unconstrained components. From this partitioned perspective, the approximate Newton step is the solution to the system of equations

$$\begin{bmatrix} F_c & F_{cu} \\ F_{uc} & F_u \end{bmatrix} \begin{bmatrix} \Delta s_c \\ \Delta s_u \end{bmatrix} = \begin{bmatrix} \nabla C_c \\ \nabla C_u \end{bmatrix}$$

where subscript c is for constrained and u for unconstrained. Solving the second equation in this system

$$F_{uc}\Delta s_c + F_u\Delta s_u = \nabla C_u \quad (1.17)$$

for Δs_u by substituting

$$\Delta s_k = d(s_k)$$

where $d(s_k)$ represents the directed distance from the trust region center s_k to the constraint boundary. This can be shown to be equivalent to solving

$$R_k \Delta s = \alpha_k$$

subject to $\Delta s_c = d(s_c)$. Since R_k is upper triangular, this can be solved using the SRIS (backward sweep method) discussed above, except that the constrained values are substituted for the constrained states as the backward sweep is done.

When the unconstrained step in (1.17) is computed, there is no guarantee that it produces a step contained inside the trust region. So the algorithm checks if any components of the new estimate exceed the trust region boundaries. If so, the constraint flag of the component most exceeding the trust region is set, and the increment Δs is pulled in radially just until all components lie within the trust region. Then the constrained minimization is solved subject to the additional new constraint. We thus continue to add constraints until a solution is found for (1.17) which produces a point within the trust region.

At this point we have a constrained minimization step Δs possibly subject to many constraints, but all of the constraints may not have been necessary. When computing the Gauss-Newton step, a quadratic approximation to the cost function at the point s_k is minimized. The quadratic approximation at s_k is

$$C_F(s) = C(s_k) + \nabla C(s_k)\Delta s + \frac{1}{2}\Delta s^T F(s_k)\Delta s$$

Figure 1.6: Flowchart for the Trust Region Algorithm

where the information matrix $F(s_k)$ is used to approximate the second derivative (Hessian matrix). If the minimum occurs in the interior of the trust region, then the gradient of $C_F(s)$ should be zero at the minimum. If the minimum occurs on the boundary of the trust region, then the gradient at the minimizing point should be orthogonal to the constraint surface and pointing into the interior of the trust region. This last property can be checked by computing the cost function gradient at the potential constrained minimum and checking the components corresponding to the constrained states. At a constrained state, the gradient component should be positive at a lower boundary and negative at an upper boundary in order to be pointing into the trust region interior. If this is not true, the constraint is released by setting the corresponding logic vector component back to zero. If any constraints are released, then the constrained minimum is recomputed with the new set of constraints. If no constraints can be released, then Δs is the constrained minimum step. This completes the constrained minimization loop.

To decide if the potential new estimate $s_{k+1} = s_k + \Delta s$ is acceptable, the acceptance test for s_{k+1} is

$$C(s_{k+1}) \leq C(s_k) + \alpha \nabla C(s_k)^T (s_{k+1} - s_k)$$

where $\alpha = 10^{-4}$ has numerically been shown to work well. This test takes into account that if the gradient at s_k , in the direction toward s_{k+1} , is very negative, then there should be a significant decrease in the cost function value. To perform this test, the SRIS computes $C(s_{k+1})$, but R_{k+1} and α_{k+1} are not computed to minimize unnecessary calculations until a satisfactory step has been found.

If the acceptance test fails, then r is reduced and we try again. The minimum of a one-dimensional quadratic approximation to $C(s)$ is used, where the approximation is in the direction of s_{k+1} , agrees with $C(s_k)$ and $C(s_{k+1})$, and has derivative $\nabla C(s_k)^T (s_{k+1} - s_k)$ at s_k . The minimum is at $s_k + \tau(s_{k+1} - s_k)$ where

$$\tau = \frac{-\nabla C(s_k)^T (s_{k+1} - s_k)}{2[C(s_{k+1}) - C(s_k) - \nabla C(s_k)^T (s_{k+1} - s_k)]}.$$

This minimum is used to reduce r , but should not be followed too blindly. Thus the following method is used to reduce r when the acceptance test fails:

If $0 \leq \tau \leq 0.1$, replace r by $(0.1)r$.

If $0.1 \leq \tau \leq 0.5$, replace r by τr .

If $0.5 \leq \tau$, replace r by $(0.5)r$.

If the trust region reduction is performed 10 times, it is assumed that no satisfactory step will be found and the program terminates.

In the case where s_{k+1} passes the acceptance test, a test for convergence is made. If convergence has been obtained, then the algorithm is terminated. The trust region loop is also terminated with an error code if the maximum number of iterations has been reached.

