submitted to *Geophys. J. Int.*

Minimum relative entropy, Bayes, and Kapur

Allan D. Woodbury ⋆

University of Manitoba, 15 Gillson Street, Winnipeg, Manitoba, R3T 2N2, Canada.

(woodbur@cc.umanitoba.ca)

SUMMARY

The focus of this paper is to illustrate important philosophies on inversion and the similarly and differences between Bayesian and minimum relative entropy (MRE) methods. The development of each approach is illustrated through the general-discrete linear inverse. MRE differs from both Bayes and classical statistical methods in that knowledge of moments are used as "data" rather than sample values. MRE, like Bayes, presumes knowledge of a prior probability distribution (pdf) and produces the posterior pdf itself. MRE attempts to produce this pdf based on the information provided by new moments. It will use moments of the prior distribution only if new data on these moments is not available. It is important to note that MRE makes a strong statement that the imposed constraints are are exact and complete. In this way, MRE is maximally uncommitted with respect to unknown information. In general, since input data are known only to within a certain accuracy, it is important that any inversion method should allow for errors in the measured data. The MRE approach can accommodate such uncertainty and in new work described here, previous results are modified to include a Gaussian prior. A variety of MRE solutions are reproduced under a number of assumed moments and these include second-order central moments. Various solutions of Jacobs and van der Geest (1991) were repeated and clarified. Menke's weighted minimum length solution was shown to have a basis in information theory, and the classic least squares estimate is shown as a solution to MRE under the conditions of more data than unknowns and where we utilize the observed data and their associated noise. An example inverse problem involving a gravity survey over a layered and faulted zone is shown. In all cases the inverse results match quite closely the actual density profile, at least in the upper portions of the profile. The similar results to Bayes presented in are a reflection of the fact that the MRE posterior pdf, and its mean are constrained not by $\mathbf{d} = \mathbf{Gm}$ but by its first moment $E(\mathbf{d} = \mathbf{Gm})$, a weakened form of the constraints. If there is no error in the data then one should expect a complete agreement between Bayes and MRE and this is what is shown. Similar results are shown when second moment data is available (for example posterior covariance equal to zero). But dissimilar results are noted when we attempt to derive a Bayesian like result from MRE. In the various examples given in this paper, the problems look similar but are, in the final analysis, not equal. The methods of attack are different and so are the results even though we have used the linear inverse problem as a common template.

1 INTRODUCTION

Arguably, since the advent of Taratola's work (eg. Tarantola, 1987) it has become common to consider both measured data and unknown model parameters as uncertain. In a probabilistic inverse approach, the ultimate goal is the posterior probability density function (pdf), updated from some previous level of knowledge. Generally, although not exclusively, we are concerned with the expected values (or some other mode) of the posterior pdf, together with appropriate confidence limits. A detailed discussion of information-based methods is given in Ulrych and Sacchi (2006), specifically Bayesian inference, maximum entropy, and minimum relative entropy (MRE).

Entropy maximization (MaxEnt) is a general approach of inferring a probability distribution from constraints which do not uniquely characterize that distribution. Applications of this method have met with considerable success in a variety of fields (eg. Kapur 1989; Buck and Macauly, 1991). An often studied use for an entropy measure is as a penalty term or ”norm”.

⋆ Department of Civil Engineering

In this way the entropy of a model can be maximized subject to non-linear constraints imposed the observed data.

The related, but more general entropic principle is that of minimum relative entropy (MRE). An axiomatic foundation for MRE has been given by Shore and Johnson (1980) and Johnson and Shore (1983). The MRE principle is perhaps less well known than MaxEnt and was first introduced by Kullback (1959) as a method of statistical inference. The MRE approach to inversion was originally developed by Shore (1981) as an extension of Burg's (1975) method of maximum entropy spectral analysis (see also Ulrych and Bishop, 1975).

In the hydrological and geophysical literature our first publications on MRE attracted a certain amount of attention (Woodbury and Ulrych, 1993; 1996; Woodbury et al., 1998a). Our (1998) work was followed up with studies of small 'toy' problems, die experiments and an exhaustive comparison of MRE with Maximum Entropy, Bayesian and SVD solutions (Woodbury and Ulrych, 1998b). Finally, thesis and other papers (Neupauer, 1999; Neupauer et al., 2000; Ulrych and Woodbury, 2003) were produced in which detailed comparisons between Tikhonov regularization (Provencher, 1982ab; TR) and MRE were made.

Ulrych et al. (2000) used entropic principles for a classical problem; that is finding moments and distributions from sample data. They showed that probability density functions that are estimated from L-moments are superior estimates to those obtained using sample central moments (C-moments) and the principle of maximum entropy. Woodbury (2004) detailed a general purpose computer program that produces a univariate pdf from a series of constraints and a prior probability. Some guidelines for the selection of the prior were presented.

From a Bayesian perspective, Ulrych et al. (2001) summarized some of the concepts central to that approach to inverse problems. Of course, there are many examples in the geophysical literature on Bayes, many of which are referred to in Ulrcyh and Sacchi (2006) and Scales et. al. (2001). Some of the most often referred to works include Duijndam (1988) and Bretthorst (1988). These references are by no means exhaustive and only serve the purpose of recognizing the importance of Bayes theorem in problems of inference.

Empirical Bayesian techniques have also undergone development and application in Geophysics (e.g. Woodbury, 2007). There has always been controversy surrounding the use of, perhaps, arbitrary priors in Bayes Theorem. For example, the form of the pdf may be known or highly suspected but the actual statistical parameters embedded into the prior pdf, such as the mean, the variance and so on may not be well known and difficult to estimate. Noise levels in the data may also be unknown. The idea behind the empirical Bayes approach is that the prior is based on information contained in the input data (Ulrych et al.,2001).

What is perhaps less obvious from the published literature is how these methods are distinct from classical methods of statistical inference, such as maximum likelihood or least squares. How do Bayesian methods compare to entropic techniques, and under what circumstances should one use entropic rather than Bayesian methods? This current effort is organized as follows. Minimum relative entropy (MRE), and Bayesian solutions for inverse problems are detailed and related. The focus of the paper is to illustrate important philosophies on inversion and the similarly and differences between the various solution methods. While there exists a considerable body of published works comparing maximum entropy and Bayes, there are few such works related to Bayes and MRE. Much of the existing work can be attributed to Kapur and coauthors, but some of the derivations were left incomplete. This paper will attempt to fill in those gaps, step by step in comparing inverse solutions on the general linear underdetermined model.

2 BAYESIAN APPROACH TO INVERSION

In this section we will review the classic Bayesian solution to the linear inverse problem. Much of this material will likely be well known to the reader but it is appropriate here to repeat this for the sake of completeness and understanding of the notation used. Here, we rely on linear transformations and Gaussian assumptions for probabilities. This step makes the following expectations and analysis tractable.

In the discrete linear-inverse case, one needs to start with a forward modeling problem and this can be written as

$$\mathbf{d} = \mathbf{G}\mathbf{m} \tag{1}$$

where $\mathbf{d}$ is a $(n \times 1)$ vector of theoretically predicted data and $\mathbf{G}$ $(m \times n)$ is a linear transformation (kernel matrix) from model to data space. In the case of observations though, the true "data" are corrupted by noise, for example

$$\mathbf{d}^* = \mathbf{G}\mathbf{m} + \mathbf{e} \tag{2}$$

where $\mathbf{e}$ is an $(n \times 1)$ vector of unknown error terms and $\mathbf{m}$ $(m \times 1)$ vector of model parameters, "the model". In the case where there are more data than unknowns we can, of course, use the classic least-squares approach, but this subject will not be covered in the present work. Here we focus on more interesting cases where more unknowns are sought than data values observed. Consequently, this ill-posed deterministic problem is recast in terms of a problem in statistical inference; one in which can be attacked from the viewpoint of updating probabilities.

