Solving Inverse Problems by Bayesian Iterative Inversion of a Forward Model with Applications to Parameter Mapping using SMMR Remote Sensing Data

Daniel T. Davis, Zhengxiao Chen, Jenq-Neng Hwang, Leung Tsang
Department of Electrical Engineering, FT-10
University of Washington
Seattle, Washington 98195
davisd@ee.washington.edu
and
Eni Njoku
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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 additional informative constraints to the problem solution using Bayesian methodology. Bayesian modeling gains much of its power from its ability to isolate and incorporate causal models as conditional probabilities. As causal models are accurately represented by forward models, we convert implicit functional models into data driven forward models represented by neural networks, to be used as engines in a Bayesian modeling setting. 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 first apply these 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. Next, microwave brightness temperatures obtained from the Scanning Multichannel Microwave Radiometer (SMMR) over the African continent are inverted. The values of soil moisture, surface air temperature and vegetation moisture retrieved from the inversion produced contours that agree with the expected trends for that region.

1 Inverse Problems In Remote Sensing

Remote sensing problems are of the general class of inverse problems, where we have a measurement vector $\mathbf{m}$ from which we wish to infer the parameter vector $\mathbf{x}$ that gave rise to it. The inverse problem is difficult for the following reasons. First, the inverse mapping is very often a many-to-one mapping, with more than one parameter $\mathbf{x}$ which could account for the observed measurement $\mathbf{m}$. Second, the relation between remote sensing measurements and the medium parameters is highly nonlinear. In the past, the simplifying approximation of single scattering is used so that the scattering measurements are linearly related to the medium geophysical parameters, allowing easy inversion of parameter values. Third, the linear inverse problem is often in the form of a Fredholm equation of the first kind, making the method ill-conditioned. Various techniques, such as the regularization method and the Backus-Gilbert inverse techniques have been used to obtain a stable solution [1, 2, 3]. Fourth, the amount of remote sensing measurements is enormous so that it is desirable that the parameter mapping can be done in a speedy manner. Fifth, past solutions of inverse problems merely consisted of matching the remote sensing measurements to the scattering model without using other information sources.

Satellite remote sensing has the additional feature that a whole set of $\{\mathbf{m}\}$, 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 perhaps 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. In addition, there often exist certain ground truth values for any particular problem. This ground truth information takes a couple of distinctive forms. One form is model ground truth. In general, $\mathbf{m} = \phi(\mathbf{x})$ will be modeled according to an approximate analytic model $\hat{\phi}$. We obtain model ground truth when we have experimental measurements of $\mathbf{x}$ linked to the resulting $\mathbf{m}$. Similarly, we have contour ground truth (*cgt*) when we know the parameters $\mathbf{x_j}$ for particular locations $\mathbf{p_j}$. Bayesian methodology allows meaningful and rigorous incorporation of each of these information sources into the inverse problem solution.

Recently, neural network processes have been used for parameter retrieval [4, 5, 6]. Scattering models are used to train the neural network to the nonlinear relations between parameters and computed brightness temperatures and scattering coefficients. Once the neural network is trained, parameter inversion can be performed speedily. The parameter vector $\mathbf{x}$ and the measurement vector $\mathbf{m}$ are related by a physical process $\mathbf{m} = \phi(\mathbf{x})$. In initial work with neural networks, an explicit inverse process was used [6, 4, 7]. This assumes that for a measurement $\mathbf{m}$, a unique inverse $\phi^{-1}(\mathbf{m})$ exists that will yield $\mathbf{x}$. This does not work well where the mapping is many-to-one. To deal with the many-to-one mapping, we later used an iterative constrained inversion technique [8, 5].

In the iterative constrained inversion, we use the multilayer perceptron (MLP) neural network to facilitate solving the inverse problem. The MLP is first trained to match the input-output relationship of a physical model, with input of geophysical parameters and output of measurements. With the trained MLP, a parametric representation is established so that the gradients of the geophysical parameters with respect to the measurements can be taken, giving a means to perform a directed search of the input space to locate a set of geophysical parameters which could account for the given set of measurements. The process of searching the input space of a MLP using the gradient of the output error is known as iterative inversion [9, 10].

Iterative inversion allows use of many different cost functions, and the imposition of many different constraints. The use of a neural network also gives the opportunity to incorporate model ground truth data into the training of the model, giving us a synthesis of theoretical and measured data in the trained forward model [11]. Iterative inversion of a trained forward model was shown superior to using an explicit inverse to perform the inverse problem [5].

While iterative inversion gives improved performance over an explicit inverse, it is in some ways unsatisfactory. The mechanisms available to control the iterative inversion process, the starting point, cost function, and stopping criteria, appear inadequate for the multitude of informational sources we would like to include in the process. Iterative inversion of a neural network has no straightforward way to take model mismatch error and sensor error into account, while incorporation of a priori knowledge, such as contour ground truth data and smoothness constraints, can only be performed on a very ad hoc basis. Finally, the mechanisms we have at our disposal in the method of iterative inversion of a neural network have no physical significance, so they can only be adjusted through ad hoc means.

In this paper, we use a Bayesian approach of parameter retrieval. A Bayesian approach was first introduced by Besag [12] in the context of an image restoration problem. In the Bayesian approach, the parameter retrievals are performed by maximizing the posterior probability. The posterior probability is broken down into smaller, physically meaningful conditional probabilities. Bayesian modeling gains much of its power from its ability to isolate and incorporate causal models as conditional probabilities. As causal models are accurately represented by forward models, we convert implicit functional models into data driven forward models that are represented by neural networks. These are then used in a Bayesian modeling setting. In Section 2, we describe the Bayesian model in terms of conditional probabilities that take into account ground truth information and parameters determined at neighboring sites. In Section 3, we describe the methods of formulating the different conditional probabilities in the Bayesian model. In Section 4, we show examples of parameter retrieval based on synthetic data and show that the Bayesian method has superior performance to that of iterative inversion. In Section 5, we perform parameter retrieval using SMMR (Scanning Multichannel Microwave Radiometer) data taken over Africa. The microwave emission model of Kerr and Njoku [13] is used to train the neural network and various conditional probabilities are presented. The parameter mapping of soil moisture, vegetation moisture, and surface air temperature agree with expected trends in Africa.

2 Bayesian Iterative Inversion Using A Neural Network

A Bayesian probabilistic framework allows incorporation of various informational sources in a rigorous and physically meaningful way. We use a *maximum a posterior* (MAP) framework based on Bayesian analysis to estimate the optimal inverse geophysical parameters iteratively for the remote sensing applications. We iteratively search for the "optimal" inverse geophysical parameters $\mathbf{x}$ given the measurements $\mathbf{m}$, the trained neural network representation $\hat{\phi}$ (trained by the data from the theoretical electromagnetic scattering model), the limited amount of ground-truth data, the local continuity property of parameters (usually described as a Markov random field neighborhood formulation), the sensor noise characteristics of remote sensing measurements, etc. The Bayesian framework allows incorporation of many information sources, providing additional constraints for the ill posed inverse problem.

The framework has a close relationship with previous work of Besag on Bayesian methods applied to an image restoration problem [12]. 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 inversion, we want to find the $\{\mathbf{x}\}$ which maximizes the posterior probability $f(\{\mathbf{x}\}|\{\mathbf{m}\})$.

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

$$\begin{aligned} f(\{\mathbf{x}\}|\{\mathbf{m}\}) &= f(\mathbf{x}_i, \{\mathbf{x}_{j\neq i}\}|\mathbf{m}_i, \{\mathbf{m}_{j\neq i}\}) \\ &= f(\mathbf{x}_i|\{\mathbf{x}_{j\neq i}\}, \mathbf{m}_i, \{\mathbf{m}_{j\neq i}\}) f(\{\mathbf{x}_{j\neq i}\}|\mathbf{m}_i, \{\mathbf{m}_{j\neq i}\}) \end{aligned} \tag{1}$$

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

$$f(\mathbf{x}_i|\{\mathbf{x}_{j\neq i}\}, \mathbf{m}_i, \{\mathbf{m}_{j\neq i}\}) = f(\mathbf{x}_i|\{\mathbf{x}_{j\neq i}\}, \mathbf{m}_i) \tag{2}$$

If we let $\mathbf{x}_{s/i}$ denote the set of parameter vectors associated with the neighboring sites of the i^{th} site, and assume a standard Markov random field assumption that $\mathbf{x}_i$ is conditionally independent of any $\{\mathbf{x}_{j\neq i}\}$ not contained in the neighborhood set $\mathbf{x}_{s/i}$ (a standard Markov random field assumption), we may conclude that

$$f(\mathbf{x}_i|\{\mathbf{x}_{j\neq i}\}, \mathbf{m}_i) = f(\mathbf{x}_i|\mathbf{x}_{s/i}, \mathbf{m}_i) \tag{3}$$

