Stochastic Modelling of Uncertainty in PDEs and Multilevel Monte Carlo

Robert Scheichl

Institute for Applied Mathematics &
Interdisciplinary Center for Scientific Computing
Universität Heidelberg

Tutorial on *Uncertainty Quantification –*
Efficient Methods for PDEs with Random Coefficients

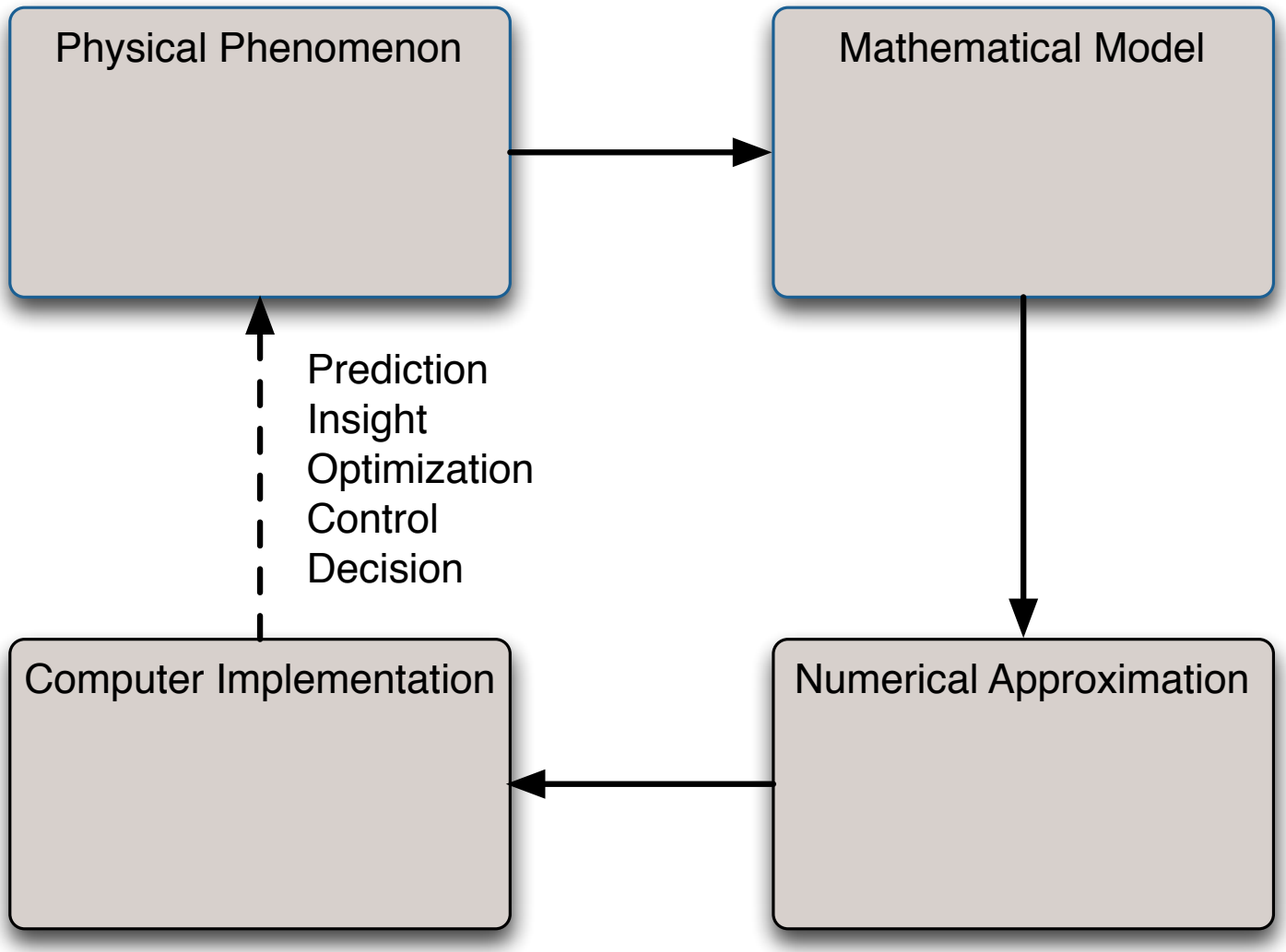
as part of RICAM Special Semester on
"Multivariate Algorithms and their Foundations in Number Theory"

Linz, December 14, 2018

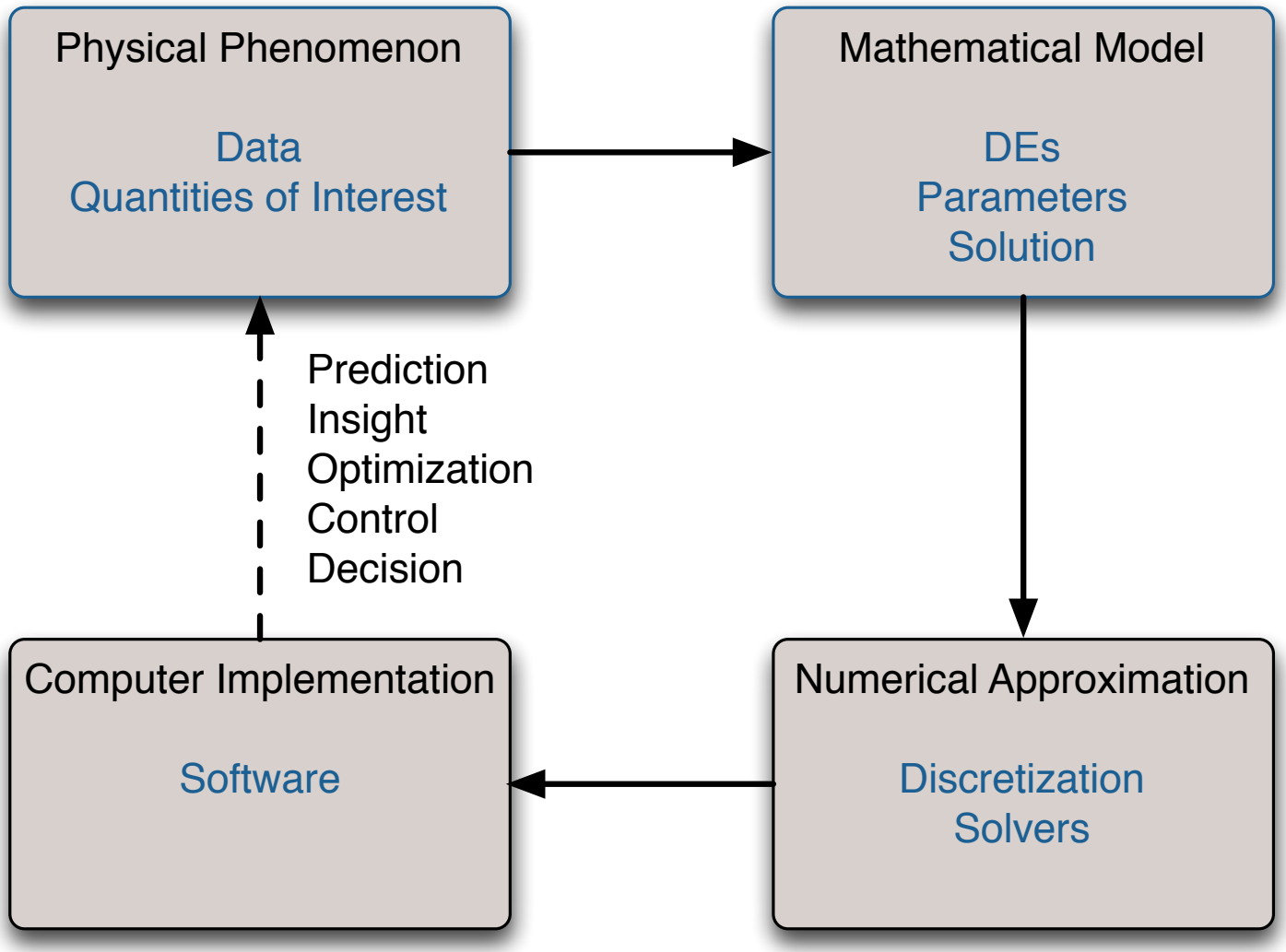
Outline

- **Motivation:** data uncertainty & stochastic modelling
- **Model problem:** subsurface flow and radwaste disposal
- What are the **computational/mathematical challenges**?
- The *Curse of Dimensionality*
- Monte Carlo methods
- **Multilevel Monte Carlo**
- **Numerical Analysis** for PDEs with random coefficients

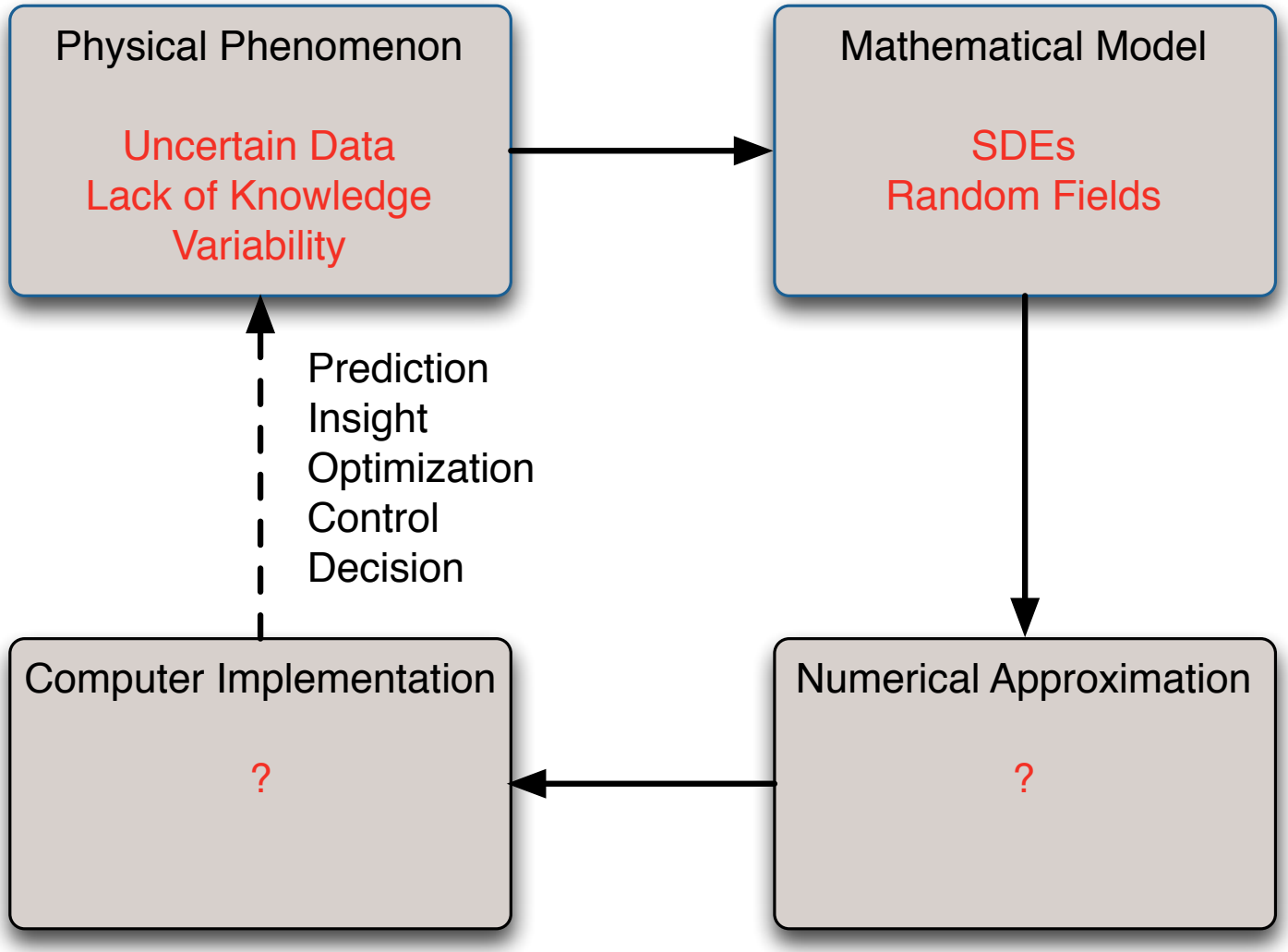
UQ and the Scientific Computing Paradigm



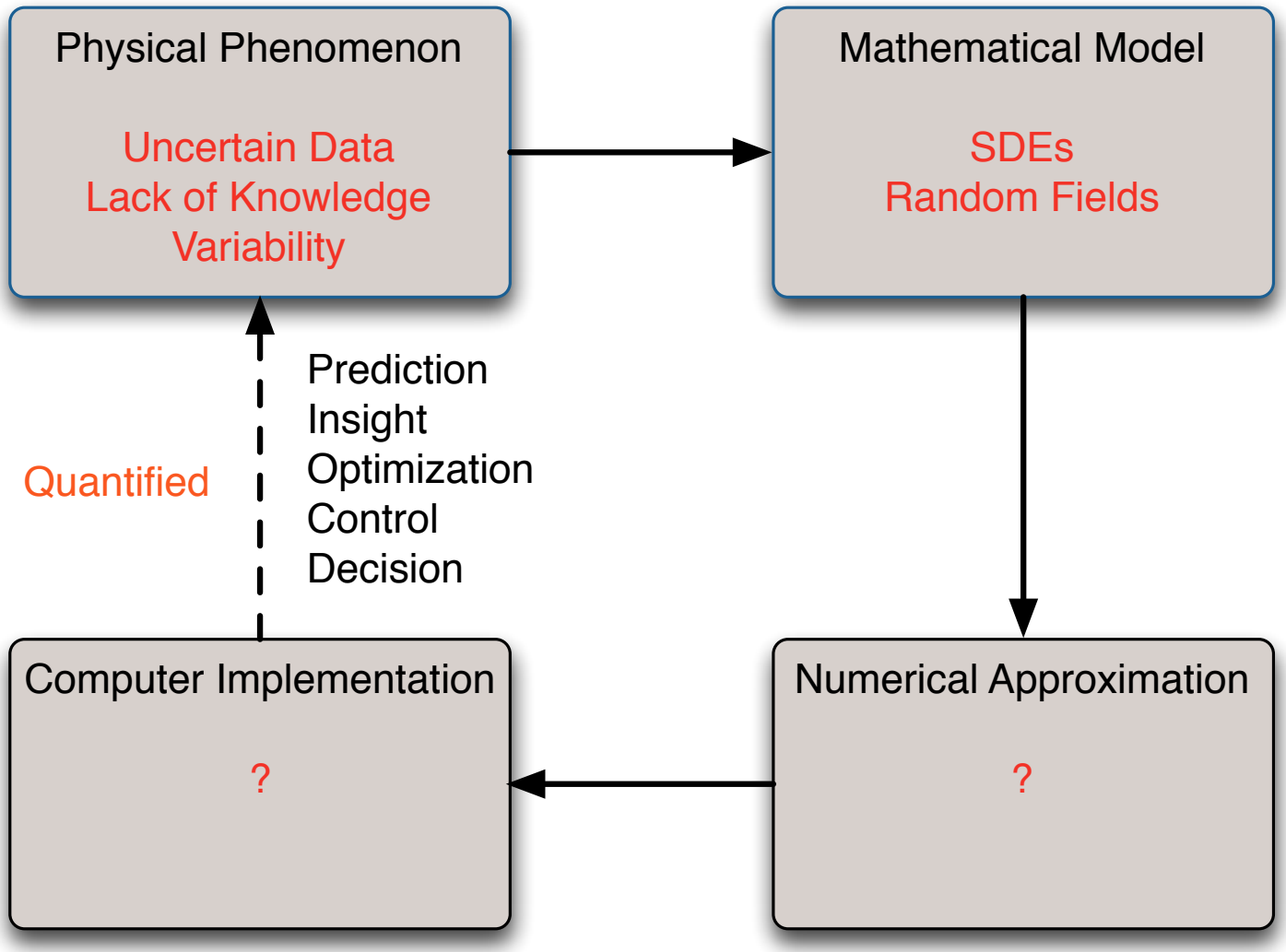
UQ and the Scientific Computing Paradigm



UQ and the Scientific Computing Paradigm



UQ and the Scientific Computing Paradigm



Stochastic Modelling – Many Reasons

- **lack of data** (e.g. data assimilation for weather prediction)
- **data uncertainty** (e.g. uncertainty quantification in subsurface flow)
- **parameter identification** (e.g. Bayesian inference in engineering)
- **unresolvable scales** (e.g. atmospheric dispersion modelling)
- **high dimensionality** (e.g. stochastic simulation in systems biology)

Stochastic Modelling – Many Reasons

- **lack of data** (e.g. data assimilation for weather prediction)
- **data uncertainty** (e.g. uncertainty quantification in subsurface flow)
- **parameter identification** (e.g. Bayesian inference in engineering)
- **unresolvable scales** (e.g. atmospheric dispersion modelling)
- **high dimensionality** (e.g. stochastic simulation in systems biology)

Input: best knowledge about system (PDE), input parameter statistics, measured ouput data with error statistics,...

Stochastic Modelling – Many Reasons

- **lack of data** (e.g. data assimilation for weather prediction)
- **data uncertainty** (e.g. uncertainty quantification in subsurface flow)
- **parameter identification** (e.g. Bayesian inference in engineering)
- **unresolvable scales** (e.g. atmospheric dispersion modelling)
- **high dimensionality** (e.g. stochastic simulation in systems biology)

Input: best knowledge about system (PDE), input parameter statistics, measured ouput data with error statistics,...

Output: statistics of QoIs or of entire state space, e.g.

- Data misfit or rainfall at some location in weather prediction
- Flow or 'breakthrough' time at radioactive waste repository
- Production rate of an oil reservoir
- Amount of ash over Heathrow
- Certification of carbon fibre composite wing in aeronautical engin.

Stochastic Modelling – Many Reasons

- **lack of data** (e.g. data assimilation for weather prediction)
- **data uncertainty** (e.g. uncertainty quantification in subsurface flow)
- **parameter identification** (e.g. Bayesian inference in engineering)
- **unresolvable scales** (e.g. atmospheric dispersion modelling)
- **high dimensionality** (e.g. stochastic simulation in systems biology)

Input: best knowledge about system (PDE), input parameter statistics, measured ouput data with error statistics,...

Output: statistics of QoIs or of entire state space, e.g.

- Data misfit or rainfall at some location in weather prediction
- Flow or 'breakthrough' time at radioactive waste repository
- Production rate of an oil reservoir
- Amount of ash over Heathrow
- Certification of carbon fibre composite wing in aeronautical engin.

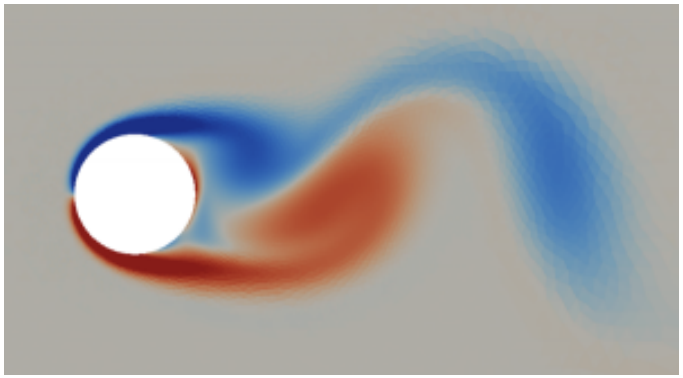
But: Often data very sparse/uncertain $\Rightarrow$ Need good physical models!

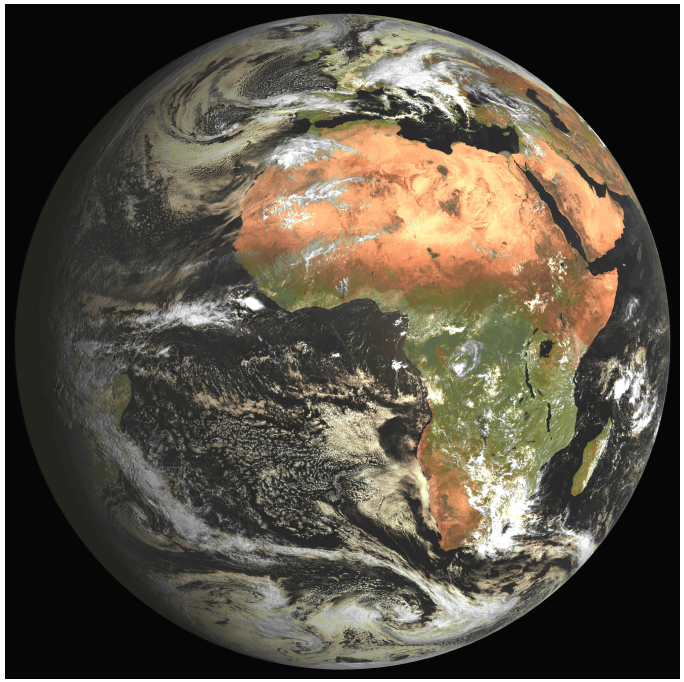
Examples – PDEs with Random Coefficients

- **Navier–Stokes** (e.g. flow around wing, weather forecasting):

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = -\nabla p + \mu\nabla^2\mathbf{v} + f \quad \text{in} \quad \Omega$$

$$\text{subject to IC} \quad v(x,0) = v_0(x) \quad + \quad \text{BCs}$$



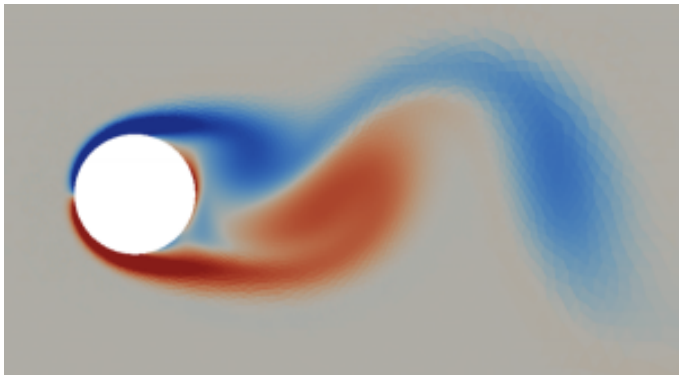


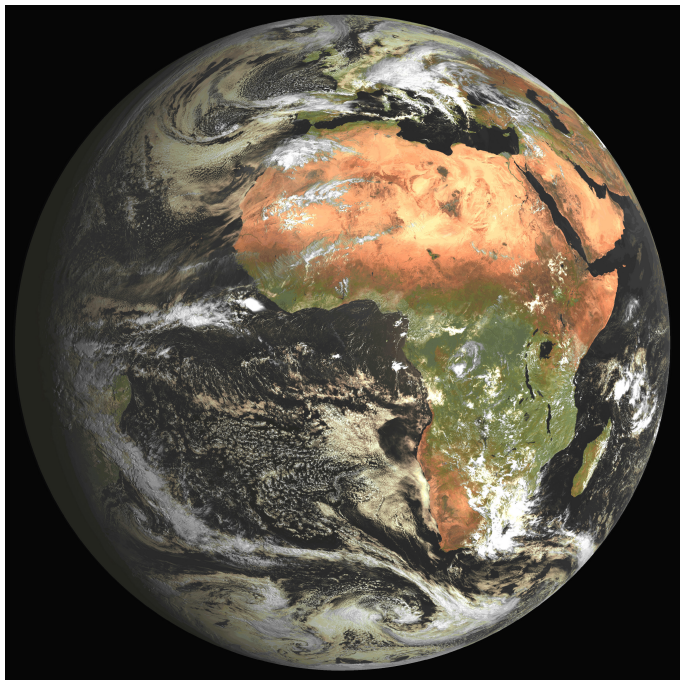
Examples – PDEs with Random Coefficients

