A MATLAB Implementation of the Minimum Relative Entropy Method for Linear Inverse Problems

Roseanna M. Neupauer[1] and Brian Borchers[2]

[1] Department of Earth and Environmental Science [2] Department of Mathematics

New Mexico Institute of Mining and Technology, Socorro, NM 87801, USA

borchers@nmt.edu

August 23, 2000

Abstract

The minimum relative entropy (MRE) method can be used to solve linear inverse problems of the form $\mathbf{Gm} = \mathbf{d}$, where $\mathbf{m}$ is a vector of unknown model parameters and $\mathbf{d}$ is a vector of measured data. The MRE method treats the elements of $\mathbf{m}$ as random variables, and obtains a multivariate probability density function for $\mathbf{m}$. The probability density function is constrained by prior information about the upper and lower bounds of $\mathbf{m}$, a prior expected value of $\mathbf{m}$, and the measured data. The solution of the

inverse problem is the expected value of $\mathbf{m}$, based on the derived probability density function. We present a MATLAB implementation of the MRE method. Several numerical issues arise in the implementation of the MRE method and are discussed here. We present the source history reconstruction problem from groundwater hydrology as an example of the MRE implementation.

Keywords Maximum Entropy, Minimum Relative Entropy, Inverse Problems

1 Introduction

In this paper, we present a MATLAB implementation of the minimum relative entropy (MRE) method to solve linear inverse problems of the form

$$\mathbf{Gm} = \mathbf{d}, \tag{1}$$

where $\mathbf{G}$ is a matrix of size N by M, $\mathbf{m}$ is a vector of length M containing unknown model parameters, and $\mathbf{d}$ is a data vector of length N. The vector $\mathbf{d}$ consists of measured data, and typically includes measurement error. The matrix $\mathbf{G}$ is typically very badly conditioned and often underdetermined, making a conventional least squares approach impractical.

In the MRE method, the unknown model parameters, $\mathbf{m}$, are treated as random variables, and the solution to the inverse problem is obtained from the multivariate probability density function (pdf) of $\mathbf{m}$ (Woodbury, Ulrych, 1993). The pdf is selected

in a two-step process. First, we generate a prior distribution whose entropy is maximized subject to constraints imposed by the lower and upper bounds of $\mathbf{m}$ ($\mathbf{l}$ and $\mathbf{u}$, respectively) and by an estimate, $\mathbf{s}$, of the expected value of $\mathbf{m}$ (Woodbury, Ulrych, 1993). This prior distribution ensures that the solution of the inverse problem is within the specified bounds on $\mathbf{m}$, but does not guarantee that Eq. (1) is satisfied.

In the second step, we select a posterior distribution for $\mathbf{m}$, whose entropy is minimized relative to the prior distribution and subject to the constraint imposed by the measured data $\mathbf{d}$. The mean of the posterior pdf, $\hat{\mathbf{m}}$, can be used as a "solution" to the inverse problem. The 5th and 95th percentiles of the posterior distribution define a 90% "Bayesian" probability interval, similar to the 90% confidence interval in classical statistical methods. We can investigate the properties of the solution by using the posterior distribution to randomly generate many realizations of $\mathbf{m}$.

The goal of this paper is to present the MRE method in the context of a general discrete linear inverse problem, to discuss some important issues in the implementation of the method, and to describe our MATLAB implementation of the MRE method. We also present an example taken from groundwater hydrology.

2 The MRE Method

In this section, we describe the generation of the prior distribution, $p(\mathbf{m})$, and the posterior distribution, $q(\mathbf{m})$, of $\mathbf{m}$. The elements of $\mathbf{m}$ are assumed to be independent, so $p(\mathbf{m}) = \prod_{i=1}^{M} p(m_i)$ and $q(\mathbf{m}) = \prod_{i=1}^{M} q(m_i)$, where M is the number of model parameters.

In this discussion, we assume that the lower bound is zero. As we show later, non-zero lower bounds can easily be incorporated with a change of variables. For a continuous random variable with finite upper bounds, zero lower bounds, and a finite expected value within the bounds, it can be shown that the maximum entropy distribution is a multivariate truncated exponential distribution (Kapur, Kesavan, 1992; Woodbury, Ulrych, 1993). Using this fact we can derive the prior distribution

$$\begin{aligned} p(m_i) &= \frac{\beta_i \exp(-\beta_i m_i)}{1 - \exp(-\beta_i U_i)} \text{ for } \beta_i \neq 0 \\ p(m_i) &= \frac{1}{U_i} \text{ for } \beta_i = 0 \ ; \\ p(\mathbf{m}) &= \prod_{i=1}^{M} p(m_i) \ , \end{aligned} \tag{2}$$

where U_i is the upper bound of parameter m_i, and β_i is a Lagrange multiplier whose value is determined by satisfying the expected value constraint, $\int m_i p(m_i) dm_i = s_i$, where s_i is the prior expected value of m_i for $i = 1, 2, \ldots, M$. By integrating Eq. (2), we obtain the expected value equation that is used to evaluate β_i

$$\frac{-(\beta_i U_i + 1)\exp(-\beta_i U_i) + 1}{\beta_i \left[1 - \exp(-\beta_i U_i)\right]} = s_i \ . \tag{3}$$

Details of this derivation can be found in Neupauer (1999).

