SAND2017-2073PE

Optimal Experimental Design and Uncertainty Propagation in Physical Systems

Xun (Ryan) Huan

Sandia National Laboratories
Combustion Research Facility

February 23, 2017

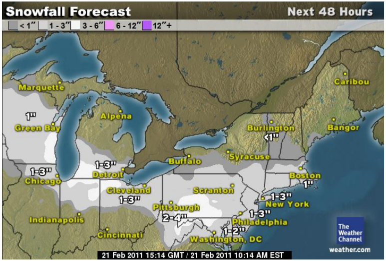
About me

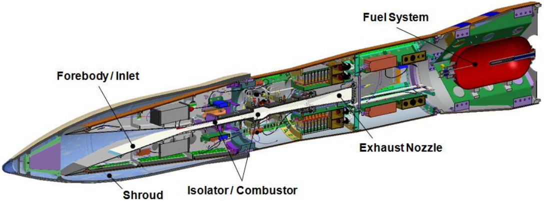
Postdoc at Sandia National Labs, California

- Ph.D. in Computational Science and Engineering, MIT, 2015
- S.M. in Aerospace Engineering, MIT, 2010
- B.A.Sc. in Engineering Science (Aerospace), Toronto, 2008

Research focus: uncertainty quantification

Uncertainty in engineering and science applications





Uncertainty is everywhere

"Uncertainty is everywhere and you cannot escape from it."

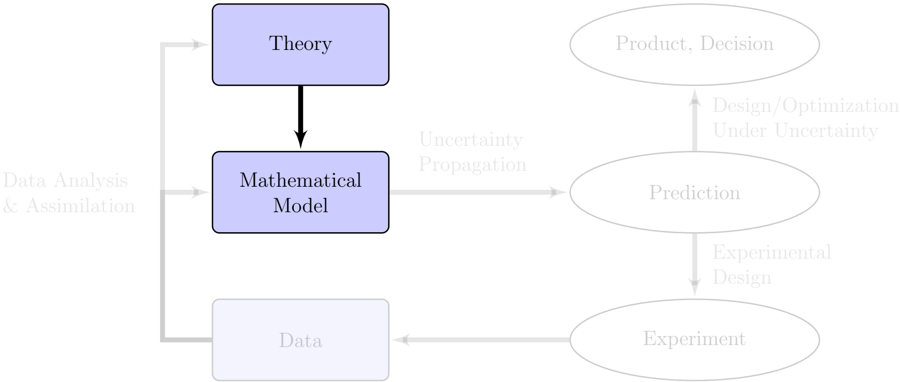
— Dennis Lindley

Uncertainty quantification (UQ) provides a systematic approach to:

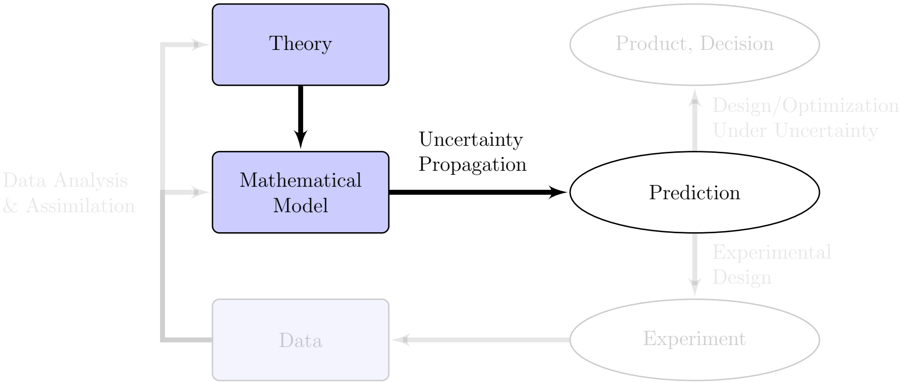
- characterize
- incorporate
- propagate
- reduce

...uncertainty for computational and real-world applications

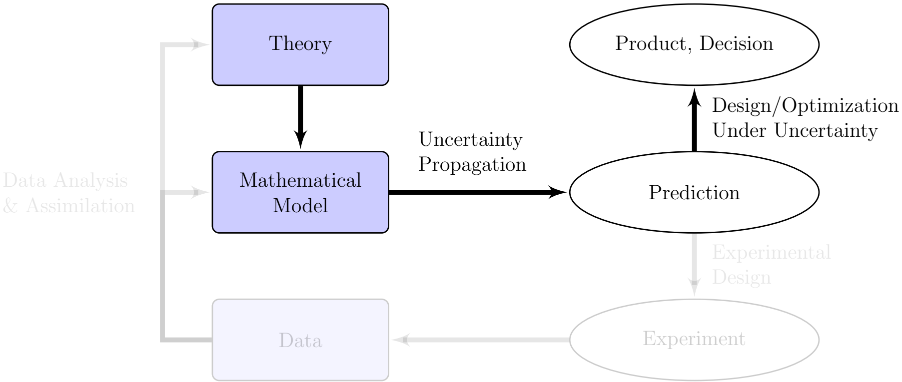
UQ in engineering and science: big picture



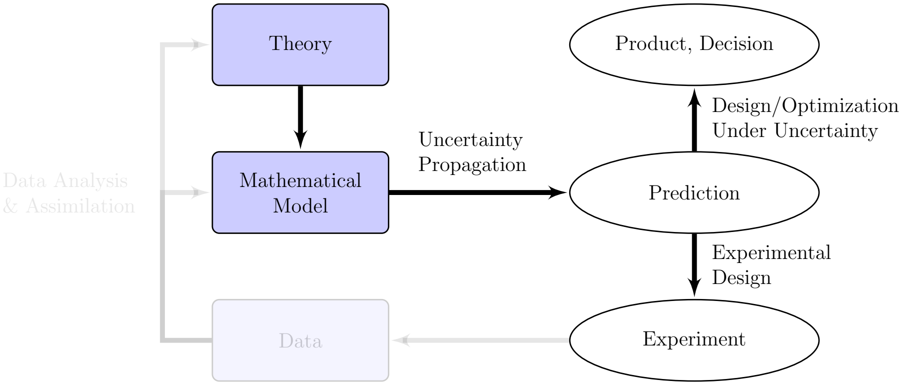
UQ in engineering and science: big picture



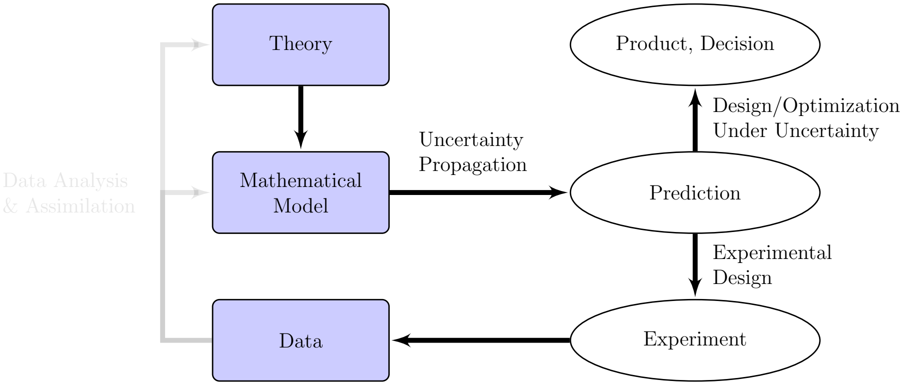
UQ in engineering and science: big picture



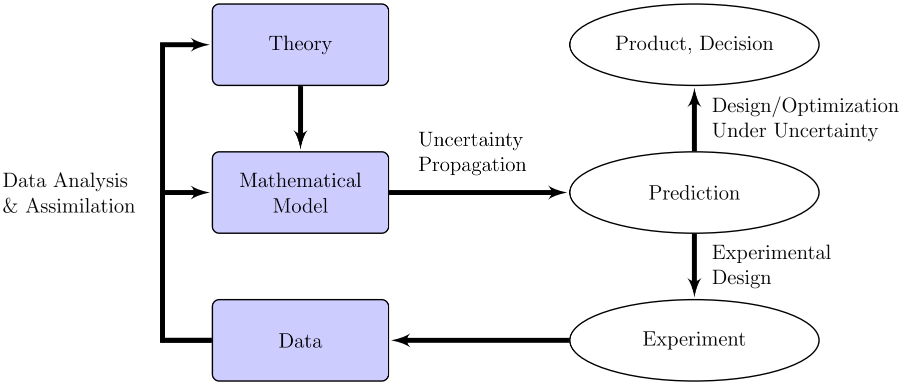
UQ in engineering and science: big picture



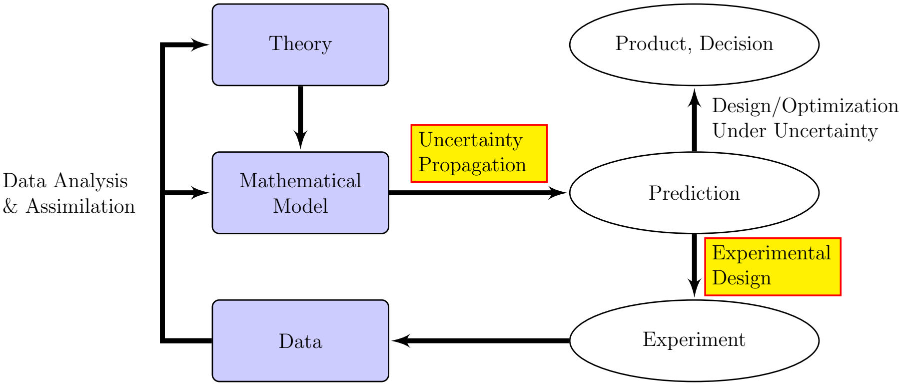
UQ in engineering and science: big picture



UQ in engineering and science: big picture



UQ in engineering and science: big picture



Outline

1. Sequential Optimal Experimental Design
   - Formulation
   - Numerical Methods
   - Results
2. Uncertainty Propagation in Scramjet Computations
3. Summary and Future Work

Outline

1. Sequential Optimal Experimental Design
   - Formulation
   - Numerical Methods
   - Results
2. Uncertainty Propagation in Scramjet Computations
3. Summary and Future Work

Some experiments are more useful than others

Experiments are:

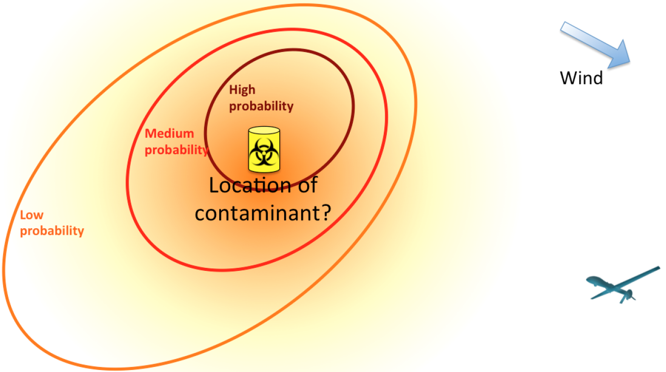
- Expensive
- Time-consuming
- Delicate to perform

Experimental design helps address:

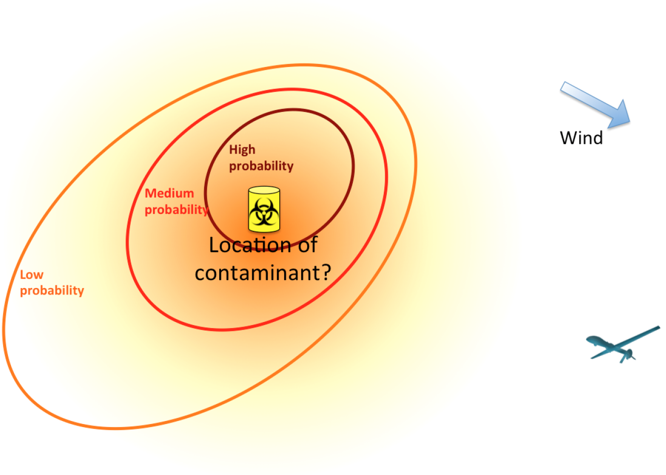
- Under what **conditions** to perform the experiment?
- **What** to measure?
- **Where** to measure?
- **When** to measure?

