

SOLVING INVERSE PROBLEMS BY BAYESIAN ITERATIVE INVERSION OF A FORWARD MODEL WITH GROUND TRUTH INCORPORATION*

Daniel T. Davis, Jenq-Neng Hwang
Information Processing Laboratory
Department of Electrical Engr., Box # 352500
University of Washington
Seattle, Washington 98195
davisd@pierce.ee.washington.edu, hwang@ee.washington.edu

ABSTRACT

Inverse problems have been often considered ill-posed, i.e., the statement of the problem does not thoroughly constrain the solution space. In this paper we take advantage of this lack of information by adding informative constraints to the problem solution using Bayesian methodology. Remote sensing problems afford opportunities for inclusion of ground truth information, prior probabilities, noise distributions, and other informative constraints within a Bayesian probabilistic framework. We apply Bayesian methods to a synthetic remote sensing problem, showing that the performance is superior to a previously published method of iterative inversion of neural networks. In addition, we show that the addition of *ground truth* information, naturally included through Bayesian modeling, provides a significant performance improvement.

1. INTRODUCTION

Remote sensing problems are of the general class of inverse problems, where we have a measurement vector \mathbf{m} (e.g., multispectral active or passive microwave measurements) arising from some physical process $\phi()$ acting on a (geophysical) parameter vector \mathbf{x} (e.g., temperatures, moistures, vegetation indices, etc), and we wish to infer the parameter vector \mathbf{x} from the observed measurement vector \mathbf{m} .

Satellite remote sensing has the additional feature that a whole set of measurements $\{\mathbf{m}\}$, over some region denoted by positions $\{\mathbf{p}\}$, are to be inverted to their resulting $\{\mathbf{x}\}$. Figure 1 details the different quantities and information sources available in a remote sensing problem. The parameter vector \mathbf{x} and the measurement vector \mathbf{m} are related by some physical process $\mathbf{m} = \phi(\mathbf{x})$, or more commonly by $\mathbf{m} = \phi(\mathbf{x}) + \mathbf{n}$, where \mathbf{n} denotes some channel or sensor noise vector from the physical process. Remote sensing problems are especially ripe for Bayesian methods because the \mathbf{x}_i are in general not independent, i.e., they vary smoothly according to their positions \mathbf{p}_i .

There often exist certain *ground truth* values for any particular problem. This ground truth information takes a couple of distinctive forms. We have contour ground truth (*cgt*) when we know the parameters \mathbf{x}_j for particular locations \mathbf{p}_j . We have model ground truth when we have a limited number of data pairs of \mathbf{x} linked to the resulting \mathbf{m} , preferably having a greater accuracy than that provided by the approximate analytic model $\hat{\phi}$ of the underlying physical process is $\mathbf{m} = \phi(\mathbf{x})$. Bayesian methodology allows meaningful and rigorous incorporation of each of these information sources into the inverse problem solution.

Neural networks have been used for geophysical parameter retrieval, and in initial work with neural networks, an explicit inverse process was used [14, 2, 8]. To deal with the many-to-one mapping problems, we later used an iterative constrained inversion technique on the forward mapping, which more accurately represents the functional relationship $\mathbf{m} = \phi(\mathbf{x})$ [15, 3].

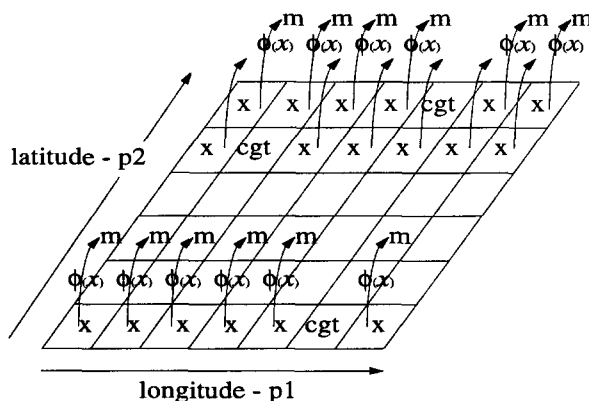


Figure 1. Information sources available in a remote sensing problem: contour ground truth points (*cgt*), measurements \mathbf{m} , knowledge of the physical process $\mathbf{m} = \phi(\mathbf{x}) + \mathbf{n}$, knowledge of the noise \mathbf{n} , and positions $(\mathbf{p}_1, \mathbf{p}_2)$.

