Uncertainty Quantification and Probabilistic Risk Analysis in Subsurface Hydrology

Daniel M. Tartakovsky

Department of Mechanical & Aerospace Engineering

University of California, San Diego

Uncertainty & Risk Assessment

- Epistemic uncertainty: "the uncertainty attributable to incomplete knowledge about a phenomenon that affects our ability to model it"

- Aleatory uncertainty: "the uncertainty inherent in a nondeterministic (stochastic, random) phenomenon"

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

Challenges of Probability

> "In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics that sometimes yield reasonable judgments and sometimes lead to severe and systematic errors" (Kahneman and Tversky, 1973;1982)

Evolutionary psychology (the evolution of the human brain did not require the concept of probability)

Neuropsychology (the human brain is wired to operate with whole numbers),

Probability and statistics are not taught properly from the point of view of human cognition (Keeler and Steinhorst, 2001)

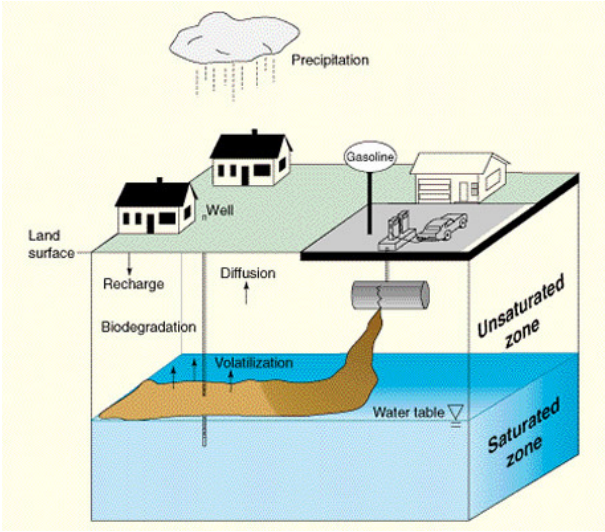
Probabilistic Risk Assessment (PRA)

- The Three Mile Island nuclear power plant accident, 1979
- The Space Shuttle Challenger disaster, 1986
- U.S. Congress, Natl. Res. Council (also, NRC, 1997):
  - Risk assessment must be probabilistic
  - Probabilities must be subjective
  - "The main focus of PRA should be on uncertainty quantification"
- PRA should answer the following three questions:
  - What can happen?
  - How likely is it to happen?
  - Given that it occurs, what are the consequences?

Risk of Contamination

Modern "science-based predictions" should be probabilistic. In subsurface hydrology, they must account for

- Model or structural uncertainty
  - geologic models
  - physical (geochemical, etc) models
  - their mathematical representations
- Parametric uncertainty
  - sparse data, etc.



toxics.usgs.gov

Probabilistic Risk Assessment (PRA) is an ideal framework for uncertainty quantification

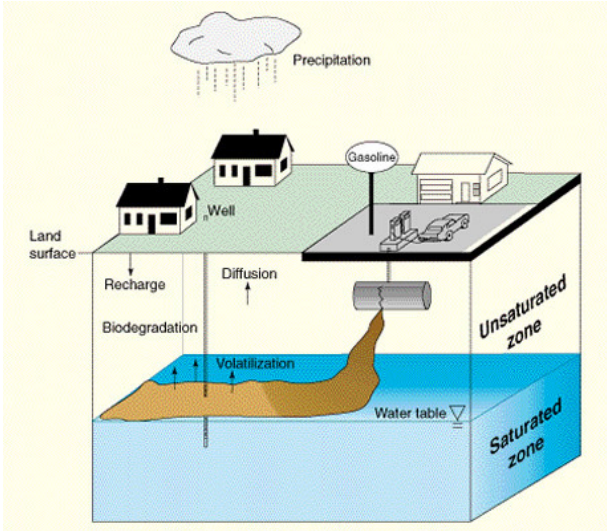
Algorithm for Probabilistic Risk Assessment

- Identification of a system's components
- Construction of a fault tree or a binary decision diagrams
- A cut set representation of a fault tree using Boolean operators
- Computation of the probability of system's failure