Today's talk: design of **multiple** experiments

Planning measurements: batch (non-sequential) design



Planning measurements: sequential design



Sequential experimental design is less developed

Batch experimental design:

- Linear: Fisher information matrix (e.g., A-, D-optimal)
- Nonlinear: advances beyond linearization and Gaussianization
- Information-based experimental design [Lindley 56]

Greedy (myopic) design:

- Repeated application of batch design [Solonen 12, Drovandi 14, Kim 14]
- *Not optimal*

Dynamic programming:

- Fully optimal description (has 1. feedback, 2. forward looking)
- Thus far limited to discrete variables, special problem and solution structures [Carlin, Bradley 98, Brockwell 03, Berry 02]
- Only simple objectives

Contribution and scope

Contribution:

Develop a mathematical framework and numerical tools to find optimal sequential experimental designs in a computationally feasible manner

Scope:

- Finite number of experiments
- Nonlinear and physically realistic models
- Continuous parameter, design, and data spaces of multiple dimensions
- Bayesian treatment of uncertainty
- Non-Gaussian distributions
- Information measure objective (design for parameter inference)

Outline

Core components of general sequential design formulation

Experiment: $k = 0, \ldots, N-1$, total N experiments; $N < \infty$

State: $x_k = [x_{k,b},\, x_{k,p}]$ all information needed for optimal future designs

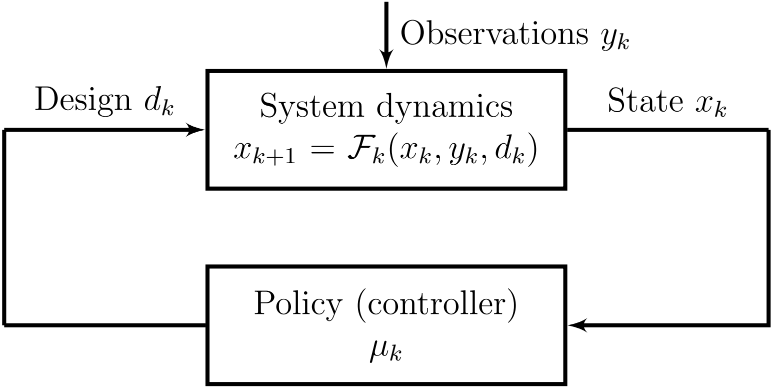
- *Belief state*: $x_{k,b}$ current state of uncertainty
- *Physical state*: $x_{k,p}$ deterministic design-relevant variables

Design: $d_k = \mu_k(x_k)$
seek good *policy* $\pi \equiv \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$

Observations: y_k distributed according to likelihood $f(y_k|\theta, d_k)$
(e.g., $y_k = G(\theta, d_k) + \epsilon$, with ϵ Gaussian)

System dynamics: $x_{k+1} = \mathcal{F}_k(x_k, y_k, d_k)$ state evolution

Sequential design exhibits a closed-loop behavior



The sOED problem: find optimal policy that maximizes the expected total reward

Stage reward: $g_k(x_k, y_k, d_k)$ **Terminal reward:** $g_N(x_N)$

The sequential optimal experimental design (sOED) problem:

Find π^* where

$$\pi^* = \underset{\pi=\{\mu_0,\ldots,\mu_{N-1}\}}{\operatorname{argmax}} \mathbb{E}_{y_0,\ldots,y_{N-1}|\pi}\left[\sum_{k=0}^{N-1} g_k\left(x_k, y_k, \mu_k(x_k)\right) + g_N\left(x_N\right)\right]$$

$$\begin{aligned} s.t. \qquad & x_{k+1} = \mathcal{F}_k(x_k, y_k, d_k), \forall k \\ & \mu_k(x_k) \in \mathcal{D}_k, \forall x_k, k \end{aligned}$$

Difficult to solve directly, involves optimization of a functional

The sOED problem in dynamic programming (DP) form

Re-express using Bellman's Principle of Optimality [Bellman 53]

Dynamic programming form (Bellman equations): (e.g., [Bertsekas 05])

$$J_k(x_k) = \max_{d_k \in \mathcal{D}_k} \mathbb{E}_{y_k|x_k,d_k}\left[g_k(x_k, d_k, y_k) + J_{k+1}\left(\mathcal{F}_k(x_k, d_k, y_k)\right)\right]$$

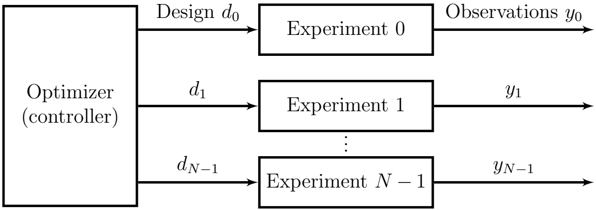
$$J_N(x_N) = g_N(x_N)$$

$k = 0, \ldots, N-1$; $J_k(x_k)$ are value functions

- A set of smaller tail subproblems
- Optimal policy functions implicitly in argmax: $d_k^* = \mu_k^*(x_k)$
- "Curse of dimensionality": exponential scenario growth from recursion
- Large body of approximate methods: *approximate dynamic programming* (e.g., [Bertsekas 96, Kaelbling 96, Sutton 98, Powell 11])

Batch (non-sequential) design is a special case of the sOED problem, and thus suboptimal

- Has no feedback
- Designs all experiments concurrently as a batch
- Finds optimal *designs* (vectors) rather than a policy



$$\left\{d_0^*, \ldots, d_{N-1}^*\right\} = \underset{d_0,\ldots,d_{N-1}}{\operatorname{argmax}} \mathbb{E}_{y_0,\ldots,y_{N-1}|d_{0:N-1}} \left[\sum_{k=0}^{N-1} g_k\left(x_k, y_k, d_k\right) + g_N\left(x_N\right) \right]$$

Greedy (myopic) design is a special case of the sOED problem (DP form), and thus suboptimal

- Uses feedback
- Considers the *next experiment* only
- Has no future effects

$$\begin{aligned} J_k(x_k) &= \max_{d_k \in \mathcal{D}_k} \mathbb{E}_{y_k|x_k,d_k} \left[g_k(x_k, y_k, d_k) + \cancel{J_{k+1}\left(\mathcal{F}_k(x_k, d_k, y_k)\right)} \right] \\ J_N(x_N) &= g_N(x_N) \end{aligned}$$

subject to $x_{k+1} = \mathcal{F}_k(x_k, y_k, d_k)$

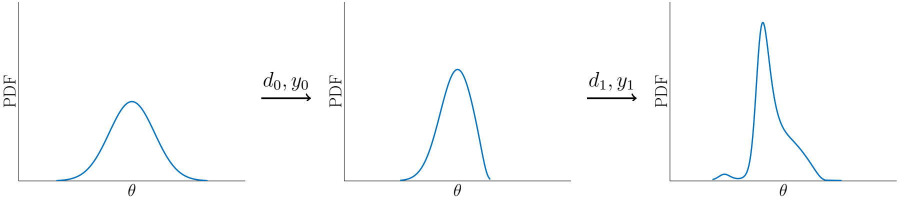
Sequential Bayesian inference

For the k-th experiment:

$$\overbrace{f(\theta|y_k, d_k, I_k)}^{\textit{posterior}} = \frac{\overbrace{f(y_k|\theta, d_k, I_k)}^{\textit{likelihood}}\overbrace{f(\theta|I_k)}^{\textit{prior}}}{\underbrace{f(y_k|d_k, I_k)}_{\textit{evidence}}}$$

θ — parameters to infer
I_k — information from previous experiments, $I_k \equiv \{d_0, y_0, \dots, d_{k-1}, y_{k-1}\}$

Conceptually: belief state is posterior random variable $x_{k,b} = \theta | I_k$



Information gain objective for parameter inference

We choose to use total information gain at end of all experiments (Kullback-Leibler (KL) divergence from final posterior to prior)

$$\begin{aligned} g_k(x_k, d_k, y_k) &= \text{reflects experimental cost} \\ g_N(x_N) &= D_{KL}(f(x_{N,b})||f(x_{0,b})) = \int_{\mathcal{H}} f(x_{N,b}) \ln \left[\frac{f(x_{N,b})}{f(x_{0,b})} \right] d\theta \end{aligned}$$

Corresponding system dynamics:

- *Belief state*: Bayes' Theorem

$$f(x_{k+1,b}) = \frac{f(y_k|\theta, d_k, I_k) f(x_{k,b})}{f(y_k|d_k, I_k)}$$

- *Physical state*: physical process

Outline

1. Sequential Optimal Experimental Design
   - Formulation
   - Numerical Methods
   - Results
2. Uncertainty Propagation in Scramjet Computations
3. Summary and Future Work

Represent a policy using one-step lookahead form

One-step lookahead policy representation: (e.g., [Bertsekas 05])

$$\mu_k(x_k) = \underset{d_k \in \mathcal{D}_k}{\operatorname{argmax}} \, \mathbb{E}_{y_k|x_k,d_k} \left[g_k(x_k, y_k, d_k) + \tilde{J}_{k+1}\left(\mathcal{F}_k\left(x_k, y_k, d_k\right)\right) \right]$$

Approximate value functions using **linear architecture**:

$$\tilde{J}_k(x_k) = r_k^\top \phi_k(x_k)$$

ϕ_k features (selected from heuristics), r_k weights

Approximate value iteration (backward induction with regression):

$$\tilde{J}_k(x_k) = \mathcal{P}\left\{ \max_{d_k \in \mathcal{D}_k} \mathbb{E}_{y_k|x_k,d_k}[g_k(x_k, d_k, y_k) + \tilde{J}_{k+1}\left(\mathcal{F}_k(x_k, d_k, y_k)\right)] \right\}$$

Start with $\tilde{J}_N(x_N) \equiv g_N(x_N)$, and proceed backwards $k = N-1, \ldots, 1$

$\mathcal{P}$: regression operator, samples from **exploration** and **exploitation**

Belief state representation

Conceptually: belief state is posterior random variable

How to numerically represent it . . .

