A Probabilistic Characterization of Aleatoric and Epistemic Uncertainty in Solutions to Stochastic Inverse Problems Using Machine Learning Surrogate Models

Tim Wildey[♣], Tian Yu Yen[♣], Troy Butler[♦]

[♣] Sandia National Laboratories
Center for Computing Research
Scientific Machine Learning Department

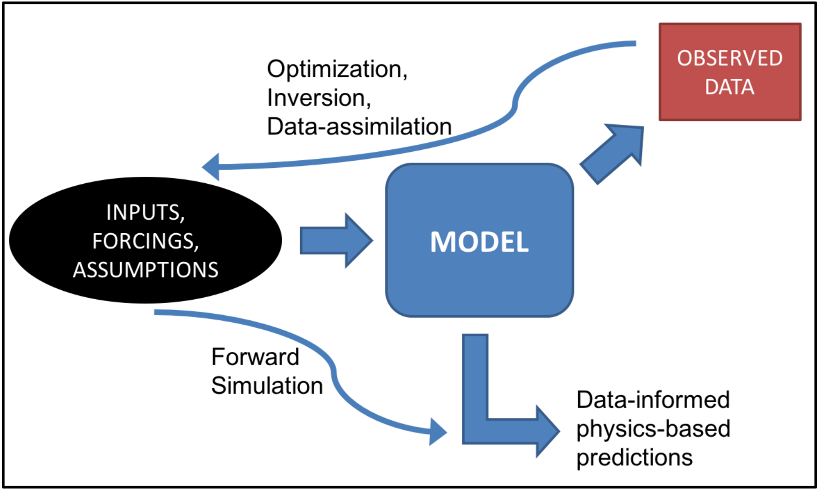
[♦] University of Colorado Denver
Department of Mathematics and Statistics

SIAM Conference on Uncertainty Quantification
April 12-15, 2022

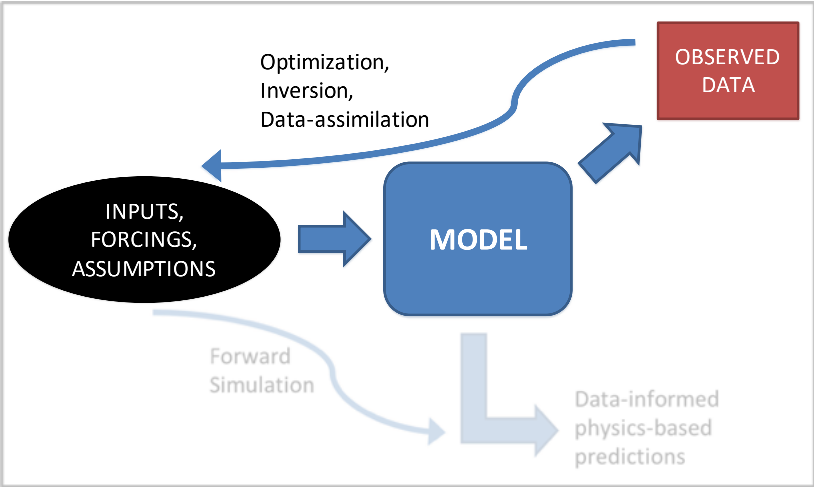
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. SAND2021-4672 C

boratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LL
Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE

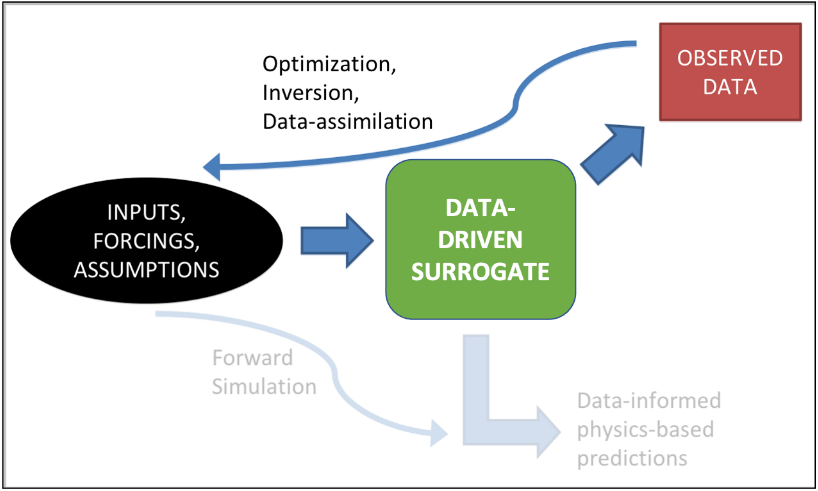
Data-informed Physics-Based Predictions



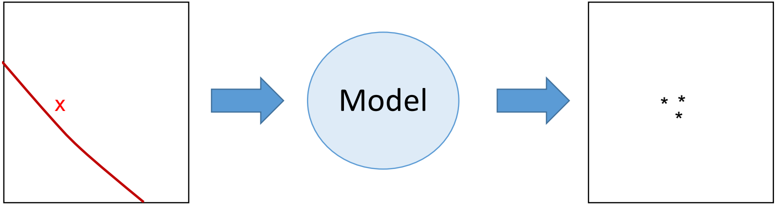
Data-informed Physics-Based Predictions



Data-informed Physics-Based Predictions



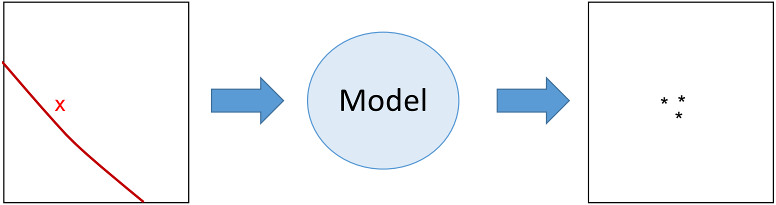
A Deterministic Inverse Problem



Problem

Given some observed data, find $\lambda \in \Lambda$ that best predicts the data.