Identification of System's Components

Basic events:

- Spill occurs (SO)
- Natural attenuation fails (NA)
- Remediation effort fails (RE)



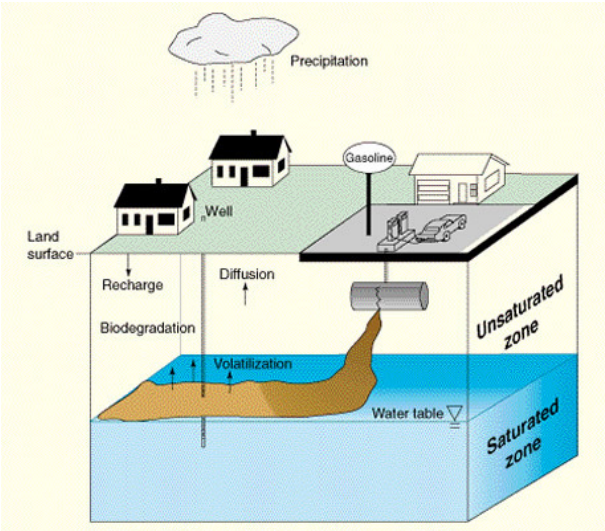
Goal: assess the probability of aquifer contamination (AC)

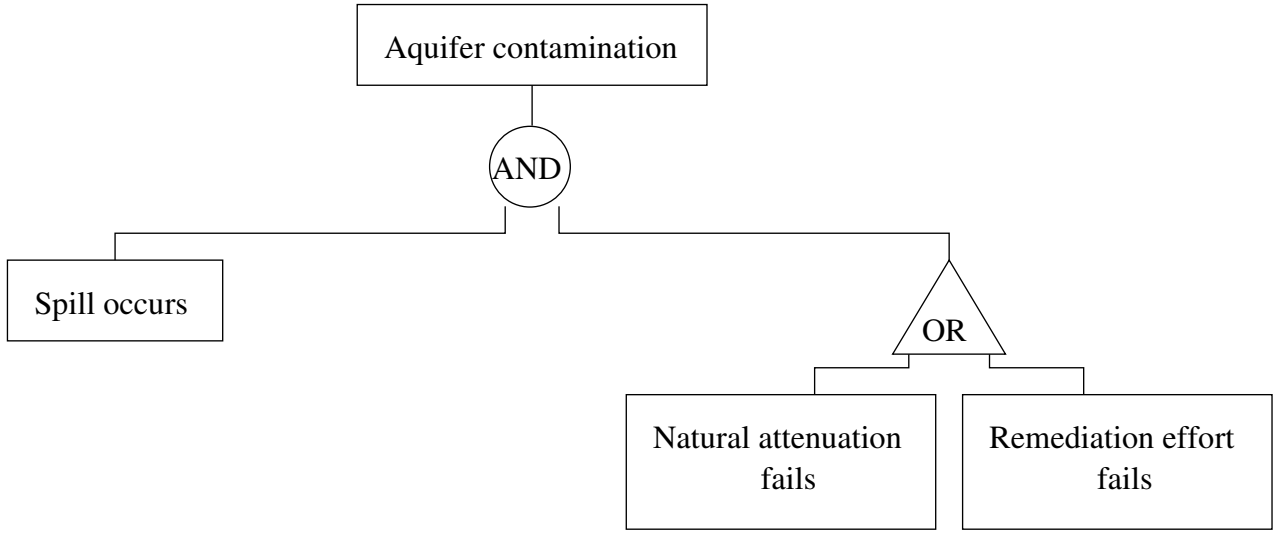
The Excess Lifetime Cancer Risk (ELCR) factor,

$$\mathrm{ELCR} = \alpha C, \qquad \alpha = \frac{\mathrm{IR} \times \mathrm{EF}}{365\, days \times \mathrm{BW}}$$