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

By using Bayes theorem, we convert $f(\mathbf{x}_i|\mathbf{m}_i, \mathbf{x}_{s/i})$ into a number of smaller, physically meaningful conditional probabilities:

$$\begin{aligned} \max_{\forall \mathbf{x}_i} f(\mathbf{x}_i|\mathbf{m}_i, \mathbf{x}_{s/i}) &= f(\mathbf{m}_i, \mathbf{x}_{s/i}|\mathbf{x}_i)\ f(\mathbf{x}_i)/\ f(\mathbf{m}_i, \mathbf{x}_{s/i}) \\ &= f(\mathbf{m}_i|\mathbf{x}_i)\ f(\mathbf{x}_{s/i}|\mathbf{x}_i)\ f(\mathbf{x}_i)/\ f(\mathbf{m}_i, \mathbf{x}_{s/i}) \\ &\propto f(\mathbf{m}_i|\mathbf{x}_i)\ f(\mathbf{x}_{s/i}|\mathbf{x}_i)\ f(\mathbf{x}_i) \\ &\propto f(\mathbf{m}_i|\mathbf{x}_i)\ f(\mathbf{x}_i|\mathbf{x}_{s/i}) \end{aligned}$$

where $\propto$ denotes proportional to. In (4), since the maximization is with respect to $\mathbf{x}_i$, any terms without $\mathbf{x}_i$ can be dropped out.

Note that we are now left with a simple maximization problem on $f(\mathrm{x}_i|\mathbf{m}_i, \mathrm{x}_{s/i})$. 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.

3 Construction of the Different Conditional Probabilities in the Bayesian Model

The three probability distributions, the sensor noise and model mismatch distribution $f(\mathbf{m}_i|\mathrm{x}_i)$, the neighborhood distribution $f(\mathrm{x}_{s/\ i}|\mathrm{x}_i)$, and the prior distribution $f(\mathrm{x}_i)$, when multiplied together, are proportional to $f(\mathrm{x}_i|\mathbf{m}_i, \mathrm{x}_{s/i})$, and so allow us to iteratively update the x_i. Figure 2 illustrates the functioning of the different distributions. The neural network $\hat{\phi}(\mathrm{x}_i)$ operates within the sensor noise and model mismatch distribution. In the following, we describe how these three probability distributions can be constructed.

3.1 The Neighborhood Distribution $f(\mathrm{x}_{s/i}|\mathrm{x}_i)$

The neighborhood distribution, $f(\mathrm{x}_{s/i}|\mathrm{x}_i)$, can adopt the standard Markov random field (MRF) modeling under the Gibbs distribution formulation [14] or by the probabilistic neural network modeling proposed by Hwang [15]. 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 [16].

The distributions $f(\mathrm{x}_{s/i}|\mathrm{x}_i)$ and $f(\mathrm{x}_i)$ are used in combination, since modeling the full distribution $f(\mathrm{x}_i|\mathrm{x}_{s/i})$ would prove very difficult for the modeler if one wants to set all parameters according to physically meaningful intuition. In the simulations of this paper, $\mathrm{x}_{s/i}$ is the collective set of parameters $\{\mathrm{x}_j\}$ associated with the eight neighbors whose geographical locations $\{\mathbf{p}_j\}$ are adjacent to that of the i-th site $\mathbf{p}_i$, and $f(\mathrm{x}_{s/i}|\mathrm{x}_i)$ is modeled as a product of independent Gaussian $f(\mathrm{x}_j|\mathrm{x}_i)$, with mean $\mu_{mm} = \mathrm{x}_i$ and covariance matrix $\mathbf{\Sigma_{mm}}$ for each point x_j in $\mathrm{x}_{s/i}$.

3.2 The Prior $f(\mathrm{x}_i)$

The probability $f(\mathrm{x}_i)$ is called the *prior distribution* of x_i. It contains information of the *a priori* probability distribution of the parameters at the site i. Generally, $f(\mathrm{x}_i)$ depends on the geographical position $\mathbf{p}_i$ of site i. In our simulations, we model $f(\mathrm{x}_i)$ as a Gaussian with mean μ_p and covariance $\mathbf{\Sigma}_p$.

3.3 The Sensor Noise and Model Mismatch Distribution $f(\mathbf{m}_i|\mathrm{x}_i)$

$f(\mathbf{m}_i|\mathrm{x}_i)$ is the conditional probability of the measurement $\mathbf{m}_i$ given the parameters x_i. Let $\phi(\mathrm{x}_i)$ be the true physical process that would give $\mathbf{m}_i$ in the absence of sensor noise. Thus $\mathbf{m}_i = \phi(\mathrm{x}_i) + n$, where n denotes noise in the remote sensing process. However, $\phi(\mathrm{x}_i)$ is difficult to obtain in practice requiring exact solution of Maxwell's equations for very complicated environments of random media and random rough surface. Instead we approximate $\phi(\mathrm{x}_i)$ by a simpler electromagnetic microwave emission model. This electromagnetic model is further used to train a neural network. The resulting neural network representation is denoted by $\hat{\phi}(\mathrm{x}_i)$. Thus there would be a model mismatch error of $\phi(\mathrm{x}_i) - \hat{\phi}(\mathrm{x}_i)$. We model the measurement noise n and the model mismatch error $\phi(\mathrm{x}_i) - \hat{\phi}(\mathrm{x}_i)$ as independent Gaussian processes. Thus the conditional probability $f(\mathbf{m}_i|\mathrm{x}_i)$ is a Gaussian distribution on the measurement $\mathbf{m}_i$ with a mean of the model output $\hat{\phi}(\mathrm{x}_i)$ and a covariance matrix $\mathbf{\Sigma}$ for which $\mathbf{\Sigma} = \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2$. $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ denote covariance matrices for sensor error, $\mathbf{m}_i - \phi(\mathrm{x}_i)$, and model mismatch error, $\phi(\mathrm{x}_i) - \hat{\phi}(\mathrm{x}_i)$, respectively.

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

4 Simulation Results based on Synthetic Data

To test the comparative performance of the proposed Bayesian iterative inversion versus conventional iterative inversion, an artificial landscape of 2-d geophysical parameters $\{\mathbf{x}_i\}$ is created at locations $\{\mathbf{p}_i\}$. They are run through an artificial function $\phi(\mathbf{x})$ and corrupted by Gaussian noise to produce the associated 2-d measurements $\{\mathbf{m}_i\}$. Each output measurement is created with independent Gaussian noise of zero mean and standard deviation 0.01, approximately 1% of the full range of each coordinate of $\{\mathbf{m}_i\}$.

Figure 3 and Figure 4 display the original function $\phi(\mathbf{x})$, and the original landscape of parameters, $\{\mathbf{x}_i\}$ at locations $\{\mathbf{p}_i\}$, respectively.

A 2-layer MLP is trained as a forward model $\mathbf{m} = \hat{\phi}(\mathbf{x})$, training from the geophysical parameters to the measurements.

The various parameters for the Bayesian iterative inverse method are set according to statistics appropriately taken from the given landscape and noise distribution. $\mathbf{\Sigma}_1$ is set according to the noise of standard deviation 0.01, $\mathbf{\Sigma}_2$ is set to zero (no model mismatch error), giving $\mathbf{\Sigma} = \mathbf{\Sigma}_1$. The mean μ_p and covariance matrix $\mathbf{\Sigma}_p$ of the prior $f(\mathbf{x}_i)$ are calculated from the $\{\mathbf{x}_i\}$, as is $\mathbf{\Sigma}_{mm}$, the covariance matrix of the neighborhood distribution.

The initial starting points for both iterative processes are calculated from a weighted average of the 10 contour ground truth points (the ten $\mathbf{x}_j$ corresponding to 10 randomly selected contour ground truth locations $\mathbf{p}_j$) weighted according to distance to each contour ground truth point, as in [5]. The synthetic landscape is then reconstructed using the iterative inversion of the neural network and Bayesian iterative inversion. The results are displayed in Figure 5 and Figure 6, with the error between the reconstructed and original landscapes labeled at the top of each plot. We see that Bayesian iterative inversion, which takes advantage of a prior probability of $\mathbf{x}$, noise statistics, Markov random field properties, and contour ground truth, has superior performance to iterative inversion of a neural network.

