Title:

Use of Data Fusion to Optimize Contaminant Transport Predictions

Author:

*Edward Van Eeckhout
Energy and Environmental Analysis (TSA-4)*

Submitted to:

*General Distribution
October 1997*

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; therefore, the Laboratory as an institution does not endorse the viewpoint of a publication or guarantee its technical correctness.

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.

Contents

1.	Introduction.....	1
2.	Hydrogeologic Data Fusion--What Is It?.....	2
3.	Cost Comparison of Data Fusion to Traditional Methodologies	3
3.1	Site 1: Two Areas at the DOE Savannah River Site	4
3.2	Site 2: A DoD Massachusetts Site.....	4
3.3	Site 3: A DoD California Site	5
4.	Discussion	5
5.	Conclusions.....	6
6.	Appendix A: Data Fusion Background.....	7
6.1	Hydrogeological Foundation.....	7
6.2	Geophysical Perspective	10
6.3	Data Assimilation.....	11
6.4	Data Fusion	12
6.5	Markov Random Field (MRF)	13
6.6	Square Root Information Smoother (SRIS).....	13
6.7	Summary	14
7.	References.....	15

Figures

Figure 1.	Original data fusion characterization envisioned.....	1
Figure 2.	Data fusion modeling structure.....	3
Figure A1.	Data fusion role in decision analysis.....	8
Figure A2.	Remedial decision process supported by data fusion.....	8
Figure A3.	Data fusion uses Bayesian statistics.....	12

1. Introduction

The original data fusion workstation, as envisioned by Coleman Research Corp, was constructed under funding from DOE (EM-50) in the early 1990s. The intent was to demonstrate the viability of fusion and analysis of data from various types of sensors for waste site characterization, but primarily geophysical. The types of data originally envisioned are shown in Figure 1 below.

Figure 1. Original data fusion characterization envisioned.

This overall concept changed over time and evolved more towards hydrogeological (groundwater) data fusion after some initial geophysical fusion work focused at Coleman. This initial geophysical fusion platform was tested at Hanford and Fernald, and the later hydrogeological fusion work has been demonstrated at Pantex (CRC, 1995c), Savannah River (CRC and HydroGeoLogic, 1995d), the U.S. Army Letterkenny Depot, a DoD Massachusetts site and a DoD California site. The hydrogeologic data fusion "package" has been spun off to a company named Fusion and Control Technology, Inc. (<http://www.clark.net/pub/fuserdp/>). This package is called the Hydrological Fusion And Control Tool (Hydro-FACT) and is being sold as a product that links with the software package, MS-VMS (MODFLOW-SURFACT Visual Modeling System), sold by HydroGeoLogic, Inc. (<http://www.hgl.com/>). MODFLOW is a USGS development, and is in the public domain.

Since the government paid for the data fusion development at Coleman, the government and their contractors have access to the data fusion technology in this hydrogeologic package for certain computer platforms, but would probably have to hire FACT (Fusion and Control Technology, Inc.) and/or HydroGeoLogic for some level of software and services. Further discussion in this report will concentrate on the hydrogeologic fusion module that is being sold as Hydro-FACT, which can be linked with MS-VMS.

2. Hydrogeologic Data Fusion--What Is It?

We've all been involved in some trial and error experiences throughout our lives. Whether it was sighting a rifle, changing the ingredients for a recipe, or perhaps sinking drillholes to find water, we've all experienced it. Thus it is reasonable that various methodologies should be sought that could do trial and error computations for us, particularly in the mathematical environment. Such is the case with this methodology, which really is a sophisticated iterative first guessing technique for a specialized saturated flow groundwater modeling code. There is really nothing particularly new about this general concept, but the software code development for this specific module is different from those developed previously, particularly given the necessary iterations for thousands of parameters (including hydrologic heads, boundary conditions, conductivities (or transmissivities), recharge values, and leakages) and the necessary detail required.

