SAND2015-8935PE

Bayesian inference of the permeability field of a binary medium

Jaideep Ray

Sandia National Laboratories, Livermore CA

jairay@sandia.gov

https://www.sandia.gov/~jairay

With contributions from

S. McKenna, S. Lefantzi and B. van Bloemen Waanders

Introduction

- Estimation of fields (2D/3D) from limited data quite common
  - E.g., estimation of material properties
- Limited data
  - Could be a property of the the field measured at a few locations and/or
  - Observation of a dynamical process affected by the material
    - E.g. flow breakthrough times, modulation of waves through it etc.
- Estimation of properties often posed as a model-fitting problem
  - Called an inverse problem
  - Requires a model or parameterization of the field being inferred from data
    - Called the Random Field Model (RFM)
    - Often has a large number of parameters
  - A model that, given a realization from the RFM, simulates the dynamics being observed
    - Called the forward problem
    - Can be computationally expensive

Uncertainties in inferences

- Because of limited observations, the (large number of) parameters in RFM cannot be estimated accurately
  - Also, a priori, we don't know the subset of parameters that can be informed by the observations
- Two ways out:
  - Way # 1: Keep all parameters in the RFM and estimate them using Bayesian inference
    - Bayesian inference estimates parameters as a joint probability density function (PDF); capture the uncertainty very concisely
    - Requires one to pose a statistical inverse problem
    - Require sampling methods such as Markov chain Monte Carlo (MCMC) to solve
    - Forward model invokes $O(10^4)$ – $O(10^6)$ times
  - Way # 2: Use shrinkage regression to simplify the RFM (set "un-estimatable" parameters to zero / nominal / prior-belief values
    - Does not capture the uncertainty in the parameters that *are* estimated

Bayesian estimation of a binary medium's permeability field

J. Ray, S. A. McKenna, B. van Bloemen Waanders and Y. M. Marzouk, "Bayesian reconstruction of binary media with unresolved fine-scale spatial structures" in *Advances in Water Resources* , 44:1--19, 2012.

S. A. McKenna, J. Ray, Y. Marzouk and B. van Bloemen Waanders, "Truncated multiGaussian fields and effective conductance of binary media", in *Advances in Water Resources* , 34:617-626, 2011.

The problem setup

- Aim: Given a material with spatially variable properties, estimate structural properties at all scales from *sparse measurements*
- Slight relaxation:
  - Need to know large-scale variations/structures accurately
  - Need to know statistics of the fine structures
- Given: measurements/observations which are impacted by both the fine & coarse structures
- Why? Materials with random & multiscale structures abound and cannot be imaged/measured at all scales
  - Geophysical materials are random & multiscale (geological strata, soil properties etc)
  - Mesoscale O(1μ) electrochemical & catalytic processes at fuel cell anodes
  - Material degradation/aging – e.g., "bubbles" in explosive "cook-off"

Challenges in estimation

- *Never enough data to infer fine & coarse scales simultaneously*
  - If possible to observe / image all scales, why bother to infer anything?
  - Corollary: inferences are always done with incomplete data
- Most inferential methods are iterative
  - Propose, compare with observations, reject/accept
  - Involve a forward model that links the objects of inference with the observables
- So even if a gigantic model resolving all scales is available, can't be used in a inferential setting (aka inverse problem)
  - Takes too long
  - Plus, never enough observables to inform the gigantic model's gigantic d.o.f
- *Net result: Inferences are always uncertain*
  - Due to the use of simplified models and incomplete observations
  - So how to capture the uncertainty?

Inference in a binary medium

- Given: A porous medium with 2 phases
  - A low permeability matrix
  - With fine, high-permeability inclusions
  - Inclusions are unevenly distributed in the domain
  - Domain is rectangular – 1.5 x 1.0
- Scale separation: Impose a 30 x 20 grid on domain
  - Inclusions are 1/10th the grid-block size
    - fine scale variable, δ
  - Each grid-block has an inclusion proportion (***F(x)***)
    - Resolved on the 30 x 20 mesh; coarse scale variable
- Impact: Permeability in a grid-block affected by both fine- and coarse-scale variables
  - $\boldsymbol{k} = \mathcal{K}_{\mathit{eff}}(\, \boldsymbol{F(x)},\ \delta\,)$

Figure of F(x) with mesh

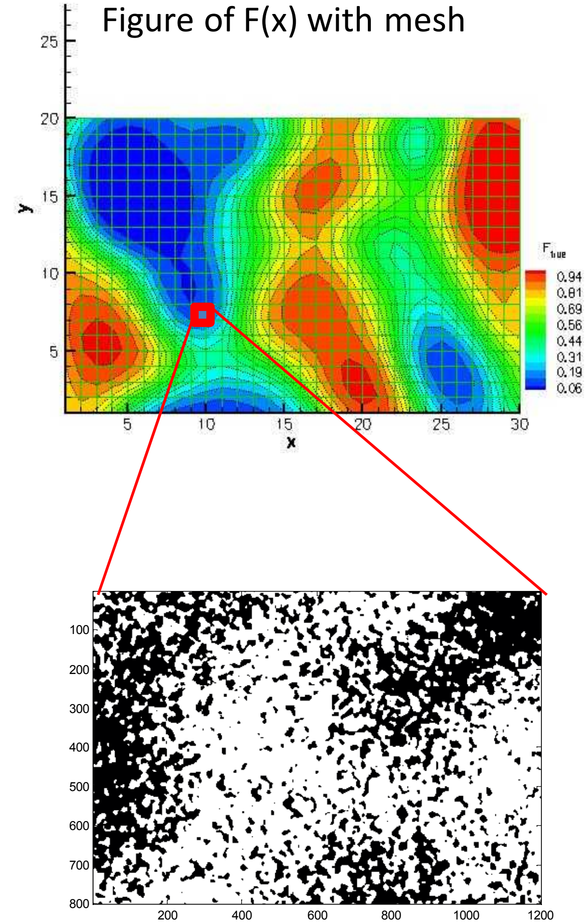
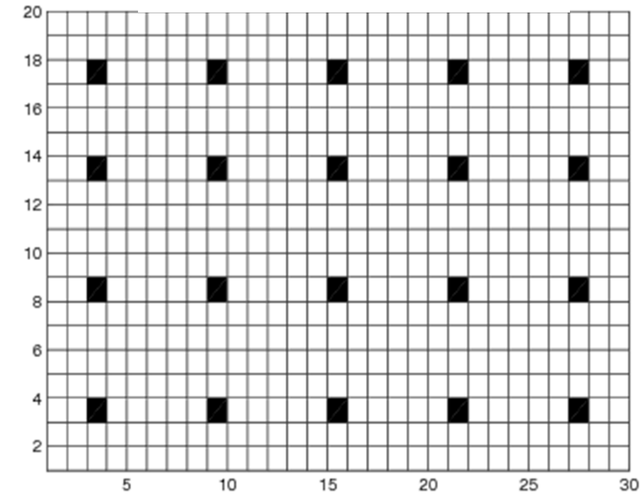


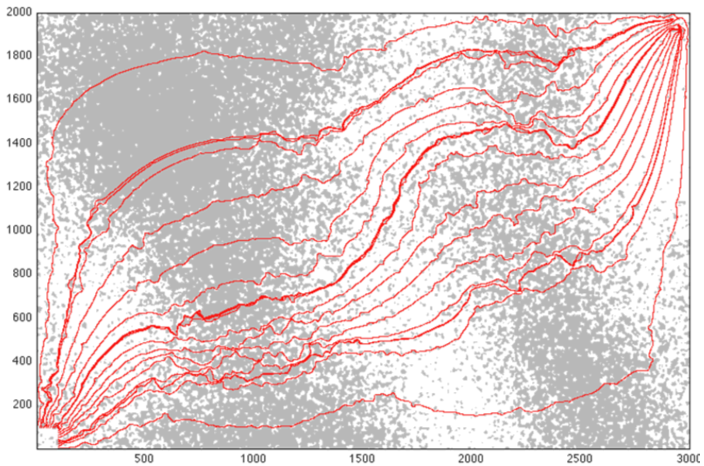
Figure of inclusions (white) in a grid-box

Informative observations

- Consider a set of 20 grid-blocks with sensors
  - $\{k^{obs}\}$ given info on $\{F, \delta\}$ at the sensors
  - OK for inferring structures > inter-sensor spacing
- Water-flood experiment for finer structures
  - What is this?
  - Inject water at one corner, pump it out at the diagonally opposite corner
  - Flow impacted by structures at all scales
  - Water breakthrough time at sensors $\{t^{obs}\}$ contain the integrated impact of multiscale structures
- Teasing out the contributions of the fine- and coarse-scale to $\{t^{obs}\}$ could allow inference of both scales
  - But how?

Location of sensors





Picture of pathlines through the binary medium. Inclusions in white

Recap, and an idea for inference

- Permeability ***k(x)*** = $\mathcal{K}_{eff}$ (***F(x)***, δ)
  - But we don't know what the functional form of $\mathcal{K}_{eff}$ is
- Breakthrough time ***t*** = $\mathfrak{M}$ (***k(x)***)
  - But we have only 20 measurements of **t, {*t*^{obs}}**
  - And 30 x 20 = 600 grid-blocks of unknown ***F*** and δ
- The idea
  - Model #1: Develop a "pointwise" model for ***k*** = $\mathcal{K}_{eff}$ (***F***, δ) in a grid-block
    - Subgrid model
  - Model #2: Develop a parameterized model for ***F*** to describe its spatial variation
    - Have a about 20 – 30 parameters in it – reduced order modeling of ***F(x)***
  - With 20 {***k***^{obs}} and 20 {***t***^{obs}}, should be able to infer all unknowns
    - 20-30 parameters for ***F(x)*** and one δ
- Caution
  - With 40 observations, none of these parameters will be estimated well
    - Fine, but how inaccurate are the estimations?