- for general non-Gaussian continuous random variables
- in a finite-dimensional manner
- to easily perform Bayesian inference repeatedly in evaluating

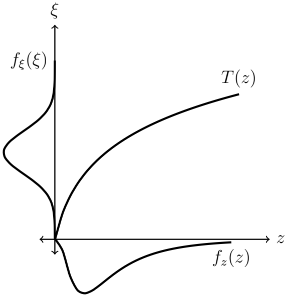
$$\tilde{J}_k(x_k) = \mathcal{P}\left\{ \max_{d_k \in \mathcal{D}_k} \mathbb{E}_{y_k|x_k,d_k} \left[g_k(x_k, d_k, y_k) + \tilde{J}_{k+1}\left(\mathcal{F}_k(x_k, d_k, y_k)\right) \right] \right\}$$

Traditional approaches:

- Gaussian approximation and model linearization
- Gridding or functional approximation of its PDF or CDF
- Non-parametrics (with particle filter, MCMC)

Often expensive and some do not scale well to multiple dimensions. We seek an approach that can quickly perform **many** Bayesian inferences.

Transport map transforms *distributions* (e.g., [Villani 08])



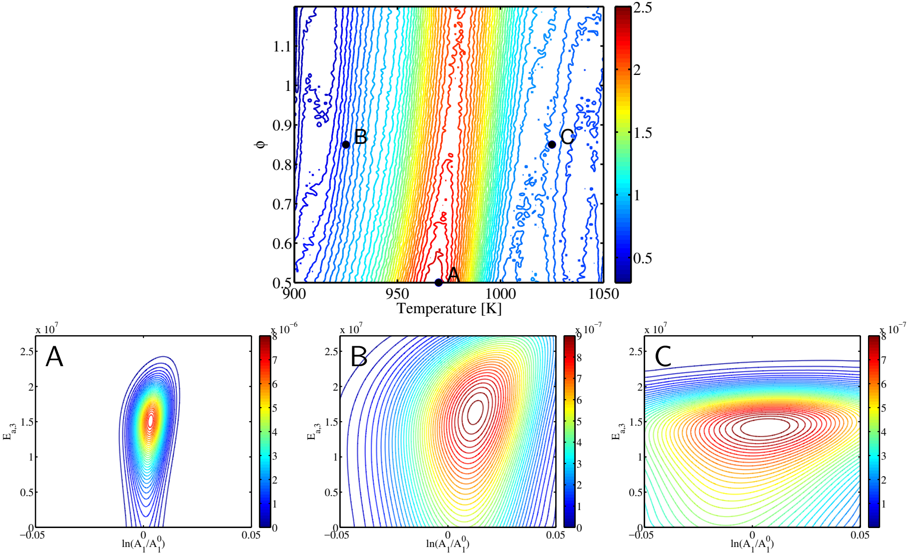
- $\xi \sim$ reference distribution, $z \sim$ target distribution
- Equivalence in distribution $\xi \overset{i.d.}{=} T(z)$
- Knothe-Rosenblatt (KR) maps: defined by conditional distributions, is triangular and monotone, exists and is unique [Rosenblatt 52, Carlier 10]
- Easy to construct from samples: convex optimization problem
- Target *joint* distribution for fast approx Bayesian inference

Final algorithm for sOED

1. **Set parameters**
2. **Initial exploration**
3. **Make joint map**
4. Iterate to refine . . .
   - (a) **Exploration**
   - (b) **Exploitation**
   - (c) **Approximate value iteration**
5. **Extract final policy parameterization**

Outline

Shock tube experiments for combustion kinetics



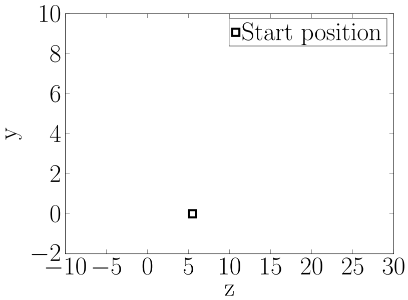
1D source inversion problem: problem settings

$$y_k = \frac{s}{\sqrt{2\pi}2\sqrt{(0.3 + Dt)}} \exp\left(-\frac{\| \theta + d_w(t) - z_{k+1} \|^2}{2(4)\,(0.3 + Dt)}\right) + \epsilon_k$$

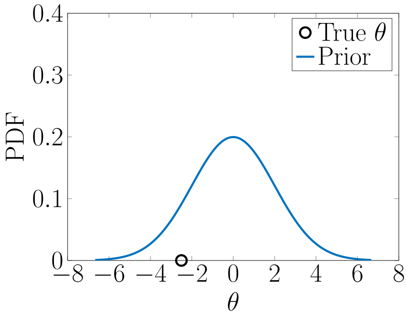
- 2 experiments
- $\theta \sim \mathcal{N}(0, 2^2)$ starting location: 5.5
- Strong wind blows to the right after first experiment
- Quadratic movement penalty

1D source inversion problem: state and plume evolution

An example scenario:



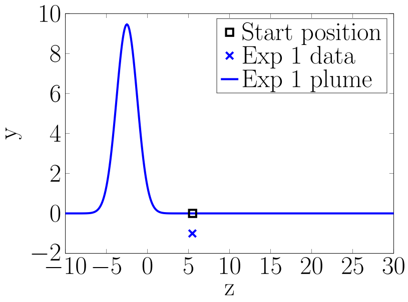
physical state and plume



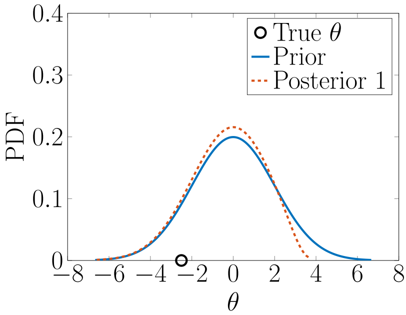
belief state

1D source inversion problem: state and plume evolution

An example scenario:



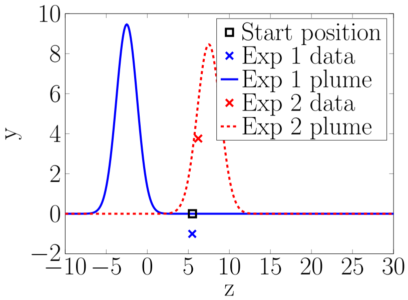
physical state and plume



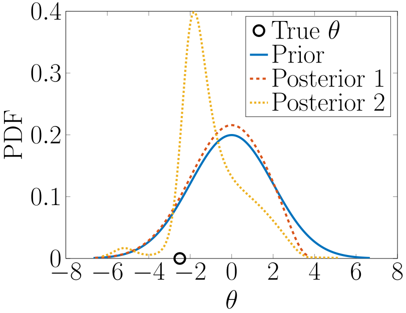
belief state

1D source inversion problem: state and plume evolution

An example scenario:



physical state and plume

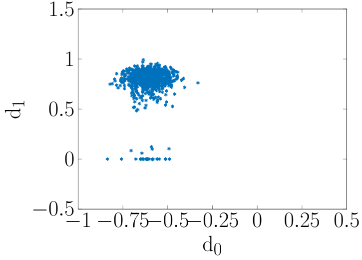


belief state

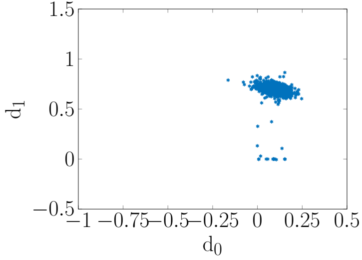
1D source inversion problem: case 1 advantages of sOED over greedy design

Greedy design does not account for future wind conditions

Expected reward: greedy (0.07), sOED (0.15)



greedy design



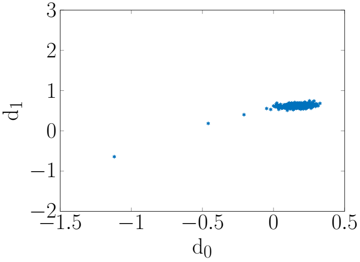
sOED

1D source inversion problem: case 2 advantages of sOED over batch design

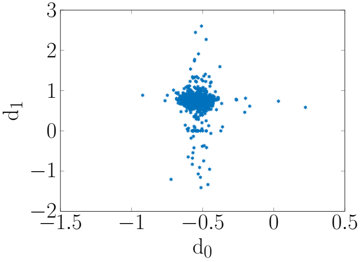
A more precise instrument available only if prior variance < 3

Batch design does not have feedback

Expected reward: batch (0.15), sOED (0.26)



batch design

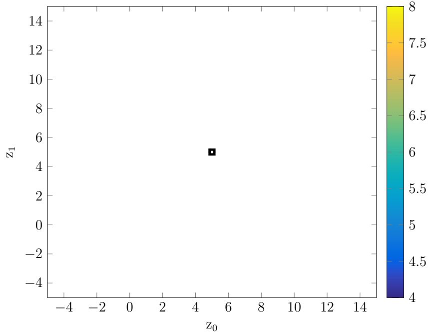


sOED

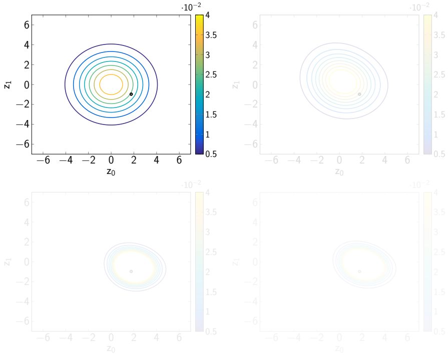
2D source inversion problem: trajectory example #1

An example scenario:

physical state and plume



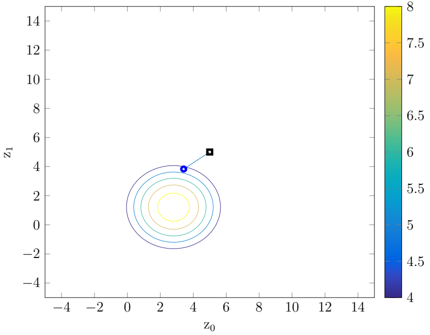
belief states



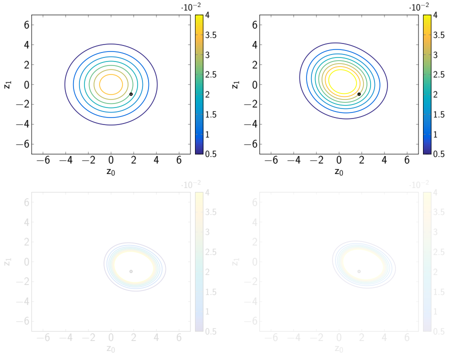
2D source inversion problem: trajectory example #1

An example scenario:

physical state and plume



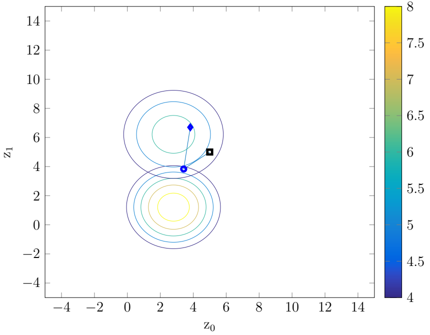
belief states



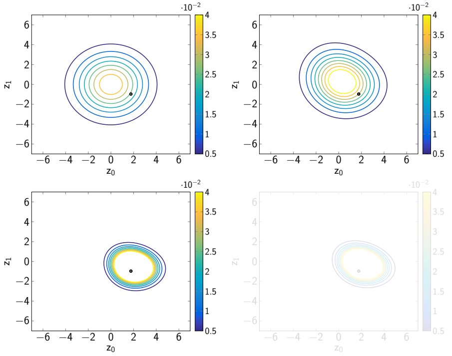
2D source inversion problem: trajectory example #1

