Optimal experimental design for Bayesian inverse problems governed by PDE models with uncertainty

with application to subsurface flow and tsunami equations

Karina Koval

Partially funded by the NSF via grant RTG/DMS – 1646339.

Motivation

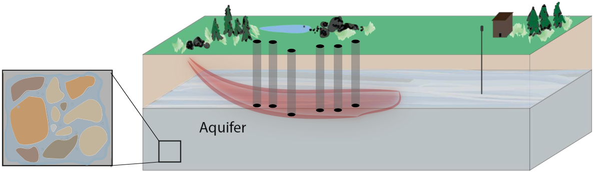
$$\boxed{\boldsymbol{d} = \mathcal{F}(m) + \boldsymbol{\eta}}$$

- Models for real-world phenomena involve unknown parameters, m
- Accurate estimation of parameters relies on informative data, $\boldsymbol{d}$ (*inverse problem*)
- Data collection limited due to cost or physical constraints
- **Optimal experimental design (OED):**
  Design of experimental conditions for parameter inference problems governed by PDE – where and what to measure/observe?

Contaminants in groundwater

Inverse problem:

Given concentration readings, infer source of contamination

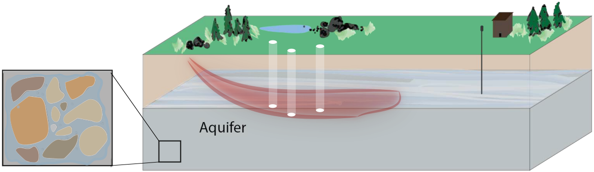


Graphic from sciencefriday.com

Contaminants in groundwater

Inverse problem:
Given concentration readings, infer source of contamination

OED problem:
Where to drill wells to *optimally* infer initial source in event of contamination

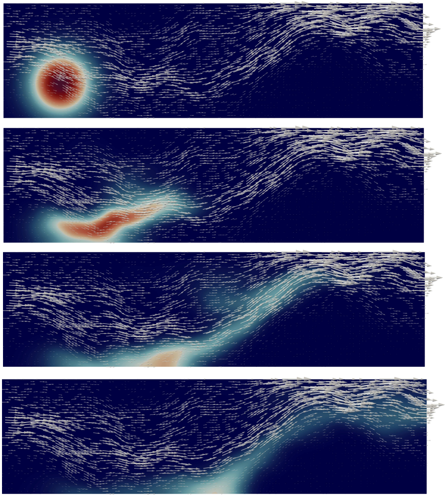


Graphic from sciencefriday.com

Contaminants in groundwater - governing model

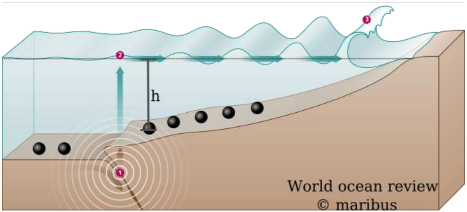
$$
\begin{aligned}
u_t - \kappa \Delta u + \boldsymbol{v} \cdot \nabla u &= 0 \\
u(\cdot, T_i) &= m \\
&+ BCs
\end{aligned}
$$

- $u(\boldsymbol{x}, t)$: concentration
- $\kappa > 0$: diffusion coefficient
- $\boldsymbol{v}(\boldsymbol{x}, t)$: advection velocity field
- m: unknown initial concentration



Earthquake-generated tsunamis

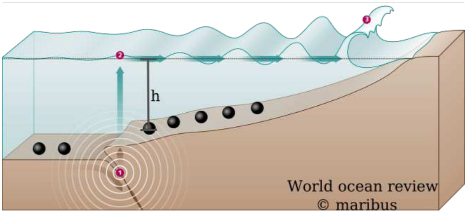
- Tsunamis generated by earthquakes beneath ocean floor
- Earthquake $\rightsquigarrow$ ocean floor deformation $\rightsquigarrow$ tsunami waves
- Tsunami warning relies on knowledge of bathymetry change
- Cannot measure this, but can measure water depth



Earthquake-generated tsunamis

- Tsunamis generated by earthquakes beneath ocean floor
- Earthquake $\rightsquigarrow$ ocean floor deformation $\rightsquigarrow$ tsunami waves
- Tsunami warning relies on knowledge of bathymetry change
- Cannot measure this, but can measure water depth

Inverse problem: Given water level measurements, reconstruct ocean floor deformation

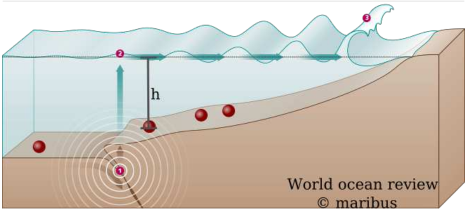


Earthquake-generated tsunamis

- Tsunamis generated by earthquakes beneath ocean floor
- Earthquake $\rightsquigarrow$ ocean floor deformation $\rightsquigarrow$ tsunami waves
- Tsunami warning relies on knowledge of bathymetry change
- Cannot measure this, but can measure water depth

Inverse problem: Given water level measurements, reconstruct ocean floor deformation

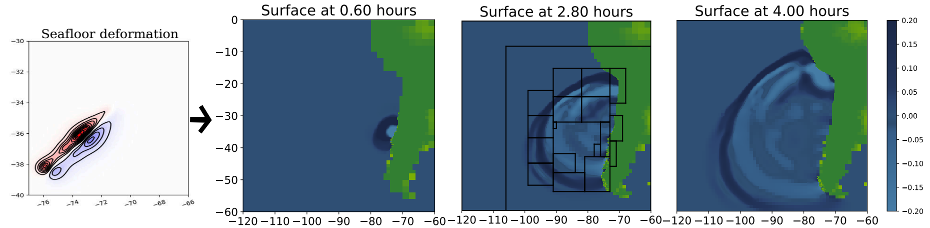
OED problem: Place sensors for optimal tsunami source reconstruction and accurate tsunami forecasting



Earthquake-generated tsunamis - governing equations

$$\begin{bmatrix} h \\ u \\ v \end{bmatrix}_t + \begin{bmatrix} u \\ \frac{u^2}{h} + \frac{1}{2}gh^2 \\ \frac{uv}{h} \end{bmatrix}_x + \begin{bmatrix} v \\ \frac{uv}{h} \\ \frac{v^2}{h} + \frac{1}{2}gh^2 \end{bmatrix}_y = - \begin{bmatrix} 0 \\ ghB_x \\ ghB_y \end{bmatrix}$$

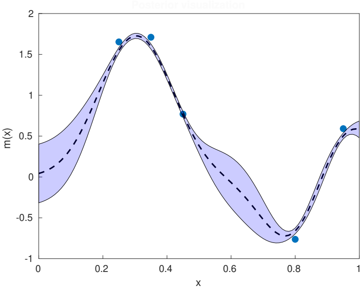
- $h(x, y, t)$: water depth
- $u(x, y, t)$ and $v(x, y, t)$: fluid momentum
- $B(x, y)$: bathymetry
- Models wave propagation due to bathymetry change



(Simulations performed in GeoClaw)

Gaussian process regression

Measure $m(x_i)$ directly at points x_i



$$d_i = m(x_i) + \eta_i$$

- **Goal:** Determine distribution for $m(x)$ given noisy data $\boldsymbol{d} \in \mathbb{R}^s$

- **Gaussian process**: probabilistic approach to regression problems
- Uses (noisy) data $\boldsymbol{d}$ to update prior knowledge about m
  $\rightsquigarrow$ posterior distribution

Gaussian regression review

$$\boxed{d_i = m(x_i) + \eta_i}$$

Goal: determine $\boldsymbol{m}^* = [m(x_1^*), \ldots, m(x_n^*)]$ for $x_i^* \in D^*$