Let us note, however, that the Bayesian methodology used does not entirely avoid the many to one problems inherent in inverse problems, it only reduces the likelihood of many to one problems by introducing more informational constraints through probability distributions. Note the large error in the neighborhood of $(p1, p2) \approx (1, 0)$. It appears that for points in that region, the iterative search process became trapped in minima significantly removed from the "true" minima. Applying statistical search methods instead of gradient descent would certainly help alleviate those problems. However, in the general case, when two different parameter sets account equally well for the measurement set, according to the criterion of selection used, there is by definition no meaningful way to choose between the two parameter sets. Hence the importance of introducing many informational constraints as possible into the problem formulation, thus eliminating as many spurious solutions as possible.

Finally, while the results are not displayed here, an explicit inverse was applied to this problem, and it produced extremely poor performance.

5 Application to Parameter Retrieval using SMMR Data over Africa

The problem of recovering geophysical parameters from microwave measurements is next examined. A multi-layer perceptron (MLP) neural network was built, and then trained with data produced from a passive radiative transfer model to give $\hat{\phi}(\mathbf{x}_i)$.

5.1 Microwave Emission Model

The model used to characterize the microwave emission was developed by Kerr and Njoku [13]. This model is a radiative transfer model that takes soil moisture, soil temperature, and vegetation moisture as input parameters and produces as output dual polarized brightness temperatures at 6.6, 10.7, 18, and 37 GHz. These frequencies correspond to those recorded by the satellite Scanning Multichannel Microwave Radiometer (SMMR).

The microwave model depicts the earth's surface as a three layer entity: a soil layer, a vegetation layer, and an atmosphere layer. The brightness temperatures observed by the satellite can originate in emission from any of the three layers. This emission can travel upward towards the satellite, experiencing attenuation as it travels through the layers, or (in the case of atmosphere or canopy emission), it can travel downward and become reflected upward off the soil boundary, also undergoing attenuation as it travels. The components contributing to the satellite observed microwave brightness temperatures are shown in Figure 7. The radiation components are 1) upward atmospheric emission, 2) upward soil-surface emission attenuated through the canopy and atmosphere, 3) downward atmospheric emission reflected off the soil surface and attenuated by the atmosphere and vegetation canopy, 4) upward canopy emission attenuated by the atmosphere. 5) downward canopy emission reflected off soil surface and attenuated through canopy and atmosphere. Thus, the brightness temperature can be written as a sum of these five terms,

$$T_{BP} = T_1 + T_2 + T_3 + T_4 + T_5 \tag{4}$$

where

$$\begin{aligned}
T_1 &= T_{au} && (5)\\
T_2 &= \exp(-\tau_{au})\,\epsilon_{sp}\,T_s\,\exp(-\tau_c) && (6)\\
T_3 &= \exp(-\tau_{au})\,(T_{ad}+T_{sky})\,r_{sp}\,\exp(-2\tau_c) && (7)\\
T_4 &= T_c(1-\omega)\,(1-\exp(-\tau_c))\,\exp(-\tau_{au}) && (8)\\
T_5 &= T_c(1-\omega)\,(1-\exp(-\tau_c))\,\exp(-\tau_c)\,\exp(-\tau_{au})\,r_{sp} && (9)
\end{aligned}$$

and

- τ_{au} and τ_{ad} are upward and downward atmospheric opacities
- T_{au} and T_{ad} are upward and downward atmospheric radiation components
- T_{sky} is cosmic background radiation
- r_{sp} and ϵ_{sp} are surface reflectivity and emissivity, and $\epsilon_{sp} = 1 - r_{sp}$
- T_s is a weighted soil temperature including the effects of subsurface temperature profiles.
- T_c, τ_c and ω are canopy temperature, opacity and single-scattering albedo

In this paper, four frequencies were used (6.6, 10.7, 18 and 37GHz), and for these frequencies the atmosphere was assumed to have no effect $\tau_{au} = \tau_{ad} = T_{au} = T_{ad} = T_{sky} = 0$. With these assumptions, the resulting brightness temperature is

$$T_{BP} = \underbrace{\epsilon_{sp}\,T_s\,\exp(-\tau_c)}_{T_2} + \underbrace{T_c\;(1-\omega)\,(1-exp(-\tau_c))\,exp(-\tau_c)\,r_{sp}}_{T_5} + \underbrace{T_c\,(1-\omega)\,(1-\exp(-\tau_c))}_{T_4}. \tag{10}$$

To compute the satellite brightness temperature using the above equation, the vegetation state must be specified in terms of its opacity and temperature, while the soil must be given in terms of its emissivity and weighted temperature T_s. Where canopy exists, T_c is assumed equal to the surface air temperature T_{airs}. The microwave parameters for canopy opacity were expressed in terms of geophysical parameters by using a form introduced by Kirdiashev [17] and adjusted by Kerr and Njoku [13] to correlate with empirical data published in the literature :

$$\tau_c = A\,f\,W\,\epsilon_{sw}^{''}\sec(\theta)/.3 \tag{11}$$

where A is a canopy model coefficient related to vegetation type, f is the microwave frequency, W is the water content per unit area (kg/m^2), $\epsilon_{sw}^{''}$ is the imaginary part of the dielectric constant of saline water in the vegetation, and θ is the microwave viewing angle. The expression for canopy brightness temperature can also incorporate a measure of canopy cover to take into account the fractional coverage of vegetation in semiarid regions.

Freq.	T_s	(1-ω_h)	(1-ω_v)	Q	h
6.6	$T_{air\,s} - 8.5$	0.875	0.95	0.0	1.7
10.7	$T_{air\,s} - 5.5$	0.87	0.953	0.14	1.5
18	$T_{air\,s} - 3.0$	0.887	0.957	0.26	0.7
37	$T_{air\,s} - 1.5$	0.903	0.965	0.31	0.01

Table 1: The chosen values of the surface air temperature $T_{air\,s}$ and soil temperature T_s, single scattering albedo ω_v and ω_h, the polarization factor Q, and roughness height parameter h.

The remaining microwave parameter, soil reflectivity, can likewise be expressed for polarizations v and h:

$$r_{sp} = (Qr_{oq} + (1 - Q)r_{op})\ \exp(-h\cos^2\theta) \tag{12}$$

where $(p, q) = (v, h)$ and $p \neq q$, and Q is a polarization coupling factor, h is a roughness height parameter , and r_{op} and r_{oq} are the Fresnel reflectivities of the p and q polarizations, respectively.

The resulting form of the model allowed specification of three geophysical parameters as input parameters: vegetation moisture, soil moisture, and surface air temperature (assumed equal to canopy temperature). Where there is a vegetation canopy, the canopy temperature is set equal to the surface air temperature. The output of the model generated four dual polarized brightness temperatures (6.6, 10.7, 18, and 37 GHz). Thus, the model provided a three input, 8 output (four frequencies $\times$ two polarizations) system for characterization of the African continent.

The differences between $T_{air\,s}$ and T_s are due to the fact that the soil thermal emission arises from beneath the soil surface. The difference is larger for lower frequencies because of the deeper penetration into the soil. The roughness parameters Q and h are chosen so that the brightness temperature measurements are in agreement with a few sites over the region of the Sahara desert which do not have vegetation cover. In the past, other parameters Q and h have been used [18], but they were taken over a much smaller region, whereas the SMMR data are taken over a much larger footprint requiring a different set of Q and h. The vegetation scattering albedo values of ω_h and ω_v are chosen so that the brightness temperature measurements agree with a few sites over the Zaire region which is covered with forests and has no soil effect on the brightness temperatures.

5.2 Training the Network

To train the network, input-output sets were needed to determine the network weights. These sets were generated by varying all three input parameters (soil moisture, surface air temperature, and vegetation moisture) over their physical range, and utilizing the microwave model to obtain corresponding brightness temperatures. For Africa it was assumed that the range for soil moisture was from 0.03 to 0.3 g/cm^2, surface air temperature ranged from 4^0 to 40^0 C, and vegetation moisture varied between 0.01 and 10 kg/m^2. Ten discrete values were processed for soil moisture and surface air temperature, selected according to a linear scale along the parameter ranges, and normalized to lie between 0 and 1 for input to the neural network. Because of the greater dynamic range of vegetation moisture, 16 discrete values were selected according to a logscale along the parameter range, with the logscale values linearly normalized to lie between 0 and 1 for input to the neural network. We thus produced $10 \times 10 \times 16 = 1600$ input parameter vectors. The microwave emission model produced the 8 brightness measurements, for 2 polarizations at 4 frequencies, corresponding to the input parameter vectors. The resulting brightness measurements ranged roughly between 205K and 305K, and were also normalized to lie between 0 and 1 for use with the neural network. We thus created 1600 normalized input-output pairs to train the neural network.