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

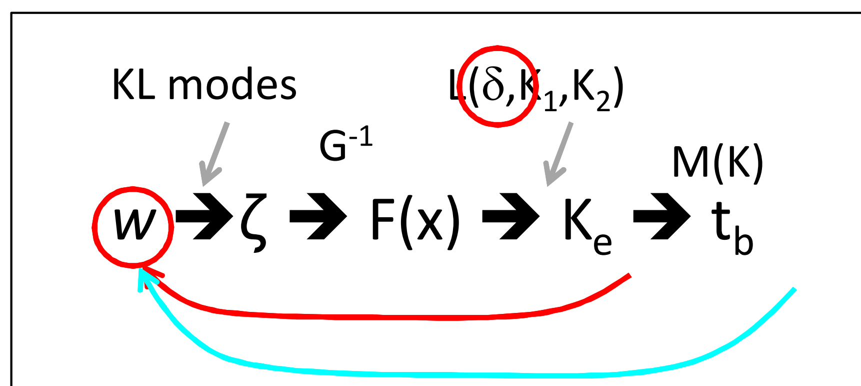
$$K_e = L(F(x), \delta, K_1, K_2)$$

Link function provides K at the coarse scale

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

Flow model operating on fine scale K provides travel times

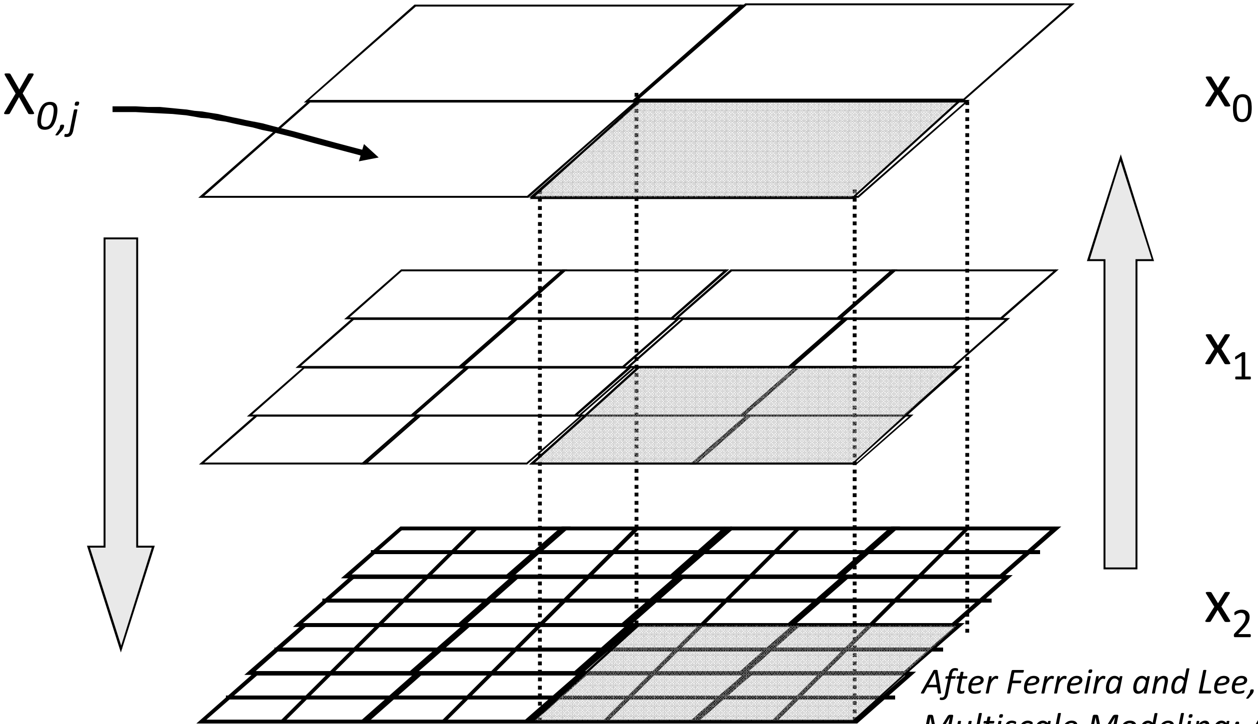
$$d_i = \{K(x)^0, t_b^1\} \qquad i = 1, \ldots, Ns$$

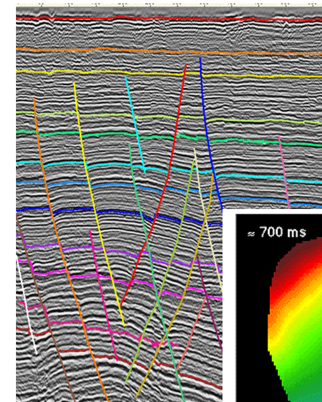


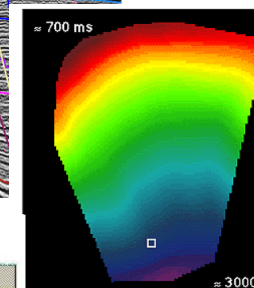
Model # 1 – a sub-grid model

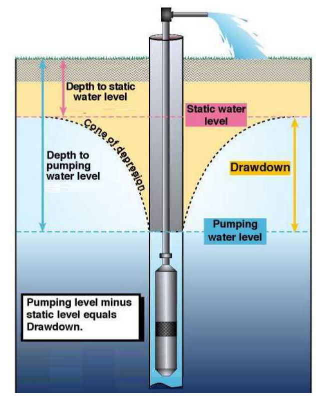
Data collected at one level informs values at other levels

Multiscale random fields with averaging "link" between them

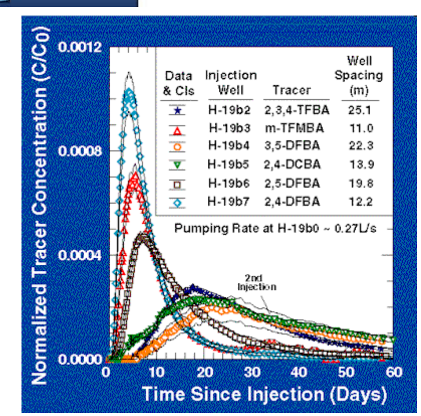








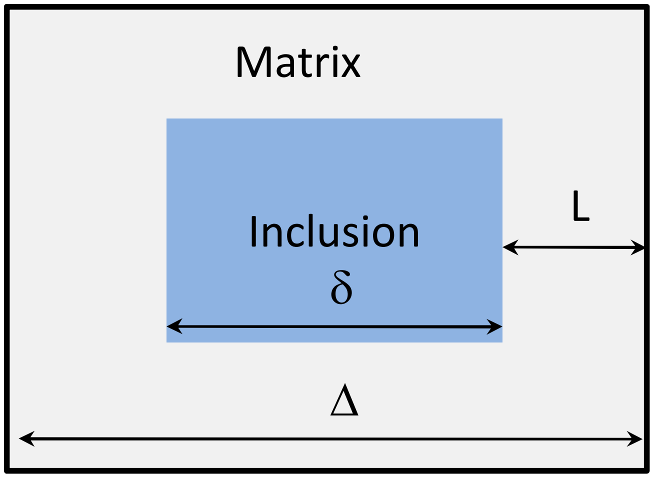
After Ferreira and Lee, 2007, Multiscale Modeling: A Bayesian Perspective



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

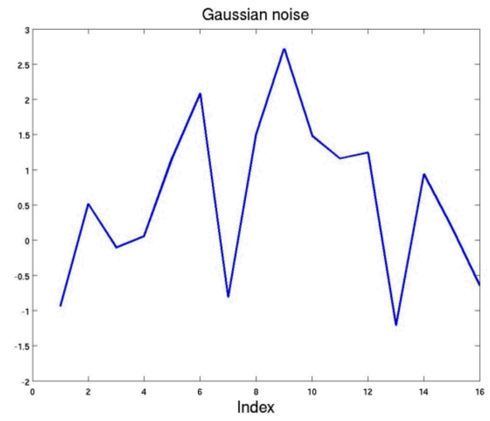
Upscaling problem

- We need: $k = \mathcal{K}(F, \delta)$
- Knudby's theory, restricted to rectangular inclusions of size d
  - $k = \mathcal{K}_{\text{Knudby}}(F, \delta, L/\Delta)$
    - L = flow path in the matrix
- Problem: Our inclusions are arbitrarily shaped
- Questions:
  - Can we create a field of arbitrary inclusions, given F and δ?
  - Can we find L in such cases? Just the expected value.
  - Can we do so analytically, without actually creating a field and instantiating an inclusion-in-matrix field?
- Subgrid modeling, but solely geometric

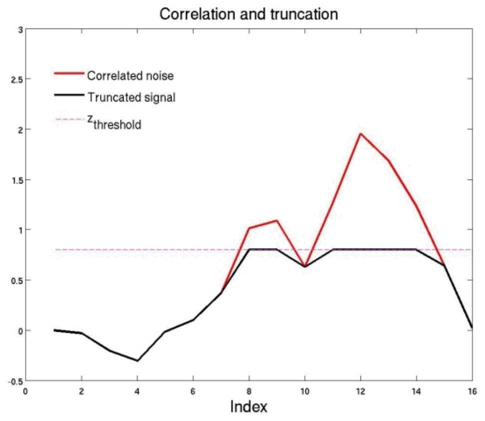


Subgrid geometric modeling

- Consider a grid-block divided into 100 x 100 *grid-cells*
- Initialize a 100 x 100 white-noise field
- Convolve with a Gaussian kernel with FWHM of δ
  - Creates a correlated field with correlation length δ
- Truncate at a level $z_{threshold}$
  - Flat sections are inclusions!
  - $Z_{threshold}$ decides the inclusion proportion ***F*** in the grid-block
- The theory of truncated pluriGaussian fields provides analytical expressions for expected values
  - Number of inclusions
  - Total area in the inclusions
  - These are explicit functions of F and δ



