Posterior Predictive Modeling Using Multi-Scale Stochastic Inverse Parameter Estimates

Sean A. McKenna
Geoscience Research and Applications Group
Sandia National Laboratories, Albuquerque, New Mexico
samcken@sandia.gov

Jaideep Ray
Predictive Simulation R&D Department
Sandia National Laboratories, Livermore, CA

Bart van Bloemen Waanders
Applied Math and Applications Department
Sandia National Laboratories, Albuquerque, New Mexico

Youssef Marzouk
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology, Cambridge, Massachusetts

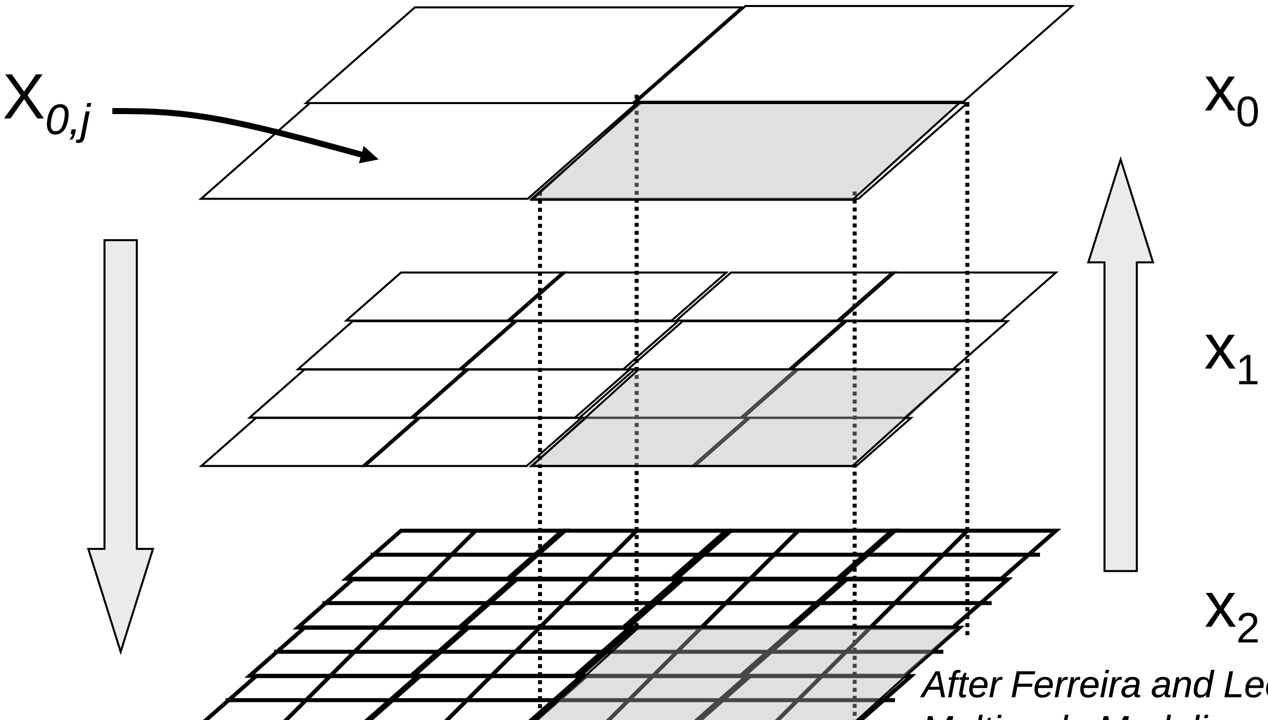
Sandia National Laboratories

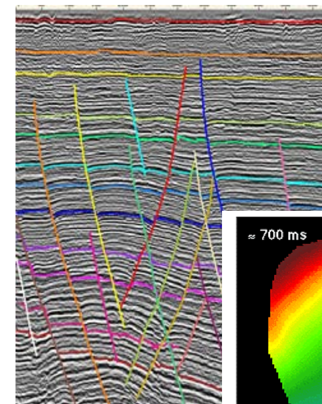
This work was funded by Sandia National Laboratories Laboratory Directed Research and Development program. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's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

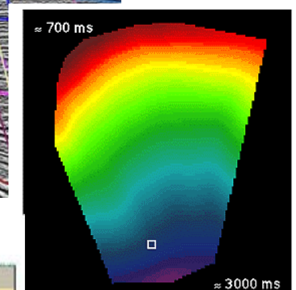
Multi-Scale Modeling Motivation

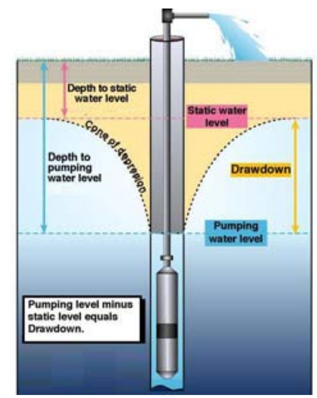
Data collected at one level informs values at other levels

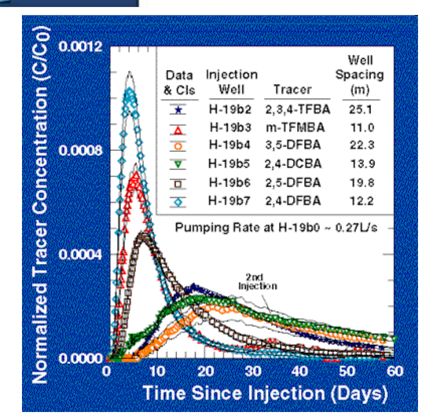
Multiscale random fields with averaging “link” between them











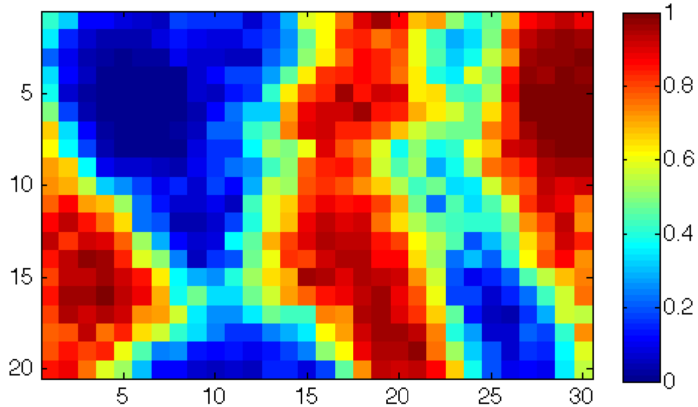
After Ferreira and Lee, 2007, Multiscale Modeling: A Bayesian Perspective

Infer statistical summaries of the fine-scale, conditional on the observations at two scales, and generate fine-scale realizations that could plausibly reproduce them