Given: data $\boldsymbol{d} = [d_1, \ldots, d_s]$ at $x_1, \ldots, x_s$ for $x_i \in D_{\boldsymbol{d}}$

Gaussian regression review

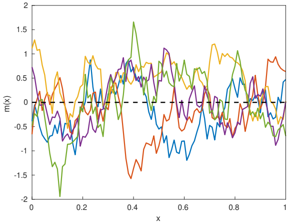
$$d_i = m(x_i) + \eta_i$$

Goal: determine $\boldsymbol{m}^* = [m(x_1^*), \ldots, m(x_n^*)]$ for $x_i^* \in D^*$

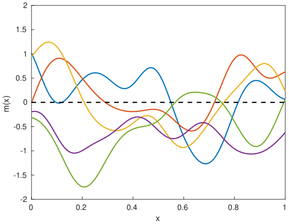
Given: data $\boldsymbol{d} = [d_1, \ldots, d_s]$ at $x_1, \ldots, x_s$ for $x_i \in D_{\boldsymbol{d}}$

- Assume $m \sim \mathcal{N}(0, \mathcal{C}_{\text{pr}})$ (*prior*) and $\eta_i \sim \mathcal{N}(0, \sigma_n^2)$
- $\mathcal{C}_{\text{pr}}$ defined through *covariance function* $c(x, y)$

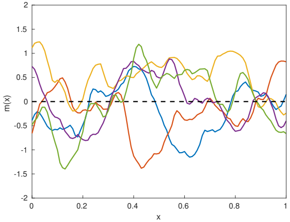
Absolute exponential



Squared exponential



General



Gaussian regression review

$$\boxed{d_i = m(x_i) + \eta_i}$$

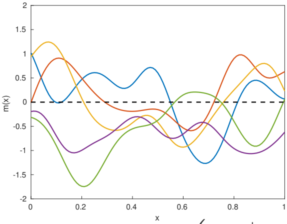
Goal: determine $\boldsymbol{m}^* = [m(x_1^*), \ldots, m(x_n^*)]$ for $x_i^* \in D^*$

Given: data $\boldsymbol{d} = [d_1, \ldots, d_s]$ at $x_1, \ldots, x_s$ for $x_i \in D_{\boldsymbol{d}}$

- Assume $m \sim \mathcal{N}(0, \mathcal{C}_{\text{pr}})$ (*prior*) and $\eta_i \sim \mathcal{N}(0, \sigma_n^2)$
- $\mathcal{C}_{\text{pr}}$ defined through *covariance function* $c(x, y)$

Absolute exponential

Squared exponential



General

$$c^{SE}(x, y) = \sigma^2 \exp\left(-\frac{1}{2}\frac{|x-y|^2}{l^2}\right)$$

Gaussian regression review

$$\boxed{d_i = m(x_i) + \eta_i}$$

Goal: determine $\boldsymbol{m}^* = [m(x_1^*), \ldots, m(x_n^*)]$ for $x_i^* \in D^*$

Given: data $\boldsymbol{d} = [d_1, \ldots, d_s]$ at $x_1, \ldots, x_s$ for $x_i \in D_{\boldsymbol{d}}$

- Assume $m \sim \mathcal{N}(0, \mathcal{C}_{\text{pr}})$ (*prior*) and $\eta_i \sim \mathcal{N}(0, \sigma_n^2)$

$\rightsquigarrow$ Joint multivariate distribution:

$$\begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{m}^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{C}_{\text{pr}}(D_{\boldsymbol{d}}, D_{\boldsymbol{d}}) + \sigma_n^2 \mathbf{I} & \mathbf{C}_{\text{pr}}(D_{\boldsymbol{d}}, D^*) \\ \mathbf{C}_{\text{pr}}(D^*, D_{\boldsymbol{d}}) & \mathbf{C}_{\text{pr}}(D^*, D^*) \end{bmatrix}\right),$$

where, e.g., $\mathbf{C}_{\text{pr}}(D^*, D_{\boldsymbol{d}}) \in \mathbb{R}^{n \times s}$ with $[\mathbf{C}_{\text{pr}}(D^*, D_{\boldsymbol{d}})]_{ij} = c^{SE}(x_i^*, x_j)$

Gaussian regression review

$$\boxed{d_i = m(x_i) + \eta_i}$$

Goal: determine $\boldsymbol{m}^* = [m(x_1^*), \ldots, m(x_n^*)]$ for $x_i^* \in D^*$

Given: data $\boldsymbol{d} = [d_1, \ldots, d_s]$ at $x_1, \ldots, x_s$ for $x_i \in D_{\boldsymbol{d}}$

- Assume $m \sim \mathcal{N}(0, \mathcal{C}_{\text{pr}})$ (*prior*) and $\eta_i \sim \mathcal{N}(0, \sigma_n^2)$

$\rightsquigarrow$ Joint multivariate distribution:

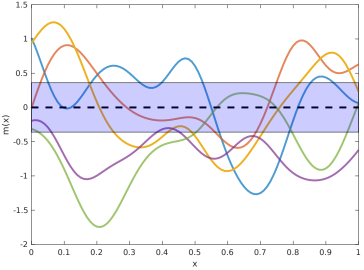
$$\begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{m}^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{C}_{\text{pr}}(D_{\boldsymbol{d}}, D_{\boldsymbol{d}}) + \sigma_n^2 \mathbf{I} & \mathbf{C}_{\text{pr}}(D_{\boldsymbol{d}}, D^*) \\ \mathbf{C}_{\text{pr}}(D^*, D_{\boldsymbol{d}}) & \mathbf{C}_{\text{pr}}(D^*, D^*) \end{bmatrix}\right),$$

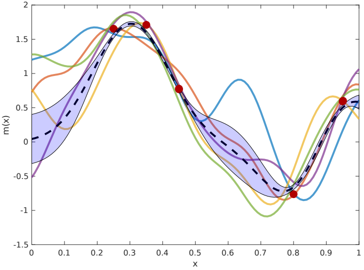
Bayesian inference $\rightsquigarrow$ *posterior* $\boldsymbol{m}^* | \boldsymbol{d} \sim \mathcal{N}(\boldsymbol{m}_{\text{post}}, \mathbf{C}_{\text{post}})$:

$$\boldsymbol{m}_{\text{post}} = \mathbf{C}_{\text{pr}}(D^*, D_{\boldsymbol{d}}) \left[\mathbf{C}_{\text{pr}}(D_{\boldsymbol{d}}, D_{\boldsymbol{d}}) + \sigma_n^2 \mathbf{I}\right]^{-1} \boldsymbol{d}$$

$$\mathbf{C}_{\text{post}} = \mathbf{C}_{\text{pr}}(D^*, D^*) - \mathbf{C}_{\text{pr}}(D^*, D_{\boldsymbol{d}}) \left[\mathbf{C}_{\text{pr}}(D_{\boldsymbol{d}}, D_{\boldsymbol{d}}) + \sigma_n^2 \mathbf{I}\right]^{-1} \mathbf{C}_{\text{pr}}(D_{\boldsymbol{d}}, D^*)$$

Gaussian regression – samples and variance

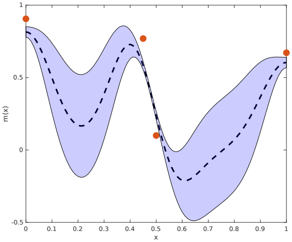


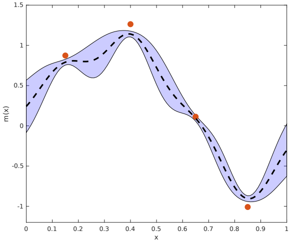


Prior:
$\boldsymbol{m}^* \sim \mathcal{N}\left(\mathbf{0}, \mathbf{C}_{\mathrm{pr}}(D^*, D^*)\right)$

Posterior:
$\boldsymbol{m}^* | \boldsymbol{d} \sim \mathcal{N}(\boldsymbol{m}_{\mathrm{post}}, \mathbf{C}_{\mathrm{post}})$

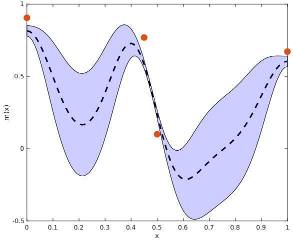
OED for Gaussian regression

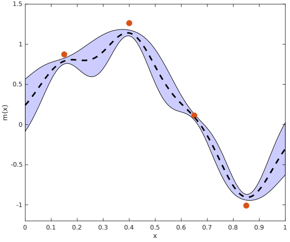




- Can observe $\boldsymbol{d}$ at a limited number of locations of our choice
- How to choose these locations to optimally infer $\boldsymbol{m}^*$?