A Deterministic Inverse Problem

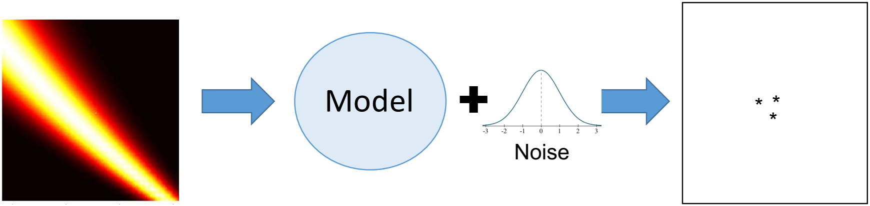


Problem

Given some observed data, find $\lambda \in \Lambda$ that best predicts the data.

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

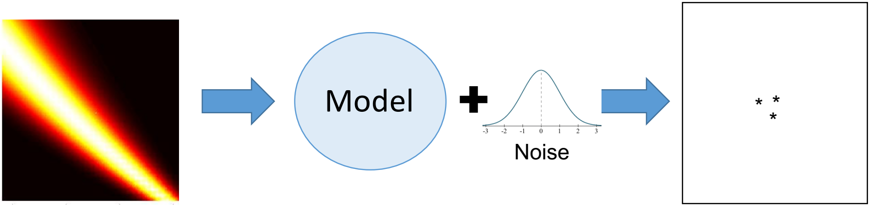
A Stochastic Inverse Problem



Problem

Given some observed data and an assumed noise model, find the parameters that are most likely to have produced the data.

A Stochastic Inverse Problem

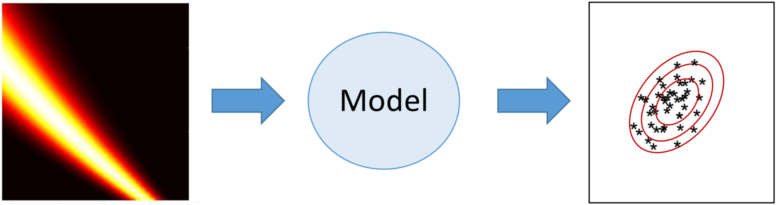


Problem

Given some observed data 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.

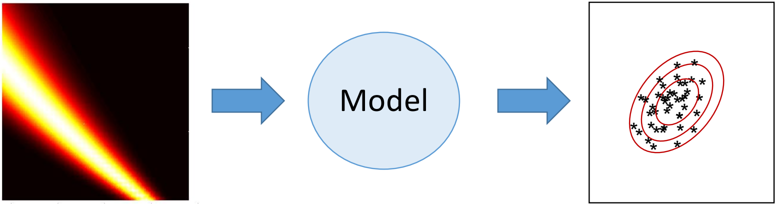
A Different Stochastic Inverse Problem



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. A **observed/target probability measure** on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, denoted $\mathbb{P}_{\mathcal{D}}^{\text{obs}}$, with density $\pi_{\mathcal{D}}^{\text{obs}}$ (typically from experimental data)
4. An **initial probability measure** on $(\Lambda, \mathcal{B}_{\Lambda})$, denoted $\mathbb{P}_{\Lambda}^{\text{init}}$, with density $\pi_{\Lambda}^{\text{init}}$ (typically from prior beliefs or expert knowledge)

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. A **observed/target probability measure** on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$, denoted $\mathbb{P}_{\mathcal{D}}^{\text{obs}}$, with density $\pi_{\mathcal{D}}^{\text{obs}}$ (typically from experimental data)
4. An **initial probability measure** on $(\Lambda, \mathcal{B}_{\Lambda})$, denoted $\mathbb{P}_{\Lambda}^{\text{init}}$, with density $\pi_{\Lambda}^{\text{init}}$ (typically from prior beliefs or expert knowledge)

We need to compute:

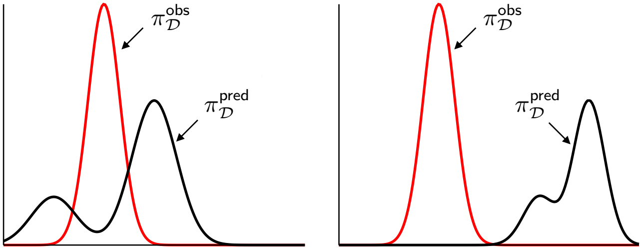
1. The **push-forward of the initial density** through the model.

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

A Key Assumption

Predictability Assumption

We assume that the observed probability measure, $\mathbb{P}_{\mathcal{D}}^{\text{obs}}$, is absolutely continuous with respect to the push-forward of the initial, $\mathbb{P}_{\mathcal{D}}^{\text{pred}}$.



Good Initial

Bad Initial
(Cannot predict all observations)

A Solution to the Stochastic Inverse Problem

Theorem

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and an observed probability measure, $\mathbb{P}_\mathcal{D}^{obs}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, the probability measure $\mathbb{P}_\Lambda^{up}$ on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

A Solution to the Stochastic Inverse Problem

Theorem

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and an observed probability measure, $\mathbb{P}_\mathcal{D}^{obs}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, the probability measure $\mathbb{P}_\Lambda^{up}$ on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

Corollary

The updated measure of Λ is 1.

A Solution to the Stochastic Inverse Problem

Theorem

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and an observed probability measure, $\mathbb{P}_\mathcal{D}^{obs}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, the probability measure $\mathbb{P}_\Lambda^{up}$ on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

Corollary

The updated measure of Λ is 1.

Theorem

$\mathbb{P}_\Lambda^{up}$ *is stable with respect to perturbations in* $\mathbb{P}_\mathcal{D}^{obs}$ *and in* $\mathbb{P}_\Lambda^{init}$.

