Bayesian Integration of Hydrogeological and Geophysical Data for Site Characterization: Theory and Application to the LLNL Superfund Site

Souheil Ezzedine [1] **and Yoram Rubin**[2]

[1] Weiss Associates, ERD. Lawrence Livermore National Laboratory, L530. 7000 East Avenue, Livermore, CA-94550.
[2] Civil and Environmental Engineering. University of California, Berkeley. 440 Davis Hall, Berkeley, CA-94720.

Abstract. A stochastic Bayesian approach for combining well logs and geophysical surveys for enhancing subsurface characterization is presented. The main challenge we face is in creating the bridge to link between ambiguously related geophysical surveys and well data. The second challenge is imposed by the disparity between the scale of the geophysical survey and the scale of the well logs. Our approach intends to integrate and transform the well log data to a form where it can be updated by the geophysical survey and this tends to be a convoluted process. Our approach starts with generating images of the lithology, conditional to well logs. Each lithology image is then used as the basis for generating a series of shaliness images, conditional to well logs data. Shaliness images are converted to resistivity images using a site-specific petrophysical model relating between shaliness, resistivity and lithology, to create the necessary interface with the crosswell resistivity survey. The lithology and resistivity images are then updated using crosswell electromagnetic resistivity surveys. We explored the limits of the approach through synthetic surveys of different resolutions and error levels, employing the relationships between the geophysical and hydrological attributes which are weak or non-linear or both. The synthetic surveys closely mimic the conditions at the LLNL Superfund site. We show that the proposed stochastic Bayesian approach improves hydrogeological site characterization even when using low-resolution resistivity surveys.

1 Introduction

Research in the area of site characterization has focused mainly on development of inverse algorithms [McLaughlin and Townley, 1996; Ginn, Cushman, 1990; Yeh, 1986]. While our understanding of the problem has improved, the problem is generally considered as yet unsolved with no fully proven technique, there are clear ideas of where the weak points are and what remedies might be. Most inverse techniques rely on point measurements such as permeability, head, concentration, and use geostatistical techniques for interpretation and integration of different data for subsurface characterization. Few works have gone beyond and incorporated additional information such as geophysical surveys with well log measurements for enhancing the quality of subsurface characterization measurements (Hubbard and Rubin, [2000]). The primary motivation has been the recognition that geophysical surveys offer unique opportunities for enhancing crosswell interpolation, and are particularly promising in cases of data scarcity.

Hyndman et al. [1994] developed an inversion algorithm that employs both seismic crosswell travel times and solute tracer concentration to estimate the inter-well geology and therefore the hydraulic parameters. Sheets and Hendricks [1995] used regression techniques to build a site-specific petrophysical relationship between the soil water content estimated from borehole neutron probes and the bulk electrical conductivity of the soil estimated from electromagnetic induction surveys. This model was then used for mapping the soil water content. Daily et al. [1992] conducted an infiltration experiment to build a site-specific regression model between the resistivity and moisture, and showed the potential capability of electrical resistivity tomography to monitor capillary barriers performance and flow in the vadose zone. Doyen [1988] used cokriging to estimate porosity from surface seismic data and well logs. Cassiani et al. [1998] included seismic tomography data and sonic data using a geostatistical approach to improve the estimation of the hydraulic conductivity. Lucet and Mavko [1991] combined crosswell seismic tomography, logs and petrophysical relationships between porosity, velocity and clay content to estimate porosity and lithology. Rubin et al. [1992], and Copty and Rubin [1995] used a Bayesian approach and maximum likelihood principles to combine seismic velocity with sparsely measured hydraulic conductivity and pressure for the purpose of mapping the spatial distribution of the hydraulic conductivity. Hubbard et al. [1997] used a similar approach to incorporate the spatial distribution of dielectric constant obtained from ground penetrating radar to estimate soil

saturation and permeability in the vadose zone in the case of bimodal spatially distributed hydraulic conductivity distribution. Hubbard et al. [1999] combined acoustic tomography with borehole data to estimate the spatial covariances of the log-conductivity. More recently, Chen et al. [2001] used Bayesian paradigm based on the normal linear regression model to estimate the permeability from geophysical tomographic data.

A few observations based on these studies are as follows: (i) No universal methods or petrophysical models are available for converting geophysical attributes to hydrogeological ones; (ii) The most challenging problem is tying well-logging measurements to the geophysical surveys. This issue involves problems of scale disparity and inconsistencies in the methods of data acquisition and interpretation, which, sometimes, lead to dramatically different results.

The present paper investigates the use of geophysical data and surveys for mapping lithology and soil properties in the subsurface using a Bayesian approach. The study focuses primarily on the issues and problems associated with the assimilation of weakly or non-linearly correlated data, which are characterized by different spatial resolutions, in a geologically complex environment. The paper includes 5 sections. Section 2 introduces the LLNL superfund site and presents a geostatistical and a petrophysical analysis of the data. Section 3 outlines in detail our approach for data interpretation, principles and application. Section 4 introduces the synthetic electromagnetic survey, and Section 5 discusses Bayesian updating of pre-simulated lithology and resistivity random fields and evaluates the effectiveness of the proposed approach. Section 6 summarizes all findings.

