The Consistent Bayesian Approach for Stochastic Inverse Problems

SAND2018-12334C

Tim Wildey†, Troy Butler‡, John Jakeman†

† Sandia National Laboratories[1]
Center for Computing Research
Optimization and UQ Group

‡ University of Colorado Denver
Department of Mathematics and Statistics

AMS Fall Southeastern Sectional Meeting
University of Arkansas, Fayetteville, AR
November 3-4, 2018

SAND2018-****

[1] Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

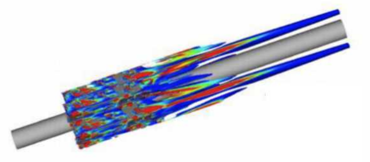
Outline

1. Motivation
2. Stochastic Inverse Problems
   - Problem formulation
   - A measure-theoretic solution
   - An illustrative example
3. Comparison with the standard Bayesian formulation
4. A natural verification procedure
5. Applications
   - Resistive Magnetohydrodynamics
   - Additive Manufacturing
   - Tokamak Disruption Mitigation
6. Conclusions and Future Work

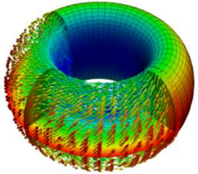
Outline

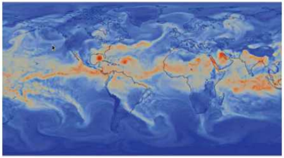
Motivation

Flow in Nuclear Reactor (Turbulent CFD)

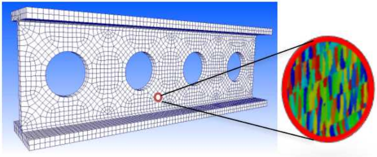


Tokamak Equilibrium (MHD)





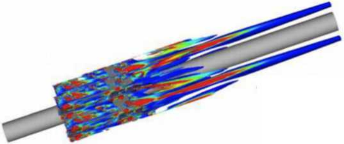
Climate Modeling



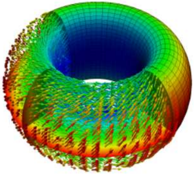
Multi-scale Materials Modeling

Motivation

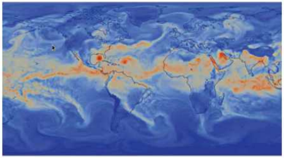
Flow in Nuclear Reactor (Turbulent CFD)



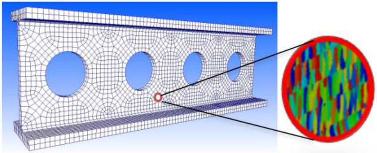
Tokamak Equilibrium (MHD)



We are working to develop **data-informed** models ...

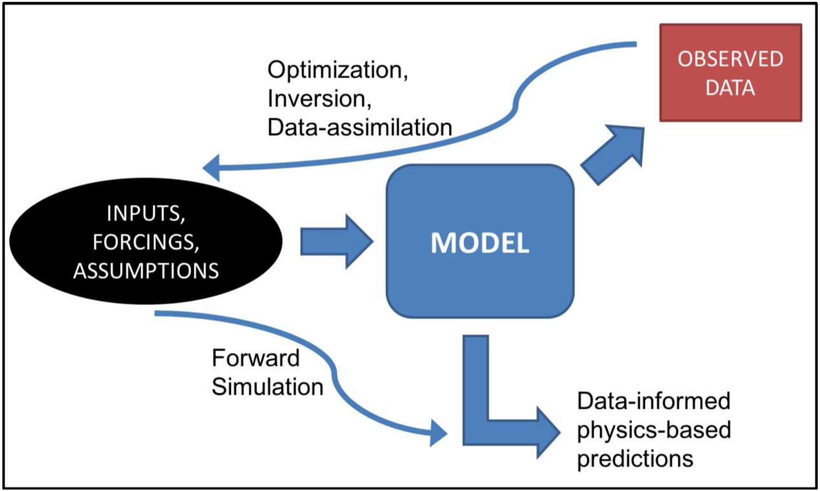


Climate Modeling



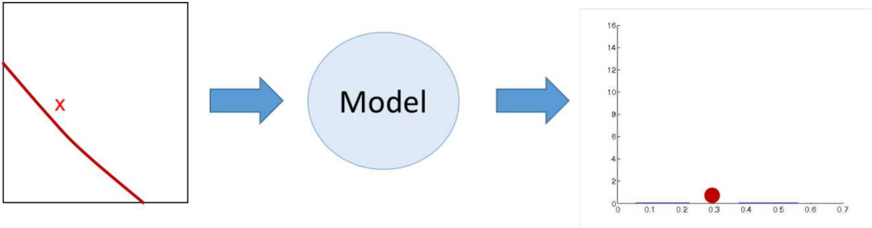
Multi-scale Materials Modeling

Data-informed Physics-Based Predictions



Outline

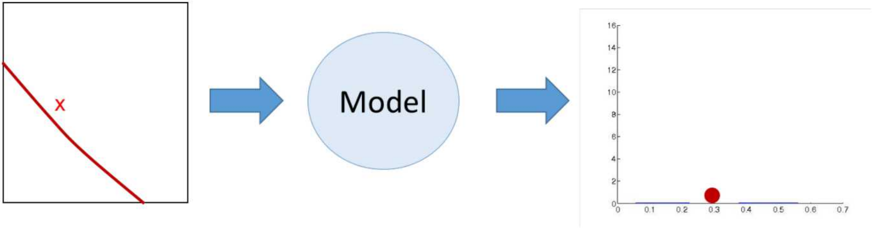
A Deterministic Inverse Problem



Problem

Given a deterministic observation, $\hat{Q}$, find $\lambda \in \Lambda$ such that $Q(\lambda) = \hat{Q}$.

