Stochastic spectral methods for Bayesian inference in inverse problems

Youssef Marzouk & Habib Najm
Sandia National Laboratories
Livermore, CA 94551 USA

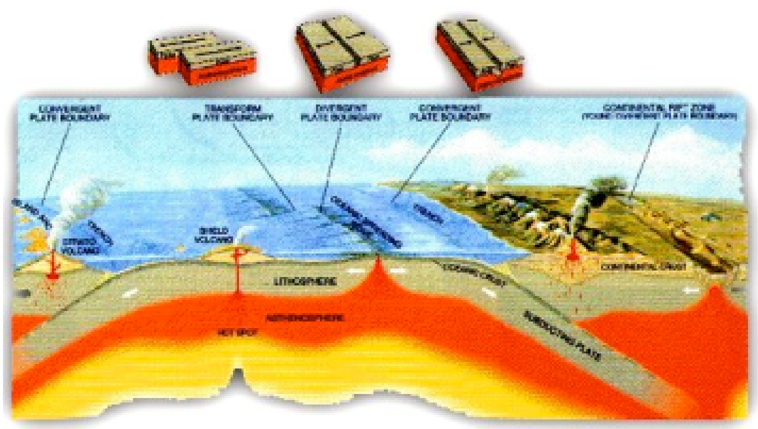
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

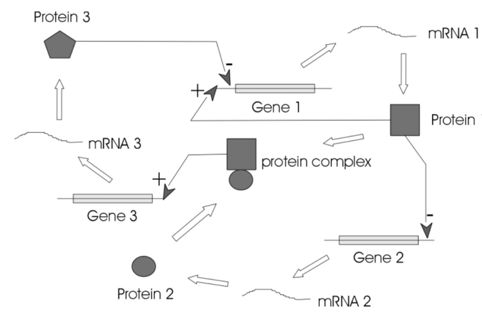
A Department of Energy National Laboratory

Inverse problems

From ***indirect*** observations to physical parameters and models…



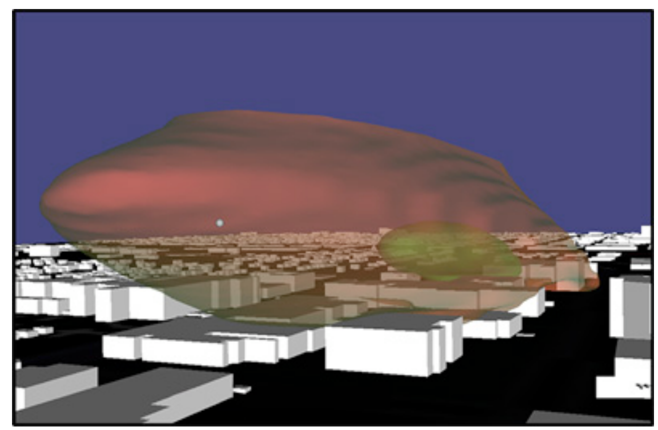
geophysics



chemical reaction networks

[J. Gebbert, U Köln]

medical imaging & tomography

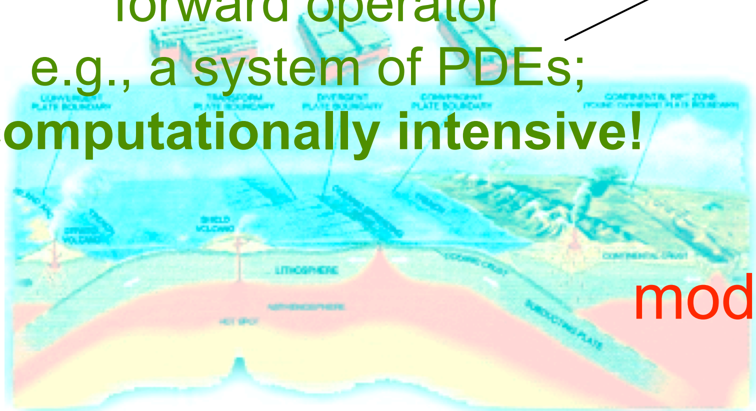


source inversion
(environmental dispersion)

Inverse problems

$$G(\boldsymbol{m}) \approx \boldsymbol{d}$$

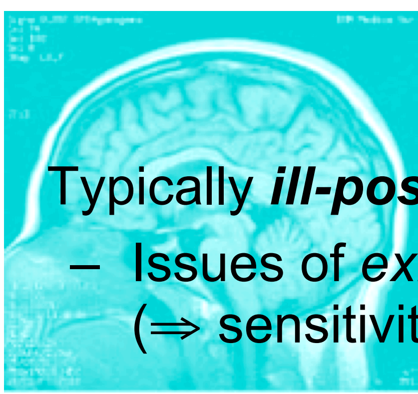
forward operator
e.g., a system of PDEs;
computationally intensive!



observed data

model parameters/inputs

Given a set of data *d*, estimate *m*



- Typically ***ill-posed***:
  - Issues of *existence, uniqueness,* and *stability* (⇒ sensitivity to noise)

Statistical approaches

- **Bayesian inference** for inverse problems
  - The model ***m*** is now a random variable/field
  - Apply Bayes’ theorem:

$$\underbrace{p(m|d)}_{\text{posterior density}} = \frac{\overbrace{p(d|m)}^{\text{likelihood function L(m)}}\,\overbrace{p(m)}^{\text{prior density}}}{\underbrace{\int p(d|m)\,p(m)\,dm}_{\text{evidence (here just a normalizing const)}}}$$

- The **posterior density** $\pi(m) \equiv p(m|d)$ **is** the full Bayesian solution to the inverse problem
  - Not just a single value for *m*, but a probability density
  - A complete description of uncertainty

Bayesian inference for IPs

likelihood function L(m)

prior density

$$\underbrace{p(m|d)}_{\text{posterior density}} = \frac{p(d|m)\,p(m)}{\int p(d|m)\,p(m)\,dm}$$

posterior density

- **Likelihood** function: $L(m) \equiv p(d|m)$
  (how well does the model support the data?)
  - *Example:* deterministic forward problem $\boldsymbol{G(m)}$ additive measurement + model error $\boldsymbol{\eta} \sim p_\eta(\cdot)$

$$\boxed{d = G(m) + \eta} \quad \longrightarrow \quad L(m) = p_\eta\left(G(m) - d\right)$$

  - Simple choice: $\boldsymbol{\eta} \sim N(0, \sigma^2 \mathbf{I})$

Bayesian inference for IPs

likelihood function L(m)

prior density

$$p(m|d) = \frac{p(d|m)\,p(m)}{\int p(d|m)\,p(m)\,dm}$$

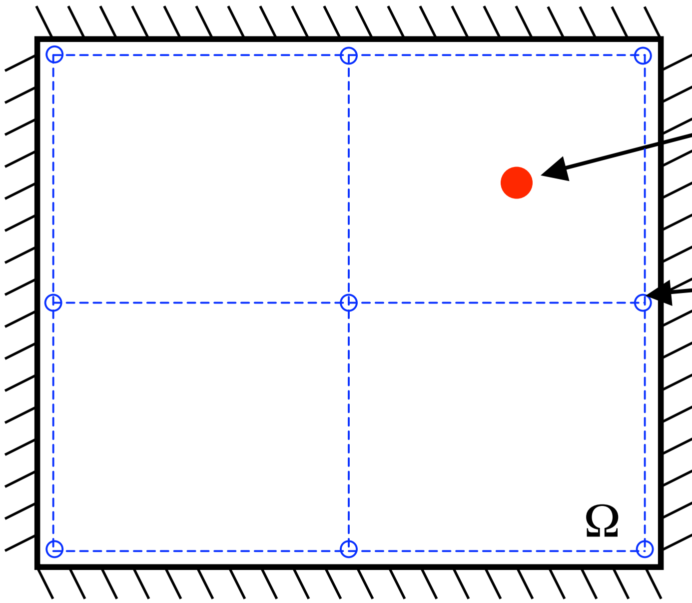
posterior density

- **Prior** *$p_m(m)$*:
  - Incorporates additional information (e.g., physical constraints, smoothness, structure, expert judgment)
  - No regularization parameter *per se*
- Use shorthand: $\pi_{m|d}(m) \propto p(d|m)\,p_m(m)$
- Useful extension— hyperparameters θ, ϕ:

$$p(m,\phi,\theta|d) \propto p(d|m,\phi)\,p(m|\theta)\,p(\theta)\,p(\phi)$$

Example: source inversion

I. Transient diffusion problem—



source described by location parameters m = χ, active for t $\in$ [0,0.2]; strength *s*, width σ

Data from M sensors on a regular grid; $d = \{c_{t1}, c_{t2}, \ldots\}_{i=1\ldots M}$

$\Omega = [0,1]\times[0,1]$

$$\frac{\partial c}{\partial t} = \nabla^2 c + \frac{s}{2\pi\sigma^2}\exp\left(-\frac{|\boldsymbol{\chi}-\boldsymbol{x}|^2}{2\sigma^2}\right)\left[1-H(t-\tau)\right]$$

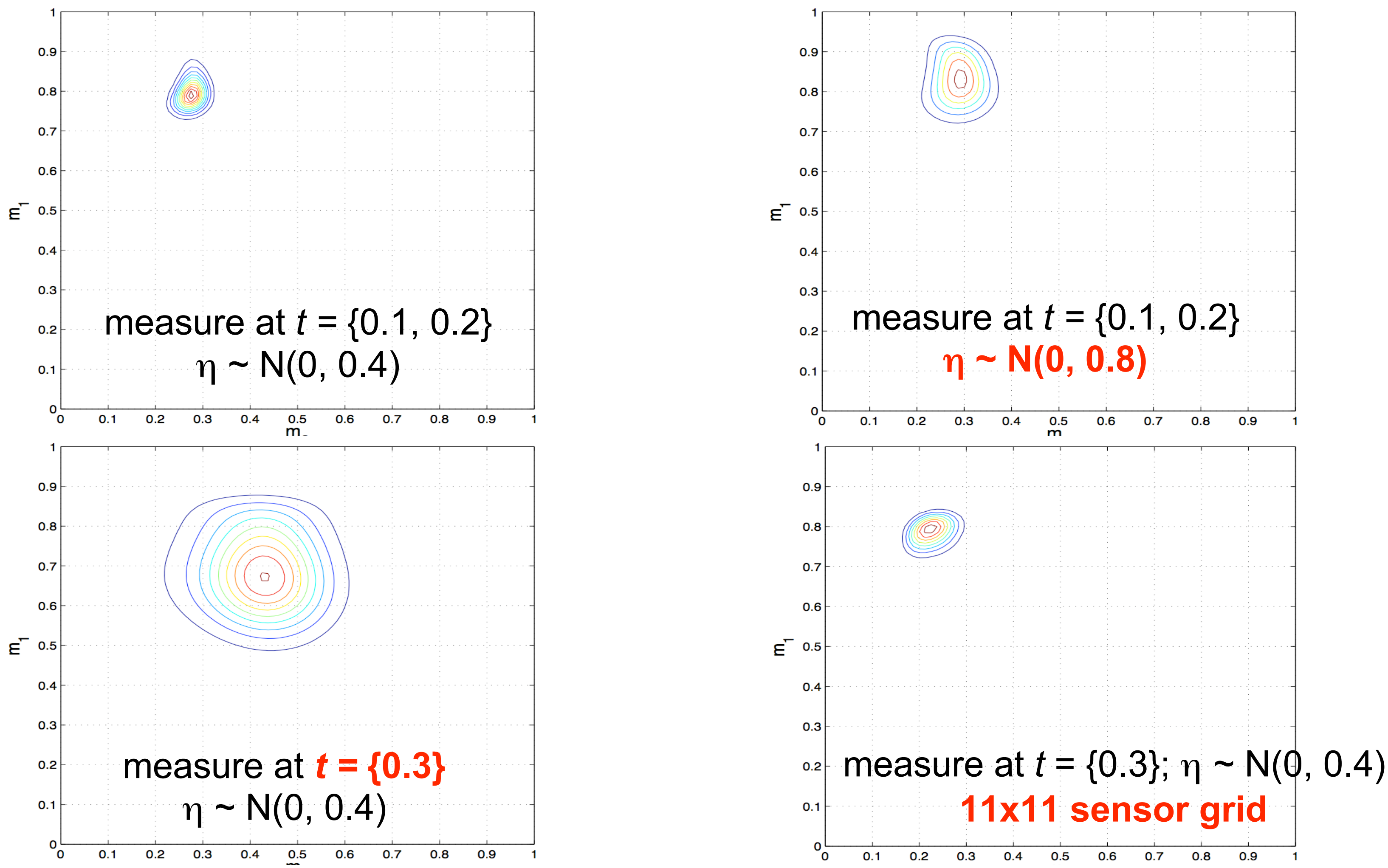
$\nabla c \cdot \hat{n} = 0 \; on \; \partial\Omega, \quad c(\boldsymbol{x},0) = 0$