2 Site Description, Sources of Data, and Geostatistical Analysis

2.1 Lawrence Livermore Superfund Site

Volatile organic compounds (VOC) were used at the LLNL superfund site (Figure 1) as solvents when the site was an active Naval Air Force Base in the 1940's. Fuel petroleum hydrocarbons associated with gasoline spills have also contaminated the underlying aquifer. The VOCs are mainly Trichloroethylene (TCE), Tetrachloroethylene (PCE) and Chloroform [Noyes, 1991].

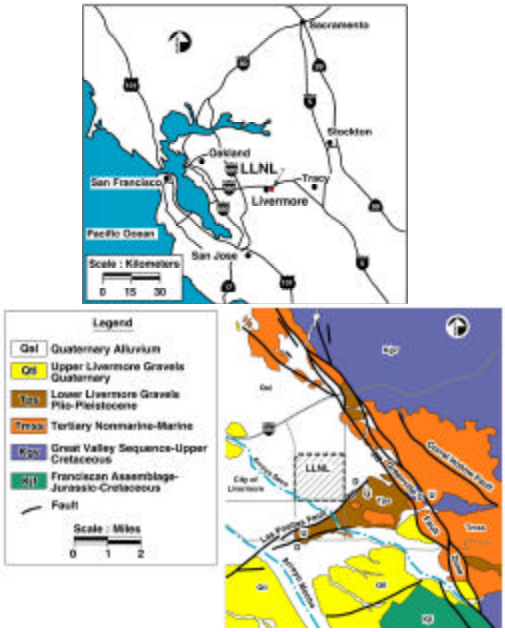

Figure 1: Aerial view of the location of LLNL and the complex geological setting in its vicinity.

The site is located in an unconsolidated alluvial basin. The hydrogeology of the area is very complex, but a considerable amount of geological, geophysical, hydraulic and geochemical data is available. These data provide a unique opportunity to study the relationship between hydraulic conductivity and sediment texture. Our analysis is focused on the treatment facility D (TFD) shown in Figure 2. The boreholes used in the present analysis, are depicted in Figure 3. The contaminants are distributed within a thick, complex sequence of unconsolidated alluvial sediments [Blake et al., 1995]. A hydrostratigraphic analysis has been conducted to divide sequence of layers into hydrostratigraphic units (HSUs).

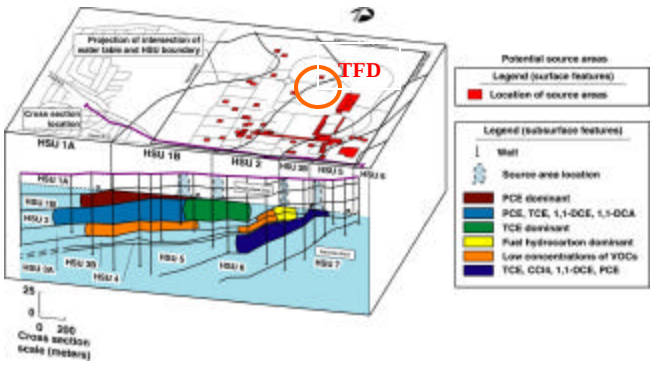

Figure 2: Site map of LLNL showing Hydrostratigraphic Units (HSUs) and contoured volatile organic compounds (VOCs). Our analysis is focused on the Treatment Facility D (TFD) shown on the top right-hand side.

2.2 Lithological and Geophysical Raw Data

We focus our efforts on the cross-section between wells 1206-1208-1205-1252-1250-1251-1254 at HSU2 (Figure 3a). Spatial statistics are inferred from all available data and wells in HSU2. Types of data collected along the wells include geophysical well log data and lithology. These various data are characterized by different vertical spatial resolution along the boreholes, varying from 3*cm* to 15*cm*. Geophysical well log data collected at the site include: induction resistivity, short and long normal resistivity, spontaneous potential, single point resistance, guard resistivity, and gamma ray, among others. Lithology logs were classified as gravel, clay, sand, silt, and their mixtures totaling 16 different lithologies. Because we are particularly interested in mapping the high and low hydraulic conductivity zones, only two main classes will be used: 1) silt, which includes all silts, clays and their mixtures, and 2) sand, which includes all sands, gravels and their mixtures.

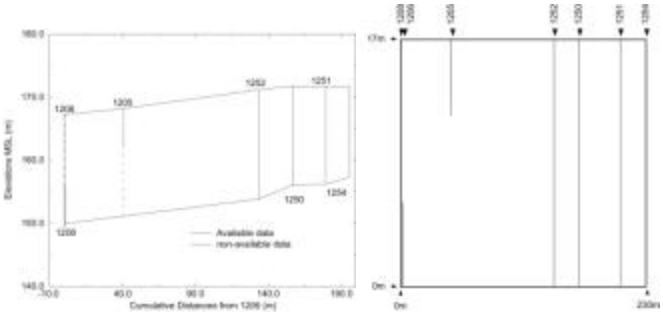

Figure 3. a) Vertical cross-section over HSU2. Vertical dash lines represent non-available data. Continuous vertical lines represent available data along the wells. b) Vertical cross-section of the present study over HSU2 and only through wells 1206-1208-1205-1252-1250-1251-1254. Distances are reported from well 1206 and along the cross-section.