The neural network chosen for the problem was a multilayer perceptron (MLP). A treatment of the details of training a MLP can be found in [19]. The MLP used two layers of sigmoid neurons, fully interconnected between layers, with 40 hidden neurons. The weights were randomly initialized with dynamic range inversely proportional to the fan-in neuron size. The MLP was trained by the standard back-propagation learning algorithm based on gradient descent search with momentum [20].

5.3 The SMMR Data

We used SMMR data obtained between the sixth and tenth days of January, 1982. The brightness temperatures were from the African continent. The SMMR imaging system is a five frequency dual polarized radiometer with a conical cross-track scan providing a constant local incidence angle of 50^0. Details of the instrument's design and calibration can be found in [21] and [22]. The SMMR data were obtained as calibrated, gridded brightness temperatures on TCT map tapes from the Goddard Space Flight Center. The data were gridded on a 0.5 degree latitude-longitude grid at all frequencies, although the inherent spatial resolution of the data varied from approximately 30 km at 37 Ghz to 150 km at 6.6 GHz. Each element in the grid corresponds to a brightness temperature for a particular frequency and polarization in tenths of a degree Kelvin.

5.4 Setting the Physical Parameters of Bayesian Iterative Inversion

The Bayesian methodology requires estimates of physical parameters to describe the various probabilities and conditional probabilities.

For the prior probability $f(\mathbf{x}_i)$, the mean μ_p and covariance matrix $\mathbf{\Sigma}_p$ were estimated from contour diagrams of vegetation type, average rainfall, and average temperature [24]. The means of the geophysical parameters are taken as estimated in Figure 8, and the means are location dependent. The covariance is constant over all points, independent in each input dimension, and based on a standard deviation of 0.1 in the normalized input parameters.

The covariance matrix associated with the neighborhood distribution $f(\mathbf{x}_{s/i}|\mathbf{x}_i)$ was also set based on a standard deviation of 0.1 in each of the normalized input parameters, with each input parameter considered independent of the others.

The covariance of the sensor error and model mismatch distribution $\mathbf{\Sigma}$ was similarly set to correspond to independent noise in the 8 brightness channels. This time, we allowed for two magnitudes of standard deviation, normalized values of .02 and .05 (which correspond to 2 K and 5 K). A smaller standard deviation 2 K represents more confidence on the electromagnetic model. A larger standard deviation will put more weight on the prior distribution $f(\mathbf{x}_i)$ and the neighborhood distribution $f(\mathbf{x}_i|\mathbf{x}_{s/i})$.

5.5 Performing Bayesian Iterative Inversion

The algorithm of iterated conditional modes maximizes $f(\{\mathbf{x}\}|\{\mathbf{m}\})$, iteratively selects each site $\mathbf{p}_i$, and estimate the parameters $\mathbf{x}_i$ which maximize the posterior probability $f(\mathbf{x}_i|\mathbf{m}_i, \mathbf{x}_{s/i})$.

Each $\mathbf{x}_i$ was initialized to be equal to it's prior mean μ_p (recall that the prior means are location dependent). Since the prior distribution indicates our expectations before we receive any data, it is only natural to use it as the initial value.

After initialization, the algorithm selected each successize location $\mathbf{p}_i$, and maximized $f(\mathbf{x}_i|\mathbf{m}_i, \mathbf{x}_{s/i})$ by gradient descent. When all $\mathbf{p}_i$ had been visited, called one completed iteration, the algorithm started back at the first $\mathbf{p}_i$. The process was stopped when the total change summed over all $\mathbf{p}_i$ between one completed iteration and the next was less than 1% of the change between the initial values and the first completed iteration.

5.6 Results and Discussions

In Figures 7,8, and 9, we show the results of applying the Bayesian iterative inversion to the SMMR data over Africa . In Figure 7, we show the prior distribution for the three parameters: (a) vegetation moisture, (b) surface air temperature, and (c) soil moisture. All three parameters are normalized to the range of 0 to 1. The actual physical values of the ranges are as follows. The moisture content of vegetation is between 0.01 kg/m^2 to 10 kg/m^2 , log to the base 10 scale, i.e $v_m = (\log v + 2)/3$, where v is the physical value of vegetation moisture in kg/m^2, while v_m is the map value normalized between 0 and 1. For surface air temperature, the physical value range is between 4 degrees centigrade and 40 degrees centigrade. Thus $T = 36T_m + 4$, where T is the physical surface air temperature in centigrade and T_m is the map value normalized between 0 and 1. Similarly for soil moisture, the range of physical value is between 0.03 gm/cm^3 to 0.3 gm/cm^3. Thus $s = 0.27s_m + 0.03$ where s

and s_m are respectively the physical value (gm/cm^3) and the map value. In Figure 8 and 9, we show respectively the retrieved maps for standard deviation of 5 K and 2 K for the brightness temperature noise respectively.

Figure 7 shows the mean values of the prior distribution. They are estimated from published maps of vegetation type, rainfall, and January temperature. However, we make the geophysical distributions coarse for the purpose of examining how the retrieved maps of Bayesian iteration method compare with the prior. A rough estimate was taken of each 5^0 latitude by 5^0 longitude square area according to the geophysical maps, and the result was smoothly interpolated between the centers of the squares areas. In Figure 7a and 7c, we show respectively the mean of the prior distribution for vegetation moisture and soil moisture. They are assigned high values for the forests of Zaire, and small values for the Sahara desert. The Namib Desert is assigned small value of soil moisture. In figure 7b, we show the assigned values of surface air temperature for the prior distribution. In January, the surface air temperature is moderate in central Africa and decreases towards the Sahara desert.

In Figure 8, we show the retrieved maps based on 5 K noise. This means that the computed brightness temperatures of the microwave emission model are assumed to be correct to within 5K Gaussian noise. Using 5K noise instead of 2K noise means relatively more weight is attached to the prior than to the microwave thermal emission model. Although the retrieved maps follow trends with the prior, the retrieved maps bring out finer scale features than those provided by the coarse prior geographical distribution. Because the smoothing effects of the neighborhood distribution and the prior are weighted more for 5K noise versus 2K noise, the retrieved maps for 5K noise have more gradual changes of parameters from one region to another when compared with the retrieved maps for 2K noise. On comparison with published maps, we conclude that some of the encouraging features of the retrieved map include (a) on the average, the retrieved map shows a lower temperature than the prior (b) the retrieved map shows a higher temperature for the Kalahari desert, (c) the microwave emission model is for land rather than for ocean and lake region. Those regions have much lower brightness temperatures. Thus the retrieved map shows lower temperatures for the Lake Victoria, Lake Tanganyika regions as well as the coastal regions. (d) it maintains the same high temperature for Sudan, (e) it maintains practically zero vegetation moisture for Sahara with a slight increase in soil moisture, (f) it shows small values of vegetation moisture for the Namib Desert and the Kalahari desert, an improvement over the prior, (g) it show small soil moisture for a more extended region around the Namib Desert and the Kalahari desert (h) it shows even smaller soil moisture for the savannasa in Somalia (i),it shows even higher vegetation moisture for Zaire and Congo. (j) it shows more vegetation moisture in Zambia.

In Figure 9, we show the retrieved maps based on 2K model noise. This means that more weight is attached to the model than to the prior and neighborhood smoothing. The retrieved maps show substantial difference from the prior .On comparison with the prior maps and the 2K maps, we conclude the following. (a) the vegetation moisture map follows similar trend to the 5K case but with more extreme values. It shows large vegetation moisture in Zaire, Congo, Cameroon and the Ivory Coast. It shows very small value for the desert regions of Sahara , Namib and Kalahari, as well as Somalia, Ethiopia , Kenya and the inland region of Tanzania. (b) the temperature map follow similar trend to the 5K case except for higher temperature in the Sahara and lower temperature for the Kalahari desert, (c) the soil moisture shows the largest disagreement with the 5K case. It gives much lower values for Botswana, Nambia and the Western Part of South Africa. However, it gives apparently erroneous values of larger soil moisture for the Sahara region. Since the 2K case favors the microwave emission model and the sahara has little vegetation, this suggests that the roughness model or the roughness parameters may not be satisfactory.

The results of application of the Bayesian approach to SMMR data over Sahara indicate the inversion of parameters from the measured brightness temperatures do give improved values over the prior distribution, in particular, for the 5K case. However, when more weight is given to the model as in the 2K case, it can give non-satisfactory results for some regions. This indicates that a more accurate electromagnetic scattering and emission model may be necessary in order to give better results. On the other hand, the retrieval can also be improved with more information sources that can be built into the Bayesian approach.

