Enabling Predictive Simulations for Design and Decision Making under a Limited Budget

Ilias Bilionis
School of Mechanical Engineering
Purdue University
predictivesciencelab.org

PREDICTIVE
SCIENCE LABORATORY

Talk Objectives

- Focus is on the following UQ tasks: uncertainty propagation, model calibration, and optimization.

- Quantify epistemic uncertainty on **any** UQ task induced by restrictions on the number of simulations: “the **small-n problem**”.

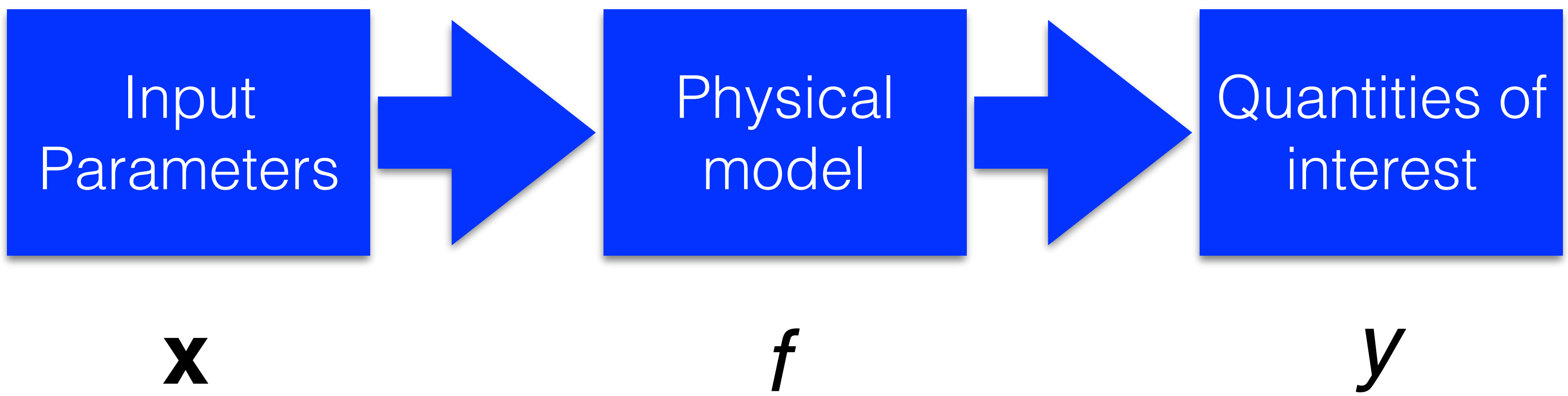
- Suggest new simulations that are “maximally informative/valuable” for a desired task.

This is collage of:

- Bilionis, I. and N. Zabaras (2012). "Multi-output local Gaussian process regression: Applications to uncertainty quantification." Journal of Computational Physics **231**(17): 5718-5746.
- Bilionis, I. and N. Zabaras (2012). "Multidimensional adaptive relevance vector machines for uncertainty quantification." SIAM Journal on Scientific Computing **34**(6): B881-B908.
- Bilionis, I., et al. (2013). "Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification." Journal of Computational Physics **241**: 212-239
- Bilionis, I. and N. Zabaras (2014). "Solution of inverse problems with limited forward solver evaluations: a Bayesian perspective." Inverse Problems 30(1).
- Kristensen, J., Bilionis, I. and N. Zabaras (2016). "Adaptive Simulation Selection for the Discovery of the Ground State Line of Binary Alloys with a Limited Computational Budget." Journal of Computational Physics (under review).
- Bilionis, I. and N. Zabaras (2016 (?)). Bayesian uncertainty propagation using Gaussian processes. Handbook of Uncertainty Quantification. R. Ghanem, D. Higdon and H. Owhadi, Springer.
- Pandita P. and Bilionis I., Extended Expected Improvement for Design Optimization Under Uncertainty (to be submitted in 2016).

PREDICTIVE
SCIENCE LABORATORY

Motivation



We'll think about it as a mathematical function:

$$y = f(\mathbf{x})$$

PREDICTIVE SCIENCE LABORATORY

Some of the Problems of Uncertainty Quantification

- Uncertainty propagation:

$$p(x)\stackrel{f}{\rightarrow}p(y)$$

- Model calibration:

$$y\stackrel{f}{\rightarrow}p(x \mid y)$$

- (Multi-objective) optimization under uncertainty:

$$\max_x \mathbf{E}[f_i(x)], i=1,\dots,m$$

PREDICTIVE
SCIENCE LABORATORY

Why are these problems difficult?

- High computational cost of models.
- High-dimensionality of inputs/outputs.
- Fusion of information from multiple sources.
- Quantification of model-form uncertainties.
- Heteroscedastic (input-dependent) noise.
- …

The Surrogate Idea

- Do a finite number of simulations.
- Replace model with an approximation:

$$y \approx \hat{f}(\mathbf{x})$$

- The surrogate is usually cheap to evaluate.
- Solve the UQ problem with the surrogate.

PREDICTIVE SCIENCE LABORATORY

Classic Approach to Surrogates

- Usually

$$\hat{f}(\mathbf{x}) = \sum_{j=1}^{M} w_j \phi_j(\mathbf{x})$$

- with weights by looking at :

$$\mathcal{D} = \left\{\left(x^{(n)}, f\left(x^{(n)}\right)\right)\right\}_{n=1}^{N}$$

- using either a quadrature rule (orthogonal basis), least squares, or machine learning techniques.

PREDICTIVE
SCIENCE LABORATORY

Examples of Surrogates

- generalized polynomial chaos
- Fourrier expansions
- splines
- wavelets
- neural networks
- support vector machines
- compressive sensing
- …

PREDICTIVE
SCIENCE LABORATORY

Issues of (Classic) Surrogates

- inability to quantify epistemic uncertainties due to limited number of observations
- high-dimensionality
- rare events
- ….

PREDICTIVE
SCIENCE LABORATORY

The Bayesian Approach

- Put prior on functions.
- Evaluate model output on a finite set of inputs.
- Compute the posterior on functions.
- Use posterior to quantify epistemic uncertainty on anything.

"Most people, even Bayesians, think that this sounds crazy when they first hear about it".

-Persi Diaconis (1988)

PREDICTIVE
SCIENCE LABORATORY

Some History

- (Poincaré, 1912): interpolating a real function, **first Gaussian process**.
- (O'Hagan, 1987; Diaconis, 1988): **integration**.
- (O'Hagan, 1991; Kennedy et al., 1996; Kennedy, 1996; Minka, 2000): **Bayesian quadrature**.
- (Haylock et al., 1996; O'Hagan et al., 1999; Oakley et al., 2002): **Uncertainty propagation**.
- Bilionis and Zabaras (4 pubs in 2012-2014): **General uncertainty propagation**. Summary soon in springer chapter on "Bayesian Uncertainty Propagation".
- **Probabilistic numerics** (Hennig et al, 2015), general principle. Applications:
  - **Sensitivity analysis** (Oakley et al., 2004; Becker et al., 2012; Daneshkhah et al., 2013).
  - **Model calibration** (Bilionis et al. 2014).
  - **Linear algebra** (Hennig, 2015).
  - **Ordinary differential equations** (Skilling, 1992; Graepel, 2003; Calderhead et al., 2009; Chkrebtii et al. 2013; Korostil et al., 2013; Barber, 2014; Hennig et al., 2014; Schober et al., 2014, …)
  - **Optimization** (Hennig et al., 2012; Hennig, 2013)
  - …

PREDICTIVE
SCIENCE LABORATORY

Priors on Functions

Prior on functions =~ Gaussian process

PREDICTIVE
SCIENCE LABORATORY

Priors on Functions

A Gaussian process needs two ingredients:

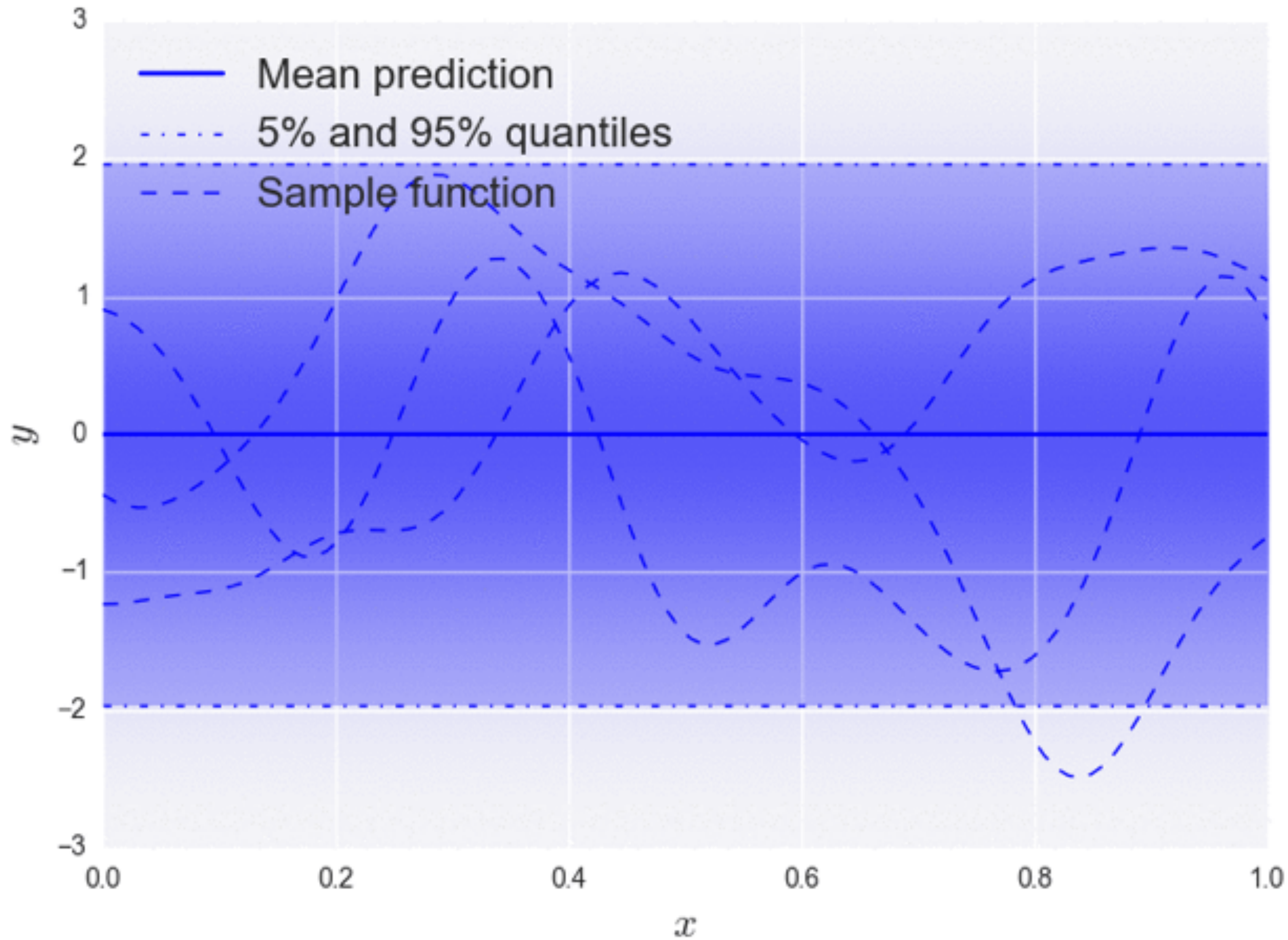
- a mean function
- a covariance function

It uses them to define a probability measure on the space of functions.