2.3 Geostatistical well logs analysis

Since the HSUs are not horizontal and are not defined by constant thickness (Figure 2), the vertical coordinates were normalized by the average thickness of the HSU which is ~17*m* (Figure 3b). An indicator semivariogram was used to characterize the spatial variability of the lithologies based on a binary representation for sand and silt. Semivariograms have also been used to characterize the spatial variability of gamma ray (G) and resistivity (R). The reason for presenting the statistics of these 3 variables is in the fact they form the basis of our method for utilizing the resistivity survey.

2.3.1 Lithology Indicator Semivariograms: Adopting an indicator coding of 0 for sands and 1 for silts, a geostatistical analysis of the lithologies spatial distribution was performed. Figures 4a-b depict the vertical and horizontal indicator semivariograms. The volume fractions of silts, *p*, and sands, (1-*p*), are respectively 0.48 and 0.52. The sills of the semivariograms are 0.25, equal to the theoretical value of the variance of the population which is *p*(1-*p*). The theoretical models fitted to the data are exponential with a range of 1.5*m* in the vertical direction, and 30*m* in the horizontal direction.

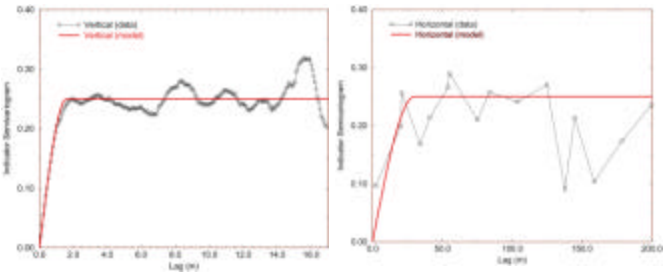

Figure 4: Indicator experimental and theoretical semivariograms: a) vertical direction, b) horizontal direction. Theoretical semivariograms are exponential.

2.3.2 Resistivity Measurements Analysis: Since a crosswell electromagnetic resistivity survey is considered at the LLNL site, well log resistivities were chosen as the primary means for tying and correlating the tomographic survey with other soil properties. Semivariograms of induction resistivity, guard resistivity and short and long resistivities were investigated; yet well-defined, long-range patterns of spatial correlation were not identifiable. Despite the apparent lack of spatial correlation of the resistivity we found that the combination of induction resistivity with gamma ray and lithology logs offers an opportunity for indirect projection of gamma ray and lithology pairs into resistivity. Our choice of induction resistivity as the primary candidate for correlation with the resistivity survey is based on the excellent quality of the induction log measurements at the LLNL.

2.3.3 Gamma-Ray Analysis: Gamma ray logs measure naturally occurring gamma emissions around the borehole. Gamma-ray response decreases from shales and clays, to siltstone, to sandy siltstone, to sandstones and gravels. Conversion of gamma ray measurements to shaliness helps to remove inconsistencies in the data introduced by using different measurement tools and calibration techniques [Serra, 1986]. Figures 5a,b depict the vertical and the horizontal semi-variograms of the shaliness and the fitted models. The best fit was found to be a Gaussian model with a nugget equal to 0.011 and ranges of 2.5*m* in the vertical direction and 25*m* in the horizontal direction. Integral scales are set to 1.46*m*(~1.5*m*) and 14.43*m*(~14.5*m*) in the vertical and the horizontal directions.

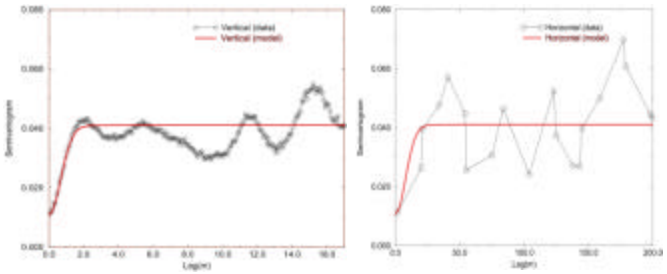

Figure 5: Shaliness experimental and theoretical semivariograms: a) vertical direction, b) horizontal direction. Theoretical semivariograms are Gaussian.

2.3.4 Shaliness vs. Resistivity Relationship: Figure 6a displays a crossplot of the resistivity and shaliness. Two main clusters are shown, corresponding to the different lithologies. It suggests that resistivity/shaliness pairs are useful for lithology identification. The overlap between the sand and silt clusters indicates that a unique identification of lithology based on resistivity and shaliness is not possible for all pairs.