An example scenario:

physical state and plume



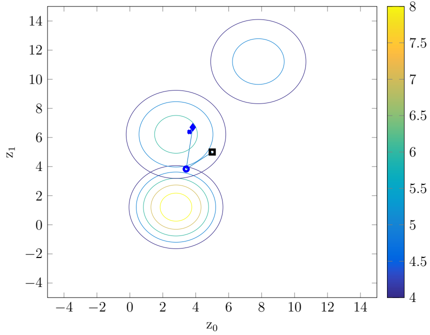
belief states



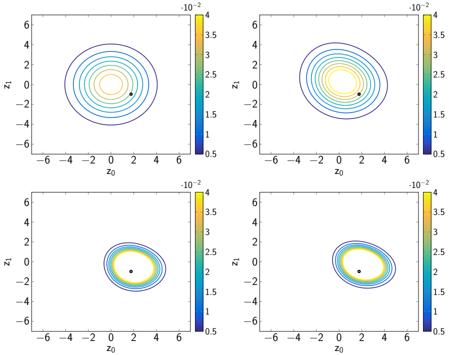
2D source inversion problem: trajectory example #1

An example scenario:

physical state and plume



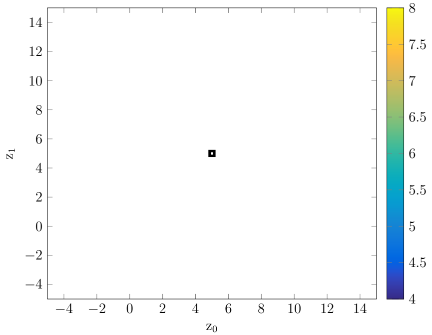
belief states



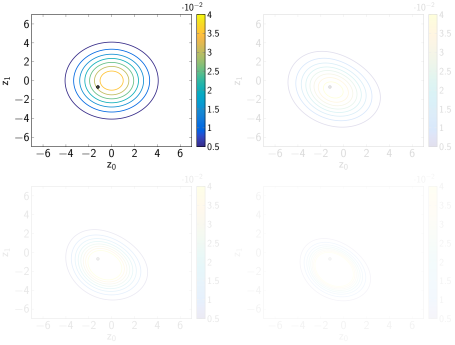
2D source inversion problem: trajectory example #2

Another example scenario:

physical state and plume



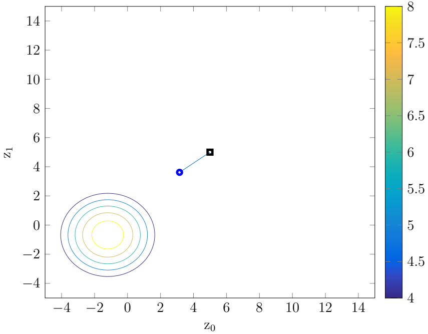
belief states



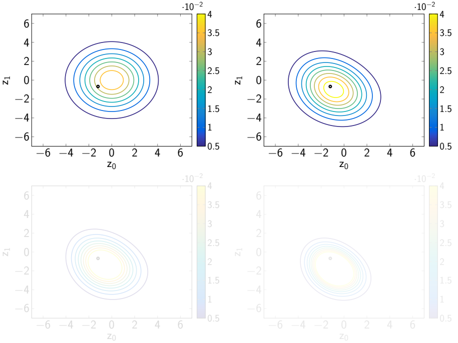
2D source inversion problem: trajectory example #2

Another example scenario:

physical state and plume



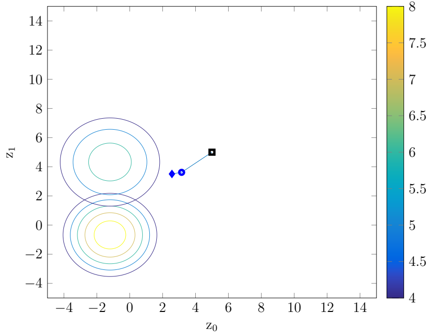
belief states



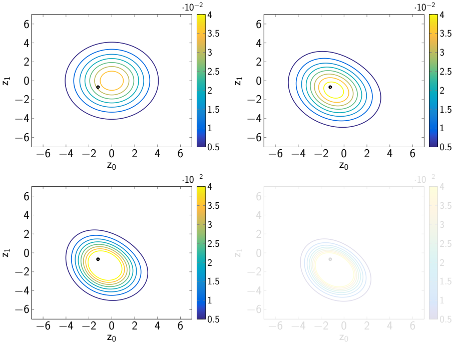
2D source inversion problem: trajectory example #2

Another example scenario:

physical state and plume



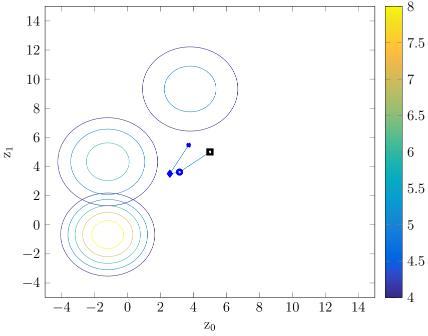
belief states



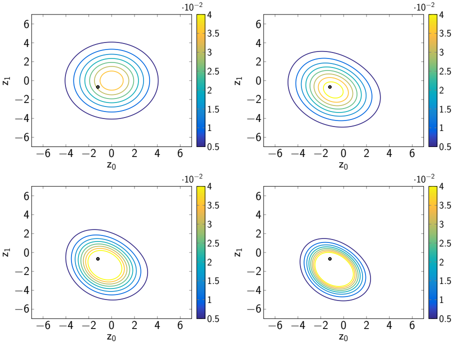
2D source inversion problem: trajectory example #2

Another example scenario:

physical state and plume



belief states



Summary

- Formulated the sequential optimal experimental design (sOED) problem rigorously (has 1. feedback, 2. forward looking)
- Developed new numerical methods to solve the sOED problem in a computationally-feasible manner, using
  - approximate dynamic programming
  - transport maps
- Demonstrated computational effectiveness on realistic applications

References:

1. Huan & Marzouk, "Simulation-based optimal Bayesian experimental design for nonlinear systems," *Journal of Computational Physics*, 232(1):288-317, 2013.
2. Huan & Marzouk, "Gradient-Based Stochastic Optimization Methods in Bayesian Experimental Design," *International Journal of Uncertainty Quantification*, 4(6):479-510, 2014.
3. Huan & Marzouk, "Sequential Bayesian Optimal Experimental Design via Approximate Dynamic Programming," *arXiv: 1604.08320*, 2016.

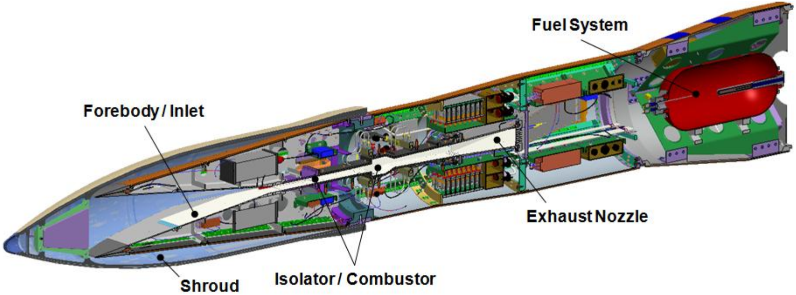
Outline

HIFiRE-II Scramjet

Development of scramjet[1] engine involves

- flow simulations
- uncertainty quantification (UQ)
- design optimization

We focus on the HIFiRE-II[2] configuration:



[1] supersonic combusting ramjet

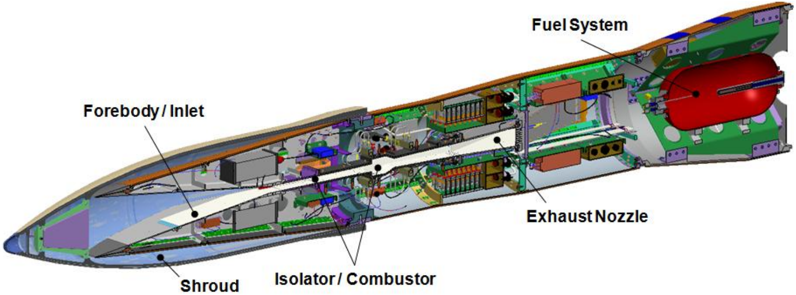
[2] Hypersonic International Flight Research and Experimentation-II

HIFiRE-II Scramjet

Development of scramjet[1] engine involves

- flow simulations
- uncertainty quantification (UQ)
- design optimization

We focus on the HIFiRE-II[2] configuration:

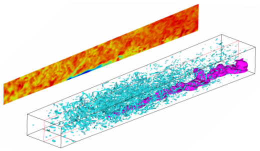


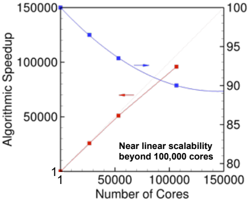
[1] supersonic combusting ramjet

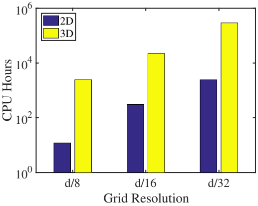
[2] Hypersonic International Flight Research and Experimentation-II

Direct UQ would be intractable

RAPTOR: LES solver by Oefelein *et al.* at Sandia [Oefelein 06]







Highly-scalable but still **very expensive**

Major challenges: for uncertainty quantification

- Many uncertain parameters (high stochastic dimension)
- Expensive simulations

$\Rightarrow$ direct exploration of parameter space **intractable!**

Global sensitivity analysis: Sobol indices

Global sensitivity analysis (GSA) [Saltelli 04, Saltelli 08]

- For a given quantify of interest (QoI) ...
- Variance of QoI decomposed into contributions from each parameter
- Sobol indices rank parameters by their contributions [Sobol 03]

$$\text{Total effect} \qquad S_{T_i} = \frac{\mathbb{E}_{\lambda_{\sim i}}\left[\mathrm{Var}_{\lambda_i}\left(f(\lambda)|\lambda_i\right)\right]}{\mathrm{Var}\left(f(\lambda)\right)}$$

S_{T_i} small $\Rightarrow$ low impact parameter $\Rightarrow$ fix value (i.e. dim. eliminated)

How to compute?