6 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 solves some of the issues in inverse problems in remote sensing. The many-to-one mapping problem is dealt with by introducing more informational constraints through probability distributions. The neural network training can deal with nonlinear relations between parameters and measurements. The maximization of various conditional probabilities serves as a smoothing process so that the inversion is stable. Since the forward mapping has been used to train the neural networks, directed search in the neural network can be performed speedily.

This paper represents a an effort at application of Bayesian iterative inversion to retrieval of geophysical parameters. Enhancement to the method fall into two categories: application specific refinements and algorithmic enhancements. Problem specific refinements would include more accurate and meaningful formulation of priors, such as enforcing relationships between vegetation moisture and soil moisture, and altering the physical model to relax assumptions and include other parameters, such as atmospheric emission and absorption. Algorithmic enhancements include changing the optimization method from gradient descent to more sophisticated methods, such as Gauss Newton or Gibbs sampling.

References

[1] G. Backus and F. Gilbert, "Uniqueness in the Inversion of Inaccurate Gross Earth Data," Phil. Trans. R. Soc. London Ser. A266, pp. 123-192, 1970.

[2] A. N. Tikonov. "Regularization of incorrectly posed problems", Dokl. Akad. Nauk SSSR 153, 49; Soviet Math Dokl. 4, 145-147.

[3] A. Ishimaru, "Wave Propagation and Scattering in Random Media", vol. 2, Academic Press, 1978.

[4] 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

[5] 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.

[6] 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.

[7] M. S. Dawson, A. K. Fung. 'Neural Networks and Their Applications to Parameter Retrieval and Classification", IEEE Geoscience and Remote Sensing Sensing Society Newsletter, pp 6-14, September 1993.

[8] 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.

[9] J. N. Hwang and C.H. Chan, "Iterative constrained inversion and its applications", 24th. Conf. on Information Systems and Sciences, Princeton, pp. 754-759, March 1990.

[10] A. Linden and J. Kindermann. "Inversion of multilayer nets", In Proc. Int'l Joint Conf. on Neural Networks, II 425-430, Washington D.C., June 1989.

[11] J. N. Hwang, D. T. Davis, L. Tsang. "Locally tuned neural networks for ground truth incorporation", IEEE International Geoscience and Remote Sensing Symposium, pp. 1064-1066, Houston, Texas, May 1992.

[12] J. Besag. "On the statistical analysis of dirty pictures", J. R. Statist. Soc., series B, (48):259-302, 1986

[13] Y. Kerr and E. Njoku, "A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from sopspace", IEEE. Trans on Geoscience and Remote Sensing, vol. 28 may 1990, p384-393

[14] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images" IEEE Trans. on PAMI, 6:721-741, 1984.

[15] J.N. Hwang and T.Y. Chen. "Gibbs Sampling for Textured Image Segmentation via Neural Network MRF Modeling", IEEE Int'l Conference on Neural Networks, pp. 1702-1707, San Francisco, CA, March 1993.

[16] J.N. Hwang, S.R. Lay, A. Lippmann. "Unsupervised learning for multivariate probability density estimation: radial basis and projection pursuit", IEEE Int'l Conference on Neural Networks, San Francisco CA, March 1993. (To appear in IEEE Transactions on Signal Processing, in press.)

[17] K. P. Kirdyashev, A. A. Chukhlantsev, and A. M. Shutko,"Microwave radiation of the Earth's surface in the presence of a vegetation cover", Radio Eng. Electron Phys., vol.24, pp. 37-44, 1979

[18] T. Mo, T. J. Schmugge, and J. R. Wang. "Calculations of the Brightness Temperature of Rough Soil Surfaces: Bare Field", IEEE Trans. Geosci. Remote Sensing, vol. GE-25, pp. 47-54, 1987.

[19] S. Haykin, Neural Networks: A Comprehensive Foundation, MacMillan College Publishing Co. Inc., New York, 1994.

[20] 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.

[21] P. Gloersen and F.T.Barath, "A scanning multichannel microwave radiometer for Nimbus-7 and Seasat-A", IEEE J. Oceanic Eng., vol. OE-2, pp. 172-180, 1977

[22] E. G. Njoku, J. M. Stacey, and F. T. Barath, "The Seasat SMMR: Instrument description and performance", IEE J. Oceanic Eng., vol OE-5, pp. 100-115, 1980

[23] E. G. Njoku, E. J. Christensen, and R. E. Cofield, "The Seasat scanning multichannel microwave radiometer (SMMR): Antenna pattern corrections - development and implementation", IEEE J. Oceanic Eng.,vol. OE-5, pp. 125-137, 1980

[24] Encyclopedia Britannica Macropedia. Vol. 1, 15^{th} Ed., 1977, pp. 188-193.

Figure 1: Information sources available in a remote sensing problem.

Figure 2: The Pieces of the Bayesian Model: the sensor noise and model mismatch distribution $f(\mathbf{m}_i|\mathbf{x}_i)$, the neighborhood distribution $f(\mathbf{x}_{s/i}|\mathbf{x}_i)$, and the prior distribution $f(\mathbf{x}_i)$.

Figure 3: Artificial Function Modeling a "Physical Process"

Figure 4: Artificial Landscape of "Geophysical Parameters"

Figure 5: Landscape Reconstructed by Iterative Inversion of a Neural Network

Figure 6: Landscape Reconstructed by Bayesian Iterative Inversion of a Neural Network

Figure 7: Components contributing to the satellite observed microwave brightness temperatures.

Figure 8: Priors of geophysical parameters.

Figure 9: Reconstruction using $\mathbf{\Sigma}$ corresponding to an error standard deviation of 5K.

Figure 10: Reconstruction using $\mathbf{\Sigma}$ corresponding to an error standard deviation of 2K.

