

# SANDIA REPORT

SAND2007-5975

Unlimited Release

Printed September 2007

## Markov Models and the Ensemble Kalman Filter for Estimation of Sorption Rates

Eric D. Vugrin, Sean A. McKenna, Kay W. Vugrin

Prepared by  
Sandia National Laboratories  
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Approved for public release; further dissemination unlimited.



**Sandia National Laboratories**

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

**NOTICE:** This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from  
U.S. Department of Energy  
Office of Scientific and Technical Information  
P.O. Box 62  
Oak Ridge, TN 37831

Telephone: (865) 576-8401  
Facsimile: (865) 576-5728  
E-Mail: [reports@adonis.osti.gov](mailto:reports@adonis.osti.gov)  
Online ordering: <http://www.osti.gov/bridge>

Available to the public from  
U.S. Department of Commerce  
National Technical Information Service  
5285 Port Royal Rd  
Springfield, VA 22161

Telephone: (800) 553-6847  
Facsimile: (703) 605-6900  
E-Mail: [orders@ntis.fedworld.gov](mailto:orders@ntis.fedworld.gov)  
Online ordering: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



# **Markov Models and the Ensemble Kalman Filter for Estimation of Sorption Rates**

Eric D. Vugrin

Performance Assessment and Decision Analysis Department  
Carlsbad Programs Group  
Sandia National Laboratories  
4100 National Parks Highway  
Carlsbad, NM 88220  
edvugri@sandia.gov

Sean A. McKenna

Geohydrology Department  
P.O. Box 5800, MS 0735  
Sandia National Laboratories  
Albuquerque, NM 87185-0735  
samcken@sandia.gov

Kay White Vugrin

Performance Assessment and Decision Analysis Department  
Carlsbad Programs Group  
Sandia National Laboratories  
4100 National Parks Highway  
Carlsbad, NM 88220  
kwvugri@sandia.gov

## **Abstract**

Non-equilibrium sorption of contaminants in ground water systems is examined from the perspective of sorption rate estimation. A previously developed Markov tran-

sition probability model for solute transport is used in conjunction with a new conditional probability-based model of the sorption and desorption rates based on breakthrough curve data. Two models for prediction of spatially varying sorption and desorption rates along a one-dimensional streamline are developed. These models are a Markov model that utilizes conditional probabilities to determine the rates and an ensemble Kalman filter (EnKF) applied to the conditional probability method. Both approaches rely on a previously developed Markov-model of mass transfer, and both models assimilate the observed concentration data into the rate estimation at each observation time. Initial values of the rates are perturbed from the true values to form ensembles of rates and the ability of both estimation approaches to recover the true rates is examined over three different sets of perturbations. The models accurately estimate the rates when the mean of the perturbations are zero, the unbiased case. For the cases containing some bias, addition of the ensemble Kalman filter is shown to improve accuracy of the rate estimation by as much as an order of magnitude.

# Contents

Introduction .....	9
Problem Set Up .....	11
Physical Set Up .....	11
Temporal and Spatial Discretization .....	11
Markov Process Model of Sorption-Desorption and Rate Estimating Models ...	12
Sorption Rate Estimation Model .....	13
Desorption Rate Estimation Model .....	14
Rate Estimation .....	15
Implementation of the EnKF .....	16
Numerical Simulations .....	18
Results .....	20
LOW Simulation Results .....	20
MED Simulation Results .....	21
HIGH Simulation Results .....	22
Conclusions .....	27

# Figures

1	Discretization of a Five Zone Streamline . . . . .	12
2	Logic Flow for Calculating Markov Rates . . . . .	16
3	Rate REs for LOW Perturbations . . . . .	21
4	Rate REs for MED Perturbations . . . . .	24
5	Rate REs for HIGH Simulations . . . . .	25

## Tables

1	Possible Pathways for Solute Particle to Remain in Sorbed State for 1 Out of the 4 First Time Steps . . . . .	14
2	True Rates Used for Observation Calculations . . . . .	18
3	LSEs for LOW Simulations . . . . .	22
4	LSEs for MED Simulations . . . . .	23
5	MED Concentrations for $N_e = 150$ . . . . .	23
6	HIGH Concentrations for $N_e = 150$ . . . . .	25
7	LSEs for HIGH Simulations . . . . .	26



# Introduction

Mass transfer of solute between the mobile and immobile domains in permeable media is strongly controlled by the processes of diffusion and sorption and the development of models that parameterize these processes has been a longstanding area of research (e.g., *van Genuchten and Wierenga* [1976]). More recent work [*Brusseau et al.*, 1991; *Chen and Wagenet*, 1997; *Cunningham and Roberts*, 1998; *Goltz and Roberts*, 1986; *Haggerty et al.*, 2000; *Jaekel et al.*, 1996; *Miller and Weber*, 1986; *Pedit and Miller*, 1994; *Quinodoz and Valocchi*, 1993] has focused on identifying heterogeneous rates of mass transfer that are due to local scale variability in the diffusion and sorption properties of the permeable media. As pointed out by *Haggerty et al.* [2004], variability in sorption strength increases variation in mass transfer time scales. Here we focus on the estimation of varying sorption and desorption rates. These rates tend to slow the average travel time of a solute through the permeable media and also influence mass transfer by increasing the capacity of the immobile zone to retain solute as well as decreasing the mass-transfer rate between the two zones. Techniques are developed that accommodate real-time data assimilation for the estimation of sorption and desorption rates.

Our focus here on real-time assimilation, as opposed to off-line inverse parameter estimation, is driven by the eventual goal of developing parameter estimation techniques that can utilize natural stimuli in the updating of models. Interest in using natural stimuli as energy sources for the investigation of natural systems is increasing with relevant examples found in ground water hydrology and seismology (e.g., *Xiang and Yeh*, [2005]; *Gret et al.*, [2005]). Here we initiate development of such techniques using a relatively simple example of a known contaminant source employing a previously developed particle tracking based model of non-equilibrium sorption processes.

Particle-based approaches provide an efficient means of modeling solute transport where mass-transfer between mobile (advective) and immobile domains is an important process. Relative to Eulerian approaches, particle-based approaches have the advantage of eliminating numerical dispersion and can often be computationally efficient in strongly heterogeneous media (e.g., *Tompson*, [1993]; *LaBolle et al.*, [1996]). Particularly, Markov models are an attractive choice for modeling the sorbed and non-sorbed phases of the particle with transition probabilities representing the sorption and desorption rates [*Andricevic and Fofoula-Georgiou*, 1991; *Mishra et al.*, 1999; *Valocchi and Quinodoz*, 1989].