- Monte Carlo estimators [Saltelli 02, Saltelli 10] still prohibitive for LES
- **Our plan:** construct affordable surrogate models via polynomial chaos expansion (PCE)

Polynomial chaos expansions

A QoI (output) random variable can be expanded as follows:

$$f(\lambda(\xi)) = \sum_{\beta \in \mathcal{J}} c_\beta \Psi_\beta(\xi)$$

- c_β: PCE coefficients
- ξ: reference random vector (e.g., uniform, Gaussian)
- Ψ_β: multivariate orthonormal polynomial (e.g., Legendre, Hermite)
- β: multi-index, reflects order of polynomial basis

Orthonormality property
$\Rightarrow$ extract Sobol indices analytically from coefficients (no Monte Carlo!):

$$S_{T_i} = \frac{1}{\mathrm{Var}\,(f(\lambda))} \sum_{\beta \in \mathcal{J}: \beta_i > 0} c_\beta^2 \qquad \text{where} \qquad \mathrm{Var}\,(f(\lambda)) = \sum_{0 \neq \beta \in \mathcal{J}} c_\beta^2$$

Sparse polynomial chaos expansions

Non-intrusive regression to compute expansion coefficients $Gc = f$:

$$\underbrace{\begin{bmatrix} \Psi_{\beta^1}(\xi^{(1)}) & \cdots & \Psi_{\beta^N}(\xi^{(1)}) \\ \vdots & & \vdots \\ \Psi_{\beta^1}(\xi^{(M)}) & \cdots & \Psi_{\beta^N}(\xi^{(M)}) \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} c_{\beta^1} \\ \vdots \\ c_{\beta^N} \end{bmatrix}}_{c} = \underbrace{\begin{bmatrix} f(\lambda(\xi^{(1)})) \\ \vdots \\ f(\lambda(\xi^{(M)})) \end{bmatrix}}_{f},$$

Challenges:

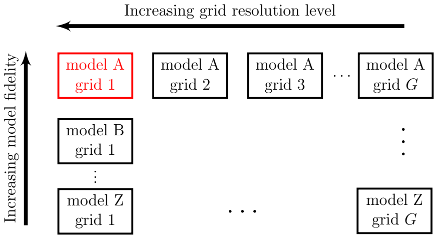
- Few LES flow solves (data), many PCE basis (e.g. total-order degree 3 in 24 dimensions: 2925 terms)
- Extremely under-determined system ($N \gg M$)

Our approach: use compressed sensing to find sparse solution (LASSO)

$$\min_c \frac{1}{2} \| c \|_1 + \tau \| Gc - f \|_2^2$$

discover and retain only basis terms with high magnitude coefficients

Multilevel and multifidelity forms



Telescopic sum:

$$f_L(\lambda) = f_0(\lambda) + \sum_{\ell=1}^{L} f_{\Delta_\ell}(\lambda)$$

- ℓ indicates different grid levels or fidelity of models
- Δ_ℓ indicates difference between models ℓ and $\ell - 1$

Function approximation:

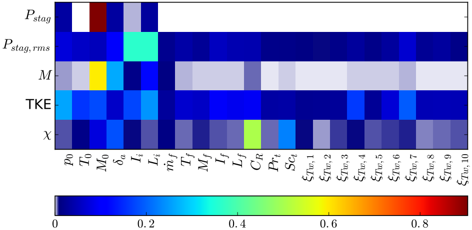
$$f_L(\lambda) \approx \hat{f}_L(\lambda) = \hat{f}_0(\lambda) + \sum_{\ell=1}^{L} \hat{f}_{\Delta_\ell}(\lambda)$$

Unit problem: 24 parameters

Parameter	Range	Description
	Inlet boundary conditions	
p_0	[1.406, 1.554] MPa	Stagnation pressure
T_0	[1472.5, 1627.5] K	Stagnation temperature
M_0	[2.259, 2.761]	Mach number
δ_a	[2, 6] mm	Boundary layer thickness
I_i	[0, 0.05]	Turbulence intensity magnitude
L_i	[0, 8] mm	Turbulence length scale
	Fuel inflow boundary conditions	
$\dot{m}_f$	[6.633, 8.107] $\times 10^{-3}$ kg/s	Mass flux
T_f	[285, 315] K	Static temperature
M_f	[0.95, 1.05]	Mach number
I_f	[0, 0.05]	Turbulence intensity magnitude
L_f	[0, 1] mm	Turbulence length scale
	Turbulence model parameters	
C_R	[0.01, 0.06]	Modified Smagorinsky constant
Pr_t	[0.5, 1.7]	Turbulent Prandtl number
Sc_t	[0.5, 1.7]	Turbulent Schmidt number
	Wall boundary conditions	
T_w	Expansion in 10 params of $\mathcal{N}(0,1)$	Wall temperature represented via Karhunen-Loève expansion

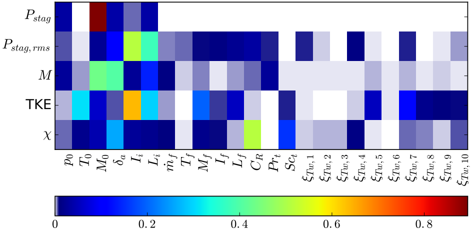
- 2D runs: 1939 (coarse grid), 79 (fine grid)
- 3D runs: 46 (coarse grid), 11 (fine grid)

Unit problem: total sensitivity



Multilevel expansion of:

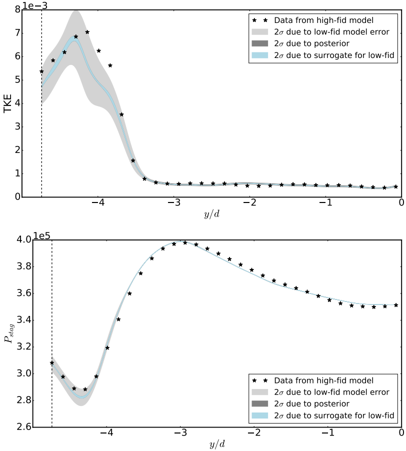
$$\hat{f}_{2D,d/16} = \hat{f}_{2D,d/8} + \hat{f}_{\Delta_{2D,d/16-2D,d/8}}$$



Multifidelity expansion of:

$$\hat{f}_{3D,d/8} = \hat{f}_{2D,d/8} + \hat{f}_{\Delta_{3D,d/8-2D,d/8}}$$

Posterior predictives in model QoIs

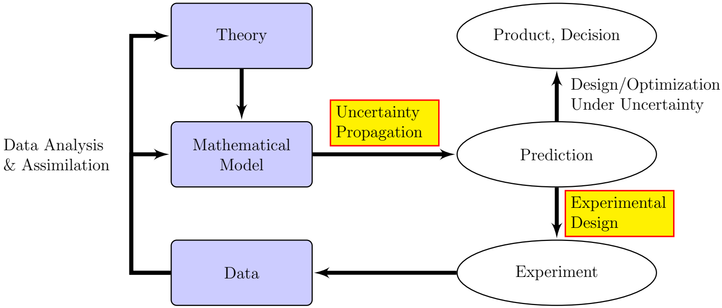


Reference:

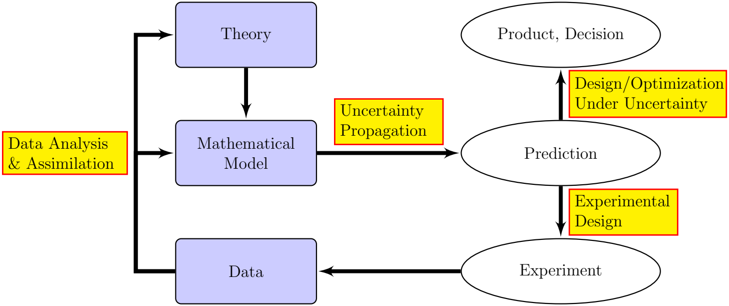
X. Huan, C. Safta, K. Sargsyan, G. Geraci, M. S. Eldred, Z. P. Vane, G. Lacaze, J. C. Oefelein, and H. N. Najm, "Global Sensitivity Analysis and Quantification of Model Error for Large Eddy Simulation in Scramjet Design," *AIAA paper 2017-1089*, 2017.

Outline

Summary and vision



Summary and vision



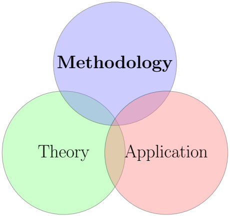
Vision: develop comprehensive UQ capability to

1. design experiments for acquiring data
2. analyze and assimilate data for improving model and theory
3. propagate uncertainty for making predictions
4. optimize and design in the presence of uncertainty

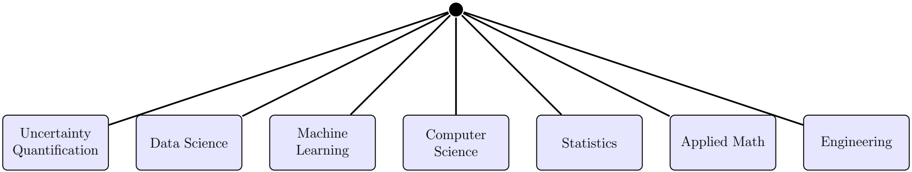
for applications with complex and realistic physics-based models

Summary and vision

Focus on **methodology** development



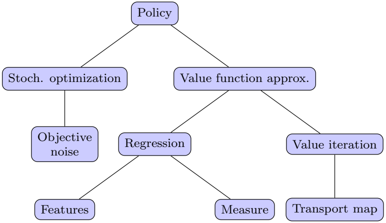
Naturally **interdisciplinary**



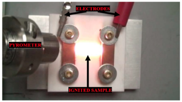
Many **collaboration** opportunities

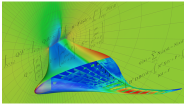
Ideas for future work

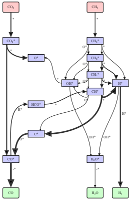
Adaptive numerical methods



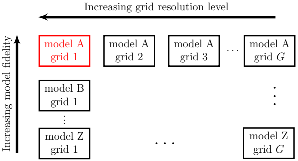
UQ for expensive, high-dimensional models



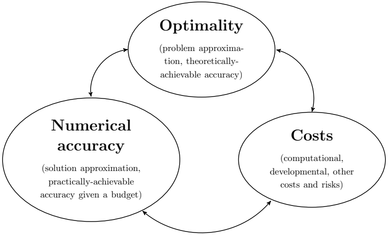




Multi-model management for experimental design and statistical inference



Hybrid frameworks for experimental design



Student education program

Strong foundational courses from multiple departments, for example:

- V&V, machine learning, data science, data mining, statistical methods in data mining, applied Bayesian statistics, statistical inference, optimization, stochastic analysis, stochastic control theory, etc.

New course ideas:

- uncertainty quantification
- statistical data analysis
- optimal experimental design / decision-making under uncertainty
- model reduction
- optimization under uncertainty
- inverse problems

Acknowledgment

- Dr. Habib Najm, Prof. Youssef Marzouk
- Sandia National Laboratories[3]
- Defense Advanced Research Projects Agency (DARPA)
  Enabling Quantification of Uncertainty in Physical Systems (EQUiPS) Program
- National Science Foundation (NSF)
- Air Force Office of Scientific Research (AFOSR)
  Computational Mathematics Program
- U.S. Department of Energy
  Office of Science, Office of Advanced Scientific Computing Research (ASCR)
- KAUST Global Research Partnership
- Natural Sciences and Engineering Research Council of Canada (NSERC)

[3]Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Thank You!