- **Navier–Stokes** (e.g. flow around wing, weather forecasting):

$$\rho(\omega)\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = -\nabla p + \mu(\omega)\nabla^2\mathbf{v} + f \quad \text{in} \quad \Omega$$

$$\text{subject to IC} \quad v(x,0) = v_0(x) \quad + \quad \text{BCs}$$



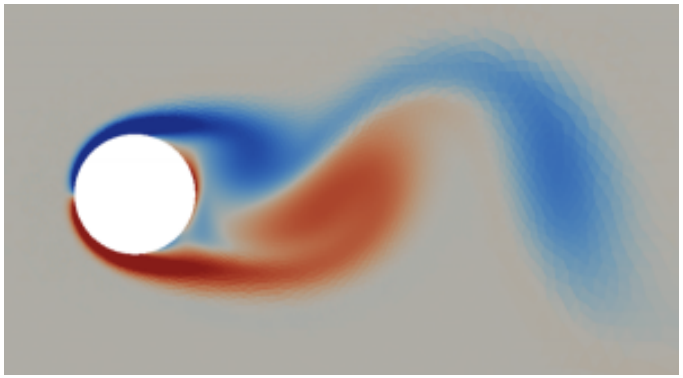


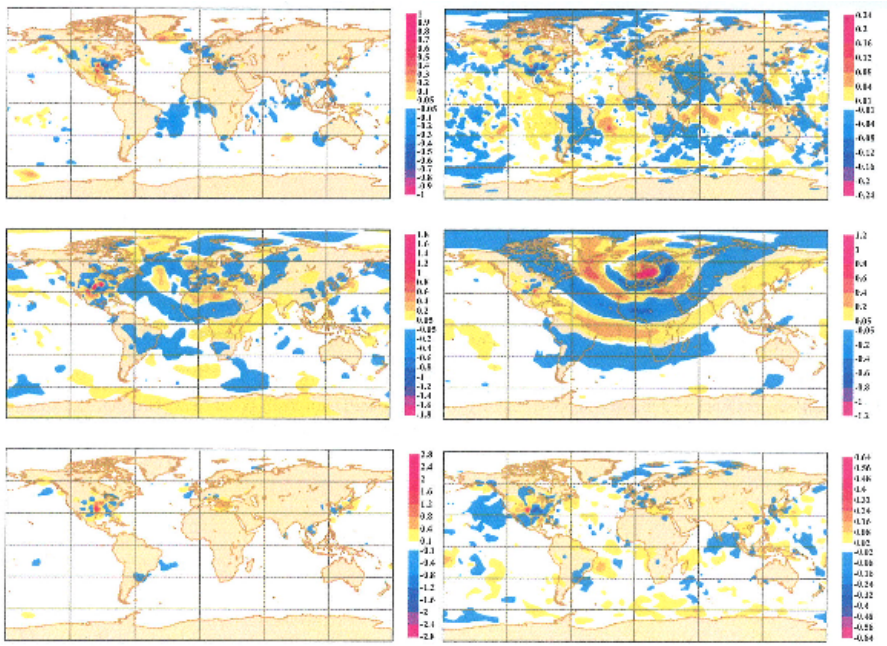
Examples – PDEs with Random Coefficients

- **Navier–Stokes** (e.g. flow around wing, weather forecasting):

$$\rho(\omega)\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = -\nabla p + \mu(\omega)\nabla^2\mathbf{v} + \mathbf{f}(x,\omega) \quad \text{in} \quad \Omega(\omega)$$

$$\text{subject to IC} \quad v(x,0) = v_0(x,\omega) \quad + \quad \text{BCs}$$





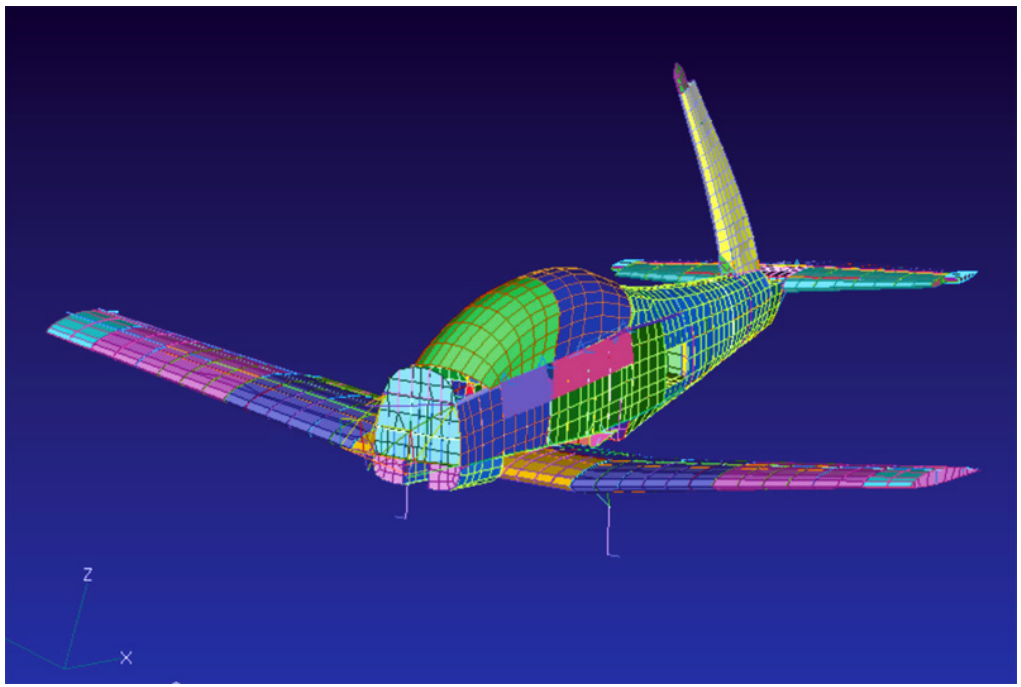
uncertain ICs $\rightarrow$ **data assimilation**

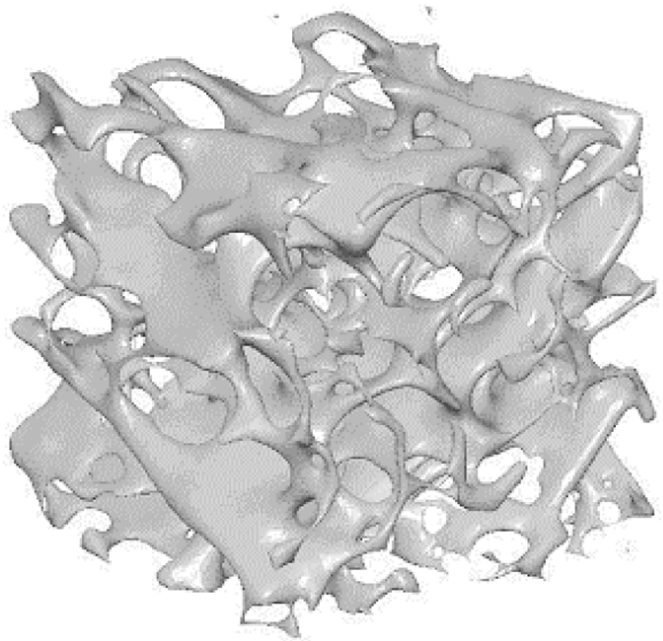
Examples – PDEs with Random Coefficients

- **Structural Mechanics** (e.g. composites, tires or bone):

$$\nabla \cdot \left(\overline{\overline{C}} : \frac{1}{2} \left[\nabla \mathbf{u} + \nabla \mathbf{u}^T \right] \right) + \mathbf{F} = 0 \quad \text{in} \quad \Omega$$

subject to BCs



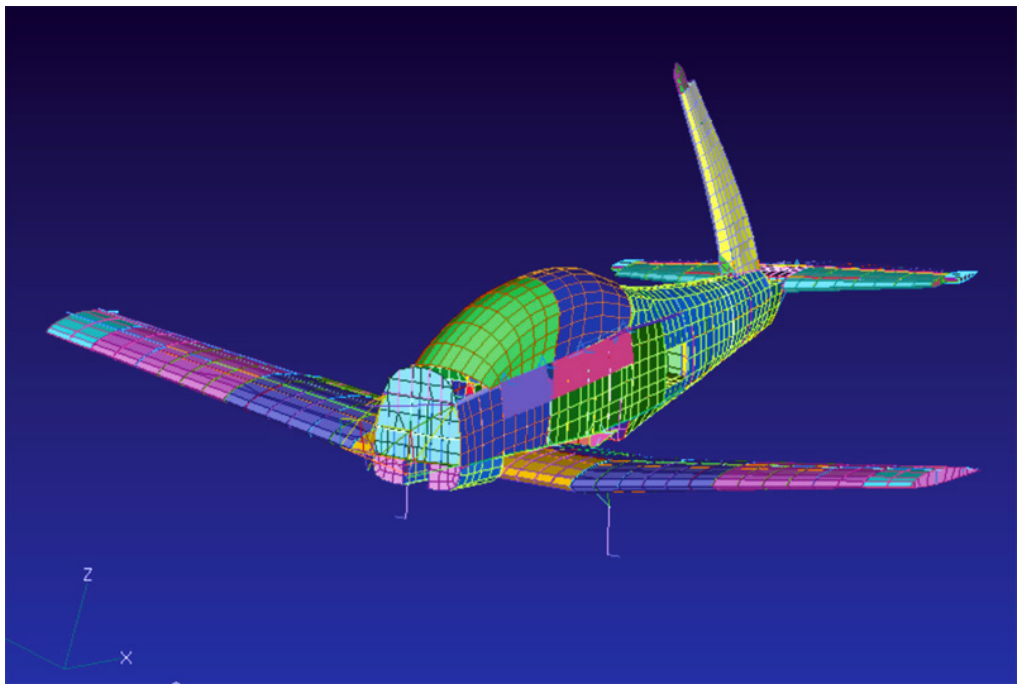


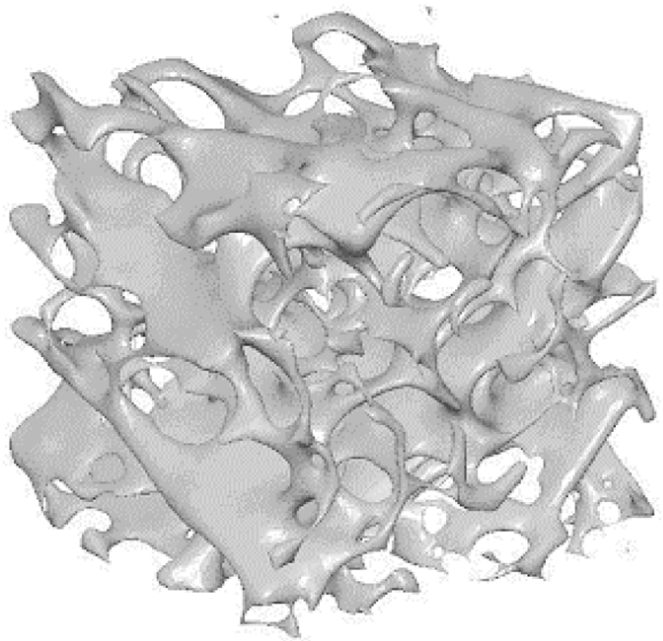
Examples – PDEs with Random Coefficients

- **Structural Mechanics** (e.g. composites, tires or bone):

$$\nabla \cdot \left(\overline{\overline{C}}(x,\omega) : \frac{1}{2}\left[\nabla \mathbf{u} + \nabla \mathbf{u}^T \right] \right) + \mathbf{F} \;=\; 0 \quad \text{in} \quad \Omega$$

subject to BCs



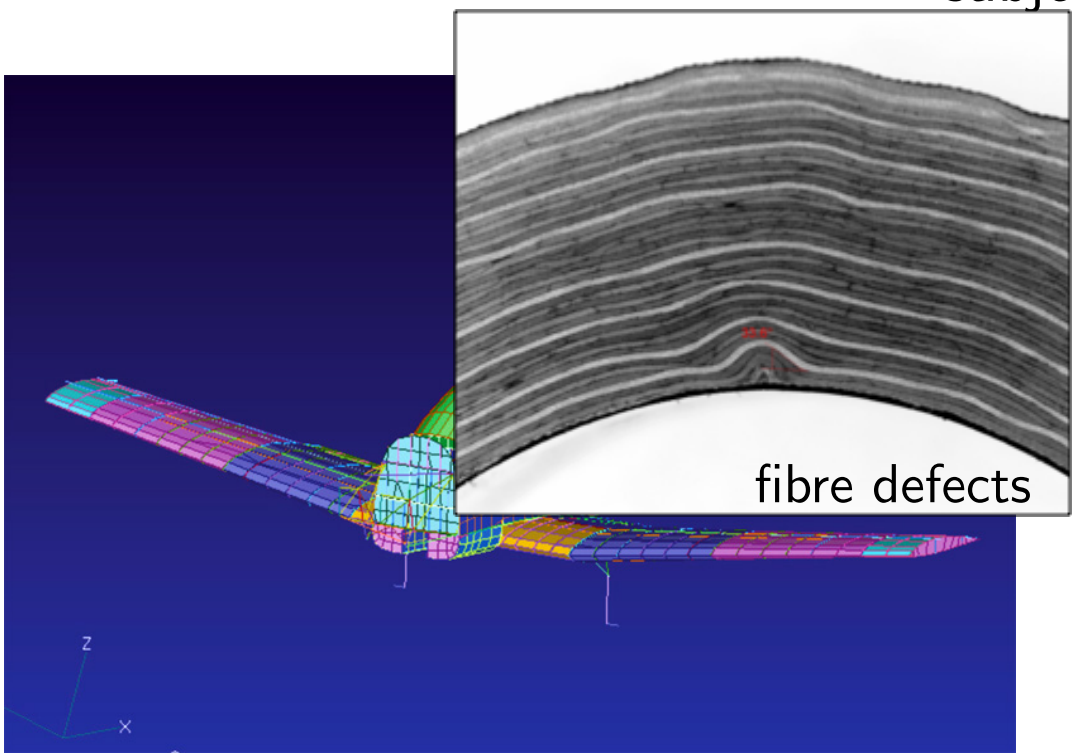


Examples – PDEs with Random Coefficients

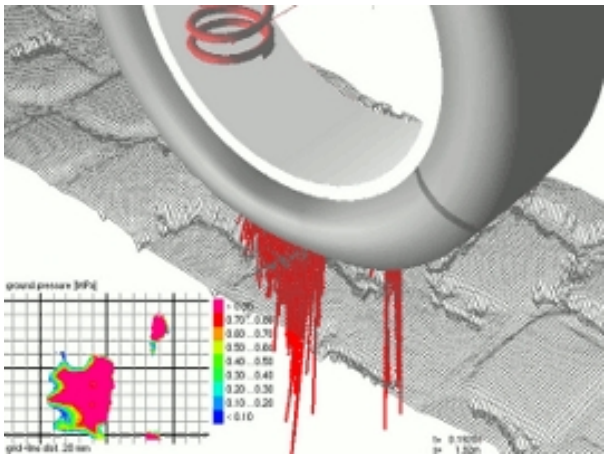
- **Structural Mechanics** (e.g. composites, tires or bone):

$$\nabla \cdot \left(\overline{\overline{C}}(x,\omega) : \frac{1}{2}\left[\nabla \mathbf{u} + \nabla \mathbf{u}^T\right] \right) + \mathbf{F}(x,\omega) \;=\; 0 \quad \text{in} \quad \Omega(\omega)$$

subject to BCs

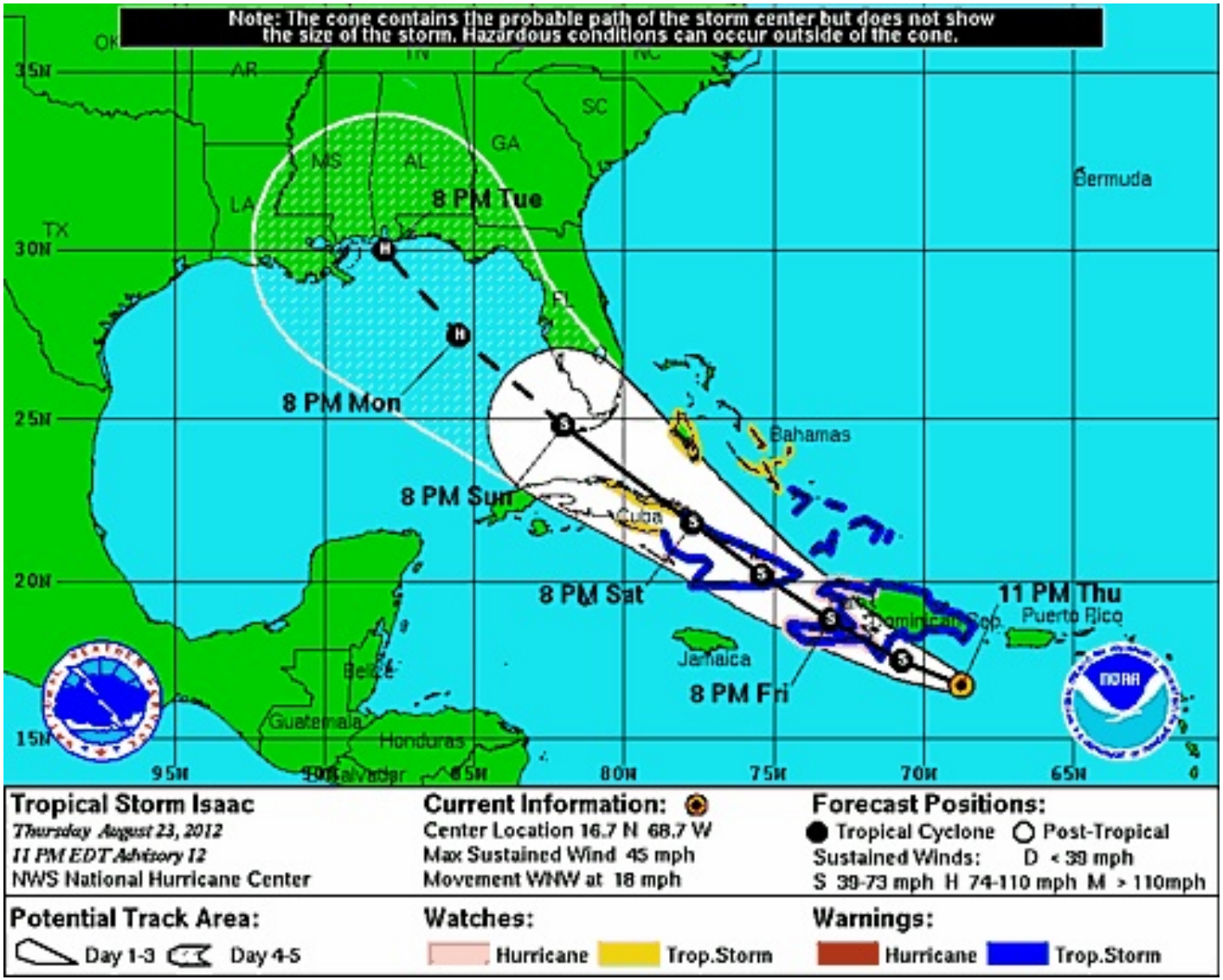


contact on rough surface



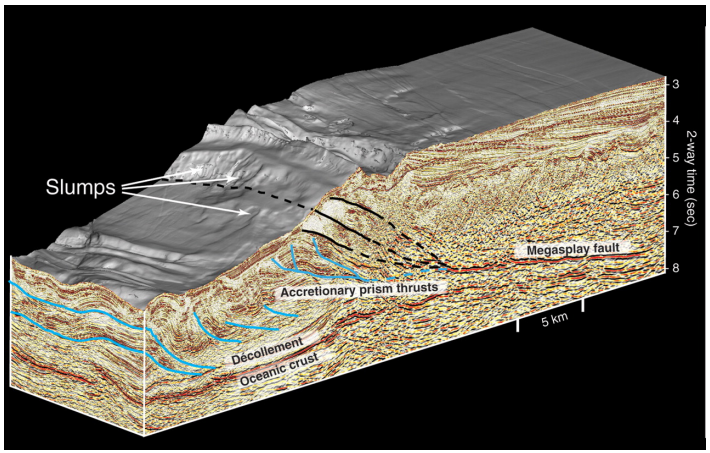
Predicted storm path with *uncertainty cones*

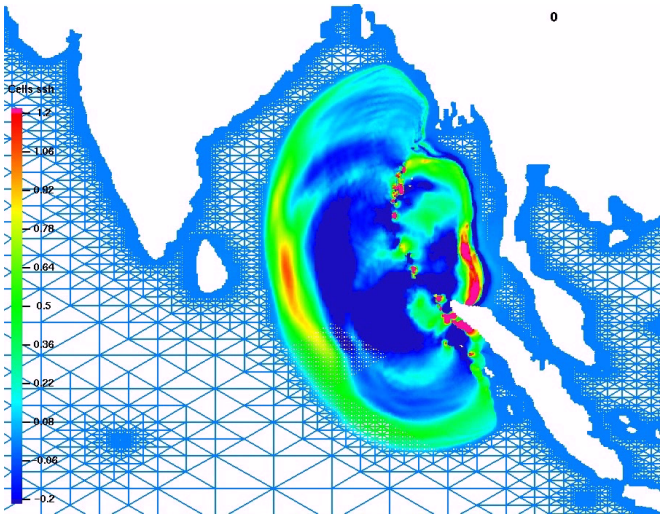
Source: National Hurricane Center, USA



Examples

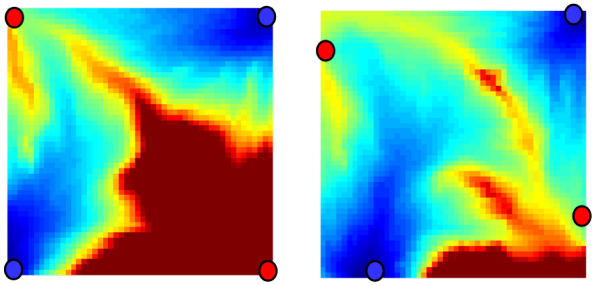
- **Waves in Heterogeneous Media** (e.g. seismic or tsunami):



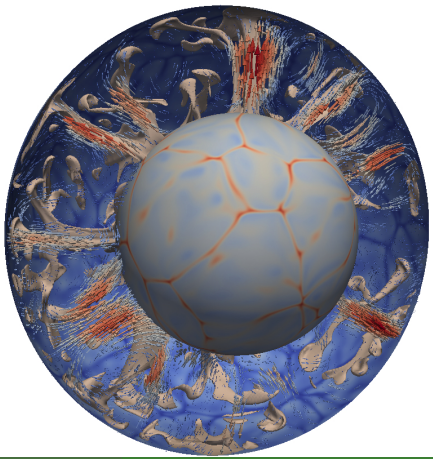


[Behrens et al]

- **Subsurface Fluid Flow** or **mantel convection**:



optimal well placement



[Gmeiner et al]

A Case Study: Radioactive Waste Disposal

Deep geological disposal – favoured by nearly all countries with radwaste

- An area where UQ has played central role in last 30 years.
- Storage in containers in tunnels, several hundred meters deep, in stable geological formations.
- Several barriers: chemical, physical, **geological**.
- Containment must be assured for **at least 10,000 years**
  (uncertainties due to technological complexity & long time scales).
- Main escape route for radionuclides: groundwater pathway.