To obtain the posterior distribution, $q(\mathbf{m})$, we minimize its entropy relative to the prior distribution, $p(\mathbf{m})$. The entropy to be minimized is

$$H(q,p) = \int_{\mathbf{m}} q(\mathbf{m}) \ln \left[\frac{q(\mathbf{m})}{p(\mathbf{m})}\right] d\mathbf{m} \ , \tag{4}$$

where $p(\mathbf{m})$ is given in Eq. (2). This minimization is subject to two constraints—the

normalization requirement $\int_{\mathbf{m}} q(\mathbf{m})d\mathbf{m} = 1$, and the requirement that the posterior mean solution fit the data within a specified tolerance (Johnson, Shore, 1994; Woodbury, Ulrych, 1996)

$$||\mathbf{d} - \mathbf{G}\hat{\mathbf{m}}||^2 \leq \xi^2\epsilon^2 \ , \tag{5}$$

where $||\cdot||$ denotes the L_2 norm, $\hat{\mathbf{m}}$ is the mean of the posterior distribution, ϵ is the measurement error, and ξ is a parameter that depends on the assumed error model. This constrained optimization problem is solved by the method of Lagrange multipliers by minimizing

$$\phi = H(q,p) + \mu\left[\int_{\mathbf{m}} q(\mathbf{m})d\mathbf{m} - 1\right] + \gamma\left[\sum_{j=1}^{N}\left(\sum_{i=1}^{M} g_{ji}\hat{m}_i - d_j\right)^2 - \xi^2\epsilon^2\right\} \ , \tag{6}$$

where μ and γ are Lagrange multipliers. We have replaced the data inequality constraint in Eq. (5) with an equality constraint. If the initial guess of the solution does not satisfy the data constraint, the estimate will be modified until the data constraint is satisfied, which will first occur when $||\mathbf{d} - \mathbf{G}\hat{\mathbf{m}}||^2 = \xi^2\epsilon^2$.

This objective function, ϕ, is minimized relative to $q(\mathbf{m})$ when the following equality holds (Johnson, Shore, 1984):

$$0 = \ln\left[\frac{q(\mathbf{m})}{p(\mathbf{m})}\right] + 1 + \mu + \sum_{j=1}^{N}\lambda_j\left(\sum_{i=1}^{M} g_{ji}m_i - d_j\right) \ , \tag{7}$$

where $\lambda_j = 2\gamma\left(\sum_{i=1}^{M} g_{ji}m_i - d_j\right)$ are Lagrange multipliers on the individual measured data points. In terms of the Lagrange multipliers, λ_j, the data constraint in Eq. (5) can be rewritten as $||\boldsymbol{\lambda}||^2 = 4\gamma^2\xi^2\epsilon^2$, showing that $\gamma = ||\boldsymbol{\lambda}||/(2\xi\epsilon)$. With this definition of γ

and the definition of λ_j above, the data constraint holds when $\boldsymbol{\lambda}$ satisfies the nonlinear system of equations

$$\mathbf{d} - \mathbf{G}\hat{\mathbf{m}}(\boldsymbol{\lambda}) + \frac{\xi\epsilon}{||\boldsymbol{\lambda}||}\boldsymbol{\lambda} = 0 \; . \tag{8}$$

These Lagrange multipliers constrain the individual entries of model solution vector so that the norm of the residuals between the measured and predicted data satisfies the (equality) data constraint.

The resulting posterior distribution takes the form

$$\begin{aligned} q(m_i) &= \frac{a_i \exp(-a_i m_i)}{1 - \exp(-a_i U_i)} \quad \text{for } a_i \neq 0 \\ q(m_i) &= \frac{1}{U_i} \quad \text{for } a_i = 0 \; ; \\ q(\mathbf{m}) &= \prod_{i=1}^{M} q(m_i) \; , \end{aligned} \tag{9}$$

where $a_i = \beta_i + \sum_{j=1}^{N} \lambda_j g_{ji}$. Details of the derivation are given in Woodbury, Ulrych (1996) and Neupauer (1999). The solution to this inverse problem is $\hat{\mathbf{m}}$, given by

$$\begin{aligned} \hat{m}_i &= \frac{1 - (a_i U_i + 1)\exp(-a_i U_i)}{a_i \left[1 - \exp(-a_i U_i)\right]} \quad \text{for } a_i \neq 0 \\ \hat{m}_i &= \frac{U_i}{2} \quad \text{for } a_i = 0 \; . \end{aligned} \tag{10}$$

The uncertainty in the MRE solution can be expressed with probability levels of $q(\mathbf{m})$. For example, the 90% probability interval has as its bounds the 5th and 95th percentile probability levels, which are given by the values of m_i such that $\int_0^{m_i} q(m_i')\, dm_i' = 0.05$ and $\int_0^{m_i} q(m_i')\, dm_i' = 0.95$, respectively.