A Deterministic Inverse Problem

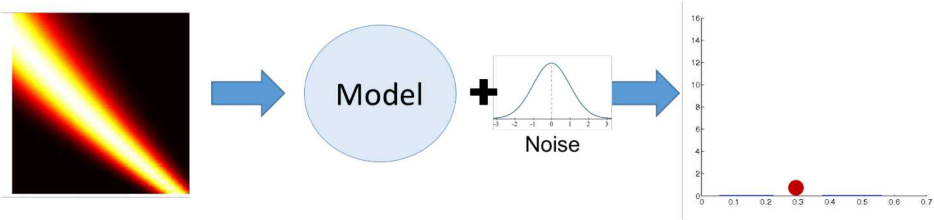


Problem

Given a deterministic observation, $\hat{Q}$, find $\lambda \in \Lambda$ such that $Q(\lambda) = \hat{Q}$.

- Solutions may not be unique without additional assumptions.
- Requires solving several deterministic forward problems.

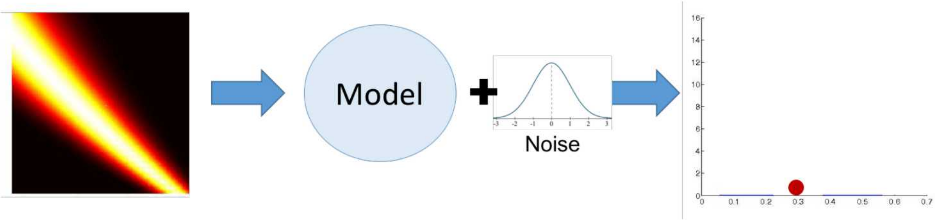
A Stochastic Inverse Problem



Problem

Given a deterministic observation, $\hat{Q}$, and an assumed noise model, find the parameters that are most likely to have produced the data.

A Stochastic Inverse Problem

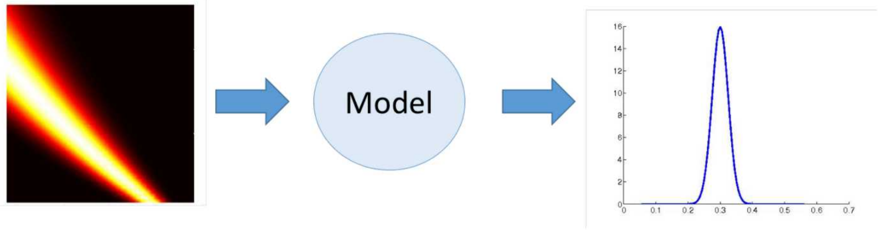


Problem

Given a deterministic observation, $\hat{Q}$, and an assumed noise model, find the parameters that are most likely to have produced the data.

- Solutions may not be unique without additional assumptions.
- Requires solving several deterministic forward problems.

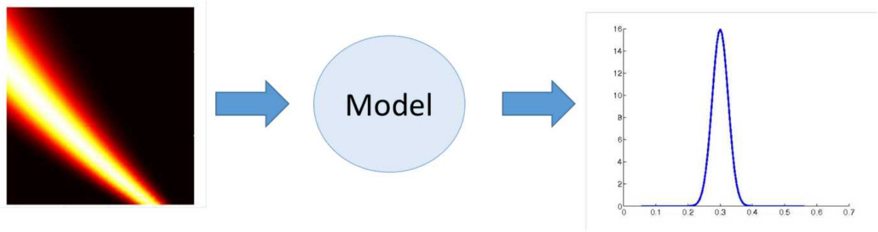
A Different Stochastic Inverse Problem



Problem

Given a probability density on observations, find a probability density on Λ such that the push-forward matches the given density on the observed data.

A Different Stochastic Inverse Problem



Problem

Given a probability density on observations, find a probability density on Λ such that the push-forward matches the given density on the observed data.

- Solutions may not be unique without additional assumptions.
- **We only need to solve a single stochastic forward problem.**

Notation

We assume we are given:

1. A finite-dimensional **parameter space**, Λ.
2. A **parameter-to-observation/data map**, $Q : \Lambda \to \mathcal{D} = Q(\Lambda)$
3. An **observed probability measure** on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, denoted $P_{\mathcal{D}}^{\text{obs}}$, that has a density, $\pi_{\mathcal{D}}^{\text{obs}}$.
4. A **probability measure** on $(\Lambda, \mathcal{B}_{\Lambda})$, denoted $P_{\Lambda}^{\text{prior}}$, that has a density, $\pi_{\Lambda}^{\text{prior}}$.