→ Measurement **noise/error:** $\eta_i \sim N(0,0.4)$

Priors: $\chi = (m_0, m_1) \sim U(0,1)$

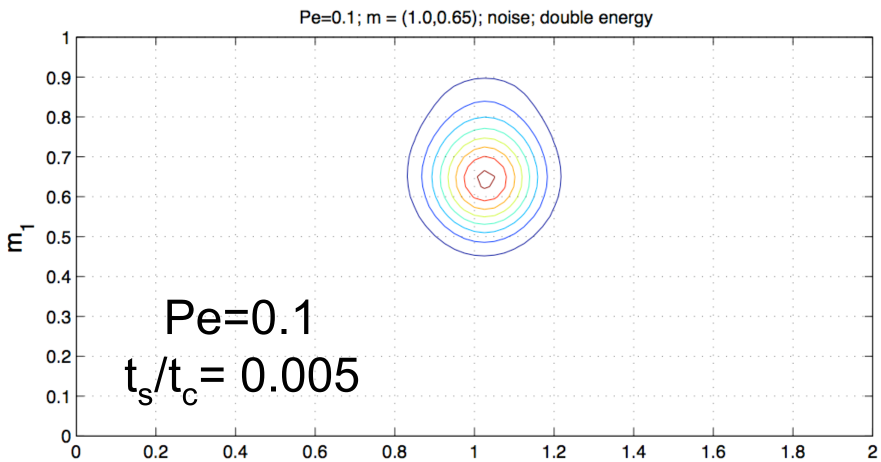
Posterior density

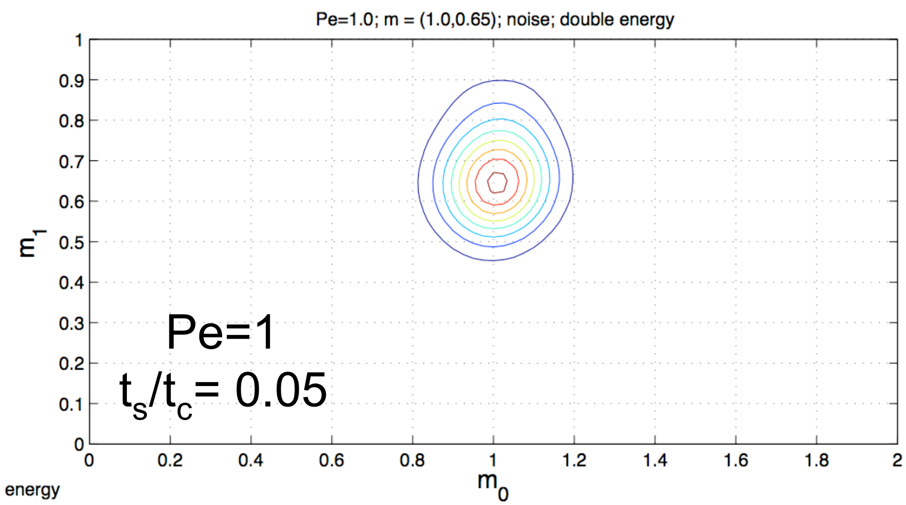
- Posterior reflects data noise/error, length & time scales of measurement, and the underlying physical process:

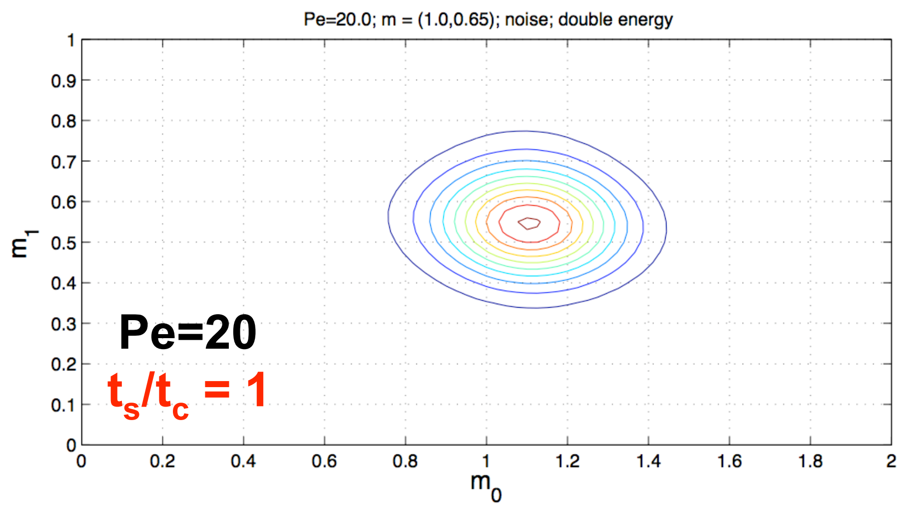


Advection-diffusion problems

- Vary the velocity; measure always at the *same* fraction of the diffusion timescale *(t/t_d=0.10)*; source active for $t/t_d \in [0, 0.05]$







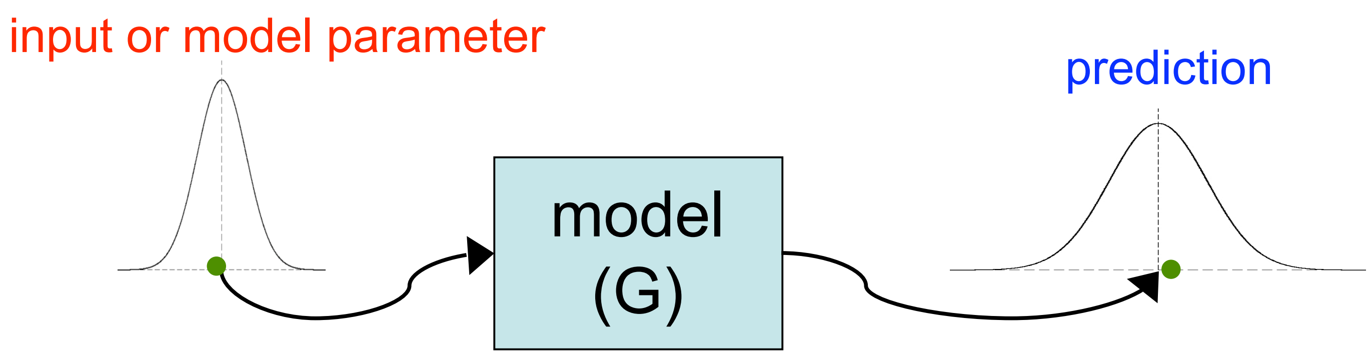
∴ posterior broadens in the streamwise direction for $t_s \geq t_c$

Outline

1. Bayesian approach to inverse problems
2. Accelerating Bayesian inference via stochastic spectral methods
3. Inference of spatiotemporal fields
4. Conclusions

Computational challenges

- **What information to extract from the posterior $\pi(m|d)$?**
  - Posterior means, variances, higher moments:
    $E_\pi f = \int f(m)\pi(m)dm$
  - Marginal distributions $\pi(m_i)$
- Use quadrature/cubature or **sampling** (MCMC)
  - Posterior evaluations are **expensive**
    (forward problems are **simulation-based!**)
  - Posteriors are often **high-dimensional**
- Similar challenges faced in “forward” UQ…



Spectral rep'n of random variables

- **Efficient UQ** relies on the **polynomial chaos** expansion (PCe)
  - Let X be a real second-order random process, defined on a probability space (Ω, U, P)

PCe: $$X(\omega) = \sum_{k=0}^{\infty} x_k \Psi_k\left(\xi_{i_1}, \xi_{i_2}, \dots\right)$$

  - $\left\{\xi_i(\omega)\right\}_{i=1}^{\infty}$ are i.i.d. random variables (e.g., Gaussians)
  - $\left\{\Psi_k(\xi)\right\}$ are orthogonal multivariate polynomials (e.g., Hermite)

$$\left\langle \Psi_i \Psi_j \right\rangle = \int_{\Omega} \Psi_i(\xi) \Psi_j(\xi) \, dP(\omega) = \delta_{ij} \left\langle \Psi_i^2 \right\rangle$$

- Orthogonality wrpt measure on ξ determines polynomial (Ψ) family: (Hermite + Gaussian, Legendre + uniform, Laguerre + Gamma, etc)

- Truncate at polynomial order p & "stochastic dimension" n : $\Rightarrow$ size of basis: $P + 1 = \dfrac{(n+p)!}{n!\,p!}$

Propagating uncertainty w/PCe

⇒ Given a PCe for *X*: $X = \sum_{k=0}^{P} x_k \Psi_k(\xi)$

+ governing equations $O(Y, X) = 0$

→ How to obtain a PCe of Y? $Y = \sum_{k=0}^{P} y_k \Psi_k(\xi)$

- **Galerkin method ("intrusive projection"):**

$$\langle O(Y,X)\ \Psi_l \rangle = \left\langle O\left(\sum_{k=0}^{P} y_k \Psi_k, \sum_{j=0}^{P} x_j \Psi_j \right) \Psi_l \right\rangle = 0 \quad \text{for} \quad l = 0 \ldots P$$

⇒ leads to P+1 coupled problems (reformulation of governing equations). But solve only ONCE!

= efficient propagation of uncertainty from X to Y

Propagating uncertainty w/PCe

- ODE example:

$$\frac{du}{dt} = \lambda u \qquad (\lambda \text{ uncertain})$$

  – Introduce PCes:

$$\lambda = \sum_{k=0}^{P} \lambda_k \Psi_k(\xi), \quad u = \sum_{k=0}^{P} u_k \Psi_k(\xi)$$

  – Apply Galerkin projection:

$$\frac{du_i}{dt} = \sum_{p=0}^{P} \sum_{q=0}^{P} \lambda_p u_q C_{pqi}, \quad i = 0 \ldots P$$

where $C_{pqi} = \left\langle \Psi_p \Psi_q \Psi_i \right\rangle \big/ \left\langle \Psi_i^2 \right\rangle$ are known coefficients

- **In practice—** elementary functions implemented in a **library** for "stochastic arithmetic." Pseudo-spectral construction & other approaches for non-polynomial funcs. [Debusschere *et al* SIAM 2004]

PCe in Bayesian inference

- Write a PCe for ***m*** ~ **prior**
- Propagate prior uncertainty through the forward problem

$$m_i = \sum_{k=0}^{P} \tilde{m}_{ik} \Psi_k(\xi) \rightarrow \boxed{G} \rightarrow \tilde{G}_j(\xi) = \sum_{k=0}^{P} g_{jk} \Psi_k(\xi)$$

→ result: $\tilde{G}(\xi)$, a stochastic spectral representation of forward model predictions → (compute g_{jk} only **once!**) ←

→ Use this as a ***surrogate*** for the forward model in the likelihood function. No repeated forward problem solutions!

PCe in Bayesian inference

- Likelihood function with PCe:

$$L(\xi) = \prod_i p_\eta\left(d_i - \tilde{G}_i(\xi)\right)$$

- **More generally:** PCe is a change of variables ***m*** = ***g(ξ)***:

$$\int_M f(m)L(m)p_m(m)dm \quad = \quad \int_{\tilde{\Xi}} f\left(g(\xi)\right)L\left(g(\xi)\right)p_m\left(g(\xi)\right)\left|\det(Dg)\right|d\xi$$

$$\mathrm{E}_{\pi_m} f \quad = \quad \mathrm{E}_{\pi_\xi}\left(f \circ \mathbf{g}\right)$$

leads to

– $\pi_\xi(\xi) \propto \tilde{L}_\xi(\xi)\, p_m\left(g(\xi)\right)\left|\det(Dg)\right|$ is a **surrogate posterior**

– ***g*** is an invertible/differentiable map from $\tilde{\Xi} \subseteq \Xi$ to the range of ***m***

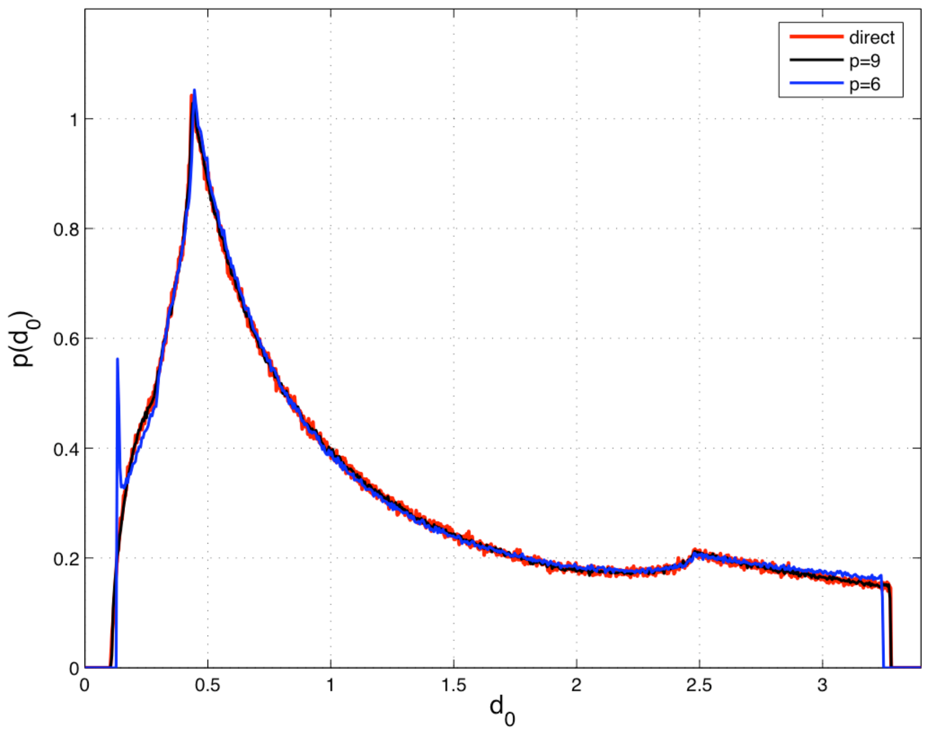
– Choice of ***g***: do not require ***m* = *g(ξ)* ~ prior**

What is the uncertainty that you should propagate through a forward model in order to invert it?