3 Implementation Issues

We wrote a MATLAB program to implement the MRE method, following the approach of the previous section. Several numerical issues arose that may not be apparent from the previous discussion, and are discussed here. The code is available at ftp.iamg.org.

3.1 Non-zero Lower Bounds

The code was written with lower bounds of zero. For non-zero lower bounds, the problem can be re-scaled by defining $\hat{\mathbf{m}} = \hat{\mathbf{m}}_o + \mathbf{l}$, where $\hat{\mathbf{m}}$ is the true solution, $\hat{\mathbf{m}}_o$ is the corresponding model solution for a zero lower bound, and $\mathbf{l}$ is the vector of lower bounds. We can solve for $\hat{\mathbf{m}}_o$ using the MATLAB routine, with the data modified as $\mathbf{d}_L = \mathbf{d} - \mathbf{G}\mathbf{l}$, where $\mathbf{d}_L$ is the data vector used in the MRE program, $\mathbf{d}$ is the true data vector, and $\mathbf{G}$ is the matrix of kernel values. The upper bounds and expected values must be replaced by $U_i - L_i$ and $s_i - L_i$, respectively, where L_i is the lower bound of model parameter m_i. After $\hat{\mathbf{m}}_o$ is computed with the MRE method, the true solution is obtained from $\hat{\mathbf{m}} = \hat{\mathbf{m}}_o + \mathbf{l}$ (Woodbury, Ulrych, 1996).

3.2 Error Models

The data constraint (Eq. 5) and the system of equations used to obtain the values of the Lagrange multipliers λ_j (Eq. 8) contain a parameter ξ that depends on the error model. We consider two possible error models–an additive error model with $d_j = d_j^0 + \epsilon_a \delta_j$ and a multiplicative error model with $d_j = d_j^0 + d_j^0 \epsilon_m \delta_j$, where ϵ_a and ϵ_m are the standard deviations of the additive and multiplicative error, respectively, δ_j is a standard normal

random number, d_j^0 is the true value of the data, and d_j is the measured value.

If the additive error model is used, then $\xi^2 = N$; and with the multiplicative error model, we use $\xi^2 = ||\mathbf{d}||^2$. The MRE implementation allows both additive and multiplicative components in the error model by replacing Eq. (8) with

$$\mathbf{d} - \mathbf{G}\hat{\mathbf{m}}(\boldsymbol{\lambda}) + \frac{\boldsymbol{\lambda}}{||\boldsymbol{\lambda}||}\left(\sqrt{N}\epsilon_a + ||\mathbf{d}||\epsilon_m\right) = 0 \; . \tag{11}$$

3.3 Prior Distribution

The prior distribution is calculated from Eq. (2), which is written in terms of the Lagrange multipliers, β_i. The values for each β_i are determined individually from Eq. (3) by using the bisection method (Gill et al., 1981). For certain values of β_i, Eq. (3) is subject to numerical error, which can be avoided with asymptotic approximations.

If $\beta_i = 0$, Eq. (3) is indeterminate. In the limit as $s_i \to U_i/2$, $\beta_i \to 0$. Thus, if $s_i = U_i/2$, the MRE program assigns $\beta_i = 0$. If $s_i \approx U_i/2$, β_i is small but non-zero, and Eq. (3) is subject to numerical error. To avoid this error, we solve for β_i using an asymptotic approximation to Eq. (3) for $|\beta_i|$ small ($|\beta_i| < 10^{-4}$ in the MRE code):

$$s_i \approx \frac{12U_i - 8\beta_i U_i^2 + 3\beta_i^2 U_i^3}{24 - 12\beta_i U_i + 4\beta_i^2 U_i^2 - \beta_i^3 U_i^3} \; . \tag{12}$$

If $s_i \approx U_i$, then $\beta_i \to -\infty$, and Eq. (3) is subject to numerical error. To avoid this error, we use the asymptotic approximation of $\beta_i = -1/(U_i - s_i)$ for $s_i \approx U_i$ in our MRE algorithm. If $s_i \approx 0$, then $\beta_i \to \infty$. For this case, the MRE algorithm sets β_i to the upper limit of the starting interval of the bisection method, and no further approximations are needed.