For details: [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

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ *on* $(\Lambda, \mathcal{B}_\Lambda)$ *and an observed probability measure,* $\mathbb{P}_\mathcal{D}^{obs}$, *on* $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, *the probability measure* $\mathbb{P}_\Lambda^{up}$ *on* $(\Lambda, \mathcal{B}_\Lambda)$ *defined by*

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

A Solution to the Stochastic Inverse Problem

Theorem

Given an initial probability measure, $\mathbb{P}_\Lambda^{init}$ on $(\Lambda, \mathcal{B}_\Lambda)$ and an observed probability measure, $\mathbb{P}_\mathcal{D}^{obs}$, on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$, the probability measure $\mathbb{P}_\Lambda^{up}$ on $(\Lambda, \mathcal{B}_\Lambda)$ defined by

$$\mathbb{P}_\Lambda^{up}(A) = \int_\mathcal{D} \left(\int_{A \cap Q^{-1}(q)} \pi_\Lambda^{init}(\lambda) \frac{\pi_\mathcal{D}^{obs}(Q(\lambda))}{\pi_\mathcal{D}^{pred}(Q(\lambda))} d\mu_{\Lambda,q}(\lambda) \right) d\mu_\mathcal{D}(q), \ \forall A \in \mathcal{B}_\Lambda$$

solves the stochastic inverse problem.

The updated density is:

$$\pi_\Lambda^{\text{up}}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda) \frac{\pi_\mathcal{D}^{\text{obs}}(Q(\lambda))}{\pi_\mathcal{D}^{\text{pred}}(Q(\lambda))}.$$

- Both $\pi_\Lambda^{\text{init}}$ and $\pi_\mathcal{D}^{\text{obs}}$ are given.
- Computing $\pi_\mathcal{D}^{\text{pred}}$ requires a forward propagation of the initial density.

A Parameterized Nonlinear System

Example

Consider a parameterized nonlinear system of equations:

$$\begin{aligned} \lambda_1 u_1^2 + u_2^2 &= 1, \\ u_1^2 - \lambda_2 u_2^2 &= 1 \end{aligned}$$

- Quantity of interest is the second component: $Q(\lambda) = u_2$.
- Given $\pi_{\mathcal{D}}^{\text{obs}} \sim N(0.3, 0.025^2)$.
- Given a uniform initial density.
- Use 10,000 samples from the initial and a standard KDE to approximate the push-forward.
- Use standard rejection sampling to generate samples from π_Λ^{up}.

A Parameterized Nonlinear System

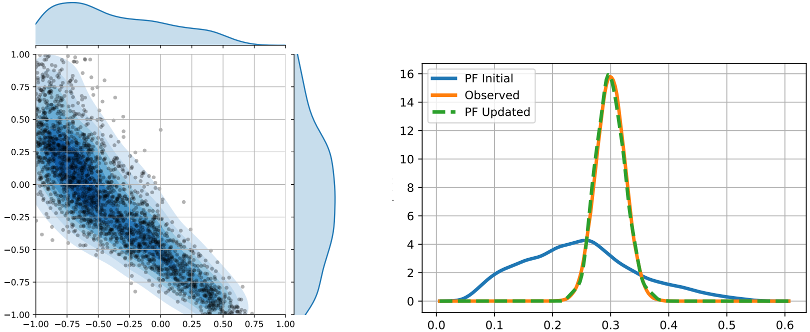
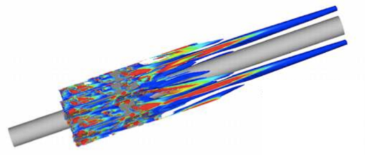


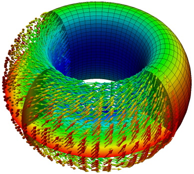
Figure: Samples from the updated density (left) and a comparison of $\pi_{\mathcal{D}}^{\text{obs}}$, $\pi_{\mathcal{D}}^{\text{pred}}$ and push-forward of the updated density (right).

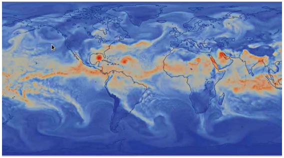
Why do we care about approximate models?

Flow in Nuclear Reactor (Turbulent CFD)

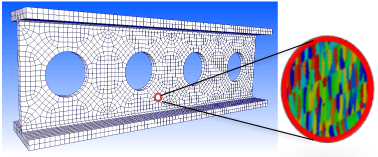


Tokamak Equilibrium (MHD)





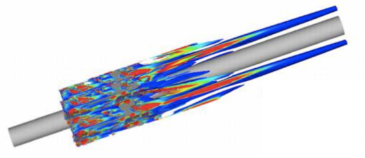
Climate Modeling



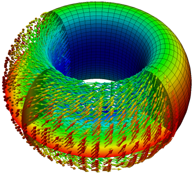
Multi-scale Materials Modeling

Why do we care about approximate models?

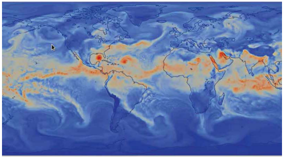
Flow in Nuclear Reactor (Turbulent CFD)



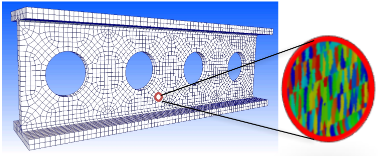
Tokamak Equilibrium (MHD)



All are computationally expensive and require some form of approximation ...



Climate Modeling



Multi-scale Materials Modeling

Convergence of Inverse Solutions

Recall that the updated density is given by

$$\pi_\Lambda^{\text{up}}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda)\frac{\pi_\mathcal{D}^{\text{obs}}(Q(\lambda))}{\pi_\mathcal{D}^{\text{pred}}(Q(\lambda))}$$

The updated density using a surrogate model, $Q_S(\lambda)$, is given by

$$\pi_\Lambda^{\text{up},S}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda)\frac{\pi_\mathcal{D}^{\text{obs}}(Q_S(\lambda))}{\pi_\mathcal{D}^{\text{pred},S}(Q_S(\lambda))}$$

Convergence of Inverse Solutions

Recall that the updated density is given by

$$\pi_\Lambda^{\text{up}}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda)\frac{\pi_\mathcal{D}^{\text{obs}}(Q(\lambda))}{\pi_\mathcal{D}^{\text{pred}}(Q(\lambda))}$$

The updated density using a surrogate model, $Q_S(\lambda)$, is given by

$$\pi_\Lambda^{\text{up},S}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda)\frac{\pi_\mathcal{D}^{\text{obs}}(Q_S(\lambda))}{\pi_\mathcal{D}^{\text{pred},S}(Q_S(\lambda))}$$

Theorem (B.J.W. SISC 2018b)

Under the the assumptions in [B.J.W., 2018b], $Q_S(\lambda) \to Q(\lambda)$ in $L^\infty(\Lambda) \implies \pi_\Lambda^{up,S}(\lambda) \to \pi_\Lambda^{up}(\lambda)$ in $L^1(\Lambda)$.

