SANDIA REPORT

SAND2011-6811
Unlimited Release
Printed October 2011

Bayesian Data Assimilation for Stochastic Multiscale Models of Transport in Porous Media

J. Ray, S. Lefantzi, K. Klise, L. Salazar, S. A. McKenna, B. van Bloemen Waanders, M. D. Parno, Y. M. Marzouk

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

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

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

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
Online ordering:	http://www.doe.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
Online ordering:	http://www.ntis.gov/ordering.htm

SAND2011-6811
Unlimited Release
Printed October 2011

Bayesian Data Assimilation for Stochastic Multiscale Models of Transport in Porous Media

J. Ray, S. Lefantzi, Sandia National Laboratories,
P. O. Box 969, Livermore CA 94551
and
S. A. McKenna, K. Klise, L. Salazar, B. van Bloemen Waanders,
Sandia National Laboratories, P. O. Box 5800, Albuquerque NM 87185-0751
and
M. D. Parno and Y. M. Marzouk
Massachusetts Institute of Technology, Cambridge, MA 02139
{jairay,slefant,kaklise,luksala,samcken,bartv@sandia.gov}
{mparno,ymarz@mit.edu}

Abstract

We investigate Bayesian techniques that can be used to reconstruct field variables from partial observations. In particular, we target fields that exhibit spatial structures with a large spectrum of lengthscales. Contemporary methods typically describe the field on a grid and estimate structures which can be resolved by it. In contrast, we address the reconstruction of grid-resolved structures as well as estimation of statistical summaries of subgrid structures, which are smaller than the grid resolution. We perform this in two different ways (a) via a physical (phenomenological), parameterized subgrid model that summarizes the impact of the unresolved scales at the coarse level and (b) via multiscale finite elements, where specially designed prolongation and restriction operators establish the interscale link between the same problem defined on a coarse and fine mesh. The estimation problem is posed as a Bayesian inverse problem. Dimensionality reduction is performed by projecting the field to be inferred on a suitable orthogonal basis set, viz. the Karhunen-Loève expansion of a multiGaussian. We first demonstrate our techniques on the reconstruction of a binary medium consisting of a matrix with embedded inclusions, which are too small to be grid-resolved. The reconstruction is performed using an adaptive Markov chain Monte Carlo method. We find that the posterior distributions of the inferred parameters are approximately Gaussian. We exploit this finding to reconstruct a permeability field with long, but narrow embedded fractures (which are too fine to be grid-resolved) using scalable ensemble Kalman filters; this also allows us to address larger grids. Ensemble Kalman filtering is then used to estimate the values of hydraulic conductivity and specific yield in a model of the High Plains Aquifer in Kansas. Strong conditioning of the spatial structure of the parameters and the non-linear aspects of the water table aquifer create difficulty for the ensemble Kalman filter. We conclude with a demonstration of the use of multiscale stochastic finite elements to reconstruct permeability fields. This method, though computationally intensive, is general and can be used for multiscale inference in cases where a subgrid model cannot be constructed.

Acknowledgment

This work was supported Sandia National Laboratories' LDRD (Laboratory Directed Research and Development) funds, sponsored by the Computational and Information Sciences Investment Area. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Contents

List of Figures

List of Tables

This page intentionally left blank

Chapter 1

Introduction

The estimation/reconstruction of field variables from sparse observations has long been of practical interest. The field of geostatistics (e.g., [2]) seeks to reconstruct such fields, often in 3D, from sparse observations of the same, whereas seismic inversions seek to estimate rock strata from acoustic reflections.The fields being inferred are described on a grid and spatial structures which are well resolved on the grid (and for which information is available in the observations) are estimated. Often, the range of spatial scales that are inferred is not very large.

Observations hold the key to the type of scales that can be resolved. These observations may be of the field being reconstructed (but at a sparse set of points); we will refer to these as "static" data, since for the purposes of this study, we will only infer static fields. However, one may also observe dynamical processes that are dependent on the field of interest; we will refer to these as "dynamic" data. In many cases, the dynamic data contain information about a separate set of scales, vis-à-vis static data. In such cases, it is advantageous to combine the two types of observations and expand the range of lengthscales that can be inferred. This leads to *multiscale* reconstruction/inversion/estimation of fields. Multiscale inversions tend to be both mathematically complex and computationally expensive; at the very least, they require a parametrized, *low dimensional random field model* for the field being inferred, a *dynamic forward model* that connects the field to the observed dynamics and an *"inverse solver"* to estimate optimal parameters so that observations are reproduced with minimal errors.

Zonation methods [3, 4, 5, 6] are a good example of how multiscale permeability (or hydraulic conductivity) fields may be inferred. One discretizes a domain with a very coarse grid (often just a single grid-block) and infers an effective permeability from dynamic data (generally pressure measurements or tracer breakthrough times from a pump test[1]). The grid is then locally refined (by dividing into two) and the field re-estimated. This refinement proceeds locally, creating a grid hierarchy, till further refinement does not lead to a

[1]A test where a fluid, generally water, is injected into a domain from one end and produced/extracted/pumped out from the other

better fit with data. The fitting of the field to data is performed using an optimization algorithm and the net result is a *single, best-fit* field, with spatial structures, reconstructed on meshes with variable resolution. Unfortunately, the use of a deterministic optimization method does not quantify the reconstruction uncertainty. The uncertainty is due to both the sparsity of observations and the shortcomings of the forward model that connects the dynamic observations (pressure, breakthrough times) to the permeability field.

A rather different approach, that addresses uncertainty in the reconstructions, involves multiscale finite elements [7, 8, 9]. The approach starts with a scale separation imposed by a coarse grid. The field estimation (in this case, a permeability field too) is performed using production history as the data and a Darcy flow model. A Markov chain Monte Carlo (MCMC) method is used to investigate coarse permeability fields' fit to data; promising configurations are re-tested by projecting onto a fine mesh, recomputing the dynamic predictions and comparing with observations. The fine-coarse dichotomy allows efficient exploration of permeability fields with MCMC and the coarse model while the fine-mesh computations enforce accuracy. The reconstruction is performed as a set of accepted (or possible) permeability field configurations which are conditioned on the observations; these are deemed to be samples from a *posterior distribution* of permeability fields. Such a probabilistic reconstruction method allows simple computations of the uncertainty in the estimation.

These contemporary approaches are limited to estimations (a.k.a. inversions) on a grid; no multiscale method addresses the estimation of *subgrid* structures which cannot be resolved on a mesh. Neither do they target large problems (big grids) when an accurate quantification of the uncertainty in the inversion is required; the inherent sequential nature of MCMC has largely been responsible for this lack of scalability. Consequently, multiscale inversions have yet to target real-life problems.

In this study, we will investigate multiscale inversion *where subgrid structures are reconstructed* from sparse observations, and the algorithms are scaled to a real-life problem. The key paradigm underlying all our work is Bayesian inversion (or Bayesian inverse problems), with a focus on quantifying the uncertainty in our estimations. Modeling innovations lie in deriving models that link across scales, as well as extensions of scalable algorithms to be applicable to multiscale inversion problems. Multiscale inversion, including estimation of summaries of subgrid structures, raises two issues:

1. *Cross-scale linking:* Since we will reconstruct subgrid structures, an additional model (called the *link function* or *subgrid model*) will be required. The subgrid model will summarize the impact of unresolved scales at the *macroscale* i.e., the scale which is resolved by the grid. We will demonstrate two ways of doing so (1) using a subgrid model based on the physics of the problem and (2) a competing approach, where upscaling is performed by multiscale finite elements (MsFEM). The MsFEM method does not require an explicit subgrid model and consequently is the only possible approach when a subgrid model cannot be created. In both cases, the models will contain parameters which are reflective of subgrid lengthscales; the estimation of

these parameters establishes a summary for the unresolved structures.

2. *Scalability:* The use of MCMC has limited the application of multiscale methods to small problems, when uncertainty quantification (UQ) of the inferred field is desired. Ensemble Kalman Filters (EnKF) are a scalable means of performing inversions but suffer from two restrictions: (a) they assume that the field being reconstructed is a multiGaussian and (b) the data being used for reconstruction are time-variant. In this study, we will first develop methods to identify if reconstructed fields can be safely assumed to be multiGaussians and thereafter develop EnKF extensions to allow the assimilation of both dynamic and static data in the EnKF method. Scalability will be demonstrated in two ways: (1) by demonstrating an inversion on a large grid and (2) performing a permeability field inversion for the Ogallala - High Plains aquifer in Kansas. The later will investigate the ability of our methods to address the complexities of real-life inversion problems.

This investigation is structured as follows. We begin with the estimation of a binary field i.e., a field where a high permeability material is embedded in a low permeability one. The inclusions are too small to be resolved by a grid; further, their proportion varies in space. Observations include permeability measurements at a handful of sites in a 2D domain, as well as tracer breakthrough times. We reconstruct the proportionality distribution as well as obtain an estimate of the (subgrid) inclusions' size. In Chapter 2, we develop the subgrid model, from physical arguments; in Chapter 3, we formulate and solve a Bayesian inverse problem where the subgrid model is used to reconstruct realizations of the binary field conditioned on the observations. The reconstructions will use MCMC as the "inverse solver", and we will investigate how different the posterior distributions are from a multiGaussian. In Chapter 4, we describe a very different way of upscaling, using multiscale finite elements (MsFEM) to provide the cross-scale link. In Chapter 5 we address scalability, developing a modified EnKF to assimilate both static and dynamic data to reconstruct a fractured medium (a field where the width of high-permeability fractures are much smaller than a grid-block, but the length may be as large as the domain of interest). Finally, in Chapter 6, we apply the EnKF to solve a real-life reconstruction problem (the Ogallala - High Plains aquifer).

The individual sections are self-contained. Each contains its own literature review and the inverse problem is rederived in each case (the derivations have little in common, except a Bayesian approach with a focus on the use of priors to assimilate exogenous information). We proceed up the length, problem size and complexity scales in successive sections, carrying over lessons that can be used for simplifications at a larger/coarser scale. All estimations are performed with sparse observations i.e., our estimates have a significant degree of uncertainty in them. However, the probabilistic (Bayesian) approach is adopted throughout and each step we provide an uncertainty bound on the quantities that we infer.

This page intentionally left blank

Chapter 2

An Analytical Subgrid Model for Porous Binary Random Media

Determination of a single effective property value from an assemblage of materials is a long standing research problem in a number of scientific and engineering fields. Here we focus on the development of an effective conductivity value from a mixture of two materials (binary media) with distinct conductances. A simple conceptualization of the binary medium as inclusions of a high/low conductivity material within a continuous matrix of material having the opposite conductivity serves for discussion here.

Effective properties of materials composed of mixtures of two component materials have been the subject of study for heat conduction, electrical conductivity, magnetic permeability, and electrical permittivity [10, 11, 12, 13, 14, 15, 16, 17, 18]. Hashin and Shtrikman [17] demonstrate the mathematically analogous nature of calculations for effective values of the conductance terms in these varied fields. An extensive amount of work for binary media has focused on defining the theoretical bounding values for the effective properties of the medium [13, 17, 19, 20].

The same approaches to effective medium equations hold for calculation of effective permeability, or hydraulic conductivity, in steady-state flow through porous media. Binary models of conductivity are widely applied in subsurface flow through porous media particularly for representation of permeability patterns in fluvial deposits (e.g., [15, 21, 22, 23, 24]). Additionally, fractured media are often characterized using a binary permeability model where the fractures represent strongly anisotropic, high conductivity inclusions embedded within a matrix of low conductance. Fractured media can be represented as linear or planar conductive elements within a less conductive background using discrete fracture representations [25, 26] or as representation of fractured zones within continuum models [27, 28, 29].

Previous work on binary fields in the context of flow through porous media has emphasized development of expressions for the effective conductance of the field as a function of

the proportion of the high/low, conductivity phase. Effective medium theories (EMT) for binary assemblages have focused on using the conductivities and proportion of the two materials to determine an effective conductivity value [14, 18, 30, 31]. Initial development of these theories used spherical inclusions and more recent work has incorporated additional information on the shapes of the inclusions [10, 11, 32, 33, 34, 35].

Thorough reviews of variations of the EMT-based approaches with comparison to other methods can be found in: [19] and [36]. EMT-based approaches assume non-interaction between inclusions and therefore do not utilize information on the connectivity or interaction of either phase with other inclusions of the same phase. In testing against numerical results, EMT-based approaches work best when the inclusion fraction is less than 50 percent (see [15, 19]).

The effective conductivity formula developed by [36] incorporates information on connectivity of high conductivity inclusions through the average path length within the low conductivity material. This novel approach motivates exploration of various measures of connectivity and different techniques for estimating the mean path length between inclusions. A model for these mean distances tied to the geometry of binary media resulting from truncation of multiGaussian (mG) fields is proposed herein and the behavior of this model is compared to previously developed expressions for effective media and numerical results. We limit comparisons to other approaches developed for binary media that use information on the modal conductances, proportions and geometry of the phases to calculate effective conductivity. Other techniques that require full knowledge of a fine scale field in order to complete the upscaling such as renormalization [37], anisotropic effective medium approximation [38] and wavelet coarsening [39] are not considered here.

A number of numerical techniques are available for simulation of binary random fields. Indicator geostatistical techniques [40] with spatial variation defined through a variogram provide an efficient means of generating stochastic realizations of binary fields [41, 42, 43]. Alternatively, indicator simulation approaches can be based on transition probabilities between indicator classes [44, 45]. Typical applications of geostatistical simulation techniques are focused on generation of fields with more than two classes, multiple indicator simulation, but they can also be used for the generation of binary fields. Less common approaches for generating spatial binary fields include object-based and Boolean models [46], generation of periodic media [16, 47] and pluriGaussian and truncated mG fields [46, 48].

Development of excursion set theory applied to truncated mG fields over the past 15 years has been driven by developments in medical imaging and astrophysics [49, 50, 51]. In particular, calculation of the expected values of the total excursion area, number of distinct excursions and the average excursion size over a threshold value can be calculated from definition of the mG field and knowledge of the threshold value ([52, 53]). Excursion set theory is applicable to truncation with a single threshold or multiple thresholds that produce multiphase fields (e.g.,[54]). Phillips and Wilson [55] proposed mean threshold crossing distances to estimate correlation lengths of permeability. However, in contrast to the wide application of mG random fields in hydrogeology, use of excursion sets from truncated fields for characterization and modeling of heterogeneous media in groundwater studies

has been limited.

We parametrize a form of distance-based upscaling using point-process theory and properties of truncated Gaussian fields to develop an expression for the effective conductance of binary media. This new expression, truncated Gaussian distance-based upscaling (TG-DBU) differs significantly from the existing distance-based upscaling in that calculation of the effective conductivity does not require instantiation of the binary field. This aspect provides a distance-based calculation of the effective conductivity that is efficient enough for iterative parameter estimation. Section 2.1 summarizes distance-based upscaling and the salient aspects of point-process theory and truncated mG fields. Section 2.2 combines these three elements into an expression for effective conductance for isotropic inclusions within a background matrix. In Section 2.3, this new expression is compared to distance-based upscaling using full knowledge of the binary field as well as an existing analytical solution and numerical solutions. Section 2.4 compares various distance measures and examines the behavior of the average distance between inclusions in the neighborhood of the percolation threshold.

2.1 Estimation of Effective Conductivity

Development of the TG-DBU procedure is motivated by the goal of estimating the effective conductance of a binary medium created from thresholding a Gaussian random field without instantiation of that field. The two modal permeabilities, K_1 and K_2, are considered known. Given the threshold at which the field is truncated, and size of the Gaussian kernel used to create the field as defined by the full-width at half-maximum (FWHM) parameter, the effective permeability is estimated.

2.1.1 Distance-Based Upscaling

The distance-based upscaling (DBU) approach developed by [36] utilizes an estimate of the mean flowpath length between inclusions within the background (matrix) material as a measure of phase connectivity. Through application of a phase-change theorem, the DBU approach applies to high or low conductivity inclusions within a matrix of the opposing material. The DBU approach [36] serves as the foundation for our estimation approach and is briefly outlined here.

The basis of DBU is conceptualization of each inclusion as a rectangular object of dimensions (B_x,B_y) centered within a larger rectangular block, having the same orientation, of dimensions (L_x,L_y). Fixed pressure boundary conditions on each end of the block and no-flow boundary conditions on the opposite sides create steady, one-dimensional flow along the x-direction. Knudby et al [36] identified an approximate linear relationship between the inverse of the effective conductivity of the block $(K_B)^{-1}$ and the relative shape of the

inclusion $(B_x/L_x)(B_y/L_y)$ and used this relationship along with harmonic and arithmetic conductivity bounds to develop an expression for $(K_B)^{-1}$:

$$\frac{1}{K_B} = \left(\frac{1}{K_A} - \frac{1}{K_H}\right)\frac{R - p_1}{\frac{1}{p_1} - p_1} + \frac{1}{K_H} \tag{2.1}$$

where K_A and K_H are the arithmetic and harmonic mean conductivities, respectively, p_1 is the proportion of high permeability material and:

$$R = \frac{B_x/L_x}{B_y/L_y} = \frac{1}{p_1}\left(\frac{B_x}{L_x}\right)^2 \in \left[p_1, \frac{1}{p_1}\right] \tag{2.2}$$

The expression for $(K_B)^{-1}$ can also be cast as a weighted mean of K_A and K_H.

$$\frac{1}{K_B} = \rho\frac{1}{K_A} + (1-\rho)\frac{1}{K_H} \tag{2.3}$$

where ρ is the relative inclusion shape, R, normalized by p_1:

$$\rho = \frac{R - p_1}{\frac{1}{p_1} - p_1} \in [0,1] \tag{2.4}$$

The distance-based component of the DBU method enters as a normalized average distance, D_{norm}, of the flow in the background medium within the block:

$$D_{norm} = \frac{B_x - L_x}{L_x} \tag{2.5}$$

The normalized inclusion shape, ρ, can be restated using D_{norm}:

$$\rho = \frac{2D_{norm} - D_{norm}^2}{1 - p_1^2} \tag{2.6}$$

Expansion of these relationships from a single inclusion within a single block to a field of inclusions requires calculation of average values across the field for the inclusion dimensions, $\overline{B}_x$, $\overline{B}_y$ and block dimensions, $\overline{L}_x$, $\overline{L}_y$. A key element of this development is determination of the average distances between inclusions along the direction of flow; $\overline{D} = \overline{L} - \overline{B}$. The block domain is conceptualized as a virtual permeameter centered on each inclusion within the field. $\overline{R}$ then represents the average relative inclusion shape and $\overline{D}_{norm}$ is the normalized average distance between inclusions along the direction of flow weighted by the area of each connecting inclusion.

In the DBU approach, the average distance, $\overline{D}$, is calculated as a weighted average using distances, D, and inclusion areas A as measured directly on the binary field:

$$\overline{D} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} D_{i,j}A_iA_j}{\sum_{j=1}^{n}\sum_{i=1}^{n} A_iA_j} \tag{2.7}$$

with the average block dimension in the direction of flow, x, calculated as:

$$\overline{B} = \frac{\sum_{j=1}^{n} B_{x,i}A_i}{\sum_{i=1}^{n} A_i} \tag{2.8}$$

The normalized average distance between inclusions is: $\overline{D}_{norm} = \overline{D}/\overline{L}$. Use of these spatial averages renders the block conductivity estimate, K_B, as an effective conductivity, K_{eff} for the domain.

A strong advantage of the DBU is the incorporation of the phase interchange theorem [12]. This theorem provides a relationship between K_{eff} of a field with low conductivity inclusions in a high conductivity matrix (low-in-high, $LinH$) and the generally easier-to-estimate K_{eff} of a complementary field of high conductivity inclusions in a low conductivity matrix (high-in-low, $HinL$). The fixed head and no-flow boundaries are rotated 90 degrees and applied to the complementary field. The fluxes, Q, through the two fields are related by:

$$Q_{LinH}Q_{HinL} = K_1K_2(\Delta H)^2 \tag{2.9}$$

where ΔH is the pressure drop across both fields from the prescribed boundary conditions. The product of the effective conductivities for each field is equal to the product of the two modal conductivities in the binary field:

$$K_{eff(HinL)}K_{eff(LinH)} = K_1K_2 \tag{2.10}$$

A key advantage of the phase interchange theorem is that it enables the calculation of the $LinH$ case for any geometry for which the $HinL$ solution is available.

Knudby et al [36] demonstrate accurate estimation of K_{eff} for a range of simulated fields created with Poisson placement of ellipses or rectangles as well as those created with transition probability-based geostatistical simulation. Within these fields, the ratios of the two conductivities range from 100 to $10,000$. The DBU results are also compared with several other effective value approaches.

2.1.2 Excursion Sets and Kernel Size

Calculation of K_{eff} with the DBU method requires both creation of the binary field and calculation of all inclusion sizes and distances between proximal inclusions. Image processing algorithms are available for these calculations; however, the computational expense of these algorithms is non-trivial. Here, we develop an approach for estimation of K_{eff} based on

DBU that estimates the average inter-inclusion distance without explicit creation or processing of the binary field. This new approach relies on properties of truncated Gaussian fields to estimate the inclusion sizes and the mean distance between them and provides an analytical expression for K_{eff}.

The model for spatially correlated multiGaussian (mG) fields is based on a Gaussian kernel:

$$G(x,y) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} d\Sigma^{-1} d^T\right) \tag{2.11}$$

where d is the distance vector containing distances d_x and d_y from any location (x,y) to the origin of the Gaussian function x_0, y_0 (here (0,0) for the standard normal distribution). In this work, the covariance matrix, $\Sigma = \sigma^2 I$, (where I is the identity matrix) is diagonal for the specific case of the kernel being aligned with the grid axes. Convolution of an uncorrelated mG field with a Gaussian kernel creates a realization of a spatially correlated random field. A discretized uncorrelated mG field (e.g., as described on a mesh) can be created by simply sampling values at the mesh points i.i.d. from a standard normal.

The spatial correlation of the mG field is defined by the FWHM of the Gaussian kernel used to create the Gaussian field. The FWHM parameter is commonly used as a spatial measure in image processing:

$$FWHM = \sigma\sqrt{8\ln(2)} \tag{2.12}$$

where σ is the standard deviation of the Gaussian kernel. Truncation of a Gaussian field at a threshold u defines the $u - level$ excursion set:

$$X_u = \left\{x \in R^D : Y(x) \geq u\right\} \tag{2.13}$$

and the variogram, of the random set X_u can be calculated:

$$\gamma_u(h) = \frac{1}{\pi} \int_0^{\arcsin(\sqrt{\gamma(h)/2})} \exp\left(-\frac{u^2}{2}(1+\tan^2(t))\right) dt \tag{2.14}$$

at lag spacings h. Variogram models that are linear at the origin (e.g., exponential, hyperbolic) cause the perimeter of X_u to be infinite (see [46], Section 16.1) and we restrict our work here to Gaussian kernel functions.

Three related properties of the truncated Gaussian field (following [52]) are:

- N, the number of pixels above the truncation threshold, u,
- m, the number of distinct regions (inclusions) above the threshold, and
- n, the number of pixels in each region,

with expectation relationship $E[N] = E[m]E[n]$. For threshold value, u, the number of cells

above that threshold, N, is provided by the Gaussian cdf and the size of the domain, S:

$$E[N] = S\int_u^\infty (2\pi)^{-1/2} e^{-z^2/2} dz \tag{2.15}$$

The Euler Characteristic, EC, in $D = 2$ represents the number of connected objects in the domain minus the total number of holes within those objects. Therefore EC goes to 0.0 at $u = 0$ and EC becomes negative when $u < 0.0$ as the truncated field represents a single domain-spanning object containing a large number of holes. In 2D, the absolute value of EC is the number of distinct inclusions of either phase within the opposite phase and is used here to determine $E[m]$.

$$E[m] = |EC| = |(2\pi)^{-(D+1)/2} W^{-D} u^{D-1} e^{u^2/2}| \tag{2.16}$$

where W is an alternative measure of the spatial correlation of the mG field defined as a fraction of the FWHM:

$$W = FWHM/\sqrt{4\ln(2)} \tag{2.17}$$

For a given threshold, u, the average object area is found from the expectation relationship:

$$E[n] = E[N]/E[m] = E[N]/|EC| \tag{2.18}$$

Figure 2.1 compares a direct calculation of EC using the Matlab Image Processing toolbox [56] with estimates made using Equation 2.16 across a range of u values increasing from left to right. The corresponding binary fields (500 x 500 cells) are also shown for several representative threshold values.

2.2 Reduced Model Estimation

The properties of the truncated Gaussian field are used with the DBU method to develop an approximation for K_{eff} of a binary field. These estimations are done as a function of the proportion $[0,1]$ of the high permeability phase (p_1) as defined by the threshold, u. A critical component of the DBU approach is $\overline{D}_{norm}$. We employ a combination of spatial point process theory and use of FWHM as a characteristic distance of the truncated field to estimate $\overline{D}_{norm}$ and refer to this approach as TG-DBU. The development here is for isotropic fields.

At u values near $-\infty$ or $+\infty$, the distances between centroids of inclusions are approximated as the distribution of nearest neighbor distances, d, from a Poisson point process (e.g., [57]) with an intensity $\lambda = |EC|/S$:

$$F(d) = 1.0 - \exp(-\pi\lambda d^2) : d \geq 0. \tag{2.19}$$

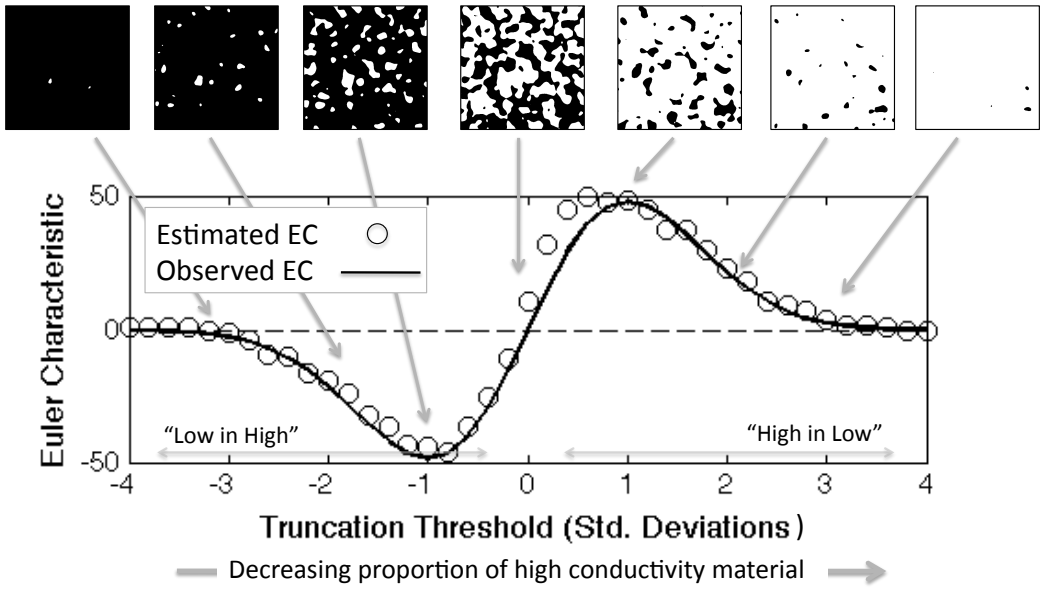


Figure 2.1: Observed (calculated) and estimated Euler characteristic for a truncated mG field as a function of u. The corresponding binary fields are shown for select thresholds. Regions of high conductivity are colored black.

The estimated average distance between inclusion centroids, $\overline{D}^*$ is:

$$\overline{D}^* = \sqrt{\frac{S}{|EC|}\frac{1}{\pi}} \tag{2.20}$$

This approximation only holds at the extreme values of u (see [52]) as the distances between inclusion centroids overestimate the distance between inclusion edges as the average inclusion size, approximated as $E[n]$, increases. The value of $\overline{D}^*$ is adjusted to account for the inclusions having non-zero area by subtracting twice the average object radius $\overline{D}^* = \overline{D}^* - 2\sqrt{E[n]/\pi}$.

The DBU calculates distances between objects in the downstream direction only. The nearest neighbor distance calculation is adjusted to account for this preferential search direction through incorporation of a half angle, θ, that constrains the search for objects in the +/- 90 degree directions at u = 0.0 and with θ decreasing as u moves to the extreme values:

$$\theta = -2\pi(p_1 - 0.5)^2 + \pi/2 \tag{2.21}$$

This expression defines an exponential distribution for the variable θx^2 (after [57], page 34).

The geometry and connectivity of the binary patterns in truncated Gaussian fields vary considerably as u increases from $-\infty$ to $+\infty$ (Figure 2.1). Near the extreme values of u,

the field is composed of independent high/low conductivity inclusions in a matrix of the opposite material. As u moves towards 0.0, the inclusions begin to coalesce forming larger inclusions with shapes that are roughly approximated by overlapping circles. At u values even closer to 0.0, the inclusions begin to span the domain and at $u = 0.0$, there is no distinction between what is background and what is inclusion.

Conceptually, for the case of isotropic Gaussian fields the calculated value of $\overline{D}$ will never go to zero. As u moves towards 0.0 from either extreme, the $\overline{D}$ calculation changes from that of distances between isolated independent inclusions to distances between a few isolated inclusions and a main inclusion composed of several inclusions that were isolated at lower u values and finally to distances from one portion of a domain spanning inclusion across holes to another portion of that same inclusion.

At $u = 0.0$ ($p_1 = 0.50$) the average flow distance within the low permeability background should be equal to the FWHM distance. This assertion is due to the FWHM being the expected size of both the inclusions and the background matrix at this threshold.

This conceptualization provides the final piece of the effective conductivity approximation. For a given value of u, or the corresponding value of p_1, $\overline{D}$ is estimated as the maximum of the average distance between inclusion edges and the $FWHM$:

$$\overline{D}^* = \max\left[\sqrt{\frac{S}{|EC|}\frac{1}{\pi}} - 2\sqrt{E(n)/\pi}, FWHM\right] \tag{2.22}$$

and used with Equation 2.1 to calculate K_B. This formulation is referred to as the *basic model* in the remainder of this paper. Figure 2.2 compares the results of the basic model against effective conductances calculated numerically using MODFLOW-2005, [58]. Harmonic averaging is used to calculate internodal conductances within MODFLOW. An ensemble of 30 mG fields are created on a 500×500 grid (with cells of unit size) with a convolution kernel of FWHM of 37.7 length-units (σ = 16.0). These mG fields are transformed to binary fields through truncation at thresholds uniformly spaced from p_1 = 0.04 to 0.96. Additionally, thresholds of u = -2.5 and 2.5 are used to create the minimum and maximum p_1 values for each field: 0.0062 and 0.9938. This process results in truncation of each field at 26 unique thresholds. For each of the 26 u threshold values, the average numerical result across 30 fields (780 evaluations) is shown. We define the ratio: K_1/K_2 using the log10 conductivity values: $\kappa = \log_{10}(K_1) - \log_{10}(K_2)$ and show results for κ = 2 and 4 in Figure 2.

The basic model utilizes an exceedingly simple parameterization of the mean distance between inclusions and produces relatively accurate estimates of the effective conductance. The basic model estimates are less than a factor of 2 (100 percent error) away from the numerical results for the case of $\kappa = 2$ for all values of p_1 with the most accurate results for $p_1 < 0.50$. The basic model tends to overestimate the numerical results at $p_1 > 0.50$. For the case of $\kappa = 4$, the basic model strongly overestimates the numerical results at $p_1 > 0.50$.

Several extensions to the basic model distance calculations are incorporated for the final TG-DBU model. At low values of p_1, the basic model underestimates the numerical results

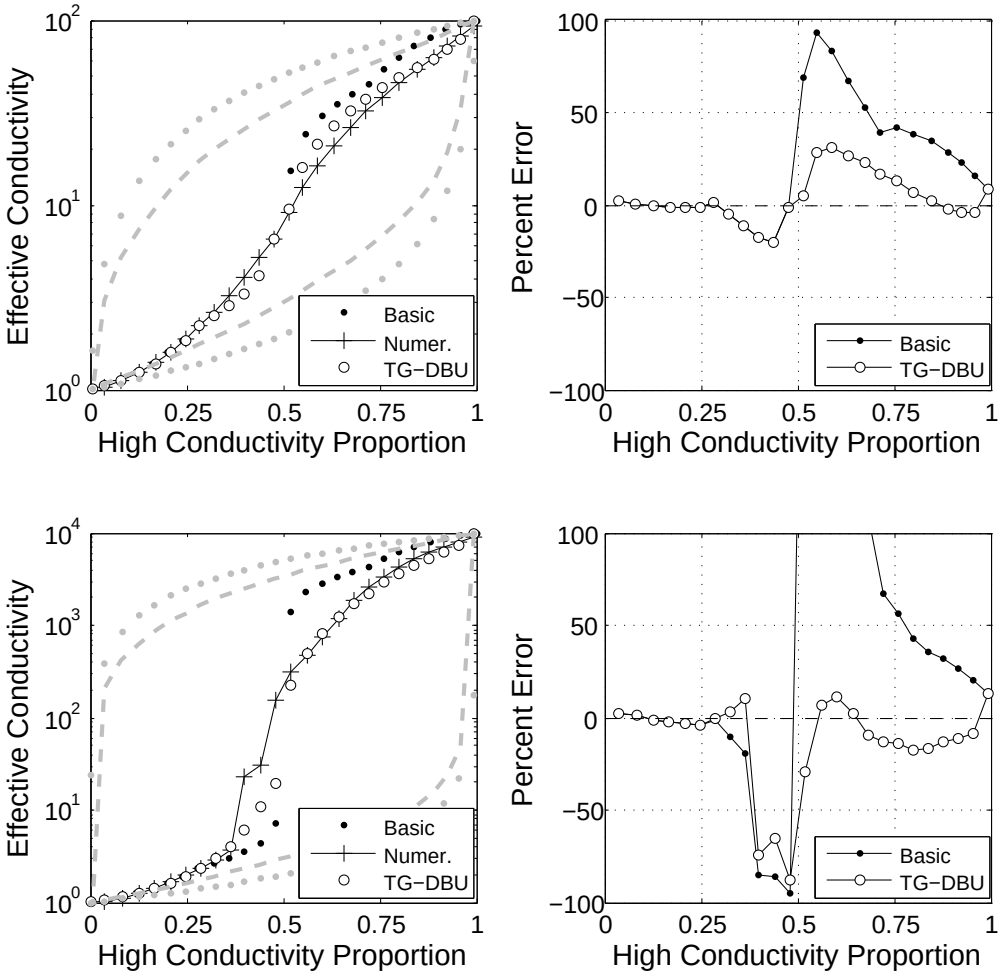


Figure 2.2: Comparison of effective conductivity values estimated with the basic model and the TG-DBU approaches to numerical results. The percent error of the effective conductivity solutions relative to the numerical results are shown in the right hand images. Results for two and four orders of magnitude difference in the modal conductivities are shown in the top and bottom rows, respectively. Results are for Gaussian fields created with an FWHM of 37.7 (σ = 16.0) length-units. The gray dots and dashed lines indicate the limiting values of arithmetic and harmonic averages and the Hashin-Shtrikman bounds.

and smaller distance values are needed to minimize this error. Additionally, the degree of underestimation increases with increasing κ (Figure 2.2). At levels of p_1 above 0.50, the basic model overestimates the numerical results and, due to application of the phase change theorem at these higher proportions, the distances must also be decreased in this region. The correction here must also be a function of κ.

The extended distance calculation is:

$$\overline{D}^* = \begin{cases} \max\left[\sqrt{\frac{S}{|EC|}\frac{1}{\pi}} - 2\sqrt{E(n)/\pi}, \frac{FWHM}{(\kappa-1.0)}\right] & \text{for } p_1 < 0.50 \\ \max\left[\sqrt{\frac{S}{|EC|}\frac{1}{\pi}} - 2\sqrt{E(n)/\pi}, FWHM \times (1.0-p_1)^{\kappa-1}\right] & \text{otherwise.} \end{cases} \tag{2.23}$$

The extended distance calculations significantly improve the ability of the TG-DBU model to estimate the effective conductivity (Figure 2.2). For the case of $\kappa = 2$ the maximum error is reduced to less than 40 percent and for the $\kappa = 4$ case, the maximum error is less than 100 percent with the largest improvement occurring at $p_1 > 0.50$.

2.3 Comparison to Analytical and Numerical Results

The TG-DBU estimated effective conductivity values are compared to existing models for values of p_1 in $[0, 1.0]$ and for κ values of 2 and 4. Visual comparisons and calculations of the percent relative error between the estimated values and numerical results are examined for two inclusion sizes. Comparisons are made to a self-consistent effective medium approximation [34, 59] that also employs the phase interchange theorem. Additionally, a series of binary fields are created from truncation of mG fields and used as input to the DBU approach of [36] as well as numerical calculation of effective conductivity.

The self consistent approximation to solution for an effective medium [34], [59] uses the phase-interchange theorem to provide effective conductivity estimates across all proportions of high/low conductivity material. Equation 14 of [34] provides an analytic solution for the effective conductivity in a 2D domain with circular (isotropic) inclusions.

$$K_{eff} = (K_1 - K_2)(p_1 - 1/2) + 1/2\sqrt{(1-2p_1)^2(K_1-K_2)^2 + 4K_1K_2} \tag{2.24}$$

The DBU approach of [36] as outlined in Section 2 is applied to each binary field. The same fields are also used as input to numerical calculations done with MODFLOW-2005 ([58]). For each inclusion size, DBU and numerical results are calculated on 30 fields at each of 26 thresholds.

For any geometrical combination of two materials with separate conductances, the arithmetic and harmonic averages provide the upper and lower bounds on the resulting effective conductivity. These averages are used frequently for problems involving flow through

porous media and are also known as the Wiener bounds within statistical physics. Tighter bounds on the effective conductivity can be defined when information on the geometry of the inclusions is available. For a binary material composed of circular inclusions of one material within another, the Hashin-Shtrikman bounds [17] provide tighter limits on the effective conducitivity estimates (see also [19]) and [36]. Both sets of bounds are calculated as reference for the different effective conductivity calculations.

Results comparing the model developed here to results of effective media theory (EMT) through the self-consistent approach [34, 59], the DBU approach [36] and numerical results are shown in Figure 2.3. These results were created from fields with a FWHM of 37.7 length-units. The TG-DBU and the EMT results are calculated independently of the actual binary field and require the phase proportions, the two modal conductances and the inclusion shape as inputs. The TG-DBU also utilizes the FWHM as an input. The DBU and numerical results are dependent on the actual binary fields, and for these results, each value in Figure 2.3 represents the average conductance calculated over 30 realizations. The deviations of the DBU estimates from the numerical estimates at proportions just above 0.50 appear to be an artifact of the distance calculations done here on the truncated mG fields and are not a function of the DBU technique [36].

The right-hand side of Figure 2.3 shows the percent error of the three estimators relative to the numerical results. The axes are limited to +/- 100 percent, or a factor of +/- 2, for calculations where the modal conductivities vary by factors of 100 (top) and 10,000 (bottom). Relative to the self-consistent approach, the two distance-based upscaling techniques better capture the effective conductivity at proportions of the high conductance phase $p_1 > 0.50$.

For all three approaches, errors are highest at or near $p_1 = 0.50$. This proportion corresponds to the percolation threshold for both a square lattice and square tiles in 2D [18] and represents the change point where the high conductivity phase becomes fully connected across the domain. Percolation theory and the percolation threshold have been developed for systems with no spatial correlation and are applied here where the ratio of FWHM to domain size is small (i.e., < 0.10). For these calculations, the size of the FWHM relative to the domain size is 0.075.

The approaches examined in this study define an effective conductivity for the domain. Effective properties are meaningful in cases where the domain size is much larger than the correlation length of the random field contained within the domain. This condition is also the definition of an ergodic field and a rule of thumb is that an effective property can be assigned to a domain when the correlation length is ≤ 0.10 of the domain size. In cases where the domain is discretized into smaller cells, or blocks, and the correlation length exceeds this limit relative to the cell size, a *block* property is assigned. Additional details on effective versus block properties are provided in [19].

As the FWHM increases, development of phase connection across the domain will occur at lower proportions of that phase. Figure 2.4 shows the results of calculations for a FWHM of 73.4 length-units ($\sigma = 32.0$), or 0.15 of the domain size.

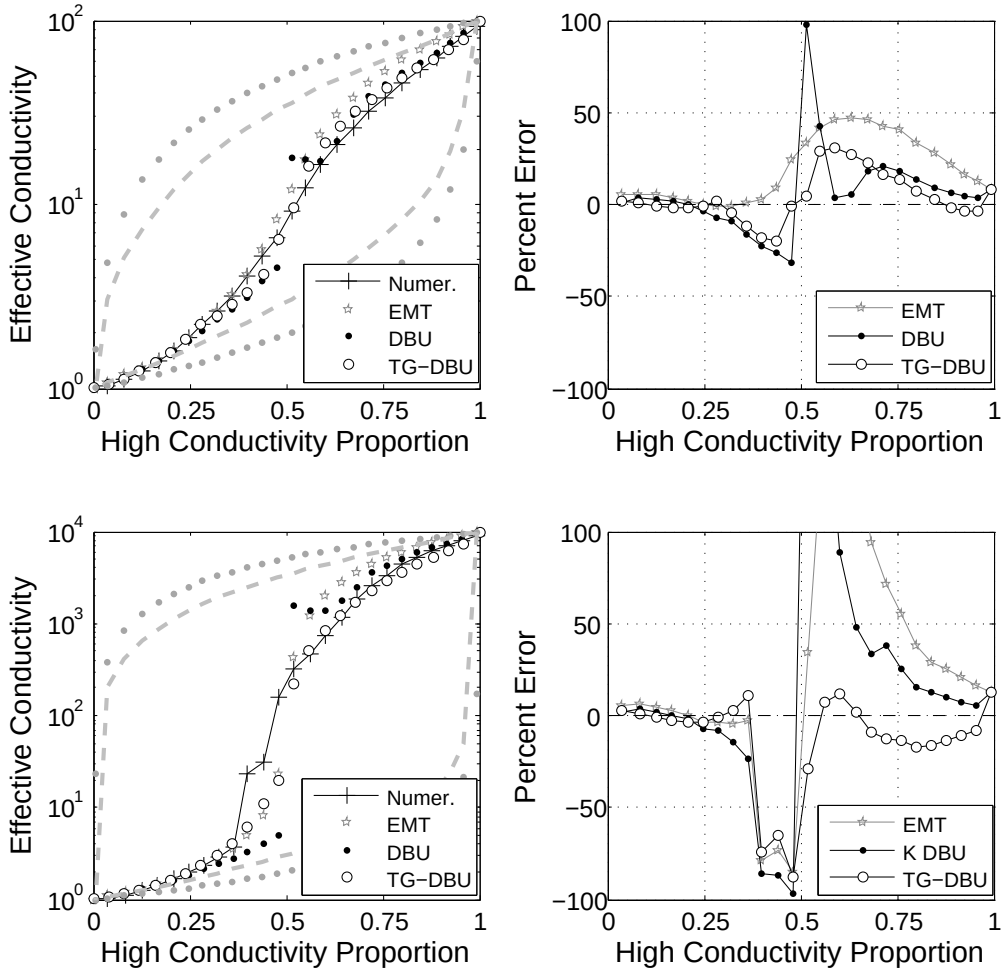


Figure 2.3: Comparison of TG-DBU results, to the EMT solution, the DBU solution and numerical results. The harmonic and arithmetic average bounds (gray dots) and the narrower Hashin-Shtrikman bounds (dashed gray lines) are also shown. The percent error of the effective conductivity solutions relative to the numerical results are shown in the right hand images. Results for two and four orders of magnitude difference in the modal conductivities are shown in the top and bottom rows, respectively. Results are for Gaussian fields created with an FWHM of 37.7 (σ = 16.0) length-units.

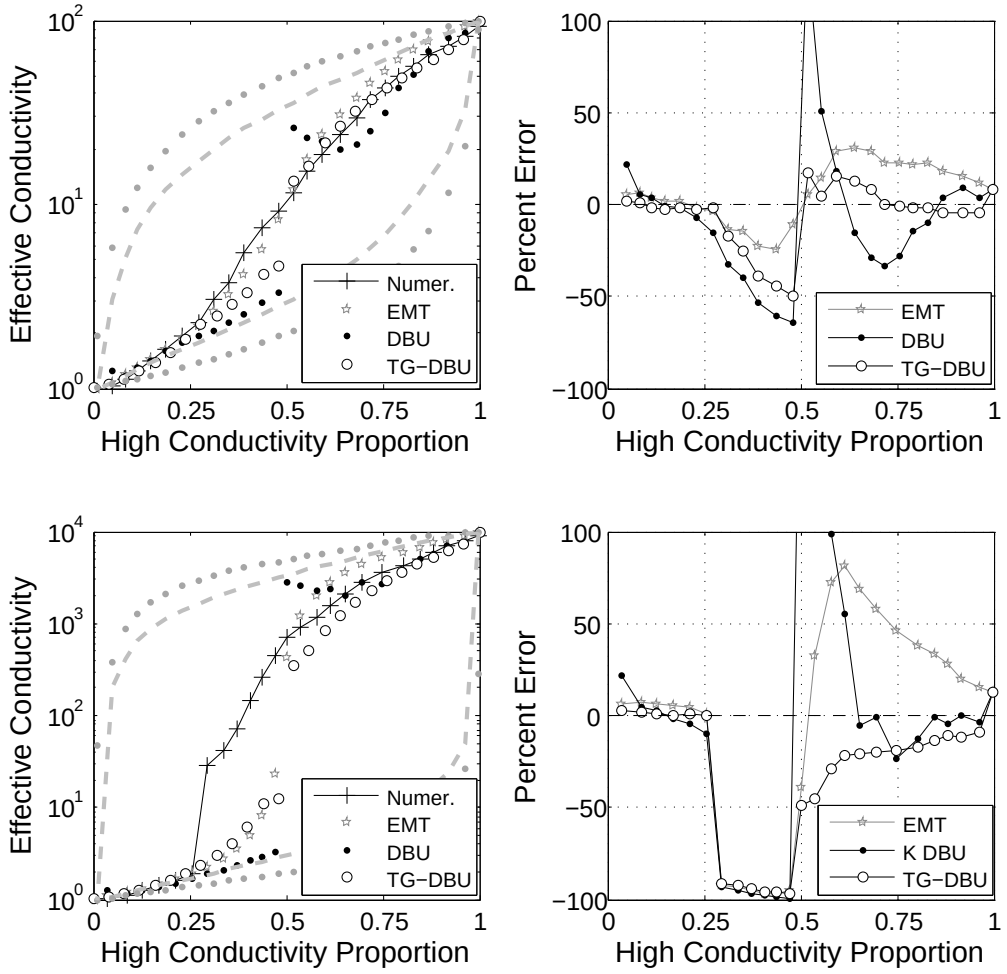


Figure 2.4: Comparison of TG-DBU results, to the EMT solution, the DBU solution and numerical results. The harmonic and arithmetic average bounds (gray dots) and the narrower Hashin-Shtrikman bounds (dashed gray lines) are also shown. The percent error of the effective conductivity solutions relative to the numerical results are shown in the right hand images. Results for two and four orders of magnitude difference in the modal conductivities are shown in the top and bottom rows, respectively. Results are for Gaussian fields created with an FWHM of 73.4 (σ = 32.0) length-units.

The increase in the relative size of the inclusions decreases the proportion of high-conductivity material necessary to create a connected phase across the domain and the effective medium techniques tend to underestimate the numerical conductivity beginning at approximately $p_1 = 0.30$. This underestimation is particularly apparent in the $\kappa = 4$ results. All techniques examined are able to create reasonable estimates of the block conductivity for $p_1 > 0.6$. These results are motivation for future work to improve block conductivity estimates by incorporating percolation threshold behavior into the TG-DBU formulation.

2.4 Results and Discussion

The value of $\overline{D}$ is a key feature of the DBU and TG-DBU approaches and final effective conductances are sensitive to these values. Distance calculations are explored further in Fig. 2.5 and compared to average distances across the background material as calculated along streamlines.

The basic model and the DBU model derive average distances from purely geometrical considerations and these values do not change as a function of the ratio of the logarithms of the permeabilities of the two components of the binary medium, κ (Fig. 2.5). The average distances along streamlines are also quite stable across the change in κ while the TG-DBU approach explicitly incorporates the κ value into the average distance calculation. With the exception of the TG-DBU model, the FWHM serves as an excellent approximation of a lower limit on the distance values calculated by the different approaches.

The average distances from the streamline values are the largest of all calculated values. Examination of streamlines in truncated binary fields shows that streamlines crossing the background material occurs when flow is nearly normal to the direction of the average gradient (Fig. 2.6). This observation is contrary to the development of the DBU and TG-DBU that limit the search across background material to other inclusions located in the downgradient direction. This observation of high local gradients creating flows in directions nearly normal to the average gradient is consistent with field observations and numerical model results for hydraulic gradient monitoring networks [60, 61]. For the isotropic fields examined here, the FWHM is an excellent approximation of $\overline{D}$ at p_1 near 0.50 both along the direction of the gradient and orthogonal to it.

The average distances calculated by the DBU method are weighted by the sizes of the inclusions on either end of the travel distance. In contrast, the TG-DBU employs a single average inclusion size, $E[n]$ for a given p_1, thus weighting all distances equally. Fig. 2.5 indicates that longer distances are generally connected to larger inclusions of high permeability material and are more highly weighted in the DBU approach relative to the TG-DBU calculations. The distribution of distances calculated along streamlines are similarly skewed towards larger values.

The basic model and the DBU approach are developed as a function of p_1 and the average

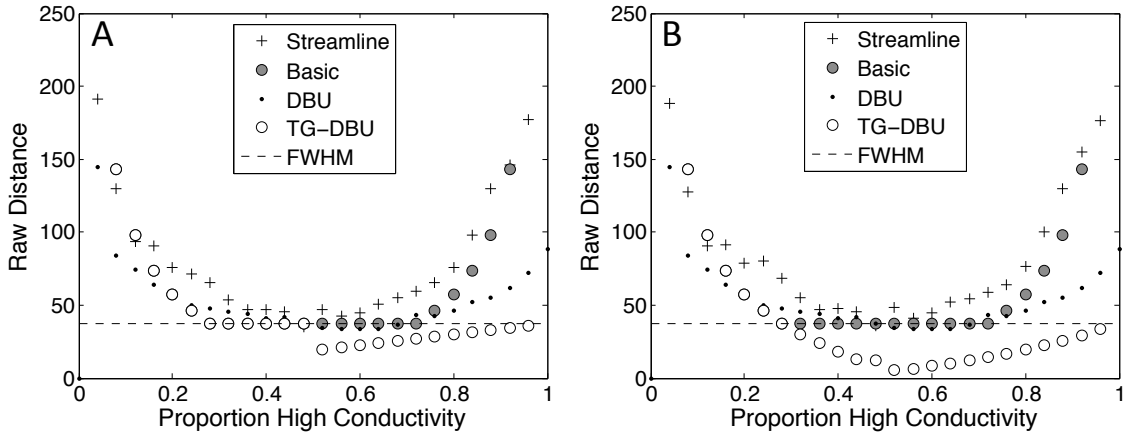


Figure 2.5: Comparison of average distance estimates for different calculation approaches. The DBU and Streamline results are average values calculated over 30 realizations. Results are for Gaussian fields created with κ values of 2 (A) and 4 (B) and an FWHM of 37.7 (σ = 16.0) length-units.

distances between objects. However, effective conductivity is not solely a function of the geometric arrangement of the inclusions. The κ value influences the average distance taken by flowpaths across the lower conductance material and simulations show that changes in κ have the largest impact on flow paths at p_1 values near 0.50. As an example, Fig. 2.6 shows significant changes in the flow path locations for the same binary field at κ values of 2.0 and 4.0.

The average streamline distances (Fig. 2.5) are nearly unchanged from $\kappa = 2$ to $\kappa = 4$, yet Fig. 2.6 shows significant changes in the locations of the streamlines on the same field for the two different κ values. The calculations of average distances for Fig. 2.5 do not include a calculation at exactly $p_1 = 0.50$ (0.48 and 0.52 are the closest). The simulation results in Fig. 2.6 are at exactly $p_1 = 0.50$ and show differences in the average streamline distances with values of 25.4 and 34.9 for the κ values of 2 and 4, respectively.

The impact of the κ value and the initiation of a percolating cluster precludes accurate application of distance-based upscaling techniques for estimation of block-scale properties. Fig. 2.7 shows both $\overline{D}$ calculated along streamlines and the effective conductance calculated numerically for four different truncated Gaussian fields. The streamline-based $\overline{D}$ values are normalized by the FWHM and the numerical K_{eff} is normalized by the geometric mean conductivity. The region around $p_1 = 0.50$ is highlighted with flow and streamline solutions at p_1 increments of 0.008.

For each simulation with $\kappa = 4$, there is a significant increase in the effective conductance at the percolation threshold. This increase is not evident in the $\kappa = 2$ results. The average

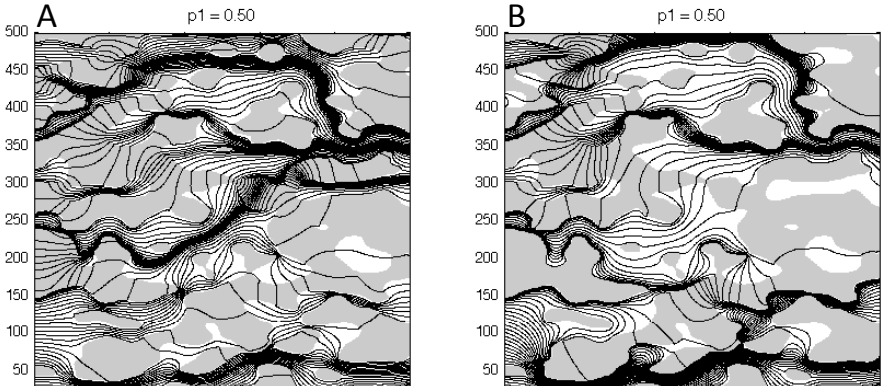


Figure 2.6: Comparison of flowpath locations for the same binary field with κ values of 2 (A) and 4 (B). Results are for Gaussian fields created with an FWHM of 37.7 (σ = 16.0) length-units and a threshold of u = 0.00 (p_1 = 0.50). In both images, flow is from left to right.

streamline distances between high permeability inclusions are not a strong function of p_1 and show gently decreasing values from p_1 = 0.40 to 0.60. These values are well approximated at p_1 values near 0.50 by the FWHM (ratio of 1.0) for both κ values. The vertical lines in Fig. 2.7 indicate the location of the percolation threshold and vary from $p_1 <=$ 0.43 to near 0.57 in these four example fields.

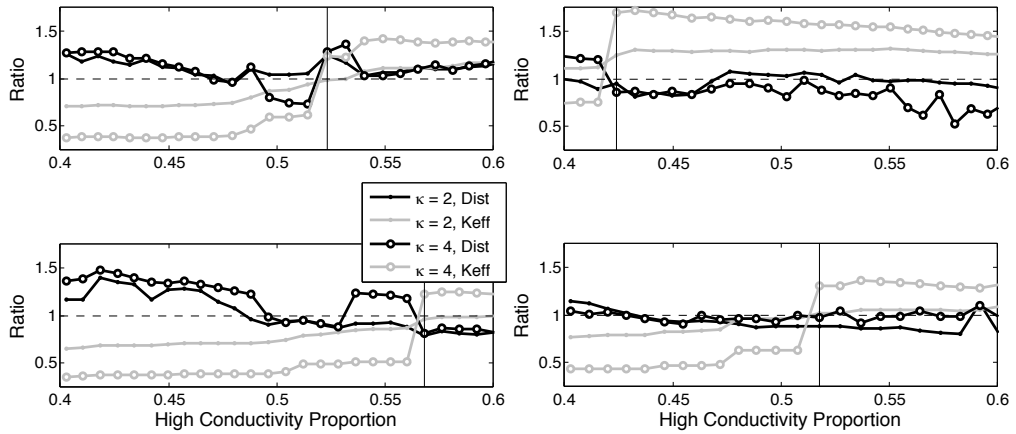


Figure 2.7: Effective conductances and average streamline lengths for p_1 values near 0.50. Results for four different fields are shown. The vertical black line denotes the percolation threshold. Effective conductances are normalized by the geometric mean conductance. Average streamline lengths are normalized by the FWHM.

2.5 Conclusions

This paper presents truncated mG fields as a flexible means of creating simulated binary media and then extends distance-based upscaling to directly utilize properties of the truncated mG fields for calculation of effective conductivity values. Excursion set theory provides techniques for estimation of the number of inclusions, and average inclusion size from knowledge of the threshold and the kernel size (FWHM). These results are coupled with point process theory and distance-based upscaling (DBU) to develop a robust estimator of the effective conductivity of binary media. This new approach is called Truncated Gaussian-Distance Based Upscaling (TG-DBU). TG-DBU is based on expectation relationships and does not require instantiation of the binary field for estimation of the effective conductance.

TG-DBU is unique among upscaling approaches considered in that the kernel/inclusion size parametrized as the FWHM is a direct input to the upscaling function. We introduce the FWHM as a characteristic length for this upscaling and demonstrate its applicability for estimation of distances between inclusions across a broad range of p_1. Extensions to the geometrically-derived basic model that account for deviations in the estimated effective conductivities near the percolation threshold and account for the impact of κ on flow path distances between inclusions result in the TG-DBU model. Comparison of TG-DBU with numerical, DBU and EMT approaches demonstrates the accuracy of TG-DBU and shows results that are at least as accurate as the other techniques for all values of p_1 for the fields examined.

Understanding the role of the average distance between inclusions and the sensitivity of this measure to other parameters is critical for further development of any distance-based upscaling techniques including TG-DBU. The impact of p_1, κ and the percolation threshold on numerical calculations of effective conductivity and distances between inclusions along streamlines were examined. Results show that local gradients normal to the direction of the average gradient cause streamlines to traverse the background material in a direction orthogonal to the average flow direction. For isotropic media examined here, the FWHM value provides a robust approximation of the average streamline distance in the background material at p_1 values near 0.50 and these results are not significantly impacted by percolation behavior or the value of κ. These results indicate that for anisotropic media where a maximum and minimum FWHM are used to define the Gaussian kernel, the FWHM normal to the flow direction will provide the best estimate of the average distance between inclusions. Effective conductances calculated across the percolation threshold indicate that the effective conductivity is a strong function of the κ value when $\kappa = 4$, but at $\kappa = 2$ crossing the percolation threshold has little effect on the resulting effective conductance. For the $\kappa = 2$ results, the geometric mean permeability serves as a reasonable estimate of K_{eff} on both sides of the percolation threshold. For the 30 fields examined, the percolation threshold is reached at p_1 values ranging from less than 0.40 to greater than 0.60 indicating that estimation of effective conductance values is possible, but that detailed knowledge of the field geometry and percolation threshold are necessary for estimation of block-scale

properties.

This page intentionally left blank

Chapter 3

Probabilistic Estimation of Binary Fields Using a Subgrid Model

A binary medium is one that can be adequately described in terms of inclusions of one material phase embedded in another. The two materials can have vastly different properties, and the effective behavior of the binary medium arises as a nonlinear interaction of the two constituents. Further, the proportions of the two materials need not be distributed evenly, and for situations with uneven proportions, inclusions of the less common material will exist within a continuous matrix of the other material. Binary media exhibit complex behaviors and rich dynamics. Examples of binary media are fractured rock (with the fractures acting as high permeability inclusions), sandstone with embedded shale, and engineered composite materials. In many cases, the domain size of the binary medium (henceforth, the coarse-scale) may be a few orders of magnitude larger than the size of the inclusions and it is impractical to measure the inclusions individually. However, the length-scale contrast between the inclusion and domain size is not large enough that the inclusions can simply be homogenized i.e., a mean value for the effect of the inclusion cannot be used, but a more detailed characterization of the fine-scale is required. We call such behavior "multiscale".

This multiscale behavior raises the possibility that it may be possible to infer the characteristics of the latent fine-scale from a judicious set of measurements which inform on both the coarse and fine scales. In order to do so, a *link function* $\mathcal{L}$, that locally captures the effect of fine-scale processes and structure at the coarse-scale is required. If this function is parametrized by structural/dynamical properties of the fine-scale, it is possible to construct statistical summaries for the fine-scale. Realizations of the fine-scale that are equally consistent with observations at both scales can then be created.

We demonstrate inversion with a binary medium where the permeabilities of the two materials, $\mathcal{K}_l$ and $\mathcal{K}_h$, are known. The difference in permeabilities is parametrized as $\kappa = \log_{10}(\mathcal{K}_h/\mathcal{K}_l)$. We assume that we have noisy measurements of the coarse-scale effective log-permeability of the medium, $\mathbf{k}^{(obs)}$, at a few locations; they provide localized insights into the coarse-scale structure and are referred to as *static data*. We also assume that we

have noisy measurements of breakthrough times, $\mathbf{t}_b^{(obs)}$, of a tracer at the same locations, from a tracer test[1]; they capture the integrated effect of fine-scale variations and are referred to as *dynamic data*. The proportion of the high permeability phase $\mathbf{F}(\mathbf{x})$ is assumed to vary in space in an unknown but smooth manner with a known covariance structure. $\mathbf{F}(\mathbf{x})$ is resolved on a relatively coarse grid and describes the spatial variation at the domain-scale. We use the grid to impose a separation of scales, and internal to a grid-box, the inclusions are modeled using a correlated Gaussian field truncated at a threshold consistent with $\mathbf{F}(\mathbf{x})$. The characteristic length of the Gaussian field, δ, is defined below (and is termed a sub-grid or fine-scale variable). We adopt the convention of referring to variations on the grid as being "large" or "small"; we reserve "fine" or "subgrid" for variations that are smaller than a grid-box. Together, $\mathbf{F}(\mathbf{x})$ and δ constitute a characterization of the fine-scale binary medium and their estimation from data $\mathbf{d} = \{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ constitutes a statistical summary of the fine-scale. A model for generating fine scale structures given $\mathbf{F}(\mathbf{x})$ and δ, provides realizations of the latent, unresolved fine-scale consistent with the observations.

The estimation of $\mathbf{F}(\mathbf{x})$ is posed as a Bayesian inverse problem, predicated on a forward problem $\mathcal{M}(\mathbf{K})$ that predicts the breakthrough times $\mathbf{t}_b$ at the observation locations. $\mathbf{K}_e(x)$, the effective log-permeability field, is obtained from $\mathbf{F}(\mathbf{x})$ and δ via a link function i.e. $\mathbf{K}_e(x) = \mathcal{L}(\mathbf{F}(\mathbf{x}), \delta, \kappa)$. Retention of δ in the inverse problem will allow us to evaluate the information content in the observations regarding δ, though we do not expect it to be large. To reduce the dimensionality of the inversion, we develop a reduced-order model (ROM) of $\mathbf{F}(\mathbf{x})$, based on a Karhunen-Loève (KL) decomposition of a 2D field. This allows us to generate $\mathbf{F}(\mathbf{x})$, and consequently $\mathbf{K}_e(x) = \mathcal{L}(\mathbf{F}(\mathbf{x}), \delta, \kappa)$, in a parametric manner by varying δ and $\mathbf{w} = \{w_i\}, i = 1 \ldots M$, where M is the number of Karhunen-Loève modes retained in the ROM and w_i are their weights. Associated predictions of $\mathbf{t}_b$ are obtained using $\mathcal{M}(\mathbf{K})$. The inverse problem is solved by sampling over the $(\mathbf{w}, \delta)$ space using an adaptive Markov Chain Monte Carlo (MCMC) technique and constructing a joint posterior probability density distribution $P(\mathbf{w}, \delta|\mathbf{d})$ from the samples. The use of MCMC allows us to construct posterior distributions of arbitrary topologies, unlike other inversion techniques e.g., Kalman smoothers and filters, which constrain $P(\mathbf{w}, \delta|\mathbf{d})$ to be Gaussian. $P(\mathbf{w}, \delta|\mathbf{d})$ is thereafter used in posterior predictive checks, to construct fine-scale realizations of the binary medium, gather statistics on $\mathbf{K}_e(x)$ and $\mathbf{t}_b$ at the observation locations and gauge the quality of the fit of the model to data.

This paper brings a number of innovations to the problem of estimating an upscaled field variable e.g., permeability fields.

1. *Incorporation of the effect of unresolved scales:* Existing methods for estimating multiscale fields (e.g., zonation; see review in Sec. 3.1.1) involve grid-refinement under various guises – no attempt is made to incorporate structures unresolved by the grid (i.e., subgrid). In contrast, we use a statistical model to capture the effect

[1] We define a tracer test as follows: Water is transported through a porous medium by pumping under a steady pressure gradient. Once a velocity field has been established in the porous medium, a non-reactive tracer is injected; the advective travel time to arrive at various measurement points is called the *breakthrough* time. The exact configuration for the test is described in Sec. 3.3

of subgrid structures in our estimation procedure. The use of a model that has a physical, but subgrid, lengthscale as one of its parameters makes the formulation fundamentally multiscale.

2. *The use of approximate expansions to reduce dimensionality in upscaling problems:* While Gaussian process models have been used to regularize field variables (as reviewed in Sec. 3.1.1), few have used Karhunen-Loève expansions to reduce the dimensionality of a multiscale inverse problem. Further we use a Gaussian processes to model a *latent* variable from which the object of inference, $\mathbf{F}(\mathbf{x})$, and the observed field, $\mathbf{K}_e(x)$, are obtained via mappings. Thus, while the latent field variable is constrained to be smooth (so that it can be modeled as a Gaussian), this constraint is not necessarily extended to $\mathbf{F}(\mathbf{x})$ or $\mathbf{K}_e(x)$. In our case, $\mathbf{K}_e(x)$ contains sharp gradients. This is in contrast to existing studies where the object of inference (which is also the observed field) is modeled as a Gaussian and thus required to be a smooth function.

3. *The use of adaptive, general-purpose MCMC schemes:* Our use of a Karhunen-Loève expansion reduces the problem of estimating a field $\mathbf{F}(\mathbf{x})$ to that of inferring the values of a few parameters which are independent *a priori*. This allows us to use adaptive, general-purpose (and efficient!) MCMC schemes (and software packages) in a straightforward manner. In contrast, existing MCMC-based field-estimation methods (reviewed in Sec. 3.1.1) use specialized blocking schemes and need to retain spatial correlations in their MCMC block proposals. It is unclear how one would use general-purpose, adaptive MCMC software packages in such contexts.

The chapter is structured as follows. In Sec. 3.1 we review literature on the key elements of the research presented here. In Sec. 3.2 we describe the forward model $\mathcal{M}(\mathbf{K})$, the link function $\mathcal{L}$, and models used for reducing the dimensionality of the inverse problem. In Sec. 3.3, we pose the inverse problem and test the inversion technique on a problem with various types of data. In Sec. 3.5, we instantiate realizations of the binary fine-scale from the inferences and test their predictive skill using transport simulations i.e., breakthrough times. We draw our conclusions in Sec. 3.7.

3.1 Literature Review

The estimation of field variables (permeabilities, hydraulic conductivities, etc), from limited data, per se, have long been topics of active research in hydrology; see [62, 63] for recent reviews. In this section we restrict ourselves to reviewing existing literature on the multiscale (or multilevel) inference of log-permeability modeled as random fields and adaptive MCMC samplers.

3.1.1 Estimation of Random Fields

The use of random fields to regularize a spatial variable has been explored within the context of inferring log-permeability fields. Lee *et al.* [64] considered the estimation of spatially dependent permeability by modeling it as a random field. They estimated the permeability in each grid-block of the discretized field from dynamic data obtained from an "inverted 9 spot test". Two separate prior models were adopted for the random field to regularize the problem – a Markov random field (MRF) model with an unknown precision parameter (i.e., the precision parameter was also inferred when estimating the permeability) and a model based on Gaussian processes (GP) with a known variogram and mean permeability in the rectangular domain. Neither of the two models reduced the dimensionality of the problem i.e., the number of parameters being estimated was equal to the size of the Cartesian mesh, which ranged between 32^2 and 64^2. This work required specialized updating schemes to improve mixing in the Metropolis-Hastings sampler used to construct the posterior distribution of the permeability field. About 50,000 samples were required. MRF priors were also used by Wang *et al.* [65] when estimating the initial (spatial) distribution of a contaminant within a porous medium. The contaminant was transported by groundwater flows and time-variant concentration measurements were available at a few locations. They employed a hierarchical Bayesian formulation to estimate the concentration distribution as well as the precision of the MRF model and the variance of the measurement error. As in [64], no attempt was made to reduce the dimensionality of the inference problem. Fu and Gómez-Hernández [9, 66] present a more recent example of the use of MCMC with blocked-updating when inferring log-permeability fields. Unlike Lee *et al.* [64] where a red-black decomposition of grid-blocks was used to update the log-permeability field (modeled as a MRF) in the MCMC, they used a multiGaussian representation for the object of inference and devised a specialized technique for constructing the proposal within the MCMC. In particular, they divided the grid-blocks in the Cartesian mesh into concentric "strips", which were updated together as a block; the proposals for the blocks were obtained by kriging.

The need for explicit regularization can be eliminated if one can represent the field to be estimated using a low-dimensional model. If the object of inference can be modeled as a multivariate Gaussian field, a truncated Karhunen-Loève expansion can be used. In [67], Li *et al.* consider estimation of the log-conductivity distribution of a reservoir from measurements of the steady-state hydraulic heads by preserving 400 terms in the Karhunen-Loève expansion. Application of Karhunen-Loève expansion to provide a reduced-order model for a random field was also used by Marzouk *et al.* in [68] to estimate a log-diffusivity field within the context of a 1D thermal transport problem, with time-dependent temperature measurements at a few sensor points. They found that the Karhunen-Loève expansion based on the *prior* covariance of the log-diffusivity field was remarkably accurate for representing its *posterior* distribution, conditioned on the observations. In [69], Jafarpur and Mclaughlin compare the use of a Karhunen-Loève transform versus the discrete cosine transform (DCT) and find the latter to be more advantageous. In [70] they couple the reduced order DCT model to an ensemble Kalman filter to infer permeability fields as well as

reservoir states via history matching. In [71], they use the discrete cosine bases as a *sparse* representation for the log-permeability field and infer their value *as well as the sparsity pattern* via history matching.

Multiscale/multilevel inversion techniques explicitly recognize the existence of more than one scale and employ different types of data across scales. These techniques generally involve solving the inverse problem at different levels of discretizations , (i.e., on a multilevel mesh), with conditioning relations (i.e., upscaling and downscaling functions) to link scales together [72, 73]. Multiscale solutions primarily differ in the complexity of the conditioning relations and whether the multiscale inference requires iteration between scales. Other techniques run separate MCMC chains using coarse and fine-scale models, with periodic swaps of parameters between them [74, 9], or use a fast coarse-scale (or approximate model) as a preconditioner/filter for proposals (inside an MCMC chain), prior to computing the posterior with a finely-resolved forward model [7, 8, 9]. Note that these methods require explicit definition of coarse, fine and if necessary, intermediate scales. On the other hand, "zonation" methods [3, 4, 5, 6] adopt a continuous-level-of-detail approach to inference and in the process combine elements of dimensionality reduction and multiscale inference.

We borrow the approach adopted in [7, 8, 68] and use Karhunen-Loève expansions to reduce the dimensionality of the inverse problem. We apply them here to the spatially varying $\mathbf{F}(\mathbf{x})$ which is then input to the link function providing an effective permeability informed from both scales. Instead of using a multilevel inversion technique, we assume that a sufficient contrast exists between the resolution at which we perform the inference and the size of individual fine-scale/subgrid structures that a complete description of the fine-scale is not very useful. However, we adopt the practice, common in multiscale inversion, of enforcing scale separation using a grid.

3.1.2 Adaptive Markov Chain Monte Carlo Techniques

In recent years, MCMC techniques have been increasingly used to fit models to observations [75], since they allow estimation of parameters while simultaneously quantifying the uncertainty in the estimate. Further, they place no restrictions on the probability distributions of the estimated parameters. Metropolis-Hastings (MH) samplers [75] are commonly used since they place no restrictions on the kind of models, the type of likelihood expressions or the priors used in posing the inverse problem. Blockwise updates, when a number of (or all) parameters are updated at once are typically used when estimating field quantities [9, 64, 75, 76, 77, 78].

Adaptive Metropolis [76, 79] (AM) is a variation of the MH sampler which uses a *global* adaptive strategy to perform online tuning of the current proposal to increase mixing and acceptance rates. AM starts with a pre-specified proposal density but periodically recalculates an empirical posterior covariance based on the samples collected up to that point. The covariance asymptotically resembles that of the posterior. The technique is neither Marko-

vian nor reversible and in [76] the authors identify the conditions under which AM will recover the desired stationary distribution. Delayed rejection [80, 81, 82] is a *local* MH variation that combines different proposals. An MH sampler is started with a rather large proposal covariance. When a proposal is rejected, the initial covariance is scaled down by a uniform factor, and tried again rather than simply advancing in the sample path (hence delayed rejection, DR). DR has been shown to outperform MH [79]. Delayed Rejection Adaptive Metropolis (DRAM), the MCMC technique used here, is an amalgamation of the AM and DR [77]. DRAM is non-Markovian and provably ergodic, i.e., it yields asymptotically unbiased estimators [77].

3.2 Models Used in the Inverse Problem

In this section we describe the models used in our inverse problem. We first review the link function $\mathcal{L}$ that summarizes the impact of subgrid structures, followed by Karhunen-Loève expansions of random fields which are used to reduce the dimensionality of the inverse problem. Finally, we describe $\mathcal{M}(\mathbf{K})$, a porous media transport model which serves as the forward problem in the inversion.

3.2.1 The Link Function $\mathcal{L}$

The key to our multiscale inference procedure is a recently developed *link function* ($\mathcal{L}(\mathbf{F}(\mathbf{x}), \delta, \kappa)$), a statistical model that estimates the effect of unresolved inclusions on the log-permeability of a grid-block. Full details of this truncated Gaussian- distance-based upscaling (TG-DBU) model can be found in [83]. Note that $(\mathbf{F}(\mathbf{x}), \delta)$ is an incomplete specification of the fine-scale structures that may reside within a grid-block and consequently an infinite number of realizations (i.e., an ensemble) of the fine-scale may be conditioned to it. The statistical model constructs a representative value for this distribution, which can then serve as a deterministic approximation for the ensemble.

Consider a binary medium with component permeabilities $\mathcal{K}_l$ and $\mathcal{K}_h$. Consider, too, a rectangular inclusion embedded in a "box" of the binary medium containing both the inclusion and the matrix. Knudby *et al.* [36] provide a model for the effective permeability of the "box", as a function of the inclusion and box parameters, $\mathcal{K}_l$, $\mathcal{K}_h$ and D_{norm}, an average normalized distance traveled by the flow in the matrix between inclusions. In the context of a random medium, a mean $\overline{D_{norm}}$ is more meaningful and can be calculated easily from a fully-resolved realization of the medium. Instead, we construct a model for $\overline{D_{norm}}$ that does not require a random binary field to be instantiated.

We consider a spatially correlated multiGaussian (mG) field with a correlation length of δ created by initializing an uncorrelated mG field, with values sampled i.i.d. (independent and identically distributed) from a standard normal, and convolving it with a Gaussian

kernel. The kernel is aligned with the discretization of the field, i.e. a diagonal covariance matrix $\sigma^2 I$ where I is the identity matrix. This symmetric Gaussian kernel is fully characterized by its full-width-at-half-maximum (FWHM), $\delta = \sigma\sqrt{8\ln 2}$. The FWHM also characterizes the spatial correlation of the mG field and is a representative length-scale for the inclusions. If the correlated mG field is truncated at a threshold $w, -\infty \leq w \leq \infty$, we define a $w-level$ excursion set as $X_w = \{x \in R^d : Y(x) \geq w\}$. The expression for the corresponding variogram can be found in [83].

Following [52], analytical expressions for the expected values of the following quantities - N, the number of cells above the truncation threshold w, the number m of distinct regions (inclusions) above the threshold and the number n of pixels in each region are given by:

$$\begin{aligned} \frac{\mathbb{E}[N]}{S} &= \frac{1}{2\pi}\int_w^\infty \exp(-z^2)dz \\ \mathbb{E}[m] &= |EC| = \frac{\exp(w^2/2)}{(2\pi)^{3/2}}\left(\frac{\delta}{\sqrt{4\ln(2)}}\right)^{-2}|w|S \\ \mathbb{E}[n] &= \frac{\mathbb{E}[N]}{|EC|} \end{aligned} \tag{3.1}$$

$\mathbb{E}[x]$ denotes the expectation of x. Here, EC, the Euler Characteristic in 2D, represents the number of connected objects in the domain minus the number of holes in those objects. The EC tends to 0 as w tends to 0, which in turn denotes $\mathbf{F}(x) = 0.5$. For $w < 0$, $EC < 0$. S is the area of the domain. The relationships in Eq. 3.1 are applied at the scale of each grid block (i.e., S = area of coarse-scale grid block).