IR = human ingestion rate, EF = exposure frequency, BW = average body weight

Fault Tree Ananlysis





Minimal Cut Sets: {SO, NA} and {SO, RE}

Cut Set Representation: AC $= \mathrm{SO} \cdot \mathrm{NA} + \mathrm{SO} \cdot \mathrm{RE}$

Probability of aquifer contamination at time $t = T$:

$$P[\mathrm{AC}] = P[\mathrm{SO} \cap \mathrm{NA}] + P[\mathrm{SO} \cap \mathrm{RE}] - P[\mathrm{SO} \cap \mathrm{NA} \cap \mathrm{RE}]$$

Probability Models

Probability of aquifer contamination at time $t = T$:

$$P[\mathrm{AC}] = P[\mathrm{SO} \cap \mathrm{NA}] + P[\mathrm{SO} \cap \mathrm{RE}] - P[\mathrm{SO} \cap \mathrm{NA} \cap \mathrm{RE}]$$

Operational approximations:

- Rare vent approximation: $P[\mathrm{AC}] \approx P[\mathrm{NA}] + P[\mathrm{RE}]$
- Common cause approximation:

$$P[\mathrm{NA} \cap \mathrm{RE}] \approx P[\mathrm{NA}|\mathrm{PF}]P[\mathrm{PF}] + P[\mathrm{NA}]P[\mathrm{RE}]P[\mathrm{PF}']$$

- "Markov jump process" approximation

Probability Models: Markov Jump Process

Assumptions:

1. State transitions form a Markov process

2. Transition times are direction-independent, $P_{\sigma'\sigma''}[0 < \tau < t''] = F_{\sigma'}(t'')Q_{\sigma'\sigma''}$

State	State description
U	The site is uncontaminated
S	A spill has occurred
R	The site is undergoing remediation
N	The site is undergoing natural attenuation
C	The aquifer is contaminated

Transition	Outcome of the transition
US	The site is contaminated by a spill
SR	The contaminant escapes the waste site
SU	The spill is contained on site
RN	Remediation fails
NC	Natural attenuation fails
RU	Remediation succeeds
NU	Natural attenuation succeeds

"Markov jump process" approximation:

$$P_{UC}(0, t_{crit}) = Q_{US}Q_{SR}Q_{RN}Q_{NC}\mathcal{I}(t_{crit})$$

$Q_{\sigma'\sigma''}$ – probability of the state transition *ever* occurring

$F_\sigma(t_\sigma) = \int_0^{t_\sigma} q_\sigma \mathrm{e}^{-q_\sigma \tau_\sigma} \mathrm{d}\tau_\sigma$ – distribution of the transition time

UCSD

Computation of Probabilities

Stochastic methods for quantification of

- Model uncertainty
  - Fine-scale simulations of coarse-grain models
  - Bayesian maximum entropy approach
  - Maximum likelihood Bayesian averaging
- Geologic / geometric uncertainty
  - PDEs on random domains
- Parametric uncertainty
  - PDF methods
  - Moment equations
  - Stochastic FEMs

Uncertainty in Reactive Systems

A (reversible) chemical reaction between n species $A_1, A_2, \ldots, A_n$:

$$\alpha_1 A_1 + \alpha_2 A_2 + \ldots + \alpha_m A_m \rightleftharpoons \alpha_{m+1} A_{m+1} + \ldots + \alpha_n A_n$$

Model: The concentration $C_i(t) \equiv [A_i]$ satisfies a rate equation

$$\frac{\mathrm{d}C_i}{\mathrm{d}t} = F_i(C_1, C_2, ...C_n), \qquad i = 1, \ldots, n$$

Model (and aleatory) uncertainty: Imperfect knowledge about the functional forms of F_i $(i = 1, \ldots, n)$.

Parametric uncertainty: Imperfect knowledge about the coefficients entering the functions F_i $(i = 1, \ldots, n)$ and/or initial concentrations.

Quantification of Model Uncertainty

Master equation: PDF of collisions between the molecules of A_i

Modified Gillespie algorithm: Reaction PDF $P(\tau, \mu)$ for reaction μ to occur in the infinitesimal time interval $[t + \tau, t + \tau + \Delta\tau]$ given a certain state at time t.