Extensions to convergence in L^p have also been developed recently [Butler, Wildey, Zhang, IJUQ, 2022].

Does this include data-driven models?

Theorem (W. Zhang Thesis 2021)

Suppose $Q \in C(\Lambda)$ *and the assumptions in [B.J.W., 2018b] are satisfied. Then* **there exists** *a sequence of single hidden layer Neural Networks defined on* Λ *such that (amongst other results):*

$$\pi_{\Lambda}^{up,S}(\lambda) \rightarrow \pi_{\Lambda}^{up}(\lambda) \text{ in } L^1(\Lambda).$$

Similar results can be shown for Neural Networks with arbitrary depth and fixed width by combining this result with the UAT from [Zhou et al 2017].

Does this include data-driven models?

Theorem (W. Zhang Thesis 2021)

Suppose $Q \in C(\Lambda)$ and the assumptions in [B.J.W., 2018b] are satisfied. Then **there exists** *a sequence of single hidden layer Neural Networks defined on Λ such that (amongst other results):*

$$\pi_{\Lambda}^{up,S}(\lambda) \to \pi_{\Lambda}^{up}(\lambda) \text{ in } L^1(\Lambda).$$

Similar results can be shown for Neural Networks with arbitrary depth and fixed width by combining this result with the UAT from [Zhou et al 2017].

Does this help me add **error bars or confidence intervals to my solution** to the stochastic inverse problem?

Does this include data-driven models?

Theorem (W. Zhang Thesis 2021)

Suppose $Q \in C(\Lambda)$ and the assumptions in [B.J.W., 2018b] are satisfied. Then **there exists** *a sequence of single hidden layer Neural Networks defined on Λ such that (amongst other results):*

$$\pi_{\Lambda}^{up,S}(\lambda) \to \pi_{\Lambda}^{up}(\lambda) \text{ in } L^1(\Lambda).$$

Similar results can be shown for Neural Networks with arbitrary depth and fixed width by combining this result with the UAT from [Zhou et al 2017].

Does this help me add **error bars or confidence intervals to my solution** to the stochastic inverse problem?

No, but let's see what we can do ...

Back to verification for physics-based models

To paraphrase a quote from the movie Shrek:

Verification is like an onion ...

Back to verification for physics-based models

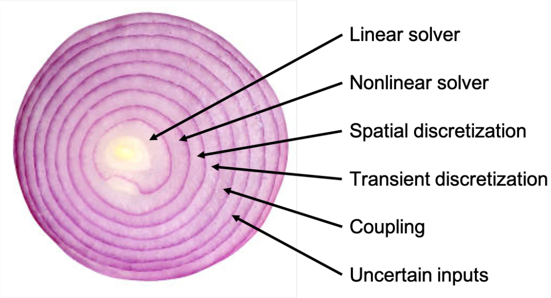
To paraphrase a quote from the movie Shrek:

Verification is like an onion … it stinks.

Back to verification for physics-based models

To paraphrase a quote from the movie Shrek:

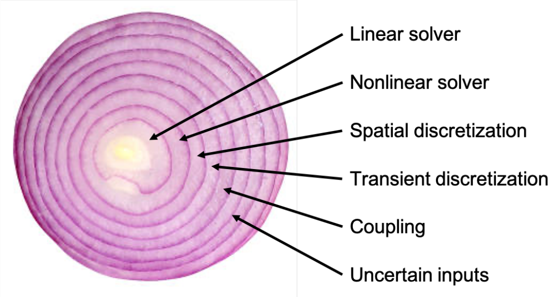
Verification is like an onion ... it stinks. Also, it has layers.



Back to verification for physics-based models

To paraphrase a quote from the movie Shrek:

Verification is like an onion ... it stinks. Also, it has layers.



We need to consider a variety of approaches to quantify the various sources of error/uncertainty.

Error Estimates for Surrogates of Deterministic Physics-based Models

If we assume:

- we have a QoI from a physics-based model,
- we build a surrogate approximation of the QoI.

Error Estimates for Surrogates of Deterministic Physics-based Models

If we assume:

- we have a QoI from a physics-based model,
- we build a surrogate approximation of the QoI.

Then, the error in point-wise evaluations of the surrogate model are due to:

- interpolation or extrapolation of the surrogate model
- biased training data from using discretized physics-based models

Error Estimates for Surrogates of Deterministic Physics-based Models

If we **also** assume:

- we have an adjoint for the physics-based model,

Then, we can use a generalization of adjoint-based techniques to estimate the error in **point-wise evaluations** of the surrogate model [Butler, Dawson, W. 2011].

Error Estimates for Surrogates of Deterministic Physics-based Models

If we **also** assume:

- we have an adjoint for the physics-based model,

Then, we can use a generalization of adjoint-based techniques to estimate the error in **point-wise evaluations** of the surrogate model [Butler, Dawson, W. 2011].

Such error estimates are higher-order and can be used to:

- Define an improved surrogate model [Butler, Dawson, W. 2013]
- Drive adaptivity in the surrogate model [Jakeman, W. 2015]
- Decompose errors into various contributions [Bryant, Prudhomme, W. 2015]
- Derive better MCMC sampling strategies [Butler, Dawson, W. 2015]
- Estimate errors in probabilities of rare events [Butler, W. 2018]

Error Estimates for Surrogates of Deterministic Physics-based Models

If we **also** assume:

- we have an adjoint for the physics-based model,

Then, we can use a generalization of adjoint-based techniques to estimate the error in **point-wise evaluations** of the surrogate model [Butler, Dawson, W. 2011].