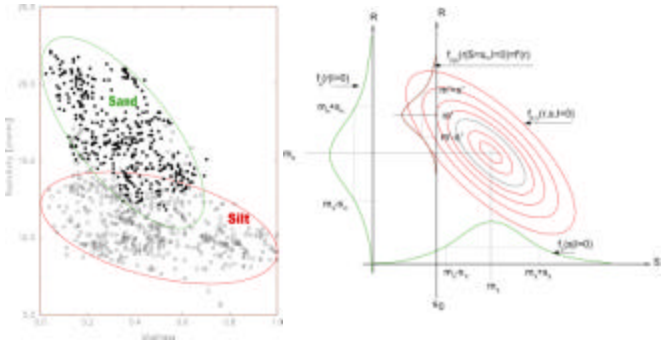

Figure 6: a) True petrophysical relationship between shaliness and resistivity plotted from available data at the wells crossing the HSU2. b) Generic scheme for constructing conditional (prior) resistivity pdfs to lithology and shaliness

The main reason for the overlap between the two clusters is data reduction: the original lithology classification consisted of 16 members, which we have reduced to only two.

Despite the ambiguous interpretation of several pair combinations, it appears that this crossplot is a good analytical tool. This is one of the fundamental results of our analysis so far since it suggests a systematic approach for tying the survey with well logging information. Figure 6a was obtained using all well log data within the HSU2. The use of shaliness instead of gamma-ray activity considerably improved the clustering analysis. This analysis has been applied to other HSUs as well, and we observed a behavior similar to Figure 6a.

3 Bayesian Data Assimilation

Ideally, the geophysically measured attributes correlate well with the hydrogeological ones and the conversion of the geophysical attributes to hydrogeological ones is straightforward. In realistic situations, however, the conversion is convoluted and non-unique. Hence, a conceptual, data-driven approach for lithology mapping based on the well log data is developed. The proposed approach is general in its basic principles, but not universal since the employed petrophysical models are site specific. The general approach is stochastic given the large uncertainty associated with crosswell interpolation, with the petrophysical models and with the interpretation of the geophysical surveys.

Our approach consists of sequentially generating a series of collocated attributes. At the basis of the hierarchy, images of the lithology are generated, conditional to well logs. Each lithology image serves then as the basis for generating a series of shaliness images conditional to well data. The shaliness images are then used to correlate the survey resistivity with the hydrogeological attributes obtained experimentally. The series of generated images all have in common the well data and the same underlying spatial structure and hence they are all physically plausible. The variations between

the images constitute a measure of the spatial variability and estimation uncertainty. We focus here on estimating resistivity, but it can be converted to conductivity using well-known or site-specific models [Mavko et al., 1998; Daily et al. 1992].

3.1 Outline of the approach

After the exploratory data analysis is performed as previously described, we proceed in four steps as described below.

a- Generation of the lithology images using sequential indicator simulation (SIS). The lithology is defined through an indicator variable I according to: i=1 if $\mathbf{x}$ is located in a silt body, 0 otherwise. Note that boldface letters denote vectors, i.e., $\mathbf{x}$ is the location coordinates vector. Lower-case i is a realization of the spatial random function (SRF) I. I is characterized through its expected value conditional to the borehole data, $p^c = \mathbf{E}^c\{I\} = \mathbf{E}\{I|\text{measurements}\}$, with a superscript "c" denoting conditional. Since I is binary, p^c is statistically exhaustive. Its spatial variability is defined through the semivariogram and is shown in Figures 4a-b. These statistics are the cornerstone of the SIS algorithm [Deutsch and Journel, 1998] adopted here.

b- Generation of shaliness images. This step is similar in principle to the previous one. The differences are in the fact that (i) the shaliness S is not a binary variable and (ii) the pattern of spatial variability of the shaliness may be different between the sand and silt lithologies, i.e. $\gamma_{S|i}$, the semivariogram of the shaliness S, depends on the lithology i=0 or 1. SGS algorithm [Deutsch and Journel, 1998] is adopted here to generate shaliness images. Shaliness S is defined by its mean $m_{S|i}$, its semivariogram $\gamma_{S|i}$ and its covariance, $Cov_{S|i}$, for a given facies i.

c- Computing the resistivity prior pdf. Once $\mathbf{x}$ is identified as being either sand or silt and is assigned a shaliness value, a prior pdf for the resistivity $f_{R(\mathbf{x})}(r|I=i, S=s)$ can be defined through Figure 6a. R and S denote the SRF of the resistivity and the shaliness, respectively, and r and s denote their realizations. Figure 6b illustrates the joint pdf of R and S given I=0 (i.e., sand lithology) and the marginals $f_R(r|I=0)$ and $f_S(s|I=0)$. Conditioning further on $S=s_0$ leads to $f_{R|S}(r|S=s_0, I=0)$, which is our Bayesian prior. Scarcity of data led us to condition on ranges of S values rather than on single values. These pdfs are the Bayesian prior pdfs of the resistivity, and hence our stochastic estimation for the resistivity R at $\mathbf{x}$ in case no additional data become available through surveying.

d- Updating $f_{R(\mathbf{x})}(r|I=i, S=s)$ based on crosswell electromagnetic resistivity survey $\rho(\mathbf{x})$. Defining $f_{R(\mathbf{x})}(r|I=i, S=s) = f'_{R(\mathbf{x})}(r)$ for brevity, and given a collocated survey resistivity $\rho(x)$, the posterior pdf $f'_{R(\mathbf{x})}(r|\rho)$ can be defined through Bayes' rule

$$f'_{R(x)}(r \mid ?) = C_R L(? \mid r) f'_{R(x)}(r), \quad \text{and}$$

$$C_R = \left(\int_{-\infty}^{+\infty} L(? \mid r) f'_{R(x)}(r)\, dr \right)^{-1}$$