The distance between the inclusion centroids is modeled as a Poisson distribution, with a mean given by $\sqrt{(S/(\pi|EC|))}$. The mean flow travel distance through the matrix between inclusions $\overline{D_*}$ is obtained by subtracting off a mean inclusion size from the inter-centroidal distance i.e., $\overline{D_*} \approx \sqrt{(S/(\pi|EC|))} - 2\sqrt{\mathbb{E}[n]/\pi}$. This approximation for the edge to edge distances between adjacent inclusions holds when there are few inclusions of either material (i.e., low/high values of w), but breaks down when as w approaches 0.5. At these intermediate w values, $\overline{D_*}$ is estimated as the FWHM in a "basic" model. A more complex model that includes the impact of κ on $\overline{D_*}$ is utilized here:

$$\overline{D_*} \approx \overline{D^*} = \begin{cases} \max\left(\sqrt{\frac{S}{\pi|EC|}} - 2\sqrt{\frac{\mathbb{E}[n]}{\pi}}, \frac{\delta}{\kappa-1}\right) & \text{if } \mathbf{F} \leq 0.5 \\ \max\left(\sqrt{\frac{S}{\pi|EC|}} - 2\sqrt{\frac{\mathbb{E}[n]}{\pi}}, \delta(1-\mathbf{F})^{\kappa-1}\right) & \text{otherwise} \end{cases} \tag{3.2}$$

where $\kappa = \log_{10}(\mathcal{K}_h/\mathcal{K}_l)$. The normalized version $\overline{D_{norm}}$ is given by

$$\overline{D_{norm}} = \frac{\overline{D^*}}{\sqrt{S/(\pi|EC|)}}. \tag{3.3}$$

Additional details on this model can be found in [83]. This expression allows us to model the effect of subgrid (unresolved) fine-scale structures on the permeability, as a function of $\mathbf{F}(\mathbf{x})$, δ and κ . Given $\mathbf{F} = \mathbb{E}[N]/S$, we evaluate w, $|EC|$ and $\mathbb{E}[n]$ using Eq. 3.1. Thereafter, with knowledge of δ and κ, we estimate $\overline{D^*}$ from Eq. 3.2.

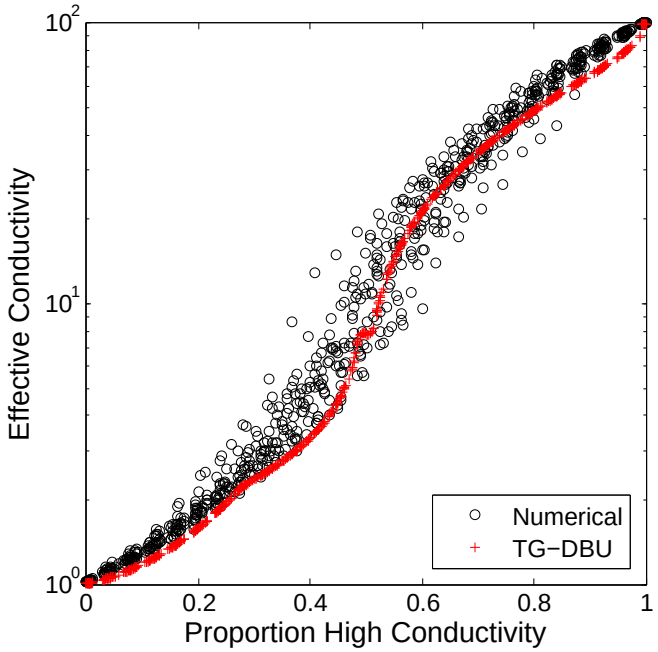


Figure 3.1: Plot of $\mathbf{K}_e(x)$ calculated using the link function $\mathcal{L}$, (labeled: TG-DBU) compared to $\mathbf{K}_e(x)$ calculated numerically from 30 random binary fields as a function of $\mathbf{F}(\mathbf{x})$. These results were calculated for a square domain with length, Δ = 500 units and δ is 37.7 (δ/Δ = 0.075).

Note that this subgrid model provides a log-permeability $\mathbf{K}_e(x) = \mathcal{L}(\mathbf{F}, \delta, \kappa)$ as a point approximation of the distribution of log-permeabilities consistent with $\mathbf{F}$, δ and κ. In Fig. 3.1, we plot the effective log-permeabilities for $\mathcal{K}_l = 1$, $\mathcal{K}_h = 100$ and δ of 37.7 grid-cells (in a domain of 500×500 grid-cells) as predicted by our model, $\mathcal{L}$. We also plot the numerically evaluated log-permeabilities corresponding to 30 binary field realizations having the same $\{\mathbf{F}(\mathbf{x}), \delta\}$ with points, forming a cloud around $\mathbf{K}_e(x) = \mathcal{L}(\mathbf{F}(\mathbf{x}), \delta, \kappa)$. Note that the model predicts $\mathbf{K}_e(x)$ for all $\mathbf{F}(\mathbf{x})$. The break in the log-permeability predictions $\mathcal{L}(\mathbf{F}(\mathbf{x}), \delta, \kappa)$ at $\mathbf{F} \approx 0.5$ is due to application of the phase interchange theorem [12] as part of $\mathcal{L}$ that provides consistent upscaling of inclusions of either material within the other. Additionally, this break also accounts for non-linear percolation effects i.e., the high permeability phase exists in a connected region large enough to span the domain, see [83].

The scaling function, $\mathcal{L}$, used here is unique among binary media scaling functions in that it incorporates $\mathbf{F}(\mathbf{x})$, δ and κ into estimates $\mathbf{K}_e(x)$. However, the impact of δ on the resulting $\mathbf{K}_e(x)$ values is limited to areas of $\mathbf{F}(\mathbf{x})$ near 0.50 under the expected conditions of $\delta/\Delta \ll 1$, where Δ is the size of the region where the impact of inclusions will be modeled using $\mathcal{L}$. Practically, Δ denotes the grid-box size of the mesh that we will use in this work to

discretize a domain, and thus $\mathcal{L}$ acts as a subgrid model for the structures that the grid will not resolve. The impact of δ increases for when $\delta/\Delta \to 1$, but that violates the assumption that inclusions are too small to be resolved by the mesh – in practice, we take it to mean $\delta/\Delta \approx 0.1$. The practical ramification of this varying impact is that δ is difficult to estimate. If $\mathbf{F}(\mathbf{x})$ in a domain shows large variation, it may be close to 0.5 only at a few spots, where the impact of δ may be felt. Thus the observations may not be very informative regarding δ.

In the rest of the paper, we will use $\mathbf{K}_e(x)$ as a deterministic approximation for the log-permeabilities and model the discrepancy between observations of log-permeability and model predictions as simple i.i.d. Gaussians (homoscedastic errors). While the choice of the error model (Gaussian versus a more involved one) does not detract from the general characteristics of an inference procedure, it does impact the accuracy of the estimates/inferences.

3.2.2 Karhunen-Loève Expansions of Random Fields

In order to reduce the dimensionality of our inverse problem, we will model fields with a truncated Karhunen-Loève series. Let $R(\mathbf{x},\omega)$ be a real-valued random field with zero mean, finite second moments and covariance function that is continuous in $\mathcal{D}\times\mathcal{D}$. $\omega \in \Omega$, where Ω is a sample space and $R(\mathbf{x},\omega)$ can be considered to be a collection of real-valued random variables, indexed by $\mathbf{x}\in\mathcal{D}$, $\mathcal{D}$ being a bounded spatial domain. Then, the Karhunen-Loève expansion of $R(\mathbf{x},\omega)$ can be written as $R(\mathbf{x},\omega) = \sum_{i=1}^{\infty} w_i(\omega)\sqrt{\lambda_i}\phi_i(\mathbf{x})$. This equality holds in the pointwise and mean-square sense; convergence is in $L^2(\Omega)$ for all $\mathbf{x}\in\mathcal{D}$. Further, if $R(\cdot)$ is Gaussian and almost surely continuous, then the convergence is uniform in $\mathcal{D}$ with probability 1 [84]. λ_i and $\phi_i(\mathbf{x})$ are the eigenvalues and eigenfunctions of the covariance kernel $C(\mathbf{x},\mathbf{y})$

$$\int_{\mathcal{D}} C(\mathbf{x}_1,\mathbf{x}_2)\phi_i(\mathbf{x}_2)d\mathbf{x}_2 = \lambda_i\phi_i(\mathbf{x}_1). \tag{3.4}$$

Since $R(\cdot)$ is assumed Gaussian, the covariance kernel $C(\mathbf{x},\mathbf{y})$ is symmetric and positive semi-definite and so, by [85], $C(\mathbf{x}_1,\mathbf{x}_2) = \sum_{i=1}^{\infty}\lambda_i\phi_i(\mathbf{x}_1)\phi_i(\mathbf{x}_2)$ where $\phi_i(\mathbf{x})$ are continuous functions and form an orthonormal system in $L^2(\mathcal{D})$. Also, $w_i \sim \mathsf{N}(0,1)$ and independent of each other.

A multiGaussian field $R(\mathbf{x},\omega)$ can be approximated using a Karhunen-Loève expansion as $R_M(\cdot)$ and its covariance function can be represented as

$$R_M(\mathbf{x},\omega) = \sum_{i=1}^{M} w_i(\omega)\sqrt{\lambda_i}\phi_i(\mathbf{x}), \qquad C_M(\mathbf{x}_1,\mathbf{x}_2) = \sum_{i=1}^{M}\lambda_i\phi_i(\mathbf{x}_1)\phi_i(\mathbf{x}_2)$$

The total variance or "energy" of $R_M(\cdot)$ is given by

$$\int_{\mathcal{D}} E(R_M(\mathbf{x},\omega)^2)d\mathbf{x} = \int_{\mathcal{D}} C_M(\mathbf{x},\mathbf{x})d\mathbf{x} = \sum_{i=1}^{M}\lambda_i^2 \tag{3.5}$$

3.2.3 The Transport Model $\mathcal{M}(\mathbf{K})$

$\mathcal{M}(\mathbf{K})$ is a 2D Darcy-flow model for the transport of an inert tracer through a saturated porous medium by an incompressible, single-phase fluid. Given a coarse-scale log-permeability field $\mathbf{K}$, appropriate initial and boundary conditions (including a steady-state pressure gradient, a fluid source and a sink inside $\mathcal{D}$), the model calculates a steady state velocity field and advects a tracer (treated as a passive scalar) through it to obtain breakthrough times $\mathbf{t}_b$ at a set of N_s "sensor" locations inside $\mathcal{D}$.

As shown in Fig. 3.2(a), we consider a 2D domain $\mathcal{D}$ with no-flow boundary conditions imposed on $\partial\mathcal{D}$. We consider a log-permeability field $\mathbf{K}$ defined on $\mathcal{D}$. An incompressible fluid, with viscosity μ is pumped in at the lower left corner and pumped out at an equal rate at the upper right. In this problem, we will ignore the effect of gravity. Therefore, via Darcy's model for porous media flows, the velocity $\mathbf{v}$ is given by

$$\mathbf{v} = \frac{\mathcal{K}}{\mu}\nabla p, \qquad \nabla\cdot\mathbf{v} = \nabla\cdot\frac{\mathcal{K}}{\mu}\nabla p = \frac{q}{\rho} \tag{3.6}$$

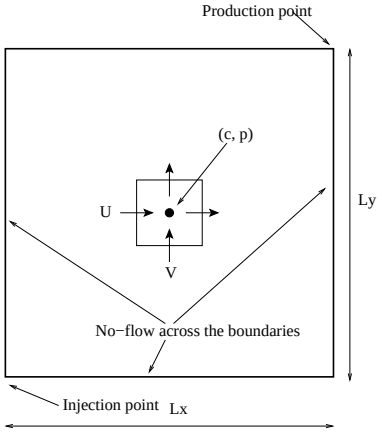
where p is the pressure field defined on $\mathcal{D}$, q is the strength of the source/sink and ρ is the density of the fluid. The equation is solved using the second-order finite-volume scheme described in [86], on a uniform mesh. Two-point flux approximations are used, and the permeability at the interface of adjacent grid-blocks are estimated by a harmonic average. A solution of these equations, for a injection-production well pair in a binary medium, is shown in Fig. 3.2(b). The light areas denote high-permeability material and the dark regions are low-permeability. A few streamlines, flowing from bottom left to top right, are plotted in Fig. 3.2(b). The local permeability is considered isotropic, and the permeability $\mathcal{K}$ is modeled as a 2D field rather than a full tensor. The proportion of high-permeability material $\mathbf{F}(\mathbf{x})$ and the true effective (upscaled) log-permeability are shown in Fig. 3.2 (c) and (d). Note that the $\mathbf{K}_e(x)$ field shows a rougher distribution and significantly more structure than $\mathbf{F}(\mathbf{x})$.

Solving Eq. 3.6 for p yields the velocity $\mathbf{v}$ at the centers of the grid-block edges i.e., the velocities are obtained on a staggered mesh. This is used to advect the passive scalar, using the model in Eq. 3.7,

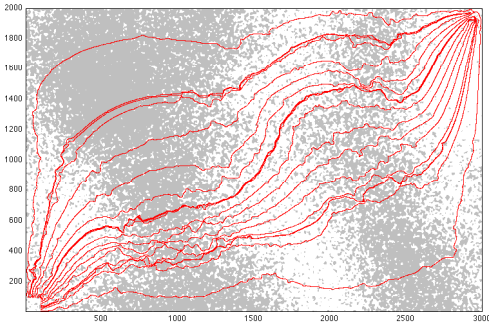
$$\frac{\partial c}{\partial t} + \mathbf{v}\cdot\nabla c = q_c \tag{3.7}$$

where $c(\mathbf{x})$ is the concentration field of a tracer and q_c, non-zero only at the lower left and upper right corners of $\mathcal{D}$, is the source and sink for the tracer. The concentration of the tracer at the source, c_{source}, is set to 1 and $q_{c,source}$ is set equal to the flux of the fluid. The tracer flux at the sink, $q_{c,sink}$, is obtained by multiplying the fluid outflow with the local tracer concentration, $c_{sink}(t)$, which increases in time until it reaches the source value of 1.0. Eq. 3.7 was solved on the same Cartesian mesh as Eq. 3.6, using an upwind second-order, finite-volume scheme [87]. Tracer concentrations were monitored at two sets of sensor locations (set A with 20 sensors and B with 34) as shown in Fig. 3.2(e) and (f); the time t_b at which $c = 0.5$ was achieved at any given sensor and was denoted as its breakthrough

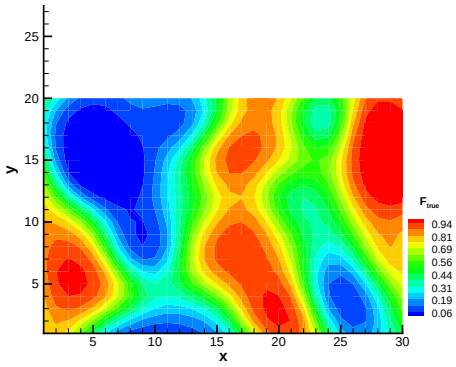
time. Note that Eq. 3.7 does not model pore-scale dispersion or molecular diffusion of the tracer.



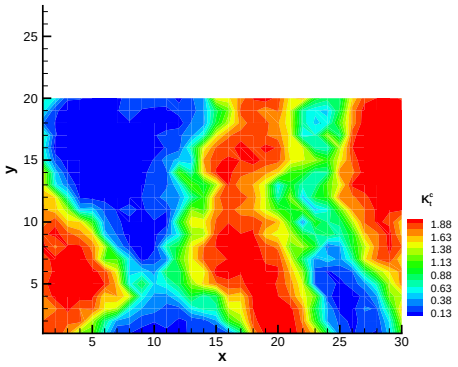
(a) Domain



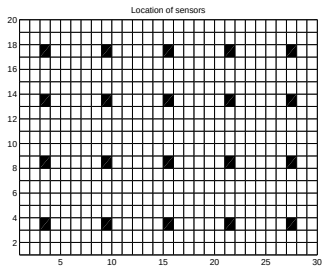
(b) Fine-scale



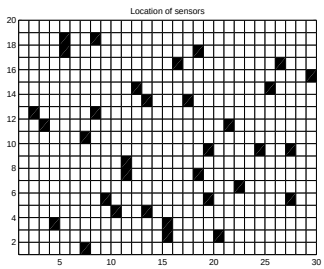
(c) Proportionality field $\mathbf{F}(\mathbf{x})$



(d) The true $\log_{10}(\mathbf{K}_e(x))$ field



(e) SSA



(f) SSB

Figure 3.2: (a) The 2D rectangular domain $\mathcal{D}$ with associated dimensions. No-flow boundary conditions are defined on $\partial\mathcal{D}$. A grid-block showing the collocation for pressure p and the velocities (u, v) in the x- and y-directions is shown. (b) We show a schematic of a binary medium higher permeability (white) and lower permeability (gray). Individual inclusions of either material in the other are visible. (c) The true $\mathbf{F}_c^t(\mathbf{x})$ field for the proportion of inclusions. (d) The true upscaled log-permeability field $\mathbf{K}_c^t$. (e) The 30×20 coarse-scale computational mesh with the locations of 20 sensors (sensor-set A, SSA). (f) We show the locations of the 34 sensors in sensor-set B (SSB).

3.3 Multiscale Inference

In this section, we pose and solve an inverse problem using the models described in Sec. 3.2. Specifically, we explore the ability of the observations to constrain the inference of structures at two spatial scales, the relative contributions of the static and dynamic data and the effect of increasing the number of observations. We also demonstrate sampling of the posterior distribution to generate fine-scale realizations which are consistent with the observations.

3.3.1 Posing the Inverse Problem

Consider that the domain $\mathcal{D}$ is 2D, rectangular, of size $L_x \times L_y$, and discretized by a Cartesian mesh of resolution $N_x \times N_y$. This is also the grid used for imposing scale separation. Consider a spatially variable field $\zeta(\mathbf{x})$, $-\infty \leq \zeta(\mathbf{x}) \leq \infty$ defined on $\mathcal{D}$. We model $\zeta(\mathbf{x})$ as a random field using GP, i.e., the discrete form of $\zeta(\mathbf{x})$, the vector ζ, is a random variable with multivariate Gaussian distribution and a known covariance matrix Γ. We further specify that the inclusion proportion $\mathbf{F}(\mathbf{x})$ is an analytical function of $\zeta(\mathbf{x})$, and it, in turn, governs the predicted effective log-permeability $\mathbf{K}_e(x)$ and breakthrough time $\mathbf{t}_b$ via models $\mathcal{L}(\mathbf{F}(\mathbf{x}),\delta)$ and $\mathcal{M}(\mathbf{K})$. These relationships are summarized in Eq. 3.8

$$\begin{aligned}
\zeta &\sim \mathsf{N}(0,\Gamma), \\
\Gamma_{ij} &= C(\mathbf{x}_i,\mathbf{x}_j) = a\exp(-|\mathbf{x}_i-\mathbf{x}_j|^2/b^2), \\
\mathbf{F}(\mathbf{x}) &= \frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\zeta(\mathbf{x})}{\sqrt{2}}\right)\right), \\
\mathbf{K}_e(x) &= \mathcal{L}(\mathbf{F}(\mathbf{x}),\delta,\kappa), \\
\mathbf{t}_b &= \mathcal{M}(\mathbf{K}_e(x))
\end{aligned} \tag{3.8}$$

where Γ_{ij}, an element of the covariance matrix Γ, denotes the correlation between gridblocks i and j on the mesh. The scale a and range b of the covariance kernel are assumed known. Note that the analytical transformation $\zeta \mapsto \mathbf{F}(\mathbf{x})$ is required to map $\mathbf{F} \in [0,1]$ to $\zeta \in [-\infty,\infty]$, so that a Gaussian model may be used. The inference does not depend on the particular transformation in Eq. 3.8; any transformation that allows the use of Gaussian fields (and consequently, a Karhunen-Loève decomposition) may be used.

Given a set of noisy data $\mathbf{d} = \{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ at a set of N_s "sensor" points, we wish to infer the posterior distribution $P(\mathbf{K}_e(x),\mathbf{F}(\mathbf{x}),\delta|\mathbf{d})$. Using Bayes' formula, this can be written as

$$\underbrace{P(\mathbf{K}_e(x),\mathbf{F}(\mathbf{x}),\delta|\mathbf{d})}_{Posterior} \propto \underbrace{P(\mathbf{d}|\mathbf{K}_e(x),\mathbf{F}(\mathbf{x}),\delta)}_{Likelihood}\underbrace{\pi(\mathbf{K}_e(x),\mathbf{F}(\mathbf{x}),\delta)}_{Prior} \tag{3.9}$$

where $\pi(\cdot)$ denotes our prior belief regarding the distribution of a variable. We model the

discrepancy between the data $\mathbf{d}$ and model predictions as Gaussian residuals, i.e.,

$$\begin{aligned}\mathbf{k}^{(obs)} - \mathbf{K}_e(x) &= \mathbf{k}^{(obs)} - \mathcal{L}(\mathbf{F}(\mathbf{x}), \delta, \kappa) = \mathbf{e}_k \sim \mathsf{N}(\mu_k, \Gamma_k), \\ \mathbf{t}_b^{(obs)} - \mathbf{t}_b &= \mathbf{t}_b^{(obs)} - \mathcal{M}(\mathbf{K}_e(x)) = \mathbf{e}_t \sim \mathsf{N}(\mu_t, \Gamma_t),\end{aligned} \tag{3.10}$$

where μ_k, μ_t are the means of the discrepancy between observed and model (i.e., predicted) log-permeabilities and breakthrough times and Γ_k, Γ_t the corresponding error covariances respectively. Under these error-modeling assumptions, Eq. 3.9 reduces to

$$\begin{aligned}P(\mathbf{K}_e(x), \mathbf{F}(\mathbf{x}), \delta | \mathbf{d}) &\propto P(\mathbf{d} | \mathbf{K}_e(x), \mathbf{F}(\mathbf{x}), \delta) \pi(\mathbf{K}_e(x), \mathbf{F}(\mathbf{x}), \delta) \\ &\propto \exp\left(-[\mathbf{e}_k - \mu_k]^T \Gamma_k^{-1} [\mathbf{e}_k - \mu_k]\right) \exp\left(-[\mathbf{e}_t - \mu_t]^T \Gamma_t^{-1} [\mathbf{e}_t - \mu_t]\right) \\ &\quad \pi(\mathbf{K}_e(x), \mathbf{F}(\mathbf{x}), \delta),\end{aligned} \tag{3.11}$$

We assume that the discrepancies $\mathbf{e}_k$ and $\mathbf{e}_t$ are independent and model them as i.i.d. Gaussians with constant standard deviations i.e.,

$$\Gamma_k = \sigma_k^2 \mathbf{I}; \qquad \Gamma_t = \sigma_t^2 \mathbf{I}.$$

where $\mathbf{I}$ is the identity matrix. This is equivalent to stating that there are no systematic discrepancies between observations and model predictions as a function of $\mathbf{F}(\mathbf{x}), \delta$ and κ. Given the relationship between $\mathbf{K}_e(x), \mathbf{F}(\mathbf{x})$ and ζ (Eq. 3.8), and the modeling assumptions regarding errors, Eq. 3.11 can be compactly written in terms of ζ as

$$\begin{aligned}P(\zeta, \delta | \mathbf{d}) &\propto P(\mathbf{d} | \zeta, \delta) \pi(\zeta) \pi(\delta) \\ &\propto \exp\left(-\frac{[\mathbf{e}_k(\zeta, \delta) - \mu_k]^T [\mathbf{e}_k(\zeta, \delta) - \mu_k]}{\sigma_k^2}\right) \\ &\quad \exp\left(-\frac{[\mathbf{e}_t(\zeta, \delta) - \mu_t]^T [\mathbf{e}_t(\zeta, \delta) - \mu_t]}{\sigma_t^2}\right) \\ &\quad \pi(\zeta) \pi(\delta).\end{aligned} \tag{3.12}$$

Here, we have also assumed that the prior distributions of δ and ζ are independent.

Solving the inverse problem in Eq. 3.12 would require us to infer each of the elements of ζ (though constrained by Γ); the dimensionality of the inverse problem is $N_x \times N_y$, which can be extremely large depending on the mesh. In order to reduce the dimensionality of the inverse problem, we appeal to the discussion in Sec. 3.2.2 and construct a low-dimensional model of $\zeta(\mathbf{x})$ using a Karhunen-Loève expansion.

$$\zeta_M = \sum_i^M w_i \sqrt{\lambda_i} \phi_i, \quad \text{with} \quad w_i \sim \mathsf{N}(0, 1), \tag{3.13}$$

where λ_i are the eigenvalues of the covariance matrix Γ, ϕ_i the corresponding eigenvectors, and w_i the weights which are modeled as i.i.d standard normals due to the GP model for ζ. In this expansion, we retain the M-largest eigenmodes. Note that by Eq. 3.5, the variance of the approximate field is less than the original one, and inferences will tend to be smoother than the true field. The link between the Karhunen-Loève weights $\mathbf{w}$ and $\mathbf{t}_b$ can be succinctly written as

$$\mathbf{w} \xrightarrow{\text{KL modes}} \zeta \xrightarrow{\text{Eq. 3.8}} \mathbf{F}(\mathbf{x}) \xrightarrow{\mathcal{L},\delta,\kappa} \mathbf{K}_e(x) \xrightarrow{\mathcal{M}(\mathbf{K})} \mathbf{t}_b \tag{3.14}$$

Replacing ζ in Eq. 3.12 with its Karhunen-Loève expansion (Eq. 3.13), we get

$$\begin{aligned} P(\mathbf{w},\delta|\mathbf{d}) \quad \propto \quad & \exp\left(-\frac{[\mathbf{e}_k(\mathbf{w},\delta)-\mu_k]^T[\mathbf{e}_k(\mathbf{w},\delta)-\mu_k]}{\sigma_k^2} - \frac{[\mathbf{e}_t(\mathbf{w},\delta)-\mu_t]^T[\mathbf{e}_t(\mathbf{w},\delta)-\mu_t]}{\sigma_t^2}\right) \\ & \pi(\delta)\prod_{l=1}^{M}\exp(-w_l^2). \end{aligned} \tag{3.15}$$

Here, the prior on $\mathbf{w}$, $\pi(\mathbf{w})$, has been expressed in terms of its independent elements, w_i, whose priors are standard normals. The objects of inference are the M elements of $\mathbf{w}$ and δ. Note that the dimensionality of the inverse problem is no longer directly dependent on the mesh used for the inversion.

We will assume that the porosity of the binary medium is a constant in space and time, and is known. The basis for this assumption is the observation that variations in $\mathbf{K}$ for most porous media are much larger than variations in porosity.

3.3.2 Solving the Inverse Problem

In this section, we solve Eq. 3.15 to develop a multi-dimensional posterior distribution from which we develop realizations of both the fine- and coarse-scale fields that are consistent with the data. We first describe how we develop the "ground-truth" binary field (fine-scale), followed by the generation of the synthetic data that serve as observations, the development of the posterior distribution $P(\mathbf{w},\delta|\mathbf{d})$ and finally, by an analysis of the inversion methodology.

3.3.2.1 Development of the "Ground-Truth" Fine-Scale Binary Field

The domain $\mathcal{D}$ is dimensioned as $L_x = 1.5, L_y = 1.0$. The covariance of the ζ field is initialized with $a = 1.0, b = 0.1(L_x^2 + L_y^2)^{1/2}$ (see Eq. 3.8). ζ is modeled as a multiGaussian field and a realization is obtained on a coarse 30×20 mesh of grid-blocks on $\mathcal{D}$. The "true" high-permeability proportions on the coarse mesh, $\mathbf{F}_c^t(\mathbf{x})$, are calculated using the

transformation in Eq. 3.8. Each grid-block is thereafter further refined into 100×100 *grid-cells*, leading to a 3000×2000 grid-cell discretization of $\mathcal{D}$.

We generate a random binary field on the 3000×2000 mesh per the procedure in Sec. 3.2.1 and [83]. We initialize the mesh with white noise and convolve it with a Gaussian kernel with $\sigma = 5.0$ ($\delta = 11.774$) grid-cells. $\mathbf{F}(\mathbf{x})$, the proportion of high-permeability material in each grid-block (equivalent to $\mathbb{E}[N(x)]/S$ in Eq. 3.1), is used to calculate the threshold w (Eq. 8). The $w - level$ excursion set creates the random inclusions in each grid-block. Repeated over all coarse-scale grid-blocks, we obtain the fine-scale binary medium on a 3000×2000 grid. The high and low permeability materials are assigned permeabilities of $\mathcal{K}_h = 100$ and $\mathcal{K}_l = 1$ ($\kappa = 2$).

3.3.2.2 Generation of Synthetic Data

The synthetic data consist of measurements $\mathbf{k}^{(obs)}$ of effective log-permeability of the coarse grid-blocks containing the sensors. Data for two sets of sensors, SSA and SSB (see sensor locations in Figure. 3.2, e and f), are available. The permeabilities are calculated empirically by solving a permeameter boundary condition problem using the true fine-scale permeabilities for each of the coarse grid-blocks. For each grid-block, a time-independent pressure difference is imposed in one direction, and zero-outflow in the perpendicular one to calculate a flow-rate. The resulting flux is used to calculate the effective grid-block permeability in that direction. The directions of pressure difference and zero-outflow are then exchanged to obtain the permeability in the perpendicular direction. The two permeabilities calculated by this procedure are similar and the geometric mean of the permeabilities is taken as the "upscaled" permeability of the binary medium in that grid-block. Flow simulations for each of the grid-blocks in the 30×20 mesh are performed with MODFLOW-2005 [58]. We refer to the resulting log-permeability field as the true, upscaled log-permeability field, $\mathbf{K}_c^t$. Those values that correspond to the sensor grid-blocks in the SSA and SSB sets form the static data, $\mathbf{k}^{(obs)}$. I.i.d. Gaussian observation errors ($\sim \mathsf{N}(0, 0.1)$) are added to them.

The original random binary field, on the 3000×2000 mesh, is then subjected to a tracer transport simulation, as described in Sec. 3.2.3. MODPATH [88] is a Lagrangian particle tracking method operating on flux fields calculated in MODFLOW and is used here to simulate transport through the binary porous medium. A fluid is injected via a cross pattern of five wells within the single coarse-scale grid-block in the lower-left corner and extracted via a similar configuration of pumping wells on the top-right coarse grid-block to create a steady-state flow field. To calculate the breakthrough times, we reverse the steady-state velocity field. Then, for each coarse-scale sensor grid-block, we release 121 particles from a uniformly spaced 11×11 configuration of locations and advect those particles back to the injection wells using the reversed velocity field. The time it takes for half the particles (median of the distribution) released in a particular grid-block to reach the lower-left (injection) grid-block is taken as the breakthrough time. Repeated for each of the sensor grid-blocks, we obtain the dynamic data $\mathbf{t}_b^{(obs)}$. I.i.d. Gaussian observation errors ($\sim \mathsf{N}(0, 10^{-3})$) are

added to them after non-dimensionalization (non dimensional time = (raw time)*(injection rate)/(domain pore volume)). Thus the breakthrough times contain the effect of the fully resolved binary medium on the advective transport dynamics.

3.3.2.3 The Error Model

We assume that along with $\mathbf{k}^{(obs)}$, both $\mathbf{F}^*$ and δ^* too can be measured at the grid-blocks with sensors, allowing us to estimate $\mathbf{K}_e(x)^* = \mathcal{L}(\mathbf{F}^*, \delta^*, \kappa)$ there. μ_k and σ_k are calculated empirically as the mean and standard deviation of $\mathbf{k}^{(obs)} - \mathbf{K}_e(x)^*$ evaluated at the observation points. μ_t is set to zero. The standard deviation for the breakthrough time, σ_t, is set to 5% of the maximum non-dimensional breakthrough time observed in the SSA set of sensors. These values are used in Eq. 3.12 and 3.15.

3.3.2.4 Generation of the Posterior Distribution $P(\mathbf{w}, \delta|\mathbf{d})$

We solve Eq. 3.15 using the log form of δ, i.e. $\ln(\delta)$ which allows us to model its prior using a truncated Gaussian

$$\ln(\delta) \sim \mathsf{N}(\ln(10), 2) \text{ if } 1 < \delta < 100.$$

Outside these limits, the prior is set to zero, thus loosely modeling the inclusions to be larger than a fine-scale grid-cell, but smaller than a (coarse) grid-block. 30 terms were retained in the Karhunen-Loève expansion of the spatial field i.e., $M = 30$ in Eq. 3.13. 10^6 samples were taken using the adaptive MCMC sampler (DRAM, [77]) and the chain was checked for mixing and burn-in using the metric in [89] (as implemented in the `mcgibbsit` package [90] in R [91]) by monitoring the 5^{th} and 95^{th} percentiles as well as the median. 10,000 samples were retained by thinning the chain and this sample set is used for developing posterior distributions of the objects of inference. We checked the impact of thinning by computing the percentiles with a chain thinned to 100,000 samples (instead of 10,000); no significant change was observed.

In Fig. 3.3 we plot the probability density functions (PDF) for w_1, w_{15}, w_{30} and $\ln(\delta)$, by marginalizing over the thinned samples. The three weights, w_1, w_{15}, w_{30}, correspond to Karhunen-Loève modes which are representative of large, medium and small-scale (but resolved) structures. Their joint PDFs are also plotted. We see that the posterior distribution for w_1, w_{15}, and w_{30} are roughly Gaussian, though that does not hold true for $\ln(\delta)$ (which is a truncated Gaussian). The median value of δ is found to be 9.837 grid-cells (compared to the true value of 11.774) with a 90% credibility interval of (1.4–70.11) grid-cells. The inter-quartile range is (3.76–25.6) grid-cells. The approximately Gaussian (posterior) distributions for w_{15} and w_{30} are centered around 1, unlike the priors which are standard normals (centered at 0.0). The scatter plots show that the samples of these four parameters are not correlated with each other; the correlations between the (posterior of the) weights of adjacent Karhunen-Loève modes weaken very quickly and are insignificant for modes which are at least 5 modes apart.

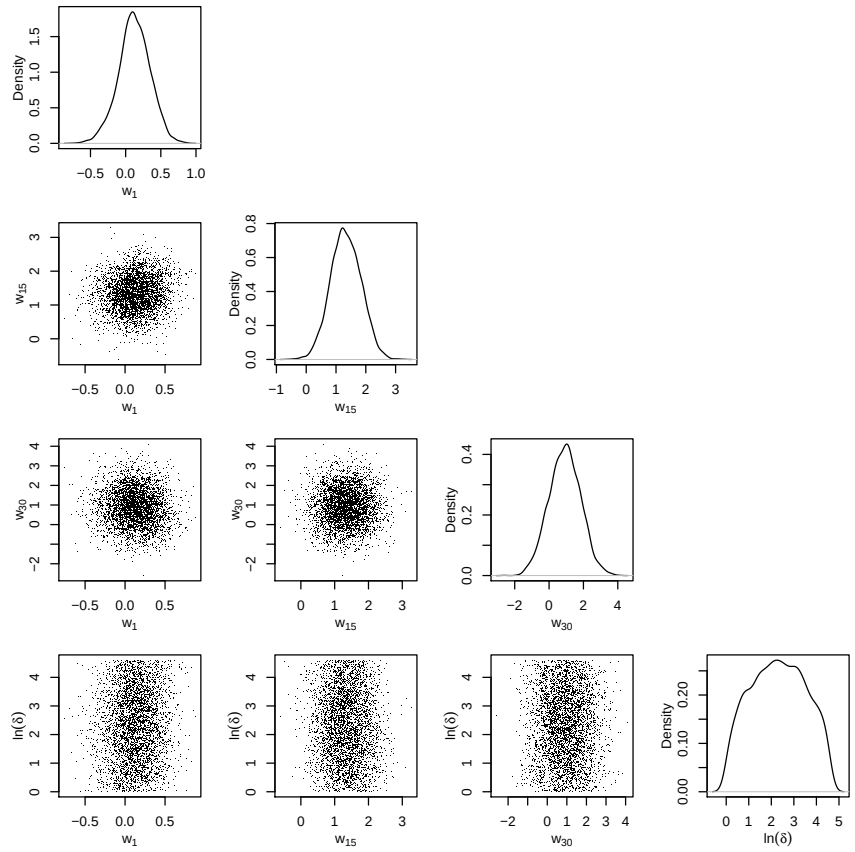


Figure 3.3: Plot of the joint and marginal probability densities of w_1, w_{15}, w_{30} and $\ln(\delta)$ from the posterior distribution. The joint distributions show very little correlations between the Karhunen-Loève modes at the large, medium and small scales, as well as their correlations with δ. Also, PDFs of the individual parameters are roughly Gaussian, with the exception of δ.

3.3.3 Results and Discussion

In this section, we analyze the solution of the inverse problem. We first perform the inversion using both the static and dynamic data, and then repeat using the static and dynamic data individually. These steps are done to explore the contribution of each type of data to the inference.

3.3.3.1 Assessment of $\mathbf{F}(\mathbf{x})$ and $\mathbf{K}_e(x)$ Inferences

We use the samples from $P(\mathbf{w}, \delta|\mathbf{d})$ (Eq. 3.15) to generate realizations of $\mathbf{F}_i(\mathbf{x})$ and $\mathbf{K}_{e,i}, i = 1 \ldots 10,000$. The first 500 samples are discarded as "burn-in". In Fig. 3.4, in the top row, we plot their expected values $\overline{\mathbf{F}} = \mathbb{E}[\mathbf{F}(\mathbf{x})]$ and $\overline{\mathbf{K}_e(x)} = \mathbb{E}[\mathbf{K}_e(x)]$ in color while the "true" values (as plotted in Fig. 3.2, middle row) are plotted as dotted contours. For $\overline{\mathbf{F}}$, we see that the shade plot and the contours are similar with respect to large scale structures,

though there are differences at smaller scales. On the other hand, the $\overline{\mathbf{K}_e(x)}$ plot at the top right shows significant differences between the shade and contour plots. This arises mostly because of the sharp gradients in $\mathbf{K}_e(x)$ and the approximations inherent in the $\mathcal{L}(\mathbf{F}(\mathbf{x}), \delta)$ model for the log-permeability. In the middle row of Fig. 3.4, we plot the grid-block-wise standard deviation of the 9,500 samples of $\mathbf{F}_i(\mathbf{x})$ and $\mathbf{K}_{e,i}$, along with the locations of the sensors in SSA. Standard deviations (the uncertainty in the inferences) reach minimum values at the measurement points and increase in the poorly instrumented regions. The regions with the largest uncertainties are concentrated in regions of high gradients. The bottom row shows the difference between the true and average values i.e. $\varepsilon_F = \mathbf{F}_c^t(\mathbf{x}) - \overline{\mathbf{F}}$ and $\varepsilon_K = \mathbf{K}_c^t - \overline{\mathbf{K}_e(x)}$. Regions of low errors are generally at the sensor locations. In the plot for ε_K, the largest errors correspond strongly with the "wall" of high gradients in $\mathbf{K}_e(x)$ near $x = 25$. The error approaches 50% in this region.

We repeat the experiment above using only the static data $\mathbf{k}^{(obs)}$ and plot the $\overline{\mathbf{F}} = \mathbb{E}[\mathbf{F}(\mathbf{x}))]$ and $\overline{\mathbf{K}_e(x)} = \mathbb{E}[\mathbf{K}_e(x)]$ in Fig. 3.5, top row. Again, 9,500 samples are used. Comparing the shade plot with the true field (plotted as contours), we see that the inversion captures the large structures but not the smaller details. Their equivalents, obtained from inversions that used only the dynamic data $\mathbf{t}_b^{(obs)}$ are plotted in Fig. 3.5, bottom row; they bear little resemblance to the true field. We omit the standard deviations, ε_F and ε_K for both the cases, but they may be found in [92]. The standard deviation for the static-data-only inversion shows much the same behavior as in Fig. 3.4, i.e., they are minimum at sensor points.

3.3.3.2 Assessment of $\{\mathbf{w}, \delta\}$ Inferences

The marginalized PDFs for w_1, w_{15}, w_{30} and $\ln(\delta)$ for all three combinations of conditioning data are shown in Fig. 3.6. We also plot the prior distributions (using symbols) for comparison. The top left figure shows that the observations are most informative about w_1 (i.e., the difference between the prior and posterior is the largest), the Karhunen-Loève mode corresponding to the largest structures. The grid spacing of the sensors in SSA is smaller than the length-scale of the first Karhunen-Loève mode and provides dense sampling of it. Consequently, the distribution of w_1 obtained using just the static data is almost identical to that obtained from using both static and dynamic data. Relative to the static data, the dynamic data contributes very little to the inference of the lower (larger scale) Karhunen-Loève modes - the posterior for w_1 in Fig 3.6 (top left) is barely different from the prior. For intermediate Karhunen-Loève modes e.g., w_{15}, (top right in Fig. 3.6) both the static and dynamic data contribute to the inference, though the contribution of the static data is larger; the posterior developed from static data alone is quite close that developed using $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$. At the small scales i.e., w_{30} (Fig. 3.6, bottom left), the dynamic data contributes little - the posterior distributions obtained from the $\mathbf{k}^{(obs)}$-only and $\mathbf{t}_b^{(obs)}$-only inversions are almost the same as the prior, whereas the inversion conditioned jointly on $(\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)})$ is somewhat informative. Thus, dynamic data is informative, *only after* the larger/coarser scales have been accounted for by the static data; by itself, $\mathbf{t}_b^{(obs)}$ does not have the information content to resolve both scales. Transport occurs preferentially through

high permeability regions, which can be affected by the smallest scales; since breakthrough times are the integrated effect of the travel times of the tracer, one may naively expect that the effect of small scale variations are easily captured there. However, breakthrough times are also affected by the larger Karhunen-Loève modes and deconvolving the impact of the smaller and larger lengthscales cannot be done without $\mathbf{k}^{(obs)}$, which uniquely capture the coarse-scale structures. The inference with the SSA sensors provides very little information about δ, regardless of the kind of data ($(\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)})$ versus $\mathbf{t}_b^{(obs)}$) used. This is because the impact of δ is felt only near the percolation threshold ($\mathbf{F}(\mathbf{x}) \approx 0.5$), which occurs only at a few spots in the entire domain. Consequently, the posterior distribution for $\ln(\delta)$ is almost the same as the prior.

3.3.3.3 Impact of the Number of Sensors

In Fig. 3.7 we compare the marginalized posteriors for w_1, w_{15}, w_{30} and $\ln(\delta)$ as obtained with sensor-sets SSA and SSB, both using the combined static and dynamic data. The second set, which has more than half as many sensors, results in a PDF for w_1 which is sharper than that obtained with SSA, while the PDF for $\ln(\delta)$ is almost unchanged and not very different from the prior. Also, the PDFs for w_{15} and w_{30} change significantly when recomputed using SSB, indicating that the information content of $\mathbf{t}_b^{(obs)}$ with respect to finer scale structures collected by SSA and SSB could be different; the marginals developed using the SSB sensor set are steeper.

In order to analyze the model fits arising from SSA and SSB measurements, we conduct a posterior predictive check (PPC) using $\mathbf{t}_b^{(obs)}$. Note that since inferences were drawn using both $\mathbf{k}^{(obs)}$ and $\mathbf{t}_b^{(obs)}$, good fits with $\mathbf{k}^{(obs)}$ may compensate for bad fits with $\mathbf{t}_b^{(obs)}$. In Fig. 3.8, we plot the results of PPC performed using inferences drawn from both SSA and SSB. We use the realizations $\mathbf{F}_i(\mathbf{x}), i = 500 \ldots 10,000$, to generate the corresponding coarse-scale $\mathbf{K}_{e,i} = \mathcal{L}(\mathbf{F}_i(\mathbf{x}), \delta_i, \kappa)$; thereafter $\mathbf{t}_{b,i} = \mathcal{M}(\mathbf{K}_{e,i})$ at the SSA- and SSB-sets of sensors. We plot the median, the 1^{st} and the 99^{th} percentile of the breakthrough times. The spread of the replicates of $\mathbf{t}_b$ obtained with SSB (right figure) is generally smaller than those obtained with SSA, indicating a reduction in predictive uncertainty.

We next evaluate the predictive ensemble of breakthrough times (plotted for SSA and SSB in Fig. 3.8) quantitatively using the continuous rank probability score (CRPS), the mean absolute error (MAE) and the interval score (IS) [93, 94]. These metrics are defined for each sensor / observation; we report metrics averaged over all sensors (in SSA or SSB). 950 samples from the predictive ensemble (rather than the full 9500 ensemble members) were used for the purpose; the sensitivity of the metrics to the number of samples was checked by repeating the calculations with double and half the number of samples. The interquartile range was used for calculating the IS. We also perform the same experiment at four sensors in a "testing set". The sensors in this set are randomly placed and are distinct from the sensors in SSA and SSB. Further, the observations in the "testing set" sensors are not used in the inversion. In Table 3.1, we tabulate the CRPS, MAE and IS for

the two predictive ensembles in Fig. 3.8 as well as the predictions at the "testing set" of sensors. Results indicate that predictions at the SSB sensors are more accurate (i.e., similar to observations). Further, the inversion performed using SSB data results in a model that has a higher predictive skill at the "testing set" of sensors. Thus the extra observations in the SSB set lead to a more accurate estimate of $\{\mathbf{w}, \delta\}$, which, in turn, results in a more predictive ensemble, as plotted in Fig. 3.8 (right).

Ensemble	Predictions at own sensors			Predictions at "testing set" sensors		
	CRPS	MAE	IS	CRPS	MAE	IS
SSA	0.047	0.0618	0.240	0.054	0.077	0.077
SSB	0.046	0.0588	0.193	0.042	0.056	0.048

Table 3.1: CRPS, MAE, and IS for the predictive ensembles of breakthrough times developed from the SSA and SSB sensor sets (see Fig. 3.8). CRPS, MAE and IS have units of breakthrough times i.e., they are non-dimensional. The left half of the table contains metrics evaluated at the locations of the sensors themselves; the right half contains metrics that evaluate the predictive skill of the model, calibrated using SSA or SSB observations, at the "testing set" of sensors. Both sets of metrics show that inversion using the SSB set of observations leads to a more accurate predictive model.

Finally, we address the question of whether the change in the posterior density of w_{15} and w_{30} in Fig. 3.6, when the inference is conducted using $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ (vis-à-vis the estimation using just $\{\mathbf{k}^{(obs)}\}$), is due to the larger number of observations that are used in the former or due to the different *types* of information in $\mathbf{k}^{(obs)}$ and $\mathbf{t}_b^{(obs)}$. In Fig. 3.9, we plot the posterior densities of w_{15} and w_{30} as obtained using the SSA and SSB sets of sensors, computed using just $\{\mathbf{k}^{(obs)}\}$ observations (i.e., static-data only) as well as conditioned jointly on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ (i.e., static and dynamic data). If the difference in the estimation of w_{15} using $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ versus $\{\mathbf{k}^{(obs)}\}$ (both obtained from SSA sensors) had been due to the number of observations ($\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ has 20 observations of each type), then the posterior distribution computed using just the static data obtained from the SSB sensor set (34 observations) should be similar to that computed from the 40 $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ observations from the SSA set (20 observations each of$\{\mathbf{k}^{(obs)}\}$ and $\{\mathbf{t}_b^{(obs)}\}$). However, Fig. 3.9 (left) shows quite the opposite trend, as the posterior computed from just the static data from the SSB sensors set *moves away* from that computed using $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ from the SSA sensor set. Further, since the larger number of observations in SSB resolve w_{15} better (vis-à-vis SSA), the static-data only estimate of w_{15} is quite close to the posterior obtained from using both $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ from the SSB sensors. Thus, the difference in the posterior distributions is not due to the extra observations in $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$; rather it is due to the different *types* of information in $\{\mathbf{k}^{(obs)}\}$ and $\{\mathbf{t}_b^{(obs)}\}$. In Fig. 3.9 (right), we see that the posterior densities of w_{30} obtained using just the static data are rather similar, regardless of whether the SSA or the SSB sensors were used; further, they are both very similar to the

prior. This is because neither of the two sensor sets are dense enough to resolve w_{30}. However, when $\mathbf{t}_b^{(obs)}$ are used in the inference, the posterior densities differ from those obtained using just the static data, particularly for the SSB sensors. Furthermore, we do not see a trend; had the posterior distribution of w_{30} been solely dependent on the number of observations, the static-only estimate from SSB should have occupied a position between the static-only estimate of w_{30} drawn from SSA sensor set and the SSA estimate conditioned on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$. Clearly, this is not the case in Fig. 3.9 (right).

3.3.4 Summary

To summarize, we have developed an inference scheme, predicated on a multiscale link function, to infer coarse-scale features and summaries of fine-scale structures of a random binary field from a combination of static and dynamic observations. We can estimate a spatially variable proportion $\mathbf{F}(\mathbf{x})$ of the high-permeability material from observations that are obtained on the coarse-scale but are nevertheless informative about the fine-scale. We see that these observations are not very informative regarding δ; its posterior distribution is not very different from the prior. This is because the impact of δ is small (if it is much smaller than a grid block) and is limited to regions were $\mathbf{F}(\mathbf{x}) \approx 0.5$. We see that the static data can inform on the large-scale features mainly because the distribution of sensors is sufficient to resolve such structures. The dynamic data, which is an integrated measure of the effect of small variations in the permeability field is key to estimating smaller (but nevertheless resolved) structures. Increasing the number of sensors leads to estimates with lower uncertainties.

We have also explored how the static ($\mathbf{k}^{(obs)}$) and dynamic ($\mathbf{t}_b^{(obs)}$) data contribute to the estimation of $\{\mathbf{w}, \ln(\delta)\}$. By itself, $\mathbf{t}_b^{(obs)}$ is not sufficiently informative to estimate both the large-scale structures and the smaller details. However, if the large-scale structures are constrained/estimated using $\mathbf{k}^{(obs)}$, the smaller details can be inferred from $\mathbf{t}_b^{(obs)}$. Thus joint inversions on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ result in more accurate estimations of $\{\mathbf{w}, \ln(\delta)\}$ not only because of the larger number of observations involved, but rather because of the *type* of information. Since $\mathbf{k}^{(obs)}$ is informative about the larger lengthscales in the domain while $\mathbf{t}_b^{(obs)}$ is impacted most by the smaller scales that contort the flowpaths in the porous medium, joint inversion conditions the estimates to *multiscale* data.

In Sec. 3.5, we explore the robustness of the inference. Since the aim of reconstructing a permeability field is generally to use it to predict transport phenomena, we will subject the reconstructions conditioned on multiscale data (as well as those obtained individually from static and dynamic data) to posterior predictive tests.

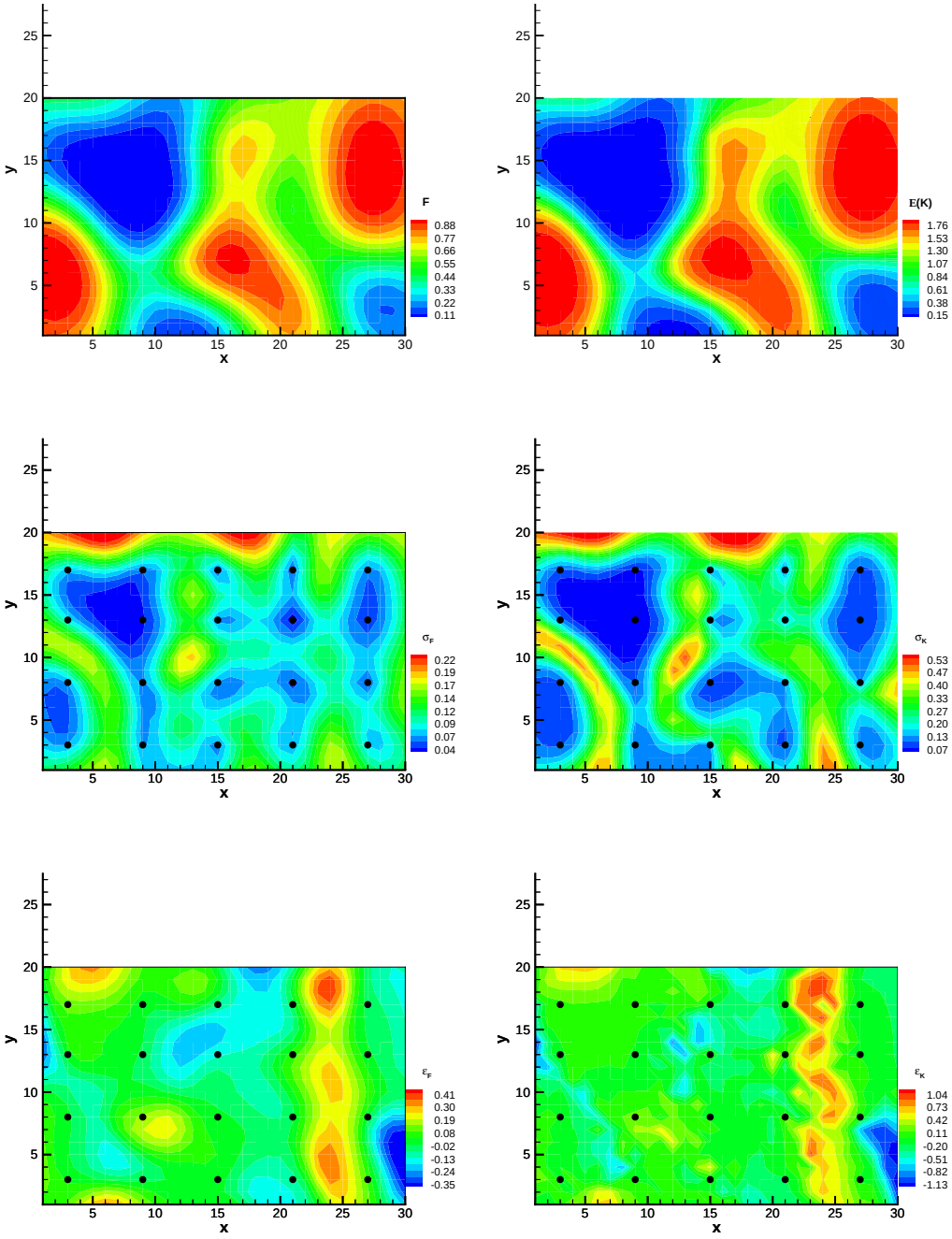


Figure 3.4: Results from the inference, computed using 9500 samples from the posterior distribution. In the top row, we plot the means $\overline{\mathbf{F}} = \mathbb{E}[\mathbf{F}(\mathbf{x})]$ and $\overline{\mathbf{K}_e(x)} = \mathbb{E}[\mathbf{K}_e(x)]$; they bear a strong resemblance to the true values in Fig. 3.2 (middle row) and also plotted as overlaid contours. In the middle row, we plot the pointwise standard deviation of $\mathbf{F}_i(\mathbf{x})$ and $\mathbf{K}_{e,i}$; they are smallest at the sensor locations. In the bottom row, we plot the errors $\varepsilon_F = \mathbf{F}_c^t(\mathbf{x}) - \overline{\mathbf{F}}$ and $\varepsilon_K = \mathbf{K}_c^t - \overline{\mathbf{K}_e(x)}$, which show large values in regions of high gradients.

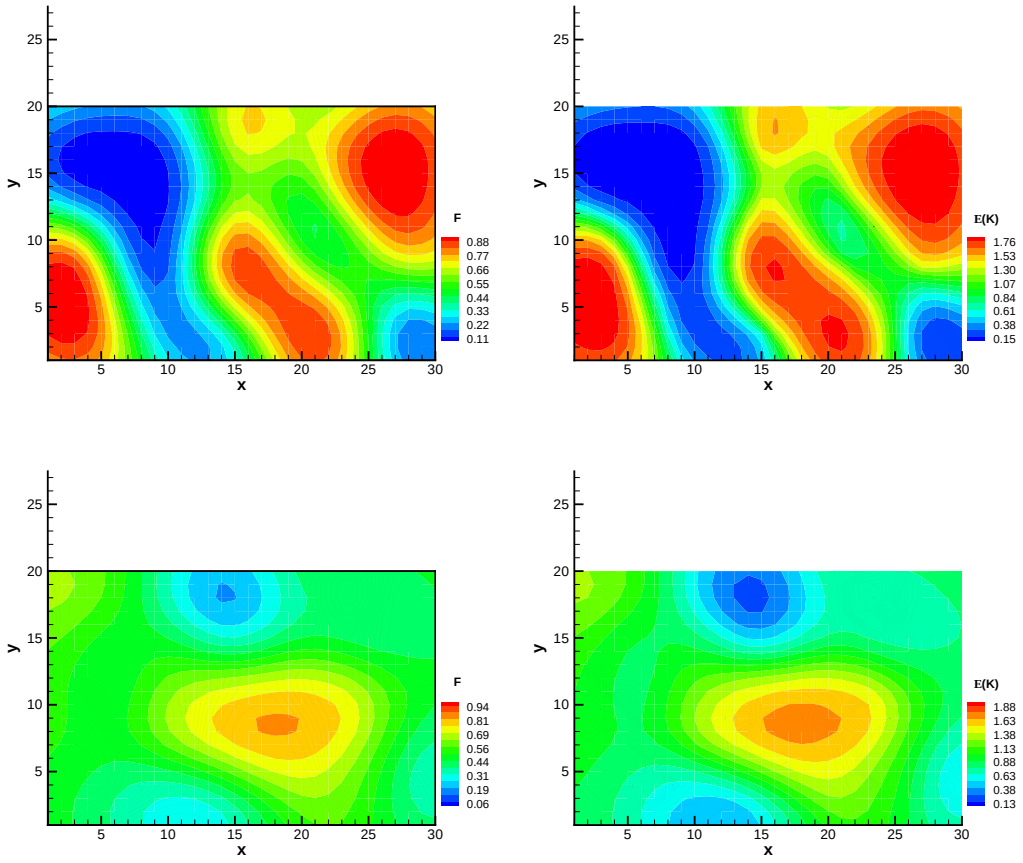


Figure 3.5: Top: Results (the means $\overline{\mathbf{F}}$ and $\overline{\mathbf{K}_e(x)}$) from an inversion using only static data. The true values are plotted as contours. Bottom: Their counterparts, calculated using only dynamic data. Computations were done using 9,500 samples as in Fig. 3.4. We see that the static-data-only inversion in the top row are slightly worse than the results in Fig. 3.4; they miss the finer details and are overly smooth. The inversions in the bottom row bear little resemblance to the true $\mathbf{F}_c^t(\mathbf{x})$ and $\mathbf{K}_c^t$ plotted as contours.

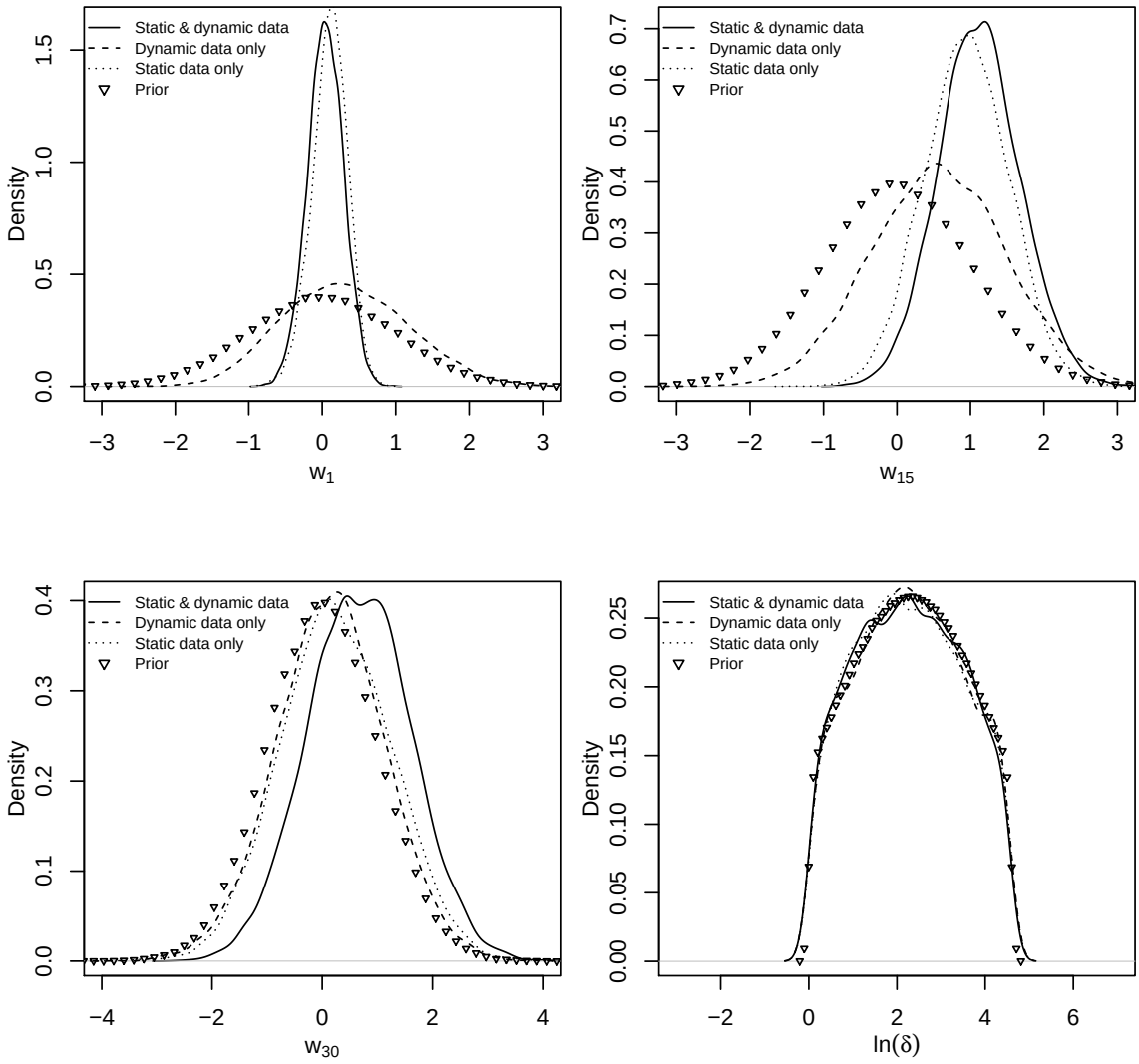


Figure 3.6: Marginalized PDFs of w_1, w_{15}, w_{30} and $\ln(\delta)$ as inferred from $\mathbf{k}^{(obs)}$-only (dotted lines), $\mathbf{t}_b^{(obs)}$-only (dashed lines) and jointly from $(\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)})$ (solid lines). The priors are plotted with ∇ for comparison. Note that the prior and posterior densities for $\ln(\delta)$, plotted in the bottom right figure are truncated at $\ln(\delta) = 0, 4.6$ but the kernel density estimates used to create plots smooth them near the truncation limits. The true values of w_1, w_{15} and w_{30} are 0.127, 0.814 and 1.607 respectively.

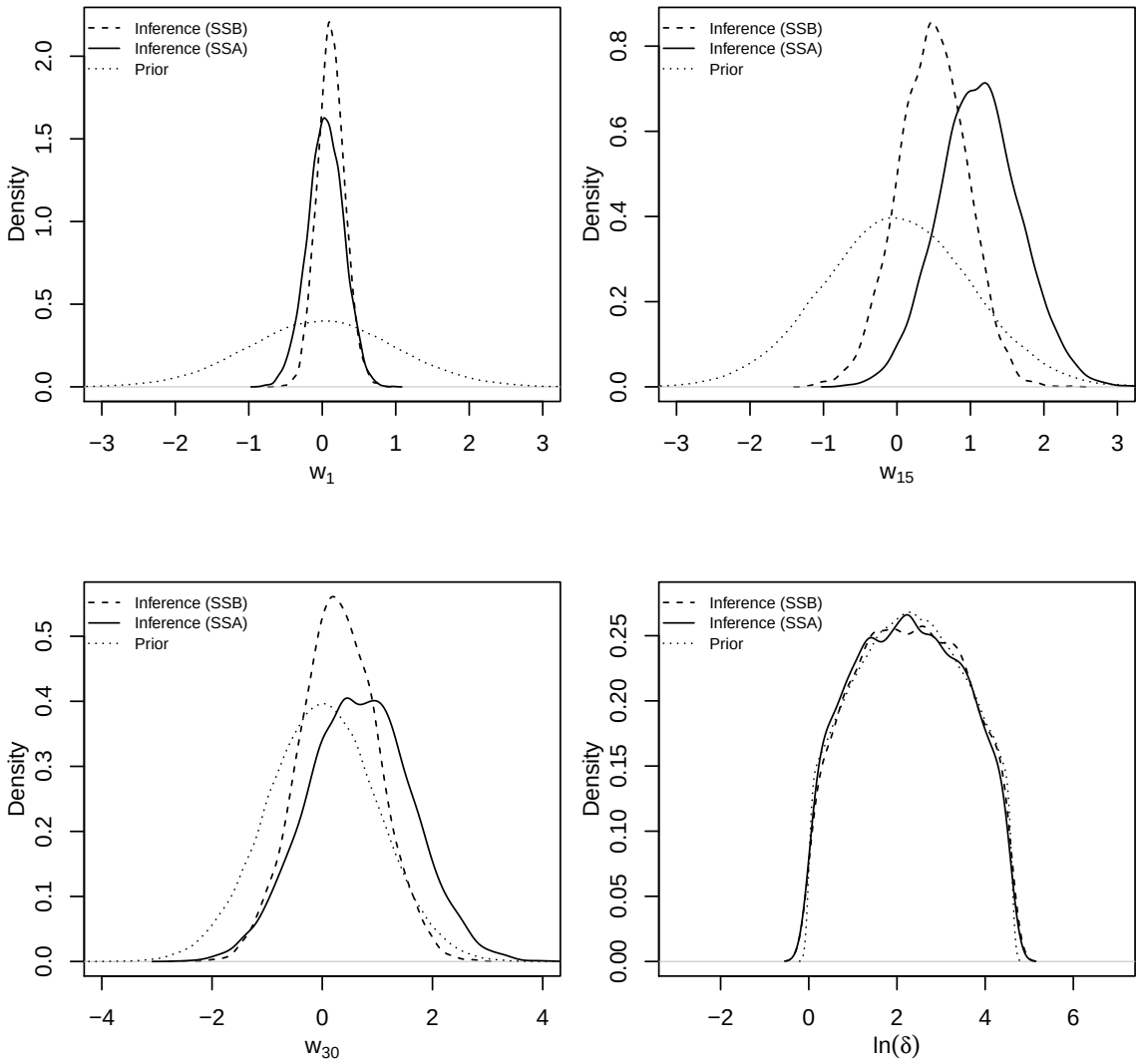


Figure 3.7: Prior and marginalized posterior values of the Karhunen-Loève mode weights w_1, w_{15}, w_{30} and $\ln(\delta)$ as computed using the sensor-sets SSA (solid line) and SSB (dashed line). We see that the posteriors for w_1 are similar, indicating that the higher sensor density of SSB collects little extra information on them. However, the PDFs for w_{15} and w_{30} are quite different, indicating that the uncertainty in their values may have been underestimated in both cases. The posterior density for $\ln(\delta)$ (bottom right) obtained from either set of sensors is not very different from the priors. The true values of w_1, w_{15} and w_{30} are 0.127, 0.814 and 1.607 respectively.

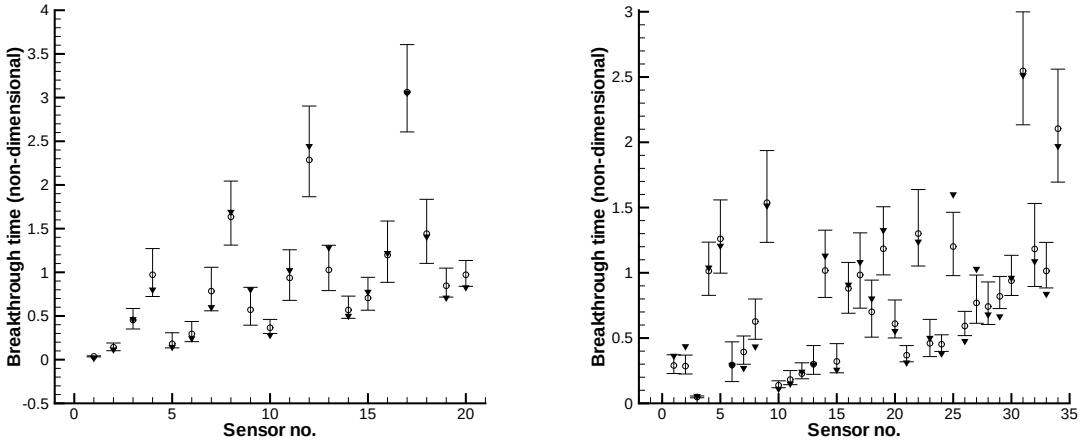


Figure 3.8: Results from the posterior predictive check for breakthrough times conducted by generating coarse-scale $\mathbf{K}_{e,i}$ and using $\mathcal{M}(\mathbf{K})$ to generate the breakthrough times. The filled, inverted triangles are the observations at the sensors. The median breakthrough times from the posterior predictive checks are plotted with open symbols and the error bars denote the 1^{st} and 99^{th} percentiles of the breakthrough time distribution. Results for SSA observations (Left) and SSB observations (Right).

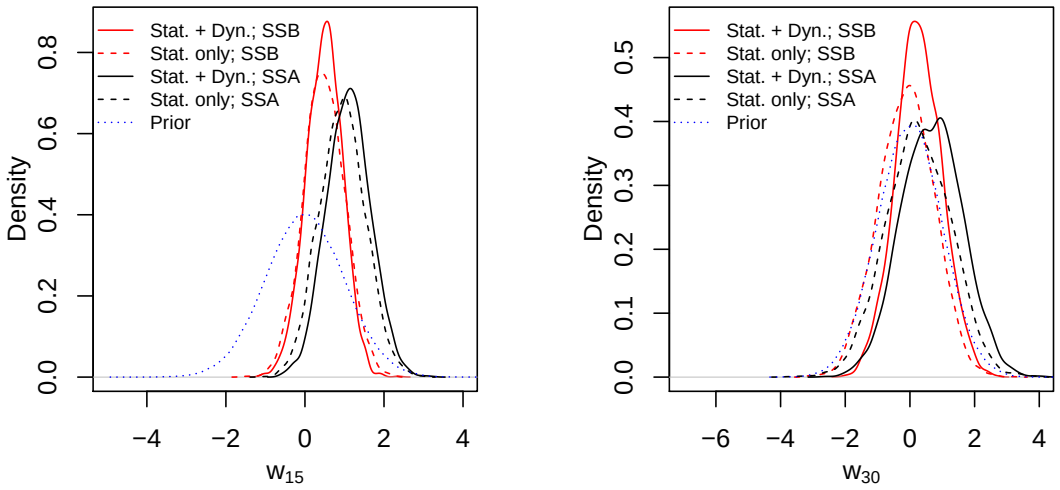


Figure 3.9: Posterior densities of w_{15} and w_{30} obtained using just the static observations $\mathbf{k}^{(obs)}$ (dashed lines) as well as static and dynamic data $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ (solid lines). The priors are plotted with dots. Estimates drawn from SSA sensors are in black; their SSB counterparts are in red. Left: The w_{15} density drawn solely from SSB $\mathbf{k}^{(obs)}$ (34 observations) is quite different from the density obtained from SSA observations (of any type). Instead, it shows a marked similarity to the posterior distribution obtained from SSB $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$. Right: The w_{30} distributions obtained from static data only are very similar to the prior, regardless of the sensor set; the distribution becomes informative (different from the prior) only when $\mathbf{t}_b^{(obs)}$ are included. Thus, it is the *type* of data, rather than the quantity, that renders the posterior densities informative. The true values of w_{15} and w_{30} are 0.814 and 1.607 respectively.

3.4 Estimation of δ

In Sec. 3.3, we addressed the estimation of the Karhunen-Loève modes describing the spatial distribution and δ, the size of the inclusions. While the weights of the KL modes could be estimated (the prior and posterior distributions of the weights were different), the prior and posterior distributions for δ were almost identical.

The reason why the observations carried little information regarding δ can be seen in Fig. 3.10 (a). We see that for small δ ($\delta/\Delta \leq 0.3$, where the model described in Chp. 2 is valid), the two components of the binary medium have to be present in roughly equal quantities ($0.4 < \mathbf{F} < 0.6$) for the medium to be sensitive to the size of the inclusions. In a mesh where $\mathbf{F}$ varied from 0 to 1, only a small fraction of the grid-blocks would display effective permeabilities that were informative on δ; in the limit of sparse observations, the effect of these grid-blocks on the estimate of δ could be missed. In Fig. 3.10 (b), we show a histogram of the $\mathbf{F}(\mathbf{x})$ in the grid; less than a sixth of the grid-blocks have $\mathbf{F}$ that lie in the limits mentioned above.

We investigate further whether δ can be estimated under conditions where the domain's permeability is more sensitive to δ. We do so by creating a new proportionality field ($\mathbf{F}^*$) parametrically from the old one ($\mathbf{F}$)

$$\mathbf{F}^* = 0.5 + \gamma(\mathbf{F} - 0.5).$$

Note that γ describes the range, centered at 0.5, of $\mathbf{F}^*$, the proportion of high permeability material. We generate two such domains, with $\gamma = 0.2, 0.5$, whose distributions (histograms) are plotted in Fig. 3.10 (c) and (d). We see that in the first case, the entire domain has its high-permeability proportion in the range [0.4, 0.6], with the effective permeability in all grid-blocks sensitive to δ. In the second case, only a fraction of the grid-blocks lie within the [0.4, 0.6] range. Note that as the range of $\mathbf{F}^*$ shrinks, the domain becomes more homogeneous and shows only slow, large scale variations in $\mathbf{K}$.

We then subject the binary medium, generated using $\gamma = 0.2$, to the same inversion as described in Sec. 3.3 with SSA set of sensors. Inversions were done with $\{\mathbf{k}^{(obs)}\}$ and then with $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$. The results are in Fig. 3.11. We see that the lack of small-scale spatial structure in $\mathbf{F}^*$ (and consequently $\mathbf{K}_e(x)$) allows the static observations to be very informative. Since the medium is so spatially homogeneous (and has little in the nature of higher Karhunen-Loève modes), the ability of $\{\mathbf{t}_b^{(obs)}\}$ to capture the fine-scale structure of a domain is not put to much use; thus inferences with $\{\mathbf{k}^{(obs)}\}$ and $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ are very similar. Further, in Fig. 3.11, we plot the PDF of $\ln(\delta)$; we see it is close to the true figure of 2.45. However, there is a distinct bias in the estimate. This is due to the error in the subgrid model for $\mathbf{K}_e(x)$. As seen in Fig. 3.1, for a given δ, the subgrid model underestimates permeability; within the context of inversion, the subgrid model $\mathcal{L}$ compensates for its under-prediction error by estimating a larger δ.

We now consider the case of $\gamma = 0.5$ i.e., $0.25 \leq \mathbf{F}^* \leq 0.75$, which show a spatial variation with a richer range of lengthscales. The same inversion as above is performed for 3 sensor

sets – SSA (20 sensors), SSB (34 sensors) and SSE (54 sensors, Fig. 3.12). The estimates of $\mathbf{w}$ and δ obtained with $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ are in Fig. 3.13. We see that the SSA set of sensors provide a good estimate of the MAP (maximum a posteriori) value of w_1; adding more sensors narrows the PDF. w_{15} is not well estimated with the SSA sensors, and adding more sensors leads to a shift in the PDF. For w_{30}, we see the PDF narrow, but it is unclear whether this is due to static or dynamic data (a question that we address in detail below). The estimates of δ show no trend (with increasing sensors) and are far from the true figure.

We next investigate how informative the $\{\mathbf{k}^{(obs)}\}$ are, versus $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$, in the inference of the $\gamma = 0.5$ cases. In Fig. 3.14, we plot the PDFs for w_{30} and $\ln(\delta)$ as obtained using $\{\mathbf{k}^{(obs)}\}$ and $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$. We see that the contribution of the dynamic data collected with the SSA sensor set is small (the posterior distributions constructed with $\{\mathbf{k}^{(obs)}\}$ and $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ are similar, and almost the same as the prior). However, with more sensors, the posterior distributions become markedly different from the prior (which is expected), but the contribution of the dynamic data is much more than the static data, reinforcing the finding in Sec. 3.3.3 that the dynamic data contains information about the small scale structures that the static observations do not have. In the case of δ, we see much the same behavior i.e., the contribution of the dynamic data increases with the number of sensors, but the sequence of PDFs obtained with SSA, SSB and SSE sensors do not display a trend, nor do they get closer to the truth. We conjecture that the impact of δ on the permeability in the $\gamma = 0.5$ case is small enough that it is masked by the model errors/shortcomings of the link function $\mathcal{L}$ (this conjecture is further investigated below). We check the accuracy of the inference done with SSA, SSB and SSE sensors by conducting posterior predictive tests and calculating the CRPS and MAE at the sensors and the "testing set" of sensors. These are tabulated in Table 3.2. We see that there is little difference in the scores calculated from the posterior predictive tests of the breakthrough times at the sensors for the SSA and SSB sensor sets, but a clear improvement when the SSE sensor set is used. With regard to the "testing set" sensors, there is no clear trend. This is because the accuracy of breakthrough time predictions at the testing set sensors affected by (1) the accuracy of the inferred permeability field and (2) the proximity of the testing set locations to a sensor where an observation may be available. As the sensor density increases, both the contributors to predictive accuracy change, but it is unclear which of the two causes play a greater part in the change in CRPS and MAE. Thus, the CRPS and MAE scores at the testing set do not show a trend and are not a good indicator of the accuracy of the estimated permeability field.

We finally check the impact of dynamic data on the estimation of δ. We argue that δ cannot be inferred from the dynamic observations because its impact on breakthrough times is small; consequently, if the samples from the posterior distribution had their δ component replaced by draws from the prior, and used in a predictive test for the breakthrough time, their accuracy would not degrade significantly. We henceforth refer to this predictive test for breakthrough times as the "perturbed" predictive test. In Fig. 3.15, we plot the median, 1st and 99th percentiles for the breakthrough times calculated from a posterior

Table 3.2: CRPS and MAE for the predictive ensembles of breakthrough times developed from the SSA,SSB and SSE sensor sets (see Fig. 3.8 and 3.12) for the $\gamma = 0.5$ case. CRPS and MAE have units of breakthrough times i.e., they are non-dimensional. The left half of the table contains metrics evaluated at the locations of the sensors themselves; the right half contains metrics that evaluate the predictive skill of the model, calibrated using SSA, SSB or SSE observations, at the "testing set" of sensors. We see that the accuracy of the predictions at the SSA and SSB sets of sensors are about the same, but improved (smaller) for SSE. Predictions for the "testing set" sensors do not show any systematic behavior for reasons explained in the text.