Notation

We assume we are given:

1. A finite-dimensional **parameter space**, Λ.
2. A **parameter-to-observation/data map**, $Q : \Lambda \to \mathcal{D} = Q(\Lambda)$
3. An **observed probability measure** on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, denoted $P_{\mathcal{D}}^{\text{obs}}$, that has a density, $\pi_{\mathcal{D}}^{\text{obs}}$.
4. A **probability measure** on $(\Lambda, \mathcal{B}_{\Lambda})$, denoted $P_{\Lambda}^{\text{prior}}$, that has a density, $\pi_{\Lambda}^{\text{prior}}$.

We need to compute:

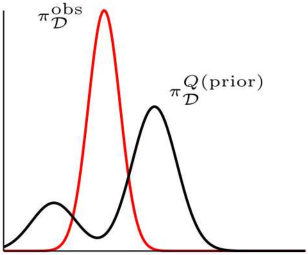
1. The **push-forward of the prior** through the model.

- In other words, **we need to solve a forward UQ problem using the prior.**
- We use $\pi_{\mathcal{D}}^{Q(\text{prior})}$ to denote this push-forward density.

A Key Assumption

Predictability Assumption

We assume that the observed probability measure, $P_{\mathcal{D}}^{\text{obs}}$, is absolutely continuous with respect to the push-forward of the prior, $P_{\mathcal{D}}^{Q(\text{prior})}$.



Good Prior

Bad Prior
(Cannot predict all observations)

A Solution to the Stochastic Inverse Problem

Theorem (Butler, Jakeman, Wildey, SISC, 2018)

The probability measure P_Λ^{post} *on* $(\Lambda, \mathcal{B}_\Lambda)$ *defined by*

$$P_\Lambda^{post}(A) = \int_{\mathcal{D}} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{prior}(\lambda) \frac{\pi_{\mathcal{D}}^{obs}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q(prior)}(Q(\lambda))} \, d\mu_{\Lambda,q}(\lambda) \right) d\mu_{\mathcal{D}}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

A Solution to the Stochastic Inverse Problem

Theorem (Butler, Jakeman, Wildey, SISC, 2018)

The probability measure P_Λ^{post} *on* $(\Lambda, \mathcal{B}_\Lambda)$ *defined by*

$$P_\Lambda^{post}(A) = \int_{\mathcal{D}} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{prior}(\lambda) \frac{\pi_{\mathcal{D}}^{obs}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q(prior)}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_{\mathcal{D}}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

Corollary (BJW., SISC 2018)

The posterior measure of Λ *is 1.*

A Solution to the Stochastic Inverse Problem

Theorem (Butler, Jakeman, Wildey, SISC, 2018)

The probability measure P_Λ^{post} *on* $(\Lambda, \mathcal{B}_\Lambda)$ *defined by*

$$P_\Lambda^{post}(A) = \int_{\mathcal{D}} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{prior}(\lambda) \frac{\pi_{\mathcal{D}}^{obs}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q(prior)}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_{\mathcal{D}}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

Corollary (BJW., SISC 2018)

The posterior measure of Λ *is 1.*

Theorem (BJW., SISC 2018)

P_Λ^{post} *is stable with respect to perturbations in* $P_{\mathcal{D}}^{obs}$.

A Solution to the Stochastic Inverse Problem

Theorem (Butler, Jakeman, Wildey, SISC, 2018)

The probability measure P_Λ^{post} *on* $(\Lambda, \mathcal{B}_\Lambda)$ *defined by*

$$P_\Lambda^{post}(A) = \int_{\mathcal{D}} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{prior}(\lambda) \frac{\pi_{\mathcal{D}}^{obs}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q(prior)}(Q(\lambda))} \, d\mu_{\Lambda,q}(\lambda) \right) d\mu_{\mathcal{D}}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

Corollary (BJW., SISC 2018)

The posterior measure of Λ *is 1.*

Theorem (BJW., SISC 2018)

P_Λ^{post} *is stable with respect to perturbations in* $P_{\mathcal{D}}^{obs}$.

All the details can be found in: "Combining Push-forward Measures and Bayes' Rule to Construct Consistent Solutions to Stochastic Inverse Problems", BJW. SISC 40 (2), 2018.

A Solution to the Stochastic Inverse Problem

Theorem (Butler, Jakeman, Wildey, SISC, 2018)

The probability measure P_Λ^{post} on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$P_\Lambda^{post}(A) = \int_{\mathcal{D}} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{prior}(\lambda) \frac{\pi_{\mathcal{D}}^{obs}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q(prior)}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_{\mathcal{D}}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

The posterior density is:

$$\pi_\Lambda^{\text{post}}(\lambda) = \pi_\Lambda^{\text{prior}}(\lambda) \frac{\pi_{\mathcal{D}}^{\text{obs}}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q(\text{prior})}(Q(\lambda))}.$$

- Both $\pi_\Lambda^{\text{prior}}$ and $\pi_{\mathcal{D}}^{\text{obs}}$ are given.
- Computing $\pi_{\mathcal{D}}^{Q(\text{prior})}$ requires a forward propagation of the prior.
- Given $\pi_{\mathcal{D}}^{Q(\text{prior})}$, we can investigate different posteriors for different $\pi_{\mathcal{D}}^{\text{obs}}$.