Bayesian inference supposes that an observer can define a personal prior probability-density function (pdf) about some random variable $\mathbf{m}$. This pdf, $p(\mathbf{m})$, can be defined on the basis of personal experience or judgment and Bayes' rule quantifies how this personal pdf can be changed on the basis of measurements. Consider a vector of observed random variables, $\mathbf{d}^*$. If the conditional pdf of $\mathbf{d}^*$ given the "true" value of $\mathbf{m}$ is given by $\Phi(\mathbf{d}^* \mid \mathbf{m})$ then Bayes' rule states that

$$\Phi(\mathbf{m} \mid \mathbf{d}^*) = \frac{\Phi(\mathbf{d}^* \mid \mathbf{m})p(\mathbf{m})}{\int_K \Phi(\mathbf{d}^* \mid \mathbf{m})p(\mathbf{m})d\mathbf{m}} \tag{3}$$

$\Phi(\mathbf{m} \mid \mathbf{d}^*)$ is the conditional pdf of $\mathbf{m}$, given $\mathbf{d}^*$, and $\Phi(\mathbf{d}^* \mid \mathbf{m})$ represents the conditional pdf from forward modeling. If a likelihood function can be defined (i.e., the forward model exists), and there is a compatibility between observed results and a prior understanding of model parameters, (i.e., $\Phi(\mathbf{d}^* \mid \mathbf{m}) > 0$ for some $\mathbf{m}$ where $p(\mathbf{m}) > 0$) then Bayes' rule implies that the resulting posterior pdf exists and is unique (Tarantola, 1987, p53).

For Bayes we must define a likelihood function which is a conditional pdf of the data, given the "true" model $\mathbf{m}$. In other words we need a pdf for $(\mathbf{d}^* - \mathbf{d} = \mathbf{e})$. If the physics is correct then the pdf for $\mathbf{e}$ only reflects measurement error. Classically, it is assumed that $\mathbf{e} = N(\mathbf{0}, \mathbf{C}_d)$; i.e. normally distributed with mean zero and covariance $\mathbf{C}_d = E[(\mathbf{d} - \mathbf{d}^*)(\mathbf{d} - \mathbf{d}^*)^T]$. Therefore, we can write the Bayes likelihood as (see Tarantola, 1987, p68):

$$\Phi(\mathbf{d}^* \mid \mathbf{m}) = ((2\pi)^{nd}|\mathbf{C}_d|)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{d}^* - \mathbf{Gm})^T \mathbf{C}_d^{-1}(\mathbf{d}^* - \mathbf{Gm})\right] \tag{4}$$

where n is the length of vector $\mathbf{d}^*$. If the prior distribution of the model is also assumed to be Gaussian then the prior pdf for $\mathbf{m}$ is given the form:

$$p(\mathbf{m}) = ((2\pi)^{nm}|\mathbf{C}_p|)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{m} - \mathbf{s}))^T \mathbf{C}_p^{-1}(\mathbf{m} - \mathbf{s}))\right] \tag{5}$$

Here, $\mathbf{s}$ is the mean value and $\mathbf{C}_p$ is the covariance matrix which of course describes the variance and correlation of the parameters. Using (3, 4, and 5) the resulting posterior pdf from a Bayes analysis of the linear inverse problem with Gaussian priors and likelihood is:

$$\begin{aligned}\Phi(\mathbf{m} \mid \mathbf{d}^*) \quad &= \quad C_1 \exp[-\frac{1}{2}(\mathbf{d}^* - \mathbf{Gm})^T \mathbf{C}_d^{-1}(\mathbf{d}^* - \mathbf{Gm}) \\ &\quad -\frac{1}{2}(\mathbf{m} - \mathbf{s})^T \mathbf{C}_p^{-1}(\mathbf{m} - \mathbf{s})]\end{aligned} \tag{6}$$

Note that the mode of a Gaussian distribution is equal to its mean. So to find the mean and covariance of this pdf, we simply have to find the maximum of an objective function defined as

$$J(\mathbf{m}) \quad = \quad -\frac{1}{2}(\mathbf{d}^* - \mathbf{Gm})^T \mathbf{C}_d^{-1}(\mathbf{d}^* - \mathbf{Gm})$$

$$-\frac{1}{2}(\mathbf{m}-\mathbf{s})^T\mathbf{C}_p^{-1}(\mathbf{m}-\mathbf{s}) \tag{7}$$

Expanding out the above results in

$$\begin{aligned} J(\mathbf{m}) \;\; = \;\; & -\frac{1}{2}[\mathbf{m}^T(\mathbf{C}_p^{-1}+\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G})\mathbf{m}-\mathbf{m}^T(\mathbf{C}_p^{-1}\mathbf{s}+ \\ & \mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{d}^*)-(\mathbf{s}^T\mathbf{C}_p^{-1}+\mathbf{d}^{*T}\mathbf{C}_d^{-1}\mathbf{G})\mathbf{m}+ \\ & \mathbf{s}^T\mathbf{C}_p^{-1}\mathbf{s}+\mathbf{d}^{*T}\mathbf{C}_d^{-1}\mathbf{d}^*] \end{aligned} \tag{8}$$

Letting

$$(\mathbf{C}_p^{-1}+\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G})=\mathbf{C}_q^{-1}=\mathbf{A} \tag{9}$$

and

$$(\mathbf{C}_p^{-1}\mathbf{s}+\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{d}^*)=\mathbf{b} \tag{10}$$

Results in

$$\begin{aligned} J(\mathbf{m}) \;\; = \;\; & -\frac{1}{2}[\mathbf{m}^T\mathbf{A}\mathbf{m}-\mathbf{m}^T\mathbf{b}-\mathbf{b}^T\mathbf{m}+\mathbf{s}^T\mathbf{C}_p^{-1}\mathbf{s}+ \\ & \mathbf{d}^{*T}\mathbf{C}_d^{-1}\mathbf{d}^*] \end{aligned} \tag{11}$$

Now taking the derivative of J with respect to $\mathbf{m}$, recognizing that $\mathbf{A}$ is symmetric and setting the result to zero determines $<\mathbf{m}>$, the posterior mean value;

$$\mathbf{A}<\mathbf{m}>=\mathbf{b} \tag{12}$$

$$<\mathbf{m}>=\mathbf{A}^{-1}\mathbf{b} \tag{13}$$

Expanding,

$$<\mathbf{m}>=(\mathbf{C}_p^{-1}+\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G})^{-1}(\mathbf{C}_p^{-1}\mathbf{s}+\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{d}^*) \tag{14}$$

$$\mathbf{C}_q=(\mathbf{C}_p^{-1}+\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G})^{-1} \tag{15}$$

which also can be written as:

$$<\mathbf{m}>=\mathbf{s}+\mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T+\mathbf{C}_d)^{-1}(\mathbf{d}^*-\mathbf{G}\mathbf{s}) \tag{16}$$

$$\mathbf{C}_q=\mathbf{C}_p-\mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T+\mathbf{C}_d)^{-1}\mathbf{G}\mathbf{C}_p \tag{17}$$

Here, $<\mathbf{m}>$ and $\mathbf{C}_q$ are expected value and covariance of the posterior pdf,

$$\begin{aligned} \Phi(\mathbf{m}\mid\mathbf{d}^*) \;\; = \;\; & ((2\pi)^{nm}|\mathbf{C}_q|)^{-\frac{1}{2}} \\ & \exp\left[-\frac{1}{2}(\mathbf{m}-<\mathbf{m}>)^T\mathbf{C}_q^{-1}(\mathbf{m}-<\mathbf{m}>)\right] \end{aligned} \tag{18}$$

These results are well known (see Tarantola, 1987). Note in this case $\mathbf{d}^* \neq \mathbf{G} < \mathbf{m} >$. In other words, the average value of the predicted data $\bar{\mathbf{d}}$ from the average value of the model $< \mathbf{m} >$ is not equal to the observed data $\mathbf{d}^*$. We note that the expected value of the model (16) could also be written as

$$< \mathbf{m} >= \mathbf{s} + \mathbf{G}_*^{-1}(\mathbf{d}^* - \mathbf{G}\mathbf{s}) \tag{19}$$

where $\mathbf{G}_*^{-1} = \mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T + \mathbf{C}_d)^{-1}$, a generalized inverse matrix of $\mathbf{G}$.

