Modeling of physical systems underspecified by data

Daniel M. Tartakovsky
Department of Mechanical & Aerospace Engineering
University of California, San Diego

Outline

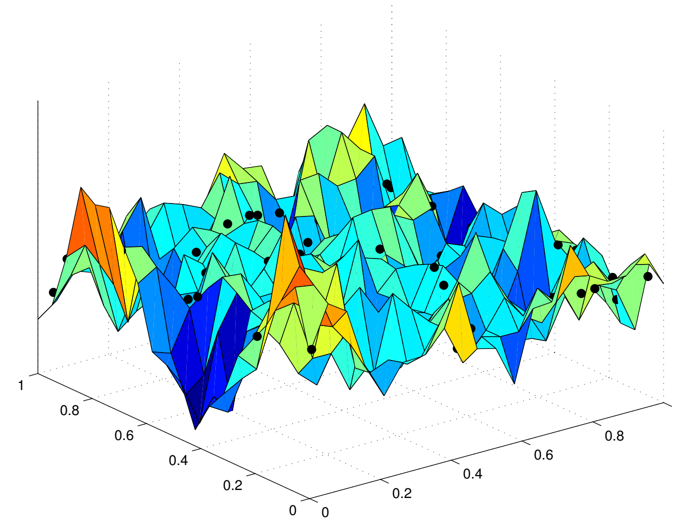
1. Physical systems & Stochastic PDEs
2. Data-Driven Domain Decompositions (D^4) for SPDEs
3. Implementation of D^4
   (a) Data analysis & image segmentation
   (b) Closure approximations for SPDEs
   (c) PDEs on random domains
4. Effective parameters for heterogeneous composites
5. Conclusions

1.1 Background

- Wisdom begins with the acknowledgment of uncertainty—of the *limits* of what we know. David Hume (1748), *An Inquiry Concerning Human Understanding*

- Most physical systems are fundamentally stochastic (Wiener, 1938; Frish, 1968; Papanicolaou, 1973; Van Kampen, 1976):

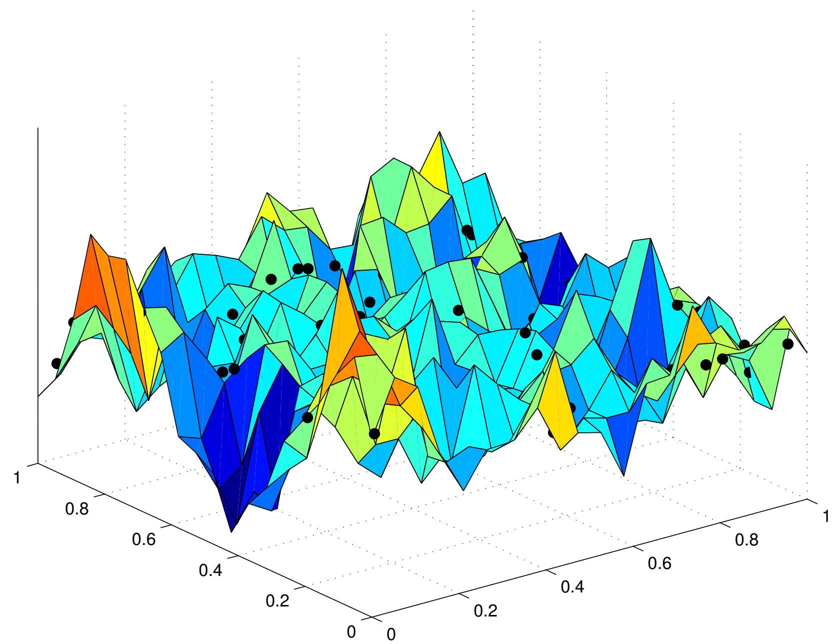
  - Model & Parameter uncertainty
    - Heterogeneity
    - Lack of sufficient data
    - Measurement noise
      - Experimental errors
      - Interpretive errors



- Randomness as a measure of ignorance

1.2 Probabilistic Framework

- Parameter,
  $p(\mathbf{x}) : \mathcal{D} \in \mathbb{R}^d \rightarrow \mathbb{R}$
- Random field,
  $p(\mathbf{x}, \omega) : \mathcal{D} \times \Omega \rightarrow \mathbb{R}$
- Ergodicity,
  $\langle p \rangle \approx \frac{1}{||\mathcal{D}||} \int_{\mathcal{D}} p \mathrm{d}\mathbf{x}$



parameter $p(x_1, x_2)$

- Governing equations become stochastic

$$\begin{cases} \mathcal{L}(\mathbf{x}; u) = f(\mathbf{x}), & \mathbf{x} \in \mathcal{D} \\ \mathcal{B}(\mathbf{x}; u) = g(\mathbf{x}), & \mathbf{x} \in \partial\mathcal{D} \end{cases} \Longrightarrow \begin{cases} \mathcal{L}(\mathbf{x}, \omega; u) = f(\mathbf{x}, \omega), & \mathbf{x} \in \mathcal{D} \\ \mathcal{B}(\mathbf{x}, \omega; u) = g(\mathbf{x}, \omega), & \mathbf{x} \in \partial\mathcal{D} \end{cases}$$

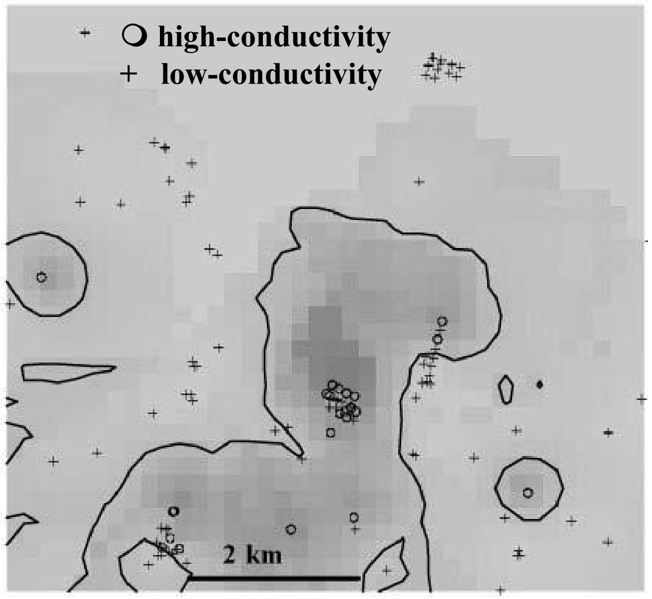
UCSD

1.3 Stochastic PDEs & UQ

- Consider a physical process

$$\begin{cases} \mathcal{L}(\mathbf{x}; u) = f(\mathbf{x}), & \mathbf{x} \in \mathcal{D} \\ \mathcal{B}(\mathbf{x}; u) = g(\mathbf{x}), & \mathbf{x} \in \partial\mathcal{D} \end{cases}$$