If convergence has not been obtained, the algorithm adjusts the trust region radius and goes back to the beginning of the trust region loop. The trust region radius is adjusted according to how well the quadratic model of the cost function $C_F(s)$ is approximating the

true cost function $C(s)$ at the new state estimate. If the error in the approximation is relatively small,

$$|C(s_{k+1}) - C_F(s_{k+1})| \leq (0.1)C(s_{k+1})$$

then the trust region radius is doubled. If the quadratic approximation error is relatively large,

$$|C(s_{k+1}) - C_F(s_{k+1})| \geq (0.75)C(s_{k+1})$$

then the trust region radius is halved. If neither of these conditions is met, the trust region radius stays the same for the next iteration.

1.8.4 Constrained Step Halving Optimization

Our out-of-memory version of the trust region algorithm has a tendency to “thrash” on the disk when the trust region radius is small. This can cause excessively long execution times. Although the trust region algorithm is more robust, the constrained step-halving optimization algorithm may be preferred when thrashing occurs. This algorithm is a Gauss-Newton type of method that takes into account physical constraints. Figure 1.7 shows the flowchart for using the step-halving optimization algorithm. Starting with the current state estimate s_k , the SRIS computes the cost function $C(s_k)$, and computes R_k from which the gradient and the information matrix can be computed.

Next the Newton step factor $STEP = \sqrt{\eta}/\text{RMS}$ is computed, where RMS is the residual mean square (the average of the squared residuals) and η is typically 1.5 to 3. The step factor will be used to control the size of the increment Δs , so that when the residual mean square is large, the increment is limited to a small step.

Next the SRIS computes the Gauss-Newton step Δs . If any physical constraints are violated by using Δs for the increment, then the constrained minimum on the surface of unconstrained states is computed. This is performed by the same method as in the trust region algorithm and was discussed in the last section. Now using the constrained step, the Newton step factor is used to limit the step size, by resetting Δs to STEP times the constrained step for the unconstrained states only. If any new constraints have now been violated, these steps are repeated.

Once this loop has been executed without violating any new constraints, then the potential new state estimate is $s_{k+1} = s_k + \Delta s$. For s_{k+1} to be acceptable, it must yield a reduction in the cost function. The cost function value at s_{k+1} is obtained by running the SRIS without computing the information array R_{k+1} . If the cost function has been reduced, then s_{k+1} is accepted as the new state estimate. The algorithm stops if the maximum number of iterations has been reached, or if convergence as discussed in Section ?? has been obtained. If neither of these is the case, then $k + 1$ is set to k and the loop is repeated.

If the cost function has not been reduced, then s_{k+1} is not an acceptable new estimate, so the Newton step factor is reduced and we try again. The step factor is halved, and s_{k+1} is recomputed using the new STEP. However, the step factor cannot be halved more than eight times, or the program will terminate with an error message.

Figure 1.7: Flowchart for the Step-Halving Optimization Algorithm

1.9 Post-Execution Validation

We assume that the fusion user has carefully followed all our suggestions for pre-screening the data, has chosen the fusion input values with care, and has a fusion solution which did not abort with serious error messages. The user now wants to check the validity of the solution.

The first metric which should be checked is the convergence test. If the convergence test metric is less than the user specified threshold, the message "Successful Convergence" should be printed near the end of output. This metric is computed as

$$\epsilon = \frac{1}{n}(s_{k+1} - s_k)^T (P(s_{k+1}))^{-1} (s_{k+1} - s_k) \quad (1.18)$$

where s_{k+1} is the state estimate at the current iteration, $P(s)$ is the state error covariance matrix computed by the fusion ($P^{-1} = F$), and n is the number of states. The program actually computes ϵ using a different equation, but it is equivalent to (1.18) if states are not constrained. ϵ represents an average squared change in state estimates normalized by the uncertainty in the estimates. Since we would like the change in state estimates between iterations to be "small" compared to the final uncertainty in the estimate, a typical ϵ convergence threshold might be 0.1.

In some nonlinear problems, the optimization can never obtain a value of ϵ as small as 0.1. This may happen because the computed value of $P(s)$ is based on an approximation. If the convergence fails because the trust region radius was reduced too many times, the final estimate may still be good. The user should check the least squares cost function and ϵ summary at the end of the run. If the cost function had been decreasing steadily in early iterations, but finally reached a plateau, and ϵ was less than 10, then the results may be good. If the convergence test failed and it seems that the residuals are large, a possible culprit is poor initial state estimates. One could also consider changing the parameters to be estimated, changing the minimum or maximum parameter limits, changing the editing threshold for the measurements, tuning spatial continuity or model noise, hypothesizing discontinuities, and adding more measurements or data types.