1d white noise field



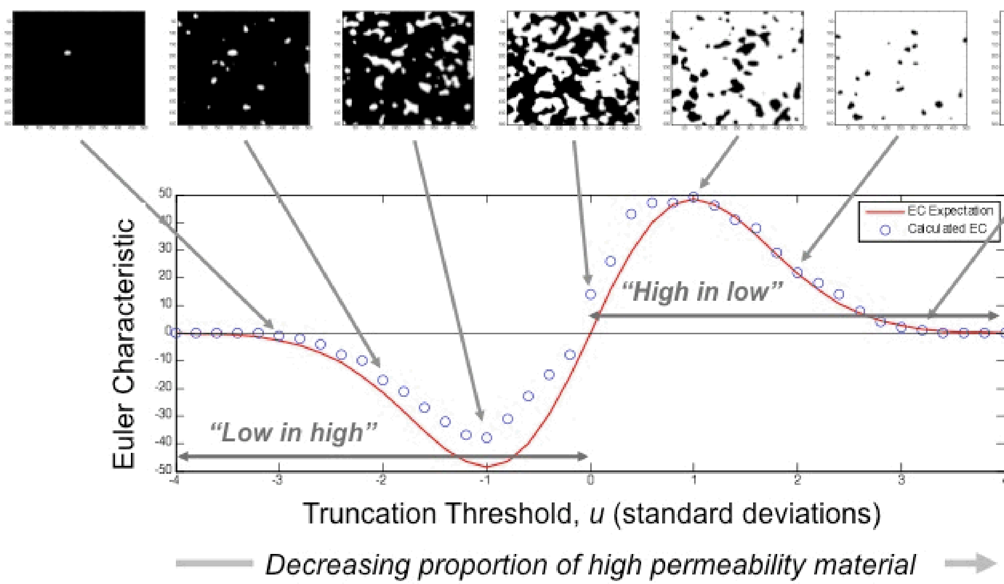
Truncated, correlated field

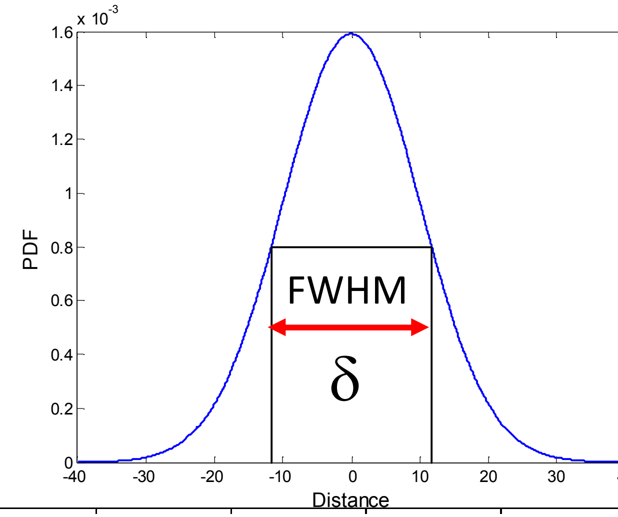
Subgrid upscaling with Knudby

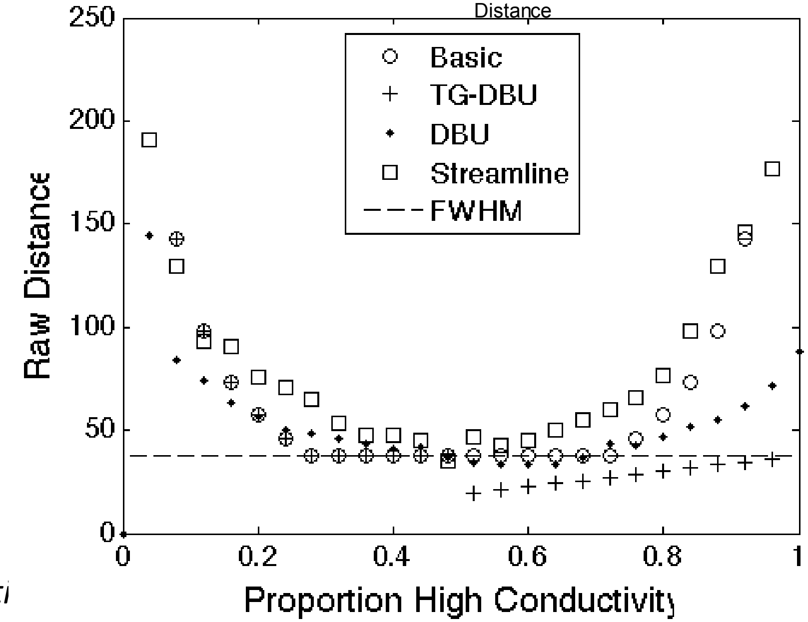
- If {***F***, δ} specified for each grid-block, we can analytically predict
  - Number of inclusions and total area of the inclusions
  - Ditto, area per inclusion
- Assume that the inclusions are round
  - Inclusion radius can be calculated
- Assume that the centroids of the inclusions are distributed per a Poisson point process
  - Expected value of inter-inclusion distance obtained
- Expected value of flowpath length in matrix *L* can be calculated
- Plug into $\mathcal{K}_{\text{Knudby}}$ and you're done
  - Not quite, but that's the rough outline of the subgrid model

Linking Function

Binary mixtures are modeled using truncated Gaussian fields







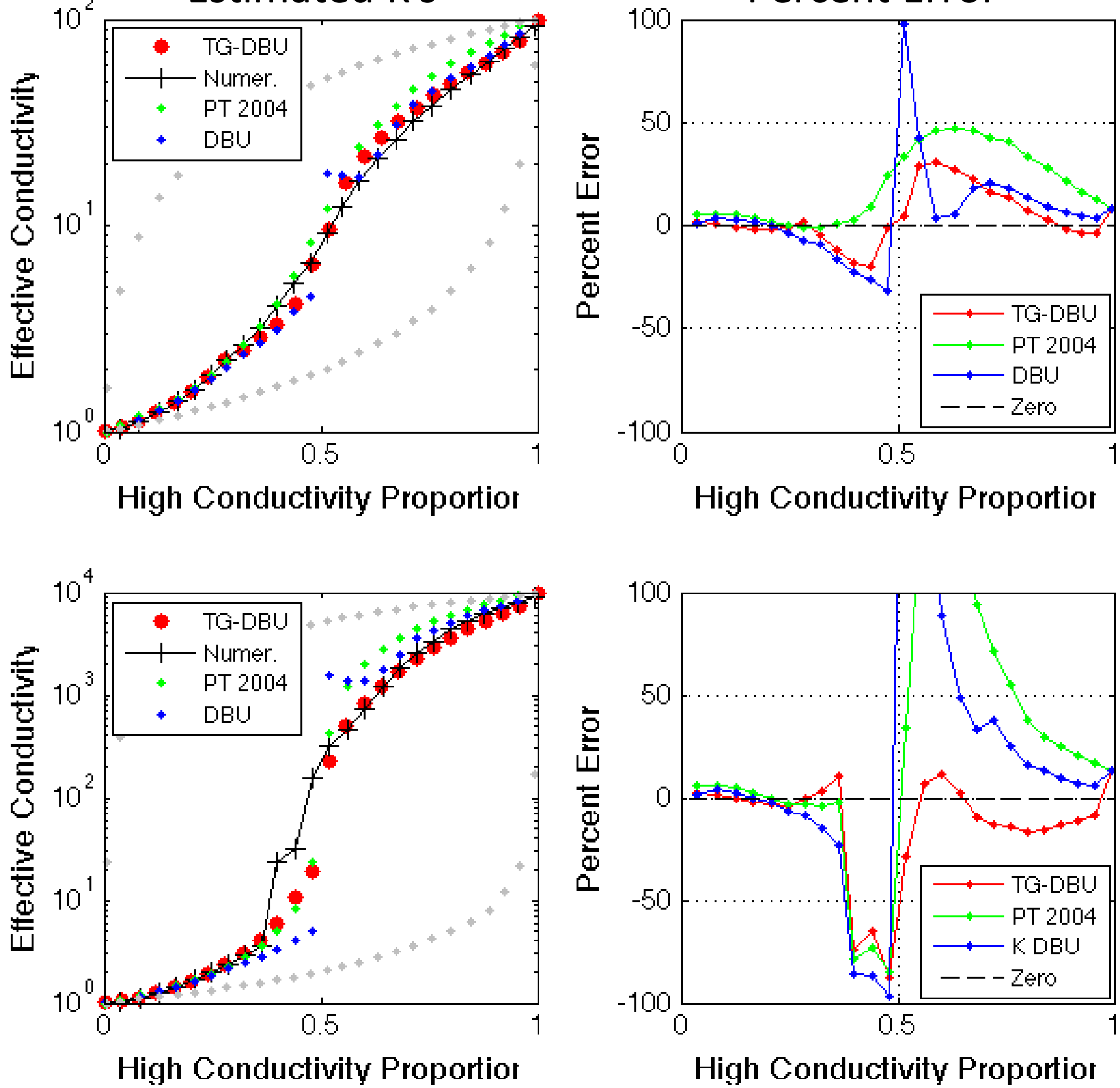
New upscaling function uses proportions (tied to truncation threshold) and average estimated distances between inclusions to estimate upscaled effective permeability

McKenna, et al., (in review), Truncated MultiGaussian Fields and Effecti
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*

Model #2: Reduced order modeling of *ζ(x)*

- ***ζ(x)*** varies in space and is described on a 30 x 20 mesh
  - Don't want to infer all 600 values
  - But ***ζ(x)*** is smooth – can't we exploit this to make a lower-dimension model?
- Model ***ζ(x)*** as a 600 variate Gaussian
  - Smoothness guaranteed
  - Assume correlation function known (~ exp(- x^2)) i.e. covariance Γ of multiGaussian is known
- Any multiGaussian can be expanded in a Karhunen-Loeve series
  - We'll truncate at 30 terms
  - ***Φ (x; Γ)*** are called KL modes; w_i are the weights

$$\zeta(x) = \sum_{i=1}^{30} w_i \sqrt{\lambda_i(\Gamma)} \Phi(x;\Gamma)$$

- Inferring ***ζ(x)*** means inferring w_i

Posing the inverse problem

- Given: $\{\boldsymbol{k}^{obs}, \boldsymbol{t}^{obs}\}$ at 20 sensors
- Models:
  - $\boldsymbol{F}$ = sum of KL modes with unknown weights w_i
  - $\boldsymbol{k} = \mathcal{K}_{\mathit{eff}}(F, \delta)$ – the subgrid model
  - $t = \mathfrak{M}(\boldsymbol{k(x)})$ - Darcy flow model, solved using finite-difference method
- Infer weights w_i, i = 1 … 30 and δ
  - Develop distributions for these quantities, not point values
- Generating synthetic $\{\boldsymbol{k}^{obs}, \boldsymbol{t}^{obs}\}$
  - Start with a “ground-truth” binary medium on a 3000 x 2000 mesh
  - Push water through it and measure breakthrough times at 20 sensors – $\{\boldsymbol{t}^{obs}\}$
    - Done with MODFLOW, a Lagrangian code distributed by USGS
  - Superimpose a coarse 30 x 20 mesh
    - Pick out the grid-blocks with sensors
    - Solve a 1D flow equation in each and estimate effective grid-block permeability – $\{\boldsymbol{k}^{obs}\}$

Bayesian Inverse Problem

- Objects of inference, $\Theta = \{F(x), \delta\} = \{w_i, i = 1 \ldots 30, \delta\}$
- Bayesian inverse problem

$$-2 \log \pi(\Theta) \propto \frac{\{\boldsymbol{t}^{obs} - \mathcal{M}(\Theta)\}^2}{\sigma^2{}_T} + \frac{\{\boldsymbol{k}^{obs} - \mathcal{K}_{eff}(\Theta)\}^2}{\sigma_K{}^2} + \frac{\{\Theta - \Theta_p\}^2}{\sigma^2{}_\Theta}$$

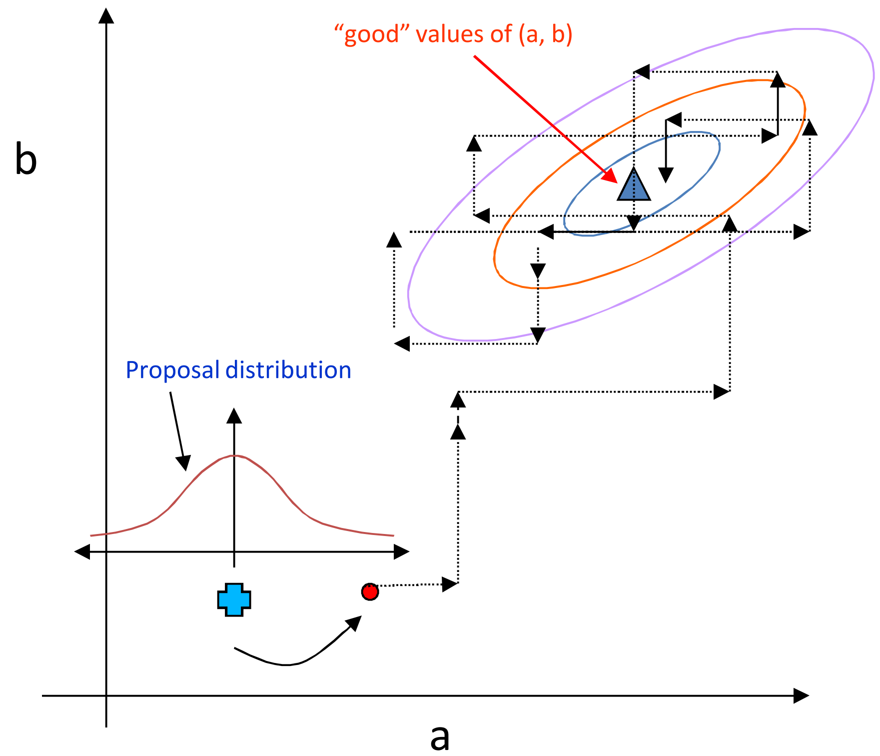
  - $\mathcal{M}$, Darcy flow model to relate Θ to breakthrough times $\{\boldsymbol{t}^{obs}\}$
  - $\mathcal{K}_{eff}$, subgrid model to relate Θ to observed permeability at certain sampling points
  - Θ_p, prior beliefs regarding the values of Θ
  - $\sigma_{\{K, T\}}$, std. dev. of various measurement errors
- $\pi(\Theta)$ evaluated by Markov Chain Monte Carlo sampling
  - Particular algorithm called DRAM

What is MCMC?