Ensemble	Predictions at own sensors		Predictions at "testing set" sensors	
	CRPS	MAE	CRPS	MAE
SSA	0.040	0.048	0.125	0.124
SSB	0.042	0.051	0.066	0.081
SSE	0.036	0.044	0.110	0.121

predictive test (filled symbols) and its "perturbed" counterpart (open symbols) where $\ln(\delta)$ in the posterior distribution is replaced by draws from the prior. Calculations are for SSA sensors and $\gamma = 0.5$. Observations are plotted in red for comparison. We see that both the tests have similar medians but the "perturbed" predictive tests are more dispersed. This is particularly true for sensors which are far from the injection point (generally sensors with higher indices). This is because long flowpaths allow the accumulation of inaccuracies introduced by replacing $\ln(\delta)$ in the posterior distribution by draws from the prior, causing larger errors which are easily visible in the figure. However, the figure tells us little about the actual distribution of breakthrough times from the two predictive tests. Consequently, we tabulate the CRPS and MAE scores for the "perturbed" predictive test (for SSA, as well as SSB and SSE) set of sensors in Table 3.3. Comparing with Table 3.2, we see that replacing $\ln(\delta)$ in the posterior distribution by draws from the prior hardly made any difference in the predicted breakthrough times at the sensors (for any of the sensor sets), underscoring the weak effect of δ on breakthrough times. Thus, while "perturbing" $\ln(\delta)$ in the posterior does lead to more a dispersed breakthrough-time distribution (Fig. 3.15), the distributions are, in fact, quite similar.

The lack of sensitivity of the breakthrough times to $\ln(\delta)$ also provides an explanation of the lack of any trend in the PDFs of $\ln(\delta)$ in Fig. 3.14. While the PDFs obtained with $\{\mathbf{k}^{(obs)}\}$ and $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ are different, the discrepancy between them is due to the model errors/structural errors between the flow model employed for the inference and MODFLOW, which was used to generate the synthetic observations. The dependence of the PDF on the model errors, rather than on the essential physics of the problem, results in a lack of any trend in the inferences as more observations are added. However, the "perturbed" posterior predictive test can be used as an easy check to ascertain (1) whether observations are sensitive to an inferred quantity and (2) the relative contribution of model errors (vis-à-vis the

Table 3.3: CRPS and MAE of breakthrough times at the sensors, calculated by replacing $\ln(\delta)$ in the posterior with draws from the prior. These number are averages calculated over all the sensors in the sensor sets. Compared to Table. 3.2, we see that "perturbing" $\ln(\delta)$ in the posterior has hardly any effect on the breakthrough times at the sensors.

Sensor set	CRPS	MAE
SSA	0.039	0.047
SSB	0.041	0.052
SSE	0.035	0.044

information content in sparse observations) in calculating the posterior distribution.

To conclude, we can infer δ in some cases where the proportion of inclusions is such that the permeability is sensitive to δ over large portions of the domain. However, for the case of $\delta/\Delta \ll 1$, such a situation is rare (e.g., the $\gamma = 0.2$ case); in most cases, the impact of δ on the permeability is very small. When δ is inferred in such a case, the PDFs obtained do not show any systematic trend as the number of sensors is increased; this is because the PDFs are affected more by the model errors rather than the data. The sensitivity of the dynamic data to δ can be gauged by performing a posterior predictive test, where δ in the posterior is replaced by draws from the prior and comparing the distribution of breakthrough times at the sensors with that obtained from a conventional posterior predictive test. Using CRPS and MAE, we identified that the two distributions were very similar, indicating their insensitivity to δ. Thus δ cannot be inferred with any degree of accuracy; furthermore, the ramifications of a wrong estimation on the predictive capability of the inferred permeability field (with respect to the breakthrough time) is also small. Increasing the number of sensors does not help significantly in this regard.

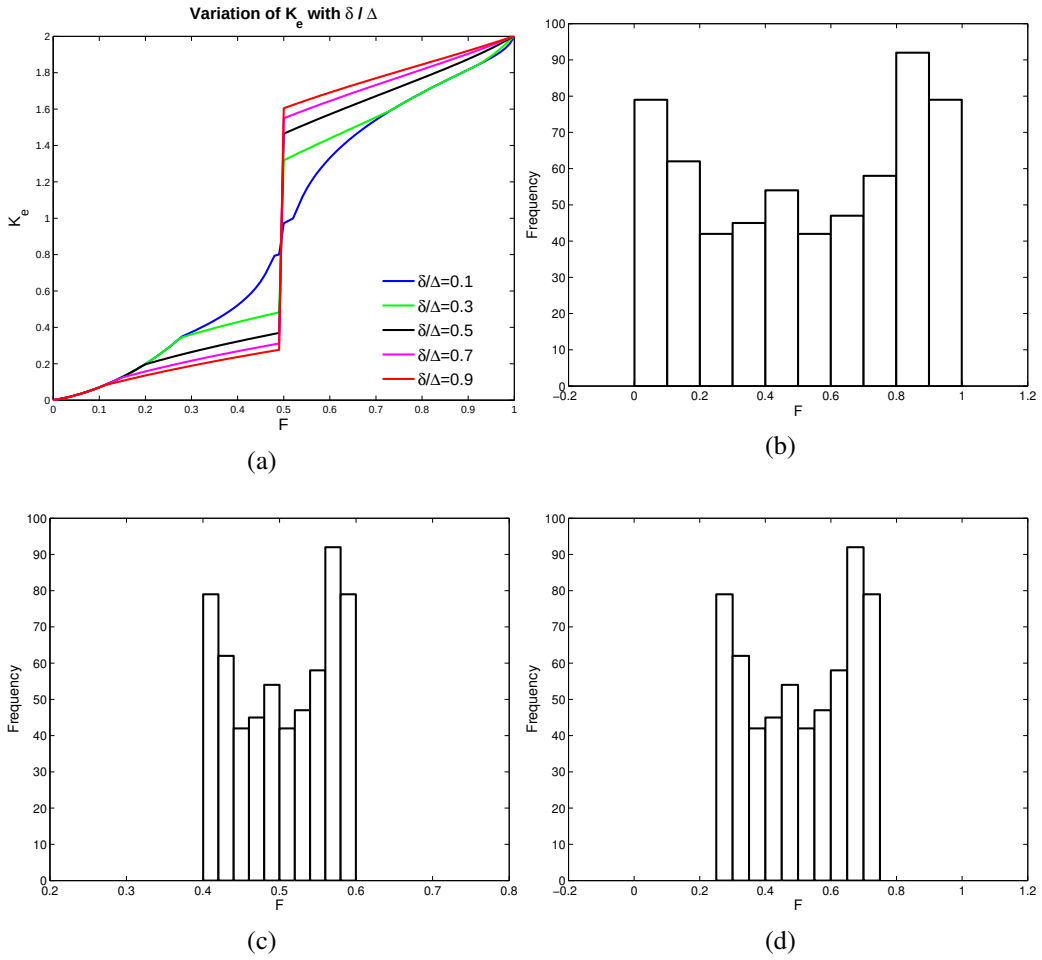


Figure 3.10: Top left: $\mathbf{K}_e(x)$ versus $\mathbf{F}$ for a set of δ/Δ. We see that for small inclusion sizes i.e. $\delta/\Delta < 0.3$, the permeability $\mathbf{K}_e(x)$ is affected by δ only if $0.4 \leq \mathbf{F} \leq 0.6$. Top right: The distribution of $\mathbf{F}$ in the entire domain (all the 600 grid-boxes) for the problem addressed in Sec. 3.3. Only about a sixth of the grid-boxes are affected by δ. Below, we plot the modified proportionality $\mathbf{F}^*$ for cases corresponding to $\gamma = 0.2$ (c) and $\gamma = 0.5$ (d). γ indicates the range of $\mathbf{F}^*$, centered at 0.5. Note that both these cases a more spatially homogeneous compared to the problem in Sec. 3.3.

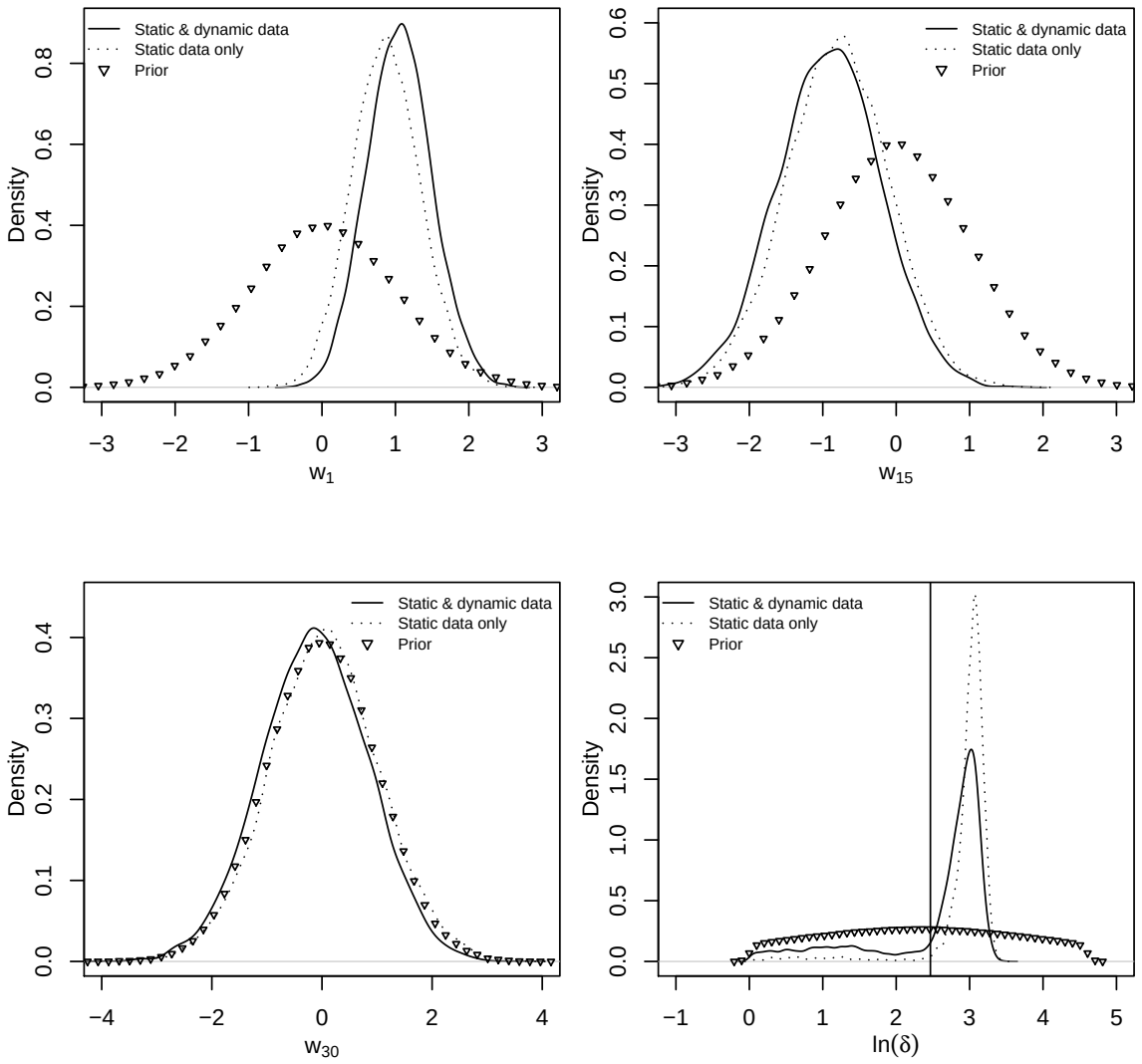


Figure 3.11: Marginalized PDFs of w_1, w_{15}, w_{30} and $\ln(\delta)$ as inferred from $\mathbf{k}^{(obs)}$-only (dotted lines), and jointly from $(\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)})$ (solid lines). The priors are plotted with ∇ for comparison. $\gamma = 0.2$. Note that the prior and posterior densities for $\ln(\delta)$, plotted in the bottom right figure are truncated at $\ln(\delta) = 0, 4.6$ but the kernel density estimates used to create plots smooth them near the truncation limits. We see that the dynamic data contributes little; the rather homogeneous medium and the lack of small-scale structure allows an accurate estimation from the static data. However, δ is estimated with little uncertainty, but a distinct (overestimate) bias. The true value of $\ln(\delta)$ is shown by the vertical line.

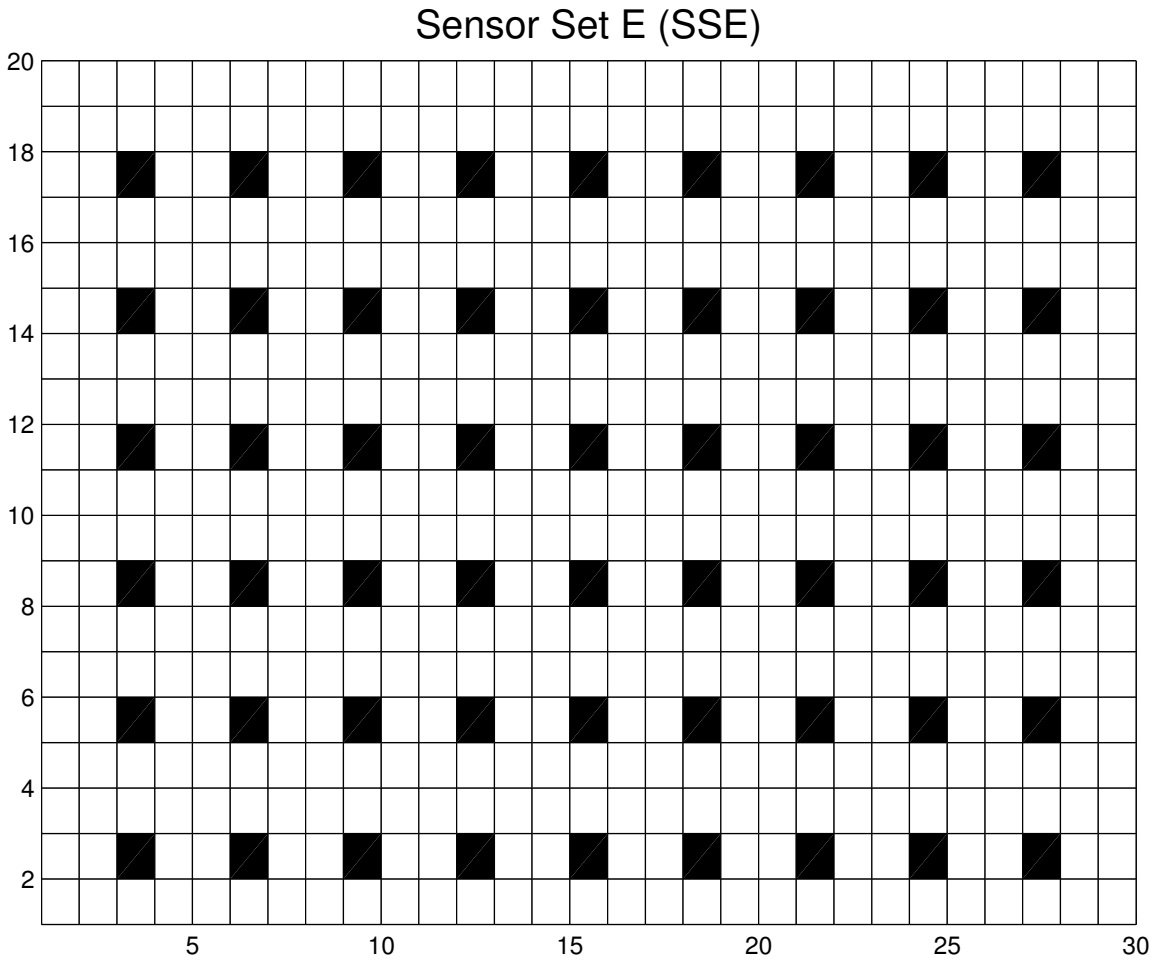


Figure 3.12: The SSE set of sensors, with a total of 54 sensors.

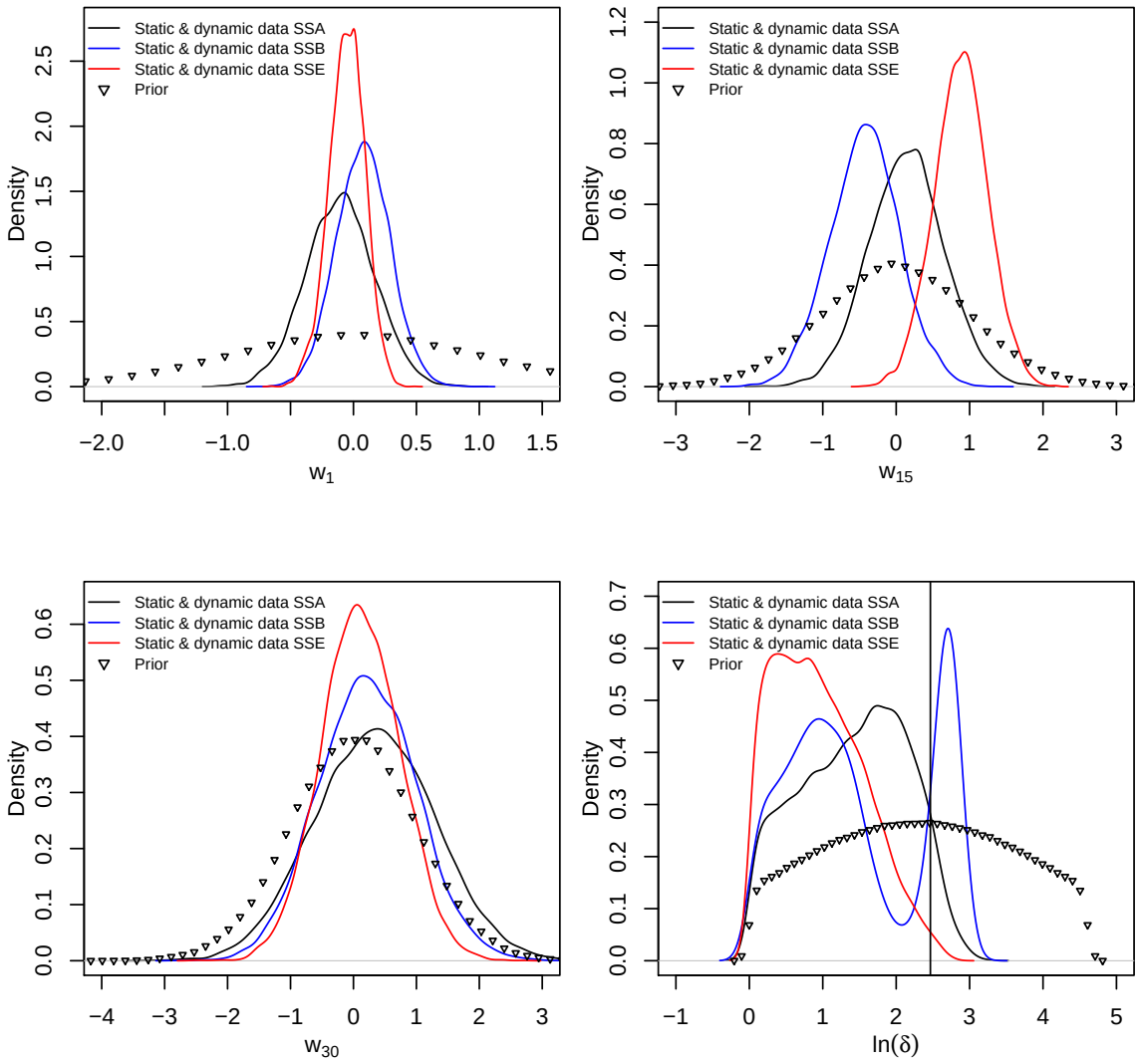


Figure 3.13: Marginalized PDFs of w_1, w_{15}, w_{30} and $\ln(\delta)$ as inferred jointly from $(\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)})$ for SSA (black line), SSB (blue) and SSE (red) . The priors are plotted with ∇ for comparison. $\gamma = 0.5$. Note that the prior and posterior densities for $\ln(\delta)$, plotted in the bottom right figure are truncated at $\ln(\delta) = 0, 4.6$ but the kernel density estimates used to create plots smooth them near the truncation limits. We see that the SSA sensors are sufficient to estimate w_1; i.e., adding more sensors makes a small improvement in the PDF. The PDF for w_{15} changes significantly with the number of sensors, while that for w_{30} does not. The PDFs for $\ln(\delta)$ show no trend. The true values of w_1, w_{15} and w_{30} are 0.127, 0.814 and 1.607 respectively.

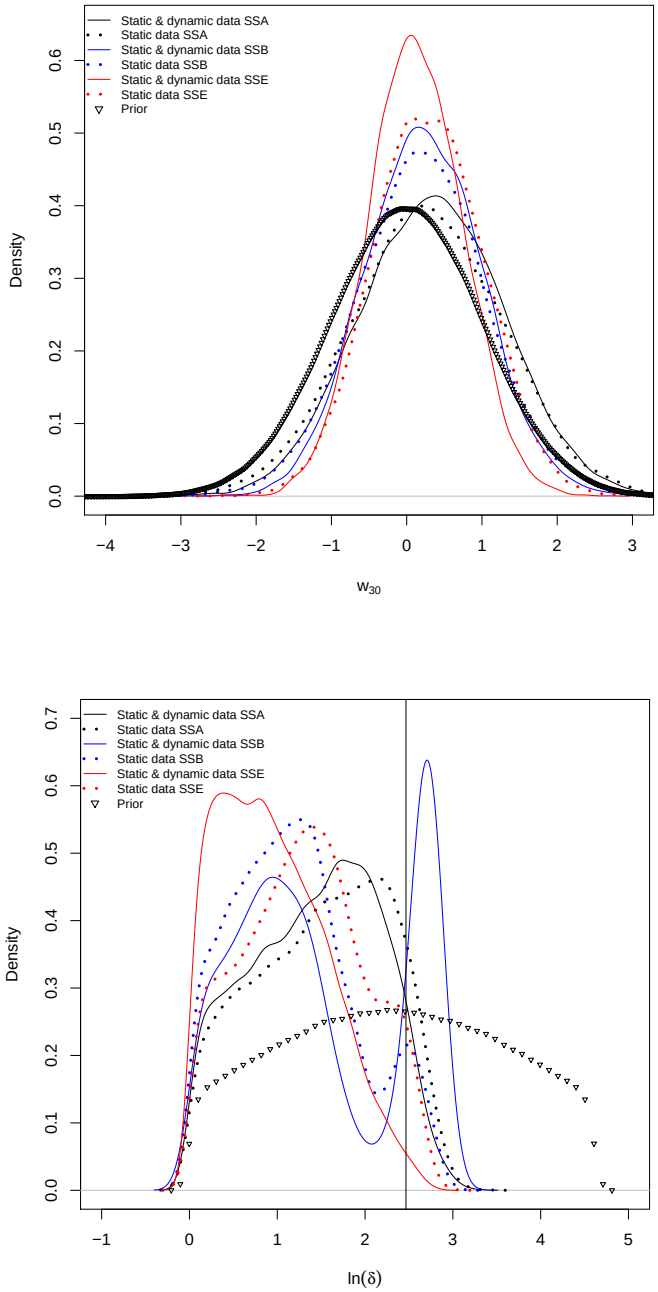


Figure 3.14: Estimates of w_{30} (above) and $\ln(\delta)$ (below) inferred jointly from $(\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)})$ (solid lines) and from $\mathbf{k}^{(obs)}$-only (dotted), for SSA (black), SSB (blue) and SSE (red) sensor sets. We see that the dynamic data $\mathbf{t}_b^{(obs)}$ results in a sharpening of the PDF; further, as more sensors are brought to bear, the sharper the PDF gets. In addition, the contribution of the dynamic data to sharpening the PDF increases with the number of sensors. In case of $\ln(\delta)$, the contribution of the dynamic data increases with the number of sensors, but the results do not get closer to the truth, or even follow a trend, as more sensors are added. The true value of $\ln(\delta)$ is plotted with the vertical line. $\gamma = 0.5$. The true values of w_{30} is 1.607.

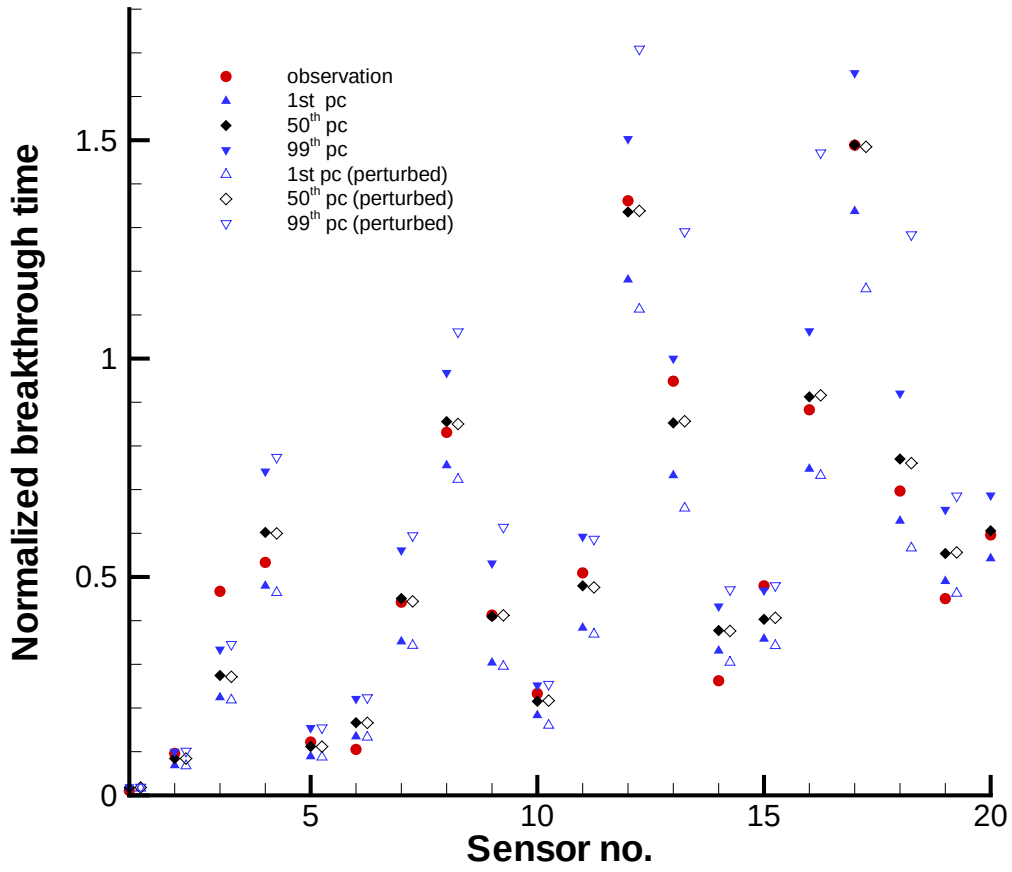


Figure 3.15: Comparison of the median, 1st and 99th percentiles for the breakthrough times calculated the posterior predictive test (filled symbols) and its "perturbed" counterpart (open symbols) where $\ln(\delta)$ in the posterior distribution is replaced by draws from the prior. Calculations are for SSA sensors and $\gamma = 0.5$. Observations are in red. The y-axis denotes the normalized (non-dimensional) breakthrough times. We see that both the tests have similar medians but the "perturbed" posterior predictive tests are more dispersed. This is particularly true for sensors which are far from the injection point, where the inaccuracies introduced by simply sampling $\ln(\delta)$ from the prior provide a large integrated impact. The sensor indices for the "perturbed" test have been shifted on the horizontal axis so that they may be easily compared.

3.5 Posterior Predictive Modeling

The discussion in Sec. 3.3 focused on the probabilistic reconstruction of the proportion of high permeability $\mathbf{F}(\mathbf{x})$ and a length scale δ on a coarse 30×20 mesh. However, the link function $\mathcal{L}$ allows us to also construct realizations of the fine-scale binary medium, given $\{\mathbf{w}, \delta\}$. In this section, we explore the difference between fine-scale binary permeability fields developed by conditioning $P(\mathbf{w}, \delta|\mathbf{d})$ jointly on static and dynamic data (multiscale data) and those reconstructed from only one type of data. We assess them based on their ability to reproduce the observed breakthrough times at the SSA sensors. $P(\mathbf{w}, \delta|\mathbf{d})$ is used to develop realizations of the *fine-scale* binary medium (on a 3000×2000 mesh) as outlined in Sec. 3.3.2, which is then used in flow simulations using MODFLOW and advective transport using MODPATH. In this section, we will use the posterior distribution for $\{\mathbf{w}, \delta\}$ as developed in Sec. 3.4 with $\gamma = 0.5$ (i.e., $0.25 \leq \mathbf{F}(\mathbf{x}) \leq 0.75$) and plotted in Figs. 3.13- 3.15.

Fig. 3.16 shows example realizations of the binary medium which are consistent with the multiscale observations i.e., they were developed from $\{\mathbf{w}, \delta\}$ conditioned on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$. The white and gray regions are the high and low permeability phases, respectively. A single particle track from the injection wells to each of the SSA sensors is also shown. We choose 8 $\{\mathbf{w}, \delta\}_i$ samples from the posterior and use them to develop the corresponding $\mathbf{F}_i(\mathbf{x})$ using Eq. 3.14. Then, using the procedure (based on excursion sets of multiGaussian [mG] fields) described in Sec. 3.2.1, we develop the corresponding fine-scale binary field on a 3000×2000 mesh. The true binary field is plotted in the middle of the figure (image (e)). The variation between realizations is due to both the stochastic nature of the construction process, where white noise is convolved with a Gaussian kernel, and the sampled value of δ, which is different for each realization. Of these eight examples, realization (g) has the largest δ value and realization (f) has the smallest.

Two sets of $\mathbf{k}^{(obs)}$ and $\mathbf{t}_b^{(obs)}$ conditioning data ($\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}, \{\mathbf{k}^{(obs)}\}$) were used, and for each set, 1000 fine-scale binary realizations were created. For each of the SSA sensors and for each realization, the median travel time between the injector and the sensor is determined. Recall that sensor locations are defined at the grid-block scale and that for each location a total of 121 particles are tracked between the injector and the sensor. The median time from these 121 particles is extracted at each sensor location and designated as the breakthrough time for that location. A distribution of breakthrough times across the 1000 realizations is then created for each sensor location.

As a measure of the robustness of the parameter inference, the estimated fine-scale fields are used in a separate set of flow simulations. The parameter inference was done using data obtained on the true fine-scale field with an injection well in the lower left corner and an extraction well in the upper right corner of the domain. The 1000 fine-scale fields created from these inferred values are also used as input to a flow field with the injector in the upper left corner and extraction in the lower right corner. This second flow configuration is referred to as the "flipped" configuration.

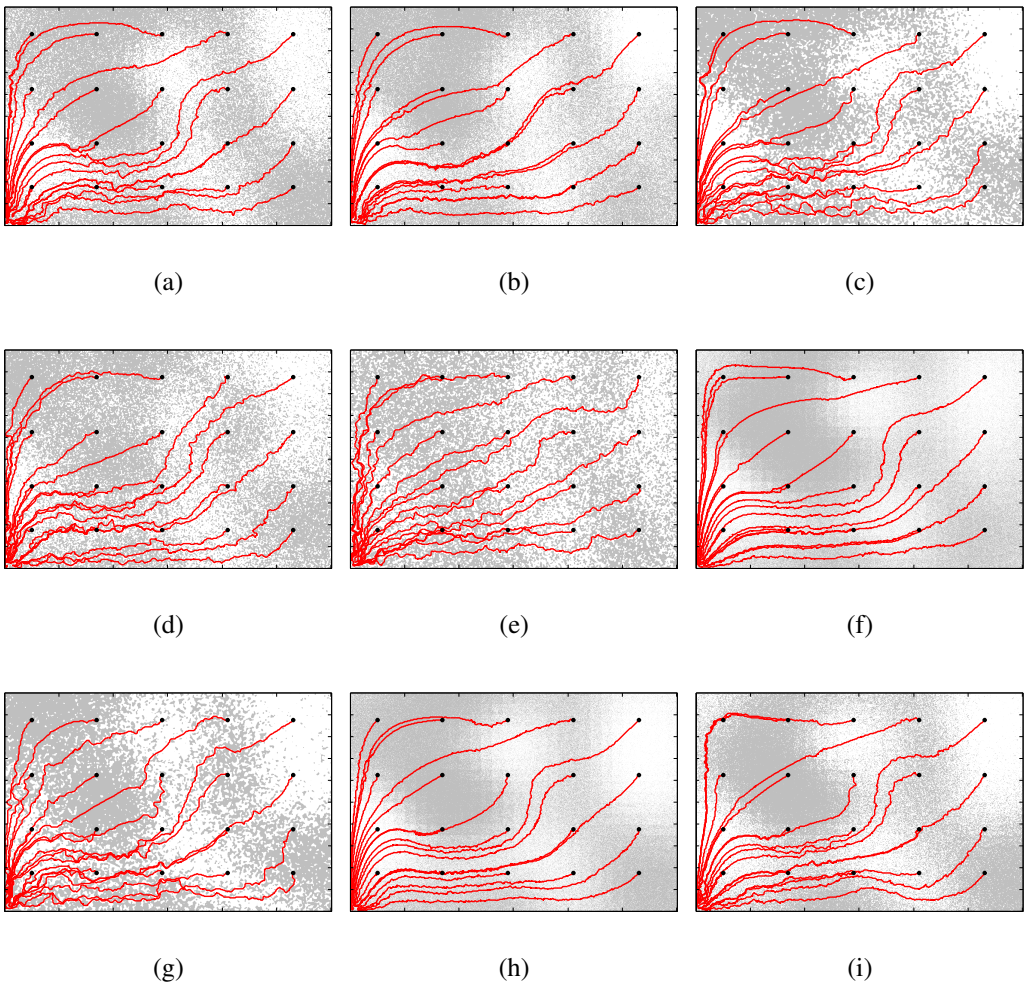


Figure 3.16: Fine-scale realizations of the random binary field conditioned on the observations and showing a single flowpath from the injection wells to each SSA sensor. These were developed from the posterior distribution of $\{\mathbf{w}, \delta\}$ using the mG-based technique described in Sec. 3.2.1. The center image (e) is the ground truth fine-scale realization. White indicates high permeability and gray is low permeability. Variation in the binary patterns are due to the stochastic nature of the process and the variation in the estimated δ parameter between realizations.

Fig. 3.17 shows a comparison of the breakthrough time CDFs (cumulative distribution function) for the two different condition data sets and the two flow configurations against the true breakthrough times for example sensor location 14 (coordinates: 21,13). For both flow configurations, both distributions capture the true breakthrough time (accurate) with the "Fine and Coarse" (alternatively, inferred jointly from $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$) set being the narrowest (most precise) for both flow configurations. Under both flow configurations, The "Coarse Only" (alternatively, inferred solely from $\{\mathbf{k}^{(obs)}\}$) data produces a breakthrough time distribution that is less precise but also less biased relative to the true breakthrough times.

Results show in Fig. 3.17 are abstracted to show the results for all 20 sensor locations. Fig. 3.18 (left) shows the precision and bias for all 20 SSA sensors and provides perspective on how these quantities change across the model domain. For each sensor location, the red "+" indicates both the location of the sensor and the relative location of the true breakthrough time. The distribution of breakthrough times from each combination of $\mathbf{k}^{(obs)}$ and $\mathbf{t}_b^{(obs)}$ is shown as a circle with black being $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ and green being $\{\mathbf{k}^{(obs)}\}$). The center of the circle is offset from the true breakthrough time by the amount of bias, calculated here as the difference between the median of the distribution and the true value. Underestimation shifts the center of the circle to the left and over estimation to the right. The center 95% of the breakthrough time distribution defines the radius of the circle. Both the radius and the bias offset are calculated in units of dimensionless travel time. The axes scales in Fig. 3.18 serve as both geographic and travel time measures to simultaneously define both the sensor locations and the relationship of the median breakthrough time distributions to the true breakthrough time. At each location, the order of the circles is set such that the smallest (tightest distribution) is at the front and the largest is at the back. Fig. 3.18 (right) shows results for the flipped configuration.

Examination of Fig. 3.18 (right) shows that both combinations of $\mathbf{k}^{(obs)}$ and $\mathbf{t}_b^{(obs)}$ create accurate distributions of breakthrough times for both the original and flipped flow configurations (all circles contain the true value denoted by the "+"). Additionally, estimated fields that are conditioned jointly on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ create the most precise distribution (black circle is smallest and therefore on the top) for the majority of the SSA locations in both flow configurations. Bias is relatively small at all locations as shown by all three circles being approximately centered on the true value. The spatial patterns shown Fig. 3.18 are consistent with the flow patterns. The breakthrough time distributions are most precise along the diagonal between the injector and producer where the majority of the flow takes place and least precise along the upper and lower boundaries where the flow fraction is quite low (see Fig. 3.2, b). The amount of bias is also smallest along the diagonal and greatest along the top and bottom boundaries. Results of the flipped configuration (Fig. 3.18, right) show decreasing precision towards the right end of the domain.

The results summarized in Fig. 3.18, show that the bias tends to be smallest in the $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ case where the fine-scale realizations were inferred jointly from static and dynamic observations. This difference in bias between the results of the two data sets increases for sensors which are further way from the diagonal connecting the injector and the producer, which

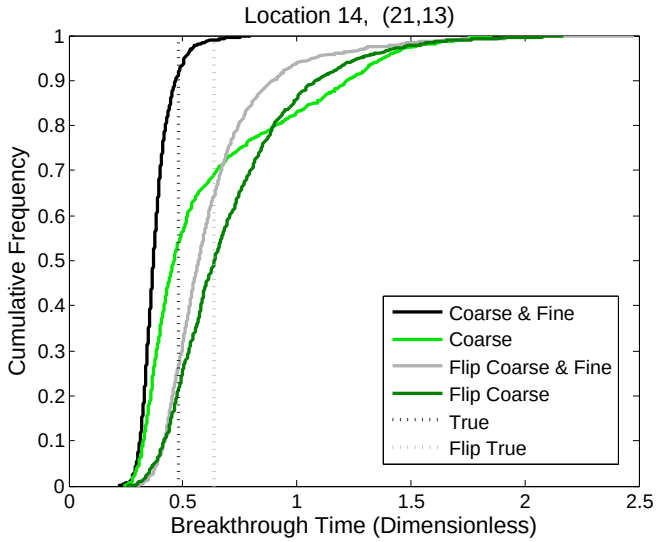


Figure 3.17: Breakthrough time CDFs, each created from 1000 samples, are shown for an example location (sensor 14). Distributions for both data sets: "Coarse & Fine" refer to realizations conditioned jointly on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$, and "Coarse only" refer to inferences using $\{\mathbf{k}^{(obs)}\}$ and the original and flipped flow configurations are shown.

forms the main travel path for the fluid. Ray *et al.* [92] provide additional results for fields where $\mathbf{F}(\mathbf{x})$ was allowed to vary from 0.0 to 1.0 including tabulated values of the bias and the variability of the breakthrough times as a function of sensor location and the conditioning data combination.

Fig. 3.18 is focused on the breakthrough time at the SSA sensor locations. However, since the ground truth in this study is known, it is possible to examine the quality of the travel time estimates for every location in the model domain. Fig. 3.19 (left) shows distributions of the average absolute error (AAE) between the estimated and true breakthrough time. Each average is calculated over all 600 coarse-scale grid blocks for a single realization and the distribution of the averages is then determined across all 1000 realizations. Lower values in the distribution come from realizations that better fit the true breakthrough time at all locations. The two distributions created from the joint $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ conditioning data produce the lowest average AAE values, although for the original flow configuration, the difference between the two conditioning data sets is minimal. The addition of the fine-scale $\{\mathbf{t}_b^{(obs)}\}$ data produces a significant reduction in the average breakthrough time AAE values for the flipped flow configuration. The median dimensionless travel time AAE values range from approximately 0.20 to 0.30. (Fig. 3.19).

All comparisons shown above consider the breakthrough time values from the injector to

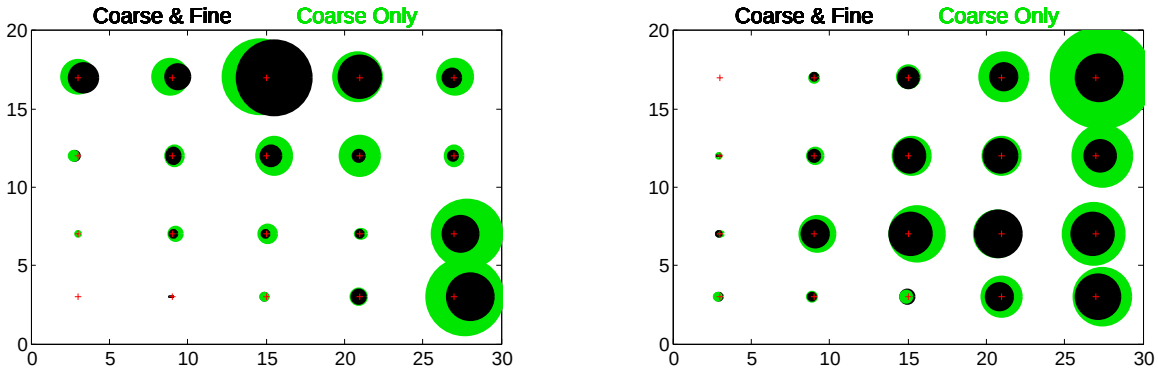


Figure 3.18: Comparison of the median travel times to the SSA sensors for two conditioning data sets and two flow configurations. The circle plots summarize the breakthrough time distributions and compare them to the true breakthrough times for all sensor locations for the original (left) and flipped (right) flow configurations: Details are in the text. "Coarse & Fine" refer to realizations conditioned jointly on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$, "Coarse only" refer to inferences using on $\{\mathbf{k}^{(obs)}\}$. The red "+" is the location of sensors and the relative location of the true breakthrough time.

one or more coarse grid cells. These comparisons are consistent with the conditioning data, $\mathbf{t}_b^{(obs)}$, used in the inverse estimates. However, additional information can be gained from the full distribution of travel times to each coarse grid block. The two-sample Kolmogorov-Smirnov (KS) test is used to compare the true travel time distribution to the distribution calculated on each estimated field for each coarse-scale grid block. The KS test is a non-parametric test of the difference between two distributions. The KS test statistic is the maximum vertical distance, D between two CDFs:

$$D_{i,j} = \sup_x |F_{i,n}(x) - F_{j,n'}(x)| \tag{3.16}$$

Here the i^{th} and j^{th} distributions have the same number of travel times, 121, for all comparisons ($n = n'$). Values of the D statistic range from 0, when there is no difference between the distributions, to 1.0 when the values of distributions do not overlap at all. The D statistic can be used in KS test where the null hypothesis of the KS test is that both samples come from the same underlying population. Here, for each fine-scale realization, we compare the modeled and observed travel time distribution at every location within the model domain and calculate the test statistic, D. The average value of D across all 600 locations is retained for each of the 1000 realizations.

Fig. 3.19 (right) shows the resulting distributions of the average D value for these sets of

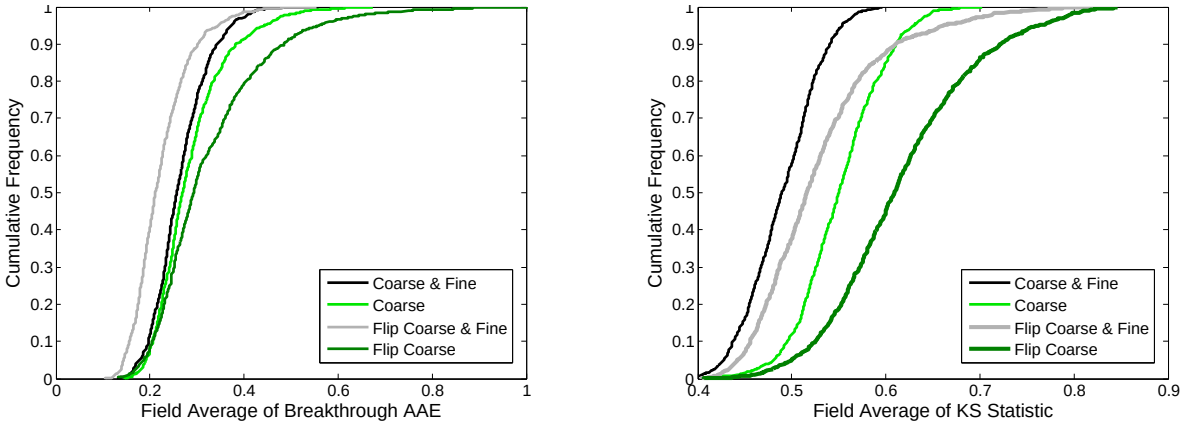


Figure 3.19: Average absolute errors (AAE) between median travel times from the estimated and true travel time distributions for all locations in each field (left) and distribution of average KS statistic values calculated by comparing the full distribution of estimated and true travel time distributions at all locations (right).

1000 realizations. The smallest values of D occur when both $\mathbf{k}^{(obs)}$ and $\mathbf{t}_b^{(obs)}$ are used as conditioning data and this result holds for both flow configurations. The distributions of D when only $\mathbf{k}^{(obs)}$ data are used have average D values that are consistently 0.06 to 0.07 higher than the corresponding values created using both $\mathbf{k}^{(obs)}$ and $\mathbf{t}_b^{(obs)}$ for the original flow configuration (compare black and bright green lines in Fig. 3.19, right). There is also a nearly consistent offset to higher D values for the flipped configuration of approximately 0.10.

A third flow configuration is used as an additional test of the robustness of the multiscale inferences. The fine-scale fields constructed from the inferred parameters are used as input to flow and transport calculations with "permeameter" boundary conditions (i.e., zero-flux boundaries on the two long edges of the domain and fixed pressure boundaries on the short ends of the domain). The fixed pressures are set such that the same average gradient as produced with the injection and extraction wells in the other two flow configurations is also maintained in the permeameter flow conditions. This flow configuration is representative of ambient conditions in a ground water aquifer when all pumping has been stopped. The same example realizations and the ground truth field as shown in Fig. 3.16 are shown in Fig. 3.20 with particle tracks calculated under the permeameter flow configuration.

The same analysis as done above for inferences conditional on both the $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ and $\{\mathbf{k}^{(obs)}\}$ data sets are repeated for the permeameter flow configuration. Additionally, the y-coordinate at which each particle exits the model domain along the downstream (right-hand) boundary is recorded and these results are also analyzed. Similar to Fig. 3.17, Fig. 3.21 (left) shows the CDF's of the median travel times at sensor location 14, but in Fig. 3.21 these are the travel times from sensor location 14 to the downstream boundary. Also note that the X-axis has values of log10 dimensionless travel time. The CDFs of the

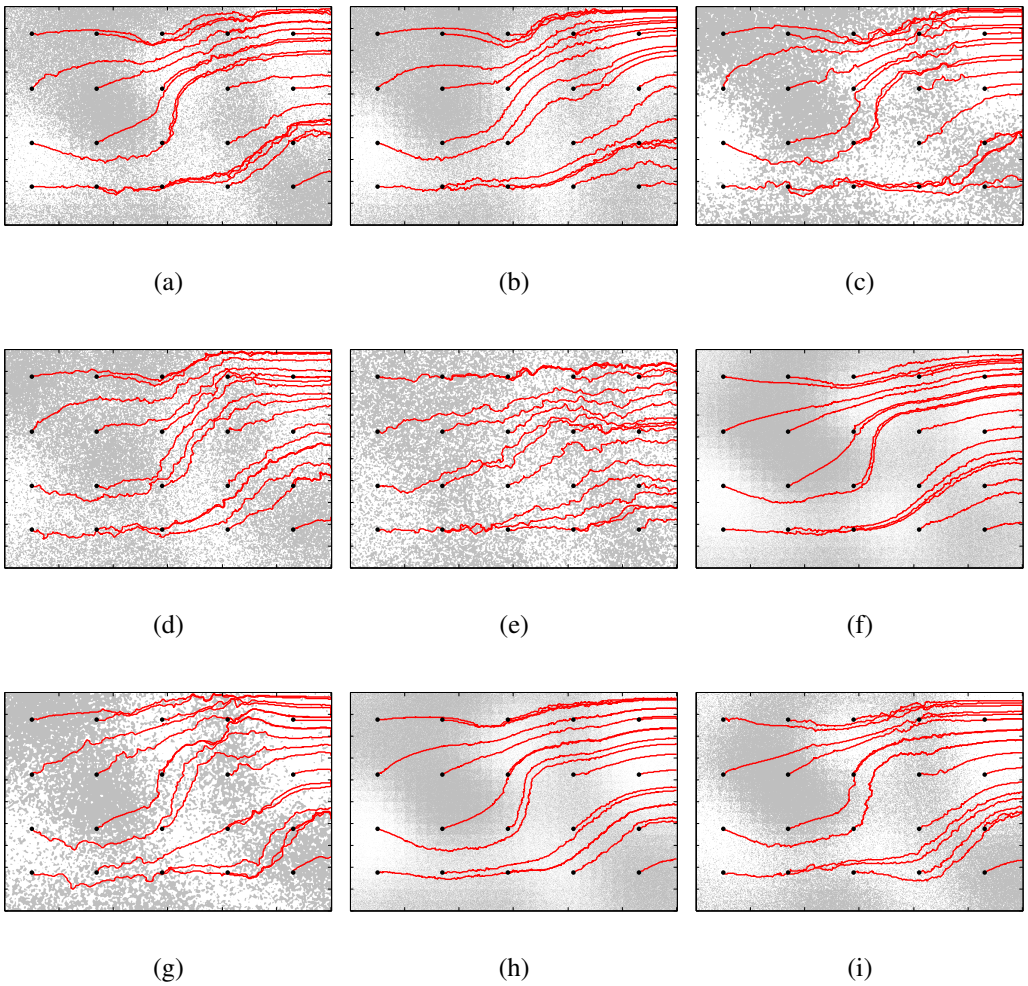


Figure 3.20: Fine-scale realizations of the random binary field conditioned on the observations under the permeameter flow configuration and showing a single flowpath from each SSA sensor to the downstream boundary. These are the same fine-scale realizations shown in Fig. 3.16. The center image (e) is the ground truth fine-scale realization. White indicates high permeability and gray is low permeability.

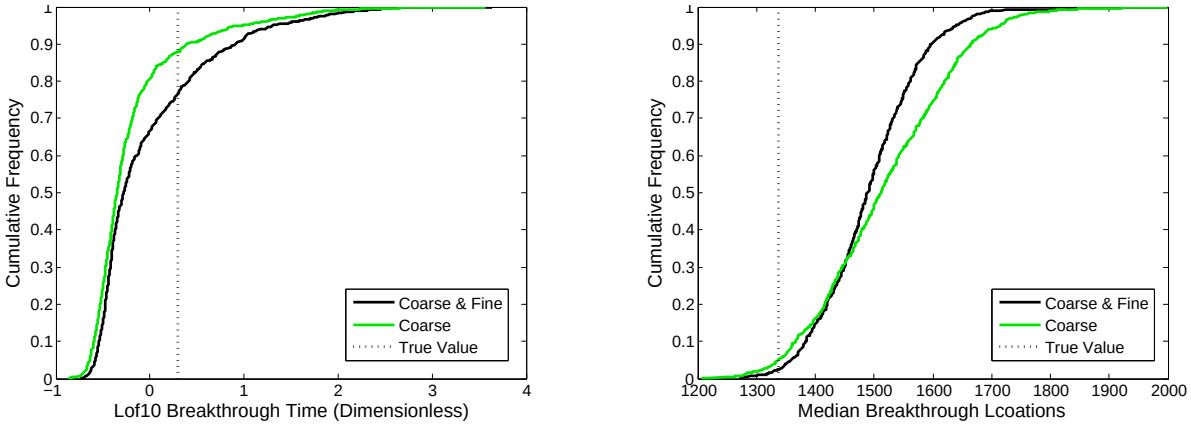


Figure 3.21: Permeameter flow configuration results. Breakthrough time CDFs (left) for example location (sensor 14) for both sets of conditioning data. The right image shows the CDF's of the median y-coordinate where the particle exits the downstream boundary. Each CDF is constructed from flow and transport simulations on 1000 fine-scale fields. The true breakthrough time and median exit coordinate values are shown for comparison.

particle exit location (Y-coordinate on the downstream boundary) for particles released at sensor location 14 are shown in the right image. It is noted that the particle exit coordinates can range from 0 to 2000 for these fine-scale simulations. All distributions are accurate, and there is little difference in the results created by using the two different sets of conditioning data. Given the large variability, the predicted breakthrough times, arising from permeability fields conditioned on $\{\mathbf{k}^{(obs)}\}$ as well as $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$, for sensor 14, are roughly of the same accuracy and quality.

The results in Fig. 3.22 show that, for the permeameter boundary condition (BC), contrary to the results summarized in Fig. 3.18, the variability of the posterior predictions is not a strong function of the sensor location within the domain, not all posterior predictions are accurate and the variability as shown by the size of the circles is larger. In order to create Fig. 3.22 it was necessary to use the middle 95 percent of the log10 travel time distributions to fit all of the circles into the domain. Therefore, the circle sizes are not directly comparable with those in Fig. 3.18. Despite the increased width of the distributions, two of them do not capture the true value: the Coarse Only ($\{\mathbf{k}^{(obs)}\}$, green) distribution at sensor location (14,7) and the Fine and Coarse ($\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$, black) distribution at sensor location (21,3). In the majority of the cases, the Coarse Only ($\{\mathbf{k}^{(obs)}\}$) data produce the most precise distributions, green circles lie on top of black circles, indicating that the addition of fine-scale breakthrough times from a pumping configuration does not improve predictions of travel time under an ambient flow configuration. This is because the small-scale structures at the inflow and outflow of the permeameter configuration are not measured very well by $\{\mathbf{t}_b^{(obs)}\}$, which are obtained by a flow configuration that is largely diagonal (bottom-left to top-right). The lack of accuracy at the inflow and outflow (left and right boundaries of the

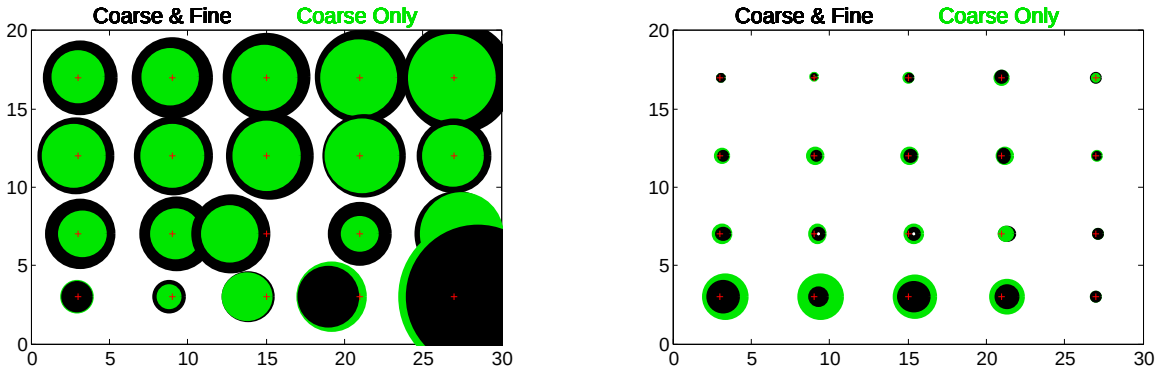


Figure 3.22: Comparison of the median travel times from the SSA sensors for two conditioning data sets ($\{\mathbf{k}^{(obs)}\}$ versus $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$) under the permeameter flow configuration (left). The median particle exit coordinates are shown in the right image. The circle plots summarize posterior predictive distributions and compare them to the true values for all sensor locations. "Coarse & Fine" refer to realizations conditioned jointly on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$, "Coarse only" refer to inferences using on $\{\mathbf{k}^{(obs)}\}$.

domain) can be seen in Fig. 3.18, where the predicted $\mathbf{t}_b$ at the top left and bottom right show large uncertainties.

The circles in the right image of Fig. 3.22 show the distribution of the exit coordinates. The values here can range from 0 to 2000 and the circles radii were scaled by dividing the width of the middle 95 percent of each distribution by 1000. The lowest row of sensors has the largest variability in exit coordinates while this variability is otherwise relatively constant across the domain. Contrary to the travel time results, using both Fine and Coarse scale data, $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$, increases the precision of the posterior predictions (black circles on top of green circles). All posterior exit coordinate distributions are accurate. Interestingly, the least precise exit coordinate distributions are the most precise travel time distributions (compare the lower row of circles in the left and right plots of Fig. 3.22).

Fig. 3.23 shows distributions of AAE as averaged across each realization for both set of conditioning data and for travel time (left image) and exit coordinate (right image). The left side of Fig. 3.23 is similar to the left side of Fig. 3.19 but the longer travel times necessitate using the log10 transform for the X-axis in Fig. 3.23 . In both images of Fig. 3.19, there is little difference in the error distributions between the two different conditioning data sets, although the Coarse and Fine ($\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$, black) data set produces the largest AAE values for the breakthrough times (left image). Again, the difference between the fields conditional on $\{\mathbf{k}^{(obs)}\}$, versus $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$, is very small.

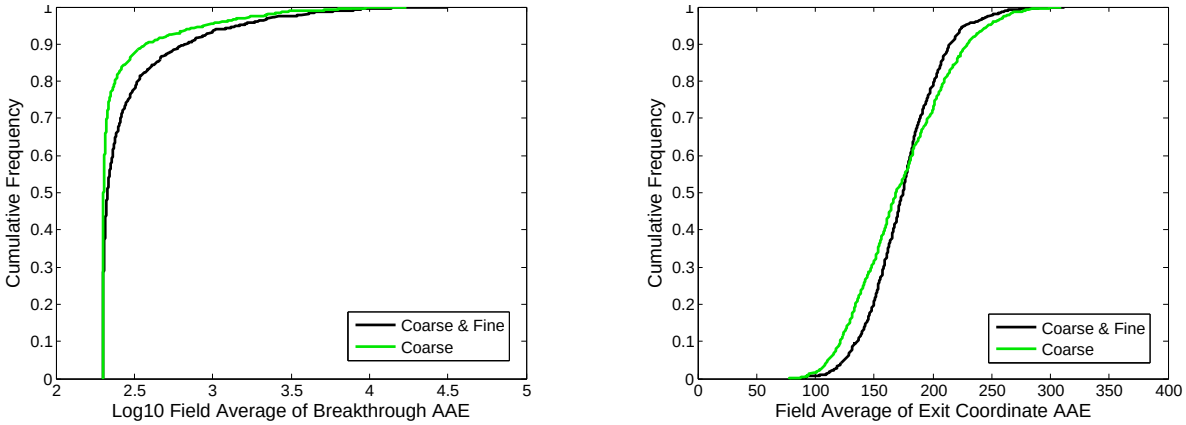


Figure 3.23: Average absolute errors (AAE) between log10 median travel times from the estimated and true travel time distributions (left) and distribution of average absolute errors (AAE) between median exit coordinates from the estimated and true exit coordinate distributions (right). These results are for the permeameter flow configuration and both graphs are created from all locations in each field.

Fig. 3.24 compares the KS statistics for the travel time distributions (left) and exit coordinate distributions (right). Each distribution represents 1000 average value, each average calculated across the 600 locations on the coarse-scale field. From Fig. 3.24, it is obvious that adding fine-scale conditioning data to the parameter inference does not make a significant difference in the travel time or exit location distributions under the permeameter flow configuration. The KS statistics for the exit coordinates are both larger and more precise than for the travel times. Comparison of the left and right images of Fig. 3.24 shows a median KS statistic of approximately 0.6 for the travel times and about 0.85 to 0.87 for the exit coordinates.

To summarize, we sampled the distribution $P(\mathbf{w},\delta|\mathbf{d})$, developed in Sec. 3.4, to construct 1000 realizations of the fine-scale binary medium on a 3000×2000 mesh and predict breakthrough times at the SSA sensors via simulation. These calculations were performed for the two different $P(\mathbf{w},\delta|\mathbf{d})$ conditioned on $\{\mathbf{k}^{(obs)},\mathbf{t}_b^{(obs)}\}$, and $\{\mathbf{k}^{(obs)}\}$) to gauge the impact of multiscale data in the estimation of $\mathbf{K}_e(x)$. Additionally, the degree to which the estimations are robust to changes in flow conditions was addressed by using the estimated fields in two additional flow configurations, flipped and permeameter, that are different from the flow configuration used to create the observations used in the parameter inference.

Posterior predictive model evaluations using 1000 realizations created from each combination of conditioning data clearly showed that while both data combinations produce accurate results, using data collected on both scales i.e., $\{\mathbf{k}^{(obs)},\mathbf{t}_b^{(obs)}\}$, as opposed to a single scale ($\{\mathbf{k}^{(obs)}\}$, creates predictions that are most precise and closest to actual values from the ground truth. This result holds for predictions made using the original flow configuration as well as when using the flipped flow configuration. However, when permeameter

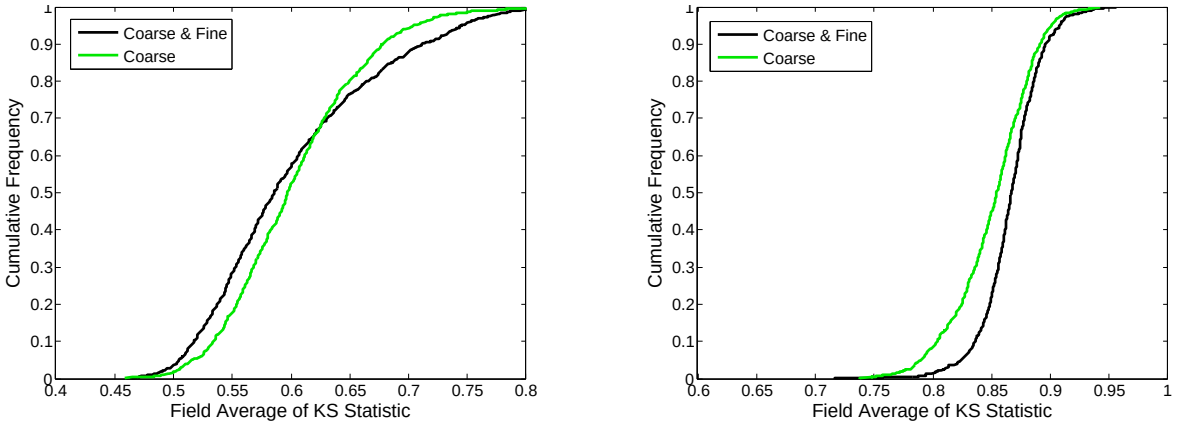


Figure 3.24: Distributions of average KS statistics between median travel times from the estimated and true travel time distributions (left) and distribution of average KS statistics between median exit coordinates from the estimated and true exit coordinate distributions (right). These results are for the permeameter flow configuration and both graphs are created from all locations in each field.

boundary conditions are applied to the fine-scale fields, the predictions are essentially independent of the data conditioning set. This is because $\mathbf{t}_b^{(obs)}$ contains little information on the small-scale structure along the left and right boundaries of the domain – Fig. 3.16 show that few path lines emanating from the injector travel along the top left and bottom right corners of the domain. This lack of information results in less accurate inferences in those regions (alternatively, the posterior predictive tests for $\mathbf{t}_b$ show large spreads, Fig. 3.18). However, these are the very regions which form the inflow and outflow of the permeameter flow configurations, leading to large spreads in breakthrough times and locations (Fig. 3.22). Thus using $\{\mathbf{t}_b^{(obs)}\}$ to infer the permeability fields conferred no advantage, in terms of accuracy and predictive skill, in the permeameter flow configuration.

3.6 Multiscale Inference with Structural Errors

In Sec. 3.3 we demonstrated the solution of the inverse problem for inclusion distribution and size under the assumption that the discrepancy between the predictions of our models ($\mathbf{t}_b = \mathcal{M}(\mathbf{K})$ and $\mathbf{K}_e(x) = \mathcal{L}(\mathbf{F}(\mathbf{x}), \delta)$) and the observations $\mathbf{d}$ as Gaussians (Eq. 3.10). However, in most real-life cases, the discrepancy is dominated by model errors i.e., the inability of $\mathcal{M}(\mathbf{K})$ and $\mathcal{L}(\mathbf{F}(\mathbf{x}), \delta)$ to capture reality. Fig. 3.1 compares the predictions of effective log-permeability made by $\mathcal{L}$ versus those computed numerically using MODFLOW-2005, for the same set of random binary media realizations. As is clear, the scatter of MODFLOW-2005 estimates around the $\mathcal{L}$ predictions is neither symmetric nor Gaussians; imposing a normal distribution on the discrepancy, while sufficient for demonstrating an algorithm, cannot be justified when accuracy of inference is a concern. Thus, a new error model needs to be devised and used to derive a counterpart to Eq. 3.11. It is clear from Fig. 3.1 that the discrepancy in $\mathbf{K}_e(x)$ will be dependent on $\mathbf{F}$ i.e., an expression for $P(\Delta\mathbf{K}|\mathbf{F})$ will have to be developed, where $\Delta\mathbf{K} = \mathbf{K}_c^t - \mathcal{L}(\mathbf{F}, \delta)$. Also, for ease of use with MCMC, a unimodal distribution for $P(\Delta\mathbf{K}|\mathbf{F})$ would be desirable.

3.6.1 Construction of the Error Model for the Link Function

We develop a model to evaluate $P(\Delta\mathbf{K}|\mathbf{F})$ based on kernel density estimation [95]. In Fig. 3.25, top left, we plot $\mathbf{K}_c^t$ and $\mathbf{K}_e(x)$ as a function of $\mathbf{F}$, for δ = 38 grid-cells. Since $(\mathbf{F}, \delta)$ is an incomplete description of the fine-scale, many fine-scale realizations, each with its own $\mathbf{K}_c^t$ are consistent with such a specification, leading to a cloud of points (in black); the crosses indicate $\mathbf{K}_e(x) = \mathcal{L}(\mathbf{F}, \delta)$. A joint distribution of $(\Delta\mathbf{K}, \mathbf{F})$, plotted as a scatter plot, is shown in Fig. 3.25, top right; this serves as our starting point for developing a kernel-smoothed model for the joint probability density function $P(\Delta\mathbf{K}, \mathbf{F})$.

We assume that the samples plotted in Fig. 3.25 (top right) are drawn from an unknown joint distribution which we approximate with $P(\Delta\mathbf{K}, \mathbf{F})$ as

$$\begin{aligned} P(\Delta\mathbf{K}, \mathbf{F}) &= \frac{1}{S}\sum_{k=1}^{S} \mathcal{K}\left(\frac{\Delta\mathbf{K} - \Delta\mathbf{K}_k}{h_{\mathbf{K}}}\right) \mathcal{K}\left(\frac{\mathbf{F} - \mathbf{F}_k}{h_{\mathbf{F}}}\right) \\ &= \frac{1}{S}\frac{1}{h_{\mathbf{K}} h_{\mathbf{F}}} \sum_{k=1}^{S} \mathcal{K}(\Delta\mathbf{K} - \Delta\mathbf{K}_k)\, \mathcal{K}(\mathbf{F} - \mathbf{F}_k) \end{aligned} \tag{3.17}$$

where S is the number of $(\Delta\mathbf{K}, \mathbf{F})$ samples, $\mathcal{K}$ is an Epanetchnikov kernel and $\{h_{\mathbf{K}}, h_{\mathbf{F}}\}$ are the bandwidths in the two directions. The bandwidths are calculated using a plug-in technique [96]. The evaluation of $P(\Delta\mathbf{K}, \mathbf{F})$, per Eq. 3.17, involves a sum over all samples, which can be expensive; instead we employ an approximate summation method using KD-trees [97]. The samples are collated in to boxes numbering O(log(S)) and stored in a KD-tree. An approximate expression for the contribution of kernels in a given box to

$P(\Delta\mathbf{K},\mathbf{F})$ is computed and the summation is done over $O(\log(S))$ boxes. The novelty of the scheme lies in the hierarchical manner in which kernels are collated in boxes, which are further refined/sub-divided to meet an accuracy criterion. Some boxes are shown in Fig. 3.25, top right. The joint probability density distribution $P(\Delta\mathbf{K},\mathbf{F})$ resulting from the kernel-smoothed approximation is shown in Fig. 3.25, lower left.

The evaluation of $P(\Delta\mathbf{K}|\mathbf{F})$ is performed simply by

$$P(\Delta\mathbf{K}|F) = \frac{P(\Delta\mathbf{K},\mathbf{F})}{P(\mathbf{F})} \tag{3.18}$$

where $P(\mathbf{F})$ is calculated by numerically integrating out $\Delta\mathbf{K}$ in $P(\Delta\mathbf{K},\mathbf{F})$. Computations involving KDE were performed using the MATLAB$^{\text{TM}}$ code in [98]. Fig. 3.25, lower right shows $P(\Delta\mathbf{K}|\mathbf{F})$ for different values of $\mathbf{F}$.

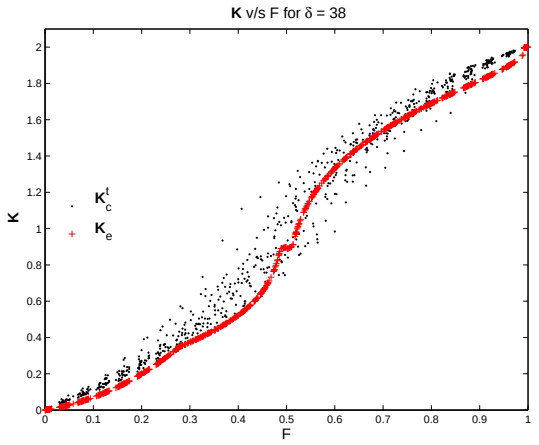
Note that the bandwidths $\{h_{\mathbf{K}}, h_{\mathbf{F}}\}$ obtained using plug-in bandwidth estimation had to be over-smoothed by a factor of 2 to ensure that $P(\Delta\mathbf{K}|F)$ did not have largely disconnected modes (since they prevent MCMC chains from mixing).

In order to construct a model for $P(\Delta\mathbf{K}|\mathbf{F})$, valid for arbitrary $\mathbf{F}$ and δ, the exercise was repeated for $\delta = 9, 15, 21, 28, 38, 56$ and 75 grid-cells. Models for $P(\Delta\mathbf{K}|\mathbf{F},\delta)$ were developed for each. Thereafter, $P(\Delta\mathbf{K}|\mathbf{F},\delta)$ for arbitrary values of δ were obtained by adopting that of the nearest known δ.

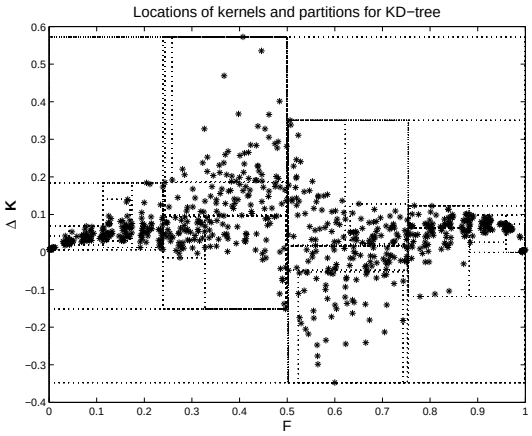
Kernel density estimates of probability density distributions are generally applicable only in the region where samples exist and can be misleading away from it. Our use of Epanetchnikov kernels (which have compact support) ensures that at large $\Delta\mathbf{K}$, $P(\Delta\mathbf{K}|\mathbf{F},\delta)$ will evaluate to zero, in keeping with the absence of samples. This will be reflected when devising an expression for $P(\mathbf{d}|\mathbf{K}_e(x),\mathbf{F},\delta)$ in Eq. 3.11 i.e., the expression will evaluate to zero in large parts of the multi-dimensional $(\mathbf{w},\delta)$ space, as will the posterior distribution. This can pose a problem when exploring the space of the posterior density using a MCMC sampler - large regions of zero gradient in the posterior will not allow the chain to move to more promising regions. We ameliorate this situation by adding an extra Gaussian kernel $\mathcal{K}_G$ to Eq. 3.17, i.e.

$$P(\Delta\mathbf{K},\mathbf{F}) = \frac{1-\alpha}{S}\sum_{k=1}^{S}\mathcal{K}\left(\frac{\Delta\mathbf{K}-\Delta\mathbf{K}_k}{h_{\mathbf{K}}}\right)\mathcal{K}\left(\frac{\mathbf{F}-\mathbf{F}_k}{h_{\mathbf{F}}}\right) + \alpha\mathcal{K}_G(\Delta\mathbf{K}-\mu_k,\Gamma_k) \tag{3.19}$$

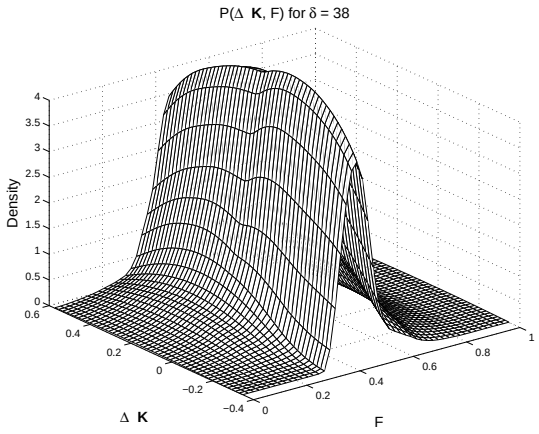
where α is chosen so that only 10% of the probability mass is contributed by the Gaussian kernel. The infinite support of the Gaussian allows the posterior to be calculated anywhere in the $(\mathbf{w},\delta)$ space and provides the gradient required by the MCMC chain to move into the high probability regions of the posterior. In the regions in $(\mathbf{w},\delta)$ space where $\Delta\mathbf{K}$ is small, the kernel density estimate for the model errors' probability density dominate, allowing the inference to reflect their effect (uncertainty due to model errors), rather than the measurement noise.



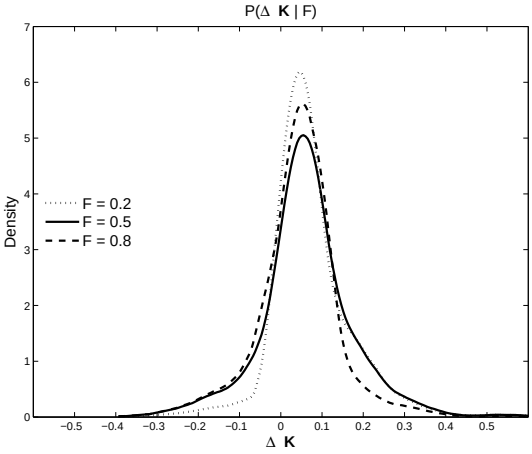
(a) $\mathbf{K}_c^l$ and $\mathcal{L}(\mathbf{F},\delta)$ versus $\mathbf{F}$



(b) $\Delta\mathbf{K}$ versus $\mathbf{F}$



(c) Joint PDF $P(\Delta\mathbf{K},\mathbf{F})$



(d) $P(\Delta\mathbf{K}|\mathbf{F})$

Figure 3.25: Stages in the construction of the conditional PDF $P(\Delta\mathbf{K}|\mathbf{F},\delta)$. Top left: We create multiple realizations of the fine-scale random binary media field and plot their true upscaled log-permeability (as a cloud) and the model prediction $\mathbf{K}_e(x) = \mathcal{L}(\mathbf{F},\delta)$, for $\delta = 38$ grid-cells. Top right: we show the samples and the boxes from Level 4 of the KD-tree holding the boxed samples. Bottom left: We plot the kernel density estimate of $P(\Delta\mathbf{K},\mathbf{F})$ for $\delta = 38$ grid-cells, constructed using Epanetchnikov kernels (after over-smoothing the bandwidths to remove isolated modes). Bottom right: We plot the conditional distribution $P(\Delta\mathbf{K}|\mathbf{F})$, for $\delta = 38$ grid-cells, for various values of $\mathbf{F}$.

3.6.2 Adapting and Solving the Inverse Problem for Model Errors

In this section, we incorporate the expression for $P(\Delta\mathbf{K}|\mathbf{F})$ (Eq. 3.18), but calculated from the "ameliorated" joint density distribution in Eq. 3.19, into a formulation of an inverse problem (similar to Eq. 3.11) and solve it to infer fine-scale properties.