Residence time τ, during which no reactions occur, depends upon the total molecular population of all reacting species and reflects the randomness of collisions.

A constant deterministic value to τ corresponds to standard reaction rate equations

$$\frac{\mathrm{d}C_i}{\mathrm{d}t} = F_i(C_1, C_2, ...C_n), \qquad i = 1, \ldots, n$$

Quantification of Model Uncertainty (cntd.)

Modified Gillespie algorithm:

1. Compute the total number of reacting pairs of molecules available for each reaction a_i, and compute their sum $a_0 = \sum a_i$
2. Generate random numbers r_1 and r_2 on the uniform unit interval and m uniformly random on the interval $[1, 10]$
3. Compute $\tau = -m a_0^{-1} \ln r_1$
4. Determine which reaction μ occurs by taking μ to be that integer for which $\sum_{j=1}^{\mu-1} a_j < r_2 a_0 \leq \sum_{j=1}^{\mu} a_j$
5. Update time by τ and molecular levels for reaction μ (decrease reactants by 1 and increase products by 1)
6. Repeat steps 1-5 until either of the reactant population goes to zero or steady state is reached

Example: Neptunium Ion Exchange

Reacting system:

$$NpO_2^+ + \{tAl - Na\} \rightarrow \{tAl - NpO_2\} + Na^+$$

$$\{tAl - NpO_2\} + Na^+ \rightarrow NpO_2^+ + \{tAl - Na\}$$

$$Ca^{2+} + 2\{tAl - Na\} \rightarrow \{2tAl - Ca\} + 2Na^+$$

$$\{2tAl - Ca\} + 2Na^+ \rightarrow Ca^{2+} + 2\{tAl - Na\}$$

Standard deterministic model:

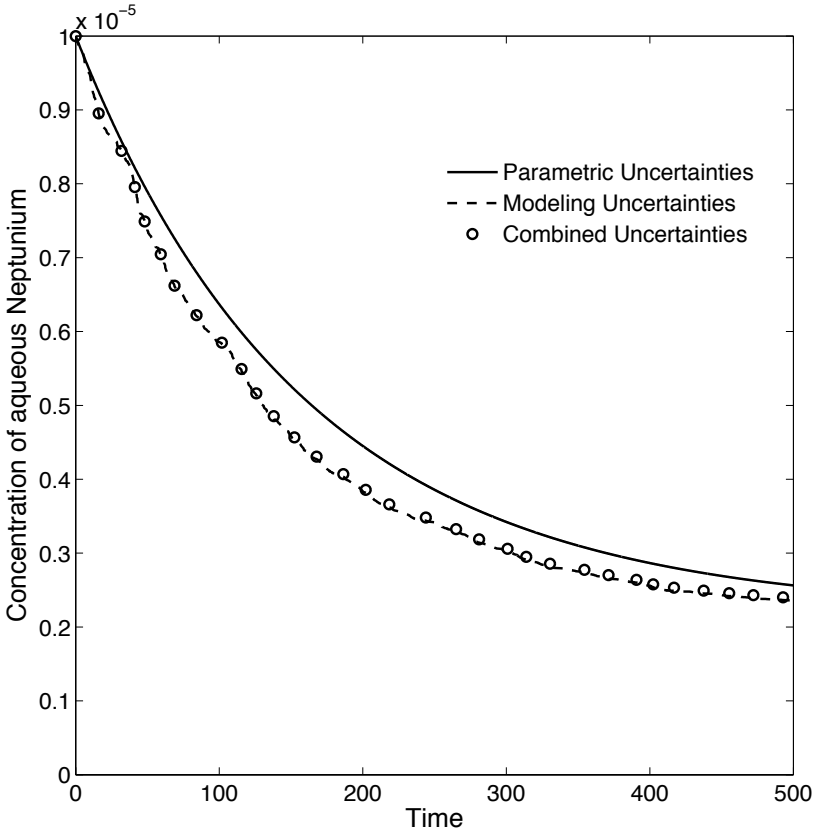
$$\frac{\mathrm{d}C_1}{\mathrm{d}t} = -k_1C_1C_4 + k_2C_2C_3 - 2k_3C_1^2C_6 + 2k_4C_2^2C_5,$$