Here we examine the specific problem of non-equilibrium sorption processes occurring in series within discrete zones along a single streamline. A different pair of forward and reverse sorption rates is defined for each discrete zone. This problem is analogous to solute transport through a column experiment with a series of different materials or field-scale transport through different geochemically distinct hydrofacies (e.g., vertical transport through a sedimentary sequence).

The Kalman filter [*Kalman*, 1960] is a well known technique for prediction of system behavior. The Kalman filter is a “predictor-corrector” method that uses observa-

tion/measurement data to correct model predictions, and correction of model predictions is determined by tracking of error statistics in the form of an error covariance matrix [Welch and Bishop, 2003]. The Kalman filter can be applied when system variables, or the states of the system, are modeled with a linear stochastic difference equation (e.g., Assouline, [1993]). The Kalman filter is well-suited for rapid updating of model predictions as it only requires information about the system for the previous time step, not all previous times, in order to predict system behavior at a future time step. This feature of the Kalman filter can provide a significant time and computational savings over other techniques, such as parameter estimation, that require information from all previous times.

A limitation of the Kalman filter is that it is only applicable to linear systems, but the extended Kalman filter algorithm has been applied to nonlinear systems. The extended Kalman filter uses linearization techniques to approximate error statistics and incorporates these statistics into the Kalman filter framework [Welch and Bishop, 2003]. However, the extended Kalman filter has had limited success for “highly” nonlinear systems. The ensemble Kalman filter (EnKF), developed specifically for nonlinear systems, was first introduced by Evensen [1994] and has been used for a variety of applications (see Evensen [2003] for a thorough listing of EnKF applications). The EnKF uses Monte Carlo sampling to create an ensemble of input parameters, and error statistics are estimated with the ensemble of model predictions.

The fundamental goal for tracer tests in aquifers is the determination of in-situ aquifer properties based on experimental results. Here we examine the specific issue of accurately estimating the spatially varying sorption and desorption rates from the concentration measurements obtained along the flow path. The robustness of the estimation approach is tested by varying the amount of bias in the initial estimates of the rates. Furthermore, if these rates can be estimated accurately, can this estimation occur within an assimilation framework where each new measurement in time is used to update previous estimates. This paper describes and tests a set of models and techniques that we have developed for accurate, real-time estimation of spatially varying sorption and desorption rates.

# Problem Set Up

## Physical Set Up

We consider one-dimensional fluid flow. A solute is injected into a fluid with a Dirac pulse at time zero at the upstream end of a streamline. The solute is initially in an aqueous phase, and it transitions between an aqueous phase and a sorbed phase as it travels along the flow path. At the local scale, an Eulerian discretization of the 1-D advective-dispersive equation with rate-limited sorption and desorption is given as:

$$\frac{\partial C}{\partial t} + \frac{\partial S}{\partial t} = D \frac{\partial^2 C}{\partial x^2} - v \frac{\partial C}{\partial x}, \quad (1)$$

where  $C$  is the aqueous concentration ( $M/L^3$ ),  $D$  is the hydrodynamic dispersion coefficient ( $L^2/T$ ),  $v$  is the pore velocity ( $L/T$ ) and  $S$  is the sorbed concentration per volume of aquifer material ( $M/L^3$ ). Under the case of non-equilibrium, linear sorption examined here,

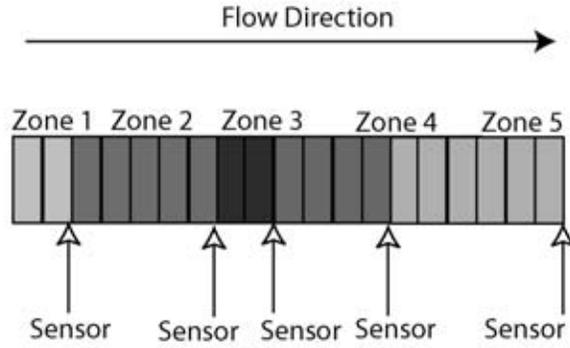
$$\frac{\partial S}{\partial t} = k_f C - k_r S, \quad (2)$$

where  $k_f$  and  $k_r$  are the sorption (forward) and desorption (reverse) rate coefficients such that  $K_d = k_f/k_r$  under the conceptual model of linear sorption. The local scale transport equation is solved using appropriate boundary and initial conditions.

Here we consider that the rates of sorption and desorption vary in space but not time. Sensors are placed at intervals of varying length along the streamline, and these sensors measure the concentration of the solute at their respective locations as the tracer passes. For the purposes of this article, the concentration measured is relative to the initial concentration at the influent end of the flow path, resulting in a normalized unitless quantity.

## Temporal and Spatial Discretization

Time is discretized in uniform step sizes of length  $\Delta t$ , and the streamline is divided into  $N_z$  zones of varying length. The rates of sorption and desorption are assumed constant within each discrete zone. Each zone is further subdivided into subintervals of uniform length  $\Delta x = v\Delta t$  where  $v$  denotes the fluid velocity. Furthermore, we assume that each sensor is positioned at the end of a zone and takes a concentration reading at the end of each time step. Figure 1 provides an example of how a streamline with five zones might be discretized.



**Figure 1.** Discretization of a Five Zone Streamline

## Markov Process Model of Sorption-Desorption and Rate Estimating Models

*Andricevic and Foufoula-Georgiou* [1991] model the sorption-desorption process as a Markov process with constant rates, and *Mishra et al.* [1999] use this framework to establish a model for spatially varying rates. In each of these models, a set of probabilities describes the likelihood of a particle of solute transitioning between the sorbed and aqueous phases. We use the following notations for these probabilities:

1.  $\hat{p}(j)$  - the probability that a particle in the aqueous phase stays in the aqueous phase for one time step in zone  $j$ .
2.  $[1 - \hat{p}(j)]$  - the probability that a particle in the aqueous phase becomes sorbed for one time step in zone  $j$ .
3.  $\tilde{p}(j)$  - the probability that a particle in the sorbed phase transitions to the aqueous phase for one time step in zone  $j$ .
4.  $[1 - \tilde{p}(j)]$  - the probability that a particle in the sorbed phase stays in the sorbed phase for one time step in zone  $j$ .