We start with the assumption that the measurement errors are far smaller than the model error / discrepancy $\Delta\mathbf{K}$ as defined above. In such a case

$$\mathbf{k}^{(obs)} \approx \mathbf{K}_c^t = \mathcal{L}(\mathbf{F}(\mathbf{x}), \delta) + \Delta\mathbf{K}.$$

Let $k_j^{(obs)}$ be the measured log-permeability in the grid-box containing sensor j and $\mathbf{F}(\mathbf{x}_j)$ be the corresponding inclusion proportion. Assuming that the model errors, conditioned on the local $\mathbf{F}(\mathbf{x}_j)$ are i.i.d,

$$P(\mathbf{k}^{(obs)}|\mathbf{F}(\mathbf{x}), \delta) = \prod_{j=1}^{N_s} P(k_j^{(obs)}|\mathbf{F}(\mathbf{x}_j), \delta) = \prod_{j=1}^{N_s} P(\Delta\mathbf{K}_j|\mathbf{w}, \delta) \tag{3.20}$$

where $\Delta\mathbf{K}_j$ is the observation-prediction mismatch of log-permeabilities at sensor j (but now attributed to model errors rather than measurement errors, as was done in Eq. 3.11). We model the mismatch between observed and predicted breakthrough times as before, i.e., as measurement errors modeled using i.i.d. Gaussians, implying that the model $\mathcal{M}(\mathbf{K})$ is an accurate representation of tracer transport. Thus the counterpart to Eq. 3.12, incorporating model errors, is

$$\begin{aligned} P(\mathbf{w}, \delta|\mathbf{d}) \quad \propto \quad & \exp\left(-\frac{[\mathbf{e}_t(\mathbf{w}) - \mu_t]^T[\mathbf{e}_t(\mathbf{w}) - \mu_t]}{\sigma_t^2}\right) \prod_{j=1}^{N_s} P(\Delta\mathbf{K}_j|\mathbf{w}, \delta) \\ & \pi(\delta) \prod_{l=1}^{M} \exp(-w_l^2) \end{aligned} \tag{3.21}$$

Eq. 3.21 is solved using the same techniques and for the same problem as Sec. 3.3. However, we restrict ourselves to the SSA set of sensors.

The new model for observation-prediction discrepancy in $\Delta\mathbf{K}$ will allow the MCMC chain to visit regions of the log-permeability ($\mathbf{K}$) space that may not have been allowed by the Gaussian model used in Sec. 3.3. However, this exploration will also be modified by the Gaussian error model used for $\mathbf{t}_b$. We examine the effect of these error models on $\{\mathbf{w}, \delta\}$ individually.

In Fig. 3.26, we plot the posterior PDFs for w_1 and w_{30}, as inferred solely from static data, as Q-Q plots. Posterior distributions, as obtained in Sec. 3.3, are also plotted. We see that the KDE error model constructed in Sec. 3.6.1 makes a difference only for the low mode (w_1) whereas the high mode (w_{30}) is unaffected. This is in keeping with the conclusions

drawn in Sec. 3.3. Note, however, that the posterior PDF for w_1, obtained with the KDE model is different from the one obtained with the Gaussian model of Sec. 3.3; for instance, the Q-Q plot is not a straight line and so the posterior is no longer a Gaussian.

We now investigate the effect of $\mathbf{t}_b$ by incorporating dynamic data into our inference. In Fig. 3.27 we plot the posterior PDFs for the Karhunen-Loève weights w_1, w_{15}, w_{30} and $\ln(\delta)$ as computed using model errors. The plots with solid lines were inferred using both static and dynamic data, whereas the ones with dotted lines were inferred using static data only. We see that like our results in Sec. 3.3 the static data determines the large-scale structures - the posteriors for w_1 obtained from the static observations only is very similar to the one obtained from static and dynamic observations i.e. the inclusion of $\mathbf{t}_b$ contributes very little information. On the other hand, as described in the previous paragraph the posterior and prior densities for w_{30} in the static-observations-only case are indistinguishable (Fig. 3.26) and the posterior in the lower left sub-figure of Fig. 3.27 is due almost entirely to $\mathbf{t}_b$ observations. $\ln(\delta)$ seems very little affected by the inclusion of dynamic data as well as the model errors. We also plot the posterior PDFs for the same variables as obtained in Sec. 3.3 (i.e., from Fig. 3.7, but only for the SSA set of sensors). We see a substantial difference in the posterior PDF for the larger modes, but the higher/finer modes are very similar. Thus the effect of KDE model for errors is most felt by inference variables that have global effect e.g. w_1, (and are impacted by static data) whereas those with local effect are dominated by the information content of the breakthrough times (where the error model is the same as in Sec. 3.3). By the same argument, if the inference were to be done with static data only, the posterior density for w_{30} would be the same as the prior (and completely unaffected by the change in the error model between Sec. 3.3 and this section), whereas the largest change would be seen in w_1. This is corroborated in Fig. 3.26. Also, comparing the width of the PDFs for the objects of inference, we do not see much of decrease in width, i.e., while the posterior distributions are different for all the objects of inference, there has not been a reduction of uncertainty. However, whether or not there is an improvement in the goodness of fit (i.e., in the accuracy of the inference) will be evaluated next.

We perform posterior predictive checks (PPC) using the posterior distribution of $\{\mathbf{w}, \delta\}$. The distribution of breakthrough times obtained at the 20 sensors of the SSA set are plotted in Fig. 3.28 (the "error bars" correspond to the 1^{st} and 99^{th} percentiles of the 9,500 breakthrough times obtained using $\{\mathbf{w}, \delta\}$ samples from the posterior distribution). The observed breakthrough times are plotted as triangles whereas the circles are the median of the PPC. The corresponding values from Sec. 3.3 are plotted in red. We see that the KDE error model did not contribute to the reduction of uncertainty (as mentioned in the last paragraph) since the widths of the error bars obtained with the KDE error model and Sec. 3.3's Gaussian model are about the same; however, the error bars are slightly shifted and only 2 sensor readings (as opposed to 3 in Sec. 3.3.2) are outside the bounds. This indicates a slight improvement in the goodness-of-fit.

To summarize, we developed a KDE-based representation for the model error and performed the inference of $\{\mathbf{w}, \delta\}$ based on the assumption that model errors dominated measurement errors. We obtained inferences which were different from those obtained

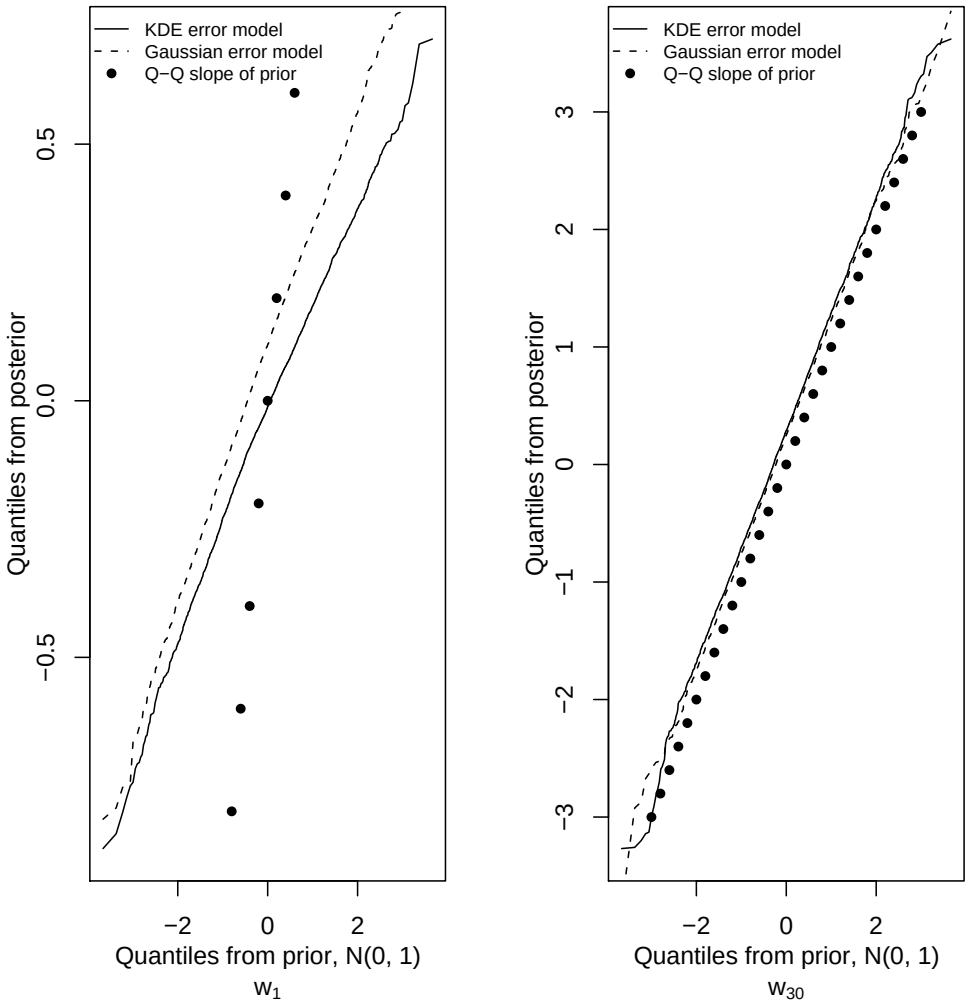


Figure 3.26: Q-Q plot for the posterior distribution of w_1 (left) and w_{30} (right). The solid line denote the Q-Q plot for inference developed with the KDE error model described in this section; the dashed line denote those that used the Gaussian model of Sec. 3.3. Only static data was used. The prior is plotted with dots. We note that the KDE model results in a plot that is not a straight line (for w_1); also the plots are different when the KDE and Gaussian models are used. Right, we see that the static data made no difference to the inference of w_{30}; the posterior is indistinguishable from the prior.

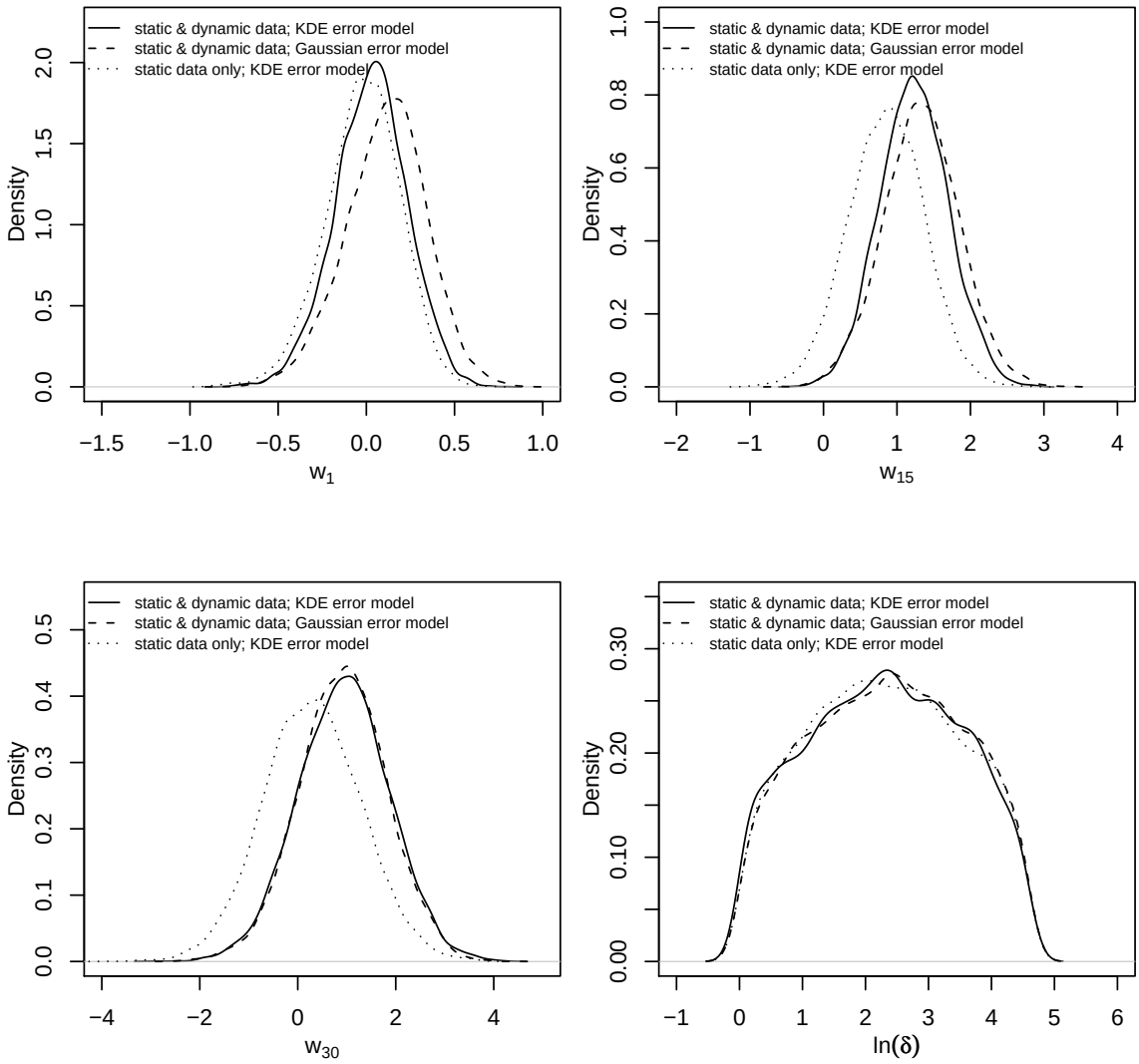


Figure 3.27: Posterior PDFs for $\{w_1, w_{15}, w_{30}, \ln(\delta)\}$ using the KDE error model, obtained with static data only (dotted lines), as well as static and dynamic data (solid line). We see that dynamic data affects the posterior for the higher Karhunen-Loève modes, as in Sec. 3.3. Also the impact of the static data is limited to the lower Karhunen-Loève modes. We see that the impact of dynamic data on $\ln(\delta)$ is low. We also plot the posterior PDF obtained using the Gaussian error model in Sec. 3.3. We see that at the finer (higher) Karhunen-Loève modes, which are impacted mostly by dynamic data, have posteriors that are relatively insensitive to the choice of KDE versus Gaussian data model.

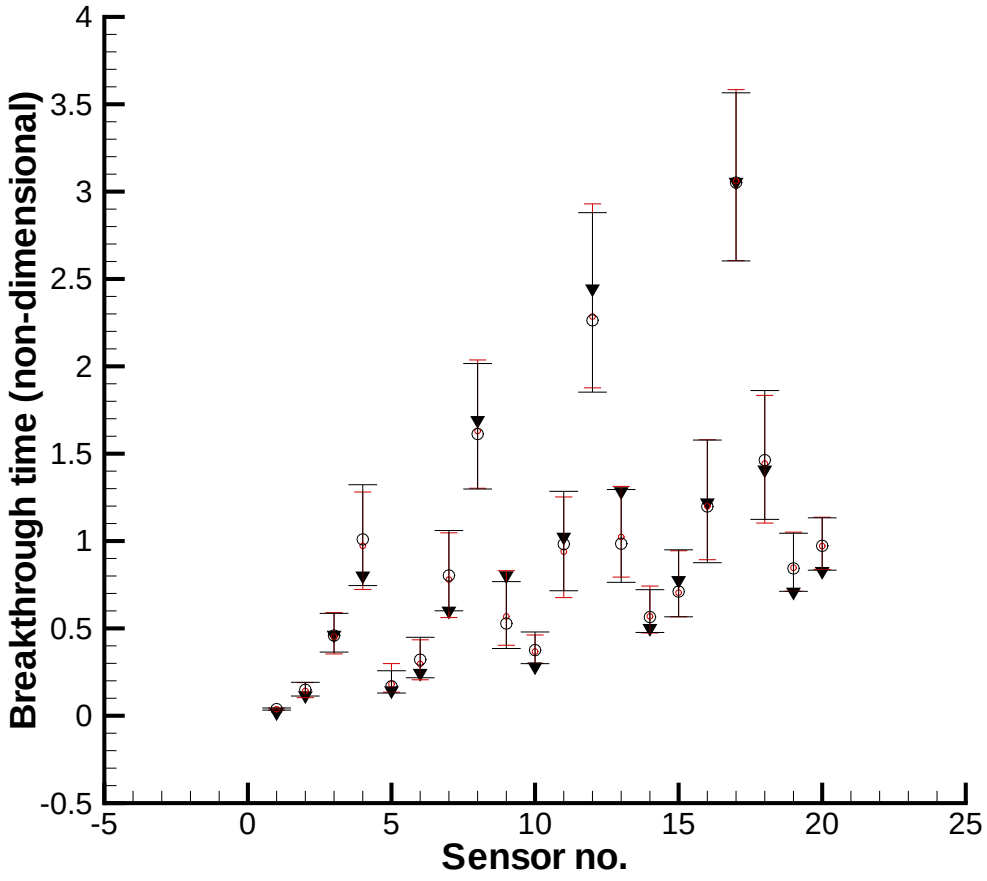


Figure 3.28: Distributions of the tracer breakthrough times at the 20 SSA sensors, obtained via posterior predictive checks using the posterior density for $\{\mathbf{w}, \delta\}$. Those plotted in black were developed using the KDE error model; the ones in red use the Gaussian model described in Sec. 3.3. The observations are in black triangles; the circles are medians. The error bars indicate the 1^{st} and 99^{th} percentiles. We see that the width of the error bars are about the same irrespective of the error model (KDE/Gaussian) used; i.e. our choice of models did not reduce the uncertainty in the inference. However, only 2 observations fall outside the black error bars, whereas 3 do for the red ones, indicating a slight improvement in the goodness-of-fit when using the KDE error model.

in Sec. 3.3, but these were restricted to variables which had global effect; the posterior for higher Karhunen-Loève modes, which are controlled by the breakthrough times, were hardly affected. A local variable like $\ln(\delta)$ also was unaffected. The reason for the change was the ability of the MCMC chain to explore parts of the $\mathbf{K}$ space which was not possible under the Gaussian error model in Sec. 3.3. While we did obtain a better fit of the model to data, as indicated by the PPC, it was marginal. The reason for this lies in our restrictive model for $\mathbf{K}_e(x) = \mathcal{L}(\mathbf{F}(\mathbf{x}), \delta)$; the KDE model for $\Delta\mathbf{K}$ shows that the $\mathbf{K}$-space is quite involved and may not be easily captured by the Karhunen-Loève -mode based reduced model. This inability to capture the spatial complexity of $\mathbf{K}$ also results in poor predictions for $\mathbf{t}_b$.

A solution to this problem of the rigidity/limited range of $\mathcal{L}$ lies in actually using the KDE model for $\Delta\mathbf{K}$, along with $\mathcal{L}$, to construct proposals for $\mathbf{K}_c^t$. These $\mathbf{K}_c^t$ could be compared with observations $\mathbf{k}^{(obs)}$ (under a measurement error assumption) and also used to generate $\mathbf{t}_b$ and compared to $\mathbf{t}_b^{(obs)}$, again under a tight measurement error model. This would undoubtedly improve the fit of the model to data, but will also result in a far more difficult inference problem - apart from inferring $\{\mathbf{w}, \delta\}$, we would also have to infer a $\mathbf{K}_c^t$ field under a suitable smoothness model. This, much harder, inference problem will be addressed in a subsequent publication.

3.7 Conclusions

We have developed a multiscale, statistical technique to reconstruct characteristics of a random, porous binary medium from partial observations. The binary medium consists of high and low permeability material in spatially varying proportions. The geometry of the medium consists of inclusions of the less abundant material embedded in a matrix of the other. The uneven spatial distribution of the two materials, $\mathbf{F}(\mathbf{x})$, and a characteristic inclusion lengthscale, δ, are the objects of inference; effective permeability $\mathbf{K}_e(x) = \mathcal{L}(\mathbf{F}(\mathbf{x}), \delta, \kappa)$ is calculated from a known dependence on them. The inclusions are too small to be resolved on the mesh. The observations consist of measurements of log-permeability ($\mathbf{k}^{(obs)}$) and breakthrough times ($\mathbf{t}_b^{(obs)}$) of a tracer from a tracer test at a set of sensor points. The log-permeability measurements inform on the large-scale variations in the domain of interest; the breakthrough times are governed strongly by the flow paths through the medium and are informative on the small-scale structures.

The reconstruction is posed as a Bayesian inverse problem, predicated on fitting a transport model to the data. The transport model is formulated at the coarse-scale, but with a statistical subgrid model that incorporates the impact of the fine, unresolved scales. The subgrid model (alternatively, the *link* function between the scales) is parametrized with the high-permeability material proportion, $\mathbf{F}(\mathbf{x})$, and characteristic length, δ. The inverse problem is regularized by expressing $\mathbf{F}(\mathbf{x})$ using Gaussian processes; its dimensionality is reduced by expanding $\mathbf{F}(\mathbf{x})$ in terms of a truncated Karhunen-Loève series. Fitting the model to data yields a joint distribution of the Karhunen-Loève weights, inferring characteristics of the variation of $\mathbf{F}(\mathbf{x})$ and $\mathbf{K}_e(x)$ in the domain. This distribution is realized without any approximations, using an adaptive MCMC sampler.

We find that $\mathbf{F}(\mathbf{x})$ obtained by jointly conditioning on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ is far more accurate than if it is obtained solely from $\mathbf{k}^{(obs)}$ or $\mathbf{t}_b^{(obs)}$. Inversion based on $\mathbf{k}^{(obs)}$ captures the large-scale variation correctly; however, the permeability field distribution so reconstructed predicts breakthrough times poorly since it lacks smaller details/structures responsible for contorted flow paths. The inversions based solely on $\mathbf{t}_b^{(obs)}$ are extremely poor, due to inability of these data to constrain large structures. However, when the two are put together, $\mathbf{k}^{(obs)}$ constrains the larger structures while $\mathbf{t}_b^{(obs)}$ is used to constrain the smaller ones. Thus the improved inferences (when conditioned jointly on $\mathbf{k}^{(obs)}$ and $\mathbf{t}_b^{(obs)}$) are not just due to more plentiful observations, but rather due to the different *types* of information in $\mathbf{k}^{(obs)}$ and $\mathbf{t}_b^{(obs)}$. Since the information is derived from different scales (large for $\mathbf{k}^{(obs)}$ and small for $\mathbf{t}_b^{(obs)}$), the inversion is fundamentally multiscale.

We found that the observations were not informative about δ, due to the limited nature of its impact on the permeability. For inclusions smaller than about a third of a grid block, the impact on permeability occurs close to the percolation threshold ($\mathbf{F}(\mathbf{x}) \approx 0.5$); this condition holds in only a few spots in the domain $\mathcal{D}$. This insensitive nature makes estimation of δ from sparse observations almost infeasible, unless the entire domain is close to the perco-

lation threshold. However, when $\mathbf{F}(\mathbf{x})$ in the domain was constrained to vary between tight bounds around $\mathbf{F}(\mathbf{x}) = 0.5$, δ could be inferred. However, the ability to infer δ was quickly lost as the bounds were loosened, and the posterior distribution of δ became entirely dominated by the model/structural errors in the inference. We devised a test (a "perturbed" posterior predictive test") to identify if, in a given problem, the posterior distribution of δ was being determined by structural errors or the essential physics of the problem. We also conducted a preliminary test to check the feasibility of performing inversions that explicitly include structural errors. We found that for our problem, heteroscedastic structural errors can be represented using a mixture of kernels (Epanetchnikov, in our case), leading to a slight improvement in the predictive skill of the inferred permeability fields.

Plots of marginalized posteriors of the objects of inference show that they are approximately Gaussian. While the priors used for them are Gaussian, the posterior shapes are surprising, given that both the fluid transport and the link function are strongly nonlinear.

We use the inferences of $\mathbf{F}(\mathbf{x})$ and δ to reconstruct realizations of the resolved binary medium on a fine mesh. We check the predictive skill of the ensemble of realizations by performing transport simulations with them, in an effort to recreate the observations from which they were inferred. We find that fine-scale realization developed by conditioning jointly on $\{\mathbf{k}^{(obs)}, \mathbf{t}_b^{(obs)}\}$ (i.e., "multiscale" realizations) generally have a greater predictive skill that those inferred with one type of information; further, they are more robust in situations where the flow patterns are very different from those used in the estimation. Further, we also identified a flow configuration where including $\{\mathbf{t}_b^{(obs)}\}$ in the inference conferred no advantage since they were not very informative on the small-scale features in important parts of the domain. Thus, in some cases, observations / sensor configurations may have to be crafted to the inference problem at hand to be properly informative.

Chapter 4

Multiscale Inference in Elliptic Problems using Multiscale Finite Elements

4.1 Multiscale Background

Imagine you are on a beach staring out at the ocean. You see the big waves rolling in and carrying surfers towards the shore, but you will also see smaller cross waves and little riffles that seem to move independently of the big waves. Also, if you were to sit in that spot long enough, the tide may come in and soak your blanket. Each of these observations represents a different length scale. Over short lengths, the riffles and small cross waves are observed. On a slightly larger time scale are the surfing waves, and on an even longer scale is the tide. This is a prime example of a multiscale system. Scale here effectively refers to the correlation length of the system.

It turns out that like the ocean, many natural systems exhibit multiscale behavior. Examples include permeability fields in subsurface flow as well as reactions in chemical kinetics. Fig. 4.1 gives an example of multiscale behavior in groundwater flow. Plotted is water depth below the surface elevation at a USGS monitoring station in Idaho. The first plot clearly shows yearly correlation lengths while the second plot[1] shows distinct daily patterns. As an aside, the daily patterns may in fact be due to an interesting effect called an earth tide. This occurs as the moon passes over a point on the earth and causes a dilation force on the bedrock. The force slightly stretches the media, opening up more space and reducing the pressure. Details of this phenomena can be found in [99] and the references therein.

Taking advantage of multiscale features when developing solution methods and sampling methods can dramatically improve efficiency. By concentrating on the scales of interest, it is sometimes possible to reduce much of the computational effort required for simulation. For example, in the ocean example mentioned above, certain applications may only be

[1] Note that the second plot is based on provisional data.

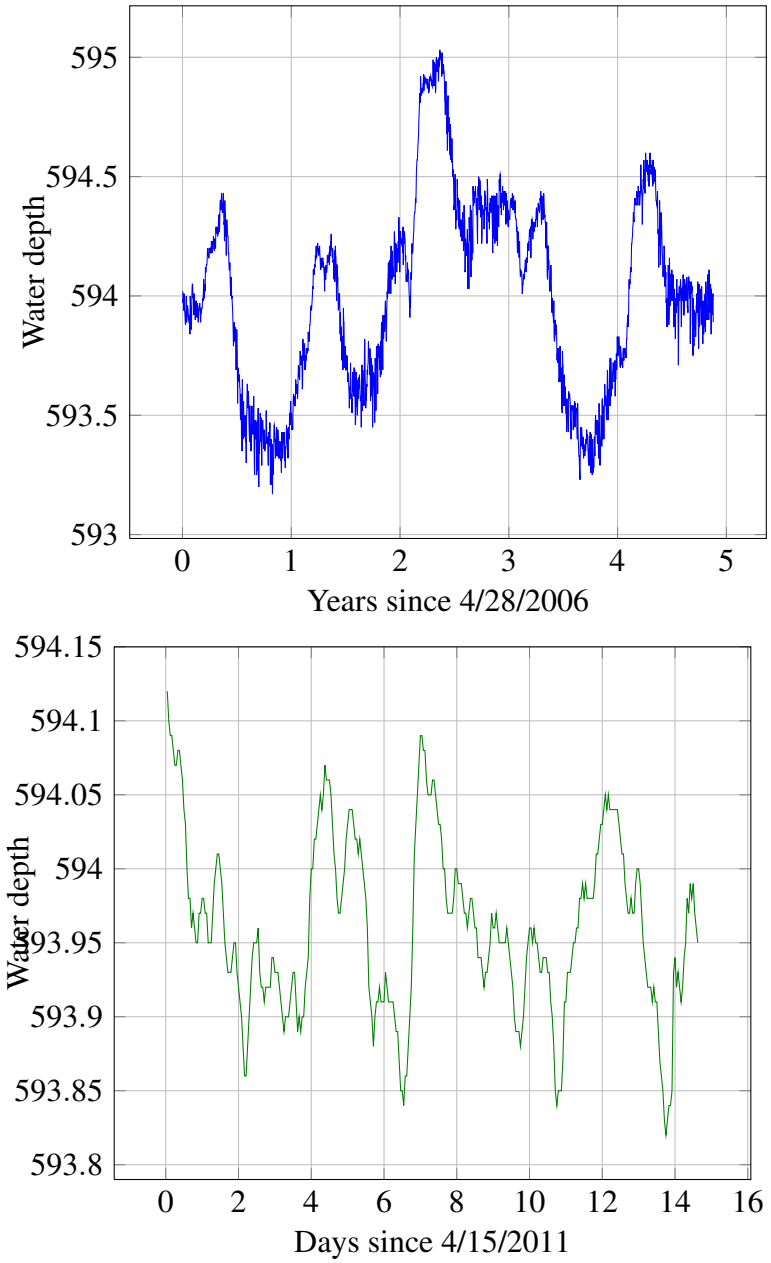


Figure 4.1: Example of groundwater observations with multiscale behavior. Not only are yearly patterns present, but daily length scales exist as well. Data comes from USGS well measurements in Bingham County, Idaho.

concerned with tidal effects while other may be heavily dependent on mid-sized waves and their erosion effects. This chapter will begin with a discussion of multiscale simulations in groundwater flow and will later turn to current multiscale sampling strategies. Before proceeding, we would like to emphasize that the multiscale tools developed here are in no way restricted to porous media flow. Many models with multiscale behavior can fit into this framework, especially in situations where the model output is much smoother than model input.

4.1.1 Multiscale Simulation Methods

4.1.1.1 Model Elliptic Equation

The focus of this work is on applications in porous media flow. A basic model used as a building block for many more sophisticated simulations is the pressure equation. To see where this model comes from, consider the conservation of mass for fluid flow through a small volume of porous media as shown in Fig. 4.2.

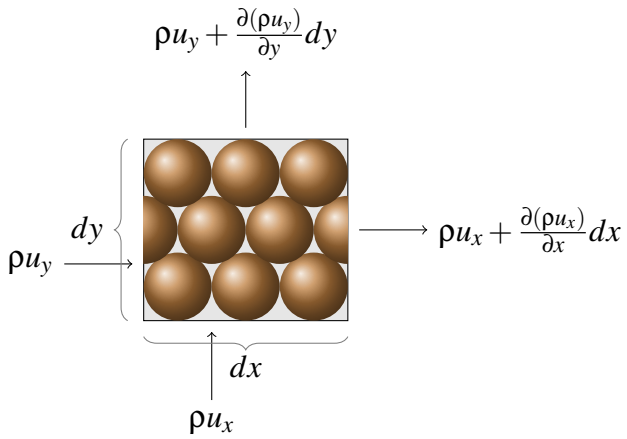


Figure 4.2: Conservation of mass in a porous media. When coupled with Darcy's equation as a constitutive law, this gives the familiar pressure equation.

Using the simple relationship

$$\frac{\partial M}{\partial t} = \text{Mass In} - \text{Mass Out}$$

with the mass flows in Fig. 4.2, leads to

$$\frac{\partial(\rho\phi)}{\partial t} dxdy = -\frac{\partial(\rho u_x)}{\partial x} dxdy - \frac{\partial(\rho u_y)}{\partial y} dxdy$$

where ρ is the density of the fluid, u_x is the inflow velocity in the x direction, u_y is the inflow velocity in the y direction. The term $\frac{\partial(\rho u_x)}{d} x$ is the rate of change in the x direction times the x distance traveled. Since we are working in the limit as $dx \to 0$, no higher order terms are needed. Also, ϕ is the porosity of the media. This term is needed because the fluid only occupies the pore space, so $\rho\phi dxdy$ is the mass in this representative volume and $\frac{\partial(\rho\phi)}{\partial t} dxdy$ is the mass rate of change. Assume the density is constant in time. Thus, we have

$$\begin{aligned} \rho \frac{\partial \phi}{\partial t} + \rho \nabla \cdot (u) &= 0 \\ \Rightarrow \frac{\partial \phi}{\partial t} + \nabla \cdot (u) &= 0 \end{aligned}$$

where the area terms $dxdy$ have been cancelled out. Note that these equations are on the continuum scale, no longer are the definitions only for the representative volume element. Assuming that ϕ is a linear function of pressure p, i.e $\phi = pC_t$ where C_t is the compressibility of the media, we have

$$C_t \frac{\partial p}{\partial t} + \nabla \cdot u = 0 \tag{4.1}$$

Obviously this equation is underdetermined; a constitutive law is needed. Here, Darcy's law is used to complete the system:

$$u = -\frac{\tilde{k}}{\mu}(\nabla p) \tag{4.2}$$

where $\tilde{k}$ is the permeability of the media and μ is the dynamic viscosity of the fluid. The previous two equations constitute the conservative form of the pressure equation. In this work, conservation is not critical.[2] The system considered here is the steady state system after non-dimensionalization, given by:

$$-\nabla \cdot (k \nabla p) = 0 \tag{4.3}$$

where k is now a spatially varying field representing the scaling of an intrinsic permeability field. From here on, k will be referred to simply as the permeability. Clearly, the pressure is a nonlinear function of the permeability. Thus, to use the predicted pressure from Eq. 4.3 as a tool for engineering design or water resource planning, not only do boundary conditions need to be tailored to the region of interest, but the permeability, $k(x)$, needs to be characterized as well. Typically, very few direct measurements of permeability exist and indirect observations of pressure must be used in conjunction with Eq. 4.3 to infer the permeability field. In this work, it will be assumed that no direct observations exist and m observations of the pressure $p(x)$ have been taken at a limited number of locations, $\{x_1, x_2, ..., x_m\} \in D$, where D is the spatial domain of interest. This vector of observations will be denoted d. Ill-posedness of the inverse problem arises when estimating $k(x)$ from d because the mapping $p(x) = G(k(x))$ defined by Eq. 4.3 acts as a nonlinear lowpass filter, removing high frequency effects of $k(x)$ from $p(x)$. This high frequency information cannot be recovered during inference.[3] This means that potentially large, high frequency changes in permeability will result in only minor pressure changes. Fig. 4.3 demonstrates this property for a one dimensional system and a two dimensional layer of the SPE10 dataset [100]. The pressure is found using the two $\log(k)$ fields on the left. In one dimension, the high frequency blue field is just the green field plus some correlated noise. Clearly the high frequency addition does not have a large impact on the pressure in the right plot. Intuitively it seems that because the pressure field is smoother, and can thus be more easily represented, only some of the information in the permeability field should be needed to construct the pressure field. That is, only certain scales of k have a significant impact on p.

[2]Conservation is not critical in this study because we are using the pressure equation as a proof of concept for our multiscale inference methodology. However, in many situations, especially when transport or saturation equations are also used, a conservative solution to the pressure equation is vital. It should be noted the methodology proposed herein can also be applied when using Mixed Finite Element Methods, ensuring a conservative solution.

[3]For continuous fields, if $k(x)$ has continuous derivatives of order m, $p(x)$ will have $m+1$ continuous derivatives.

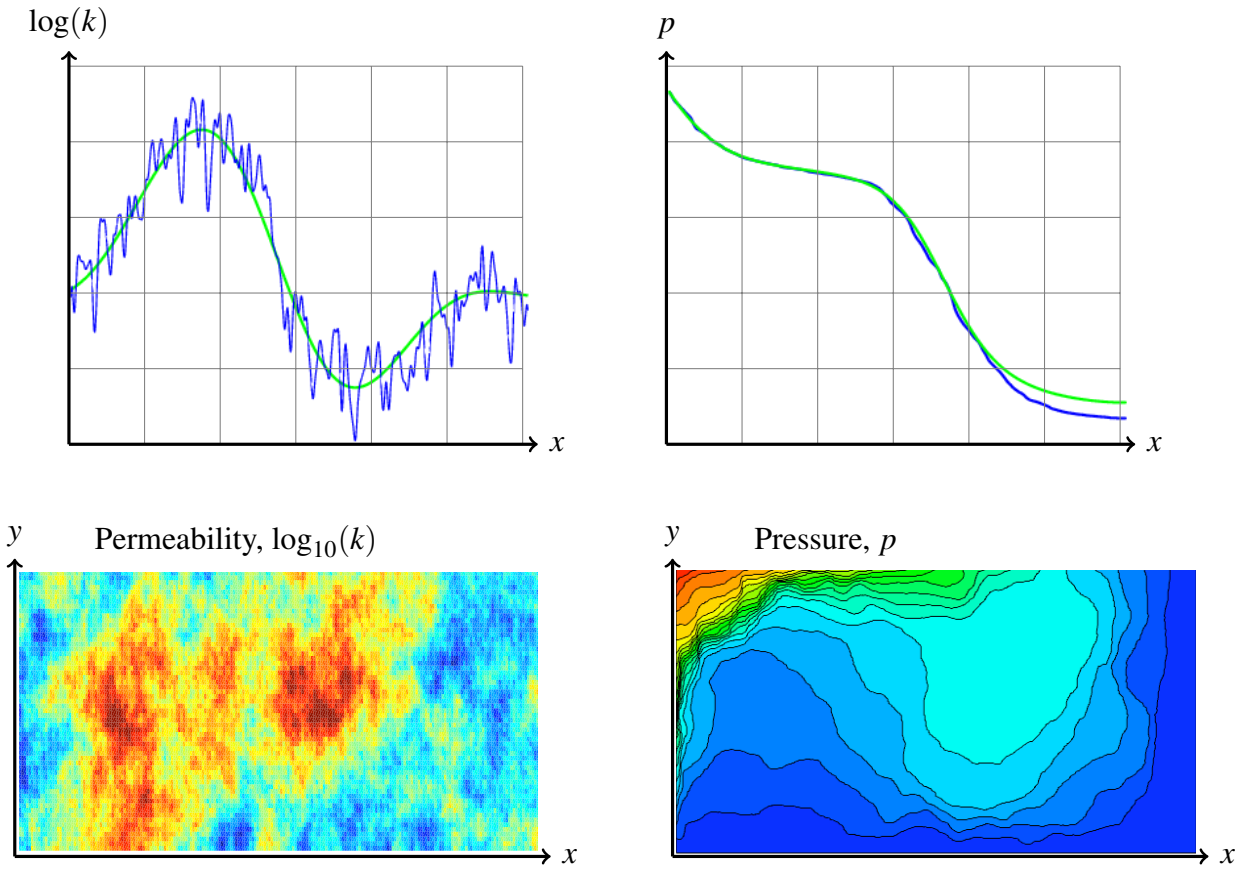


Figure 4.3: Example of multiscale field and corresponding pressures. The one dimensional case has a Dirichlet condition on the left and a homogeneous Neumann condition on the right. In the two dimensional case, all boundaries are fixed with Dirichlet conditions. Clearly, the pressure fields are much smoother than the permeability fields. The one dimensional case shows the smoothing effect of the elliptic operator and the two dimensional setting again shows the relative smoothness of the pressure field compared to the permeability field.

In large-scale (regional) simulations, the meshes needed to resolve fine scale features can be prohibitively large for performing global pressure solves. However, being much smoother than permeability, it seems reasonable to solve for pressure on a more computationally tractable coarse mesh. There are several methods of doing this. Often the most straightforward approach, known as upscaling, is to solve the pressure equation on a coarse mesh with a representative coarse permeability field. Once the coarse permeability has been found, the pressure equation is then solved on the coarse mesh. Note that upscaling the permeability field is an artificial coarsening of the system. Only pressure can be accurately represented on the coarse scale. Furthermore, with upscaling, the pressure is only represented on the coarse mesh, even though some fine scale features may exist and there is no way to find an approximate fine scale pressure field based on the coarse solution.

The desire to solve the pressure equation on a coarse mesh but maintain some fine scale

features is met in variational methods such as Multiscale Finite Element Methods, Variational Multiscale methods, and heterogeneous multiscale methods. The idea behind all of these solution strategies is to implicitly coarsen the pressure equation through a variational minimization. These methods allow the pressure to be approximated on a fine mesh, while only solving a small linear system for a coarse representation of the pressure. The following sections describe and contrast upscaling and these variational methods.

4.1.1.2 Upscaling

In the ocean example, upscaling could be defined as smoothing or averaging the waves into a quantity on the tidal scale. In the porous media setting, upscaling is just taking the average fine scale permeability over a coarse element. Choosing an appropriate average is not trivial. Analytic averages such as the arithmetic or harmonic mean would provide a simple form for the coarse permeability, but do not introduce any physics into the operation. To see that choosing the arithmetic mean is not a good choice, let $p = G(k)$ represent the nonlinear mapping from the permeability to pressure. We would like the coarse pressure solution to be the average of the fine scale pressure, but

$$\mathbb{E}[p] = \mathbb{E}[G(k)] \neq G(\mathbb{E}[k])$$

Clearly, taking the arithmetic average of k will not give the average pressure as desired and more sophisticated techniques need to be employed for upscaling to be effective. Nevertheless, the harmonic mean and arithmetic mean provide bounds on the effective permeability. As mentioned in the review paper, [101], and the references within, these means provide the so-called Wiener bounds:

$$\mu_h \leq k_{eff} \leq \mu_a \tag{4.4}$$

where k_{eff} is the effective permeability on a coarse element, μ_h is the harmonic mean over the coarse element, and μ_a is the arithmetic mean. Fig. 4.4 shows an example of various means for a layer of the SPE10 dataset. The averages were taken over 10 fine cells in the x-direction and 5 fine cells in the y-direction. After some inspection, it is apparent that on each coarse cell, the geometric mean lies between the harmonic and arithmetic means. In fact, many tighter bounds on the effective mean can be proved under various circumstances. See [101] for more information.

The Wiener bound is not tight, so to overcome the nonlinear relationship between k and p, a nonlinear upscaling based on model physics needs to be used. On each coarse element, a few local fine scale solves can be used to characterize the effective permeability. Farmer gives a good review of these methods in [102]. To give a general idea of the method, consider a two dimensional problem with quadrilateral coarse elements of size $h_x \times h_y$. In order to build a coarse permeability tensor, on each coarse element, two local solutions of the steady state pressure equation could be found. The first would enforce homogeneous Neumann (no-flow) boundary conditions on the top and bottom boundaries, with specified pressures at the other boundaries, and the second would switch the flow direction, with the Neumann conditions on the left and right. Once the pressure has been found, the total

flux, Q, through one of the Dirichlet boundaries can be found and Darcy's law will give the effective permeability:

$$k_{eff} = \frac{Qh_x}{A\Delta P} \tag{4.5}$$

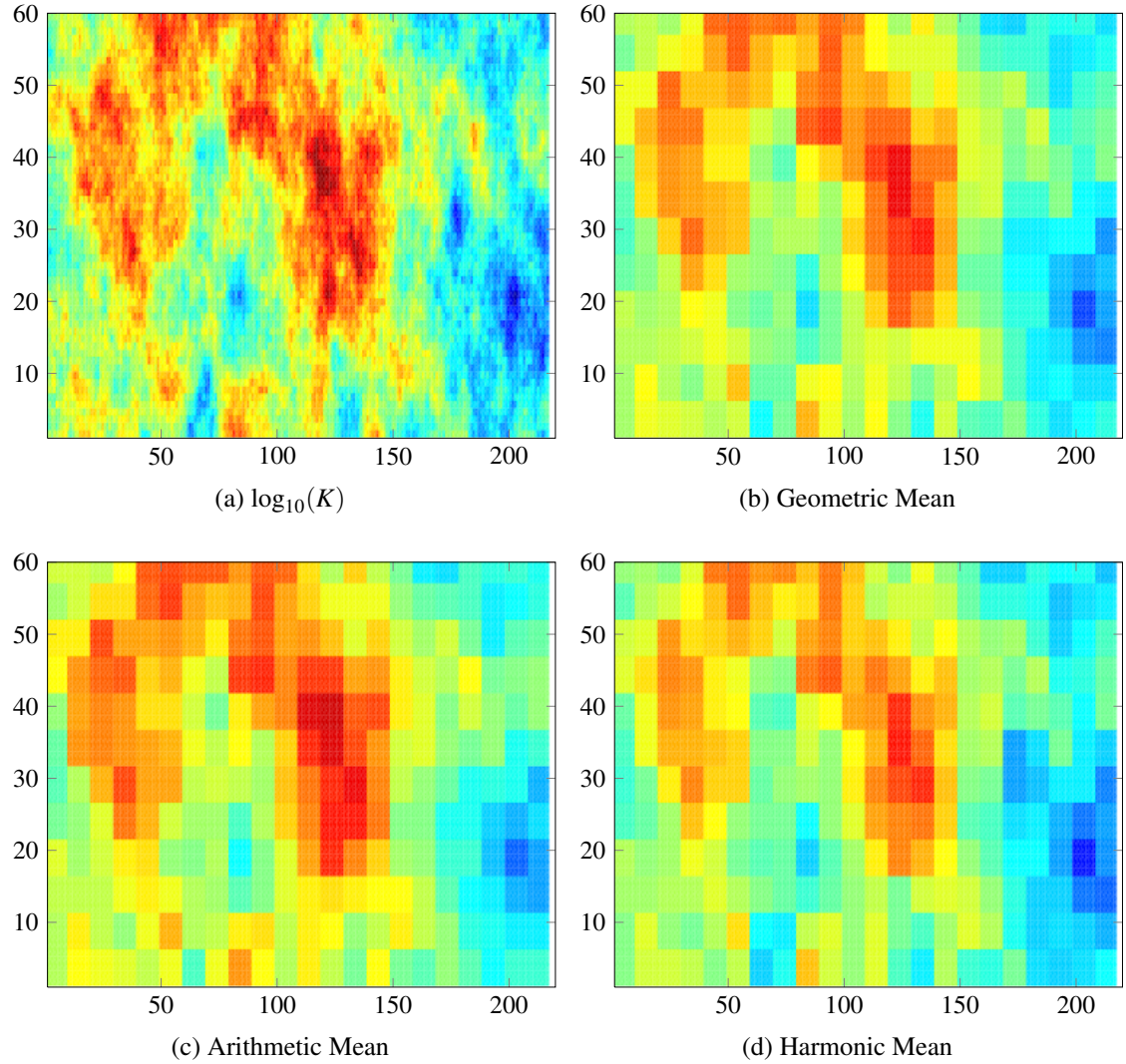
where k_{eff} is the effective permeability in the x-direction, Q is the computed flow rate, h_x is the horizontal size of the element, $A = h_x h_y$ is the element area, and ΔP is the pressure drop induced by the Dirichlet boundary conditions. In 1961, Warren and Price in [103] first introduced this idea and various extensions have become prevalent in petroleum engineering. In one dimension, choosing appropriate boundary conditions is trivial and the upscaled permeability can perform quite well. However, in higher dimension, not only does the computational cost increase because local flow simulations need to be computed to fill in the effective permeability tensor, but the choice of boundary conditions for the local solves is also not obvious.

In a fairy-tale world, computed effective permeabilities would not depend on the choice of local boundary conditions. However, we have not fallen into a rabbit hole and the boundary conditions can significantly impact the computed effective permeabilities. The methods for alleviating this dependence discussed in [102] include slightly increasing the computational domain size to reduce boundary effects, or using an approximate global solution to choose the boundary conditions. Additionally, some more recent publications, see [104] and [105], couple the mesh-generation process with the upscaling procedure to choose a mesh that reduces the boundary effects. The issue of appropriate boundary conditions for local solves is not unique to upscaling and will be discussed further in section 4.1.1.5 in the context of the Multiscale Finite Element Method (MsFEM).

The upscaling methods discussed above are useful when the fine scale permeability field is known. However, stochastic upscaling methods also exist when the fine scale permeability is represented as a stochastic field. Kitanidis provides an introduction to stochastic upscaling methods in [106]. The goal of stochastic upscaling is to use information about the fine scale distribution of k to develop effective permeabilities. The main approach discussed in [106] is based on small perturbation theory, where an assumption of small variance leads to the analytic form:

$$k_{eff} = k_g \exp\left(-\frac{\sigma^2}{n} + \frac{\sigma^2}{2}\right) \tag{4.6}$$

Here, k_g is the geometric mean of of the stochastic field k, σ^2 is the variance of $\log(k)$, and n is the domain dimension. Notice that in this case, the effective permeability is no longer stochastic. The randomness only exists on the fine scale. This upscaling operation is a method for taking a log-normal stochastic k field and computing an effective permeability that represents the general structure of the stochastic field. In section 4.2.3, we will take a different approach in which the coarse quantities are also represented through probability distributions. It will also become clear that a stochastic coarse representation can be useful in a multiscale inference setting.



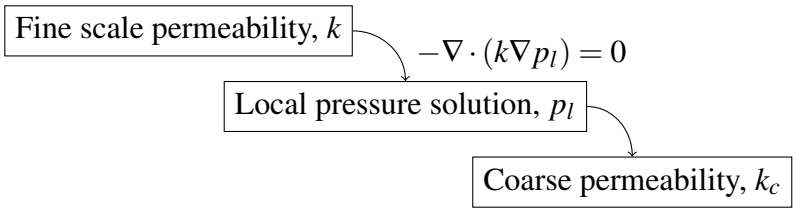
(a) $\log_{10}(K)$ (b) Geometric Mean

(c) Arithmetic Mean (d) Harmonic Mean

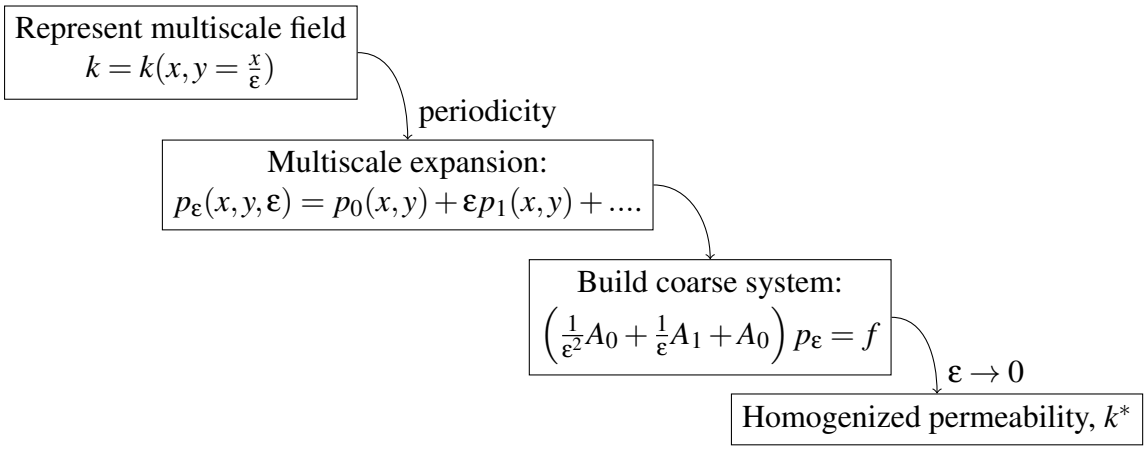
Figure 4.4: Example of geometric, arithmetic, and harmonic averaging. After inspection, note that the geometric means visually seems to be between the arithmetic and harmonic means, as required by the Wiener bounds. This field is $\log_{10}(K)$ for the 10th layer of the popular SPE10 dataset.

4.1.1.3 Homogenization

A good upscaling method introduces a physical model (the pressure equation here) to generate a coarse representation of the permeability. The process of upscaling to the coarse scale can be represented graphically as Fig. 4.5. An alternative analytic approach is homogenization. The object of homogenization is to find a multiscale expansion of the solution field whose coefficients can be found by solving a homogenized (coarsened) system. The expansion considers the action of the elliptic operator on a rapidly oscillation permeability field. The homogenization process is also outlined in Fig. 4.5. In this setting, the perme-



(a) Upscaling Process



(b) Homogenization Process

Figure 4.5: Comparison of Homogenization process and upscaling process. While upscaling builds a coarse permeability and uses the same governing equation, homogenization finds a homogenized permeability by putting constraints on the coarse operator.

ability is represented as a periodic field with period ε and the pressure equation becomes

$$A^\varepsilon p = -\nabla \cdot [k(x/\varepsilon)\nabla p] = f \tag{4.7}$$

The operator A^ε represents the pressure equation (an elliptic operator) built from a permeability field with period ε. Let $y = x/\varepsilon$ represent a "fast" coordinate. The goal is then to find a solution expansion of the form

$$p_\varepsilon(x) = p_0(x,y) + \varepsilon p_1(x,y) + \varepsilon^2 p_2(x,y) + ... \tag{4.8}$$

The introduction of the dependence of p on y requires the use of a new total derivative, given by:

$$\nabla = \nabla_x + \frac{1}{\varepsilon}\nabla_y \tag{4.9}$$

Using this in the pressure equation gives:

$$-(\nabla_x + \frac{1}{\varepsilon}\nabla_y) \cdot \left[k(y)(\nabla_x p + \frac{1}{\varepsilon}\nabla_y p)\right] = f \tag{4.10}$$

which after some algebra gives the compact form:

$$A^\varepsilon p = \frac{1}{\varepsilon^2}A_0 + \frac{1}{\varepsilon}A_1 + A_2 \tag{4.11}$$

where

$$\begin{aligned} A_0 &= -\nabla_y \cdot (k(y)\nabla_y) \\ A_1 &= -\nabla_y \cdot (k(y)\nabla_x) - \nabla_x \cdot (k(y)\nabla_y) \\ A_2 &= -\nabla_x \cdot (k(y)\nabla_x) \end{aligned}$$

The pressure equation is now

$$\left(\frac{1}{\varepsilon^2}A_0 + \frac{1}{\varepsilon}A_1 + A_0\right)(p_0(x,y) + \varepsilon p_1(x,y) + \varepsilon^2 p_2(x,y) + ...) = f \tag{4.12}$$

Collecting terms, we find a system of equations:

$$A_0 p_0 = 0 \tag{4.13}$$
$$A_0 u_1 + A_1 u_0 = 0 \tag{4.14}$$
$$A_0 u_2 + A_1 u_1 + A_2 u_0 = f \tag{4.15}$$

Following [107], it is possible to show that p_0 is independent of y, and Eq. 4.14 can then be written as

$$-\frac{\partial}{\partial y_i}\left(a_{ij}(y)\frac{\partial}{\partial y_j}\right)p_1 = \left(\frac{\partial}{\partial y_i}a_{ij}(y)\right)\frac{\partial p}{\partial x_j}(x) \tag{4.16}$$

Now define χ_j as the solution to

$$\frac{\partial}{\partial y_i}\left(a_{ij}(y)\frac{\partial}{\partial y_j}\right)\chi_j = -\frac{\partial}{\partial y_i}a_{ij}(y) \tag{4.17}$$

After some technical arguments and looking at the limit $\varepsilon \to 0$, [107], shows that the homogenized equation becomes

$$-\frac{\partial}{\partial x_i}\left(k^*_{ij}\frac{\partial}{\partial x_j}\right)p = f \tag{4.18}$$

Note, this is just the usual pressure equation with a rigorous choice of effective permeability! In fact, the homogenized permeability, k^*, is given by

$$a^*_{ij} = \frac{1}{|Y|}\left(\int_Y (a_{ij} - a_{ik}\frac{\partial \chi_j}{\partial y_k})dy\right) \tag{4.19}$$

The literature on homogenization theory is vast and only the tip of the iceberg of homogenization for elliptic problems was introduced here. An interested reader can find a thorough discussion of homogenization and its relationship with averaging in [108].

4.1.1.4 Variational Methods

Homogenization and upscaling reduce the degrees of freedom in an elliptic solve and thus reduce the computational cost of the global solve. The idea is that being smoother than the permeability, computing pressure only requires lumped information about the permeability field. The lumped effective permeability however, is a fictitious field developed simply for numerical reasons. The vast literature on upscaling shows that it can be effective, but a more satisfying methodology would maintain the fine scale representation of k and only develop a coarse approximation to p, the field that truly exists on a smoother scale. Multiscale variational methods provide just such an alternative. Additionally, the variational approaches can often provide a systematic way of projecting the coarse pressure to a fine scale pressure that exhibits important features seen in a full fine scale solution. Several variational multiscale methods exist, including the Variational Multiscale Method (VMM) [109, 110], Heterogeneous Multiscale Methods (HMM) [111, 112], Subgrid upscaling [113], Multiscale finite element methods [107, 114], and the Multiscale finite volume method [115, 116]. All of these methods are very similar and on some specific problems can be equivalent. In fact, some papers, such as [117] do not even provide a clear distinction between these methods. Subtleties aside, the basic idea is to perform a limited number of local pressure solves to build a set of basis functions for use in a coarse scale (often called global) coupling. Here, we will focus on the multiscale finite element method which uses the local basis functions in a coarse scale Galerkin finite element formulation.

4.1.1.5 Multiscale Finite Element Methods

The formulation here follows closely the introduction in [107]. For more general information and useful background information on the standard finite element approaches, see [118].

Let Ω be the domain of interest, where the pressure equation is to be solved. Consider a coarse triangulation T_h of Ω into finite elements. For each element of T_h, a fine mesh can also be created. A simple choice is to use quadrilateral elements for the coarse grid and then either quadrilaterals or simplex elements on the fine grid. Let p lie in a function space X, usually chosen as $H_0^1(\Omega)$. Then the usual weak form reads: *find $u \in X$ such that:*

$$a(u,v) = b(v)\ \forall v \in X$$

The weak form of the pressure equation is given by

$$a(u,v) = \int_\Omega k(x)\nabla u \cdot \nabla v dx = \int_\Omega f v dx = b(v)\ \forall v \in X \tag{4.20}$$

From here, most finite element methods will discretize the system by defining nodal basis functions, ϕ_i^0 , (often linear) and only consider $v \in W_h = \text{span}\{\phi_i^0\}$. However, it is possible to choose these basis functions such that fine scale information is embedded into the coarse scale weak formulation. Consider the basis function ϕ_i defined on the support of ϕ_i^0. Since

ϕ_i^0 is a nodal basis function, the support is over the elements containing node i as a vertex. Let K be one of these coarse elements and define ϕ_i to satisfy:

$$-\nabla \cdot (k(x)\nabla\phi_i) = 0 \qquad x \in K \tag{4.21}$$

$$\phi_i = \phi_i^0 \qquad x \in \partial K \tag{4.22}$$

where ∂K is the coarse element boundary. Clearly, the MsFEM basis function, ϕ_i matches ϕ_i^0 on the element boundaries, but solves the homogeneous pressure equation on the element interior. Thus, effects of the fine scale k on the pressure are embedded into the coarse scale formulation through ϕ_i. Fig. 4.6 shows an example of an MsFEM basis function over one coarse element. In the case of Fig. 4.6, the coarse elements are quadrilaterals, so the figure only shows one quarter of a complete nodal basis function.

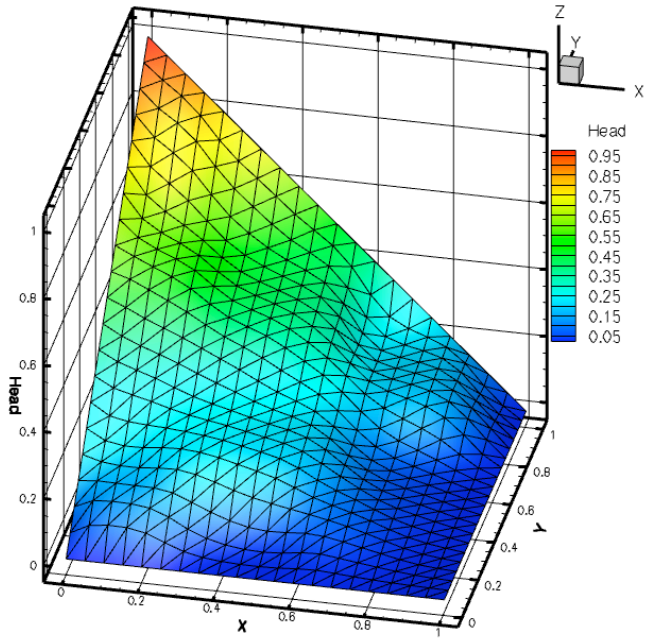


Figure 4.6: Illustration of MsFEM Basis function. This illustrates the basis function evaluated over one coarse element.

Let $P_h = \text{span}\{\phi_i\}$ be the space spanned by the MsFEM basis functions. To discretize the system using MsFEM, we first represent the pressure in terms of the MsFEM basis functions:

$$p_h = \sum_i p_i \phi_i \tag{4.23}$$

and use the following weak formulation:

$$\sum_K \int_K k\nabla p_h \nabla v_h dx = \int_\Omega v_h dx \;\forall v_h \in P_h \tag{4.24}$$

This is a Galerkin projection; however, a Petrov-Galerkin projection could also be used. Either form will yield a linear system:

$$A_c p_{\text{nodal}} = b \tag{4.25}$$

where p_{nodal} is a vector of basis function weights, A_c is the coarse stiffness matrix, and b represents source terms. Elements of A_c are given by

$$a_{ij} = \sum_K \int_K k \nabla \phi_j \nabla \phi_i dx \tag{4.26}$$

The integrals inside the summation will be referred to as *elemental integrals*. As pointed out by [107], in implementation it is more straightforward to use the relationship:

$$A_c = D^T A_f D$$

than the integral form. Here, column i of D is a vector holding the discrete basis function ϕ_i, and A_f is the global fine scale stiffness matrix. A_f would be the result of using a standard finite element approach on the global fine scale mesh. Boundary conditions for Eq. 4.25 can be implemented identically to standard finite element approaches. Once the coarse pressure p_{nodal} has been found, projecting this solution back to the fine scale is trivial:

$$p_f^* = D p_{\text{nodal}} \tag{4.27}$$

Since P_h is spanned by a relatively small number of basis functions compared to a full fine scale formulation, solving the resulting linear system is much more computationally tractable. However, local solves are still required to find ϕ_i. The local solves are easily parallelized. Local solves on each element are independent of all other elements. Furthermore, the basis functions do not need to be recomputed when the source terms in f are changed, a useful feature when considering MsFEM as the solver in a well-design framework, or other application where many solves are needed but only f changes between solves.

In the definition of the MsFEM basis functions the constraint $\phi_i = \phi_i^0$ on ∂K was used as the boundary conditions of the local homogeneous pressure equation. However, it is well known that these boundary conditions do not represent any fine scale heterogeneities and significant errors can be produced on problems without a large scale separation in the permeability field. Possible solutions include oversampling, [107, 119], where the local solve is performed on a computational domain larger than a single coarse element and then truncated, or solving one dimensional problems along the boundary, as in [115, 116]. Here we use the latter method and first solve a one dimensional problem along each edge of K. For ϕ_i in K, the pressure is set to 1 at node i and zero elsewhere for the one dimensional solves. The one dimensional solutions are then used as Dirichlet conditions in the two dimensional solve for ϕ_i on the interior of K.

In practice, even these boundary conditions can introduce error because of a lack of scale separation in the permeability field. Errors may still exist, coming into play when the coarse element size resonates with lengthscales of the permeability. Overcoming this requires the use of limited global information which can require a global fine scale solve. See [107] for a detailed discussion of global information in MsFEM as well as a thorough error analysis based on homogenization theory.

4.1.2 Multiscale Sampling Methods

So far, this chapter has focused solely on multiscale solution strategies for the pressure equation. However, using multiscale strategies within a sampling method such as MCMC can also be advantageous. This section will give a brief overview of current multiscale MCMC methods and point out where there is room for improvement.

4.1.2.1 Coupled Metropolis Sampling

Following the derivations in [74] and [120], consider a case where we have two representations of the permeability field, k_1 is a fine field, and k_2 is a coarsened field. The fine scale field is donned with a Gaussian prior:

$$k_1 \sim N(\mu, \Sigma_1)$$

Furthermore, a linear equation is used to transfer information from the fine scale k_1 to the coarse field k_2:

$$k_2 = A_1 k_1 \tag{4.28}$$

The linear link A_1, usually represents some type of linear average, [120]. As already mentioned in the upscaling section, a linear relationship between fine and coarse permeability fields does not satisfy the problem physics. This is not a killer in the coupled Metropolis setting discussed in [120]. In that case, a linear map is necessary and any errors in the mapping are alleviated by introducing an additive error term to the upscaling. The new multiscale inference procedure outlined in the next chapter avoids this issue by allowing for a nonlinear relationship between scales.

In a two scale problem, the multiscale Metropolis method proposed by Higdon in [121] has two MCMC chains simultaneously exploring the posterior distributions of k_1 and k_2. In porous media, each scale will have a likelihood distribution, $\pi(d|k_1)$ and $\pi(d|k_2)$, where d are pressure observations. The coarse likelihood will require a simulation using a coarse discretization of the pressure equation with permeability k_2. Similarly, the fine scale likelihood will require a more expensive pressure equation solve using a fine grid with k_1. Occasionally, the chains will swap information through an interscale swap proposal, $q(k_1', k_2'|k_1^n, k_2^n)$. The proposal for k_2 is simply taken as the linear upscaling, Ak_1. Furthermore, k_1 and k_2 are assumed to both have normal densities, so the fine proposal will simply be the prior for k_1 conditioned on $Ak_1 = k_2$. So, we have:

$$q(k_1', k_2'|k_1^n, k_2^n) = q(k_1'|Ak_1 = k_2^n) q(k_2'|k_2 = Ak_1) \tag{4.29}$$

The density $q(k_2'|k_2 = Ak_1^n)$ is a Dirac measure at Ak_1^n. Using this proposal in the Metropolis-Hastings rule gives

$$\alpha = \min\left\{1, \frac{\pi(d|k_2')\pi(k_2')\pi(d|k_1')\pi(k_1')q(k_1^n, k_2^n|k_1', k_2')}{\pi(d|k_2)\pi(k_2)\pi(d|k_1)\pi(k_1)q(k_1', k_2'|k_1^n, k_2^n)}\right\} \tag{4.30}$$

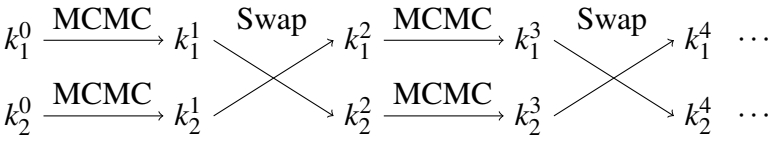


Figure 4.7: Illustration of the Multiscale Metropolis MCMC method. Here, steps labeled with MCMC correspond to steps with standard single scale proposals. It is also possible to have more than one MCMC step before a swap is performed.

Fig. 4.7 shows a graphical representation of the Coupled Metropolis MCMC.

Multiscale Metropolis increases MCMC effectiveness through better mixing of the fine scale MCMC chain and can help the chain explore more of the fine scale parameter space. However, each step of the chain still requires a fine scale simulation to evaluate the fine scale likelihood. Like computing MsFEM basis functions offline for a well-field design problem, we would like to avoid any fine scale simulations beyond offline preprocessing. The basic idea behind Metropolis-Coupled MCMC is that a coarse representation of k is easier to sample because it reduces the problem dimension. This is useful concept that will be expanded later to the point where knowing the coarse quantities alone is sufficient to predict the pressure and no fine scale solves are needed during the MCMC routines.

4.1.2.2 Proposal Filtering with MsFEM

Taking one step closer to avoiding fine scale simulation altogether is the work on two stage proposals in [122] and expanded to Langevin type proposals in [123]. Conceptually, these methods use a coarse multiscale finite volume (MsFV) simulation to filter out bad proposals before any fine scale simulations need to be performed. The papers are focused on saturation simulations, so the major computational cost is in solving fine scale saturation equations. Thus, a coarse MsFV pressure is used with a crudely averaged coarse saturation solve. While some errors in the coarse solve exist, the information is still valuable for filtering out bad proposal points. Consider a coarse forward model $G^*(k)$ and a likelihood function $\pi^*(k')$ that depends only on outputs from the coarse forward model. Now, let k' be a proposal from a standard Gaussian proposal mechanism. During a preconditioning step, this proposal is accepted as the *true* proposal with probability

$$\alpha^*(k^n,k') = \min\left\{1, \frac{q(k^n|k')\pi^*(k')}{q(k'|k^n)\pi^*(k^n)}\right\} \tag{4.31}$$

where $q(k|k^n)$ is the usual proposal density. This means that the final, fine scale, proposal comes from the density:

$$Q(k|k^n) = \alpha^*(k^n,k)q(k|k^n) + \left(1 - \int \alpha(k^n,k)q(k|k^n)dk\right)\delta_{k^n}(k) \tag{4.32}$$

Using this proposal in the Metropolis-Hastings rule gives a final acceptance probability of

$$\alpha = \min\left\{1, \frac{Q(k^n|k)\pi(k)}{Q(k|k^n)\pi(k^n)}\right\} \tag{4.33}$$

The advantage of this approach is that when k' is rejected in the initial stage, no fine scale saturation solve is necessary and a significant computational savings is experienced. In [123], this procedure is generalized to a Langevin MCMC case where the posterior gradient is approximated using only coarse scale solves. Unfortunately, the advantages in these methods are not as useful in the steady state single flow scenario considered here because some fine scale simulations are still required after a proposed point passes the initial rejection stage. As will be shown in the next chapter, when only considering the pressure equation, reasonable assumptions and some offline preprocessing allow for a complete decoupling of the fine and coarse scales. This simultaneously reduces the MCMC parameter dimension and the computational expense of a forward evaluation.

4.2 Multiscale Inference Framework

4.2.1 Conditional Independence

For steady state single phase flow based on the pressure equation, solving for pressure with MsFEM can be broken into two stages. In the first stage, local solves are used to build a coarse scale stiffness matrix and in the second stage, this stiffness matrix is used to compute the pressure. If desired, the coarse pressure can also be projected onto the fine scale using the same basis functions. In the forward problem, information moves from permeability to coarse stiffness matrix, to pressure. Thus, when solving the inference problem, information should travel from pressure observations, through the coarse scale stiffness matrix, to the permeability field. The previous work in [122, 123] discussed above, takes partial advantage of this to pre-reject bad proposals based only on a coarse approximation. Here we take further advantage of the MsFEM structure to eliminate the need for any global fine scale solves. We focus on a static inference problem where steady state pressure observations, d, are given. The goal is then to characterize a permeability field on the fine scale, k, conditioned on the data. That is, we wish to find the density $\pi(k|d)$. As usual Bayes' rule will be used to obtain the posterior in terms of a likelihood and prior:

$$\pi(k|d) \propto \pi(d|k)\pi(k) \tag{4.34}$$

In the MsFEM setting, the intermediate coarse scale stiffness matrix can also be included in the inference. Thus, the joint posterior of the stiffness matrix A and the permeability k can be represented as:

$$\pi(k,A|d) \propto \pi(d|k,A)\pi(k,A)$$

Here, the notation $\pi(k,A|d)$ refers to the joint distribution of the permeability values and the entries in the stiffness matrix. Note that in some instances, the number of variables

representing A can be reduced by considering the distribution of elemental integrals. This topic will be discussed later. Expanding the joint prior with the law of total probability gives

$$\pi(k,A|d) \propto \pi(d|k,A)\pi(k|A)\pi(A) \tag{4.35}$$

At this point it may seem awkward to place a prior, $\pi(A)$, on the stiffness matrix entries but we will see that this is a critical step in decoupling the fine and coarse scales. Now, concentrate on the likelihood term $\pi(d|k,A)$. This is the probability of the data given both the permeability and the coarse scale stiffness matrix. However, as we know from MsFEM, only the stiffness matrix is needed to construct the pressure, p. This is an important observation: when using MsFEM, the stiffness matrix A is "sufficient" for p. Sufficiency is an important notion in statistics; see [124] for more information or [125] for a discussion in the context of Jeffreys' conditioning. Since d is an observation of p, we have that A is sufficient to describe d. Probabilistically, this idea corresponds to the conditional independence of d and k given A:

$$\pi(d,k|A) = \pi(d|A)\pi(k|A) \tag{4.36}$$

which implies

$$\pi(d|k,A) = \pi(d|A) \tag{4.37}$$

Using this in Eq. 4.35 gives

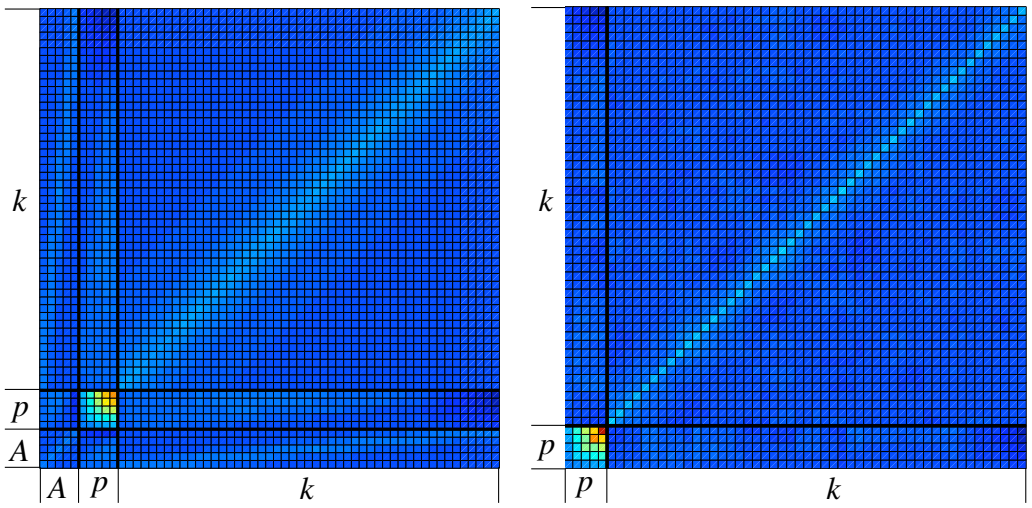
$$\pi(k,A|d) \propto \pi(d|A)\pi(k|A)\pi(A) \tag{4.38}$$

Fig. 4.8 validates the conditional independence assumption. 500000 samples were generated from a prior on k and the corresponding entries of the stiffness matrix and pressure were found. In the joint covariance, three regions are marked. The first columns and last rows are the entries in the stiffness matrix, the next group inward is the pressure, and the large section filling the upper right is the permeability. In the joint covariance, faint correlations can be seen between all the fields. However, looking at the joint conditional distribution $\pi(k,p|A^*)$, no significant correlations are seen between the pressure and permeability. In order to perform the conditioning, a particular stiffness matrix, A^*, needed to be used. Several choices were tested and all produced similar results. Note the pressure in Fig. 4.8 was computed with a fine scale solve. Similar results for MsFEM show that this is not a feature of the solver being used, but a consequence of A being sufficient to describe p.

4.2.2 Multiscale MCMC Formulation

In order to characterize the new target distribution $\pi(k,A|d)$, MCMC will be used. In order to sample the joint density, a joint proposal, $q(k',A'|k,A)$ will be used. In this setting the Metropolis-Hastings acceptance probability becomes:

$$\alpha = \min\{1,\gamma\}$$



(a) Joint Covariance (b) Conditioned Covariance

Figure 4.8: Validation of conditional independence assumption.

where

$$\begin{aligned}\gamma &= \frac{\pi(d|k',A')\pi(k',A')q(k,A|k',A')}{\pi(d|k,A)\pi(k,A)q(k',A'|k,A)} \\ &= \frac{\pi(d|A')\pi(k'|A')\pi(A')q(k|A)q(A|A')}{\pi(d|A)\pi(k|A)\pi(A)q(k'|A')q(A'|A)}\end{aligned} \tag{4.39}$$

The second equality comes from simply expanding the joint distributions into conditionals and using the conditional independence assumption. This equation looks much more difficult to use than the usual expression when we want $\pi(k|d)$. This equation has both coarse stiffness matrix terms and permeability terms. However, if we could sample exactly from $\pi(k|A)$, the corresponding portion of the proposal could be exact and major simplifications would ensue. Assume therefore

$$q(k|A) = \pi(k|A)$$

then, after some simplification, we have

$$\gamma = \frac{\pi(d|A')\pi(A')q(A|A')}{\pi(d|A)\pi(A)q(A'|A)} \tag{4.40}$$

which is nothing more than a standard Metropolis-Hastings rule in just the low dimensional A! Using this acceptance probability, the relatively low dimensional coarse scale stiffness matrix could be inferred without ever needing the high dimensional fine scale field, k. Alas, there is no free lunch. The two major assumptions made during this derivation need to be addressed. Specifically,

- a means of evaluating the prior density for the entries in A needs to be found.

- a method to sample from the conditional distribution $\pi(k|A)$ needs to be developed

A prior on A is needed to perform the coarse sampling. Additionally, sampling from $\pi(k|A)$ will allow posterior realizations of k to be generated from each sample of the stiffness matrix. The proposed multiscale sampling strategy is summarized in Fig. 4.9. In this figure, data is used with a prior on A to generate posterior samples with standard MCMC tools such as DRAM. Then, a nonlinear iterative conditioning procedure is used to generate posterior samples of k for each sample of A. The next section will discuss how a prior for A can be developed offline from a prior on k. The following section will then introduce a new nonlinear conditioning approach that will allow samples of $\pi(k|A)$ to be generated. Numerical results and analysis will be left to the next chapter.

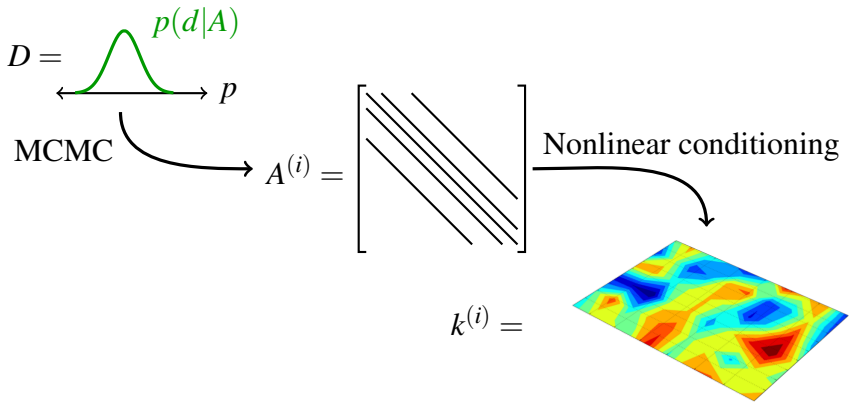


Figure 4.9: Illustration of full multiscale inference using procedure using MsFEM.