  – System parameters $p_i = p(\mathbf{x}_i)$, $i = 1, \ldots, N$
  – System states $u_j = u(\mathbf{x}_j)$, $j = 1, \ldots, M$



- Modeling steps

1. Use data to construct a probabilistic description of $\{p(\mathbf{x}, \omega)\}$
2. Solve SPDEs to obtain a probabilistic description of $u(\mathbf{x}, \omega)$
3. Assimilate $u_j = u(\mathbf{x}_j)$ to refine prior distributions

1.4a Statistical Methods for UQ

- (Brute-force) Monte Carlo methods
  - Convergence rate (CR): $1/\sqrt{N}$
  - CR is independent of the number of random variables
- Monte Carlo based methods
  - Quasi MC (QMC)
  - Markov chain MC (MCMC)
  - Importance sampling (Fishman, 1996)
- Variance reduction techniques
  - Problematic when the number of RVs is large
- Response Surface Methods (RSM)
  - Interpolation in the state space reduces the number of realizations
  - Problematic when the number of RVs is large

1.4b Stochastic Methods for UQ

- "Indirect" methods
  - Fokker-Planck equations
  - Moments equations

- "Direct" methods
  - Interval analysis: Maximum output bounds
  - Operator-based methods
    - Weighted integral method (Takada, 1990; Geodatis, 1991)
    - Neumann expansion (Shinozuka, 1988)

- Polynomial chaos expansions
  - Grounded in rigorous mathematical theory of Wiener (1938)
  - Arbitrary inputs: Generalized polynomial chaos

1.5 Modeling Dichotomy

- Real systems are characterized by
  - Non-stationary (statistically inhomogeneous)
  - Multi-modal
  - Large variances
  - Complex correlation structures
- Standard SPDE techniques require
  - Stationarity (statistically homogeneity)
  - Small variances
  - Simple correlation structures (Gaussian, exponential)
  - Uni-modality
- Our goal is
  - to incorporate realistic statistical parameterizations
  - to enhance predictive power
  - to improve computational efficiency

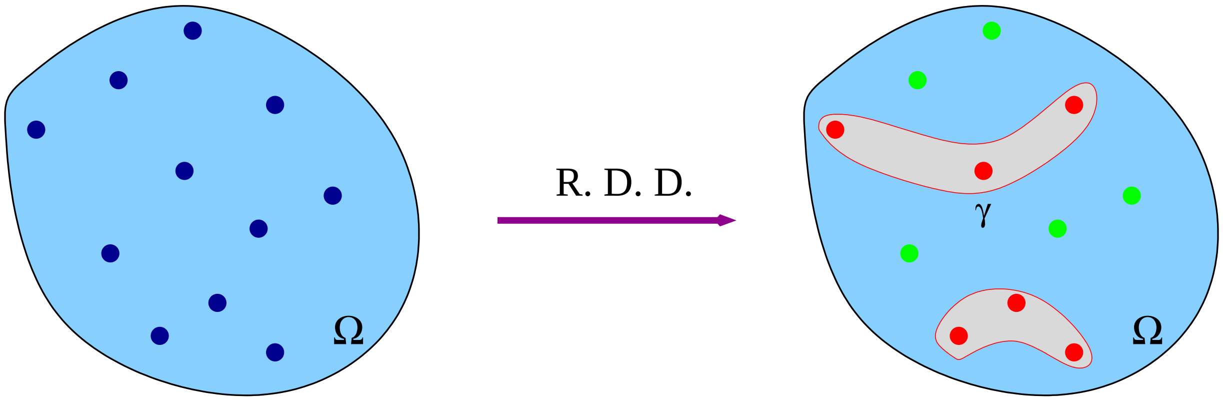
1.5 Reasons for Modeling Dichotomy

- Perturbation expansions
  - Small variance σ_p^2
- Spectral methods / polynomial chaos expansions
  - large correlation length l_p
  - uni-modality
- Mapping closures
  - "Nice" parameter distributions, e.g., Gaussian
- Stochastic upscaling (homogenezation)
  - Regularity requirements, e.g., spatial periodicity
  - Small σ_p^2
- Monte Carlo simulations
  - N increases with variance σ_p^2
  - Resolution depends on σ_p^2 and l_p

Outline

1. Physical systems & Stochastic PDEs
2. Data-Driven Domain Decompositions (D^4) for SPDEs
3. Implementation of D^4
4. Effective parameters for heterogeneous composites
5. Conclusions

2. Data-Driven Domain Decomposition



$f(\{\Pi\})$ $\qquad$ $f(\{\Pi\}, \Gamma)$

Multi-modal distributions
High variances
Complex correlation structures

Uni-modal distributions
Low variances
Simple correlation structures

$$\sigma_p^2 = Q_1 \sigma_{p_1}^2 + Q_2 \sigma_{p_2}^2 + Q_1 Q_2 \left[\langle p_1 \rangle - \langle p_2 \rangle \right]^2$$

2.1 Strategy for Domain Decompositions

- Step 1: Decomposition of the parameter space (image processing techniques; probability maps)

- Step 2: Conditional statistics (noise propagation; closures)

$$\int \mathcal{L}_{\{\Pi\}} u \; f(\{\Pi\}|\gamma) \; d\{\Pi\} \quad \rightarrow \quad \langle u|\gamma \rangle$$

- Step 3: Averaging over random geometries

$$\langle u \rangle = \int \langle u|\Gamma \rangle \; f(\Gamma) \; d\Gamma$$

2.2 Implementation of Domain Decompositions

1. Probabilistic decomposition of the parameter space
   - Geostatistical reconstruction of internal geometries
   - (nonstationary) Bayesian spatial statistics
   - Statistical Learning Theory (Support Vector Machines)
   - Risk-based parameterization
2. Conditional statistics from SPDEs
   - Perturbation expansions
   - Polynomial chaos expansions
   - Collocation methods
   - Nonlinear Gaussian mappings
3. PDEs on random domains
   - Monte Carlo simulations
   - Stochastic mappings
   - Perturbation expansions

Outline

1. Physical systems & Stochastic PDEs

2. Data-Driven Domain Decompositions (D^4) for SPDEs

3. Implementation of D^4

   3.1 Decomposition of the parameter space
   - Spatial statistics (geostatistics)
   - MCMC with Metropolis sampling
   - Support Vector Machines