In this paper, we apply Bayesian modeling to the inverse parameter retrieval problem. A Bayesian approach was first introduced by Besag [1] in the context of an image restoration problem. In the Bayesian approach, the parameter retrievals are performed by maximizing a posterior probability. The posterior probability is broken down into smaller, physically meaningful conditional probabilities.

In Section 2., we describe the Bayesian model for remote sensing problems in terms of conditional probabilities that take into account ground truth information and parameters determined at neighboring sites. The Bayesian framework allows for a Bayesian Iterative Inversion of a neural network, shown to have superior performance to that of iterative inversion [6, 4]. We evaluate the Bayesian Iterative Inversion method including the use of both model and contour ground truth.

In Section 3., we compare the performance of Bayesian Iterative Inversion technique with and without ground truth, to the performance of Iterative Inversion and the use of an explicit inverse. The Bayesian methods are found clearly superior, with the incorporation of a limited number of ground truth points significantly enhancing the performance of the Bayesian Iterative Inversion method.

2. BAYESIAN ITERATIVE INVERSION

We use a *maximum a posterior* (MAP) framework based on Bayesian analysis to estimate the optimal inverse geophysical parameters for remote sensing applications.

2.1. Derivation of the Bayesian Framework

The framework has a close relationship with previous work of Besag on Bayesian methods applied to an image restoration problem [1]. Let i be the index of the sites in the area of interest. Let \mathbf{x}_i and \mathbf{m}_i , respectively be the geophysical parameters and measurements at the i^{th} site. The sets $\{\mathbf{x}\}$ and $\{\mathbf{m}\}$ denote the parameters and measurements at the sites of interest. Let $f(\{\mathbf{x}\}|\{\mathbf{m}\})$ be the conditional probability of the set of parameters $\{\mathbf{x}\}$ given the set of measurements $\{\mathbf{m}\}$. In Bayesian

*THIS RESEARCH WAS PARTIALLY SUPPORTED THROUGH GRANTS FROM NASA UNDER CONTRACT NO. NAGW-1702 AND NO. NAGW-4251.

inversion, we want to find the $\{x\}$ which maximizes the posterior probability $f(\{x\}|\{m\})$.

This may be performed iteratively if we note that by applying a simple Bayesian analysis,

$$\begin{aligned} f(\{x\}|\{m\}) &= f(x_i, \{x_{j \neq i}\} | m_i, \{m_{j \neq i}\}) \\ &= f(x_i | \{x_{j \neq i}\}, m_i, \{m_{j \neq i}\}) f(\{x_{j \neq i}\} | m_i, \{m_{j \neq i}\}) \end{aligned} \quad (1)$$

where the second term is independent of x_i , so maximization of $f(x_i | \{x_{j \neq i}\}, m_i, \{m_{j \neq i}\})$ will always monotonically increase $f(\{x\}|\{m\})$. If we consider each x_i as the physical cause of each associated m_i , we may write:

$$f(x_i | \{x_{j \neq i}\}, m_i, \{m_{j \neq i}\}) = f(x_i | \{x_{j \neq i}\}, m_i). \quad (2)$$

If we let $x_{s/i}$ denote the set of parameter vectors associated with the neighboring sites of the i^{th} site, and assume a standard Markov random field (MRF) assumption [9] that conditioned on $x_{s/i}$, i.e., x_i is conditionally independent of any $\{x_{j \neq i}\}$ not contained in the neighborhood set $x_{s/i}$, we may conclude that

$$f(x_i | \{x_{j \neq i}\}, m_i) = f(x_i | x_{s/i}, m_i). \quad (3)$$

From Eqs. (1)-(3), it follows that to maximize $f(\{x\}|\{m\})$, it is sufficient to iteratively select each site p_i , and estimate the parameters x_i which maximize the posterior probability $f(x_i | m_i, x_{s/i})$. The method of maximizing $f(\{x\}|\{m\})$ through iterative maximization of each $f(x_i | m_i, x_{s/i})$ is called iterated conditional modes (ICM) [1].