A Case Study: Radioactive Waste Disposal

Deep geological disposal – favoured by nearly all countries with radwaste

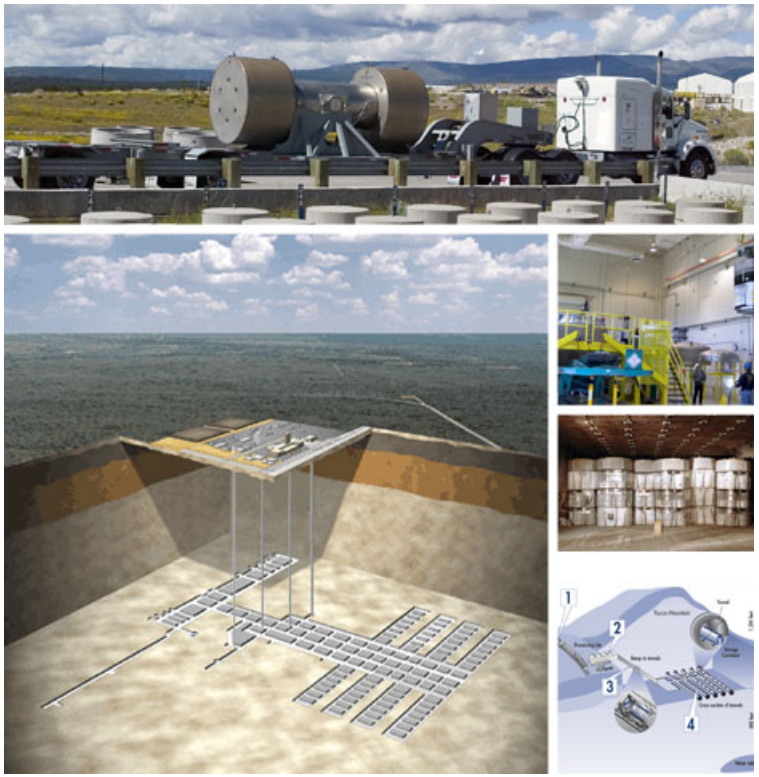
- An area where UQ has played central role in last 30 years.
- Storage in containers in tunnels, several hundred meters deep, in stable geological formations.
- Several barriers: chemical, physical, **geological**.
- Containment must be assured for **at least 10,000 years**
  (uncertainties due to technological complexity & long time scales).
- Main escape route for radionuclides: groundwater pathway.
- **Assessing safety** of potential sites of utmost importance

  long timescales $\rightarrow$ **modelling essential!**
- **Key aspect:** How to quantify uncertainties in the models?

A Case Study: Radioactive Waste Disposal

WIPP – Waste Isolation Pilot Plant

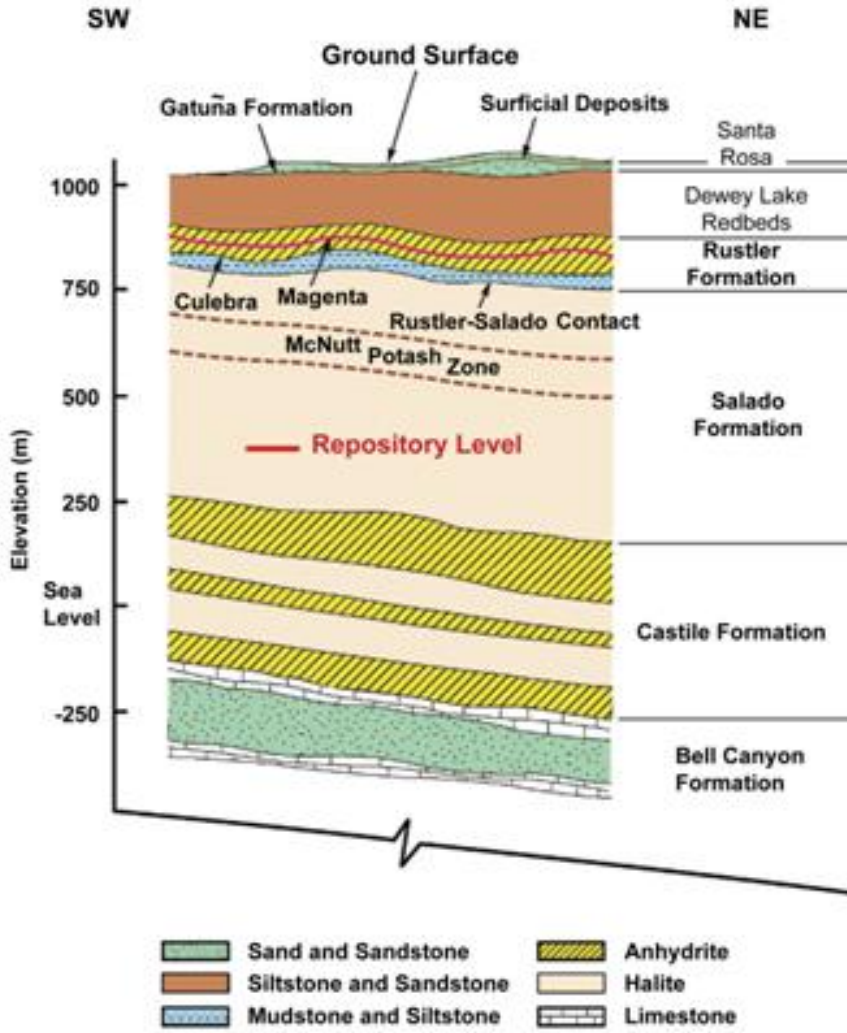
- US DOE repository for radioactive waste situated near Carlsbad, NM (fully operational since 1999).
- Extensive site characterisation and performance assessment since 1976, in particular compliance certification by US EPA (every 5 years)
- Lots of publicly available data. http://www.wipp.energy.gov



A Case Study: Radioactive Waste Disposal

WIPP – Waste Isolation Pilot Plant

- US DOE repository for radioactive waste situated near Carlsbad, NM (fully operational since 1999).
- Extensive site characterisation and performance assessment since 1976, in particular compliance certification by US EPA (every 5 years)
- Lots of publicly available data. http://www.wipp.energy.gov
- Repository at 655m depth in bedded evaporites (primarily halite, a salt)
- Most transmissive rock layer: Culebra Dolomite (principal pathway)

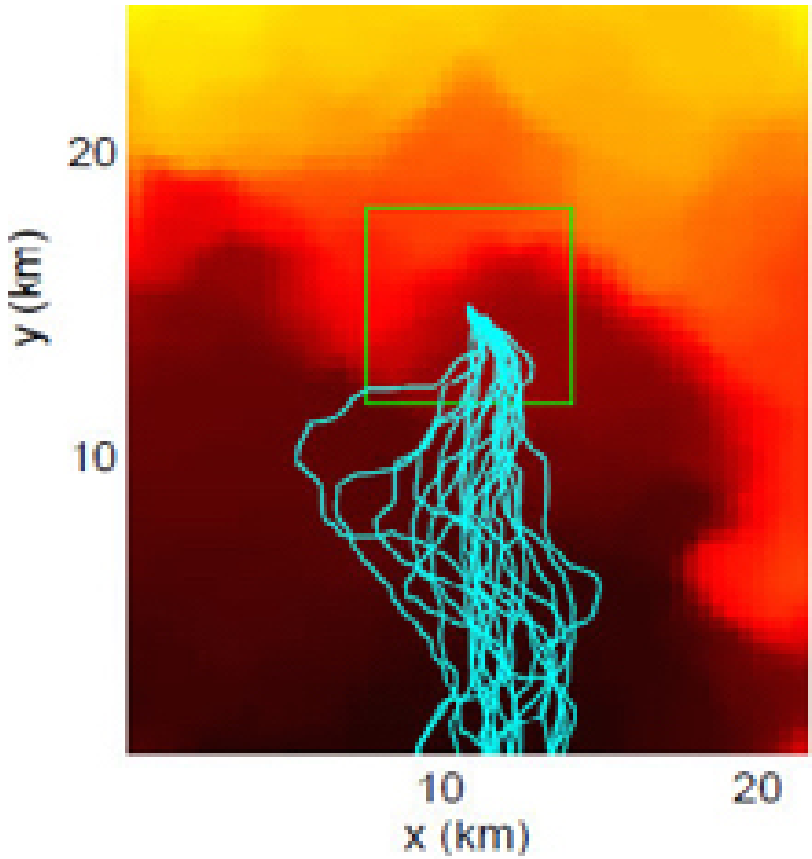


A Case Study: Radioactive Waste Disposal

WIPP UQ Scenario

Important WIPP UQ Scenario:

1. Borehole accidentally drilled into the repository.
2. Radionuclides released into Culebra Dolomite and transported by groundwater.
3. Estimate travel time from release point to boundary of the region.
   (to a good approximation flow is 2D)



Model Problem: Uncertainty in Subsurface Flow

(e.g. risk analysis of radwaste disposal)

$$\textbf{Darcy's Law:} \qquad \vec{q} + k\,\nabla p = \vec{f}$$

$$\textbf{Incompressibility:} \qquad \nabla \cdot \vec{q} = g$$

$$+ \quad \textbf{Boundary Conditions}$$

Model Problem: Uncertainty in Subsurface Flow

(e.g. risk analysis of radwaste disposal)

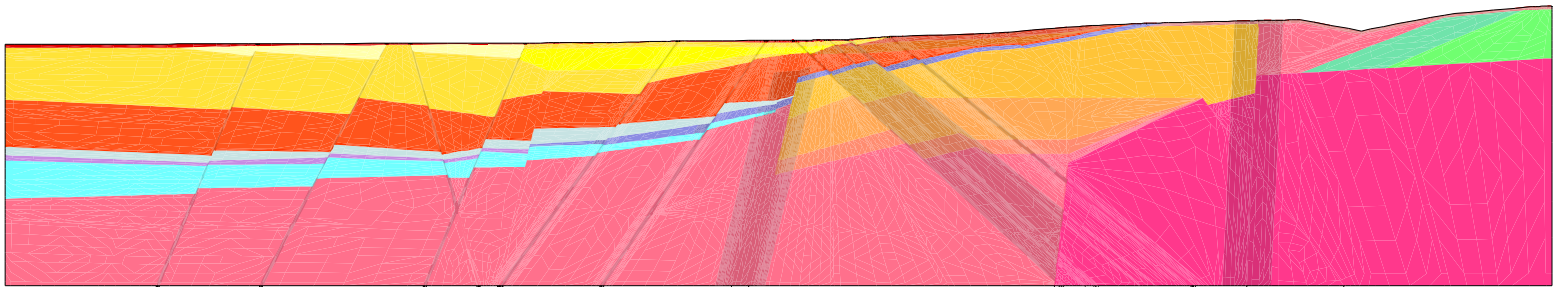
Darcy's Law: $\vec{q} + k(x,\omega)\nabla p = \vec{f}(x,\omega)$

Incompressibility: $\nabla \cdot \vec{q} = g(x,\omega)$

$+$ **Boundary Conditions**

Uncertainty in $k \Longrightarrow$ **Uncertainty** in p & $\vec{q}$

Stochastic Modelling!



Geology at Sellafield (former potential UK radwaste site) ©NIREX UK Ltd.

Stochastic Modelling of Uncertainty (simplified)

Model uncertain conductivity tensor k as a random field $k(x, \omega)$

Stochastic Modelling of Uncertainty (simplified)

Model uncertain conductivity tensor k as a random field $k(x, \omega)$

- $k(x, \omega)$ isotropic, scalar
- $\log k(x, \omega)$ = **Gaussian** field

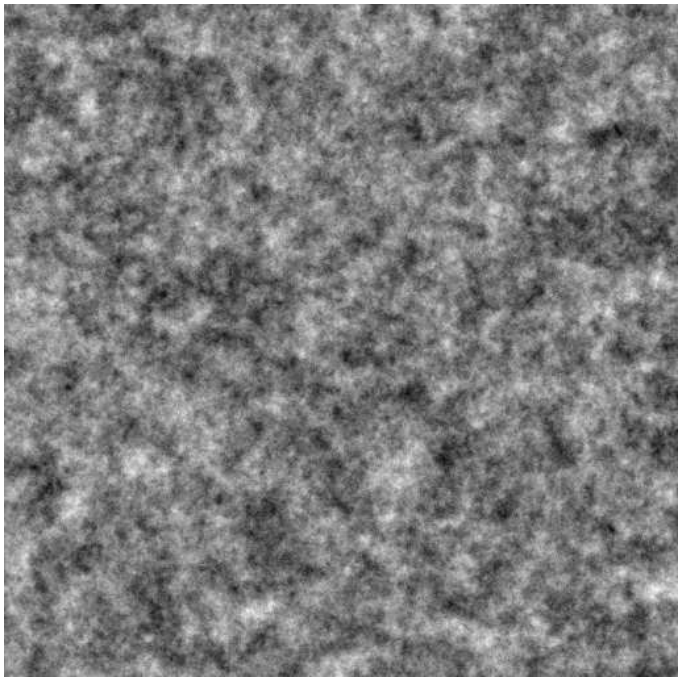
  with isotropic covariance (e.g. Matérn):

$$R(x, y) := \sigma^2 \rho_\nu \left(\tfrac{1}{\lambda} \|x - y\| \right)$$

- Usually: $\sigma^2 > 1$, $\lambda < 1$ & $\nu < 1$

  e.g. $\rho_{1/2}(r) = \exp(-r)$ or $\rho_\infty(r) = \exp(-r^2)$

- Compute statistics (eg. moments) of functionals of p and $\vec{q}$, e.g.
  - point evaluations, norms, averages
  - **breakthrough time** (to boundary)



realisation ($\lambda = \frac{1}{64}$, $\sigma^2 = 8$)

contrast: $\max_{x,y} \frac{k(x)}{k(y)} = \mathcal{O}(10^9)$

(Reasonably good fit to some field data [Gelhar, 1975], [Hoeksema et al, 1985])

A Case Study: Radioactive Waste Disposal

Example: Matérn Family of Covariance Kernels

$$c(\mathbf{x},\mathbf{y}) = c_{\boldsymbol{\theta}}(r) = \frac{\sigma^2}{2^{\nu-1}\,\Gamma(\nu)} \left(\frac{2\sqrt{\nu}\,r}{\lambda}\right)^{\nu} K_\nu \left(\frac{2\sqrt{\nu}\,r}{\lambda}\right), \quad r = \|\mathbf{x}-\mathbf{y}\|_2$$

K_ν : modified Bessel function of order ν

Parameters $\boldsymbol{\theta} = (\sigma^2, \lambda, \nu)$

σ^2 : variance

λ : correlation length

ν : smoothness parameter

A Case Study: Radioactive Waste Disposal

Example: Matérn Family of Covariance Kernels

$$c(\mathbf{x},\mathbf{y}) = c_{\boldsymbol{\theta}}(r) = \frac{\sigma^2}{2^{\nu-1}\,\Gamma(\nu)} \left(\frac{2\sqrt{\nu}\,r}{\lambda}\right)^{\nu} K_\nu \left(\frac{2\sqrt{\nu}\,r}{\lambda}\right), \quad r = \|\mathbf{x}-\mathbf{y}\|_2$$

K_ν : modified Bessel function of order ν

Parameters $\boldsymbol{\theta} = (\sigma^2, \lambda, \nu)$

σ^2 : variance

λ : correlation length

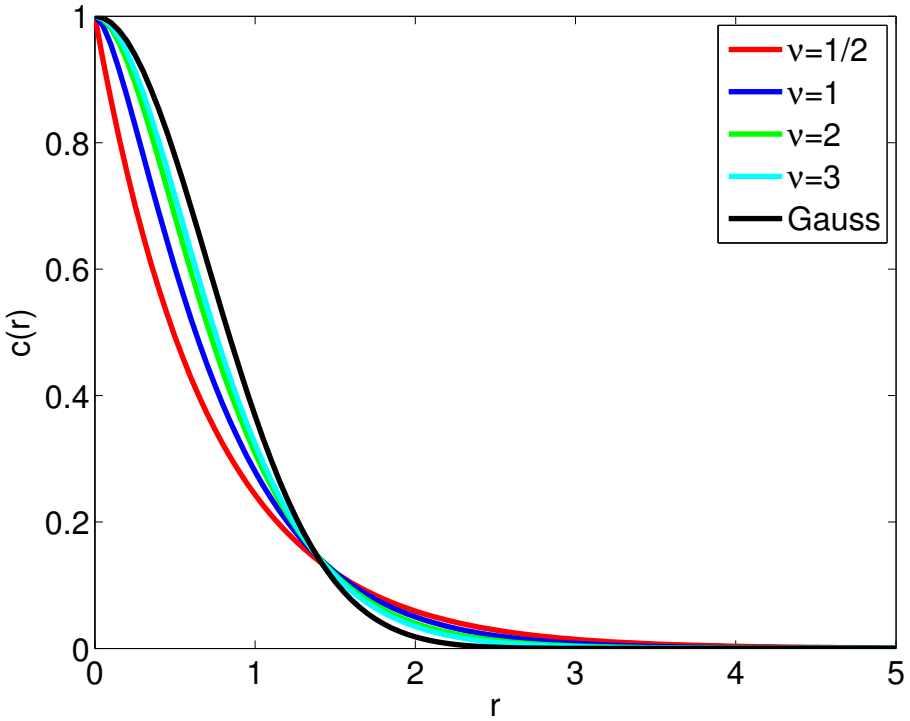
ν : smoothness parameter

Special cases:

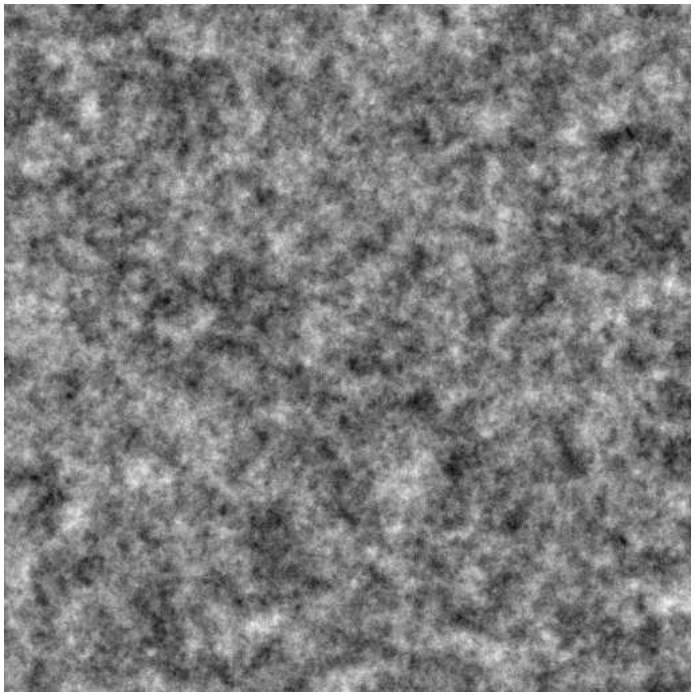
$\nu = \frac{1}{2}$:	$c(r) = \sigma^2 \exp(-\sqrt{2}r/\lambda)$	exponential covariance
$\nu = 1$:	$c(r) = \sigma^2 \left(\frac{2r}{\lambda}\right) K_1 \left(\frac{2r}{\lambda}\right)$	Bessel covariance
$\nu \to \infty$:	$c(r) = \sigma^2 \exp(-r^2/\lambda^2)$	Gaussian covariance

A Case Study: Radioactive Waste Disposal

Matérn Covariance Functions



$\sigma^2 = 1,\ \lambda = 1$



$\sigma^2 = 8,\ \lambda = 1/64,\ \nu = 1/2$

Smoothness: Realisations $Z(\cdot, \omega) \in C^\eta(D)$ (Hölder), for any $\eta < \nu$.

A Case Study: Radioactive Waste Disposal

Sampling from Z – Karhunen-Loève expansion

Since $c(\mathbf{x}, \mathbf{y})$ is symmetric positive semidefinite and continuous, the covariance operator

$$C : L^2(D) \to L^2(D), \quad (Cu)(\mathbf{x}) = \int_D u(\mathbf{y}) c(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{y}$$

is selfadjoint, compact, nonnegative. Hence, the eigenvalues $\{\mu_m\}_{m \in \mathbb{N}}$ form a non-increasing sequence, accumulating at most at 0.

Karhunen-Loève expansion (converges in $L^2_{\mathbf{P}}(\Omega; L^\infty(D))$):

$$Z(x, \omega) = \overline{Z}(\mathbf{x}) + \sum_{m=1}^{\infty} \sqrt{\mu_m} \, \phi_m(\mathbf{x}) \, Y_m(\omega)$$

where $\{\phi_m\}_{m \in \mathbb{N}}$ are normalised eigenfunctions and $Y_m \sim N(0, 1)$ i.i.d.

A Case Study: Radioactive Waste Disposal

Sampling from Z – Karhunen-Loève expansion

Since $c(\mathbf{x}, \mathbf{y})$ is symmetric positive semidefinite and continuous, the covariance operator

$$C : L^2(D) \to L^2(D), \quad (Cu)(\mathbf{x}) = \int_D u(\mathbf{y}) c(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{y}$$

is selfadjoint, compact, nonnegative. Hence, the eigenvalues $\{\mu_m\}_{m \in \mathbb{N}}$ form a non-increasing sequence, accumulating at most at 0.

Truncated Karhunen-Loève expansion (converges in $L^2_{\mathbf{P}}(\Omega; L^\infty(D))$):

$$Z_s(x, \omega) = \overline{Z}(\mathbf{x}) + \sum_{m=1}^{s} \sqrt{\mu_m} \, \phi_m(\mathbf{x}) \, Y_m(\omega)$$

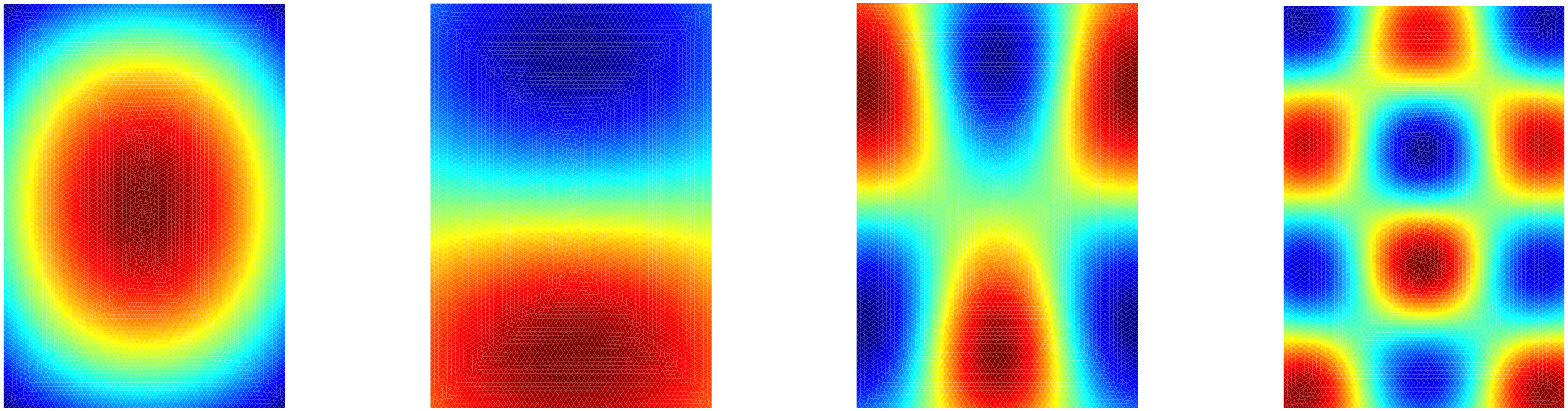
where $\{\phi_m\}_{m \in \mathbb{N}}$ are normalised eigenfunctions and $Y_m \sim N(0, 1)$ i.i.d.