   3.2 Conditional Statistics

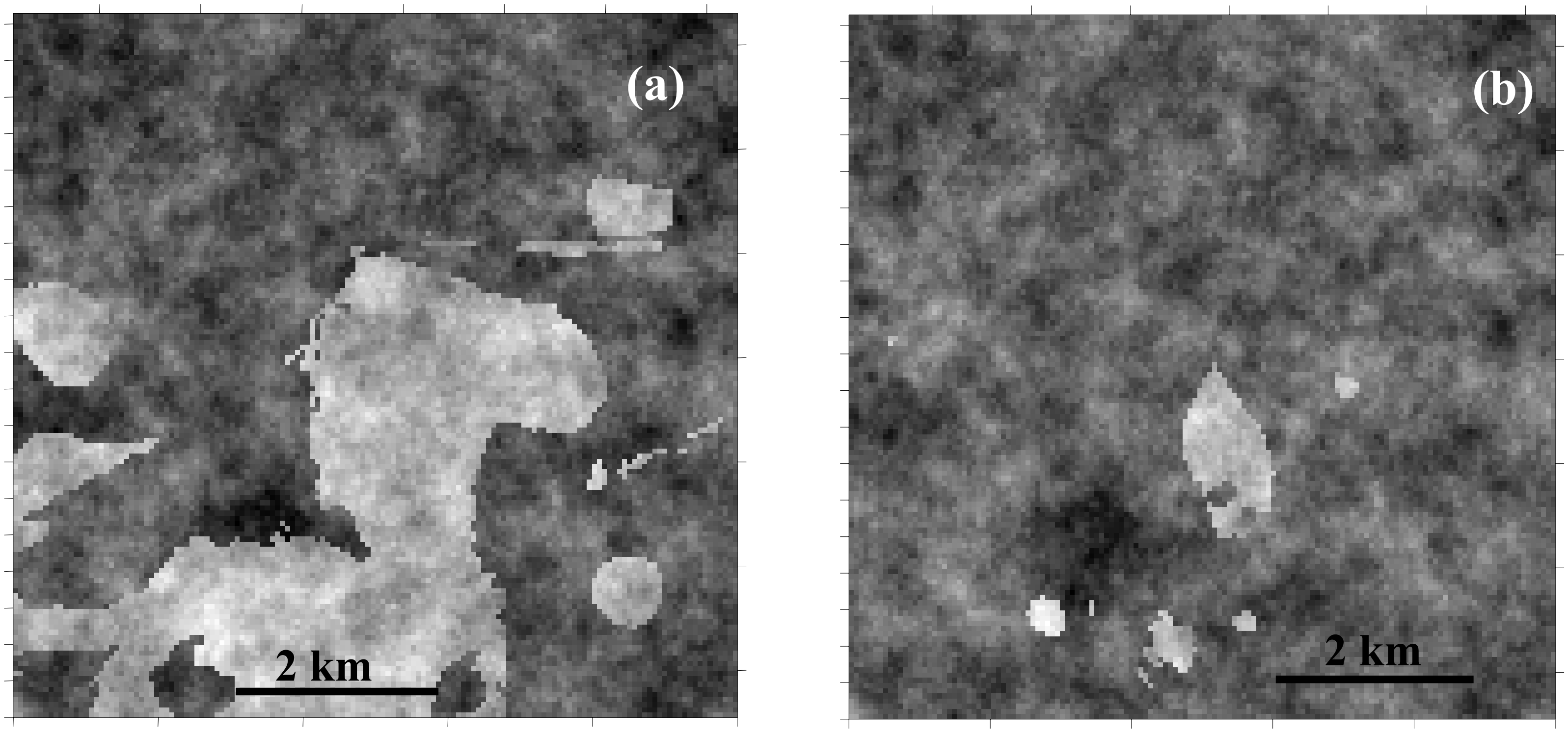
   3.3 Averaging over random geometries

4. Effective parameters for heterogeneous composites

5. Conclusions

UCSD

Probabilistic Reconstruction of Facies



Mean boundary (a) and the boundary with 95% probability (b)

Outline

1. Physical systems & Stochastic PDEs

2. Data-Driven Domain Decompositions (D^4) for SPDEs

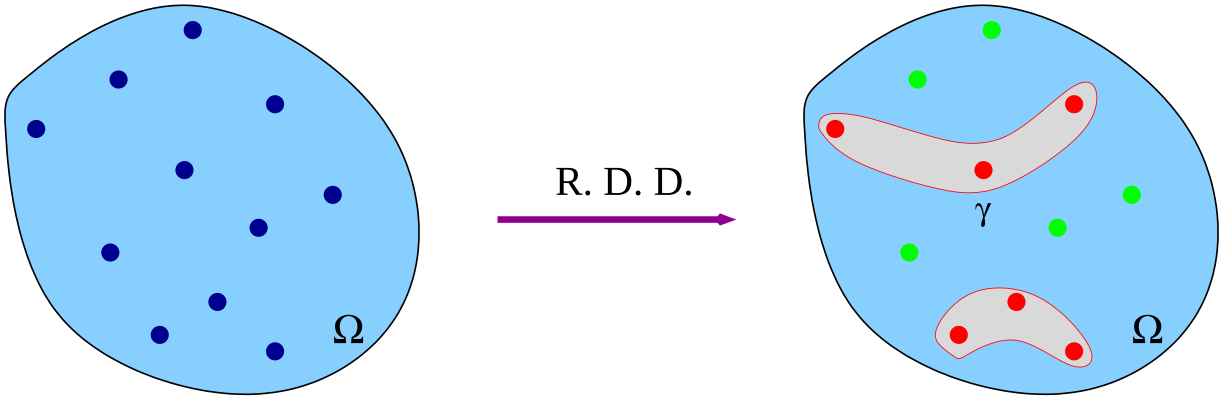
3. Implementation of D^4
   - 3.1 Decomposition of the parameter space
   - 3.2 Conditional Statistics
     - Perturbation solutions
     - Polynomial chaos expansions
     - Stochastic collocation on sparse grids
   - 3.3 Averaging over random geometries

4. Effective parameters for heterogeneous composites

5. Conclusions

Conditional Statistics

$$\begin{cases} \mathcal{L}(\mathbf{x},\omega;u) = f, & \mathbf{x} \in \mathcal{D} \\ \mathcal{B}(\mathbf{x},\omega;u) = g, & \mathbf{x} \in \partial\mathcal{D} \end{cases}, \qquad p = p(\mathbf{x},\omega), \quad \omega \in \Omega$$



$$\begin{cases} \mathcal{L}(\mathbf{x},\omega;u) = f, & \mathbf{x} \in \mathcal{D} \\ \mathcal{B}(\mathbf{x},\omega;u) = g, & \mathbf{x} \in \partial\mathcal{D} \\ \text{continuity conditions}, & \mathbf{x} \in \gamma \end{cases}, \qquad p = \begin{cases} p_1(\mathbf{x},\omega), & \mathbf{x} \in \mathcal{D}_1, \omega \in \Omega_1 \\ p_2(\mathbf{x},\omega), & \mathbf{x} \in \mathcal{D}_2, \omega \in \Omega_2 \end{cases}$$

