Leveraging Domain Expertise in Bayesian Experimental Design

Vikram V. Garg

March 10, 2022

Overview

What is Experimental Design ?

- Find sets of experiments that provide most information about targeted parameters.
- Where and when to make measurements ?
- Which variables to interrogate ?
- What experimental conditions are to be choosen ?

Example

- $-D\Delta u + \mathbf{V} \cdot \nabla u + \gamma u(1-u) = 0.$
- A bad experiment would be insensitive to errors in the inferred value of diffusivity.

Goals

- Maximize the value of data for inference and prediction
- Explore impact of observables on information gain
- Conditions under which to repeat experiments

Tools

- Bayesian description of data assimilation
- Information theoretic measure of information gain
- Computational Model: Physics or Data based or both

Bayes' rule

- $p(\theta \mid \mathbf{y}, \mathbf{d}) = \frac{p(\mathbf{y}|\theta,\mathbf{d})p(\theta)}{p(\mathbf{y}|\mathbf{d})}$
- θ: Parameter to be inferred
- $\mathbf{d}$: Experimental conditions
- $\mathbf{y}$: Data obtained from realization of $\mathbf{d}$

Bayes' rule

- $p(\theta \mid \mathbf{y}, \mathbf{d}) = \frac{p(\mathbf{y}|\theta,\mathbf{d})p(\theta)}{p(\mathbf{y}|\mathbf{d})}$
- θ: Parameter to be inferred
- $\mathbf{d}$: Experimental conditions
- $\mathbf{y}$: Data obtained from realization of $\mathbf{d}$

Information gain

- Measure difference between two densities
- Kullback-Leibler (KL) divergence:

$$D_{KL}(A||B) = \int_{-\infty}^{\infty} p_A(x) log\left(\frac{p_A(x)}{p_B(x)}\right) dx$$

- Relative entropy, represents information gain

Utility Function

- KL divergence from prior to posterior in current context
- Function of conditions $\mathbf{d}$ and realizations $\mathbf{y}$
- $u(\mathbf{d},\mathbf{y}) = D_{KL}(p(\theta \mid \mathbf{y},\mathbf{d}) || p(\theta)) = \int\limits_{-\infty}^{\infty} p(\theta|\mathbf{y},\mathbf{d}) log(\frac{p(\theta|\mathbf{y},\mathbf{d})}{p(\theta)})\, d\theta$

Utility Function

- KL divergence from prior to posterior in current context
- Function of conditions $\mathbf{d}$ and realizations $\mathbf{y}$
- $u(\mathbf{d},\mathbf{y}) = D_{KL}(p(\theta \mid \mathbf{y},\mathbf{d})||p(\theta)) = \int_{-\infty}^{\infty} p(\theta|\mathbf{y},\mathbf{d}) log(\frac{p(\theta|\mathbf{y},\mathbf{d})}{p(\theta)})\, d\theta$

Expected Utility

- Maximize utility function over all possible data $\rightarrow$ Expected information gain at conditions $\mathbf{d}$
- $U(\mathbf{d}) = \int_{Y} \left(\int_{\Theta} (\log(p(\mathbf{y}|\theta,\mathbf{d}) - log(p(\mathbf{y}|\mathbf{d}))p(\theta)\, d\theta \right) p(\mathbf{y}|\theta,\mathbf{d}) d\mathbf{y}.$

Utility Function

- KL divergence from prior to posterior in current context
- Function of conditions $\mathbf{d}$ and realizations $\mathbf{y}$
- $u(\mathbf{d}, \mathbf{y}) = D_{KL}(p(\theta \mid \mathbf{y}, \mathbf{d}) || p(\theta)) = \int_{-\infty}^{\infty} p(\theta | \mathbf{y}, \mathbf{d}) log(\frac{p(\theta|\mathbf{y},\mathbf{d})}{p(\theta)})\, d\theta$

Expected Utility

- Maximize utility function over all possible data $\rightarrow$ Expected information gain at conditions $\mathbf{d}$
- $U(\mathbf{d}) = \int_{Y} \left(\int_{\Theta} (\log(p(\mathbf{y}|\theta, \mathbf{d}) - log(p(\mathbf{y}|\mathbf{d})) p(\theta)\, d\theta \right) p(\mathbf{y}|\theta, \mathbf{d}) d\mathbf{y}.$
- Optimization problem: Find $\mathbf{d}^* = \arg\max U(\mathbf{d})$

What makes obtaining $\mathbf{d}^*$ hard ?

- Design space can be massive.
- Likelihood $p(\mathbf{y}|\theta, \mathbf{d})$ can be expensive or infeasible to evaluate.
- Prior $p(\theta)$ can be difficult to specify and sample from.

What makes obtaining $\mathbf{d}^*$ hard ?