3 MINIMUM RELATIVE ENTROPY THEORY

Minimum relative entropy (MRE) is an information-theoretic method of problem solving. Its roots lie in probability theory and in an abstract way deals with information measures in probability spaces (Kapur and Kessavan, 1992). MRE was classically dervived by Shore and Johnson (1980) and follows directly from four basic axioms of consistent inference which are: uniqueness, invariance, system independence and subset independence. Stated in other words, if a problem can be solved in difference ways, each path has to lead to the same answer.

Consider a system having a set of possible states. Let $\mathbf{x}$ be a state and $q^\dagger(\mathbf{x})$ its unknown multivariate probability density function (pdf). Note that $\mathbf{x}^T = (x_1, x_2, x_3, \ldots, x_K)$ and the integrals noted below are multiple integrals over each of the x_i. The pdf must satisfy a normalizing constraint

$$\int q^\dagger(\mathbf{x})d\mathbf{x} = 1 \tag{20}$$

Now, let us suppose that there is prior information on $q^\dagger(\mathbf{x})$ in the form of a pdf, $p(\mathbf{x})$, and new information exists, say expectations of the form of $j = 1, M$ expected value constraints

$$\int q^\dagger(\mathbf{x})f_j(\mathbf{x})d\mathbf{x} = \bar{f}_j \tag{21}$$

or bounds on these values. It is important to note here that MRE makes a strong statement that the constraints are exact and complete. That is, the M values are the only only ones operating and these are known exactly. The task at hand then is to choose a distribution $q(\mathbf{x})$ out of the infinite possibilities that is in some way the best estimate of $q^\dagger$, consistent with this new information. The solution is to minimize $H(q,p)$, the entropy of $q(\mathbf{x})$ relative to $p(\mathbf{x})$, where

$$H(q,p) = \int q(\mathbf{x}) \ln\left[\frac{q(\mathbf{x})}{p(\mathbf{x})}\right] d\mathbf{x} \tag{22}$$

subject to the constraints of the form of equations (20) and (21). The posterior estimate $q(\mathbf{x})$ has the form (Shore and Johnson, 1980; Woodbury and Ulrych, 1993)

$$q(\mathbf{x}) = p(\mathbf{x}) \exp\left[-1 - \mu - \sum_{j=1}^{M} \lambda_j f_j(\mathbf{x})\right] \tag{23}$$

where μ and the λ_j are Lagrange multipliers determined from equations (20) and (21). If the final goal is the posterior pdf, (23) is the form of the solution and one has to determine the Lagrange multipliers (Johnson, 1983). For an inverse problem, we are likely concerned with the expected values of the posterior pdf, $q(\mathbf{x})$ together with the appropriate confidence limits.

Now for the geophysical inverse problem, we have to minimize

$$H = \int q(\mathbf{m}) \ln \frac{q(\mathbf{m})}{p(\mathbf{m})} d\mathbf{m} \tag{24}$$

subject to any number of moments but for the sake of argument here, the first two central moments:

$$\int q(\mathbf{m})\mathbf{m} d\mathbf{m} = < \mathbf{m} > \tag{25}$$

$$\int q(\mathbf{m})[(< \mathbf{m} > -\mathbf{m})(< \mathbf{m} > -\mathbf{m})^T] d\mathbf{m} = \mathbf{C}_q \tag{26}$$

along with the normalizing constraint

$$\int q(\mathbf{m}) d\mathbf{m} = 1 \tag{27}$$

The MRE solution to this problem is (Kapur, 1989)

$$\begin{aligned} q(\mathbf{m}) &= ((2\pi)^{nm}|\mathbf{C}_q|)^{-\frac{1}{2}} \\ &\quad \exp\left[-\frac{1}{2}(\mathbf{m}- < \mathbf{m} >)^T \mathbf{C}_q^{-1} (\mathbf{m}- < \mathbf{m} >)\right] \end{aligned} \tag{28}$$

The reader will notice that in order for us to obtain this solution we would have to first specify moments of the posterior pdf, namely $\mathbf{C}_q$ and $< \mathbf{m} >$ which we do not normally have. Instead, given the theoretical relationship $\mathbf{d} = \mathbf{G}\mathbf{m}$, we can rearrange the above constraints into a form that is more convenient for the data we actually observe and then solve the MRE minimization. Rearranging (25) yields

$$\int q(\mathbf{m})(\bar{\mathbf{d}} - \mathbf{G}\mathbf{m}) d\mathbf{m} = 0 \tag{29}$$

where $\mathbf{G} < \mathbf{m} >= \bar{\mathbf{d}}$ and

$$\int q(\mathbf{m})[(\bar{\mathbf{d}} - \mathbf{G}\mathbf{m})(\bar{\mathbf{d}} - \mathbf{G}\mathbf{m})^T] d\mathbf{m} = \mathbf{R} = \mathbf{G}\mathbf{C}_q\mathbf{G}^T \tag{30}$$

In the above, $\mathbf{R}$ is the covariance of $\mathbf{d}$, $E[(\mathbf{d} - \bar{\mathbf{d}})(\mathbf{d} - \bar{\mathbf{d}})^T]$. If $p(\mathbf{m})$ is multivariate and Gaussian;

$$p(\mathbf{m}) = ((2\pi)^m|\mathbf{C}_p|)^{-\frac{1}{2}}$$

$$\exp\left[-\frac{1}{2}(\mathbf{m}-\mathbf{s}))^T\mathbf{C}_p^{-1}(\mathbf{m}-\mathbf{s}))\right] \tag{31}$$

then $q(\mathbf{m})$ is (Kapur et al., 1994):

$$\begin{aligned} q(\mathbf{m}) \quad &= \quad C_2 \times \exp\left[-\frac{1}{2}((\mathbf{m}-\mathbf{s})^T\mathbf{C}_p^{-1}(\mathbf{m}-\mathbf{s}))\right] \\ &\quad \times \exp\left[-\lambda_0 - \boldsymbol{\lambda}^{\boldsymbol{T}}(\bar{\mathbf{d}}-\mathbf{G}\mathbf{m}) - (\bar{\mathbf{d}}-\mathbf{G}\mathbf{m})^T\mathbf{D}^{-1}(\bar{\mathbf{d}}-\mathbf{G}\mathbf{m})\right] \end{aligned} \tag{32}$$

where C_2 is a normalizing constant, and λ_0, $\boldsymbol{\lambda}$ and the matrix $\mathbf{D}^{-1}$ are Lagrange multipliers that have to be determined from the constraints (27, 29, 30). In the subsequent sections it will be shown how the multipliers can determined for specific cases.

3.1 Updating prior with only first moment information

For the MRE solution, we want to minimize

$$\int q(\mathbf{m}) \ln \frac{q(\mathbf{m})}{p(\mathbf{m})} d\mathbf{m} \tag{33}$$

subject to:

$$\int q(\mathbf{m}) d\mathbf{m} = 1 \tag{34}$$

along with

$$\int q(\mathbf{m})(\bar{\mathbf{d}} - \mathbf{G}\mathbf{m}) d\mathbf{m} = 0 \tag{35}$$

That is, we want to update the prior, based only on a new mean value constraint. As before we assume a Gaussian prior, and replacing the expected value of the data $\bar{\mathbf{d}}$ for the observations $\mathbf{d}^*$ that we actually obtain yields

$$\begin{aligned} q(\mathbf{m}) \quad &= \quad C_3 \times \exp\left[-\frac{1}{2}((\mathbf{m}-\mathbf{s})^T\mathbf{C}_p^{-1}(\mathbf{m}-\mathbf{s}))\right] \\ &\quad \times \exp\left[-\lambda_0 - \boldsymbol{\lambda}^{\boldsymbol{T}}(\mathbf{d}^* - \mathbf{G}\mathbf{m})\right] \end{aligned} \tag{36}$$