Hydro-FACT automatically finds the parameter values which produce the best fit to available measurements. This is done iteratively, and the package will arrive at potential solutions within minutes, hours, or weeks, depending upon the platform, size and extent of problem, and quality of the data. This contrasts with hours, weeks, or months for arriving at a solution obtained using the manual trial and error method (or the solution may never converge). Further, Hydro-FACT provides a measure of model sensitivity by computing the standard deviations of the errors in the estimates of the parameter.

However, the software and computer can't do everything. An experienced analyst is still needed to decide if the results are reasonable and, if not, modify the model at the structural or conceptual level. One vision of the data fusion modeling structure is shown in Figure 2.

The software system envisioned is (Hydro-FACT User's Guide, 1997):

1. a preprocessor for inputting grids from MODFLOW,
2. the hydrological estimation programs,
3. a graphical interface, and
4. several utility programs to facilitate analysis.

The principal value-added feature of this package appears to be the hydrological estimation programs, number 2 above. In fact, this entire module could be viewed as a preprocessor to MODFLOW data input.

With Hydro-FACT linked to MS-VMS, it is probable that the package will sell for \$4,000 to \$5,000 per site (Hydro-FACT accounting for about \$2,900 of the cost). It is intended to execute

on a personal computer, but is being ported to workstations as well. There may be an additional price structure for workstations, but those structures are not available at this time.

Further background on the data fusion technique is presented in Appendix A.

Figure 2. Data fusion modeling structure.

3. Cost Comparison of Data Fusion to Traditional Methodologies

The benefits of this particular data fusion methodology over manual trial and error modeling (using an expert) come mainly from the ability to rapidly combine diverse sources of information to quantify and reduce uncertainty. However, the benefits are very site dependent. For a very complex site, savings might be on the order of man-months of time (or allowing convergence to occur where none might occur manually), whereas for a simpler site the savings might be on the order of man-weeks. If one is trained in the manual trial and error method, then the training time to perform the work using Hydro-FACT might be on the order of a month. Both the traditional and data fusion methodology would require the usage of computer technologies.

Direct savings may result from reduced labor costs, a reduction of monitoring wells, and a reduction in treatment wells. Indirect savings may result from better demonstration of regulatory compliance, better match to measured data and prior knowledge, and a more quantifiable parameter uncertainty. Since the savings are so site dependent, let us review some sites where this technology has been used, and what little comparative data that is known. These results are estimated from the Hydro-FACT User's Guide, September 1997, pp. 5-6 and personal communication with Dave Porter and personnel at the sites.

3.1 Site 1: Two Areas at the DOE Savannah River Site

Characterisitics of areas

- extremely heterogeneous conductivity field
- large number of wells: 228 at one site and 246 at the other
- boundary conditions poorly defined

Computation results

- data fusion root-mean-square fit (RMS) to data was 1 foot and 1.3 feet at the two sites; manual trial and error fit was 3 feet at one of the sites
- site personnel estimated 75% less effort than manual calibration

Estimated cost savings per area

- if it took four months to perform the calibration manually, the data fusion could do it in one month: three months savings at \$20,000 per month (using \$240,000 as the cost per man-year) = \$60,000
- however, the code cost \$5,000 and it took one month to train, so the net savings =

\$35,000 per Type 1 stand-alone area

(yielding more accurate results plus quantified uncertainty)

- for further areas, the code cost and training time would not be relevant

3.2 Site 2: A DoD Massachusetts Site

Characterisitics of areas

- site characterized by sand, silt, and clay overlying bedrock
- 337 available water level measurements
- variable boundary conditions and numerous ponds

Computation results

- data fusion RMS fit to data was .7 feet; manual fit was 2.3 feet
- Hydro-FACT solved the problem within a week; manual method took two months

Estimated cost savings per area

- if it took eight weeks to perform the calibration manually and data fusion could do it in one week: seven weeks savings at \$5,000 per week = \$35,000
- however, the code cost \$5,000 and it took four weeks to train, so the net savings =