A Parameterized Nonlinear System

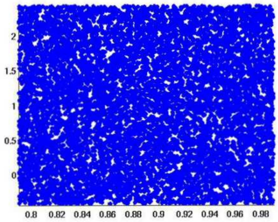
Example

Consider a parameterized nonlinear system of equations:

$$\begin{aligned} \lambda_1 x_1^2 + x_2^2 &= 1, \\ x_1^2 - \lambda_2 x_2^2 &= 1 \end{aligned}$$

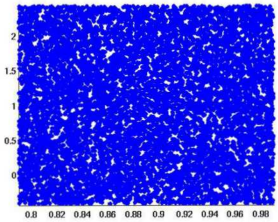
- The quantity of interest is the second component: $q(\lambda) = x_2$.
- Assume that we observe $q(\lambda) \sim N(0.3, 0.025^2)$.
- We consider a uniform prior.
- We use 10,000 samples from the prior and a standard KDE to approximate the push-forward of the prior.

A Parameterized Nonlinear System

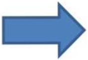


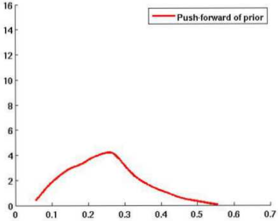
Prior

A Parameterized Nonlinear System



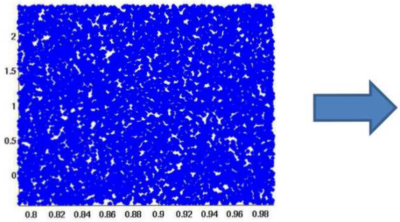
Prior



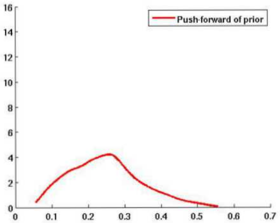


Push-forward of Prior

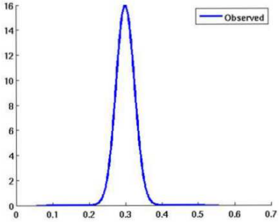
A Parameterized Nonlinear System



Prior

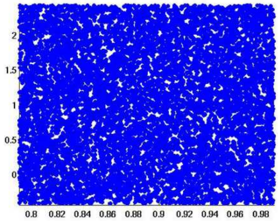


Push-forward of Prior

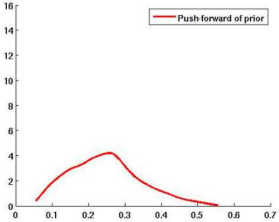


Observed density

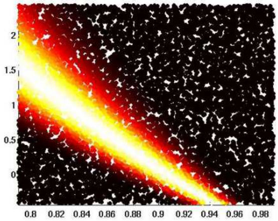
A Parameterized Nonlinear System



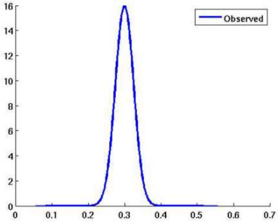
Prior



Push-forward of Prior

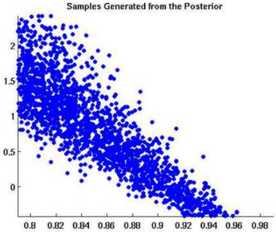


Posterior



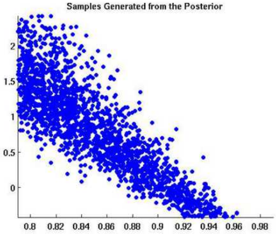
Observed density

A Parameterized Nonlinear System

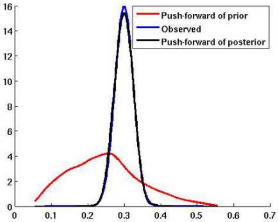


Samples from the posterior

A Parameterized Nonlinear System

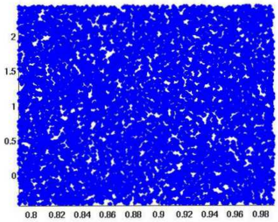


Samples from the posterior

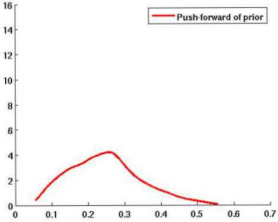


Observed and push-forward densities in $\mathcal{D}$

Exploring Different Observed Densities

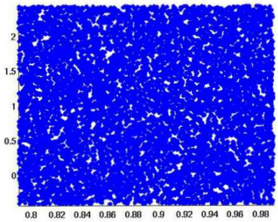


Prior

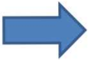


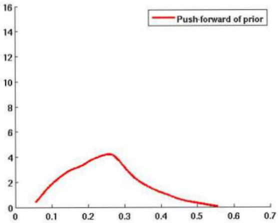
Push-forward of Prior

Exploring Different Observed Densities

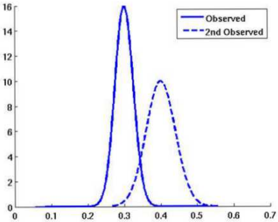


Prior



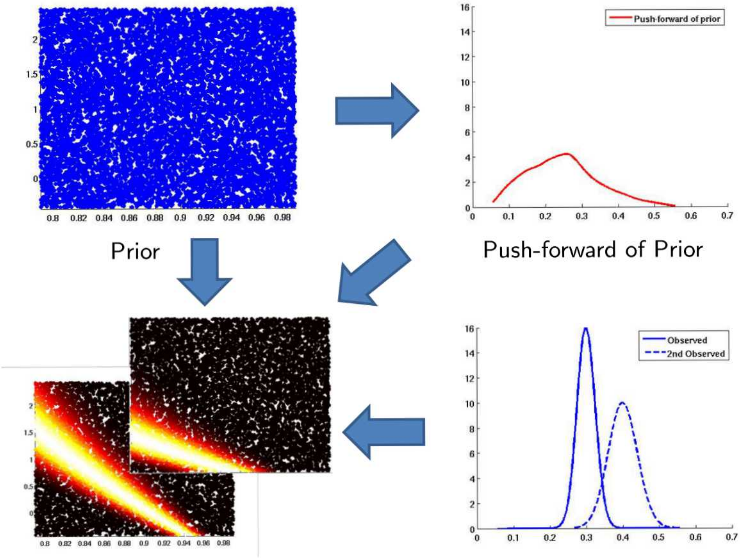