We write: $$f(\cdot) \sim p(f(\cdot)) = \mathrm{GP}\left(f(\cdot) \mid m(\cdot), k(\cdot,\cdot)\right)$$

PREDICTIVE
SCIENCE LABORATORY

Priors on Functions



PREDICTIVE
SCIENCE LABORATORY

Priors on Functions

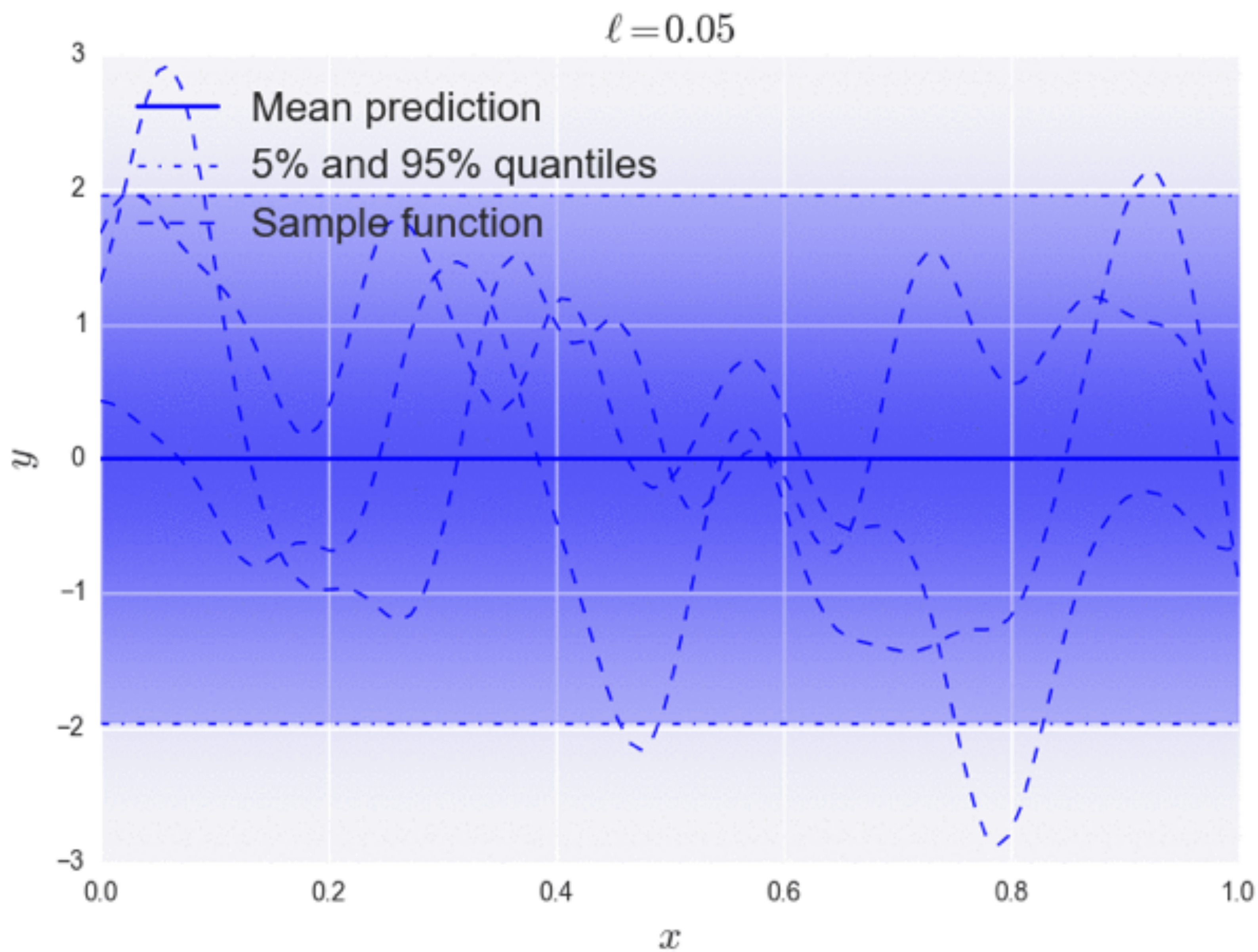
- We write:

$$f(\cdot) \sim p(f(\cdot)) \sim \mathrm{GP}\left(f(\cdot) \mid m(\cdot), k(\cdot,\cdot)\right)$$

- and we interpret:
  - *m(x)*: What do I think *f(x)* could be?
  - *k(x, x)*: How sure am I about my expectation of *f(x)*?
  - *k(x, x')*: How similar are *f(x)* and *f(x')*?