Defining $\mathbf{A} = \mathbf{C}_p^{-1}$ and $\mathbf{b} = \mathbf{C}_p^{-1}\mathbf{s}$ yields the objective function

$$\begin{aligned} J(\mathbf{m}) \quad &= \quad -\frac{1}{2}[\mathbf{m}^T\mathbf{A}\mathbf{m} - \mathbf{m}^T\mathbf{b} - \mathbf{b}^T\mathbf{m} + \mathbf{s}^T\mathbf{C}_p^{-1}\mathbf{s} + \\ &\quad -2\boldsymbol{\lambda}^{\boldsymbol{T}}\ (\mathbf{G}\mathbf{m} - \mathbf{d}^*)] \end{aligned} \tag{37}$$

Taking the derivative of J with respect to $\mathbf{m}$ and setting the result to zero yields

$$\mathbf{A} < \mathbf{m} >= \mathbf{b} - \mathbf{G}^T\boldsymbol{\lambda} \tag{38}$$

The next step is to find the $\boldsymbol{\lambda}$ values. We know from the second MRE constraint

$$\int q(\mathbf{m})(\mathbf{d}^* - \mathbf{Gm})d\mathbf{m} = 0 \tag{39}$$

or $\mathbf{d}^* = \mathbf{G} < \mathbf{m} >$. Substituting the above expression for the mean value of the model results in

$$\mathbf{GA}^{-1}(\mathbf{b} + \mathbf{G}^T\boldsymbol{\lambda}\ ) = \mathbf{d}^* \tag{40}$$

and solving for the λ's

$$\boldsymbol{\lambda}\ = (\mathbf{GA}^{-1}\mathbf{G}^T)^{-1}[\mathbf{d}^* - \mathbf{GA}^{-1}\mathbf{b}] \tag{41}$$

Finally substituting this value back into our expression for the mean value

$$< \mathbf{m} >= \mathbf{A}^{-1}\left[\mathbf{b} + \mathbf{G}^T\boldsymbol{\lambda}\ \right] \tag{42}$$

$$< \mathbf{m} >= \mathbf{A}^{-1}(\mathbf{b} + \mathbf{G}^T(\mathbf{GA}^{-1}\mathbf{G}^T)^{-1}[\mathbf{d}^* - \mathbf{GA}^{-1}\mathbf{b}]) \tag{43}$$

$$< \mathbf{m} >= \mathbf{A}^{-1}\mathbf{b} + \mathbf{A}^{-1}\mathbf{G}^T(\mathbf{GA}^{-1}\mathbf{G}^T)^{-1}[\mathbf{d}^* - \mathbf{GA}^{-1}\mathbf{b}] \tag{44}$$

or, with the definitions of $\mathbf{A} = \mathbf{C}_p^{-1}$ and $\mathbf{b} = \mathbf{C}_p^{-1}\mathbf{s}$

$$< \mathbf{m} >= \mathbf{s} + \mathbf{C}_p\mathbf{G}^T(\mathbf{GC}_p\mathbf{G}^T)^{-1}[\mathbf{d}^* - \mathbf{Gs}] \tag{45}$$

and it is noted that the posterior covariance is not updated;

$$\mathbf{C}_q = \mathbf{C}_p \tag{46}$$

This is identical to the results of Jacobs and van der Geest (1991). Note that $\mathbf{C}_p\mathbf{G}^T(\mathbf{GC}_p\mathbf{G}^T)^{-1}$ could be again be referred to as $\mathbf{G}_*^{-1}$, a generalized inverse of $\mathbf{G}$. Equation (45) is of the same form as a weighted minimum length solution (see Menke, 1989, p. 54). Therefore, the WML solution has a basis in information theory.

3.2 Updating prior with first and second moment information

For the MRE solution, we have to minimize (33) subject to (34) along with the mean value constraint (35) and

$$\int q(\mathbf{m})[(\bar{\mathbf{d}} - \mathbf{Gm})(\bar{\mathbf{d}} - \mathbf{Gm})^T]d\mathbf{m} = \mathbf{R} \tag{47}$$

$q(\mathbf{m})$ is of the form (32) and again, the Lagrange multipliers determined from the constraints. We can then proceed in the same way as before;

Letting

$$(\mathbf{C}_p^{-1} + \mathbf{G}^T\mathbf{D}^{-1}\mathbf{G}) = \mathbf{C}_q^{-1} = \mathbf{A} \tag{48}$$

and

$$(\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{G}^T\mathbf{D}^{-1}\bar{\mathbf{d}}) = \mathbf{b} \tag{49}$$

We now have

$$\begin{aligned} J(\mathbf{m}) \;&=\; -\frac{1}{2}[\mathbf{m}^T\mathbf{A}\mathbf{m} - \mathbf{m}^T\mathbf{b} - \mathbf{b}^T\mathbf{m} + \mathbf{s}^T\mathbf{C}_p^{-1}\mathbf{s} + \\ &\quad \bar{\mathbf{d}}^T\mathbf{D}^{-1}\bar{\mathbf{d}} - 2\boldsymbol{\lambda}^{\boldsymbol{T}}\,(\mathbf{G}\mathbf{m} - \bar{\mathbf{d}})] \end{aligned} \tag{50}$$

Now taking the derivative of J, and setting the result to zero determines $<\mathbf{m}>$, as

$$\mathbf{A} < \mathbf{m} > -\mathbf{b} - \boldsymbol{\lambda}^{\boldsymbol{T}}\;\mathbf{G} = 0 \tag{51}$$

$$\mathbf{A} < \mathbf{m} >= \mathbf{b} + \boldsymbol{\lambda}^{\boldsymbol{T}}\;\mathbf{G} \tag{52}$$

$$< \mathbf{m} >= \mathbf{A}^{-1}\left[\mathbf{b} + \boldsymbol{\lambda}^{\boldsymbol{T}}\;\mathbf{G}\right] \tag{53}$$

Expanding these terms out yields

$$< \mathbf{m} >= (\mathbf{C}_p^{-1} + \mathbf{G}^T\mathbf{D}^{-1}\mathbf{G})^{-1}(\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{G}^T\mathbf{D}^{-1}\bar{\mathbf{d}} - \mathbf{G}^T\boldsymbol{\lambda}\;) \tag{54}$$

$$\mathbf{C}_q = (\mathbf{C}_p^{-1} + \mathbf{G}^T\mathbf{D}^{-1}\mathbf{G})^{-1} \tag{55}$$

Rearranging,

$$< \mathbf{m} >= \mathbf{C}_q(\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{G}^T\mathbf{D}^{-1}\bar{\mathbf{d}} - \mathbf{G}^T\boldsymbol{\lambda}\;) \tag{56}$$

The covariance can be written as

$$\mathbf{C}_q = \mathbf{C}_p - \mathbf{C}_p\mathbf{G}^T[\mathbf{G}\mathbf{C}_p\mathbf{G}^T + \mathbf{D}]^{-1}\mathbf{G}\mathbf{C}_p \tag{57}$$

Let us first examine a limiting case. Suppose we impose a second moment constraint condition that $\mathbf{C}_q = 0$. This means that $\mathbf{R} = 0$, and then $\mathbf{D}$ must also equal zero in the above and

$$\mathbf{C}_q = \mathbf{C}_p - \mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T)^{-1}\mathbf{G}\mathbf{C}_p \to 0 \tag{58}$$

Using the results from (44), we also arrive with

$$< \mathbf{m} >= \mathbf{s} + \mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T)^{-1}[\mathbf{d}^* - \mathbf{G}\mathbf{s}] \tag{59}$$

which is identical to the results of Jacob and van Der Geest (1991) for $\mathbf{R} = 0$, assuming $\bar{\mathbf{d}} = \mathbf{d}^*$. In (58) $\mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T)^{-1} = \mathbf{G}_*^{-1}$. The important lesson to be learned at this stage is that the second moment information in the posterior is not updated unless specifically required.