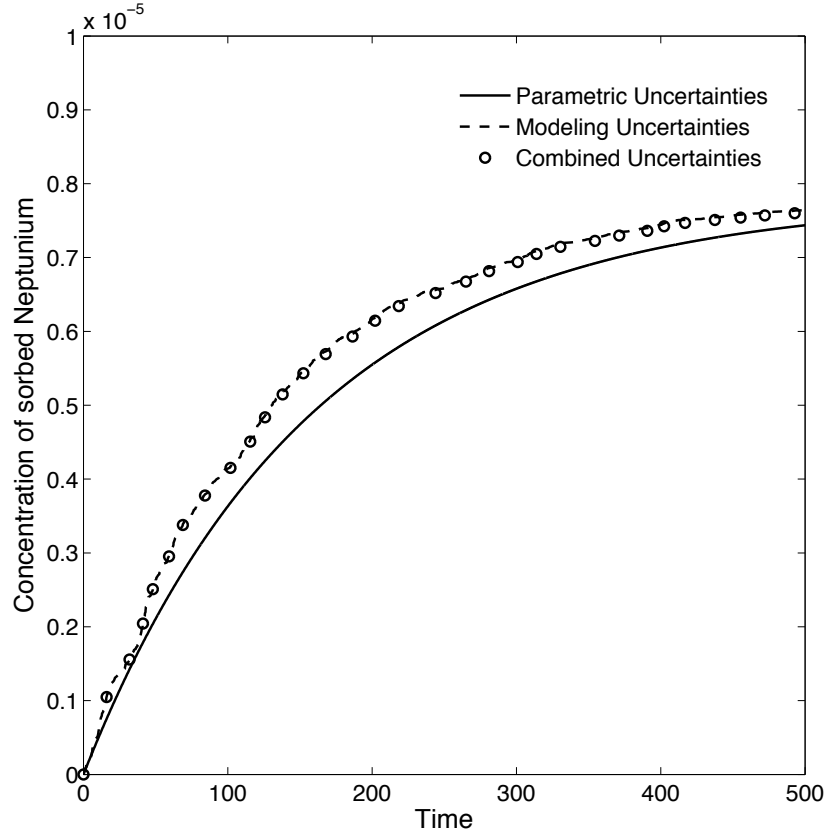
$$\frac{\mathrm{d}C_2}{\mathrm{d}t} = k_1C_1C_4 - k_2C_2C_3 + 2k_3C_1^2C_6 - 2k_4C_2^2C_5,$$

$$\frac{\mathrm{d}C_3}{\mathrm{d}t} = k_1C_1C_4 - k_2C_2C_3, \qquad \frac{\mathrm{d}C_4}{\mathrm{d}t} = -k_1C_1C_4 + k_2C_2C_3,$$

$$\frac{\mathrm{d}C_5}{\mathrm{d}t} = k_3{C_1}^2C_6 - k_4C_2^2C_5, \qquad \frac{\mathrm{d}C_6}{\mathrm{d}t} = -k_3{C_1}^2C_6 + k_4C_2^2C_5$$