[Ang and Tang, 1975]:

where $L(\rho|r)$ is the likelihood function, and C_R is a normalized factor. In general, ρ is defined over a support volume larger than the support volume of r. In the case of a high resolution geophysical survey $\rho(\mathbf{x}) \rightarrow r(\mathbf{x})$ and Bayesian updating is unnecessary. This, however, is not generally the case and the alternative is to update $f'_{R(\mathbf{x})}(r)$ given ρ. Typically we are interested in R representative of a block of scale ~1m while ρ is defined by blocks of scale ~3m or greater. The inference of the likelihood function, $L(\rho|r)$, is critical for the successes of the updating process. Once $f'_R(r|\rho)$ is defined, a realization of R at $\mathbf{x}$ can be drawn. The whole process is repeated for all $\mathbf{x}$ until a complete image of the resistivity field is completed. Similarly, the lithology images can be improved through the resistivity survey despite the non-linear and non-unique relationship displayed in the cross-plot. Our approach calls for Bayesian updating of p^c as well, through the relationship: $p^{c'} = C_I\, L(\rho|I)\, p^c$, where $L(\rho|I)$ is the likelihood function, of a similar nature to $L(\rho|r)$, only relating ρ to I rather than R. C_I is a normalized factor similar to C_R.

3.2 The Synthetic "True" Database

The concept outlined in Section 3.1 is demonstrated here using a synthetic example,

generated to simulate closely the conditions of HSU2. Figure 7a depicts a realization of HSU2 lithology conditional to the lithology observed at the wells. The field is 230*m* and 17*m*, in the horizontal and the vertical direction, respectively. Realizations of the shaliness (not shown here) and resistivity fields, generated based on 3.1.b-c, conditional to borehole data, are depicted in Figure 7b using spatial statistics described in Section 2.

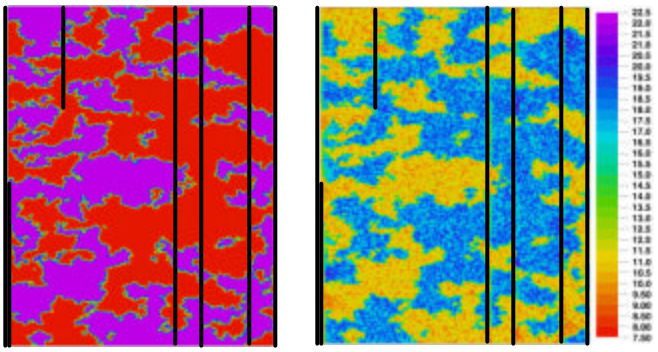

Figure 7: "True" geological setting. a) Sequential Indicator Simulation of the lithology conditional to borehole data. Black and white colors represent sand and silt, respectively. b) True resistivity random field built by projecting the "true" shaliness field using the petrophysical relationship depicted on Figure 6a.

4 Electromagnetic Surveying

Field EM surveying is a complex mapping of the detailed, high resolution $R(\mathbf{x})$ distribution into a low resolution $\rho(\mathbf{x})$ field. An electromagnetic survey was conducted at LLNL, but final results are not yet available. To explore the Bayesian updating approach, synthetic surveys of the resistivity are simulated. Under reasonable approximation of low variability of the resistivity between the sand and the silt bodies (see Figure 6a), the EM wave propagation problem can be reduced to an electric current diffusion problem. Identical problems have been considered in fluid flow in porous media and electric currents [Dagan, 1989; Abramovich and Indelman, 1995]. Borrowing from their results, and considering 2D survey, the electrical conductivity κ_b (inverse of electrical resistivity) of a block, which covers n_x by n_z small-scale blocks, where n_x is the number of blocks in x direction and n_z in the *z* direction, is given by the geometric mean. The geometric mean is applicable for blocks, which are large relative to the characteristic length scale of resistivity heterogeneity, which is the case of the present study. We shall consider $n_x=n_z=3,6,9$. Figures 8a-c show results of synthetically surveying the resistivity field shown in Figure 7b using different resolutions. As the resolution decreases, small-scale details become obscure and fuzzy, and the range of resistivity values detected narrows.

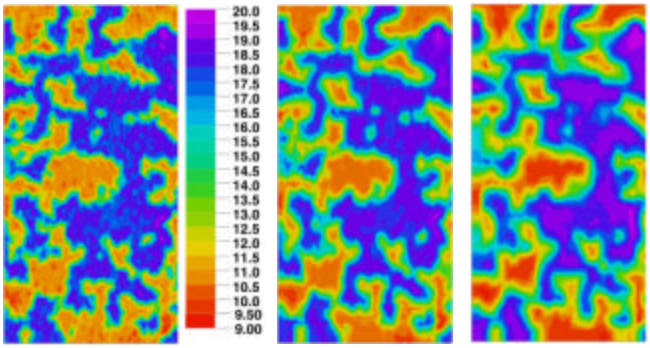

Figure 8a-c: Examples of the resistivity surveys obtained by geometric averaging of the "true" resistivity field (Figure 7b) over a) 3, b) 6, and c) 9 small scale blocks in the horizontal and vertical directions.