Figure 1

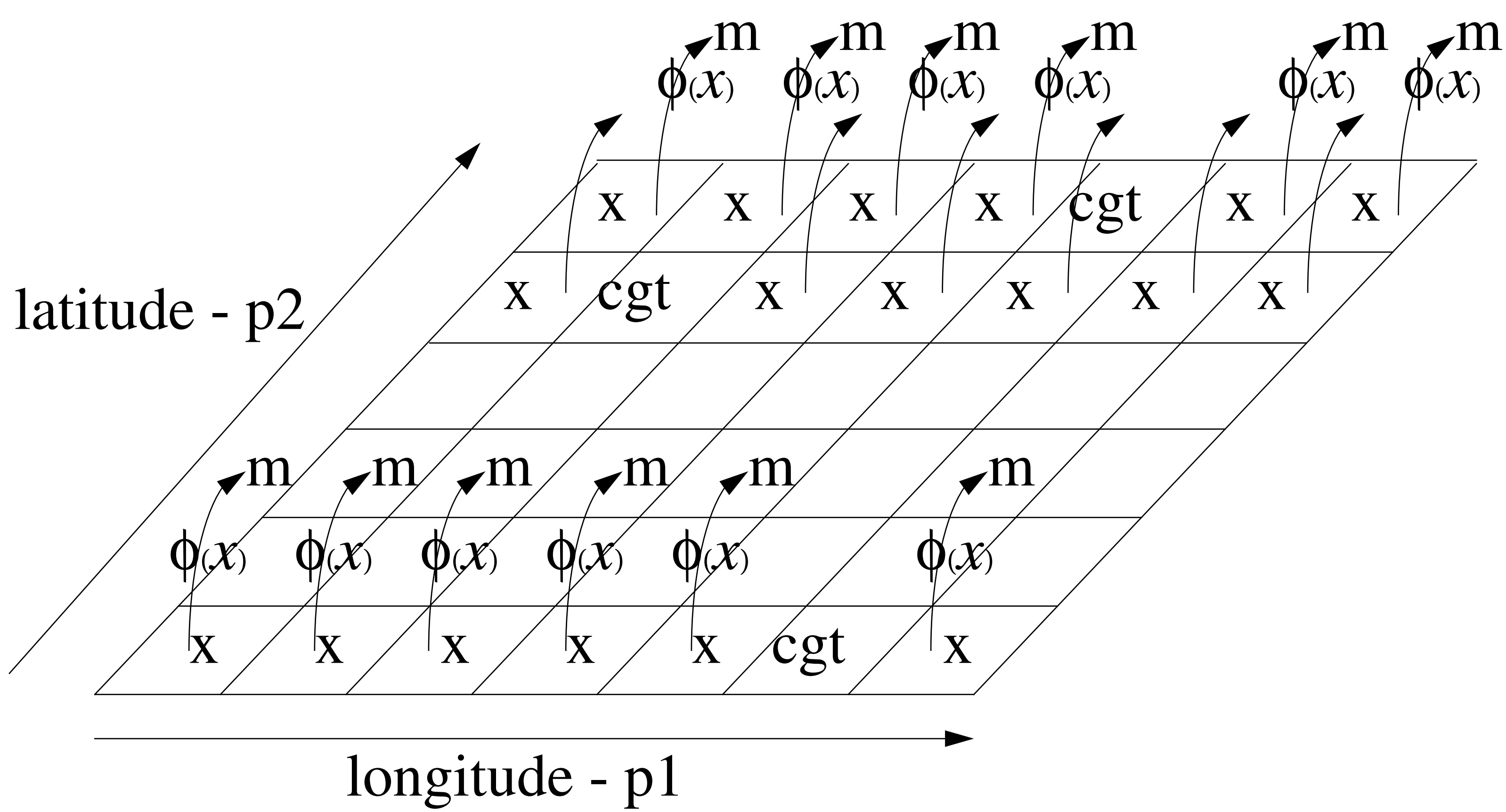
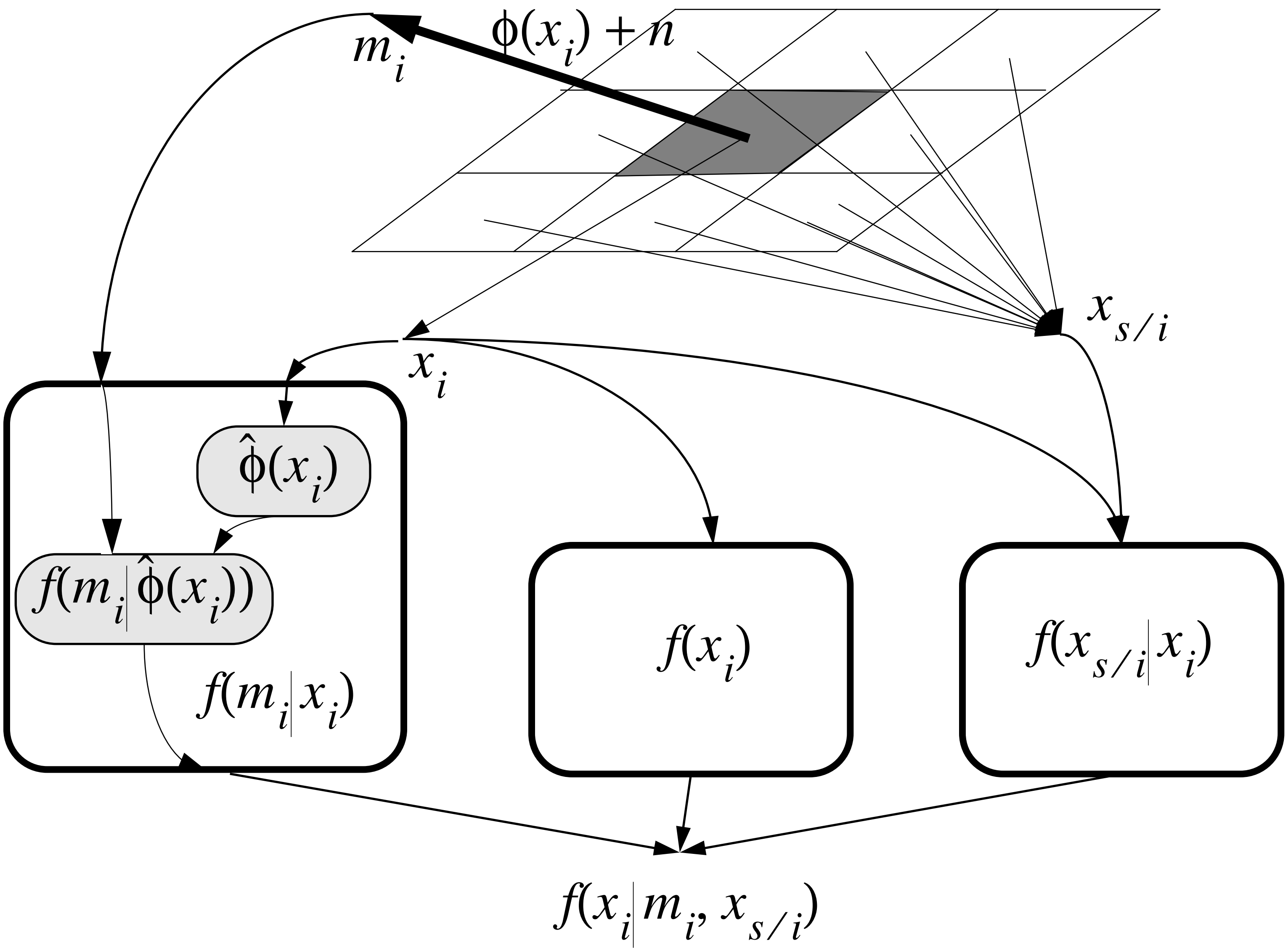


Figure 2



$\phi(x_i) + n$
m_i
$x_{s/i}$
x_i
$\hat{\phi}(x_i)$
$f(m_i|\hat{\phi}(x_i))$
$f(m_i|x_i)$
$f(x_i)$
$f(x_{s/i}|x_i)$
$f(x_i|m_i, x_{s/i})$