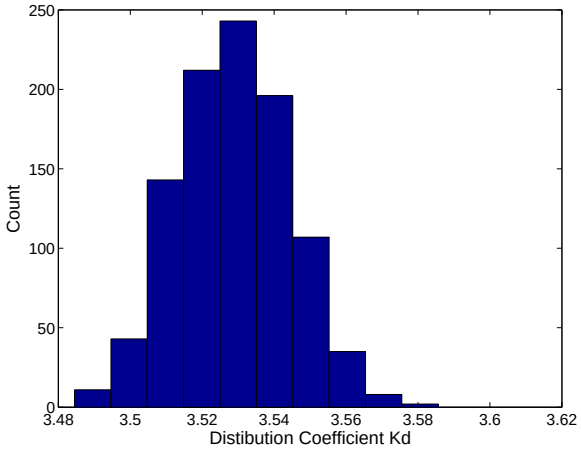
Neptunium Ion Exchange: UQ



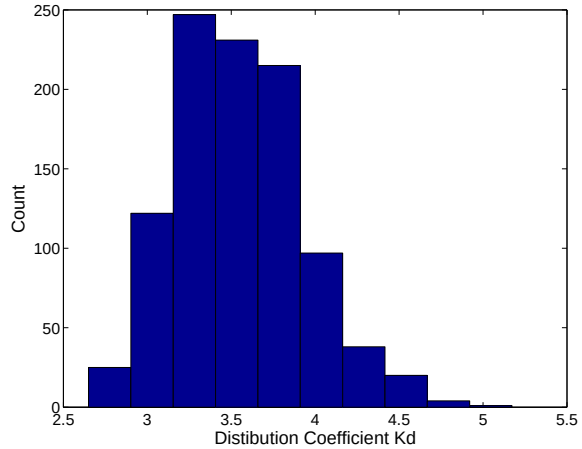


Neptunium Ion Exchange: UQ

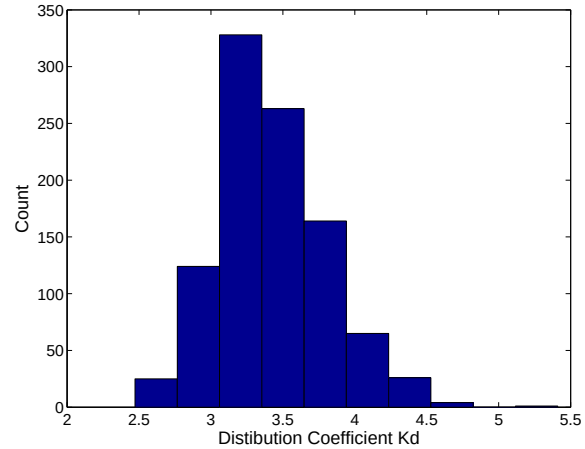
Distribution coefficient $K_d = C_3/C_4$:



Parametric U



Model U



Joint U

Reactive transport (*instantaneous* adsorption):

$$\omega R \frac{\partial C}{\partial t} = \nabla \cdot \left(\tilde{\mathbf{D}} \nabla C \right) - \nabla \cdot (\mathbf{v} C), \qquad R_c = 1 + \frac{1-\omega}{\omega} \rho_s K_d$$

Quantification of Parametric Uncertainty

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

Strategy for Random Domain Decomposition

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

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

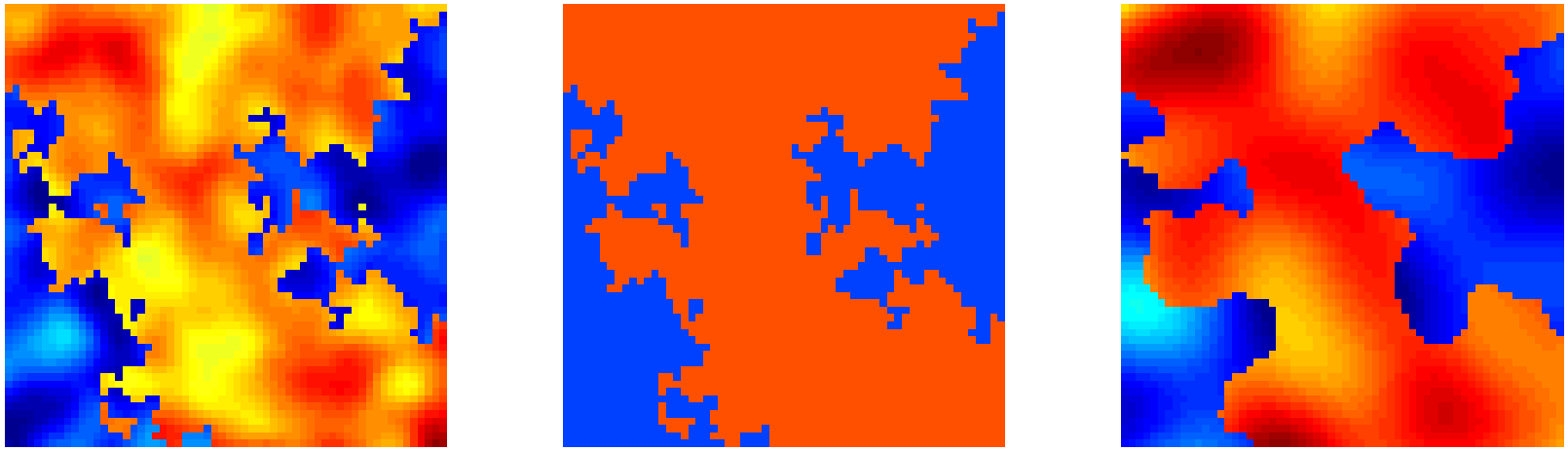
$$\int \mathcal{L}_{\{\Pi\}} u \; f(\{\Pi\}|\gamma) \; d\{\Pi\} \quad \rightarrow \quad \langle u|\gamma \rangle$$