Such error estimates are higher-order and can be used to:

- Define an improved surrogate model [Butler, Dawson, W. 2013]
- Drive adaptivity in the surrogate model [Jakeman, W. 2015]
- Decompose errors into various contributions [Bryant, Prudhomme, W. 2015]
- Derive better MCMC sampling strategies [Butler, Dawson, W. 2015]
- Estimate errors in probabilities of rare events [Butler, W. 2018]

Have not been used to estimate errors in data-consistent inversion ...

Error Estimates for Data-consistent Solutions

Suppose we are given

- A surrogate model, $Q_S(\lambda) \approx Q(\lambda)$.
- A set of samples (not training data), $\{\lambda_i\}_{i=1}^N$, generated from $\pi_\Lambda^{\text{init}}$, where we want to evaluate $Q_S(\lambda)$.
- An estimate of the error $e_i \approx Q(\lambda_i) - Q_S(\lambda_i)$

Error Estimates for Data-consistent Solutions

Suppose we are given

- A surrogate model, $Q_S(\lambda) \approx Q(\lambda)$.
- A set of samples (not training data), $\{\lambda_i\}_{i=1}^N$, generated from $\pi_\Lambda^{\text{init}}$, where we want to evaluate $Q_S(\lambda)$.
- An estimate of the error $e_i \approx Q(\lambda_i) - Q_S(\lambda_i)$

Then, we can defined the **improved surrogate approximation**:

$$Q_{S+}(\lambda_i) = Q_S(\lambda_i) + e_i,$$

and the **improved data-consistent solution**:

$$\pi_\Lambda^{\text{up},S+}(\lambda_i) = \pi_\Lambda^{\text{init}}(\lambda_i) r_{S+}(\lambda_i), \quad r_{S+}(\lambda_i) = \frac{\pi_\mathcal{D}^{\text{obs}}(Q_{S+}(\lambda_i))}{\pi_\mathcal{D}^{\text{pred},S+}(Q_{S+}(\lambda_i))}$$

Error Estimates for Data-consistent Solutions

The **improved ratio**, $r_{S+}(\lambda_i)$, can be used to estimate the error in the updated density in the total variation metric:

$$\int_\Lambda \left|\pi_\Lambda^{\text{up}}(\lambda) - \pi_\Lambda^{\text{up},S}(\lambda)\right| \; d\mu_\Lambda \approx \int_\Lambda \left|\pi_\Lambda^{\text{up},S+}(\lambda) - \pi_\Lambda^{\text{up},S}(\lambda)\right| \; d\mu_\Lambda$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} |r_{S+}(\lambda_i) - r_S(\lambda_i)|$$

We can also it to evaluate the **reliability** in the updated density on a point-wise basis.

A simple example

Consider the following partial differential equation,

$$\begin{cases} -\nabla \cdot (\mathsf{K} \nabla u) + \mathsf{b}(\lambda_1, \lambda_2, x) \cdot \nabla u = g(x), & x \in \Omega = (0,1) \times (0,1) \\ u = 0, & x \in \partial\Omega \end{cases}$$

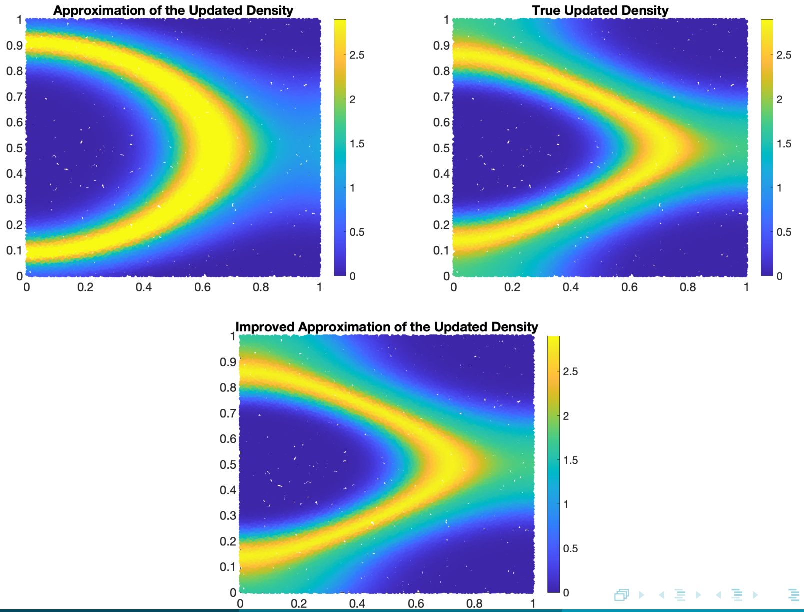
The quantity of interest is a mollified point-evaluation:

$$Q(\lambda) = \frac{100}{\pi} e^{-100(x_1 - 0.5)^2 - 100(x_2 - 0.5)^2}$$

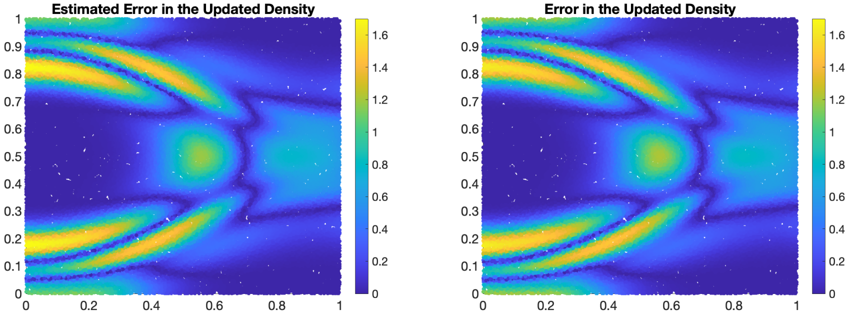
Discretization details:

- Finite element on 50×50 mesh,
- Surrogate approximation is 3$^{\text{rd}}$-order pseudo-spectral approximation
  - Implies the error estimate is 6$^{\text{th}}$-order
- $\pi_\Lambda^{\text{init}}$ is uniform on $[0,1]^2$
- Use 50,000 samples evaluated using surrogate to approximate push-forward

A simple example



A simple example



True TV error = 0.4002

Estimated TV error = 0.4017

This is nice, but ...

This approach is not very useful for models built from data!

This is nice, but ...