The prior distribution, $p(\mathbf{m})$, is shown in Eq. (2) in terms of the Lagrange multipliers, β_i. If $\beta_i \ll -1$ ($\beta_i < -10^2$ in the MRE code), the expressions shown in Eq. (2) cannot be evaluated numerically, and the prior distribution is approximated by

$$p(m_i) = \begin{cases} -2\beta_i \left[1 + \beta_i \left(U_i - m_i\right)\right] & U_i + 1/\beta_i \leq m_i \leq U_i \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{13}$$

3.4 Posterior Distribution

3.4.1 Lagrange Multipliers

The posterior distribution is calculated from Eq. (9), which is written in terms of $a_i = \beta_i + \sum_{j=1}^{N} \lambda_j g_{ji}$, where λ_j are Lagrange multipliers. The values of $\boldsymbol{\lambda}$ ($\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_N]^T$, where T denotes transpose) are calculated using the Newton-Raphson method with a line search to solve Eq. (8), using $F(\boldsymbol{\lambda}) = 0$ where

$$F(\boldsymbol{\lambda})_j = d_j - \sum_{i=1}^{M} g_{ji} \hat{m}_i(\boldsymbol{\lambda}) + \xi\epsilon \frac{\lambda_j}{||\boldsymbol{\lambda}||} \ . \tag{14}$$

Here d_j is the j^{th} measured data point, $\hat{m}_i$ is the expected value of m_i (Eq. 10), and $\xi\epsilon = \sqrt{N}\epsilon_a + ||\mathbf{d}||\epsilon_m$. We begin with $\lambda_j = 1$. If this satisfies the data constraint (Eq. 5), we use $\lambda_j = 1$ in Eq. (10) to obtain the solution to the inverse problem. Otherwise, we use the Newton-Raphson method with a line search to iteratively solve for the zeroes of Eq. (14) using

$$\boldsymbol{\lambda}^k = \boldsymbol{\lambda}^{k-1} - \left(\frac{\partial \mathbf{F}}{\partial \boldsymbol{\lambda}} \bigg|_{\boldsymbol{\lambda}^{k-1}} \right)^{-1} \mathbf{F}^{k-1} \tag{15}$$

where the superscripts denote the iteration, and the terms of the Jacobian matrix, $\partial \mathbf{F}/\partial \boldsymbol{\lambda}$, are

$$\frac{\partial F_j}{\partial \lambda_l} = -\sum_{i=1}^{M} g_{ji} \begin{bmatrix} \partial \hat{m}_i \\ \partial a_i \end{bmatrix} g_{li} + \frac{\xi \epsilon}{||\boldsymbol{\lambda}||} \begin{bmatrix} \delta_{jl} & \lambda_j \lambda_l \\ & ||\boldsymbol{\lambda}||^2 \end{bmatrix}, \tag{16}$$

where $l = 1, 2, \ldots, N$, δ_{jl} is the Kronecker delta, and

$$\begin{matrix} \partial \hat{m}_i \\ \partial a_i \end{matrix} = \frac{a_i^2 U_i^2 \exp\left(-a_i U_i\right) - \left[1 - \exp(-a_i U_i)\right]^2}{a_i^2 \left[1 - \exp(-a_i U_i)\right]^2}. \tag{17}$$

Eq. (14) is sufficiently non-linear that the step size calculated in Eq. (15) is not necessarily optimal. Therefore, we use the Newton-Raphson method to calculate the optimal search direction; then, we use a univariate golden section search (Gill et al., 1981) in that direction to calculate the optimal step length. The process is repeated until $||\mathbf{F}||/(1 + ||\mathbf{d}||)$ is less than a user-specified tolerance. We have found that the condition number of the Jacobian matrix can be high when $\epsilon = 0$. To reduce the condition number, we perform row-scaling on the Jacobian matrix. It is also possible to substantially reduce the condition number by including a small amount of noise, at a level that is negligible compared to the true data values.

3.4.2 Asymptotic Approximations

The posterior distribution, $q(\mathbf{m})$, in Eq. (9), and the expressions for $\hat{m}_i$ (Eq. 10) and $\partial m_i/\partial a_i$ (Eq. 17) are all written in terms of $a_i = \beta_i + \sum_{j=1}^{N} \lambda_j g_{ji}$. For certain values of β_i and a_i, these expressions are subject to numerical error. Several techniques, such as asymptotic approximations or scaling, can be used to avoid numerical error. In our

implementation, we used asymptotic approximations, which we present here.

We obtained Eq. (9) by minimizing the entropy of the posterior distribution relative to the prior distribution shown in Eq. (2). However, if $\beta_i \ll -1$, an asymptotic approximation to the prior distribution is used (Eq. 13), and the resulting the asymptotic approximation of the posterior distribution is

$$q(m_i) = \begin{cases} \frac{6\beta_i^2}{(b_i - 3\beta_i)}[1 + \beta_i(U_i - m_i)]e^{b_i(U_i - m_i)} & U_i + 1/\beta_i \leq m_i \leq U_i \ , \\ \\ 0 & \text{otherwise} \ , \end{cases} \tag{18}$$

where $b_i = \sum_{j=1}^{N} \lambda_j g_{ji}$.