4.2.3 Upscaling the Prior

The prior distribution on the permeability field is derived from expert opinions, a priori site characterizations, and an overall geologic understanding of the medium. However, this knowledge does not directly transfer to a prior on A. It is not clear how a knowledge of sediment formations will affect A. Consider the case when k is a log-normal random process. That is $k = \exp[Y]$ where $Y \sim GP(\mu_y, \Sigma_Y)$. The covariance of Y at two locations from a two point covariance kernel. The kernel maps the spatial distance between two permeability regions to a correlation. Multiplying by a stationary variance then gives the covariance. A general covariance kernel may follow an exponential form:

$$\text{Cov}[Y(x_i), Y(x_j)] = \sigma \exp\left[-\frac{1}{p}\left(\frac{\|x_i - x_j\|_2}{L}\right)^p\right] \tag{4.41}$$

where p is a power parameter, L is a length scale, and σ is the variance. When $p = 1$ we have an exponential kernel, and $p = 2$ is a Gaussian kernel. Exponential kernels tend to be much rougher than Gaussian kernels and usually need to be represented on a finer mesh. Representing the prior with a covariance kernel such as this is convenient; only 3

parameters and the mean are needed to completely describe a high dimensional distribution. This section attempts to demonstrate that with some offline Monte Carlo preprocessing, a reference mapping can be built that takes the kernel parameters for Y to parameters describing a log-normal distribution for the elemental integrals. In this setting, Y will be a discrete dimensional field described by a Multivariate Gaussian distribution. This is a simply a discretization of Gaussian process definition. Begin by considering a one dimensional Petrov-Galerkin formulation of MsFEM on a one dimensional mesh with nodes at $x_0, x_1, ...$ Construction of the coarse scale stiffness matrix in this case is given by:

$$a_{ij} = \int_{x_{i-1}}^{x_i} k(x) \frac{\partial \phi_i}{\partial x} \frac{\partial \phi_j^0}{\partial x} dx + \int_{x_i}^{x_{i+1}} k(x) \frac{\partial \phi_i}{\partial x} \frac{\partial \phi_j^0}{\partial x} dx \tag{4.42}$$

where ϕ_i is the MsFEM basis function defined as 1 at node i and 0 at x_{i-1} and x_{i+1}. Within these two elements, ϕ_i satisfies:

$$\frac{\partial}{\partial x}\left(k(x)\frac{\partial \phi_i}{\partial x}\right) = 0$$

so

$$k(x)\frac{\partial \phi_i}{\partial x} = c$$

for some constant c. Dividing both sides by $k(x)$ and integrating over $[x_{i-1}, x_i]$ we see that

$$\begin{aligned} \int_{x_{i-1}}^{x_i} \frac{\partial \phi_i}{\partial x} dx &= \int_{x_{i-1}}^{x_i} \frac{c}{k(x)} dx \\ \Rightarrow \phi_i|_{x_{i-1}}^{x_i} &= \int_{x_{i-1}}^{x_i} \frac{c}{k(x)} dx \\ \Rightarrow c &= \frac{1}{\int_{x_{i-1}}^{x_i} \frac{1}{k(x)} dx} \end{aligned} \tag{4.43}$$

An equivalent process can be repeated over $[x_i, x_{i+1}]$ to obtain

$$\begin{aligned} a_{ij} &= \int_{x_{i-1}}^{x_i} \frac{1}{\int_{x_{i-1}}^{x_i} \frac{1}{k(x)} dx} \frac{\partial \phi_j^0}{\partial x} dx + \int_{x_i}^{x_{i+1}} \frac{1}{\int_{x_i}^{x_{i+1}} \frac{1}{k(x)} dx} \frac{\partial \phi_j^0}{\partial x} dx \\ &= \frac{1}{\int_{x_{i-1}}^{x_i} \frac{1}{k(x)} dx} \int_{x_{i-1}}^{x_i} \frac{\partial \phi_j^0}{\partial x} dx + \frac{1}{\int_{x_i}^{x_{i+1}} \frac{1}{k(x)} dx} \int_{x_i}^{x_{i+1}} \frac{\partial \phi_j^0}{\partial x} dx \end{aligned} \tag{4.44}$$

Using nodal linear functions for the test basis functions ϕ_i^0 simplifies this expression to:

$$a_{i,i-1} = -\frac{1}{\int_{x_{i-1}}^{x_i} \frac{1}{k(x)} dx} \tag{4.45}$$

$$a_{i,i+1} = -\frac{1}{\int_{x_i}^{x_{i+1}} \frac{1}{k(x)} dx} \tag{4.46}$$

$$a_{i,i} = \frac{1}{\int_{x_{i-1}}^{x_i} \frac{1}{k(x)} dx} + \frac{1}{\int_{x_i}^{x_{i+1}} \frac{1}{k(x)} dx} \tag{4.47}$$

Denote the coarse element from x_{i-1} to x_i as element i, then the elemental integral for element i is given by:

$$e_i = \frac{1}{\int_{x_{i-1}}^{x_i} \frac{1}{k(x)} dx} \tag{4.48}$$

Note that the stiffness matrix entries are composed of the elemental integrals. Clearly, $a_{i,i} = e_i + e_{i+1}$, $a_{i,i-1} = -e_i$, and $a_{i,i+1} = -e_{i+1}$. Thus, a prior on the stiffness matrix, $\pi(A)$ is equivalent to having a prior on the elemental integrals: $\pi(e)$. This holds in one dimension, in higher dimensions more care is required.

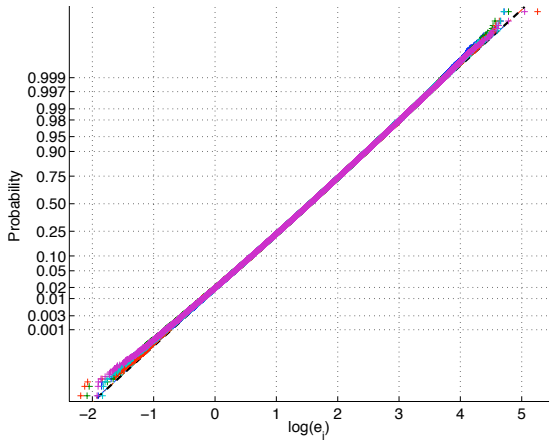
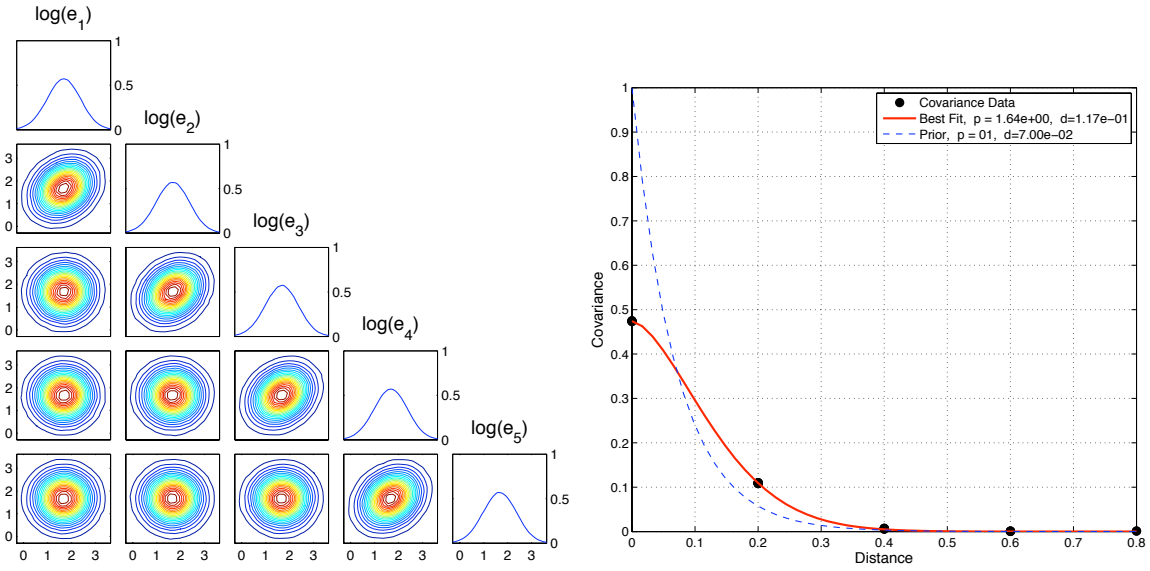


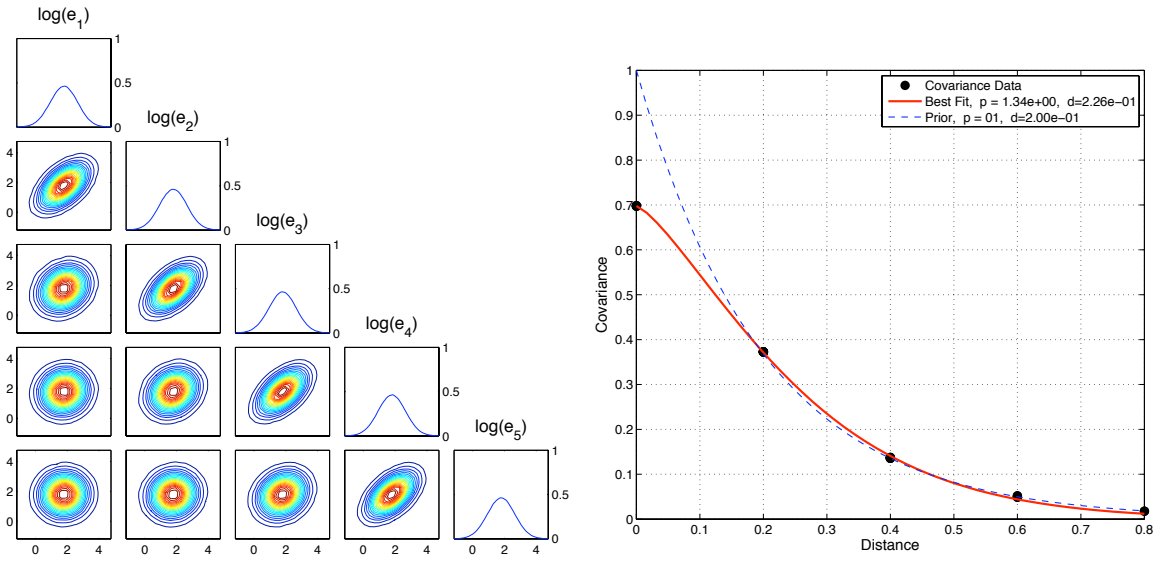
Figure 4.10: Quantile-Quantile plot comparing sample quantiles to true log-normal quantiles. The distribution is an exact fit when the scatter points are exactly linear.

Let k be defined as a piecewise constant field in one dimension following a log-normal distribution, $k = \exp[Y]$ for $Y \sim N(\mu_Y, \Sigma_{YY})$. Furthermore, let Σ_{YY} be described through an exponential covariance kernel of the form in Eq. 4.41. Taking 400,000 Monte Carlo of Y and computing e with Eq. 4.48 shows that e is approximately joint log-normally distributed as well. Figs. 4.11 and 4.12 show the joint distribution of the elemental integrals when 5 coarse elements were used with 20 fine elements for each coarse element for a total of 100 fine elements over $[0,1]$. The fitted covariance kernels are also shown. The densities pass the eyeball test of normality, Fig. 4.10 also shows that these e are well approximated

by log-normal distribution. Through the majority of parameter space, the sample quantiles match what is expected of a log-normal density. However, the tail probabilities are off, as you would expect in an approximation and with a finite number of samples. Each color corresponds to a different elemental integral. The plot only shows a few dimensions but similar results are found in all dimensions.



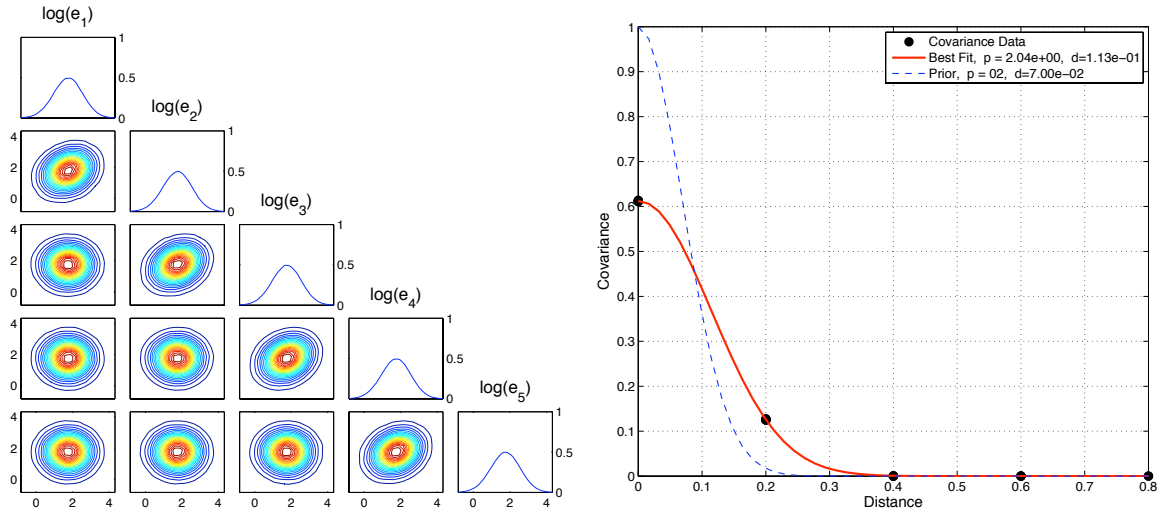
(a) $p = 1, d = 0.07$



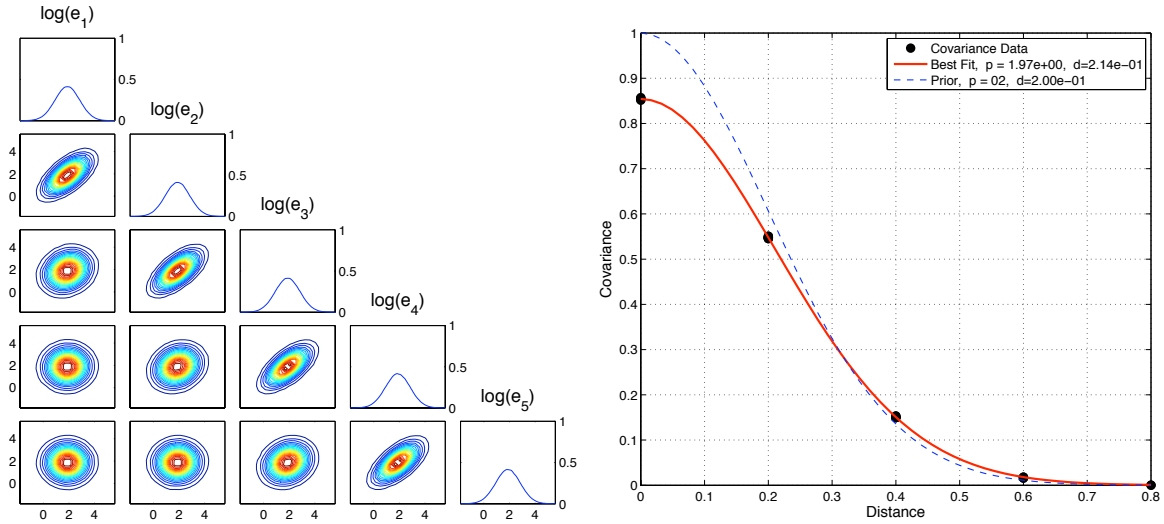
(b) $p = 1, d = 0.2$

Figure 4.11: Distribution of $\log(e)$ based on Monte-Carlo sampling of a piecewise constant log-normal k field with exponential prior.

Therefore, for a prior parametrized by a given length scale, power, and variance, the distribution of elemental integrals can be well approximated by a log-normal distribution of a similar form to the prior. The fitted parameters for $\pi(log(e))$ were found using maximum



(a) $p = 2, d = 0.07$



(b) $p = 2, d = 0.2$

Figure 4.12: Distribution of $\log(e)$ based on Monte-Carlo sampling of a piecewise constant log-normal k field with Gaussian prior.

likelihood estimation. The optimization was done with a simple implementation of the Levenberg-Marquadt algorithm with constant Marquadt parameter λ.

It is infeasible to perform the Monte Carlo sampling and fitting each time a prior for e is needed. However, using a database of previously run samples provides an empirical way to map the parameters describing $\pi(k)$ to parameters characterizing $\pi(e)$. All the work to build the database can be done before an inference is required and reused each time the algorithm is run. Performing the same fitting procedure used in Figs. 4.11 and 4.12, the surfaces in Fig. 4.13 were constructed. Some sampling error is visible, especially in the power plots. These mappings are critical to this multiscale framework; given the parameters of the prior, the corresponding parameters of the elemental integrals can be trivially found. Distances

between elemental integrals are the distances between the center of each coarse element. It may seem unusual to represent a vector quantity e as a spatially dependent parameter. However, in a case where k was a continuous random process, the elemental integrals could be interpreted as the reciprocal of a moving average of $\frac{1}{k(x)}$ over a length $\delta x = x_i - x_{i-1}$. The elemental integrals used in the coarse stiffness matrix are then snapshots of a continuously varying random process.

Fig. 4.13 shows the coarse length scale is only a function of the fine correlation length. For significant correlation lengths, the coarse variance is also only a function of the fine variance. However, as the fine scale correlation length d_f goes to zero, so does the coarse variance. Conceptually, this is expected. When $\sigma_f \to 0$, Y becomes white noise on top of μ_Y, so integration on each coarse element will give approximately the same quantity, and $\text{var}[\log(e)] \to 0$.

4.2.3.1 Proof of Positive Definiteness in One Dimension

Using the information in Fig. 4.13, a coarse scale prior distribution is defined in terms of the prior on k. Using the log-normal $\pi(e)$, samples of the stiffness matrix can easily be generated. It can also be shown that these samples of A are guaranteed to be positive definite in the one dimensional setting. Here we will show that in one dimensional, the coarse scale stiffness matrix is identical to a matrix derived with linear basis functions and a coarse representation of k. Thus, coming from a standard finite element discretization of an SPD operator, we conclude that by construction, any sample of $\pi(e)$ will create a SPD stiffness matrix A.

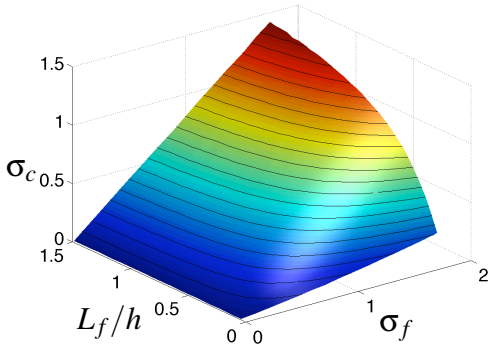
Using linear basis functions on the coarse grid in a standard Galerkin projection; the following elemental integrals, $\tilde{e}$, are created:

$$\tilde{e}_i = \int_{x_{i-1}}^{x_i} \tilde{k}(x) \frac{\partial \phi_i^0}{\partial x} \frac{\partial \phi_{i-1}^0}{\partial x} dx \tag{4.49}$$

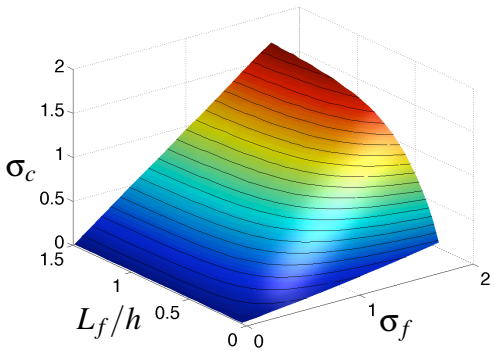
Note that a different permeability field $\tilde{k}$ was used here. The field $\tilde{k}$ is assumed to be piecewise constant on the *coarse* grid, whereas the field k used in the MsFEM formulation is piecewise constant on the *fine* grid. From standard finite element theory, using Eq. 4.49 is guaranteed to produce a symmetric positive definite stiffness matrix when $\tilde{k} > 0$. Thus, to guarantee A will be symmetric positive definite for any sample of $\pi(e)$ it suffices to show there exists a field $\tilde{k}$ such that $\tilde{e}_i = d_i$ for any e.

Proof. Let e_i^* be a sample of $\pi(e)$. On each element, choose

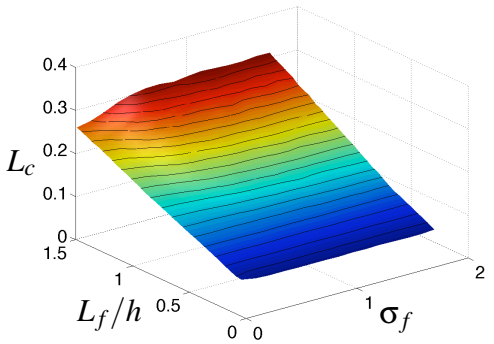
$$\tilde{k}_i = -h^2 e_i^* \tag{4.50}$$



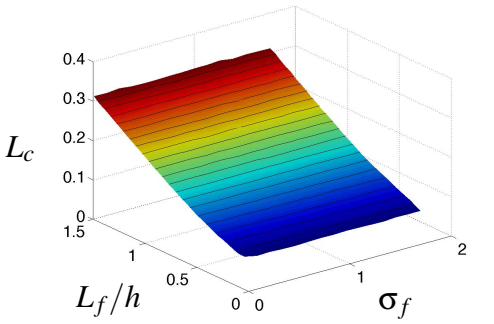
(a) Coarse σ_c with exponential kernel in $\pi(Y)$



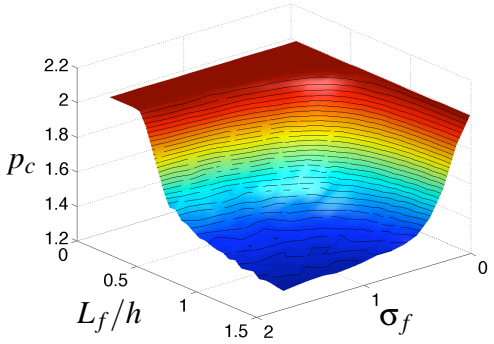
(b) Coarse σ_c with Gaussian kernel in $\pi(Y)$



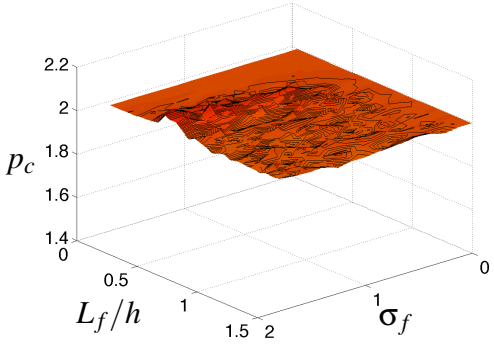
(c) Coarse d_c with exponential kernel in $\pi(Y)$



(d) Coarse L_c with Gaussian kernel in $\pi(Y)$



(e) Coarse p with exponential kernel in $\pi(Y)$



(f) Coarse p with Gaussian kernel in $\pi(Y)$

Figure 4.13: Mapping from $\pi(Y)$ to $\pi(log(e))$ for exponential and Gaussian covariance kernels in $\pi(Y)$. The fine scale correlation length, L_f, has been scaled by the coarse element size h to allow for arbitrary discretizations. Notice the x axis of the power plots is in the opposite direction to the other plots.

where $h = x_i - x_{i-1}$. Then

$$\tilde{e}_i = \int_{x_{i-1}}^{x_i} -h^2 e_i^* \frac{\partial \phi_i^0}{\partial x} \frac{\partial \phi_{i-1}^0}{\partial x} dx \tag{4.51}$$

$$= -h^2 e_i^* \int_{x_{i-1}}^{x_i} \frac{\partial \phi_i^0}{\partial x} \frac{\partial \phi_{i-1}^0}{\partial x} dx \tag{4.52}$$

$$= e_i^* \frac{h^2}{h^2} \tag{4.53}$$

The last step comes from the fact that ϕ_i^0 and ϕ_{i-1}^0 are linear with slopes $\frac{1}{h}$ and $\frac{-1}{h}$ respectively. Thus, the MsFEM stiffness matrix from any sample of $\pi(e)$ is equivalent to a coarse Galerkin discretization of the pressure equation on a permeability field $\tilde{k}$ and is guaranteed to be symmetric positive definite. □

In higher dimensions, it is not clear whether a similar proof can be found. Currently, our only solace is that a proposal of e resulting in a non-SPD stiffness matrix may result in an inaccurate pressure and have a very small probability of being accepted.

4.2.4 Iterative Nonlinear Conditioning

At first glance, generating a sample of $\pi(k|A)$ may seem just as difficult as sampling a posterior. Indeed, using Bayes' rule would give

$$\pi(k|A) \propto \pi(A|k)\pi(k)$$

which is no easier to compute than $\pi(k|d)$. However, for each sample $A^{(i)}$, we are conditioning on a point density, which means that $\pi(k|A)$ will only have positive measure on a low dimensional manifold constrained by $A^{(i)}$. The constraint comes from the nonlinear relationship map between k to A. Sampling from $\pi(k|A)$ can be seen as sampling from a slice of the prior on k. Fig. 4.14 illustrates this idea.

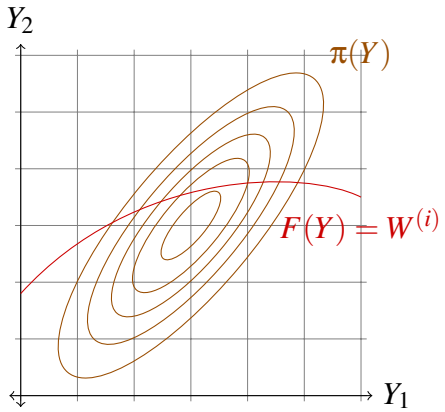


Figure 4.14: Illustration of $\pi(k|A)$ as a constraint on the prior. We wish to sample from the "slice" of $\pi(k)$ along the constraint.

Ultimately, the goal is to characterize k through A. To do this, it will be convenient to work with Gaussian distributions; however, k and A are distributed log-normally. Thus, we will instead consider $Y = \log(k)$ and $W = \log(e)$ in the following analysis.

4.2.4.1 Linear Conditioning

Let $F(Y)$ be the mapping from Y to W. Assume for the moment that this mapping is linear: $F(Y) = FY$. In this case, the joint distribution of Y and W is given by

$$\begin{bmatrix} W \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} F\mu_Y \\ \mu_Y \end{bmatrix}, \begin{bmatrix} F\Sigma_{YY}F^T & F\Sigma_{YY} \\ \Sigma_{YY}F^T & \Sigma_{YY} \end{bmatrix}\right) \tag{4.54}$$

The dimension of W is equal to the number of coarse elements in the system and will be significantly smaller than the dimension of Y, which is equal to the total number of fine elements. The goal here is to express Y in terms of W. To do this, we can use the Schur complement of the joint covariance, which can also be viewed here as a block Gaussian elimination. We obtain:

$$\begin{aligned} I\!E[Y|W] &= \mu_Y + \Sigma_{YY}F^T\left(F\Sigma_{YY}F^T\right)^{-1}(W - F\mu_Y) && (4.55) \\ \mathrm{Cov}[Y|W] &= \Sigma_Y - \Sigma_{YY}F^T\left(F\Sigma_{YY}F^T\right)^{-1}F\Sigma_{YY} && (4.56) \end{aligned}$$

which is more commonly written as

$$\begin{aligned} I\!E[Y|W] &= \mu_Y + K(W - F\mu_Y) && (4.57) \\ \mathrm{Cov}[Y|W] &= \Sigma_{YY} - KF\Sigma_{YY} && (4.58) \end{aligned}$$

where $K = \Sigma_{YY}F^T\left(F\Sigma_{YY}F^T\right)^{-1}$ is often called the Kalman gain. Clearly, in the linear Gaussian case a simple analytic form exists for $\pi(Y|W)$ in terms of the prior mean and covariance. Alternatively, independent samples of $\pi(Y|W)$ can be generated by taking samples of the prior, and using a modified version of Eq. 4.55 to move the sample onto the constraint. The moments of the moved samples are guaranteed to converge to the analytic mean and covariance using the linear update equation:

$$Y_a^* = Y^* + \Sigma_{YY}F^T\left(F\Sigma_{YY}F^T\right)^{-1}(W - FY^*) \tag{4.59}$$

The following proof shows that sampling from the prior and then moving those samples to the constraint with Eq. 4.59 will converge in distribution to $\pi(k|A)$.

Proof. Let $Y \sim N(\mu, \Sigma)$ be a multivariate Gaussian random variable with dimension N. Given a linear operator F of size $M \times N$ with $M \leq N$, define the multivariate Gaussian $W = FY$. Thus, $I\!E[Y|W = b]$ is defined in Eq. 4.55 and $\mathrm{Cov}[Y|W = b]$ is given by Eq. 4.56. Now, consider the linear operation:

$$Z = Y + F\Sigma(F\Sigma F^T)^{-1}(b - FY)$$

where Z is a multivariate Gaussian random variable of dimension N. The expectation of Z is:

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}[Y + F\Sigma(F\Sigma F^T)^{-1}(b - FY)] \\
&= \mathbb{E}[Y] + F\Sigma(F\Sigma F^T)^{-1}\mathbb{E}[(b - FY)] \\
&= \mathbb{E}[Y] + F\Sigma(F\Sigma F^T)^{-1}(b - F\mathbb{E}[Y]) \\
&= \mu + F\Sigma(F\Sigma F^T)^{-1}(b - F\mu)
\end{aligned}$$

and the covariance of Z is:

$$\begin{aligned}
\mathrm{Cov}[Z] &= \mathbb{E}[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])^T] \\
&= \mathbb{E}\left[\left(Y + F\Sigma(F\Sigma F^T)^{-1}(b - FY) - \mu - F\Sigma(F\Sigma F^T)^{-1}(b - F\mu)\right)\right. \\
&\quad \left.* \left(Y + F\Sigma(F\Sigma F^T)^{-1}(b - FY) - \mu - F\Sigma(F\Sigma F^T)^{-1}(b - F\mu)\right)^T\right] \\
&= \mathbb{E}\left[\left(Y - \mu - F\Sigma(F\Sigma F^T)^{-1}FY + F\Sigma(F\Sigma F^T)^{-1}F\mu\right)\right. \\
&\quad \left.* \left(Y - \mu - F\Sigma(F\Sigma F^T)^{-1}FY + F\Sigma(F\Sigma F^T)^{-1}F\mu\right)^T\right] \\
&= \mathbb{E}\left[\left(Y - \mu - F\Sigma(F\Sigma F^T)^{-1}F(Y - \mu)\right)\left(Y - \mu - F\Sigma(F\Sigma F^T)^{-1}F(Y - \mu)\right)^T\right] \\
&= \mathbb{E}\left[\left(Y - \left(\mu + F\Sigma(F\Sigma F^T)^{-1}(FY - F\mu)\right)\right)\left(Y - \left(\mu + F\Sigma(F\Sigma F^T)^{-1}(FY - F\mu)\right)\right)^T\right] \\
&= \mathbb{E}\left[\left(Y - \mathbb{E}[Y|FY]\right)\left(Y - \mathbb{E}[Y|FY]\right)^T\right] \\
&= \mathrm{Cov}[Y|FY]
\end{aligned}$$

The mean and covariance match the mean and covariance of W. Therefore, since the mean and covariance of a multivariate Gaussian completely describe the distribution, $Z \stackrel{d}{=} W$. This implies sampling from Y and then updating the samples is equivalent to sampling from W since the update is linear. □

4.2.4.2 Nonlinear Conditioning

The idea of taking a sample from the prior and moving it to a representative location is not unique to this work. This concept is also fundamental to the Ensemble Kalman Filter (EnKF) [126]. In a dynamic problem, the EnKF works by running several samples (an ensemble) through the forward model until an observation time is reached. Then the observation is incorporated by using Eq. 4.59 on the entire ensemble. The samples are then again propagated through the forward model until another observation is reached. The basic EnKF assumes a linear observation function.[4] That is, even if the forward simulation is nonlinear, the relationship between simulation output and the observational data is linear. Interestingly, sampling $\pi(k|A)$ is identical to a incorporating observations in the EnKF with

[4] Additionally, the EnKF assumes Gaussianity of the ensemble members when incorporating observations.

a nonlinear observation function. To surmount the nonlinearity, we follow years of mathematical tradition and simply approximate the nonlinear $F(Y)$ with a linearization around Y_0:

$$F(Y) \approx F(Y_0) + J(Y_0)(Y - Y_0) \tag{4.60}$$

where $J(Y_0)$ is the Jacobian matrix of $F(Y)$ evaluated at Y_0. The constraint, $F(Y) = W$ can now be approximated with

$$F(Y_0) + J(Y_0)(Y - Y_0) = W \tag{4.61}$$
$$\Rightarrow J(Y_0)Y = W + J(Y_0)_0 - F(Y_0) \tag{4.62}$$

and a linear update can be performed:

$$\begin{aligned} Y_1 &= Y_0 + \Sigma_{YY} J^T \left(J \Sigma_{YY} J^T\right)^{-1} (W + JY_0 - F(Y_0) - JY_0) \\ &= Y_0 + \Sigma_{YY} J^T \left(J \Sigma_{YY} J^T\right)^{-1} (W - F(Y_0)) \end{aligned} \tag{4.63}$$

where $J = J(Y_0)$. In this situation, Y_1 is an approximation to a sample of $\pi(Y|W)$. The linear approximation gets Y_1 closer to satisfying $F(Y) = W$, but will not in general satisfy the constraint. To obtain a better approximation, the linearization process is repeated at Y_1 and another linear conditioning step can be performed. This iteration is repeated until a point Y_n satisfying $\|F(Y_n) - W\|_2 < \varepsilon$ is reached for some small ε. Linearization errors are significant in this process and only an approximate sample of $\pi(k|A)$ will be generated. The errors can be corrected after the constraint is reached through additional measures such as MCMC. However, samples more closely representing $\pi(k|A)$ will require less correction and less MCMC will be required. Use of MCMC is discussed more thoroughly in later sections. For now, the focus remains on approximately getting to the constraint.

Several variations of the simple iterative linearization introduced above can be used. In fact, the linear update step can be viewed as a forward Euler discretization of the nonlinear ordinary differential equation

$$\frac{\partial y}{\partial t} = \Sigma_{YY} J^T(y) \left(J(y) \Sigma_{YY} J^T(y)\right)^{-1} (W - F(y)) \tag{4.64}$$

with initial condition $y(t = 0) = Y_0$. By using higher order integration techniques, the linearization error can be reduced. In fact, as the timestep $\Delta t \to 0$, the high order terms originally truncated in Eq. 4.61 become negligible and virtually no approximation error exists. However, computational costs will require the use of a more moderate Δt, and some linearization errors will undoubtedly be introduced.

Using high order methods to integrate Eq. 4.64 is similar to the mollified Ensemble Kalman Filters, where data is incorporated into the system gradually using a continuous time derivation of the EnKF [127]. Another related concept is the idea of nonlinearly constrained Kalman Filters. Sampling from $\pi(Y|W)$ is nearly identical to the EnKF update equation in

that setting. The nonlinear constrained EnKF has been thoroughly discussed in [128] and more recently in [129]. The idea proposed by Geeter et al. is to replace the hard constraint $F(Y) = W$ with a series of weak constraints $J(Y_{n-1})Y_n = W + \varepsilon_Y$ for some $\varepsilon \sim N(0, \eta_n \Gamma_0)$. Where

$$\Gamma_0 = \alpha J \Sigma_{YY} J^T$$

This follows from the fact that linearly conditioning a Gaussian random variable on observations with variance σ is equivalent to conditioning the random variable on the same observation 10 times with observation variance 10σ. In the presence of this article noise, the update equation becomes

$$Y_{n+1} = Y_n + \Sigma_{YY} J^T \left(J \Sigma_{YY} J^T + \eta \Gamma_0 \right)^{-1} (W - F(Y_n)) \tag{4.65}$$

Where the observation error $\eta\Gamma_0$ is meant to roughly approximate the linearization error at each step. As the iteration progress and $F(Y_n)$ approaches W, the incremental stepsize $Y_{n+1} - Y_n$ will decrease and a decrease in linearization error will ensue. To incorporate this into the artificial noise, [128] proposes shrinking η with

$$\eta_n = \alpha \exp(-n)$$

In true EnKF fashion, a covariance is approximated at each step, replacing Σ_{YY} in subsequent position updates. The complete method for approximately sampling $\pi(Y|W)$ is outlined in algorithm 1.

Algorithm 1 Smoothly constrained Kalman filter sampling

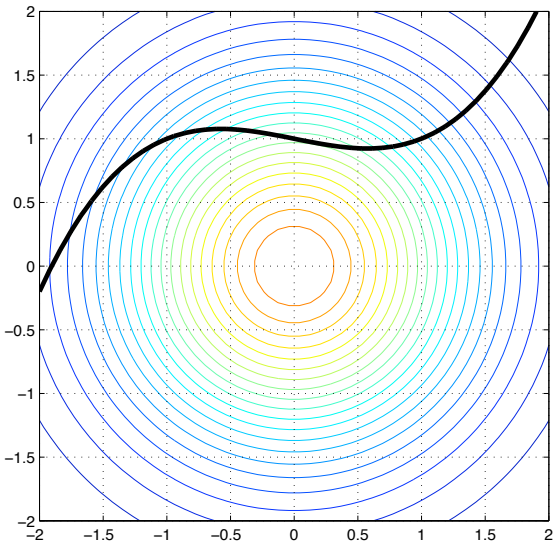
Require: Prior distribution, $\pi(Y) \sim N(\mu_Y, \Sigma_{YY})$
Error parameter, η
Nonlinear constraint $F(Y) = W$, Stopping tolerance δ.

1: Draw a sample, Y_0 from $\pi(Y)$
2: Set initial covariance, $\Sigma_0 = \Sigma_{YY}$
3: Set initial error, $\Gamma_0 = \alpha J(Y) \Sigma_0 J^T(Y)$;
4: $n = 0$
5: **while** $\|F(Y_n) - W\|_2 > \delta$ **do**
6: $\quad \Gamma = \Gamma_0 \exp(-n)$
7: $\quad n = n + 1$
8: $\quad Y_{n+1} = Y_n + \Sigma_n J^T(Y_n) \left(J(Y_n) \Sigma_n J^T(Y_n) + \Gamma \right)^{-1} (W - F(Y_n))$
9: $\quad \Sigma_{n+1} = \Sigma_n - \Sigma_n J^T(Y_n) \left(J(Y_n) \Sigma_n J^T(Y_n) + \Gamma \right)^{-1} J(Y_n) \Sigma_n$
10: **end while**
11: Return Y_{n+1} as approximate sample of $\pi(Y|W)$

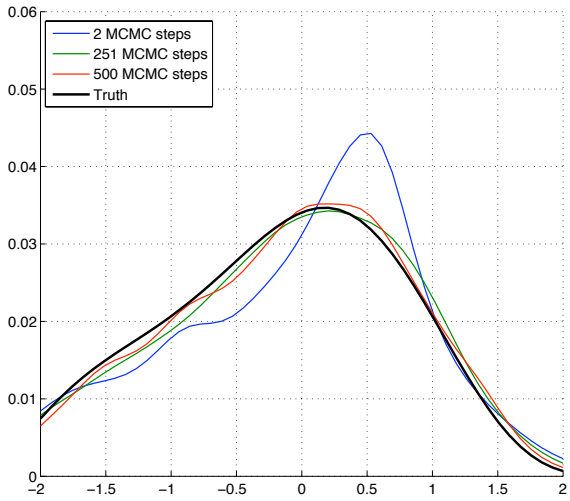
Fig. 4.15 shows a comparison of this smoothly constrained Kalman filter (SCKF) approach with a 4th-5th order adaptive Runge-Kutta integration of Eq. 4.64. The prior distribution is simply $N(0, I)$ in two dimensions with a constraint

$$y = 0.2x^3 - 0.2x + 1$$

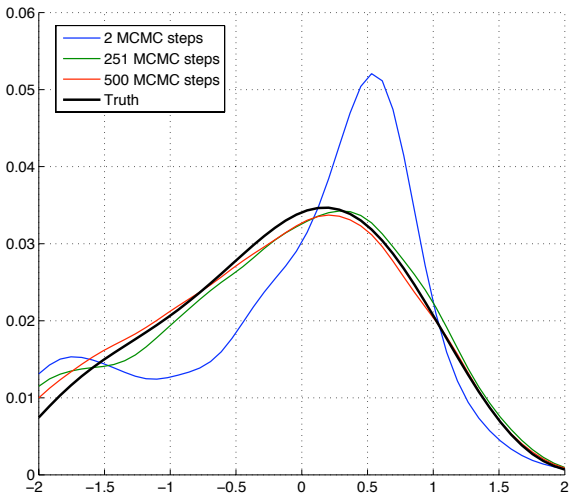
Clearly, the SCKF method provides a better initial approximation compared to the Runge-Kutta solver. However, after a few hundred MCMC steps along the constraint, any error is erased and the samples represent the true posterior shown in black. In general, as the problem becomes more nonlinear, more MCMC correction will be needed to ensure the final sample represents $\pi(Y|W)$.



(a) Nonlinear constraint



(b) SCKF with $\alpha = 0.01$



(c) ODE45

Figure 4.15: Comparison of high order ODE integration with SCKF for sampling from a constraint. The densities are plotted over x.

The iterative conditioning procedure can also be viewed as Monte Carlo integration. For uncorrelated jointly Gaussian random variables, x_1 and x_2, marginalizing is equivalent to conditioning: $\pi(x_1|x_2) = \pi(x_1)$. This can be generalized to correlated random variables by introducing a linear transformation. Specifically, marginalization in linearly transformed coordinates is equivalent to conditioning in the original coordinates. This can be used to

cast "sample and move" linear conditioning as Monte-Carlo integration. Begin by transforming Y to an uncorrelated vector of random variables, z, with $Y = \mu_Y + Lz$ where L is a matrix square root (e.g. Cholesky factorization) of Σ_{YY}

$$L^T L = V^T \Sigma_k V$$

The goal now is to find $\pi(z|FLz = W)$. Where $z \sim N(0, I)$. Consider the singular value decomposition of FL,

$$FL = USV^T$$

where V contains right eigenvectors, S is the diagonal matrix of singular values, and U contains the left eigenvectors. Since F has fewer rows than columns, some singular values will be zero. Using the SVD in the conditional density gives

$$\begin{aligned} \pi(z|FLz = W) &= \pi(z|USV^T z = W) \\ &= \pi(z|SV^T z = U^T W) \end{aligned} \tag{4.66}$$

Next, introduce the transformed coordinates, $x = V^T z$. The new problem becomes to find $\pi(x|Sx = W')$ where $W' = U^T A$ and the prior on x is

$$x \sim N(V^T \mu_k, V^T V)$$

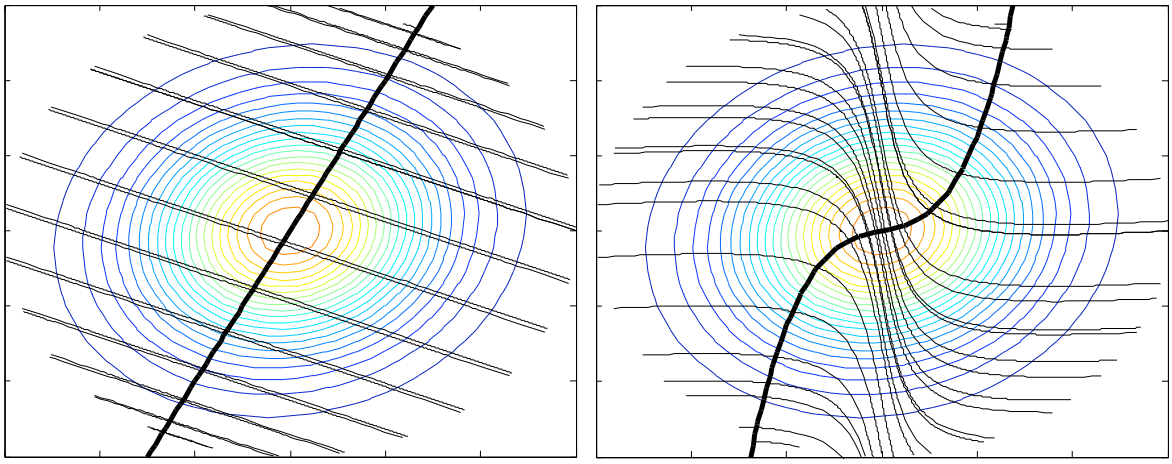
which is equivalent to

$$x \sim N(V^T \mu_k, I)$$

Note that x is uncorrelated and S is diagonal. Also observe that rows of S corresponding to the zero singular values are completely zero. Now, consider the set Z of all nonzero rows of S. The conditional distribution can then be decomposed as

$$\pi(x|S'x = W'') = \prod_{i \notin Z} \pi(x_i) \tag{4.67}$$

where $x_i = \frac{W''_i}{S'_i} \forall i \in Z$. Thus, conditioning in the original coordinates, $\pi(Y|W)$ is simply marginalizing over x_i for $i \in Z$. Back to sampling from $\pi(Y|W)$. It is clear that for a linear map, sampling a Y and moving it to Y^* with Eq. 4.59 is a projection of the point onto a subspace. Many samples of $\pi(Y)$ can be mapped to x space and then moved to the constraint. This process eliminates all Z directions. By elimination, we mean those coordinates have been marginalized out of the final distribution. The sample and move strategy can now be seen as a Monte-Carlo integration over the Z directions. In the nonlinear case, the singular value decomposition of the Jacobian is not constant over the entire parameter space, so the integration directions will not be lines. Instead, the integration directions will be given by Eq. 4.64. Fig. 4.16 illustrates the linear and nonlinear directions along which Monte-Carlo integration is taking place. Any sample of the prior that lands on one of these lines will move to the same location on the constraint. In the linear case, moving along a line to the constraint represents a projection onto the null space of F. The norm in this projection is defined by the prior covariance. Thus, moving along a line to a nonlinear constraint is a



(a) Linear Conditioning (b) Nonlinear Conditioning

Figure 4.16: Illustration of integration paths for nonlinear conditioning and linear conditioning

projection to the nullspace of $F(Y)$ using a warped nonlinear metric. In fact, this nonlinear projection might be related in some way to geodesics and Riemannian manifolds.

Fig. 4.16 also gives insight into why the nonlinear conditioning approximations in Fig. 4.15 show larger peaks than the true posterior. Exact integration of $\pi(Y)$ along each line in 4.16 with a numerical integration rule will give correct density evaluations along the constraint. However, moving random samples of the prior along the integration path results in more samples where the integration paths converge towards one another, resulting in an overestimation of the density. By incorporating some artificial noise into the system the SCKF results in some diffusion, alleviating the impact of converging integration paths.

4.2.4.3 MCMC Along A Constraint

The approximate nonlinear conditioning approaches generate a point Y^* that is an approximate sample of $\pi(Y|W)$. Linear approximation errors as well as the convergence of integration paths will need to be corrected by performing MCMC on the constraint itself. The constraint can be seen as a nonlinear manifold in a high dimensional space. Moving around such as manifold is non-trivial. Here, we instead consider a slight relaxation by replacing the constraint $F(Y) = W$ with $\|F(Y) - W\|_2 \leq \varepsilon$ for a small tolerance ε. A likelihood function for satisfying this constraint is given by

$$\pi(W|Y) = \begin{cases} 0 & \|F(Y) - W\|_2 > \varepsilon \\ 1 & \|F(Y) - W\|_2 \leq \varepsilon \end{cases} \tag{4.68}$$

Using Bayes' rule, the density we need to sample is

$$\pi(Y|W) \propto \pi(W|Y)\pi(Y) \tag{4.69}$$

A naive Metropolis-Hastings or DRAM approach could be used here. However, for small ε these methods will mix poorly and many evaluations of $F(Y)$ will be needed. We can take advantage of the fact that more information is at our disposal then is typically available for Metropolis-Hastings or DRAM. During the nonlinear conditioning, Jacobian approximations to $F(Y)$ were used in an approximate linear conditioning step. During the MCMC, this same information will be available and can be used to build efficient proposal densities adapted to the local topology of $F(Y)$. In fact, an isotropic proposal covariance can be adapted to local structure through the same equations used for nonlinear conditioning. In the absence of noise, updating the covariance $\Sigma = \sigma^2 I$ is done through

$$\Sigma_a = \Sigma - \Sigma J^T(Y_n)\left(J(Y_n)\Sigma J^T(Y_n)\right)^{-1} J(Y_n)\Sigma$$

Within the MCMC framework, Σ, should be chosen like a standard Metropolis-Hastings proposal. The updated covariance Σ_a will then be tuned to local effects of $F(Y)$. In cases where large global correlations exists, that information should be incorporated into Σ. Only local adaptation will be introduced here.

The linear conditioning update causes Σ_a to become rank deficient. Positive variance will only exist on the plane spanned by the rows of $J(Y_n)$. A Gaussian proposal with this covariance will result in a non-ergodic chain and also cripple the chain's ability to move around any regions of the parameter space where $F(Y)$ has curvature in directions perpendicular to the rowspace of $J(Y_n)$. Thus, much like the SCKF and Levenberg-Marquadt algorithm, artificial noise is introduced into the system:

$$\Sigma_{prop} = \Sigma - \Sigma J^T(Y_n)\left(J(Y_n)\Sigma J^T(Y_n) + \lambda I\right)^{-1} J(Y_n)\Sigma \tag{4.70}$$

where λ controls the width of the proposal in the nullspace of $J(Y_n)$. The choice of λ is problem dependent. Generally, λ should be less than ε to stay within the constraint tolerance but large enough to account for the linearization error and allow the chain to scoot around high curvature regions of $F(k)$. Fig. 4.17 shows two examples of the conditioned proposal. In both cases, $\lambda = 0.01$. The gray outline is the isotropic Gaussian density, Σ and the colored contours show the distribution updated with Jacobian information.

Table 4.1 shows the acceptance rates of the above MCMC algorithm using a cubic constraint with $Y \sim N(0,I)$ as a prior in two dimensions. The bold entries represent the proposal variance that minimizes error between the true density and nonlinear conditioned estimate. The estimate comes from iteratively conditioning and correction 500 samples of the prior. Note that even though smaller proposals have a higher acceptance rate, the larger proposals the entire parameter space more effectively. Additionally, for small λ, the MCMC iteration tended to get trapped in regions between high curvature. The linear approximation used in the proposal creation, prevented the chain from moving through these regions in a finite number of samples, even though asymptotic convergence to $\pi(Y \mid \|F(Y) - W\|_2 < \varepsilon)$ is guaranteed.

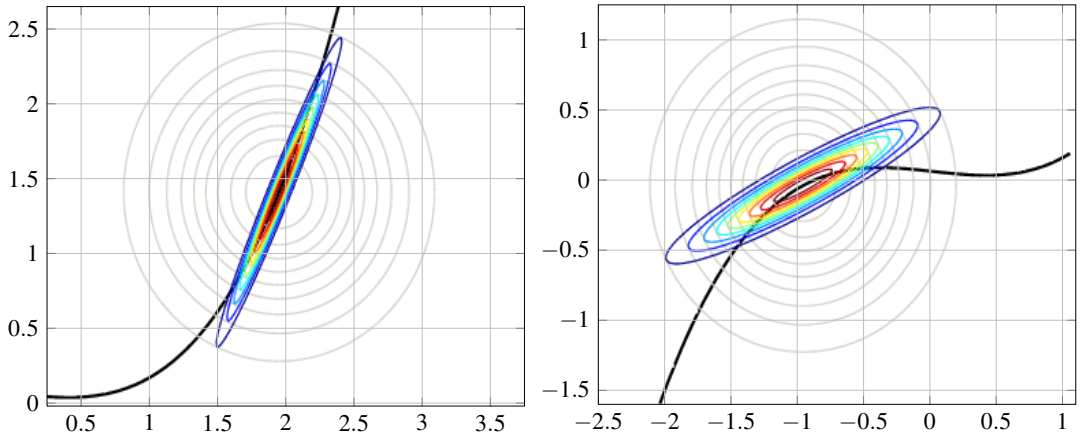


Figure 4.17: Illustration of conditioned proposals in two dimensions on a cubic constraint. In both cases, $\lambda = 0.01$.

Table 4.1: Comparison of acceptance rates for MCMC on cubic constraint with $\varepsilon = 1e-2$. The bold rates are the combinations that result in the smallest density reconstruction error from 500 nonlinear conditioning runs and 500 MCMC steps.

		Isotropic Variance, σ			
		1.00E-03	1.00E-02	1.00E-01	1.00E+00
λ	5.00E-04	*38.55%*	**31.92%**	19.96%	0.0741
	1.00E-03	32.04%	**24.16%**	16.35%	6.41%
	5.00E-03	24.41%	13.01%	**9.25%**	4.10%

4.2.5 Multiscale Framework Overview

A summary of the entire multiscale inference framework presented here is given in algorithm 2. Each iteration of the procedure is more computationally expensive than a single MCMC iteration. However, W and e have many fewer dimensions than Y and k. Furthermore, less correlation exists in W compared to Y. These two changes allow MCMC to more efficiently sample the posterior. Furthermore, the majority of the computational cost in the multiscale framework comes from the iterative nonlinear conditioning, which is embarrassingly parallel. Once MCMC has generated samples of e, all samples of k can be generated simultaneously. More precisely, no communication is needed between iterations of the loop in algorithm 2. When sufficient processing power is available, a new thread can be created for each step of the coarse MCMC chain to perform the nonlinear conditioning. It is also possible to only generate a fine scale realization for every k MCMC steps. This can be useful when a user sees the MCMC chain is not mixing quickly and has limited computational resources.

Algorithm 2 Basic summary of multiscale inference algorithm

Require: Prior distribution, $\pi(Y) \sim N(\mu_Y, \Sigma_{YY})$ and $\pi(W) \sim N(\mu_W, \Sigma_{WW})$
Pressure observations d
Number of samples N_s
Nonlinear mapping from Y to W, $F(Y) = W$
1: Generate N samples of $\pi(W|d)$, $W^1, W^2, ..., W^N$ using DRAM
2: **for** $i = 1...N$ **do**
3: Use SCKF or ODE integration to obtain Y^i satisfying $F(Y^i) = W^i$
4: Correct Y^i with M steps of MCMC on the constraint
5: **end for**
6: Return $\exp(Y^1), ..., \exp(Y^N)$ as samples of $\pi(k|d)$

4.2.6 Implementation Details

4.2.6.1 Jacobian in One Dimension

To perform nonlinear conditioning, the Jacobian of $F(Y)$ is needed. Recall the mapping

$$W_i = \log\left[\frac{1}{\int_{x_{i-1}}^{x_i} \exp(-Y)dx}\right]$$

Let Y be piecewise constant on each fine element. Let N_c be the number of fine elements and N_f be the number of fine elements in each coarse element. Thus, Y_{ij} is the log permeability in the jth fine element of the ith coarse element. Using this notation, the mapping becomes:

$$W_i = \log\left[\frac{1}{\delta x \sum_{j=1}^{N_f} \exp(-Y_{ij})}\right] \tag{4.71}$$

where $\delta x = x_i - x_{i-1}$. The derivative of W_i with respect to Y_{ij} is then

$$\frac{\partial W_i}{\partial Y_{(ij)}} = \frac{1}{\exp(Y_{ij})\left(\sum_{k=1}^{N_f}\exp(-Y_{ik})\right)} \tag{4.72}$$

The analytic form in one dimension makes computing the Jacobian trivial. In higher dimensions, where no clear analytic form exists, more effort is required.

4.2.6.2 Other Notes

As an implementation note. In cases where $J(k)$ has a row of very small values, the term $J(Y)\Sigma_k J^T(Y)$ can be nearly singular. In this case, a nugget needs to be added to diagonal, $J(Y)\Sigma_{YY}J^T(Y)+\varepsilon I$. This is essentially adding some artificial width to the constraint, placing a lower bound on the eigenvalues of $J(Y)\Sigma_k J^T(Y)+\varepsilon I$ and reducing the condition number. The value of ε is usually very small, $\approx 10^{-10}$.

4.2.7 Generalization to Other Multiscale Settings

The framework presented here uses MsFEM to map fine scale quantities to coarse scale quantities. However, the framework is not restricted to this case. Any method that provides a coarse quantity sufficient to represent p can apply, even upscaling could be used. MsFEM can be viewed as a local reduced order modeling technique. Other methods that represent the solution by a limited number of basis functions could also be used in this framework. For the multiscale framework to be effective, the coarse parameter should be sufficiently low dimensional for efficient MCMC and the mapping should not be extremely nonlinear. That is, the mapping should at least be continuous to ensure that MCMC can be efficient at correcting for linearization errors acquired during the iterative conditioning.

4.3 Numerical Results

4.3.1 Test Cases

This section illustrates the multiscale framework presented in the previous chapter on several one dimensional test problems. Initially, a comparison is done with single scale inference to validate the multiscale approach. Then several examples are given to demonstrate the framework's applicability in a wide variety of situations. A summary of the additional problems can be found in Table 4.4.

4.3.1.1 Problem Description

All example problems share the same type of boundary conditions, Dirichlet on the left and Neumann on the right. However, each problem consists of a different type of permeability field and varying amounts of data. A fine scale simulation using a true permeability field is combined with additive noise to generate the data. The variance of this noise is 0.01 for all situations, which is approximately 1% of the pressure range for each case. This low noise level was used to ensure that the likelihood has a strong impact and the posterior is sufficiently different than the prior.

In the multiscale framework, DRAM is used for sampling the elemental integrals before iterative conditioning is performed with the SCKF and constrained MCMC. Within DRAM, the initial proposal size was manually tuned based on the well founded squint test and practical experience.

The first test is a verification that the multiscale method produces similar results to standard MCMC approaches. The first example contains a relatively smooth permeability field without a wide scale separation. This is important for practical applications of MCMC. The parameter space is already 50 dimensional in this smooth case and would need to be increased for a permeability with large scale separation. An accurate DRAM characterization in the high dimensional setting would require a prohibitive number of samples.

The remaining tests illustrate the effectiveness of the multiscale inference framework on more complicated permeability fields. A summary of all the problems can be found in Table 4.4. As an implementation note, coarse nodes need to be placed at the observation points because the MsFEM basis functions are never explicitly created. In the ensuing examples we have the luxury of creating the data, so it is easy to place observations at coarse node locations. In real situations this is not the case and either the mesh needs to be constructed with knowledge of the observation locations, or interpolation error must be introduced by projecting the data onto known basis functions instead of the MsFEM basis functions.

The standard error of a posterior estimate for a fixed number of samples is minimized when the posterior samples are uncorrelated. However, an MCMC chain generates correlated samples. Thus, in comparing MCMC methods, the autocorrelation of samples is a useful measure of performance. In fact, an integrated autocorrelation can be used to compute the effective sample size of a chain. This quantity represents how many *independent* samples would be needed to have essentially the same amount of information as the correlated MCMC chain. In the comparisons below, effective sample size (ESS) of a correlated sequence, Θ, with length N is defined as

$$ESS(\Theta) = \frac{N}{1 + 2\sum_{k=1}^{\infty} \rho_k(\Theta)} \tag{4.73}$$

where $\rho_k(\theta)$ is the autocorrelation of the sequence at lag k. This quantity is computed using all samples after an initial burn-in of the chain. Obviously the sum has to be truncated to

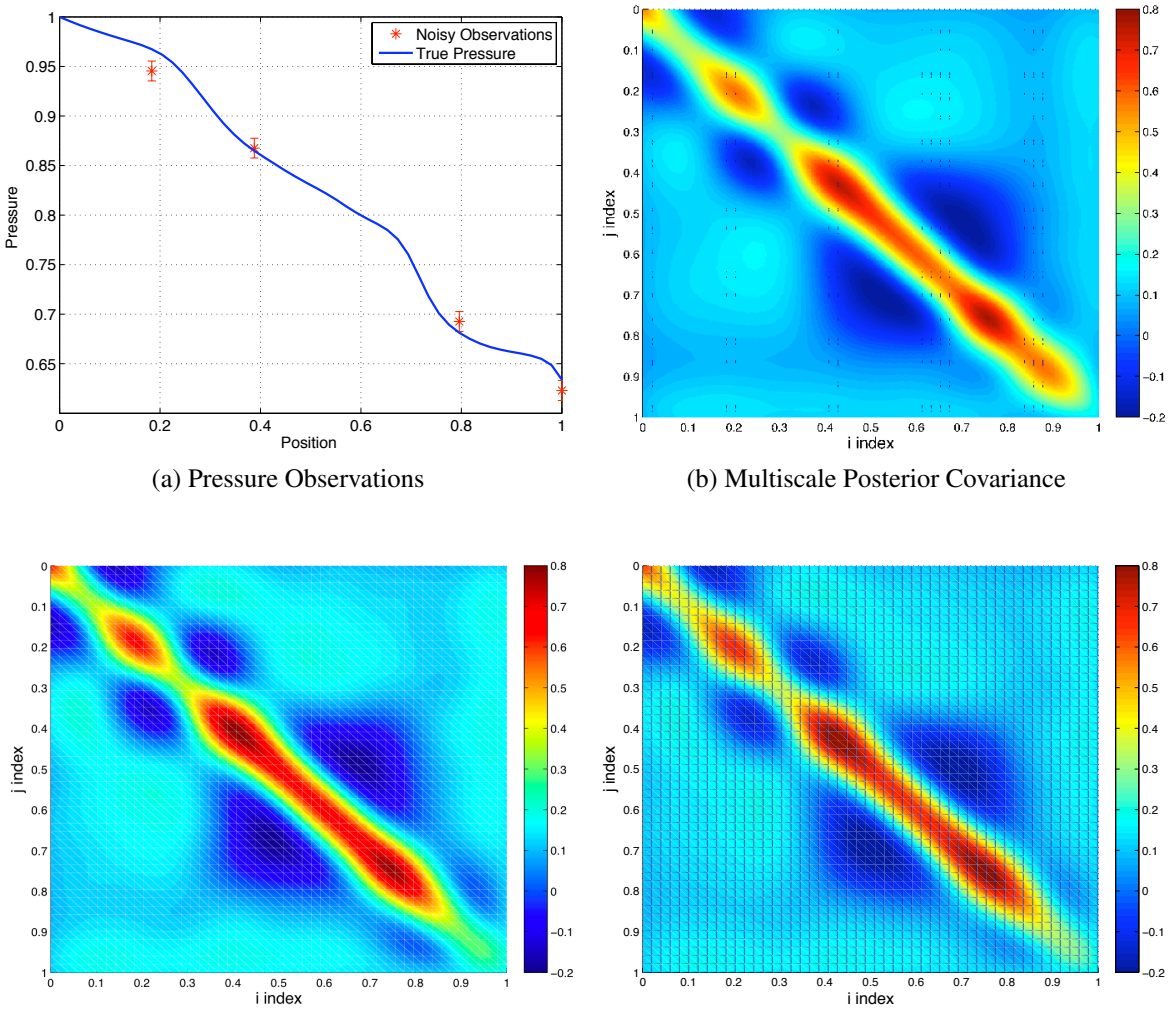
a finite lag. Here, the maximum lag, n, is chosen so that $\rho_k(\Theta) \leq 0.05$ for all $k > n$. The effective sample size can be computed for each dimension of the chain so a range of effective sample sizes is obtained. Unless otherwise noted by Max ESS, or Min ESS, just *ESS* will refer to the median effective sample size over all dimensions.

4.3.1.2 Verification with Single Scale Inference Results

Here we compare three different methods. The base case is single scale inference for k using DRAM for sampling with a standard FEM forward solver with linear basis functions on the fine scale. The second single scale approach again uses DRAM for sampling but MsFEM as the forward solver. These methods are compared with the multiscale framework presented in the last chapter. The single scale methods operate with 50 elements spread over $[0,1]$ and the multiscale approach uses 5 coarse elements with 10 fine elements in each coarse element. Fig. 4.18 shows the true pressure with all 4 noisy observations used in this test as well as the posterior covariances obtained with each method. Fig. 4.19 concentrates on the posterior mean and variance.

Clearly, all three methods are in fairly good agreement. The mean fields are nearly identical until the high permeability region on the right of the domain and the variances are quite close except at a few locations. The multiscale method slightly overestimates the posterior mean but captures the same covariance trend as the single scale methods. One possibility is the single scale MCMC chains have not sufficiently converged. While possible, this is unlikely. Each single scale chain was run for 2×10^6 steps, achieving a minimum effective sample size of 3400. A more likely explanation is in the approximate prior distribution on $\pi(e)$ and errors in the nonlinear iterative conditioning. Decreasing the step size of the nonlinear conditioning can help improve the result and the number of MCMC steps along the constraint so that no error in the iterative conditioning exists. However, performing enough MCMC samples so the chain "forgets" its initial point on the constraint is intractable. Thus, the iterative conditioning method should be chosen and tuned to best approximate the generation of samples from $\pi(k|e)$. While some errors are introduced, the computational savings and applicability of the multiscale framework to problems with large scale separation outweigh this relatively small posterior error.

Although the effect is small, another possible error source also needs to be mentioned. With the strongly correlated prior in this test case, the update matrix from Eq. 4.59 can become ill-conditioned and cause issues during the SCKF stage and constrained MCMC stage. In some cases, the ill-conditioning causes the permeability to go far into the tails of the parameter space. It is possible to reject these samples and try again, as in rejection sampling. However, sometimes the error is not as extreme and catching the inaccurate sample is not possible. This effect may be slightly contributing to the overestimation seen in Fig. 4.19. The ill-conditioned update occurs predominantly in smooth cases where the prior places a strong restriction on the posterior. With better conditioned prior covariances, this issue becomes insignificant. The smooth example here was used for efficient MCMC comparison but is not the application of interest. The multiscale framework will be most



(a) Pressure Observations

(b) Multiscale Posterior Covariance

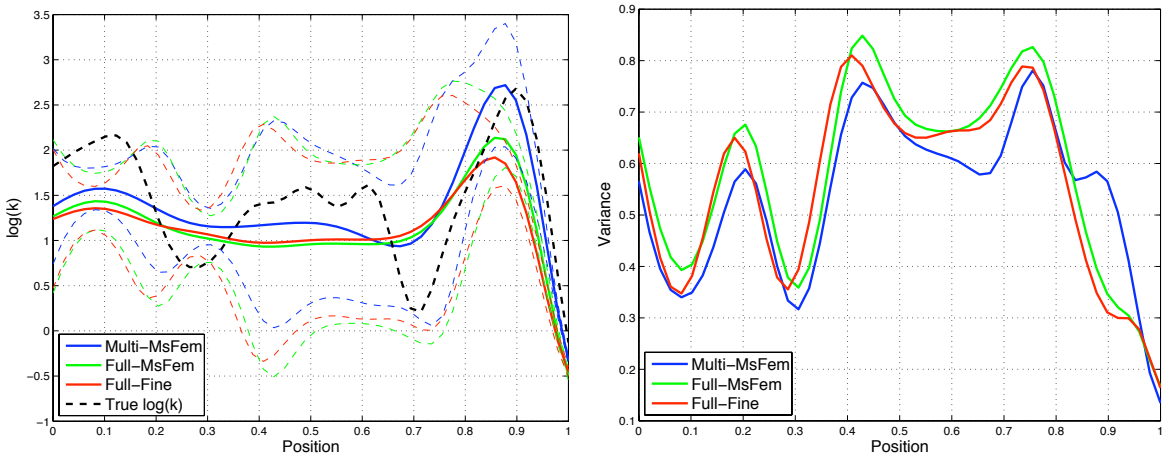
(c) Single scale FEM Covariance

(d) Single scale MsFEM Covariance

Figure 4.18: Location of pressure observations and posterior covariance comparison using the fully multiscale approach, single scale with FEM forward solver, and single scale with MsFEM forward simulation.

applicable to problems with weak priors and when the permeability field is expected to have short length scales. On average, this reduces the condition number of the linear update and the conditioning errors are less likely to have any impact on the posterior.

In addition to the posterior summaries, Table 4.2 shows a performance comparison. The ratio of effective sample size to the total number of samples is shown in the first table while the second table displays how many effectively independent samples were produced each second. Note that the coarse MsFEM chain was thinned by a factor of 10 before performing the iterative conditioning. The timings were performed on a MacBook with 2.4Ghz Core2Duo and 2GB of 1067 MHz DDR3 memory. Focusing on ESS/N, it is clear that the multiscale method generates more than an order of magnitude more effectively



(a) Mean Comparison

(b) Variance Comparison

Figure 4.19: Comparison of posterior mean and variance between standard single scale approaches and the multiscale framework. The left plot shows the posterior means $\pm 2\sigma^2$

Table 4.2: Comparison of single scale performance with multiscale performance. ESS is the effective sample size, so ESS/N is the effective number of samples related to the total number of fine scale samples generated and ESS/S is the number of effective samples per second.

Method	Min ESS/N	Max ESS/N	Median ESS/N
FEM Forward	1.7×10^{-3}	9.5×10^{-3}	7.0×10^{-3}
MsFEM Forward	5.6×10^{-3}	9.3×10^{-3}	7.2×10^{-3}
Multiscale	1.7×10^{-1}	3.1×10^{-1}	2.5×10^{-1}
	Min ESS/s	Max ESS/s	Median ESS/s
FEM Forward	0.26	1.54	1.38
MsFEM Forward	2.03	3.34	2.60
Multiscale	1.12	2.05	1.64

independent samples than the single scale approaches. Not only does the coarse scale chain more effectively explore the parameter space and reduce correlations between samples, but additional randomness is introduced into the system by starting the iterative conditioning from a random sample of the prior. This additional randomness dramatically reduces the chain correlation.

Now concentrating on the timing, the benefit of the multiscale framework is not as prevalent. While still outperforming the standard FEM methods, the gap is not as convincing as the ESS/N. However, the single scale MsFEM timing is dependent on the analytical form of the elemental integrals in one dimension. In higher dimensions, MsFEM time is expected to drop closer to the standard FEM approach. These runs were computed in a sequential fashion, each iterative conditioning step was performed one after another, but an additional benefit of the multiscale procedure is its easy parallelization onto nearly as many

nodes as available. For example, even for a moderate 80 core cluster, the parallel ESS/s would be over 120. In addition to easy parallelization, the multiscale framework allows the inexpensive coarse chain to generate many more samples of the elemental integrals than is needed to characterize the poster of k. The coarse chain can be downsamples to reduce the number of iterative conditioning steps required.

The example problem here is a smooth problem using a Gaussian kernel prior used solely for comparison of the multiscale framework with standard single scale approaches. Smooth priors can be handled efficiently by standard single scale sampling approaches with a dimensionality reduction technique like Karhunen-Loeve expansion. However, many interesting problems do not have smooth permeability fields and can have large scale separations. While problematic for single scale methods, the multiscale framework is ideally suited for this type of problem. The following three test problems illustrate this fact. For a summary of the problems, see Table 4.3.

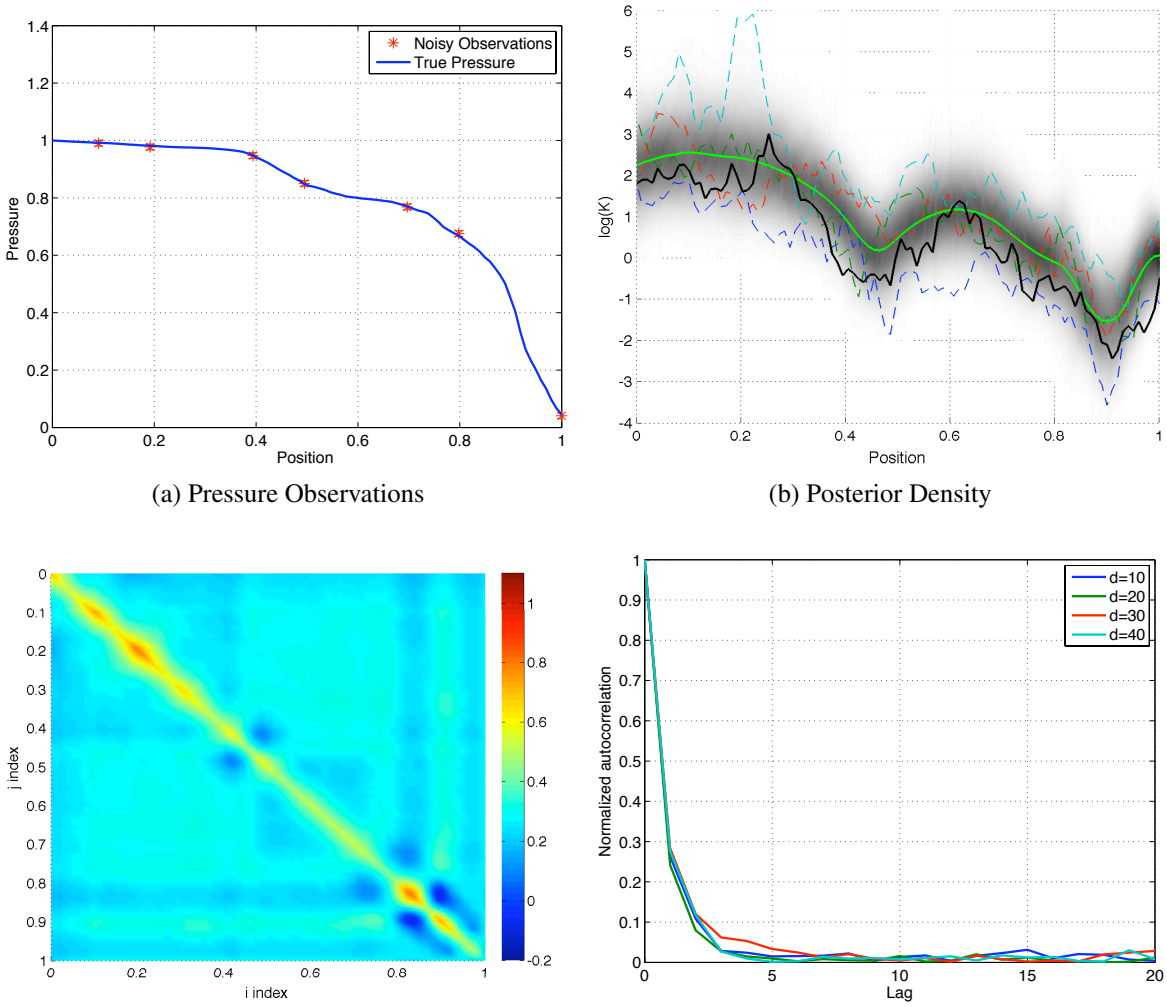
4.3.1.3 Draw from Exponential Prior

Table 4.3: Summary of multiscale inference tests. K_c is the number of coarse elements, K_f is the number of fine elements in each coarse element. d is the correlation length of the prior, p is the prior power, σ is the prior variance, μ_Y is the prior mean, N_d is the number of data points.

Trial Name	K_c	K_f	d	p	σ	μ_Y	N_d
Exp. Prior	10	10	0.2	1	1	1	7
SPE Layer 1	11	20	0.075	2	3	3	6
SPE Layer 5	11	20	0.17	1	2	4	11

This test uses an exponential prior covariance kernel on a 100 dimensional permeability field to create a more difficult problem beyond the single scale MCMC regime. In order to characterize the field, the number of coarse elements was increased to 10. The fine mesh has 10 fine elements in each coarse element. The number of observations was also increased from 4 to 8 as shown in Fig. 4.20. In terms of inference results, in the posterior density plot, the thick black line is the true permeability, the dashed lines are posterior samples, the thick green line is the posterior mean, and fuzzy shading is a more complete visualization of the posterior density. Darker shading implies a larger density. The autocorrelation plot is for the chain of permeability samples generated by sequentially taking each step of the elemental integral chain and performing iterative conditioning.