By using Bayes theorem, we convert $f(x_i | m_i, x_{s/i})$ into a number of smaller, physically meaningful conditional probabilities:

$$\begin{aligned} \max_{x_i} f(x_i | m_i, x_{s/i}) &= f(m_i, x_{s/i} | x_i) f(x_i) / f(m_i, x_{s/i}) \\ &= f(m_i | x_i) f(x_{s/i} | x_i) f(x_i) / f(m_i, x_{s/i}) \end{aligned} \quad (4)$$

$$\propto f(m_i | x_i) f(x_{s/i} | x_i) f(x_i) \quad (5)$$

$$\propto f(m_i | x_i) f(x_{s/i} | x_i) f(x_i) \quad (6)$$

where \propto denotes proportional to. Since the maximization is with respect to x_i , any terms without x_i can be dropped out.

Note that we are now left with a simple maximization problem on $f(x_i | m_i, x_{s/i})$. A vast array of search techniques may be brought to bear, including simulated annealing, Gibbs sampling, conjugate gradient, gradient descent, and gradient descent with momentum. The simulations in this paper are performed with gradient descent with momentum.

2.2. Construction of the Different Conditional Probabilities in the Bayesian Model

The three probability distributions, the sensor noise and model mismatch distribution $f(m_i | x_i)$, the neighborhood distribution $f(x_{s/i} | x_i)$, and the prior distribution $f(x_i)$, when multiplied together, are proportional to $f(x_i | m_i, x_{s/i})$, and so allow us to iteratively update the x_i . Figure 2 illustrates the relationship of the different distributions. The neural network $\hat{\phi}(x_i)$ operates within the sensor noise and model mismatch distribution.

The neighborhood distribution, $f(x_{s/i} | x_i)$, can adopt the standard Markov random field (MRF) modeling under the Gibbs distribution formulation [9] or by the probabilistic neural network modeling proposed by Hwang [11]. Once a system is up and running which reproduces terrains from measurements, the reconstructed terrains could be used to generate these densities through statistical density estimation [10].

In the simulations of this paper, $x_{s/i}$ is the collective set of parameters $\{x_j\}$ associated with the eight neighbors whose geographical locations $\{p_j\}$ are adjacent to that of the i -th site p_i , and $f(x_{s/i} | x_i)$ is modeled as a product of independent Gaussian $f(x_j | x_i)$, with mean $\mu_{mm} = x_i$ and covariance matrix Σ_{mm} for each point x_j in $x_{s/i}$.

The probability $f(x_i)$ is called the *prior distribution* of x_i . It contains information of the *a priori* probability distribution of the parameters at the site i . In our simulations, we model

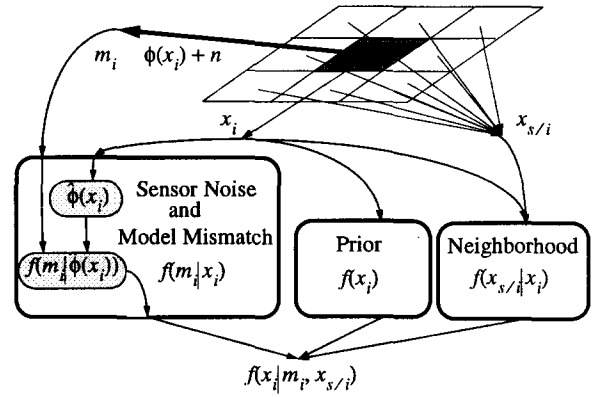


Figure 2. The Pieces of the Bayesian Model: the sensor noise and model mismatch distribution $f(m_i | x_i)$, the neighborhood distribution $f(x_{s/i} | x_i)$, and the prior distribution $f(x_i)$.

$f(x_i)$ as a Gaussian with mean μ_p and covariance Σ_p . These parameters of the prior, μ_p and Σ_p , can be made to vary according to position p , giving site dependent priors, as in [6]. In this paper, μ_p and Σ_p will be constant over all positions p .

$f(m_i | x_i)$, the sensor noise and model mismatch distribution, is the conditional probability of the measurement m_i given the parameters x_i . Let $\phi(x_i)$ be the true physical process that would give m_i in the absence of sensor noise. Thus $m_i = \phi(x_i) + n$, where n denotes noise in the remote sensing process.