OED for Gaussian regression





- Can observe $\boldsymbol{d}$ at a limited number of locations of our choice
- How to choose these locations to optimally infer $\boldsymbol{m}^*$?

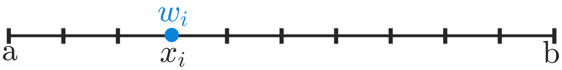
Requires:

1. Incorporation of design
2. Description of "optimal" design
3. Incorporation of cost constraints

1. Definition and incorporation of design

Design definition is problem specific

For 1D Gaussian regression:



- Assume grid of s possible measurement locations, $x_i \in [a, b]$
- Assign binary weight w_i to measurement at location x_i

$$w_i = \begin{cases} 1 \implies & \text{use measurement at } x_i \\ 0 \implies & \text{ignore measurement at } x_i \end{cases}$$

- $\boldsymbol{w} = [w_1, \ldots, w_s]$

1. Definition and incorporation of design

Design-dependent model:

$$\boxed{\boldsymbol{d}(\boldsymbol{w}) = \mathbf{W}\,(\boldsymbol{m} + \boldsymbol{\eta})}$$

- $\mathbf{W} := \mathbf{W}(\boldsymbol{w}) \in \mathbb{R}^{k(\boldsymbol{w}) \times s}$:

$$\begin{bmatrix} \boldsymbol{d}(\boldsymbol{w}) \\ \boldsymbol{m}^*(\boldsymbol{w}) \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{W}\left(\mathbf{C}_{\mathsf{pr}}(D_{\boldsymbol{d}}, D_{\boldsymbol{d}}) + \sigma_n^2 \mathbf{I}\right)\mathbf{W}^T & \mathbf{W}\mathbf{C}_{\mathsf{pr}}(D_{\boldsymbol{d}}, D^*) \\ \mathbf{C}_{\mathsf{pr}}(D^*, D_{\boldsymbol{d}})\mathbf{W}^T & \mathbf{C}_{\mathsf{pr}}(D^*, D^*) \end{bmatrix}\right)$$

1. Definition and incorporation of design

Design-dependent model:

$$\boxed{\boldsymbol{d}(\boldsymbol{w}) = \mathbf{W}\,(\boldsymbol{m} + \boldsymbol{\eta})}$$

- $\mathbf{W} := \mathbf{W}(\boldsymbol{w}) \in \mathbb{R}^{k(\boldsymbol{w}) \times s}$:

$$\begin{bmatrix} \boldsymbol{d}(\boldsymbol{w}) \\ \boldsymbol{m}^*(\boldsymbol{w}) \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{W}\left(\mathbf{C}_{\mathsf{pr}}(D_{\boldsymbol{d}}, D_{\boldsymbol{d}}) + \sigma_n^2\mathbf{I}\right)\mathbf{W}^T & \mathbf{W}\mathbf{C}_{\mathsf{pr}}(D_{\boldsymbol{d}}, D^*) \\ \mathbf{C}_{\mathsf{pr}}(D^*, D_{\boldsymbol{d}})\mathbf{W}^T & \mathbf{C}_{\mathsf{pr}}(D^*, D^*) \end{bmatrix}\right)$$

$\rightsquigarrow$ design-dependent posterior:

$$\boldsymbol{m}^* | \boldsymbol{d}(\boldsymbol{w}) \sim \mathcal{N}(\boldsymbol{m}_{\mathsf{post}}(\boldsymbol{w}), \mathbf{C}_{\mathsf{post}}(\boldsymbol{w}))$$

2. Description of "optimal" design

Goal: Choose measurement locations to minimize posterior "uncertainty"

- level of "uncertainty" measured by $\phi(\boldsymbol{w}) := \phi(\mathbf{C}_{\mathsf{post}}(\boldsymbol{w}))$

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}\in\{0,1\}^s} \phi(\boldsymbol{w})$$

2. Description of "optimal" design

Goal: Choose measurement locations to minimize posterior "uncertainty"

- level of "uncertainty" measured by $\phi(\boldsymbol{w}) := \phi(\mathbf{C}_{\text{post}}(\boldsymbol{w}))$

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}\in\{0,1\}^s}{\arg\min}\, \phi(\boldsymbol{w})$$

- $\lambda_1(\boldsymbol{w}) \leq \lambda_2(\boldsymbol{w}) \leq \ldots \leq \lambda_n(\boldsymbol{w})$ eigenvalues of $\mathbf{C}_{\text{post}}(\boldsymbol{w})$

Many choices for ϕ...

A-optimal: $\phi^A(\boldsymbol{w}) = \operatorname{trace}\left[\mathbf{C}_{\text{post}}(\boldsymbol{w})\right] = \sum_{i=1}^n \lambda_i(\boldsymbol{w})$

D-optimal: $\phi^D(\boldsymbol{w}) = \det\left[\mathbf{C}_{\text{post}}(\boldsymbol{w})\right] = \prod_{i=1}^n \lambda_i(\boldsymbol{w})$

E-optimal: $\phi^E(\boldsymbol{w}) = \lambda_n(\mathbf{C}_{\text{post}}(\boldsymbol{w}))$

3. Incorporation of cost constraints into optimization

$$\boldsymbol{w}^* \quad = \quad \operatorname*{arg\,min}_{\boldsymbol{w} \in \{0,1\}^s} \quad \phi^A(\boldsymbol{w})$$

- Trivial solution: $w_i = 1$ for $i = 1, \ldots, s$
- Real-world applications, measurements often costly

3. Incorporation of cost constraints into optimization

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}\in\{0,1\}^s}{\arg\min} \quad \phi^A(\boldsymbol{w})$$

- Real-world applications, measurements often costly

$\Longrightarrow$ introduce **cost constraints**:

3. Incorporation of cost constraints into optimization

$$\boldsymbol{w}^*_{\mathsf{opt}} = \underset{\substack{\boldsymbol{w}\in\{0,1\}^s \\ \text{s.t. } \sum_{i=1}^s w_i = k}}{\arg\min} \quad \phi^A(\boldsymbol{w})$$

- Real-world applications, measurements often costly

$\Longrightarrow$ introduce **cost constraints**:

1. Direct combinatorial search $\rightsquigarrow$ global optimal $\boldsymbol{w}^*_{\mathsf{opt}}$

- Requires $\binom{s}{k}$ evaluations of $\phi^A(\boldsymbol{w}) = \operatorname{trace}\left[\mathcal{C}_{\mathsf{post}}(\boldsymbol{w})\right]$

3. Incorporation of cost constraints into optimization

$$\boldsymbol{w}_{\mathsf{G}}^{*} \approx \boldsymbol{w}_{\mathsf{opt}}^{*} = \underset{\substack{\boldsymbol{w} \in \{0,1\}^s \\ \text{s.t. } \sum_{i=1}^{s} w_i = k}}{\arg\min} \quad \phi^A(\boldsymbol{w})$$

- Real-world applications, measurements often costly

$\Longrightarrow$ introduce **cost constraints**:

1. Direct combinatorial search $\rightsquigarrow$ global optimal $\boldsymbol{w}_{\mathsf{opt}}^{*}$

- Requires $\binom{s}{k}$ evaluations of $\phi^A(\boldsymbol{w}) = \operatorname{trace}\left[\mathcal{C}_{\mathsf{post}}(\boldsymbol{w})\right]$

2. Greedy approach

- Simple to implement, less ϕ^A evaluations but still many, suboptimal

3. Incorporation of cost constraints into optimization

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}\in[0,1]^s}{\arg\min} \quad \phi^A(\boldsymbol{w}) + \gamma\psi(\boldsymbol{w})$$

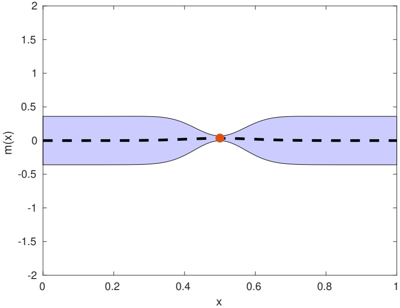
- Real-world applications, measurements often costly

$\Longrightarrow$ introduce **cost constraints**:

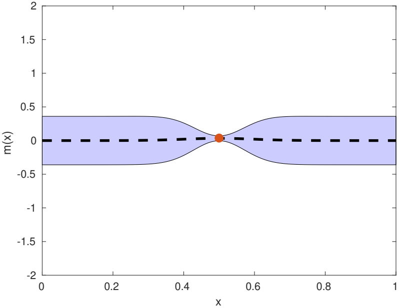
1. Direct combinatorial search $\rightsquigarrow$ global optimal $\boldsymbol{w}^*_{\text{opt}}$
   - Requires $\binom{s}{k}$ evaluations of $\phi^A(\boldsymbol{w}) = \operatorname{trace}[\mathcal{C}_{\text{post}}(\boldsymbol{w})]$
2. Greedy approach
   - Simple to implement, less ϕ^A evaluations but still many, suboptimal
3. Relaxation + sparsification
   - \# of ϕ^A evaluations does not grow with \# of sensors, gradients of ϕ^A needed, indirect control of sparsity, suboptimal

Choosing a design for Gaussian regression

Greedy



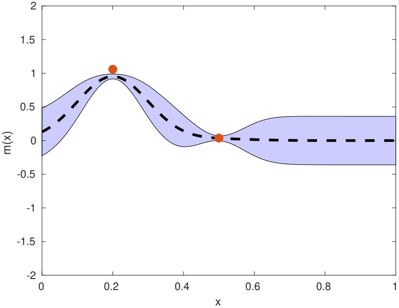
Optimal



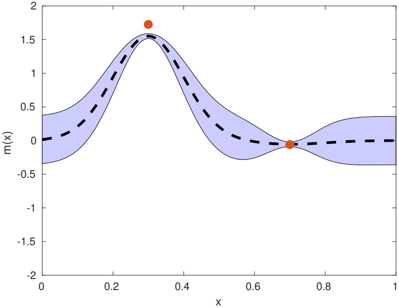
Choosing locations one at a time is simple but suboptimal

Choosing a design for Gaussian regression

Greedy

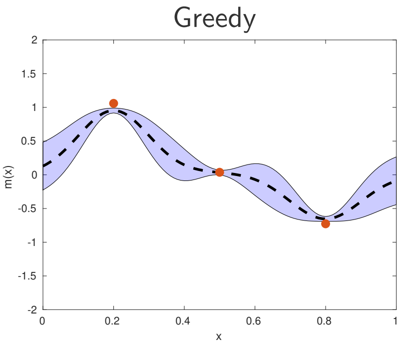


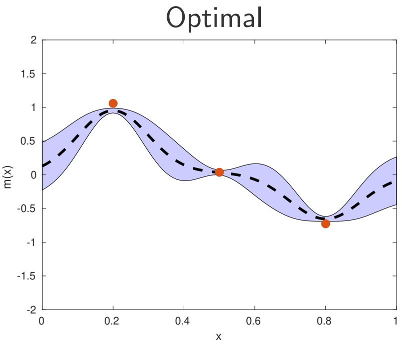
Optimal



Choosing locations one at a time is simple but suboptimal

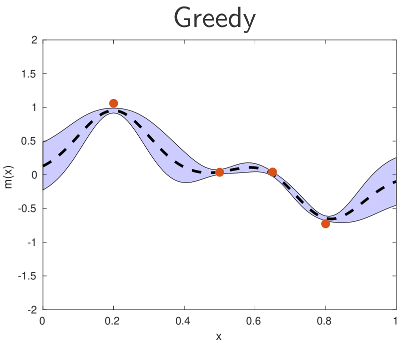
Choosing a design for Gaussian regression

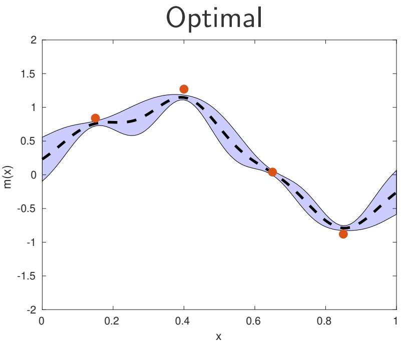




Choosing locations one at a time is simple but suboptimal

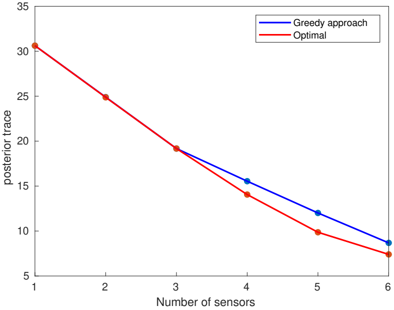
Choosing a design for Gaussian regression





Choosing locations one at a time is simple but suboptimal

Choosing a design for Gaussian regression



Choosing locations one at a time is simple but suboptimal

Infinite-dimensional Bayesian inverse problems

Introduce non-trivial *parameter-to-observable map* $\mathcal{F} : \mathcal{H} \to \mathbb{R}^d$

$$\boxed{\boldsymbol{d} = \mathcal{F}(m) + \boldsymbol{\eta}}$$

- $\mathcal{F}$: PDE solve + spatiotemporal observation operator
- $m \sim \mu_0 = \mathcal{N}(0, \mathcal{C}_{\text{pr}})$, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}})$

Goal: Infer posterior measure for m given *indirect* noisy measurements $\boldsymbol{d}$

Infinite-dimensional Bayesian inverse problems

Introduce non-trivial *parameter-to-observable map* $\mathcal{F} : \mathcal{H} \to \mathbb{R}^d$

$$\boxed{\boldsymbol{d} = \mathcal{F}(m) + \boldsymbol{\eta}}$$

- $\mathcal{F}$: PDE solve + spatiotemporal observation operator
- $m \sim \mu_0 = \mathcal{N}(0, \mathcal{C}_{\text{pr}})$, $\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\text{noise}})$

Goal: Infer posterior measure for m given *indirect* noisy measurements $\boldsymbol{d}$

Bayes' rule $\rightsquigarrow$ posterior law on m:

$$\frac{d\mu_{\text{post}}^{\boldsymbol{d}}}{d\mu_0} \propto \pi_{\text{like}}(\boldsymbol{d}|m), \quad \pi_{\text{like}}(\boldsymbol{d}|m) \propto \exp\left[-\frac{1}{2}\|\mathcal{F}(m) - \boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}}\right]$$

Infinite-dimensional *linear* Bayesian inverse problems

Introduce non-trivial *parameter-to-observable map* $\mathcal{F} : \mathcal{H} \to \mathbb{R}^d$

$$\boxed{\boldsymbol{d} = \mathcal{F}\, m + \boldsymbol{\eta}}$$

- $\mathcal{F}$: PDE solve + spatiotemporal observation operator
- $m \sim \mu_0 = \mathcal{N}(0, \mathcal{C}_{\text{pr}})$, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}})$