PREDICTIVE SCIENCE LABORATORY

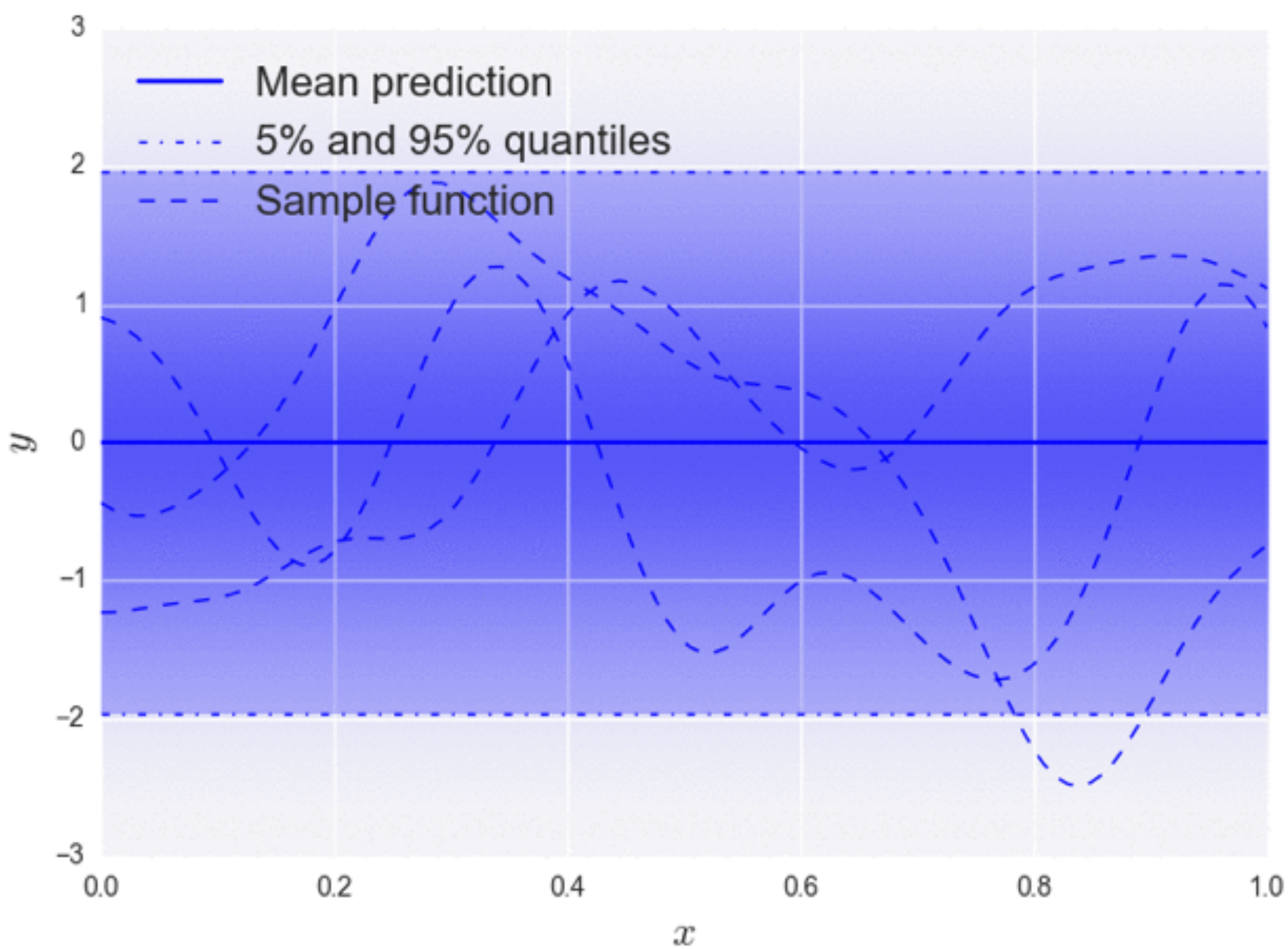
Changing the length scale



PREDICTIVE
SCIENCE LABORATORY

The samples are as smooth as the covariance

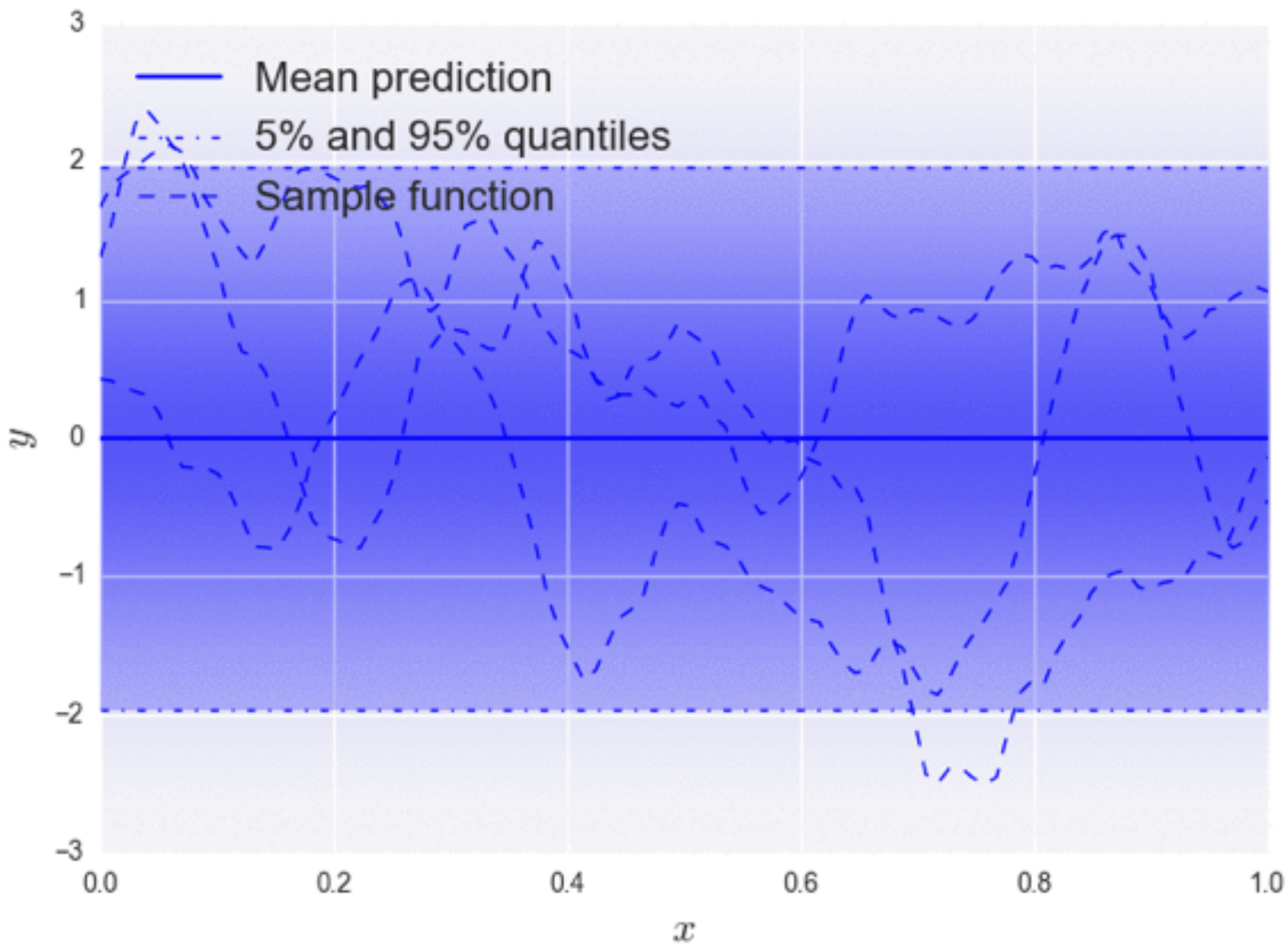
Infinitely smooth SE covariance



PREDICTIVE
SCIENCE LABORATORY

The samples are as smooth as the covariance

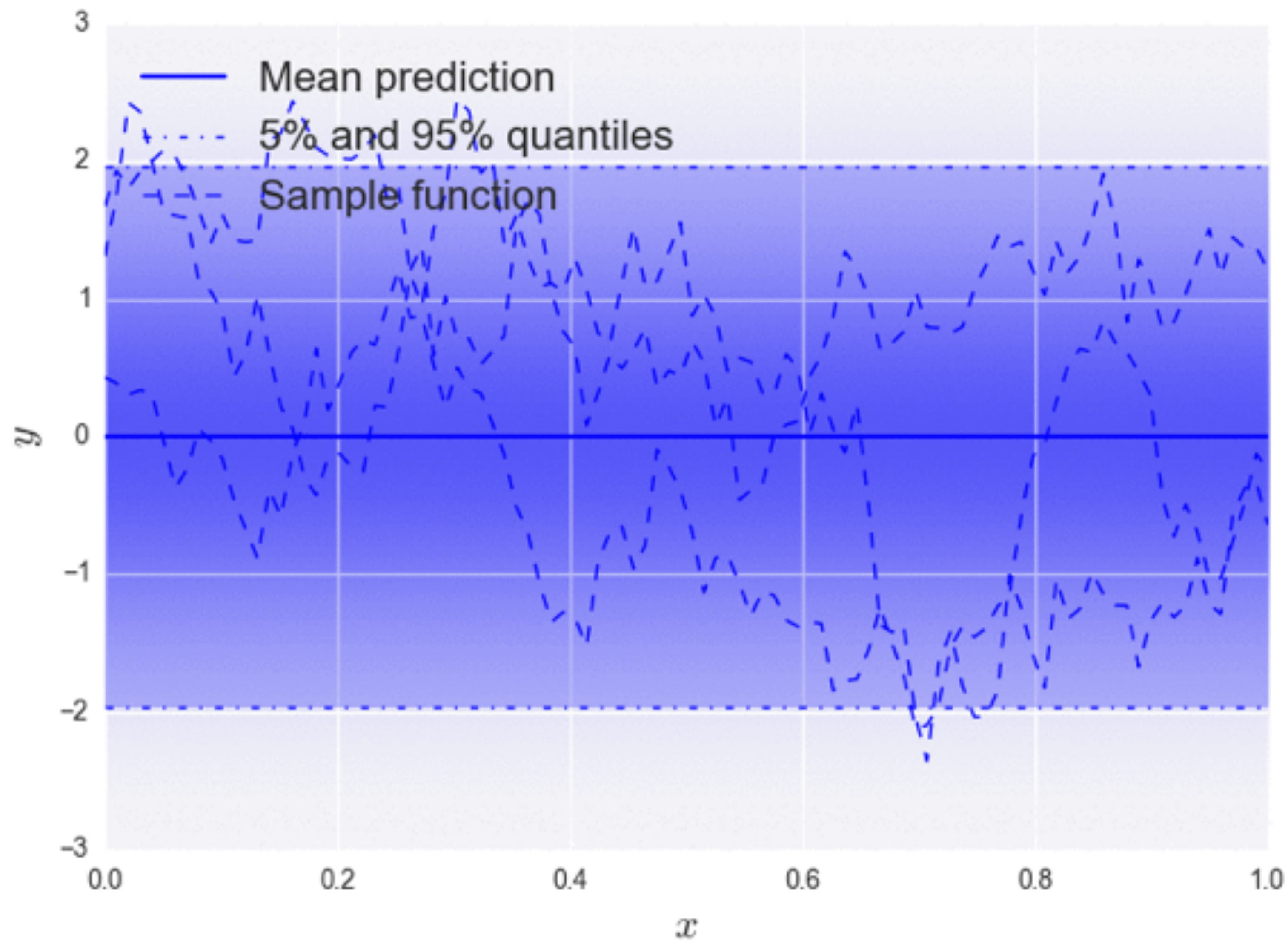
Matern 2-3, 2 times differentiable



PREDICTIVE
SCIENCE LABORATORY

The samples are as smooth as the covariance

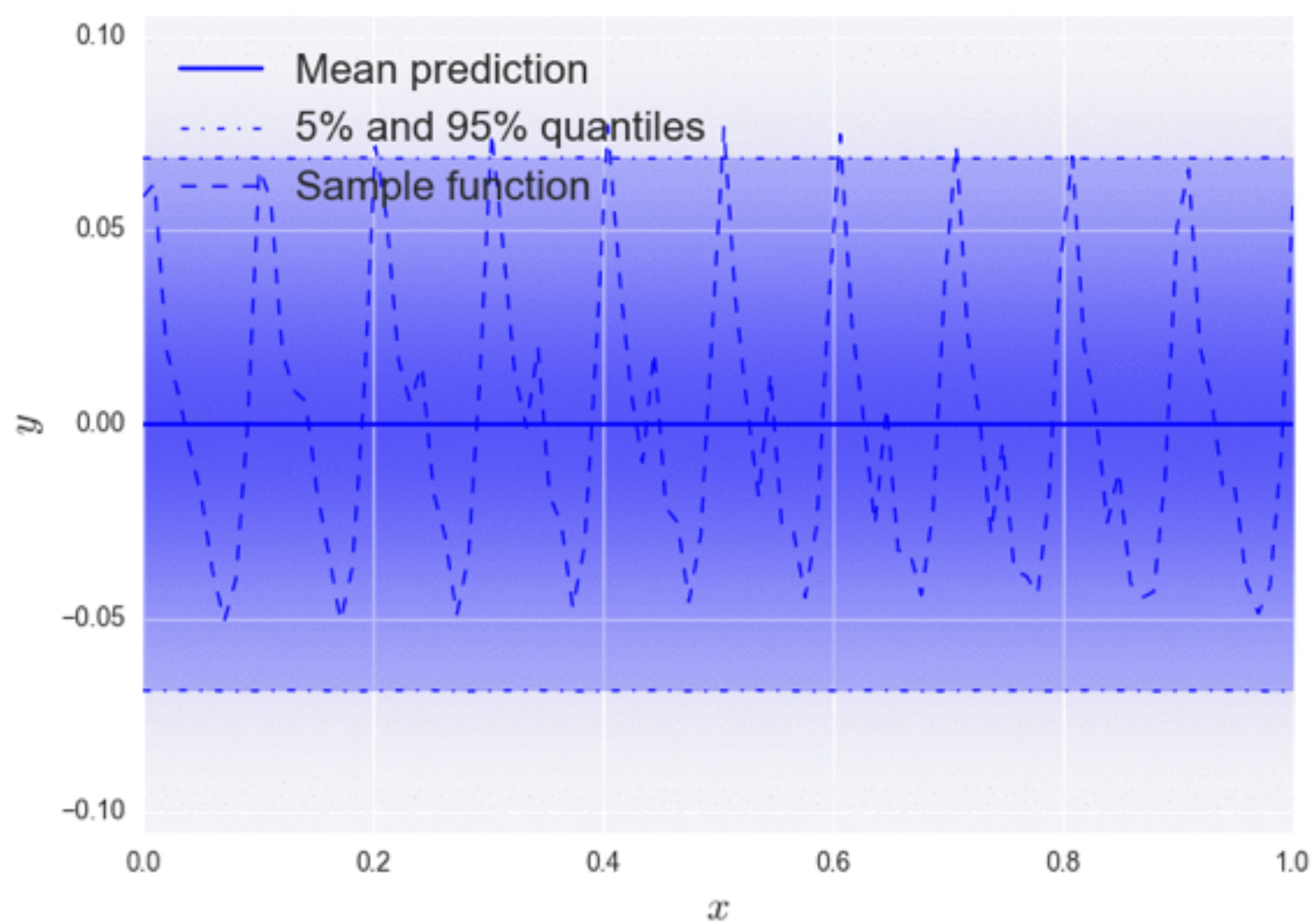
Exponential, continuous, nowhere differentiable



PREDICTIVE
SCIENCE LABORATORY

Invariances may be built-into covariance functions

Periodic Exponential, period = 0.1



PREDICTIVE
SCIENCE LABORATORY

What about known physics?

$$\mathcal{L}[f](\mathbf{x}) = g(\mathbf{x})$$

$$\mathcal{B}_i[f](\mathbf{x}) = g_i(\mathbf{x}), i = 1,\dots,m$$

If all operators are linear (even if stochastic), then it is possible to constrain the GP prior to approximately satisfy them.