Supplement

Supplemental Slides

Optimal Experimental Design

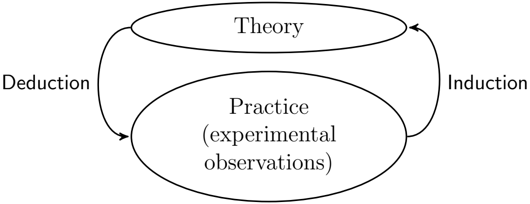
open-loop design

- classical linear design theory [Atkinson & Donev (1992)]
- Bayesian experimental design [Chaloner & Verdinelli (1995)]
- nonlinear models [Box & Lucas (1959), Lindley (1956, 1972), Sebastiani & Wynn (1997, 2000), Ryan (2003), Müller & Parmigiani (1993, 1995), Clyde *et al.* (1995), Loredo (2010)]

closed-loop design

- greedy (rolling-horizon of 1 experiment) [Gautier & Pronzato (1998), Negoescu *et al.* (2011), Solonen *et al.* (2012)]
- dynamic programming approach [Müller (2006), Brockwell & Kadane (2003), Lewis & Berry (1994), Christen & Nakamura (2003)]

Experiments are an integral part of learning



"...science is a means whereby learning is achieved, not by mere theoretical speculation on the one hand, nor by the undirected accumulation of practical facts on the other, but rather by a motivated iteration between **theory** *and* **practice** *..."* [Box 76]

— *George E. P. Box*

Inference can be done by conditioning the joint map

Bayes' theorem: posterior is simply conditioning the joint distribution

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} = \frac{f(y,\theta)}{f(y)}$$

For one experiment: KR map from (d, y, θ) to $\xi \sim \mathcal{N}(0, I)$ is:

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = \begin{bmatrix} T_d(d) \\ T_{y|d}(d, y) \\ T_{\theta|y,d}(d, y, \theta) \end{bmatrix} = \begin{bmatrix} \Phi^{-1}\left(F(d)\right) \\ \Phi^{-1}\left(F(y|d)\right) \\ \Phi^{-1}\left(F(\theta|y, d)\right) \end{bmatrix}$$

KR map of posterior given realizations d^* and y^* is:

$$T_{\theta|y^*,d^*}(\theta) = \Phi^{-1}\left(F(\theta|y^*, d^*)\right)$$

This is precisely $T_{\theta|y,d}$ conditioned on d^* and y^*: $T_{\theta|y,d}(d^*, y^*, \theta)$!

Map construction [Parno 15]

- Approximate using linear parameterization $\xi = T(z;\gamma)$ (we use a linear architecture of monomial polynomial basis functions)
- Construct $T(z;\gamma)$ by minimizing KL using M samples from target

For i-th dimension of the multivariate map:

$$\min_{\gamma_i} \sum_{m=1}^{M} \left[0.5\, T_i^2(z^{(m)};\gamma_i) - \ln \left. \frac{\partial T_i(z;\gamma_i)}{\partial z_i} \right|_{z^{(m)}} \right]$$

$$s.t. \qquad \left. \frac{\partial T_i(z;\gamma_i)}{\partial z_i} \right|_{z^{(m)}} \geq \lambda_{min} > 0, m = 1, \ldots, M$$

Numerically attractive properties:

- Dimensions are separable
- Convex optimization
- Model-independent (only uses samples)

A good joint map also implies good posterior maps

Question: are the resulting posterior maps accurate?

Theorem: ([Huan 15])

- If we construct a joint map that is *optimal* in an *approximation* subspace
- then the resulting posterior maps from conditioning on d_k and y_k are also optimal *on average*
- with respect to the probability measure of d_k and y_k

A good joint map also implies good posterior maps

Theorem

Let the optimal joint map be

$$T^*_{1:n} = \underset{T_{1:n}\in\mathcal{T}_{1:n}}{\operatorname{argmin}}\, D_{KL}\left(f_{1:n}(\cdot)||\tilde{f}_{1:n}(\cdot;T_{1:n})\right)$$

where $\tilde{f}_{1:n}$ is the target density induced by candidate map $T_{1:n}$. Then, for any $0 < j \leq n$, the dimension-truncated "head" map is also the optimal map for dimensions $(1:j)$, in the sense

$$T^*_{1:j} = \underset{T_{1:j}\in\mathcal{T}_{1:j}}{\operatorname{argmin}}\, D_{KL}\left(f_{1:j}(\cdot)||\tilde{f}_{1:j}(\cdot;T_{1:j})\right),$$

where $\mathcal{T}_{1:j} \subseteq \mathcal{T}_{1:n}$ is its first j-dimensional truncation.

Corollary

For each $j = 1, \ldots, n$, the component map is optimal in an expected sense:

$$T^*_j = \underset{T_j\in\mathcal{T}_j}{\operatorname{argmin}}\, \mathbb{E}_{z_{1:(j-1)}}\left[D_{KL}\left(f_{j|1:(j-1)}(\cdot|z_{1:(j-1)})||\tilde{f}_{j|1:(j-1)}(\cdot|z_{1:(j-1)};T_j)\right)\right].$$

Extension to multiple experiments

$k=0$	1	2	$\cdots$
$\xi_{d_0} = T_{d_0}(d_0)$	$\xi_{d_0} = T_{d_0}(d_0)$	$\xi_{d_0} = T_{d_0}(d_0)$	
$\xi_{y_0} = T_{y_0}(d_0, y_0)$	$\xi_{y_0} = T_{y_0}(d_0, y_0)$	$\xi_{y_0} = T_{y_0}(d_0, y_0)$	
$\xi_{\theta_0} = T_{\theta_0}(d_0, y_0, \theta)$	$\xi_{d_1} = T_{d_1}(d_0, y_0, d_1)$	$\xi_{d_1} = T_{d_1}(d_0, y_0, d_1)$	
	$\xi_{y_1} = T_{y_1}(d_0, y_0, d_1, y_1)$	$\xi_{y_1} = T_{y_1}(d_0, y_0, d_1, y_1)$	
	$\xi_{\theta_1} = T_{\theta_1}(d_0, y_0, d_1, y_1, \theta)$	$\xi_{d_2} = T_{d_2}(d_0, y_0, d_1, y_1, d_2)$	
		$\xi_{y_2} = T_{y_2}(d_0, y_0, d_1, y_1, d_2, y_2)$	
		$\xi_{\theta_2} = T_{\theta_2}(d_0, y_0, d_1, y_1, d_2, y_2, \theta)$	

- Posterior map after Bayesian inference on $k+1$ experiments is the n_θ-dimensional $T_{\theta_k|d_0^*,y_0^*,\ldots,d_k^*,y_k^*}(\theta)$
- Components grouped by the red rectangular boxes are identical; concatenate unique parts and construct overall map in one shot
- Map constructed using samples of trajectory simulation (exploration and exploitation)

Final algorithm for sOED

1. **Set parameters:** Select VFA features, exploration measure, L, R_0, R, T
2. **Initial exploration:** Simulate R_0 exploration trajectories, without inference
3. **Make exploration joint map:** Make T_{explore} from these samples
4. For $\ell = 1, \ldots, L$
   (a) **Exploration:** Simulate R exploration trajectories, with inference using T_{explore}, store states visited $\mathcal{X}^{\ell}_{k,\text{explore}} = \{x_k^r\}_{r=1}^R$
   (b) **Exploitation:** If $\ell > 1$, simulate T exploitation trajectories by evaluating $d_k = \text{argmax}_{d_k'} \mathbb{E}_{y_k|x_k,d_k'} \left[g_k(x_k, y_k, d_k') + \tilde{J}^{\ell-1}_{k+1}(\mathcal{F}_k(x_k, y_k, d_k')) \right]$, with inference using T_{explore}, store states visited $\mathcal{X}^{\ell}_{k,\text{exploit}} = \{x_k^t\}_{t=1}^T$
   (c) **Approximate value iteration:** Update $\tilde{J}^{\ell}_k$ functions from new regression points $x_k^{rt} \in \{\mathcal{X}^{\ell}_{k,\text{explore}} \cup \mathcal{X}^{\ell}_{k,\text{exploit}}\}$ by evaluating $t_k^{rt}(x_k^{rt}) = \max_{d_k'} \mathbb{E}_{y_k|x_k,d_k'} \left[g_k(x_k^{rt}, y_k, d_k') + \tilde{J}^{\ell}_{k+1}(\mathcal{F}_k(x_k^{rt}, y_k, d_k')) \right]$ or terminal reward at all regression points, with inference using T_{explore}
5. **Extract final policy parameterization:** $\tilde{J}^L_k, \forall k$

Linear-Gaussian problem: problem settings

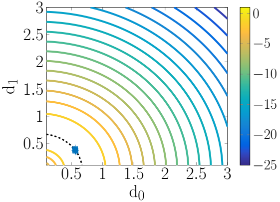
$$y_k = \theta d_k + \epsilon_k$$

- $\theta \sim \mathcal{N}(0, 3^2)$ $\quad d_k \in [0.1, 3]$
  $\epsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0, 1^2)$ $\quad g_k = 0$
- Conjugate family, all posteriors are Gaussian
- $N = 2$ experiments
- Additional terminal reward component (target variance)

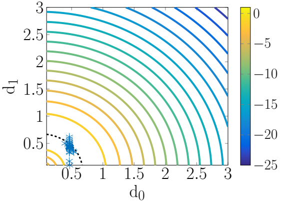
$$g_N = D_{KL}(f(x_{N,b})||f(x_{0,b})) - 2(\ln \sigma_N^2 - \ln 2)^2$$

- Compare state representations:
  - **analytic** (mean and variance)
  - PDF on a **grid** (50 grid points)
  - **map** (total order polynomial degree 3, made with 10^5 samples)

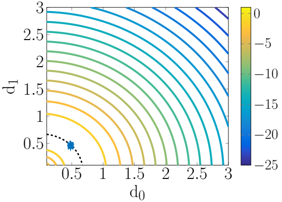
Linear-Gaussian problem: d_0, d_1 on exact expected utility



analytic



grid



map

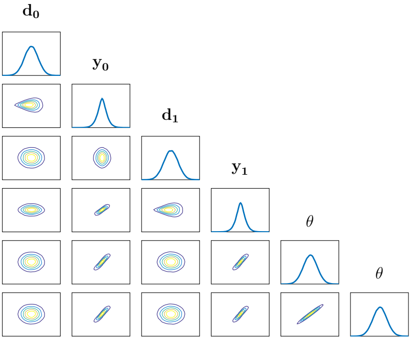
Linear-Gaussian problem: expected total rewards

Expected rewards from 1000 simulated trajectories:

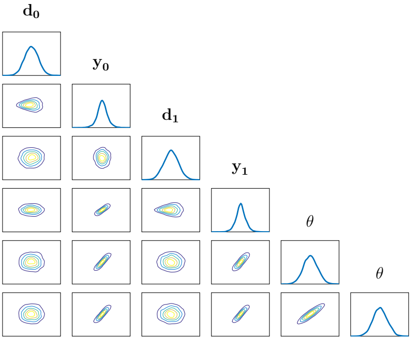
exact	analytic	grid	map
0.7833	0.79	0.76	0.79

- All values have standard error ± 0.02
- Optimal policy for linear-Gaussian problem with constant noise variance is state-independent and non-unique ("exchangeability" between d_0 and d_1)
- Excellent agreement between analytic, grid, and map methods
- In comparison, exploration policy expected reward is -8.5