Push-forward of Prior



Observed densities

Exploring Different Observed Densities



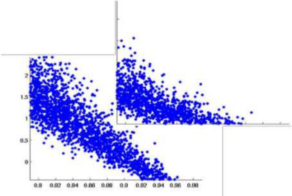
Prior

Push-forward of Prior

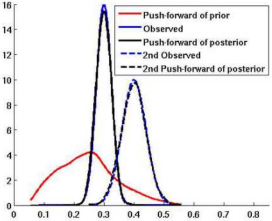
Posteriors

Observed densities

Exploring Different Observed Densities



Samples from the posteriors



Observed and push-forward densities in $\mathcal{D}$

Outline

Relationship with Statistical Bayesian Inference

Using Bayes theorem we can define a different posterior density [Stuart 2010; Gelman et al 2013; Jaynes 1998, ...]:

$$\tilde{\pi}_{\Lambda}^{\text{post}}(\lambda|q) = \pi_{\Lambda}^{\text{prior}}(\lambda)\frac{\pi(q|\lambda)}{C}.$$

Relationship with Statistical Bayesian Inference

Using Bayes theorem we can define a different posterior density [Stuart 2010; Gelman et al 2013; Jaynes 1998, ...]:

$$\tilde{\pi}_{\Lambda}^{\text{post}}(\lambda|q) = \pi_{\Lambda}^{\text{prior}}(\lambda)\frac{\pi(q|\lambda)}{C}.$$

Example

Let $\Lambda = [-1, 1]$ and consider the simple nonlinear map

$$q(\lambda) = \lambda^p, \quad p = 1, 3, 5, \ldots .$$

Here, p is not uncertain and are used to vary the nonlinearity.

Relationship with Statistical Bayesian Inference

Using Bayes theorem we can define a different posterior density [Stuart 2010; Gelman et al 2013; Jaynes 1998, ...]:

$$\tilde{\pi}_{\Lambda}^{\text{post}}(\lambda|q) = \pi_{\Lambda}^{\text{prior}}(\lambda)\frac{\pi(q|\lambda)}{C}.$$

Example

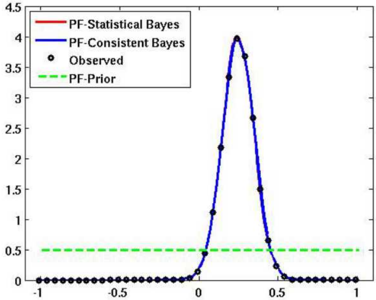
Let $\Lambda = [-1, 1]$ and consider the simple nonlinear map

$$q(\lambda) = \lambda^p, \quad p = 1, 3, 5, \ldots.$$

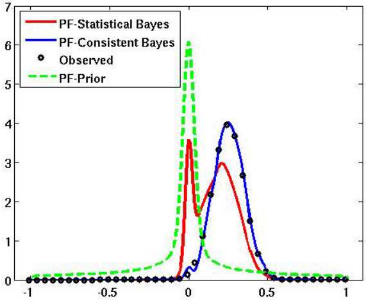
Here, p is not uncertain and are used to vary the nonlinearity.

- Assume a uniform prior and the observed density is given by $\pi_{\mathcal{D}}^{\text{obs}} \sim N(0.25, 0.1^2)$.
- For the statistical Bayesian approach, we use an observed value of $\hat{q} = 0.25$ and assume a Gaussian noise model $\eta \sim N(0, 0.1^2)$.