Figure 3

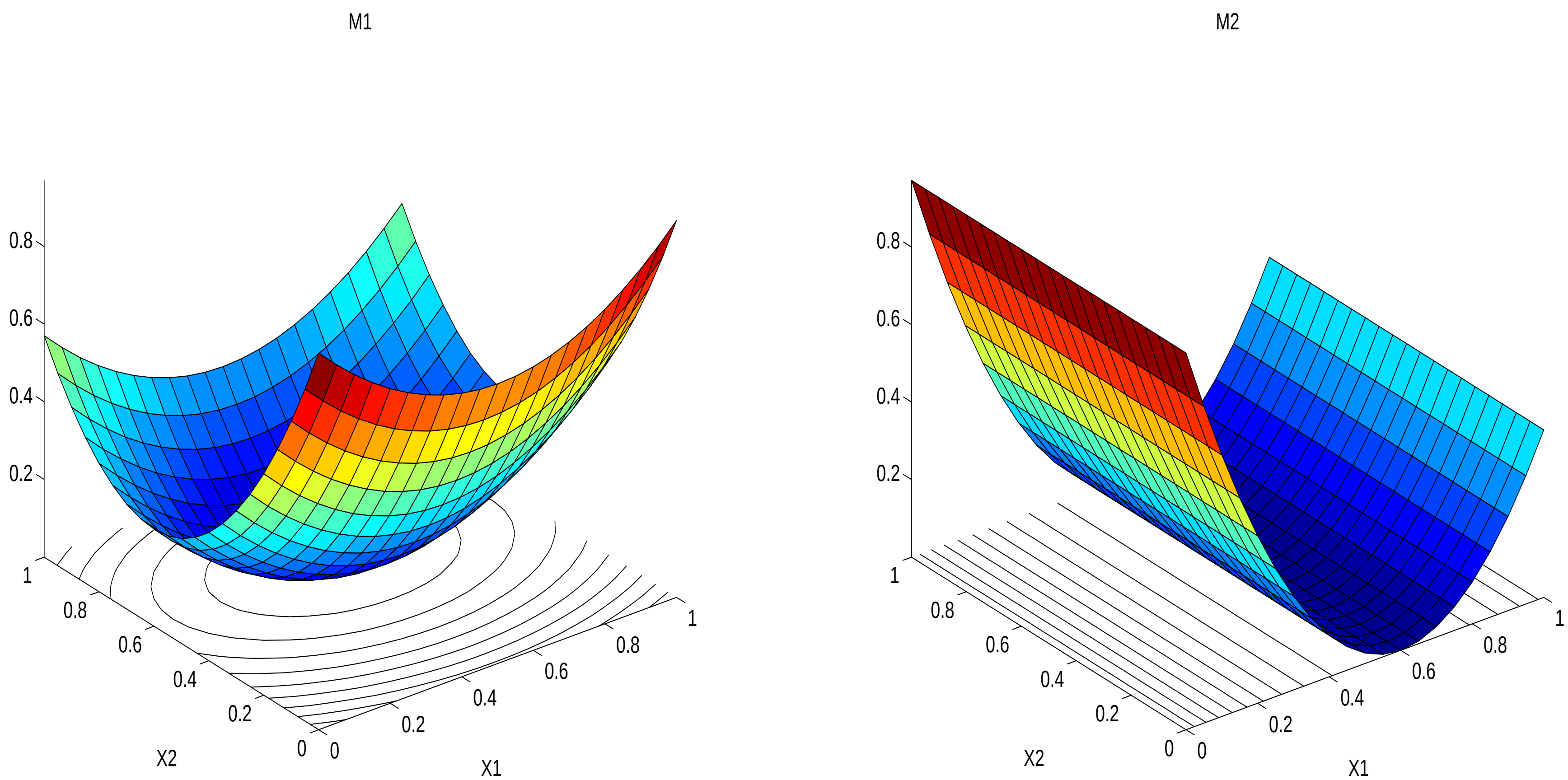


Figure 4

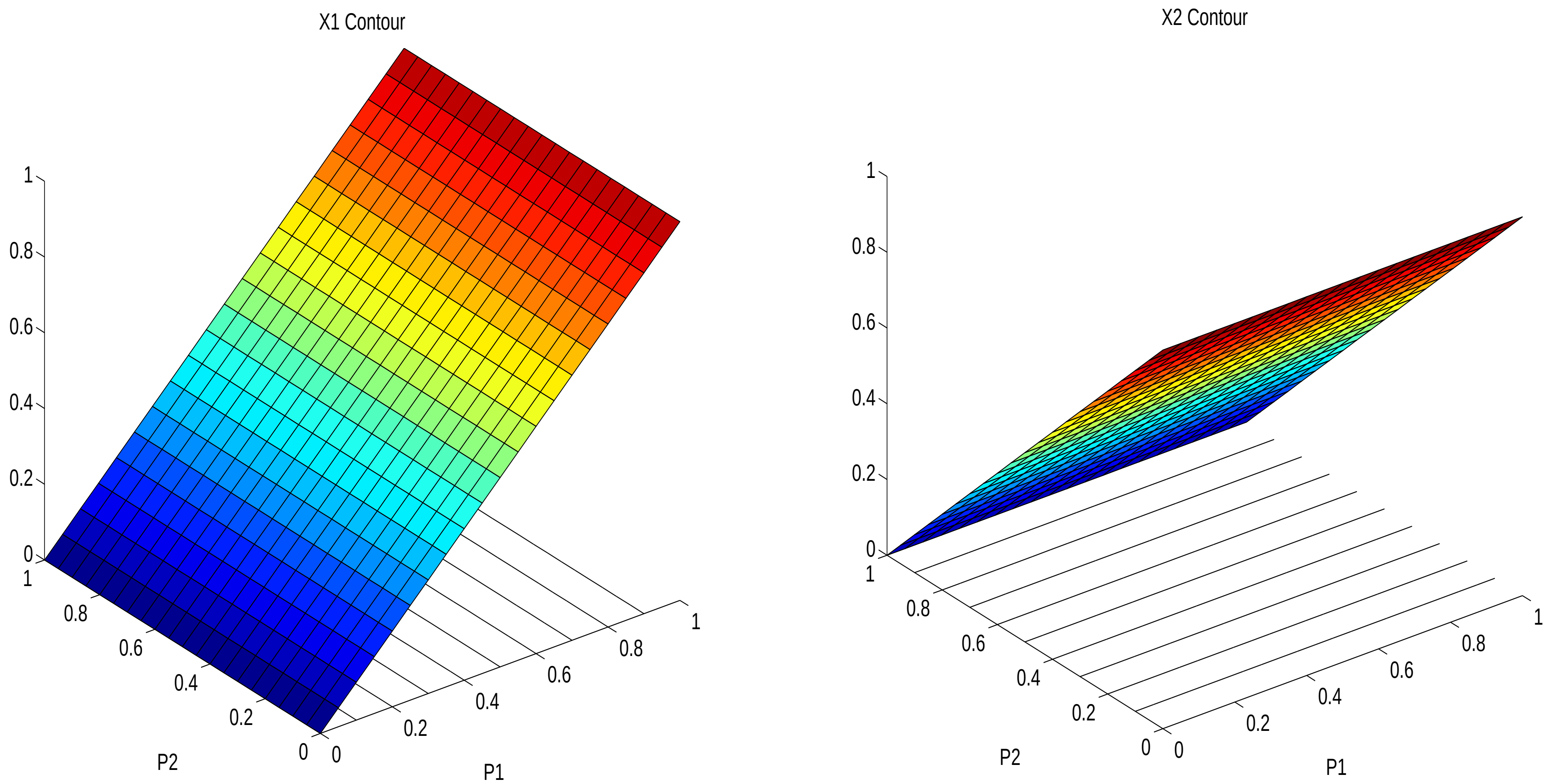
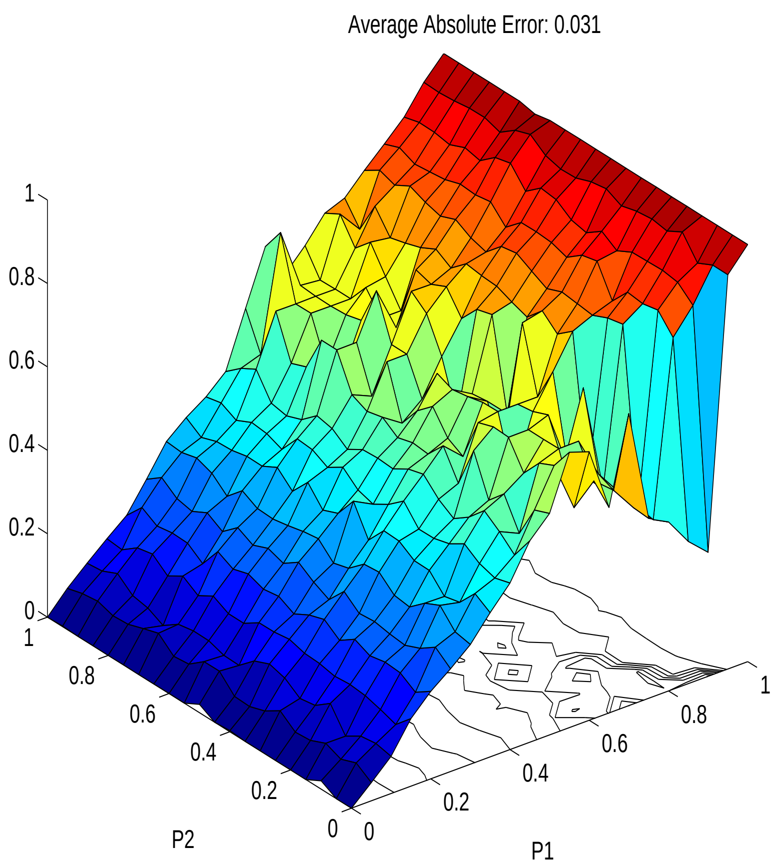