We denote the solute's rates of sorption and desorption by  $k_f(j)$  and  $k_r(j)$ , respectively, with the distribution coefficient,  $K_d(j) = k_f(j)/k_r(j)$ , and the probabilities  $\hat{p}(j)$  and  $\tilde{p}(j)$  are defined in terms of the rates:

$$\hat{p}(j) = 1 - k_f(j)\Delta t, \text{ and} \quad (3)$$

$$\tilde{p}(j) = k_r(j)\Delta t. \quad (4)$$

These definitions require that the time discretization is fine enough so that the probabilities are bounded between 0 and 1. Derivation of these probabilities is discussed by *Andricevic and Foufoula-Georgiou* [1991].

These probabilities, in conjunction with the following assumptions, form the basis of our particle tracking model:

1. At time  $t = 0$  all solute particles are in the aqueous phase at the the influent end of the streamline.
2. The decision about sorption/desorption is made at the beginning of each time step.
3. A solute particle advances one spatial step per time step and only if it is in the mobile aqueous phase.
4. Following *Mishra et al.* [1999a], it is assumed that no dispersion takes place. However, this model is developed for 1-D transport along streamlines and dispersion could be simulated using multiple streamlines or hybrid Lagrangian-Eulerian techniques [*DiDonato et al.*, 2003] or additional Markov processes [*Mishra and Gutjahr*, 1999].

## Sorption Rate Estimation Model

*Mishra et al.* [1999] provide the initial basis for developing the model for sorption rate estimation. Whereas *Mishra et al.* [1999] assume explicit knowledge of the rates to determine the concentration of solute at the end of the streamline, we attempt the inverse problem: using the concentrations to estimate the rates.

Let  $z_j$  denote the number of subintervals comprising the  $j^{th}$  zone, and

$$t_j = \sum_{i=1}^j z_i.$$

That is,  $t_j$  denotes the total cumulative number of subintervals in the first  $j$  zones . Since a particle can advance at most one spatial step per time step, the concentration readings from the  $j^{th}$  sensor are zero for all times less than  $t = t_j\Delta t$ . In order for a particle to reach the  $j^{th}$  sensor at the  $t_j$ th time step, the particle must remain in the aqueous phase for all  $t_j$  time steps. Thus, the probability that a solute particle reaches the  $j^{th}$  sensor precisely at time  $t = t_j\Delta t$  is

$$\prod_{i=1}^j [\hat{p}(i)]^{z_i}. \quad (5)$$

When tracking a large number of particles, the concentration reading from the  $j^{th}$  sensor after  $t_j$  time steps is approximately equal to the probability defined in (5). We define  $\mathbf{C}(t)$  to be the the  $N_Z \times 1$  vector whose  $j$ th entry,  $\mathbf{C}_j(t)$ , is the concentration reading from the  $j$ th sensor at time  $t$ . Equating the probability in (5) with the concentration  $\mathbf{C}_j(t_j\Delta t)$  and application of (3) leads to the following model for estimating  $k_f(j)$  :

$$k_f(j) = (\Delta t)^{-1} \begin{cases} 1 - [\mathbf{C}_j(t_j \Delta t)]^{1/z_j}, & j = 1 \\ 1 - \frac{[\mathbf{C}_j(t_j \Delta t)]^{1/z_j}}{\left\{ \prod_{i=1}^{j-1} [\hat{p}(i)]^{z_i} \right\}^{1/z_j}}, & j \neq 1. \end{cases} \quad (6)$$

## Desorption Rate Estimation Model

We introduce a simple example to motivate development of the desorption rate model. Consider a streamline that traverses two zones, and the first zone contains three subintervals. For the first four time steps, a particle must sorb for precisely one time step and remain in the aqueous phase for all other time steps to be measured by the first sensor during the fourth time step. Table 1 indicates the different pathways and associated probabilities that would allow a particle to contribute to  $\mathbf{C}_1(4\Delta t)$ . Note that the sensors only measure the particles in solution.

Table 1 shows that the probability that a particle contributes to  $\mathbf{C}_1(4\Delta t)$  is

$$P_T(1) = 3 [1 - \hat{p}(1)] \tilde{p}(1) [\hat{p}(1)]^2. \quad (7)$$

Replacing  $P_T(1)$  with  $\mathbf{C}_1(4\Delta t)$  in (7) and application of (4) leads to the following model for estimation of the desorption rate  $k_r(1)$ :

$$k_r(1) = \frac{\mathbf{C}_1(4\Delta t)}{3\Delta t [1 - \hat{p}(1)] [\hat{p}(1)]^2}. \quad (8)$$

We use similar logic to calculate the desorption rate for a zone consisting of  $z_j$  intervals.

The probability that a particle is in solution at the  $j^{\text{th}}$  sensor at the  $(t_j + 1)$ th time step is

$$P_T(j) = \left\{ \prod_{k=1}^j [\hat{p}(k)]^{z_k} \right\} \sum_{i=1}^j \left\{ \frac{z_i [1 - \hat{p}(i)] \tilde{p}(i)}{\hat{p}(i)} \right\}. \quad (9)$$

**Table 1.** Possible Pathways for Solute Particle to Remain in Sorbed State for 1 Out of the 4 First Time Steps

$t =$	0	$\Delta t$	$2\Delta t$	$3\Delta t$	$4\Delta t$	Probability
	A	S	A	A	A	$[1 - \hat{p}(1)] \tilde{p}(1) [\hat{p}(1)]^2$
	A	A	S	A	A	$[1 - \hat{p}(1)] \tilde{p}(1) [\hat{p}(1)]^2$
	A	A	A	S	A	$[1 - \hat{p}(1)] \tilde{p}(1) [\hat{p}(1)]^2$

This probability includes the probability that the particles could transition into the sorbed phase in previous zones. To estimate the desorption rate, our model requires that we know what fraction of the particles sorb in the  $j^{th}$  zone. In practice, one would not know this quantity, but we can estimate it using a conditional probability. Given that a particle has remained in the aqueous phase for all but one of the first  $t_j + 1$  time steps, the conditional probability that the solute particle remains in the aqueous phase while traveling through the first  $(j - 1)$  zones is