In practice, approximate Z by truncated KL expansion, using only s terms.

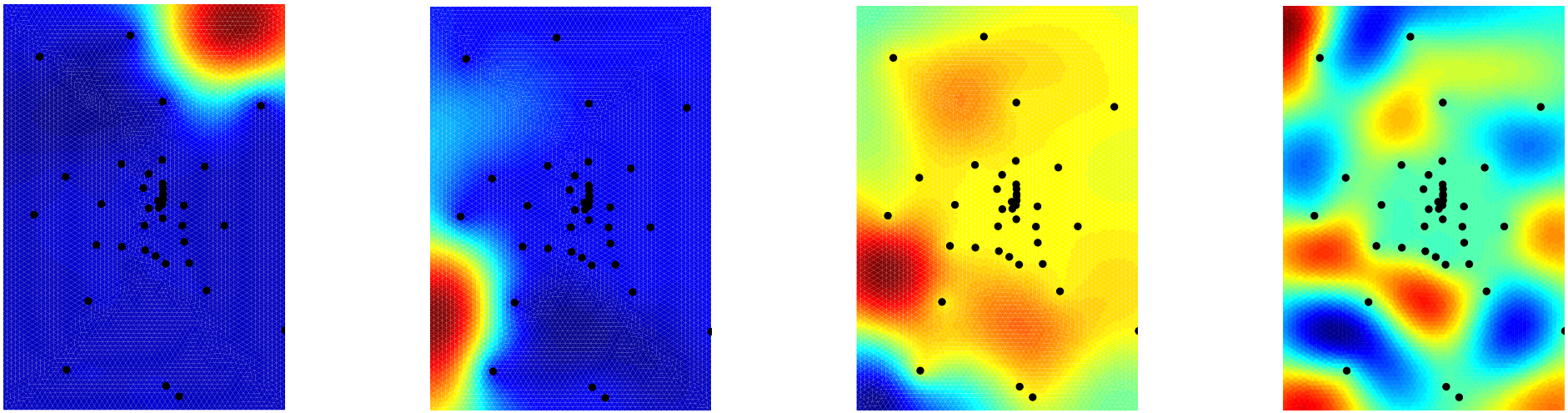
A Case Study: Radioactive Waste Disposal

WIPP KL modes for $Z = \log k$ conditioned on 38 transmissivity observations

$\phi_m(\mathbf{x})$ unconditioned, $m = 1, 2, 9, 16$



$\phi_m(\mathbf{x})$ conditioned on data, $m = 1, 2, 9, 16$



Computational Challenges

Simulating PDEs with Highly Heterogeneous Random Coefficients

$$-\nabla \cdot (k(\mathbf{x},\omega)\nabla p(\mathbf{x},\omega)) = f(\mathbf{x},\omega), \quad \mathbf{x} \in D \subset \mathbb{R}^d, \ \omega \in \Omega \text{ (prob. space)}$$

Computational Challenges

Simulating PDEs with Highly Heterogeneous Random Coefficients

$$-\nabla \cdot (k(\mathbf{x},\omega)\nabla p(\mathbf{x},\omega)) = f(\mathbf{x},\omega), \quad \mathbf{x} \in D \subset \mathbb{R}^d, \ \omega \in \Omega \text{ (prob. space)}$$

- **Sampling** from random field $\log k(\mathbf{x},\omega)$ (correlated Gaussian):
  - truncated Karhunen-Loève expansion of $\log k$ (see above)
  - matrix factorisation, e.g. circulant embedding (FFT)
  - via pseudodifferential "precision" operator (PDE solves)

Computational Challenges

Simulating PDEs with Highly Heterogeneous Random Coefficients

$$-\nabla \cdot (k(\mathbf{x},\omega)\nabla p(\mathbf{x},\omega)) = f(\mathbf{x},\omega), \quad \mathbf{x} \in D \subset \mathbb{R}^d, \ \omega \in \Omega \text{ (prob. space)}$$

- **Sampling** from random field $\log k(\mathbf{x},\omega)$ (correlated Gaussian):
  - truncated Karhunen-Loève expansion of $\log k$ (see above)
  - matrix factorisation, e.g. circulant embedding (FFT)
  - via pseudodifferential "precision" operator (PDE solves)
- **High-Dimensional Quadrature**:
  - Monte Carlo
  - Multilevel Monte Carlo (**RS**), Quasi-Monte Carlo (**DN**)
  - stochastic Galerkin/collocation (sparse grids, low-rank) (**AN**)

Computational Challenges

Simulating PDEs with Highly Heterogeneous Random Coefficients

$$-\nabla \cdot (k(\mathbf{x},\omega)\nabla p(\mathbf{x},\omega)) = f(\mathbf{x},\omega), \quad \mathbf{x} \in D \subset \mathbb{R}^d, \ \omega \in \Omega \text{ (prob. space)}$$

- **Sampling** from random field $\log k(\mathbf{x},\omega)$ (correlated Gaussian):
  - truncated Karhunen-Loève expansion of $\log k$ (see above)
  - matrix factorisation, e.g. circulant embedding (FFT)
  - via pseudodifferential "precision" operator (PDE solves)
- **High-Dimensional Quadrature**:
  - Monte Carlo
  - Multilevel Monte Carlo (**RS**), Quasi-Monte Carlo (**DN**)
  - stochastic Galerkin/collocation (sparse grids, low-rank) (**AN**)
- **Solve** large number of **multiscale** deterministic PDEs:
  - Efficient discretisation & FE error analysis (mesh size h)
  - Multigrid Methods, AMG, DD Methods (robustness?)

Computational Challenges

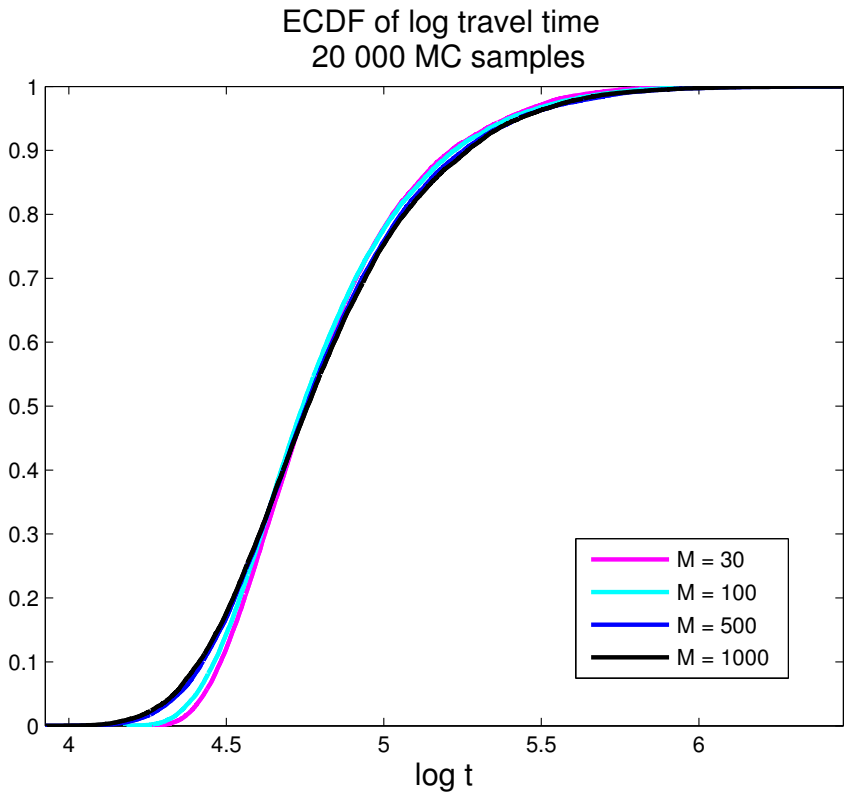
Why is it computationally so challenging?

- Low regularity (global): $k \in C^\eta,\ \eta < \nu < 1 \implies$ **fine** mesh $h \ll 1$
- Large σ^2 & exponential $\implies$ **high** contrast $k_{\max}/k_{\min} > 10^6$
- Small $\lambda \implies$ **multiscale** + **high** stochast. dimension $s > 100$

Computational Challenges

Why is it computationally so challenging?

- Low regularity (global): $k \in C^\eta,\ \eta < \nu < 1 \implies$ **fine** mesh $h \ll 1$
- Large σ^2 & exponential $\implies$ **high** contrast $k_{\max}/k_{\min} > 10^6$
- Small $\lambda \implies$ **multiscale** + **high** stochast. dimension $s > 100$



Source: Ernst et al, 2014 ($s = M$)

Spatial Discretisation – Finite Elements

Write PDE (subject to $p|_{\partial D} \equiv 0$) in weak form: $p(\cdot,\omega) \in H_0^1(D)$ s.t.

$$\int_D \nabla v \cdot (k(x,\omega)\nabla p(x,\omega))\,\mathrm{d}x = \int_D f(x,\omega) v\,\mathrm{d}x, \quad \forall v \in H_0^1(D).$$

- $\exists! p(\cdot,\omega) \in H_0^1(D)$ a.s. in $\omega \in \Omega$ (Lax-Milgram pathwise).

Spatial Discretisation – Finite Elements

Write PDE (subject to $p|_{\partial D} \equiv 0$) in weak form: $p(\cdot,\omega) \in H_0^1(D)$ s.t.

$$\int_D \nabla v \cdot (k(x,\omega)\nabla p(x,\omega))\,\mathrm{d}x = \int_D f(x,\omega) v\,\mathrm{d}x, \quad \forall v \in H_0^1(D).$$

- $\exists! p(\cdot,\omega) \in H_0^1(D)$ a.s. in $\omega \in \Omega$ (Lax-Milgram pathwise).
- Let $V_h \subset H_0^1(D)$ be the space of continuous, piecewise linear FEs w.r.t. a mesh $\mathscr{T}_h$ with mesh width $h > 0$.
- Find $p_h(\cdot,\omega) \in V_h$ that satisfies weak form for all $v_h \in V_h$.

Spatial Discretisation – Finite Elements

Write PDE (subject to $p|_{\partial D} \equiv 0$) in weak form: $p(\cdot,\omega) \in H_0^1(D)$ s.t.

$$\int_D \nabla v \cdot (k(x,\omega)\nabla p(x,\omega))\,\mathrm{d}x = \int_D f(x,\omega) v\,\mathrm{d}x, \quad \forall v \in H_0^1(D).$$

- $\exists! p(\cdot,\omega) \in H_0^1(D)$ a.s. in $\omega \in \Omega$ (Lax-Milgram pathwise).
- Let $V_h \subset H_0^1(D)$ be the space of continuous, piecewise linear FEs w.r.t. a mesh $\mathscr{T}_h$ with mesh width $h > 0$.
- Find $p_h(\cdot,\omega) \in V_h$ that satisfies weak form for all $v_h \in V_h$.
- Write $p_h(x,\omega) := \sum_{i=1}^{M_h} P_i \varphi_i(x)$. Then this is equivalent to the random matrix system

$$A(\omega)\mathbf{P}(\omega) = \mathbf{F}(\omega)$$

with

$$A_{i,j}(\omega) := \int_D \nabla\varphi_j \cdot (k(x,\omega)\nabla\varphi_i)\,\mathrm{d}x, \quad F_i(\omega) := \int_D f(x,\omega)\varphi_i\,\mathrm{d}x$$

Computational Challenges

Standard Monte Carlo Quadrature

$$\mathbf{Y}(\omega) \in \mathbb{R}^s \overset{\text{Model}(h)}{\longrightarrow} \mathbf{P}(\omega) \in \mathbb{R}^{M_h} \overset{\text{Output}}{\longrightarrow} Q_{h,s}(\omega) \in \mathbb{R}$$

random input | state vector | quantity of interest

Computational Challenges

Standard Monte Carlo Quadrature

$$\mathbf{Y}(\omega) \in \mathbb{R}^s \overset{\text{Model}(h)}{\longrightarrow} \mathbf{P}(\omega) \in \mathbb{R}^{M_h} \overset{\text{Output}}{\longrightarrow} Q_{h,s}(\omega) \in \mathbb{R}$$

random input | state vector | quantity of interest

- Real QoI $Q(\omega)$ inaccessible (exact PDE), but we can assume

$$\mathbb{E}[Q_{h,s}] \overset{h\to 0,\ s\to\infty}{\longrightarrow} \mathbb{E}[Q] \quad \text{and} \quad |\mathbb{E}[Q_{h,s} - Q]| = \mathscr{O}(h^{\alpha}) + \mathscr{O}(s^{-\alpha'})$$

- **Standard Monte Carlo** estimator for $\mathbb{E}[Q]$: More detail below!

$$\hat{Q}^{\mathrm{MC}} := \frac{1}{N} \sum_{i=1}^{N} Q_{h,s}^{(i)}$$

where $\{Q_{h,s}^{(i)}\}_{i=1}^{N}$ are i.i.d. samples computed with Model(h)

Computational Challenges

Standard Monte Carlo Quadrature

$$\mathbf{Y}(\omega) \in \mathbb{R}^s \overset{\text{Model}(h)}{\longrightarrow} \mathbf{P}(\omega) \in \mathbb{R}^{M_h} \overset{\text{Output}}{\longrightarrow} Q_{h,s}(\omega) \in \mathbb{R}$$

random input | state vector | quantity of interest

- Real QoI $Q(\omega)$ inaccessible (exact PDE), but we can assume

  $$\mathbb{E}[Q_{h,s}] \overset{h\to 0,\ s\to\infty}{\longrightarrow} \mathbb{E}[Q] \quad \text{and} \quad |\mathbb{E}[Q_{h,s} - Q]| = \mathscr{O}(h^{\alpha}) + \mathscr{O}(s^{-\alpha'})$$

- **Standard Monte Carlo** estimator for $\mathbb{E}[Q]$: More detail below!

  $$\hat{Q}^{\mathrm{MC}} := \frac{1}{N} \sum_{i=1}^{N} Q_{h,s}^{(i)}$$

  where $\{Q_{h,s}^{(i)}\}_{i=1}^{N}$ are i.i.d. samples computed with Model(h)

- Cost per sample is $\mathscr{O}(h^{-\gamma})$ (optimal: $\gamma = d$)

Computational Challenges

Standard Monte Carlo Quadrature

- Convergence of plain vanilla MC **(mean square error)**:

$$\underbrace{\mathbb{E}\big[\big(\hat{Q}^{\mathrm{MC}} - \mathbb{E}[Q]\big)^2\big]}_{=:\,\mathrm{MSE}} = \underbrace{\frac{\mathbb{V}[Q_{h,s}]}{N}}_{\textbf{sampling error}} + \underbrace{\Big(\mathbb{E}[Q_{h,s} - Q]\Big)^2}_{\textbf{model error (“bias”)}}$$

- Typical: $\alpha = 1 \;\Rightarrow\; \mathrm{MSE} = \mathscr{O}(N^{-1}) + \mathscr{O}(h^2) \leq \mathrm{TOL}^2$, and so $h \sim \mathrm{TOL}$ and $N \sim \mathrm{TOL}^{-2}$.

Computational Challenges

Standard Monte Carlo Quadrature

- Convergence of plain vanilla MC **(mean square error)**:

$$\underbrace{\mathbb{E}\big[\big(\hat{Q}^{\mathrm{MC}} - \mathbb{E}[Q]\big)^2\big]}_{=:\,\mathrm{MSE}} = \underbrace{\frac{\mathbb{V}[Q_{h,s}]}{N}}_{\textbf{sampling error}} + \underbrace{\Big(\mathbb{E}[Q_{h,s} - Q]\Big)^2}_{\textbf{model error (“bias”)}}$$

- Typical: $\alpha = 1 \;\Rightarrow\; \mathrm{MSE} = \mathscr{O}(N^{-1}) + \mathscr{O}(h^2) \leq \mathrm{TOL}^2$, and so $h \sim \mathrm{TOL}$ and $N \sim \mathrm{TOL}^{-2}$.

- Using optimal PDE solver: $\mathrm{Cost} = \mathscr{O}(Nh^{-d}) = \mathscr{O}(\mathrm{TOL}^{-(d+2)})$

  (e.g. for $\mathrm{TOL} = 10^{-3}$: $h \sim 10^{-3}$, $N \sim 10^6$ and $\mathrm{Cost} = \mathscr{O}(10^{12})$ in 2D!!)

Quickly becomes **prohibitively expensive !**

Computational Challenges

Numerical Experiment with standard Monte Carlo

$D = (0,1)^2$, unconditioned KL expansion, $Q = \| - k \frac{\partial p}{\partial x_1} \|_{L^1(D)}$
using mixed FEs and the AMG solver `amg1r5` [Ruge, Stüben, 1992]

Computational Challenges

Numerical Experiment with standard Monte Carlo

$D = (0,1)^2$, unconditioned KL expansion, $Q = \| - k \frac{\partial p}{\partial x_1} \|_{L^1(D)}$
using mixed FEs and the AMG solver `amg1r5` [Ruge, Stüben, 1992]

- Numerically observed FE-error: $\approx \mathscr{O}(h^{3/4}) \implies \alpha \approx 3/4$.
- Numerically observed cost/sample: $\approx \mathscr{O}(h^{-2}) \implies \gamma \approx 1$.

Computational Challenges

Numerical Experiment with standard Monte Carlo

$D = (0,1)^2$, unconditioned KL expansion, $Q = \| - k \frac{\partial p}{\partial x_1} \|_{L^1(D)}$
using mixed FEs and the AMG solver `amg1r5` [Ruge, Stüben, 1992]

- Numerically observed FE-error: $\approx \mathscr{O}(h^{3/4}) \implies \alpha \approx 3/4$.
- Numerically observed cost/sample: $\approx \mathscr{O}(h^{-2}) \implies \gamma \approx 1$.
- **Total cost** to get RMSE $\mathscr{O}(\mathrm{TOL})$: $\approx \mathscr{O}(\mathrm{TOL}^{-14/3})$
  to get error reduction by a factor 2 $\rightarrow$ cost grows by a factor 25!

Computational Challenges

Numerical Experiment with standard Monte Carlo

$D = (0,1)^2$, unconditioned KL expansion, $Q = \| - k \frac{\partial p}{\partial x_1} \|_{L^1(D)}$
using mixed FEs and the AMG solver `amg1r5` [Ruge, Stüben, 1992]

- Numerically observed FE-error: $\approx \mathscr{O}(h^{3/4}) \implies \alpha \approx 3/4$.
- Numerically observed cost/sample: $\approx \mathscr{O}(h^{-2}) \implies \gamma \approx 1$.
- **Total cost** to get RMSE $\mathscr{O}(\text{TOL})$: $\approx \mathscr{O}(\text{TOL}^{-14/3})$
  to get error reduction by a factor 2 $\rightarrow$ cost grows by a factor 25!

Case 1: $\sigma^2 = 1$, $\lambda = 0.3$, $\nu = 0.5$

TOL	h^{-1}	N	Cost
0.01	129	1.4×10^4	21 min
0.002	1025	3.5×10^5	30 days

Case 2: $\sigma^2 = 3$, $\lambda = 0.1$, $\nu = 0.5$

TOL	h^{-1}	N	Cost
0.01	513	8.5×10^3	4 h
0.002	**Prohibitively large!!**		

(actual numbers & CPU times on a cluster of 2GHz Intel T7300 processors)

Computational Challenges

Alternatives – The Curse of Dimensionality

Stochastic Galerkin/collocation methods

- cost grows very fast with dimension s and polynomial order q (faster than exponential) $\rightarrow$ #stoch. DOF: $N_{\mathrm{SC}} = \mathscr{O}((s+q)!/(s!q!))$
- lower with sparse grids (Smolyak) but growth still exponential in s!
- Anisotropic sparse grids or adaptive best N-term approximation can be dimension independent with **sufficient smoothness** ($\nu \gg 1$)!
- **But** best N-term theory by does **not** extend to nonlinear problems.
- Typically requires rewriting of PDE solvers and no access to good preconditioners ("intrusive").

Computational Challenges

Alternatives – The Curse of Dimensionality

Stochastic Galerkin/collocation methods A Nouy

- cost grows very fast with dimension s and polynomial order q (faster than exponential) $\rightarrow$ #stoch. DOF: $N_{\mathrm{SC}} = \mathcal{O}((s+q)!/(s!q!))$
- lower with sparse grids (Smolyak) but growth still exponential in s!
- Anisotropic sparse grids or adaptive best N-term approximation can be dimension independent with **sufficient smoothness** ($\nu \gg 1$)!
- **But** best N-term theory by does **not** extend to nonlinear problems.
- Typically requires rewriting of PDE solvers and no access to good preconditioners ("intrusive").

Computational Challenges

Alternatives – The Curse of Dimensionality

Stochastic Galerkin/collocation methods A Nouy

- cost grows very fast with dimension s and polynomial order q (faster than exponential) $\rightarrow$ #stoch. DOF: $N_{\mathrm{SC}} = \mathcal{O}((s+q)!/(s!q!))$
- lower with sparse grids (Smolyak) but growth still exponential in s!
- Anisotropic sparse grids or adaptive best N-term approximation can be dimension independent with **sufficient smoothness** ($\nu \gg 1$)!
- **But** best N-term theory by does **not** extend to nonlinear problems.
- Typically requires rewriting of PDE solvers and no access to good preconditioners ("intrusive").

Monte Carlo methods do not suffer from **Curse of Dimensionality**, they are **"non-intrusive"** and nonlinear parameter dependence is no problem,

but plain vanilla version is **too slow!**

Alternatives?

Monte Carlo Methods

The Buffon Needle Problem

- In 1777, George Louis Leclerc, Comte de Buffon (1707–1788), French naturalist and mathematician, posed the following problem:

  Let a needle of length ℓ be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance $d > \ell$ from each other. What is the probability p that the needle will intersect one of these lines?

The Buffon Needle Problem

- In 1777, George Louis Leclerc, Comte de Buffon (1707–1788), French naturalist and mathematician, posed the following problem:

  Let a needle of length ℓ be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance $d > \ell$ from each other. What is the probability p that the needle will intersect one of these lines?

- Answer: $p = \frac{2\ell}{\pi d}$ (simple geometric arguments)

The Buffon Needle Problem

- In 1777, George Louis Leclerc, Comte de Buffon (1707–1788), French naturalist and mathematician, posed the following problem:

  Let a needle of length ℓ be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance $d > \ell$ from each other. What is the probability p that the needle will intersect one of these lines?

- Answer: $p = \frac{2\ell}{\pi d}$ (simple geometric arguments)
- Laplace later used similar randomised experiment to approximate π.

The Buffon Needle Problem

- In 1777, George Louis Leclerc, Comte de Buffon (1707–1788), French naturalist and mathematician, posed the following problem:

  Let a needle of length ℓ be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance $d > \ell$ from each other. What is the probability p that the needle will intersect one of these lines?

- Answer: $p = \frac{2\ell}{\pi d}$ (simple geometric arguments)
- Laplace later used similar randomised experiment to approximate π.
- The term "Monte Carlo method" was coined by Metropolis, Ulam, von Neumann in the Manhattan project (Los Alamos, 1946).