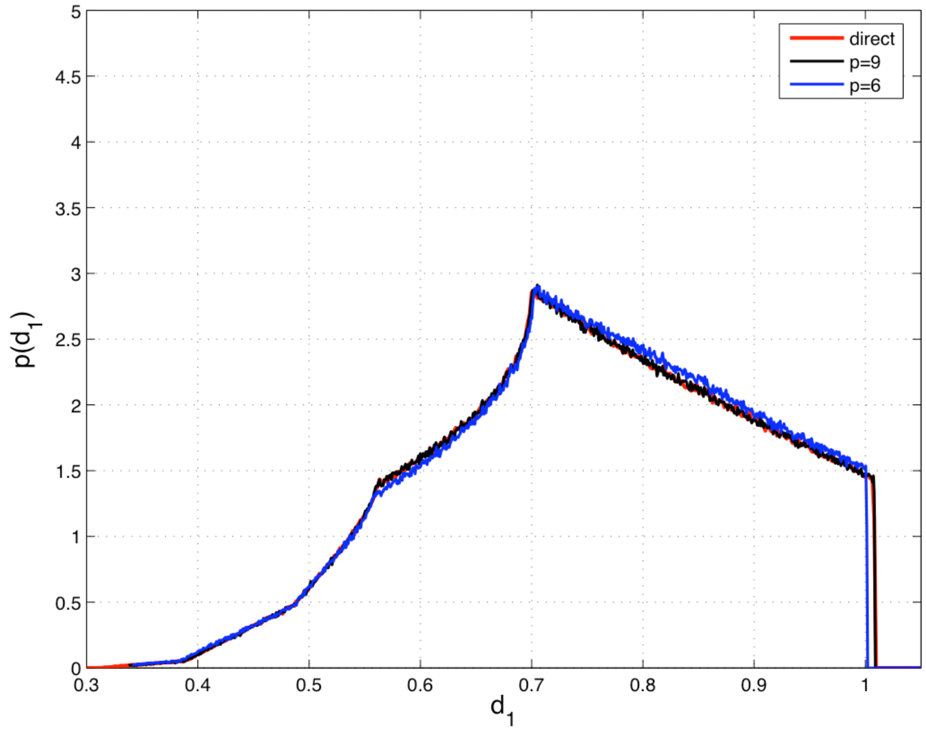
[Marzouk *et al* JCP 2007]

Stochastic spectral solution

- **Return to 2-D source inversion problem—**
  - Propagate uniform prior on source location; examine probability density of model prediction *u(x=0, y=0, t)*:



t = 0.05

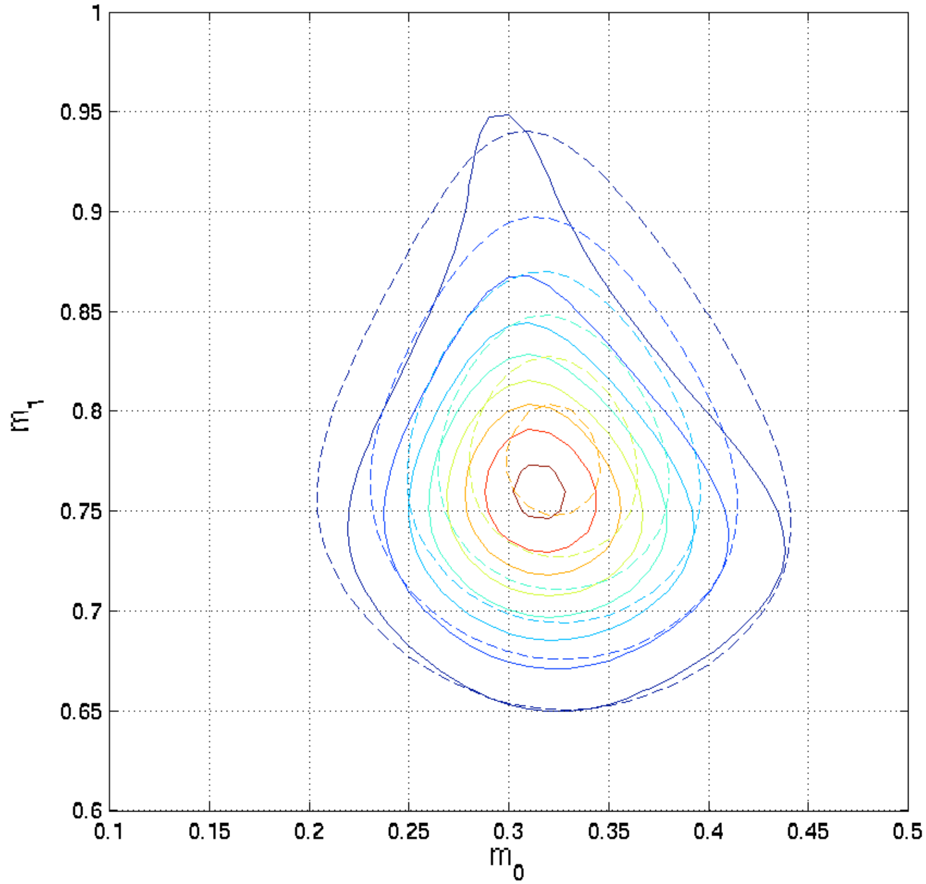


t = 0.15

- Compare times: ***sensor placement, experimental design…***

Posterior density

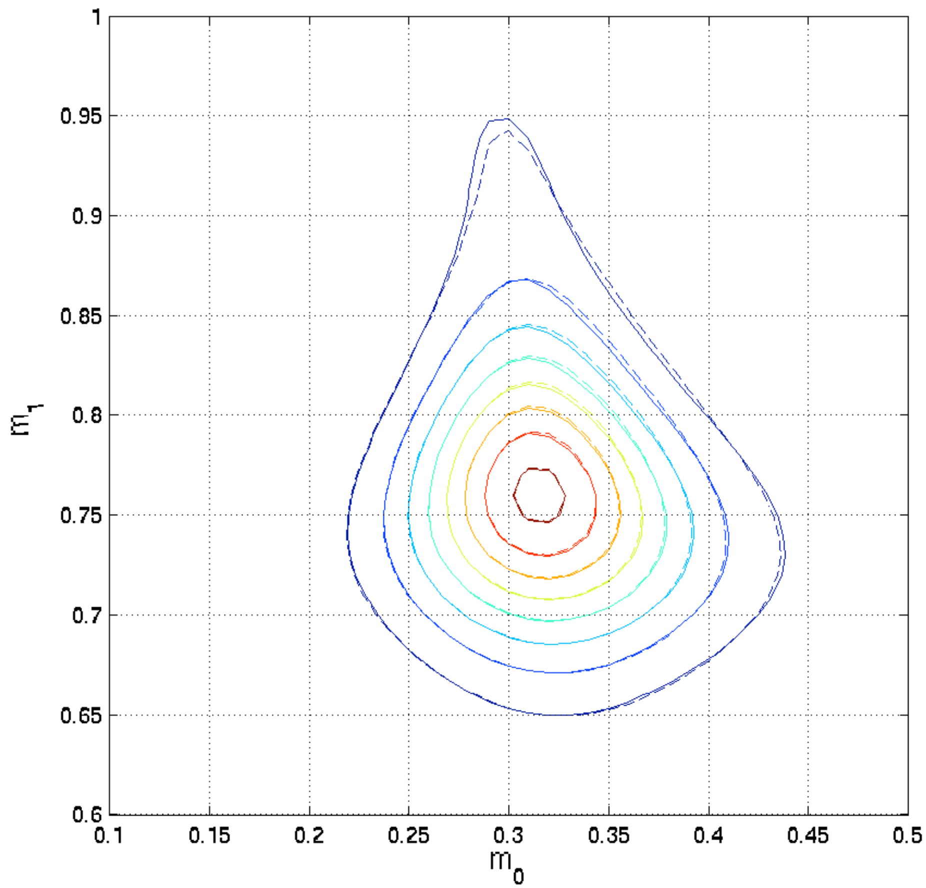
- 3×3 grid of sensors; measure at t = {0.05, 0.15}; ***d*** from noisy observations of a source at (x,y) = (0.25,0.75).



p=3 (dashed) vs direct (solid)

Posterior density

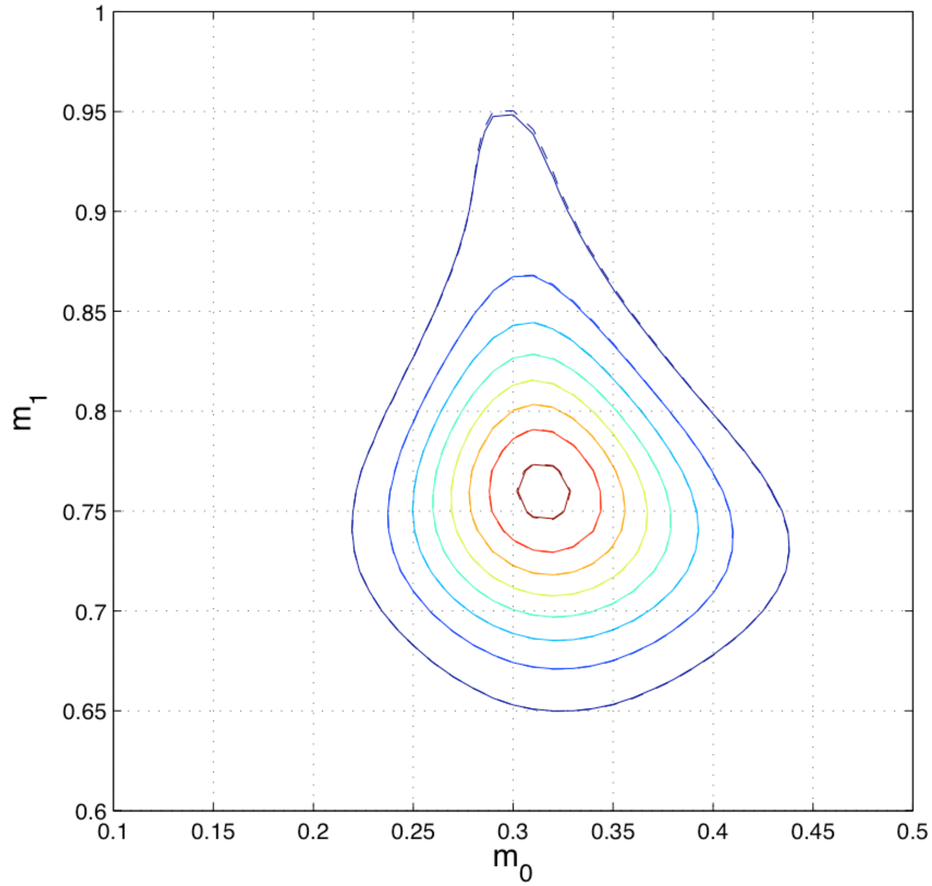
- 3×3 grid of sensors; measure at $t = \{0.05, 0.15\}$; $\boldsymbol{d}$ from noisy observations of a source at (x,y) = (0.25,0.75).



p=6 (dashed) vs direct (solid)

Posterior density

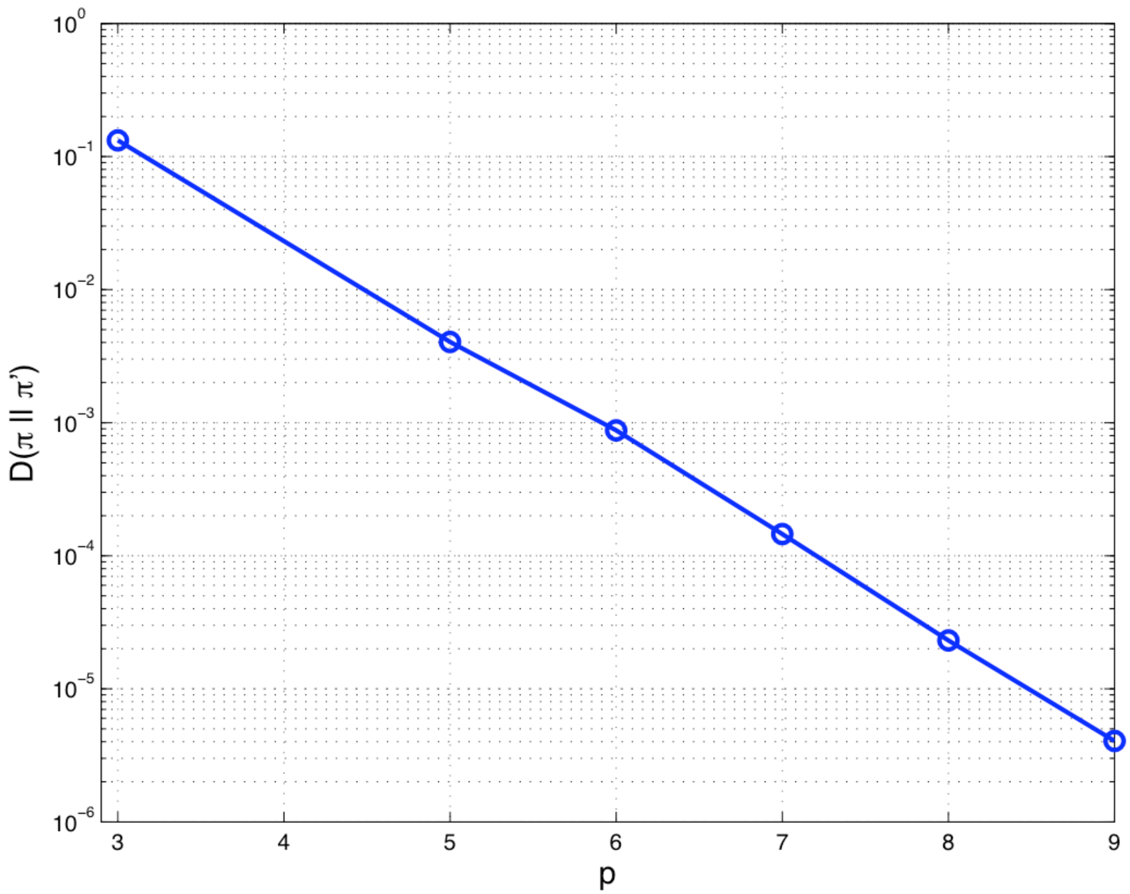
- 3×3 grid of sensors; measure at $t = \{0.05, 0.15\}$; $\boldsymbol{d}$ from noisy observations of a source at (x,y) = (0.25,0.75).



p=9 (dashed) vs direct (solid)

Posterior density

- 3×3 grid of sensors; measure at t = {0.05, 0.15}; ***d*** from noisy observations of a source at (x,y) = (0.25,0.75).