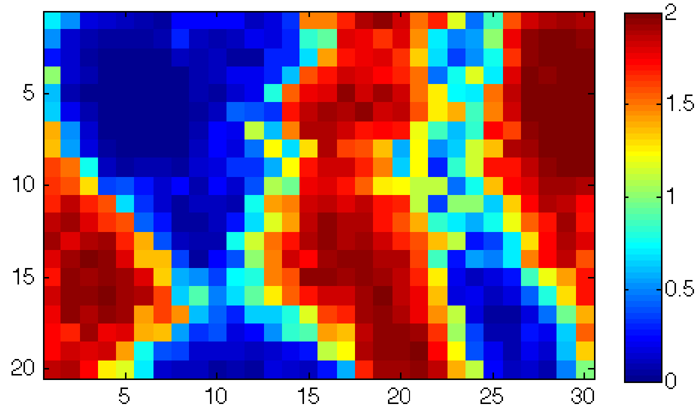
Two Scales

Model domain 3x2km, Coarse Scale: 30x20 cells, Continuous variables

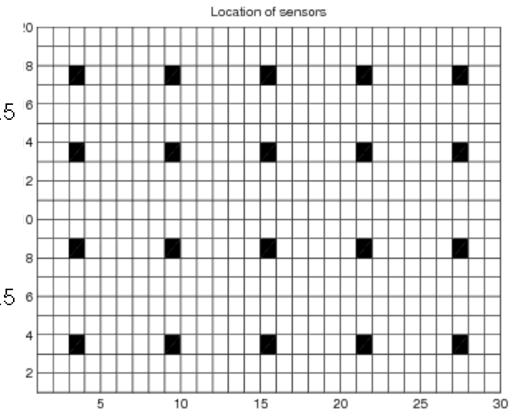
True F field



True Coarse K field

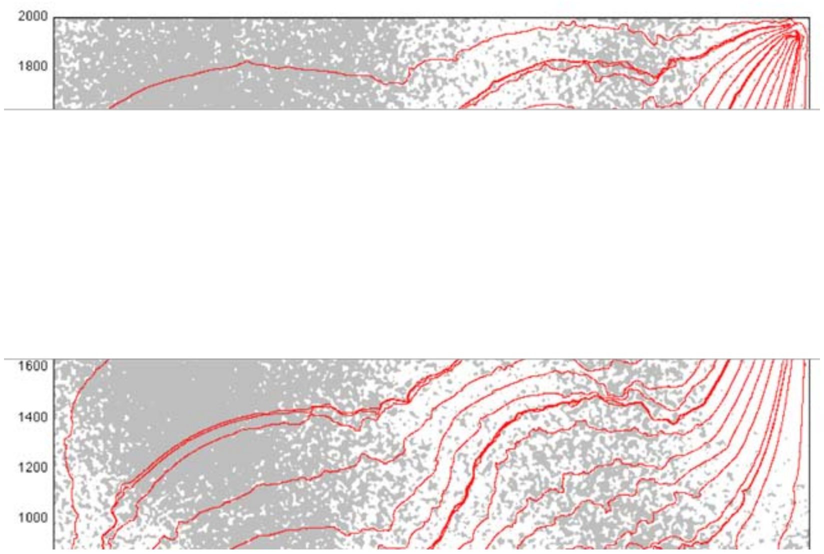


20 Well Locations



F = proportion of high conductivity

True Fine K field



<u>Fine Scale:</u>
Binary Media
3000x2000 cells
Measured travel times to 20 sensors

Injector in lower left
Producer in upper right

True binary fine-scale K field with example particle tracks

Inversion

$\zeta \sim N(0,\Gamma)$ multiGaussian process – defines spatially varying proportion field

$$\Gamma_{ij} = C(x_i, x_j) = a\exp(-|x_i - x_j|^2 / b^2)$$

$$F(x) = \frac{1}{2}\left(1 + erf\left(\frac{\zeta(x)}{\sqrt{2}}\right)\right)$$

Definition of Gaussian cdf provides transform between ζ and F

$$K_e = L(F(x), \delta, K_1, K_2)$$

Link function provides K at the coarse scale

$$t_b^0 = M(K_e)$$

Flow model operating on fine scale K provides travel times

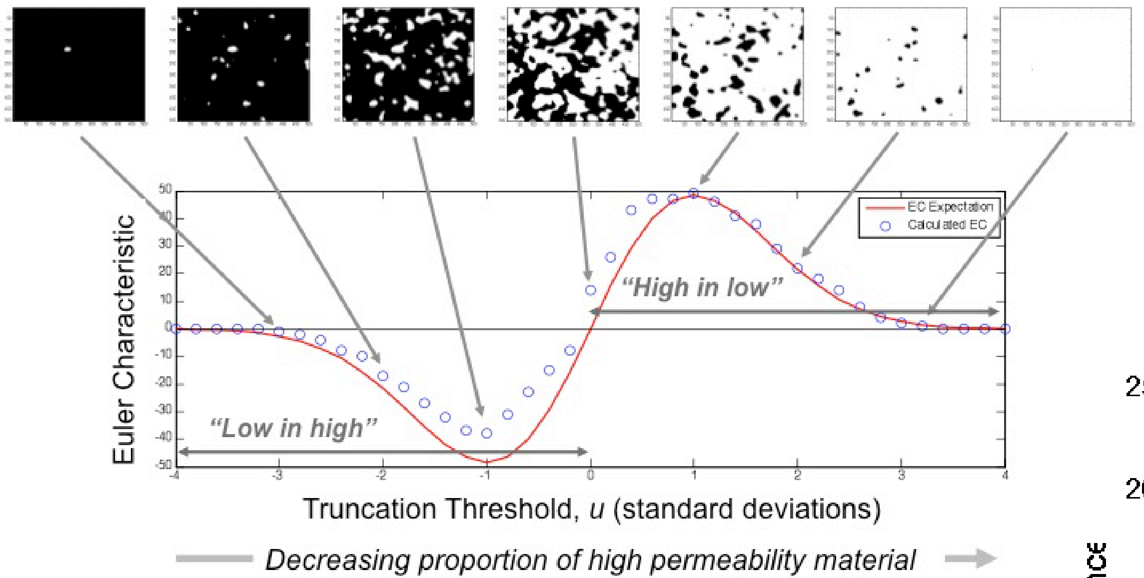
$$d_i = \{K(x)^0, t_b^1\} \qquad i = 1, \ldots, Ns$$