- Step 3: Averaging over random geometries

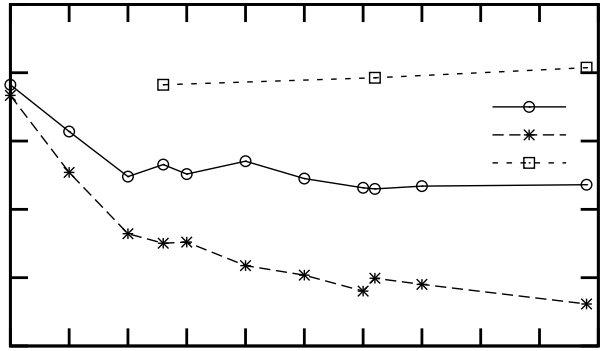
$$\langle u \rangle = \int \langle u|\Gamma \rangle \; f(\Gamma) \; d\Gamma$$

Step 1: Parameter Space Decomposition

Parameter field:



- Indicator kriging
- Statistical learning theory (SVM)
- Nearest neighbor classification



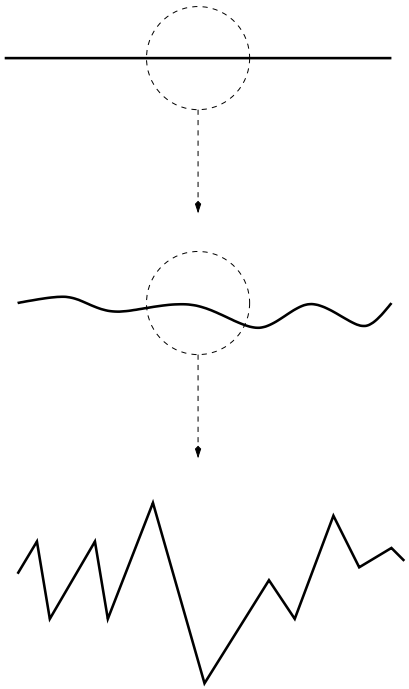
Steps 2-3: PDEs on Random Domains

- Find $u = u(x;\omega) \in \overline{D} \times \Omega \to \mathbb{R}$,

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

- Probability space $(\Omega, \mathcal{A}, \mathcal{P})$
  - Ω is a set of events
  - $\mathcal{A} = 2^{\Omega}$ is the σ-algebra
  - $\mathcal{P}$ is a probability measure

- Physical space: $D(\omega) \in \mathbb{R}^d$
  - $\partial D(\omega)$ is sufficiently smooth



Scales of roughness

Assumption: The problem is well-posed $\mathcal{P}$-a.e. in Ω.