- Design space can be massive.
- Likelihood $p(\mathbf{y}|\theta, \mathbf{d})$ can be expensive or infeasible to evaluate.
- Prior $p(\theta)$ can be difficult to specify and sample from.

Every challenge also an opportunity (to do math).

Overview

Combustion Kinetics[2]

- Use shock tube experiment to interrogate hydrogen-oxygen reaction
- Shock wave spikes temperature and pressure and triggers reaction.

Combustion Kinetics[2]

- Use shock tube experiment to interrogate hydrogen-oxygen reaction
- Shock wave spikes temperature and pressure and triggers reaction.

Mathematical Model of Reaction(s)

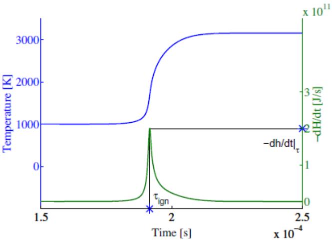
- Conservation of energy and mass
- Constitutive relation: $k_{f,m} = A_m T^{b_m} exp(\frac{-E_{a,m}}{R_u T})$
- Want to infer parameters A_1 and $E_{a,3}$

Design variables

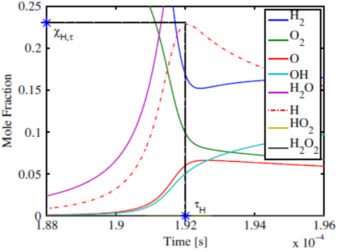
- Initial temperature T_0
- Fuel-oxidizer equivalence ratio ϕ
- What temperature should the experiment be performed at, and what should be the relative amount of fuel and oxidizer ?

Selected observables for the combustion problem. Note that $dh/dt < 0$ when enthalpy is released or lost by the system.

Observable	Explanation
τ_{ign}	Ignition delay, defined as the time of peak enthalpy release rate
τ_O	Characteristic time in which peak X_O occurs
τ_H	Characteristic time in which peak X_H occurs
τ_{HO_2}	Characteristic time in which peak X_{HO_2} occurs
$\tau_{H_2O_2}$	Characteristic time in which peak $X_{H_2O_2}$ occurs
$\frac{dh}{dt}\big\|_\tau$	Peak value of enthalpy release rate
$X_{O,\tau}$	Peak value of X_O
$X_{H,\tau}$	Peak value of X_H
$X_{HO_2,\tau}$	Peak value of X_{HO_2}
$X_{H_2O_2,\tau}$	Peak value of $X_{H_2O_2}$



(a) τ_{ign}



(b) τ_{H} and $X_{\mathrm{H},\tau}$

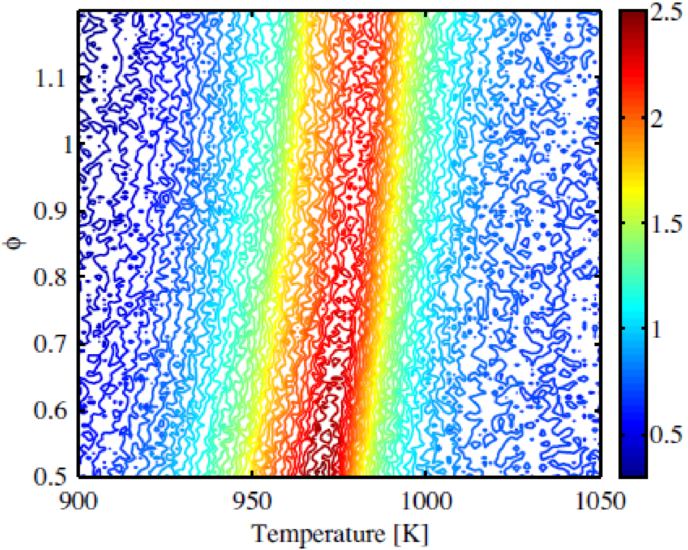
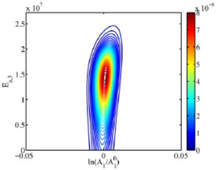
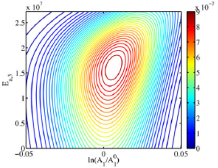


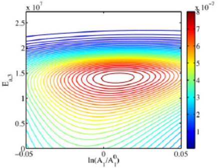
Figure: Utility contours with all observables



(a) Design A full ODE model



(c) Design B full ODE model



(e) Design C full ODE model

Figure: A(975,0.5), B(925,0.85), C(1025,0.85)

Overview

Ocean Turbulent Mixing Viscosity

- Governing equation:
  $$\frac{\partial \phi}{\partial t} + \nabla \cdot (\mathbf{u}\phi + \mathbf{u}^*\phi) = \nabla \cdot \kappa_\sigma \nabla_\sigma \phi + \frac{\partial\left(\kappa_z(D_{kr})\frac{\partial \phi}{\partial z}\right)}{\partial z}$$ [1].
- ϕ tracer, $\mathbf{u}$ from hydrodynamics solver.
- Parameter of interest: Turbulent mixing viscosity: $D_{kr}(\mathbf{x})$

Sources of Complexity

- Infinite dimensional parameter, expensive forward evaluations.
- Need to avoid unphysical realizations of D_{kr} which lead to sample rejections.
- Expert knowledge to inform prior and reduce computational burden.