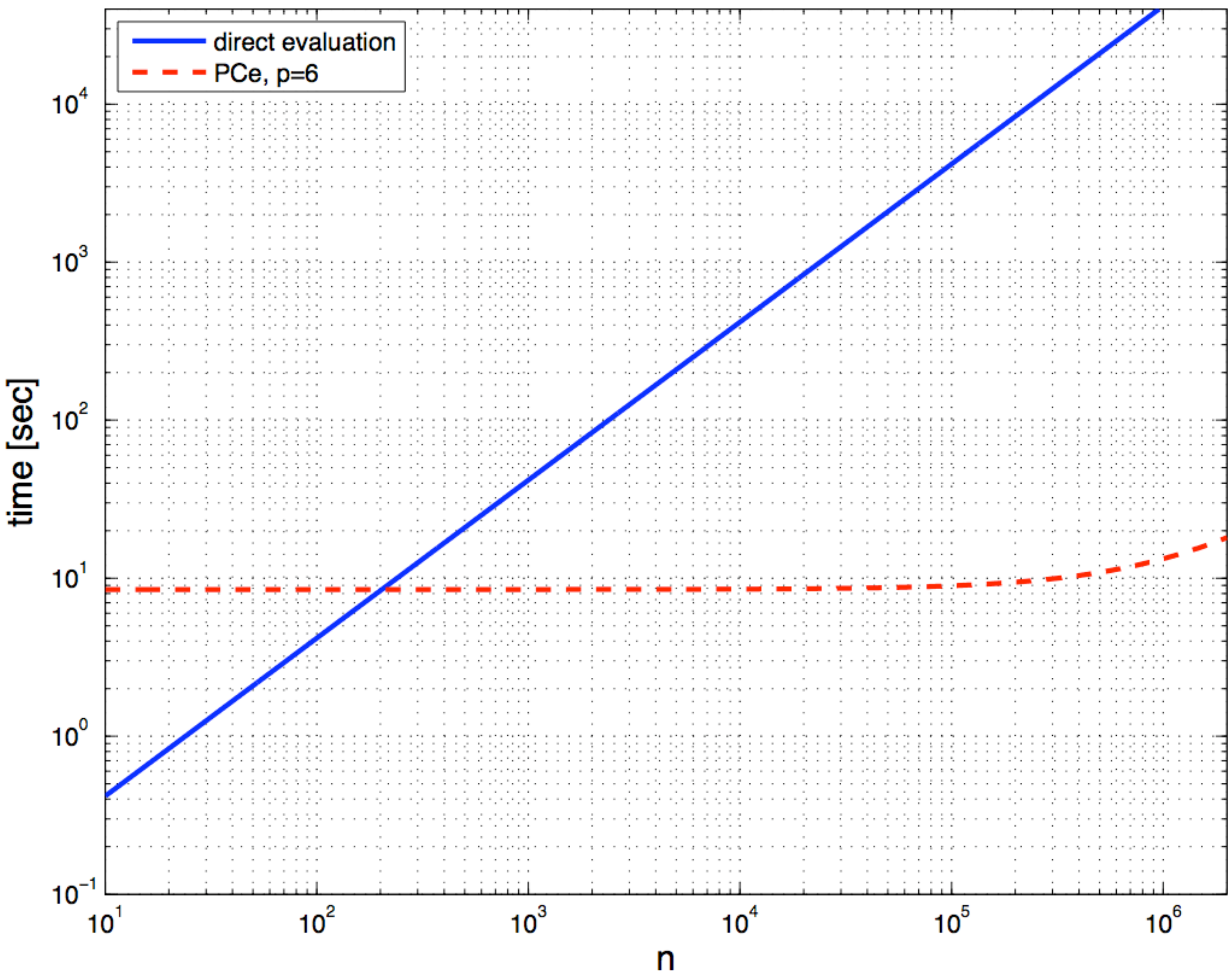
convergence:

Kullback-Leibler distance $D\left(\tilde{\pi}\|\pi\right)$ versus p

$$D\left(\tilde{\pi}\|\pi\right) = \int \tilde{\pi}(m) \log \frac{\tilde{\pi}(m)}{\pi(m)}\, dm$$

Speedup in sampling

- Total computational time *vs* number of samples



Per-sample cost reduced by 3–4 orders of magnitude!!

Outline

1 Bayesian approach to inverse problems

2 Accelerating Bayesian inference via stochastic spectral methods

3 Inference of spatiotemporal fields

- Prior models, regularization, and dimensionality reduction

4 Conclusions

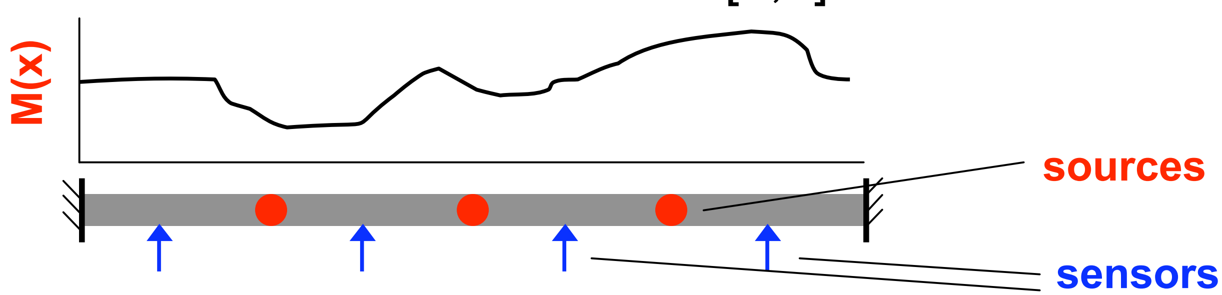
Bayesian IPs for spat-temp fields

$$\boldsymbol{d} = \boldsymbol{G}(M) + \boldsymbol{\eta}$$

- Let *M(**x**)* be a field quantity, $\boldsymbol{x} \in R^d$
- Bayesian approach:
  - *M(**x**,ω)* is a stochastic process
  - We want to explore/evaluate *p(M|**d**)*
- Example:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\boxed{\nu(x)}\frac{\partial u}{\partial x}\right) + \sum_{i=1}^{N} \frac{s_i}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(l_i - x)^2}{2\sigma_i^2}\right)\left[1 - H(t - \tau_i)\right]$$

  - M(x) = log ν(x) is an unknown log-diffusivity; infer from a few sources/sensors in the domain *D* = [0,1]



GP priors and regularization

- Forward problem $\nu(x) \mapsto u(x_i, t_i)$ is nonlinear
- Estimation of $\nu(x)$ from sparse data is *ill-posed*
- Encode additional information on M(x) = log ν(x) in the prior distribution:
  - Gaussian process prior $M \sim GP(\mu, C)$
  - Great flexibility via choice of covariance kernel
  - Here, employ a stationary Gaussian covariance kernel:

$$C(x_1, x_2) = C(x_1 - x_2) = \theta \exp\left(|x_1 - x_2|^2 / 2L^2 \right)$$

  - Magnitude θ of the prior covariance is a hyperparameter:

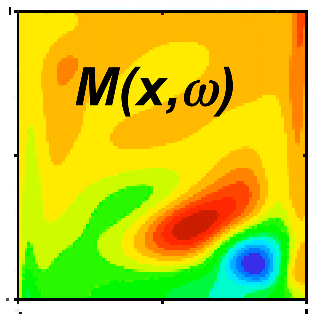
$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-\alpha-1} \exp\left(-\frac{\beta}{\theta} \right)$$

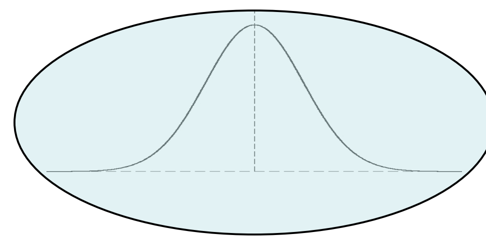
(inverse-Gamma hyperprior)

  - MAP estimate with GP prior equivalent to introducing RKHS-norm regularization penalty with reproducing kernel $K(\cdot,\cdot)$ = covariance $C(\cdot,\cdot)$

Bayes + PCe approach

- Big picture of the computational approach:
  - Endow *M*(***x***,ω) with an appropriate prior and hyperpriors (**Gaussian process** prior + …)
  - PCe to propagate prior uncertainty in *M*(***x***) through the forward model
  - MCMC sampling from the surrogate posterior





G

$$\rightarrow \tilde{G}_j(\xi) = \sum_{k=0}^{P} g_{jk} \Psi_k(\xi)$$

- Representing *M*(***x***):
  - Should we just use $\mathbf{M}_n$ (e.g., pixels)?
  - How many stochastic degrees of freedom in *M*? Minimize dimensionality?
  - What basis functions?

Dimensionality reduction

- Let $M \sim GP(\mu, C)$
- Introduce the **Karhunen-Loève expansion** & truncate:

$$M(\boldsymbol{x},\omega) = \mu(\boldsymbol{x}) + \sum_{i=1}^{K} \sqrt{\lambda_i}\, c_i(\omega)\, \phi_i(\boldsymbol{x})$$

  - λ_i and $\phi_i(\boldsymbol{x})$ are eigenvalues/eigenfunctions of $C(\boldsymbol{x}_1,\boldsymbol{x}_2)$

$$\int_D C(\boldsymbol{x}_1,\boldsymbol{x}_2)\phi_i(\boldsymbol{x}_2)d\boldsymbol{x}_2 = \lambda_i \phi_i(\boldsymbol{x}_1)$$

  - $c_i \sim N(0,1)$
- Transform the inverse problem **(dim = *K* << *n*)**

$$d = \tilde{G}(\boldsymbol{c}) + \eta$$

  - $c_i \sim N(0,1)$ is now the prior on **c**
  - Obtain posterior density $p(\boldsymbol{c}|\boldsymbol{d})$
- With hyperparameter: prior is $c_i \sim N(0,\theta)$

Computational approach

- Explore the full *K+1*-dimensional posterior:

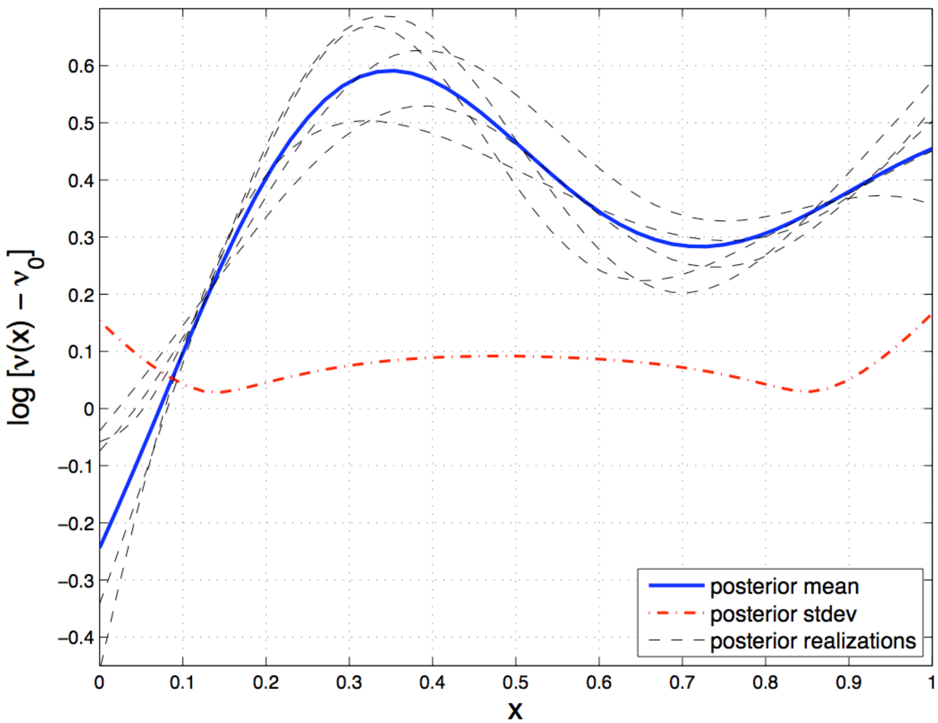
$$p(\boldsymbol{c},\theta|\boldsymbol{d}) \propto p_\eta\left(\boldsymbol{d} - G(M_K(\boldsymbol{c}))\right) \cdot \theta^{-K/2} \prod_{i=1}^{K} \exp\left(-c_i^2/2\theta\right) \cdot \theta^{-1-a} \exp\left(-\beta/\theta\right)$$

- Single-component MCMC scheme:
  - Gibbs update for θ: sample from full conditional $p(\theta|\boldsymbol{c})$
  - Random-walk Metropolis updates for $\boldsymbol{c}$
  - Cholesky factorization $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$ for grid-based case; $\mathbf{m} = \mathbf{L}\mathbf{z}$
- Numerical details
  - Nystrom method to solve integral equation for λ_i and $\phi_i(\boldsymbol{x})$
  - Explicit RKC scheme for deterministic *and* stochastic forward problems

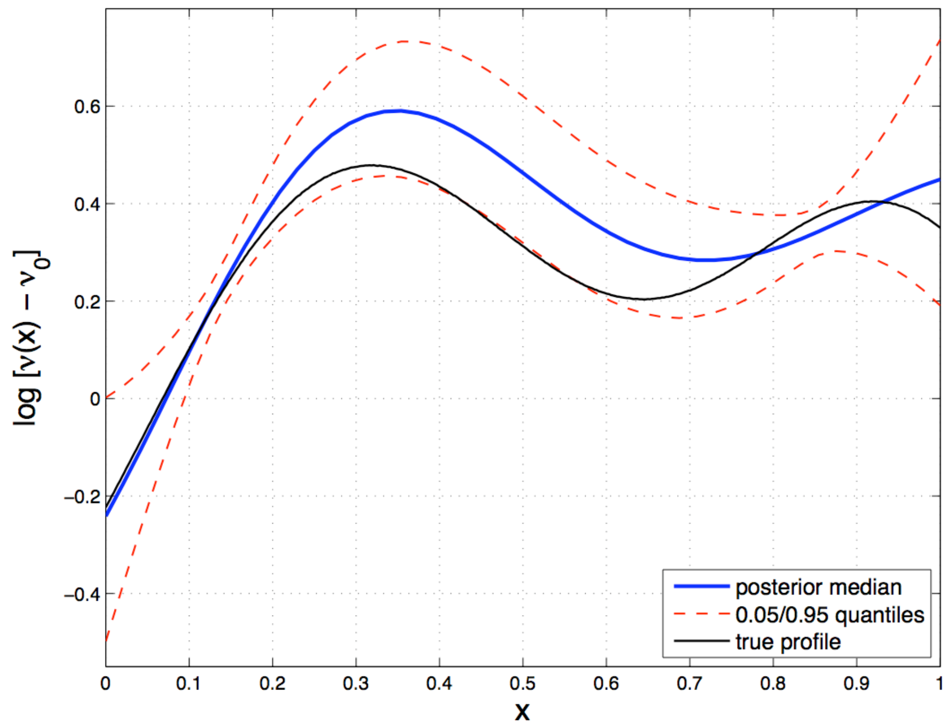
- Solve inverse problem in three ways: *(1)* grid-based, $p(\mathbf{z}_n, \theta\,|\,\boldsymbol{d})$; *(2)* K-L, $p(\boldsymbol{c}_K, \theta\,|\,\boldsymbol{d})$; *(3)* K-L + PCe, $p(\xi_K, \theta\,|\,\boldsymbol{d})$