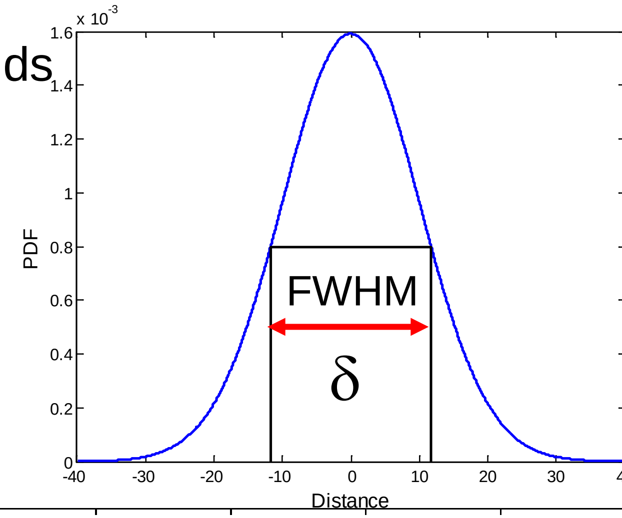
KL modes $\qquad$ $L(\delta, K_1, K_2)$

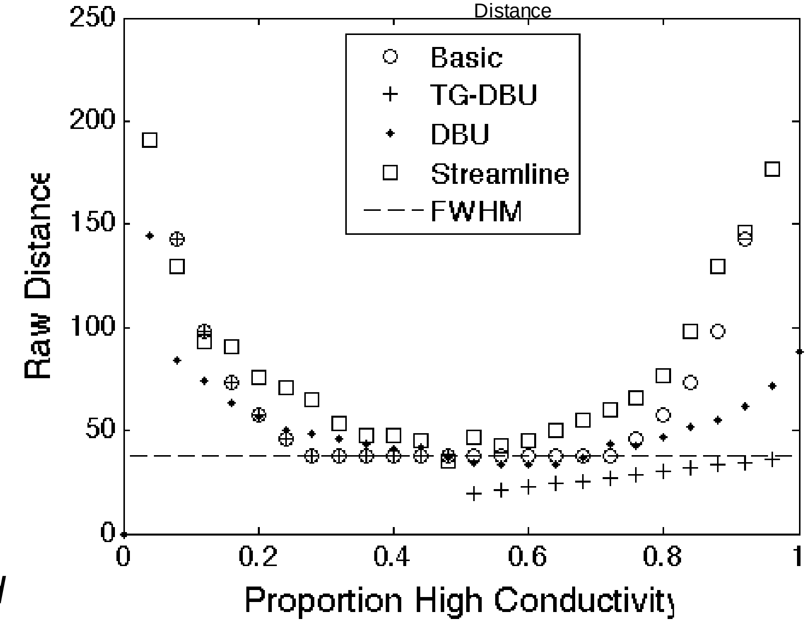
$$w \rightarrow \zeta \xrightarrow{G^{-1}} F(x) \rightarrow K_e \xrightarrow{M(K)} t_b$$

Linking Function

Binary mixtures are modeled using truncated Gaussian fields





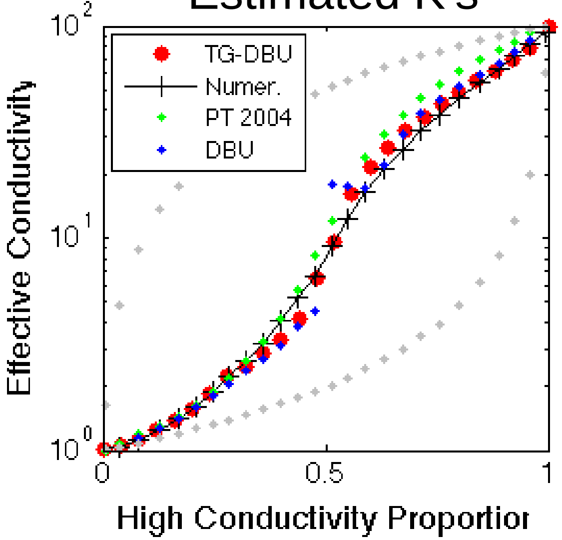


New upscaling function uses proportions (tied to truncation threshold) and average estimated distances between inclusions to estimate upscaled effective permeability

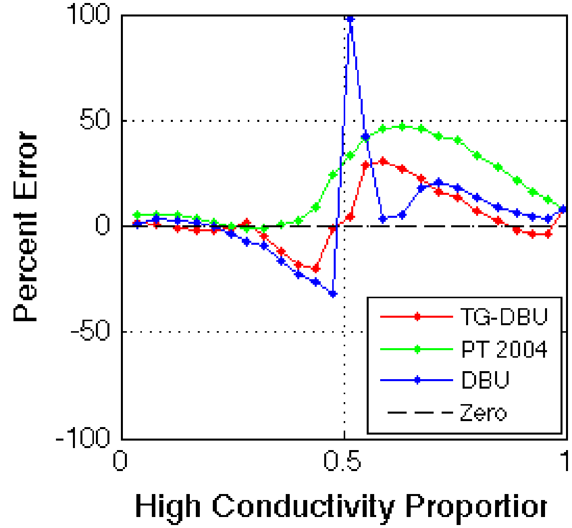
McKenna, et al., (in review), Truncated MultiGaussian Fields and Effective Conductance of Binary Media, Submitted: November, 2010

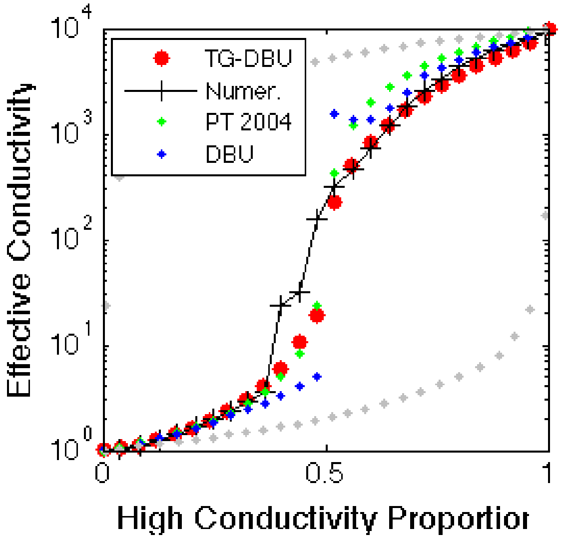
Link Function Results

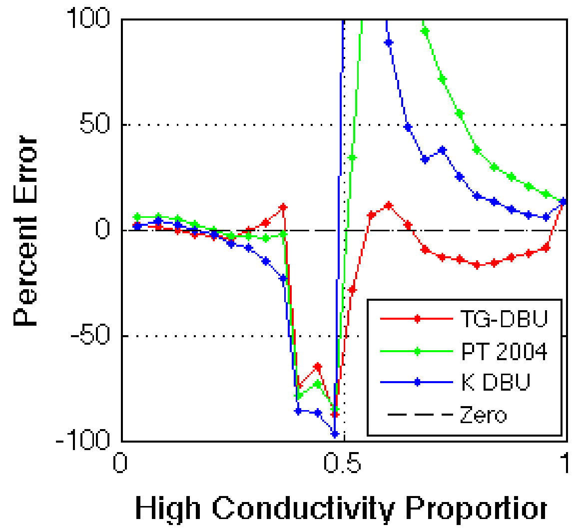
Estimated K's



Percent Error







New function is TG-DBU (Truncated Gaussian – Distance Based Upscaling)

Results compare well with DBU and another EMT-based approach

Numerical results are the average of 30 realizations

For results shown today, model errors are assumed mean zero and i.i.d.