The Buffon Needle Problem

THE ROYAL SOCIETY

Ants estimate area using Buffon's needle

Eamonn B. Mallon* and Nigel R. Franks

Centre for Mathematical Biology, and Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK

We show for the first time, to our knowledge, that ants can measure the size of potential nest sites. Nest size assessment is by individual scouts. Such scouts always make more than one visit to a potential nest before initiating an emigration of their nest mates and they deploy individual-specific trails within the potential new nest on their first visit. We test three alternative hypotheses for the way in which scouts might measure nests. Experiments indicated that individual scouts use the intersection frequency between their own paths to assess nest areas. These results are consistent with ants using a 'Buffon's needle algorithm' to assess nest areas.

Keywords: ants; colony emigration; individual-specific pheromones; *Leptothorax*; nest sites; rules of thumb

Proceedings of the Royal Society of London, 2000

Basic Monte Carlo Simulation

- Given a sequence $\{X_k\}$ of i.i.d. copies of a given random variable X, basic MC simulation uses the estimator

$$\mathbf{E}[X] \approx \frac{S_N}{N}, \qquad S_N = X_1 + \cdots + X_N.$$

- By the *Strong Law of Large Numbers*: $\frac{S_N}{N} \to \mathbf{E}[X]$ a.s.

Basic Monte Carlo Simulation

- Given a sequence $\{X_k\}$ of i.i.d. copies of a given random variable X, basic MC simulation uses the estimator

$$\mathbf{E}[X] \approx \frac{S_N}{N}, \qquad S_N = X_1 + \cdots + X_N.$$

- By the *Strong Law of Large Numbers*: $\frac{S_N}{N} \to \mathbf{E}[X]$ a.s.

- Also, for any measurable function f: $\frac{1}{N}\sum_{k=1}^{N} f(X_k) \to \mathbf{E}[f(X)]$ a.s.

Basic Monte Carlo Simulation

- Given a sequence $\{X_k\}$ of i.i.d. copies of a given random variable X, basic MC simulation uses the estimator

$$\mathbf{E}[X] \approx \frac{S_N}{N}, \qquad S_N = X_1 + \cdots + X_N.$$

- By the *Strong Law of Large Numbers*: $\frac{S_N}{N} \to \mathbf{E}[X]$ a.s.

- Also, for any measurable function f: $\frac{1}{N}\sum_{k=1}^{N} f(X_k) \to \mathbf{E}[f(X)]$ a.s.

- If $\mathbf{E}[X] = \mu$ and $\mathbf{Var}[X] = \sigma^2$, then (via the *Central Limit Theorem*)

$$\mathbf{E}[S_N] = N\mu, \quad \mathbf{Var}[S_N] = N\sigma^2 \quad \text{and} \quad S_N^* = \frac{S_N - N\mu}{\sqrt{N}\sigma} \to \mathrm{N}(0,1)$$

  i.e. the estimate is unbiased, its variance is σ^2/N and the distribution of the normalised RV S_N^* becomes Gaussian as $N \to \infty$.

Monte Carlo Convergence Statements

1. Since

$$\mathbf{E}\left[\left(\frac{S_N}{N}-\mu\right)^2\right] = \mathbf{Var}\,\frac{S_N}{N} = \frac{\sigma^2}{N} \to 0,$$

we have mean square convergence of S_N/N to μ.

Monte Carlo Convergence Statements

1. Since
$$\mathbf{E}\left[\left(\frac{S_N}{N}-\mu\right)^2\right] = \mathbf{Var}\,\frac{S_N}{N} = \frac{\sigma^2}{N} \to 0,$$
we have mean square convergence of S_N/N to μ.

2. Also *Chebyshev's Inequality* implies, for any $\epsilon > 0$,
$$\mathbf{P}\left\{\left|\frac{S_N}{N}-\mu\right| > N^{-1/2+\epsilon}\right\} \leq \frac{\sigma^2}{N^{2\epsilon}},$$
i.e. for any $\epsilon > 0$, the probability of the error being larger than $N^{-1/2+\epsilon}$ converges to zero as $N \to \infty$.

Monte Carlo Convergence Statements

1. Since
$$\mathbf{E}\left[\left(\frac{S_N}{N}-\mu\right)^2\right] = \mathbf{Var}\,\frac{S_N}{N} = \frac{\sigma^2}{N} \to 0,$$
we have mean square convergence of S_N/N to μ.

2. Also *Chebyshev's Inequality* implies, for any $\epsilon > 0$,
$$\mathbf{P}\left\{\left|\frac{S_N}{N}-\mu\right| > N^{-1/2+\epsilon}\right\} \leq \frac{\sigma^2}{N^{2\epsilon}},$$
i.e. for any $\epsilon > 0$, the probability of the error being larger than $N^{-1/2+\epsilon}$ converges to zero as $N \to \infty$.

3. If $\rho := \mathbf{E}\left[|X-\mu|^3\right] < \infty$, then the *Berry-Esseen Inequality* gives
$$\left|\mathbf{P}\{S_N^* \leq x\} - \Phi(x)\right| \leq \frac{\rho}{2\sigma^3\sqrt{N}},$$
where Φ denotes *cumulative density function (CDF)* of $\mathrm{N}(0,1)$.

Quasi-Monte Carlo Methods

In quasi-Monte Carlo (QMC) methods, samples are not chosen randomly, but special (deterministic) number sequences, known as low-discrepancy sequences, are used instead. (discrepancy = measure of equidistribution)

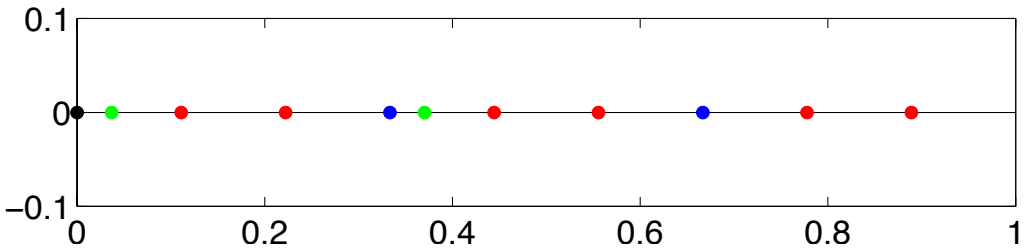
Quasi-Monte Carlo Methods

In quasi-Monte Carlo (QMC) methods, samples are not chosen randomly, but special (deterministic) number sequences, known as low-discrepancy sequences, are used instead. (discrepancy = measure of equidistribution)

Example: The van der Corput sequence is a low-discrepancy sequence for the unit interval. The first 11 numbers are

$$\{x_n\}_{n=1}^{11} = \left\{0, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}\right\}.$$

(for base 3, it is given by $x_n = \frac{k}{3^j}$, where j increases monotonically and k runs through all nonnegative integers such that $k/3^j$ is an irreducible fraction < 1. The ordering in k is obtained by representing k in base 3 and reversing the digits)



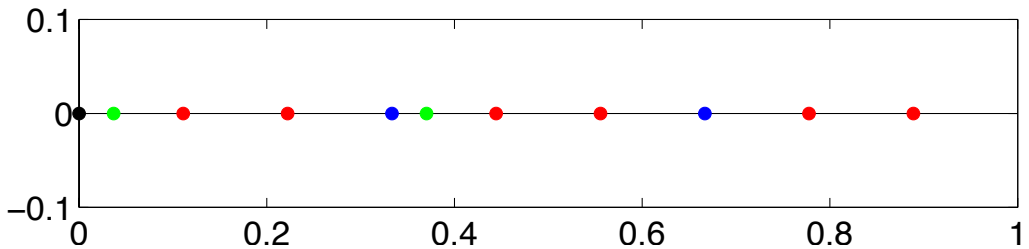
Quasi-Monte Carlo Methods

In quasi-Monte Carlo (QMC) methods, samples are not chosen randomly, but special (deterministic) number sequences, known as low-discrepancy sequences, are used instead. (discrepancy = measure of equidistribution)

Example: The van der Corput sequence is a low-discrepancy sequence for the unit interval. The first 11 numbers are

$$\{x_n\}_{n=1}^{11} = \left\{0, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}\right\}.$$

(for base 3, it is given by $x_n = \frac{k}{3^j}$, where j increases monotonically and k runs through all nonnegative integers such that $k/3^j$ is an irreducible fraction < 1. The ordering in k is obtained by representing k in base 3 and reversing the digits)



Convergence rate improved from $\mathcal{O}(N^{-1/2})$ to $\mathcal{O}(N^{-2})$!

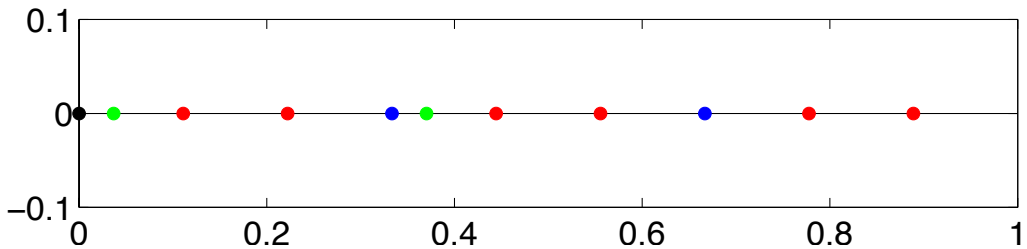
Quasi-Monte Carlo Methods

In quasi-Monte Carlo (QMC) methods, samples are not chosen randomly, but special (deterministic) number sequences, known as low-discrepancy sequences, are used instead. (discrepancy = measure of equidistribution)

Example: The van der Corput sequence is a low-discrepancy sequence for the unit interval. The first 11 numbers are

$$\{x_n\}_{n=1}^{11} = \left\{0, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}\right\}.$$

(for base 3, it is given by $x_n = \frac{k}{3^j}$, where j increases monotonically and k runs through all nonnegative integers such that $k/3^j$ is an irreducible fraction < 1. The ordering in k is obtained by representing k in base 3 and reversing the digits)



Convergence rate improved from $\mathcal{O}(N^{-1/2})$ to $\mathcal{O}(N^{-2})$!

D Nuyens

Variance Reduction: Antithetic Sampling

The constant in the MC convergence rate is the variance σ^2 of the RV X. By designing an equivalent MC approximation with lower variance, we can expect faster convergence.

Variance Reduction: Antithetic Sampling

The constant in the MC convergence rate is the variance σ^2 of the RV X. By designing an equivalent MC approximation with lower variance, we can expect faster convergence.

- To approximate $\mathbf{E}[X]$ by standard MC, we draw independent samples $\{X_k\}_{k=1}^N$ of X and compute the sample average S_N/N.

Variance Reduction: Antithetic Sampling

The constant in the MC convergence rate is the variance σ^2 of the RV X. By designing an equivalent MC approximation with lower variance, we can expect faster convergence.

- To approximate $\mathbf{E}[X]$ by standard MC, we draw independent samples $\{X_k\}_{k=1}^N$ of X and compute the sample average S_N/N.
- Now, let $\{\tilde{X}_k\}_{k=1}^N$ be a second set of samples of X with sample average $\tilde{S}_N/N$.

Variance Reduction: Antithetic Sampling

The constant in the MC convergence rate is the variance σ^2 of the RV X. By designing an equivalent MC approximation with lower variance, we can expect faster convergence.

- To approximate $\mathbf{E}[X]$ by standard MC, we draw independent samples $\{X_k\}_{k=1}^N$ of X and compute the sample average S_N/N.
- Now, let $\{\tilde{X}_k\}_{k=1}^N$ be a second set of samples of X with sample average $\tilde{S}_N/N$.
- Since both averages converge to $\mathbf{E}[X]$, so does $\frac{1}{2}(S_N/N + \tilde{S}_N/N)$.
- When X_k and $\tilde{X}_k$ are negatively correlated they are called antithetic samples, and the approximation $\frac{1}{2N}(S_N + \tilde{S}_N)$ is a more reliable approximation of $\mathbf{E}[X]$ than $\frac{1}{2N}S_{2N}$.

Variance Reduction: Antithetic Sampling

Theorem

Let the two sequences of RVs $\{X_k\}$ *and* $\{\tilde{X}_k\}$ *be identically distributed with*

$$\mathbf{Cov}(X_j, X_k) = \mathbf{Cov}(\tilde{X}_j, \tilde{X}_k) = 0 \qquad \text{for } j \neq k.$$

Then

$$\mathbf{Var}\left[\frac{S_N + \tilde{S}_N}{2N}\right] = \mathbf{Var}\left[\frac{S_{2N}}{2N}\right] + \frac{1}{2}\mathbf{Cov}\left(\frac{S_N}{N}, \frac{\tilde{S}_N}{N}\right) \leq \mathbf{Var}\left[\frac{S_N}{N}\right].$$

Variance Reduction: Antithetic Sampling

Theorem

Let the two sequences of RVs $\{X_k\}$ *and* $\{\tilde{X}_k\}$ *be identically distributed with*

$$\mathbf{Cov}(X_j, X_k) = \mathbf{Cov}(\tilde{X}_j, \tilde{X}_k) = 0 \qquad \text{for } j \neq k.$$

Then

$$\mathbf{Var}\left[\frac{S_N + \tilde{S}_N}{2N}\right] = \mathbf{Var}\left[\frac{S_{2N}}{2N}\right] + \frac{1}{2}\mathbf{Cov}\left(\frac{S_N}{N}, \frac{\tilde{S}_N}{N}\right) \leq \mathbf{Var}\left[\frac{S_N}{N}\right].$$

- **Worst case:** Variance of average of N samples and N antithetic samples no better than variance of N independent samples.

Variance Reduction: Antithetic Sampling

Theorem

Let the two sequences of RVs $\{X_k\}$ *and* $\{\tilde{X}_k\}$ *be identically distributed with*

$$\mathbf{Cov}(X_j, X_k) = \mathbf{Cov}(\tilde{X}_j, \tilde{X}_k) = 0 \qquad \text{for } j \neq k.$$

Then

$$\mathbf{Var}\left[\frac{S_N + \tilde{S}_N}{2N}\right] = \mathbf{Var}\left[\frac{S_{2N}}{2N}\right] + \frac{1}{2}\mathbf{Cov}\left(\frac{S_N}{N}, \frac{\tilde{S}_N}{N}\right) \leq \mathbf{Var}\left[\frac{S_N}{N}\right].$$

- **Worst case:** Variance of average of N samples and N antithetic samples no better than variance of N independent samples.
- **Best case:** S_N/N and $\tilde{S}_N/N$ negatively correlated; therefore variance of N samples and N antithetic samples is smaller than variance of $2N$ indepependent samples.

Easy Example: Predator-Prey Dynamical System

Consider the popular model of the dynamics of two interacting populations

$$\dot{\mathbf{u}} = \begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{bmatrix} = \begin{bmatrix} u_1(1-u_2) \\ u_2(u_1-1) \end{bmatrix}, \quad \mathbf{u}(0) = \mathbf{u}_0.$$

Let the initial condition $\mathbf{u}_0$ be uncertain and modeled as a (uniform) random vector $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, where Γ denotes the square

$$\Gamma = \overline{\mathbf{u}}_0 + [-\epsilon, \epsilon]^2.$$

Easy Example: Predator-Prey Dynamical System

Consider the popular model of the dynamics of two interacting populations

$$\dot{\mathbf{u}} = \begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{bmatrix} = \begin{bmatrix} u_1(1-u_2) \\ u_2(u_1-1) \end{bmatrix}, \quad \mathbf{u}(0) = \mathbf{u}_0.$$

Let the initial condition $\mathbf{u}_0$ be uncertain and modeled as a (uniform) random vector $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, where Γ denotes the square

$$\Gamma = \overline{\mathbf{u}}_0 + [-\epsilon, \epsilon]^2.$$

- **Goal:** estimate $\mathbf{E}[u_1(T)]$ at time $T > 0$.
- Denote by $\mathbf{u}_h = \mathbf{u}_h(\omega)$ the **explicit Euler** approximation after M time steps of length $h = \frac{T}{M}$ starting with initial data $\mathbf{u}_0 = \mathbf{u}_0(\omega)$.
- Define the QoI $Q = \phi(\mathbf{u}(T)) = u_1(T)$ for $\mathbf{u} = [u_1, u_2]^\mathsf{T}$ and estimate $\mathbf{E}[Q_h]$ using the MC method just described, where $Q_h = \phi(\mathbf{u}_h)$.

Easy Example: Predator-Prey Dynamical System

Consider the popular model of the dynamics of two interacting populations

$$\dot{\mathbf{u}} = \begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{bmatrix} = \begin{bmatrix} u_1(1-u_2) \\ u_2(u_1-1) \end{bmatrix}, \quad \mathbf{u}(0) = \mathbf{u}_0.$$

Let the initial condition $\mathbf{u}_0$ be uncertain and modeled as a (uniform) random vector $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, where Γ denotes the square

$$\Gamma = \overline{\mathbf{u}}_0 + [-\epsilon, \epsilon]^2.$$

- **Goal:** estimate $\mathbf{E}[u_1(T)]$ at time $T > 0$.
- Denote by $\mathbf{u}_h = \mathbf{u}_h(\omega)$ the **explicit Euler** approximation after M time steps of length $h = \frac{T}{M}$ starting with initial data $\mathbf{u}_0 = \mathbf{u}_0(\omega)$.
- Define the QoI $Q = \phi(\mathbf{u}(T)) = u_1(T)$ for $\mathbf{u} = [u_1, u_2]^\mathsf{T}$ and estimate $\mathbf{E}[Q_h]$ using the MC method just described, where $Q_h = \phi(\mathbf{u}_h)$.
- Can easily make this problem **high dimensional** by adding uncertain coefficients (= environment) **!**

Easy Example: Predator-Prey Dynamical System

Analysis of Monte Carlo Estimator

- Denote the Monte Carlo estimator for $\mathbf{E}[Q_h]$ by

$$\widehat{Q}_h := \widehat{Q}_{h,N} = \frac{1}{N} \sum_{k=1}^{N} Q_h^{(k)}$$

(i.e the average over N samples $\{Q_h^{(k)}\}_{k=1}^N$ of Q_h with step size h)

- Expect better approximations for N large and h small.

Easy Example: Predator-Prey Dynamical System

Analysis of Monte Carlo Estimator

- Denote the Monte Carlo estimator for $\mathbf{E}[Q_h]$ by

$$\widehat{Q}_h := \widehat{Q}_{h,N} = \frac{1}{N} \sum_{k=1}^{N} Q_h^{(k)}$$

(i.e the average over N samples $\{Q_h^{(k)}\}_{k=1}^N$ of Q_h with step size h)

- Expect better approximations for N large and h small.
- Error with N samples and $M = T/h$ time steps:

$$e_{h,N} = |\mathbf{E}[Q] - \widehat{Q}_h| \leq \underbrace{|\mathbf{E}[Q] - \mathbf{E}[Q_h]|}_{\text{discretisation error}} + \underbrace{|\mathbf{E}[Q_h] - \widehat{Q}_h|}_{\text{Monte Carlo error}}$$

Easy Example: Predator-Prey Dynamical System

Analysis of Monte Carlo Estimator

- Denote the Monte Carlo estimator for $\mathbf{E}[Q_h]$ by

$$\widehat{Q}_h := \widehat{Q}_{h,N} = \frac{1}{N}\sum_{k=1}^{N} Q_h^{(k)}$$

(i.e the average over N samples $\{Q_h^{(k)}\}_{k=1}^N$ of Q_h with step size h)

- Expect better approximations for N large and h small.
- Error with N samples and $M = T/h$ time steps:

$$e_{h,N} = |\mathbf{E}[Q] - \widehat{Q}_h| \leq \underbrace{|\mathbf{E}[Q] - \mathbf{E}[Q_h]|}_{\text{discretisation error}} + \underbrace{|\mathbf{E}[Q_h] - \widehat{Q}_h|}_{\text{Monte Carlo error}}$$

- Can show also that the mean square error satisfies (with equality!)

$$\mathbf{E}\left[\left(\mathbf{E}[Q] - \widehat{Q}_h\right)^2\right] = \left(\mathbf{E}[Q - Q_h]\right)^2 + \frac{\mathbf{Var}[Q_h]}{N}$$

Problem Sheet!

Easy Example: Predator-Prey Dynamical System

Discretisation Error (Bias)

- Explicit Euler discretisation error (with some constant $K > 0$):

$$\|\mathbf{u}(T) - \mathbf{u}_h\| \leq Kh.$$

Easy Example: Predator-Prey Dynamical System

Discretisation Error (Bias)

- Explicit Euler discretisation error (with some constant $K > 0$):

$$\|\mathbf{u}(T) - \mathbf{u}_h\| \leq Kh.$$

- ϕ Lipschitz-continuous with constant $L = 1$:

$$|\phi(\mathbf{u}(T)) - \phi(\mathbf{u}_h)| \leq KLh.$$

Easy Example: Predator-Prey Dynamical System

Discretisation Error (Bias)

- Explicit Euler discretisation error (with some constant $K > 0$):

$$\|\mathbf{u}(T) - \mathbf{u}_h\| \leq Kh.$$

- ϕ Lipschitz-continuous with constant $L = 1$:

$$|\phi(\mathbf{u}(T)) - \phi(\mathbf{u}_h)| \leq KLh.$$

- Therefore

$$|\mathbf{E}[Q] - \mathbf{E}[Q_h]| = |\mathbf{E}[Q - Q_h]| \leq KLh.$$

Easy Example: Predator-Prey Dynamical System

Balancing discretisation and MC error

- Using the Berry-Esseen bound with $\mathbf{Var}[Q_h] = \sigma_h^2$ we get (see the **Problem Sheet**)

$$\mathbf{P}\left(\left|\mathbf{E}\left[Q_h\right] - \widehat{Q}_{h,N}\right| \leq \frac{1.96\sigma_h}{\sqrt{N}}\right) > 0.95 + \mathscr{O}(N^{-1/2})$$

Easy Example: Predator-Prey Dynamical System

Balancing discretisation and MC error

- Using the Berry-Esseen bound with $\mathbf{Var}[Q_h] = \sigma_h^2$ we get (see the **Problem Sheet**)

$$\mathbf{P}\left(\left|\mathbf{E}\left[Q_h\right] - \widehat{Q}_{h,N}\right| \leq \frac{1.96\sigma_h}{\sqrt{N}}\right) > 0.95 + \mathscr{O}(N^{-1/2})$$

- Combined with the discretisation error (using triangle inequality):

$$\mathbf{P}\left(e_{h,N} \leq 2KLh + \frac{1.96\sigma_h}{\sqrt{N}}\right) > 0.95 + \mathscr{O}(N^{-1/2}).$$

Easy Example: Predator-Prey Dynamical System

Balancing discretisation and MC error

- Using the Berry-Esseen bound with $\mathbf{Var}[Q_h] = \sigma_h^2$ we get (see the **Problem Sheet**)

$$\mathbf{P}\left(\left|\mathbf{E}\left[Q_h\right] - \widehat{Q}_{h,N}\right| \leq \frac{1.96\sigma_h}{\sqrt{N}}\right) > 0.95 + \mathscr{O}(N^{-1/2})$$

- Combined with the discretisation error (using triangle inequality):

$$\mathbf{P}\left(e_{h,N} \leq 2KLh + \frac{1.96\sigma_h}{\sqrt{N}}\right) > 0.95 + \mathscr{O}(N^{-1/2}).$$

- Balancing discretization and MC errors, i.e.