This approach is not very useful for models built from data!

Data-driven models tend to have **many** sources of error/uncertainty:

- Discretization/architecture (epistemic)
- Sparse/uninformative data (epistemic)
- Noisy data (aleatoric)
- Optimization/solver variability (aleatoric)
- Extrapolation/OoD (epistemic)

This is nice, but ...

This approach is not very useful for models built from data!

Data-driven models tend to have **many** sources of error/uncertainty:

- Discretization/architecture (epistemic)
- Sparse/uninformative data (epistemic)
- Noisy data (aleatoric)
- Optimization/solver variability (aleatoric)
- Extrapolation/OoD (epistemic)

From [Hüllermeier and Waegeman 2021]:

> *... a trustworthy representation of uncertainty is desirable and should be considered as a key feature of any machine learning method ...*

Bayesian [Neal 2012; Gal et al 2016; ...] and ensemble-based [Lakshminarayanan et al 2017; Ashukha et al 2021, ...] approaches are the most common.

From [Abdar et al 2021]:

> *... ensemble methods have a great ability to deal with uncertainty ...*

We use a combination of ensemble-based approaches ...

Using the proper ensemble for DCI

Suppose we compute an ensemble of data-driven surrogate models, $\left\{Q_S^{(i)}(\lambda)\right\}_{i=1}^{M}$.

Let $\overline{g}$ denote an ensemble-averaged quantity, e.g.,

$$\overline{Q}_S(\lambda) = \frac{1}{M}\sum_{i=1}^{M} Q_S^{(i)}(\lambda)$$

Using the proper ensemble for DCI

Suppose we compute an ensemble of data-driven surrogate models, $\left\{ Q_S^{(i)}(\lambda) \right\}_{i=1}^{M}$.

Let $\overline{g}$ denote an ensemble-averaged quantity, e.g.,

$$\overline{Q}_S(\lambda) = \frac{1}{M} \sum_{i=1}^{M} Q_S^{(i)}(\lambda)$$

How can we construct a data-consistent measure/density?

Using the proper ensemble for DCI

Suppose we compute an ensemble of data-driven surrogate models, $\left\{ Q_S^{(i)}(\lambda) \right\}_{i=1}^{M}$.

Let $\overline{g}$ denote an ensemble-averaged quantity, e.g.,

$$\overline{Q}_S(\lambda) = \frac{1}{M} \sum_{i=1}^{M} Q_S^{(i)}(\lambda)$$

How can we construct a data-consistent measure/density?

Each member of the ensemble can be used to compute a data-consistent solution:

$$\pi_\Lambda^{\text{up},S,i}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda) \frac{\pi_\mathcal{D}^{\text{obs}}(Q_S^{(i)}(\lambda))}{\pi_\mathcal{D}^{\text{pred},S,i}(Q_S^{(i)}(\lambda))}$$

Using the proper ensemble for DCI

Suppose we compute an ensemble of data-driven surrogate models, $\left\{Q_S^{(i)}(\lambda)\right\}_{i=1}^{M}$.

Let $\overline{g}$ denote an ensemble-averaged quantity, e.g.,

$$\overline{Q}_S(\lambda) = \frac{1}{M}\sum_{i=1}^{M} Q_S^{(i)}(\lambda)$$

How can we construct a data-consistent measure/density?

Each member of the ensemble can be used to compute a data-consistent solution:

$$\pi_\Lambda^{\text{up},S,i}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda)\frac{\pi_\mathcal{D}^{\text{obs}}(Q_S^{(i)}(\lambda))}{\pi_\mathcal{D}^{\text{pred},S,i}(Q_S^{(i)}(\lambda))}$$

But we need to be careful with ensemble averages ...

Using the proper ensemble for DCI

A few options:

1. Use the ensemble-averaged updated density (ratio),

$$\pi_{\Lambda}^{\text{up},S}(\lambda) = \pi_{\Lambda}^{\text{init}}(\lambda)\bar{r}_S(\lambda)$$

Using the proper ensemble for DCI

A few options:

1. Use the ensemble-averaged updated density (ratio),

$$\pi_{\Lambda}^{\text{up},S}(\lambda) = \pi_{\Lambda}^{\text{init}}(\lambda)\overline{r}_S(\lambda)$$

2. Use the ensemble averaged observed and predicted densities,

$$\pi_{\Lambda}^{\text{up},S}(\lambda) = \pi_{\Lambda}^{\text{init}}(\lambda)\frac{\overline{\pi}_{\mathcal{D}}^{\text{obs}}}{\overline{\pi}_{\mathcal{D}}^{\text{pred}}}$$

where $\overline{\pi}_{\mathcal{D}}^{\text{obs}}$ and $\overline{\pi}_{\mathcal{D}}^{\text{pred}}$ are understood to be averaged for each λ.

Using the proper ensemble for DCI

A few options:

1. Use the ensemble-averaged updated density (ratio),

$$\pi_{\Lambda}^{\text{up},S}(\lambda) = \pi_{\Lambda}^{\text{init}}(\lambda)\overline{r}_S(\lambda)$$

2. Use the ensemble averaged observed and predicted densities,

$$\pi_{\Lambda}^{\text{up},S}(\lambda) = \pi_{\Lambda}^{\text{init}}(\lambda)\frac{\overline{\pi}_{\mathcal{D}}^{\text{obs}}}{\overline{\pi}_{\mathcal{D}}^{\text{pred}}}$$

   where $\overline{\pi}_{\mathcal{D}}^{\text{obs}}$ and $\overline{\pi}_{\mathcal{D}}^{\text{pred}}$ are understood to be averaged for each λ.