H53E-1073: *The Effect of Error Models in the Multiscale Inversion of Binary Permeability Fields*

Bayesian Inference

$$P(\zeta,\delta \mid d) \propto P(d \mid \zeta,\delta)\pi(\delta)$$

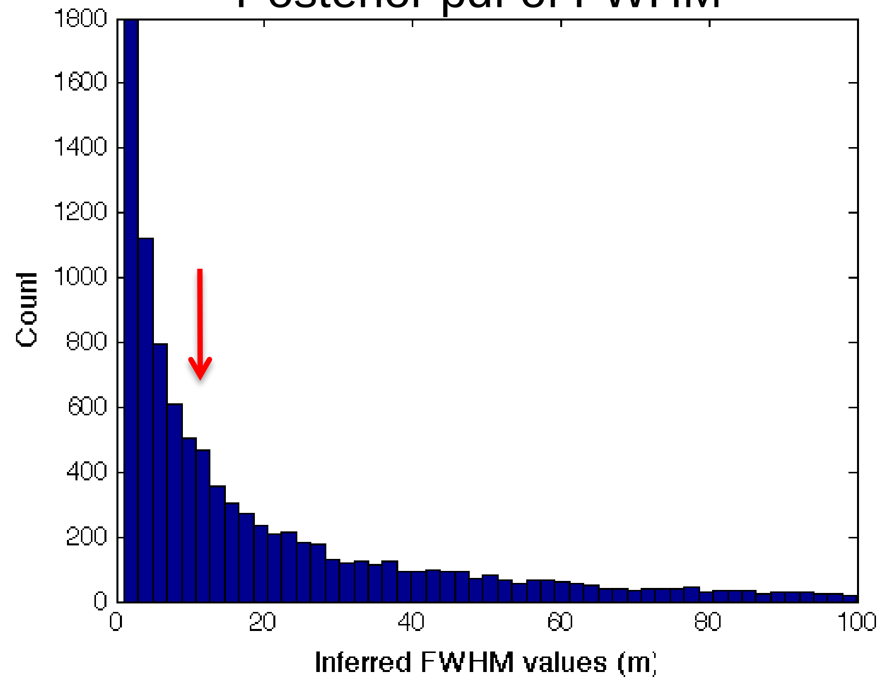
$$P(\zeta,\delta \mid d) \propto \exp\left(-\frac{[e_k(\zeta)-\mu_k]^T[e_k(\zeta)-\mu_k]}{\sigma_k^{2Ns}}\right)\exp\left(-\frac{[e_t(\zeta)-\mu_t]^T[e_t(\zeta)-\mu_t]}{\sigma_t^{2Ns}}\right)\pi(\zeta)\pi(\delta)$$

$$P(w,\delta \mid d) \propto \exp\left(-\frac{[e_k(w)-\mu_k]^T[e_k(w)-\mu_k]}{\sigma_k^{2Ns}}\right)\exp\left(-\frac{[e_t(w)-\mu_t]^T[e_t(w)-\mu_t]}{\sigma_t^{2Ns}}\right)\pi(\delta)\prod_{i=1}^{M}\exp(-w_i^2)$$

Parameterize the Gaussian process: ζ, using Karhunen-Loeve decomposition with 30 coefficients, w's

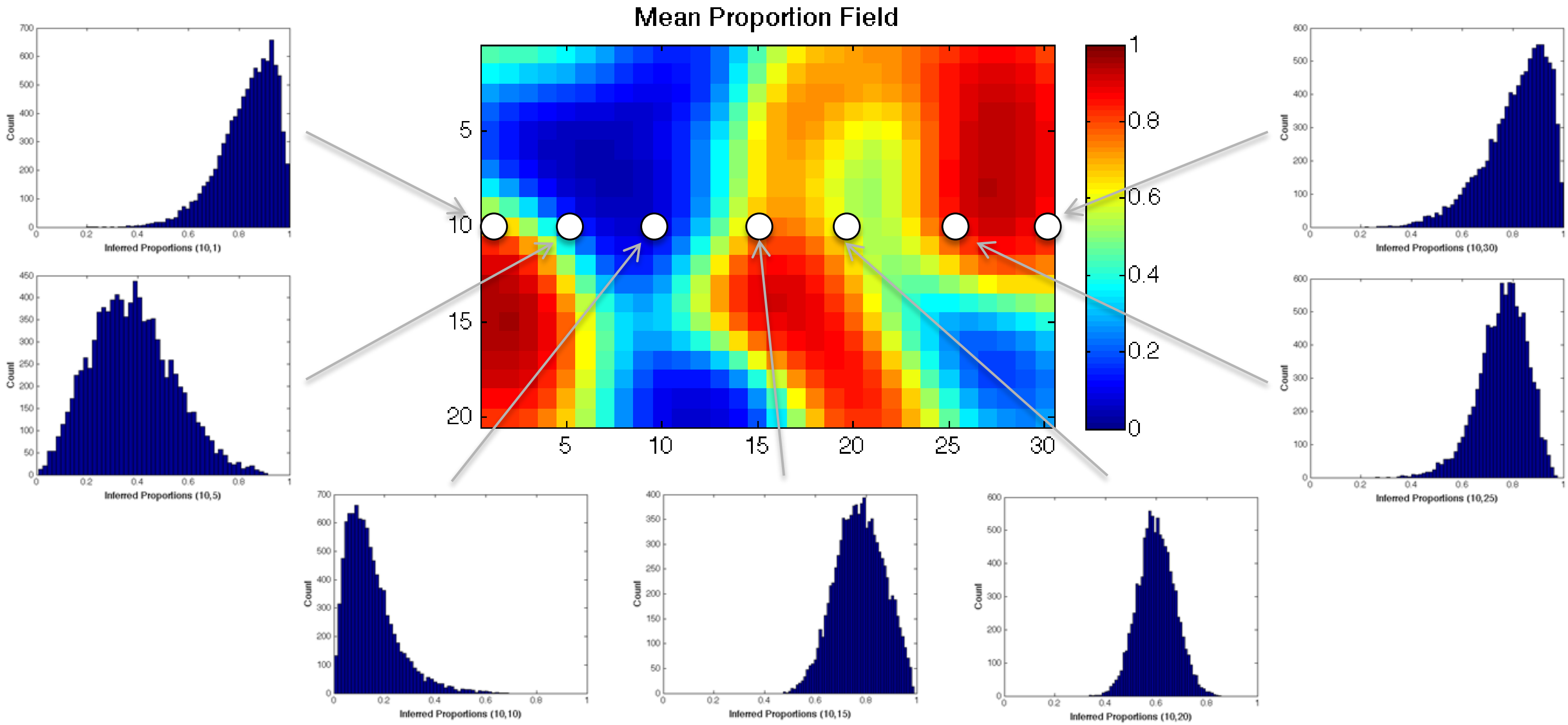
Use MCMC with delayed rejection adaptive Metropolis (DRAM) sampling to estimate 10,000 realizations of the 30 KL coefficients and the single FWHM parameter