Goal: Infer posterior measure for m given *indirect* noisy measurements $\boldsymbol{d}$

Bayes' rule $\rightsquigarrow$ posterior law on m:

$$\frac{d\mu_{\text{post}}^{\boldsymbol{d}}}{d\mu_0} \propto \pi_{\text{like}}(\boldsymbol{d}|m), \quad \pi_{\text{like}}(\boldsymbol{d}|m) \propto \exp\left[-\frac{1}{2}\|\mathcal{F}\, m - \boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}}\right]$$

- For linear $\mathcal{F}$, $m|\boldsymbol{d} \sim \mathcal{N}(m_{\text{post}}, \mathcal{C}_{\text{post}})$ with

$$\boxed{\mathcal{C}_{\text{post}} = \left(\mathcal{F}^*\boldsymbol{\Gamma}^{-1}_{\text{noise}}\mathcal{F} + \mathcal{C}_{\text{pr}}^{-1}\right)^{-1}}$$

OED for infinite-dimensional Bayesian inverse problems

As before:

- Grid of s possible sensor locations for measurement collection at r times $\implies d = rs$ observations
- Differentiate between designs through *(block-)diagonal* $\mathbf{W} \in \mathbb{R}^{d\times d}$

Other design definitions possible

Design enters through the likelihood:

$$\pi_{\mathsf{like}}(\boldsymbol{d}|m) \propto \exp\left[-\frac{1}{2}\|\mathcal{F}m - \boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\mathbf{W}}}\right]$$

- $\boldsymbol{\Gamma}^{-1}_{\mathbf{W}}$ depends on noise model
- For *uncorrelated* noise, e.g., $\boldsymbol{\Gamma}_{\mathsf{noise}} = \sigma_n^2\mathbf{I}$, $\boldsymbol{\Gamma}^{-1}_{\mathbf{W}} := \frac{1}{\sigma_n^2}\mathbf{W}$

OED for infinite-dimensional Bayesian inverse problems

Design enters through the likelihood:

$$\pi_{\text{like}}(\boldsymbol{d}|m) \propto \exp\left[-\frac{1}{2\sigma_n^2}\|\mathcal{F}m - \boldsymbol{d}\|^2_{\mathbf{W}}\right]$$

$\rightsquigarrow$ design-dependent posterior measure $m|\boldsymbol{d}(\boldsymbol{w}) \sim \mathcal{N}(m_{\text{post}}(\boldsymbol{w}), \mathcal{C}_{\text{post}}(\boldsymbol{w}))$

$$\boxed{\mathcal{C}_{\text{post}}(\boldsymbol{w}) = \left(\sigma_n^{-2}\mathcal{F}^*\mathbf{W}\mathcal{F} + \mathcal{C}_{\text{pr}}^{-1}\right)^{-1}}$$

Infinite-dimensional A-optimality criterion and challenges

- *Infinite-dimensional* A-optimality criterion defined by:

$$\phi^A(\boldsymbol{w}) = \mathrm{trace}\left[\mathcal{C}_{\mathsf{post}}(\boldsymbol{w})\right] = \mathrm{trace}\left[\left(\sigma_n^{-2}\mathcal{F}^*\mathbf{W}\mathcal{F} + \mathcal{C}_{\mathsf{pr}}^{-1}\right)^{-1}\right]$$

Finding A-optimal designs is challenging:

- Requires many evaluations of ϕ^A
- Computing trace of ∞-dimensional, PDE-dependent operator
- Finding global or greedy optimal is too expensive

Making OED computationally tractable

1. **Approximate the trace**
   - Use *randomized trace estimation*
   - **Alternatively**: reformulate ϕ^A to reduce dimensionality

Making OED computationally tractable

1. **Approximate the trace**
   - Use *randomized trace estimation*
   - **Alternatively**: reformulate ϕ^A to reduce dimensionality
2. **Eliminate PDEs**
   - Exploit low-rank structure of $\mathcal{F}$
   - Approximate $\mathcal{F}$ with truncated SVD using matrix-free algorithms

Making OED computationally tractable

1. **Approximate the trace**
   - Use *randomized trace estimation*
   - **Alternatively**: reformulate ϕ^A to reduce dimensionality
2. **Eliminate PDEs**
   - Exploit low-rank structure of $\mathcal{F}$
   - Approximate $\mathcal{F}$ with truncated SVD using matrix-free algorithms
3. **Enforce sparse designs with sparsity-inducing penalty ψ**

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}\in[0,1]^s}{\arg\min}\, \phi^A(\boldsymbol{w}) + \gamma\psi(\boldsymbol{w})$$

   - $\psi(\boldsymbol{w}) \approx \|\boldsymbol{w}\|_0$, the number of non-zero weights

Making OED computationally tractable

1. **Approximate the trace**
   - Use *randomized trace estimation*
   - **Alternatively**: reformulate ϕ^A to reduce dimensionality
2. **Eliminate PDEs**
   - Exploit low-rank structure of $\mathcal{F}$
   - Approximate $\mathcal{F}$ with truncated SVD using matrix-free algorithms
3. **Enforce sparse designs with sparsity-inducing penalty** $\psi_{\varepsilon(i)}$

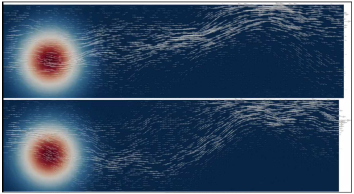
$$\boxed{\boldsymbol{w}^*{}_{\varepsilon(i)} = \underset{\boldsymbol{w}\in[0,1]^s}{\arg\min}\, \phi^A(\boldsymbol{w}) + \gamma \psi_{\varepsilon(i)}(\boldsymbol{w})}$$

- $\psi(\boldsymbol{w}) \approx \|\boldsymbol{w}\|_0$, the number of non-zero weights
- ℓ_0-sparsification, $\psi_{\varepsilon(i)} \to \|\cdot\|_0$ as $i \to \infty$

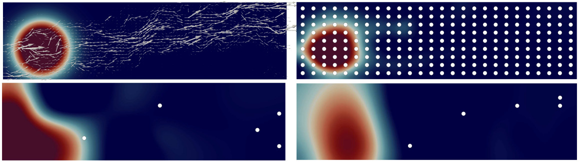
Motivating example

Many models for real-world phenomena have uncertain inputs

Ex: Contaminant source identification in groundwater flow:



- Designs need to work well for all realizations of uncertainty



OED under uncertainty

Common OED assumptions:

- exact knowledge of model equations
- no other sources of uncertainty

Aim of this work:

- Formulation of OED under irreducible uncertainty
- Mathematical structure and computational challenges

OED for Bayesian linear inverse problems under uncertainty

Incorporate uncertainty into $\mathcal{F}$, $\mathcal{F} : (\Omega, \mathcal{G}, P) \to \mathcal{L}(\mathcal{H}, \mathbb{R}^d)$

$$\boxed{\boldsymbol{d}(\xi) = \mathcal{F}(\xi)m + \boldsymbol{\eta}}$$

Likelihood depends on uncertainty:

$$\pi_{\mathsf{like}}(\boldsymbol{d}|m) \propto \exp\left[-\frac{1}{2\sigma_n^2}\|\mathcal{F}(\xi)m - \boldsymbol{d}\|_{\mathbf{W}}^2\right]$$

$\implies$ posterior depends on uncertainty:

$$\mu_{\mathsf{post}}^{\boldsymbol{d}} = \mathcal{N}(m_{\mathsf{post}}(\xi, \boldsymbol{w}), \mathcal{C}_{\mathsf{post}}(\xi, \boldsymbol{w}))$$

$$\boxed{\mathcal{C}_{\mathsf{post}}(\xi, \boldsymbol{w}) = \left(\frac{1}{\sigma_n^2}\mathcal{F}^*(\xi)\mathbf{W}\mathcal{F}(\xi) + \mathcal{C}_{\mathsf{pr}}^{-1}\right)^{-1}}$$