In the results from Fig. 4.20, the coarse chain was downsamples by 50 before performing the iterative conditioning. Little correlation can be seen within the chain. This is a major advantage over single scale methods, here the level of downsampling can be adaptively chosen to maximize the information after analyzing the correlation of the coarse chain. Thus, with a fixed amount of computational power, a user can be confident that as much information as possible about the posterior was gathered. While not always a good measure



(a) Pressure Observations

(b) Posterior Density

(c) Posterior Covariance

(d) Autocorrelation of k chain

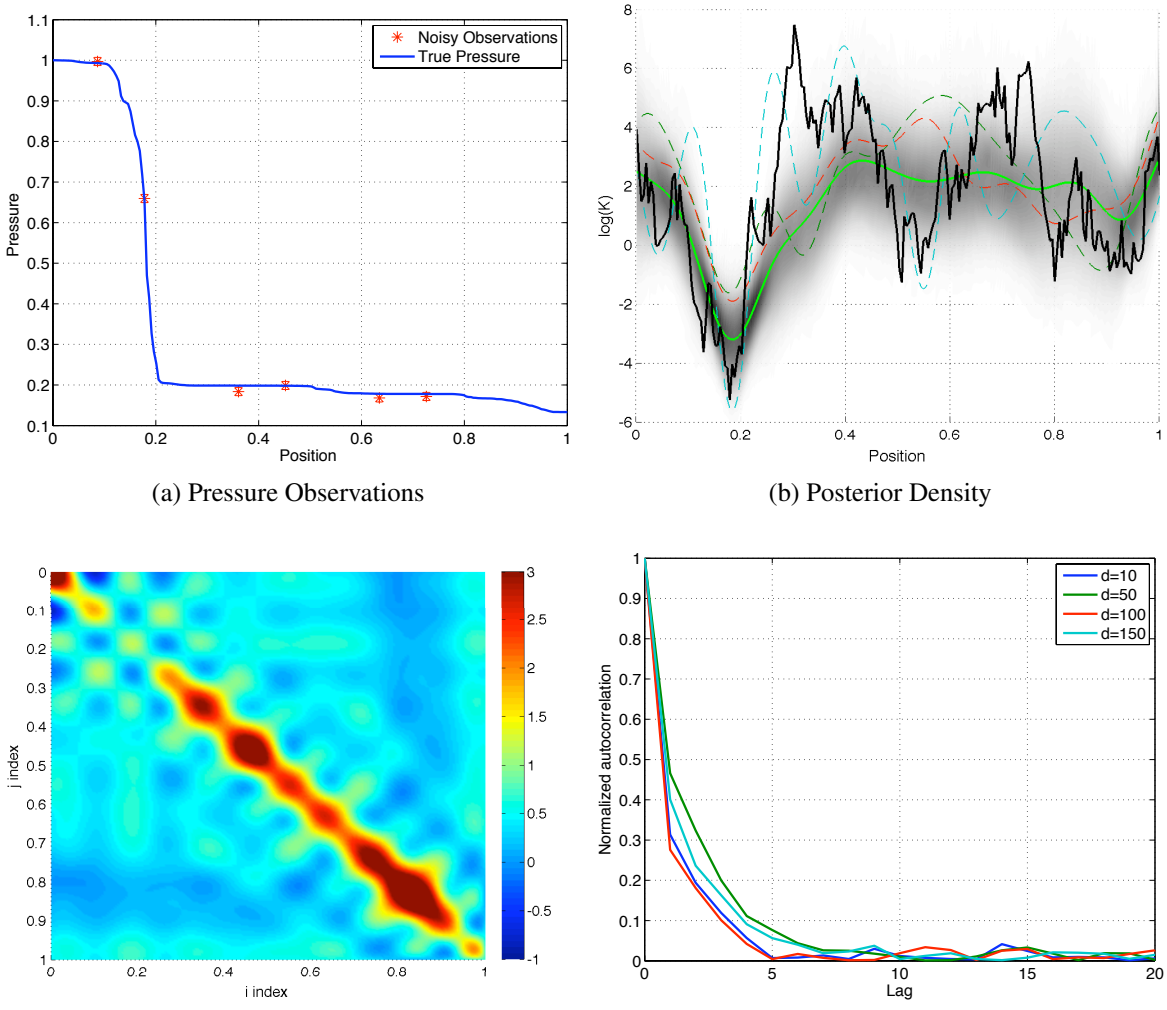
Figure 4.20: Results for multiscale inference using a draw from the prior. The prior kernel is defined by $p = 1$, $d = 0.2$, $\sigma = 1$. Twenty fine elements were used in each coarse element and 5 coarse elements were used.

of performance, it is encouraging to see the posterior mean agreeing quite well with the true permeability field. Also, note the relative smoothness of the posterior mean compared to the true field.

4.3.1.4 First Layer of SPE10

In some sense, the above test problems were "cheating." The permeability field was drawn from the prior, which means the prior is related to the true field. This violates the philosophical definition of prior information and does not stress the multiscale framework in the same way as a realistic problem. To take a small step towards reality, this test considers

a slice of the SPE10 dataset. The SPE10 dataset is a representative reservoir description developed by the society of petroleum engineers. It consists of 85 layers of permeability and porosity information with 60×220 cells. In this test, row 25 of layer 1 is used as the true permeability field. The prior used here was adapted from a geostatistical characterization of the entire dataset developed by Sean McKenna at Sandia National Laboratory. See Table 4.3 for a complete prior definition. The results are shown in Fig. 4.21.



(a) Pressure Observations

(b) Posterior Density

(c) Posterior Covariance

(d) Autocorrelation of k chain

Figure 4.21: Results for multiscale inference using a slice of layer 1 of the SPE 10 dataset. The chain was downsampled by 50 before iterative conditioning.

The permeability field in this test has a very different structure than the prior realizations used before. There is clearly several inherent scales on a slowly varying field superimposed with high frequency oscillations. Interesting to note is the significant impact of a low permeability area around $x = 0.2$ on the pressure. The small permeability causes a significant drop in the pressure. Being a dominant feature of the entire pressure field, matching the data is very sensitive to matching this drop. The posterior results demonstrate this with less

uncertainty (darker shading) near $x = 0.2$. Additionally, throughout the domain the true log permeability is covered by the posterior, showing that with additional observations, the variance could be reduced and the posterior could collapse on the true field.

Even in this one dimensional problem, the fine scale inference problem has 220 dimensions. DRAM would require millions of samples to attempt searching this space. However, in this multiscale framework, only the 11 elemental integral dimensions are searched with DRAM and an acceptance ratio of 23% is achieved. Furthermore, the lack of significant autocorrelation shows that most samples are nearly uncorrelated. In a similar fashion to the previous test, the info the inexpensive coarse MCMC chain was run longer and downsampled to provide less correlation between samples of k. The correlation plot in figure 4.21 was constructed after downsampling the coarse chain by a factor of 50.

4.3.1.5 Fifth Layer of SPE10

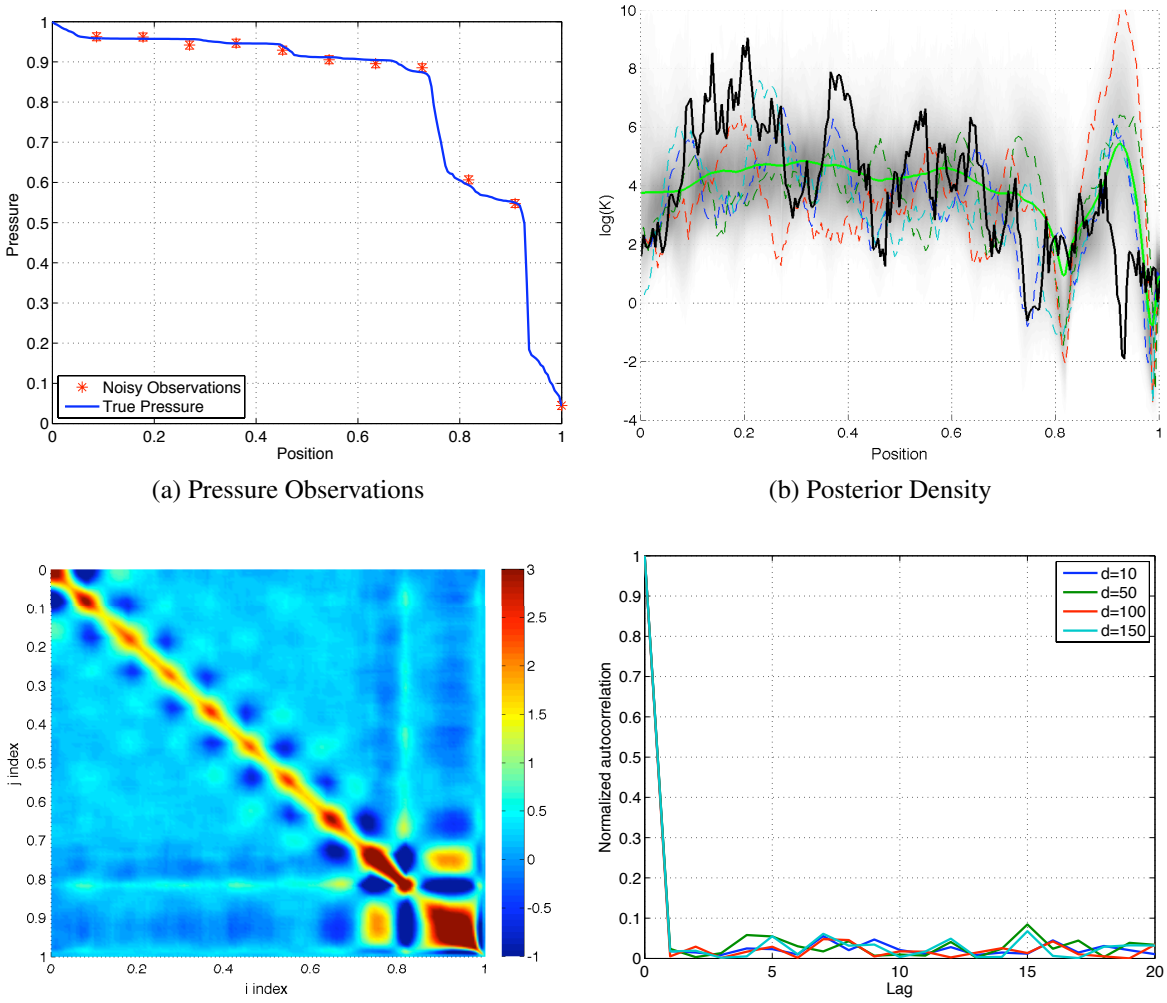
The fifth layer of the SPE10 dataset has different characteristics than the first. Here, row 45 of layer 5 is used which has a generally negative slope with a medium correlation length pattern and some high frequency components. An exponential prior is used and the number of observations is increased to 11. For a more complicated field like this one with a shorter lengthscale, more data is required for the posterior to be significantly different than the prior. Fig. 4.22 shows the results. The pressure is dominated by the two distinct drops near the right of the domain. The pressure is again sensitive to the permeability values in these areas. The posterior captures this behavior as expected. There are two low permeability areas in the posterior with a small variance. These bound the highly uncertain area around $x = 0.9$. While the true permeability does not have as large a peak at $x = 0.9$ as the posterior mean, the multiscale method samples effectively enough to capture the uncertainty in this region. The smaller true permeability is well within the high probability region of the posterior.

4.3.1.6 Summary of Test Cases

Table 4.4: Summary of multiscale inference performance. ESS is the effective sample size, so ESS/N is the effective number of samples related to the total number of fine scale samples generated and ESS/S is the number of effective samples per second. Additionally, acceptance is the acceptance rate of the coarse MCMC chain.

Trial Name	ESS/N	ESS/S	Acceptance
Exp. Prior	0.2267	2.11	0.16
SPE Layer 1	0.028	0.17	0.23
SPE Layer 5	0.067	0.082	0.12

These test cases stress the multiscale inference framework with high dimensionality and realistic permeability fields with scale separation. Table 4.4 summarizes the test problems



(a) Pressure Observations

(b) Posterior Density

(c) Posterior Covariance

(d) Autocorrelation of k chain

Figure 4.22: Results for multiscale inference using a slice of layer 5 of the SPE 10 dataset. The chain was downsampled by 50 before iterative conditioning.

not compared with MCMC. MCMC methods could be applied on these problems but the comparison between the MCMC results and multiscale results would be moot because the MCMC would not be able to sufficiently explore the posterior. Some of these test problems have also been studied with more observation noise, the result is as expected, more noise results in a posterior more similar to the prior.

In the verification of the multiscale method with MCMC results, its clear that in a smooth setting, the iterative conditioning process can become difficult due to poor condition numbers of the linear update. However, the ability to adaptively down sample and easily parallelize the iterative conditioning stage would result in a dramatic increase in the number of effectively independent samples generated. Some will argue that MCMC can also easily be parallelized by running multiple chains simultaneously and then combining the results.

This is an improvement, but each chain still suffers an inability to effectively search vast emptiness of large dimensional spaces. Additionally, several coarse MCMC chains could be also used in the multiscale setting, reducing the amount of down sampling required before iteratively conditioning.

4.4 Future Work and Conclusions

4.4.1 Future work

One dimensional examples were given here, but it is important to note the framework is not dependent on the spatial dimension of the parameters being estimated. The exact same formulation exists for any number of spatial dimensions. This chapter focused on the multiscale framework derivation and a discussion of the formulation intricacies with motivation using a one dimensional field. Compared to the one dimensional setting, no analytical expression exists for the Jacobian information of the two dimensional mapping from k to e. However, the Jacobian can be computed by using adjoint methods to first compute the sensitivity of the MsFEM boundary conditions to the log permeability and then computing the joint sensitivity of the elemental integrals to the permeability and boundary condition. Additionally, the number of elemental integrals per element will also increase. It should be noted however, that once Jacobian information is computed, all other parts of the algorithm remain unchanged. Implementation in higher dimensions is just one aspect of future work.

4.4.2 Use in Dynamic Problems

In addition to expanding this application of the multiscale framework to higher dimensions, the method is also applicable to time dependent parabolic equations. In the porous media context this corresponds to a time dependent pressure equation. An interesting avenue of research is the use of particle filters or the ensemble Kalman filter to dynamically update the posterior elemental integral distribution and only generate a few fine scale realizations at each timestep. This also has applications in real-time PDE constrained dynamic inference. Being sufficient to describe the coarse pressure, only elemental integrals are needed to evaluate the likelihood and evolve the ensemble forward in time. By dramatically reducing the computational cost of a likelihood evaluation, the multiscale approach would allow for real time inference. Fine scale realizations would not need to be generated in real-time for many applications and a massively parallel post-processing procedure would allow the fine scale realizations to be efficiently generated.

4.4.3 Spatial Decoupling During Iterative Conditioning

The iterative conditioning procedure does not take into account any spatial decomposition. All components of the permeability could be correlated with other permeabilities. However, in many applications, some areas of the spatial domain will not be related and further decomposition could be performed. Using knowledge of uncorrelated parameters would allow the iterative conditioning process itself to be parallelized. The multiple levels of parallelism (independent coarse samples and independent spatial regions) would allow this multiscale framework to capitalize on the advent of highly heterogeneous clusters where multiple levels of parallelism are needed for maximum efficiency.

4.5 MsFEM Inference Summary

Inference problems found in real-world applications are often characterized by large dimensional parameter spaces and noisy observations. Bayesian inference provides an intuitive paradigm for formulating the inference problem but requires sampling of a high dimension space. Especially in the case of elliptic forward models, high probability regions of the posterior distribution can be small compared to the parameter space, making traditional single scale MCMC difficult. Using a multiscale approach with a nonlinear mapping from fine parameters to coarse parameters, we showed that when the coarse parameters are sufficient to describe the model output, the scales can be decoupled and the inference process can be broken into two stages. The first stage uses standard MCMC sampling tools to characterize a posterior for the coarse quantities and the second stage takes of the coarse samples and generates a realization of the fine scale parameter that agrees with the coarse quantity.

It was shown in the context of porous media that this multiscale framework can be successfully applied for inference. By decoupling the elemental integrals from the permeability, as shown in the several test cases above, it is possible to efficiently explore a high dimensional posterior. It is important to note that the multiscale procedure used here is not limited to elliptic problems or even hydrology. In fact, the framework should be applicable to many situations involving smoothing operators (like the pressure equation) and/or nonlinear observations.

The unique decoupling of fine and coarse scales with a nonlinear mapping between the two, provides a small step towards tackling large dimensional inference problems in a Bayesian setting. The several examples of permeability estimation shown here indicate the efficacy of this decoupling and point towards a future application in currently intractable real-world inference problems.

Chapter 5

Computationally Efficient Stochastic Inference for Multiphase Flow

5.1 Introduction

Fluid flow through porous media is central to many important applications including petroleum reservoir management, filtration systems, energy storage, physiological processes and ground water transport. In most cases, the exact material properties of the porous media are unknown but yet critical to achieving accurate dynamical predictions. Field observations and laboratory measurements can guide the selection of model parameters but these cannot provide the coverage needed for accurate dynamics predictions in complicated simulations. In particular for subsurface geologic transport, material properties such as permeability, porosity, relative permeability, initial saturations are required at every point in the computational domain. This motivates the inverse problem in which the goal is to infer the material properties by making use of field or laboratory information and comparing these sparse measurements with numerical predictions of the underlying dynamics. One of the key challenges for subsurface transport inversion is to characterize the uncertainty that arises from the lack of complete knowledge of the material properties model approximations and measurement errors. Other important challenges include handling of large numbers of inversion parameters, nonlinearities in the dynamics as well as in the inversion problem, and computational requirements.

The focus of this chapter is to investigate computationally efficient inversion methods that are capable of also providing statistical characterizations. The data assimilation problem can be solved with a variety of techniques, the selection of which depends on the availability of computational resources and the need for complete statistics. On one end of the spectrum, linearity and Gaussian assumptions allow for adjoint based algorithms that solve inverse problems quite efficiently [130, 131]. However, if model or measurement errors are non-Gaussian and the solution requires a complete statistical solution, Monte Carlo algorithms are required that are very computationally demanding. To this end, we employ

Ensemble Kalman filtering (EnKF) algorithms to reconstruct the underlying permeability field in a porous media and predict the dynamics of a water-oil displacement process. The EnKF method makes use of field observations to help calibrate the simulation model and predict the dynamics. Unlike the Kalman filter in which the covariance matrix depends on an adjoint of the forward model and the algorithm assumes linearity and Gaussianity, the EnKF evolves the covariance matrix by post-processing a number of forward simulations (also known as ensembles) and incorporates nonlinear behavior from the post-processing of the ensembles. However, the Ensemble Kalman filter algorithm uses a Gaussian update step and although this algorithm was designed to handle nonlinearities as well as manage non-Gaussian statistics, it is not clear how accurate the resulting statistical moments are.

This chapter first provides a mathematical formulation to explain the connection between Bayesian theory, Kalman filters, Ensemble Kalman filters, and MCMC-based solvers. Because most of these methods are well known, we only provide a general description and refer the interested reader to other references for additional details [132, 133]. Our goal is to clarify the important details in each algorithm to provide a concise connection between approaches in addition to providing enough detail to help explain numerical results. Our numerical results consist of three parts. First, we report on a parameter study in which the sensitivities and accuracy of the EnKF algorithm are explored. Second, we apply a hybrid EnKF-MCMC approach to reconstruct permeabilities and predict saturations using a large and complicated dataset. Lastly, we show the utility of the EnKF on a complex and large inversion dataset. New contributions of this work consist of 1) sensitivity study comparing EnKF to MCMC using a porous media dataset, 2) a hybrid EnKF-MCMC algorithm that provides a mechanism to incorporate static and dynamics data, and 3) the application of EnKF on a dataset that is intractable computationally for MCMC solvers.

Large scale numerical simulation is becoming an increasingly important tool for engineering design in which complicated dynamics are being resolved. Assuming that the formulation, discretization, and implementation are verified, numerical predictions depend on a multitude of model parameters. Especially in the case of large scale problems and highly nonlinear dynamics, these underlying properties can span large ranges and pose considerable uncertainty in the final solution. Laboratory measurements or field observations of certain "state" parameters can help to calibrate simulation models and consequently motivate the "inverse" problem. Next we briefly review previous work in this area and identify which application have benefited from ensemble methods, in addition to reviewing the state of the art methods in ensemble Kalman filters.

Evenson first introduced the Ensemble Kalman filter in 1996 [134] but acknowledged the origin of the ensemble approach to be related to the theory of stochastic dynamics dating back to 1969 (Epstein, [135]) and subsequently extended by several researchers [136, 137, 138, 139]. The assimilation process however is rooted in the Kalman filter algorithm, which was named after Rudolf Kalman. Eventually, Evenson produced an authoritative text on the subject that is one of the most cited reference on the subject [132]. The EnKF has been applied to a range of applications, such as ocean and ice modeling [140], weather prediction [141], soil moisture [142], reservoir management [143, 144, 145, 146, 147, 148]

and several other relevant problems.

In this chapter our target application is porous media flows but our methods are applicable to other areas. The following papers are reviewed in some detail to provide additional context for our work. Liu and Oliver [147] evaluate the performance of gradient based minimization methods to EnKF for the problem of estimation of facies boundaries in history matching. An adjoint based LBFGS algorithm is used. The authors question the quality of the uncertainty metrics.

Gu et al. [144] apply the EnKF to solve history matching problems for 1D and 2D water flood. The focus of this paper is to accurately predict the water saturation front by injecting data from porosity, and permeability at each time period. Two issues are addressed: bimodal water saturations in a 1D problem before and after a shock, and the effects of using ensembles substantially less than the number of variables. One major issue with EnKF is that the permeability and porosities change with time and there is no guarantee that current values field match data of previous timesteps.

Gu et al [145] presents an iterative EnKF methodology to address the shortcomings of the analysis part of EnKF for highly nonlinear problems. The Gauss-Newton formulation is given for the equivalent minimization formulation to the maximization of the probability density. In this paper, the EnMLF is compared to EnKF and shows that the Gauss Newton method performs identically for a linear problem. Problem 3 compared EnKF and EnRML to an acceptance/rejection algorithm from Ripley (1987). The EnRML matched the acceptance/rejection algorithm results, whereas the EnKF did not. In the case of nonlinear problem – waterflood with a shock – EnKF and conforming EnKF did not generate feasible solutions. EnRML was always feasible and emulated the truth model much better.

Nowak [148] discusses the use of EnKF to transition to geostatistical based inversion. An extension to the original EnKF method is introduced by using an unbiased ensemble linearization (Jacobian linearization causes a bias). In the case of solute transport, the linearization causes less dispersion in the mean solution.The most compelling reason to use ensemble based methods is to avoid computationally infeasible sensitivity analysis and storage of very large autocovariance matrices of parameters.

Zhou et al. [146] present an EnKF based algorithm that addresses non-Gaussian distributions using a normal score transform. The EnKF algorithm causes the distributions to become more Gaussian as the process evolves. Although Evenson designed the EnKF to obtain a single optimal estimate of the system state, the EnKF provides as a by-product the entire ensemble of states, which can be used to assess uncertainty. In this paper log of hydraulic conductivity is used and for the state variable piezometric head is used. A normal score transformation ensures that the prior non-Gaussian marginals of the model parameters are kept throughout, in this case a bimodal pattern of log conductivities. The fluvial examples in this paper look like our "pathological" dataset. The paper describes how these fluvial datasets are generated using geostatistical tools. They show that the normal score EnKF algorithms can reconstruct the bimodal distributed material parameters. A normal score transform changes a non-Gaussian distribution to a Gaussian distribution with zero

mean and standard deviation of unity.

The remainder of the chapter is organized as follows. First Bayesian theory is reviewed to provide a common foundation for a comparison between MCMC and EnKF. The Kalman filter and EnKF derivation is then outlined. MCMC methods are explained which is followed by an introduction of our MCMC-EnKF hybrid approach. Our numerical example consists of a synthetic waterflood dataset that uses a Karhunen-Loève parametrization for permeability. Finally, the numerical section presents our results consisting of a comparison of MCMC and EnKF methods for a simple waterflood problem in addition to the reconstruction of permeability and prediction of saturations for a more complicated dataset. We end the chapter with conclusions.

5.2 Bayesian Inference Review

Bayesian inference is a flexible and statistically rigorous way to model parameters and states. The inference problem is usually formulated to statistically estimate parameters or inputs to a model, based on noisy observations of the model output. The process of predicting dynamics, in addition to propagating errors through the model, is referred to as the forward problem and can be defined as

$$d(x,t) \approx G(x,t;m) \tag{5.1}$$

where m is a set of model parameters, d is a vector of observable output quantities, x is position, t is time, and G is potentially a nonlinear function of x and t, parametrized by m. The model parameters m are unknown and need to be estimated from noisy measurements of d. To handle the observation uncertainty, both d and m will be modeled as random variables in a proper probability space.

To estimate parameters, we invoke Bayes' rule to condition the model parameters on d:

$$f(m|d) = \frac{f(d|m)f(m)}{\int f(d|m)f(m)dm} \tag{5.2}$$

where $f(m)$ is the prior density of m, which represents previously held degrees of belief about m. This information could come from physical meanings of the model parameters or simply expert opinion. In a Bayesian setting, as opposed to the frequentist paradigm, probabilities represent degrees of belief that the model parameters will take particular values. An informative prior acts as a regularization term as well, allowing for the solution of ill posed and under-determined problems. While the prior distribution represents previously held beliefs about model parameters, the likelihood function, $f(d|m)$ measures how well a particular set of model parameters match the available measurements. By using both the prior distribution and likelihood function, Bayes' rule updates previously held beliefs with the current measurements in d. In a static problem, only one update will be used because all

information is presented at once. However, Bayes rule can also be used iteratively to update the densities as each new measurement arrives. Thus, the posterior at time t becomes the prior for an update at time $t + \delta t$. This is the basis of sequential data assimilation methods which will be discussed in section 5.3.1.

Simplifying Bayes rule by disregarding the normalization constant gives:

$$f(m|d) \propto f(d|m)f(m) \tag{5.3}$$

A major challenge of using Bayes' law for inferences is sampling from the posterior density $p(m|d)$. Often posterior evaluations require a computationally expensive forward simulation to evaluate the likelihood function, making methods such as Markov chain Monte Carlo (MCMC) methods difficult to apply. MCMC methods use a random walk to construct a Markov chain with a stationary distribution equal to $f(m|d)$. These methods are guaranteed to converge to $f(m|d)$ in the limit of infinite samples and can therefore capture highly non-Gaussian and multimodal distributions. However, constraints on computation time often do not allow users to run MCMC methods for many samples. Significant research has been devoted to variants of the MCMC algorithms in an attempt to accelerate the convergence. However, the sequential nature of this random walk algorithm poses computational challenges that currently prevent consideration of inverting large numbers of parameters constrained by complex dynamics.

In Bayes rule, the likelihood function is equivalent to a probabilistic description of the measurement errors. If Gaussian noise is assumed as in (also assumes i.i.d.):

$$d = G + \eta \tag{5.4}$$

where $\eta \sim N(0, \sigma^2)$, will result in a Gaussian likelihood function

$$f(d|m) \sim \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(G_i - d_i)^2}{2\sigma^2}\right) \tag{5.5}$$

and provides one of the components of the mathematical foundation to consider less statistically robust but computationally more efficient methods such as the Kalman filter [149]. This algorithm recursively averages a prediction of a system's state with new measurements using a weighted average, also known as the covariance and represents the uncertainty associate with the predictions of the systems state. Gaussianity is assumed and the dynamics must be linear. To address the nonlinearities in G, the extended Kalman Filter makes use of Taylor's series to linearize the system. Both algorithms however require an adjoint of the system Jacobian to calculate the covariance. This can be difficult to implement and simply not tractable in complex simulators. The Ensemble Kalman Filter, on the other hand, does not depend on any linearizations or adjoint calculations. In addition it can capture some nonlinear behavior in the model G because the covariance is based on post-processing the ensemble members which depend entirely on the dynamics. The following sections provide

a mathematical formulation starting from Bayes theory to help compare solution strategies using MCMC and EnKF solvers.

5.3 Comparison Formulation

MCMC and the EnKF filter handle data in fundamentally different ways, necessitating an appropriate inference problem that can be formulated as a dynamic data assimilation problem for the EnKF and a batch assimilation (static) inference problem for the MCMC. Consider the general dynamical system

$$\frac{\partial s(x,t)}{\partial t} = G\left(s(x,t), m(x)\right) \tag{5.6}$$

where $s(x,t)$ is the system state, G is a potentially nonlinear model operator parametrized by the model parameters m. Although appropriate boundary conditions and initial conditions also exist these are not explicitly stated to maintain a concise presentation. In addition to the dynamical system, measurements of the system are available through some linear functional M:

$$M[s,m] = d + \varepsilon \tag{5.7}$$

where ε represents additive measurement noise. The goal here is to obtain probabilistic estimate of the model parameters $m(x)$ using the observations of the measurements $d+\varepsilon$. Formulating this inference problem in a Bayesian framework gives the posterior density:

$$f(m|d) \propto f(m) f(d|m) \tag{5.8}$$

Note that spatial and temporal dependencies have been dropped here to simplify notation. The EnKF is generally used to estimate the system state in addition to the model parameters, according to

$$f(s,m|d) \propto f(s,m) f(d|s,m) \tag{5.9}$$

However, attempting to estimate the state and model parameters simultaneously is intractable for MCMC on most applications due to large state dimensions on fine resolution grids. Thus, the formulation in Eq. 5.8 will be used for comparison.

The measurements in d will generally become available at several discrete times $\{t_0, t_1, ..., t_k\}$, $k > 0$. Denote these measurements as $d = \{d_0, d_1, ..., d_k\}$. MCMC will handle all of these observations at once with batch processing, and will use Eq. 5.8 directly. The EnKF on the other hand, will assimilate the data sequentially, updating the estimate of m after each observation. This requires the likelihoods and posterior density in Eq. 5.8 to be expanded in the time dimension.

For both MCMC and the EnKF, measurements are assumed to be uncorrelated in time, and the model integration assumes a Markovian process, implying that given the state and

model parameters at time t_i, the model state at $t_i + \delta t$ is completely determined. No information from $t < t_i$ is necessary. With this assumption, applying Bayes' rule to all available data gives the joint posterior density for the model parameters:

$$f(m|d) \propto f(m) \prod_{j=1}^{k} f(d_j|m) \tag{5.10}$$

5.3.1 Sequential Processing

The static description in Eq. 5.10 gives the joint posterior density after all data have been accumulated. This expression can be used in a MCMC sampling of the posterior; however, the additional sequential structure of the posterior can also be exploited to sequentially update the posterior as in EnKF. At the first time step, the intermediate density will be

$$f(m|d_1) \propto f(m) f(d_1|m) \tag{5.11}$$

and then at the second time step,

$$f(m|d_1, d_2) \propto f(m|d_1) f(d_2|m) \tag{5.12}$$

In general, the intermediate density for the first j measurements is given by

$$f(m|d_1, d_2, ..., d_j) \propto f(m|d_{j-1}) f(d_j|m) \tag{5.13}$$

This recursive update of the posterior can continue until all observations have been processed, resulting in the posterior described by Eq. 5.10. The EnKF uses this formulation with a variance minimizing update of ensemble locations to efficiently approximate the full posterior. Our numerical results (section 5.9) attempts to show that the EnKF can sufficiently approximate the posterior for porous media inference applications.

5.4 Kalman Filter

The EnKF approach depends on fundamental components from the original Kalman Filter algorithm which is described next. Consider a simple case where the operator G is linear and both the error model and prior distribution are Gaussian. The standard Kalman filter fully captures the posterior statistics, which consequently will also be Gaussian. After temporal discretization, this linear system becomes:

$$s_k^t = G s_{k-1}^t + q_{k-1} \tag{5.14}$$

where k is the index in time, the superscript t denotes that this is the true system state, and q is the error in the model dynamics. First, consider a time-step when no observations are

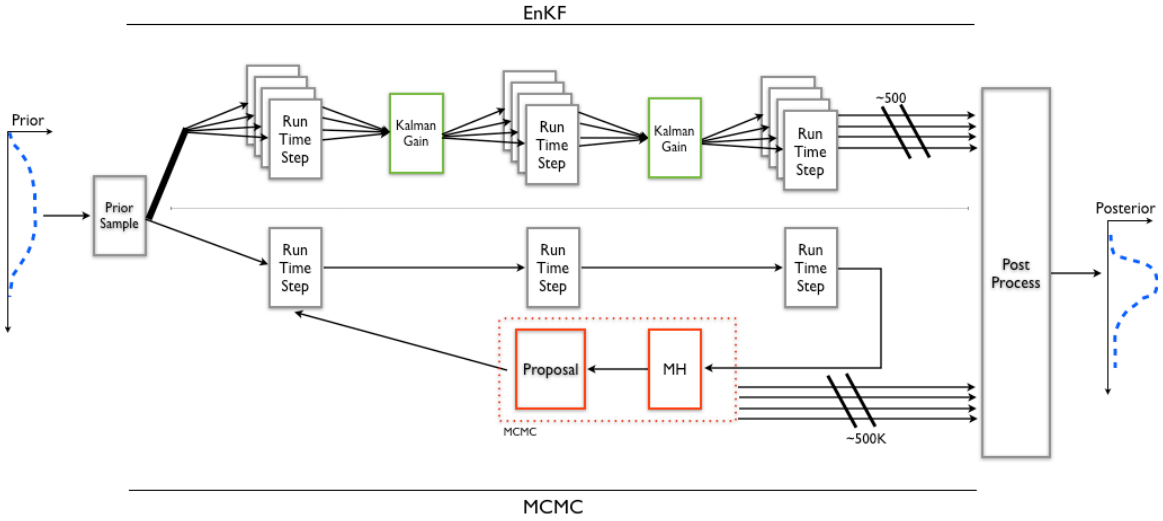


Figure 5.1: Comparison of MCMC and EnKF approach to estimating the posterior density in Eq. 5.10. Note the potential for parallelism in the EnKF.

present. In this case, the model dynamics provide the best estimate of the state at the next time step. The prediction, $s^f(t_k)$, is then given by:

$$s_k^f = Gs_{k-1}^a \tag{5.15}$$

where s^a is the best estimate from the previous time-step. For time steps without observations, $s^f = s^a = E[s]$. Thus, the Kalman filter simply integrates the mean in time to obtain the predictions. The covariance can be propagated forward by first subtracting the best numerical estimate from the unknown true field:

$$\begin{aligned} s_k^t - s_k^f &= Gs_{k-1}^t + q_{k-1} - Gs^a(t_{k-1}) && (5.16)\\ &= G\left(s_{k-1}^t - s_{k-1}^a\right) + q_{k-1} && (5.17) \end{aligned}$$

This is the model forecast error at time t_k when no new observations have occurred. Now, the model error covariance becomes

$$C_{ss}^f = \overline{\left(s_k^t - s_k^f\right)\left(s_k^t - s_k^f\right)^T} = GC_{ss}^a(t_{k-1})G^T + C_{qq}(t_{k-1}) \tag{5.18}$$

where the model state has error covariance

$$C_{ss}^a(t_{k-1}) = \overline{\left(s_{k-1}^t - s_{k-1}^a\right)\left(s_{k-1}^t - s_{k-1}^a\right)^T}$$

and model error covariance is

$$C_{qq}(t_{k-1}) = \overline{q_{k-1}q_{k-1}^T}$$

In most cases, the model G is not linear and variants of the standard Kalman filter need to be used. A popular alternative is the extended Kalman filter (EKF) which simply linearizes the nonlinear operator with a Taylor expansion around the mean, s_{k-1}^a.

The best estimate from the EKF is also computed by integrating the mean forward in time,

$$s_k^f = G(s_{k-1}^a) \tag{5.19}$$

Not surprisingly, nonlinearities in G can cause this estimate to be biased, as thoroughly discussed in [148]. The EKF covariance update step follows the usual approach when extending linear methods to nonlinear cases and uses a truncated Taylor expansion of G, to arrive at the approximate covariance update:

$$C_{ss}^f \approx {G'}_{k-1} C_{ss}^a(t_{k-1}) {G'}_{k-1}^T + C_{qq}(t_{k-1}) \tag{5.20}$$

where ${G'}_{k-1}$ is the Jacobian matrix at the previous time step.

The Kalman filter and EKF updates shown above are used when no data are available. When data are available, the sequential Bayes' formulation in Eq. 5.12 is used in addition to the model dynamics to update the mean and covariance. The "best" estimate in this case is the maximum a posteriori estimate (MAP), obtained by maximizing the log posterior distribution in Eq. 5.12.

With a Gaussian error model and a continuous state, Bayes' rule gives the posterior distribution of the model parameters and system state when observations are present:

$$\begin{aligned} f(s,m|d) \quad &\propto \quad f(s|m) f(d|s,m) \\ &= \exp \quad \left[-\frac{1}{2} \iint \left(s^f(x_1) - s(x_1) \right) W_{ss}^f(x_1,x_2) \left(s^f(x_2) - s(x_2) \right) dx_1 dx_2 \right. \\ &\qquad \left. -\frac{1}{2} \left(d - M[s] \right) W_{\varepsilon\varepsilon} \left(d - M[s] \right) \right] \end{aligned} \tag{5.21}$$

where M is a linear measurement functional relating the field s to d, W_{ss} is the functional inverse of the covariance function, and $W_{\varepsilon\varepsilon} = C_{\varepsilon\varepsilon}^{-1}$ is the inverse of the error covariance. Here, s^f is the system state integrated forward from the previous time-step as if no observations were present. The first term in the exponential is the Gaussian prior and the second term is the likelihood. In a discretized system where the likelihood is a multivariate Gaussian distribution and not a Gaussian process, the integrals would be unnecessary and $W_{ss}(x_1,x_2)$ would be replaced by the matrix $W_{ss} = C_{ss}^{-1}$. The integration simply replaces the summation that occurs when pre and post multiplying W_{ss} by vectors of discretized states. From this posterior distribution, it is clear that when data are available, the best estimate of s occurs at the maximum of this density. This is the MAP estimate for s.

To find the MAP estimate, it is generally easier to minimize the negative log posterior

$$\begin{aligned} -2\log[f(\psi,m|d)] \quad &= \quad \iint \left(\psi^f(x_1) - \psi(x_1) \right) W_{\psi\psi}^f(x_1,x_2) \left(\psi^f(x_2) - \psi(x_2) \right) dx_1 dx_2 \\ &+ \quad \left(d - M[\psi] \right) W_{\varepsilon\varepsilon} \left(d - M[\psi] \right) \end{aligned} \tag{5.22}$$

than to operate on the posterior directly. In [132] a variational derivative and the Euler-Lagrange equation approach was used to minimize this functional and find the data assimilation update step. The result is:

$$s^a = s^f + K\left(d - Ms^f\right) \tag{5.23}$$

where the Kalman gain matrix K is defined as:

$$K = C_{ss}^f M^T \left(MC_{ss}^F M^T + C_{\varepsilon\varepsilon}\right)^{-1} \tag{5.24}$$

and where $C_{\varepsilon\varepsilon}$ is the covariance of the error term ε and C_{ss}^f is:

$$C_{ss}^f = GC_{ss}^a(t_{k-1})G^T + C_{qq}(t_{k-1}). \tag{5.25}$$

Note that s^f and s^a are two estimates for the same time step, the former is the estimate before data assimilation and the latter is the analyzed estimate that includes observations. It is the covariance of the model error that differentiates the EnKF from the standard Kalman Filter, which requires the transpose (or the adjoint operator) of G and is not always easily implemented, especially in parallel code. The EnKF approach calculates this covariance by post-processing a collection (the ensemble) of forward runs per time-step.

5.5 Ensemble Kalman Filter

In addition to the adjoint implementation and the errors that arise when using a truncated Taylor expansion as an approximation to a nonlinear function, the standard Kalman filter algorithms also have to store a potentially large covariance matrix. Many realistic problems, especially in geophysics, have millions of degrees of freedom, making it impossible to store the covariance matrix required by the KF and EKF. The ensemble Kalman filter (EnKF) resolves two of the major issues with the KF and EKF: the EnKF does not need to store the covariance matrix and no linearization of the model operator is required. Instead, the EnKF uses ensemble averaging to compute the mean and covariance after each update. Randomly instantiated ensemble members are propagated between observation times with the deterministic model G. Using the ensemble members as Monte Carlo samples of the model parameters, m, and system states, s, the ensemble mean and covariance are used instead of propagating the mean and covariance directly. Data assimilation is similar to the Kalman filter, except each of the ensemble members is updated individual by replacing the mean and covariance in Eq. 5.23 with ensemble equivalents. The ensemble mean is chosen as the best estimate although this estimate may be biased. Thus,

$$s_k^f = \overline{s_k} = \frac{1}{N_e}\sum_{j=1}^{N_e} s_{j,k} = \frac{1}{N_e}\sum_{j=1}^{N_e} G(sj, k-1) \tag{5.26}$$

where the subscript s_j denotes the j^{th} ensemble member, and N_e is the ensemble size. Now, defining the ensemble error covariances as:

$$(C_{ss}^e)^f = \overline{\left(s^f - \overline{s^f}\right)\left(s^f - \overline{s^f}\right)^T} \tag{5.27}$$

$$(C_{ss}^e)^a = \overline{\left(s^a - \overline{s^a}\right)\left(s^a - \overline{s^a}\right)^T} \tag{5.28}$$

where the ensemble average defined by $\bar{\cdot}$ is defined in Eq. 5.26. To incorporate observations, ensemble members are individually updated according to the KF assimilation step:

$$s_j^a = s_j^f + (C_{ss}^e)^f M^T \left(M (C_{ss}^e)^f M^T + C_{\varepsilon\varepsilon}^e\right)^{-1} \left(d - M s_j^f\right) \tag{5.29}$$

By using the mean and covariances in this way, the ensemble Kalman filter assumes the system state and parameters have a Gaussian PDF; however, ensemble members are individually propagated through the model G and can therefore capture some of the model nonlinearity. By using the KF update during assimilation, the EnKF is maximizing a Gaussian posterior for the update step. For a highly nonlinear G, this may not be a sufficient approximation. For certain porous media flow applications as discussed below, the EnKF is an apt procedure. In different dynamics and possibly with bimodal or highly skewed distribution, EnKF may not be able to capture all nonlinearities and non-Gaussian statistics. As we will demonstrate, the EnKF is able to reconstruct permeabilities and predict saturation profiles from a complex porous media dataset and is easily extended to a hybrid approach in which other information is processed with a Markov Chain Monte Carlo method.

5.5.1 Implementation Details

In Eq. 5.29, the inverse of $\left(M(C_{\psi\psi}^e)^f M^T + C_{\varepsilon\varepsilon}^e\right)$, which we will label S, needs to be computed. Because S may not be full rank in some cases, the Moore-Penrose pseudo inverse S^+ is used, [150]. When S is full rank (i.e. there are more measurements than parameters) then the pseudo inverse is equivalent to the inverse, $C^+ = C^{-1}$. The pseudo inverse can be derived through an eigenvalue factorization:

$$S = Z\Lambda Z^T$$

where the columns of Z hold the eigenvectors and Λ is a diagonal matrix containing the eigenvalues in increasing order. The pseudo inverse is then defined as:

$$S^+ = Z\Lambda^+ Z^T$$

where Λ^+ is still diagonal but

$$\text{diag}(\Lambda^+) = \left(\lambda_1^{-1}, ..., \lambda_p^{-1}, 0, ..., 0\right)$$

where $p = \text{rank}(S)$ and the eigenvalues are in increasing order, $\lambda_i \geq \lambda_{i+1}$. In practice, the pseudo inverse is not computed directly. Instead, given that S^+ is the solution to the least squares problem $Sx = b$, the solution x can be found by more efficient and stable algorithms. Additionally, Evenson [132] presents a subspace pseudo inversion algorithm that reduces the factorization of the full covariance matrix C to factoring a matrix in the ensemble space of size N_e.

The equivalence of the EnKF and KF update equations is based on an assumption of infinite ensemble size, which ensures the ensemble statistics are equal to the true mean and covariance. In the infinite ensemble case, the EnKF will provide identical results to the KF. However, in most cases, results in finite time are desired and a finite ensemble size will be used. The finite ensemble can only provide an estimate of $\overline{\psi^f}$ and $C^e_{\psi\psi}$. In some cases, particularly with small ensemble sizes, poor estimates may allow the EnKF to make unphysical predictions. Evenson [132] calls these effects spurious correlations, and suggests a localization scheme to dampen their effects. In addition to errors in the state estimate, these spurious correlations can cause the ensemble variance to shrink and underestimate the true variance. Thus, artificial inflation of the variance has been proposed in [132].

5.5.2 Localization and Inflation

To reduce the effects of high correlations between distant points, Houtekamer et al. [151] suggest the computation of a Schur product (elementwise product) of the ensemble covariance matrix with a different covariance matrix derived from a given covariance function and correlation length. Let $C^{loc}_{\varepsilon\varepsilon}$ and C^{loc}_{ss} be the localized covariances defined as:

$$C^{loc}_{\varepsilon\varepsilon} = A_\varepsilon \cdot C^e_{\varepsilon\varepsilon} \tag{5.30}$$

$$C^{loc}_{ss} = A_\psi \cdot C^e_{ss} \tag{5.31}$$

where A_ε and A_ψ are covariance matrices with the same sizes as $C^e_{\varepsilon\varepsilon}$ and $C^e_{\psi\psi}$ respectively. These matrices are built from a correlation function $\rho(\cdot)$. In our case, a Gaussian correlation function with a length of $5\delta x$ was used, where δx is the uniform grid size.

While the localization approach will help filter out spurious correlations, the ensemble variance will still decrease. Inflation is used to alleviate this problem. An inflation factor is used to update the ensemble at each step with the expression:

$$\psi_j = \rho(\psi_j - \overline{\psi}) + \overline{\psi} \tag{5.32}$$

where ρ is the inflation factor, usually slightly larger than one (~ 1.02).

5.6 Markov Chain Monte Carlo Methods

We next describe Markov chain Monte Carlo (MCMC) methods.We highlight the computational and the statistical differences between the two methods. Additionally, EnKF is designed for transient problems and is not capable of easily incorporating static data. For our porous media problem the goal is to perform a reconstruction of permeability, as well as a prediction of dynamics, using not only observations of the dynamics, but also static measurements. This leads to our hybrid MCMC-EnKF approach.

Markov chain Monte Carlo (MCMC) methods are statistically robust methods for sampling any target probability density $\pi(x)$ as long as a function proportional to the density can be evaluated. These methods construct a Markov chain in parameter space that has $\pi(x)$ as a stationary distribution. Thus, as the Markov chain evolves, the sample statistics of the chain approach the statistics of $\pi(x)$. A popular approach is to combine Metropolis-Hastings (MH) and MCMC to update the Markov chain using a proposal distribution $q(x';x^t)$ with rejection sampling. From the current position x^t in the chain, a sample of the proposal distribution is taken and then evaluated on the target density, $\pi(x)$. The probability, α, that the chain will move to this proposed state is given by the Metropolis-Hastings rule:

$$\alpha = \min\left\{1, \frac{\pi(x')q(x^t;x')}{\pi(x^t)q(x';x^t)}\right\} \tag{5.33}$$

The chain continues to be updated in this fashion until a maximum number of iterations have been performed. The difficulty in MH-based MCMC is to choose an appropriate and efficient proposal density because if the proposal is too wide, (large variance), many proposals will be rejected and little information will be gained about $\pi(x)$. On the other hand, if the proposal is too small, the chain will only explore a limited area of the parameter space and distribution tails or multiple high probability regions will be characterized properly. Delayed rejection and the Adaptive Metropolis methods proposed in [152] and [76] respectively, are modifications to MH MCMC that attempt to overcome initially poor proposal distributions and dramatically improve mixing. In fact, combining these methods [77] can operate successfully on some problems where the basic method fails. The many other MCMC variants in the literature are beyond the scope of this paper; however, some impressive results on specific applications can be achieved, see [153, 154]. Here, only the DRAM MCMC is investigated because of its growing application to Bayesian inference across scientific and engineering disciplines [155, 156, 157].

5.7 MCMC-EnKF Hybrid Method

The major drawback of using MCMC for full dynamic inference is the large number of required forward simulations. On the other hand MCMC has the capability to fully characterize the posterior density whereas the EnKF filter implicitly assumes the posterior covariance is an accurate description . In our porous media application, typically a hybrid

approach combining the power of MCMC with the efficiency of the EnKF is a natural step forward. A hybrid method is implemented here where MCMC is used for static inversion of a permeability field using a KL decomposition, taking advantage of the very rapid likelihood evaluation. The posterior MCMC samples are then used to select initial ensemble members. Since the EnKF estimate is always a linear combination of the initial ensemble, pulling KL dimensions of the ensemble members from a good estimate is highly advantageous. The numerical results show a dramatic improvement in this hybrid over a standard EnKF implementation. Fig. 5.2 shows the general structure of the hybrid algorithm.

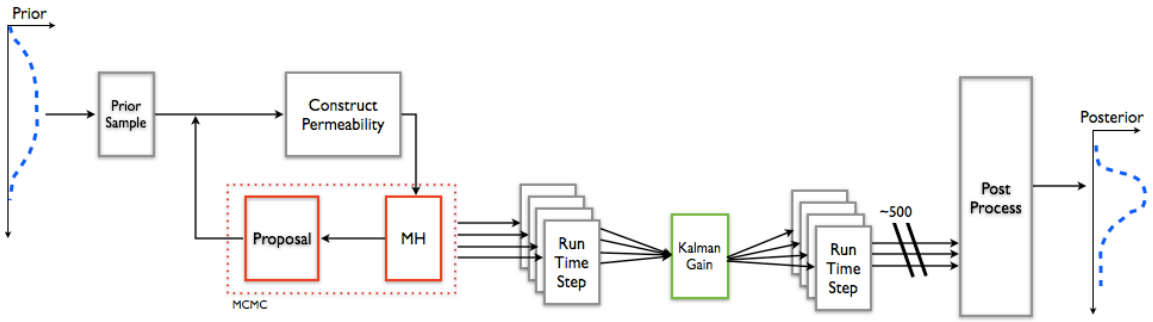


Figure 5.2: Outline of MCMC-EnKF hybrid.

5.8 An Example from Hydrology

For comparison of the EnKF and MCMC, we will consider a two phase flow problem with unknown permeability, as described by Aarnes et al. [158]. Two phases are modeled, a water phase and a hydrocarbon phase. The hydrocarbon phase consists of oil with dissolved gas and is assumed to be immiscible with the water phase. Furthermore, both liquid phases and the aquifer are assumed to be incompressible. With these restrictions, conservation of mass requires:

$$\phi \frac{\partial s_w}{\partial t} + \nabla \cdot v_w = \frac{q_w}{\rho_w} \tag{5.34}$$

$$\phi \frac{\partial s_o}{\partial t} + \nabla \cdot v_o = \frac{q_o}{\rho_o} \tag{5.35}$$

where ϕ is porosity, s_w and s_o are the saturations of water and oil respectively, v_w and v_o are the Darcy velocities, q_w and q_o are source/sink terms, and ρ_w and ρ_o are the respective phase densities. We also require

$$s_w + s_o = 1$$

Using this constraint and then combining Eq. 5.34 with Eq. 5.35, gives the mixed pressure equation:

$$\nabla \cdot v = q \tag{5.36}$$

$$v = -\left[K\lambda_w(\nabla p_w - \rho_w G) + K\lambda_o(\nabla p_o - \rho_o G)\right] \tag{5.37}$$

where $q = q_w/\rho_w + q_o/\rho_o$, $\lambda_w = k_{rw}/\mu_w$ and $\lambda_o = k_{ro}/\mu_o$ are the phase mobilities, K is the permeability tensor, and $G = -g\nabla z$. In solving this expression, a global pressure $p = p_o - p_c$ is introduced where

$$p_c = \int_1^{s_w} f_w(\xi)\frac{\partial p_{cov}}{\partial s_w}(\xi)d\xi$$

is a saturation dependent complementary pressure and $f_w = \lambda - 2/(\lambda_w + \lambda_o)$ is the fractional flow function. Introducing the total mobility $\lambda = \lambda_w + \lambda_o$, the global pressure equation is found:

$$-\nabla \cdot \left[K\lambda\nabla p - K\left(\lambda_w\rho_w + \lambda_o\rho_o\right)G\right] = q \tag{5.38}$$

The relatively permeabilities needed to compute the mobilities depend on saturations. Thus, given saturations, the pressure equation allows us to solve for p; however, the saturations are dependent on the pressure. This coupling means two additional equations relating p, s_w and s_o are needed. By definition, $s_o = 1 - s_w$, so Darcy's law,

$$v_\alpha = -K\frac{k_{r\alpha}}{\mu_\alpha}(\nabla p_\alpha - \rho_\alpha G)$$

provides the final necessary relationship for a complete model. Using Darcy's velocity in the incompressible and immiscible case gives

$$K\lambda_o\lambda_w\nabla p_{cow} = \lambda_o v_w - \lambda_w v_o + K\lambda_o\lambda_w(\rho_o - \rho_w)G \tag{5.39}$$

which can be rearranged to give the standard saturation equation:

$$\phi\frac{\partial s_w}{\partial t} + \nabla \cdot \left(f_w(s_w)\left[v + d(s_w, \nabla s_w) + g(s_w)\right]\right) = \frac{q_w}{\rho_w} \tag{5.40}$$

Note that each term $f(s)v$, $f(s)d(s,\nabla s)$, and $f(s)g(s)$ represent viscous, capillary, and gravitational forces respectively. In general the capillary forces are small compared to the other two processes. This causes the parabolic equation in Eq. 5.40 to behave almost as a hyperbolic equation, suggesting that the same discretization used on the almost elliptic pressure equation will not be applicable here. To remedy this issue, an operator splitting approach where the solution of the pressure equation is solved using the previous timestep's saturations and with this updated pressure, the saturation equation is solved. Within each saturation solve or pressure solve, a simple finite volume scheme is used for discretization and the Newton-Raphson method is used for nonlinear solves. Details of this strategy can be found in [158].

5.8.1 Karhunen-Loeve Expansion of Permeability Field

A Gaussian process prior is used for the log-permeability field. The process is usually discretized to a multivariate Gaussian of size $N_x \times N_y$. However, this is still high dimensional, making estimation difficult. To reduce the dimension, a Karhunen-Loeve expansion of the system can be used. A Karhunen-Loeve (KL) expansion is a generalization of the Fourier expansion where a process is projected onto eigenfunctions. The KL expansion of a random process, $Y(x,\omega)$, separates real space from stochastic space as follows:

$$Y(x,\omega) = \sum_{i=0}^{\infty} \sqrt{\lambda_i}\phi_i(x)z_i(\omega) \tag{5.41}$$

where λ_i and ϕ are the eigenvalues and eigenfunctions of a covariance kernel $C(x_1,x_2)$ that satisfy the integral equation:

$$\int_D C(x_1,x_2)\phi_k(x_2)dx_2 = \lambda_k\phi_k(x_1) \tag{5.42}$$

and $z_i(\omega)$ are uncorrelated random variables. In our case $Y(x,\omega)$ is assumed Gaussian and thus $z_i(\omega)$ are not only uncorrelated but are iid standard normal random variables.

A numerically robust way of solving λ_i and $\phi_i(x)$ in discretized space is to solve Eq. 5.42 using the Nystrom method. It is also possible to simply discretize Eq. 5.42 at the grid points and solve the eigenvalues and eigenvectors of the resulting covariance matrix; but in that case, only a relatively small number of the eigenvalues can be trusted as accurate.

The Nystrom method is a general method for solving integral equations similar to Eq. 5.42 by using a quadrature rule for discretization. Consider the general N point quadrature formula:

$$I[f(x)] \approx \sum_{i=1}^{N} w_i f(x_i)$$

Using this for the integral in Eq. 5.42 yields the discretized linear system:

$$\lambda_k\phi_k(x_1) = \sum_{i=1}^{N} w_i C(x_1,x_i)\phi_k(x_i) \tag{5.43}$$

By also evaluating this function at the quadrature points, we have the system:

$$\lambda_k\phi_k(x_j) = \sum_{i=1}^{N} w_i C(x_j,x_i)\phi_i(x_i) \tag{5.44}$$

This linear system can be used to solve for the eigenvalues and eigenvectors at the quadrature points. However, we are not interested in eigenvectors at the quadrature points but rather the grid points. This is simple to overcome by noticing Eq. 5.43 is also an interpolation formula. Now that λ_i and $\phi_k(x_i)$ are known, the value of $\phi_k(x_1)$ can be easily be computed by finding the covariance of x_1 and x_i. Here, a Gauss-Lagrange quadrature rule was used but in general, any quadrature scheme could be easily incorporated.

Once the KL expansion has been computed and truncated, the permeability field is completely parametrized by the random variables z_i. Since the initial process was Gaussian, all z_i are i.i.d. standard normal random variables. This is the prior density for each z_i. By trying to estimate each z_i instead of the entire field, a dramatic increase in performance can be achieved by both MCMC and the EnKF. Additional errors introduced during the KL truncation can be controlled by looking at the spectrum of λ and truncated after most of the energy has been preserved.

5.9 Numerical Results - Sensitivities

The goal of this section is to evaluate the sensitivity and statistical accuracy of the EnKF algorithms. The posterior distributions of permeability are compared by evaluating sensitivities of KL modes, observations, ensemble size, model discretization, observation noise and length scale. Given the optimal combination of parameters, the statistical accuracy is evaluated by comparing statistical moments from a MCMC solver.

The Gaussian assumption in the updating process of the EnKF imposes some statistical restrictions but it is unclear what the effects are on reconstructing permeabilities and predicting dynamics. In certain design or control problems, robust solutions will depend on the complete characterization of higher statistical moments. For instance, failure of structural members may be caused by material properties that are characterized in the tails of non-Gaussian distributions. However, in our porous media problem where the goal is ultimately to identify bypassed oil and delay the breakthrough of water at the producing well, the lack of failure modes and the presence of smoothing effects in the dynamics may suppress the importance of higher order moments. This part of the numerical studies will therefore focus on the quality of the EnKF inversions and compare moments to MCMC results.

In particular, this section will report on

- the computational efficiency of the methods by measuring the total number of required simulations,
- accuracy of the mean estimate; since artificial observations are used, the true parameters of the model are known, the root mean square error is used to measure the accuracy of MCMC and EnKF parameter estimates,
- how well the EnKF captures the posterior density compared to the MCMC posterior density; this is described qualitatively by comparing marginal posterior densities and quantitatively through the Kullback-Leibler divergence of the posterior densities.

Our numerical prototyping procedure considers the problem of estimating a permeability field $k(x)$ from noisy observations of saturation:

$$d(t) = Ms_w(t) + \varepsilon$$

at discrete times $\{t_1, t_2, t_3, ...\}$, where M is a measurement matrix and ε is additive Gaussian noise. To avoid an "inversion crime", 200 KL modes are used to generate the "truth" model and fewer modes are used for the inference. Furthermore, $k(x)$ is generated by using a Gaussian process, and a KL expansion is used to parameterize the field.

As with most numerical methods, the use of EnKF with the porous media transport model has several parameters that require tuning. To this end, a Latin Hypercube Design (LHD) strategy systematically explores the influence of the multidimensional parameter space on the reconstruction of permeability and the prediction of the waterflood dynamics. The Root Mean Square Error (RMSE) of the ensemble mean estimate, ensemble estimate variance, and posterior parameter density entropy are computed from the EnKF output as measures of the EnKF performance. All tests are performed with Karhunen-Loeve (KL) parametrized permeability fields augmented with sparse and noisy observations of a water saturation field.

The first numerical experiment is an 84 point LHD over all of the parameters with bounds as specified in Table 5.1. Ensemble estimates of the mean, variance, skewness and kurtosis are computed at each point in the computational domain for both permeabilities and saturations. A RMSE is computed to compare the reconstruction to the truth model. The LHD strategy generates statistics at a sparse set of points in the 7 dimensional tunable parameter space and to visualize general trends and relationships, kriging is used to generate continuous interpolated surfaces [159]. The kriging estimate for the RMSE fits a Gaussian process with quadratic mean to the sparse EnKF RMSE information from the LHD points. Coefficients of normalized parameters in the quadratic regression with a large magnitude indicate an important factor for EnKF tuning. Furthermore, plotting the kriging estimate shows general trends in the RMSE with respect to EnKF parameters. Table 5.2 shows the kriging coefficients for each tunable parameter and Table 5.3 show each pair of tunable parameters.

Table 5.1: Bounds used on Latin Hypercube Design of all EnKF inversion parameters and fixed values for sub-designs.

Parameter	Lower Bound	Upper Bound	Fixed
Number of Sensors (S)	10	500	50
Ensemble Size (E)	10	1500	500
Number KL Modes (K)	10	200	100
Discretization (D)	10×10	40×40	20×20
Noise Variance (N)	1e-4	0.05	1e-3
Prior Length Scale (L)	2e-2	6e-2	5e-2
Observation spacing in time (O)	1 t-step	10 t-steps	1 t-step

The magnitude of the regression coefficients indicates the relative importance of the parameter but the coefficient sign also gives useful information. For example, grid resolution has the largest effect on EnKF performance. Additionally, the negative sign on the number of sensors shows, as expected, that increasing the number of sensors will in general reduce

Table 5.2: Linear model coefficient estimates for permeabilities and saturations, as a function of various parameters. The definitions of the parameters S, E, K etc are in Table 5.1, as are their fixed values.

Parameter	Perm. Coefficient	Sat. Coefficient
S	-5.01e-02	-1.88e-01
E	-6.45e-03	2.77e-02
K	1.14e-01	1.58e-01
D	9.63e-01	5.44e-01
N	1.27e-01	3.06e-01
L	9.50e-02	1.52e-03
O	7.08e-02	-9.80e-02

Table 5.3: Permeability and saturation coefficient estimates for pairs of the LHS runs. The top half of the table are the permeability coefficients and the lower half are the saturation values; the first value on the diagonal are permeability coefficients and the second value are saturation coefficients. See Table 5.1 for the definition and details of the parameters.

	S	E	K	D	N	L	O
S	0.02/0.05	-0.002	-0.004	-0.04	0.006	-0.003	0.02
E	0.2	0.01/-0.2	-0.005	0.02	-0.01	0.006	-0.005
K	-0.1	-0.03	-0.04/-0.09	0.06	-0.01	-0.01	-0.0008
D	-0.2	-.2	0.2	0.2/0.3	0.05	0.04	0.04
N	-0.03	0.009	-0.1	0.2	-0.05/0.006	0.01	-0.01
L	0.2	-0.1	0.1	-0.2	-0.02	-0.06/0.2	0.0009
O	0.4	0.03	-0.1	0.4	-0.1	0.1	-0.02/0.1

the final estimate error. Table 5.3, shows the number of KL modes, the noise variance, and the number of sensors are also important for accurate EnKF predictions.

In a general application, the grid discretization and the observation noise structure will be predefined through the forward modeling process and scientific judgment. Beyond these parameters, the large sensitivity run shows the number of observations, the ensemble size, and the number of KL modes to be the next most important parameters. The joint effects of these parameters is studied in more detail below, using additional Latin Hypercube Designs. 24 point designs were used for each parameter pair, resulting in three additional LHDs.

Tables 5.4, 5.5, 5.7, 5.8, 5.10, and 5.11 summarize the sensitivity runs while Tables 5.6, 5.9, and 5.12 show the kriging coefficients and relative parameter importance. In the tables, red represents the largest RMSE value, green represents the lowest RMSE, purple is the largest skew, and brown is the smallest skew. The skew here is used as an indication of non-Gaussian behavior in the estimated posterior density.

In the standard Kalman filter or extended Kalman filter, no non-Gaussian behavior can

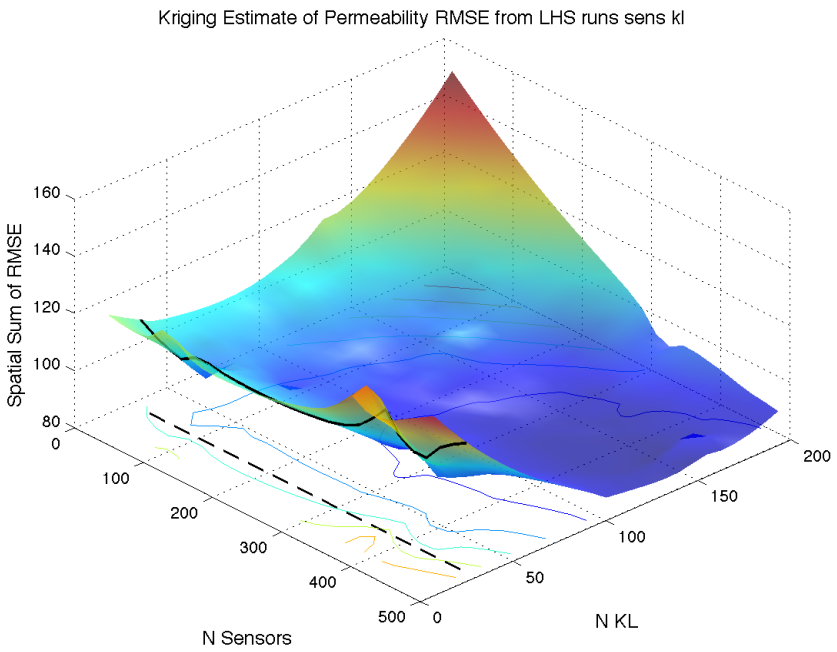


Figure 5.3: Kriging estimate comparison of sensors and KL for permeability.

be captured. While the Gaussian assumption of the EnKF for data assimilation seems to restrict the posterior to be nearly Gaussian, the nonzero skewness in the results suggests the EnKF filter can capture more statistical information about the posterior than just mean and covariance. The statistical capabilities of EnKF are analyzed more rigorously in the next section, where EnKF posterior estimates are compared to MCMC posterior estimates.

Based on Tables 5.12 and 5.6, the number of KL modes is the main indicator of EnKF performance for the permeability estimate, but is not the most important factor when predicting saturation. The saturation is most dependent on the number of sensors. The saturation can be thought of as a smoothed function of permeability, thus it is no surprise that a parameter dealing directly with saturation, the number of observations, is more important than the number of KL modes. Any errors caused by too few KL modes will be partially filtered by the parabolic saturation equation and will have a smaller influence on the saturation. On the other hand, being directly parametrized by the KL modes the permeability will be more sensitive to an inadequate basis. In a more general context this implies that the advantages of using a more sophisticated model (more KL modes in this case) depends on the goal of model calibration. For example, a user interested in accurate state predictions for oil production analysis will require a relatively low fidelity model when compared to a researcher interested in characterizing subsurface structures. In this case, the user interested in oil production would be best suited to finding more measurements or reducing measurement error, while the researcher interested in geological structure would find it more beneficial to investigate modeling assumptions (isotropy, etc...). Figs. 5.3 and 5.4 illustrate the sensitivity of the RMSE in saturation and permeability to the number of KL modes and number of sensors. Fig. 5.5 shows the L_∞ error between the true saturation and the best EnKF

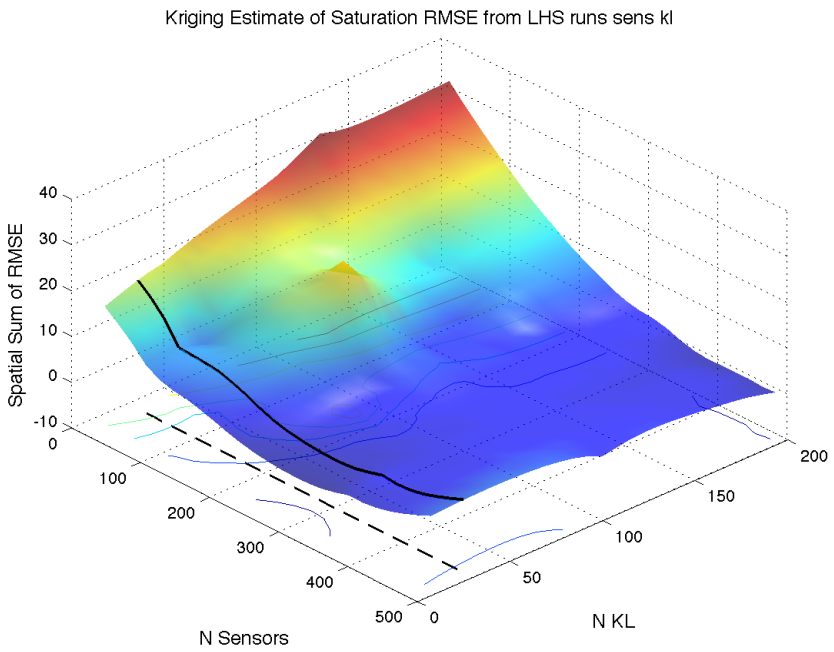


Figure 5.4: Kriging estimate comparison of sensors and KL for saturation.

prediction for a range of sensor and KL coefficients.

Fixing the number of KL modes and investigating the effects of the number of sensors and the number of ensembles shows that after a suitable ensemble size has been reached, further increasing the ensemble size has little effect on EnKF performance. Table 5.9 and Fig. 5.8 illustrate this point. Fig. 5.8 shows that after approximately 200 ensembles, the difference between the true saturation and EnKF estimated saturation is dependent mainly on the number of sensors. Here, the small contour plots show the absolute difference between the true and predicted saturation fields while the color of the scatter point represents the relative RMSE obtained at that point. Thus, when a reasonable ensemble size is used, the number of observations is the primary indicator of EnKF performance.

Figs. 5.6 and 5.7 show the sensitivity of the RMSE in saturation and permeability to the number of sensors and ensembles. Fig. 5.8 shows the L_∞ error between the true saturation and the best EnKF prediction for a range of sensors and ensembles. Figs. 5.9 and 5.10 show the sensitivity of the RMSE in saturation and permeability to the number of sensors and ensembles. Fig. 5.11 shows the L_∞ error between the true saturation and the best EnKF prediction for a range of sensors and ensembles.

The goal of statistical parameter estimation is to provide a good estimate of the parameters and also to quantify the uncertainty in that estimate. Both the EnKF and MCMC provide this information but Gaussian assumptions made during the data assimilation step in the EnKF affect the quality of the final EnKF statistical estimate. Here, the statistical quality of the EnKF posterior estimates are compared to MCMC results as an indication of EnKF

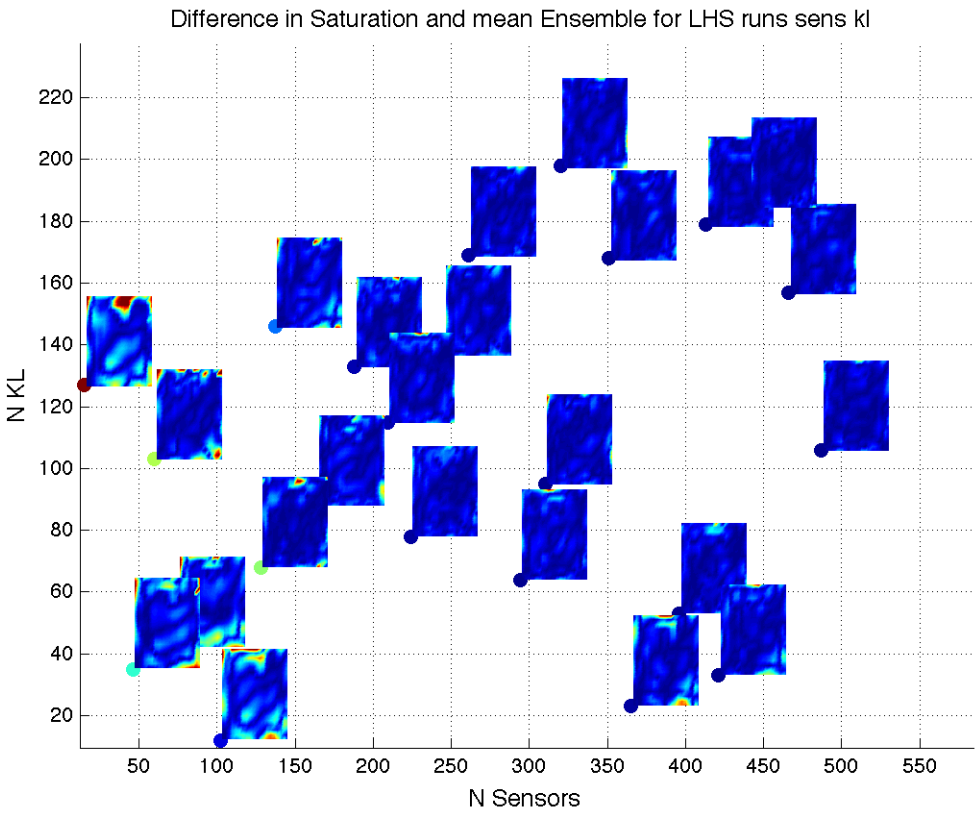


Figure 5.5: L_∞ error between true saturation and best EnKF prediction. The comparison is made between sensors and KL for permeability.

applicability to porous media inference problems. Two comparisons between the MCMC and EnKF are made, first we consider a fixed number of KL modes and use each algorithm to infer permeabilities. Secondly, we place a computational constraint on the methods and change the number of KL modes for the EnKF to demonstrate the strengths of the EnKF compared to MCMC.

In the first comparison, three runs of DRAM MCMC were performed with either 10,000 samples or 500,000 samples and initial proposal variances of 0.1 or 0.3. The EnKF filter used 500 ensembles. The reconstructed permeabilities and respective saturations for each of the MCMC runs is shown next to the EnKF results in Figs. 5.13-5.18. Clearly, the MCMC reconstructed permeability is suspect. With 30 parameters to estimate in a nonlinear model such as this, ensuring proper mixing of the DRAM chain is very difficult and the results shown in these figures is most likely related to an insufficient exploration of the parameter space.

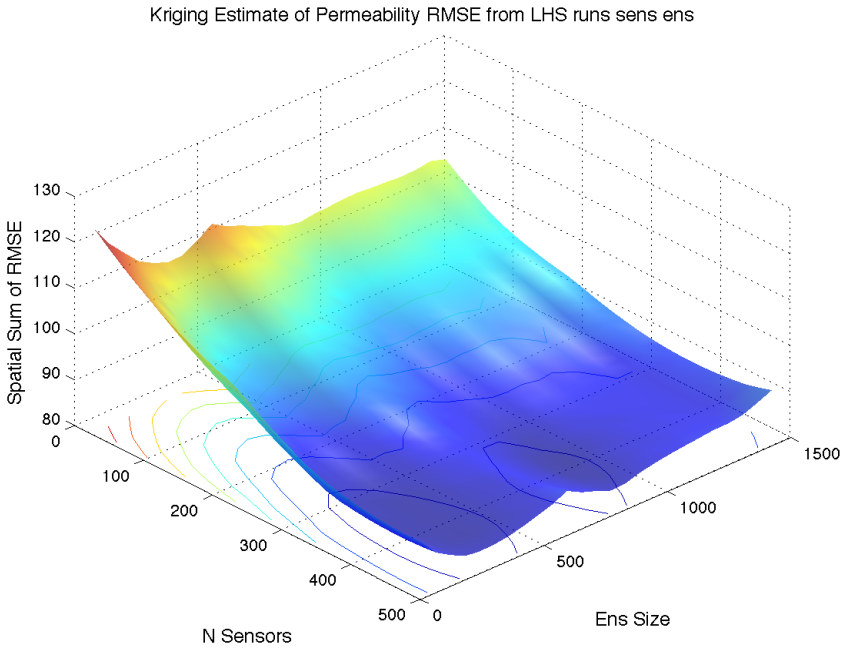


Figure 5.6: Kriging estimate comparison of sensors and ensemble size.

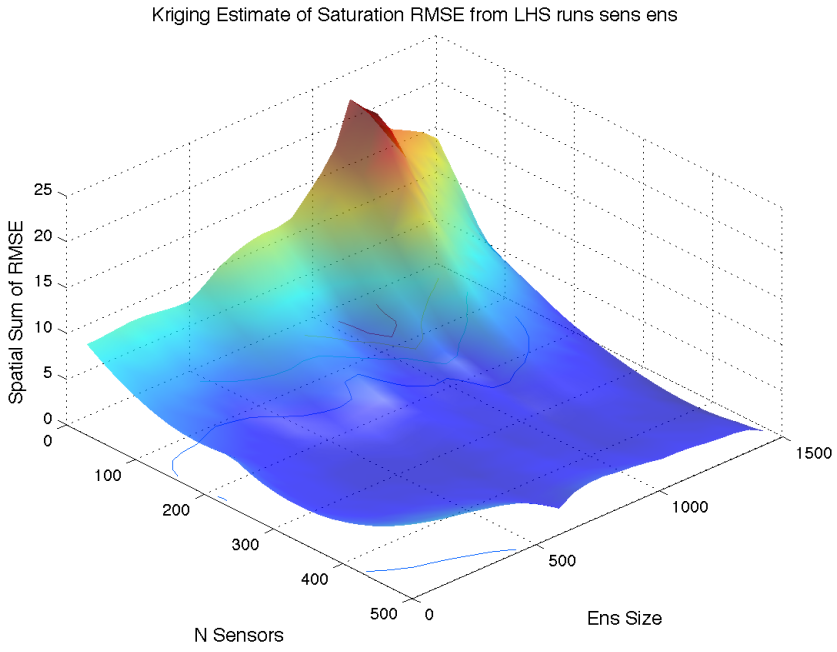


Figure 5.7: Kriging estimate comparison of sensors and ensemble size for saturation.

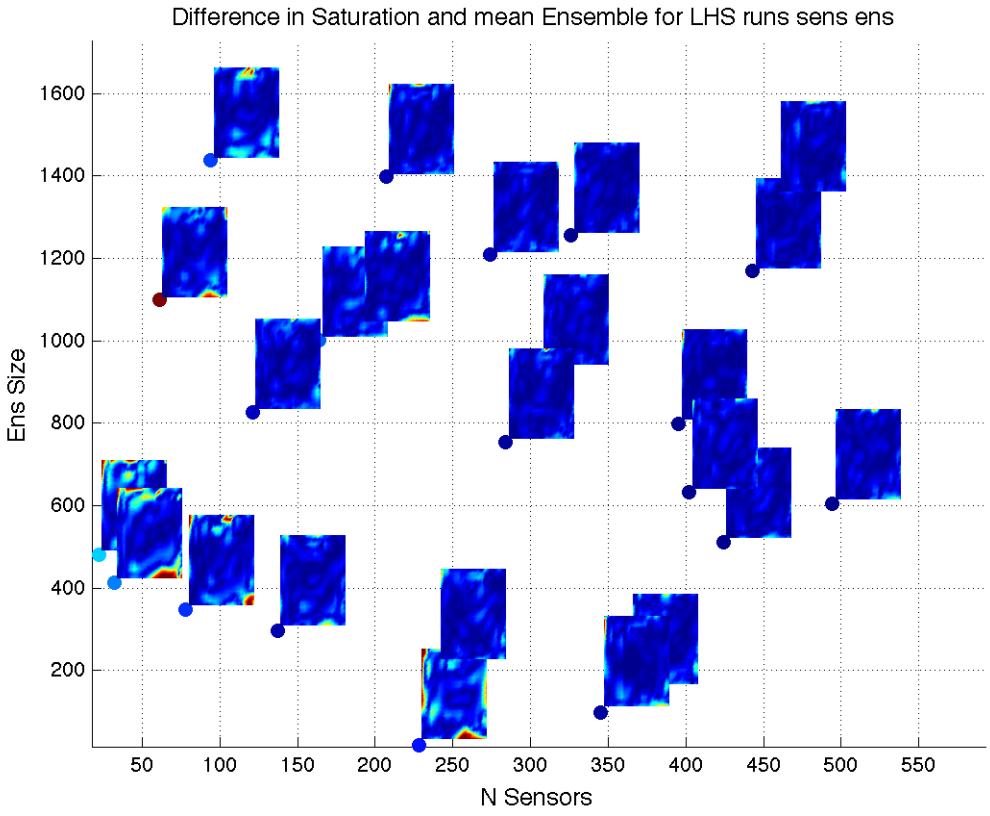


Figure 5.8: L_∞ error between true saturation and best EnKF prediction. The comparison is made between N sensors and ensemble size.