Posterior pdf of FWHM



Estimated Proportion (F) Fields

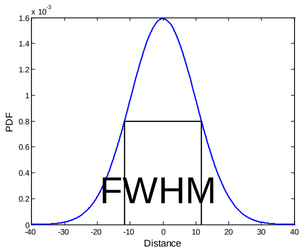
- MCMC runs met convergence diagnostics
- Results obtained with 1,500,000 iterations
  - Approximately 50 hours on workstation
  - Results in 9500 realizations of proportion field



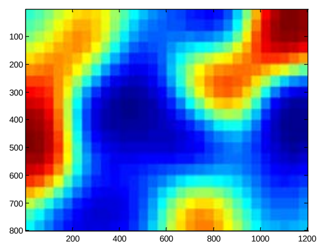
Comparison of posterior pdfs for seven points on proportion field

Posterior Evaluation

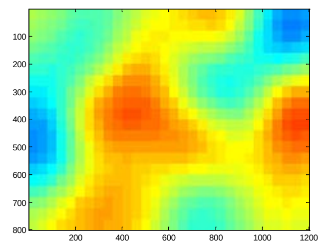
Inferred coarse-scale F fields and FWHM values provide information necessary to create fine-scale binary fields



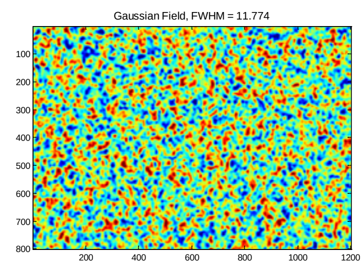
F field



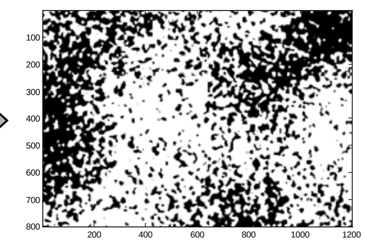
Z field



MG field



Binary field



$Z = (-1.0)*G^{-1}(f;0,1)$

If (MG - Z > 0.0), Binary = 1, else 0

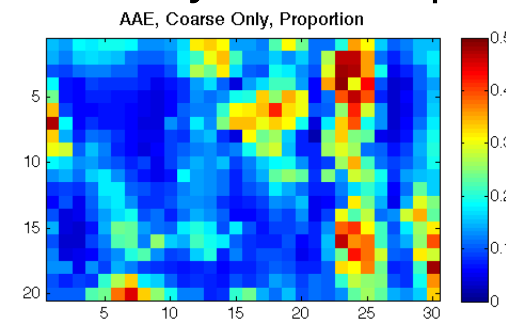
Coarse scale estimation provides the proportion of high permeability material within each coarse cell

Convolution of fine-scale uncorrelated field with estimated kernel produces smoothly varying field that is truncated to a binary field by Z-field

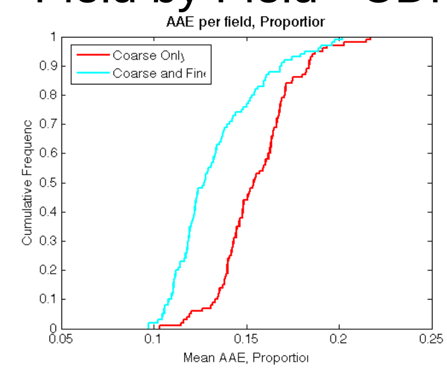
Performance Measures

	Coarse Scale	Fine Scale
Coarse Data Only	Cell by cell estimation of true F and true K Field by field estimation of true F and K	Cell by cell estimation of median travel time and over/under estimation Field by field estimation of median travel time and travel time distribution
Coarse and Fine Data	Cell by cell estimation of true F and true K Field by field estimation of true F and K	Cell by cell estimation of median travel time and over/under estimation Field by field estimation of median travel time and travel time distribution