(unpublished, let's talk offline)

PREDICTIVE
SCIENCE LABORATORY

Gaussian process regression

- Assume that we have observed:

$$\mathcal{D} = \left\{ \left(x^{(n)}, f\left(x^{(n)} \right) \right) \right\}_{n=1}^{N}$$

- and that we want to construct the posterior probability measure in the space of models:

$$f(\cdot) | \mathcal{D} \sim p(f(\cdot) | \mathcal{D})$$

PREDICTIVE
SCIENCE LABORATORY

The posterior Gaussian process

- The posterior measure is also a Gaussian process:

$$f(\cdot) | \mathcal{D} \sim \mathrm{GP}\left(f(\cdot) | \tilde{m}(\cdot), \tilde{k}(\cdot,\cdot)\right),$$

$$\tilde{m}(\mathbf{x}) = m(\mathbf{x}) + \mathbf{K}(\mathbf{x},\mathbf{X})\mathbf{K}^{-1}(\mathbf{f} - \mathbf{m}),$$

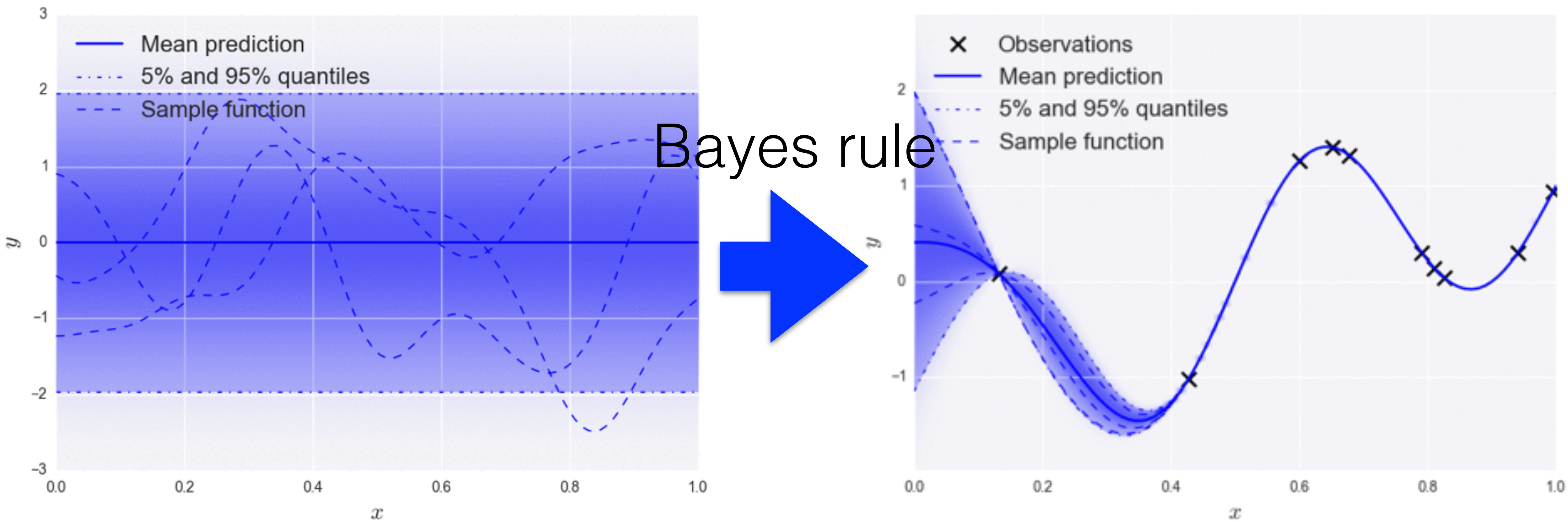
$$\tilde{k}(\mathbf{x},\mathbf{x}') = k(\mathbf{x},\mathbf{x}') - \mathbf{K}(\mathbf{x},\mathbf{X})\mathbf{K}^{-1}\mathbf{K}(\mathbf{X},\mathbf{x}')$$

- This encodes are beliefs about the model output after seeing the data.

- The **only math challenge** is drawing samples from the posterior that are analytic functions (Bilionis and Zabaras, 2016).

PREDICTIVE
SCIENCE LABORATORY

Gaussian process regression



Prior GP

Posterior GP

PREDICTIVE
SCIENCE LABORATORY

Example 1: Bayesian Uncertainty Propagation

References:

- Bilionis, I. and N. Zabaras (2012). "Multi-output local Gaussian process regression: Applications to uncertainty quantification." Journal of Computational Physics **231**(17): 5718-5746.

- Bilionis, I. and N. Zabaras (2012). "Multidimensional adaptive relevance vector machines for uncertainty quantification." SIAM Journal on Scientific Computing **34**(6): B881-B908.

- Bilionis, I., et al. (2013). "Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification." Journal of Computational Physics **241**: 212-239.

- Bilionis, I. and N. Zabaras (2016 (?)). Bayesian uncertainty propagation using Gaussian processes. Handbook of Uncertainty Quantification. R. Ghanem, D. Higdon and H. Owhadi, Springer.

PREDICTIVE
SCIENCE LABORATORY

Example 1: Uncertainty Propagation

A *statistic* Q can be thought of as a operator acting on the model f.

For example:

$$Q_\mu[f(\cdot)] := \mathbb{E}_{\mathbf{x}}[f(\mathbf{x})] = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

$$Q_v[f(\cdot)] := \mathbb{V}_{\mathbf{x}}[f(\mathbf{x})] = \int (f(\mathbf{x}) - \mathbb{E}_{\mathbf{x}}[f(\cdot)])^2 p(\mathbf{x})d\mathbf{x}$$

$$\dots$$

PREDICTIVE
SCIENCE LABORATORY

Example 1: Uncertainty Propagation

Put prior on models:

$$f(\cdot) \sim p(f(\cdot))$$

Observe:

$$\mathcal{D} = \left\{ \left(x^{(n)}, f\left(x^{(n)} \right) \right) \right\}_{n=1}^{N}$$

Find posterior on models:

$$f(\cdot) \mid \mathcal{D} \sim p(f(\cdot) \mid \mathcal{D})$$

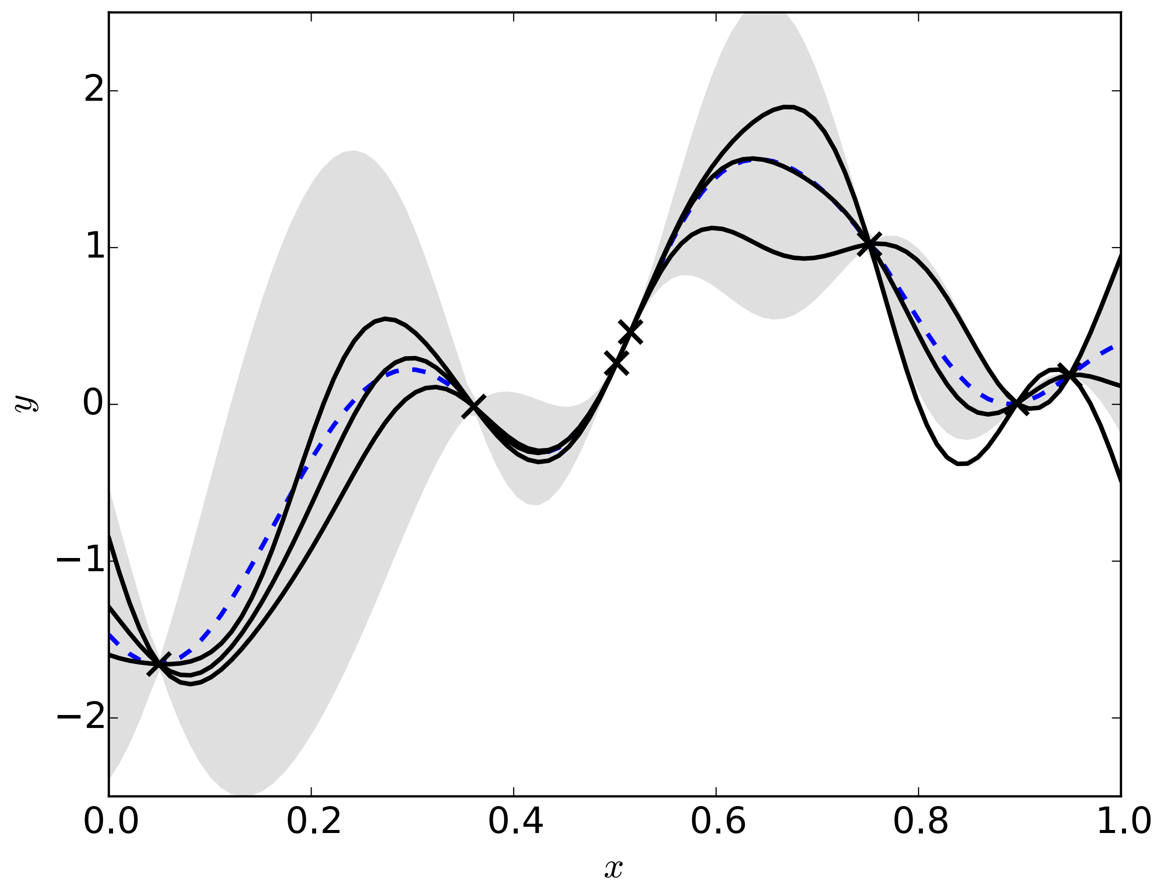
Find posterior on statistics:

If Q non-linear, only via sampling.
See (Bilionis et al., 2016) for details on sampling functions.

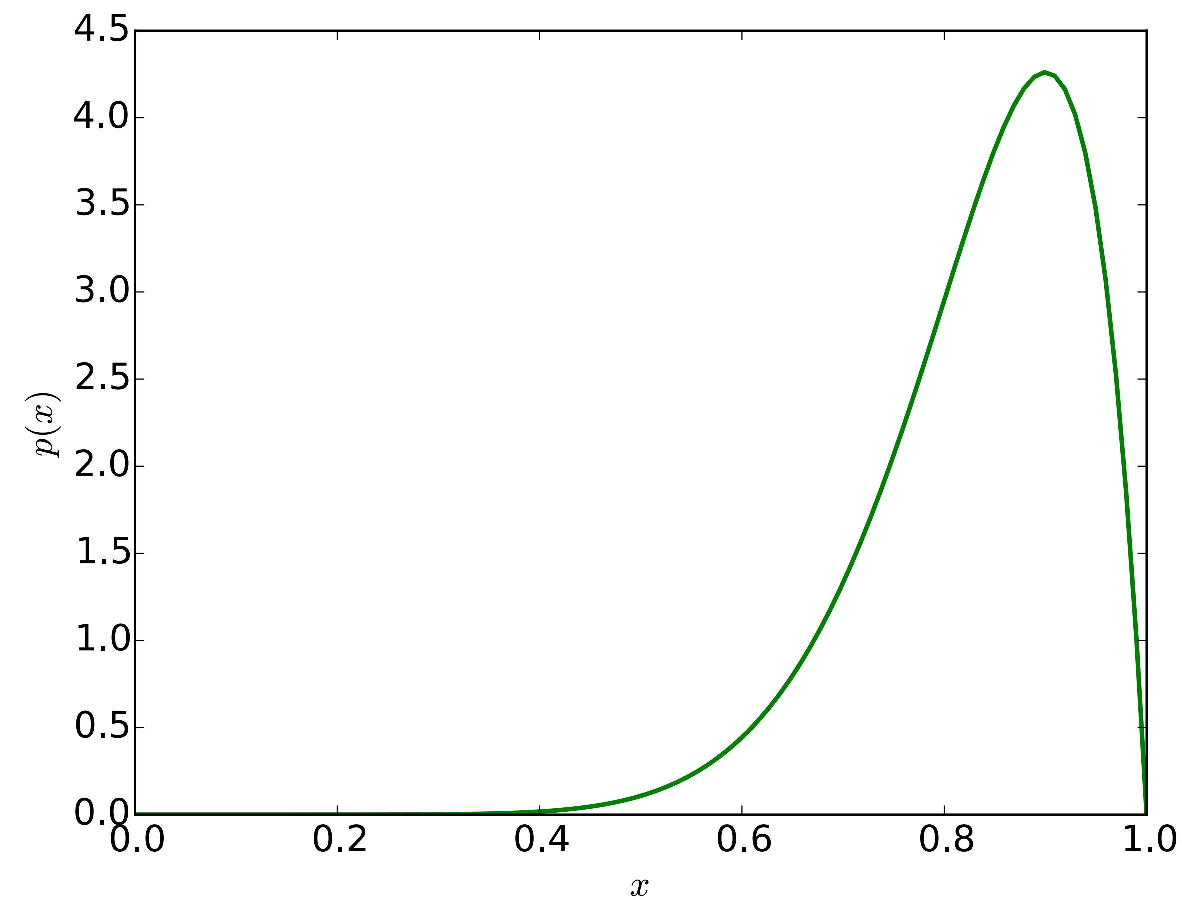
$$p(Q \mid \mathcal{D}) = \int \delta(Q - Q[f(\cdot)]) p(f(\cdot) \mid \mathcal{D}) df(\cdot).$$