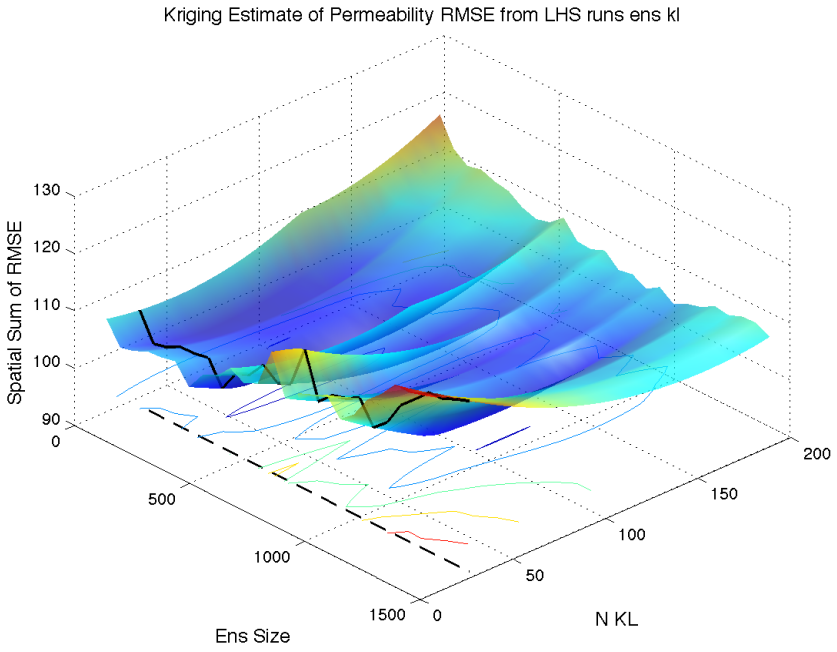


Figure 5.9: Kriging estimate comparison of ensemble size and KL.

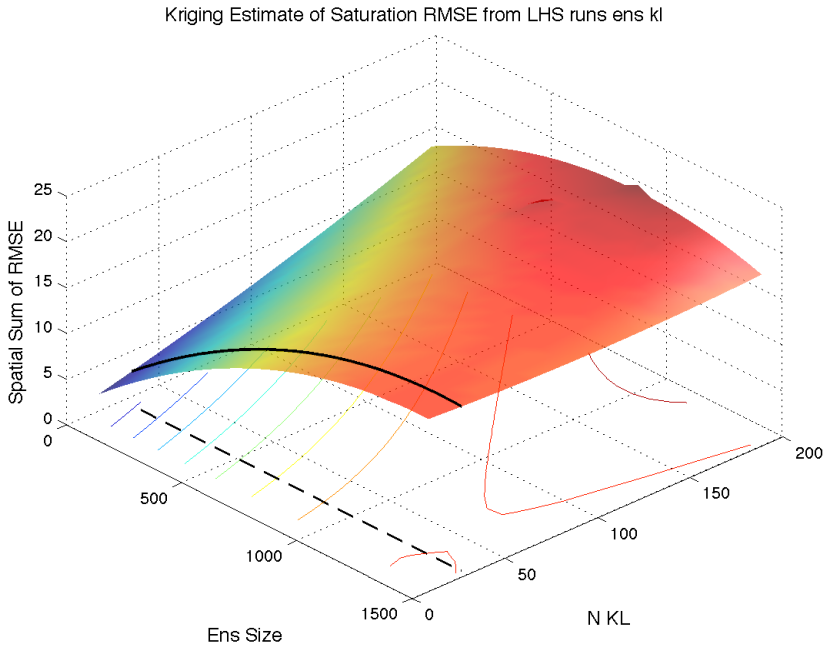


Figure 5.10: Kriging estimate comparison of ensemble size and KL for saturation.

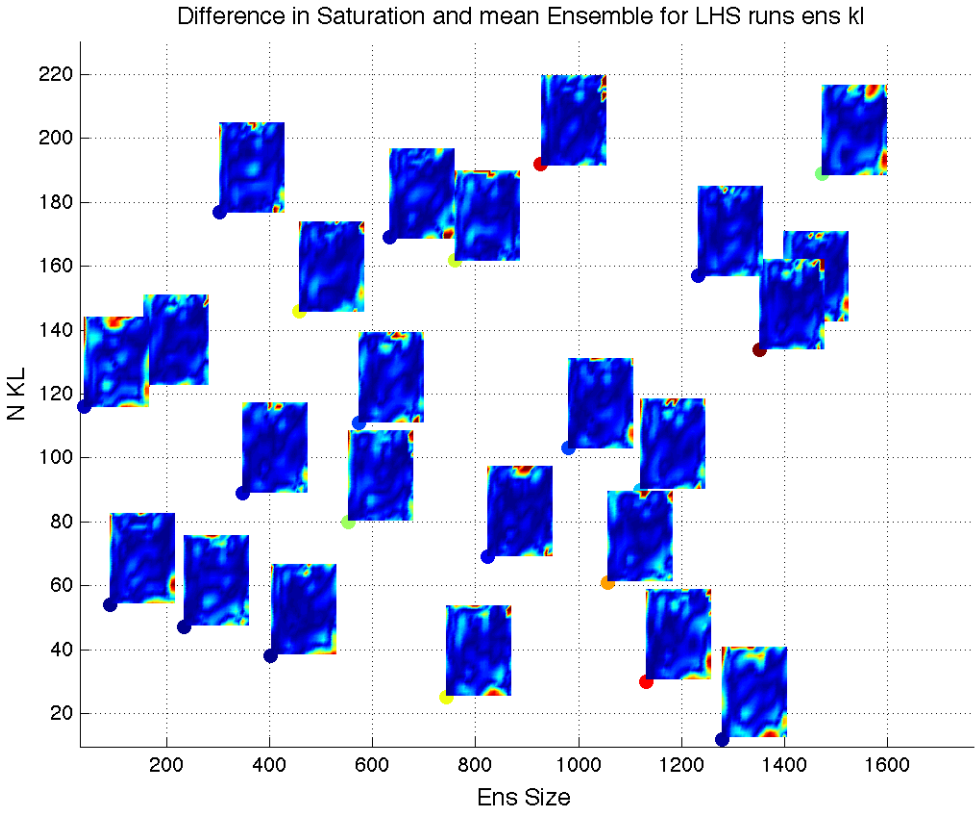


Figure 5.11: L_∞ error between true saturation and best EnKF prediction. The comparison is made between ensemble size and KL.

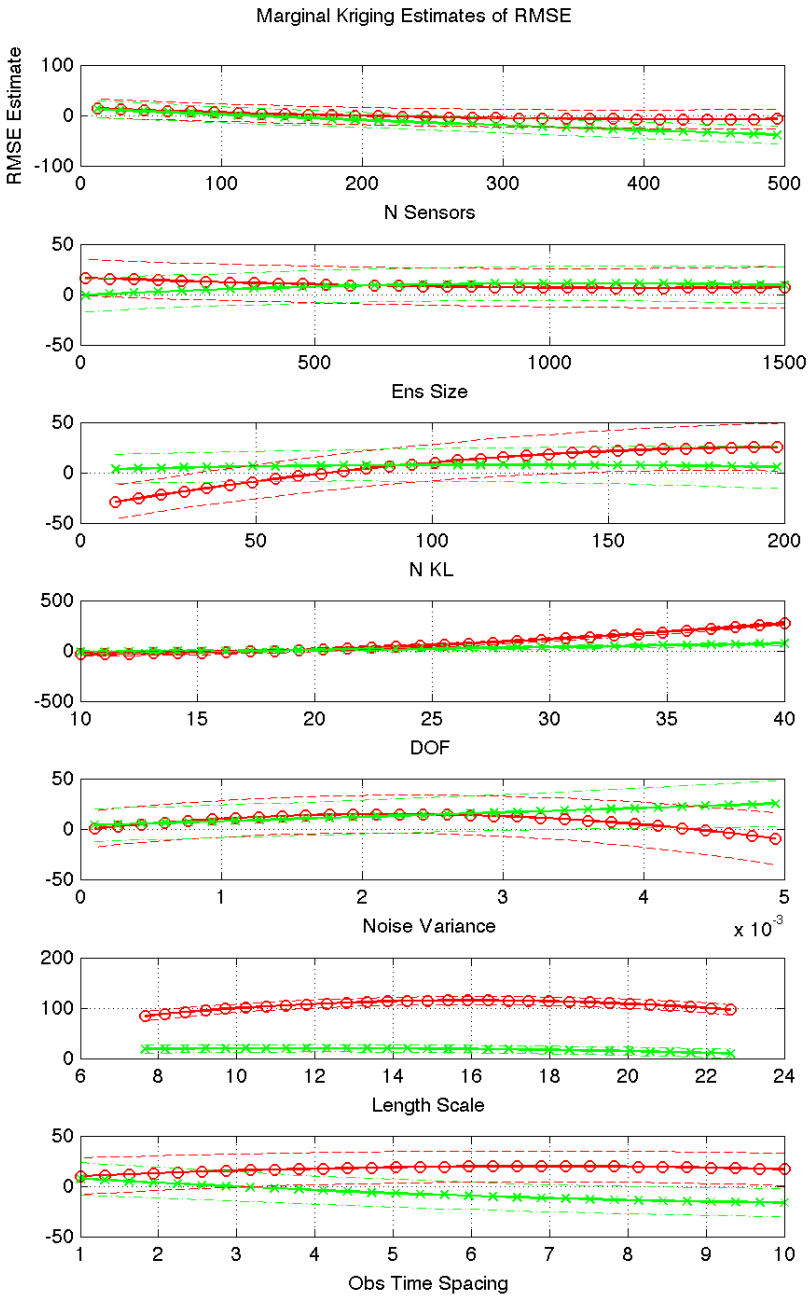


Figure 5.12: Kriging estimate of RMSE at fixed slices. Green lines with 'x' are for the saturation prediction and red lines with 'o' represent permeability RMSE. Kriging surface is made from data using LHS.

Table 5.4: Ensemble (error) statistics for permeability field as a function of the number of sensors (N Sensors) and the number of Karhunen-Loève modes (N KL). All other parameters were fixed at the values detailed in Table. 5.1.

N Sensors	N KL	$\sum$ rmse	min rmse	max rmse	min var	max var	mean skew
365	23	1.41e+02	2.72e-02	1.09e+00	6.46e-04	1.75e-02	3.05e-02
396	53	8.84e+01	7.38e-02	6.97e-01	4.47e-03	3.88e-02	1.32e-01
320	198	8.88e+01	1.24e-01	6.91e-01	1.48e-02	5.17e-02	7.99e-02
60	103	1.08e+02	1.63e-01	7.66e-01	2.32e-02	5.27e-02	7.82e-02
245	137	9.04e+01	1.35e-01	4.95e-01	1.61e-02	5.62e-02	6.77e-02
224	78	9.02e+01	1.14e-01	5.66e-01	9.67e-03	4.08e-02	7.61e-02
413	179	8.97e+01	1.18e-01	5.46e-01	1.35e-02	4.27e-02	7.97e-02
15	127	1.19e+02	1.92e-01	6.82e-01	3.55e-02	5.98e-02	3.75e-02
137	146	9.53e+01	1.45e-01	6.57e-01	2.05e-02	5.05e-02	8.69e-02
294	64	8.80e+01	8.39e-02	5.54e-01	7.03e-03	4.07e-02	6.24e-02
487	106	8.45e+01	1.04e-01	5.90e-01	9.17e-03	4.04e-02	7.75e-02
102	12	1.26e+02	3.31e-02	1.00e+00	1.26e-04	1.18e-02	5.48e-02
351	168	9.22e+01	1.28e-01	6.40e-01	1.46e-02	4.28e-02	5.28e-02
188	133	1.01e+02	1.40e-01	7.07e-01	1.53e-02	5.81e-02	6.56e-02
310	95	8.82e+01	1.03e-01	6.06e-01	1.04e-02	4.43e-02	8.50e-02
466	157	9.21e+01	1.29e-01	6.78e-01	1.17e-02	4.04e-02	8.30e-02
261	169	8.98e+01	1.24e-01	5.37e-01	1.50e-02	4.59e-02	8.17e-02
441	185	8.62e+01	1.22e-01	6.21e-01	1.30e-02	5.51e-02	7.90e-02
209	115	9.40e+01	1.25e-01	5.84e-01	1.43e-02	4.48e-02	8.11e-02
164	88	9.50e+01	1.28e-01	6.06e-01	1.24e-02	4.66e-02	1.09e-01
421	33	1.05e+02	5.16e-02	9.30e-01	1.84e-03	2.19e-02	6.52e-02
75	42	9.27e+01	1.10e-01	5.81e-01	5.99e-03	4.62e-02	9.92e-02
46	35	1.01e+02	9.39e-02	6.91e-01	4.92e-03	4.04e-02	8.19e-02
46	35	8.75e+01	1.18e-01	4.83e-01	1.09e-02	4.36e-02	9.57e-02

Table 5.5: Ensemble (error) statistics for the final saturation field as a function of the number of sensors (N Sensors) and the number of Karhunen-Loève modes (N KL). All other parameters were fixed at the values detailed in Table. 5.1.

N Sensors	N KL	∑ rmse	min rmse	max rmse	min var	max var	mean skew
365	23	3.00e+00	1.14e-12	6.92e-02	1.27e-24	4.65e-04	3.05e-02
396	53	2.26e+00	7.39e-13	4.93e-02	5.33e-25	1.14e-03	1.32e-01
320	198	2.62e+00	2.27e-07	4.73e-02	3.45e-14	2.23e-03	7.99e-02
60	103	1.85e+01	6.48e-07	8.54e+00	1.56e-13	5.90e+01	7.82e-02
245	137	2.84e+00	8.87e-13	4.10e-02	7.79e-25	1.38e-03	6.77e-02
224	78	2.60e+00	8.68e-13	4.13e-02	7.35e-25	1.65e-03	7.61e-02
413	179	2.34e+00	1.01e-12	5.06e-02	7.26e-25	1.52e-03	7.97e-02
15	127	3.30e+01	8.17e-07	1.32e+01	2.05e-13	1.15e+02	3.75e-02
137	146	9.05e+00	5.30e-07	4.94e+00	8.79e-14	2.26e+01	8.69e-02
294	64	2.50e+00	7.19e-07	4.87e-02	2.05e-13	7.31e-04	6.24e-02
487	106	2.07e+00	8.36e-13	3.62e-02	6.93e-25	1.10e-03	7.75e-02
102	12	4.66e+00	9.18e-13	9.58e-02	7.91e-25	9.12e-04	5.48e-02
351	168	2.52e+00	7.59e-13	5.73e-02	5.63e-25	2.25e-03	5.28e-02
188	133	3.21e+00	8.17e-13	1.04e-01	5.89e-25	2.00e-03	6.56e-02
310	95	2.44e+00	6.89e-08	5.55e-02	3.79e-15	6.92e-04	8.50e-02
466	157	2.17e+00	1.75e-07	3.29e-02	2.68e-14	1.08e-03	8.30e-02
261	169	2.64e+00	8.03e-13	9.65e-02	6.30e-25	9.26e-03	8.17e-02
441	185	2.14e+00	1.48e-07	4.22e-02	1.49e-14	1.52e-03	7.90e-02
209	115	3.09e+00	1.72e-08	1.23e-01	2.94e-16	1.42e-02	8.11e-02
164	88	2.81e+01	3.30e-07	2.49e+01	8.88e-14	6.04e+01	1.09e-01
421	33	2.44e+00	3.16e-07	4.78e-02	6.80e-14	4.13e-04	6.52e-02
75	42	5.23e+00	8.66e-13	2.41e-01	7.29e-25	4.96e-02	9.92e-02
46	35	1.47e+01	4.84e-07	5.03e+00	1.12e-13	2.53e+01	8.19e-02
46	35	1.79e+01	4.77e-07	1.40e+01	1.31e-13	1.03e+02	9.57e-02

Table 5.6: Linear model coefficient estimates for permeability and saturation as a function of the number of sensors and Karhunen-Loève modes.

Parameter	Perm. Coefficient	Sat. Coefficient
N Sensors	-4.18e-01	-7.14e-01
N KL	-2.20e-01	1.67e-01
(N Sensors)×(N Sensors)	3.09e-01	4.70e-01
(N Sensors)×(N KL)	-5.48e-01	-2.32e-01
(N KL)×(N KL)	6.85e-01	1.67e-01

Table 5.7: Ensemble (error) statistics for permeability field, as a function of the number of sensors (N Sensors) and the EnKF ensemble size (Ens Size). Other parameters were kept fixed at the values detailed in Table 5.1.

N Sensors	Ens Size	∑ rmse	min rmse	max rmse	min var	max var	mean skew
228	19	1.05e+02	1.77e-02	1.13e+00	1.50e-04	4.90e-03	-1.48e-01
137	296	9.49e+01	1.22e-01	5.29e-01	1.19e-02	4.07e-02	7.66e-02
274	1209	8.87e+01	1.14e-01	5.79e-01	1.21e-02	4.57e-02	1.00e-01
94	1439	9.84e+01	1.61e-01	5.21e-01	2.44e-02	5.14e-02	9.70e-02
22	480	1.15e+02	1.84e-01	5.75e-01	2.87e-02	5.98e-02	5.84e-02
164	1002	9.34e+01	1.32e-01	5.32e-01	1.50e-02	4.68e-02	1.03e-01
443	1169	8.65e+01	1.02e-01	5.15e-01	1.02e-02	4.39e-02	1.00e-01
191	1039	9.31e+01	1.36e-01	5.44e-01	1.42e-02	4.65e-02	9.66e-02
326	1256	8.65e+01	1.15e-01	5.89e-01	1.19e-02	4.44e-02	9.88e-02
364	153	8.52e+01	8.95e-02	5.87e-01	5.09e-03	2.50e-02	7.20e-02
32	413	1.09e+02	1.58e-01	6.29e-01	2.47e-02	5.80e-02	7.20e-02
494	604	8.90e+01	1.14e-01	5.45e-01	9.60e-03	4.00e-02	7.68e-02
345	99	8.67e+01	6.81e-02	5.54e-01	3.85e-03	1.74e-02	-1.53e-02
61	1099	1.03e+02	1.64e-01	6.27e-01	2.34e-02	5.50e-02	1.01e-01
240	215	8.59e+01	1.15e-01	5.61e-01	9.13e-03	3.38e-02	8.98e-02
395	799	8.23e+01	1.15e-01	4.27e-01	1.08e-02	4.65e-02	9.99e-02
424	512	8.49e+01	1.11e-01	5.21e-01	1.03e-02	4.03e-02	7.58e-02
121	826	9.46e+01	1.47e-01	5.11e-01	1.62e-02	5.03e-02	8.95e-02
78	347	1.01e+02	1.44e-01	5.89e-01	1.88e-02	4.86e-02	7.17e-02
402	632	8.62e+01	1.09e-01	5.51e-01	1.07e-02	4.17e-02	8.98e-02
459	1357	8.73e+01	1.12e-01	4.80e-01	1.22e-02	4.44e-02	1.16e-01
306	935	8.72e+01	1.16e-01	6.06e-01	1.21e-02	4.53e-02	9.37e-02
207	1399	9.54e+01	1.30e-01	5.22e-01	1.53e-02	4.98e-02	1.16e-01
207	1399	8.73e+01	1.11e-01	5.40e-01	1.22e-02	4.54e-02	9.62e-02

Table 5.8: Ensemble (error) statistics for the final saturation field, as a function of the number of sensors (N Sensors) and the EnKF ensemble size (Ens Size). Other parameters were kept fixed at the values detailed in Table 5.1.

N Sensors	Ens Size	∑ rmse	min rmse	max rmse	min var	max var	mean skew
228	19	5.23e+00	6.42e-08	1.33e-01	4.67e-16	1.70e-04	-1.48e-01
137	296	3.17e+00	8.24e-13	4.36e-02	6.76e-25	1.82e-03	7.66e-02
274	1209	3.14e+00	5.07e-07	4.40e-01	9.53e-14	1.93e-01	1.00e-01
94	1439	6.21e+00	2.53e-08	9.48e-01	6.41e-16	8.98e-01	9.70e-02
22	480	9.92e+00	1.25e-07	1.75e+00	1.39e-14	3.05e+00	5.84e-02
164	1002	7.01e+00	4.02e-07	3.70e+00	6.14e-14	1.32e+01	1.03e-01
443	1169	2.11e+00	8.02e-13	3.53e-02	6.13e-25	1.06e-03	1.00e-01
191	1039	3.89e+00	1.48e-07	2.91e-01	2.12e-14	7.63e-02	9.66e-02
326	1256	2.28e+00	5.67e-07	4.87e-02	1.29e-13	1.71e-03	9.88e-02
364	153	2.35e+00	8.45e-13	3.16e-02	5.33e-25	6.43e-04	7.20e-02
32	413	7.93e+00	8.49e-13	4.97e-01	6.92e-25	2.45e-01	7.20e-02
494	604	2.21e+00	7.93e-13	3.98e-02	5.98e-25	9.97e-04	7.68e-02
345	99	2.61e+00	7.51e-13	4.83e-02	5.50e-25	2.62e-04	-1.53e-02
61	1099	2.59e+01	8.35e-07	1.28e+01	2.19e-13	6.07e+01	1.01e-01
240	215	2.57e+00	7.05e-13	4.70e-02	4.70e-25	9.34e-04	8.98e-02
395	799	2.12e+00	8.53e-13	5.24e-02	7.24e-25	5.55e-04	9.99e-02
424	512	2.20e+00	7.77e-13	4.10e-02	6.03e-25	5.99e-04	7.58e-02
121	826	3.52e+00	8.16e-13	1.45e-01	6.58e-25	2.09e-02	8.95e-02
78	347	6.17e+00	5.40e-08	3.71e-01	2.08e-15	1.35e-01	7.17e-02
402	632	2.23e+00	3.20e-07	3.70e-02	5.66e-14	5.51e-04	8.98e-02
459	1357	2.22e+00	1.23e-07	3.78e-02	1.49e-14	1.09e-03	1.16e-01
306	935	2.39e+00	7.52e-13	4.13e-02	5.54e-25	8.05e-04	9.37e-02
207	1399	3.02e+00	7.97e-13	6.66e-02	6.05e-25	3.53e-03	1.16e-01
207	1399	2.57e+00	4.46e-07	1.51e-01	6.71e-14	1.72e-02	9.62e-02

Table 5.9: Linear model coefficient estimates for permeability and saturation, as a function of the number of sensors (N Sensors) and EnKF ensemble size (Ens Size).

Parameter	Perm. Coefficient	Sat. Coefficient
N Sensors	-8.42e-01	-6.25e-01
Ens Size	-1.79e-01	1.00e-01
(N Sensors)×(N Sensors)	3.99e-01	4.84e-01
(N Sensors)×(Ens Size)	1.78e-01	-3.01e-01
(Ens Size)×(Ens Size)	2.91e-01	1.00e-01

Table 5.10: Ensemble (error) statistics for permeability field as a function of the EnKF ensemble size (Ens Size) and the number of Karhunen-Loève modes (N KL). All other parameters are kept fixed at the values in Table 5.1.

Ens Size	N KL	∑ rmse	min rmse	max rmse	min var	max var	mean skew
760	162	1.05e+02	1.67e-01	5.72e-01	2.77e-02	6.03e-02	9.20e-02
980	103	1.05e+02	1.69e-01	6.61e-01	2.59e-02	5.35e-02	8.90e-02
1120	90	1.03e+02	1.58e-01	5.74e-01	2.36e-02	5.28e-02	1.04e-01
553	80	1.02e+02	1.33e-01	6.06e-01	1.76e-02	5.27e-02	8.63e-02
1056	61	9.99e+01	1.33e-01	5.38e-01	1.34e-02	5.29e-02	8.96e-02
574	111	1.09e+02	1.67e-01	6.87e-01	2.75e-02	5.75e-02	8.75e-02
927	192	1.05e+02	1.78e-01	6.63e-01	2.95e-02	6.01e-02	8.43e-02
1397	143	1.10e+02	1.82e-01	6.26e-01	3.00e-02	5.99e-02	8.19e-02
155	123	9.96e+01	1.49e-01	5.45e-01	1.76e-02	5.55e-02	9.34e-02
403	38	9.71e+01	9.73e-02	6.76e-01	5.75e-03	4.39e-02	1.00e-01
90	54	9.79e+01	1.02e-01	6.67e-01	7.08e-03	4.28e-02	2.12e-02
458	146	1.02e+02	1.66e-01	4.55e-01	2.61e-02	5.77e-02	7.38e-02
234	47	1.03e+02	1.07e-01	6.06e-01	6.18e-03	4.69e-02	1.02e-01
744	25	1.26e+02	6.37e-02	1.01e+00	2.07e-03	3.76e-02	8.70e-02
304	177	1.07e+02	1.60e-01	6.38e-01	2.51e-02	6.52e-02	6.46e-02
1131	30	9.89e+01	8.69e-02	6.79e-01	3.47e-03	4.22e-02	7.81e-02
1351	134	1.08e+02	1.78e-01	5.75e-01	2.81e-02	5.55e-02	7.24e-02
1279	12	1.25e+02	4.19e-02	1.11e+00	1.77e-04	1.82e-02	9.30e-02
1472	189	1.08e+02	1.82e-01	5.97e-01	3.11e-02	6.14e-02	8.67e-02
823	69	9.97e+01	1.48e-01	5.58e-01	1.83e-02	4.99e-02	9.52e-02
348	89	1.03e+02	1.49e-01	7.73e-01	1.99e-02	5.08e-02	8.58e-02
1232	157	1.07e+02	1.78e-01	5.25e-01	3.09e-02	5.76e-02	7.70e-02
633	169	1.05e+02	1.72e-01	7.39e-01	2.88e-02	5.41e-02	7.65e-02
633	169	1.11e+02	7.43e-02	7.54e-01	2.58e-03	3.62e-02	9.95e-03

Table 5.11: Ensemble (error) statistics for the final saturation field as a function of the EnKF ensemble size (Ens Size) and the number of Karhunen-Loève modes (N KL). All other parameters are kept fixed at the values in Table 5.1.

Ens Size	N KL	∑ rmse	min rmse	max rmse	min var	max var	mean skew
760	162	2.29e+01	6.84e-07	1.15e+01	1.61e-13	1.05e+02	9.20e-02
980	103	9.76e+00	2.89e-07	2.99e+00	4.24e-14	8.88e+00	8.90e-02
1120	90	1.41e+01	3.42e-07	3.80e+00	7.18e-14	1.43e+01	1.04e-01
553	80	2.12e+01	5.31e-07	9.92e+00	1.49e-13	7.11e+01	8.63e-02
1056	61	2.75e+01	6.59e-07	1.36e+01	1.73e-13	1.29e+02	8.96e-02
574	111	9.96e+00	1.54e-07	1.54e+00	2.17e-14	2.38e+00	8.75e-02
927	192	3.34e+01	1.18e-06	1.21e+01	3.13e-13	1.19e+02	8.43e-02
1397	143	6.47e+00	1.25e-12	4.09e-01	8.44e-25	1.42e-01	8.19e-02
155	123	3.19e+01	6.43e-07	1.44e+01	2.31e-13	8.97e+01	9.34e-02
403	38	4.43e+00	2.04e-08	2.11e-01	3.60e-16	4.31e-02	1.00e-01
90	54	4.15e+00	1.04e-12	1.10e-01	6.57e-25	8.37e-03	2.12e-02
458	146	2.39e+01	7.65e-07	9.31e+00	2.20e-13	5.18e+01	7.38e-02
234	47	4.58e+00	1.01e-12	2.53e-01	5.21e-25	6.40e-02	1.02e-01
744	25	2.39e+01	5.80e-07	1.04e+01	1.69e-13	9.36e+01	8.70e-02
304	177	5.83e+00	1.22e-12	3.32e-01	7.26e-25	4.51e-02	6.46e-02
1131	30	3.24e+01	6.66e-07	1.37e+01	2.93e-13	1.41e+02	7.81e-02
1351	134	3.67e+01	9.16e-07	1.11e+01	3.68e-13	1.02e+02	7.24e-02
1279	12	5.16e+00	1.16e-12	2.23e-01	1.08e-24	2.88e-02	9.30e-02
1472	189	2.02e+01	5.43e-07	1.08e+01	1.47e-13	8.06e+01	8.67e-02
823	69	7.53e+00	1.25e-12	6.06e-01	7.45e-25	2.53e-01	9.52e-02
348	89	6.19e+00	1.07e-12	5.39e-01	6.62e-25	2.33e-01	8.58e-02
1232	157	6.45e+00	1.35e-12	4.07e-01	7.25e-25	1.65e-01	7.70e-02
633	169	5.70e+00	1.33e-12	2.80e-01	8.49e-25	3.85e-02	7.65e-02
633	169	6.18e+00	1.50e-12	4.40e-01	8.00e-25	1.93e-01	9.95e-03

Table 5.12: Linear model coefficient estimates for permeability and saturation, as a function of ensemble size (Ens Size) and number of Karhunen-Loève modes (N KL).

Parameter	Perm. Coefficient	Sat. Coefficient
Ens Size	2.60e-01	2.66e-01
N KL	-7.25e-02	1.55e-01
(Ens Size)×(Ens Size)	2.41e-01	-1.40e-01
(Ens Size)×(N KL)	-3.11e-01	-9.57e-02
(N KL)×(N KL)	3.47e-01	1.55e-01

Table 5.13: Linear model coefficient estimates permeability and saturation, as a function of various parameters (as detailed in Table 5.1) and some of their second order combinations.

Parameter	Perm. Coefficient	Sat. Coefficient
N Sensors	-5.01e-02	-1.88e-01
Ens Size	-6.45e-03	2.77e-02
N KL	1.14e-01	1.58e-01
DOF	9.63e-01	5.44e-01
Noise Variance	1.27e-01	3.06e-01
Length Scale	9.50e-02	1.52e-03
Obs Time Spacing	7.08e-02	-9.80e-02
(N Sensors)×(N Sensors)	2.14e-02	5.12e-02
(N Sensors)×(Ens Size)	-2.48e-03	1.84e-01
(N Sensors)×(N KL)	-4.01e-03	-1.14e-01
(N Sensors)×(DOF)	-4.23e-02	-2.00e-01
(N Sensors)×(Noise Variance)	6.80e-03	-2.94e-02
(N Sensors)×(Length Scale)	-3.07e-03	1.52e-01
(N Sensors)×(Obs Time Spacing)	2.26e-02	3.99e-01
(Ens Size)×(Ens Size)	1.17e-02	-1.72e-01
(Ens Size)×(N KL)	-5.66e-03	-3.38e-02
(Ens Size)×(DOF)	2.45e-02	-1.50e-01
(Ens Size)×(Noise Variance)	-1.24e-02	8.95e-03
(Ens Size)×(Length Scale)	6.27e-03	-1.46e-01
(Ens Size)×(Obs Time Spacing)	-5.53e-03	3.12e-02
(N KL)×(N KL)	-4.00e-02	-9.39e-02
(N KL)×(DOF)	6.44e-02	1.57e-01
(N KL)×(Noise Variance)	-1.46e-02	-9.70e-02
(N KL)×(Length Scale)	-1.55e-02	1.24e-01
(N KL)×(Obs Time Spacing)	-8.49e-04	-9.74e-02
(DOF)×(DOF)	1.92e-01	3.07e-01
(DOF)×(Noise Variance)	5.42e-02	1.80e-01
(DOF)×(Length Scale)	4.20e-02	-1.83e-01
(DOF)×(Obs Time Spacing)	4.02e-02	-4.20e-01
(Noise Variance)×(Noise Variance)	-4.83e-02	5.51e-03
(Noise Variance)×(Length Scale)	1.41e-02	-2.19e-02
(Noise Variance)×(Obs Time Spacing)	-1.21e-02	-1.05e-01
(Length Scale)×(Length Scale)	-6.98e-02	-1.55e-01
(Length Scale)×(Obs Time Spacing)	9.29e-04	6.55e-02
(Obs Time Spacing)×(Obs Time Spacing)	-1.77e-02	1.23e-01

Table 5.14: Ensemble (error) statistics for the permeability field, as a function of the EnKF ensemble size (varied between 500 and 4500). All other parameters were kept fixed at the values in Table 5.1.

Ens Size	$\sum$ rmse	min rmse	max rmse	min var	max var	mean skew
500	1.99e+02	4.56e-02	6.29e-01	2.38e-04	2.12e-02	6.30e-02
1500	1.95e+02	4.66e-02	6.01e-01	2.82e-04	2.24e-02	6.24e-02
2500	1.92e+02	4.15e-02	6.34e-01	2.69e-04	2.20e-02	6.99e-02
3500	2.05e+02	4.33e-02	6.74e-01	2.81e-04	2.21e-02	6.53e-02
4500	1.97e+02	4.26e-02	6.72e-01	2.69e-04	2.15e-02	6.83e-02

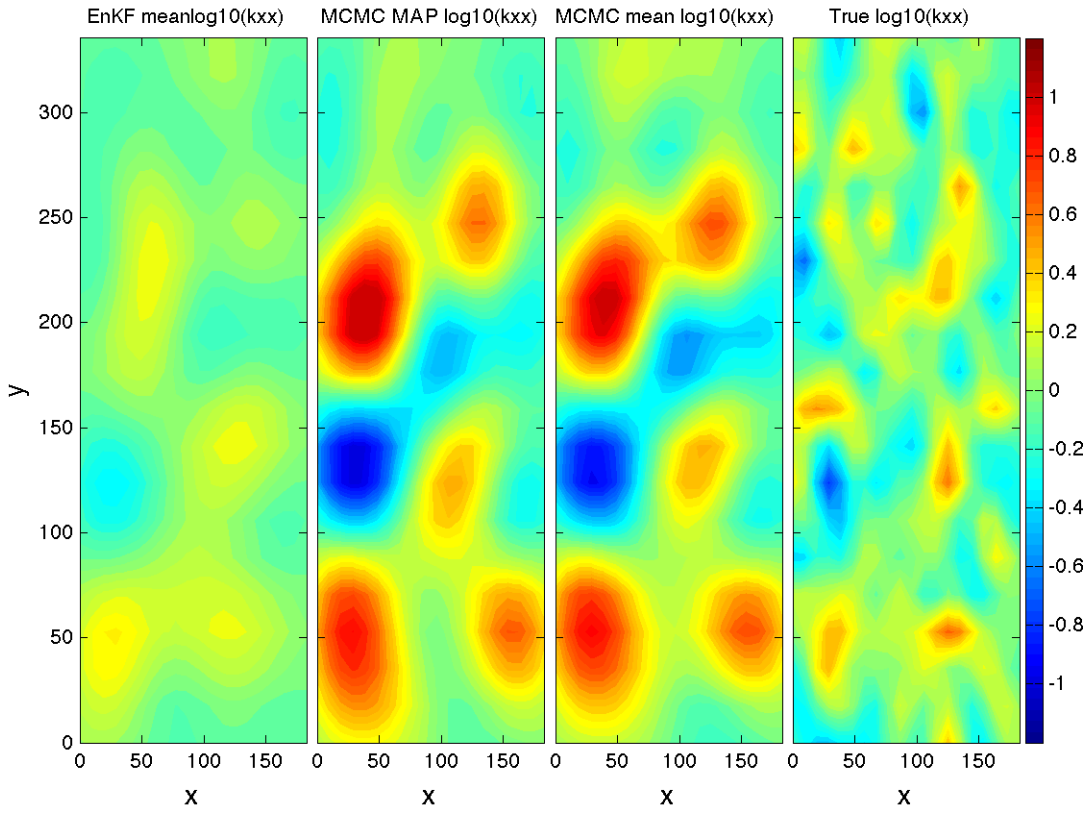


Figure 5.13: Comparison of MCMC and EnKF permeability estimates. The EnKF consisted of 500 ensemble members. The MCMC was performed with an initial proposal covariance of 0.1, and 10,000 samples were drawn.

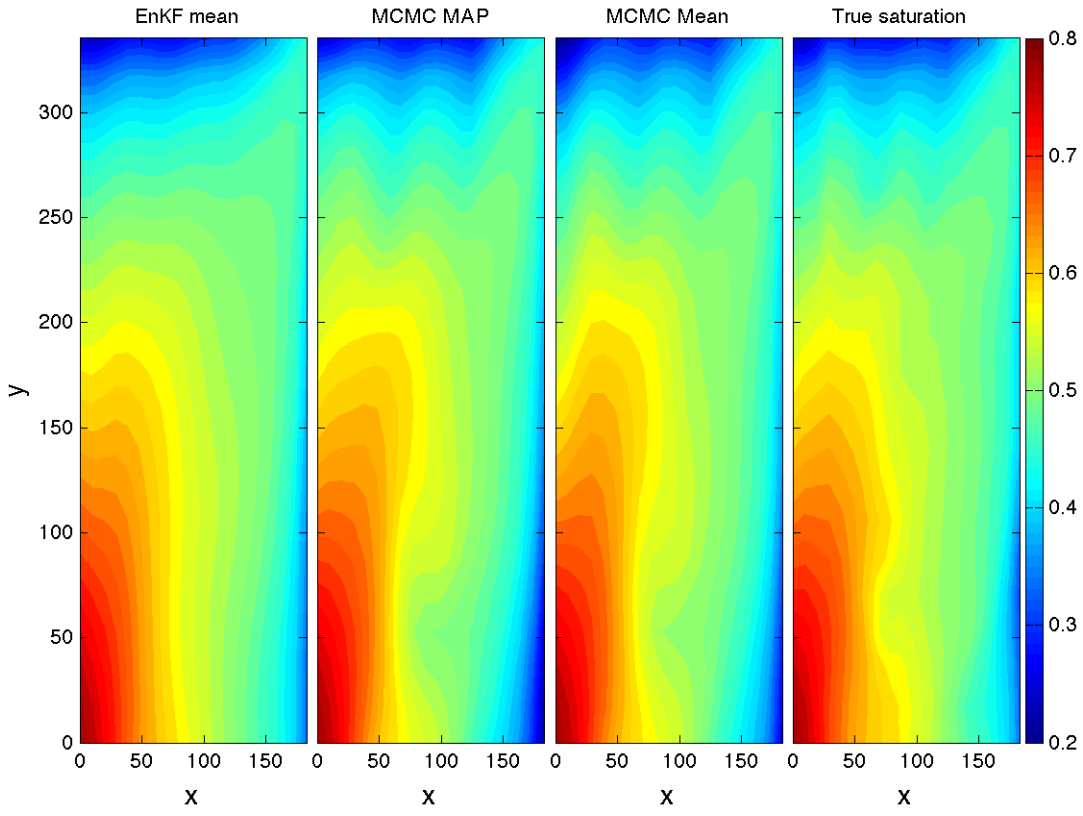


Figure 5.14: Comparison of MCMC and EnKF saturation estimates. The EnKF consisted of 500 ensemble members. The MCMC was performed with an initial proposal covariance of 0.1, and 10,000 samples were drawn.

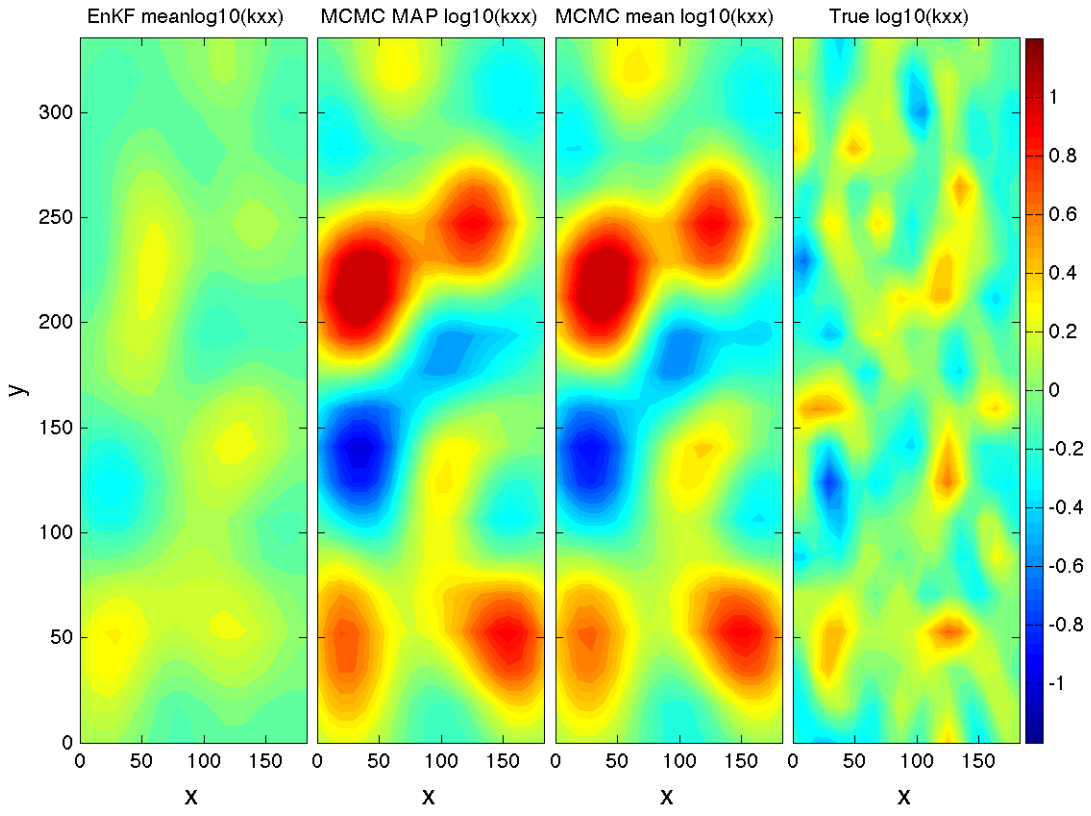


Figure 5.15: Comparison of MCMC and EnKF permeability estimates. The EnKF consisted of 500 ensemble members. The MCMC was performed with an initial proposal covariance of 0.3, and 10,000 samples were drawn.

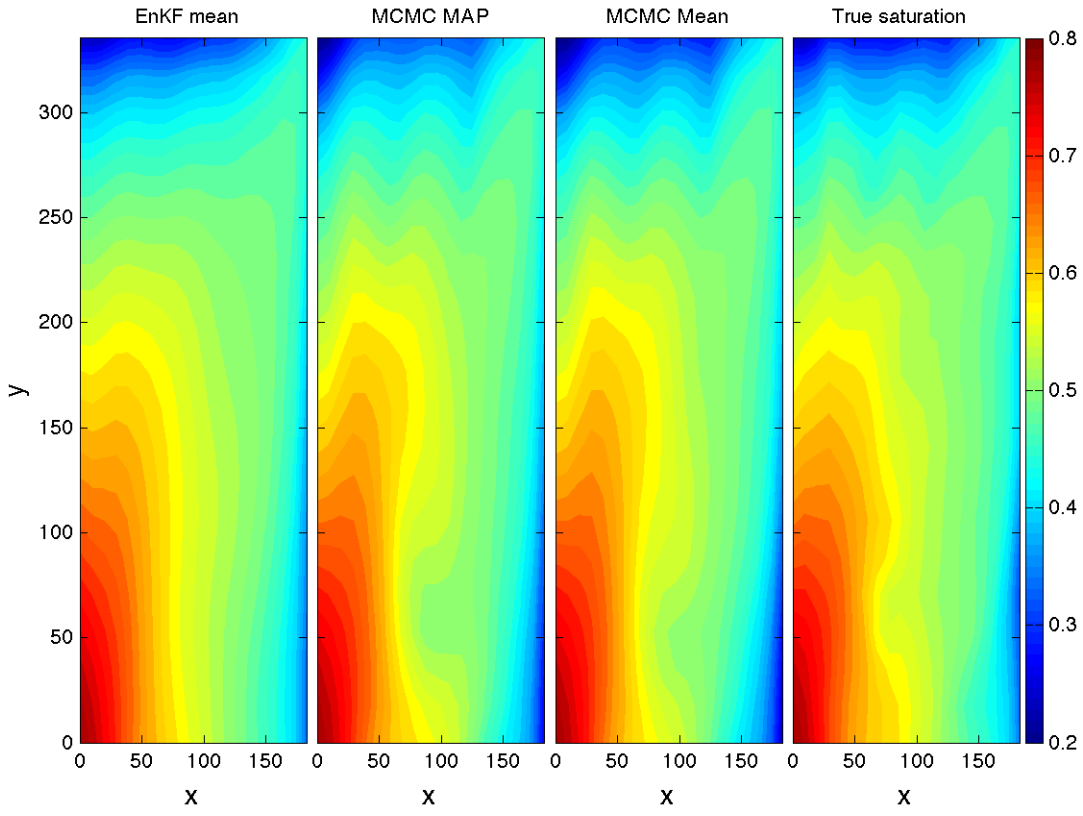


Figure 5.16: Comparison of MCMC and EnKF saturation estimates. The EnKF consisted of 500 ensemble members. The MCMC was performed with an initial proposal covariance of 0.3, and 10,000 samples were drawn.

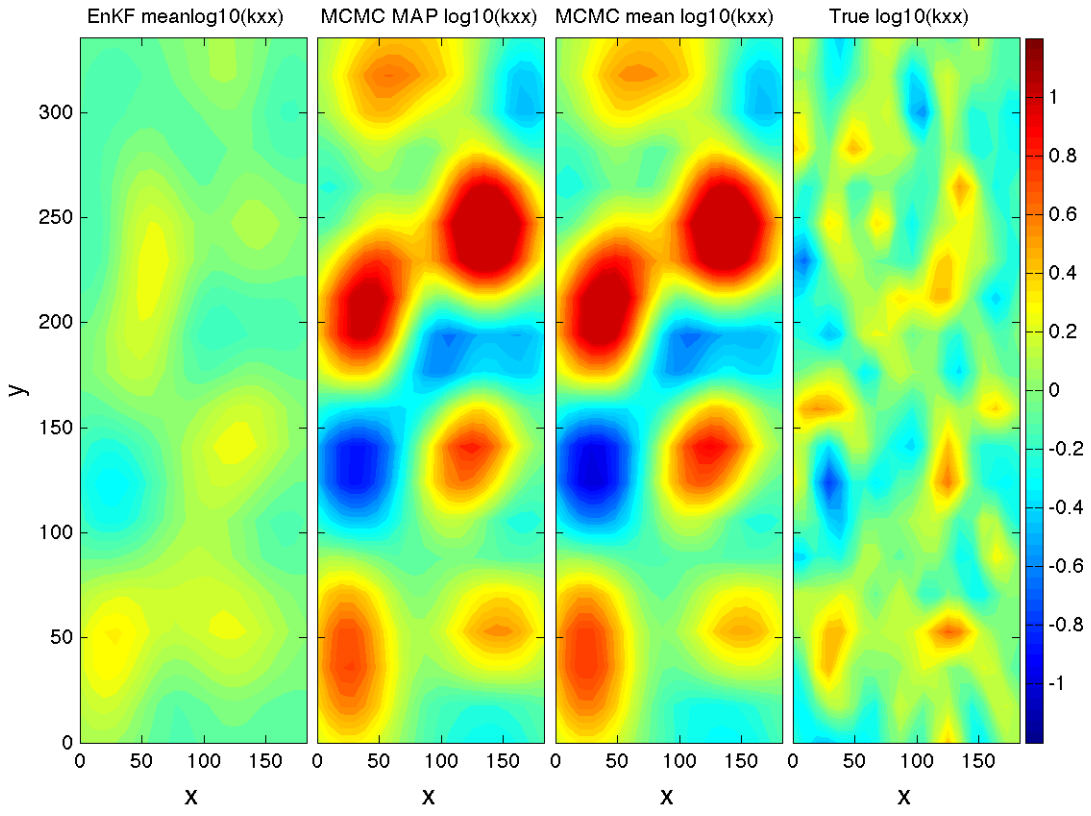


Figure 5.17: Comparison of MCMC and EnKF permeability estimates. The EnKF consisted of 500 ensemble members. The MCMC was performed with an initial proposal covariance of 0.3, and 30,000 samples were drawn.

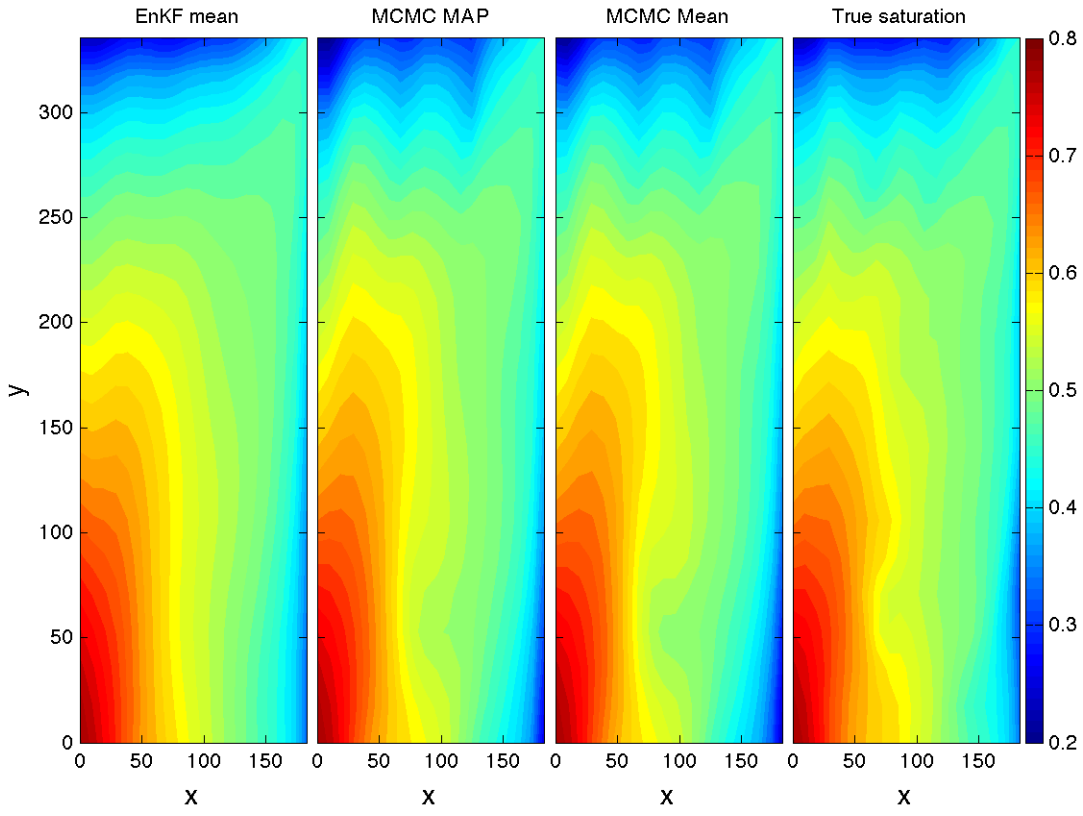


Figure 5.18: Comparison of MCMC and EnKF saturation estimates. The EnKF consisted of 500 ensemble members. The MCMC was performed with an initial proposal covariance of 0.3, and 30,000 samples were drawn.

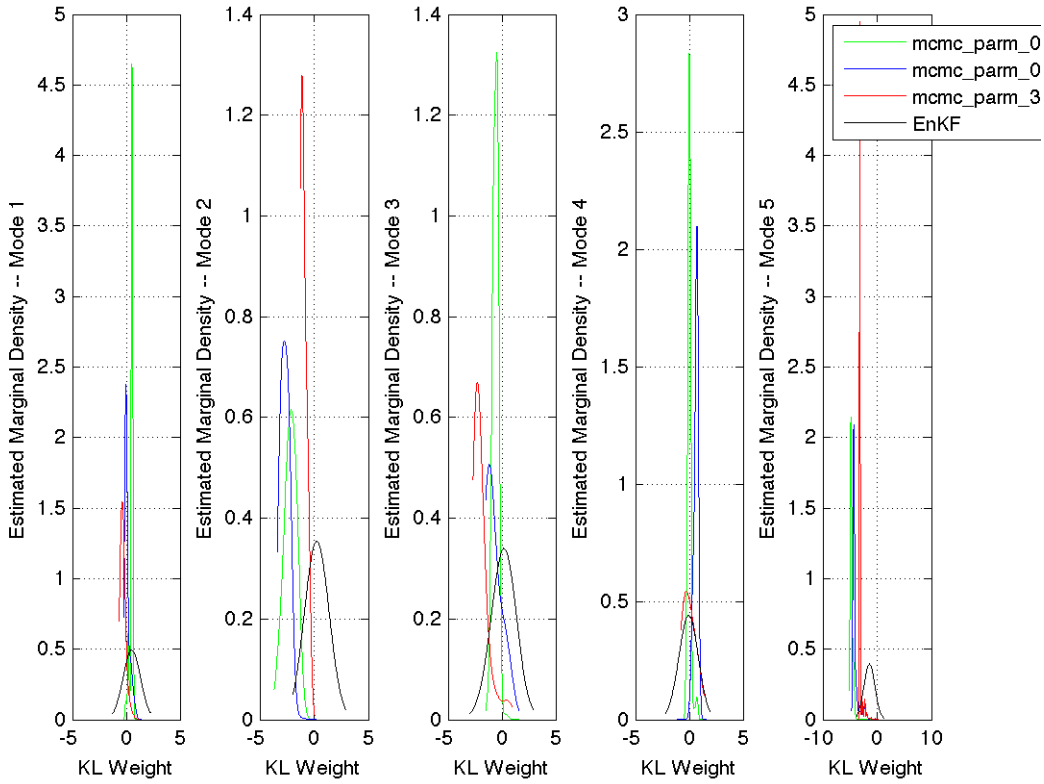


Figure 5.19: Comparison of MCMC and EnKF marginal distributions, for the three cases plotted in Figs. 5.13- 5.18.

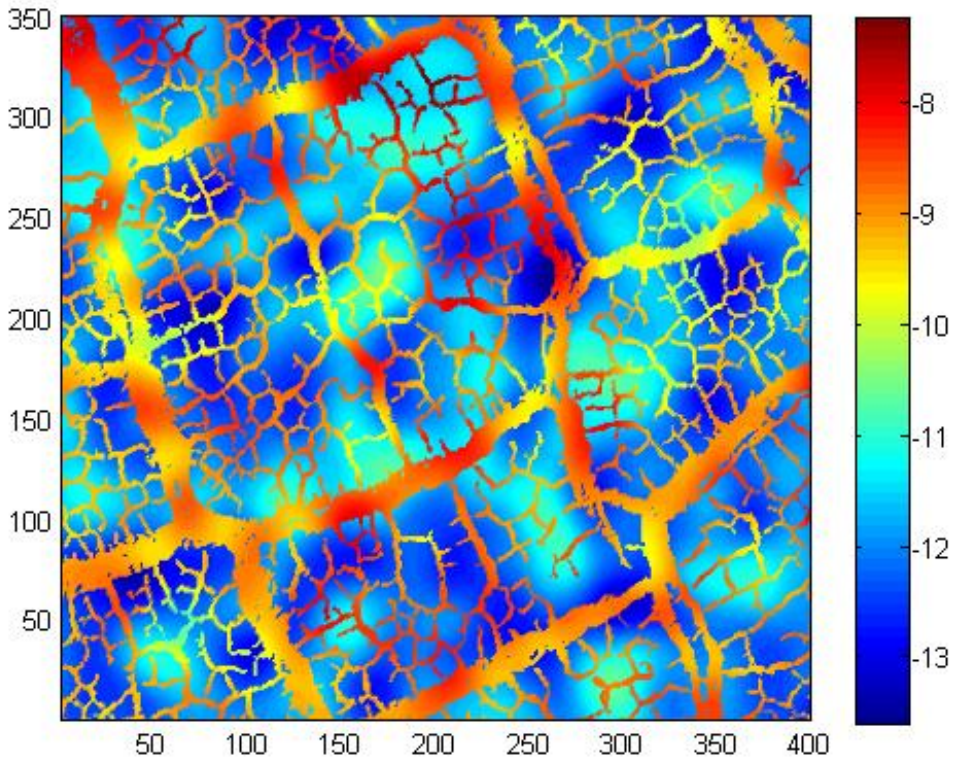


Figure 5.20: Multi-distribution dataset

5.10 Numerical Results: Multiple Distribution Example

In this section a complicated dataset is used to test the EnKF inversion capabilities (Fig. 5.20). Our original hypothesis was that the EnKF algorithms would not be able to invert for material properties of this class. However, as Figs. 5.21 and 5.22 show, the EnKF is able to reconstruct sufficient features of the permeability paths to produce an excellent water concentration prediction in comparison to the truth model. In an attempt to improve on the statistical characterization of this dataset, a hybrid method was implemented and is briefly discussed next.

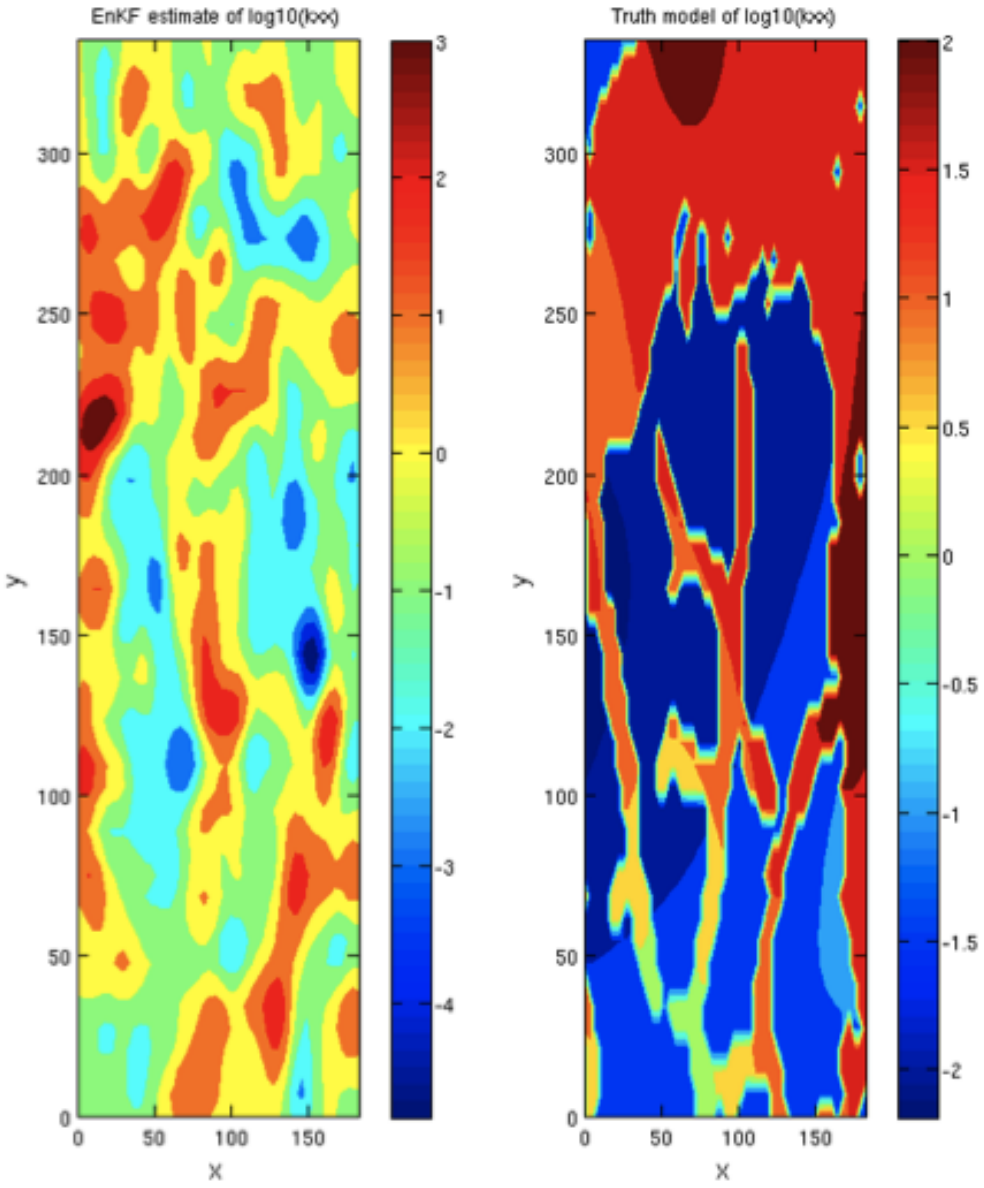


Figure 5.21: Permeability reconstruction on the left and the truth model on the right. The dots are the measurement locations.

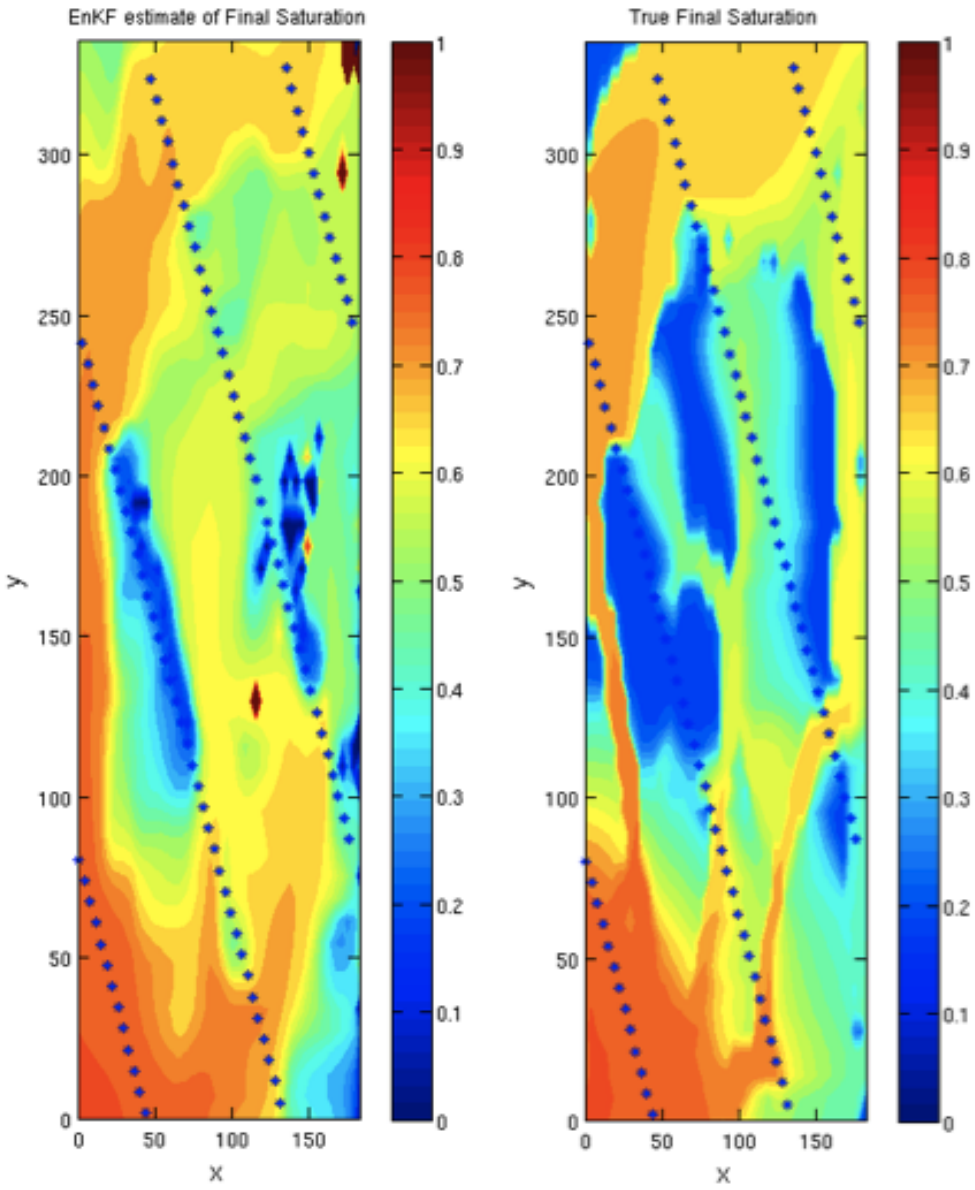


Figure 5.22: Saturation reconstruction on the left and the truth model on the right. The dots are the measurement locations.

5.11 Conclusions

Ensemble Kalman Filters (EnKF) are investigated to predict water saturations and reconstruct Karhunen and Loeve (KL) coefficient representation of permeability fields in water flooding numerical experiments. Traditionally, Kalman Filters provide computationally efficient mechanisms to solve large scale inversion problems but achieve this efficiency through Gaussianity, linear assumptions and adjoints. The EnKF approach avoids adjoints, addresses deviations from Gaussianity, and accommodates nonlinearities by direct calculation of the covariance from the ensembles. Sensitivity studies are performed to evaluate the performance of EnKF and compare to complete statistical solutions using Markov Chain Monte Carlo (MCMC) algorithms. Using a small dataset with 20 target inversion parameters (KL coefficients), the EnKF results demonstrate excellent water saturation prediction capabilities despite sparse observations and noisy data. The MCMC results show however a discrepancy in higher statistical moments. The EnKF is tested on a permeability dataset that is strongly bimodal and representative of a fracture network in less permeable matrix, capturing large and scale features of saturations. A hybrid approach is introduced to combine dynamic and static observations and shows that the inversion is not only accelerated but provides more accurate, higher statistical moments. The hybrid method consists of a MCMC inversion of static permeabilities to augment the prior information for the EnKF process, which then in turn performs a dynamic inversion of both permeabilities and water saturations.

Finally, we demonstrated a good reconstruction of saturations for a large and complicated dataset. This took 8 hours on a serial processor for 500 ensembles, which can be parallelized and theoretically computed in about one minute on a 500 core machine. MCMC would require about one million forward simulations to achieve appropriate convergence, which would require approximately 2 years of compute time on a serial processor.

Chapter 6

Application of EnKF to High-Plains Aquifer Ground Water Model

The High Plains aquifer (HPA) underlies an area of 111.4 million acres (174,000 square miles) in parts of eight States – Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, and Wyoming [160]. The term High Plains Aquifer is often used synonymously with the Ogallala Aquifer. The region overlying the HPA is one of the major agricultural regions in the world and roughly 10 percent of this area is irrigated farm land with the HPA serving as the source of the irrigation water. Water level declines began in parts of the HPA soon after the beginning of extensive groundwater irrigation in the 1950s. Water–level declines have exceeded 150 feet in some areas. The HPA is generally found at shallow depths and is a water-table (unconfined) aquifer over most of its extent. The Kansas Geological Survey (KGS) has constructed a complex MODFLOW model of Groundwater Management District No. 3 (GMD3) in a portion of the HPA in the southwest corner of Kansas. A full report on the MODFLOW model can be found in [1]. More than 15,000 lithologic logs are used to define the heterogeneous stratigraphy within this model. Eight hydraulic conductivity (K) values and five specific yield (S_y) values corresponding to different stratigraphic units within the well logs are calibrated using PEST [161]. While the GMD3 model spans the years 1944 to 2007, the calibration used in this research is based on the 18 years between 1990 and 2007. Both water level and streamflow measurements are used in the calibration. The effective K and S_y values are computed using the saturated thickness at each well. These values are used as conditioning data points to krig the effective K and S_y fields used in the MODFLOW model. With this model in place, it is possible to make projections into the future, create "what if" scenarios and overall increase the profitability of the water management effort in Kansas. To improve predictions of future aquifer sustainability, accurate estimates of aquifer parameters (K and S_y) are needed and uncertainty inherent in those estimates needs to be quantified.

Here, the Ensemble Kalman Filter (EnKF) is applied to the MODFLOW model of GMD3 to estimate K and S_y values. The hydraulic variables are updated based on observed streamflow and water levels. The GMD3 model is highly non-linear in that the predicted water

level is dependent on the estimated effective K at any location, which is a function of the predicted water level. While the EnKF is designed for use with non-linear models, the GMD3 model illustrates some challenges to this approach, primarily due to the interdependency between the calibrated hydraulic properties and the water level. The number of stratigraphic units that contribute to the effective hydraulic properties decreases with a decline in water level. Furthermore, the hydraulic properties are distributed throughout the model in a non-linear fashion. The large amount of structural information derived from the lithologic boreholes constrains the estimated K and S_y values. In this report, we demonstrate the use of EnKF with the GMD3 model using several cases. We find that the estimated K and S_y values often diverge from the calibrated model supplied by the KGS and become unstable through time using the EnKF.

6.1 Ensemble Kalman Filter Background

Details on the theory and algorithms for the Kalman filter and its extensions as used here are available in Sec. 5.5. A brief summary of the conceptual background of the Kalman filter and the EnKF is presented here for completeness. The Kalman filter technique for predicting system behavior was first introduced by Kalman [149]. Kalman filter techniques are predictor-corrector methods and use measured data to correct model predictions by tracking the error covariance matrix. Kalman filtering, however, is only applicable to linear systems. To address this limitation, linearization techniques are added to an extended Kalman filter algorithm to approximate error statistics in the Kalman filter framework [162]. The linearization techniques, however, are not successful in highly non-linear systems.

The EnKF approach was developed by Evensen [163] to predict behavior in highly non-linear systems. A thorough review of EnKF applications has been compiled by Evensen [164]. The EnKF uses an ensemble of model states (state variables) and model estimates (control variables) to predict error statistics forward in time. Initial representation of state variables is sampled from a probability distribution function, forming a state matrix. The state matrix is updated using the EnKF in an attempt to improve the model estimate as compared to physical observations. Each variable within the state matrix is updated based on an error statistic, called the Kalman Gain (KG), and the difference between predictions and physical observations. The KG is defined as follows:

$$KG = COVA(State, Control)/VAR(Control), \tag{6.1}$$

where $COVA$ is the covariance function, VAR is the variance function, $State$ is the state variable, and $Control$ is the control variable. The KG is used to update the state matrix using observational data with the following equation:

$$State_u = State_p + KG * (Control - Observation), \tag{6.2}$$

where subscript u denotes the updated state matrix, p denotes the prior state matrix, and $Observation$ is the observational data at a specified time. The value of the KG serves as

a weighting scheme to place more or less emphasis on the prior state values versus the observation data in updating the state variables. More than one observation/control data pair can be used in the EnKF by adding additional KG terms as follows:

$$State_u = State_p + \sum KG_n * (Control_n - Observation_n), \tag{6.3}$$

where n is the number of control/observational data pairs used in the EnKF. Note that one KG is calculated for each control data type used. In general, one update per time-step is carried out. Alternatively, an iterative procedure can be applied to ensure that updated state variables approximate control data within a specified tolerance. Using an iterative procedure, updates are carried out until the difference between control data and observational data is less than some specified tolerance. Once this error tolerance is met, the model is advanced to the next time-step.

6.2 Modeling Approach

The GMD3 model covers an area 100 by 150 miles and is discretized into 1 mile square grids. The model uses a 1 year time step. Several no-flow regions are defined in the model, associated with locations where bedrock is observed at the surface. The model includes approximately 5000 pumping wells distributed throughout the model domain and spatially varying recharge associated with regional precipitation. While recharge was treated as a calibration variable in the GMD3 PEST calibration done by the KGS, recharge is treated as a known parameter in this study, with time and space varying values taken from the PEST calibration. This research focuses on the 18 years between 1990 and 2007.

Stream flow and water level measurements are used to calibrate the model using the EnKF. Streamflow observations are recorded at 7 locations: Syracuse, Kendall, Deerfield, Garden City, Dodge City, Cimarron R Forgan OK, and Crooked Cr Englewood. These locations are located along the Arkansas and Cimarron rivers. Water level elevation is recorded at 122 monitoring wells. These locations are spread throughout the model domain (Fig. 6.1).

Pumping wells are included in the model to simulate the discharge associated with ground water rights (See Fig. 6.2). There are 10,367 individual water rights within the model domain. During the 18 year period, an average of 5024 wells were pumping each year. Individual pumping rates can be as high as $3.8 * 10^5$ ft^3/day

The model contains time-varying specific head boundary conditions around part of the bedrock and along the northern and southern perimeter of the area. Along the eastern and western proximity there are time varying flux boundaries. The model contains several inactive regions associated with locations where bedrock is observed at the ground surface. Recharge of the HPA is included in the MODFLOW model and includes recharge originating from precipitation, irrigation, and diverted rivers. A distribution of precipitation was applied to cells based on yearly data. The KGS implements a recharge algorithm that takes into account the different types of recharge during the calibration process. Every cell of the model area is assigned a recharge value and entered into MODFLOW (See Fig. 6.3).

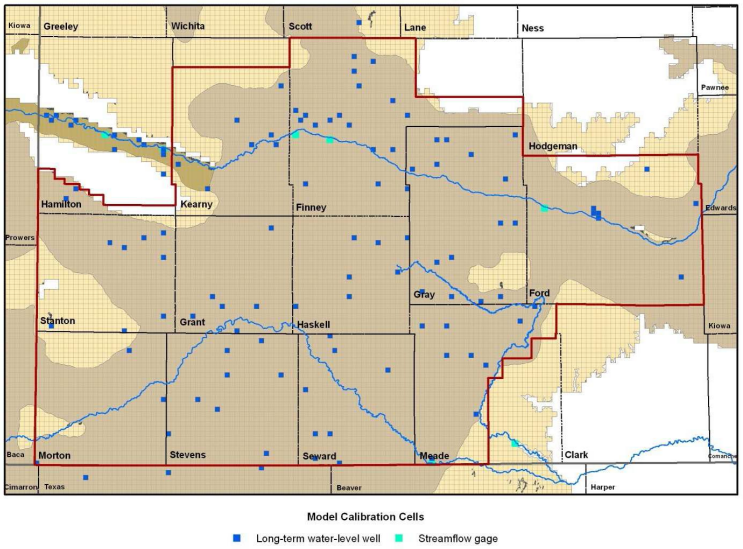


Figure 6.1: Locations of the 122 long-term water-level wells (blue cells) and 7 streamflow gages

The hydrostratigraphy of the HPA in the model domain is complex with approximately 60 different units identified from well log data. The Practical Saturated Thickness Plus (PST+) program was developed by the KGS to create heterogeneous fields that reflect well log data. PST+ is used to build hydraulic conductivity and specific storage fields based on 15,715 well logs in the model domain. Each borehole contains data of the different layers of material at their corresponding depths in the hole. The lithology of each well log is segregated into 62 categories (called synonymies). These synonymies are then aggregated into 8 K values and 5 S_y values based on similarities in their estimated K or S_y value (Fig. 6.4). An effective K and S_y value are assigned for each borehole location based on the water level at that location and the weighted average of the K and S_y values below the water table. These locations are used as conditioning points in the 2D kriging program (KB2D) [165]. A USGS supplemental field is used to fill in the outer region of the model domain for both K and S_y. The GMD3 model and the large supporting data set provides an excellent test case for examining the impact of stratigraphic conditioning data on the resulting estimates of hydraulic properties.

The GMD3 MODFLOW model was calibrated by the KGS using PEST. PEST provides a deterministic, gradient-based inversion capability. The initial and calibrated values for K and S_y are listed in Table 6.1. Overall, the calibrated values produce a good fit to observed streamflow and water levels. The Root Mean Squared Error (RMSE) is calculated for each of the 7 stream locations and 122 observation wells. RMSE is calculated using all 18 years for each location. The modeled hydrographs for the 7 stream gages (Fig. 6.5) adequately

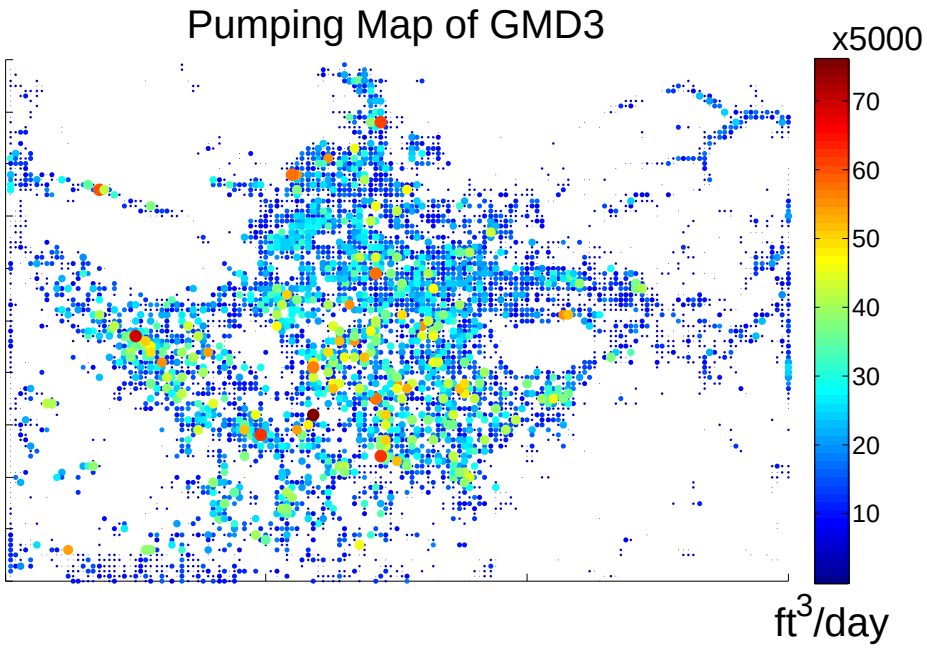


Figure 6.2: Pumping rates for the first year (1990). The magnitude of the circles radius is proportional to the amount of pumping at that particular location.