Cell by Cell = Map



Field by Field= CDF



Coarse Data Only = 20 coarse K measurements

Coarse and Fine Data = 20 coarse K measurements and 20 fine-scale travel times

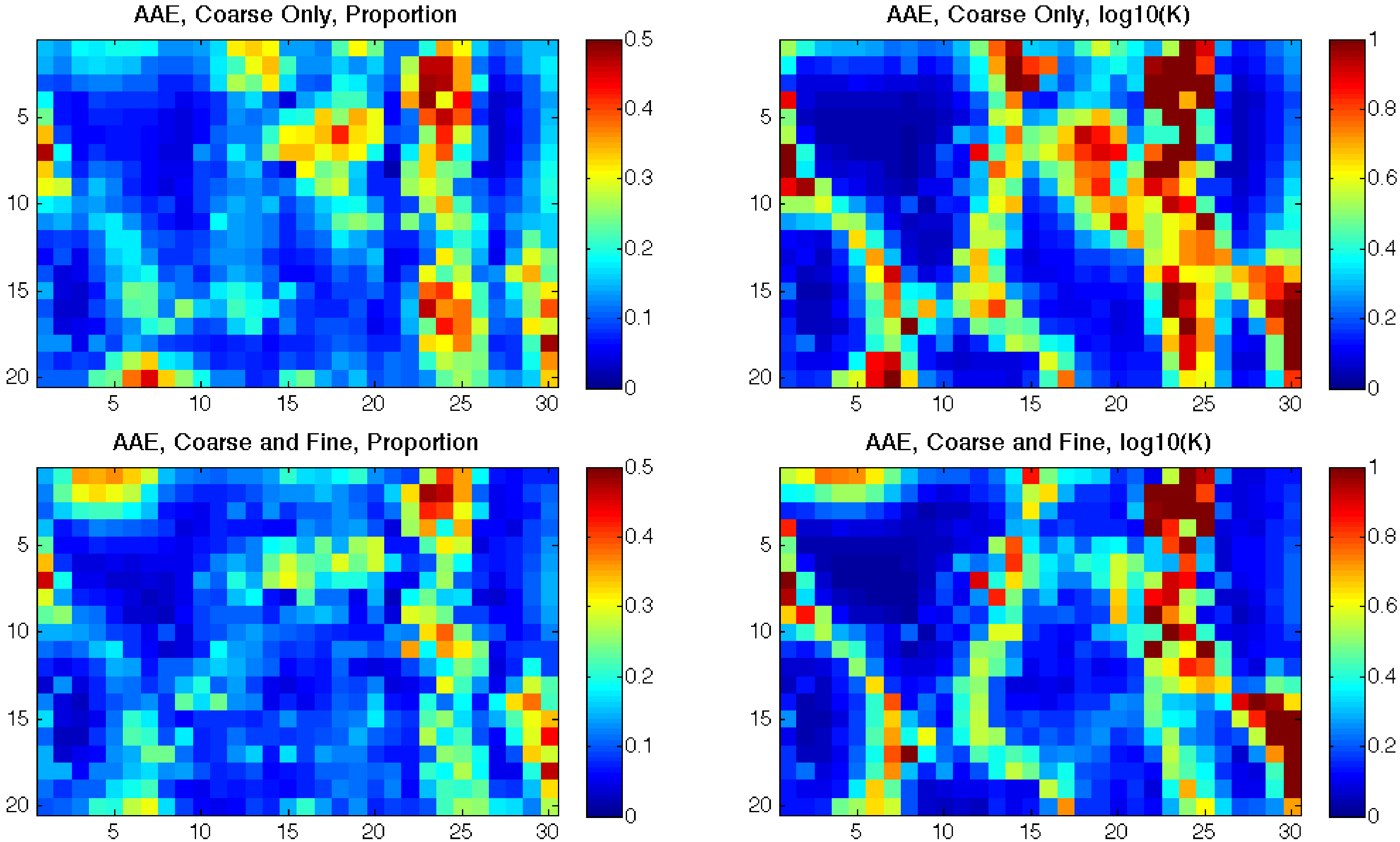
Coarse Field Estimation

Proportion Errors

Log10 (K) Errors

Coarse Data Only

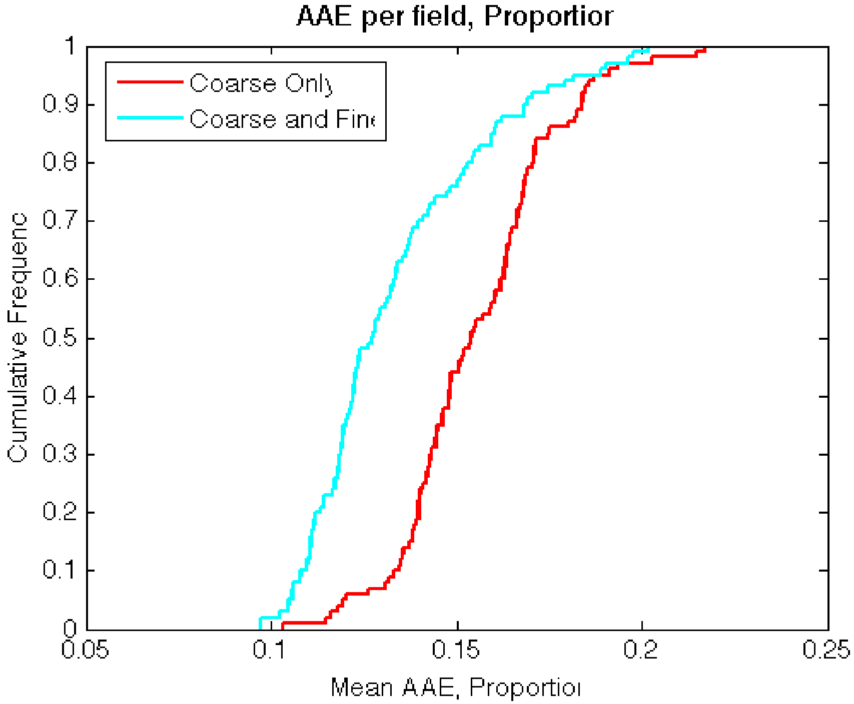
Coarse & Fine Data



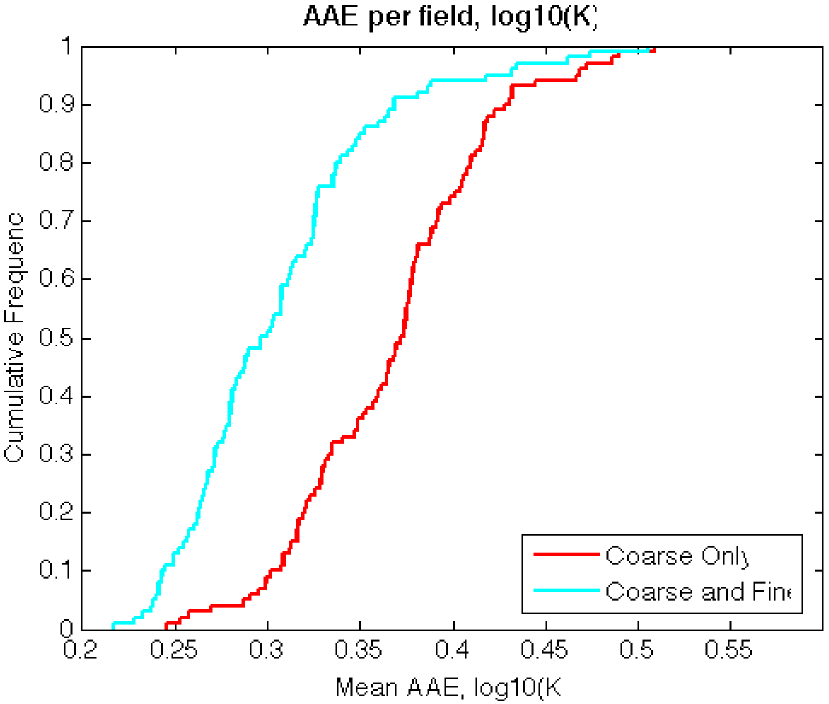
Coarse scale performance across 100 realizations evaluated at every cell