PREDICTIVE
SCIENCE LABORATORY

Example 1: Uncertainty Propagation



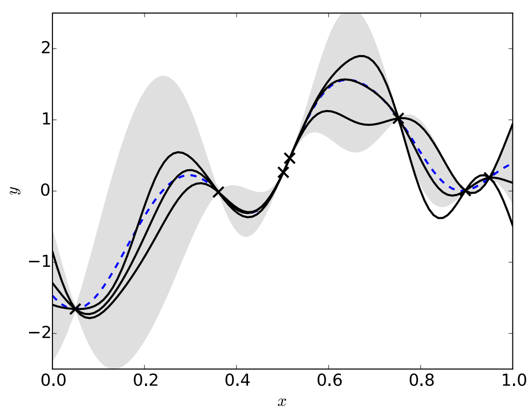
$f(\cdot) \mid \mathcal{D} \sim p(f(\cdot) \mid \mathcal{D})$



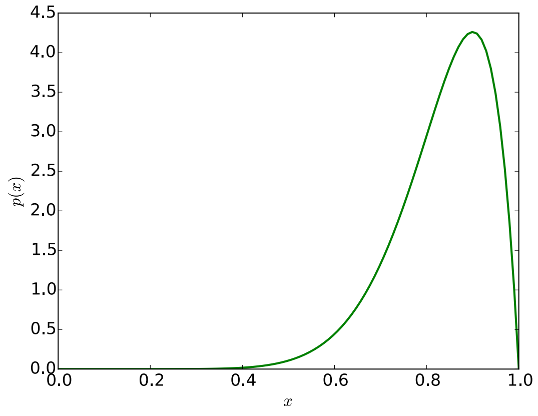
$p(\mathbf{x})$

PREDICTIVE
SCIENCE LABORATORY

Example 1: Uncertainty Propagation

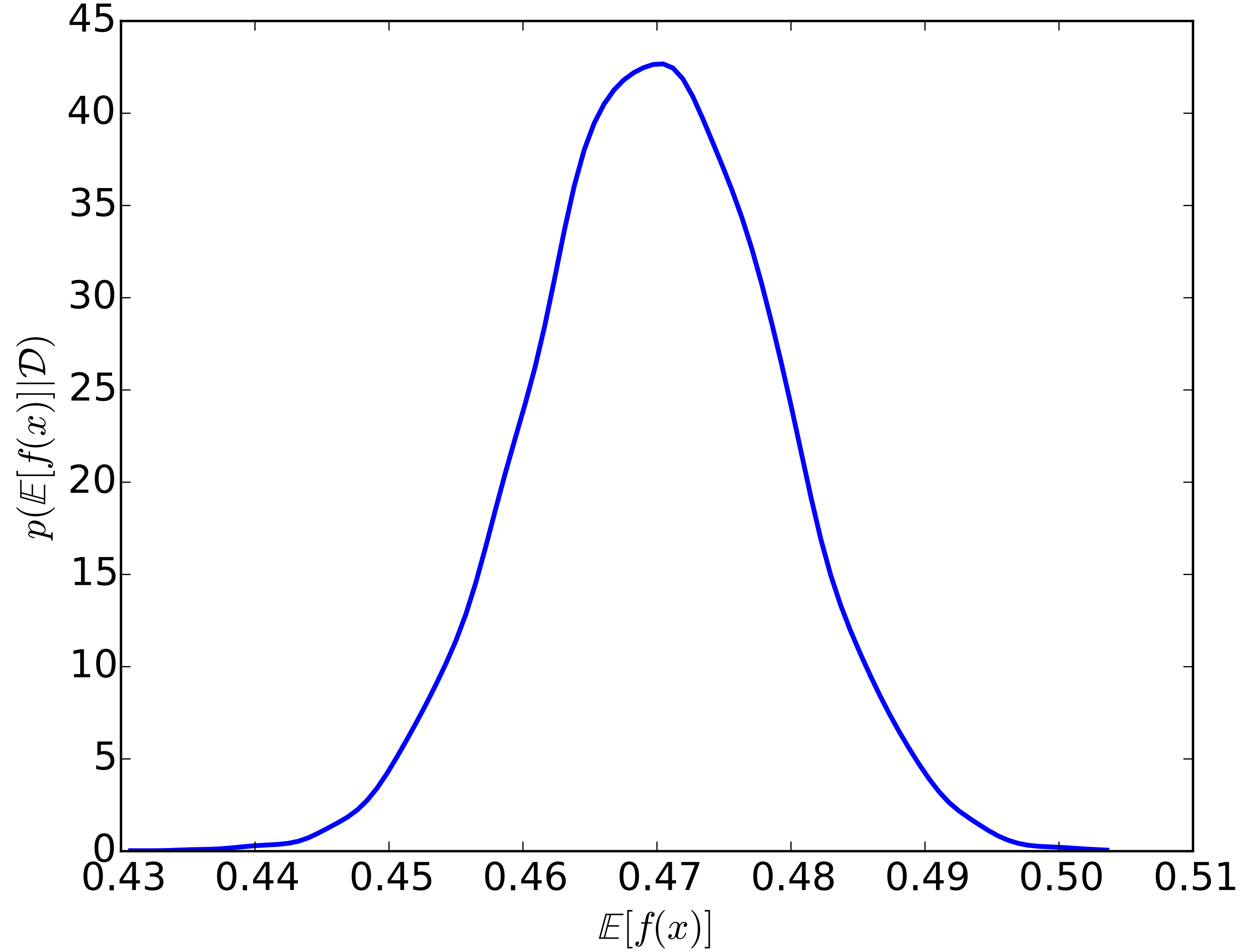


$f(\cdot) | \mathcal{D} \sim p(f(\cdot) | \mathcal{D})$



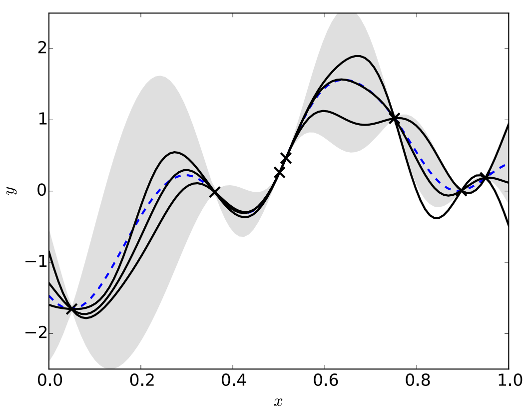
$p(\mathbf{x})$

$p(\mathbb{E}[f(\mathbf{x})] | \mathcal{D})$

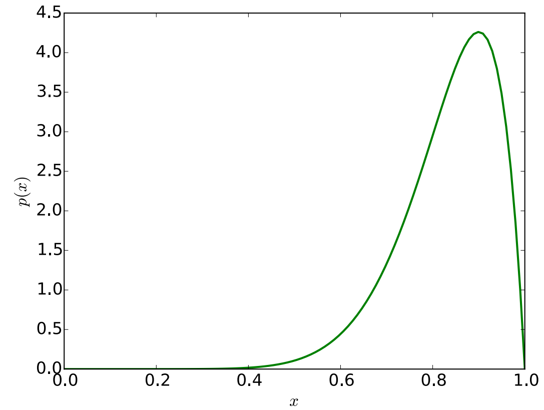


PREDICTIVE
SCIENCE LABORATORY

Example 1: Uncertainty Propagation

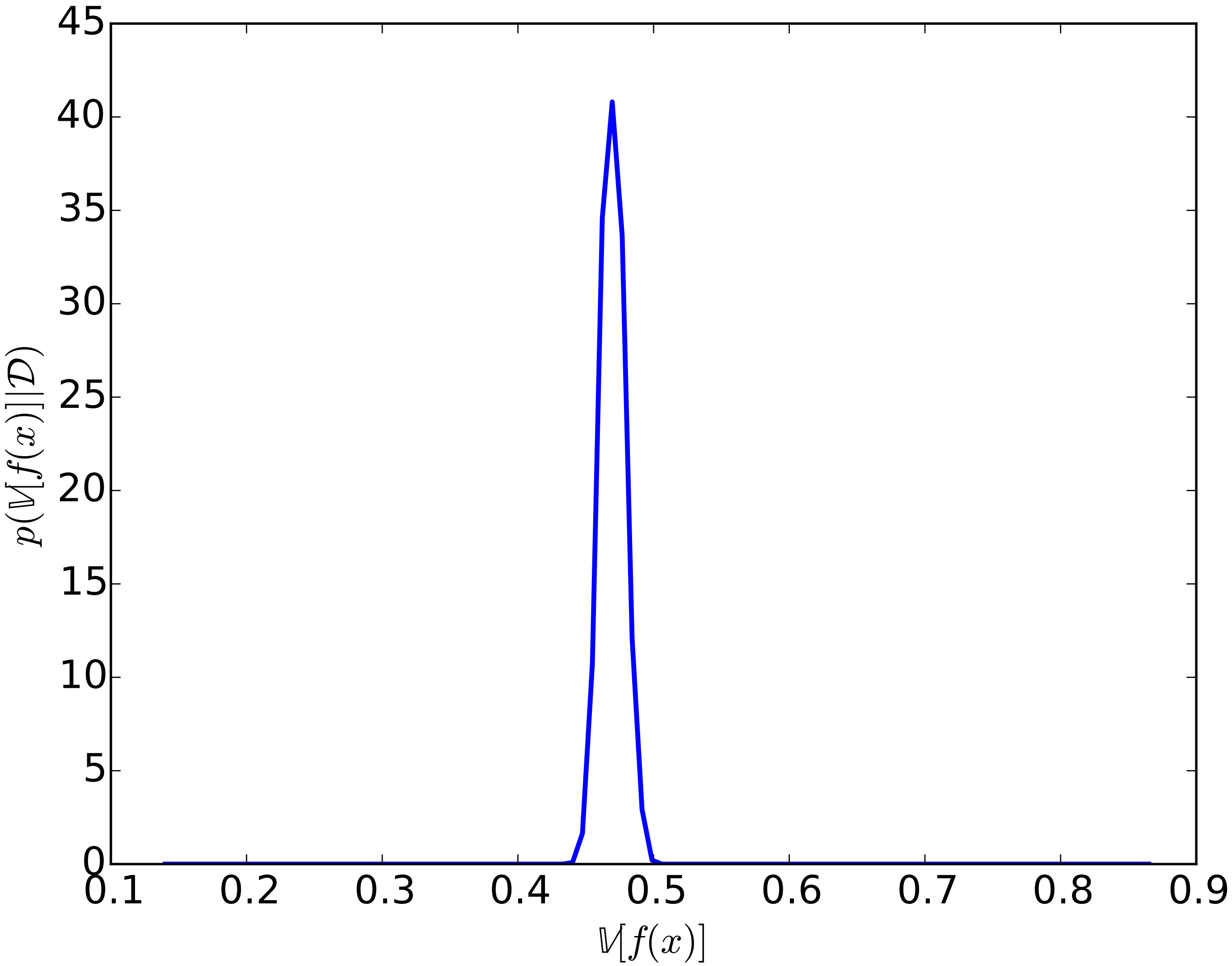


$f(\cdot) \mid \mathcal{D} \sim p(f(\cdot) \mid \mathcal{D})$



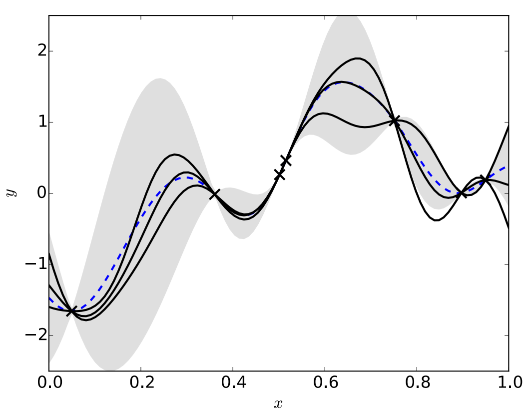
$p(\mathbf{x})$

$p(\mathbb{V}[f(\mathbf{x})] \mid \mathcal{D})$

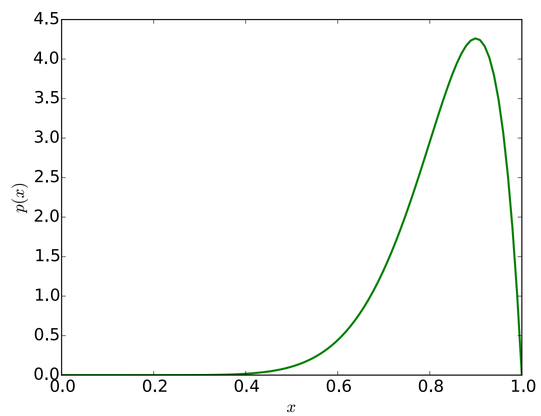


PREDICTIVE
SCIENCE LABORATORY

Example 1: Uncertainty Propagation

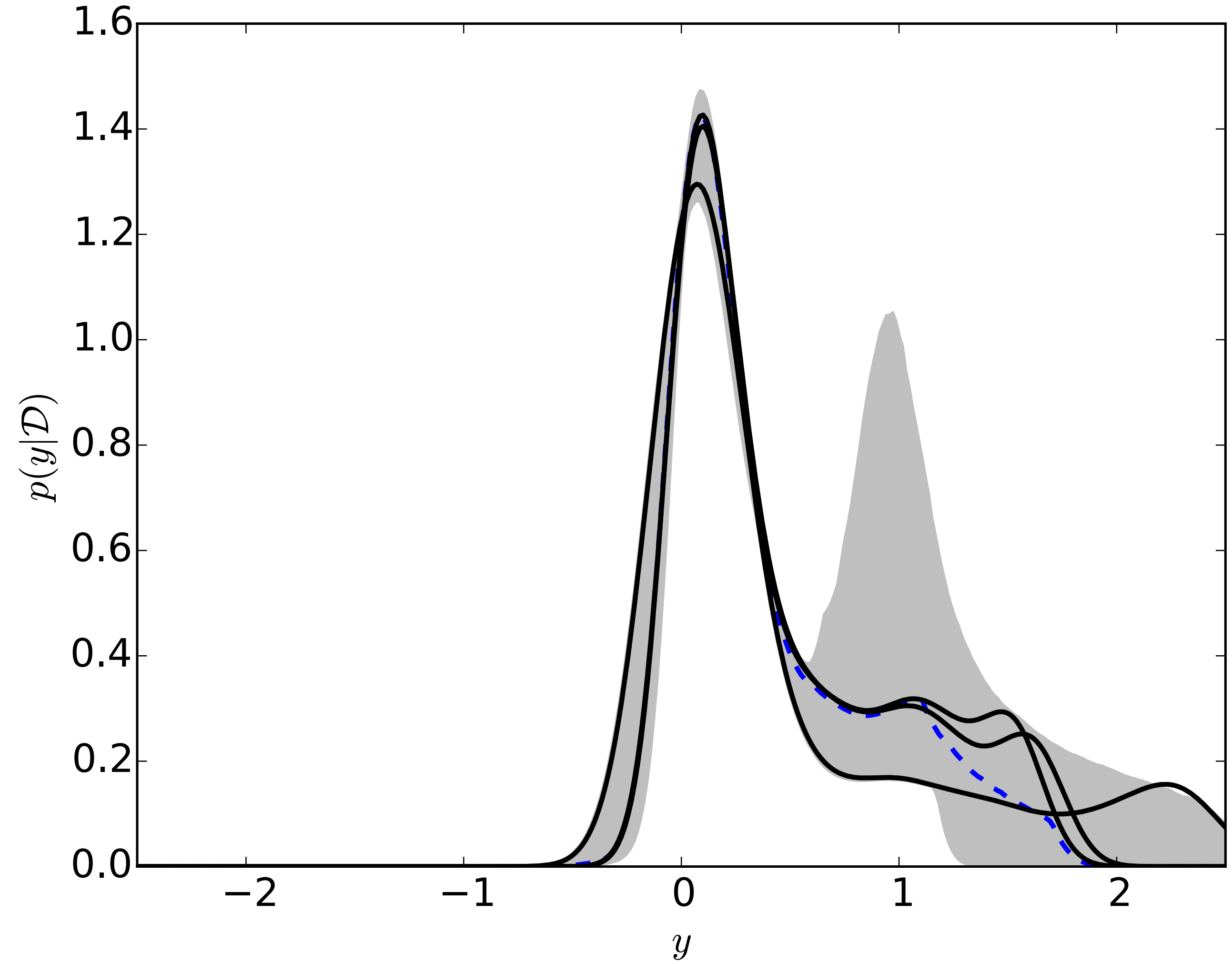


$f(\cdot) | \mathcal{D} \sim p(f(\cdot) | \mathcal{D})$



$p(\mathbf{x})$

$p(p(y) | \mathcal{D})$



You can even have characterize the uncertainty in the PDF of the output.