Figure 5


Average Absolute Error: 0.031
1
0.8
0.6
0.4
0.2
0
P2
P1

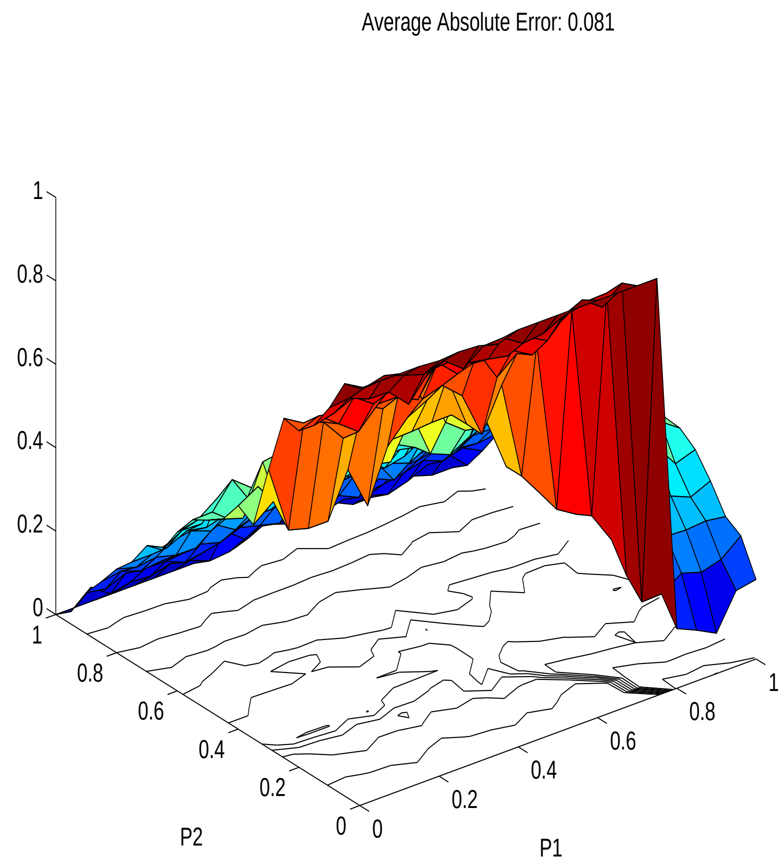
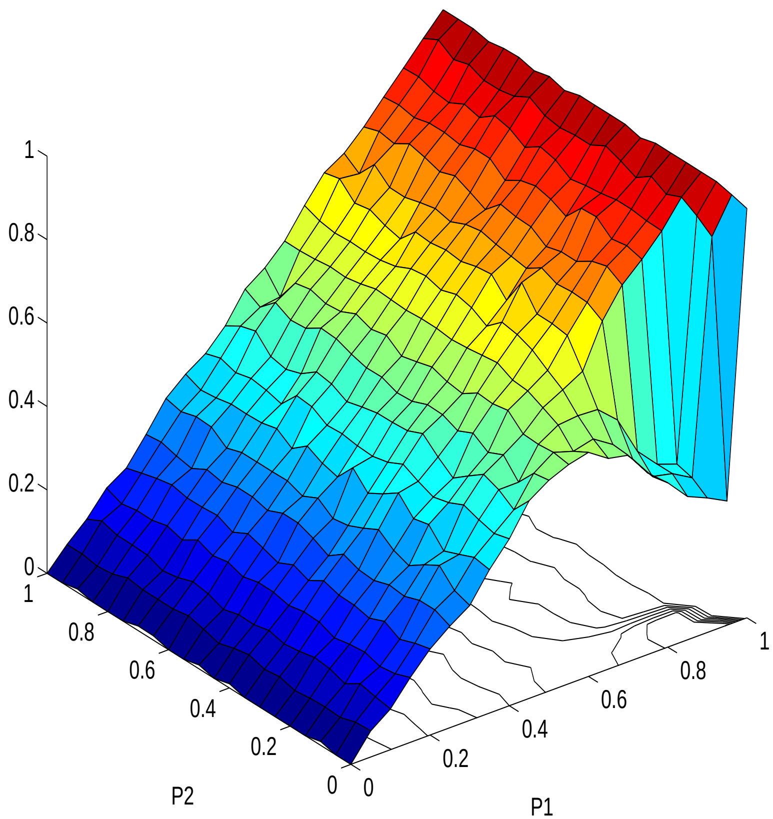

Average Absolute Error: 0.081
1
0.8
0.6
0.4
0.2
0
P2
P1

Figure 6

Average Absolute Error: 0.026



Average Absolute Error: 0.051

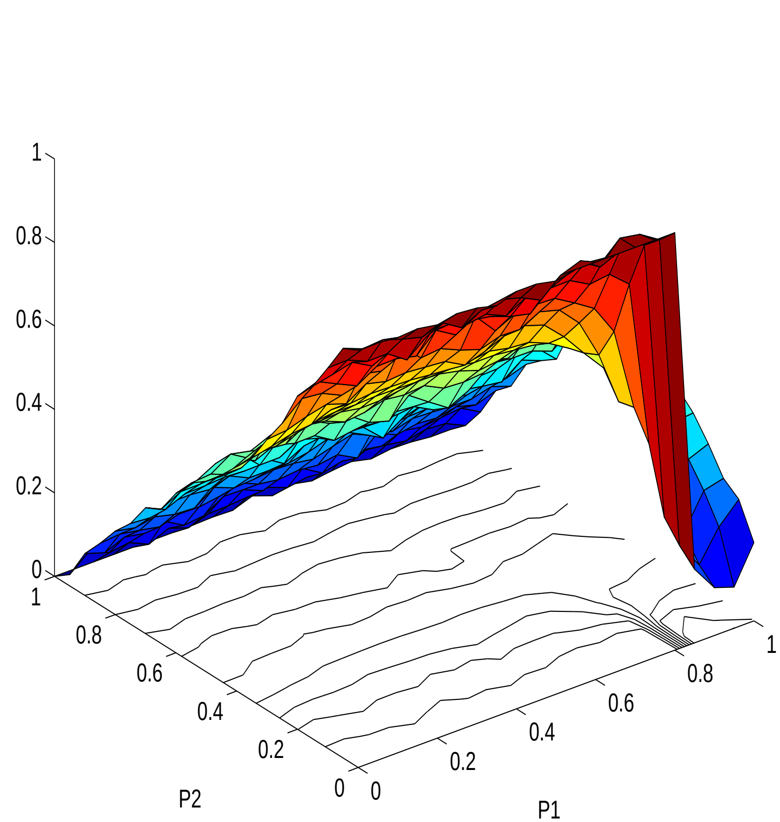
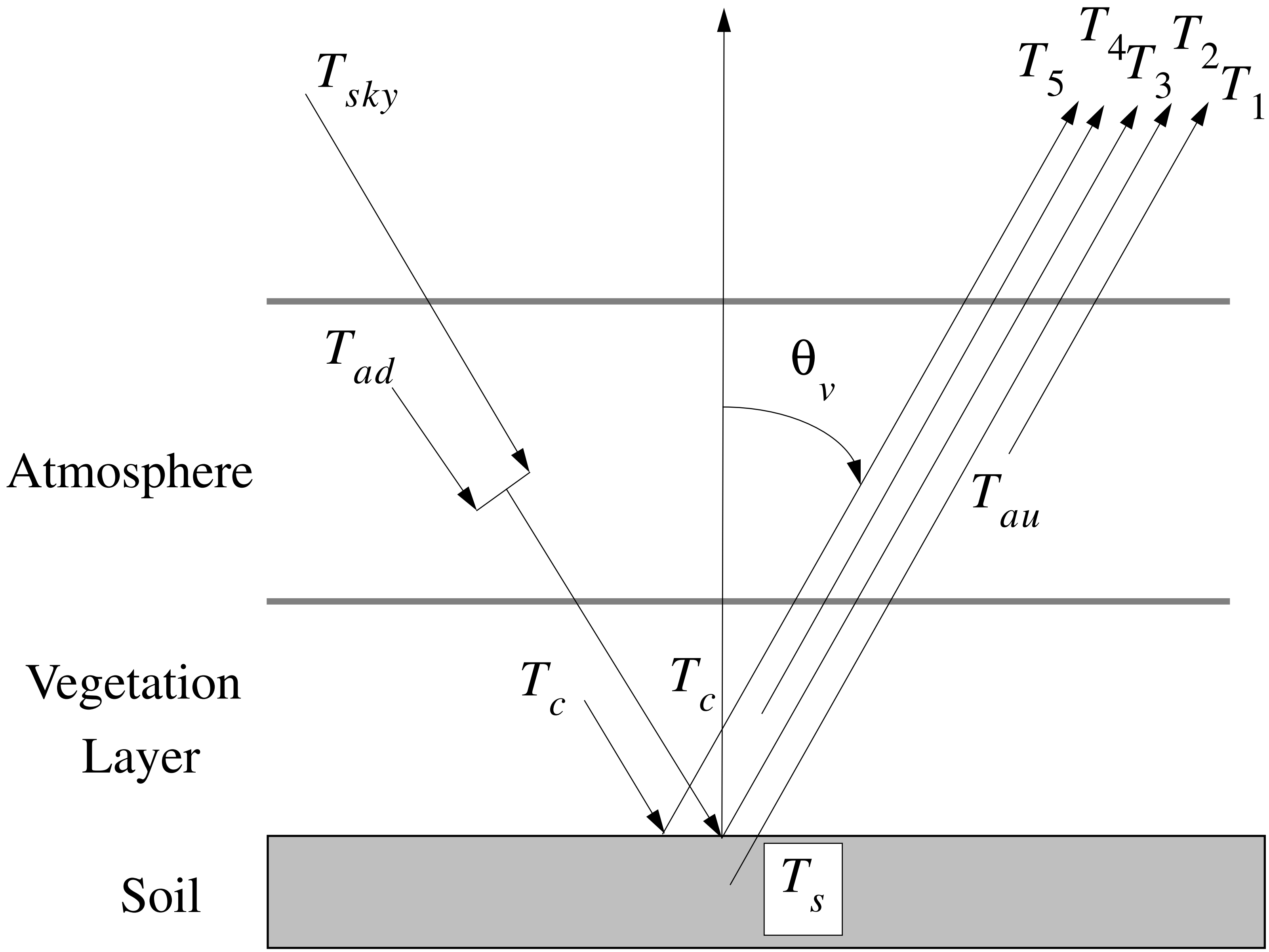


Figure 7


T_{sky}
T_5 T_4 T_3 T_2 T_1
T_{ad}
θ_v
Atmosphere
T_{au}
Vegetation
Layer
T_c
T_c
Soil
T_s

Not included in Paper. Explicit Inverse vs. Iterative Inversion of a Forward Model