Coarse Scale Evaluation

Proportion AAE per field

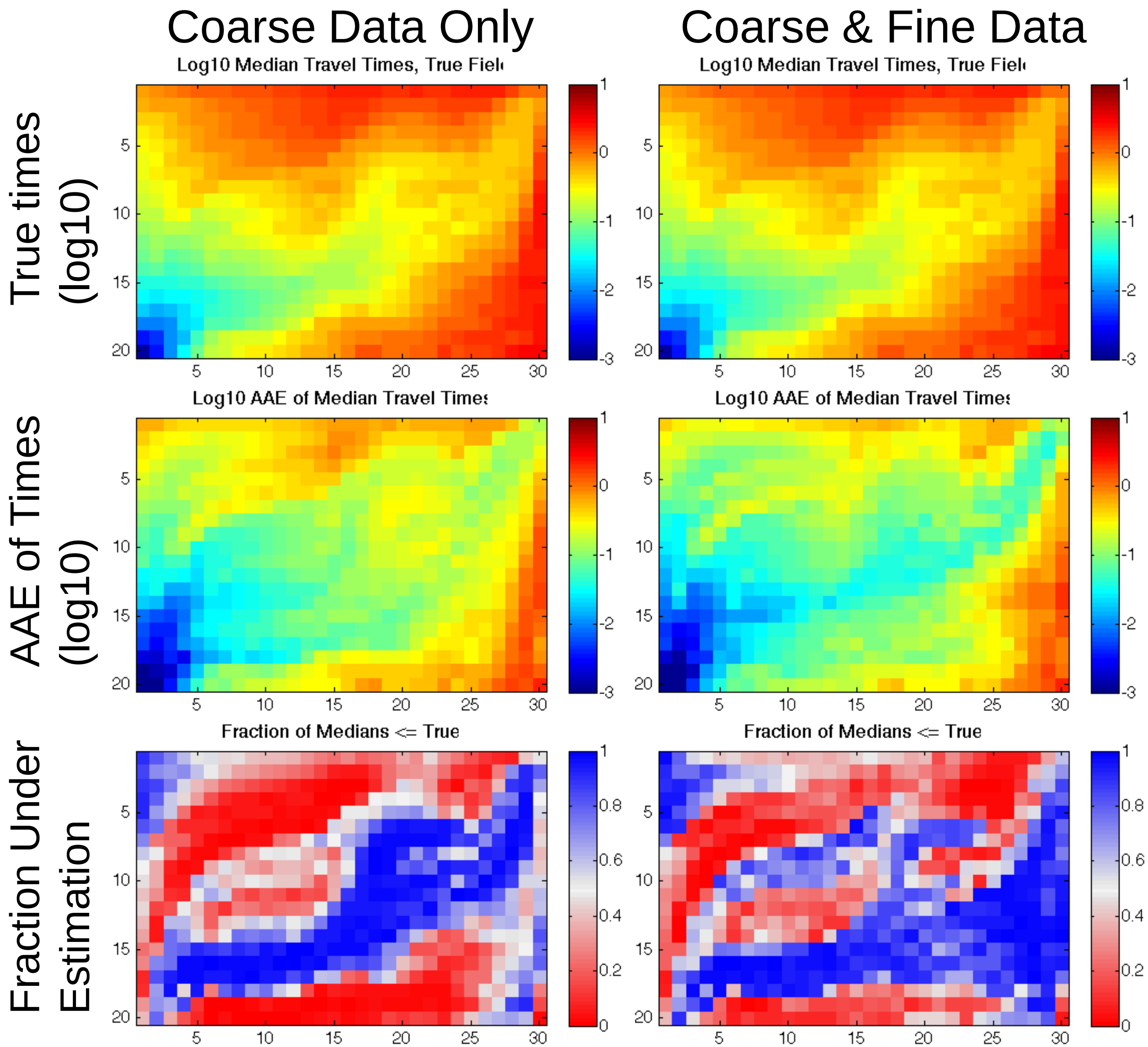


Log10(K) AAE per field



Coarse scale performance across 100 realizations evaluated for every field

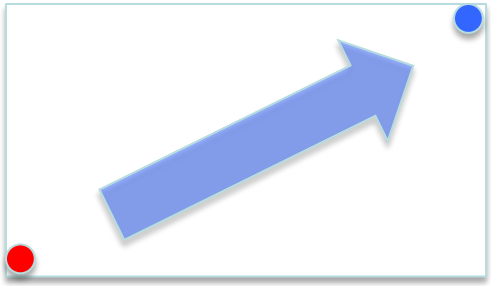
Median Travel Time Estimation



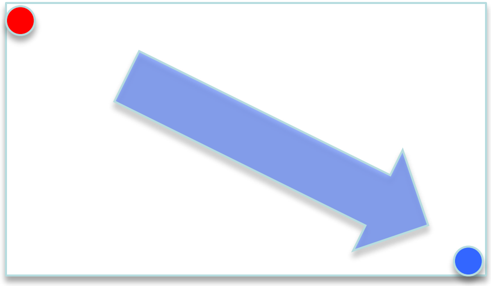
Switching Flow Direction

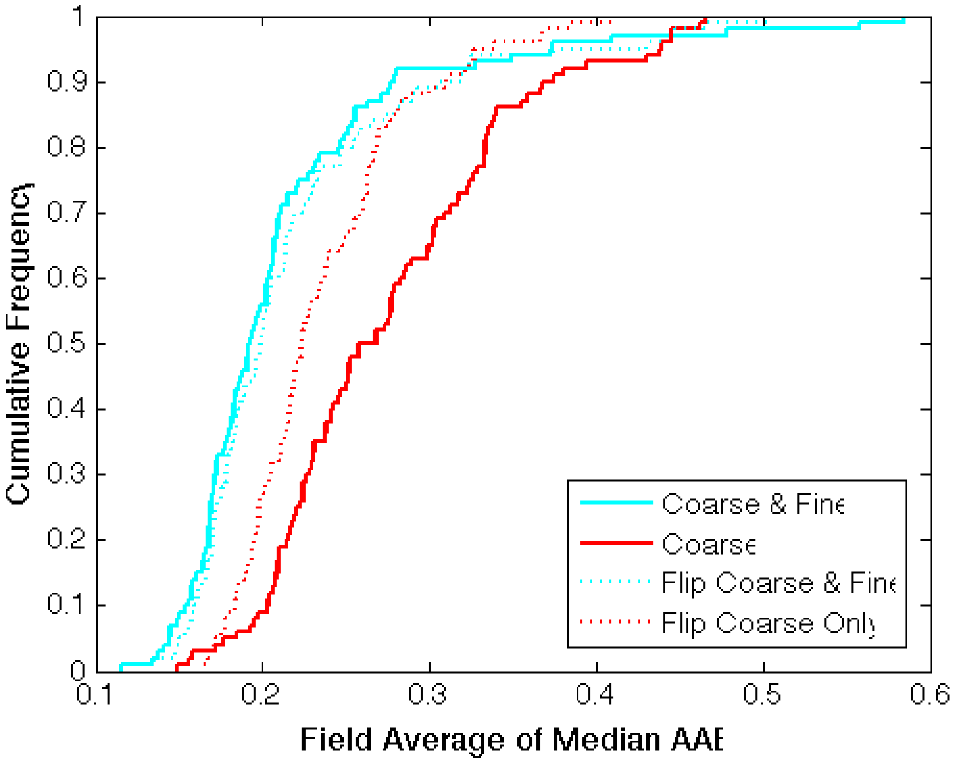
Distributions of the spatial average of the AAE of the median times (one value per realization)