PREDICTIVE
SCIENCE LABORATORY

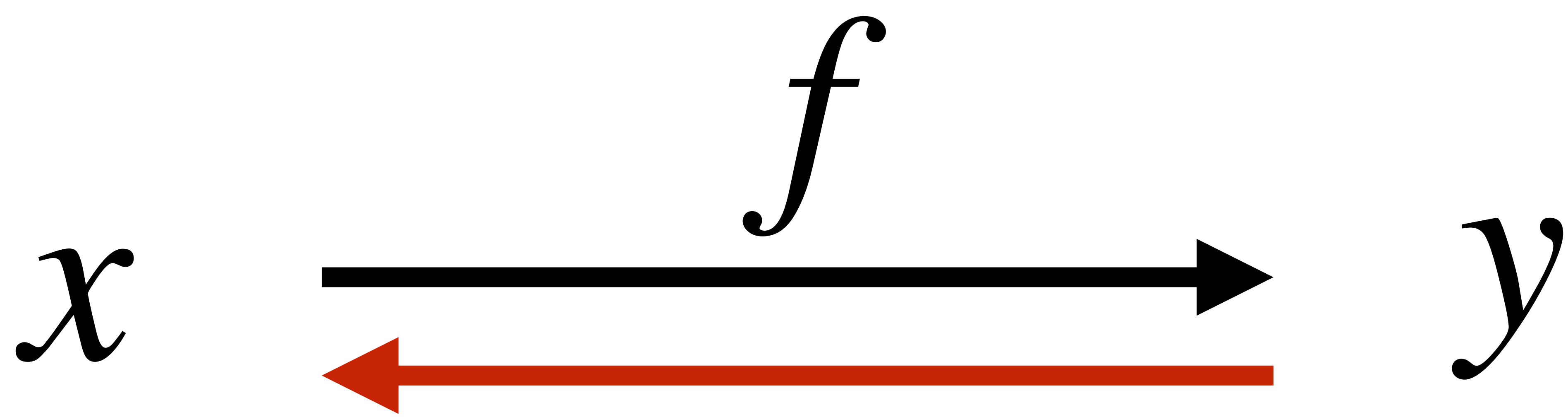
Example 2: Model Calibration

References:

- Bilionis, I. and N. Zabaras (2014). "Solution of inverse problems with limited forward solver evaluations: a Bayesian perspective." Inverse Problems 30(1).

PREDICTIVE
SCIENCE LABORATORY

Example 2: Model Calibration



$$\mathcal{D} = \left\{ \left(x^{(n)}, f\left(x^{(n)} \right) \right) \right\}_{n=1}^{N}$$

What is the best you can say about **x**?

PREDICTIVE
SCIENCE LABORATORY

Example 2: Model Calibration

$$\underbrace{p(x \mid y, f)}_{\text{Posterior}} \propto \underbrace{p(y \mid f(x))}_{\text{Likelihood}} \underbrace{p(x)}_{\text{Prior}}$$

Posterior
(knowledge)

Likelihood
(measurement)

Prior
(knowledge)

Part of modeling (GIGO)

Everything is conditional on *non-linear* and *expensive* model (simulator)

The "only" way to characterize it is through sampling.

PREDICTIVE SCIENCE LABORATORY

Example 2: Model Calibration

$$p(x \mid y, \cancel{f}\,\hat{f}) \propto p(y \mid \cancel{f}\,\hat{f}(x))\,p(x)$$

$$\mathcal{D} = \left\{\left(x^{(n)}, f\left(x^{(n)}\right)\right)\right\}_{n=1}^{N}$$