- A way of sampling from an arbitrary distribution
  - The samples, if histogrammed, recover the distribution
  - Given a starting point (1 sample), the MCMC chain will sequentially find the peaks and valleys in the distribution and sample proportionally
  - Drawback: Generating each sample requires one to evaluate the expression for the density π
- An example
  - Given: (Y^{obs}, X), a bunch of n observations
  - Believed: y = ax + b
  - Model: $y_i^{obs} = ax_i + b + \varepsilon_i$, $\varepsilon \sim \mathcal{N}(0, \sigma)$
  - We also know a range where a, b and σ might lie
    - i.e. we will use uniform distributions as prior beliefs for a, b, σ
  - For a given value of (a, b, σ), compute "error" $\varepsilon_i = y_i^{obs} - (ax_i + b_i)$
    - Likelihood of the set (a, b, σ) = $\Pi \exp(- \varepsilon_i^2/\sigma^2)$
  - Solution: $\pi (a, b, \sigma \mid Y^{obs}, X) = \Pi \exp(- \varepsilon_i^2/\sigma^2) * $ (bunch of uniform priors)

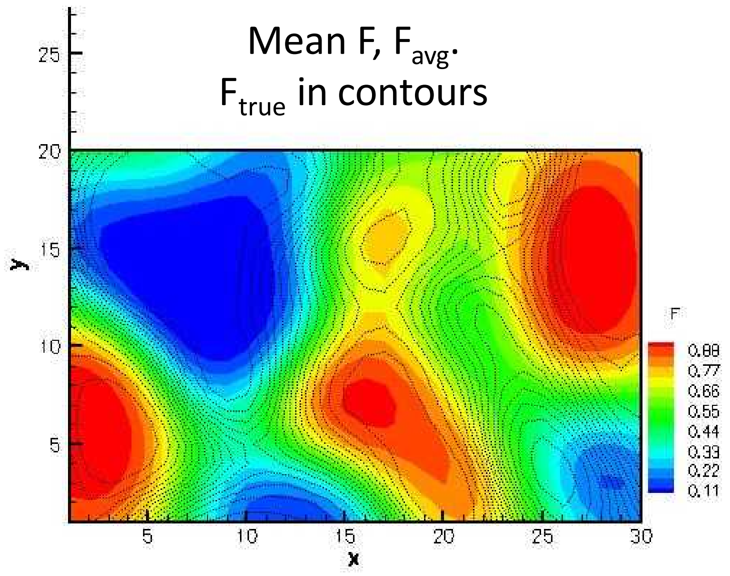
MCMC, pictorially

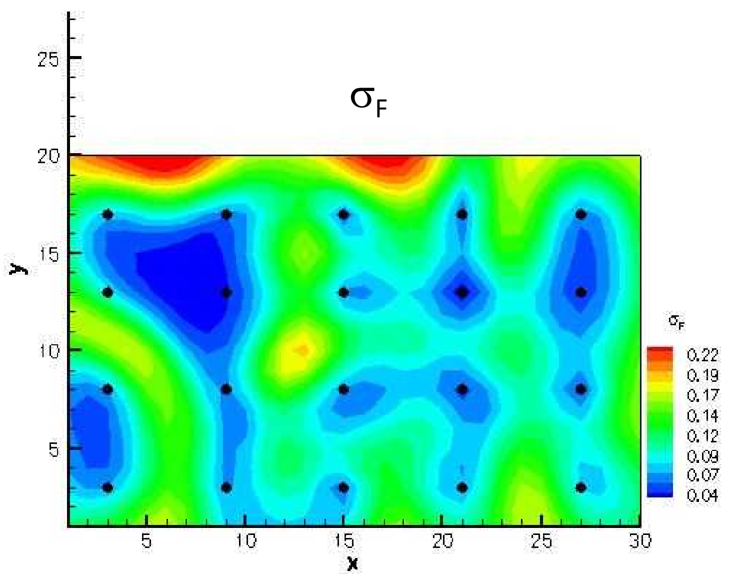
- Solution method:
  - Sample from π (a, b, σ | Y^{obs}, X) using MCMC; save them
  - Generate a "3D histogram" from the samples to determine which region in the (a, b, σ) space gives best fit
  - Histogram values of a, b and σ, to get individual PDFs for them
- Choose a starting point,
  - $P^n = (a_{curr}, b_{curr})$
- Propose a new a, $a_{prop} \sim \mathcal{N}(a_{curr}, \sigma_a)$
- Evaluate π (a_{prop}, b_{curr} | …) / π (a_{curr}, b_{curr} | …) = m
  - Accept a_{prop} (i.e. a_{curr} <- a_{prop}) with probability min(1, m)
- Repeat with b
- Loop over till you have enough samples

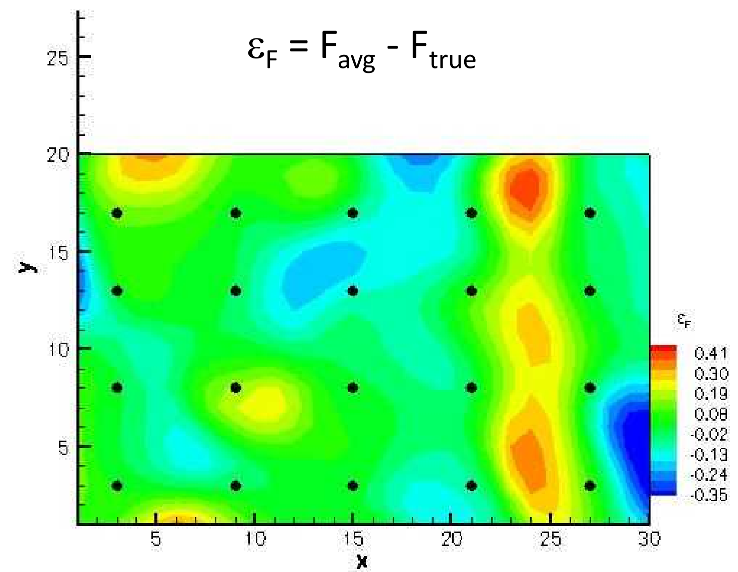


Results

- Get 10^4 samples of $\{w_i, \delta\}$
- From each $\{w_i, \delta\}$, develop 10^6 instances of ***F(x)*** and $\mathcal{K}_{eff}($ ***F(x)***, δ $)$
- Take the mean & std dev of the 10^6 ***F(x)*** instances
- Take standard deviations too