3. Use the ensemble-averaged surrogate model, $\overline{Q}_S(\lambda)$,

$$\pi_{\Lambda}^{\text{up},S}(\lambda) = \pi_{\Lambda}^{\text{init}}(\lambda)\frac{\pi_{\mathcal{D}}^{\text{obs}}(\overline{Q}_S(\lambda))}{\pi_{\mathcal{D}}^{\text{pred},S}(\overline{Q}_S(\lambda))}$$

Using the proper ensemble for DCI

A few options:

1. Use the ensemble-averaged updated density (ratio),

$$\pi_\Lambda^{\text{up},S}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda)\overline{r}_S(\lambda)$$

2. Use the ensemble averaged observed and predicted densities,

$$\pi_\Lambda^{\text{up},S}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda)\frac{\overline{\pi}_\mathcal{D}^{\text{obs}}}{\overline{\pi}_\mathcal{D}^{\text{pred}}}$$

   where $\overline{\pi}_\mathcal{D}^{\text{obs}}$ and $\overline{\pi}_\mathcal{D}^{\text{pred}}$ are understood to be averaged for each λ.

3. Use the ensemble-averaged surrogate model, $\overline{Q}_S(\lambda)$,

$$\pi_\Lambda^{\text{up},S}(\lambda) = \pi_\Lambda^{\text{init}}(\lambda)\frac{\pi_\mathcal{D}^{\text{obs}}(\overline{Q}_S(\lambda))}{\pi_\mathcal{D}^{\text{pred},S}(\overline{Q}_S(\lambda))}$$

Options 1 and 2 are not guaranteed to give a consistent measure/density.

Using the proper ensemble for DCI

We will use

1. The ensemble-averaged surrogate model, $\overline{Q}_S(\lambda)$, to estimate the data-consistent solution.
2. The ensemble of data-consistent solutions to assess the variability in the data-consistent solution.
3. A different ensemble of perturbation to assess impact of point-wise errors in ensemble-averaged model

Example

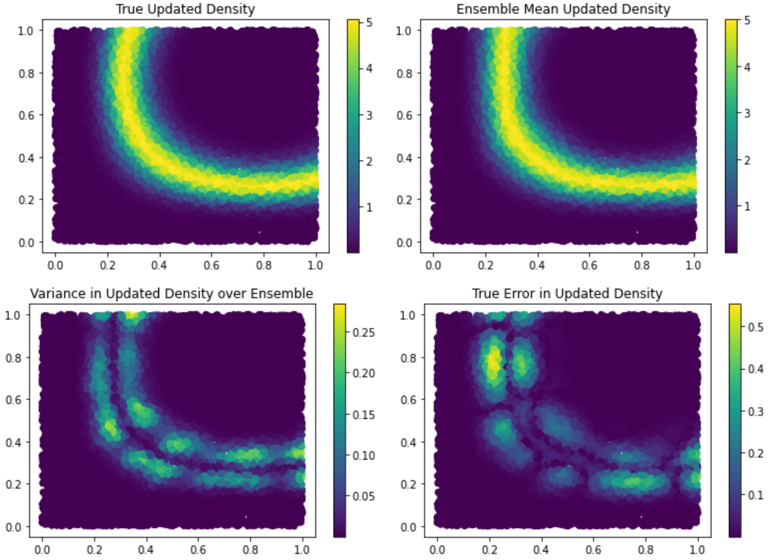
Consider the 2-dimensional map $Q : [0,1] \to \mathcal{D}$ defined by

$$Q(\lambda_1, \lambda_2) = \sin(\lambda_1)\sin(\lambda_2)$$

First, we are interested in approximating the model without noise.

- Feedforward ReLU-NN with 2 hidden layers with width of 10.
- Data-set contains uniformly distributed 1,000 samples split into 900 training and 100 test samples
- Evaluate surrogate using 10,000 uniformly distributed samples.
- Observed density is N(0.5,0.01).
- Ensemble size: 20

Example



True Updated Density
Ensemble Mean Updated Density
Variance in Updated Density over Ensemble
True Error in Updated Density
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
0.05
0.10
0.15
0.20
0.25
0.1
0.2
0.3
0.4
0.5

A perturbation ensemble to estimate interpolation error

- Suppose ϵ is an $L^\infty(\Lambda)$ error bound:

$$\sup_{\lambda \in \Lambda} \left| Q(\lambda) - \overline{Q}_S(\lambda) \right| \leq \epsilon$$

A perturbation ensemble to estimate interpolation error

- Suppose ϵ is an $L^\infty(\Lambda)$ error bound:

$$\sup_{\lambda \in \Lambda} \left| Q(\lambda) - \overline{Q}_S(\lambda) \right| \leq \epsilon$$

- Then, for each evaluation point, λ_i, we know that

$$Q(\lambda_i) \in \left[\overline{Q}_S(\lambda_i) - \epsilon, \overline{Q}_S(\lambda_i) + \epsilon \right]$$

A perturbation ensemble to estimate interpolation error

- Suppose ϵ is an $L^\infty(\Lambda)$ error bound:

$$\sup_{\lambda\in\Lambda} \left| Q(\lambda) - \overline{Q}_S(\lambda) \right| \leq \epsilon$$

- Then, for each evaluation point, λ_i, we know that

$$Q(\lambda_i) \in \left[\overline{Q}_S(\lambda_i) - \epsilon, \overline{Q}_S(\lambda_i) + \epsilon \right]$$

- We use error at test points to estimate ϵ

A perturbation ensemble to estimate interpolation error

- Suppose ϵ is an $L^\infty(\Lambda)$ error bound:

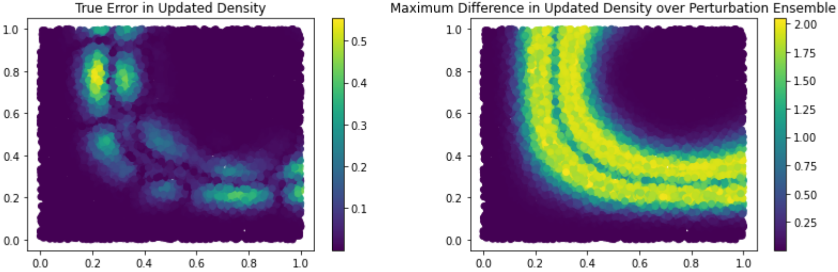
$$\sup_{\lambda \in \Lambda} \left| Q(\lambda) - \overline{Q}_S(\lambda) \right| \leq \epsilon$$

- Then, for each evaluation point, λ_i, we know that

$$Q(\lambda_i) \in \left[\overline{Q}_S(\lambda_i) - \epsilon, \overline{Q}_S(\lambda_i) + \epsilon \right]$$

- We use error at test points to estimate ϵ
- For each evaluation point, λ_i, we generate an ensemble of perturbations uniformly distributed in $[-\epsilon, \epsilon]$
- Use these to assess how interpolation errors affect $\pi_\Lambda^{\text{up}}(\lambda)$.
- Even with a small ensemble size, this should over-estimate the error in the updated density.