5 Synthetic Case Study

In our case study we investigate the cross-section shown in Figure 3b, assuming that Figures 7a-b, which were generated conditional to the borehole data, are the "true" images of that cross-section. A geophysical survey of the same cross-section is simulated. Our goal is to test the capability of the method described in Section 3 to reconstruct the base case's images while benefiting from the resistivity survey. Typical images obtained through the use of prior pdfs only are depicted in Figures 9a-b. Figures 9a and 9b will be updated following the methods outlined in 3.1c and 3.1d. Updated images will be compared to the assumed "true" images depicted in Figures 7a and 7b.

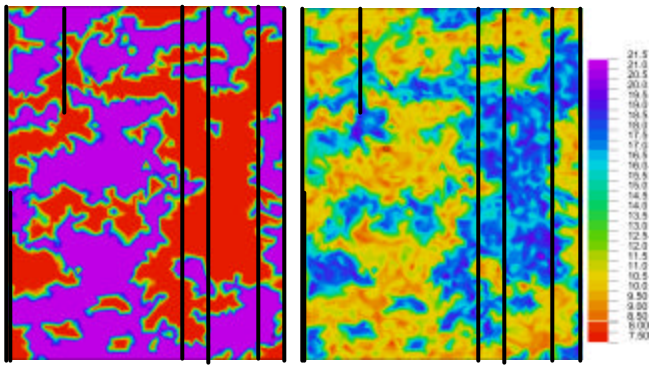

Figure 9: a) Single realization of the lithology field. Sequential Indicator Simulation (SIS) of the lithology conditional to borehole core data. Black and white colors represent sand and silt, respectively. b) Single realization of the resistivity field built by projecting the shaliness random field (not shown here) using the petrophysical relationship (Figure 6a).

5.1 Indicator Likelihood Functions and Updating the Lithology Images

To infer $L(\rho|I)$ we use the concept of "training set". The idea is to identify a portion of the survey area which will be drilled and cored post survey to yield a set of collocated measurements (ρ,i). The dimension of the training set area should be determined such that the survey represents the entire range of conditions expected over the entire surveyed area. The sampled area needs to be ergodic in terms of bivariate (ρ,i) statistics. That usually implies a dimension of several integral scales vertically, along cored wells. In the present application, the well-sampled area near well 1250 (right-hand side of Figure 3) was set to be the "training set", and the much less sampled area near well 1205 (left-hand side of Figure 3) was set as the "testing set". $L(\rho|I)$ is determined for a given $I=i$ and $\rho=\rho_0$ by scanning the set of collocated pairs (i,ρ_0) and computing the conditional probability $Prob[\rho=\rho_0|I=i]$. Bayes' rule is then used to update the lithology image (Figure 9a) using different resistivity survey resolutions. Images of the "testing set" obtained for $(n_x \times n_z)=(3x3)$, (6x6) and (9x9) resistivity surveys (Figures 10a-c) are practically of the same quality as without updating and differ only by a fraction less than 1% from the prior lithology (Figure 9a), even in case of high resolution resistivity survey (3x3).

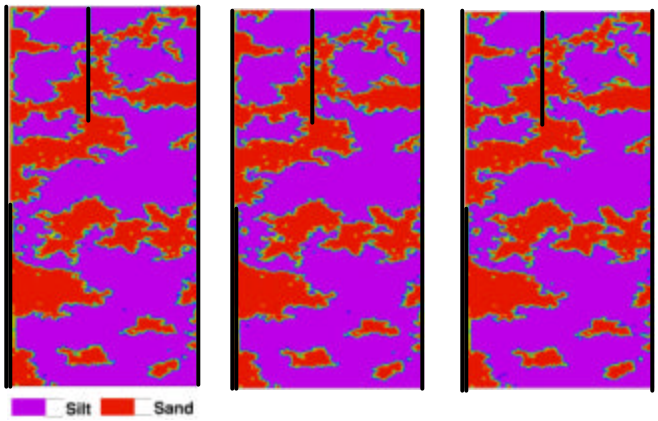

Figure 10a-c: Posterior lithology image of the "testing set", left-hand side of Figure 9a, using Bayes' rule, and a) 3x3, b) 6x6, and c) 9x9 resistivity survey (Figure 8a-c).

This outcome is a manifestation of the effect of the homogenization, which obscures the resistivity-lithology relationship. A large number of resistivity combinations can lead to the same

ρ, and hence to non-unique relationship between ρ and the lithology.

5.2 Resistivity Likelihood Functions and Resistivity Images Updating

L(ρ|r) is approximated here by L(ρ|r-δr≤ρ<r+δr) with a relatively small δr, due to the data scarcity. The effect of updating the resistivity based on surveys with different resolutions is demonstrated in Figure 11. It shows the prior and posterior resistivity pdfs at arbitrary points within the silt and sand lithologies for various resolutions of the resistivity survey. The maximum beneficial effect is obtained, not surprisingly, through the high-resolution survey, but the positive impact of conditioning on ρ is discernible even at the low-resolution surveys. The trend of reduction in impact with poorer resolution is evident and is an outcome of the diffuse and non-informative nature of the likelihood function as the discrepancy between the survey scale and the desired resolution scale increases.