Bayesian inversion

- Invert for log-diffusivity profile (13 sensors, ≈ 10% noise)



mean, stdev, realizations

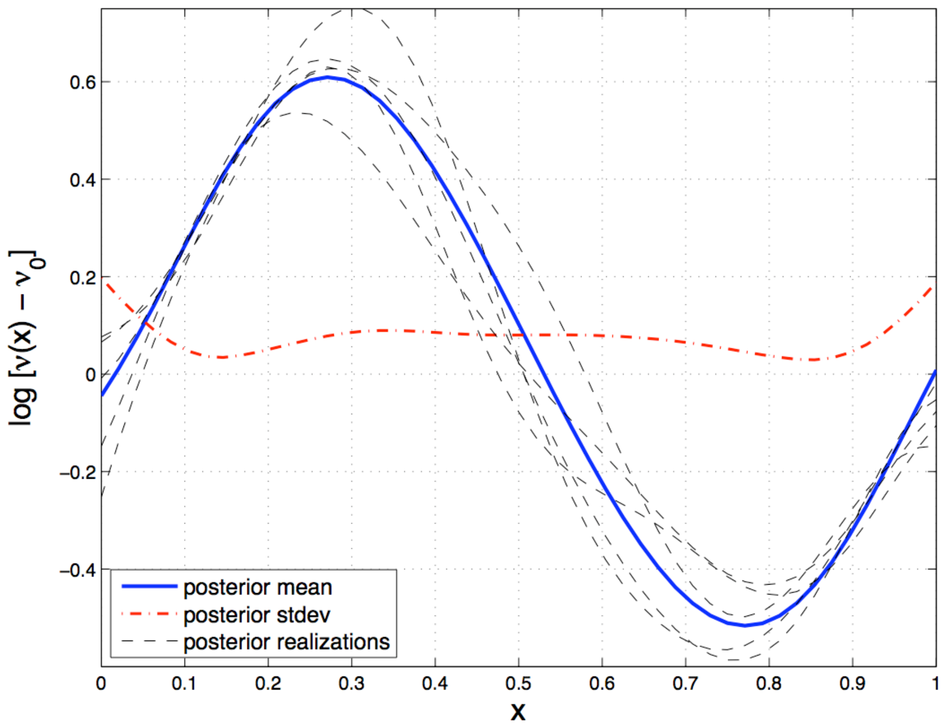


median, quantiles, true profile

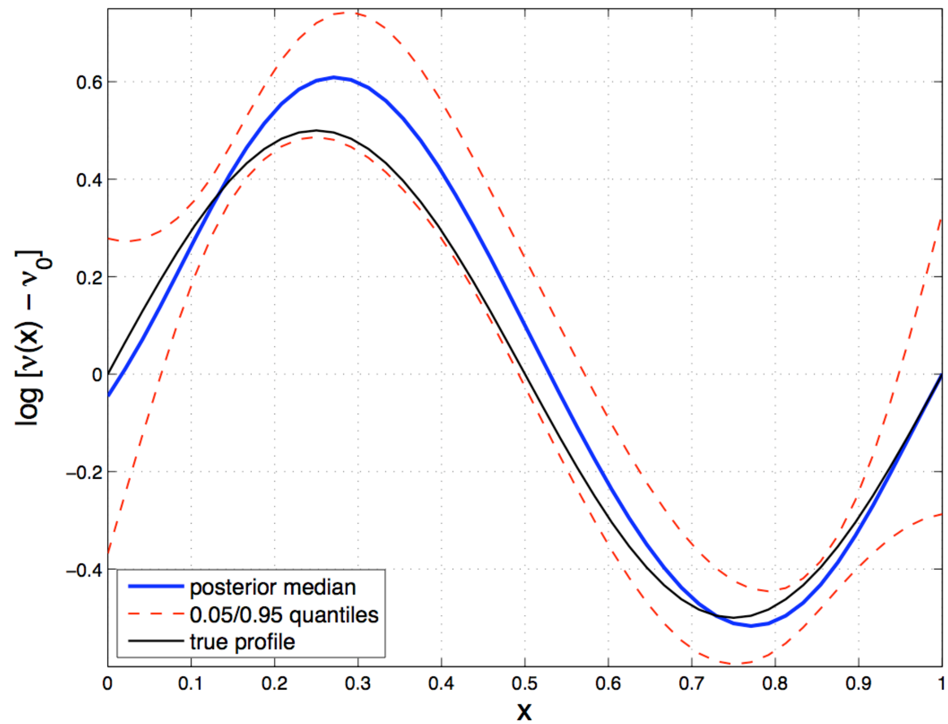
grid-based inversion — target randomly drawn from GP

Bayesian inversion

- Invert for log-diffusivity profile (13 sensors, ≈10% noise)



mean, stdev, realizations

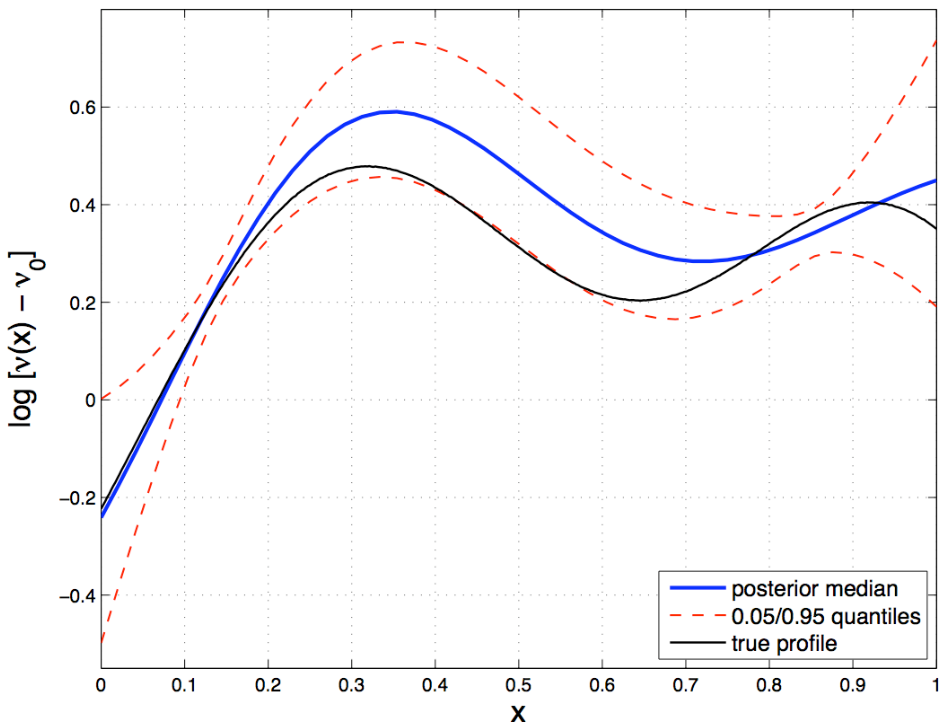


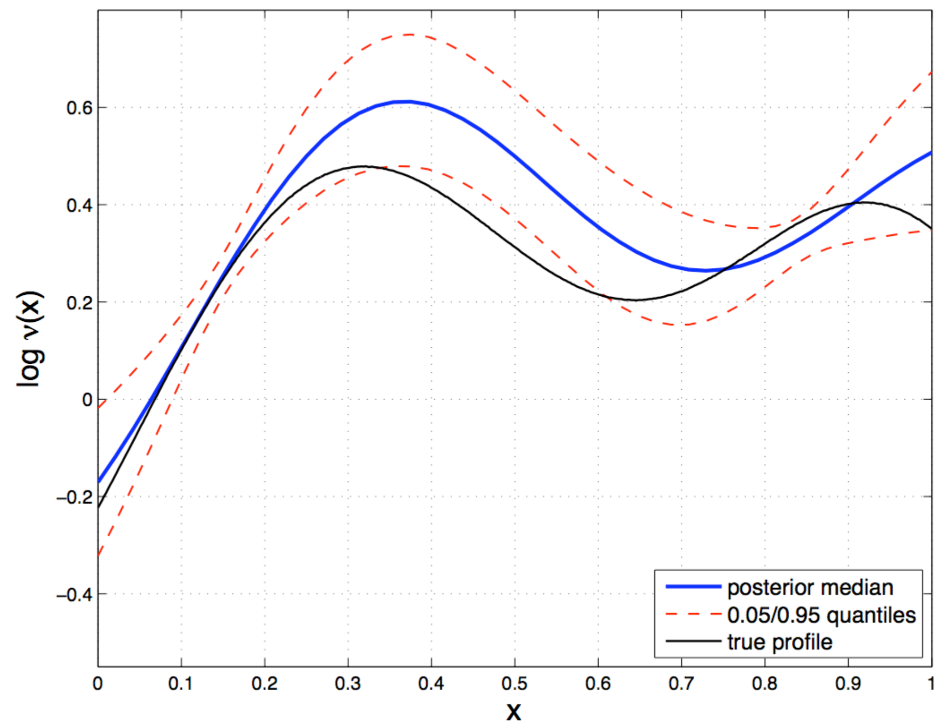
median, quantiles, true profile

grid-based inversion — sinusoidal target

Dimensionality reduction

- Invert for log-diffusivity profile (13 sensors, ≈10% noise):



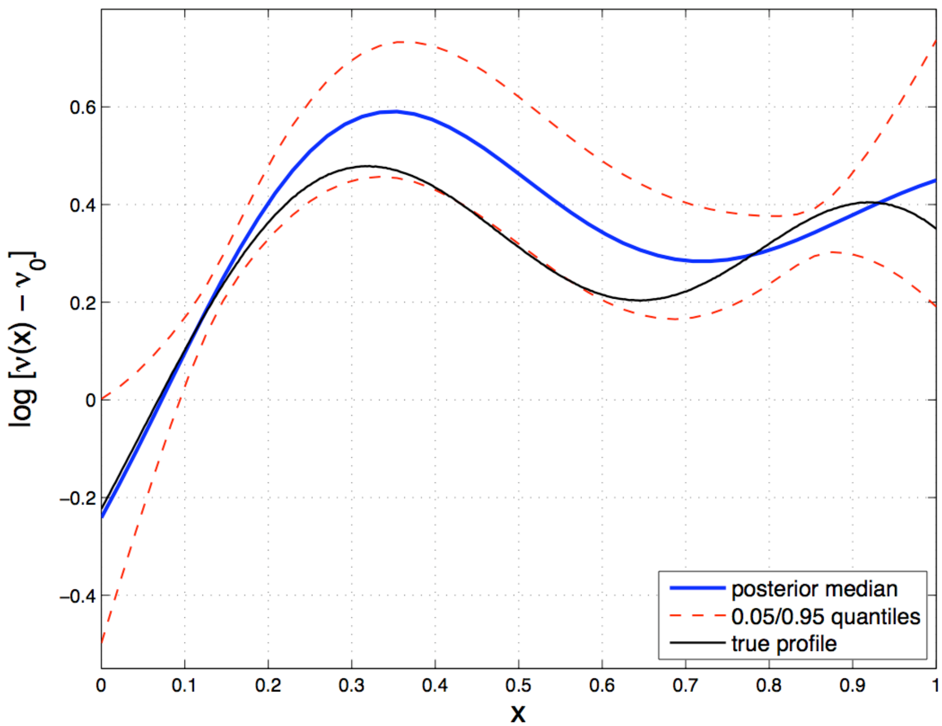


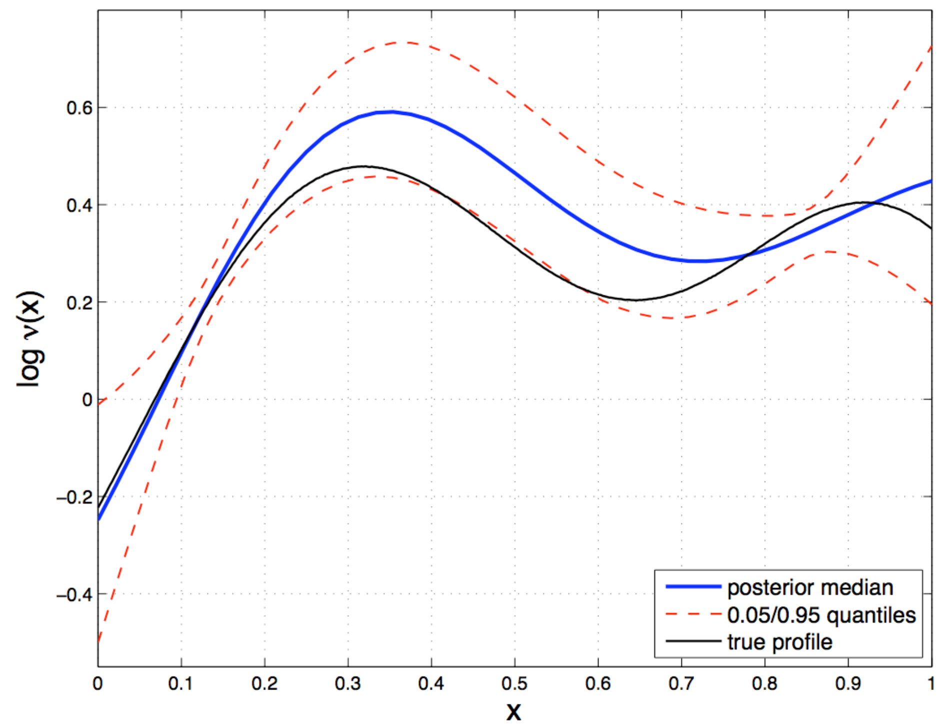
grid-based

4 K-L modes

Dimensionality reduction

- Invert for log-diffusivity profile (13 sensors, ≈10% noise):