The expression for $\hat{m}_i$ in Eq. (10) is subject to numerical error when $\beta_i \ll -1$ (i.e., when Eq. (18) is used as an asymptotic approximation of $q(m_i)$) and when $|a_i|$ is very small or very large. Using asymptotic approximations when necessary, the general expression for $\hat{m}_i$ is

$$\begin{aligned} \hat{m}_i \approx & \quad U_i - 1/a_i, & a_i \ll -1 \ \ (a_i < -10^2) \\ \hat{m}_i \approx & \quad 1/a_i, & a_i \gg 1 \ \ (a_i > 10^2) \\ \hat{m}_i \approx & \quad \frac{12U_i - 8a_iU_i^2 + 3a_i^2U_i^3}{24 - 12a_iU_i + 4a_i^2U_i^2 - a_i^3U_i^3}, & |a_i| \ll 1 \ \ (|a_i| < 10^{-4}) \\ \hat{m}_i \approx & \quad U_i + (b_i - \beta_i)/[\beta_i(b_i - 3\beta_i)], & \beta_i \ll -1 \\ \hat{m}_i = & \quad U_i/2, & a_i = 0 \\ \hat{m}_i = & \quad \frac{1 - (a_iU_i + 1)\exp(-a_iU_i)}{a_i[1 - \exp(-a_iU_i)]}, & \text{otherwise.} \end{aligned} \tag{19}$$

The expression for $\partial\hat{m}_i/\partial a_i$ in Eq. (16) is also subject to numerical error when $\beta_i \ll$

-1 and when $|a_i|$ is very small or very large. When $\beta_i \ll -1$, the partial derivative, $\partial\hat{m}_i/\partial a_i$, in Eq. (16) is replaced with $\partial\hat{m}_i/\partial b_i$. Using asymptotic approximations when necessary, the general expression for $\partial\hat{m}_i/\partial a_i$ is

$$\begin{aligned}
\frac{\partial\hat{m}_i}{\partial a_i} &\approx -U_i^2 \frac{15-15a_iU_i-8a_i^2U_i^2}{180-180a_iU_i-105a_i^2U_i^2}, && |a_i| \ll -1 \\
\frac{\partial\hat{m}_i}{\partial a_i} &\approx -1/a_i^2, && |a_i| \gg 1 \\
\frac{\partial m_i}{\partial b_i} &\approx -2/(b_i - 3\beta_i)^2 && \beta_i \ll -1 \\
\frac{\partial\hat{m}_i}{\partial a_i} &= \frac{a_i^2U_i^2\exp(-a_iU_i)-[1-\exp(-a_iU_i)]^2}{a_i^2[1-\exp(-a_iU_i)]^2}, && \text{otherwise.}
\end{aligned} \tag{20}$$

4 Example

The MRE inversion approach has been pioneered and used in groundwater hydrology for parameter estimation (Woodbury, Ulrych, 1993; Woodbury et al., 1995) and for the source history reconstruction problem (Woodbury, Ulrych, 1996; Woodbury et al., 1998a; Woodbury, Ulrych, 1998b). We present a source history reconstruction example from Woodbury, Ulrych (1996) as an illustration of our MRE algorithm. The data files are available at ftp.iamg.org.

The source history reconstruction problem involves a point source of groundwater contamination at a known location in a one-dimensional, saturated, homogeneous porous medium. The source concentration over time is unknown, but the present-day spatial distribution of the contamination plume is measured at discrete locations. Using this data with the MRE method, we reconstruct the concentration history at the contamination source.

For this example, the true source history function is

$$C_{in}(t) = \exp\left[-\frac{(t-130)^2}{2(5)^2}\right] + 0.3\exp\left[-\frac{(t-150)^2}{2(10)^2}\right] + 0.5\exp\left[-\frac{(t-190)^2}{2(7)^2}\right], \quad (21)$$

where t is in days, and C_{in} is dimensionless. This source history function is shown in Fig. 1. The solution vector, $\mathbf{m}$, is a discretized version of this curve. We discretized the source history function into 100 evenly-spaced times, at three-day intervals between $t = 0$ days and $t = 297$ days.

The governing equation of contaminant transport in groundwater is the advection dispersion equation

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x}, \quad (22)$$

where $C = C(x,t)$ is concentration at time t at location x, D is the dispersion coefficient, and v is the groundwater velocity. The initial and boundary conditions for Eq. (22) are $C(x,0) = 0$, $C(0,t) = C_{in}(t)$, and $C(x,t) \to 0$ as $x \to \infty$. For this problem, we use $v = 1$ m/d and $D = 1$ m^2/d. Using Eqs. (21) and (22), the plume at $T = 300$ days was calculated and is shown in Fig. 2. The plume was sampled at sixty points, at five-meter intervals from $x = 5$ m to $x = 300$ m; we assumed an additive noise with $\epsilon_a = 0.005$. The sampled data are also shown in Fig. 2. These data comprise the vector $\mathbf{d}$.