3.2a Perturbation Solutions

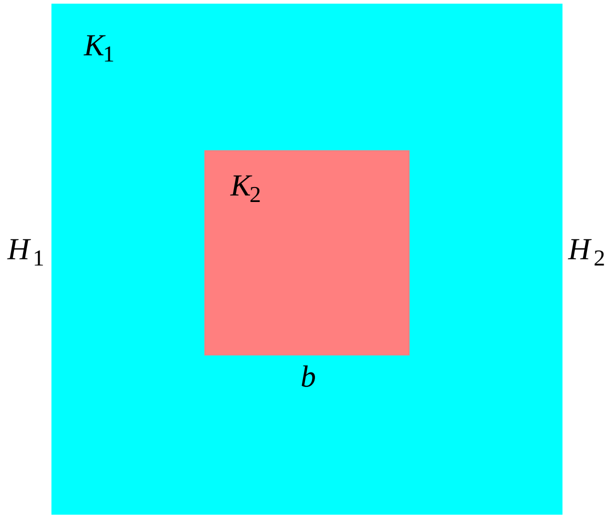
- $\nabla \cdot k(\mathbf{x}) \nabla u = Q\delta(\mathbf{x} - \mathbf{x}_0)$

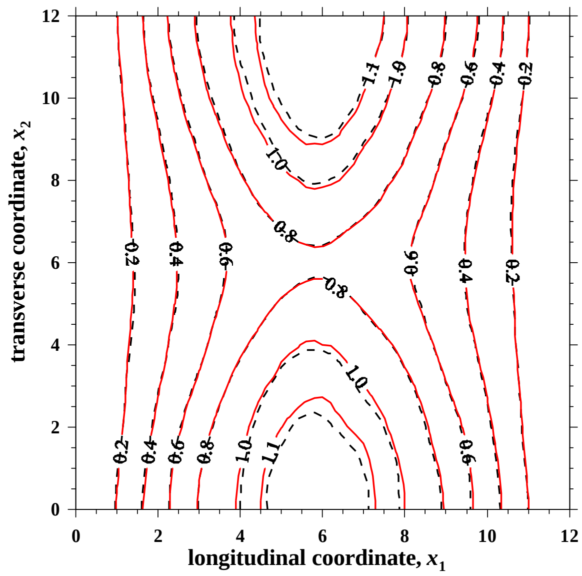
$Y_i = \ln k_i$ – Gaussian, $\overline{Y}_1 = 3\overline{Y}_2$
$\sigma^2_{Y_1} = \sigma^2_{Y_2} = 1$, $\sigma^2_Y \approx 30$
exponential correlation functions

b is log-normal

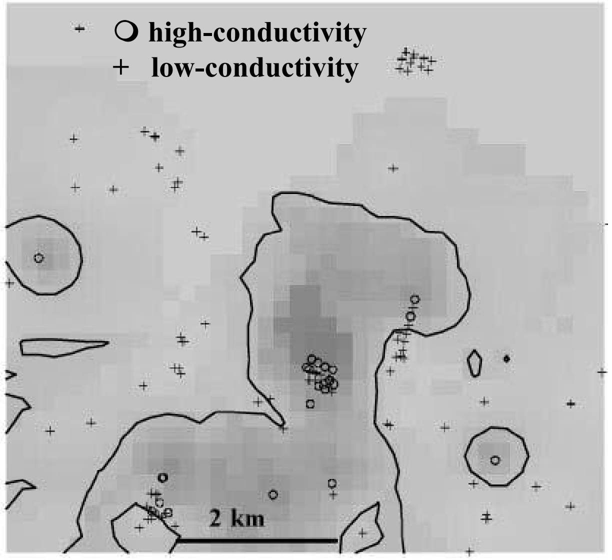
- Perturbation in $\sigma^2_{Y_i}$

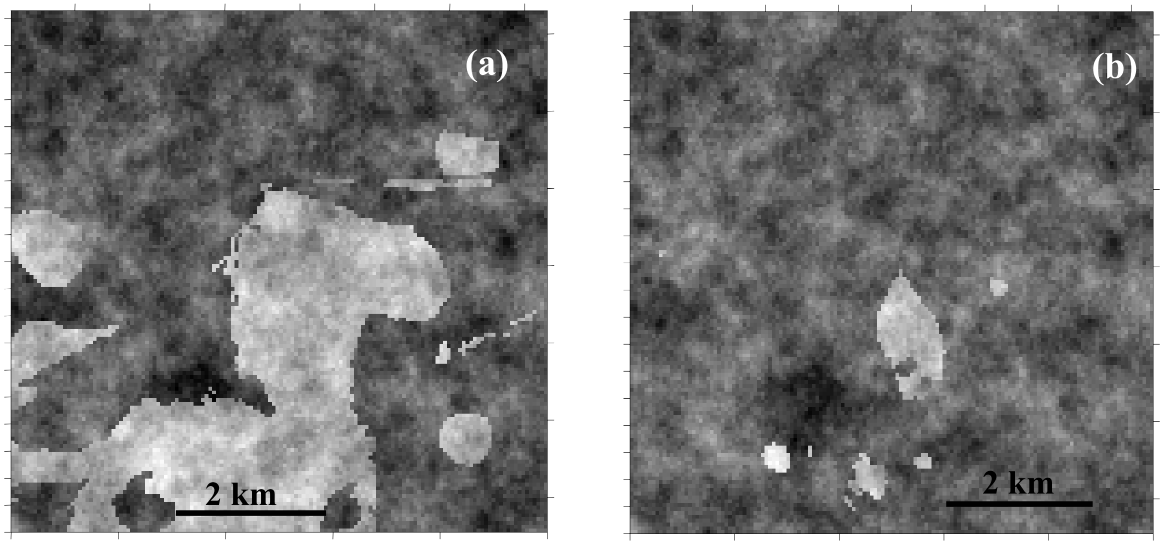
- Monte Carlo simulations for b
  - $\langle u(\mathbf{x}) \rangle$: $\mathcal{E}_u \approx 2\%$
  - $\sigma^2_u(\mathbf{x})$: $\mathcal{E}_\sigma \approx 5\%$





3.2a Perturbation Solutions (cntd)