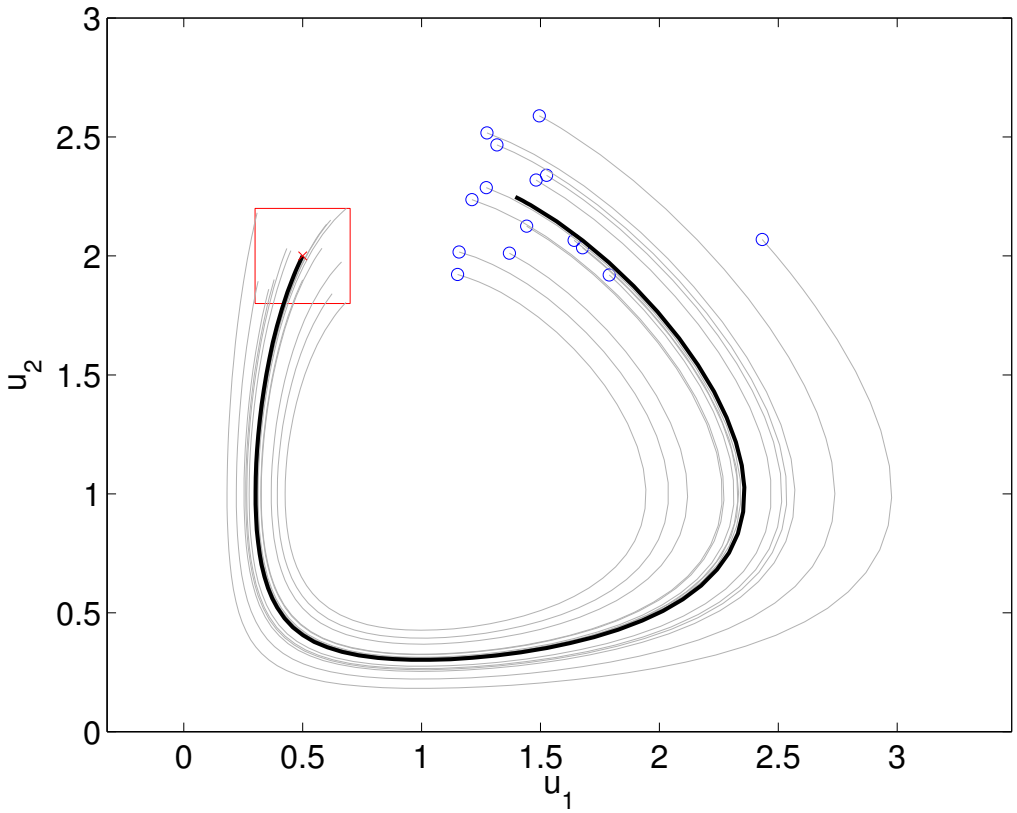
$$KLh \approx \frac{\mathrm{TOL}}{2} \quad \text{and} \quad \frac{1.96\sigma_h}{\sqrt{N}} \approx \frac{\mathrm{TOL}}{2},$$

leads to

$$h \approx \frac{\mathrm{TOL}}{2KL}, \quad N \approx \frac{16\sigma_h^2}{\mathrm{TOL}^2} \quad \text{and} \quad \mathrm{Cost} = \mathscr{O}(MN) = \mathscr{O}(\mathrm{TOL}^{-3})$$

Easy Example: Predator-Prey Dynamical System

Sample trajectories



Predator-prey system integrated from 0 to $T = 6$ with $\bar{\mathbf{u}}_0 = [0.5, 2]^T$ and $\epsilon = 0.2$. Unperturbed trajectory (black) and 15 perturbed trajectories.

(for the unperturbed trajectory $u_1(T) = 1.3942$)

Easy Example: Predator-Prey Dynamical System

Antithetic sampling

We can use **antithetic sampling** for this problem by noting that, if $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, then the same holds for the random vector

$$\tilde{\mathbf{u}}_0 := 2\bar{\mathbf{u}}_0 - \mathbf{u}_0.$$

Thus, the trajectories generated by the random initial data $\tilde{\mathbf{u}}_0$ have the same distribution as those generated by $\mathbf{u}_0$.

Easy Example: Predator-Prey Dynamical System

Antithetic sampling

We can use **antithetic sampling** for this problem by noting that, if $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, then the same holds for the random vector

$$\tilde{\mathbf{u}}_0 := 2\bar{\mathbf{u}}_0 - \mathbf{u}_0.$$

Thus, the trajectories generated by the random initial data $\tilde{\mathbf{u}}_0$ have the same distribution as those generated by $\mathbf{u}_0$.

- Let $Q_h = \phi(\mathbf{u}_h)$ be the basic samples and $\tilde{Q}_h = \phi(\tilde{\mathbf{u}}_h)$ the antithetic counterparts. Note that all pairs of samples are independent except each sample and its antithetic counterpart.

Easy Example: Predator-Prey Dynamical System

Antithetic sampling

We can use **antithetic sampling** for this problem by noting that, if $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, then the same holds for the random vector

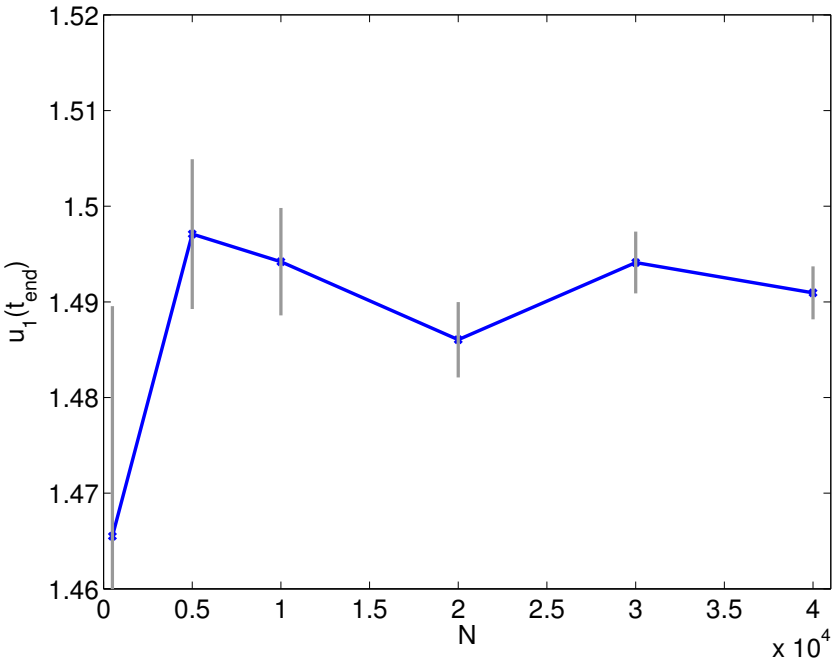
$$\tilde{\mathbf{u}}_0 := 2\bar{\mathbf{u}}_0 - \mathbf{u}_0.$$

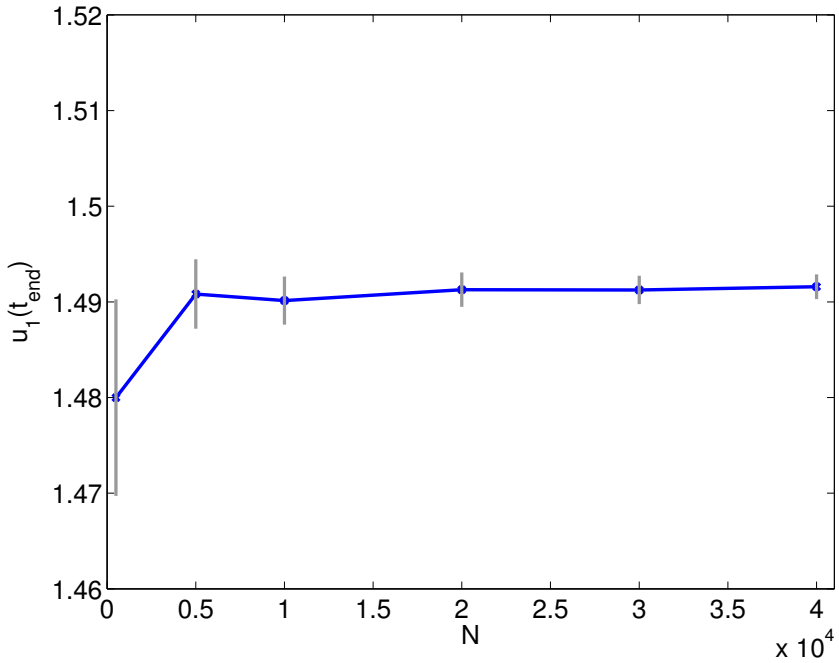
Thus, the trajectories generated by the random initial data $\tilde{\mathbf{u}}_0$ have the same distribution as those generated by $\mathbf{u}_0$.

- Let $Q_h = \phi(\mathbf{u}_h)$ be the basic samples and $\tilde{Q}_h = \phi(\tilde{\mathbf{u}}_h)$ the antithetic counterparts. Note that all pairs of samples are independent except each sample and its antithetic counterpart.
- Use $\frac{1}{2}(\widehat{Q}_{h,N} + \widehat{\tilde{Q}}_{h,N})$ instead of $\widehat{Q}_{h,2N}$ (same cost).

Easy Example: Predator-Prey Dynamical System

Numerical Experiment – Comparing standard and antithetic sampling





Estimates of $\mathbf{E}[u_1(T)]$ using standard MC with N samples (left) vs. antithetic sampling with $N/2$ pairs of samples (right) (with 95% confidence intervals)

$\mathbf{Var}[Q_h]$ and $\mathbf{Cov}(Q_h, \tilde{Q}_h)$ estimated using sample variance and covariance (resp.), i.e.

$$\frac{1}{N-1}\sum_{k=1}^{N}(Q_h^{(k)} - \widehat{Q}_{h,N})^2 \quad \text{and} \quad \frac{1}{N-1}\sum_{k=1}^{N}(Q_h^{(k)} - \widehat{Q}_{h,N})(\tilde{Q}_h^{(k)} - \widehat{\tilde{Q}}_{h,N})$$

Multilevel Monte Carlo Methods – History

- The **multilevel Monte Carlo** method is a powerful new variance reduction technique.
- First ideas for high-dimensional quadrature by Heinrich, 2000.
- Independently discovered and popularised by Giles, 2007 in the context of stochastic DEs in mathematical finance.

Multilevel Monte Carlo Methods – History

- The **multilevel Monte Carlo** method is a powerful new variance reduction technique.
- First ideas for high-dimensional quadrature by Heinrich, 2000.
- Independently discovered and popularised by Giles, 2007 in the context of stochastic DEs in mathematical finance.
- First papers in the context of UQ:
  - [Barth, Schwab, Zollinger '11]
  - [Cliffe, Giles, RS, Teckentrup '11]

Multilevel Monte Carlo Methods – History

- The **multilevel Monte Carlo** method is a powerful new variance reduction technique.
- First ideas for high-dimensional quadrature by Heinrich, 2000.
- Independently discovered and popularised by Giles, 2007 in the context of stochastic DEs in mathematical finance.
- First papers in the context of UQ:
  - [Barth, Schwab, Zollinger '11]
  - [Cliffe, Giles, RS, Teckentrup '11]
- Stochastic simulation of discrete state systems (biology, chemistry):
  - [Anderson, Higham '12], ...

Multilevel Monte Carlo Methods – History

- The **multilevel Monte Carlo** method is a powerful new variance reduction technique.
- First ideas for high-dimensional quadrature by Heinrich, 2000.
- Independently discovered and popularised by Giles, 2007 in the context of stochastic DEs in mathematical finance.
- First papers in the context of UQ:
  - [Barth, Schwab, Zollinger '11]
  - [Cliffe, Giles, RS, Teckentrup '11]
- Stochastic simulation of discrete state systems (biology, chemistry):
  - [Anderson, Higham '12], ...
- In MCMC for Bayesian inference:
  - [Hoang, Schwab, Stuart '13]
  - [Dodwell, Ketelsen, RS, Teckentrup '15]

Standard Monte Carlo – Mean-Square Error

- To estimate the expectation $\mathbf{E}[Q]$ of a quantity of interest Q, assume only approximations $Q_h \approx Q$ are computable, where $h > 0$ denotes a discretization parameter (time step size, grid size, ...) and

$$\lim_{h \to 0} \mathbf{E}[Q_h] = \mathbf{E}[Q] .$$

Standard Monte Carlo – Mean-Square Error

- To estimate the expectation $\mathbf{E}[Q]$ of a quantity of interest Q, assume only approximations $Q_h \approx Q$ are computable, where $h > 0$ denotes a discretization parameter (time step size, grid size, ...) and

$$\lim_{h \to 0} \mathbf{E}[Q_h] = \mathbf{E}[Q].$$

- More precisely, we assume error to converge in mean with rate α:

$$|\mathbf{E}[Q_h - Q]| \lesssim h^\alpha, \qquad \text{as } h \to 0, \qquad \alpha > 0.$$

(in the predator-prey case $\alpha = 1$)

Standard Monte Carlo – Mean-Square Error

- To estimate the expectation $\mathbf{E}[Q]$ of a quantity of interest Q, assume only approximations $Q_h \approx Q$ are computable, where $h > 0$ denotes a discretization parameter (time step size, grid size, ...) and

$$\lim_{h \to 0} \mathbf{E}[Q_h] = \mathbf{E}[Q].$$

- More precisely, we assume error to converge in mean with rate α:

$$|\mathbf{E}[Q_h - Q]| \lesssim h^\alpha, \qquad \text{as } h \to 0, \qquad \alpha > 0.$$

(in the predator-prey case $\alpha = 1$)

- From above (and Problem Sheet) the mean square error (MSE) is

$$\mathbf{E}\left[\left(\widehat{Q}_{h,N} - \mathbf{E}[Q]\right)^2\right] = \frac{\mathbf{Var}[Q_h]}{N} + \left(\mathbf{E}[Q_h - Q]\right)^2.$$

Standard Monte Carlo – Cost Scaling

- Denote by $\mathscr{C}(Q_h^{(k)})$ cost associated with computing one sample $Q_h^{(k)}$
  (e.g. in terms of the number of floating-point operations required)

Standard Monte Carlo – Cost Scaling

- Denote by $\mathscr{C}(Q_h^{(k)})$ cost associated with computing one sample $Q_h^{(k)}$ (e.g. in terms of the number of floating-point operations required)
- Cost typcially grows linearly or with power $\gamma \geq 1$ with $M = T/h$. We assume

$$\mathscr{C}(Q_h^{(k)}) \lesssim h^{-\gamma}, \qquad \gamma \geq 1.$$

so that $\mathscr{C}(\widehat{Q}_{h,N}) \lesssim Nh^{-\gamma}$ (in the predator-prey case $\gamma = 1$).

Standard Monte Carlo – Cost Scaling

- Denote by $\mathscr{C}(Q_h^{(k)})$ cost associated with computing one sample $Q_h^{(k)}$ (e.g. in terms of the number of floating-point operations required)
- Cost typcially grows linearly or with power $\gamma \geq 1$ with $M = T/h$. We assume

$$\mathscr{C}(Q_h^{(k)}) \lesssim h^{-\gamma}, \qquad \gamma \geq 1.$$

  so that $\mathscr{C}(\widehat{Q}_{h,N}) \lesssim Nh^{-\gamma}$ (in the predator-prey case $\gamma = 1$).
- To balance the two MSE components, bound each of them by $\frac{\mathrm{TOL}^2}{2}$, resulting in a total bound of TOL^2 for the MSE.

Standard Monte Carlo – Cost Scaling

- Denote by $\mathscr{C}(Q_h^{(k)})$ cost associated with computing one sample $Q_h^{(k)}$ (e.g. in terms of the number of floating-point operations required)
- Cost typcially grows linearly or with power $\gamma \geq 1$ with $M = T/h$. We assume

$$\mathscr{C}(Q_h^{(k)}) \lesssim h^{-\gamma}, \qquad \gamma \geq 1.$$

  so that $\mathscr{C}(\widehat{Q}_{h,N}) \lesssim Nh^{-\gamma}$ (in the predator-prey case $\gamma = 1$).
- To balance the two MSE components, bound each of them by $\frac{\mathrm{TOL}^2}{2}$, resulting in a total bound of TOL^2 for the MSE.
- This yields (since $Q_h \to Q$, we have $\mathbf{Var}[Q_h] \approx \mathbf{Var}[Q] = \text{constant}$)

$$N \geq 2\,\mathbf{Var}[Q_h]\mathrm{TOL}^{-2} \qquad \text{and} \qquad h \lesssim \mathrm{TOL}^{1/\alpha}.$$

Standard Monte Carlo – Cost Scaling

- Denote by $\mathscr{C}(Q_h^{(k)})$ cost associated with computing one sample $Q_h^{(k)}$ (e.g. in terms of the number of floating-point operations required)
- Cost typcially grows linearly or with power $\gamma \geq 1$ with $M = T/h$. We assume

$$\mathscr{C}(Q_h^{(k)}) \lesssim h^{-\gamma}, \qquad \gamma \geq 1.$$

  so that $\mathscr{C}(\widehat{Q}_{h,N}) \lesssim Nh^{-\gamma}$ (in the predator-prey case $\gamma = 1$).
- To balance the two MSE components, bound each of them by $\frac{\mathrm{TOL}^2}{2}$, resulting in a total bound of TOL^2 for the MSE.
- This yields (since $Q_h \to Q$, we have $\mathbf{Var}[Q_h] \approx \mathbf{Var}[Q] = \text{constant}$)

$$N \geq 2\,\mathbf{Var}[Q_h]\mathrm{TOL}^{-2} \qquad \text{and} \qquad h \lesssim \mathrm{TOL}^{1/\alpha}.$$

- So total cost to have $\mathrm{MSE} < \mathrm{TOL}^2$ using standard MC estimator is

$$\mathscr{C}(\widehat{Q}_{h,N}) \lesssim \mathrm{TOL}^{-2-\gamma/\alpha}$$

Multilevel Monte Carlo Methods

Main Idea

- **Key idea:** use realisations of Q_h on a hierarchy of different levels, i.e., for a sequence $h_0, \ldots, h_L$ of discretization parameters, and decompose

$$\mathbf{E}\left[Q_{h_L}\right] = \mathbf{E}\left[Q_{h_0}\right] + \sum_{\ell=1}^{L} \mathbf{E}\left[Q_{h_\ell} - Q_{h_{\ell-1}}\right] =: \sum_{\ell=0}^{L} \mathbf{E}\left[Y_\ell\right],$$

where $h_0 > 0$, $h_\ell = h_{\ell-1}/s$, for $\ell = 1, \ldots, L$, and $s \in \mathbb{N} \setminus \{1\}$.

Multilevel Monte Carlo Methods

Main Idea

- **Key idea:** use realisations of Q_h on a hierarchy of different levels, i.e., for a sequence $h_0, \ldots, h_L$ of discretization parameters, and decompose

$$\mathbf{E}\left[Q_{h_L}\right] = \mathbf{E}\left[Q_{h_0}\right] + \sum_{\ell=1}^{L} \mathbf{E}\left[Q_{h_\ell} - Q_{h_{\ell-1}}\right] =: \sum_{\ell=0}^{L} \mathbf{E}\left[Y_\ell\right],$$

  where $h_0 > 0$, $h_\ell = h_{\ell-1}/s$, for $\ell = 1, \ldots, L$, and $s \in \mathbb{N} \setminus \{1\}$.

- Given (unbiased) estimators $\{\widehat{Y}_\ell\}_{\ell=0}^{L}$ for $\mathbf{E}\left[Y_\ell\right]$, we refer to

$$\widehat{Q}_L^{\mathrm{ML}} := \sum_{\ell=0}^{L} \widehat{Y}_\ell$$

  as a multilevel estimator for Q (today use standard MC on all levels).

- All expectations $\mathbf{E}\left[Y_\ell\right]$ sampled indep. $\Rightarrow$ $\mathbf{Var}\,\widehat{Q}_L^{\mathrm{ML}} = \sum_{\ell=0}^{L} \mathbf{Var}\,\widehat{Y}_\ell$.

Multilevel Monte Carlo Methods

Multilevel Monte Carlo estimator

- If each $\widehat{Y}_\ell$ is itself a standard Monte Carlo estimator, i.e.,

$$\widehat{Y}_0 = \widehat{Y}_{0,N_0} := \frac{1}{N_0} \sum_{k=0}^{N_0} Q_{h_0}^{(k)}$$

and

$$\widehat{Y}_\ell = \widehat{Y}_{\ell,N_\ell} := \frac{1}{N_\ell} \sum_{k=0}^{N_\ell} \left(Q_{h_\ell}^{(k)} - Q_{h_{\ell-1}}^{(k)} \right), \qquad \ell = 1, \ldots, L,$$

one obtains a multilevel Monte Carlo estimator.

- The associated MSE then has the standard decomposition

$$\mathbf{E}\left[\left(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}} - \mathbf{E}\left[Q\right] \right)^2 \right] = \sum_{\ell=0}^{L} \frac{\mathbf{Var}\, Y_\ell}{N_\ell} + \mathbf{E}\left[Q_{M_L} - Q\right]^2$$

into sample variance and bias (shown as for standard MC above).

Multilevel Monte Carlo Methods

MLMC variance reduction

- Choose discretisation parameters and numbers of samples again to balance the terms in the MSE.
- The bias term is the same as for the standard MC estimator, leading again to a choice of $h_L = h \lesssim \mathrm{TOL}^{1/\alpha}$.

Multilevel Monte Carlo Methods

MLMC variance reduction

- Choose discretisation parameters and numbers of samples again to balance the terms in the MSE.
- The bias term is the same as for the standard MC estimator, leading again to a choice of $h_L = h \lesssim \mathrm{TOL}^{1/\alpha}$.
- **But why** do we get variance reduction – or rather lower cost for the same variance?

Multilevel Monte Carlo Methods

MLMC variance reduction

- Choose discretisation parameters and numbers of samples again to balance the terms in the MSE.
- The bias term is the same as for the standard MC estimator, leading again to a choice of $h_L = h \lesssim \mathrm{TOL}^{1/\alpha}$.
- **But why** do we get variance reduction – or rather lower cost for the same variance? Two reasons:

1. As we **coarsen** the problem $L \to 0$, the cost per sample decays rapidly from level to level, with $\mathscr{O}(s^\gamma)$
2. Since $Q_h \to Q$, then $\mathbf{Var}[Y_\ell] = \mathbf{Var}[Q_{h_\ell} - Q_{h_{\ell-1}}] \to 0$ as $\ell \to \infty$, allowing for smaller and smaller sample sizes N_ℓ on finer and finer levels.

Multilevel Monte Carlo Methods

Optimal sample sizes

- The cost of the MLMC estimator is

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) = \sum_{\ell=0}^{L} N_\ell \mathscr{C}_\ell, \qquad \mathscr{C}_\ell := \mathscr{C}(Y_\ell^{(k)}).$$

Multilevel Monte Carlo Methods

Optimal sample sizes

- The cost of the MLMC estimator is

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) = \sum_{\ell=0}^{L} N_\ell \mathscr{C}_\ell, \qquad \mathscr{C}_\ell := \mathscr{C}(Y_\ell^{(k)}).$$

- Treating the N_ℓ as continuous variables, we can minimise the cost of the MLMC estimator for a fixed variance

$$\sum_{\ell=0}^{L} \frac{\mathbf{Var}\, Y_\ell}{N_\ell} = \frac{\mathrm{TOL}^2}{2}$$

using the **Lagrange multiplier method**.

Multilevel Monte Carlo Methods

Optimal sample sizes

- The cost of the MLMC estimator is

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) = \sum_{\ell=0}^{L} N_\ell \mathscr{C}_\ell, \qquad \mathscr{C}_\ell := \mathscr{C}(Y_\ell^{(k)}).$$

- Treating the N_ℓ as continuous variables, we can minimise the cost of the MLMC estimator for a fixed variance

$$\sum_{\ell=0}^{L} \frac{\mathbf{Var}\, Y_\ell}{N_\ell} = \frac{\mathrm{TOL}^2}{2}$$

  using the **Lagrange multiplier method**.