Comparing Push-forwards for Linear and Nonlinear Maps

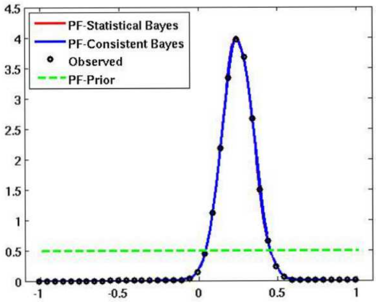


Linear map ($p = 1$)

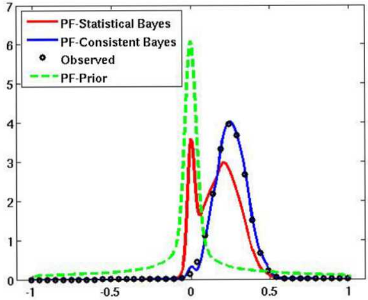


Nonlinear map ($p = 5$)

Comparing Push-forwards for Linear and Nonlinear Maps



Linear map ($p = 1$)



Nonlinear map ($p = 5$)

The statistical and consistent Bayesian formulations **solve different problems**, **have different posteriors** and **make different predictions**.

Outline

Determining Convergence

Question:

How do we know when we've generated enough samples from the posterior?

Determining Convergence

Question:

How do we know when we've generated enough samples from the posterior?

- For MCMC, one typically uses autocorrelation, mixing of chain, Gelman/Rubin statistic, etc.

Determining Convergence

Question:

How do we know when we've generated enough samples from the posterior?

- For MCMC, one typically uses autocorrelation, mixing of chain, Gelman/Rubin statistic, etc.
- For our approach, we know one thing about the posterior ...

Determining Convergence

Question:

How do we know when we've generated enough samples from the posterior?

- For MCMC, one typically uses autocorrelation, mixing of chain, Gelman/Rubin statistic, etc.
- For our approach, we know one thing about the posterior ...
- **The push-forward of the posterior matches the measure/density on the observations.**
- We can use this to assess convergence regardless of the sampling scheme.

Determining Convergence

Question:

How do we know when we've generated enough samples from the posterior?

- For MCMC, one typically uses autocorrelation, mixing of chain, Gelman/Rubin statistic, etc.
- For our approach, we know one thing about the posterior ...
- **The push-forward of the posterior matches the measure/density on the observations.**
- We can use this to assess convergence regardless of the sampling scheme.

With three caveats ...

Determining Convergence

Question:

How do we know when we've generated enough samples from the posterior?

- For MCMC, one typically uses autocorrelation, mixing of chain, Gelman/Rubin statistic, etc.
- For our approach, we know one thing about the posterior ...
- **The push-forward of the posterior matches the measure/density on the observations.**
- We can use this to assess convergence regardless of the sampling scheme.

With three caveats ...

1. Convergence will typically stagnate due to the approximation of the push-forward of the prior.
2. Cannot determine if all modes of the posterior have been identified.
3. Not sensitive to how un-informed directions are explored.

Determining Convergence

Question:

How do we know when we've generated enough samples from the posterior?

- For MCMC, one typically uses autocorrelation, mixing of chain, Gelman/Rubin statistic, etc.
- For our approach, we know one thing about the posterior ...
- **The push-forward of the posterior matches the measure/density on the observations.**
- We can use this to assess convergence regardless of the sampling scheme.

With three caveats ...

1. Convergence will typically stagnate due to the approximation of the push-forward of the prior.
2. Cannot determine if all modes of the posterior have been identified.
3. Not sensitive to how un-informed directions are explored.

In general, we have found the push-forward of the posterior to be very useful in assess the convergence of the sampling procedure.

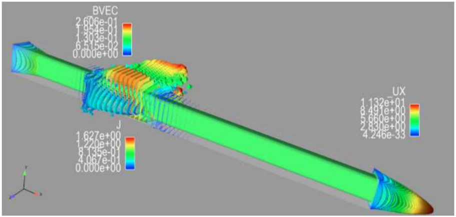
Outline

Resistive MHD with a Surrogate Model

- 3D resistive MHD generator
- VMS stabilized FEM approximation
- QoI - avg. induced magnetic energy:

$$Q = \frac{1}{2\mu_0} \int_{\Omega} (B_x^2 + B_y^2) \, d\Omega$$

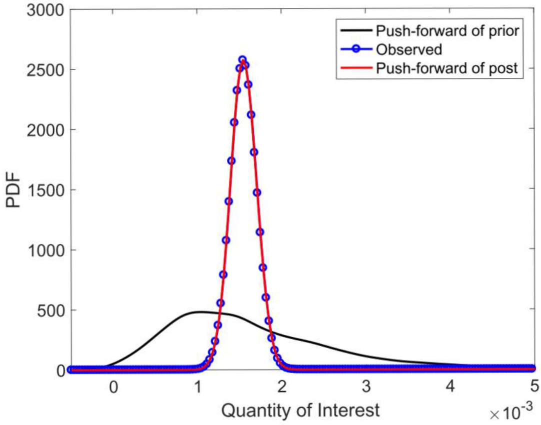
- Treat 4 input parameters as uncertain with uniform prior
- **Re-used an existing LHS study with 100 samples**
- Build surrogate model using Gaussian process regression
- 50,000 samples of surrogate used to compute push-forward of prior