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 i: Stochastic mapping

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

- Step ii: Solving a random equation in a deterministic domain

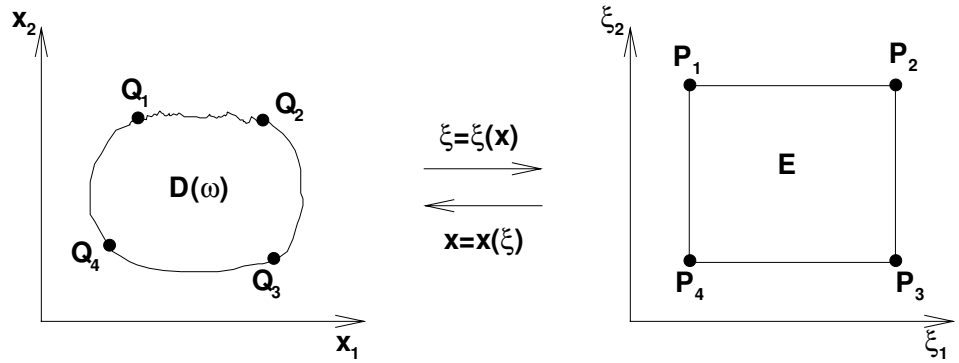
$$\begin{cases} \mathcal{L}(\xi;u;\omega) = f(\xi;\omega), & \xi \in E \\ \mathcal{B}(\xi;u;\omega) = g(\xi;\omega), & \xi \in \partial E \end{cases}$$

Step i: Random Mapping

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

- Numerical mappings, e.g.,

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



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

- Analytical mappings

Step ii: Solving Stochastic PDEs

- Stochastic PDE

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

- Generalized polynomial chaos expansions

$$u(\xi, t; \omega) = \sum_{i=1}^{N} a_i(\xi, t) \Psi_i(\omega)$$

  - Choose an orthogonal polynomial basis
  - Reduce N

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

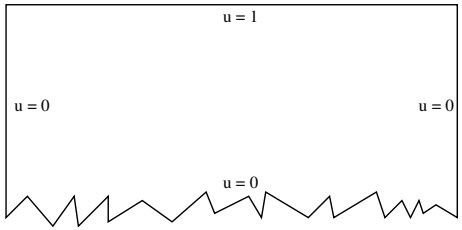
Distribution	Polynomials
Gaussian	Hermite
Gamma	Laguerre
Beta	Jacobi
Uniform	Legendre
Poisson	Charlier
Binomial	Krawtchouk
Negative binomial	Meixner
Hypergeometric	Hahn

Xiu & Karniadakis, SIAM J. Sci. Comp. (2002)

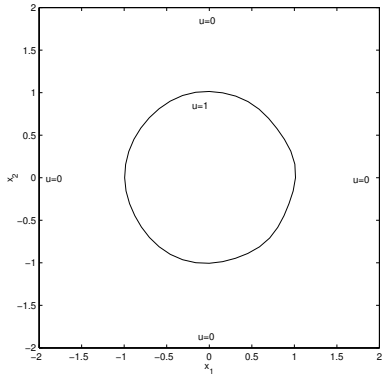
Computational examples

- Steady-state diffusion (SISC, 2006)

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

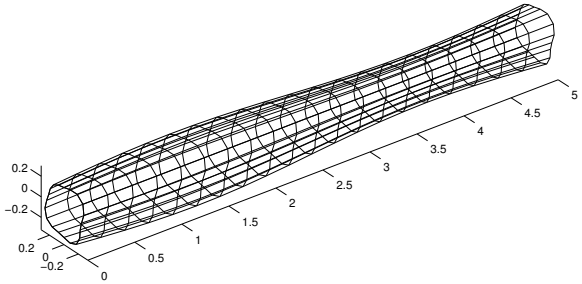


  - Rough channel
  - Rough exclusion
  - Numerical mapping



- Transport by Stocks flow (JCP, 2006)

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



  - Lubrication approximation
  - Analytical mapping

UCSD

Conclusions

- PRA frameworks provide
  - efficient mode reduction strategies
  - comprehensive treatment of structural (model) and parametric uncertainties
  - the use of subjective probabilities, i.e., the reliance on expert knowledge

- PRA is an ideal translator from the language of science to the language of decision makers