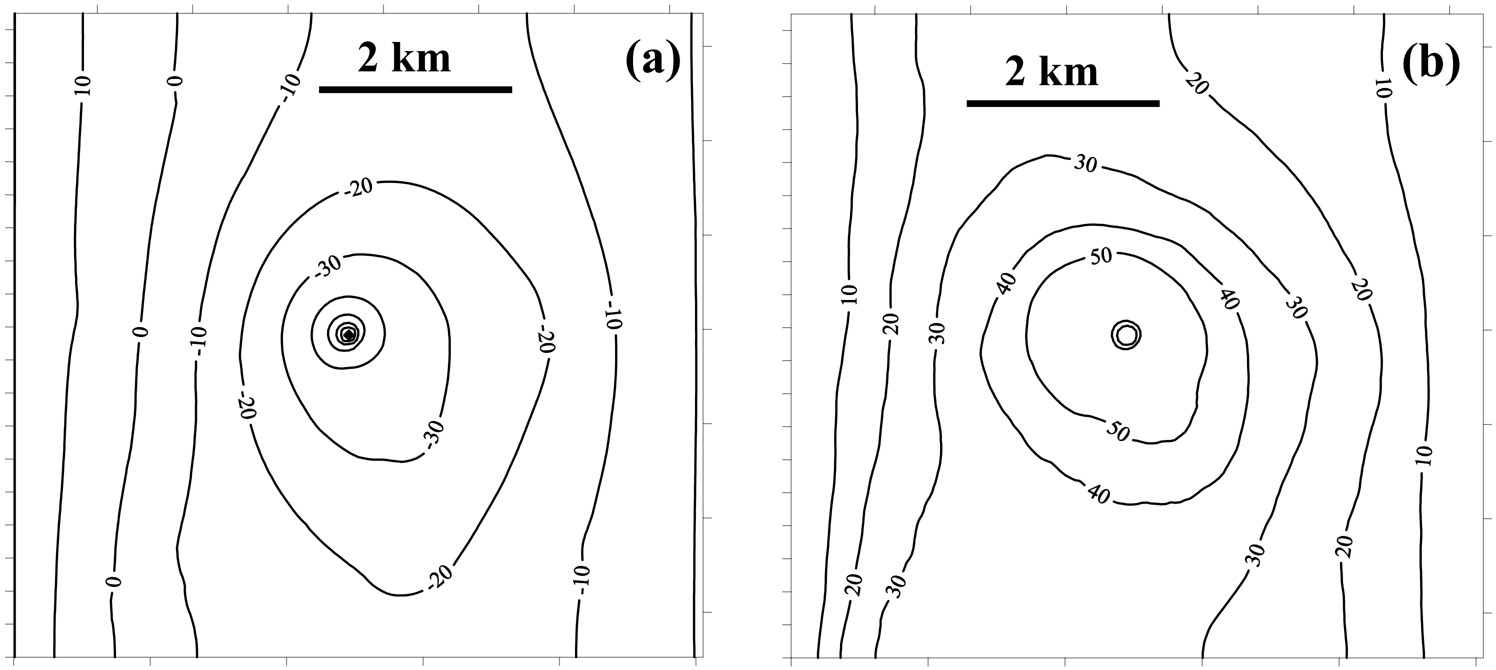




3.2a Perturbation Solutions (cntd)

Assign weights $\mathcal{W}_k$ to each boundary Γ_k, so that

$$\langle u \rangle = \int \langle u | \Gamma \rangle \, f(\Gamma) \, d\Gamma \approx \sum_k \mathcal{W}_k \langle u | \Gamma_k \rangle$$



Mean (a) and variance (b) of u

3.2b Polynomial Chaos Expansions

- Stochastic PDE

$$\begin{cases} \mathcal{L}(\mathbf{x}, \omega; u) = f(\mathbf{x}, \omega), & \mathbf{x} \in \mathcal{D} \\ \mathcal{B}(\mathbf{x}, \omega; u) = g(\mathbf{x}, \omega), & \mathbf{x} \in \partial D \end{cases}$$

- Generalized polynomial chaos expansions

$$p(\mathbf{x}, t, \omega) = \sum_{i=1}^{\infty} a_i(\mathbf{x}, t)\Psi_i(\omega)$$

  - An approximation

$$p(\mathbf{x}, t, \omega) \approx \sum_{i=1}^{N} a_i(\mathbf{x}, t)\Psi_i(\omega)$$

Correspondence between the type of the Wiener-Askey polynomial chaos and their underlying random variables.

Distribution	Polynomials
Gaussian	Hermite
Gamma	Laguerre
Beta	Jacobi
Uniform	Legendre
Poisson	Charlier
Binomial	Krawtchouk
Negative binomial	Meixner
Hypergeometric	Hahn

G. Em. Karniadakis. etc.

3.2b Polynomial Chaos Expansions (cntd)

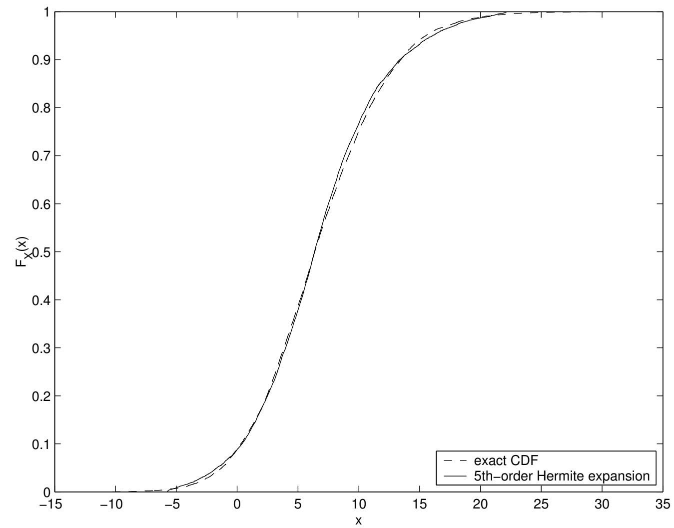
- Parameter representation

$$p(\mathbf{x}, t, \omega) = \sum_{i=1}^{N} a_i(\mathbf{x}, t)\Psi_i(\omega)$$

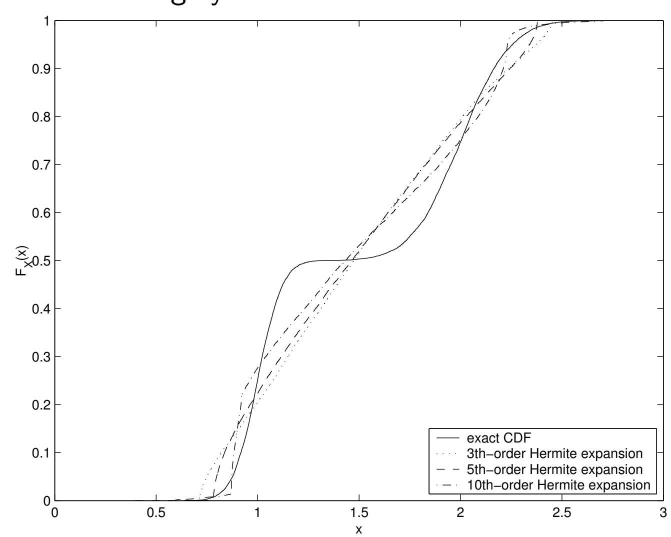
  - Choose an orthogonal polynomial basis
  - Reduce N