3.2.1 General Case: Determining Mean and Covariance

Noting the second MRE moment constraint is

$$E[(\mathbf{m}- < \mathbf{m} >)(\mathbf{m}- < \mathbf{m} >)^T] = \mathbf{C}_q \tag{60}$$

$$\int q(\mathbf{m})(\mathbf{Gm} - \bar{\mathbf{d}})(\mathbf{Gm} - \bar{\mathbf{d}})^T d\mathbf{m} = \mathbf{GC}_q\mathbf{G}^T = \mathbf{R} \tag{61}$$

The next step is to find the $\boldsymbol{\lambda}$ values. We know from the first MRE constraint.

$$\int q(\mathbf{m})(\bar{\mathbf{d}} - \mathbf{Gm})d\mathbf{m} = 0 \tag{62}$$

or $\bar{\mathbf{d}} = \mathbf{G} < \mathbf{m} >$. Substituting the above expression for the mean value of the model (58) results in

$$< \mathbf{m} >= \mathbf{C}_q(\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{G}^T\mathbf{D}^{-1} - \bar{\mathbf{d}}^T\boldsymbol{\lambda}\ ) \tag{63}$$

$$< \mathbf{m} >= (\mathbf{C}_q\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{C}_q\mathbf{G}^T\mathbf{D}^{-1} - \bar{\mathbf{d}}_q\mathbf{G}^T\boldsymbol{\lambda}\ ) \tag{64}$$

Now we know from the mean value constraint that $\mathbf{G} < \mathbf{m} >= \bar{\mathbf{d}}$ and multiplying the mean model by $\mathbf{G}$ yields

$$\mathbf{G} < \mathbf{m} >= \bar{\mathbf{d}} = \mathbf{GC}_q\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{GC}_q\mathbf{G}^T\mathbf{D}^{-1}\bar{\mathbf{d}} - \mathbf{GC}_q\mathbf{G}^T\boldsymbol{\lambda} \tag{65}$$

Solving for $\boldsymbol{\lambda}$ and substituting back into the expression for the mean value results in

$$\begin{aligned} < \mathbf{m} > \ &= \ \mathbf{C}_q\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{C}_q\mathbf{G}^T\mathbf{D}^{-1} + \mathbf{C}_q\mathbf{G}^T\mathbf{R}^{-1}\bar{\mathbf{d}} \\ &\quad -\mathbf{C}_q\mathbf{G}^T[\mathbf{GC}_q\mathbf{G}^T]^{-1}\mathbf{GC}_q\mathbf{C}_p^{-1}\mathbf{s} \\ &\quad -\mathbf{C}_q\mathbf{G}^T\mathbf{D}^{-1}\bar{\mathbf{d}}) \end{aligned} \tag{66}$$

This equation may simplify to

$$< \mathbf{m} >= \mathbf{C}_q\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{C}_q\mathbf{G}^T\mathbf{R}^{-1}\bar{\mathbf{d}} - \mathbf{C}_q\mathbf{C}_p^{-1}\mathbf{s} \tag{67}$$

and finally to

$$< \mathbf{m} >= \mathbf{C}_q\mathbf{G}^T\mathbf{R}^{-1}\bar{\mathbf{d}} \tag{68}$$

One can see here that knowledge of $\mathbf{C}_q$ is required to obtain a solution and for many problems this would not be known. If we choose to approximate it as $\mathbf{R} = \mathbf{C}_d = \mathbf{GC}_q\mathbf{G}^T$ and $\bar{\mathbf{d}} = \mathbf{d}^*$ then

$$< \mathbf{m} >= (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{d}^* \tag{69}$$

The reader should note that the above result is only valid in a "weak" sense, that is the matrix $\mathbf{C}_q = (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G})^{-1}$ in the above case (69) is likely singular for the underdetermined case and in this situation one would have to rely on the generalized inverse. We also note that replacing $\mathbf{R}$ for $\mathbf{C}_d$ involves an approximation that Bayes does not have. Equation (69) can be recognized as the classic least squares solution if there is more data than unknowns. Note also the prior covariance is ignored. This should perhaps come as no surprise to us that MRE ignores moments of the prior information if not strictly required.

We can write the equation for the mean as

$$< \mathbf{m} >= \mathbf{G}_*^{-1}\mathbf{d}^* \tag{70}$$

where $\mathbf{G}_*^{-1} = \mathbf{C}_q\mathbf{G}^T\mathbf{C}_d^{-1}$.

3.2.2 Constraining Covariance Only

If we choose to enforce the second moment constraint (63) and not the mean constraint (35), then we do not require that the data are fitted exactly. By examining equations (58, 59) the constraints on $\boldsymbol{\lambda}$ are not required. This reduces (58) and (59) to

$$< \mathbf{m} >= \mathbf{C}_q(\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{G}^T\mathbf{D}^{-1}\bar{\mathbf{d}}) \tag{71}$$

and the posterior covariance

$$\mathbf{C}_q = \mathbf{C}_p - \mathbf{C}_p\mathbf{G}^T[\mathbf{G}\mathbf{C}_p\mathbf{G}^T + \mathbf{D}]^{-1}\mathbf{G}\mathbf{C}_p \tag{72}$$

MRE requires that we know $\mathbf{C}_q$ and then given this we would have to find that matrix $\mathbf{D}$ to satisfy the constraint. If it is *assumed* that the posterior covariance is of the same form as Bayes

$$\mathbf{C}_q = \mathbf{C}_p - \mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T + \mathbf{C}_d^{-1})^{-1}\mathbf{G}\mathbf{C}_p \tag{73}$$

Then clearly $\mathbf{D} = \mathbf{C}_d^{-1}$ in (73) and (74). Using the results from (58), we also arrive with

$$< \mathbf{m} >= \mathbf{s} + \mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T + \mathbf{C}_d^{-1})^{-1}[\mathbf{d}^* - \mathbf{G}\mathbf{s}] \tag{74}$$

assuming $\bar{\mathbf{d}} = \mathbf{d}^*$.

3.3 Updating prior with first moment information and uncertain constraints

In general, since the input data are known only to within a certain accuracy, it is important that we do not satisfy the first momement constraint in a strong way, instead $\mathbf{d}^* \neq \mathbf{G} < \mathbf{m} >= \bar{\mathbf{d}}$. The inversion method should allow for errors in the measured data. The MRE approach can accommodate such uncertainty and this subject has been discussed by Johnson and Shore (1984), Ulrych et al. (1990) and Woodbury and Ulrych (1998b).

The problem is posed in the following way. Minimize $H(q,p)$, the entropy of $q(\mathbf{m})$ relative to $p(\mathbf{m})$, (34), subject to (35) and a more general form of equation (21);

$$[(\mathbf{G} < \mathbf{m} > -\bar{\mathbf{d}})^T(\mathbf{G} < \mathbf{m} > -\bar{\mathbf{d}})] \leq \epsilon^2 \tag{75}$$

or

$$(\int q(\mathbf{m})\mathbf{G}\mathbf{m}d\mathbf{m} - \mathbf{d}^*)^T(\int q(\mathbf{m})\mathbf{G}\mathbf{m}d\mathbf{m} - \mathbf{d}^*) \le \epsilon^2 \tag{76}$$

where ϵ is a known error term and replacing $\bar{\mathbf{d}}$ for $\mathbf{d}^*$. Note that ϵ is required for the analysis to proceed. If this value is not known, Ulrych and Woodbury (2003) show that it can be estimated in a variety of ways, including estimated from the data themselves, a rigorous approach using the real cepstrum and the AIC criterion. It can be shown (Johnson and Shore, 1984) that the MRE solution has the same form as equations (10) and (11) but with data constraints modified to

$$\mathbf{G} < \mathbf{m} >= \mathbf{d}^* - \epsilon\frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|} = \mathbf{d}^\dagger \tag{77}$$

for the case where the data errors are identically distributed and independent. The vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots \beta_n)$ is a vector of Lagrange multipliers. The important fact from the point of view of MRE is that the form of the solution with uncertain data is not changed; only a simple modification of the solution algorithm is required.