Modeling D_{kr}

- D_{kr} modeled as a Gaussian process.
- Need to specify the covariance for this process, $cov_{D_{kr}}(\mathbf{x}, \mathbf{y})$.

Covariance Modeling

- For a spatially distributed parameter, we need to specify covariance kernels.
- Typical kernels: stationary, isotropic, smooth and periodic
- $cov_{D_{kr}}$ non-stationary and anisotropic.

General Covariance Kernel Generation [5]

- General second order stochastic PDE:
  $(\kappa(\mathbf{x}) - \Delta)(\tau(\mathbf{x})u(\mathbf{x})) = \mathcal{W}(\mathbf{x})$
- Generalized Matern kernel:
  $\text{cov}(u(\mathbf{0}), u(\mathbf{x})) = \frac{\sigma(\tau)^2}{2^{\nu-1}\Gamma(\nu)} \left(\kappa\|\mathbf{x}\|\right)^{\nu} K_{\nu}\left(\kappa\|\mathbf{x}\|\right)$

Matern Kernel Parameters

- $\kappa(\mathbf{x})$: Inverse of the pointwise correlation length.
- $\tau(\mathbf{x})$: Inverse of the pointwise marginal variance.
- Prescribe models for κ and τ based on simulation variables, parameters.

Software Implementation

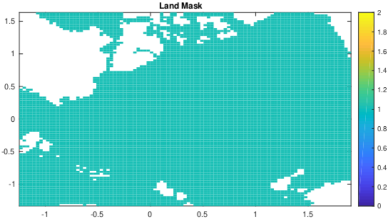
- R SPDE solver package INLA [4].
- Specify geometry, creates Finite Element mesh, generates samples with prescribed covariance structure.

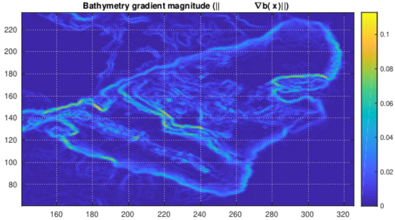
Verifying realizations

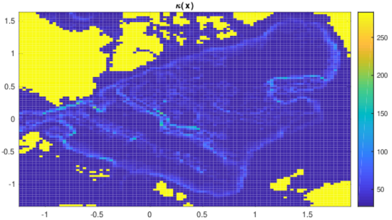
- Non physical realizations need to be rejected.
- All samples need to exhibit mixed layer produced by the interaction of the Arctic ocean's salinity with the hydrodynamics.

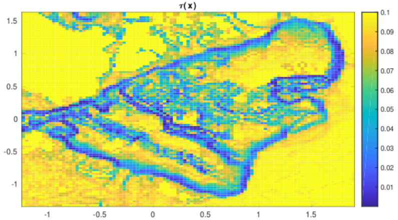
SPDE parameter specification

- Expert input: $\kappa(\mathbf{x})$ and $\tau(\mathbf{x})$ depend only bathymetry gradient.
- $\kappa(\mathbf{x}) = \kappa_m e^{c_\kappa \|\nabla b(\mathbf{x})\|}, \tau(\mathbf{x}) = \tau_m e^{c_\tau \|\nabla b(\mathbf{x})\|}$









Conclusions and Future Work

- Bayesian experimental design powerful quantitative tool for OED, especially in the presence of nonlinearities.

Conclusions and Future Work

- Bayesian experimental design powerful quantitative tool for OED, especially in the presence of nonlinearities.
- Computational burden can be alleviated by incorporating expert knowledge, computational algorithms and exploiting parallelism.

Conclusions and Future Work

- Bayesian experimental design powerful quantitative tool for OED, especially in the presence of nonlinearities.
- Computational burden can be alleviated by incorporating expert knowledge, computational algorithms and exploiting parallelism.
- Framework can be extended to sequential experiments using dynamic programming [3].

[1] Forget, G., Ferreira, D., and Liang, X. (2015). On the observability of turbulent transport rates by argo: supporting evidence from an inversion experiment. *Ocean Science*, 11(5):839.

[2] Huan, X. and Marzouk, Y. M. (2013). Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317.

[3] Huan, X. and Marzouk, Y. M. (2016). Sequential bayesian optimal experimental design via approximate dynamic programming. *arXiv preprint arXiv:1604.08320*.

[4] Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, 63(19).

[5] Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.