- Advantages:
  - Nonperturbative
  - Large fluctuations

- Drawbacks:
  - Finite correlations
  - Unimodal distributions



a highly non-Gaussian unimodal field



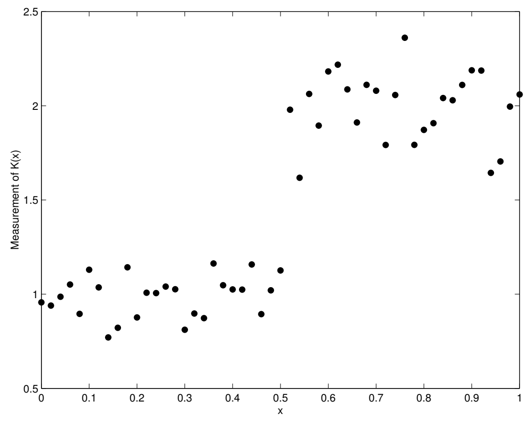
a bimodal field

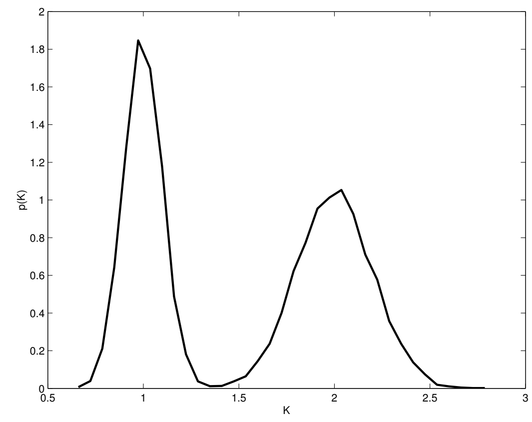
UCSD

3.2b Polynomial Chaos Expansions (cntd)

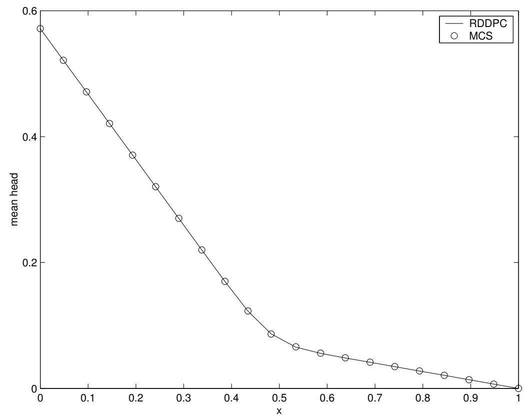
Problem: $\nabla \cdot k \nabla u = 0$

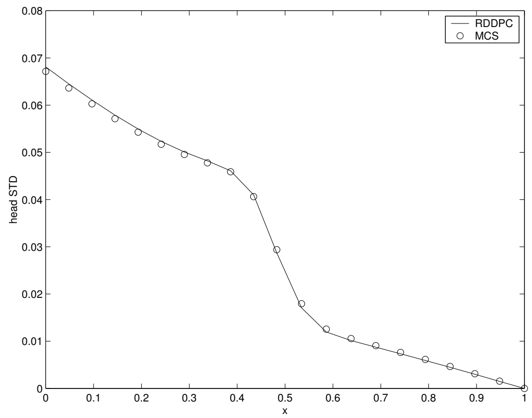
Parameter measurements $k_i = k(x_i)$:





Solution statistics: ensemble mean (a) and standard deviation (b)





3.2c Stochastic Collocation Methods

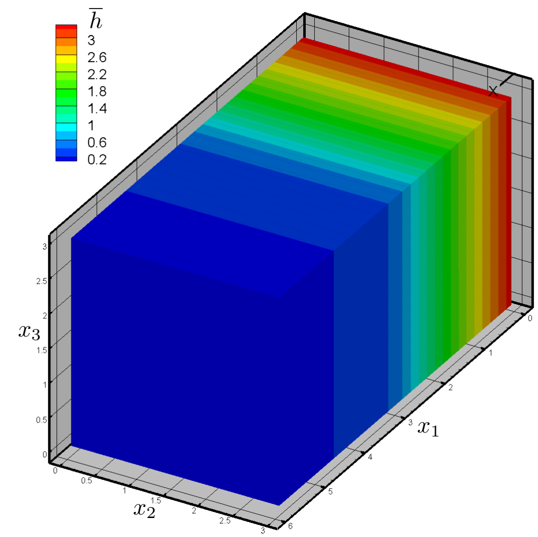
- Choice of sampling points
  - Tensor products of 1-D collocation point sets
  - Sparse grids (nested or non-nested)

- Choice of weight functions

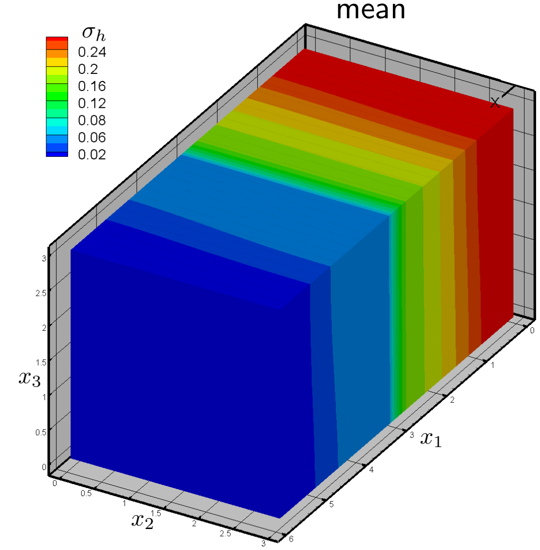
$$\vartheta(\boldsymbol{\xi}) = \delta(\xi - \xi_k)$$