Past efforts in MRE along these lines take the point of view of the prior either being uniform or with some character such as a truncated exponential (Woodbury and Ulrych, 1996). In new work described here, previous results are modified to include a Gaussian prior. In this light we can use the following results. First define our posterior pdf in the form

$$q(\mathbf{m}) = p(\mathbf{m}) \exp\left[-\alpha - \boldsymbol{\beta}^T\mathbf{G}\mathbf{m}\right] \tag{78}$$

where

$$\boldsymbol{\beta} = 2\lambda\mathbf{G} < \mathbf{m} > \tag{79}$$

If $p(\mathbf{m})$ is Gaussian as before then

$$q(\mathbf{m}) = C_3 \times \exp\left[-\frac{1}{2}((\mathbf{m}-\mathbf{s})^T\mathbf{C}_p^{-1}(\mathbf{m}-\mathbf{s}))\right] \tag{80}$$

$$\times \exp\left[-\lambda_0 - \boldsymbol{\beta}^{\boldsymbol{T}}(\mathbf{d}^\dagger - \mathbf{G}\mathbf{m})\right] \tag{81}$$

As shown earlier, the mean value for $q(\mathbf{m})$ is

$$< \mathbf{m} >= \mathbf{s} + \mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T)^{-1}[\mathbf{d}^\dagger - \mathbf{G}\mathbf{s}] \tag{82}$$

and we note $\mathbf{C}_q = \mathbf{C}_p$. Note also these results could be expressed in a general way,

$$< \mathbf{m} >= \mathbf{s} + \mathbf{G}_*^{-1}[\mathbf{d}^\dagger - \mathbf{G}\mathbf{s}] \tag{83}$$

$$\boldsymbol{\beta} = (\mathbf{G}\mathbf{C}_p\mathbf{G}^T)^{-1}[\mathbf{d}^\dagger - \mathbf{G}\mathbf{s}] \tag{84}$$

The mean value written in terms of β

$$< \mathbf{m} >= \mathbf{C}_p \left[\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{G}^T \boldsymbol{\beta} \right] \tag{85}$$

The above equation for $< \mathbf{m} >$ includes a non-linear dependency on the Lagrange multipliers β_j in the term for $\mathbf{d}^\dagger$ and an iterative sequence must used to establish the final result. Note that this result can be shown to be identical in form to that of the WML solution of (45).

4 COMPARISONS OF BAYESIAN AND MRE APPROACHES

In this section we will summarize the approaches detailed in this paper and offer a comparison between them (see Tables I and II). Seth and Kapur (1990), Kapur and Kessavan (1992), Kapur et al. (1994), Macauly and Buck (1989) and Jacobs and van der Geest (1991) enunciated some of these differences and the essence of their comparisons is repeated and expanded upon here. The essential difference between the two is given below:

(i) Bayes: This approach is different than the classical methods of first obtaining a sample, and then using sampling distribution theory to obtain estimates of the parameters. Bayes assumes some prior probability (density) which the model follows. The method then proceeds to use a sample of observations, say $x_1, x_2, \ldots$ to update the prior probability to a new (revised) posterior probability. This new probability, or probability density, incorporates the information of the sample. More data can be then taken and the updated pdf can be renewed again in a continuous way. These features of Bayes, the assumptions of prior probabilities and continuous updating in the light of new observations, are of key importance. The prior pdf can play an important role in the inversion and has been the center of many disputes that have come about as the result of the use Bayes theorem.

(ii) MRE: This approach differs from both Bayes and classical statistical methods in that knowledge of moments are used as "data" rather than sample values, $x_1, x_2, \ldots$. MRE, like Bayes, presumes knowledge of a prior probability distribution and produces the posterior pdf itself. For example, suppose we have a univariate prior pdf that is Gaussian and we know the mean μ and variance σ^2. Suppose we have new information in the form of a mean μ_1. Then the posterior pdf is also Gaussian with mean changed to μ_1 but with the variance unchanged. MRE attempts to produce a pdf based on the information provided by new moments. It will use moments of the prior distribution only if new data on these moments is not available (see Woodbury, 2004). It is important to note here that MRE makes a strong statement that the constraints are exact and complete. That is, the M values are the only only ones operating and these are known exactly. In this way, MRE is maximally uncommitted with respect to unknown information.

Some examples of Bayesian inference along with MRE solutions are given below (Tables I and II). These would appear to be at first glance identical problems and are used to illustrate important differences and similarities. For the case of a linear inverse problem and the Bayes solution, imagine that the observations are error free; that is $\mathbf{C}_d \to \mathbf{0}$ (Table I, row 4). Using the relationships

$$< \mathbf{m} >= \mathbf{s} + \mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T + \mathbf{C}_d)^{-1}(\mathbf{d}^* - \mathbf{G}\mathbf{s}) \tag{86}$$

with $\mathbf{C}_d = \mathbf{0}$ results in

$$< \mathbf{m} >= \mathbf{s} + \mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T)^{-1}(\mathbf{d}^* - \mathbf{G}\mathbf{s}) \tag{87}$$

This mean value is identical to the MRE approach of updating based on the first moment only (Table I, row 5). However, the Bayes posterior covariance with $\mathbf{C}_d = 0$ is;

$$\mathbf{C}_q = \mathbf{C}_p - \mathbf{C}_p\mathbf{G}^T(\mathbf{G}\mathbf{C}_p\mathbf{G}^T)^{-1}\mathbf{G}\mathbf{C}_p = \mathbf{C}_p - \mathbf{G}_*^{-1}\mathbf{G}\mathbf{C}_p \to 0 \tag{88}$$

which is identical to the MRE result (Table I, row 3) but an additional constraint is required. Let us take the opposite case for the Bayes solution and imagine that the observations are totally inaccurate; that is $\mathbf{C}_d \to \infty$ or $\mathbf{C}_d^{-1} \to \mathbf{0}$. Recall that for Bayes

$$< \mathbf{m} >= (\mathbf{C}_p^{-1} + \mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G})^{-1}(\mathbf{C}_p^{-1}\mathbf{s} + \mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{d}^*) \tag{89}$$

If $\mathbf{C}_d^{-1} \to \mathbf{0}$ then $\mathbf{m} = \mathbf{s}$ and the covariance $\mathbf{C}_q = \mathbf{C}_p$. (See Table I, row 6). In this limiting case if we assume that the data are totally imprecise then the Bayes posterior mean is equal to the prior mean and the prior covariance is equal to the posterior covariance. These results are not the same for the MRE approach in the mean (Table I, row 5) but identical in the covariance. Note though, that not specifying a covariance is not the same as stipulating that the covariance is infinite.

Table I, row 7 shows how we can alter the MRE approach in such a way that it includes an estimate of the error in the observed data. In this way the results for the mean value are similar in form to pure Bayes (Table I, row 1), however no change in the covariance is possible with this approach.

Figure 1 shows the result of an example inverse problem involving a gravity survey over a layered and faulted zone. The example is similar in many respects to that of Mosegaard and Tarantola (1995). The problem considered is a vertical fault extending from surface to a maximum depth of 100 km. The example has 26 layers with a thickness derived from an exponential function with a mean value of 4 km in thickness. The actual layer densities are generated out of an uncorrelated normal distribution with mean value of 3,000 kg/m^3 and a standard deviation of 500 kg/m^3. A total 19 values of the first horizontal derivative of gravity are generated at evenly spaced points (2,000 m) across the fault using equation (1) of Mohan et al. (1986). These values are corrupted with Gaussian noise of 0.25 x 10^{-9} s^{-2}. This information then, forms the kernal matrix and data set in a discrete form.

Figure 1 shows in red the actual, or true density distribution with depth. The black line shows the Bayesian results with a prior Gaussian distribution of standard deviation of 500 kg/m^3 and correlation length of 5,000 m. The blue line shows the Bayesian results with the same prior as above and correlation length of 2,000 m. In green is the equivalent MRE solution with the same prior as the blue line case above. Here, the inverse matrix in (26) was found to be singular and an SVD was used to obtain the result. In all cases the inverse results match quite closely the actual density profile, at least in the upper portions of the profile. Overall, the general trend in the results is clear and reflects a typical smearing consistent with expected value determinations. In each case the predicted model generates data that very closely matches the observations.