Parameter	Min.	Max.
Viscosity	1.0E-3	1.0E-2
Vol. source	1.0E-1	5.0E-1
Resistivity	1.0E-1	1.0E1
Density	1.0E-1	1.0E1

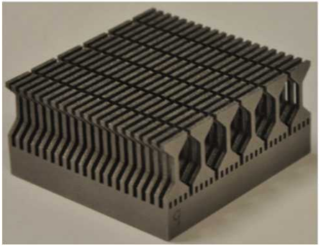
Push-forward of the Posterior Matches Observations

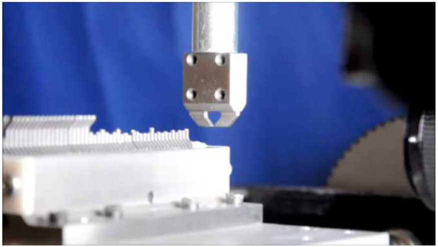
We assume a Gaussian density for the QoI with mean 1.55E-3 and 10% standard deviation.



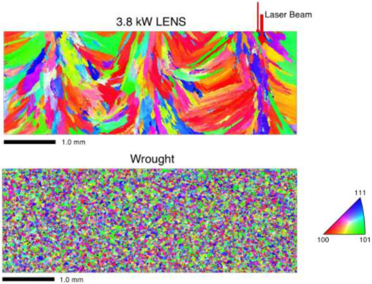
Additive Manufacturing (3D Printing)

At Sandia, We Make Dog Bones



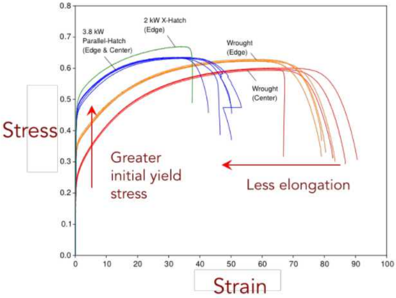


Additive Manufacturing Can Produce Extreme Properties

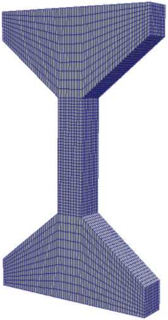


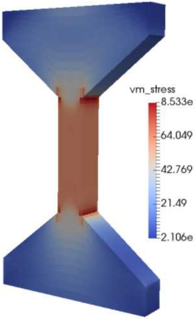
304 L

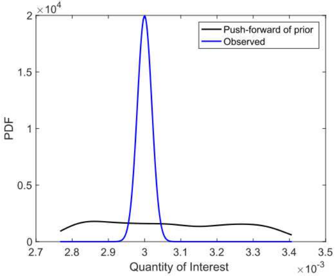
(J. Michael, SNL)



(J. Carroll, SNL)





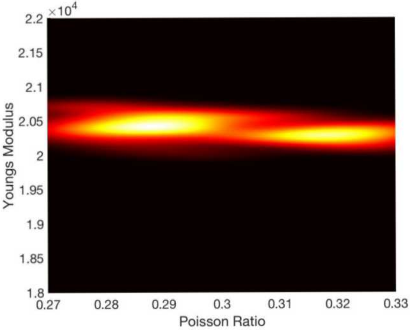


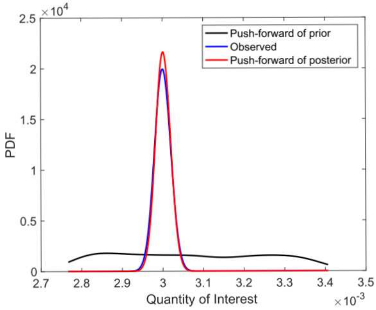
Quantity of interest is the average y-displacement in upper half of connector.

Random parameters: $\nu \sim U(0.38, 0.42)$, and $E \sim U(1.0e3, 2.0e3)$.

Evaluate model at 1,000 samples from the prior

Push-forward of the Posterior Matches Observations





Tokamak Disruption Mitigation

The tokamak is an experimental machine designed to harness the energy of fusion.

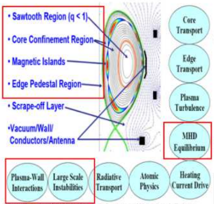
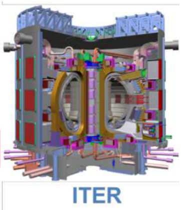


Fig. 2: Illustration of the interacting physical processes within a tokamak discharge.



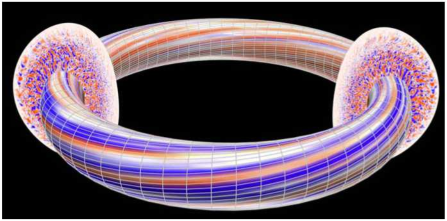


Image courtesy of W. Wang, PPPL

Disruption mitigation is essential for ITER and tokamak fusion.