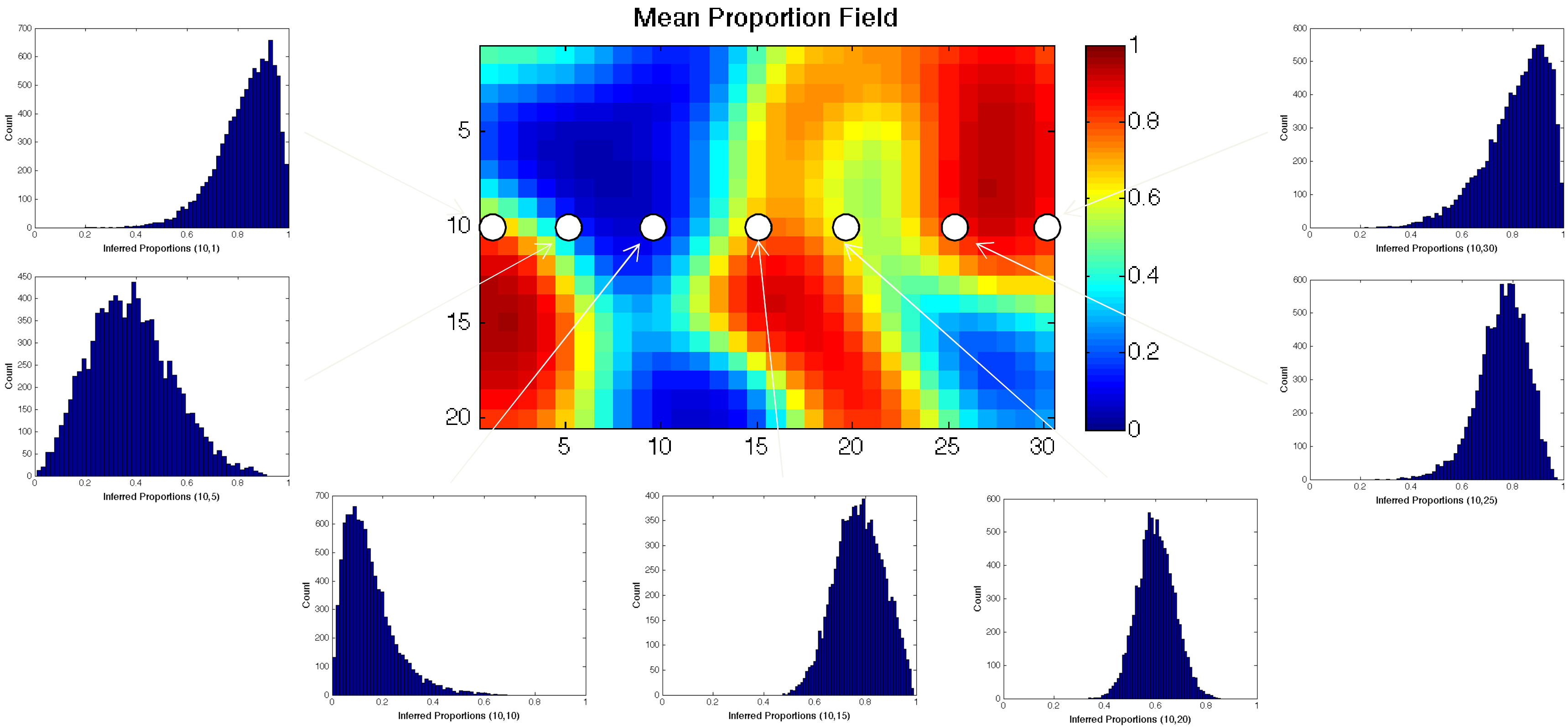






Estimated Proportion (F) Fields

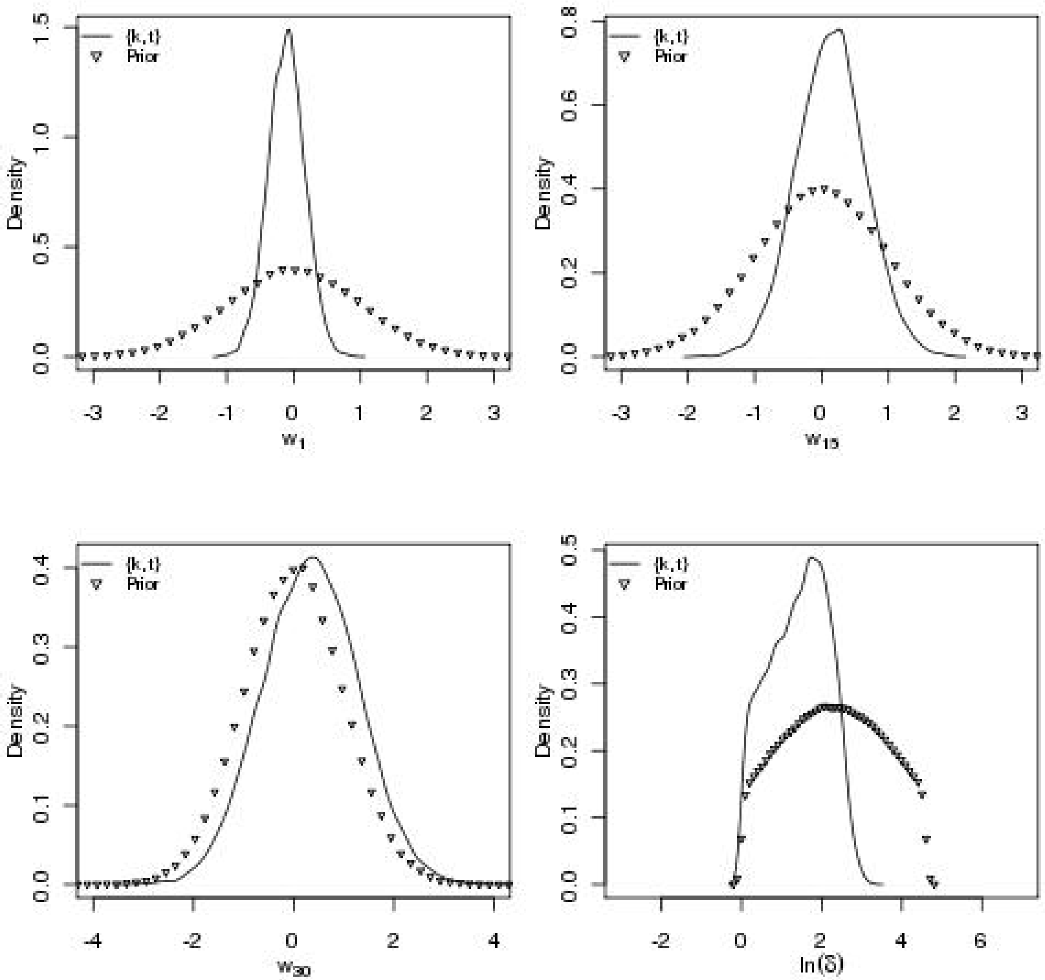
- MCMC runs met convergence diagnostics
- Results obtained with 1,500,000 iterations
  - Approximately 50 hours on workstation
  - Results in 9500 realizations of proportion field



Comparison of posterior pdfs for seven points on proportion field

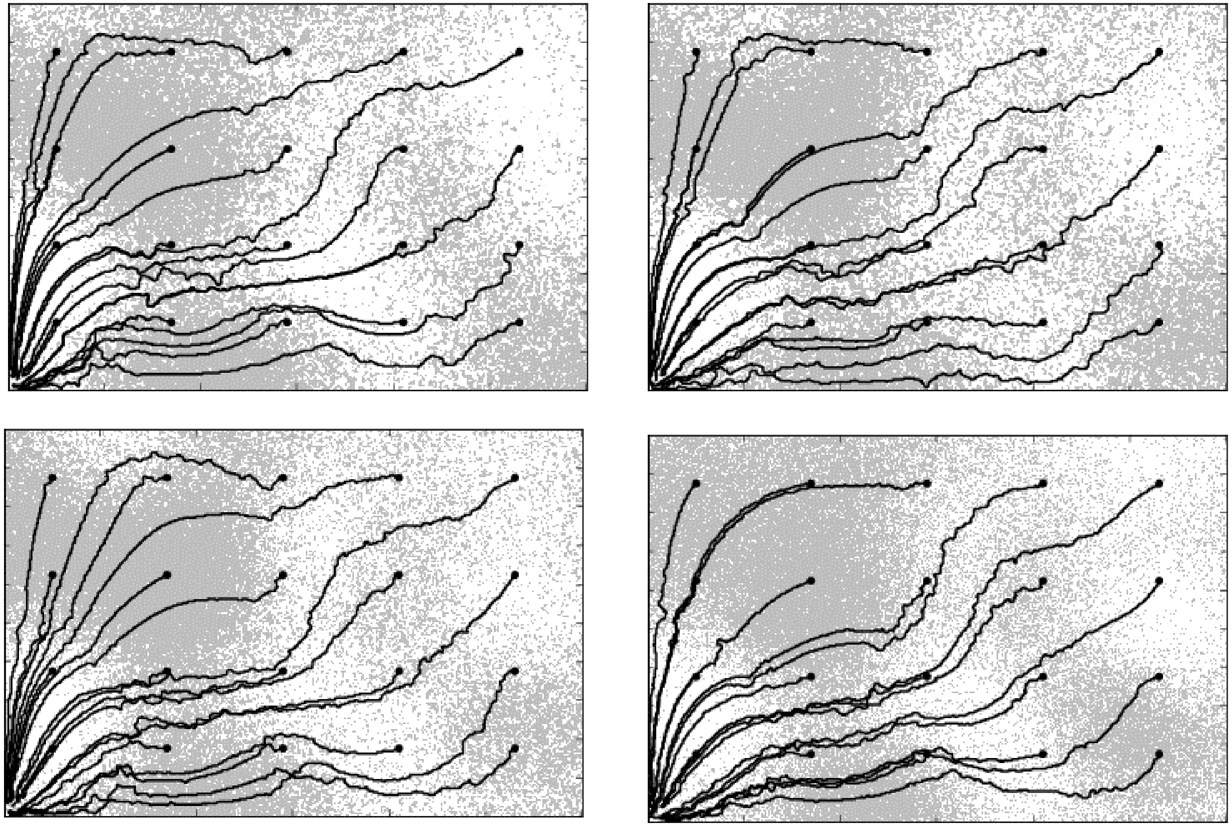
PDFS of $\{w_i, \delta\}$

- Use the 10^4 samples of $\{w_i, \delta\}$ to develop PDFs
- Take w_1, w_{15} and w_{30} as proxies for large, medium and small (but resolved) scale variations
- Inversions performed with $\{k^{obs}\}$ only also plotted
- Takeaways:
  - Large-scale structures easy to infer
  - Gets harder as we get smaller
  - Doesn't apply to inclusions



Developing fine-scale realizations

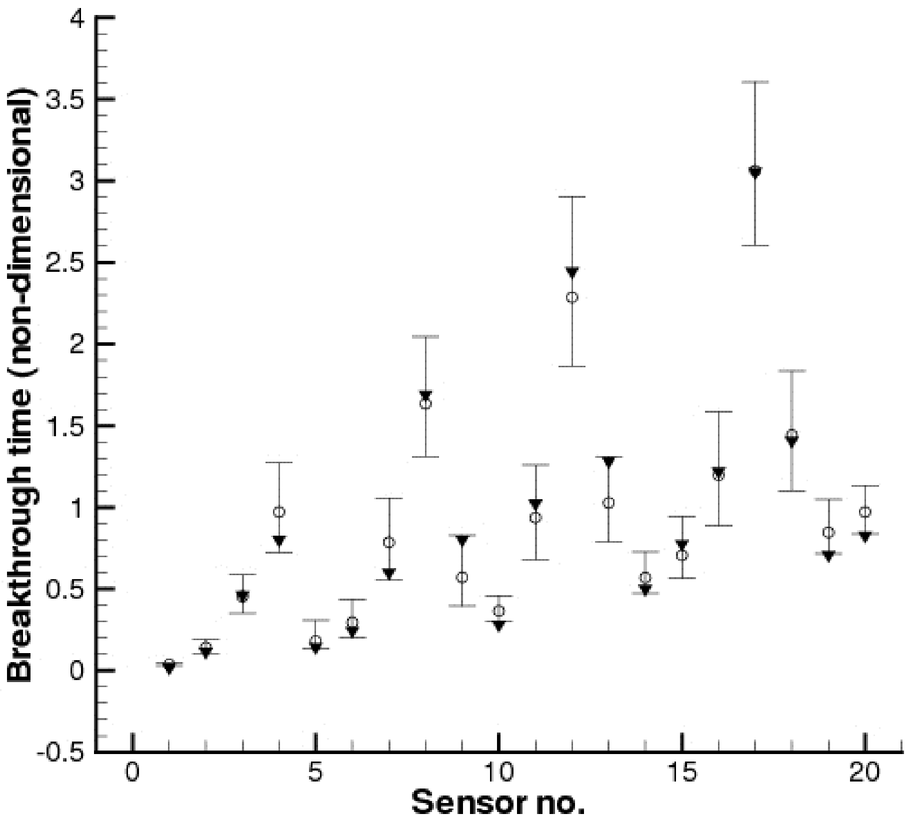
- The inferences can be used to develop fine-scale binary media



- Flow simulations can be used to obtain an ensemble of predicted breakthrough times at sensors