Learn surrogate: $\hat{f}(x)$

… and do everything with the surrogate

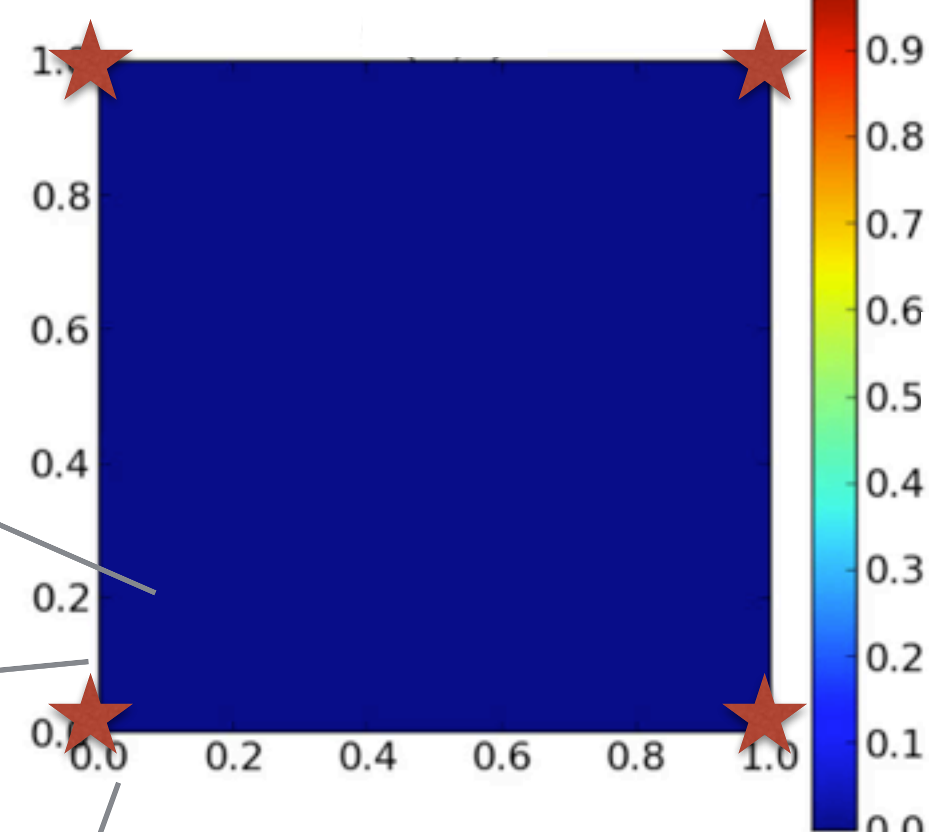
PREDICTIVE
SCIENCE LABORATORY

Example 2: Model Calibration

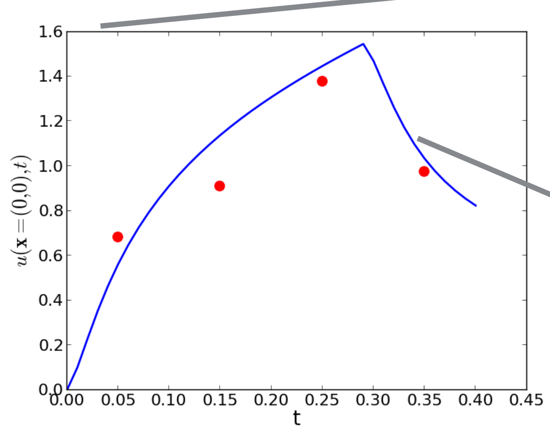
$$\frac{\partial u}{\partial t} = \nabla u + S, \text{ in } [0,1]^2$$

$$\nabla u \cdot n = 0, \text{ on } \partial[0,1]^2$$

What is the location **x** of the contamination source?



Experimental observations **y**

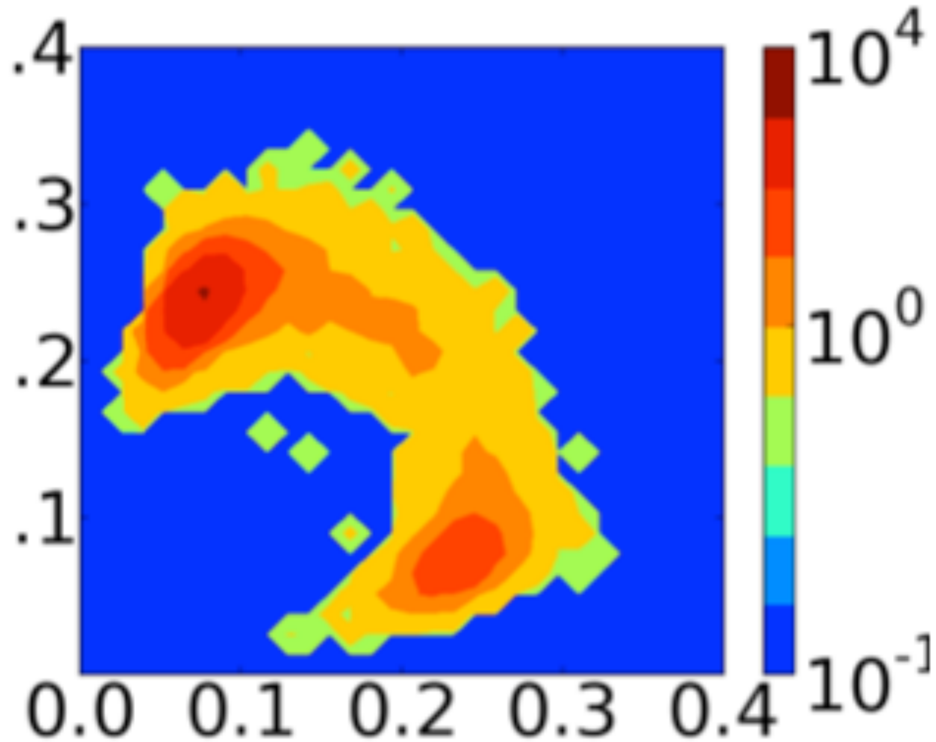


$f(x)$

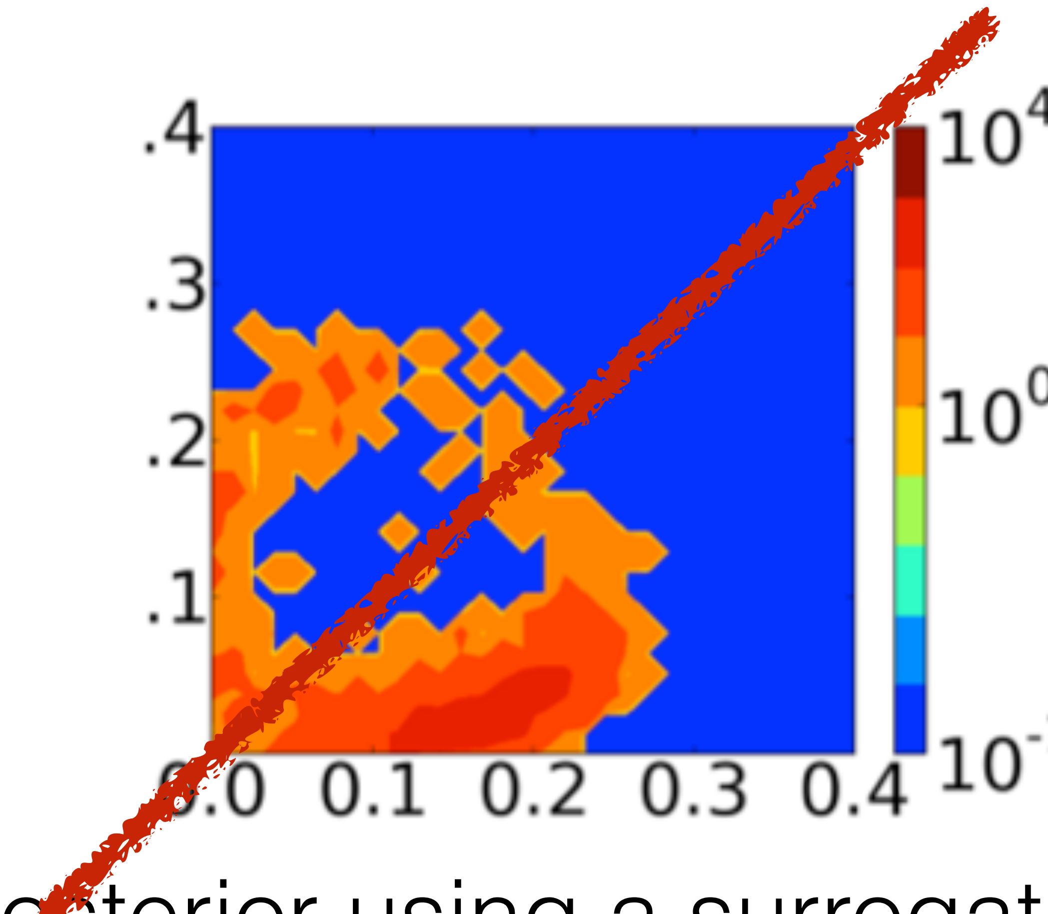
PREDICTIVE
SCIENCE LABORATORY

Example 2: Model Calibration

Example: Contaminant Source Identification



True posterior



Posterior using a surrogate based on 40 simulations

PREDICTIVE SCIENCE LABORATORY

Example 2: Model Calibration

What is the best we can say about the solution of the inverse problem given only the simulations we have already made?

$$\mathcal{D} = \left\{ \left(x^{(n)}, f\left(x^{(n)} \right) \right) \right\}_{n=1}^{N}$$

$$p(x \mid y, \mathcal{D}) = ?$$

PREDICTIVE
SCIENCE LABORATORY

Example 2: Model Calibration

Put prior on models:

$$f(\cdot) \sim p(f(\cdot))$$

Observe:

$$\mathcal{D} = \left\{(\mathbf{x}_i, y_i)\right\}_{i=1}^{N}$$

Find posterior on models:

$$f(\cdot) \mid \mathcal{D} \sim p(f(\cdot) \mid \mathcal{D})$$

Integrate model response out of the likelihood:

$$p(\mathbf{x} \mid y, \mathcal{D}) \propto \int p(y \mid \mathbf{x}, f(\mathbf{x})) p(f(\mathbf{x}) \mid \mathcal{D}) df(\mathbf{x}) p(\mathbf{x})$$

PREDICTIVE SCIENCE LABORATORY

Example 2: Model Calibration

Instead of:

$$p(\mathbf{x} \mid y, \mathcal{D}) \propto \int p(y \mid \mathbf{x}, f(\mathbf{x})) p(f(\mathbf{x}) \mid \mathcal{D}) df(\mathbf{x}) p(\mathbf{x})$$

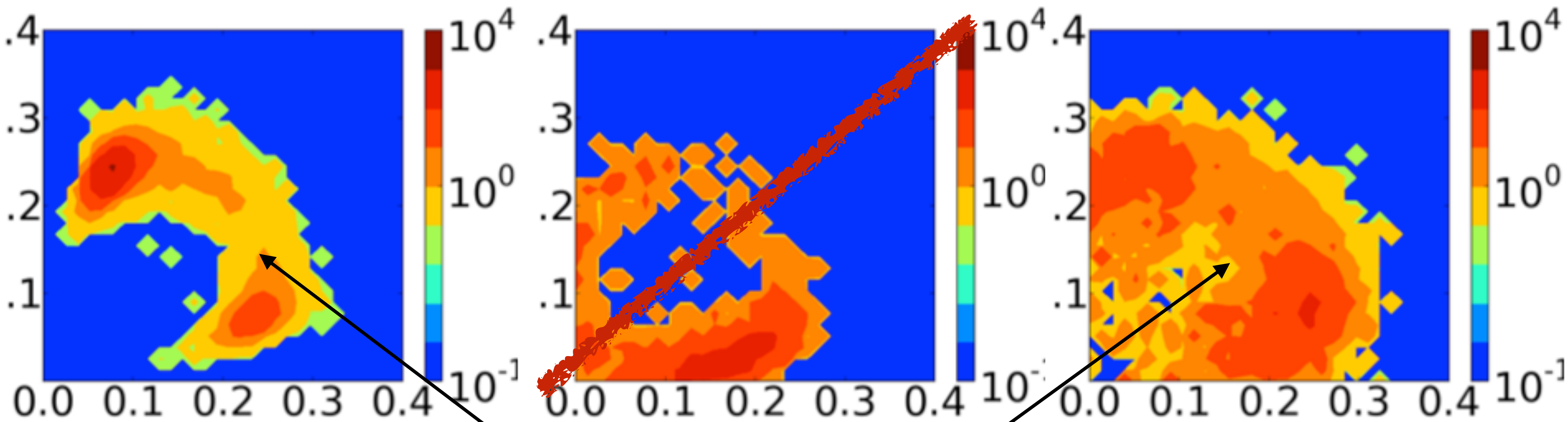
We just sample directly in the joint space:

$$p(\mathbf{x}, f(\mathbf{x}) \mid y, \mathcal{D}) \propto p(y \mid \mathbf{x}, f(\mathbf{x})) p(f(\mathbf{x}) \mid \mathcal{D}) p(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Example 2: Model Calibration