5 CONCLUSIONS

Much of the existing work comparing maximum entropy, MRE and Bayes, can be attributed to Kapur and coauthors, but some of the derivations in various works were left incomplete. This paper attempts to fill in those gaps, step by step, in comparing inverse solutions. Specifically an undetermined-discrete linear inverse problem was chosen as a common template for comparisons. It is assumed here that the "noise" in a set of observables is Gaussian and the prior probability pdf is also assumed multivariate Gaussian.

In the various examples given in this paper, the problems look similar but are, in the final analysis, not quite equal. The similar results presented in Figure 1 and Table I are a reflection of the fact that the MRE posterior pdf, and its mean is constrained not by $\mathbf{d} = \mathbf{G}\mathbf{m}$ but by its first moment $E(\mathbf{d} = \mathbf{G}\mathbf{m}]$ a weakened form of the constraints (Jacobs and van der Geest, 1991) . If there is no error in the data then one should expect a complete agreement between Bayes and MRE and this is what is shown. Similar results are shown when second moment data is available (for example posterior covariance equal to zero). But dissimilar results are noted when we attempt to derive a Bayesian like result from MRE (see section 3.2.2). The MRE pdf is still multivariate Gaussian (see 32) but different than that of Bayes (18) because that distribution does not satisfy the same constraints. We can derive the Bayes solution from MRE principles in the undetermined case if we do not insist that the predicted data $\bar{\mathbf{d}}$ are

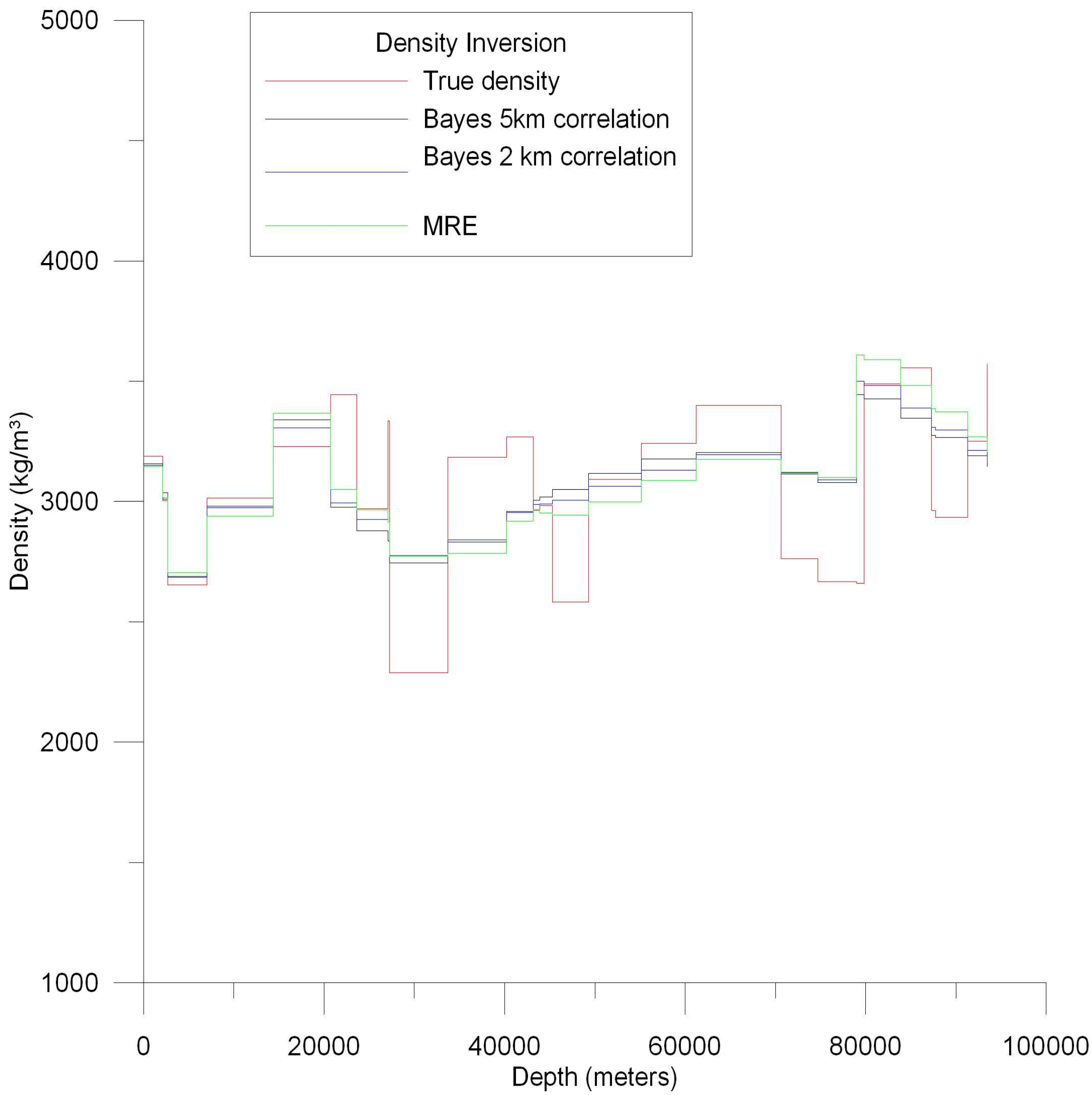

Figure 1. Comparison of linear inverse results over a layered fault.

equated to the observed data, and we only enforce a condition of known covariance $\mathbf{C}_q$ equal to what would be obtained from Bayes. If the unknown matrix of Lagrange multipliers is set equal to the observed noise error $\mathbf{C}_d^{-1}$ then we duplicate the Bayes solution. It is important to note though that this argument is circular.

In general though, since the input data are known only to within a certain accuracy, it is important that any inversion method allow for errors in the measured data. The MRE approach can accommodate such uncertainty and this subject has been discussed by Johnson and Shore (1984) and Ulrych et al. (1990). We show in this paper how a Gaussian prior can be updated with new data such that $\bar{\mathbf{d}}$ does not equal (exactly) the observed data. The important fact from the point of view of MRE is that the form of the solution with uncertain data is not changed; only a modification of the standard solution technique is required.

The classic results of the posterior solution under Gaussian priors and likelihood is repeated for clarity and completeness. A variety of MRE solutions are reproduced under a number of assumed moments and these stop at second-order central moments. Various solutions of Jacobs and van der Geest (1991) were repeated and clarified. Menke's weighted minimum length solution was shown to have a basis in information theory, and the classic least squares estimate is shown as a solution to MRE under the conditions of more data than unknowns and where we utilize the observed data and their associated noise.

One of the first questions posed in this paper was under what circumstances should one use entropic rather than Bayesian methods? The answer to that question may not be settled here. Certainly maximum entropy and MRE have had a huge success in geophysics (see Shore, 1981; Ulrych and Bishop, 1975). The MRE approach in the author's opinion is very flexible and may be much less demanding then the full Bayesian solution. MRE has shown to be ideal when constraining moments are known, for example: known second moments from autocorrelation (Shore, 1981), gross earth properties and density inversion (Woodbury and Ulrych, 1998b), and hydrologic applications (Kaplan et al., 2002; Singh, 2000). A very good case can be made for MRE in cases where little or no prior information is available (Kennedy et al., 2000) because MRE enforces conditions of independence. When some measure on the error on a set of observations is known the MRE approach can accommodate such uncertainty. Having said that, so can empirical Bayes, when say, the noise variance or other hyperparameters are unknown and have to be estimated from the data themselves.

If the prior in inverse problems has an important role then one could ask what are the ramifications associated with a particular choice? Certainly it is somewhat unpleasant to have to introduce some prior information to a problem in the first place to obtain a solution. As Menke (1989) suggested, the importance of the prior information depends on the use one plans for the results of the inversion. If one simply wants an exploration target then perhaps that choice is not important. If one plans on using the results of the inversion, specifically the errors in the estimates, then the validity of the prior assumptions (and pdfs in a probabilistic inversion) are critically important. This may be the case when the inverse problems are conducted in the realm of the environmental sciences or in engineering risk assessment. In these cases MRE may indeed play a important role in inverse problems in which it is desired to be maximally uncommitted with respect to unknown information.