- The solution to this constrained minimisation problem is

$$N_\ell \simeq \sqrt{\mathbf{Var}[Y_\ell]/\mathscr{C}_\ell}$$

  with implied constant chosen such that the total variance is $\frac{\mathrm{TOL}^2}{2}$ (which leads to the constant $\frac{2}{\mathrm{TOL}^2} \sum_\ell \sqrt{\mathscr{C}_\ell \mathbf{Var}\, Y_\ell}$)

Multilevel Monte Carlo Methods

MLMC cost

- This results in a total cost on level ℓ proportional to $\sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell}$ and therefore

$$\mathscr{C}(\widehat{Q}^{\mathsf{ML}}_{L,\{N_\ell\}}) \leq \frac{2}{\mathsf{TOL}^2} \left(\sum_{\ell=0}^{L} \sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell} \right)^2$$

For comparison, the cost for standard MC is $\mathscr{C}(\widehat{Q}^{\mathsf{SL}}_{h_L,N}) = \frac{2}{\mathsf{TOL}^2} \mathscr{C}_L \, \mathbf{Var}[Q_{h_L}]$.

Multilevel Monte Carlo Methods

MLMC cost

- This results in a total cost on level ℓ proportional to $\sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell}$ and therefore

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) \leq \frac{2}{\mathrm{TOL}^2} \left(\sum_{\ell=0}^{L} \sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell} \right)^2$$

  For comparison, the cost for standard MC is $\mathscr{C}(\widehat{Q}^{\mathrm{SL}}_{h_L,N}) = \frac{2}{\mathrm{TOL}^2} \mathscr{C}_L \, \mathbf{Var}[Q_{h_L}]$.

- If $\mathbf{Var}\, Y_\ell$ decays faster than $\mathscr{C}_\ell$ increases, the cost on level $\ell = 0$ dominates. Since $\mathbf{Var}[Q_{h_0}] \approx \mathbf{Var}[Q_{h_L}]$, the cost ratio of MLMC to MC estimation is then approximately

$$\mathscr{C}_0 / \mathscr{C}_L \eqsim s^{-L\gamma}$$

Multilevel Monte Carlo Methods

MLMC cost

- This results in a total cost on level ℓ proportional to $\sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell}$ and therefore

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) \leq \frac{2}{\mathrm{TOL}^2} \left(\sum_{\ell=0}^{L} \sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell} \right)^2$$

  For comparison, the cost for standard MC is $\mathscr{C}(\widehat{Q}^{\mathrm{SL}}_{h_L,N}) = \frac{2}{\mathrm{TOL}^2} \mathscr{C}_L \, \mathbf{Var}[Q_{h_L}]$.

- If $\mathbf{Var}\, Y_\ell$ decays faster than $\mathscr{C}_\ell$ increases, the cost on level $\ell = 0$ dominates. Since $\mathbf{Var}[Q_{h_0}] \approx \mathbf{Var}[Q_{h_L}]$, the cost ratio of MLMC to MC estimation is then approximately

$$\mathscr{C}_0 / \mathscr{C}_L \eqsim s^{-L\gamma}$$

- If $\mathscr{C}_\ell$ increases faster than $\mathbf{Var}\, Y_\ell$ decays, then the cost on level $\ell = L$ dominates, and then the cost ratio is approximately

$$\mathbf{Var}[Y_L] / \mathbf{Var}[Q_{h_L}] \eqsim \mathrm{TOL}^2$$

  (provided $\mathbf{E}\left[(Q - Q_L)^2\right] \eqsim \left(\mathbf{E}\left[Q - Q_L\right]\right)^2$, which is problem dependent).

Theorem

Let $\mathrm{TOL} < \exp(-1)$ *and assume there are constants* $\alpha, \beta, \gamma > 0$ *such that* $\alpha \geq \min\{\beta, \gamma\}$ *and, for all* $\ell = 0, \ldots, L$,

$$\text{(M1)} \quad |\mathbf{E}[Q_{h_\ell}] - \mathbf{E}[Q]| \lesssim h_\ell^\alpha,$$

$$\text{(M2)} \quad \mathbf{Var}[\widehat{Y}_\ell] \lesssim N_\ell^{-1} h_\ell^\beta,$$

$$\text{(M3)} \quad \mathscr{C}(\widehat{Y}_\ell) \lesssim N_\ell h_\ell^{-\gamma}.$$

Then there are L *and* $\{N_\ell\}_{\ell=0}^L$ *s.t.* $\mathbf{E}\left[\left(\widehat{Q}^{ML}_{L,\{N_\ell\}} - \mathbf{E}[Q]\right)^2\right] \leq \mathrm{TOL}^2$ *and*

$$\mathscr{C}(\widehat{Q}^{ML}_{L,\{N_\ell\}}) \lesssim \begin{cases} \mathrm{TOL}^{-2}, & \text{if } \beta > \gamma, \\ \mathrm{TOL}^{-2} \, |\log \mathrm{TOL}|^2, & \text{if } \beta = \gamma, \\ \mathrm{TOL}^{-2-(\gamma-\beta)/\alpha}, & \text{if } \beta < \gamma. \end{cases}$$

Multilevel Monte Carlo Methods

Adaptive MLMC Algorithm

- The optimal values of L and $N_0, \ldots, N_L$ can be estimated adaptively using the sample averages ($\widehat{Y}_\ell$) and sample variances (s_ℓ^2) of Y_ℓ.
- The sample variances s_ℓ^2 can be used directly to estimate $\mathbf{Var}[\widehat{Y}_\ell]$.

Multilevel Monte Carlo Methods

Adaptive MLMC Algorithm

- The optimal values of L and $N_0, \ldots, N_L$ can be estimated adaptively using the sample averages ($\widehat{Y}_\ell$) and sample variances (s_ℓ^2) of Y_ℓ.
- The sample variances s_ℓ^2 can be used directly to estimate $\mathbf{Var}[\widehat{Y}_\ell]$.
- To bound the bias error, we assume there exists an $h^* > 0$ such that for all $h \leq h^*$

$$h^\alpha \lesssim |\mathbf{E}[Q_h - Q]| \lesssim h^\alpha \,.$$

Multilevel Monte Carlo Methods

Adaptive MLMC Algorithm

- The optimal values of L and $N_0, \ldots, N_L$ can be estimated adaptively using the sample averages ($\widehat{Y}_\ell$) and sample variances (s_ℓ^2) of Y_ℓ.
- The sample variances s_ℓ^2 can be used directly to estimate $\mathbf{Var}[\widehat{Y}_\ell]$.
- To bound the bias error, we assume there exists an $h^* > 0$ such that for all $h \leq h^*$

$$h^\alpha \lesssim |\mathbf{E}[Q_h - Q]| \lesssim h^\alpha .$$

Then it follows (via the inverse triangle inequality) that

$$|\mathbf{E}[Q_{h_L} - Q]| \leq \frac{1}{s^\alpha - 1} \widehat{Y}_L$$

Multilevel Monte Carlo Methods

Adaptive MLMC Algorithm

- The optimal values of L and $N_0, \ldots, N_L$ can be estimated adaptively using the sample averages ($\widehat{Y}_\ell$) and sample variances (s_ℓ^2) of Y_ℓ.
- The sample variances s_ℓ^2 can be used directly to estimate $\mathbf{Var}[\widehat{Y}_\ell]$.
- To bound the bias error, we assume there exists an $h^* > 0$ such that for all $h \leq h^*$

$$h^\alpha \lesssim |\mathbf{E}[Q_h - Q]| \lesssim h^\alpha .$$

Then it follows (via the inverse triangle inequality) that

$$|\mathbf{E}[Q_{h_L} - Q]| \leq \frac{1}{s^\alpha - 1} \widehat{Y}_L$$

- Thus, a computable bias error estimate for given $h_L > 0$, which can be used to decide if h_L is sufficiently small or L should be increased.

Recall the PDE Model Problem

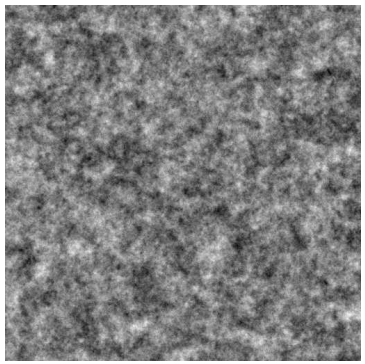
Case Study in Radioactive Waste Disposal

Recall

uncertain k $\rightarrow$ **Darcy's Law:** $\vec{q} + k\nabla p = f$, **Incompressibility:** $\nabla \cdot \vec{q} = 0$ $\rightarrow$ uncertain p, $\vec{q}$

Typical simplified model for k:

- $\log k(x,\omega)$ = isotropic, scalar **Gaussian**
  e.g. with exp. covariance $R(x,y) = \sigma^2 \exp\left(-\frac{\|x-y\|}{\lambda}\right)$
- Parametrise: $\log k(x,\omega) \approx \sum_{j=1}^{J} \sqrt{\mu_j}\psi_j(x) Y_j(\omega)$
  Karhunen-Loève expansion with $Y_j(\omega)$ i.i.d. $N(0,1)$



- **FE discretisation**: $A(\omega)\,\mathbf{P}(\omega) = \mathbf{b}(\omega)$
- QoI $Q(\omega) = \phi(p(\omega))$, e.g. particle travel time from leak to boundary

Recall the PDE Model Problem

Numerical Experiment with standard Monte Carlo

Recall

$D = (0,1)^2$, unconditioned KL expansion, $Q = \| - k \frac{\partial p}{\partial x_1} \|_{L^1(D)}$
using mixed FEs and AMG solver `amg1r5` [Ruge, Stüben, 1992]

- Num. observed FE-error: $\approx \mathscr{O}(h^{-3/4}) \Rightarrow \alpha \approx 3/4$
- Num. observed cost/sample: $\approx \mathscr{O}(h^{-d}) \Rightarrow \gamma \approx 2$
- **Total cost** to get RMSE $\mathscr{O}(\text{TOL})$: $\approx \mathscr{O}(\text{TOL}^{-14/3})$
  to get error reduction by a factor 2 $\rightarrow$ cost grows by a factor 25!

Case 1: $\sigma^2 = 1$, $\lambda = 0.3$, $\nu = 0.5$

TOL	h^{-1}	N	Cost
0.01	129	1.4×10^4	21 min
0.002	1025	3.5×10^5	30 days

Case 2: $\sigma^2 = 3$, $\lambda = 0.1$, $\nu = 0.5$

TOL	h^{-1}	N	Cost
0.01	513	8.5×10^3	4 h
0.002	**Prohibitively large!!**		

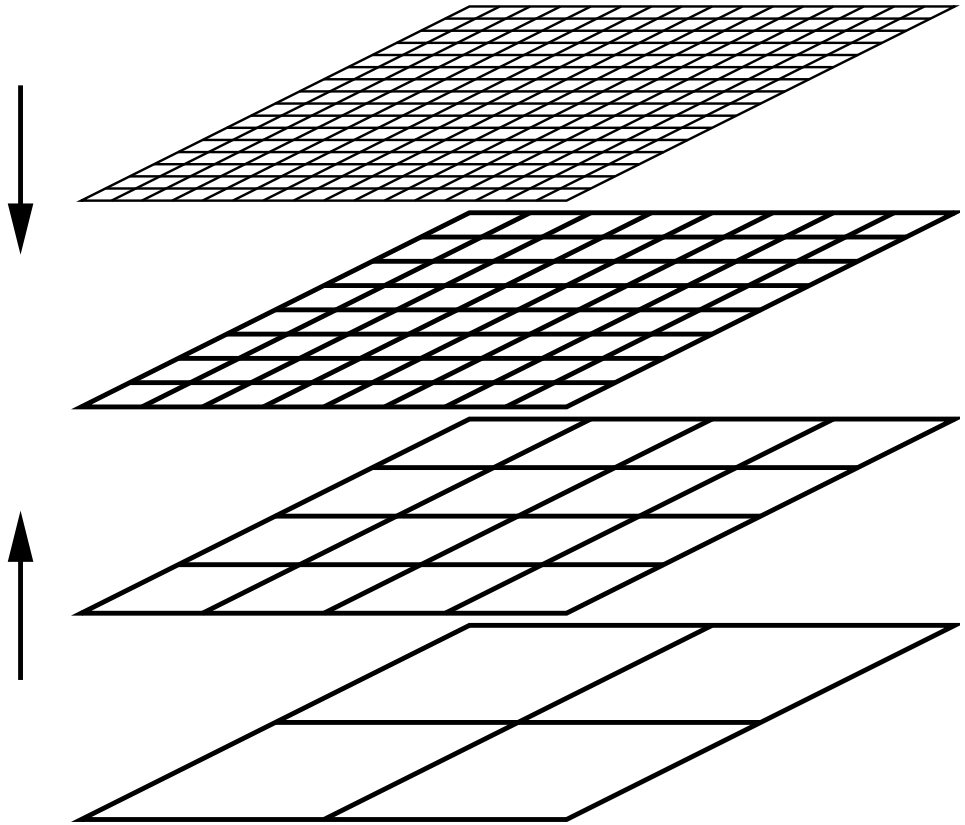
(actual numbers & CPU times on a cluster of 2GHz Intel T7300 processors)

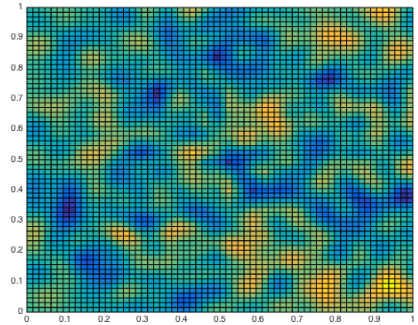
Model Hierarchy (FE mesh & KL truncation)

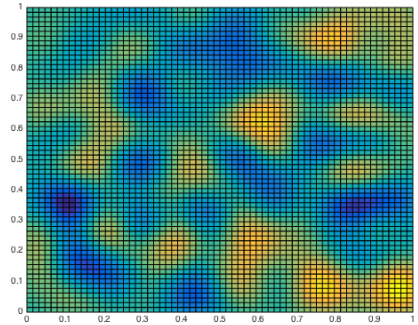
V_ℓ

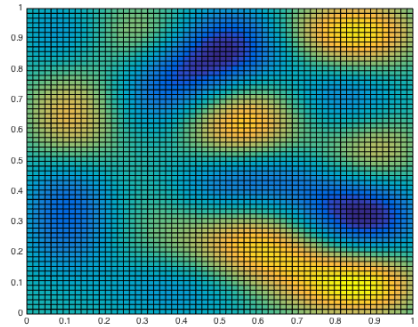
X_ℓ

L









0

Here $\alpha \approx 1$ (for smooth fctls.) and $\gamma \approx d$ (using AMG)

Multilevel Monte Carlo for the PDE Model

Theoretically predicted gains

- Assuming optimal AMG solver (i.e. $\gamma \approx d$), $\alpha \approx 1$ and $\beta \approx 2\alpha$. Then the **cost** in $\mathbb{R}^d$ is

d	**MC**	**MLMC**	cost per sample
1	$\mathscr{O}(\mathrm{TOL}^{-3})$	$\mathscr{O}(\mathrm{TOL}^{-2})$	$\mathscr{O}(\mathrm{TOL}^{-1})$
2	$\mathscr{O}(\mathrm{TOL}^{-4})$	$\mathscr{O}(\mathrm{TOL}^{-2})$	$\mathscr{O}(\mathrm{TOL}^{-2})$
3	$\mathscr{O}(\mathrm{TOL}^{-5})$	$\mathscr{O}(\mathrm{TOL}^{-3})$	$\mathscr{O}(\mathrm{TOL}^{-3})$

Multilevel Monte Carlo for the PDE Model

Theoretically predicted gains

- Assuming optimal AMG solver (i.e. $\gamma \approx d$), $\alpha \approx 1$ and $\beta \approx 2\alpha$. Then the **cost** in $\mathbb{R}^d$ is

d	**MC**	**MLMC**	cost per sample
1	$\mathscr{O}(\text{TOL}^{-3})$	$\mathscr{O}(\text{TOL}^{-2})$	$\mathscr{O}(\text{TOL}^{-1})$
2	$\mathscr{O}(\text{TOL}^{-4})$	$\mathscr{O}(\text{TOL}^{-2})$	$\mathscr{O}(\text{TOL}^{-2})$
3	$\mathscr{O}(\text{TOL}^{-5})$	$\mathscr{O}(\text{TOL}^{-3})$	$\mathscr{O}(\text{TOL}^{-3})$

Optimality (for $d = \gamma > \beta = 2\alpha$)

MLMC cost is asymptotically the same as **one deterministic solve** to accuracy TOL for $d > 1$, i.e. $\mathscr{O}(\text{TOL}^{-\gamma/\alpha})$ **!!** (only for rough problems!)

Multilevel Monte Carlo for the PDE Model

Theoretically predicted gains

- Assuming optimal AMG solver (i.e. $\gamma \approx d$), $\alpha \approx 1$ and $\beta \approx 2\alpha$. Then the **cost** in $\mathbb{R}^d$ is

d	**MC**	**MLMC**	cost per sample
1	$\mathscr{O}(\mathrm{TOL}^{-3})$	$\mathscr{O}(\mathrm{TOL}^{-2})$	$\mathscr{O}(\mathrm{TOL}^{-1})$
2	$\mathscr{O}(\mathrm{TOL}^{-4})$	$\mathscr{O}(\mathrm{TOL}^{-2})$	$\mathscr{O}(\mathrm{TOL}^{-2})$
3	$\mathscr{O}(\mathrm{TOL}^{-5})$	$\mathscr{O}(\mathrm{TOL}^{-3})$	$\mathscr{O}(\mathrm{TOL}^{-3})$

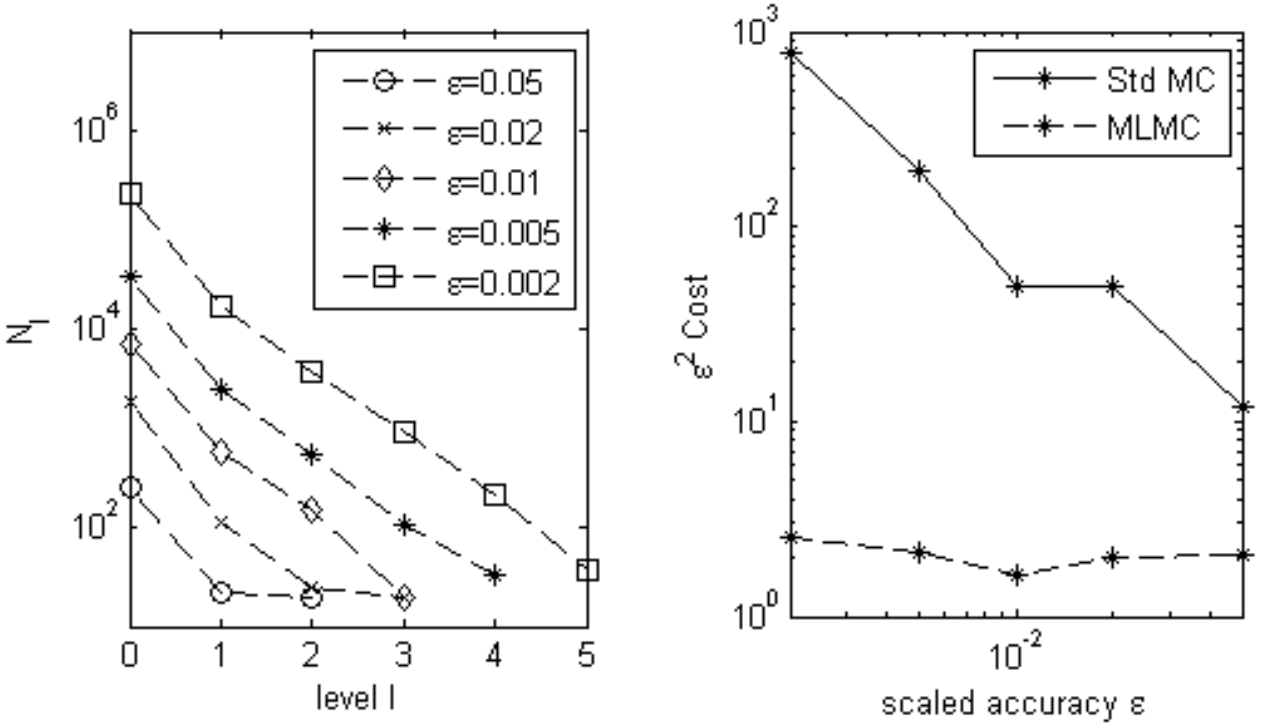
Optimality (for $d = \gamma > \beta = 2\alpha$)

MLMC cost is asymptotically the same as **one deterministic solve** to accuracy TOL for $d > 1$, i.e. $\mathscr{O}(\mathrm{TOL}^{-\gamma/\alpha})$ **!!** (only for rough problems!)

Can we achieve such huge gains in practice?

Multilevel Monte Carlo for the PDE Model

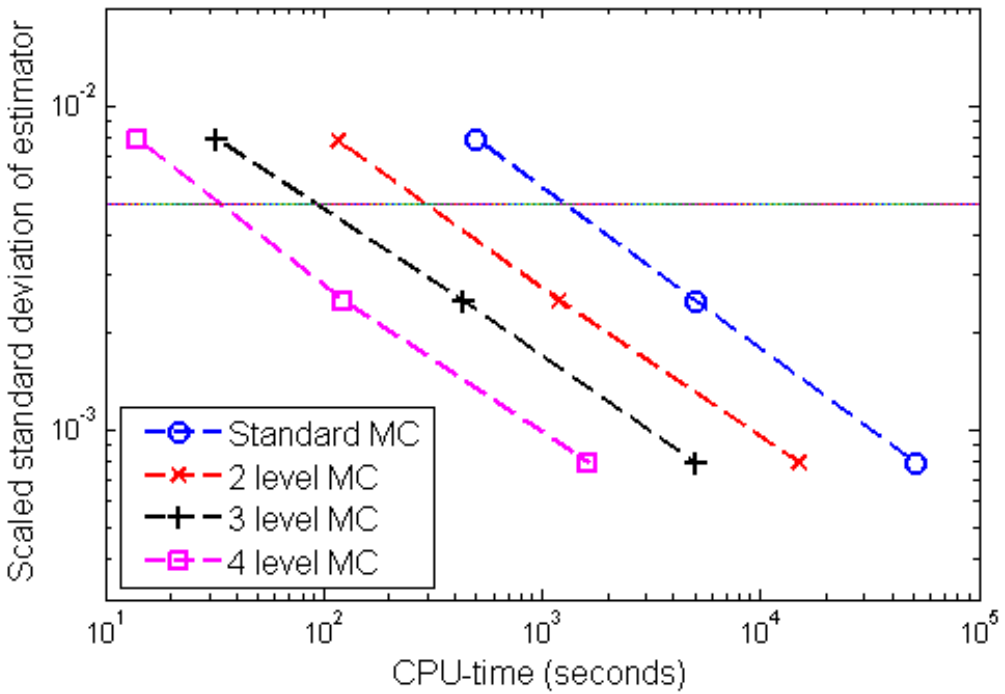
Numerical Experiments: $D = (0,1)^2$; $Q = \|p\|_{L_2(D)}$; standard FEs



$$\nu = \tfrac{1}{2}, \quad \sigma^2 = 1, \quad \lambda = 0.3, \quad h_0 = \tfrac{1}{8}$$

Multilevel Monte Carlo for the PDE Model

Numerical Experiments: $D = (0,1)^2$; $Q = \|p\|_{L_2(D)}$; standard FEs



$h_L = 1/256$ (solid line is FE-error)

`Matlab` implementation on 3GHz Intel Core 2 Duo E8400 processor, 3.2GByte RAM, with **sparse direct solver** where $\gamma \approx 1.2$

Theory: Verifying Assumptions *(M1) & (M2)*

For simplicity only shown for exponential covariance, i.e. $\nu = 1/2$

$$\int_D \nabla v \cdot (k(x,\omega)\nabla p(x,\omega))\,\mathrm{d}x = \int_D f(x,\omega) v\,\mathrm{d}x \quad \forall v \in H_0^1(D)$$

Theory: Verifying Assumptions *(M1) & (M2)*

For simplicity only shown for exponential covariance, i.e. $\nu = 1/2$

$$\int_D \nabla v \cdot (k(x,\omega)\nabla p(x,\omega))\,\mathrm{d}x = \int_D f(x,\omega) v\,\mathrm{d}x \quad \forall v \in H_0^1(D)$$

- [Barth, Schwab, Zollinger '11]: uniformly elliptic & bounded $k(\cdot,\omega) \in W^{1,\infty}(D)$ (not satisfied here!)