\$10,000 per Type 2 stand-alone area

(yielding more accurate results plus quantified uncertainty)

- for further areas, the code cost and training time would not be relevant

3.3 Site 3: A DoD California Site

Characteristics of areas

- site characterized by sand/gravel and claystone overlying volcanic rock
- 38 water level and 17 pump test measurements available
- no natural boundaries so Hydro-FACT estimated source/sink heads

Computation results

- data fusion RMS fit to data was .3 feet; manual fit unknown
- Hydro-FACT solved the problem within one day after data files created

Estimated cost savings per area

- if it took five weeks to perform the calibration manually and data fusion could do it in one week (including data file creation): four weeks savings at \$5,000 per week = \$20,000
- however, the code cost \$5,000 and it took four weeks to train, so the net savings =

-\$5,000 per Type 3 stand-alone area

(yielding more accurate results plus quantified uncertainty)

- for further areas, the code cost and training time would not be relevant, allowing the savings to be positive

Now let us make some further assumptions.

1. Let us take the 15 major sites oftentimes listed for DOE Weapons Complex Facilities: Los Alamos, Sandia, Lawrence Livermore, Hanford, Savannah River, Fernald, Idaho National Engineering and Environmental Laboratory, Paducah, Portsmouth, Rocky Flats, Oak Ridge, Mound, Pinellas, Kansas City, and Pantex.
2. Each location has two areas like Sites 1 and 2 above and four areas like Site 3.
3. Then the total projected savings across the DOE complex would be:

$$15 \times [2 \times (\text{Type 1 area costs}) + 2 \times (\text{Type 2 area costs}) + 4 \times (\text{Type 3 area costs})] \\ - 15 \times [\text{site software cost} + \text{site training cost}] =$$

$$15 \times [2 \times (\$60,000) + 2 \times (\$35,000) + 4 \times (\$20,000)] - 15 \times [\$5,000 + \$20,000] =$$

DOE COMPLEX SAVINGS: \$3.68 million

4. Discussion

Actually, the estimate of eight separate cleanup areas per DOE site may be high for some sites, but it might be low for others. One could try to get better estimates, but it probably isn't worthwhile at this point. Additional to the DOE sites, there might be savings obtained from DoD and EPA sites as well.

In terms of the phrase "more accurate results," this is a tough qualitative feature to quantify. By gaining more confidence in the results, this may in fact allow fewer wells and surface sites to be monitored with time, which would be a definite cost savings in well placement, collection and analyses on all the sites. Drilling costs may be up to \$200,000 per well, not to

mention the cost of periodically sampling. However, there is little to no actual data on this happening one way or another.

The hydrogeological data fusion was funded in the joint DOE/DoD PARAGON program for \$3.1 million. Thus to receive a good benefit to cost ratio, there would have to be a lot of sites that would use the software (unless the savings accrued is greater than that estimated). The initial estimate for benefit-to-cost ratio for the DOE complex is \$3.7 million-to-\$3.1 million, or 1.2-to-1. If one extended this to DoD and EPA sites, the benefit to cost would be higher.

5. Conclusions

It appears that many DOE sites, as well as numerous other sites that have hydrological waste problems, have, and will, benefit from the utilization of this software. The actual savings is quite dependent upon the particular scenario encountered. This particular code can be used for a great variety (but not all) situations.

The cost savings of utilizing this code over the manual trial and error methods across the DOE complex appears to be about \$3.7 million. Similar savings could be realized at other industrial or defense sites, thus increasing the benefit to cost ratio. If one assumed a similar savings for DoD and EPA sites, the total could be \$11.1 million. Thus the benefit to cost ratio would be 3.6-to-1 if the DoD and EPA is added to the DOE savings. Further savings may be realized when fewer wells are factored into the equation, although a more detailed study may find fewer sites that this technique might apply to as well.

6. Appendix A: Data Fusion Background

(taken from HydroGeoLogic and Coleman, 1996, Section 3)

The historical setting for fusion is presented in this section, 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.