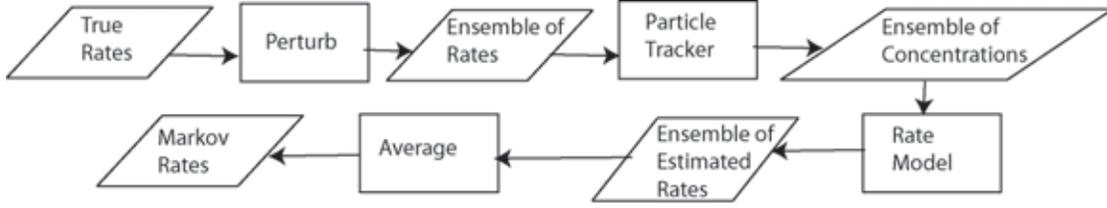
$$P_C(j) = [P_T(j)]^{-1} \times \left\{ \frac{z_j [1 - \hat{p}(j)] \tilde{p}(j)}{\hat{p}(j)} \right\} \prod_{k=1}^j [\hat{p}(k)]^{z_k}. \quad (10)$$

Replacing  $P_T(j)$  with  $C_j(t_j \Delta t + \Delta t)$  in (10) results in the following model for estimation of  $k_r(j)$ :

$$k_r(j) = \begin{cases} \frac{P_C(j) C_j(t_j \Delta t + \Delta t)}{z_j(\Delta t) [1 - \hat{p}(j)] [\hat{p}(j)]^{z_j - 1}}, & j = 1 \\ \frac{P_C(j) C_j(t_j \Delta t + \Delta t) \hat{p}(j)}{z_j(\Delta t) [1 - \hat{p}(j)] \left\{ \prod_{k=1}^j [\hat{p}(k)]^{z_k} \right\}^{-1}}, & j \neq 1. \end{cases} \quad (11)$$

## Rate Estimation

This section describes how we implement the Markov model to estimate the sorption and desorption rate for each zone. We start with a set of “true” sorption and desorption rates. We use these rates in the particle tracking solute transport model to generate a set of concentrations that are analogous to sensor readings in a tracer test. Ultimately, these are the true rates that we try to recover with the rate estimation models. We create an ensemble consisting of  $N_e$  sets of rates by perturbing each of the true sorption and desorption rates with percentages randomly sampled from a uniform distribution on  $(p_{min}, p_{max})$ . We use the notation  $k_f^i(j)$  and  $k_r^i(j)$  to denote the sorption and desorption rates, respectively, for the  $j^{th}$  zones of the  $i^{th}$  ensemble member. In order to recover the true rates that are perturbed to create the ensemble, the rates for each ensemble member are input to the particle tracking model, concentrations are calculated at each sensor, and rates are estimated using the rate models. Then, the mean rates are calculated across all ensemble members. These rates are estimated without the EnKF and are referred to, herein, as the Markov rates. Figure 2 illustrates the process for calculating the Markov rates.



**Figure 2.** Logic Flow for Calculating Markov Rates

## Implementation of the EnKF

This section outlines the procedure that we use to incorporate the EnKF with our rate estimation scheme. In addition to the ensemble of input values, the EnKF requires a set of observation measurements that are used to refine the model rate estimates. The particle tracking code is run using the true rates to recover a set of observed concentrations. We use the notation  $\bar{\mathbf{C}}_j$  to denote the vector of observed concentrations for the  $j$ th sensor, and the  $l$ th entry of  $\bar{\mathbf{C}}_j$  is the observed concentration for the  $j$ th sensor at time  $t = l\Delta t$ .

We use the following iterative procedure to estimate the true rates in the streamline:

1. For each ensemble member,
  - (a) Starting with  $j = 1$ , use the particle tracking code and  $k_f^i(j)$  and  $k_r^i(j)$  to compute the solute concentrations at all times for the  $j^{\text{th}}$  sensor,  $\mathbf{C}_j^i$ .
  - (b) Estimate the sorption rate for the  $j^{\text{th}}$  zone using  $\mathbf{C}_j^i$  and (6). Denote this estimated rate by  $\hat{k}_f^i(j)$ .
  - (c) Estimate the desorption rate for the  $j^{\text{th}}$  zone using  $\mathbf{C}_j^i$ ,  $\hat{k}_f^i(j)$ , and (11). Denote this estimated rate by  $\hat{k}_r^i(j)$ .
2. For each ensemble member we define the  $j^{\text{th}}$  state for the EnKF to be  $\mathbf{x}_j^i = \left[ \hat{k}_f^i(j) \quad \hat{k}_r^i(j) \quad \left( \mathbf{C}_j^i \right)^T \right]^T$ .

Incorporation of the time varying concentration in the definition of the state facilitates real-time assimilation of the concentration data in the rate estimation. We calculate the estimated state error covariance matrix  $\mathbf{P}_j$  with the equation

$$\mathbf{P}_j(m,n) = \sum_{i=1}^{N_e} \frac{\left[ \mathbf{x}_j^i(m) - \mu_j(m) \right] \left[ \mathbf{x}_j^i(n) - \mu_j(n) \right]}{N_e - 1}, \quad (12)$$

where the vector  $\mu_j$  denotes the average state across all ensemble members,  $\mathbf{x}_j^i(m)$  and  $\mu_j(m)$  denote the  $m$ th entries of the vectors  $\mathbf{x}_j^i$  and  $\mu_j$ , respectively, and  $\mathbf{P}_j(m,n)$  denotes the entry in the  $m$ th row and  $n$ th column of the matrix  $\mathbf{P}_j$ .