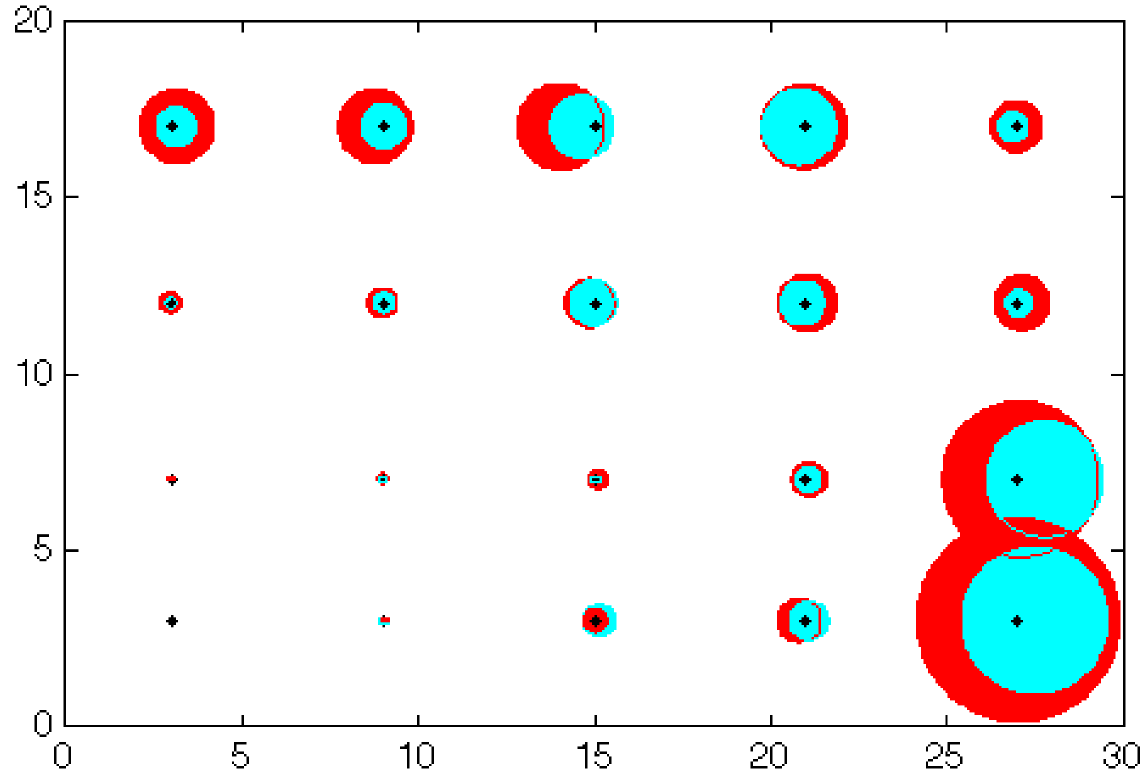
Posterior predictive checks

- Fine-scale binary media realizations (on a 3000 x 2000 mesh) can be used to calculate breakthrough times at 20 sensors
  - Did so with 1,000 realizations, not all 10^6 possible
  - Allowed us to plot 1st, 50th and 99th percentiles
  - Measurements plotted as references
- Why are some breakthrough times well predicted and others are not?



Circle plots

- Sensor: Dots
- Circles
  - Red: PPC using reconstructions using just $\{k^{obs}\}$
  - Cyan: Using $\{k^{obs}, t^{obs}\}$
- Circle radius:
  - Prop to the 95% CI of breakthrough times
- Circle center offset:
  - Prop to diff between measured and mean pred. breakthrough



- Takeaway: Further the measurement from injector/producer, bigger the uncertainty in predictions from reconstructions. Two reasons
  1. Longer breakthrough times – the % uncertainty may not be large
  2. Smaller flow rates lead to less info gathered and bigger uncertainties

Summary

- Reduced-order modeling and Bayesian inference allows us to estimate *smooth* random fields without worrying about overfitting
  - High-order KL modes have posterior densities = prior densities
  - We can quantify the uncertainty in the inference
- Due to the sub-grid model, one can infer mesh-unresolved structures
  - Requires proper data
  - Will only provide statistics of the subgrid structures
- We may also be able to generate an ensemble of fine-scale structures which are consistent with the observations
  - Made possible by the fact that we don't develop a unique/deterministic inference, but rather a distribution
- But .....
  - What if the fields are not smooth and multiGaussian RFMs don't work?
  - And what if the forward problem is expensive and $O(10^6)$ evaluations are not feasible?

What if the field is rough? Shrinkage regression

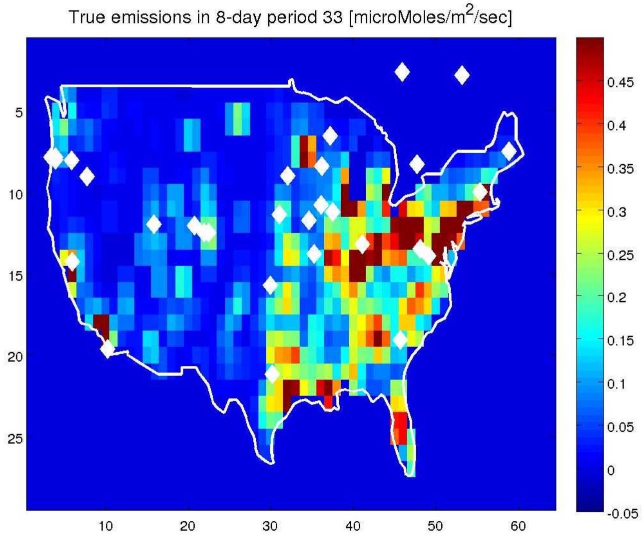
J. Ray, V. Yadav, A. M. Michalak, B. van Bloemen Waanders and S. A. McKenna, "A multiresolution spatial parameterization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions", *Geoscientific Model Development*, 7, 1901-1918, 2014.

J. Ray, J. Lee, V. Yadav, S. Lefantzi, A.M. Michalak, and B. van Bloemen Waanders, "A sparse reconstruction method for the estimation of multi-resolution emission fields via atmospheric inversion" *Geoscientific Model Development*, 8, 1259-1273, 2015.

Estimation of rough fields

- Rough fields cannot be represented easily using mGaussian models (for ζ)
  - But wavelets can. However, wavelet-based RFMs have tons(!) of parameters
    - As many as the grid-blocks in the mesh
    - But they are multi-resolution. Low index wavelets can represent large-scale structures; high index stand for fine details
- Example: Estimation of fossil-fuel CO2 (ffCO2)emission fields in US
  - Rough, and limited to where people live

- {x} is a gridded ffCO2 emission field
- $\{Y^{(obs)}\}$ = vector of CO2 concentration measurements over a year @ sensors
- $\{Y^{(obs)}\} = [H]\ \{x\} = [H]\ [\Phi]\ \{w\}$
  - H is a linear operator that maps emissions to conc. measurements
  - [F] is the matrix of wavelet bases
  - {w} are wavelet weights which are to be estimated from

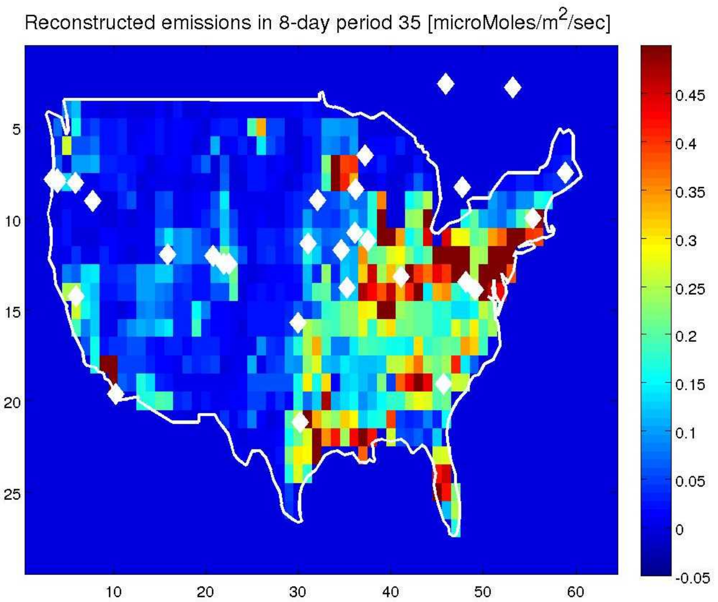


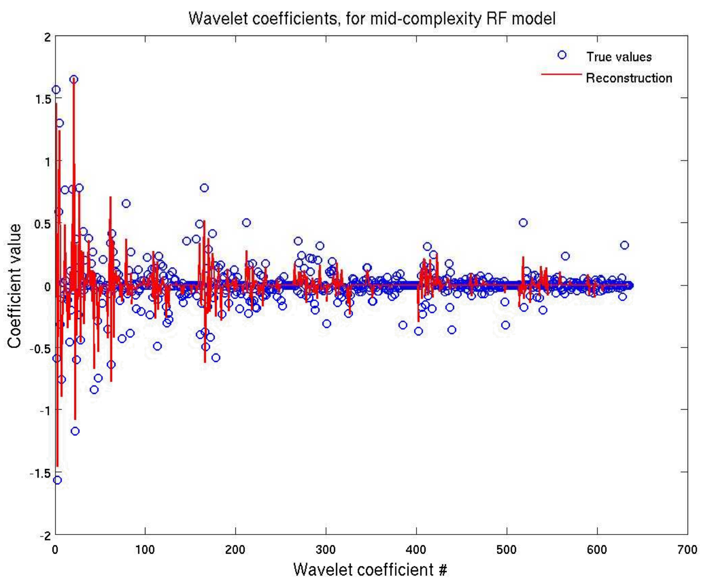
Reducing dimensionality

- Dimensionality of {w} is the same as {x} i.e., number of grid-blocks on which x is defined
  - Too many w_i to infer from $Y^{(obs)}$
- Dimensionality reduction
  - Restrict wavelets to those that inform regions where humans live
  - Use images of nightlights to do so

Estimation thru shrinkage

- How to estimate the minimum number of wavelets supported by data?
- Minimize $| Y^{(obs)} - \mathbf{[H][\Phi]}w |_2 + \lambda |w|_1$
- Choose λ based on how sparse you want {w} to be
- Will affect $| Y^{(obs)} - \mathbf{[H][\Phi]}w |_2$





- Shrinkage allows us to infer the large-scale patterns of emission field (above)
- And it zeros out the weights of many wavelets (left)

What if the forward problem is computationally expensive? Surrogates.

J. Ray, Z. Hou, M. Huang, K. Sargsyan and L. Swiler,"Bayesian calibration of the Community Land Model using surrogates" *SIAM Journal on Uncertainty Quantification*, 3(1):199-233, 2015.

J. Ray, S. Lefantzi, S. Arunajatesan and L. Dechant, "Bayesian calibration of a RANS model with a complex response surface - A case study with jet-in-crossflow configuration", 45 AIAA Fluid Dynamics Conference, Dallas, TX, June 22-26, 2015. Conference paper: AIAA-2015-2784.