6.1 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 A1 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 A1, field data are used to update and calibrate geological and parametric models and compute uncertainty of model parameters. 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 A1. 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 A2. For example, remedial simulation of flow and transport can be performed for alternative remedial actions. Contaminant concentrations uncertainty can be predicted 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 and their uncertainties.

Figure A1. Data fusion role in decision analysis.

Figure A2. Remedial decision process supported by data fusion.

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.

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 parameters. 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 improve spatial continuity models 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 improve spatial continuity models 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 and McKenna (1994) about the classical method are important. 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 process 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 postprocessing mode, rather than in one fusion 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 and McKenna (1994) 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. (1990) 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.

6.2 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 judgment 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 A1.

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 Porter, 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 solves the numerical computational burden problem of the data assimilation methods.

6.3 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, 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 while heterogeneities are viewed 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 principle models used for inverse modeling. Consequently, the Kalman filter produces a solution that honors the data within measurement error noise, the first principle 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 model errors. 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. (1991) 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.

6.4 Data Fusion

Our data fusion approach provides Bayesian inverse modeling as shown in Figure A3. It begins with prior knowledge about the state variables to be estimated in the form of first principle 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 improve prior statistics such as spatial correlation distances and standard deviations.

Figure A3. Data fusion uses Bayesian statistics.

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 reformulate the description 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 process data sequentially, 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. Statistical physics provides the replacement for causality in the concept of a Markov Random Field (MRF) which makes computations practical.

6.5 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.

6.6 Square Root Information Smoother (SRIS)

We represent 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 principle models and statistical correlations as pseudodata as if it were just more data for doing Bayesian estimation. Since the pseudodata 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 pseudodata 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 Gauss-Newton backtracking with the Trust Region method as a backup for numerical optimization (Vandergraft, 1985).

6.7 Summary

The preceding sections presented a qualitative description of data fusion theory and provided a historical setting and hydrogeological foundation for fusion modeling of groundwater systems. Further details are given in HydroGeoLogic and Coleman, 1996.

In summary, data fusion is a tool that allows the user to process geological and hydrological data from different types of invasive, noninvasive, and remote sensors to optimally estimate geologic and hydrogeological properties. The definition of data fusion is thus, "the simultaneous inversion of one or more measurement types under spatial continuity constraints in order to estimate geologic and hydrological parameters." In order to perform the data fusion; measurements from the different types of sensors (the preprocessed sensor data), models that provide the relationship between the data and the model parameters (the forward models), and the conceptual site models (developed by geologists and other scientists familiar with the site) are required. The goal is then to optimally estimate the model parameters by minimizing the difference between actual measurements and model predicted values (the residual). This is actually done by a least squares method. The model parameter estimates computed by data fusion are optimal in the sense of minimizing the sum of the squares of the residuals for the measurements, and also for the pseudomeasurements that account for the spatial continuity models, and for the prior information on model parameters.