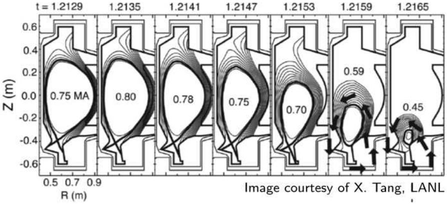
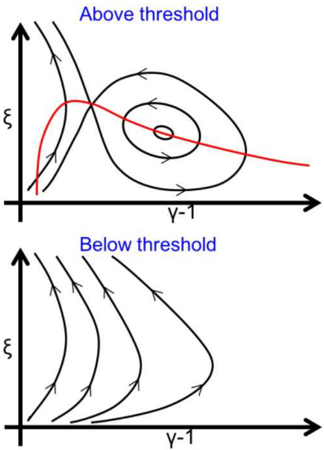


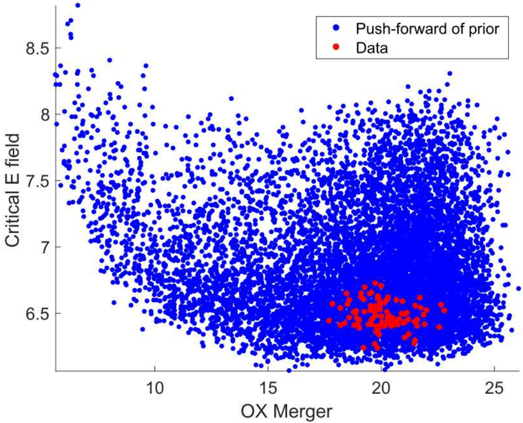
Image courtesy of X. Tang, LANL

Tokamak Disruption Mitigation

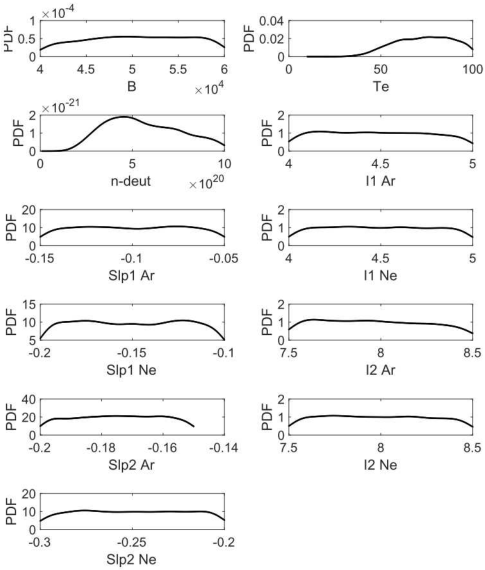
- In a disruption, plasma temperature will drop from 10 keV to a few eV in a few ms.
- This energy can be mostly channeled through runaway electrons.
- Complete avoidance is impractical
- Optimal scenario is to avoid runaway avalanche
- Semi-analytic theory of the runaway threshold recently developed (McDevitt et al. 2018)
- Provides versatile tool for determining conditions under which a large runaway population can be avoided
- Depends on the strength of the magnetic field, the electron temperature and the charge state distribution of the impurity populations.



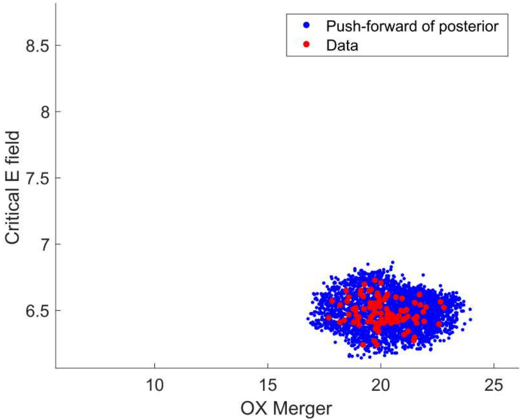
The Prior-predictive and the Data



The Marginals of the Posterior



The Posterior-predictive and the Data



Outline

Conclusions and Future Work

- Our goal is to develop **scalable, data-informed physics-based** models to make **credible predictions** and **optimize design of experiments**.
- Many approaches exist for incorporating data into a simulation.
  - Deterministic optimization, Bayesian methods, OUU, data assimilation, etc.
- The **consistent Bayesian** formulation provides a solution to a specific stochastic inverse problem.
- Approach naturally provides a **verifiable** quantity (push-forward of posterior).
- Push-forward of the prior is dominant computational expense.
- Can **leverage existing scalable/efficient approaches** for forward UQ
  - Surrogate models, e.g., Gaussian processes, polynomial chaos, sparse grid collocation, etc.
  - Multi-level and multi-fidelity methods
- Other sampling approaches, e.g., importance sampling and MCMC, are also possible

Thanks! Questions?

Acknowledgments

- T. Wildey's work was supported by the U.S. Department of Energy, Office of Science, Early Career Research Program.
- T. Butler's work was partially supported by DOE DE-SC0009286
- J. Jakeman's work was partially supported by DARPA EQUIPS

Thank you for your attention!

Questions?