We approximate $\phi(x_i)$ by a simpler model used to train a neural network, and denoted by $\hat{\phi}(x_i)$. Thus there would be a model mismatch error of $\phi(x_i) - \hat{\phi}(x_i)$. We model the measurement noise n with density $f(m_i | x_i, \phi(x_i))$ and the model mismatch error $\phi(x_i) - \hat{\phi}(x_i)$ with density $f(\phi(x_i) | x_i, \hat{\phi}(x_i))$, as independent Gaussian processes. Expanding $f(m_i | x_i, \hat{\phi}(x_i))$, we find:

$$\begin{aligned} f(m_i | x_i, \hat{\phi}(x_i)) &= \int f(m_i | x_i, \hat{\phi}(x_i), \phi(x_i)) f(\phi(x_i) | x_i, \hat{\phi}(x_i)) d\phi(x_i) \quad (7) \\ &= \int f(m_i | x_i, \phi(x_i)) f(\phi(x_i) | x_i, \hat{\phi}(x_i)) d\phi(x_i). \quad (8) \end{aligned}$$

Thus the conditional probability $f(m_i | x_i)$ is a Gaussian distribution on the measurement m_i with a mean of the model output $\hat{\phi}(x_i)$ and a covariance matrix Σ for which $\Sigma = \Sigma_{\hat{\phi}} + \Sigma_n$. $\Sigma_{\hat{\phi}}$ and Σ_n denote covariance matrices for model mismatch error, $\phi(x_i) - \hat{\phi}(x_i)$, and sensor error, $m_i - \phi(x_i)$, respectively.

Note that it is during the maximization of each individual $f(x_i | m_i, x_{s/i})$ that it becomes necessary to take the gradient of a function of $\hat{\phi}(x_i)$. This is where a neural network is applied in our algorithm.

2.3. Putting Model Ground Truth into the Bayesian Formulation

In formulating our solution to inverse remote sensing problems, we have yet to use model ground truth. Use of model ground truth requires a reformulation of the sensor noise and model mismatch distribution to incorporate two different conditional density estimates of $f(m_i | x_i)$.

With the addition of model ground truth, we have two information sources available to evaluate $f(m_i | x_i)$. The first is an approximate analytic model $\hat{\phi}(x_i)$, while the second is the limited number of model ground truth points $D = \{(x^i, m^i)\}$.

Reformulating $f(m_i | x_i)$, this time conditioning the density on $\hat{\phi}(x_i)$ and D , we estimate $f(m_i | x_i, D, \hat{\phi}(x_i))$ as:

$$\begin{aligned} f(m_i | x_i, D, \hat{\phi}(x_i)) &= C * \int f(m_i | x_i, \phi(x_i)) \\ &\quad f(\phi(x_i) | D, x_i) f(\phi(x_i) | \hat{\phi}(x_i), x_i) d\phi(x_i) \end{aligned} \quad (9)$$

where C is a normalization term, constant with respect to x_i . In other words, the two density estimates of ϕ are multiplied together and normalized to find their combined prediction. Due to the availability of very limited number (true in most remote sensing applications) of ground truth points, which make it very difficult to use standard Gaussian kernel density estimation (often called mixtures of Gaussian) [13] to create an accurate density estimate $f(\phi(x_i)|D, x_i)$, we use Expanding Gaussian Kernel (EGK) density estimation, shown in [5, 7], to have superior performance to standard Gaussian kernel density estimation as measured by the Kullback-Leibler criterion.

3. PERFORMANCE EVALUATION ON A SYNTHETIC INVERSE REMOTE SENSING PROBLEM

We now compare the performance of the different methods of inverse remote sensing parameter mapping to Bayesian Iterative Inversion. In particular, we compare the performance of an explicit inverse, iterative inversion of a trained forward model, and Bayesian Iterative Inversion, with and without ground truth information. We use the results of [5, 7] to incorporate model ground truth information into Bayesian Iterative Inversion, showing that a limited number of ground truth points can significantly enhance performance.

3.1. Definition of the Parameter Mapping and Underlying Functions

The Bayesian framework for inverse parameter mapping is fairly complete, and therefore requires special attention to detail when constructing the simulation environment. The design of the problem proceeds as follows. First, we define the original parameter mappings of x over a set of locations p . Figure 3(a) shows the contours, plotted over their 2-D positions $p = [p_1, p_2]$, for a three-dim parameter vector $x = [x_1, x_2, x_3]$.