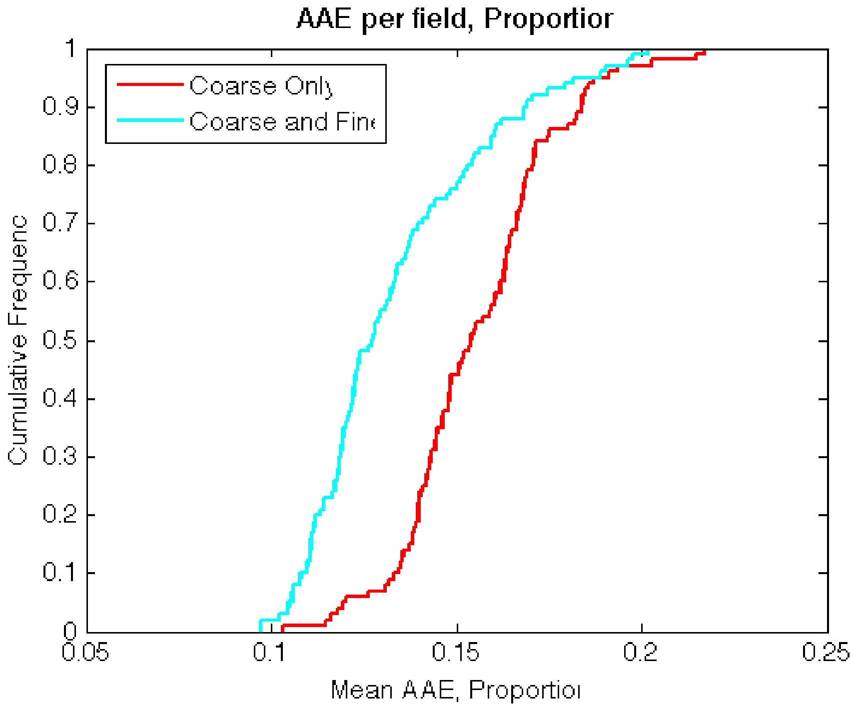
Surrogate modeling

- If the forward model $Y = \mathcal{M}(p)$ is expensive, replace it with a statistical curve-fit
  - Called emulators / surrogates / proxies
- Requires one to
  - sample the parameter space to get N samples p_i and run the the model to get N outputs $Y_i = \mathcal{M}(p_i)$
  - Runs can be done in batch; embarrassingly parallel runs
  - Fit a curve $\mathcal{M}^{(s)}(p)$ i.e., $Y_i \sim \mathcal{M}^{(s)}(p_i) + \eta$
  - Use $\mathcal{M}^{(s)}$ instead of $\mathcal{M}$
- There are many ways of doing this curve-fit
  - Review papers on surrogate models
  - Automated via software packages like DAKOTA (http://dakota.sandia.gov)
  - Can be used in MCMC estimation of really expensive climate and engineering models
- But …..
  - No guarantee that you will manage to do a curve fit OR
  - Which of the hundreds of curve-fit forms/methods will work

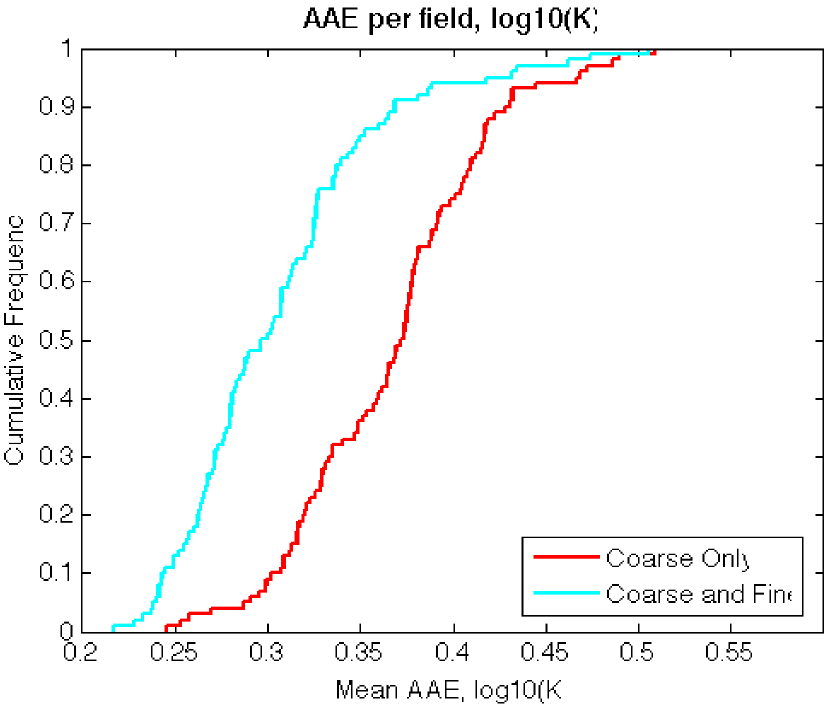
BONEYARD

Coarse Scale Evaluation

Proportion AAE per field

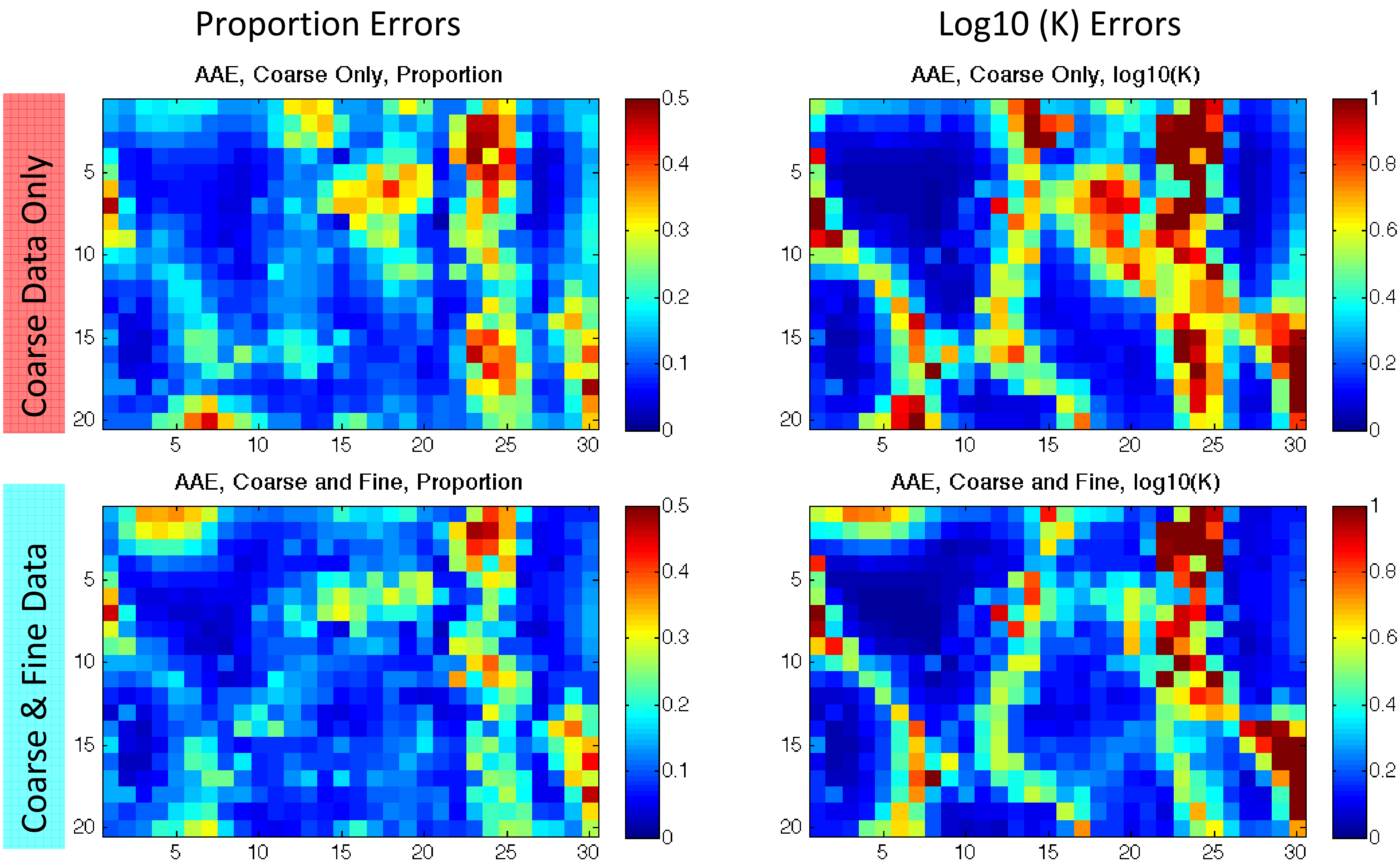


Log10(K) AAE per field



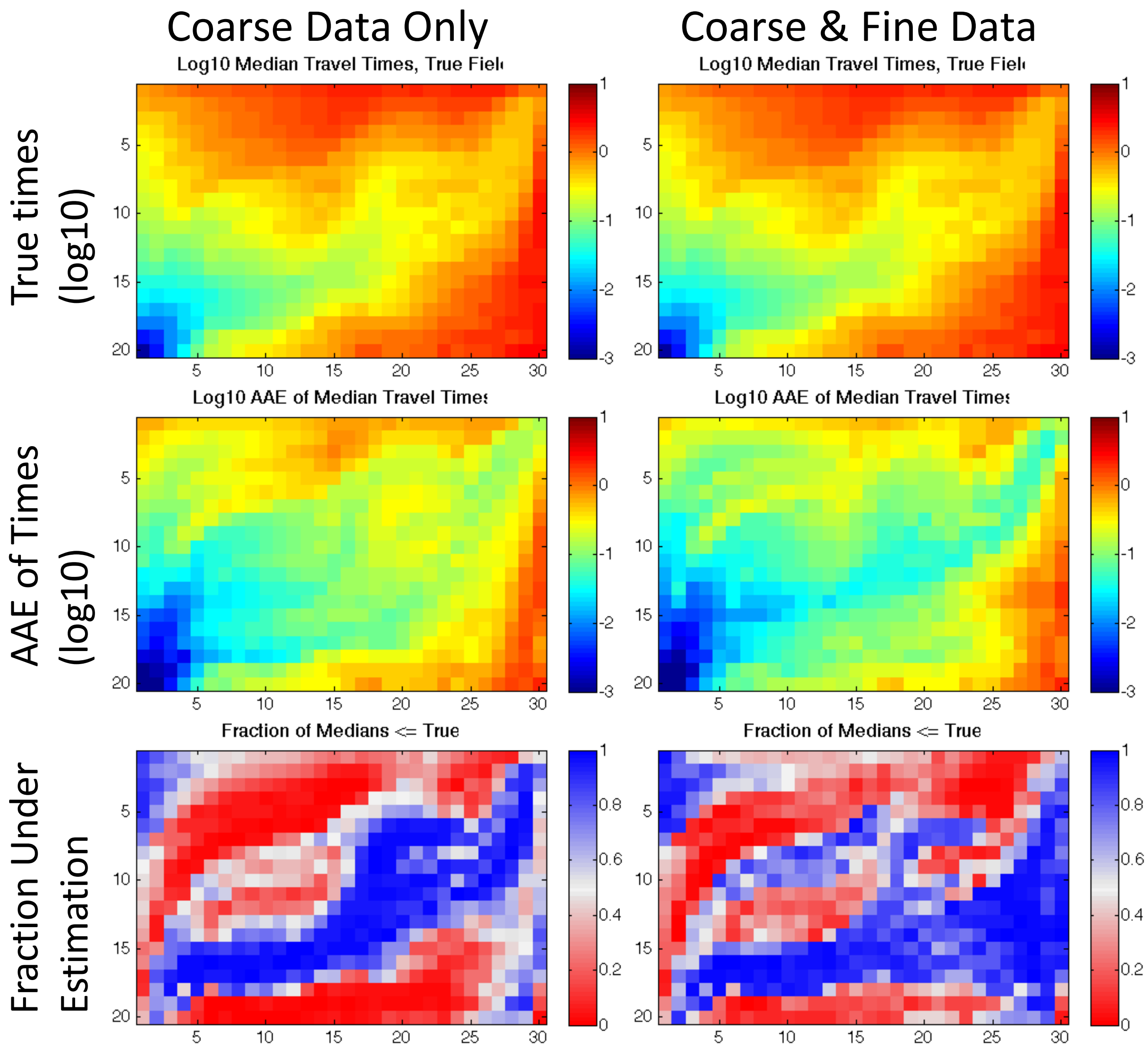
Coarse scale performance across 100 realizations evaluated for every field