- Advantage:
  - Nonintrusive

- Disadvantage
  - Can be less accurate than PCE



mean



variance

Outline

1. Physical systems & Stochastic PDEs

2. Data-Driven Domain Decompositions (D^4) for SPDEs

3. Implementation of D^4

   3.1 Decomposition of the parameter space
   3.2 Conditional Statistics
   3.3 Averaging over random geometries

4. Effective parameters for heterogeneous composites

5. Conclusions

Computational Approach

- A deterministic equation in a random domain

$$\begin{cases} L(x;u) = f(x), & x \in D(\omega) \\ B(x;u) = g(x), & x \in \partial D(\omega) \end{cases}$$

- Step 1: Stochastic mapping

$$\xi = \xi(x;\omega), \quad x = x(\xi;\omega) \qquad\qquad x \in D(\omega), \quad \xi \in E$$

- Step 2: Solving a stochastic equation in a deterministic domain

$$\begin{cases} \mathcal{L}(\xi,\omega;u) = f(\xi,\omega), & \xi \in E \\ \mathcal{B}(\xi,\omega;u) = g(\xi,\omega), & \xi \in \partial E \end{cases}$$

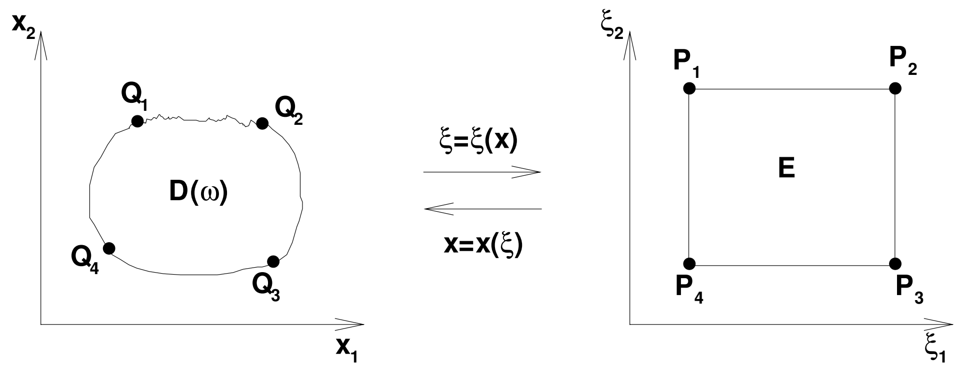
Random Mapping

- Surface parameterization
  - Karhunen-Loève representations
  - Fourier-type expansions
  - Etc.

- Numerical mappings, e.g.,

$$\nabla_\xi^2 x_i = 0 \qquad x_i|_{\partial E} = x_i|_{\partial D(\omega)}$$

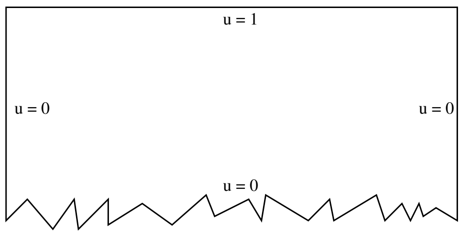
- Analytical mappings



Random mapping, $D(\omega) \to E$

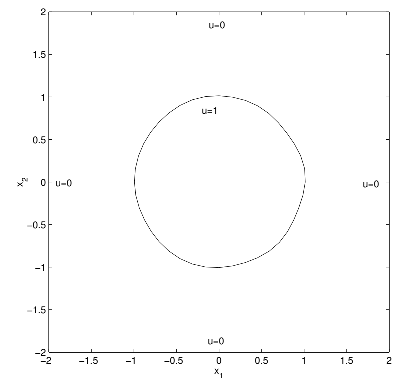
Computational examples

- Steady-state diffusion



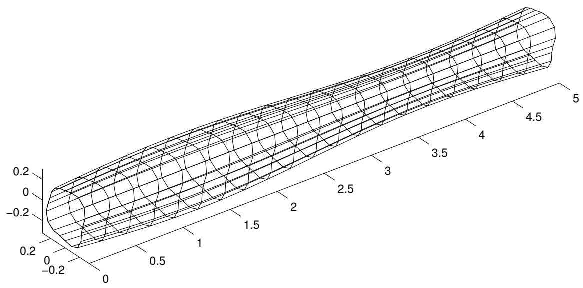
$$\nabla^2 u(x;\omega) = 0, \qquad x \in D(\omega)$$

  - Random bottom of a channel
  - Random exclusion
  - Numerical mapping



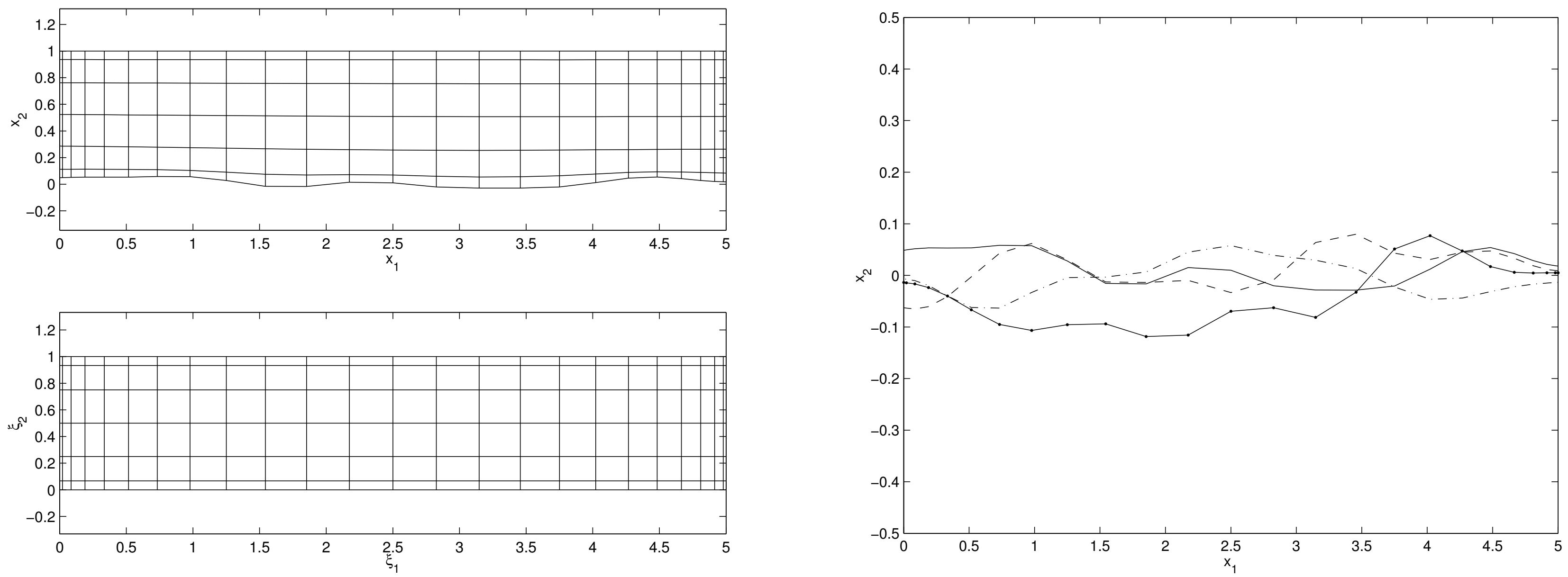
- Transport by Stocks flow in a pipe

$$\frac{\partial u}{\partial t} + \mathbf{v} \cdot \nabla u = a \nabla^2 u, \qquad x \in D(\omega)$$