The contours chosen are fairly simple, although by design they afford ample opportunity to become trapped in local minima if a search process begins too far from the true parameter x .

Next, an approximate model $\hat{\phi}()$ is created as $m = \hat{\phi}(x) = [m_1, m_2]$:

$$m_1 = 0.3(x_1 + x_2) + 0.3x_2^2 + 0.15 + n_1, \quad (10)$$

$$m_2 = 0.8(x_1 - x_3)^2 + 0.3 + n_2. \quad (11)$$

The measurement noise distributions n_1 and n_2 were chosen as independent Gaussian noise of standard deviation 0.02, thus defining Σ_n as a diagonal covariance matrix with values of 0.004 along the diagonal.

Since we thought that approximate models should be simpler in form than the true physical process $\phi(x)$, we subtracted a model mismatch term $e = [e_1, e_2]$ from $m = \hat{\phi}(x)$ to construct $\phi(x)$, $\phi(x) = \hat{\phi}(x) - e(x)$,

$$e_1 = 0.08e^{-2(x_3 - x_2)^2}, \quad (12)$$

$$e_2 = 0.09\sin[2\pi(x_1 + x_2)]. \quad (13)$$

$\Sigma_{\hat{\phi}}$ was calculated as the sample covariance from a symmetric set of data matching any $e(x)$ with a supposed negative sample $-e(x)$. The original $e(x)$ were sampled from the x parameters used in the true parameter mapping. Similarly, the prior mean μ_p and covariance Σ_p , and the neighborhood model covariance Σ_{mm} , were also calculated from the values of the true original parameter map.

J_a , as needed in the EGK method, was taken directly from the definition of $\phi(x)$, setting the appropriate terms of J_a equal to the maximum absolute value of the appropriate partial derivative of m with respect to x for $\phi(x)$ [5, 7].

20 positions p were selected at random as "contour ground truth points", yielding 20 positions with known x . Matching these known parameters x with their the corresponding measurements m , yielded 20 "model ground truth points".

The iterative methods of parameter mapping retrieval require initial starting points. We use a distance weighted average of

Table 1. Average Absolute Error for Retrieved Parameter Maps

Method	x_1	x_2	x_3
Explicit Inverse	.1722	.1227	.1367
Initial Points	.1063	.1221	.0892
Iterative Inversion	.1648	.1045	.1509
Bayes w/o GT	.1023	.1138	.0866
Bayes with GT	.0459	.0508	.0575

the contour ground truth points [3] to start the different iterative methods. To isolate the effect of incorporation of ground truth, no ground truth points are used during the initial iterative convergence of Bayesian Iterative Inversion. Eliminating the ground truth during convergence keeps them from affecting the final result through the neighborhood distribution or the sensor noise and model mismatch distribution. Since Bayesian Iterative Inversion with a small amount of model ground truth is effectively a perturbation on the results without using ground truth, we use the parameter map retrieved by Bayesian Iterative Inversion without ground truth as the initial starting point when using Bayesian Iterative Inversion with model ground truth.

2000 (x^i, m^i) training pairs were created from sampling the function $\phi(x)$. Two, two-layer MLPs with 40 hidden neurons each, were trained with backpropagation [12] to approximate $\hat{\phi}^{-1}(x)$ and $\phi(x)$, thus providing the explicit inverse, and trained forward model used in the iterative techniques, respectively.

3.2. Results of Parameter Map Retrieval

To quantitatively compare all methods, the average absolute error between the original parameter map and the retrieved parameter map, for each dimension of the parameter vector x , was tabulated in Table 1.

To qualitatively compare all the methods, the original parameter map, and all retrieved parameter maps, are plotted in Figure 3.

The explicit inverse and iterative inversion managed to produce poorer performance than the initial guesses created with the contour ground truth points. We see that we have indeed designed the problem to be difficult for the iterative techniques, although iterative inversion does capture the gross characteristics of the contours, which the explicit inverse fails to do.

Bayesian Iterative Inversion without ground truth is clearly superior to the previous methods, avoiding their pitfalls, and enhancing performance over the initial guesses. The real performance improvement comes with the incorporation of ground truth points, which cuts error rates approximately in half.