A-optimal design under uncertainty

- A-optimal design under uncertainty:

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}\in[0,1]^d}{\arg\min} \int_\Omega \mathrm{trace}\left[\mathcal{C}_{\mathsf{post}}(\xi, \boldsymbol{w})\right] P(d\xi) + \gamma\psi\left(\boldsymbol{w}\right)$$

- Minimizes expected value of average posterior variance
- Uncertainty-aware designs do well on average, but are not optimal given fixed ξ

Computational challenges

1. **Discretization of uncertainty**
2. **Efficient computation of trace**
3. **Tractable computation of optimal designs**

1. Discretization of the uncertainty

- Approximate the expected value of the average pointwise posterior variance
- Assuming we can sample $\xi_i \in \Omega$, we use SAA to approximate the integral:

$$\int_\Omega \mathrm{trace}\left[\mathcal{C}_{\mathsf{post}}(\xi, \boldsymbol{w})\right] P(d\xi) \approx \frac{1}{N}\sum_{i=1}^{N} \mathrm{trace}\left[\mathcal{C}_{\mathsf{post}}(\xi_i, \boldsymbol{w})\right]$$

2. Computation of trace, "measurement space approach"

Discretized trace (using, e.g., finite elements, $\mathcal{F} \approx \mathbf{F} \in \mathbb{R}^{d \times n}$):

$$\phi^A(\xi, \boldsymbol{w}) \approx \phi_n^A(\xi, \boldsymbol{w}) = \operatorname{trace}\left[\left(\frac{1}{\sigma_n^2}\mathbf{F}(\xi)^*\mathbf{W}\mathbf{F}(\xi) + \mathbf{C}_{\mathsf{pr}}^{-1}\right)^{-1}\right]$$

- Too expensive to compute trace exactly even after discretization
- We can rewrite $\phi_n^A(\xi, \boldsymbol{w})$ as:

$$\begin{aligned}\phi_n^A(\xi, \boldsymbol{w}) &= \operatorname{trace}\left[\mathbf{C}_{\mathsf{pr}}\right] - \operatorname{trace}\left[\frac{1}{\sigma_n^2}\mathbf{S}^{-1}(\xi, \boldsymbol{w})\mathbf{W}\mathbf{F}(\xi)\mathbf{C}_{\mathsf{pr}}^2\mathbf{F}^*(\xi)\right] \\ &= \operatorname{trace}\left[\mathbf{C}_{\mathsf{pr}}\right] - \operatorname{trace}\left[\mathbf{K}(\xi, \boldsymbol{w})\right]\end{aligned}$$

2. Computation of trace, “measurement space approach”

Discretized trace (using, e.g., finite elements, $\mathcal{F} \approx \mathbf{F} \in \mathbb{R}^{d\times n}$):

$$\phi^A(\xi, \boldsymbol{w}) \approx \phi_n^A(\xi, \boldsymbol{w}) = \text{trace}\left[\left(\frac{1}{\sigma_n^2}\mathbf{F}(\xi)^*\mathbf{W}\mathbf{F}(\xi) + \mathbf{C}_{\mathsf{pr}}^{-1}\right)^{-1}\right]$$

- Too expensive to compute trace exactly even after discretization
- We can rewrite $\phi_n^A(\xi, \boldsymbol{w})$ as:

$$\begin{aligned}\phi_n^A(\xi, \boldsymbol{w}) &= \text{trace}\left[\mathbf{C}_{\mathsf{pr}}\right] - \text{trace}\left[\frac{1}{\sigma_n^2}\mathbf{S}^{-1}(\xi, \boldsymbol{w})\mathbf{W}\mathbf{F}(\xi)\mathbf{C}_{\mathsf{pr}}^2\mathbf{F}^*(\xi)\right] \\ &= \text{trace}\left[\mathbf{C}_{\mathsf{pr}}\right] - \text{trace}\left[\mathbf{K}(\xi, \boldsymbol{w})\right]\end{aligned}$$

- Optimal design satisfies:

$$\boxed{\boldsymbol{w}^* = \underset{\boldsymbol{w}\in[0,1]^d}{\arg\min}\left[-\tfrac{1}{N}\textstyle\sum_{i=1}^N \text{trace}\left[\mathbf{K}(\xi_i, \boldsymbol{w})\right] + \gamma\psi\left(\boldsymbol{w}\right)\right]}$$

Trace of an operator in measurement space (finite)

3. Elimination of PDEs from the minimization

OEDUU objective is expensive to optimize:

$$\frac{1}{N}\sum_{i=1}^{N} \operatorname{trace}\left[\mathbf{K}(\xi_i, \boldsymbol{w})\right]$$

- $\mathbf{K}(\xi_i, \boldsymbol{w})$ depends on $\mathbf{F}_i$ and $\mathbf{F}_i^*$
- Computing trace even for one ξ_i requires many PDE solves
- Need to compute trace for each sample ξ_i

3. Elimination of PDEs from the minimization

Find a low-rank approximation to $\mathbf{F}(\xi_i)\mathbf{C}_{\mathbf{pr}}^{\frac{1}{2}} = \widetilde{\mathbf{F}}_i$

- Preconditioning promotes faster decay of eigenvalues
- Matrix-free techniques based on randomized linear algebra

Storing separate basis vectors for each $\widetilde{\mathbf{F}}_i$ is infeasible

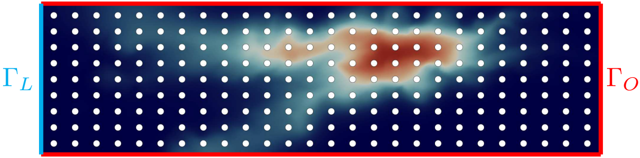
- **Solution**: find a space that captures the "effective" composite range space for all $\mathbf{F}(\xi_i)$

Find $\mathbf{Q} \in \mathbb{R}^{d\times k}$ and $\widehat{\mathbf{Q}} \in \mathbb{R}^{m\times k}$ (k small) such that $\forall\, i \in \{1,\dots,N\}$:

$$\widetilde{\mathbf{F}}_i \approx \mathbf{Q}\mathbf{Q}^*\widetilde{\mathbf{F}}_i\widehat{\mathbf{Q}}\widehat{\mathbf{Q}}^*$$

Many ways to make this more efficient...

Numerical example - subsurface flow OED



$$
\begin{aligned}
u_t - \kappa \Delta u + \boldsymbol{v}(\xi) \cdot \nabla u &= 0 \quad && \text{in } \mathcal{D} \times (T_i, T) \\
u(\cdot, T_i) &= m \quad && \text{in } \mathcal{D} \\
-\kappa \nabla u \cdot \boldsymbol{n} + \boldsymbol{v}(\xi) \cdot \boldsymbol{n} u &= 0 \quad && \text{on } \Gamma_L \times (T_i, T) \\
\kappa \nabla u \cdot \boldsymbol{n} &= 0 \quad && \text{on } \Gamma_O \times (T_i, T)
\end{aligned}
$$

- Grid of 234 sensor locations, measurements taken at $T_j \in \{\tau_1, \ldots, \tau_r\}$
- Samples $\{\boldsymbol{v}(\xi_i)\}_{i=1}^N$ of the velocity field and $T_i \sim \mathcal{U}[-1, 1]$ of initial time

$\rightsquigarrow$ Find subset of locations minimizing A-optimal criterion under uncertainty