Theory: Verifying Assumptions *(M1) & (M2)*

For simplicity only shown for exponential covariance, i.e. $\nu = 1/2$

$$\int_D \nabla v \cdot (k(x,\omega)\nabla p(x,\omega))\,\mathrm{d}x = \int_D f(x,\omega) v\,\mathrm{d}x \quad \forall v \in H_0^1(D)$$

- [Barth, Schwab, Zollinger '11]: uniformly elliptic & bounded $k(\cdot,\omega) \in W^{1,\infty}(D)$ (not satisfied here!)
- [Charrier, RS, Teckentrup '13], [Teckentrup, RS, Giles, Ullmann '14]: **not** uniformly elliptic/bounded k **&** low regularity

Theory: Verifying Assumptions *(M1) & (M2)*

For simplicity only shown for exponential covariance, i.e. $\nu = 1/2$

$$\int_D \nabla v \cdot (k(x,\omega)\nabla p(x,\omega))\,\mathrm{d}x = \int_D f(x,\omega) v\,\mathrm{d}x \quad \forall v \in H_0^1(D)$$

- [Barth, Schwab, Zollinger '11]: uniformly elliptic & bounded $k(\cdot,\omega) \in W^{1,\infty}(D)$ (not satisfied here!)
- [Charrier, RS, Teckentrup '13], [Teckentrup, RS, Giles, Ullmann '14]: **not** uniformly elliptic/bounded k **&** low regularity
- **Why more challenging?**

Theory: Verifying Assumptions *(M1) & (M2)*

For simplicity only shown for exponential covariance, i.e. $\nu = 1/2$

$$\int_D \nabla v \cdot (k(x,\omega)\nabla p(x,\omega))\,\mathrm{d}x = \int_D f(x,\omega) v\,\mathrm{d}x \quad \forall v \in H_0^1(D)$$

- [Barth, Schwab, Zollinger '11]: uniformly elliptic & bounded $k(\cdot,\omega) \in W^{1,\infty}(D)$ (not satisfied here!)
- [Charrier, RS, Teckentrup '13], [Teckentrup, RS, Giles, Ullmann '14]: **not** uniformly elliptic/bounded k **&** low regularity
- **Why more challenging?**
  - **No k^{min}, k^{max} exist** such that

    $$0 < k^{min} \leq k(\mathbf{x},\omega) \leq k^{max} < \infty \qquad \text{a.e. in } D \times \Omega.$$

    Cannot apply Lax-Milgram directly to ensure $\exists!$

Theory: Verifying Assumptions *(M1) & (M2)*

For simplicity only shown for exponential covariance, i.e. $\nu = 1/2$

$$\int_D \nabla v \cdot (k(x,\omega)\nabla p(x,\omega))\,\mathrm{d}x = \int_D f(x,\omega) v\,\mathrm{d}x \quad \forall v \in H_0^1(D)$$

- [Barth, Schwab, Zollinger '11]: uniformly elliptic & bounded $k(\cdot,\omega) \in W^{1,\infty}(D)$ (not satisfied here!)
- [Charrier, RS, Teckentrup '13], [Teckentrup, RS, Giles, Ullmann '14]: **not** uniformly elliptic/bounded k **&** low regularity
- **Why more challenging?**
  - **No k^{min}, k^{max} exist** such that

    $$0 < k^{min} \leq k(\mathbf{x},\omega) \leq k^{max} < \infty \qquad \text{a.e. in } D \times \Omega.$$

    Cannot apply Lax-Milgram directly to ensure $\exists!$
  - Only $k(\cdot,\omega) \in C^\eta(D)$ with $\eta < \frac{1}{2} \Rightarrow p(\omega,\cdot) \notin H^2(D)$

- Instead work directly (a.s. in ω) with

$$k^{min}(\omega) := \min_{\mathbf{x}} k(\mathbf{x},\omega) > 0 \quad \text{and} \quad k^{max}(\omega) := \max_{\mathbf{x}} k(\mathbf{x},\omega) < +\infty.$$

- Instead work directly (a.s. in ω) with

$$k^{min}(\omega) := \min_{\mathbf{x}} k(\mathbf{x},\omega) > 0 \quad \text{and} \quad k^{max}(\omega) := \max_{\mathbf{x}} k(\mathbf{x},\omega) < +\infty.$$

- Apply **Lax-Milgram** pathwise (a.s. in ω): $\exists!\ p(\omega) \in H^1_0(D)$.

- Instead work directly (a.s. in ω) with

$$k^{min}(\omega) := \min_{\mathbf{x}} k(\mathbf{x}, \omega) > 0 \quad \text{and} \quad k^{max}(\omega) := \max_{\mathbf{x}} k(\mathbf{x}, \omega) < +\infty.$$

- Apply **Lax-Milgram** pathwise (a.s. in ω): $\exists!\ p(\omega) \in H_0^1(D)$.

- Use **Fernique's Theorem** (for Gaussians) to show that $\forall q \in (0, \infty)$

$$1/k^{min} \in L^q(\Omega), \quad k^{max} \in L^q(\Omega), \quad \text{and} \quad k \in L^q(\Omega, C^\eta(D)).$$

- Instead work directly (a.s. in ω) with

$$k^{min}(\omega) := \min_{\mathbf{x}} k(\mathbf{x},\omega) > 0 \quad \text{and} \quad k^{max}(\omega) := \max_{\mathbf{x}} k(\mathbf{x},\omega) < +\infty.$$

- Apply **Lax-Milgram** pathwise (a.s. in ω): $\exists!\ p(\omega) \in H_0^1(D)$.

- Use **Fernique's Theorem** (for Gaussians) to show that $\forall q \in (0,\infty)$

$$1/k^{min} \in L^q(\Omega), \quad k^{max} \in L^q(\Omega), \quad \text{and} \quad k \in L^q(\Omega, C^\eta(D)).$$

Theorem (Regularity [Charrier, RS, Teckentrup '13])

If D is piecewise C^2, $\mathbb{E}[k]$ is piecewise C^η, $f \in L^{q^*}(\Omega, H^{\eta-1}(D))$, then

$$\|p(\omega)\|_{H^{1+\tau}(D)} \lesssim \frac{k^{max}(\omega)\|k(\cdot,\omega)\|^2_{C^\eta(\bar{D})}}{k^{min}(\omega)^4} \|f(\cdot,\omega)\|_{H^{\eta-1}(D)},$$

for any $\tau < \eta < 1/2$. Hence $p \in L^q(\Omega, H^{1+\tau}(D))$, for all $q < q^*$.

- Instead work directly (a.s. in ω) with

$$k^{min}(\omega) := \min_{\mathbf{x}} k(\mathbf{x},\omega) > 0 \quad \text{and} \quad k^{max}(\omega) := \max_{\mathbf{x}} k(\mathbf{x},\omega) < +\infty.$$

- Apply **Lax-Milgram** pathwise (a.s. in ω): $\exists!\ p(\omega) \in H^1_0(D)$.

- Use **Fernique's Theorem** (for Gaussians) to show that $\forall q \in (0,\infty)$

$$1/k^{min} \in L^q(\Omega), \quad k^{max} \in L^q(\Omega), \quad \text{and} \quad k \in L^q(\Omega, C^\eta(D)).$$

Theorem (Regularity [Charrier, RS, Teckentrup '13])

If D is piecewise C^2, $\mathbb{E}[k]$ is piecewise C^η, $f \in L^{q^*}(\Omega, H^{\eta-1}(D))$, then

$$\|p(\omega)\|_{H^{1+\tau}(D)} \lesssim \frac{k^{max}(\omega)\|k(\cdot,\omega)\|^2_{C^\eta(\bar{D})}}{k^{min}(\omega)^4} \|f(\cdot,\omega)\|_{H^{\eta-1}(D)},$$

for any $\tau < \eta < 1/2$. Hence $p \in L^q(\Omega, H^{1+\tau}(D))$, for all $q < q^*$.

Following "classic" proofs in [Hackbusch '92] & [Grisvard '85] ($>$ 10 pages**!**)

- **Céa's Lemma**:

$$|p(\cdot,\omega) - p_h(\cdot,\omega)|_{H^1(D)} \leq \sqrt{\frac{k^{max}(\omega)}{k^{min}(\omega)}} \inf_{v_h \in V_h} |p(\cdot,\omega) - v_h|_{H^1(D)}$$

- **Céa's Lemma**:

$$|p(\cdot,\omega) - p_h(\cdot,\omega)|_{H^1(D)} \leq \sqrt{\frac{k^{max}(\omega)}{k^{min}(\omega)}} \inf_{v_h \in V_h} |p(\cdot,\omega) - v_h|_{H^1(D)}$$

- **Best Approximation**:

$$\inf_{v_h \in V_h} |p(\cdot,\omega) - v_h|_{H^1(D)} \leq C \, \|p(\cdot,\omega)\|_{H^{1+\tau}(D)} \, h^\tau \quad C \neq C(\omega)$$

- **Céa's Lemma**:

$$|p(\cdot,\omega) - p_h(\cdot,\omega)|_{H^1(D)} \leq \sqrt{\frac{k^{max}(\omega)}{k^{min}(\omega)}} \inf_{v_h \in V_h} |p(\cdot,\omega) - v_h|_{H^1(D)}$$

- **Best Approximation**:

$$\inf_{v_h \in V_h} |p(\cdot,\omega) - v_h|_{H^1(D)} \leq C \, \|p(\cdot,\omega)\|_{H^{1+\tau}(D)} \, h^\tau \quad C \neq C(\omega)$$

- **Duality Argument** (for smooth, Fréchet-diff'ble functionals ϕ):

$$|\phi(p(\omega,\cdot)) - \phi(p_h(\omega,\cdot))| \leq$$
$$k_{\max}(\omega) \, |p(\omega,\cdot) - p_h(\omega,\cdot)|_{H^1(D)} \, |z(\omega,\cdot) - z_h(\omega,\cdot)|_{H^1(D)}$$

- **Céa's Lemma**:

$$|p(\cdot,\omega) - p_h(\cdot,\omega)|_{H^1(D)} \leq \sqrt{\frac{k^{max}(\omega)}{k^{min}(\omega)}} \inf_{v_h \in V_h} |p(\cdot,\omega) - v_h|_{H^1(D)}$$

- **Best Approximation**:

$$\inf_{v_h \in V_h} |p(\cdot,\omega) - v_h|_{H^1(D)} \leq C \, \|p(\cdot,\omega)\|_{H^{1+\tau}(D)} \, h^\tau \quad C \neq C(\omega)$$

- **Duality Argument** (for smooth, Fréchet-diff'ble functionals ϕ):

$$|\phi(p(\omega,\cdot)) - \phi(p_h(\omega,\cdot))| \leq$$
$$k_{\max}(\omega) \, |p(\omega,\cdot) - p_h(\omega,\cdot)|_{H^1(D)} \, |z(\omega,\cdot) - z_h(\omega,\cdot)|_{H^1(D)}$$

Theorem (FE Error [Charrier, RS, Teckentrup '13])

$$\left.\begin{aligned} \|p - p_h\|_{L^q(\Omega,H^1(D))} &\leq C_{\tau,q} \|f\|_{L^{q^*}(\Omega,H^{\tau-1}(D))} \, h^\tau, \\ \|p - p_h\|_{L^q(\Omega,L^2(D))} &\leq C'_{\tau,q} \|f\|_{L^{q^*}(\Omega,H^{\tau-1}(D))} \, h^{2\tau}, \end{aligned}\right\} \quad \begin{aligned} \forall\, \tau < \eta < \tfrac{1}{2}, \\ \forall\, q < q^*. \end{aligned}$$

- **Céa's Lemma**:

$$|p(\cdot,\omega) - p_h(\cdot,\omega)|_{H^1(D)} \leq \sqrt{\frac{k^{max}(\omega)}{k^{min}(\omega)}} \inf_{v_h \in V_h} |p(\cdot,\omega) - v_h|_{H^1(D)}$$

- **Best Approximation**:

$$\inf_{v_h \in V_h} |p(\cdot,\omega) - v_h|_{H^1(D)} \leq C \, \|p(\cdot,\omega)\|_{H^{1+\tau}(D)} \, h^\tau \quad C \neq C(\omega)$$

- **Duality Argument** (for smooth, Fréchet-diff'ble functionals ϕ):

$$|\phi(p(\omega,\cdot)) - \phi(p_h(\omega,\cdot))| \leq$$
$$k_{\max}(\omega) \, |p(\omega,\cdot) - p_h(\omega,\cdot)|_{H^1(D)} \, |z(\omega,\cdot) - z_h(\omega,\cdot)|_{H^1(D)}$$

Theorem (FE Error [Teckentrup, RS, Giles, Ullmann '14])

If there exists a $C_\phi \in L^{q^*}(\Omega)$ *such that* $\phi(v) \leq C_\phi(\omega)\|v\|_{H^{1-\tau}(D)}$ *a.s. in* ω *and for all* $v \in H_0^1(D)$*, then*

$$\|\phi(p) - \phi(p_h)\|_{L^q(\Omega)} \leq C''_{\tau,q} \, h^{2\tau}, \quad \forall \tau < \eta < \frac{1}{2} \text{ and } q < q^*.$$

- Thus, with $q = 1$ we get

$$\Big|\mathbb{E}\big[\phi(p) - \phi(p_h)\big]\Big| \;=\; \|\phi(p) - \phi(p_h)\|_{L^1(\Omega)} \;=\; \mathscr{O}(h^{2\tau})$$

$$\Longrightarrow \quad \boxed{\textit{(M1)}\ \text{satisfied for any}\ \alpha < 1}$$

- Thus, with $q = 1$ we get

$$\Big|\mathbb{E}\big[\phi(p) - \phi(p_h)\big]\Big| \;=\; \|\phi(p) - \phi(p_h)\|_{L^1(\Omega)} \;=\; \mathscr{O}(h^{2\tau})$$

$$\Longrightarrow \quad \boxed{\textit{(M1)}\ \text{satisfied for any } \alpha < 1}$$

- And with $q = 2$ we get

$$\mathbb{V}\big[\phi(p_h) - \phi(p_{2h})\big] \;\leq\; \|\phi(p_h) - \phi(p_{2h})\|^2_{L^2(\Omega)} \;\leq\; \mathscr{O}(h^{4\tau})$$

$$\Longrightarrow \quad \boxed{\textit{(M2)}\ \text{satisfied for any } \beta < 2}$$

- Thus, with $q = 1$ we get

$$\Big|\mathbb{E}\big[\phi(p) - \phi(p_h)\big]\Big| \;=\; \|\phi(p) - \phi(p_h)\|_{L^1(\Omega)} \;=\; \mathscr{O}(h^{2\tau})$$

$$\Longrightarrow \quad \boxed{\textit{(M1)}\ \text{satisfied for any } \alpha < 1}$$

- And with $q = 2$ we get

$$\mathbb{V}\big[\phi(p_h) - \phi(p_{2h})\big] \;\leq\; \|\phi(p_h) - \phi(p_{2h})\|^2_{L^2(\Omega)} \;\leq\; \mathscr{O}(h^{4\tau})$$

$$\Longrightarrow \quad \boxed{\textit{(M2)}\ \text{satisfied for any } \beta < 2}$$

- Hence $\boxed{\textbf{Cost} = \mathscr{O}(\mathrm{TOL}^{-\gamma})}$ **(Same as for deterministic solve!)**

- Thus, with $q = 1$ we get

$$\Big|\mathbb{E}\big[\phi(p) - \phi(p_h)\big]\Big| \;=\; \|\phi(p) - \phi(p_h)\|_{L^1(\Omega)} \;=\; \mathscr{O}(h^{2\tau})$$

$$\Longrightarrow \quad \boxed{\textit{(M1)}\text{ satisfied for any } \alpha < 1}$$

- And with $q = 2$ we get

$$\mathbb{V}\big[\phi(p_h) - \phi(p_{2h})\big] \;\leq\; \|\phi(p_h) - \phi(p_{2h})\|^2_{L^2(\Omega)} \;\leq\; \mathscr{O}(h^{4\tau})$$

$$\Longrightarrow \quad \boxed{\textit{(M2)}\text{ satisfied for any } \beta < 2}$$

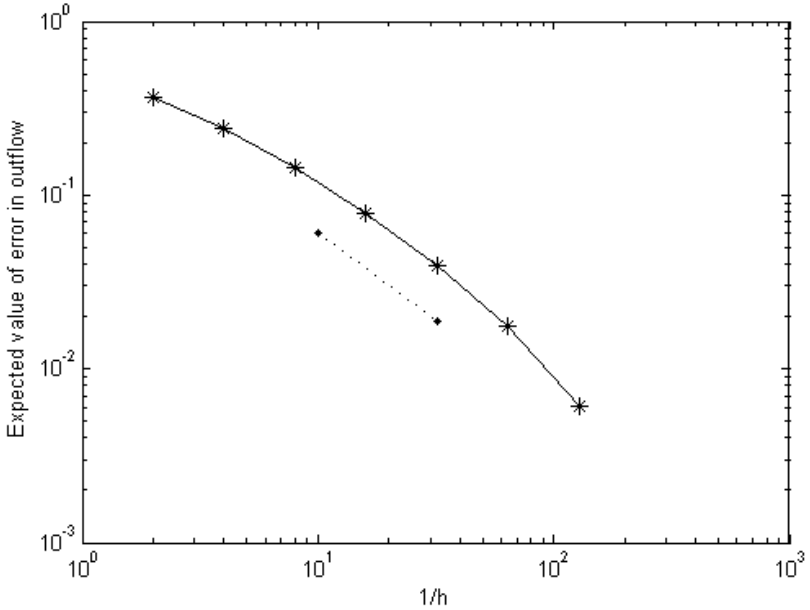
- Hence $\boxed{\textbf{Cost} = \mathscr{O}(\mathrm{TOL}^{-\gamma})}$ **(Same as for deterministic solve!)**

Hence optimal and robust deterministic solver with $\gamma = d$ **crucial**!

This is a whole talk in itself!

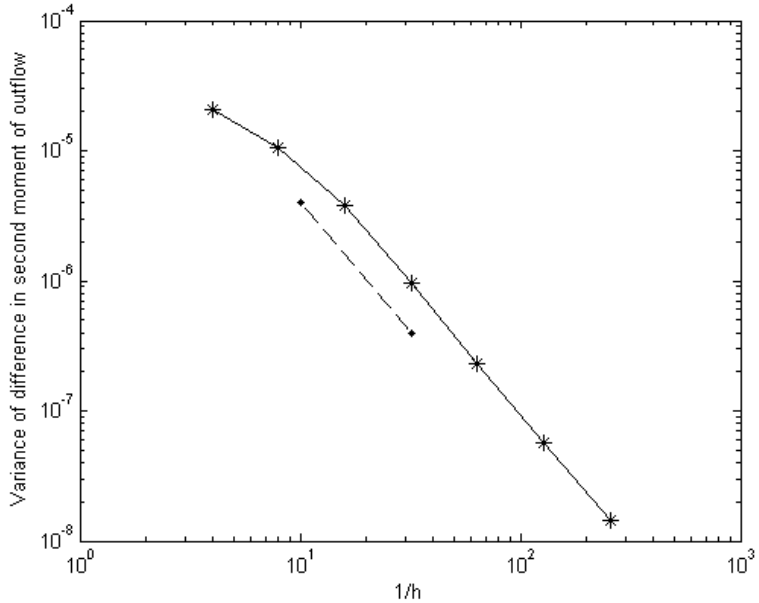
Numerical Confirmation

$D = (0,1)^2$; covariance $R(x,y) = \exp\left(-\frac{\|x-y\|_2}{0.3}\right)$; std. FEs, circulant embedding



$$\left|\mathbb{E}\left[\phi^{(1)}(p) - \phi^{(1)}(p_h)\right]\right|$$

where $\phi^{(1)}(p) := L_\omega(\Psi) - b_\omega(\Psi, v)$
given $\Psi(x) = x$ (outflow on right)



$$\mathbb{V}\left[\phi^{(2)}(p_h) - \phi^{(2)}(p_{2h})\right]$$

where $\phi^{(2)}(p) := \left(\frac{1}{|D^*|}\int_{D^*} p(x)\,\mathrm{d}x\right)^2$
(i.e. 2nd moment of p over small patch)

$$\alpha = 1 \quad \text{and} \quad \beta = 2$$

Other developments in MLMC

- discontinuous coefficients, point evaluations, mixed FEs, FVM, ...
- **Many other PDEs and applications**
- can optimise **all** parameters (not just $\{N_\ell\}$) [Haji Ali, Tempone]
- adaptivity [Von Schwerin, Tempone et al], [Kornhuber, Youett]
- variance estimation [Bierig, Chernov]
- optimal estimation of CDFs, PDFs [Giles, Nagapetyan, Ritter]
- hybrid with stochastic collocation [Tesei, Nobile et al]
- generalisation to general multilevel quadrature [Harbrecht et al]
- multilevel QMC [Kuo, Schwab, Sloan]
- multilevel SC [Teckentrup, Jantsch, Webster, Gunzburger]
- MLMCMC [Hoang, Schwab, Stuart], [Dodwell, Ketelson, RS, Teckentrup]
- etc ...

Some Key References

- Heinrich, MC complexity of global solution of integral eq., *J Complex*, 1998
- Giles, Multilevel Monte Carlo path simulation, *Oper Res*, 2008
- Giles, Waterhouse, MLQMC path simulation, Radon Series, deGruyter, 2009
- Cliffe, Giles, RS, Teckentrup, MLMC methods & applications to elliptic PDEs with random coefficients, *Comput Visual Sci*, 2011
- Barth, Schwab, Zollinger, MLMC FE method for elliptic PDEs with stochastic coefficients, *Numer Math*, 2011
- Charrier, RS, Teckentrup, FE error analysis of elliptic PDEs with random coefficients and its application to MLMC methods, *SINUM*, 2013
- Teckentrup, RS, Giles, Ullmann, Further analysis of MLMC methods for elliptic PDEs with random coefficients, *Numer Math*, 2013
- Kuo, Schwab, Sloan, MLQMC FE methods for a class of elliptic PDEs with random coefficients, *FoCoM*, 2015
- Kuo, RS, Schwab, Sloan, Ullmann, MLQMC methods for lognormal diffusion problems, *Math Comput*, 2017
- Teckentrup, Jantsch, Webster, Gunzburger, A Multilevel Stochastic Collocation Method for PDEs with Random Input Data, *JUQ*, 2015