4. CONCLUSION

Using Bayesian methodology, we exploit information relevant to an inverse problem, helping to narrow down the many in many-to-one inverse problems. Bayesian modeling gains much of its power from its ability to isolate and incorporate causal models as conditional probabilities. Remote sensing problems afford opportunities for inclusion of ground truth information, prior probabilities, noise distributions, and other informative constraints within a Bayesian probabilistic framework. The Bayesian methodology also goes a long way toward solving some of the outstanding issues with inverse problems in remote sensing. The many-to-one mapping problem is dealt with by introducing more informative constraints through probability distributions. The neural network training can deal with non-linear relations between parameters and measurements. The maximization of various conditional probabilities serves as a smoothing process so that the inversion is stable. Finally, the addition of *ground truth* information, naturally included through Bayesian modeling, provides a significant performance improvement.

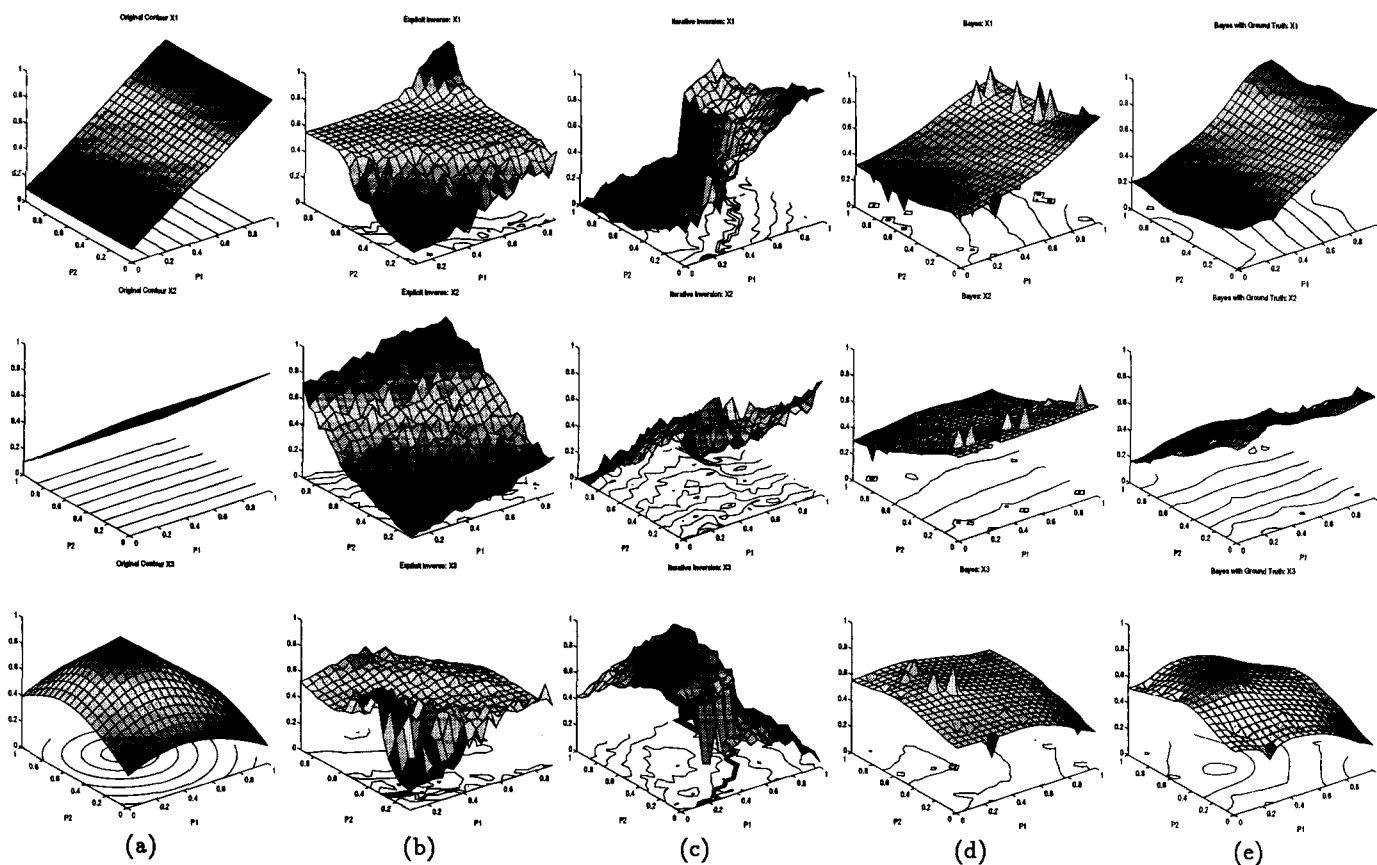


Figure 3. Plotting x versus p . (a) The original parameter map. The parameter maps retrieved by (b) an Explicit Inverse (c) Iterative Inversion (d) Bayesian Iterative Inversion (e) Bayesian Iterative Inversion with Ground Truth Incorporation.

REFERENCES

- [1] J. Besag. "On the Statistical Analysis of Dirty Pictures," *J. R. Statist. Soc., series B*, (48):259-302, 1986
- [2] Chang, A.T.C., Tsang, L., "A Neural Network Approach to Inversion of Snow Water Equivalent for Passive Microwave Measurements," *Nordic Hydrology*, 23, pp. 173-182, 1992
- [3] D. T. Davis, Z. Chen, L. Tsang, J. N. Hwang, A. T. C. Chang. "Retrieval of Snow Parameters by Iterative Inversion of a Neural Network," *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 31, Number 4, pp. 842-852, July 1993.
- [4] D. T. Davis, J. N. Hwang, L. Tsang, "Solving Inverse Problems Using Bayesian Modeling to Incorporate Information Sources," *Int'l Geo-Science and Remote Sensing Symposium*, pp. 1395-1397, Pasadena, CA, August 1994.