The elements of the matrix $\mathbf{G}$ are $g_{ji} = \Delta t f(x_j, T - t_i)$, where Δt is the temporal discretization of the source history function ($\Delta t = 3$ days); T is the sample time

($T = 300$ days), x_j is the location of the j^{th} measurement, t_i is the time corresponding to the i^{th} element of $\mathbf{m}$, and $f(x_j, T - t_i)$ is the solution of Eq. (22) for a pulse input at $x = 0$ and $t = 0$, given by

$$f(x_j, T \quad t_i) = \frac{x_j}{2\sqrt{\pi D(T - t_i)^3}} \exp \left\{ \begin{array}{c} [x_j - v(T - t_i)]^2 \\ 4D(T - t_i) \end{array} \right\} . \tag{23}$$

For this problem, we assume an upper bound on $\mathbf{m}$ of $U_i = 1.1$, and a prior expected value of $s_i = \exp\{-(t_i - 150)^2/800\}$. We solved this inverse problem using the MRE algorithm, and obtained the solution shown in Fig. 3. The results of MRE algorithm match well with the true solution, although the small peak near $t = 150$ days is not reproduced. The true solution falls within the 90% probability interval.

Using the posterior distribution, $q(\mathbf{m})$, we generated 30 realizations of the model solution vector $\mathbf{m}$. An example of one realization is shown in Fig. 4. The realization follows the same pattern as the true solution, but is less smooth. An individual realizations does not necessarily satisfy the data constraint; however, the expected value of $q(\mathbf{m})$ does. The results of the 30 realizations are plotted in Fig. 5, along with the 90% probability interval. The results show that the randomly-generated solutions fall within the 90% probability interval most of the time.

5 Summary and Conclusions

A MATLAB program for solving linear inverse problems using the minimum relative entropy method was presented. We discussed several implementation issues that may not be apparent from the MRE theory. We solved the minimization problem using the

method of Lagrange multipliers. In evaluating the Lagrange multipliers, the system of equations is sufficiently non-linear that the Newton-Raphson method fails to converge. We used the Newton-Raphson method to compute a search direction and a univariate golden section search to obtain the optimal step length. Other implementation issues included asymptotic approximations as the values of the Lagrange multipliers approach $=\infty$ or zero.

We demonstrated the MATLAB algorithm on a source history reconstruction problem from groundwater hydrology. The results show that the MRE method performs well for this simple test case. The solution to the inverse problem (i.e., the expected value of the posterior distribution) matches the true solution well, and the true solution is contained within the 90% probability levels of the posterior distribution. We have shown that individual realizations of the model solution do not necessarily satisfy the data constraint; however the expected value of the posterior distribution does satisfy the data constraint.

Acknowledgments

This research was supported in part by the Environmental Protection Agency's STAR Fellowship program under Fellowship No. U-915324-01-0. This work has not been subjected to the EPA's peer and administrative review and therefore may not necessarily reflect the views of the Agency and no official endorsement should be inferred. Allan Woodbury provided information about his earlier implementation of the minimum relative entropy inversion method. The authors acknowledge two anonymous reviewers for their valuable comments.

References

Gill, P.E., Murray, W., Wright, M.H., 1981. Practical Optimization. Academic Press, Inc., San Diego, California, 401 pp.

Johnson, R.W., and Shore, J.E., 1984. Relative-entropy minimization with uncertain constraints – Theory and application to spectrum analysis. *NRL Memorandum Report 5478*, Naval Research Laboratory, Washington, D.C., 12 pp.

Kapur, J.N., Kesavan, H.K., 1992. Entropy Optimization Principles with Applications. Academic Press, San Diego, California, 408 pp.

Neupauer, R.M., 1999. A Comparison of Two Methods for Recovering the Release History of a Groundwater Contamination Source. M.Sc. Thesis, New Mexico Institute of Mining and Technology, Socorro, New Mexico, 197 pp.

Woodbury, A.D., Ulrych, T.J., 1993. Minimum relative entropy: Forward probabilistic modeling. Water Resources Research 29 (8), 2847–2860.

Woodbury, A.D., Render, F., Ulrych, T., 1995. Practical probabilistic ground-water modeling. Ground Water 33 (4), 532–538.

Woodbury, A.D., Ulrych, T.J., 1996. Minimum relative entropy inversion: Theory and application to recovering the release history of a groundwater contaminant. Water Resources Research 32 (9), 2671–2681.