The next step in confirming the results is to check the distribution of measurement residuals and the residual RMS. The residual RMS should be approximately 1, although values from 0.3 to 3 may not be reason for great concern. Some deviation from 1 may indicate that the random measurement errors of one or more data types are smaller or larger than expected. However, large residual RMS values probably indicate modeling errors, that the assumed random measurement error is too small, or that there are data problems. A more rigorous statistical test is to compare the final cost function to the total number of measurements m . This is Chi-squared distributed with m degrees of freedom for Gaussian noise. Deviations between the cost function and m greater than three to five times $\sqrt{2m}$ are cause for concern. If the residuals are not Gaussian then the above Chi-squared test is less useful. Fusion checks to see if the residuals are Gaussian as described later in this section.

Additional information about the estimation health can be obtained by looking at the spatial distribution of measurement residuals. Fusion computes the mean, standard deviation

and RMS for each measurement type for groups of row/column blocks (typically three rows and columns). Several rows and columns are used so that the number of residuals in each block is statistically significant. Ideally the mean should be zero and both the σ and RMS should be about 1.0. If the mean is greater than $2\sigma/\sqrt{m}$ where m is the number of residuals in the block, an asterisk is printed indicating that the mean is significant.

Options are available to plot residuals for each measurement type by rows or columns. The user should look for systematic spatial patterns in the residuals which may indicate modeling problems. The estimation process attempts to assign any patterns in the measurements to states in the model. Thus the residuals should all be zero mean and white (uncorrelated) if the final model closely matches truth. Patterns which remain in the residuals indicate that there are systematic effects in the measurements which could not be fit using the given state model structure. This may be caused by systematic errors in measurements (e.g. timing bias), geologic noise (e.g. buried metal), errors in the measurement forward model (e.g. continuous variability in parameters not modeled accurately by a layer cake model, or assuming the wrong layer is causing a reflection), or defects in the conceptual site model (e.g. local anomalies such as faults, not modeled by the MRF continuity model). The determination of what modeling error caused a particular pattern in the residuals is partly science and partly art. People experienced with hydrogeology, geology, and the measurement instrumentation can often rule out some possible error sources, and quickly identify the source of the problem.

Examination of the spatial continuity pseudomeasurement residual plots may also aid this process. Since the purpose of the continuity model is to incorporate correlations between properties at different locations, large pseudo-measurement residuals indicate excessive changes in properties. This might be caused by geologic discontinuities, less spatial continuity than modeled, or errors in certain measurements. Again geologic and hydrogeologic intuition play an important role in this process.

Another potentially useful source of information is the normal probability distribution plots for the residuals. If the measurement errors are Gaussian and the models are correct, this plot should be a straight line at 45 degrees. Multiple bends in this line may indicate systematic modeling errors.

It is also wise to test the sensitivity of estimates to perturbations in input parameters. If the estimate sensitivity is high, a better choice of inputs may be possible. The results should always be checked for reasonableness. If the final model fits the data but the estimates of some parameters are outside the range known to be physically possible, then more work is needed.

If it appears that statistical parameters need to be tuned, then the statistical likelihood function provides a tuning criterion. Examining fusion estimates and residuals in the light of geological and hydrogeological judgement will provide ideas on what parameters need to be tuned and in what direction. The likelihood function will tell whether or not the statistical parameters have been improved. By minimizing the cost function in (2.1), fusion is producing an estimate that matches measured data (but not too closely or it will fit spurious measurement noise), is spatially continuous (but not too continuous or it will smooth out heterogeneities), and matches first principles models (but not too closely or it will be disturbed by model approximation errors). The statistical parameters (measurement noise

standard deviation, spatial continuity correlation distance and standard deviation, model error standard deviation) determine how fusion balances data, spatial continuity, and models. But the statistical parameters can be estimated from the data themselves when adequate data have been acquired using the statistically optimal maximum likelihood method. Parameter tuning is accomplished by selecting parameters that reduce the following tuning criterion that is proportional to the negative twice log likelihood

Tuning Criterion =

Converged value of cost function $C(s)$

- +2ln (Product of diagonals of the square root information matrix)
- 2ln (Product of diagonals of square root information matrix computed at converged estimate but without information from measured data)
- 2ln (Product of reciprocals of measurement noise standard deviations over all measured data)

The Generalized Expectation Maximization (GEM) method provides an automatic way to minimize the tuning criterion. However, we have not implemented GEM yet.

The ultimate test of a model is its predictive capability. If measurements can be left out of fusion and fusion can predict the measurements within its computed uncertainty, then there is reason for confidence in fusion results. This procedure is referred to as cross validation. Care must be taken in performing cross validation that the predictions which will influence engineering decision analysis really are being tested by the cross validation. This procedure is referred to as cross validation.

The reward for carefully initializing, executing, and validating results is a quantitative understanding of geological and hydrogeological uncertainties. Then remediation can be simulated in the computer to get it right with quantified safety margins before doing it in the earth.