3. The matrix  $\mathbf{H}$  in the Kalman filter algorithm represents the linear function that relates the state to the measurement. Commonly,  $\mathbf{H}$  is the “model” by which the physical relationships and boundary and initial conditions are applied to the state to estimate the measurement. In the Kalman filter algorithm, this relationship is defined with the following equation:

$$\mathbf{z} = \mathbf{H}\mathbf{s} + \mathbf{d}.$$

The variable  $\mathbf{z}$  represents estimate of the the measured observation,  $\mathbf{s}$  represents the state, and  $\mathbf{d}$  represents measurement error. For our specific application, the vector of concentrations  $\mathbf{C}_j^i$  is the estimated set of observed measurements, and for simplicity, we do not artificially add measurement noise (i.e.,  $\mathbf{d} = 0$ ). Thus, for our problem, we define the matrix  $\mathbf{H}$  so that

$$\mathbf{C}_j^i = \mathbf{H}\mathbf{x}_j^i. \quad (13)$$

Each ensemble member state  $\mathbf{x}_j^i$  is updated with the following equation:

$$\tilde{\mathbf{x}}_j^i = \mathbf{x}_j^i + \mathbf{K}_j [\bar{\mathbf{C}}_j - \mathbf{H}\mathbf{x}_j^i]. \quad (14)$$

The gain  $K_j$  is defined by the EnKF as

$$\mathbf{K}_j = \mathbf{P}_j \mathbf{H}^T (\mathbf{H} \mathbf{P}_j \mathbf{H}^T + \mathbf{R})^{-1}, \quad (15)$$

where  $\mathbf{R}$  is the measurement error covariance matrix. In general applications of the Kalman filter,  $\mathbf{R}$  should be determined by knowledge of the accuracy of the measurements and the methods used to make the measurements [Evensen, 2003]. In practice,  $\mathbf{R}$  is not always known explicitly. It is common practice use a matrix that has been determined through iterative testing to yield acceptable state estimates.

4. For each ensemble member, replace  $k_f^i(j)$  and  $k_r^i(j)$  with  $\tilde{\mathbf{x}}_j^i(1)$  and  $\tilde{\mathbf{x}}_j^i(2)$ , respectively.
5. Repeat steps 1 through 4 for the  $(j + 1)$ th zone until all rates are estimated.

Note that estimation of the rates for the  $j^{\text{th}}$  zone incorporates the improved estimates of the rates from the previous zones. Also, this estimation can be updated incrementally as new observational data becomes available at each time step.

## Numerical Simulations

Calculations are performed to estimate the sorption and desorption rates for a 50 meter length streamline. Sensors are assumed to be placed at distances of 10, 15, 30, 45, and 50 meters from the upstream end of the streamline, and the fluid velocity is 0.1 m/day.

The particle tracking model described previously is run once with 10,000 particles and the rates listed in Table 2 to calculate the set of observation measurements. Particle tracking simulations for rate estimations use only 2,000 particles to reduce computational times, but the smaller number of particles reduces the resolution with which the concentrations can be measured and effectively introduces measurement error. For all simulations, concentration measurements are made for a total of 600 days at time intervals of 1.0 days. For EnKF simulations, we use an  $\mathbf{R}$  matrix of the form

$$\mathbf{R} = \beta \times \mathbf{I}_{600}, \quad (16)$$

where  $\mathbf{I}_{600}$  is the 600 by 600 identity matrix, and  $\beta$  is a constant. Use of the identity matrix here assumes that measurement errors are not correlated in time and therefore can be modeled as white noise. Trial and error assessment of different definitions of  $\mathbf{R}$  also showed that (16) yielded accurate rate estimates. The value of the constant that we use depends on the simulation, and we selected values of  $\beta$  that were iteratively determined to increase performance of the EnKF. In practice, determination of this value would take into account sensor specific information on the repeatability of the measurements.

Three sets of calculations are performed to assess the performance of the rate estimation models with and without the EnKF. The calculations differed only by the percentages with which the rates in Table 2 are perturbed to create the ensembles of rates. For the first set of calculations, the perturbation percentages are randomly sampled from a uniform distribution on the interval  $(-40.0, 40.0)$ . For the second and third sets of calculations, perturbation percentages are sampled from uniform distributions on  $(0.0, 80.0)$  and  $(50.0, 80.0)$ , respectively. These simulations are referred to as LOW, MED, and HIGH, respectively, to indicate the magnitude of the perturbations. Rates are estimated for ensemble sizes of 10, 25, 50, 75, 100, and 150, and we use  $\beta$  values of  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-6}$  for LOW, MED, and HIGH simulations, respectively.

**Table 2.** True Rates Used for Observation Calculations

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
$k_f(j)$ (1/day)	5E-3	5E-3	1E-3	9E-3	1E-3
$k_r(j)$ (1/day)	5E-3	5E-6	9E-3	1E-3	1E-4
$K_d(j)$ (-)	1E+0	1E+3	1.1E-1	9E+0	1E+1

For each simulation, rates are estimated two different ways. The “Markov” rates are estimated using only the sorption and desorption rate models , and EnKF rates are estimated using the EnKF algorithm in combination with the sorption and desorption rate models. All rates discussed in the following results are the mean values calculated across all members of the ensemble.

## Results

Two metrics are used to compare performance of the Markov models and the EnKF. The first metric is the absolute value of the relative percent errors (REs) between the mean estimated rates and the true rates used to create the observed measurements. For example, the RE for the first zone's ( $j = 1$ ) sorption rate is calculated as follows

$$RE = \frac{|\text{mean}\{k_f^i(1)\} - k_f^{true}(1)|}{k_f^{true}(1)}. \quad (17)$$

However, in the general case, the true rates are unknown. A traditional method for evaluating the performance of a parameter identification algorithm is to take the parameter estimates, input them into the model, and compare the resulting predictions with a set of corresponding observed measurements. For our specific problem this entails using the means of the estimated rates as input to the particle tracking model and comparing the resulting concentrations with the observed concentration data. We use this technique and calculate the least squares error between the observed concentrations and the concentrations predicted from the estimated rates. The least squares error (LSE) is calculated with the following equation:

$$LSE = \sum_{j=1}^{N_z} [\bar{\mathbf{C}}_j - \mathbf{C}_j]^T [\bar{\mathbf{C}}_j - \mathbf{C}_j], \quad (18)$$

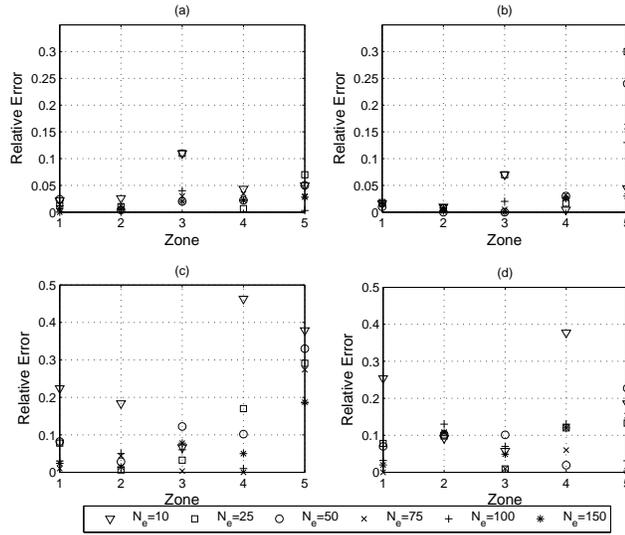
where  $\mathbf{C}(j)$  denotes the vector of concentrations at all times that result from the estimated rates and are measured at the  $j$ th sensor.