Example


True Error in Updated Density
Maximum Difference in Updated Density over Perturbation Ensemble
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

Incorporating Noise in the Data

Assumptions:

- Data is composed of **signal plus noise**
- Variability in the data is due to noise and intrinsic variability over Λ

Data-consistent approach in [Butler, W., Yen 2020] addresses this problem, but inverts into the joint Λ-noise space.

Incorporating Noise in the Data

Assumptions:

- Data is composed of **signal plus noise**
- Variability in the data is due to noise and intrinsic variability over Λ

Data-consistent approach in [Butler, W., Yen 2020] addresses this problem, but inverts into the joint Λ-noise space.

We seek a simpler deconvolutional approach:

1. Approximate the **signal** using NN regression
2. Approximate the **noise** using the residuals
3. Assuming Gaussian observations and noise, **deconvolve** the approximate noise from the observed distribution
4. Solve the inverse problem using the **deconvolved observed and NN model**
    - Pushforward of updated distribution **convolved** with noise = observed
5. Use an ensemble to characterize the variability in the updated density

Example

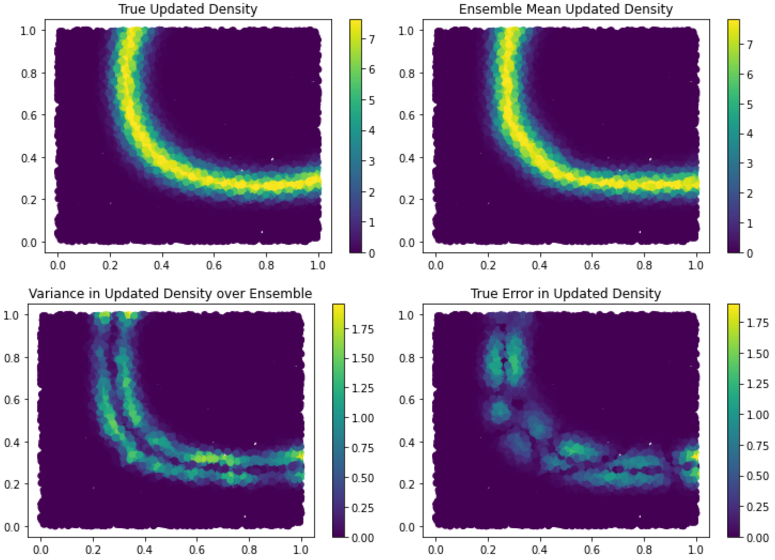
Consider the 2-dimensional map $Q : [0,1] \to \mathcal{D}$ defined by

$$Q(\lambda_1, \lambda_2) = \sin(\lambda_1)\sin(\lambda_2) + \epsilon$$

where $\epsilon \sim \mathrm{N}(0, \sigma^2_{\text{noise}})$.

- Feedforward ReLU-NN with 2 hidden layers with width of 10.
- Data-set contains uniformly distributed 1,000 samples split into 900 training and 100 test samples
- Evaluate surrogate using 10,000 uniformly distributed samples.
- Observed density is N(0.5,0.01).
- Noise is N(0.0, 0.005)

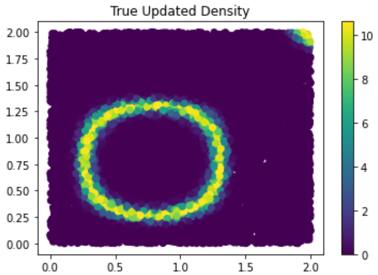
Example



True Updated Density
Ensemble Mean Updated Density
Variance in Updated Density over Ensemble
True Error in Updated Density

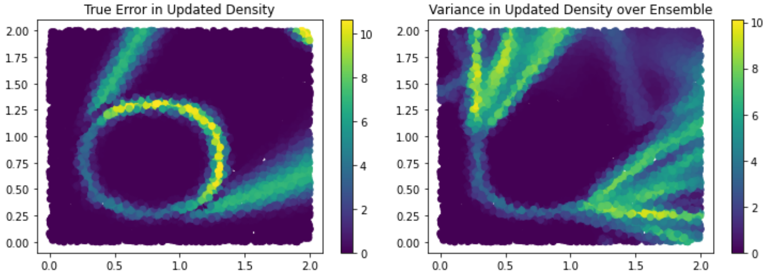
How about extrapolation (OoD) errors?

- We use the same problem trained/tested on $[0, 1]^2$
- We attempt to solve the stochastic inverse problem on $[0, 2]^2$
- Evaluate surrogate using 10,000 uniformly distributed samples on $[0, 2]^2$.



How about extrapolation (OoD) errors?

- We use the same problem trained/tested on $[0, 1]^2$
- We attempt to solve the stochastic inverse problem on $[0, 2]^2$
- Evaluate surrogate using 10,000 uniformly distributed samples on $[0, 2]^2$.



Conclusions and Future Work

- Many approaches exist for incorporating data into a model.
  - Deterministic optimization, Bayesian methods, OUU, data assimilation, etc.
- The **data-consistent inversion** approach provides an aleatoric characterization of inputs over a population/collection.
- Main computational expense is the **forward UQ problem** to obtain the push-forward of the initial density.
- We can use **data-driven models** within data-consistent inversion.
- **Errors and uncertainties** can significantly affect the solution to the inverse problem.
  - Affects the accept/reject of samples
  - Affects subsequent predictions
- If an adjoint model is available, then the affect of surrogate errors on updated density can be estimated
- We used a couple of **ensemble-based methods** to heuristically estimate a portion of the error/uncertainty.
- These **do not capture** OoD errors, but this is work in progress ...

Thanks! Questions?

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, ASCR, Early Career Research Program.

Thank you for your attention!

Questions?