Linear-Gaussian problem: joint map



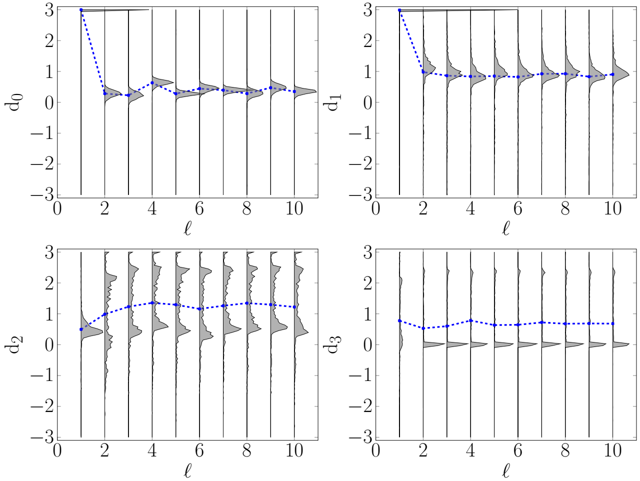
original samples



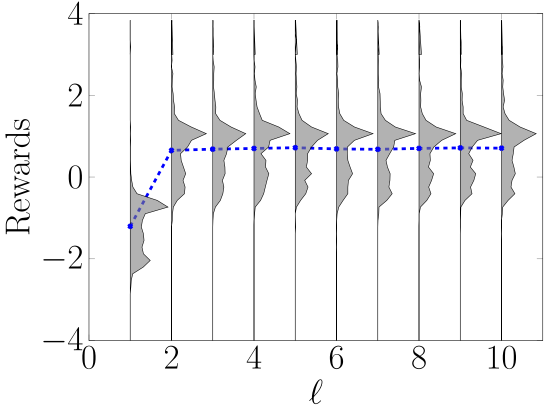
samples from map

Joint distribution even for a linear-Gaussian problem is generally not Gaussian!

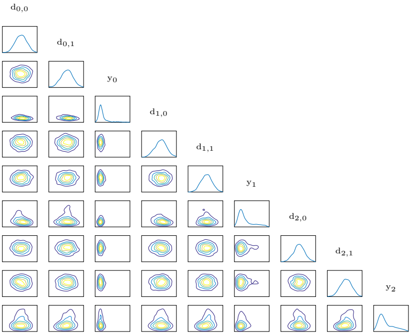
1D source inversion problem: effects of ℓ



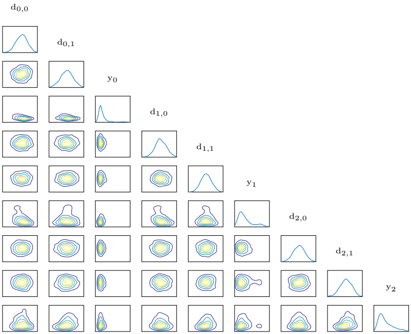
1D source inversion problem: effects of ℓ



2D source inversion problem: joint map (partial)



original samples



samples from map

Joint map does a good job in capturing some highly non-Gaussian, possibly multi-modal, behavior!

Overview

Challenges:

High number of uncertain model parameters (stochastic dimension), requiring many expensive flow solves

Would like to...

- use few runs
- incorporate less expensive simulations from low-fidelity models

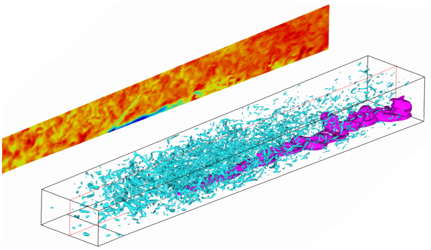
Objectives:

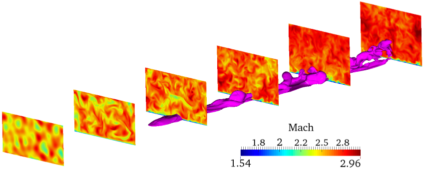
- identify influential uncertain parameters via global sensitivity analysis (GSA)
- characterize uncertainty due to model error resulted from using low-fidelity models

RAPTOR solver

RAPTOR: code by Oefelein *et al.* at Sandia [Oefelein 06]

- Fully-coupled, compressible conservation equations
- High Reynolds number, high-pressure, wide range of Mach number
- Real-fluid equation of state
- Detailed thermodynamics, transport and chemistry
- Non-dissipative, discretely conservative, staggered finite-volume
- Complex geometry treatment





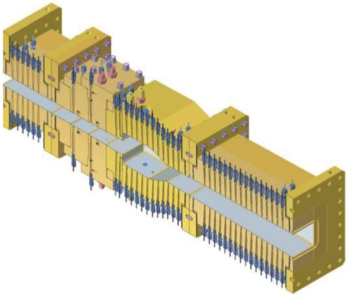
Fuel component—purple   Turbulence—blue   Mach number—cutting planes

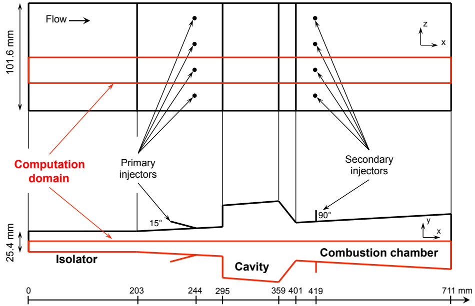
Computational domain of initial unit problem

Final: simulation of full combustor domain, match experimental setup —HIFiRE Direct Connect Rig (HDCR) [bottom-left figure]

Initial unit problem: primary injection section

- omit cavity
- no combustion
- focus on interaction of fuel jet and supersonic air crossflow



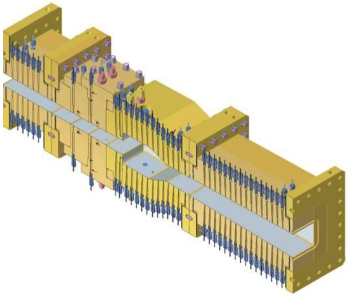


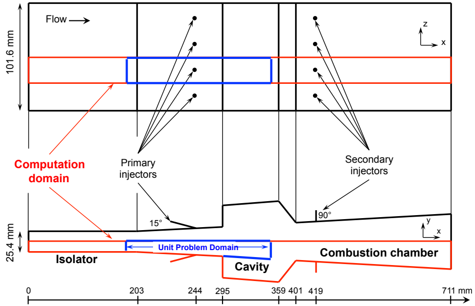
Computational domain of initial unit problem

Final: simulation of full combustor domain, match experimental setup —HIFiRE Direct Connect Rig (HDCR) [bottom-left figure]

Initial unit problem: primary injection section

- omit cavity
- no combustion
- focus on interaction of fuel jet and supersonic air crossflow





Embedded representation of model error

Motivation: different models are available
$\Rightarrow$ to use them, need to quantify the error due to model structure

Traditional additive form: [Kennedy 01]

$$q_i(s) = \rho_i f_i(s, \lambda) + \delta_i(s)$$

- Flexible for fitting model discrepancy
- Predictions do not obey governing equations
- Difficult to distinguish uncertainty contributions between model error and measurement noise
- $\delta_i(s)$ not transferable for prediction of QoIs outside calibration set

Embedding approach: [Sargsyan 15]

$$q_i(s) = f_i(s, \lambda + \delta_i(s, \alpha_i, \xi_i))$$

$\Rightarrow$ physically-meaningful predictions that satisfy governing equations

Bayesian calibration of model error term

Represent the discrepancy term δ using a PCE:

$$\lambda + \delta(\alpha, \xi) = \lambda + \sum_{\beta \neq 0} \alpha_\beta \Psi_\beta(\xi)$$

Calibrate by inferring all parameters $\tilde{\alpha} \equiv (\lambda, \alpha)$ via

Bayesian inference: $$\overbrace{p(\tilde{\alpha}|D)}^{posterior} \propto \overbrace{p(D|\tilde{\alpha})}^{likelihood} \overbrace{p(\tilde{\alpha})}^{prior}$$

Posterior explored via adaptive Markov chain Monte Carlo (MCMC)

Through use of surrogate model and likelihood approximation, we can **attribute predictive variance** to different sources:

$$\mathrm{Var}\left[q_k\right] = \underbrace{\mathbb{E}_{\tilde{\alpha}}\left[\sigma_k^2(\lambda, \alpha)\right]}_{\text{model error}} + \underbrace{\mathrm{Var}_{\tilde{\alpha}}\left[\mu_k(\lambda, \alpha)\right]}_{\text{posterior uncertainty}} + \underbrace{\sigma_{k,\mathrm{LOO}}^2}_{\text{surrogate error}}$$

Likelihood approximation, and breakdown of variance

MCMC requires likelihood evaluation $p(D|\tilde{\alpha})$, no analytical form

Enable tractable likelihood evaluation via two approximations:

1. Polynomial surrogate for QoIs, built using regression

$$q_k = f_k(\lambda + \delta(\alpha, \xi)) = \hat{f}_k(\lambda + \delta(\alpha, \xi)) + \epsilon_k$$

2. Gauss-marginal approximation to likelihood form

$$p(D|\tilde{\alpha}) \approx L_G(\tilde{\alpha}) = \frac{1}{(2\pi)^{\frac{N}{2}}} \prod_{k=1}^{N} \frac{1}{\sigma_k(\tilde{\alpha})} \exp\left[-\frac{(\mu_i(\tilde{\alpha}) - g_k)^2}{2\sigma_k^2(\tilde{\alpha})}\right]$$

$$\mu_k(\tilde{\alpha}) \approx \hat{f}_{k,0}(\tilde{\alpha}) \qquad \text{and} \qquad \sigma_k^2(\tilde{\alpha}) \approx \sum_{\beta \neq 0} \hat{f}_{k,\beta}^2(\tilde{\alpha})$$

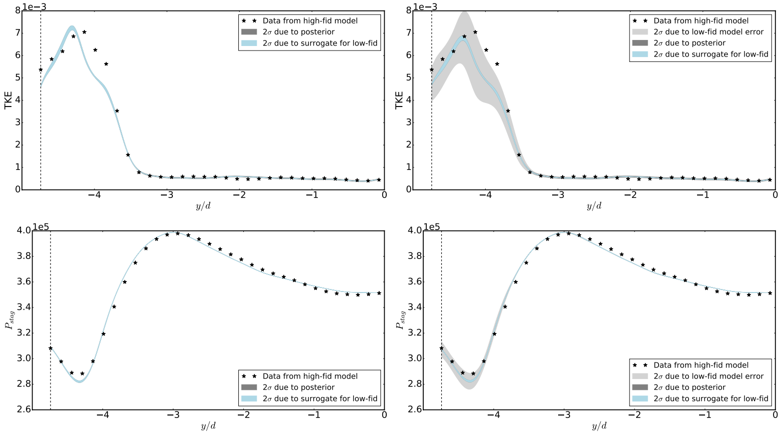
Also **enables attribution of predictive variance** to different sources:

$$\mathrm{Var}\left[q_k\right] = \underbrace{\mathbb{E}_{\tilde{\alpha}}\left[\sigma_k^2(\lambda, \alpha)\right]}_{\text{model error}} + \underbrace{\mathrm{Var}_{\tilde{\alpha}}\left[\mu_k(\lambda, \alpha)\right]}_{\text{posterior uncertainty}} + \underbrace{\sigma_{k,\mathrm{LOO}}^2}_{\text{surrogate error}}$$