Coarse Field Estimation



Coarse scale performance across 100 realizations evaluated at every cell

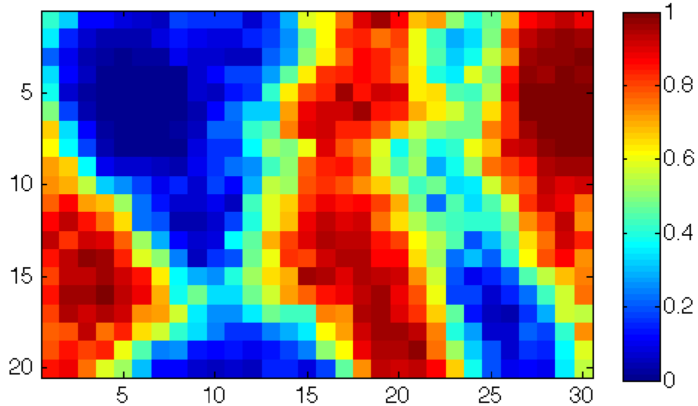
Median Travel Time Estimation



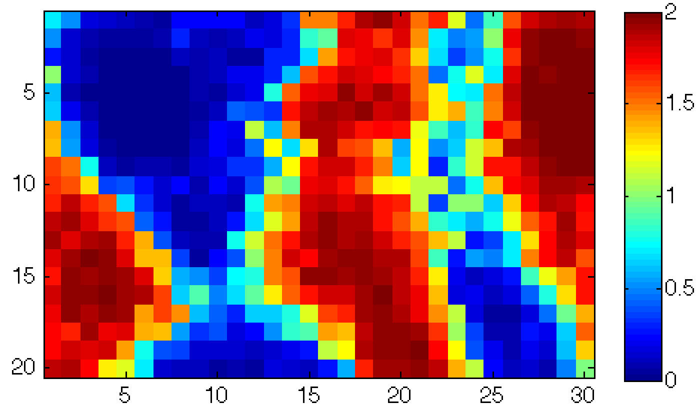
Two Scales

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

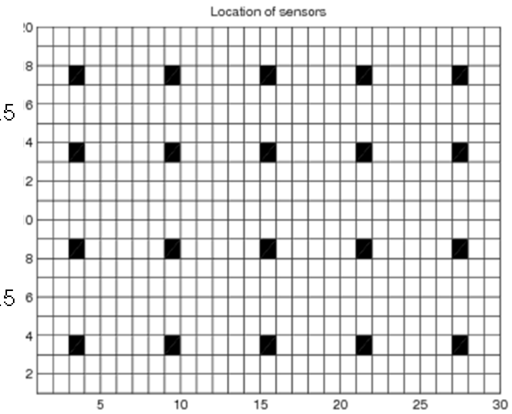
True F field



True Coarse K field

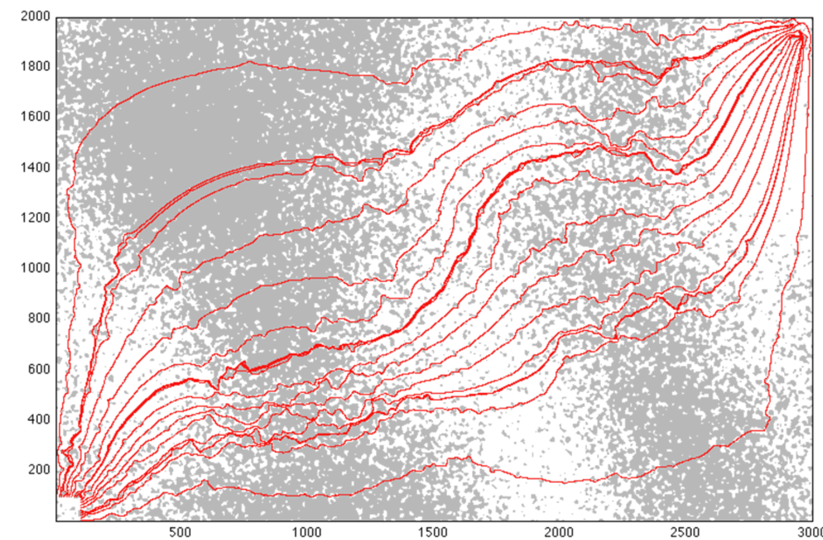


20 Well Locations



F = proportion of high conductivity

True Fine K field



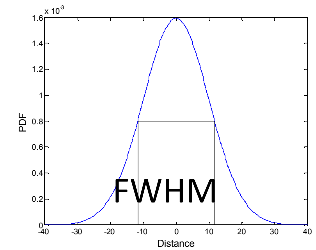
<u>Fine Scale:</u>
Binary Media
3000x2000 cells
Measured travel times to 20 sensors

Injector in lower left
Producer in upper right

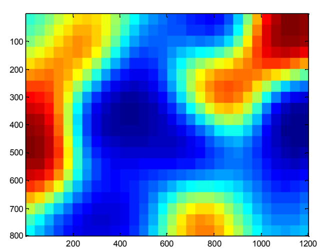
True binary fine-scale K field with example particle tracks

Posterior Evaluation

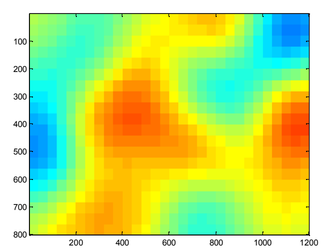
Inferred coarse-scale F fields and FWHM values provide information necessary to create fine-scale binary fields



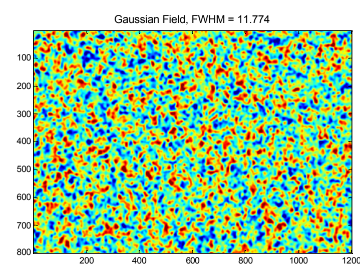
F field



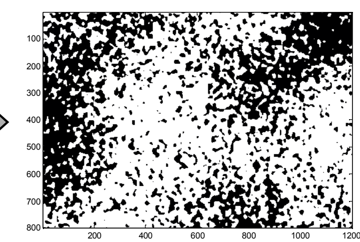
Z field



MG field



Binary field



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

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

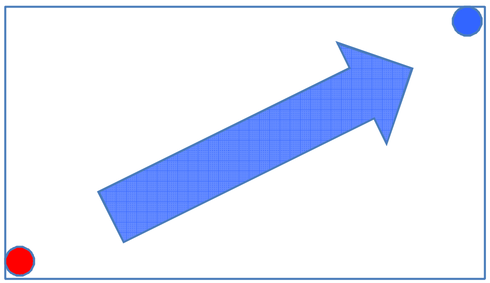
Coarse scale estimation provides the proportion of high permeability material within each coarse cell

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

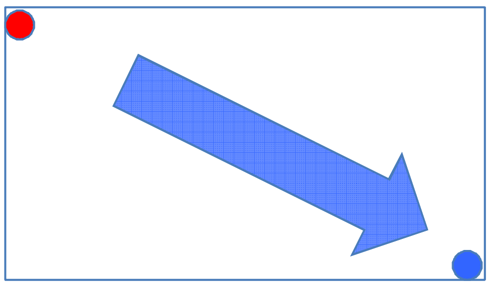
Switching Flow Direction

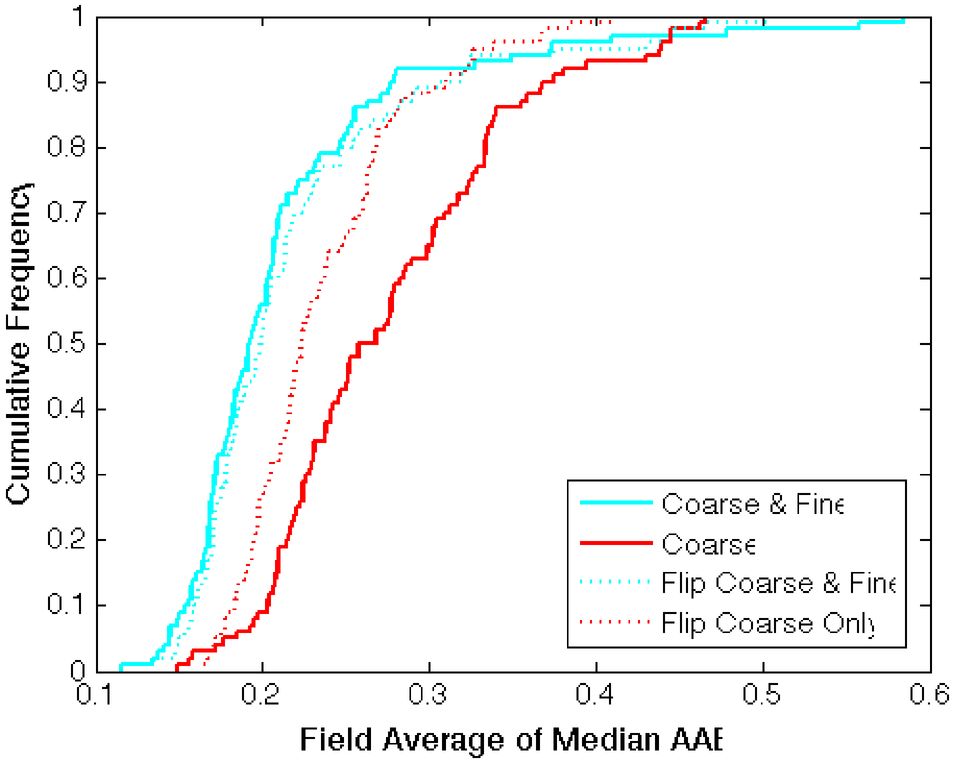
Distributions of the spatial average of the AAE of the median times (one value per realization)

Original Configuration



Flipped Configuration





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