### LOW Simulation Results

Figure 3 shows the REs for Markov and EnKF rate estimation. For the first four sorption rates, both techniques yield REs less than 0.05 for almost all ensemble sizes. The Markov REs for the fifth sorption rate are approximately 0.05 for most ensemble sizes, but the EnKF requires an ensemble size of 150 members to reduce the RE to less than 0.05 for the fifth zone. The errors in rate estimation affect rate estimation for downstream zones, so the REs are largest for the fifth zone.

Desorption rate REs estimated with both approaches are less than 0.15 for most ensemble sizes for the first four zones. For the fifth zone, Markov REs are approximately 0.2 or larger, and EnKF REs are less than 0.25. For an ensemble consisting of 150 members, the EnKF RE is less than 0.01.

In general, desorption REs are larger than sorption REs, and this can be attributed to three factors. First, estimation of a desorption rate depends on the estimated sorption rate for the same zone, and any errors in the calculation of the sorption rate are compounded in estimation of the desorption rate. Secondly, as discussed in Section , the desorption rate



**Figure 3.** Rate REs for LOW Perturbations: (a) Markov Sorption Rates, (b) EnKF Sorption Rates, (c) Markov Desorption Rates, (d) EnKF Desorption Rates

estimation model uses a conditional probability to estimate the percentage of particles that sorb in a particular zone. The conditional probability may differ slightly from the actual percentage, so its use may affect the rate estimation results. Finally, most of the desorption rates are smaller than the sorption rates, even by several orders of magnitude smaller for some zones. Large REs in the desorption rate estimates do not have as large an effect on the concentrations as do large REs in the sorption rate estimates. Since the EnKF adjusts the rate estimates based upon how well the estimated concentration matches the observed concentration, the EnKF does not correct the desorption rates as well as it corrects the sorption rates.

Table 3 shows the concentration LSEs that result when the estimated rates are input into the particle tracking model. For LOW simulations, the EnKF does not significantly reduce rate REs in comparison to the Markov REs. As a result, the EnKF LSEs in Table 3 are similar in magnitude to the Markov LSEs.

## MED Simulation Results

Markov estimates result in sorption rate REs for MED simulations that are approximately 0.40 (Figure 4), the mean MED perturbation value. EnKF sorption rate REs, when  $N_e \geq 25$ , are less than 0.1 for the first four zones, and the RE for the fifth zone is about 0.15 for an ensemble with 150 members. The lower EnKF REs indicate that the EnKF can overcome some of the MED perturbation bias, but without the EnKF, REs resemble the

**Table 3.** LSEs for LOW Simulations

$N_e$	EnKF	Markov
10	1.95E-04	2.21E-04
25	1.52E-04	1.06E-04
50	6.90E-05	1.23E-04
75	9.05E-05	7.53E-05
100	1.21E-04	5.68E-05
150	7.96E-05	2.23E-05

mean perturbation.

The range of desorption rate REs is wide in comparison to the sorption RE ranges. Markov estimates of desorption rate result in REs that tend to increase with ensemble size. EnKF desorption rate REs are generally smallest with ensembles of 25 and 50 members, but EnKF desorption rate REs for ensembles with 75 or more members are smaller than their Markov approach counterparts, with the exception of the fifth zone.

The LSEs in Table 4 confirm that the EnKF rate estimates are better than the Markov-based rate estimates. For all ensemble sizes, the LSEs resulting from EnKF concentrations are two orders of magnitude smaller than Markov LSEs.

Table 5 lists concentrations that resulted when the estimated rates were input into the particle tracking model. The observed concentrations used for the EnKF are also listed. The concentrations listed correspond to the two shortest times required for a particle to travel from solute injection site to the sensors, and because the probability that an individual particle will remain in the aqueous phase is relatively large, the first concentration represents the peak concentration. The Markov rates are consistently larger than the true rates and lead to an underprediction of peak concentrations at the sensors. This underprediction is attributed to the overestimation of the sorption rates resulting in decreased probabilities of particles in solution staying in solution.

It should be noted that increasing the ensemble size does not significantly decrease the LSEs for both the EnKF and the Markov methods. This behavior occurs because the mean MED perturbation value is 0.40, not 0.

## HIGH Simulation Results

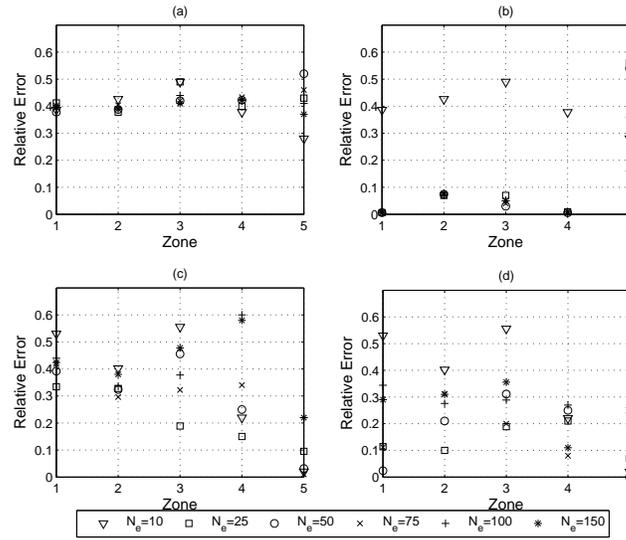
For all zones and ensemble sizes, Markov-based sorption rate REs are approximately 0.65, which is the mean value of the HIGH perturbation. EnKF sorption rate REs predicted with 10 member ensembles are similar. For larger ensembles, EnKF sorption rate REs for the first three zones are less than 0.2, and the REs for the fourth zone do not exceed 0.4. The