It is important to note that (Seth and Kapur, 1990)

> "For the same problem, different methods of estimation can lead, quite expectedly to different results. Since statistical inference is inductive, there can be no perfect solution and there can be even differences of opinion as to which one is best."

6 ACKNOWLEDGMENTS

This article is meant as a tribute to Professor Jagat Kapur who passed away in September 4, 2002 at the age of 78 in Delhi. He was a pioneer in the field of maximum entropy who sought after a deep understanding of probabilities and the commonality in statistical inference approaches. He was giant in the field and he left an outstanding legacy of teaching, research and mentorship.

The author would also like to thank Dr. Tad Ulrych of U.B.C. for his valuable assistance and suggestions on this work.

7 REFERENCES

Bretthorst, G.L., 1988,*Bayesian Spectrum Analysis and Paramter Estimation*, Lecture Notes in Statistics, 48, 220p.

Buck, B. and V. A. Macaulay, eds., 1991,*Maximum Entropy In Action*, Claredon Press, Oxford, 220p.

Burg, J.P., 1975,*Maximum entropy spectral analysis*, Ph.D. Thesis, Stanford University.

Duijndam, A.J.W., 1988, Bayesian estimation in seismic inversion: Part I, Principles, *Geop. Prosp.*, **36**, 878-898.

Jacobs, F.J. and P.A.G. van der Geest, 1991, Spiking band-limited traces with relative-entropy algorithm, *Geophysics*, **56(7)**, 1003-1014.

Johnson, R.W. , 1983,*Algorithms for single-signal and multisignal minimum cross entropy spectral analysis*, Naval Research Laboratory Report 8667, Washington, D.C..

Johnson, R.W. and J.E. Shore, 1984,*Relative-Entropy Minimization With Uncertain Constraints - Theory And Application To Spectrum Analysis*, Naval Research Laboratory Report 5478, Washington, D.C..

Johnson, R.W. and J.E. Shore, 1983, Comment and correction to "Axiomatic derivation of the principle of maximum entropy and minimum cross entropy", *IEEE Trans Information Theory*, **IT-29**, 942-943.

Kaplin, J.D. and R.E. Howitt, 2002, Estimating nonpoint source pollution: An application of a sequential entropy filter, *Water Resour. Res.*, **38(1019)**, doi:10.1029/2000WR000088 .

Kapur, J.N. and H.K. Kesavan, 1992,*Entropy Optimization Principles With Application*, Academic Press, Inc..

Kapur, J.N,., 1989,*Maximum Entropy Models In Science and Engineering*, Wiley, N.Y., 635p.

Kapur, R.N., Kesavan, H.K. and G. Baciu, 1994, Comparisons between Bayesian and entropic methods for statistical inference. In *In, Stochastic and Statistical methods in Hydrology and Environmental Engineering*, (K.W. Hipel et al. editors), Vol 3, 149-162. Kluwer Academic

Kennedy, P.L., Woodbury, A.D. and K. Wang, 2000, Minimum relative entropy theory and application to surface temperature reconstruction from borehole temperature measurements, *Geop. Res. Letters*, **27(19)**, 3081-3085.

Kullback, S., 1959,*Information Theory And Statistics*, John Wiley, 999p.

Macaulay, V.A. and B. Buck, 1989, Linear inversion by the method of maximum entropy, *Inverse problems*, **5(5)**, 859-873.

Menke, W., 1989, *Geophysical Data Analysis: Discrete Inverse Theory, Revised Edition*, Academic Press, Inc., New York.

Mohan, N.L., Anandababu, L. and S.V. Seshagiri Rao, 1986, Gravity interpretation using the Mellin transform, *Geophysics*, **51(1)**, 114-122.

Mosegaard, Klaus, and Albert Tarantola, 1995, Monte Carlo Sampling of Solutions to Inverse Problems, *J. Geophys. Res.*, **Vol. 100, No. B7**, 12,431-12,44.

Neupauer, R.M., 1999,*A Comparison of Two Methods For Recovering the Release History of a Groundwater Contamination Source*, M.S. Thesis, N. M. Inst. of Mining and Technol, Socorro, NM.

Neupauer, R.M., Borchers, B. and J.L. Wilson, 2000, Comparison of inverse methods for reconstruction the release history of a groundwater contamination source, *Water Resour. Res.*, **36(9)**, 2469-2475.

Provencher, S.W., A constrained regularization method for inverting data represented by linear algebraic or integral equations, Comput. Phys. Commun., *27*, **213-227**, 1982a.

Provencher, S.W., CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic or integral equations, Comput. Phys. Commun., *27*, **229-242**, 1982b.

Scales, J.A., Smith, M.L., and S. Treitel, 2001,*Introductory Geophysical Inverse Theory*, Samizat Press.208 pp

Seth, A.K. and J.N. Kapur, 1990, Assessment of entropic and non-entropic methods. In *Maximum Entropy And Bayesian Methods*, (P. F. Fougere), Kluwer Academic Publishers.999p

Shore, J.E., 1981, Minimum cross-entropy spectral analysis, *IEEE Trans. Acoustics Speech and Signal Processing*, **ASSP-29**, 230-237.

Shore, J.E. and R.W. Johnson, 1980, Axiomatic derivation of the principle of maximum entropy and minimum cross entropy, *IEEE Trans Information Theory*, **IT-26**, 26-37.

Singh, V.P., 2000), The entropy theory as a tool for modeling and decision making in environmental and water resources, *Water SA*, **26(1)**, 1-12.

Tarantola, A., 1987,*Inverse Problem Theory*, Elsevier, 613p..

Ulrych, T., and T.N. Bishop, 1975, Maximum entropy spectral analysis and autoregressive deconvolution, *Rev. Geophys*, **13**, 183-200.

Ulrych, T.J. and M. Sacchi, 2006, *Information-Based Inversion and Processing with Applications, Volume 36 (Handbook of Geophysical Exploration: Seismic Exploration)*, Elsevier Science, 436p..

Ulrych, T.J., Bassrei, A. and M. Lane, 1990, Minimum relative entropy inversion of 1D data with applications, *Geophys. Prosp.*, **38**, 465-487.

Ulrych, T.J., Sacchi, M. and A.D. Woodbury, 2001, A Bayes tour of inversion: A tutorial, *Geophysics*, **66(1)**, 55-69..

Ulrych, T.J., Velis, D.R., and A.D. Woodbury, 2000, L-moments and C-moments, *Stoch. Envir. Res & Risk Assess.*, **14(1)**, 50-68.

Ulrych, T. J. and A.D. Woodbury, 2003, Extension of minimum relative entropy for noisy data, *Journal of Contaminant Hydrology*, **67(1-4)**, 13-25.

Woodbury, A.D., and T. J. Ulrych, 1993, Minimum relative entropy: Forward probabilistic modeling , *Water Resour. Res.*, **29(8)**, 2847-2860.

Woodbury, A.D., and T.J. Ulrych , 1996, Minimum relative entropy inversion: theory and application to the recovery of the release history of a groundwater contaminant, *Water Resour. Res.*, **32(9)**, 2671-2681.

Woodbury, A.D., and T.J. Ulrych , 1998a, Reply to Comment on " Minimum relative entropy inversion: theory and application to the recovery of the release history of a groundwater contaminant" by Kabala and Skaggs, *Water Resour. Res.*, **34(8)**, 2081-2084.

Woodbury, A.D., and T.J. Ulrych , 1998b, Minimum relative entropy and probabilistic inversion in groundwater hydrology, *Stoch. Hydrol. Hydr.*, **12**, 317-358.

Woodbury, A.D., 2004, A FORTRAN program to produce minimum relative entropy distributions, *Computers in Geosciences*, **30**, 131-138.

Woodbury, A.D., 2007, Combining Thermal and Hydraulic Data for Hydrologic Estimation. In *Chapter 13, in "Data Integration in Subsurface Hydrology*, ((D. Hyndman editor)), American Geophysical Union Monograph Series, 253 p.. ISBN:978-0-87590-437-5