Original Configuration



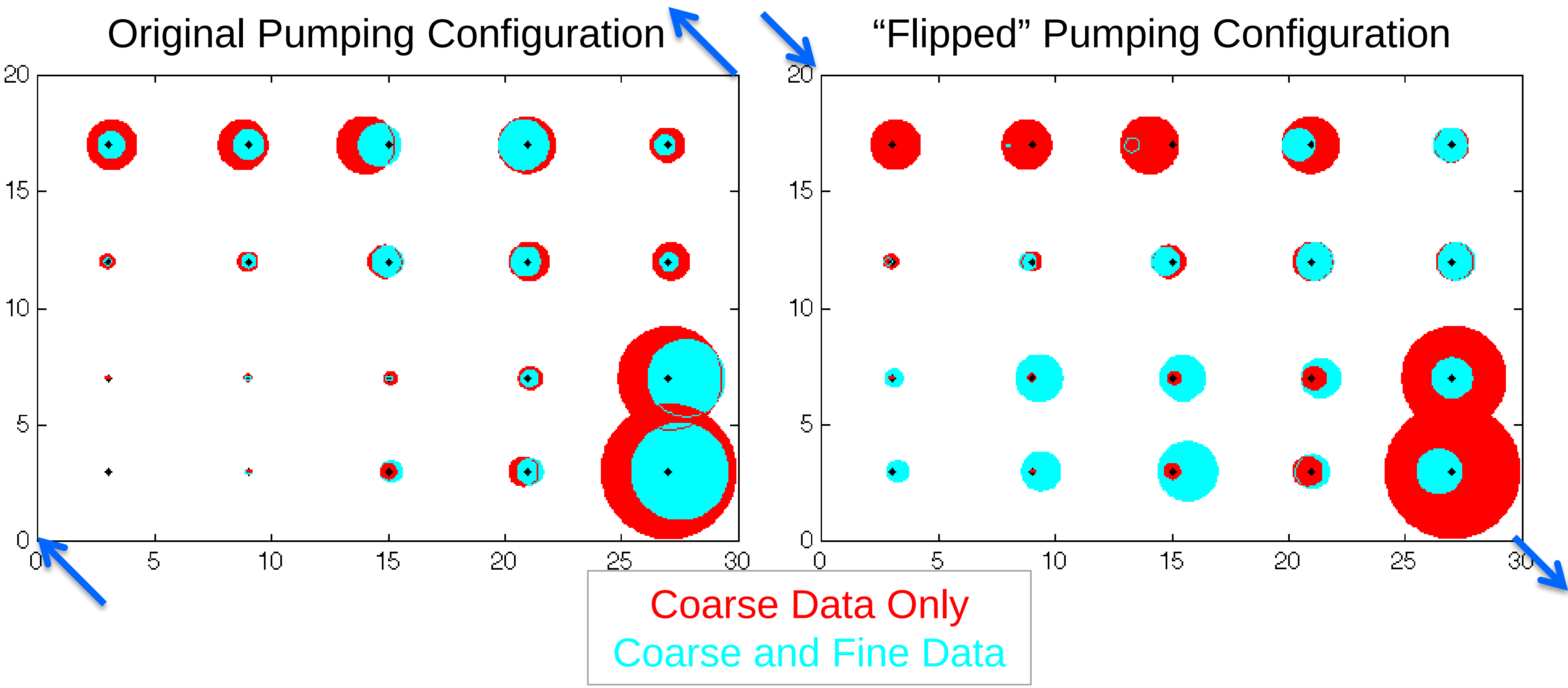
Flipped Configuration





Adding fine-scale data maintains small travel time error even for scenario of flipped source and sink locations

Accuracy and Precision at Sensors



Circle Radius = 95% Empirical CI in units of normalized time (pore volumes injected)

All distributions are accurate for the original case

All distributions using Coarse are accurate for the "flipped" case

16 of 20 are accurate when Coarse and Fine data are used

For almost all locations in both cases, adding fine-scale data decreases the CI width

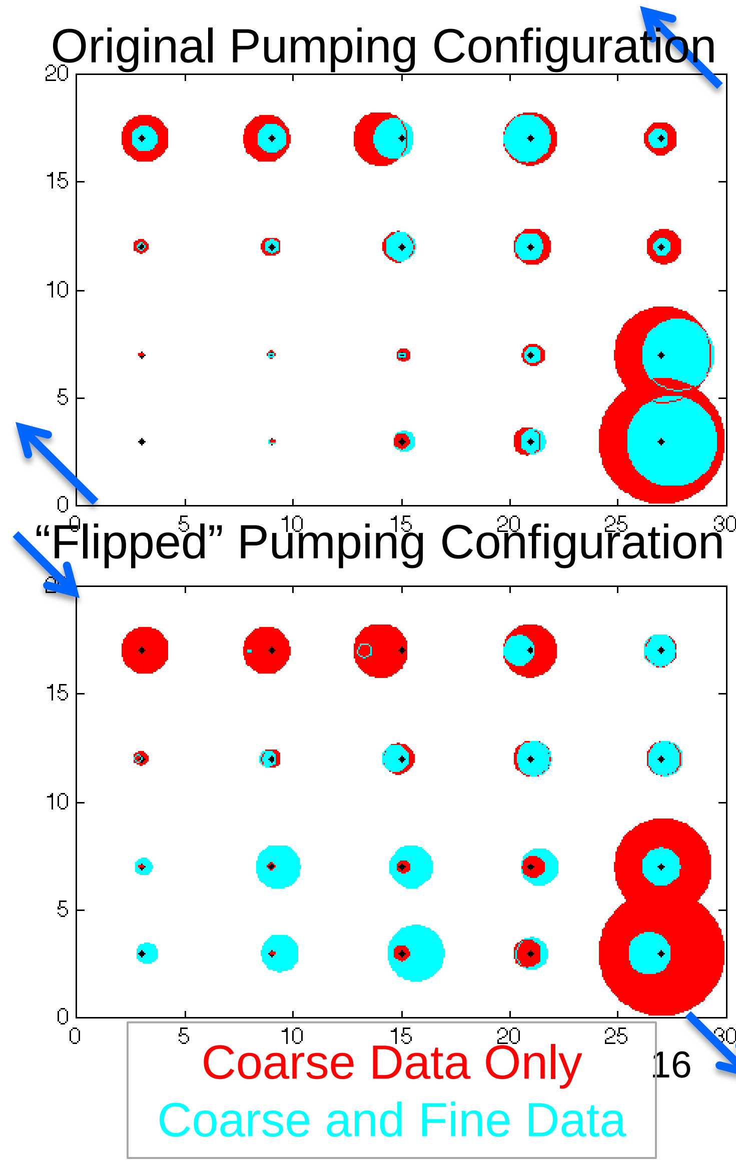
Another Look

What causes decrease in variability when fine-scale data are added?

Representation of coarse field with 30 KL coefficients is excessive – only the first 10-15 KL coefficients have posterior distributions that differ from priors. Coarse data don't impact fine-scale variability

Adding fine-scale data changes things – all 30 posteriors are significantly different than priors

Inversion with either data set is robust to changes in the locations of the source and sink



Summary

- Demonstrated approach to multi-scale stochastic inversion
  - Computationally feasible by constraining Bayesian estimation to coarse scale and limiting estimated parameters with KL decomposition
  - Link function designed to work on binary media and incorporate inclusion size directly
  - Posterior distributions are accurate (all) and precise (Coarse & Fine data)
  - Estimations are robust w.r.t. to change in flow
- Future Work:
  - Incorporate increased resolution of link function error

Posterior Predictive Modeling Using Multi-Scale Stochastic Inverse Parameter Estimates

Sean A. McKenna
Geoscience Research and Applications Group
Sandia National Laboratories, Albuquerque, New Mexico
samcken@sandia.gov

Jaideep Ray
Predictive Simulation R&D Department
Sandia National Laboratories, Livermore, CA

Bart van Bloemen Waanders
Applied Math and Applications Department
Sandia National Laboratories, Albuquerque, New Mexico

Youssef Marzouk
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology, Cambridge, Massachusetts

This work was funded by Sandia National Laboratories Laboratory Directed Research and Development program. 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-AC04-94-AL85000