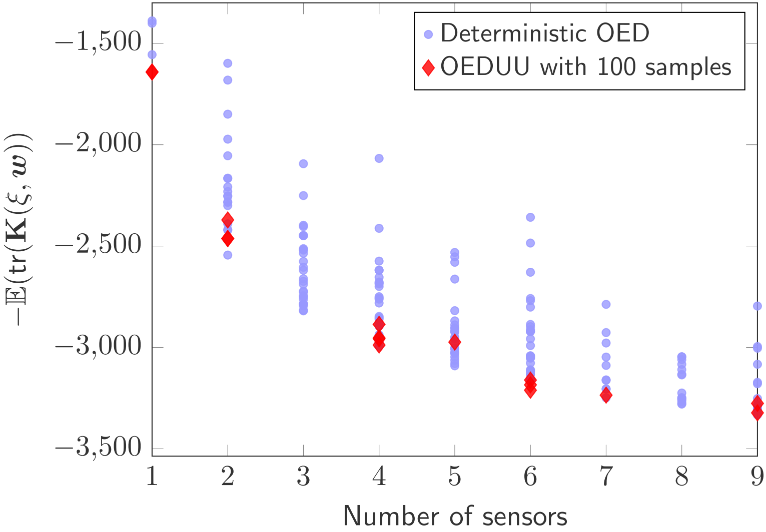
Spatial/temporal discretization: built on FEniCS and hIPPYlib (open source Python/C++ framework)

Subsurface flow OEDUU

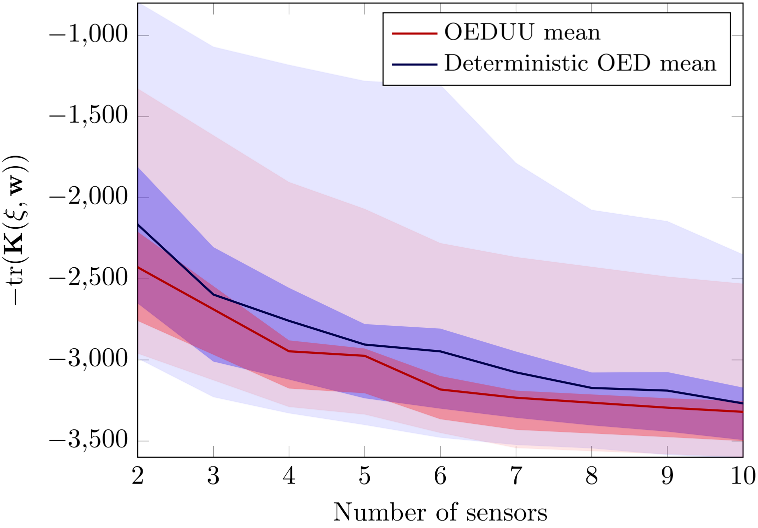
$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}\in[0,1]^d} \left[-\frac{1}{N} \sum_{i=1}^{N} \mathrm{trace}\left[\mathbf{K}(\xi_i, \boldsymbol{w})\right] + \gamma\psi(\boldsymbol{w}) \right]$$

- $N = 100$ samples for discretization of uncertainty
- ℓ_0-sparsification used to find sparse designs
- Each minimization solved with gradient-based method (projected BFGS)

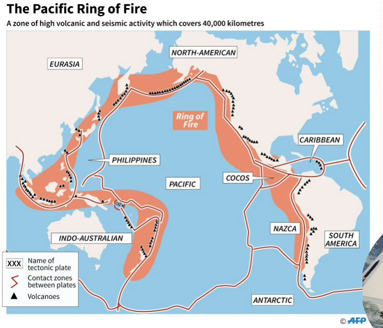
Deterministic vs. designs under uncertainty



Deterministic vs. designs under uncertainty



Earthquake-generated tsunamis



- Tsunamis generated by earthquakes beneath ocean floor at subduction zones
- Water pressure/height readings are used to detect and track tsunamis (*DART* system)
- Tsunami detection and warning relies on informative data

Earthquake-generated tsunamis

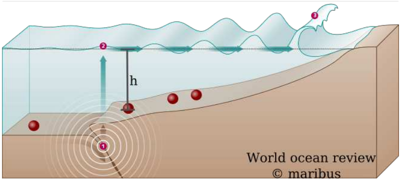
$$\boldsymbol{d} = \mathcal{G}(B) + \boldsymbol{\eta}$$

$$B \sim \mathcal{N}(B_{\text{pr}}, \mathcal{C}_{\text{pr}})$$
$$\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$$

Governing equation for $\mathcal{G}$:

$$\begin{bmatrix} h \\ u \\ v \end{bmatrix}_t + \begin{bmatrix} u \\ \frac{u^2}{h} + \frac{1}{2}gh^2 \\ \frac{uv}{h} \end{bmatrix}_x + \begin{bmatrix} v \\ \frac{uv}{h} \\ \frac{v^2}{h} + \frac{1}{2}gh^2 \end{bmatrix}_y = -\begin{bmatrix} 0 \\ ghB_x \\ ghB_y \end{bmatrix}$$

Goal: Find optimal configuration of sensors for inference of B



Gaussian approximation to posterior distribution

SWE **nonlinear** $\Longrightarrow$

1. solutions can exhibit shocks

2. non-Gaussian posterior

Gaussian approximation to posterior distribution

SWE **nonlinear** $\Longrightarrow$

1. solutions can exhibit shocks $\Longrightarrow$ difficulties for adjoint-based inversion methods
2. non-Gaussian posterior $\Longrightarrow$ difficulties for OED problem

Gaussian approximation to posterior distribution

SWE **nonlinear** $\implies$

1. solutions can exhibit shocks $\implies$ difficulties for adjoint-based inversion methods
2. non-Gaussian posterior $\implies$ difficulties for OED problem

Solutions well-approximated by linearization in deep water:

$$\mathcal{G}(B) \approx \mathcal{G}(B_0) + \mathcal{F}\left[B - B_0\right], \quad \mathcal{F} := \mathcal{G}'(B_0)$$

Linearization $\rightsquigarrow$ Gaussian approximation to posterior:

$B|\boldsymbol{d} \sim \mathcal{N}(B_{\mathsf{post}}(\boldsymbol{w}), \mathcal{C}_{\mathsf{post}}(\boldsymbol{w}))$ with

$$\boxed{\mathcal{C}_{\mathsf{post}}(\boldsymbol{w}) = \left(\tfrac{1}{\sigma_n^2}\mathcal{F}^*\mathbf{W}\mathcal{F} + \mathcal{C}_{\mathsf{pr}}^{-1}\right)^{-1}}$$

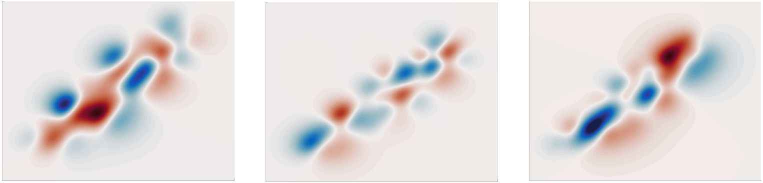
Prior on B

Bathymetry change $\hat{B} := B - B_0$ is due to a slip at a fault

- *Okada* model $\rightsquigarrow$ linear relationship between slips $\mathbf{S}$ at m slip patches and seafloor deformation

$$\hat{B} = O\mathbf{S}$$

- Prior on slips $\mathbf{S} \sim \mathcal{N}(\mathbf{0}, \theta^2 \mathbf{I})$ induces prior on $B \sim \mathcal{N}(B_0, \mathcal{C}_{\text{pr}}(O))$
- Reasonable sample seafloor deformations



Inversion for slips $\mathbf{S}$

Exploiting linear relationship between $\mathbf{S}$ and $\hat{B}$ $\rightsquigarrow$

Reformulation of inverse problem:

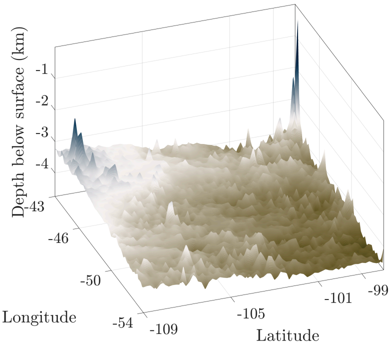
$$\boldsymbol{d} = \mathcal{G}(B_0) + \mathcal{F}\hat{B} + \boldsymbol{\eta} = \mathcal{G}(B_0) + \mathcal{F}O\mathbf{S} + \boldsymbol{\eta}$$

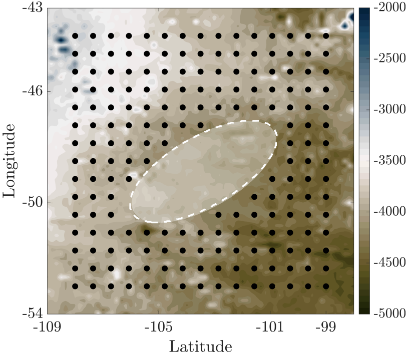
$\rightsquigarrow$ *finite-dimensional* posterior distribution for slips $\mathbf{S} \in \mathbb{R}^m$:

$$\mathbf{S}|\boldsymbol{d} \sim \mathcal{N}(\mathbf{S}_{\text{post}}(\boldsymbol{w}), \mathbf{C}_{\text{post}}(\boldsymbol{w})) \text{ with}$$

$$\boxed{\mathbf{C}_{\text{post}}(\boldsymbol{w}) = \left(\frac{1}{\sigma_n^2} O^* \mathcal{F}^* \mathbf{W} \mathcal{F} O + \theta^{-2} \mathbf{I}\right)^{-1}}$$

2D example - problem setup



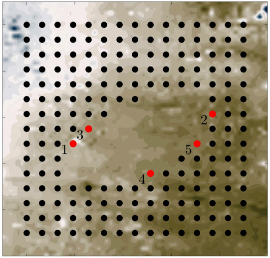


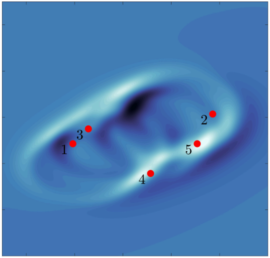
- $s = 189$ possible locations
- $r = 8$ observation times
- $m = 20$ slip patches
- GeoClaw used for numerical results

2D example - OED problem

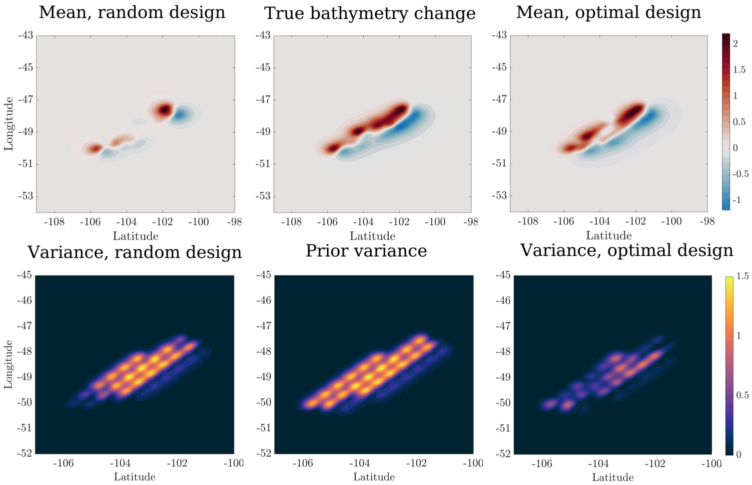
$$\boldsymbol{w}_{\mathsf{G}}^{*} \approx \underset{\substack{\boldsymbol{w}\in\{0,1\}^s \\ \text{s.t. } \sum_{i=1}^{s} w_i = k}}{\arg\min} \ \text{trace}\left[\left(\frac{1}{\sigma_n^2}(\mathcal{F}\mathcal{O})^{*}\mathbf{W}(\mathcal{F}\mathcal{O}) + \theta^{-2}\mathbf{I}\right)^{-1}\right]$$

- $\mathcal{F}\mathcal{O} \in \mathbb{R}^{d\times m}$ precomputed
- No adjoint solves needed
- Greedy optimal weight vectors computed, PDE free

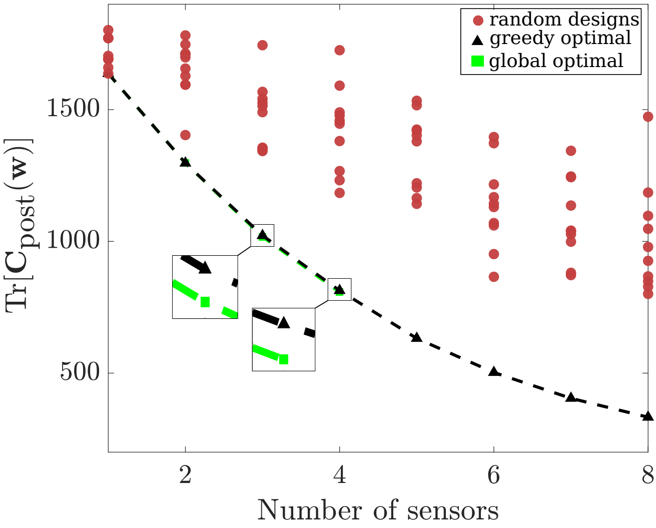




2D example - design comparisons



2D example - design comparisons



Summary and future work

Summary:

1. OED under uncertainty:

- Introduced mathematical framework for incorporation of uncertainty
- Presented "measurement space approach" formulation of OED objective
- Eliminated PDEs from minimization using *joint* basis
- Demonstrated effectiveness of OEDUU using numerical example

Summary and future work

Summary:

1. OED under uncertainty:

- Introduced mathematical framework for incorporation of uncertainty
- Presented "measurement space approach" formulation of OED objective
- Eliminated PDEs from minimization using *joint* basis
- Demonstrated effectiveness of OEDUU using numerical example

2. OED for tsunami source reconstruction:

- Formulated OED problem for deep-ocean tsunami source reconstruction using SWE
- Used Gaussian approximation to posterior through linearization
- Reformulated problem to invert for slips allowing elimination of PDEs

Summary and future work

Possible extensions:

- Alternate ways of dealing with uncertainty, e.g., stochastic approximation or Taylor expansion
- Laplace approximation to posterior
- Inclusion of uncertain parameters into tsunami model
- Incorporate OED framework into GeoClaw

Thank you!

References:

- Alexanderian, Petra, Stadler & Ghattas (2014), *A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized ℓ_0-sparsification* SISC, 36(5)
- Berger, George, LeVeque, & Mandli *The GeoClaw software for depth-averaged flows with adaptive refinement* Advances in Water Resources, 34(9) pp. 1195–1206
- Clawpack Development Team (2020), *Clawpack Version 5.7.1*, http://www.clawpack.org, doi: 10.5281/zenodo.4025432
- Koval, Alexanderian & Stadler (2020), *Optimal experimental design under irreducible uncertainty for linear inverse problems governed by PDEs* Inverse Problems, 36(7)
- Okada (1985), *Surface deformation due to shear and tensile faults in a half-space*, Bulletin of the seismological society of America, 75 pp. 1135–1154
- Rasmussen (2003), *Gaussian processes in machine learning* Springer, pp 63–71
- Stuart (2010), *Inverse problems: a Bayesian perspective* Acta numerica, 19 pp. 451–559
- Villa, Petra & Ghattas (2021). *HIPPYlib: An Extensible Software Framework for Large-Scale Inverse Problems Governed by PDEs: Part I: Deterministic Inversion and Linearized Bayesian Inference* ACM Trans. Math. Softw. 47, 2, Article 16