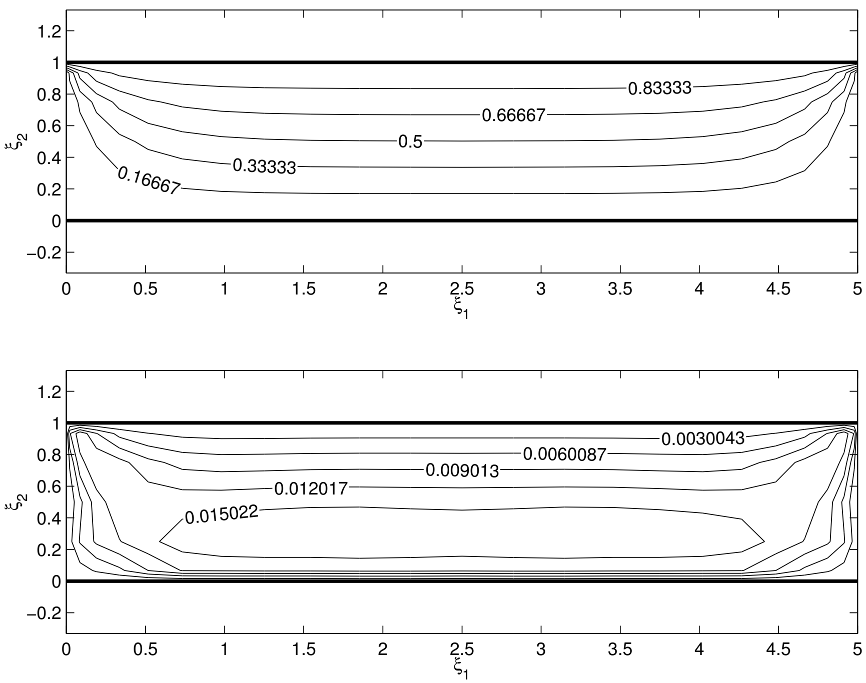
  - Lubrication approximation
  - Analytical mapping

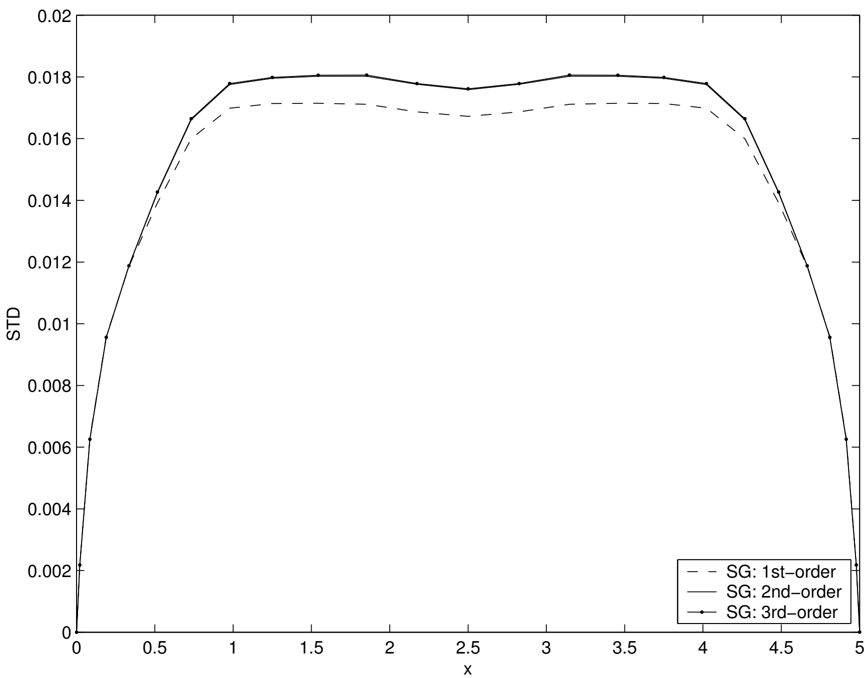
Steady Diffusion



- Random surface: exponential correlation function, $\lambda = 1$
- 10-term Karhunen-Loève representation
- 10-dimensional $(N = 10)$ random space

Steady Diffusion: Results





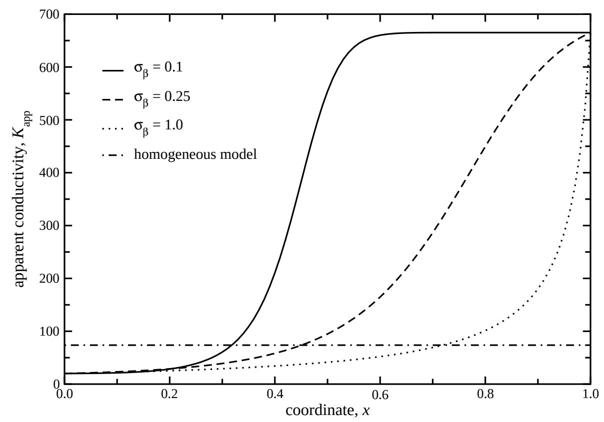
Mean & standard deviation

PCE vs. MCS

4. Effective Parameters

- Step 1: Conditional effective conductivity
  - $\mathbf{K}_{eff_i}(\mathbf{x}|\gamma) = \langle K_i \rangle \mathbf{I} + \mathbf{k}_i(\mathbf{x})$ $\qquad \mathbf{k}_i = [\mathbf{I} - \mathbf{B}_i]^{-1}\mathbf{A}_i$
  - Closure by perturbation

- Step 2: Averaging over random geometries γ



Conclusions

- Stochastic PDEs provide a natural framework for uncertainty quantification
- In real applications, SPDEs require a complex statistical parameterization
- D^4 allows for incorporation of expert knowledge and diverse data
- New concept of effective properties
- Ability to make predictions and quantify uncertainty in realistic setting