- [5] D. T. Davis, *Solving Inverse Problems using Bayesian Modeling to Incorporate Information Sources*, Ph.D. Dissertation, University of Washington, Seattle, June 1995.
- [6] D. T. Davis, Z. Chen, J. N. Hwang, L. Tsang, and E. Njoku. "Solving Inverse Problems by Bayesian Iterative Inversion of a Forward Model with Application to Parameter Mapping Using SMMR Remote Sensing Data," *IEEE Trans. on GeoScience and Remote Sensing*, 33(5):1182-1193, September 1995.
- [7] D. T. Davis, J. N. Hwang, "Expanding Gaussian Kernels for Multivariate Conditional Density Estimation," *ICASSP*, pp. VI 3525-3528, Atlanta, GA, May 1996.
- [8] M. S. Dawson, A. K. Fung. "Neural Networks and Their Applications to Parameter Retrieval and Classification," *IEEE Geoscience and Remote Sensing Society Newsletter*, pp 6-14, September 1993.
- [9] S. Geman and D. Geman. "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images," *IEEE Trans. on PAMI*, 6:721-741, 1984.
- [10] J. N. Hwang, S. R. Lay, A. Lippman "Nonparametric Multivariate Density Estimation: A Comparative Study," *IEEE Trans. on Signal Processing*, pp. 2795-2810, October 1994.
- [11] J. N. Hwang, "Textured Images Modeling and Segmentation," *The Handbook of Brain Theory and neural Networks*, ed. Michael A. Arbib, pp. 971-976, MIT Press, 1995.
- [12] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning Internal Representations by Error Propagation. Parallel Distributed Processing (PDP): Exploration in the Microstructure of Cognition," Vol. 1, Chapter 8, pp. 318-362. MIT Press, Cambridge, Massachusetts, 1986.
- [13] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall, 1986.
- [14] L. Tsang, Z. Chen, Seho Oh, R.J. Marks II, and A.T.C. Chang, "Inversion of Snow Parameters from Passive Microwave Remote Sensing Measurements by a Neural Network Trained with a Multiple Scattering Model," *IEEE Trans. on Geoscience and Remote Sensing*, vol 30, no. 5, pp. 1015-1024, Sept., 1992.
- [15] L. Zurk, D. T Davis, E. G. Njoku, L. Tsang, J. N. Hwang. "Inversion of Parameters for Semiarid Regions by A Neural Network," *IEEE International Geoscience and Remote Sensing Symposium*, pp. 1075-1077, Houston, Texas, May 1992.