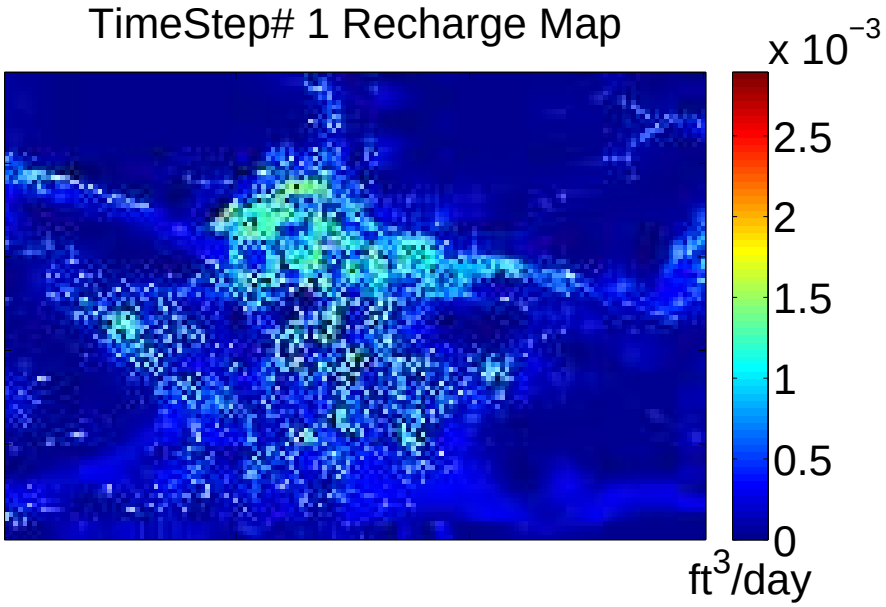


Figure 6.3: Map of model recharge for the first year in the model (1990).

predicts the observations in all but the "Cimarron R Forgan OK" station. The RMSE for each hydrograph is listed in 6.2. Water level elevations in the wells are also adequately predicted using the calibrated values in a majority of the monitoring locations. Over 50 percent of the 122 wells have a RMSE of less than 12 ft. The monitoring wells with the 10 highest RMSE values are illustrated in Fig. 6.6. The corresponding RMSE is also listed in Table 6.2. The locations and the magnitudes of the RMSE values are shown within the model domain in Fig. 6.5.

Table 6.1: Initial and calibrated (PEST) K and S_y values from the GMD3 model

Parameter	Initial Value (ft/day)	Calibrated Value (ft/day)	Parameter	Initial Value (-)	Calibrated Value (-)
K1	0.0002	0.000037	SY1	0.03	0.05
K2	0.02	0.0001	SY2	0.05	0.03
K3	0.2	4.44	SY3	0.20	0.08
K4	2	10.05	SY4	0.22	0.08
K5	20	1.00	SY5	0.25	0.05
K6	100	47.05			
K7	200	236.47			
K8	300	0.1			

To use the GMD3 MODFLOW model within the EnKF, the state variables, state matrix, observational data, and related control variables must be defined. The state variables are

Synonymy	K	SY	Synonymy	K	SY	Synonymy	K	SY
sh	0.0002	0.03	sc	0.2	0.05	fsnd	20	0.22
c	0.0002	0.03	fds	0.2	0.05	fmgsnd	20	0.22
coal	0.0002	0.03	fmds	0.2	0.05	fmsnd	20	0.22
br	0.0002	0.03	fcrsds	0.2	0.05	snd	100	0.22
rb	0.0002	0.03	ds	0.2	0.05	fcrssnd	100	0.22
r	0.0002	0.03	mds	0.2	0.05	msnd	100	0.22
sst	0.0002	0.03	gc	0.2	0.05	mcrssnd	100	0.22
ca	0.02	0.05	mcrsds	0.2	0.05	cg	100	0.22
o	0.02	0.05	crsds	0.2	0.05	crssnd	100	0.25
ts	0.02	0.05	cesd-cg	2	0.2	sg	100	0.25
fs	0.02	0.05	fss	2	0.2	fsdg	200	0.25
fsc	0.02	0.05	fmss	2	0.2	fmsdg	200	0.25
fmsc	0.02	0.05	ss	2	0.2	msdg	200	0.25
m	0.02	0.05	mss	2	0.2	sdg	200	0.25
msc	0.02	0.05	fcrsss	2	0.2	fcrssdg	200	0.25
s	0.02	0.05	mcrsss	2	0.2	mcrssdg	200	0.25
crssc	0.02	0.05	crsss	2	0.2	crssdg	200	0.25
fcrssc	0.02	0.05	u	2	0.2	fg	300	0.25
mcrssc	0.02	0.05				fmg	300	0.25
						fcrsg	300	0.25
						fcrssg	300	0.25
						g	300	0.25
						mg	300	0.25
						mcrsg	300	0.25
						crsg	300	0.25

Figure 6.4: Practical Saturated Thickness Plus (PST+) program initial K (ft/day) and S_y values for each PST+ synonymy lithology code (Table 4 in [1]). The colors denote the groupings of the synonymies into 8 K and 5 S_y values.

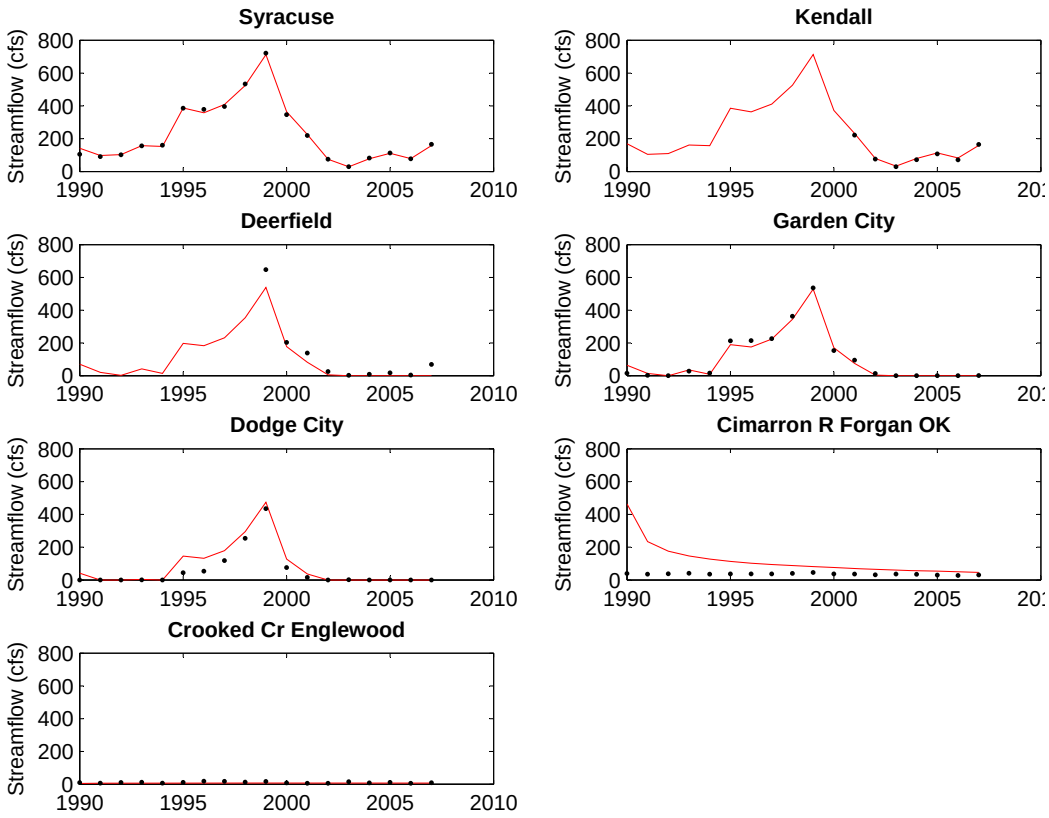


Figure 6.5: Measured (black dots) and modeled (red line) streamflow from the initial calibration of the GMD3 model with PEST.

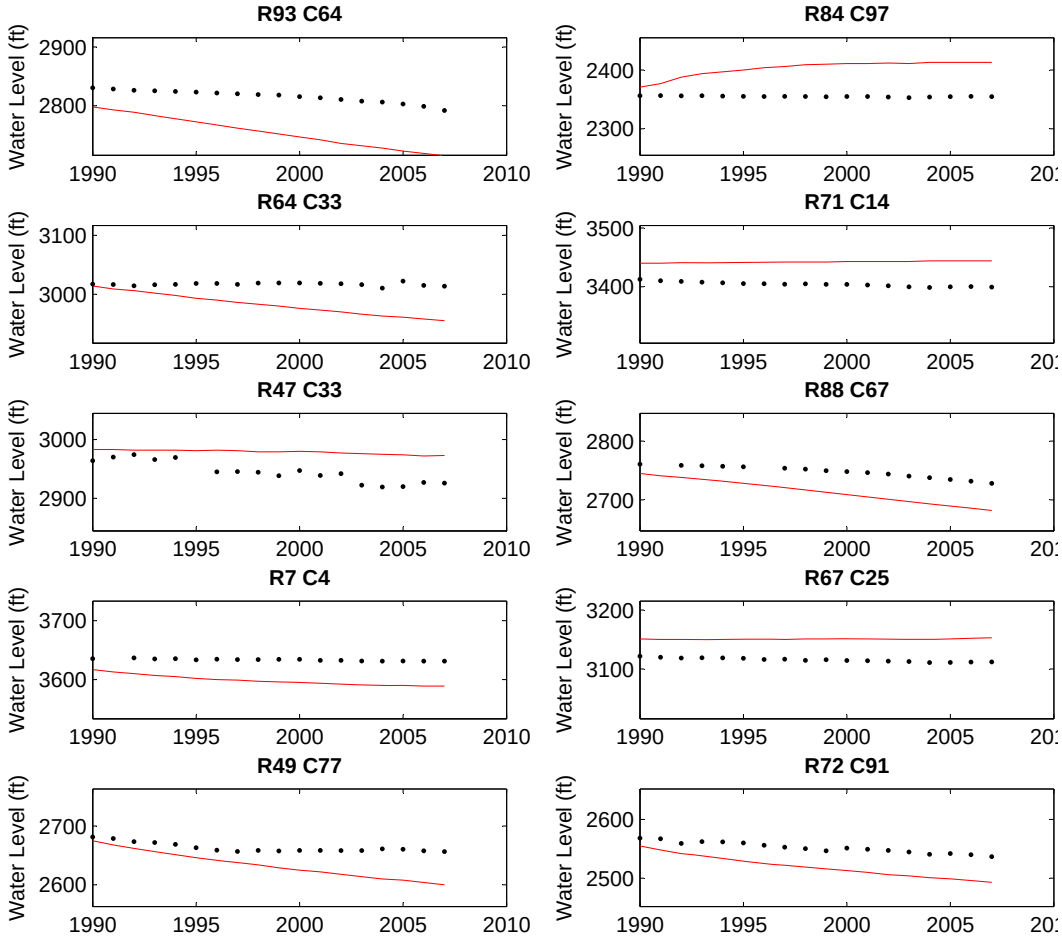


Figure 6.6: Measured (black dots) and modeled (red line) water level from the initial calibration of the GMD3 model with PEST. The 10 wells with the highest RMSE over 18 years are shown here.

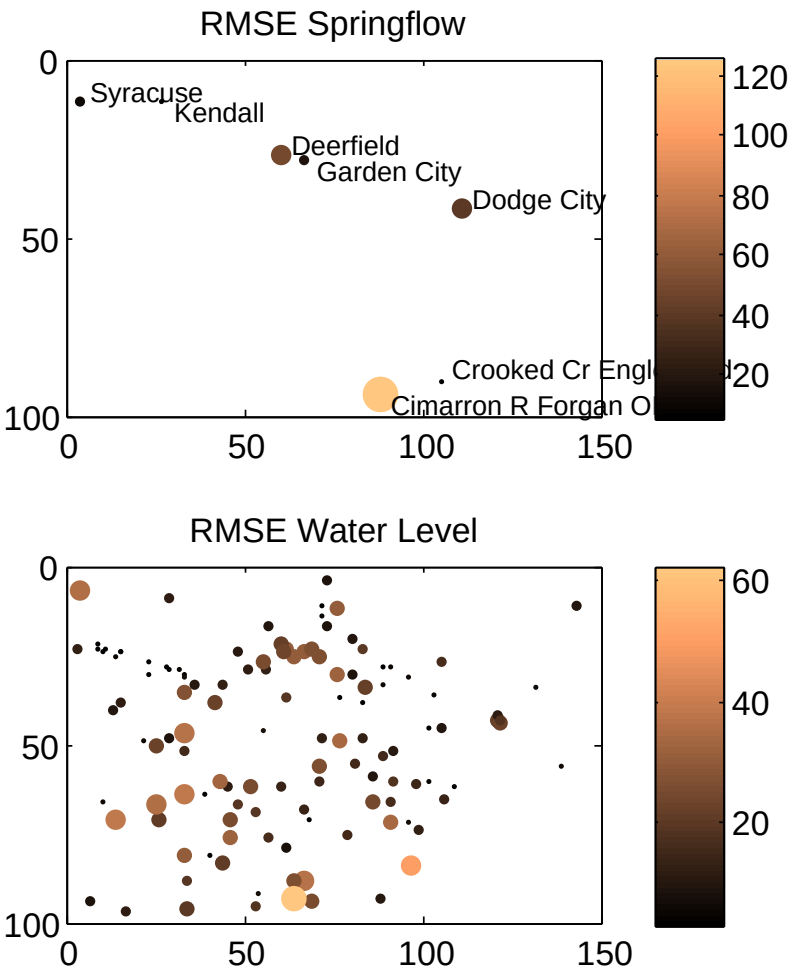


Figure 6.7: Spatial distribution of RMSE for streamflow and water level using the GMD3 calibrated model. The color scale is in units of ft^3/day (upper image) and ft (lower image).

Table 6.2: RMSE for all 7 streamflow observations and 10 highest RMSE for water level observations using the GMD3 calibrated model. The average RMSE values for the 7 streamflows and all 122 water levels are shown at the bottom.

Gage Station	RMSE (cfs)	Observation Well	RMSE (ft)
Syracuse	12.1	R93 C64	62.6
Kendall	7.8	R84 C97	49.7
Deerfield	48.3	R64 C33	38.9
Garden City	17.9	R71 C14	38.4
Dodge City	39.7	R47 C33	37.3
Cimarron R Forgan OK	125.3	R88 C67	36.6
Crooked Cr Englewood	5.1	R7 C4	36.2
		R67 C25	35.6
		R49 C77	33.8
		R72 C91	33.8
Average (7 Streams)	36.6	Average (122 Wells)	15.8

the 8 K and 5 S_y values that are used as input to the PST+ program. The state matrix is an ensemble of realizations for the state variables. An ensemble of state variables is created using a log normal distribution for each state variable and 50 realizations are included in the ensemble. For K, the mean of the ensemble is log10(0.1 ft/day) and standard deviation is 2. For S_y, the mean of the ensemble is log10(0.05) and standard deviation is 0.3. The distributions are truncated with respect to the reasonable limits on each parameter (3x10^{-}06 ft/day to 3000 ft/day for K and 0.001 to 0.25 for Sy). Two types of observational data are used: streamflow and water level. Depending on the type of observation data selected for the EnKF, either modeled streamflow and/or modeled water level is defined as the control variables.

The EnKF is used in conjunction with the GMD3 MODFLOW model and PST+ program by updating the state variables (8 K and 5 S_y values) after each every time step. A Kalman Gain (KG) is calculated for each state variable as a function of the covariance between the state variable (K or S_y) and control variable (streamflow or water level) (see Eq. 6.1). Positive correlation between the state and control variable results in a positive KG.

EnKF updates are based on 2 methods: a single update per time step, and an iterative process. With the single update method, the EnKF updates state variables once per timestep. With the iterative method, the EnKF continues to update the state variables until the control variable (modeled stream flow or water level) is within a specified tolerance of the observational data (20 cfs for streamflow or 10 feet for water level). The maximum number of iterations per time step is set to 10. Once convergence or the maximum number of time steps is met, the EnKF moves to the next time-step. Table 6.2 is used to measure how well the EnKF updates the state of the model; additionally, changes in the state matrix are tracked to see if the parameters are converging.

6.3 Results

Before applying the EnKF to the GMD3 model, we run an ensemble of state variables through the model to analyze the sensitivity of the model outputs to each of the 13 state variables. This analysis is done using the PST+ program with all 15,715 lithology boreholes as input. Using 50 realizations of independently drawn state variables from the state matrix, we note that several streamflow and water level observations change very little with changes to the K and S_y variables. Figure 6.8 and Figure 6.9 show measured and modeled streamflow and 10 of the 122 measured and modeled water levels using the 50 realizations. In general, the calibrated GMD3 model is in good agreement with observed streamflow and water levels. Using the calibrated K and S_y values, the GMD3 model has very low RMSE for streamflow at Syracuse, Kendall, and Crooked Cr Englewood. The RMSE for modeled water levels are generally below 12 ft. The streamflow at Syracuse, Kendall, and Crooked Cr Englewood, the same locations where the error is very low, are not sensitive to changes in K and S_y. There are also 19 monitoring well locations that are not very sensitive to changes in K and S_y. The EnKF updates state variables based on variability in the model outcome. When a specific observation is not sensitive to changes in the state, the covariance and variance go to zero. When the variance is small, but not zero, the KG can become very large. Ideally, state variables are not updated according to observations that are not sensitive to changes in the model. For this reason, we do not include the 3 streamflow observations (Syracuse, Kendall, Crooked Cr Englewood) and 20 water level observations (R4 C73, R7 C4, R9 C29, R11 C143, R23 C3, R23 C9, R23 C11, R23 C83, R24 C15, R24 C15, R25 C14, R27 C23, R27 C105, R29 C29, R29 C32, R30 C33, R34 C132, R40 C13, R52 C92, R97 C17) in the updating. Based on the 50 realizations used in this sensitivity analysis, the standard deviation in streamflow and water level at these locations is less than 10 cfs or 10 ft, respectively. This calculation is averaged across the 18 year simulation.

The K and S_y fields created by the PST+ program are highly conditioned. The heterogeneous fields are kriged using 15,715 boreholes. The PST+ program can be revised to accept any number of well logs that constrain heterogeneity. Before using the EnKF, we reran the sensitivity analysis above with less conditioning points. Here, we limit the number of stratigraphy boreholes to only 100 boreholes randomly selected from the 15,715 boreholes. This represents a reduction of over 99 percent in the conditioning data with just 0.6 percent of the original conditioning data remaining. Figure 6.10 illustrates the location of these selected boreholes. Figure 6.11 shows the difference in PST+ when using all 15,715 boreholes and the 100 selected locations. As expected, there is a drastic difference in the K and S_y fields estimated with 15,715 and 100 boreholes. The high frequency variations in the state variables represented with 15,715 boreholes are smoothed out into low-frequency variations when only using 100 boreholes. Suprisingly, using only 100 conditioning boreholes, streamflow and water levels change very little as compared to results using the original PST+ program. This analysis suggests that streamflow and water levels are not highly dependent on the underlying geologic structure. Streamflow and water level variability increases slightly, but the overall behavior remains the same. Initial EnKF results are run using the PST+ program with 100 boreholes.

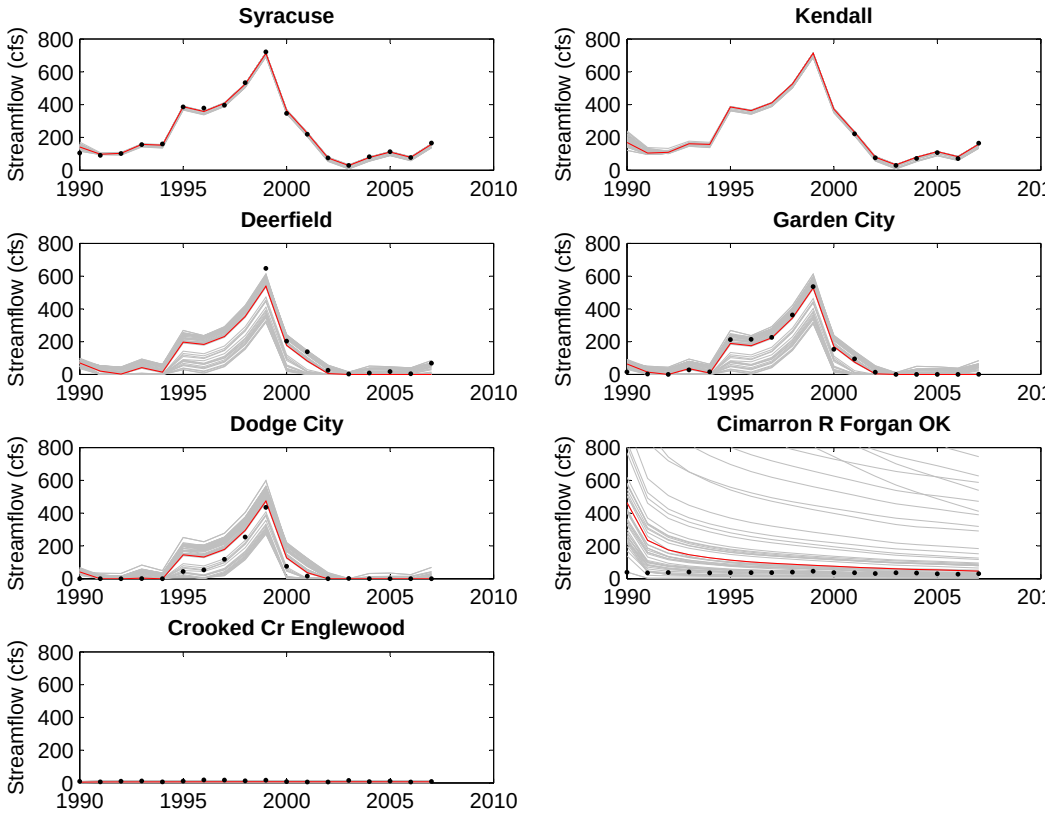


Figure 6.8: Sensitivity analysis of streamflow using 50 realizations of the K and S_y ensemble. Measured (black dots), GMD3 calibrated model (red line), and results from 50 realizations (gray lines).

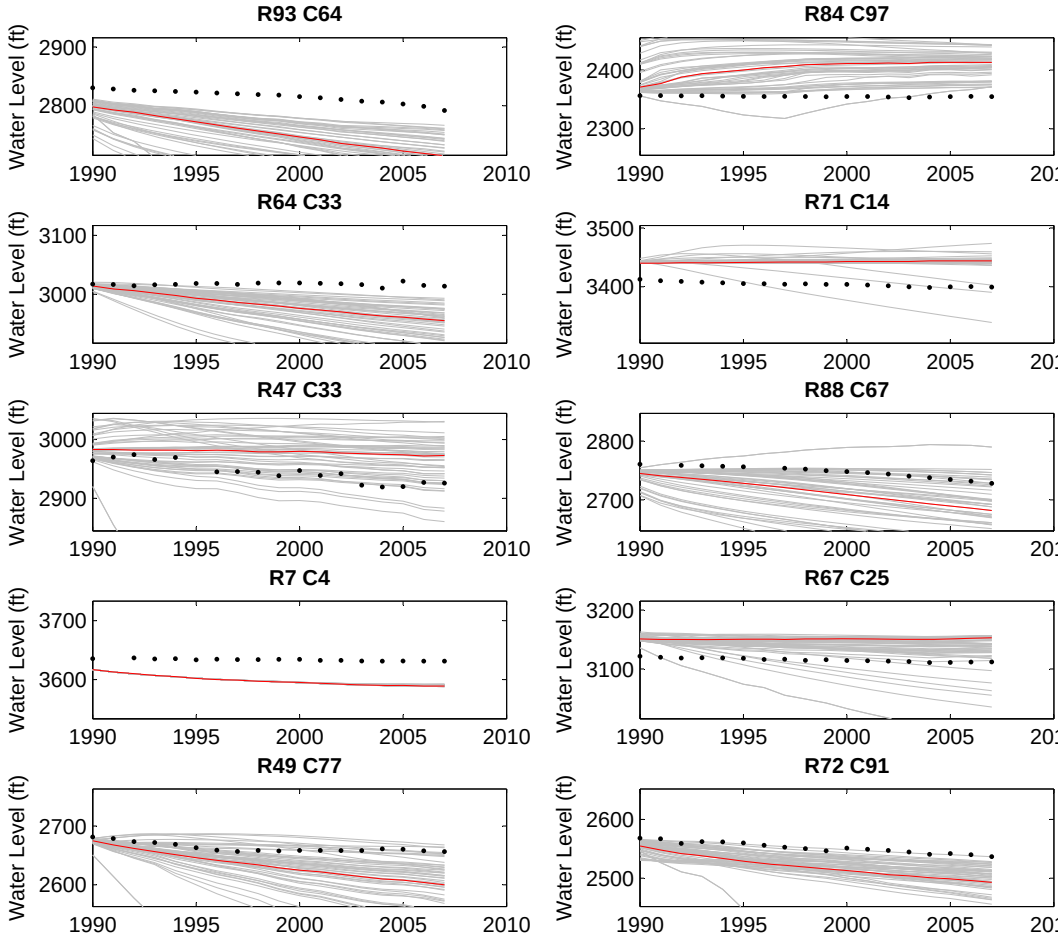


Figure 6.9: Sensitivity analysis of water level using 50 realizations of the K and Sy ensemble. Y-axis scale for water level is 200 feet for each plot. Measured (black dots), GMD3 calibrated model (red line), and results from 50 realizations (gray lines).

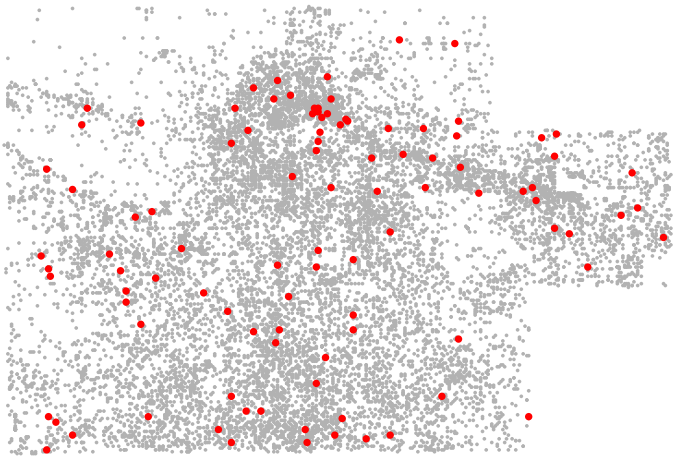


Figure 6.10: 100 randomly selected boreholes (red) superimposed on the original set of 15715 boreholes (gray) used in the PST program.

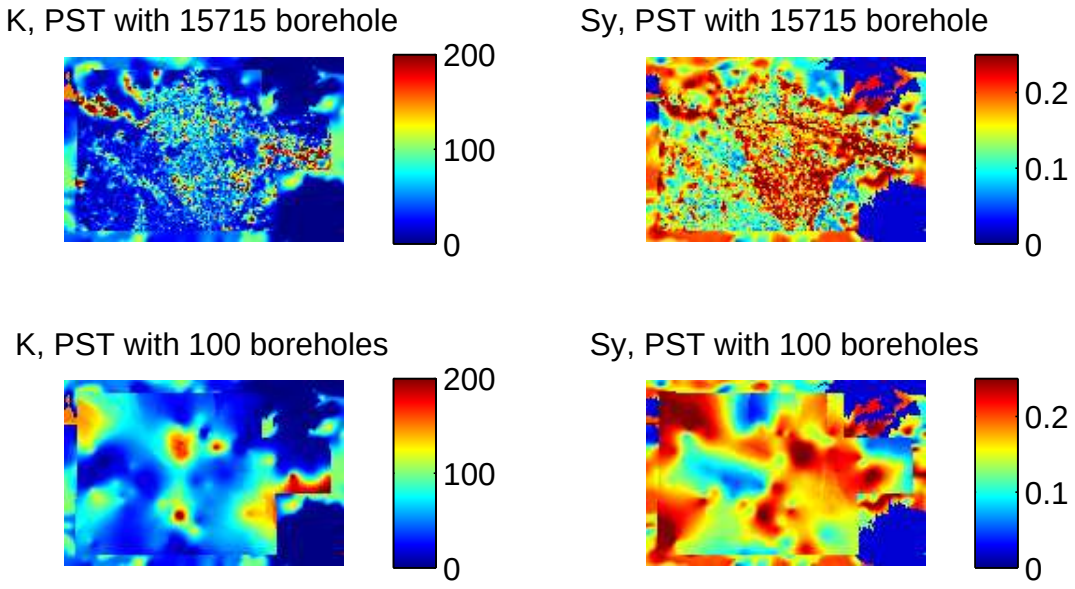


Figure 6.11: K and S_y field using 15,715 boreholes (top) and 100 randomly selected boreholes (bottom).

The EnKF can update state variables (K and S_y) based on streamflow or water level observations. Based on the sensitivity analysis, we select 4 of the 7 streams and 102 of the 122 monitoring wells that are influenced by changes to the hydraulic properties. Using a variety of control data types and single or iterative update, 6 cases are tested. The cases are:

1. *Single SF*. Single update, control data = streamflow at 4 locations.
2. *Single H*. Single update, control data = water level at 102 locations.
3. *Single SFH*. Single update, control data = streamflow at 4 locations and water level at 102 locations.
4. *Iterative SF*. Iterative update, control data = streamflow at 4 locations, maximum iterations = 10, convergence is met if average difference between model and observation is less than 20 cfs.
5. *Iterative H*. Iterative update, control data = water level at 102 locations, maximum iterations = 10, convergence is met if average difference between model and observation is less than 10 ft.
6. *Iterative SFH*. Iterative update, control data = streamflow at 4 locations and water level at 102 locations, maximum iterations = 10, convergence is met if average difference between model and observation is less than 20 cfs and 10 ft.

Table 6.3 summarizes the results from the 6 cases. In general, the EnKF does not improve the prediction of streamflow and head across all observation points. There are some cases where the EnKF improves the estimate at one or more points, but this tends to increase the error elsewhere. Using the EnKF on a large number of observation points tends to dilute the contribution from any one point. With the single update, the ensemble distribution does not narrow. While the outer limits of K and S_y migrate toward the mean, the overall ensemble is relatively unchanged. When we allow the EnKF to iterate on a single time step until the convergence is met, the ensemble starts to converge to a single value. However, the large number of observation points show that when the model improves the fit at one point, the error tends to increases elsewhere.

Results from case 4 (Iterative SF) are illustrated in Figures 6.12, 6.13 and 6.14. These figures show the streamflow and water level predictions along with changes to the K and S_y state matrix. Streamflow improves at several locations; however the water levels exaggerate the decline measured at observation wells. The K and S_y values converge steadily between 1990 and 2000. In some cases, the value is close to the calibrated value supplied by KGS. After the year 2000, the K values in particular change erratically. The feedback between water level and effective K and S_y is negatively impacting the updated ensemble. After some point, the declining water level decreases the number of stratigraphic units that influence model outcomes. The EnKF was also run using an average water level in the PST+ program. This fixes the number of stratigraphic units that contribute to the effective

heterogeneous fields through time. However, this change only reduces some of the non-linearity in the model. The effective values are still distributed in the model domain in a non-linear fashion. Results using a constant water level estimate in the PST+ program still produce errors in the range listed in Table 6.3.

6.4 Discussion

A previously calibrated groundwater model of the HPA was examined here to determine if parameter estimation using EnKF could improve on the calibration of the model to the observed data and/or provide additional insight into the variability of the estimated parameters. This application provides a contrast to the model examined in Chapter 5 in that there is a large database defining the spatial structure of the various hydro-stratigraphic units in the GMD3 model and relatively less information on the observed state variable, here groundwater levels, as compared to the more plentiful saturation observations in the Chapter 5 study.

Application of the EnKF to the GMD3 model showed that the density of the borehole data defining the hydrostratigraphy was such that it exerts a strong control on the sensitivity of observations to estimated values of K and S_y. For some locations, the observations are completely insensitive to the estimated parameters, and this independence causes degenerate covariance calculations within the EnKF framework. Removal of > 99 percent of the boreholes from the data set allowed the EnKF to estimate K and S_y, but the ensemble fit to the observed data did not converge to an accurate and stable state. It may be that more sophisticated application of the EnKF, including a means of localizing parameter changes with respect to particular observations would improve the results, but that approach is beyond the scope of this study.

The lithologic based heterogeneity and unconfined conditions of the GMD3 model results in unique non-linear conditions that create instability in the EnKF approach. Several cases were tested using the EnKF with the GMD3 model and in all cases the estimated hydraulic properties did not improve the prediction of streamflow and water level. While it was possible to remove some of the non-linearity by stabilizing the water level used in the PST+ program, the local K and S_y values are still distributed across the model. This non-linearity makes the relationship between any one parameter and the streamflow at a specific location, for example, difficult to track. Alternatively, the parameter estimation process could be modified by by removing the stratigraphic information (using no lithology information) and estimate effective K and S_y values at a number of locations throughout the model domain. Here, instead of estimating a limited number of global parameter values that are then spatially distributed according the lithologic framework, a larger number of local parameter values would be estimated. These local values would then be spatially distributed throughout the domain using kriging.

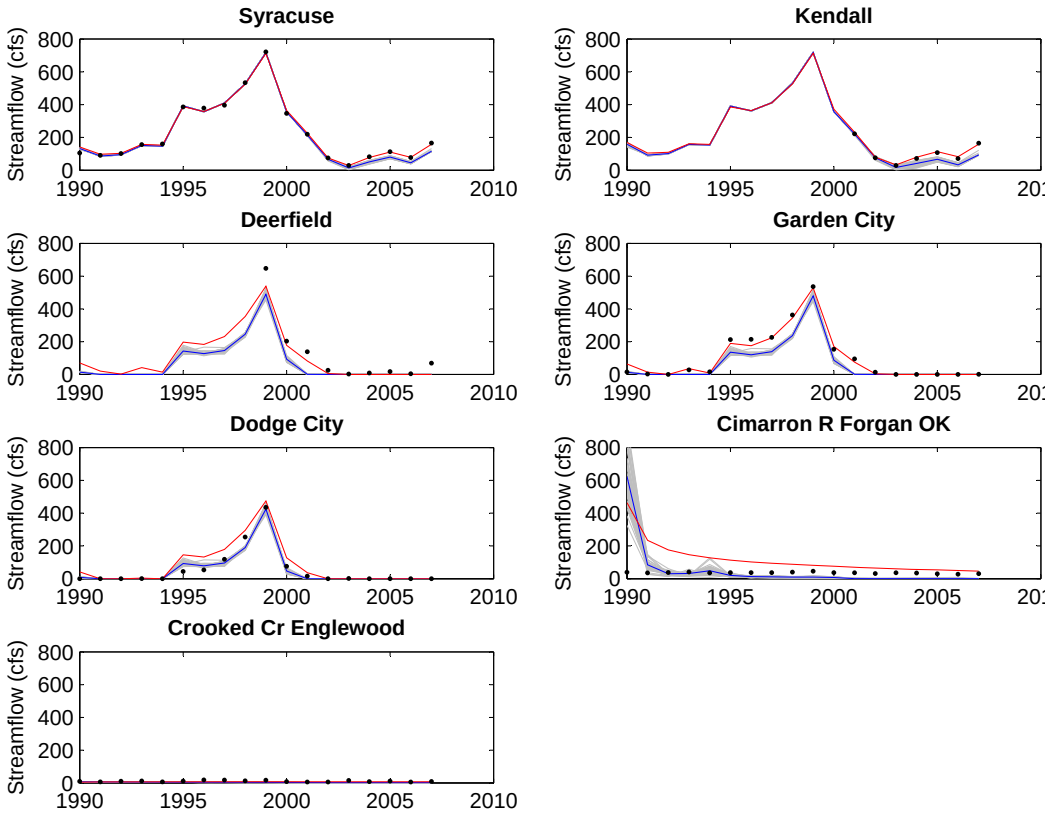


Figure 6.12: Variability in streamflow from Case 4 (Iterative SF). Measured (black dots), GMD3 calibrated model (red line), results from EnKF (gray lines), and mean streamflow from EnKF (blue line).

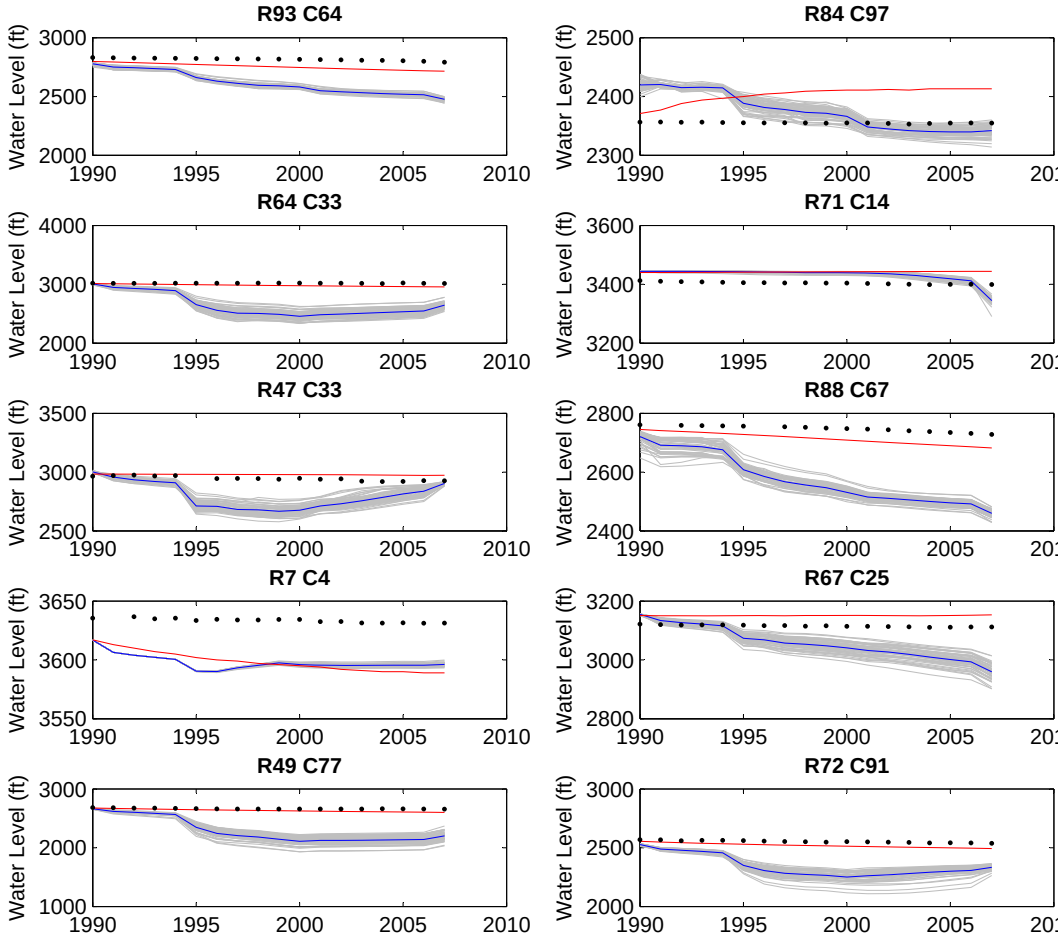


Figure 6.13: Variability in water level from Case 4 (Iterative SF). Y axis scale varies. Measured (black dots), GMD3 calibrated model (red line), results from EnKF (gray lines), and mean streamflow from EnKF (blue line).

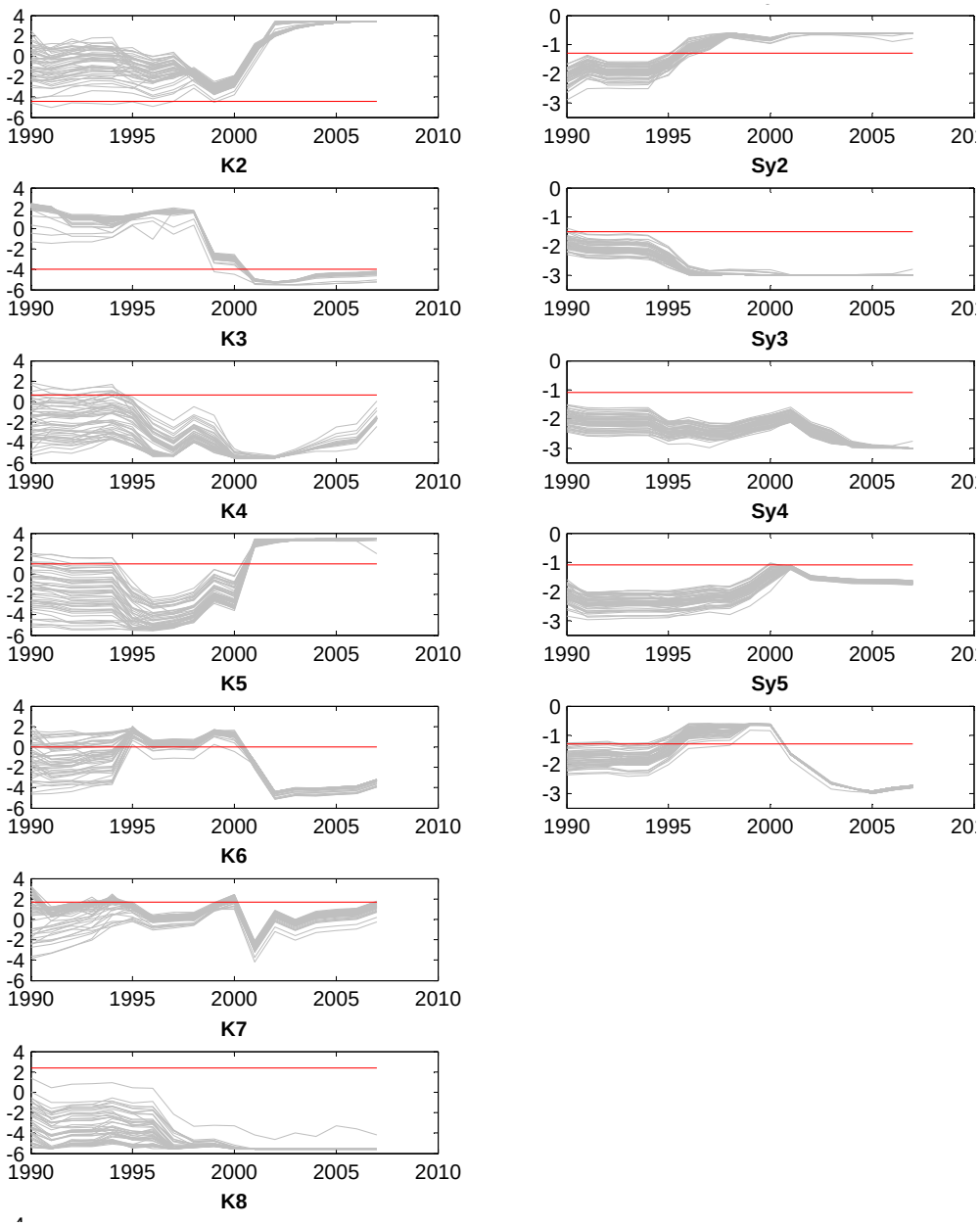


Figure 6.14: Variability in state variables (K and Sy) from Case 4 (Iterative SF). Measured (black dots), GMD3 calibrated model (red line), results from EnKF (gray lines), and mean streamflow from EnKF (blue line).

Table 6.3: Streamflow RMSE (cfs) and water level RMSE (ft) for the PEST calibration and EnKF calibrations.

Gage Station Observation Well	PEST Calibration	Case 1 Single SF	Case 2 Single H	Case3 Single SFH	Case 4 Iterative SF	Case 5 Iterative H	Case6 Iterative SFH
Syracuse	12.1	14.4	15.5	15.35	20.65	14.85	13.1
Kendall	7.8	6.7	8.05	8.05	36.25	6.95	10.0
Deerfield	48.3	57.7	34.95	50.35	82.85	82.85	61.7
Garden City	17.9	29.3	24.35	24.25	55.65	57.65	36.0
Dodge City	39.7	29.4	54.15	39.25	22.25	26.15	16.9
Cimarron R Forgan OK	125.3	162.0	170.85	171.15	140.75	152.05	123.9
Crooked Cr Englewood	5.1	5.7	6.45	5.95	8.15	6.65	6.8
Average (7 streams)	36.6	43.6	44.95	44.95	52.35	49.55	38.3
R93 C64	62.6	78.2	53.85	59.55	217.85	131.45	156.7
R84 C97	49.7	53.5	40.45	49.85	35.65	45.35	47.7
R64 C33	38.9	55.0	28.45	39.65	418.85	143.25	194.2
R71 C14	38.4	38.8	37.75	38.35	34.55	38.75	40.6
R47 C33	37.3	8.7	11.85	17.35	173.75	71.95	90.2
R88 C67	36.6	48.4	14.15	24.85	194.85	96.55	120.0
R7 C4	36.2	36.1	36.25	36.15	36.45	28.45	31.6
R67 C25	35.6	34.5	28.75	31.45	76.45	14.75	15.8
R49 C77	33.8	55.0	42.45	36.25	420.15	155.95	260.1
R72 C91	33.8	60.6	30.95	36.05	224.65	116.45	167.8
Average (122 wells)	15.8	27.75	20.25	18.75	153.85	71.15	100.3

Chapter 7

Conclusions

This report documents our investigation into the reconstruction of multiscale fields, generally permeability fields, from sparse observations. Contemporary studies on the estimation of fields do so on grids; "multiscale" typically refers to the range of structures/length-scales in the fields being reconstructed. Algorithmic approaches that seek to reduce the computational costs associated with estimating small-scale structures form bulk of the research being conducted today in the world.

In contrast, we consider the case where the field being reconstructed from sparse observations contains structures too small to be resolved by a mesh. We, too, proceed with a grid, with the aim of estimating resolved spatial structures, but also seek statistical summaries of subgrid structures. We perform this investigation first within the context of a binary medium.

We consider a low permeability matrix containing high permeability inclusions which may be present in any proportion, but individually are too small to be resolved by a grid. The proportion of inclusions varies smoothly in space, but in an unknown manner. We are provided with sparse observations of permeability at a set of "sensor" points, along with tracer breakthrough times, from an injector-producer test, at the same sensors. The aim of the investigation is to reconstruct the inclusion proportion field, estimate the inclusion size and create realizations of the binary field conditioned on data.

In Chapter 2, we construct an analytical subgrid model that summarizes the effect of dispersed inclusions at a grid-block level. The inclusions are modeled probabilistically using truncated Gaussians and their upscaling is performed by adapting an existing upscaling theory. The subgrid model is parametrized by the inclusion proportion and a proxy for the inclusion size.

In Chapter 3, this subgrid model is combined with an Eulerian single-phase porous media transport model to reconstruct the permeability field, determine the inclusion proportion distribution in space and estimate the inclusion size. The inversion is performed on a grid, but the dimensionality of the inversion, posed as a Bayesian inverse problem, is re-

duced by modeling the field being estimated with a truncated Karhunen-Loève expansion. The inverse problem is solved by constructing the posterior distribution of weights of the Karhunen-Loève modes with an adaptive Markov chain Monte Carlo (MCMC) method, thus capturing the uncertainty in the estimate. The method is tested on synthetic data.

We find that permeability measurements (static data) are sufficient for reconstructing the large-scale structures in the problem, whereas the breakthrough times (dynamic data) are informative about the small-scale structures. The size of the inclusions can only be inferred in very tightly constrained situations; generally, the estimation of the inclusion size is ruined by structural/model errors. However, the distribution of Karhunen-Loève mode weights and inclusion size can be used to create realizations of the "fine-scale" binary field, where individual inclusions are resolved. When used as an input to porous media transport models, we can perform posterior predictive tests for the breakthrough times at the sensors. We see that the binary fields conditioned on both static and dynamic data have higher predictive skill versus the ones reconstructed using just the static data; further, they are also more robust, retaining their predictive capacity in flow configurations which are different from the one used to obtain the breakthrough-time measurements. However, dynamic data is not a silver bullet, as far as estimation accuracy and predictive skills are concerned, and we provide an example where dynamic data confers very little advantage.

In the investigation above, the subgrid model played the fundamental role of providing a link across resolved and unresolved scales. In Chapter 4, we consider the case where such an explicit interscale link does not exist. We consider the case of estimating a permeability field from one of the layers in the SPE10 dataset, which is similar in structure to a fluvial bed. Multiscale finite elements are used to provide a local upscaling capability, similar to the subgrid model described in Chapter 2. We find that the recourse to multiscale finite elements increases the computational cost and complexity of the inference. In conjunction with a specialized MCMC sampler, we solve a Bayesian inverse problem for a 1D permeability field, predicated on pressure head measurements. Extensions to 2D are underway.

Our investigations in Chapter 3 revealed that (1) MCMC methods are largely unscalable and (2) the posterior distributions of Karhunen-Loève weights were approximately Gaussian, indicating the potential of using scalable Ensemble Kalman Filters (EnKF) to perform the inference. We demonstrate this in Chapter 5, using static observations of permeability and dynamic observations of saturation history from a waterflood simulation at a set of sensors (i.e., we use synthetic data). In order to incorporate the static data, we hybridized the EnKF with MCMC. The static data is assimilated first, to construct an informative initial condition for the EnKF; thereafter, the EnKF assimilates the saturation history sequentially. The hybrid EnKF was used to reconstruct a smooth permeability field as well as a "pathological" case, whose permeability distribution resembled that of a fractured medium.

Finally, the scalable EnKF was used to address a realistic permeability estimation problem in the Ogallala acquifer in Chapter 6. For the forward problem, we use a MODFLOW model of a 100×150 mile corner of the Ogallala aquifer in Kansas. The model was obtained from the Kansas Geological Survey (KGS). We estimate permeability and specific yield. Water table levels at 122 wells over 18 years, along with lithologic logs, are the ob-

servations. The EnKF-estimated permeabilities and specific yields were compared against the estimates obtained by KGS using a deterministic (optimization) approach. We found that the extreme non-linearity of the MODFLOW model (the permeability and water-table heights are non-linearly related) creates serious challenges when EnKF are used for the problem, and the posterior predictive tests sometimes failed to reproduce the measurements. However, the version of EnKF used for this purpose was a "naive" one e.g., without parameter localization, and it is expected that further modeling and EnKF enhancements, to address the non-linearity issue, will considerably improve the EnKF estimates.

In summary, it is possible to perform the stochastic reconstruction of a multiscale field; on occasion, we may even be able to estimate structures which are smaller than the resolution of the grid on which the estimation is performed. The estimation procedure is the easiest if one can construct a subgrid model that summarizes the impact of the unresolved structures at the resolved scale; in its absence, a recourse to multiscale finite elements may bear fruit. The estimation is performed, with minimum approximation, using MCMC, but the method is not scalable and one is limited to small problems. EnKF may allow us to scale to large problems, but requires significant modeling and algorithmic enhancements to address highly nonlinear problems. In all cases, observations are key, and probabilistic estimation procedures benefit when the observation schemes are crafted to be most informative on the problem at hand.

This page intentionally left blank

References

[1] G. Liu, B. Wilson, D. Whittemore, W. Jin, and J. Butler Jr. Ground-water model for southwest kansas groundwater management district No. 3. Kansas. Geological Survey Open File Report 2010-18, Kansas Geological Survey, 2010.

[2] J.-P. Chiles and P. Delfiner, editors. *Geostatistics: Modeling Spatial Uncertainty (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 1999.

[3] M. Hayek, F. Lehmann, and P. Ackerer. Adaptive multi-scale parameterization for one-dimensional flow in unsaturated porous media. *Advances in Water Resources*, 31:28–43, 2008.

[4] I. Berre, M. Lien, and T. Mannseth. A level-set corrector to an adaptive multiscale permeability prediction. *Computational Geosciences*, 11:27–42, 2007.

[5] L. K. Nielsen, H. Li, X-C Tai, S. I. Aanonsen, and M. Espedal. Reservoir description using a binary level set method. *Computing and Visualization in Science*, 13:41–58, 2010.

[6] I. Berre, M. Lien, and T. Mannseth. Multilevel parameter structure identification for two-phase porous-media flow problems using flexible representations. *Advances in Water Resources*, 32:1777–1788, 2009.

[7] Y. Efendiev, T. Hou, and W. Luo. Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM Journal for Scientific Computing*, 28(2):776–803, 2006.

[8] Y. Efendiev and T. Hou. Multiscale finite element methods for porous media flows and their applications. *Applied Numerical Mathematics*, 57(5-7):577–596, 2007. Special Issue for the International Conference on Scientific Computing.

[9] Jianlin Fu and J. J. Gómez-Hernández. A blocking Markov chain Monte Carlo method for inverse stochastic hydrogeological modeling. *Mathematical Geosciences*, 41:105–128, 2009.

[10] T. Miloh and Y. Benveniste. A generalized self-consistent method for the effective conductivity of composites with ellipsoidal inclusions and cracked bodies. *Journal of Applied Physics*, 63(3):789–796, 1988.

[11] T. Miloh and Y. Benveniste. On the effective conductivity of composites with ellipsoidal inhomogeneities and highly conducting interfaces. *Proceedings of the Royal Society of London, A*, 455:2687–2706, 1999.

[12] K. Schulgasser. On the conductivity of fiber reinforced materials. *Journal of Mathematical Physics*, 17:382–387, 1976.

[13] G. W. Milton. Bounds on the electromagnetic, elastic and other properties of two-component composites. *Physical Review Letters*, 46(8):542–545, 1981.

[14] A. H. Sihvola and J. A. Kong. Effective permittivity of dielectric mixtures. *IEEE Transactions on Geosciences and Remote Sensing*, 26(4):420–429, 1988.

[15] A. J. Desbarats. Numerical estimation of effective permeability in sand-shale formations. *Water Resources Research*, 23(2):273–286, 1987.

[16] P. A. Fokker. General anisotropic effective medium theory for the effective permeability of heterogeneous reservoirs. *Transport in Porous Media*, 44(2):205–218, 2001.

[17] Z. Hashin and S. Shtrikman. A variational approach to the theory of effective magnetic permeability of multiphase materials. *Journal Applied Physics*, 33(10):3125–3131, 1962.

[18] M. Sahimi. *Flow and transport in porous media and fractured rock*. VCH, 1995.

[19] P. Renard and G. de Marsily. Calculating equivalent permeability: a review. *Advances in Water Resources*, 20(5-6):253–78, 1997.

[20] J. Rubinstein and S. Torquato. Flow in random porous media: mathematical formulation, variational principles and rigorous bounds. *Journal of Fluid Mechanics*, 206:25–46, 1989.

[21] H. H. Haldorsen and L. W. Lake. A new approach to shale management in field scale simulation models. *57th Annual Technical Conference, Society of Petroleum Engineers*, 1982.

[22] R. W. Ritzi, D. F. Jayne, A. J. Zahradnik, A. A. Field, and G. E. Fogg. Geostatistical modeling of heterogeneity in glaciofluvial, buried-valley aquifers. *Ground Water*, 32(4):666–674, 1994.

[23] R. W. Ritzi, D. F. Dominic, N. R. Brown, K. W. Kausch, P. J. McAlenney, and M. J. Basial. Hydrofacies distribution and correlation in the miami valley aquifer system. *Water Resources Research*, 31(12):3271–3281, 1995.

[24] G. E. Fogg. Groundwater flow and sand body interconnectedness in a thick, multiple-aquifer system. *Water Resources Research*, 22(5):679–694, 1986.

[25] J.-R. De Dreuzy, P. Davy, and O. Bour. Hydraulic properties of two-dimensional random fracture networks following a power-law length distribution: 1. effective connectivity. *Water Resources Research*, 37(8):2065–2078, 2001.

[26] L. Smith and F. W. Schwartz. An analysis on the influence of fracture geometry on mass transport in fractured media. *Water Resources Research*, 20(9):1241–1252, 1984.

[27] C. D. Langevin. Stochastic ground water flow simulation with a fracture zone continuum model. *Ground Water*, 41(5):587–601, 2003.

[28] S. A. McKenna and P. C. Reeves. Fractured continuum approach to stochastic permeability modeling. In T. C. Coburn, J. M. Yarus, and R. L. Chambers, editors, *Stochastic Modeling and Geostatistics: Principles, Methods and Case Studies, Volume II, AAPG Computer Applications in Geology 5*, pages 173–186. 2006.

[29] D. M. Reeves, D. A. Benson, and M. M. Meerschaert. Transport of conservative solutes in simulated fracture networks: 1. synthetic data generation. *Water Resources Research*, 44(W05404), 2008.

[30] K. S. Mendelsohn. A theorem on the effective conductivity of a two-dimensional heterogeneous medium. *Journal of Applied Physics*, 46(11):4740–4741, 1975.

[31] G. Dagan. Models of groundwater flow in statistically homogeneous porous formations. *Water Resources Research*, 15(1):47–63, 1979.

[32] G. Dagan. *Flow and transport in porous media*. Springer-Verlag, 1989.

[33] A. D. Poley. Effective permeability and dispersion in locally heterogeneous aquifers. *Water Resources Research*, 24(11):1921–1926, 1988.

[34] S. Pozdniakov and C. F. Tsang. A self-consistent approach for calculating the effective hydraulic conductivity of a binary, heterogeneous medium. *Water Resources Research*, 40, 2004. doi:10.1029/2003WR002617. W05105.

[35] R. W. Zimmerman. Effective conductivity of a two-dimensional medium containing elliptical inhomogeneities. *Proceedings of the Royal Society*, (452):A1713–A1727, 1996.

[36] C. Knudby, J. Carrera, J. D. Bumgardner, and G. E. Fogg. Binary upscaling: The role of connectivity and a new formula. *Advances in Water Resources*, 29:590–604, 2006.

[37] P. R. King. The use of renormalization for calculating effective permeability. *Transport in Porous Media*, 4(1):37–58, 1989.

[38] S. Mukhopadhyay and M. Sahimi. Calculation of the elective permeabilities of field-scale porous media. *Chemical Engineering Science*, 55:4495–4513, 2000.

[39] F. Ebrahimi and M. Sahimi. Multiresolution wavelet coarsening and analysis of transport in heterogeneous media. *Physica A*, 316:160–188, 2002.

[40] A. R. Solow. Mapping by simple indicator kriging. *Mathematical Geology*, 18(3):335–352, 1986.

[41] A. G. Journel and F. Alaber. Non-Gaussian data expansion in the earth sciences. *Terra Nova*, 1(2):123–134, 1989.

[42] J. J. Gomez-Hernandez and R. M. Srivastava. ISIM3D: An ANSI-C three-dimensional multiple indicator conditional simulation program. *Computers and Geosciences*, 16(4):395–440, 1990.

[43] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, 1998.

[44] S. F. Carle and G. E. Fogg. Transition probability-based indicator geostatistics. *Mathematical Geology*, 28(4):453–476, 1996.

[45] S. F. Carle and G. E. Fogg. Modeling spatial variability with one and multidimensional continuous lag markov chains. *Mathematical Geology*, 29(7):891–918, 1997.

[46] C. Lantuejoul. *Geostatistical Simulation: Models and Algorithms*. Springer-Verlag, 2002.

[47] T. Harter and C. Knudby. Effective conductivity of periodic media with cuboid inclusions. *Advances in Water Resources*, 27(10):1017–1032, 2004.

[48] M. Armstrong, A. G. Galli, H. Beucher, G. Loch, D. Renard, B. Doligez, R. Eschard, and F. Geffroy. *Plurigaussian Simulations in Geosciences*. Springer, 2003.

[49] R. J Adler. On excursion sets, tube formulas and maxima of random fields. *Annals of Applied Probability*, 10(1):1–74, 2000.

[50] R. J. Adler, J. E. Taylor, and K. J. Worsley. *Applications of Random Fields and Geometry: Foundations and Case Studies*. 2009.

[51] J. E. Taylor and R. J. Adler. Euler characteristics for Gaussian fields on manifolds. *The Annals of Probability*, 31(2):533–563, 2003.

[52] K. J. Friston, K. J. Worsley, R. S. J. Frackowiak, J. C. Mazziotta, and A. C. Evan. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:210–220, 1994.

[53] K. J. Worsley, S. Marrett, P. Neelan, and A. C. Evans. Searching scale space for activation in pet images. *Human Brain Mapping*, 4(1):74–90, 1996.

[54] D. J. Nott and R. J. Wilson. Multi-phase image modelling with excursion sets. *Signal Processing*, 80:125–139, 2000.

[55] F. M. Phillips and J. L. Wilson. An approach to estimating hydraulic conductivity spatial correlation scales using geological characteristics. *Water Resources Research*, 25(1):141–143, 1989.

[56] *Matlab, 2009a, Image Processing Toolbox*. Natick, MA, USA.

[57] P. J. Diggle. *Statistical analysis of spatial point patterns, Second edition*. Oxford University Press, 2003.

[58] A. W. Harbaugh. MODFLOW-2005, the U.S.Geological Survey modular groundwater model – the ground water flow process. U.S. Geological Survey Techniques and Methods 6-A16, U.S. Geological Survey, 2005.

[59] D. A. G. Bruggeman. Berechnung verschiedener physikalischer konstanten von heterogenen substanzen. *Annals of Physics (Leipzig)*, 24:636–679, 1935.

[60] S. E. Silliman and C. Frost. Monitoring hydraulic gradients using three-point estimators. *Journal of Environmental Engineering*, 124(6):517–523, 1998.

[61] S. A. McKenna and A. Wahi. Local hydraulic gradient estimator analysis of long-term monitoring networks. *Ground Water*, pages 723–731, 2006.

[62] D. S. Oliver and Y. Chen. Recent progress on reservoir history matching: A review. *Computational Geosciences*, 15:185–221, 2011.

[63] H. J. Hendricks Franssen, A. Alcolea, M. Riva, M. Bakr, N. van der Wiel, F. Stauffer, and A. Guadagnini. A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterization of well catchments. *Advances in Water Resources*, 32:851–872, 2009.

[64] H. Lee, D. Higdon, Z. Bi, M. Ferreira, and M. West. Markov random field models of high-dimensional parameters in simulations of fluid flow in porous media. *Technometrics*, 44(3):230–241, 2002.

[65] J. Wang and N. Zabaras. A Markov random field model of contamination source identification in porous media flow. *International Journal of Heat and Mass Transfer*, 49:939–950, 2006.

[66] J. Fu and J. J. Gómez-Hernández. Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov chain Monte Carlo method. *Journal of Hydrology*, 364:328–341, 2009.

[67] W. Li and O. A. Cirpka. Efficient geostatistical inverse methods for structured and unstructured grids. *Water Resources Research*, 42, 2006. W06402.

[68] Y. M. Marzouk and H. N. Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862 – 1902, 2009.

[69] B. Jafarpour and D. McLaughlin. Efficient permeability parameterization with discrete cosine transform. In *The Proceedings of the SPE Reservoir Simulation Symposium*, number Paper SPE 106453, 2007.

[70] B. Jafarpour and D. B. McLaughlin. History matching with an ensemble Kalman filter and discrete cosine parameterization. *Computational Geosciences*, 12:227–244, 2008.

[71] B. Jafarpour, V. K. Goyal, D. B. McLaughlin, and W. T. Freeman. Compressed history matching: Exploiting transform-domain sparsity for regularization of nonlinear dynamic data integration problems. *Mathematical Geosciences*, 42:1–27, 2010.

[72] M. A. R. Ferreira, Z. Bi, M. West, H. K. H. Lee, and D. Higdon. Multi-scale modeling of 1-D permeability fields. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M Smith, and M. West, editors, *Bayesian Statistics*, volume 7, pages 519–527. Oxford University Press, 2003.

[73] S. I. Aanonsen and D. Eydinov. A multiscale method for distributed parameter estimation with application to reservoir history matching. *Computational Geosciences*, 10:97–117, 2006.

[74] D. Higdon, H. Lee, and Z. Bi. A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine scale information. *IEEE Transactions in Signal Processing*, 50(2):389–399, 2002.

[75] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1996.

[76] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242, 2001.

[77] H. Haario, M. Laine, A. Mira, and E. Saksman. DRAM-Efficient adaptive MCMC. *Statistics and Computing*, 16(4):339–354, 2006.

[78] J. Goodman and A. D. Sokal. Multigrid Monte-Carlo method: conceptual foundations. *Physical Review D*, 40:2035–2071, 1989.

[79] H. Haario, E. Saksman, and J. Tamminen. Adaptive proposal distribution for random-walk Metropolis algorithm. *Computational Statistics*, 14:375–395, 1999.

[80] L. Tierney and A. Mira. Some adaptive Monte Carlo techniques for Bayesian inference. *Statistics in Medicine*, 18:2507–2515, 1999.

[81] P. J. Green and A. Mira. Delayed rejection in reversible jump Metropolis-Hastings. *Biometrika*, 88:1035–1053, 2001.

[82] A. Mira. Ordering and improving the performance of Monte Carlo Markov Chains. *Statistical Science*, 16:340–350, 2002.

[83] S. A. McKenna, J. Ray, Y. Marzouk, and B. van Bloemen Waanders. Truncated multiGaussian fields and effective conductance of binary media. *Advances in Water Resources*, 34:617–626, 2011.

[84] R. J. Adler and J. E. Taylor. *Random fields and geometry*. Springer, 2007.

[85] M. Griogoriu. *Stochastic calculus*. Birkhauser, 2002.

[86] J. E. Aarnes, Gimse T, and K-A. Lie. *Geometric modelling, numerical simulation, and optimization*, chapter An introduction to the numerics of flow in porous media using Matlab, pages 265–306. Springer, 2007.

[87] J. B. Bell and P. Colella. A second-order projection method for the incompressible Navier-Stokes equations. *J. Comp. Phys.*, 85:257–283, 1989.

[88] D. W. Pollack. User's guide for MODPATH/MODPATH-PLOT, Version 3: A particle tracking post-processing package for modflow, the U.S. Geological Survey finite-difference ground-water flow model. U.S. Geological Survey Open-File Report 94-464, U.S. Geological Survey, 1994.

[89] A. E. Raftery and S. M. Lewis. The number of iterations, convergence diagnostics and generic metropolis algorithms. In W. R. Gilks, D. J. Spiegelhalter, and S. Richardson, editors, *Practical Markov Chain Monte Carlo*. Chapman and Hall, 1995.

[90] G. R. Warnes. *mcgibbsit: Warnes and Raftery's MCGibbsit MCMC diagnostic*, 2005. R package version 1.0.5.

[91] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[92] J. Ray, S. A. McKenna, B. van BloemenWaanders, and Y. M. Marzouk. Bayesian estimation of unresolved fine-scale properties of binary media from multiscale observations. SAND Report SAND2011-3079, Sandia National Laboratories, Livermore, CA 94550-0969, May 2011. Unclassified unlimited release; available at http://csmr.ca.sandia.gov/∼jairay/index.html.

[93] T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[94] T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007.

[95] David W. Scott. *Multivariate density estimation*. Wiley Series in Probability and Statistics, 1992.

[96] P. Hall, S. J. Sheather, M. C. Jones, and J. S. Marron. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, 78(2):263–269, 1991.

[97] A. G. Gray and A. W. Moore. Very fast multivariate kernel density estimation via computational geometry. In *Proceedings of the Joint Statistical Meeting*, 2003.

[98] A. Ihler. Kernel Density Estimation Toolbox in MATLAB (R13). `http://www.ics.uci.edu/~ihler/code/kde.html`, 2003.

[99] J.C. Marechal, M.P. Sarma, S. Ahmed, and P. Lachassagne. Establishment of earth tide effect on water-level fluctuations in an unconfined hard rock aquifer using spectral analysis. *Current Science*, 83(1), July 2002.

[100] M. Christie and M. Blunt. 2001 SPE comparative solution project, 2001.

[101] Ph. Renard and G. de Marsily. Calculating equivalent permeability: A review. *Advances in Water Resources*, 20(5–6):253–278, 1997.

[102] C.L. Farmer. Upscaling: a review. *International Journal for Numerical Methods in Fluids*, 40:63–78, 2002.

[103] J.'E. Warren and H. S. Price. Flow in heterogeneous porous media. *Society of Petroleum Engineers Journal*, pages 153–169, September 1961.

[104] C. He and L. J. Durlofsky. Structured flow-based gridding and upscaling for modeling subsurface flow. *Advances in Water Resources*, pages 1876–1892, 2006.

[105] M. Prevost, F. Lepage, L. J. Durlofsky, and J-L. Mallet. Unstructured 3d gridding and upscaling for coarse modelling of geometrically complex reservoirs. *Petroleum Geoscience*, 11:339–345, 2005.

[106] P. K. Kitanidis. Groundwater flow in heterogeneous formations. In G. Dagan and S. P. Neuman, editors, *Subsurface Flow and Transport: A Stochastic Approach*, pages 83–90. UNESCO, 1997.

[107] T. Hou and Y. Efendiev. *Multiscale Finite Element Methods: Theory and Applications*. Springer, 2009.

[108] G. A. Pavliotis and A. M. Stuart. *Multiscale Methods: Averaging and Homogenization*. Springer, 2000.

[109] T. J. R. Huges, G. R. Feijoo, L. Mazzei, and J-B. Quincy. The variational multiscale method – a paradigm for computational mechanics. *Computer methods in applied mechanics and engineering*, 166:3–34, 1998.

[110] R. Juanes and F-X. Dub. A locally conservative variational multiscale method for the simulation of porous media flow with multiscale source terms. *Computational Geosciences*, 12:273–295, 2008.

[111] W. E and B Engquist. The heterogeneous multi-scale methods. *Commun. Math. Sci.*, 1:87–133, 2003.

[112] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden. Heterogeneous multiscale methods: A review. *Communications in Computational Physics*, 2(3):367–450, 2007.

[113] T. Arbogast. Numerical subgrid upscaling of two-pahse flow in porous media. *Lecture Notes in Phyics*, 552:35–49, 2000.

[114] J. E. Aarnes and Y. Efendiev. Mixed multiscale finite element methods for stochastic porous media flows. *SIAM Journal on Scientific Computing*, 30(5):2319–2339, 2008.

[115] P. Jenney, S.H. Lee, and H.A. Tchelepi. Multi-scale finite-volume method for elliptic problems in subsurface flow simulation. *Journal of Computational Physics*, 187:47–67, 2003.

[116] P. Jenny, S.H. Lee, and H.A. Tchelepi. Adaptive fully implicit multi-scale finite-volume method for multi-phase flow and transport in heterogeneous porous media. *Journal of Computational Physics*, 217:627–641, 2006.

[117] X. Frank Xu. A multiscale stochastic finite element method on elliptic problems involving uncertainties. *Computational methods in applied mechanics and engineering*, 196:2723–2736, 2007.

[118] G. Strang and G. Fix. *An analysis of the Finite Element Method, 2nd ed.* Wellesly-Cambridge, 2008.

[119] T. Y. Hou and X. H. Wu. A multiscale finite element method for elliptic problems in composite material and porous media. *Journal of Computational Physics*, 134:169–189, 1997.

[120] M. A. R. Ferreira and H. K. Lee. *Multiscale Modeling: A Bayesian Persepective*. Springer, 2007.

[121] D. Higdon, H. Lee, and Z. Bi. A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine information. *IEEE Transactions on Signal Processing*, 50:389–399, 2002.

[122] P. Dostert, Y. Efendiev, T. Hou, and W. Luo. Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification. *Journal of Computational Physics*, 217(1):123–142, September 2006.

[123] Y. Efendiev, T. Hou, and W. Luo. Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM Journal on Scientific Computing*, 28(2):776, 2007.

[124] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis, 2nd ed.* Chapman and Hall/CRC, 2004.

[125] P. Diaconis and S. L. Zabell. Updating subjective probability. *Journal of the American Statistical Association*, 77(380):822–830, 1982.

[126] G. Evensen. *Data Assimilation: the Ensemble Kalman Filter, 2nd ed.* Springer, 2009.

[127] K. Bergemann and S. Reich. A mollified Ensemble Kalman Filter. *Quaterly Journal of the Royal Meteorological Society*, 136(651):1636–1643, July 2010.

[128] J. De Geeter, H. Van Brussel, J. De Schutter, and M. Decreton. A smoothly constrained Kalman filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10):1171–1177, 1997.

[129] C. Yang and E. Blasch. Kalman filtering with nonlinear state constraints. *IEEE Transactions on Aerospace and Electronic Systems*, 45(1):70–84, 2009.

[130] L. Biegler, O. Ghattas, M. Heinkenschloss, and B. van Bloemen Waanders, editors. *Large-Scale PDE-Constrained Optimization*. Springer-Verlag, 2003.

[131] L. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. an Bloemen Waanders, editors. *Real-Time PDE-Constrained Optimization*. SIAM, 2007.

[132] G. Evenson. *Data Assimilation: The Ensemble Kalman Filter, Second edition*. Springer-Verlag, New York, 2009.

[133] L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders, and K. Willcox, editors. *Large Scale Inverse Problems and Quantification of Uncertainty*. Wiley, 2010.

[134] G. Evenson and P. van Leeuwen. Assimilation of geosat altimeter data for the agulhas current using the ensemble Kalman filter with a quasi-geostrophic model. *Mon. Weather Rev*, 124:85–96, 1996.

[135] E. S. Epstein. Stochastic dynamics prediction. *Tellus*, 21:739–759, 1969.

[136] T. A. Gleeson. Statistical-dynamical prediction. *J. Appl. Meteor.*, 9:333–344, 1970.

[137] R. J. Fleming. On stochastic dynamic prediction. Part I: The energetics of uncertainty and the question of closure. *Mon. Wea. Rev.*, 99:851–872, 1971.

[138] R. J. Fleming. On stochastic dynamic prediction. Part II: Predictability and utility. *Mon. Wea. Rev.*, 99:927–938, 1971.

[139] E. S. Epstein and E. J. Pitcher. Stochastic analysis of meteorological fields. *Tellus Ser. A*, 29:244–257, 1971.

[140] K. A. Lisaeter, G. Evenson, and S. Laxon. Assimilating synthetic cryosat sea ice thickness in a coupled ice-ocean model. *J. Geophys. Res.*, 112, 2007.

[141] T. M. Hamill and C. Snyder. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, 129:2776–2790, 2001.

[142] R. H. Reichle, McLaughlin D. B., and D. Entekhabi. Hydrologic data assimilation with the ensemble kalman filter. *Monthly Weather Review*, 2001.

[143] A. Seiler, G. Evenson, J. A. Skjervheim, J. Hove, and J. G. Vab. *Computational Methods for Large Scale Inverse Problems and Quantification of Uncertainty*, chapter Using the EnKF for history matching and uncertainty quantification of complex reservoir models. John Wiley & Sons, 2010.

[144] Y. Gu and D. S. Oliver. The ensemble Kalman filter for continuous updating of reservoir simulation models. *Journal of Energy Resource Technolgy*, 128(70), March 2006.

[145] Y. Gu and D. Oliver. An iterative ensemble kalman filter for multiphsae fluid flow data assimilation. *SPE Journal*, 12(4), 2007.

[146] H. Zhou, J. Gomez-Hernandez, H. Fransen, and L. Li. An approach to handling non-gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources*, 34(7), 2011.

[147] N. Liu and D. Oliver. Critical evaluation of the ensemble kalman filter on history matching of geologic facies. In *Proceedings of the SPE 2005 Reservoir Simulation and Symposium held in Houston, Texas*, January.

[148] W. Nowak. Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. *Water Resources Research*, 45, 2009.

[149] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 1960.

[150] A. Ben-Israel and T.N.E. Greville. *Generalized Inverse: Theory and Applications*. Wiley, New York, 1977.

[151] P. L. Houtekamer and H. L. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129:123–137, January 2001.

[152] A. Mira. On Metropolis-Hastings algorithm with delayed rejection. *Metron*, LIX:231–241, 2001.

[153] C. Geyer and E. Thompson. Annealing Markov chain Monte Carlo with applicaitons to ancestral inference. *Journal of the American Statistical Association*, 90(431), September 1995.

[154] O. Ghattas, L. Wilcox, T. Bui-Thanh, J. Martin, G. Biros, and S. Chaillet. A PDE-Constrained optimization approach to uncertainty quantification in inverse problems, with applications to inverse scattering. *Defense Technical Information Center*, February 2010.

[155] M. Laine and J. Tamminen. Aerosol model selection and uncertainty modelling by adaptive MCMC technique. *Atmospheric Chemistry and Physics*, 8:7697–7707, 2008.

[156] K. Soetaert and T. Petzoldt. Inverse modelling, sensitivity, and Monte Carlo analysis in R using package FME. *Journal of Statistical Software*, 33(3), February 2010.

[157] M. Laine. Adaptive MCMC methods with applications in environmental and geophysical models. *Finish Meteorological Institute Contributions*, March 2008.

[158] J. Aarnes, T. Gimse, and K-A. Lie. An introduction to the numerics of flow in porous media using MATLAB.

[159] S. Lophaven, H. Nielsen, and J. Sondergaard. DACE: A Matlab kriging toolbox. Technical report, IMM Informatics and Mathematical Modeling, 2002.

[160] V. L. McGuire. Ground water depletion in the high plains aquifer water levels in some areas have declined over 150 feet. U.S. Geological Survey Fact Sheet 2007-3029, U.S. Geological Survey, 2007.

[161] J. Doherty. PEST: model-independent parameter estimation. User's guide, avaialble at: http://www.pesthomepage.org, Watermark Numerical Computing, 2010.

[162] G. Welch and G. Bishop. An introduction to the Kalman filter. Technical Report TR 95-041, Department of Computer Science, University of North Carolina, 2003.

[163] G. Evensen. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143–62, 1994.

[164] G. Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.

[165] C. V. Deutsch and A. G. Journel, editors. *GSLIB, Geostatistical software library and users guide, 2nd ed*. Oxford University Press, 1998.

DISTRIBUTION:

1	Jaideep Ray, 08954	MS 9159
1	S. A. McKenna, 06911	MS 0751
1	B. van Bloemen Waanders, 01442	MS 1318
1	Technical Library, 08944 (electronic)	MS 0899