**Table 4.** LSEs for MED Simulations

$N_e$	EnKF	Markov
10	2.96E-04	5.12E-02
25	2.06E-04	5.20E-02
50	1.57E-04	4.79E-02
75	2.05E-04	5.11E-02
100	2.71E-04	4.96E-02
150	2.39E-04	5.07E-02

**Table 5.** MED Concentrations for  $N_e = 150$ 

Sensor	Time (days)	Obs. Conc. (-)	EnKF Conc. (-)	Markov Conc. (-)
1	100	6.04E-01	6.06E-01	4.93E-01
	101	1.70E-03	1.90E-03	2.60E-03
2	150	4.69E-01	4.62E-01	3.47E-01
	151	1.50E-03	1.70E-03	1.90E-03
3	300	4.06E-01	3.96E-01	2.80E-01
	301	1.60E-03	1.80E-03	2.30E-03
4	450	1.05E-01	1.03E-01	4.12E-02
	451	5.00E-04	5.00E-04	5.00E-04
5	500	9.90E-02	9.57E-02	3.87E-02
	501	5.00E-04	5.00E-04	5.00E-04



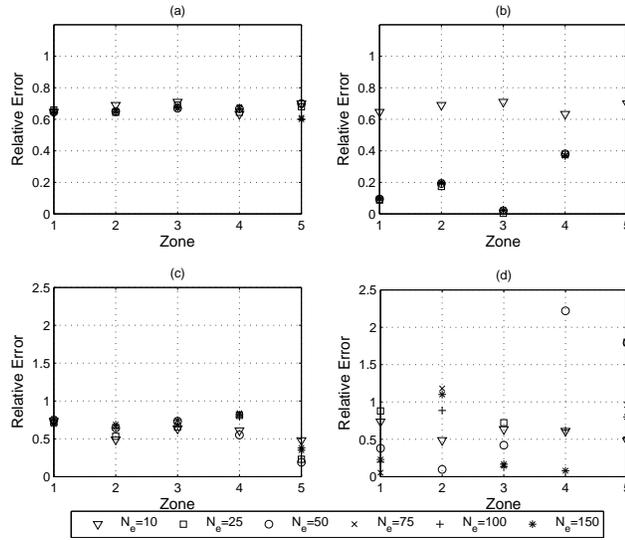
**Figure 4.** Rate REs for MED Perturbations: (a) Markov Sorption Rates, (b) EnKF Sorption Rates, (c) Markov Desorption Rates, (d) EnKF Desorption Rates

EnKF calculates sorption rates near  $0.08 \text{ day}^{-1}$  for the fifth zone, approximately eight times larger than the true sorption rate,  $0.01 \text{ day}^{-1}$ , and thus are off the vertical scale in Figure 5.

The EnKF underpredicts the fourth sorption rate by almost 40% which leads to higher overprediction of concentrations at the fourth sensor (Table 6). As a result, the EnKF overpredicts the fifth zone’s sorption rate so that the predicted concentrations more closely match the observed data.

Markov-based desorption rate REs are similar to their sorption rate counterparts. EnKF desorption rate REs vary considerably by zone and ensemble size. EnKF desorption rate REs are much smaller than the Markov REs for the first, third, and fourth zones, but they are larger for the second and fifth zones. The second and fifth rates are the smallest rates that we try to estimate (Table 2), and these zones are the shortest zones in the streamline. Hence, large relative errors that are small in magnitude have small impacts on predicted concentrations. Since the EnKF rate calculations depend upon the differences between predicted concentrations and observed concentrations, the second and fifth desorption rates are not as well estimated as other rates.

Table 7 lists the LSEs for HIGH simulations, and these differences are larger than the LSEs for the LOW and MED simulations. However, EnKF LSEs are still two orders of magnitude smaller than non-EnKF LSEs, indicating more accurate rate estimates. Because the mean HIGH perturbation value is not zero, increasing the ensemble size does not significantly decrease the LSEs.



**Figure 5.** Rate REs for HIGH Simulations: (a) Markov Sorption Rates, (b) EnKF Sorption Rates, (c) Markov Desorption Rates, (d) EnKF Desorption Rates

**Table 6.** HIGH Concentrations for  $N_e = 150$

Sensor	Time (days)	Obs. Conc. (-)	EnKF Conc. (-)	Markov Conc. (-)
1	100	6.04E-01	6.37E-01	4.35E-01
	101	1.70E-03	1.60E-03	3.70E-03
2	150	4.69E-01	4.71E-01	2.86E-01
	151	1.50E-03	1.30E-03	2.30E-03
3	300	4.06E-01	4.07E-01	2.21E-01
	301	1.60E-03	1.50E-03	2.60E-03
4	450	1.05E-01	1.77E-01	2.42E-02
	451	5.00E-04	9.00E-04	4.00E-04
5	500	9.90E-02	1.16E-01	2.22E-02
	501	5.00E-04	7.00E-04	4.00E-04

**Table 7.** LSEs for HIGH Simulations

$N_e$	EnKF	Markov
10	7.38E-03	1.11E-01
25	7.19E-03	1.10E-01
75	6.89E-03	1.09E-01
50	6.51E-03	1.09E-01
100	6.38E-03	1.09E-01
150	6.55E-03	1.09E-01

## Conclusions

Two parameter estimation models have been introduced that when implemented with a previously developed Markov model of solute transport with sorption can be used to estimate spatially varying sorption and desorption rates along a ground water streamline. It has been shown that when the ensemble of input rates has a mean perturbation value of zero, both models can estimate rates with less than five percent relative error, and addition of the EnKF to the rate estimation procedure does not significantly improve rate estimates. When mean perturbations are not zero, the EnKF can significantly improve estimation of relatively large rates and reduce concentration LSEs by two orders of magnitude, relative to the Markov model approach. The EnKF does have limitations when used in combination with the rate estimation models. The EnKF can have difficulty estimating relatively small rates that have lesser impacts upon concentration estimates. However, the improved prediction of large rates can reduce concentrations LSEs by two orders of magnitude despite the large relative errors for small rates.