Dynamic-vs-Static Smagorinsky turbulence model

Calibrate static Smagorinsky model with dynamic treatment simulations

- Calibrate using TKE profile, $\lambda = C_R$



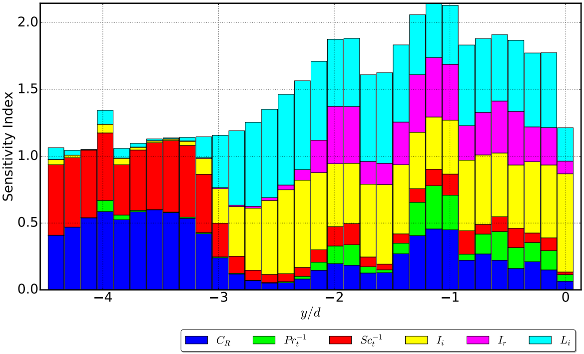
No model error treatment

Embedded model error treatment

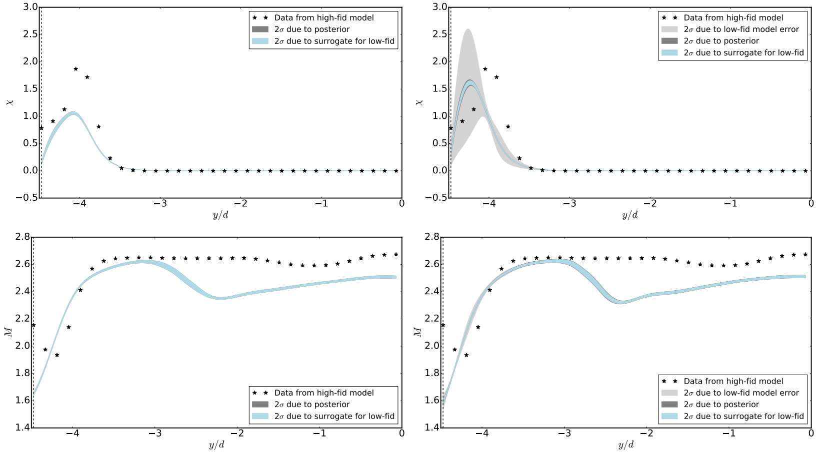
2D-vs-3D: choice of embedding parameters

Calibrate 2D model using 3D model simulations (using χ profile)

- $\lambda = (C_R, Pr_t^{-1}, Sc_t^{-1}, I_i, I_r, L_i)$
- Choose to embed δ (1st-order PCE) in C_R and Sc_t^{-1} based on GSA



2D-vs-3D: predictive quantities



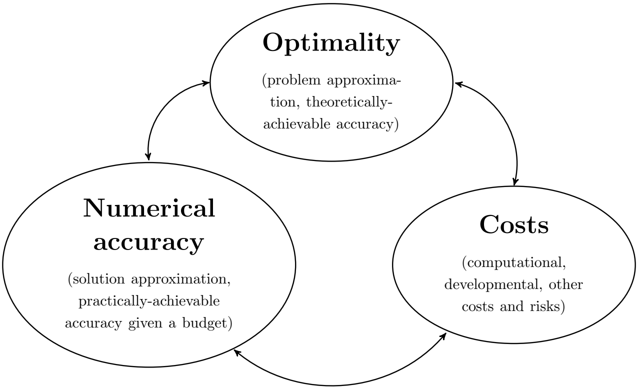
No model error treatment | Embedded model error treatment

Ideas for future work

- **Goal-oriented adaptive numerical methods:** automated refinements of component numerical approximations; quantifiable and meaningful error bounds
- **UQ for high-dimensional and expensive models:** model reduction and surrogate modeling, dimension reduction, sparse representations
- **Multi-model management in experimental design and inference:** leverage and combine existing models of different fidelities and resolutions, improve through selective experimental design and data acquisition
- **Model error with automated learning and improvement:** quantify and learn model error through experimental design, improve models through adding sub-components while preserving governing equations and physical principles
- **Practical experimental design formulation choices:** combine low-cost sub-optimal design frameworks via value of feedback and coordination

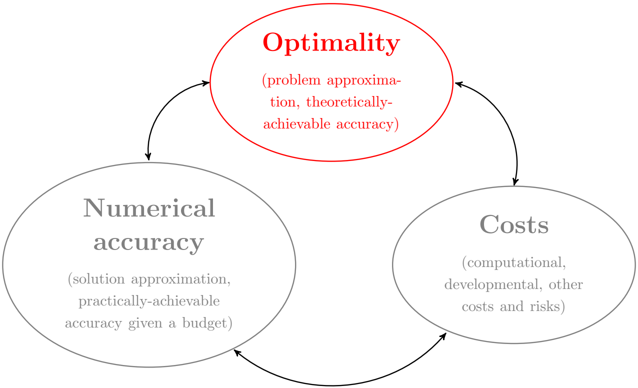
Future work: framework and methodology choices

What experimental design framework and methods to use?

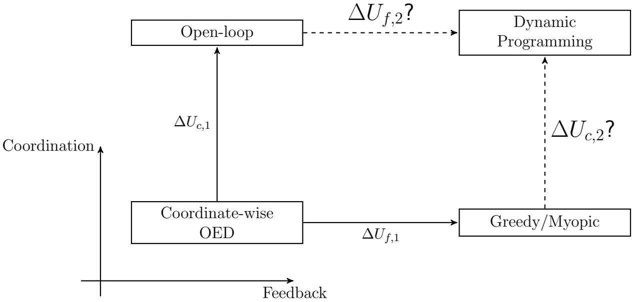


Future work: framework and methodology choices

What experimental design framework and methods to use?



Future work: value of feedback and coordination



Estimate "value of feedback" and "value of coordination": $\Delta U_{c,2}$, $\Delta U_{f,2}$

- Heuristics: use $\Delta U_{c,1}$ and $\Delta U_{f,1}$ as guesses (cheaper to obtain)
- Theoretical: stochastic bounds $U_{\text{openloop}}, U_{\text{greedy}} \leq U^* \leq U_{\text{hindsight}}$

Ultimately: find appropriate degrees and combinations of feedback and coordination for different subsets of experiments

References I

Richard Bellman.
Bottleneck Problems and Dynamic Programming.
Proceedings of the National Academy of Sciences of the United States of America, vol. 39, no. 9, pages 947–951, 1953.

Donald A. Berry, Peter Müller, Andy P. Grieve, Michael Smith, Tom Parke, Richard Blazek, Neil Mitchard & Michael Krams.
Adaptive Bayesian Designs for Dose-Ranging Drug Trials.
In Case studies in Bayesian statistics, pages 99–181. Springer New York, New York, NY, 2002.

Dimitri P. Bertsekas & John N. Tsitsiklis.
Neuro-Dynamic Programming.
Athena Scientific, Belmont, MA, 1996.

Dimitri P. Bertsekas.
Dynamic Programming and Optimal Control, Vol. 1.
Athena Scientific, Belmont, MA, 2005.

George Edward Pelham Box.
Science and Statistics.
Journal of the American Statistical Association, vol. 71, no. 356, pages 791–799, 1976.

Anthony E. Brockwell & Joseph B. Kadane.
A Gridding Method for Bayesian Sequential Decision Problems.
Journal of Computational and Graphical Statistics, vol. 12, no. 3, pages 566–584, 2003.

References II

Guillaume Carlier, Alfred Galichon & Filippo Santambrogio.
From Knothe's Transport to Brenier's Map and a Continuation Method for Optimal Transport.
SIAM Journal on Mathematical Analysis (Society for Industrial and Applied Mathematics), vol. 41, no. 6, pages 2554–2576, 2010.

P. Carlin, Bradley, Joseph B. Kadane & Alan E. Gelfand.
Approaches for Optimal Sequential Decision Analysis in Clinical Trials.
Biometrics, vol. 54, no. 3, pages 964–975, 1998.

Christopher C. Drovandi, James M. McGree & Anthony N. Pettitt.
A Sequential Monte Carlo Algorithm to Incorporate Model Uncertainty in Bayesian Sequential Design.
Journal of Computational and Graphical Statistics, vol. 23, no. 1, pages 3–24, 2014.

Xun Huan.
Numerical Approaches for Sequential Bayesian Optimal Experimental Design.
PhD thesis, Massachusetts Institute of Technology, 2015.

Leslie Pack Kaelbling, Michael L. Littman & Andrew W. Moore.
Reinforcement Learning: A Survey.
Journal of Artificial Intelligence Research, vol. 4, pages 237–285, 1996.

Marc. C. Kennedy & Anthony O'Hagan.
Bayesian calibration of computer models.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 63, no. 3, pages 425–464, 2001.

Woojae Kim, Mark A. Pitt, Zhong-Lin Lu, Mark Steyvers & Jay I. Myung.
A Hierarchical Adaptive Approach to Optimal Experimental Design.
Neural Computation, vol. 26, pages 2565–2492, 2014.

References III

Dennis V. Lindley.
On a Measure of the Information Provided by an Experiment.
The Annals of Mathematical Statistics, vol. 27, no. 4, pages 986–1005, 1956.

Joseph C. Oefelein.
Large eddy simulation of turbulent combustion processes in propulsion and power systems.
Progress in Aerospace Sciences, vol. 42, no. 1, pages 2–37, 2006.

Matthew David Parno.
Transport maps for accelerated Bayesian computation.
PhD thesis, Massachusetts Institute of Technology, 2015.

Warren B. Powell.
Approximate Dynamic Programming: Solving the Curses of Dimensionality.
John Wiley & Sons, Inc., Hoboken, NJ, 2nd edition, 2011.

Murray Rosenblatt.
Remarks on a Multivariate Transformation.
The Annals of Mathematical Statistics, vol. 23, no. 3, pages 470–472, 1952.

Andrea Saltelli.
Making best use of model evaluations to compute sensitivity indices.
Computer Physics Communications, vol. 145, no. 2, pages 280–297, 2002.

Andrea Saltelli, Stefano Tarantola, Francesca Campolongo & Marco Ratto.
Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models.
John Wiley & Sons, Chichester, United Kingdom, 2004.

References IV

Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana & Stefano Tarantola.
Global Sensitivity Analysis: The Primer.
John Wiley & Sons, Chichester, United Kingdom, 2008.

Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto & Stefano Tarantola.
Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index.
Computer Physics Communications, vol. 181, no. 2, pages 259–270, 2010.

Khachik Sargsyan, Habib N. Najm & Roger G. Ghanem.
On the Statistical Calibration of Physical Models.
International Journal of Chemical Kinetics, vol. 47, no. 4, pages 246–276, 2015.

I. M. Sobol.
Theorems and examples on high dimensional model representation.
Reliability Engineering & System Safety, vol. 79, no. 2, pages 187–193, 2003.

Antti Solonen, Heikki Haario & Marko Laine.
Simulation-Based Optimal Design Using a Response Variance Criterion.
Journal of Computational and Graphical Statistics, vol. 21, no. 1, pages 234–252, 2012.

Richard S. Sutton & Andrew G. Barto.
Reinforcement Learning: An Introduction.
The MIT Press, Cambridge, MA, 1998.

Cédric Villani.
Optimal Transport: Old and New.
Springer-Verlag Berlin Heidelberg, Berlin, Germany, 2008.