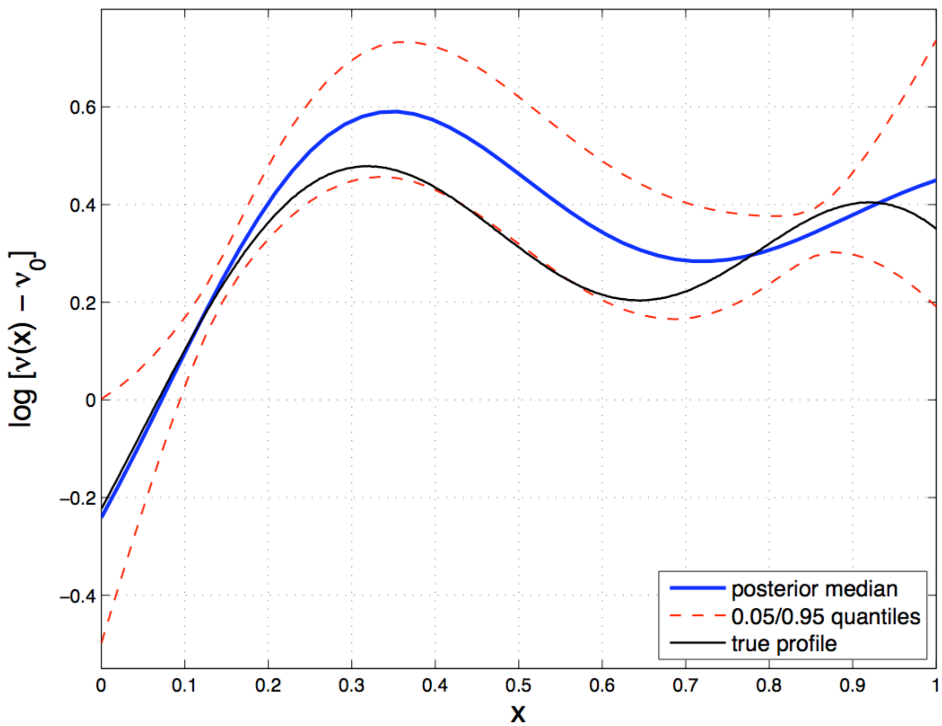


grid-based

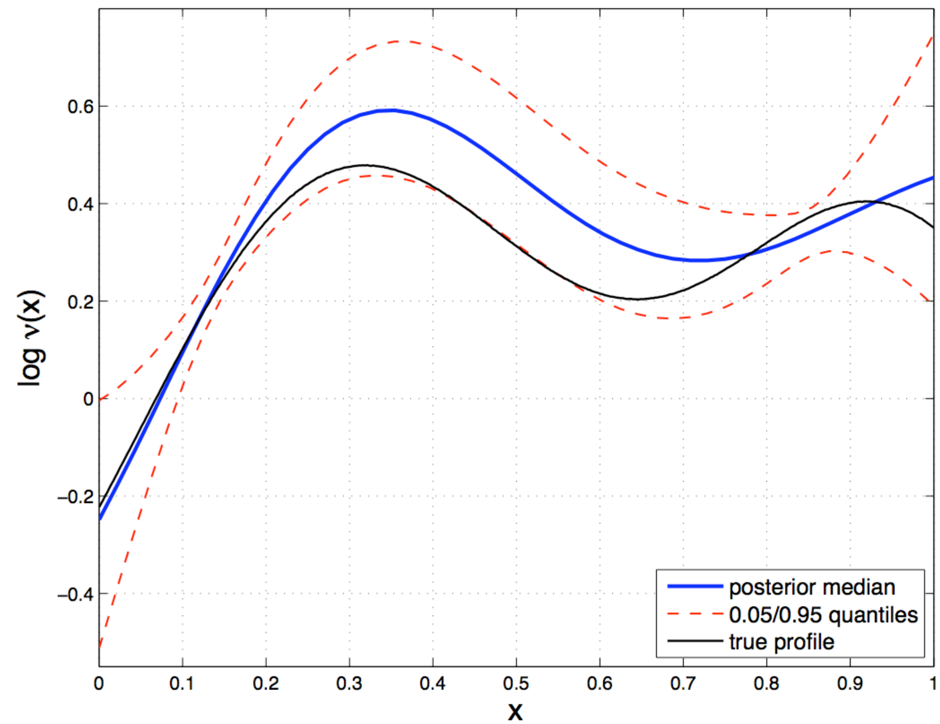
6 K-L modes

Dimensionality reduction

- Invert for log-diffusivity profile (13 sensors, ≈10% noise):



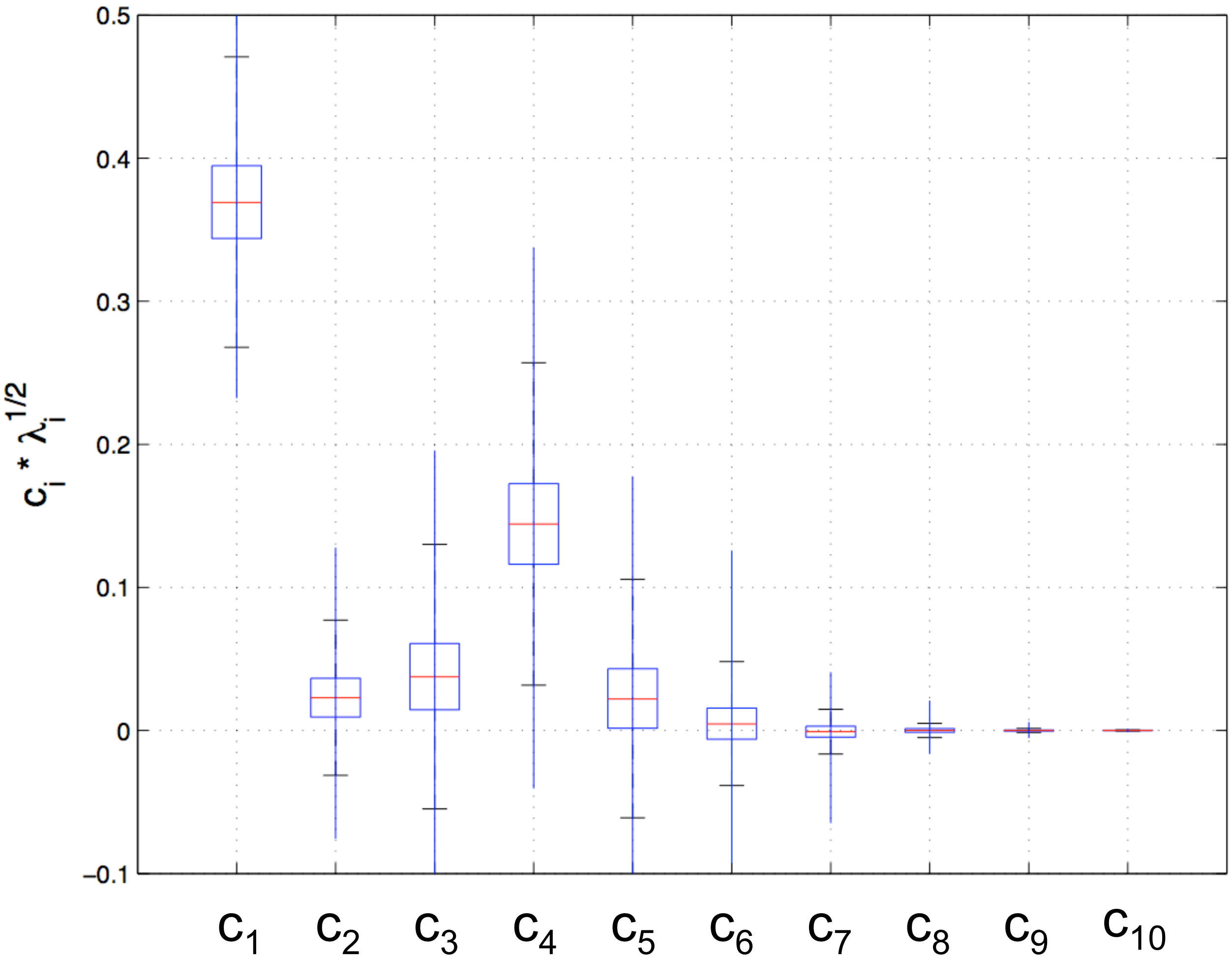
grid-based



8 K-L modes

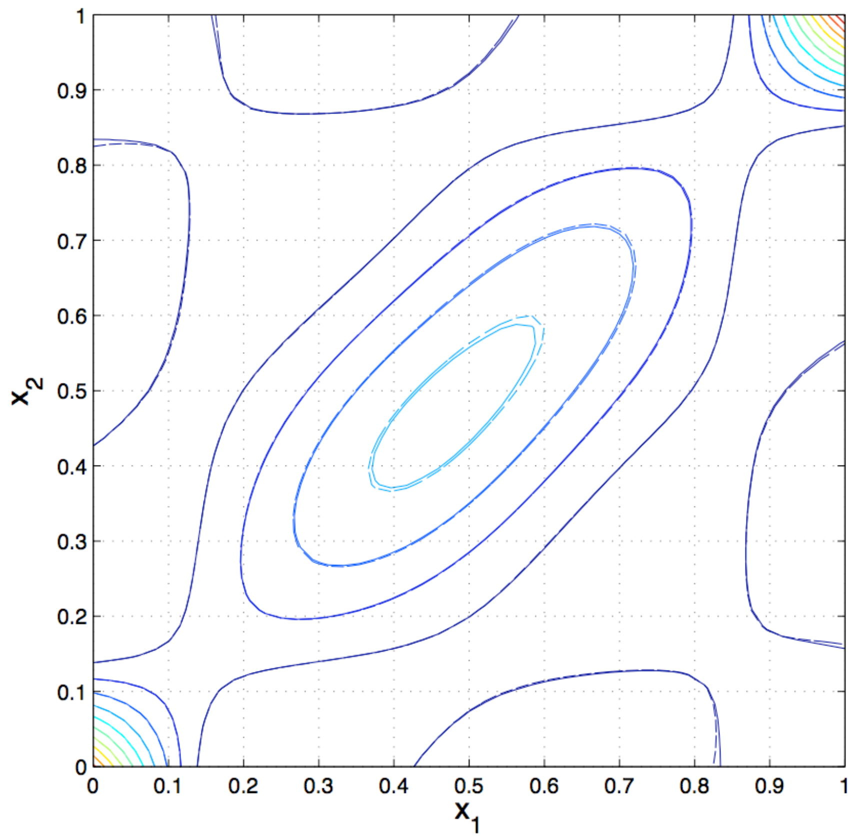
K-L based inversion

- Posterior marginals of the *scaled* KL mode strengths, $p(\lambda^{1/2} c_i \mid d)$:



K-L based inversion

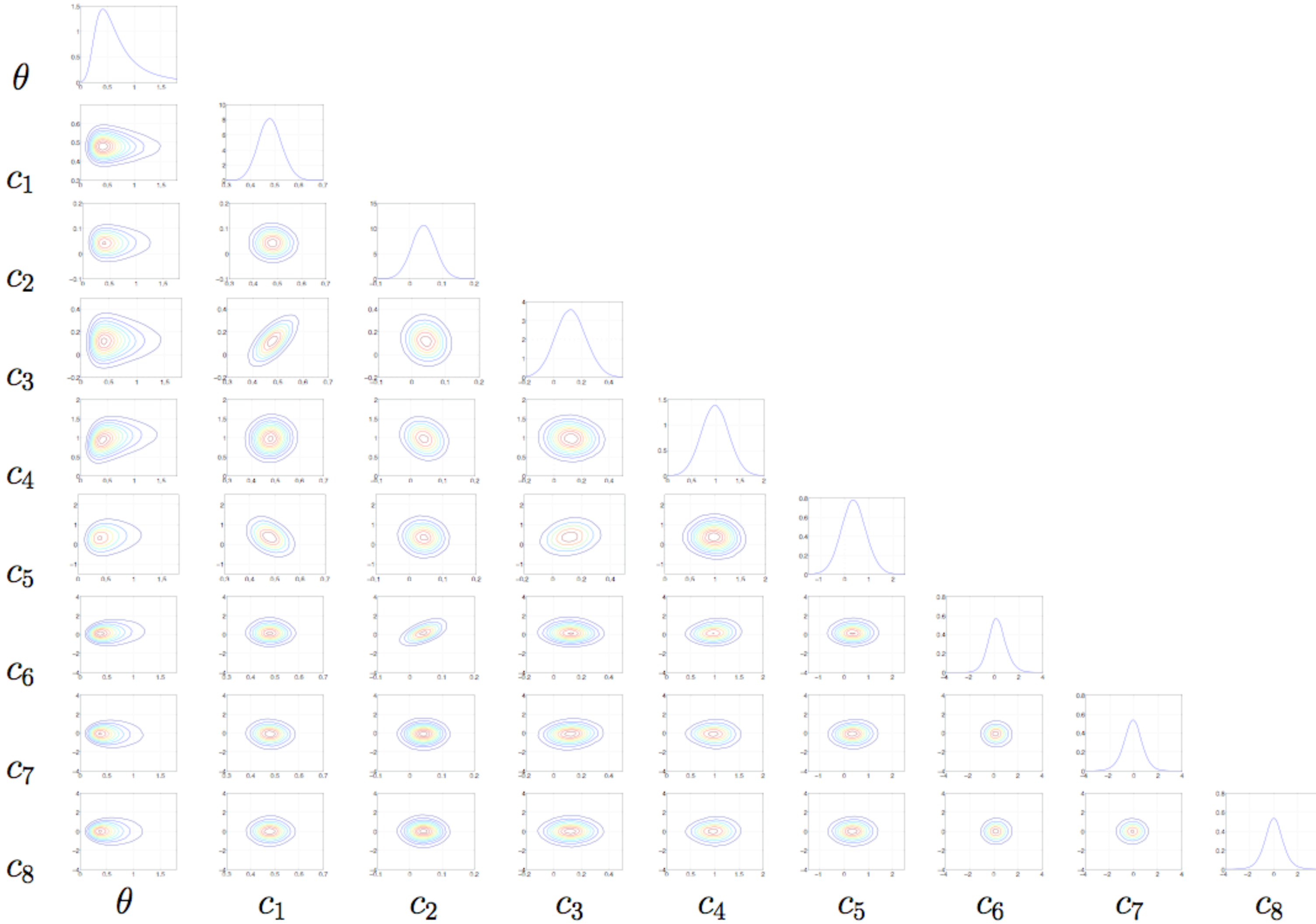
- Invert for log-diffusivity profile (13 sensors, ≈10% noise):