7. References

- Bierman, G.J., "Factorization Methods for Discrete Sequential Estimation," New York: Academic Press (1977).
- Carrera, J., and L. Glorioso, "On Geostatistical Formulations of the Ground Water Flow Inverse Problem," *Advances in Water Resources*, **14**(5), 273-283 (1991).
- Chellappa, R., and A. Jain, "Markov Random Fields: Theory and Application," Boston: Academic Press (1991).
- Clark, J.J., and A.L. Yuille, "Data Fusion for Sensory Information Processing Systems," Boston: Kluwer Academic Publishers (1990).
- Coleman Research Corporation, "Geophysical Data Fusion for Subsurface Imaging," CMR 95-293 (1995a).
- Coleman Research Corporation, "Data Fusion Workstation," CMR 195-317 (1995b).
- Coleman Research Corporation, "Groundwater Model Calibration at Pantex Using Data Fusion Modeling," CMR/95-239, prepared for DOE EM-50 (1995c).
- Coleman Research Corp. and HydroGeoLogic Inc., "Flow and Transport Model of the Savannah River Site Old Burial Grounds Using Data Fusion Modeling (DFM)," CMR/95-326, prepared for DOE EM-50 (1995d).
- Constable, S.C., R.I. Parker, and C.G. Constable, "Occam's Inversion: A Practical Algorithm for Generating Smooth Models From Electromagnetic Sounding Data," *Geophysics*, **52**(3), 289-300 (1987).
- Courtier, P., J. Derber, R. Errico, J.F. Louis, and T. Vukicevic, "Important Literature on the Use of Adjoint, Variational Methods, and the Kalman Filter in Meteorology," *Tellus*, **45A**, 342, 357 (1993).
- Freeze, R.A., J. Massmann, L. Smith, T. Sperling, and B. James, "Hydrogeological Decision Analysis: 4, The Concept of Data Worth and Its Use in the Development of Site Investigation Strategies," *Ground Water* (July - August 1992).
- Freeze, R.A., J. Massmann, L. Smith, T. Sperling, and B. James, "Hydrogeological Decision Analysis: 1, A Framework," *Ground Water* (September - October 1990).
- Geman, S., and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 721-741 (1984).
- Ghil, M., "Meteorological Data Assimilation for Oceanographers: Part I, Description and Theoretical Framework," *Dynamics of Atmospheres and Oceans*, **13**, 171-218 (1989).
- HydroGeoLogic, Inc., and Coleman Research Corporation, "Data Fusion Modeling of Lithology and Groundwater Flow at the Letterkenny Army Depot Disposal Area and Northern Southeast Industrial Area," prepared for U.S. Army Environmental Center, paginated in sections (June 1996).
- Hydro-FACT User's Guide, "A Practical Hydrological Model Calibration System for Heterogeneous Sites," Fusion and Control Technology, Inc., 101 pp. (September 1997).
- Inman, G.R., "Resistivity Inversion with Ridge Regression," *Geophysics*, **40**, No. 5, pp. 798-817 (1975).
- Massmann, J., R.A. Freeze, L. Smith, T. Sperling, and B. James, "Hydrogeological Decision Analysis: 2, Applications to Ground Water Contamination," *Ground Water* (July - August 1991).
- McLaughlin, D., "Recent Developments in Hydrologic Data Assimilation," prepared for the U.S. Report to the IUGG (June 1994).

Poeter, E.P., and S.A. McKenna, "Geostatistical Simulation and Inverse Flow Modeling to Reduce Uncertainty Associated with Flow and Transport Predictions," Proceedings of the 1994 Ground Water Modeling Conference, Fort Collins, CO, International Ground Water Modeling Center Publication (1994).

Porter, D.W., "Quantitative Data Fusion: A Distributed/Parallel Approach to Surveillance, Tracking, and Navigation Using Information Filtering," Fifth Joint Service Data Fusion Symposium, Johns Hopkins University/Applied Physics Laboratory (October 1991).

Porter, D.W., B.P. Gibbs, W.F. Jones, P.S. Huyakorn, L.L. Hamm, and G.P. Flach, "Data Fusion Modeling for Groundwater Systems," Proceedings of the Advanced Simulation of Subsurface Flow and Contaminant Transport Conference, Research Triangle Park, NC (December 1996).

Rubin, Y., G. Mayko, and J. Harris, "Mapping Permeability in Heterogeneous Aquifers Using Hydrologic and Seismic Data," *Water Resources Research*, **18**(7) (1992).

Sperling, T., R.A. Freeze, J. Massmann, L. Smith, and B. James, "Hydrogeological Decision Analysis: 3, Applications to Design of a Ground Water Control System at an Open Pit Mine," *Ground Water* (May - June 1992).

Vandergraft, J.S., "Efficient Optimization Methods for Maximum Likelihood Parameter Estimation," Proc. 24th Conf. on Decision and Control, Ft. Lauderdale, FL (December 1985).

Whittle, P., "On Stationary Processes in the Plane, *Biometrika*," Vol. 41, 434-449 (1954).