Woodbury, A.D., Sudicky, E., Ulrych, T.J., Ludwig, R., 1998a. Three-dimensional plume source reconstruction using minimum relative entropy inversion. Journal of Contaminant Hydrology 32, 131–158.

Woodbury, A.D., Ulrych, T.J., 1998b. Minimum relative entropy and probabilistic inversion in groundwater hydrology. Stochastic Hydrology and Hydraulics 12, 317–

358.

Figure Captions

Figure 1: True source history function.

Figure 2: True plume at $T = 300$ days (solid line) and sampled data (circles).

Figure 3: Inverse problem solution for the example problem. (A) MRE solution (solid line), true source history function (dashed line), prior expected value (dot-dashed line). (B) MRE solution (solid line), 5th percentile probability level (dashed line), 95th percentile probability level (dot-dashed line).

Figure 4: One realization of the model solution. solid line: realization; dashed line: true source history.

Figure 5: Results of 30 realizations of the model solution (dots) and the 90% probability interval (solid lines).

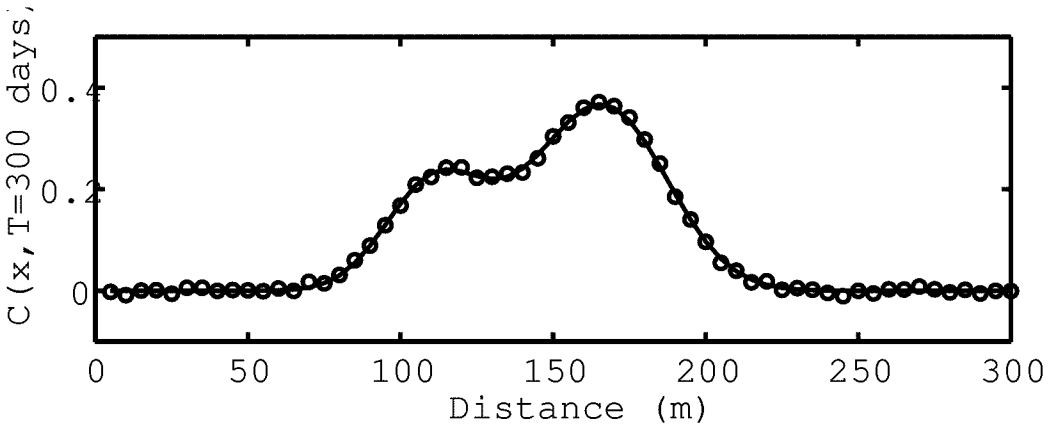

Figure 1

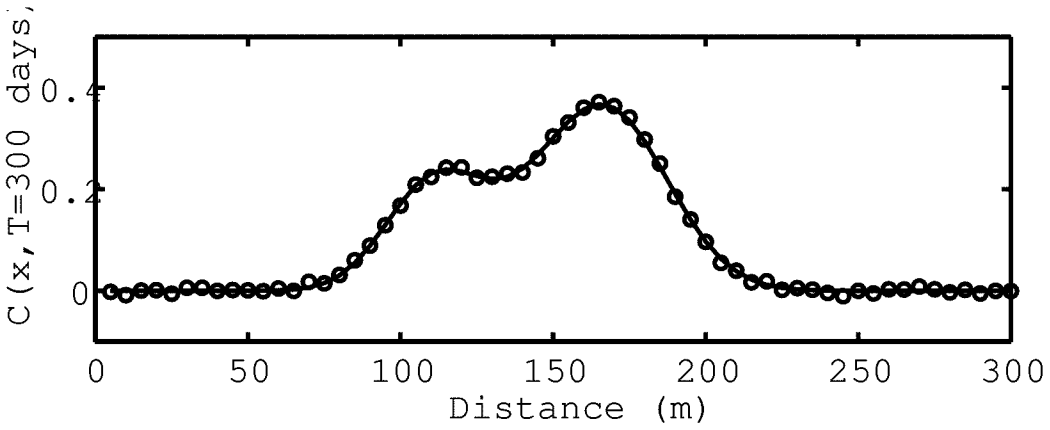

Figure 2

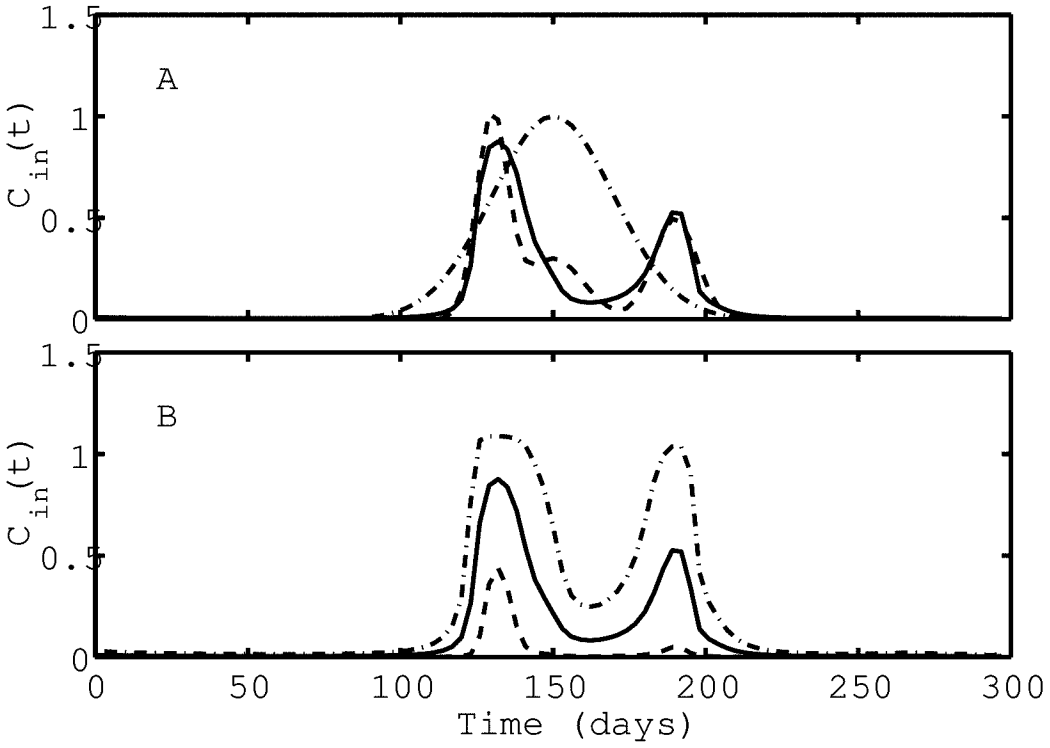

Figure 3

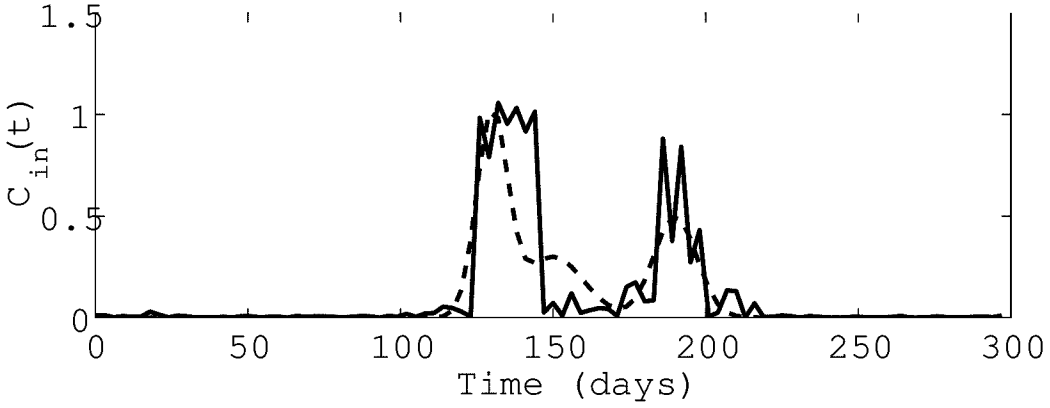

Figure 4

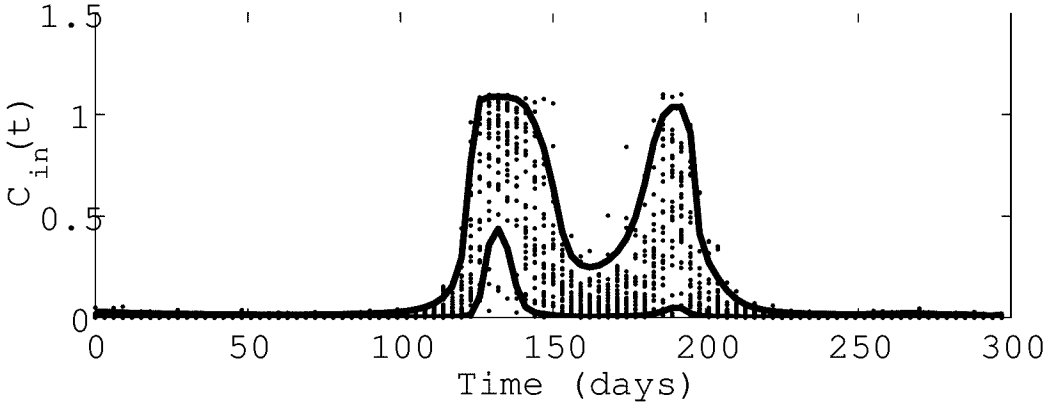

Figure 5