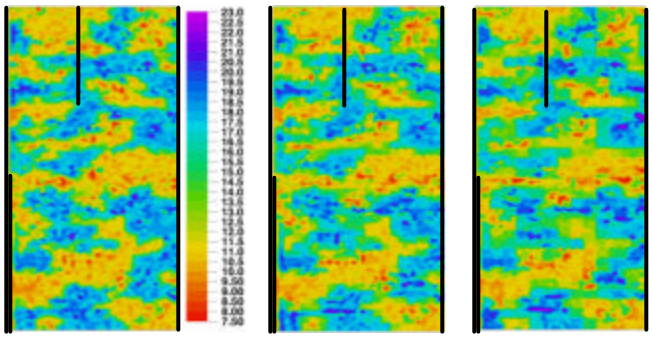

Figure 11: Posterior resistivity images of the "testing set", left side of Figure 9b, using Bayes' rule, and the resistivity surveys: a) 3x3, b) 6x6, and c) 9x9 resistivity survey (Figure 8a-c) and the posterior lithology (Figure 10a-c).

Note that conditioning R on ρ does not imply that the randomly generated values will average exactly to yield ρ unless special measures are taken. To honor precisely the surveyed value ρ, a constraint on the generated value is introduced so that the generated r values over any volume corresponding to ρ will average exactly to yield ρ. This procedure is described elsewhere (Ezzedine et al., [1999]). Figures 12a-c depict the updated resistivity fields for (n_x x n_z)=(3x3), (6x6) and (9x9) resistivity survey. These Figures should be compared to the "true" image (Figure 7b, left part), and with the image generated based on the prior pdfs, Figure 9b. It is quite obvious that the resistivity surveys have a significant positive impact, particularly at the high resolution.

5.3 Effectiveness of the Bayesian Updating

To evaluate the effectiveness of the updating procedure, we analyze the following statistic:

where k is a running index over all the points outside the wells, r is the actual resistivity (Figure 9b), m″ is the mean of the posterior pdf $f''_{R(\mathbf{x})}(\mathbf{x})$, and m' the mean of the prior pdf $f_{R(\mathbf{x})}\}(\mathbf{x})$. The ratio ℜ compares the performance of the posterior and the prior pdfs. ℜ smaller than 1 indicates a successful updating procedure. ℜ=1 is a diffuse likelihood and hence a non-informative survey. Figure 12 depicts the variation of ℜ, as a function of the resolution of the survey. For completeness, statistics were also computed for resistivity surveys of (2x2) and (12x12) block resolution. We have found that ℜ decreases with decrease in resolution, in line with Figures 8a-c.

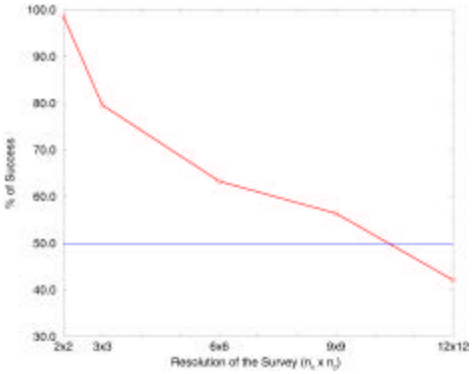

Figure 12: Percentage of number of successes, based on ℜ, of the Bayesian updating approach for different survey resolutions and different errors in the surveys.

6 Summary

We surveyed some of the problems associated with combining resistivity tomography and resistivity well logging. We analyzed the data collected at the LLNL site, and synthetically surveyed a cross-section, which was constructed to mimic closely the geology of the site.

$$\Re_k = \frac{1}{|r_k - m|}; \quad \Re_k = 1, \quad \text{Unsuccessful}$$

Our approach for data assimilation is stochastic Bayesian. We find the justification for it in the large spatial variability and data scarcity. The Bayesian approach allows us to condition estimates on what is clearly a set of complex and non-linear petrophysical models relating between different geological attributes. Our approach comprises several steps, each of which intended to explore, model and utilize the aspects of the data that are needed for relating between the tomographic data and the well logs.

Our study employed several relationships between induction resistivity, lithology, shaliness and tomographic resistivity. These relationships reflect, to a large degree, properties, which are well understood and are quite general in terms of trends. We suspect though that these relationships cannot be transported to other sites. In this sense our method does not replace nor alleviate the tedious task of data exploration. At the LLNL site, the key element is the shaliness, due to its well-defined spatial structure and its sensitivity to resistivity. The idea then is to use this variable for projecting well data and generating a prior that are both relevant for the application and can benefit from the geophysical survey.

At this stage the Bayesian approach becomes the key for data assimilation, its robustness stems from its ability to express vague relationship as probabilistic rules and to bridge over scale disparity issues. This brings us to refer to the approach presented here as a set of tools rather than as a clear road map.