Contours of posterior covariance, $C(x_1,x_2)$, random-draw target
solid = 49 grid points; dashed = 8 K-L modes

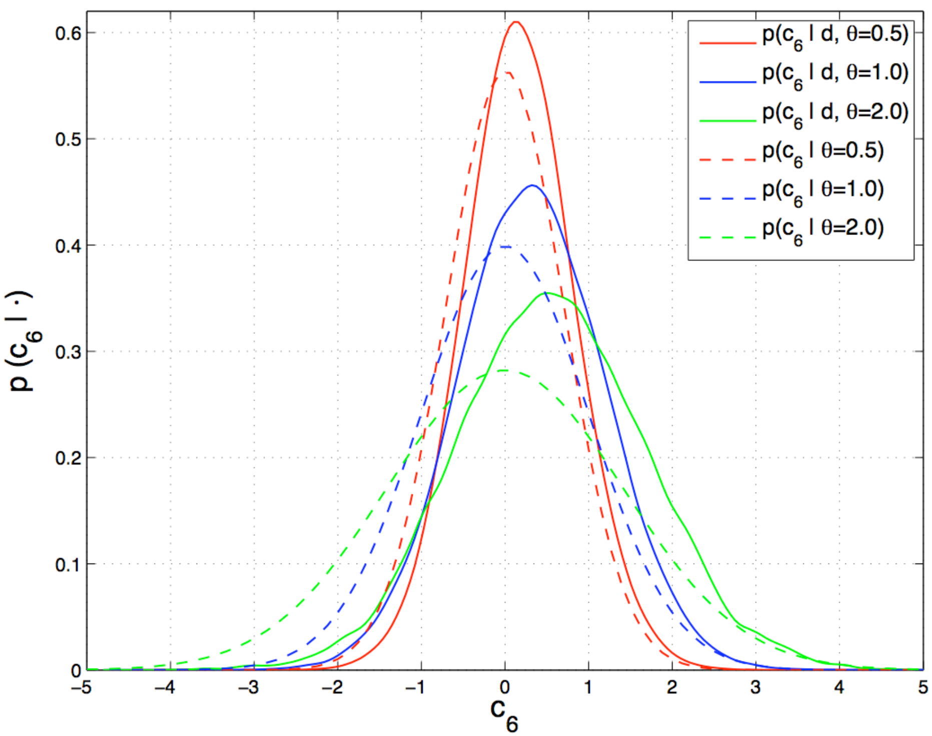
K-L based inversion

- 1-D & 2-D posterior marginals of the KL mode strengths, $p(c_i|d)$:

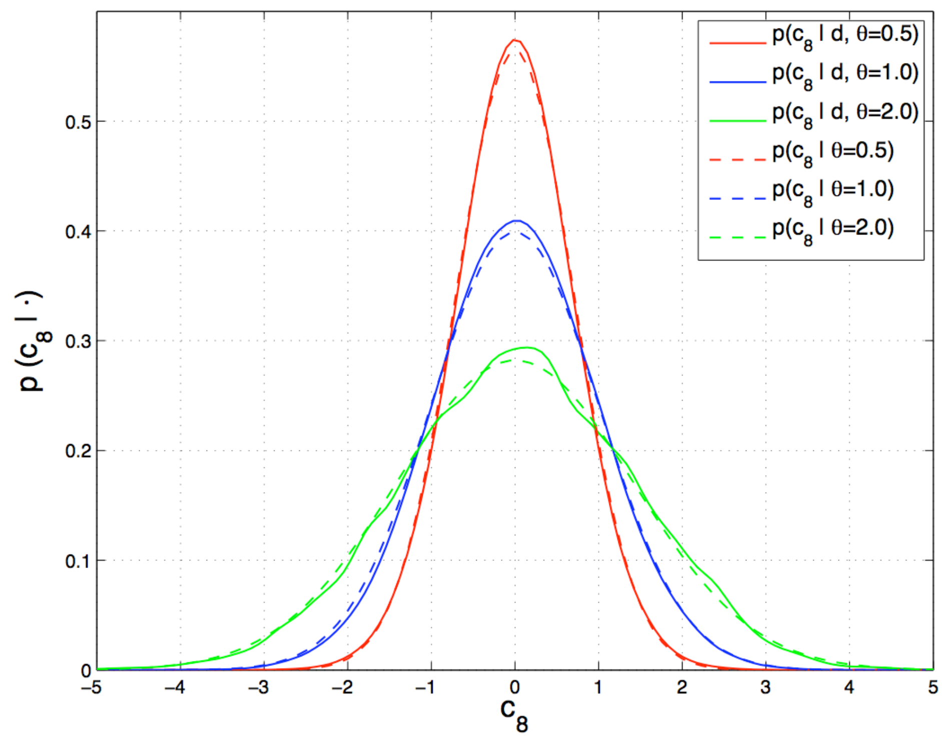


K-L based inversion

- Limiting behavior of higher-order modes



c_6



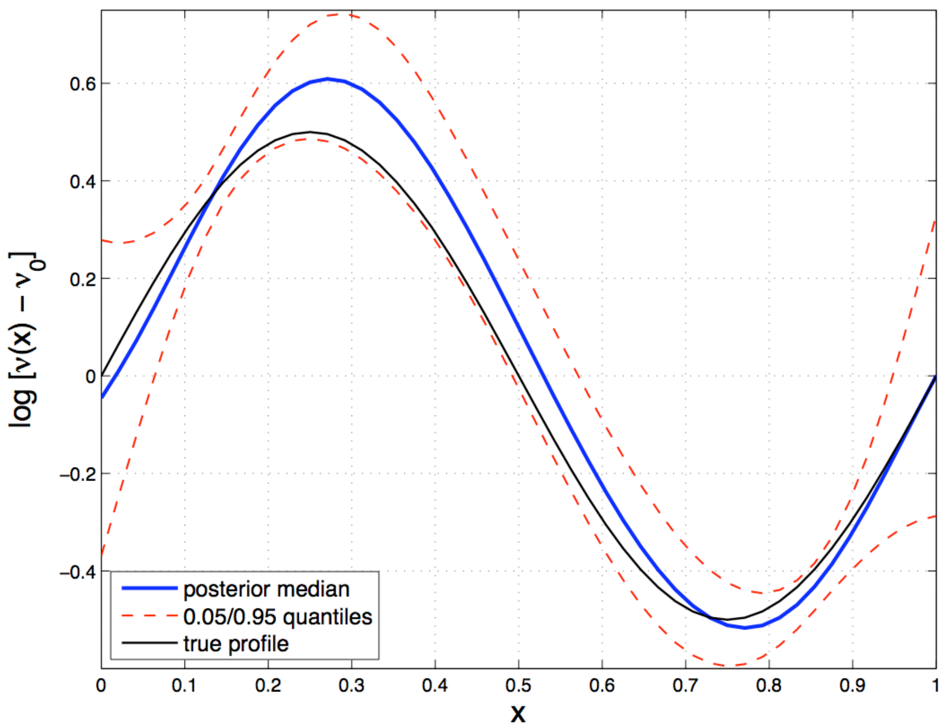
c_8

$$p\left(c_m, c_{m+1}, \ldots \middle| \boldsymbol{d}, \theta\right) \rightarrow \prod_{i \geq m} p\left(c_i \middle| \theta\right)$$

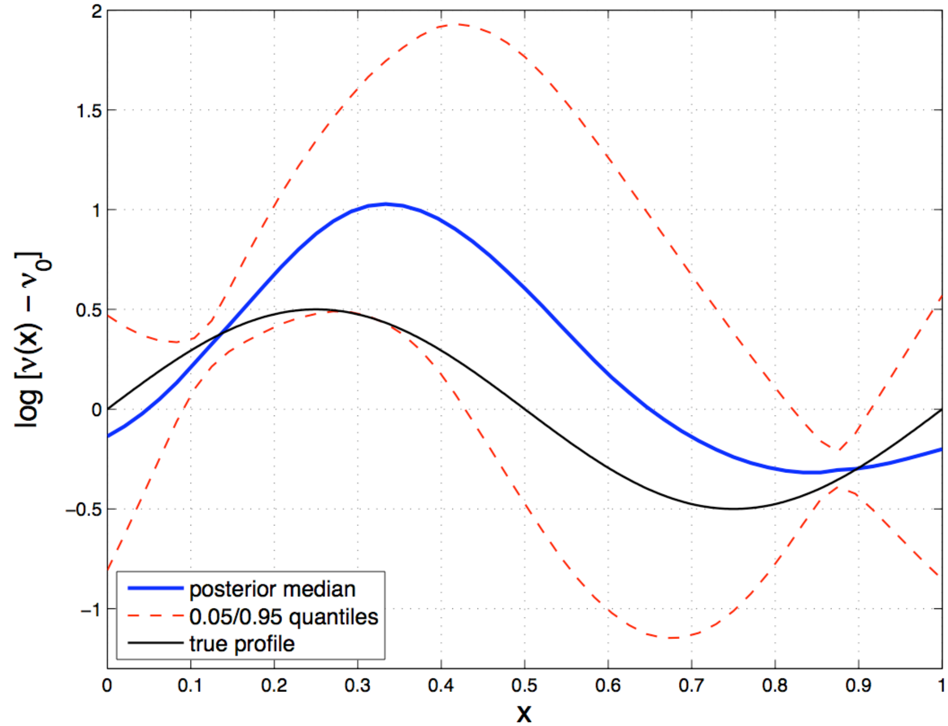
K-L based inversion

- **Coarsen the data** (same msmnt times and noise level):

posterior for sinusoidal target



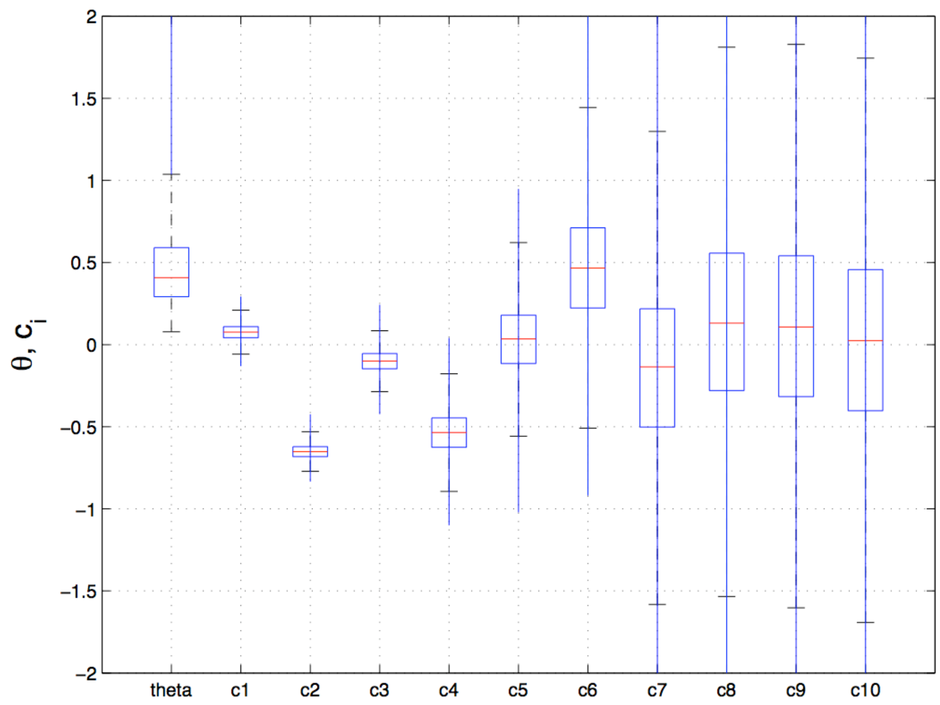
13 sensors



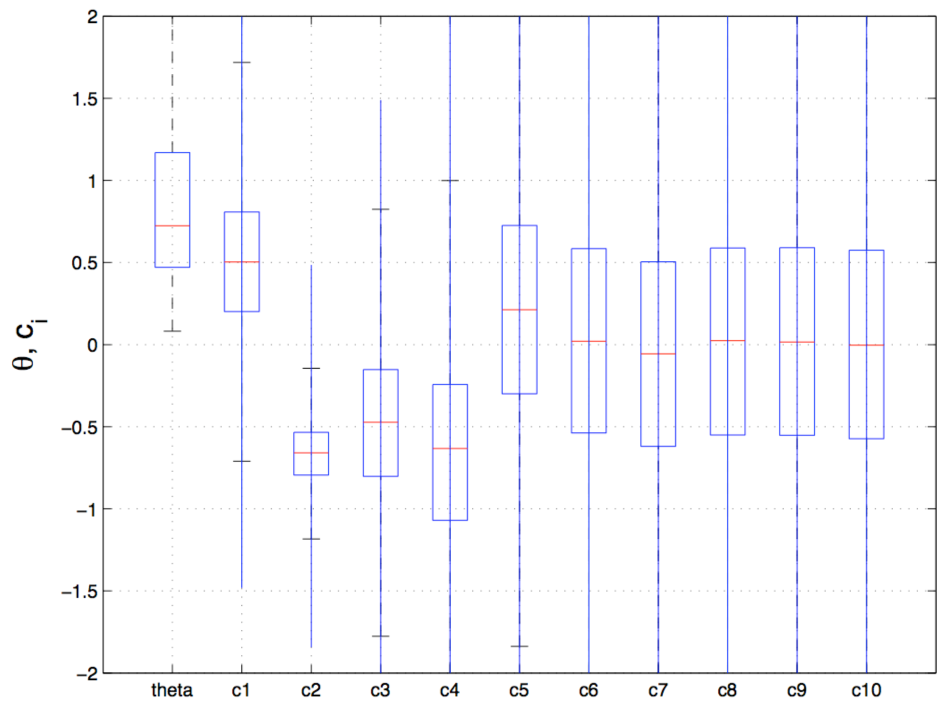
2 sensors—
only at the boundaries

K-L based inversion

- Posterior marginals of the KL mode strengths, $p(c_i|d)$, and $p(\theta|d)$:



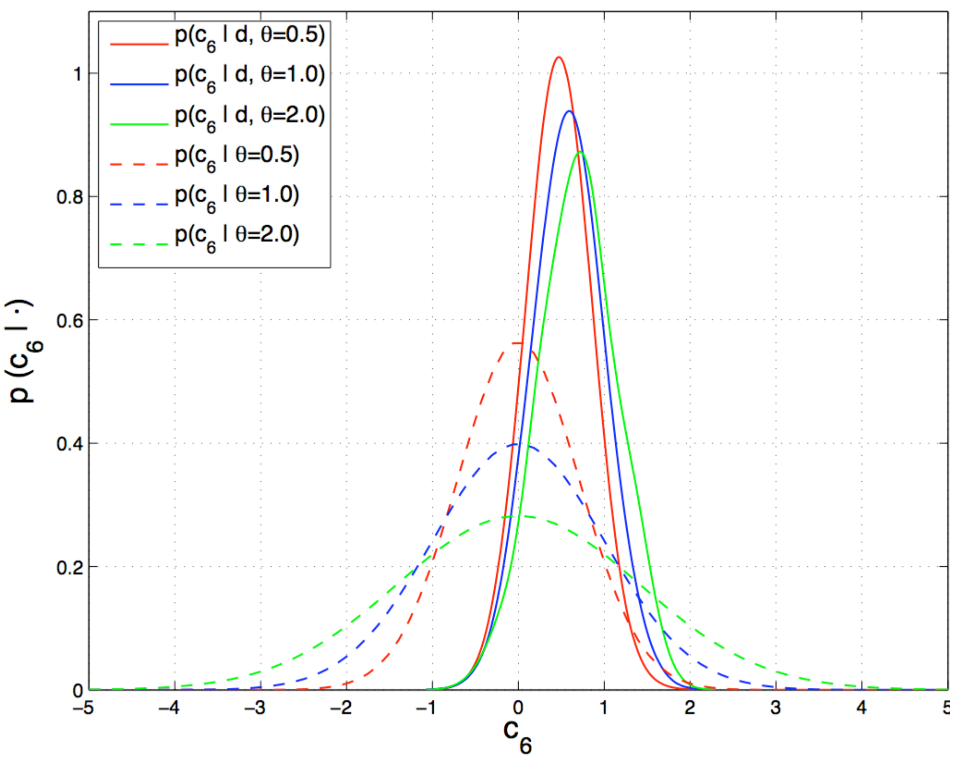
13 sensors



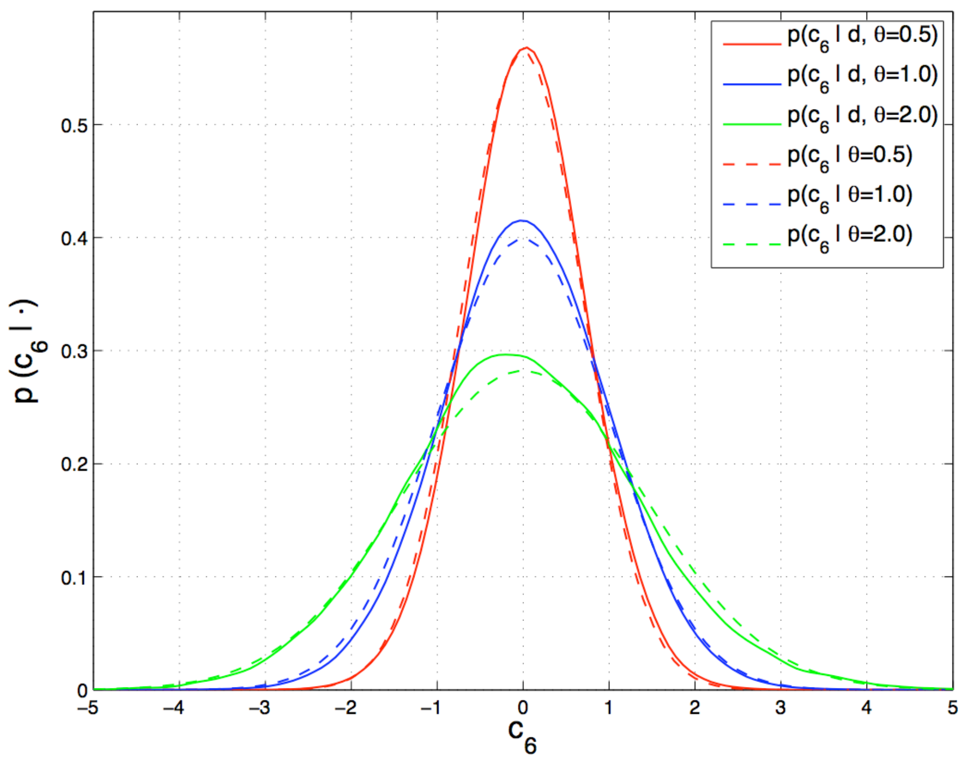
2 sensors

K-L based inversion

- Posterior distributions $p(c_i|d,\theta)$ with coarser data:



c_6, 13 sensors



c_6, 2 sensors

$$p\left(c_m, c_{m+1}, \ldots \middle| \boldsymbol{d}, \theta\right) \rightarrow \prod_{i \geq m} p\left(c_i \middle| \theta\right)$$ at smaller *m*

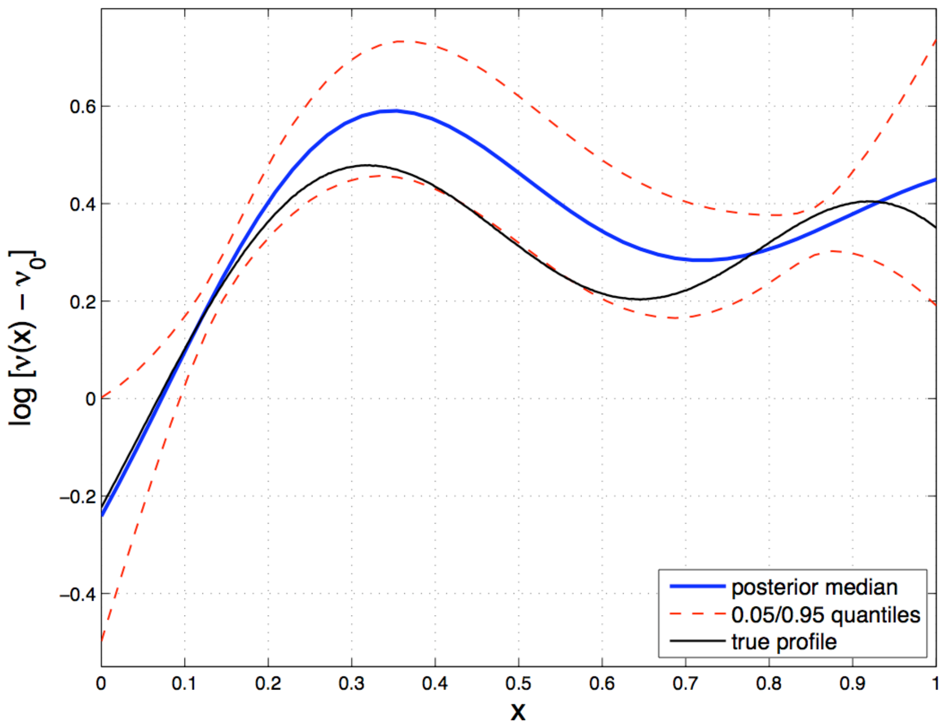
PCe + KL-based inversion

- **Further step—**accelerate with PCe:
  $\rightarrow$ put $c_i = \varpi\xi_i$, $i = 1\ldots K$, and solve the stochastic forward problem **once**
- MCMC sampling from the **surrogate posterior:**

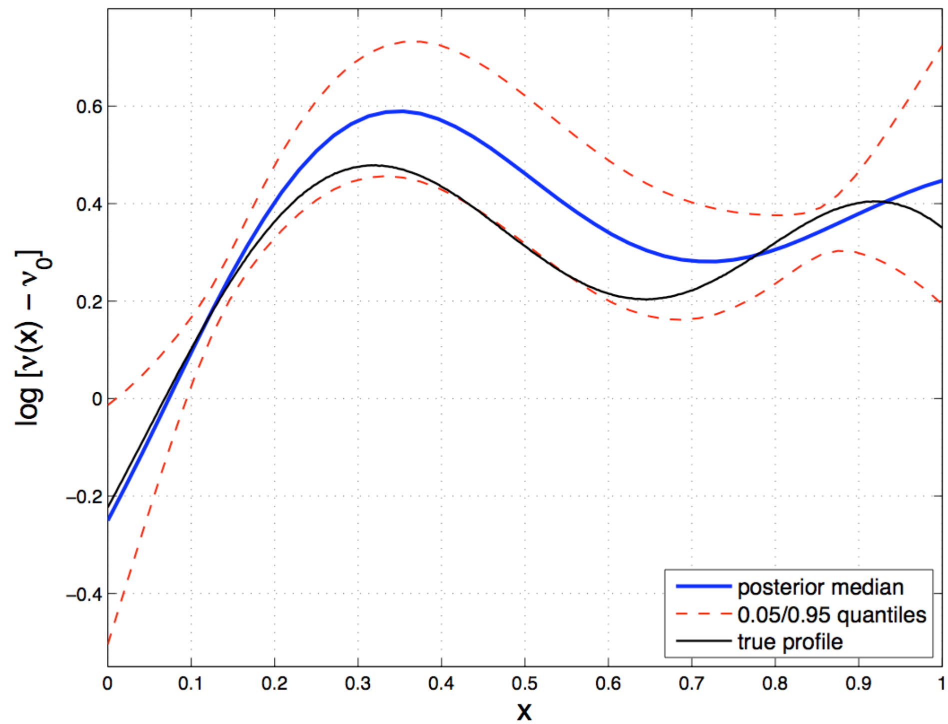
$$p(\boldsymbol{\xi},\theta|\boldsymbol{d}) \propto p_\eta\left(\boldsymbol{d} - \tilde{G}(\boldsymbol{\xi})\right) \cdot \left(\frac{\theta}{\varpi^2}\right)^{-K/2} \prod_{i=1}^{K} \exp\left(\frac{\xi_i^2}{2\theta/\varpi^2}\right) \cdot \theta^{-1-a} \exp(-\beta/\theta)$$

PCe + KL-based inversion

- **Further step—**accelerate with PCe:
  → put $c_i = \varpi\xi_i$, $i = 1\ldots K$, and solve the stochastic forward problem **once**
- MCMC sampling from the **surrogate posterior:**

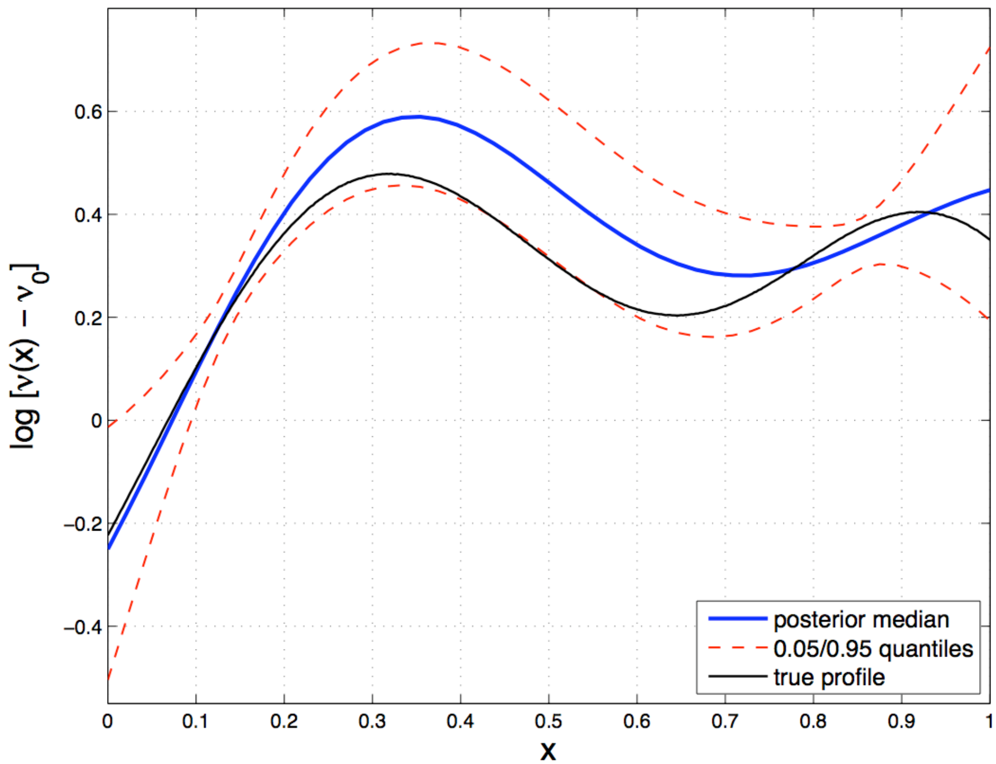


grid-based, direct forward evals



6 K-L modes, p=4

PCe + KL-based inversion



6 KL modes, 4th-order PCe

- Computational times for this example:
  (200K posterior samples)
  - solution on grid, direct MCMC sampling: 45885 s
  - 6 KL modes, direct MCMC sampling: 6418 s
  - 6 KL modes + polynomial chaos: 248 s

Conclusions

- Stochastic spectral methods for Bayesian inference
  - Efficient propagation of prior uncertainty through the forward model
  - Change of variables; *rapid sampling* of the surrogate posterior
  - Inference of spatial fields, with dimensionality reduction (basis from GP prior)
  - Enabling Bayesian inference with ***realistic physical models***
- Extensions and ongoing work
  - Larger-scale problems, complex dynamical systems— depends on advances in spectral UQ:
    - Parallelism (partitioning prior support)
    - Multi-wavelet approaches
    - Adaptive *sparse truncation* of PC bases
    - Sparse quadrature for high-dimensional problems (non-intrusive)
  - Incorporating flexible high-performance simulation frameworks (e.g., Nihilo/Sundance, with B. van Bloemen Waanders)