Development of these approaches to rate estimation provides a means of using observed concentration data to identify sorption rates that vary between discrete regions. Several extensions of this work are considered: 1) The immobile zone in any one region may be more complex than a single pair of rates and estimation of more complex sorption processes can be examined. 2) The spatial variation of rates may be continuous and parameters describing this continuous variation, e.g., spatial covariance, may be amenable to estimation with these approaches. 3) Example calculations shown here were conceptualized as estimation from tracer test results, but the real-time assimilation capability demonstrated herein may open possibilities of using long-term monitoring networks to continuously update aquifer properties from observations of natural variations in water chemistry.

## References

- [1] R. Andricevic and E. Foufoula-Georgiou. Modeling kinetic non-equilibrium using the first two moments of the residence time distribution. *Stochastic Hydrology and Hydraulics*, 5:155–171, 1991.
- [2] S. Assouline. Estimation of lake hydrologic budget terms using the simultaneous solution of water, heat and salt balances and a Kalman filtering approach: Application to Lake Kinneret. *Water Resources Research*, 29(9):3041–3048, 1993.
- [3] M. L. Brusseau, T. Larsen, and T. H. Christensen. Rate-limited sorption and nonequilibrium transport of organic chemicals in low organic carbon aquifer materials. *Water Resources Research*, 27(6):1137–1145, 1991.
- [4] W. Chen and R. J. Wagenet. Description of atrazine transport in soil with heterogeneous non-equilibrium sorption. *Soil Science Society of America Journal*, 61:360–371, 1997.
- [5] J. A. Cunningham and P. V. Roberts. Use of temporal moments to investigate the effects of non-uniform grain-size distribution on the transport of sorbing solutes. *Water Resources Research*, 34(6):1415–1425, 1998.
- [6] G. E. Di Donato, E.-O. Obi, and M. J. Blunt. Anomalous transport in heterogeneous media demonstrated by streamline transport. *Geophysical Research Letters*, 30(12):1608–1611, 2003.
- [7] G. Evensen. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res* 99(C5), 10, 143–162, 1994.
- [8] G. Evensen. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.
- [9] M. N. Goltz and P. V. Roberts. Interpreting organic solute transport data from a field experiment using physical nonequilibrium models. *Journal of Contaminant Hydrology*, 1:77–93, 1986.
- [10] A. Gret, R. Snieder, R. C. Aster, and P. R. Kyle. Monitoring rapid temporal change in a volcano with coda wave interferometry. *Geophysical Research Letters*, 32(L06304):doi:10.1029/2004GL021143, 2005.
- [11] R. Haggerty, S. A. McKenna, and L. C. Meigs. On the late time behavior of tracer test breakthrough curves. *Water Resources Research*, 36(12):3467–3479, 2000.
- [12] R. Haggerty, C. F. Harvey, C. F. v. Schwerin, and L. C. Meigs. What controls the apparent timescale of solute mass transfer in aquifers and soils? a comparison of diverse experimental results. *Water Resources Research*, 40(W01510):doi:10.1029/2002WR001716, 2004.

- [13] U. Jaekel, A. Georgescu, and H. Vereecken. Asymptotic analysis of nonlinear equilibrium solute transport in porous media. *Water Resources Research*, 32(10):3093–3098, 1996.
- [14] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 1960.
- [15] E. M. LaBolle, G. E. Fogg and A. F. B. Tompson. Random-walk simulation of transport in heterogenous porous media: local mass-conservation problem and implementation. *Water Resources Research*, 32(3):583–593, 1996.
- [16] C. T. Miller and W. J. Weber. Sorption of hydrophobic organic pollutants in saturated soil systems. *Journal of Contaminant Hydrology*, 1:243–261, 1986.
- [17] A. K. Mishra and A. Gutjahr. Transverse dispersion of a kinetically sorbing solute. *Mathematical Geology*, 31(7):771–791, 1999.
- [18] A. K. Mishra, A. Gutjahr, and H. Rajaram. Transport with spatially variable kinetic sorption: recursion formulation. *Advances in Water Resources*, 22(5):549–555, 1999.
- [19] J. A. Pedit and C. T. Miller. Heterogeneous sorption processes in subsurface systems, 1. model formulations and applications. *Environmental Science and Technology*, 28(12):2094–2104, 1994.
- [20] H. A. M. Quinodoz and A. J. Valocchi. Stochastic analysis of the transport of kinetically sorbing solutes in aquifers with randomly heterogeneous hydraulic conductivity. *Water Resources Research*, 29(9):3227–3240, 1993.
- [21] A. F. B. Tompson. Numerical simulation of chemical migration in physically and chemically heterogeneous porous media. *Water Resources Research*, 29(11):3709–3726, 1993.
- [22] A. J. Valocchi and H. A. M. Quindooz. Application of the random walk method to simulate the transport of kinetically sorbing solutes. In L. M. Abriola, editor, *Groundwater Contamination: Proceedings of the Third Scientific Assembly of the IAHS*, Baltimore, MD, 1989.
- [23] M. T. van Genuchten and P. J. Wierenga. Mass transfer studies in sorbing porous media: I analytical solutions. *Soil Science Society of America Journal*, 40:473–480, 1976.
- [24] G. Welch and G. Bishop. An introduction to the Kalman filter. Technical report, Department of Computer Science, University of North Carolina, 2003. Technical Report No. TR 95-041.
- [25] J. Xiang and T.-C. J. Yeh. Numerical simulation of river numerical simulation of river stage tomography:poster H41E-0463. 2005 Fall Meeting, American Geophysical Union, 2005.

## DISTRIBUTION:

1 Sanjay Srinivasan  
The University of Texas at Austin  
Dept. of Petroleum and Geosystems Engineering  
1 University Station C0300  
Austin, TX 78712-0228

1 MS 0735 Susan Altman, 6313  
1 MS 0735 Ray Finley, 6313  
1 MS 0701 Sid Gutierrez, 6700  
1 MS 0735 Sean McKenna, 6313  
1 MS 1395 Dave Kessel, 6710  
1 MS 1395 Moo Lee, 6711  
1 MS 0748 David G. Robinson, 6761  
1 MS 0370 Timothy Trucano, 1411  
1 MS 1318 Bart van Bloemen Waanders, 1411  
1 MS 1395 Eric Vugrin, 6711  
1 MS 1395 Kay Vugrin, 6711  
2 MS 9018 Central Technical Files, 8944  
2 MS 0899 Technical Library, 9536  
1 MS 0123 D. Chavez, LDRD Office, 1011