We have found that the benefits in estimating high-resolution subsurface resistivity given a low-resolution resistivity survey are more significant than those gained in estimating lithology. The LLNL data showed good correlation between resistivity and lithology at the small scale, but at lower resolution the correlations deteriorate. This observation is supported by the fact that the resistivity surveys were non-informative for updating the lithology images. Resistivity-shaliness-lithology relations may show perfect correlation at a fine scale but can appear to have large scatter when using a larger observation scale.

Acknowledgments. Funding for this study was provided by the NSF project EAR-9628306 and the DOE, EMSP Grant DE-FG07-96ER14726. This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48. The authors wish to thank: G. Mavko and T. Mukerji, Stanford Univ. R. Bainer, R. Blake and F. Hoffman, Lawrence Livermore National Laboratory. D. G. Hill, Weiss Associates. J. Rector, F. Morrison and R. Hatch, Univ. of California, Berkeley.

7 References

- Abramovich, B; Indelman, P. Effective Permittivity of Log-Normal Isotropic Random Media. Journal of Physics A- 28(3):693-700, 1995.
- Ang, A Hua-Sing; Tang, W. Probability concepts in engineering planning and design. Wiley, 1975.
- Blake, R.; Maley, MP.; Noyes, CM. Characterization of VOC transport in heterogeneous sediments at Lawrence Livermore National Laboratory and its impact on remediation decisions.1995 Annual Meeting - Association of Engineering Geologists; V38, abstract volume, p. 35, 1995.
- Cassiani, G; Bohm, G; Vesnaver, A; Nicolich, R.A geostatistical framework for incorporating seismic tomography auxiliary data into hydraulic conductivity. Journal of Hydrology, 206(1):58-74, 1998.
- Chen, JS; Hubbard, S; Rubin, Y. Estimating the hydraulic conductivity at the South Oyster Site from geophysical tomographic data using Bayesian techniques based on the normal linear regression model. Water Resour. Res.; 37(6):1603-13, 2001.
- Copty, N.; Rubin, Y.A stochastic approach to the characterization of lithofacies from surface seismic and well data. Water Resour. Res.; 31(7): 1673-86, 1995.
- Dagan, G. Flow and transport in porous formations. Springer-Verlag, 1989.
- Daily, W.; Ramirez, A; LaBrecque, D; Nitao, J. Electrical resistivity tomography of vadose water movement. Water Resour. Res.; 28(5):1429-42, 1992.
- Deutsch, CV.; Journel, A.GSLIB: geostatistical software library and user's guide Version 2.0, 2nd Ed. Oxford Univ. Press, 1998.
- Doyen, PM. Porosity from seismic data; a geostatistical approach. Geophysics; 53(10):1263-76, 1988.
- Ezzedine, S; Rubin, Y; Chen, JS. Bayesian method for hydrogeological site characterization using borehole and geophysical survey data: Theory and application to the Lawrence Livermore National Laboratory Superfund site. Water Resour. Res.; 35(9):2671-83, 1999.

- Ginn TR, and Cushman, JH., Inverse Methods for Subsurface Flow – A Critical Review of Stochastic Techniques. Stochastic Hydrol. and Hydraul., 4(1):1-26, 1990.
- Hubbard, SS; Rubin, Y. Hydrogeological parameter estimation using geophysical data: a review of selected techniques. J. of Contamin. Hydrol., 45(1-2):3-34, 2000.
- Hubbard, SS.; Rubin, Y.; Majer, E. Spatial correlation structure estimation using geophysical and hydrogeological data. Water Resour. Res.; 35(6):1809-25, 1999.
- Hubbard, SS.; Rubin, Y; Majer, E. Ground-penetrating-radar-assisted saturation and permeability estimation in bimodal systems. Water Resour. Res.; 33(5):971-90, 1997.
- Hyndman, DW.; Harris, JM.; Gorelick, SM. Coupled seismic and tracer test inversion for aquifer property characterization. Water Resour. Res.; 30(7):1965-77, 1994.
- Lucet, N.; Mavko, G. Images of rock properties estimated from a crosswell seismic velocity tomogram.1991 Technical program and abstracts of papers - EAEG; V53:520-21, 1991.
- Mavko, G.; Mukerji, T.; Dvorkin, J. The rock physics handbook: Tools for seismic analysis in porous media. Cambridge University Press, 1998.
- McLaughlin, D., and L.R. Townley, A Reassessment of the Groundwater Inverse Problem, Water Resour. Res., 32(5):1131-61, 1996.
- Noyes, CD. Hydrostratigraphic analysis of the pilot remediation test area, Lawrence Livermore National Laboratory. Livermore, California. MS Thesis, University of California at Davis. 163p, 1991.
- Rubin, Y.; Mavko, G.; Harris, J. Mapping permeability in heterogeneous aquifers using hydrologic and seismic data. Water Resour. Res.; 28(7):1809-16, 1992.
- Serra, O. Fundamentals of well-log interpretation. Developments in petroleum science; 15A-B.Elf Aquitaine, 1984.
- Sheets, KR.; Hendricks, JMH. Noninvasive soil water content measurement using electromagnetic induction. Water Resour. Res.; 31(10):2401-09, 1995.
- Yeh, W. W-G., Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem. Water Resour. Res., 22(1):95-108, 1986.