Using 40 forward model evaluations:



True posterior

Replace by surrogate

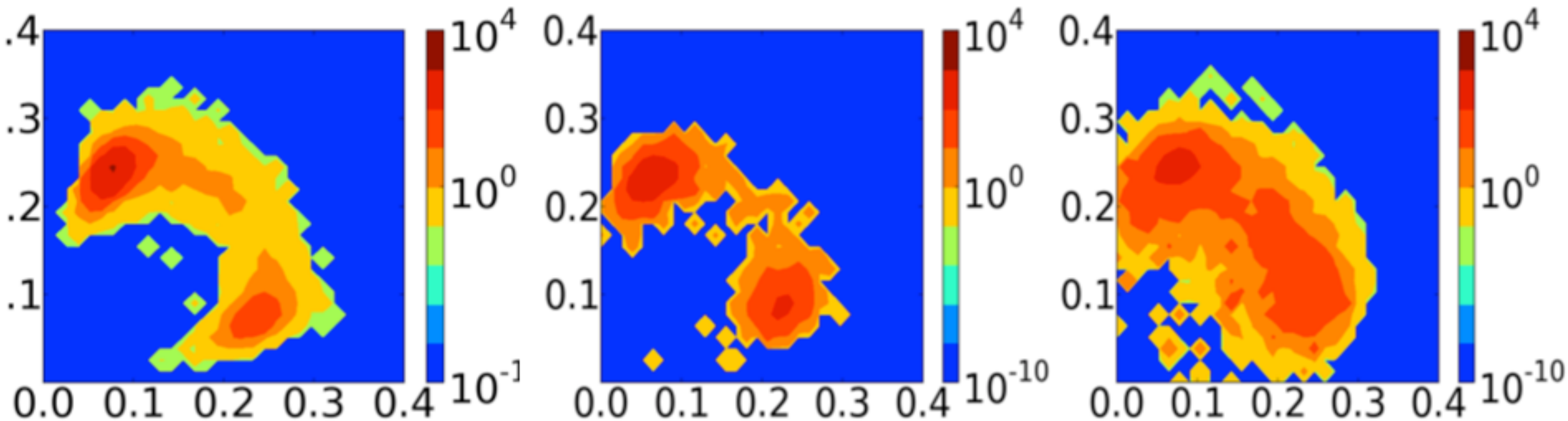
Limited data formulation

Difference is epistemic uncertainty due to limited data

PREDICTIVE
SCIENCE LABORATORY

Example 2: Model Calibration

Using 80 forward model evaluations:



True posterior

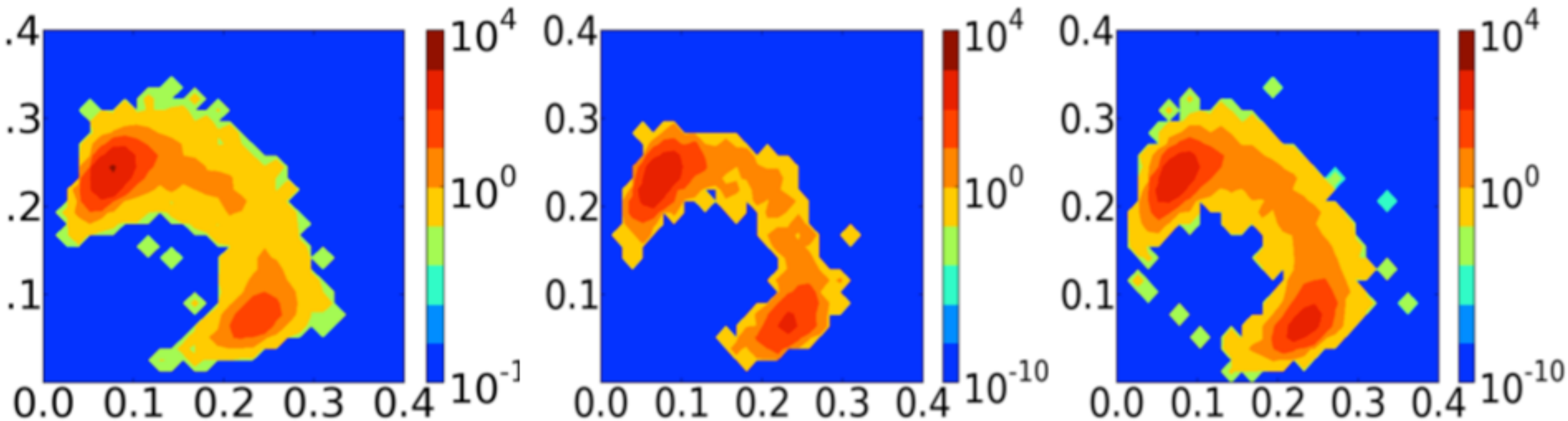
Replace by surrogate

Limited data formulation

PREDICTIVE
SCIENCE LABORATORY

Example 2: Model Calibration

Using 160 forward model evaluations:



True posterior | Replace by surrogate | Limited data formulation

PREDICTIVE
SCIENCE LABORATORY

Example 3: Optimization

References:

- Pandita P. and Bilionis I., Extended Expected Improvement for Design Optimization Under Uncertainty (to be submitted in 2016).

Example 3: Optimization

$$\mathbf{x}^* = \mathrm{argmin}_{\mathbf{x}}\, \mathbf{E}[f(\mathbf{x})]$$

What can you say about the solution of this problem with 5-10 evaluations of f(x)?

PREDICTIVE
SCIENCE LABORATORY

Example 3: Optimization

Put prior on models:

$$f(\cdot) \sim p(f(\cdot))$$

Observe:

$$\mathcal{D} = \left\{\left(x^{(n)}, f\left(x^{(n)}\right)\right)\right\}_{n=1}^{N}$$

Find posterior on models:

$$f(\cdot) \mid \mathcal{D} \sim p(f(\cdot) \mid \mathcal{D})$$

Find posterior of the location of the maximum, etc.:

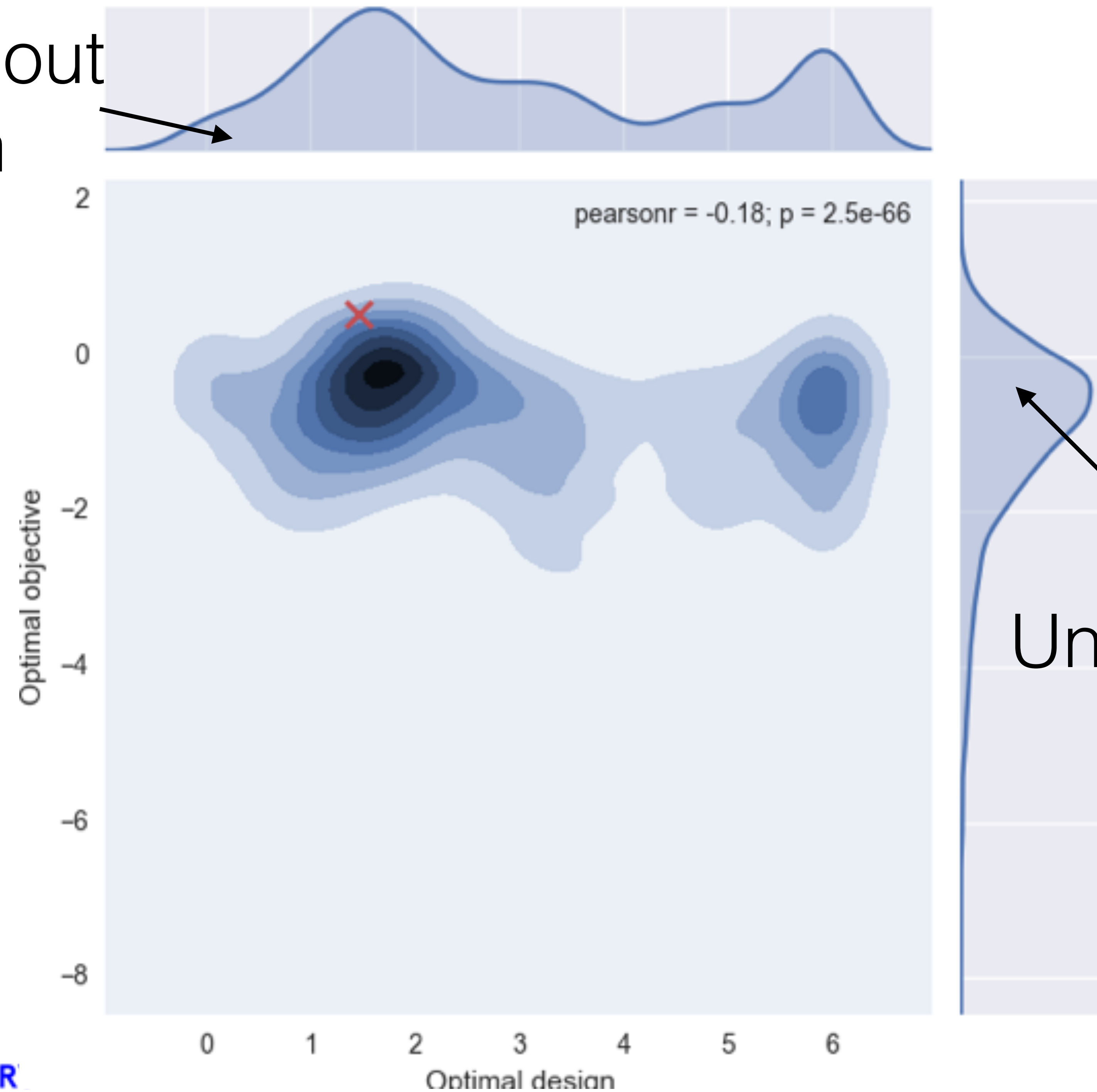
$$p(\mathbf{x}^* \mid \mathcal{D}) = \int \delta(\mathbf{x}^* - \mathrm{argmax}_{\mathbf{x}} \mathbf{E}[f(\mathbf{x})]) p(f(\cdot) \mid \mathcal{D}) df(\cdot)$$

$$p(f^* \mid \mathcal{D}) = \int \delta(\mathbf{x}^* - \max_{\mathbf{x}} \mathbf{E}[f(\mathbf{x})]) p(f(\cdot) \mid \mathcal{D}) df(\cdot)$$

PREDICTIVE
SCIENCE LABORATORY

Example 3: Optimization

Uncertainty about best design $p(\mathbf{x}^* | \mathcal{D})$



Uncertainty about best value of objective $p(f^* | \mathcal{D})$

PREDICTIVE SCIENCE LABORATORY

Example 3: Optimization

Pareto front (1,000 samples)

Multi-objective optimization:

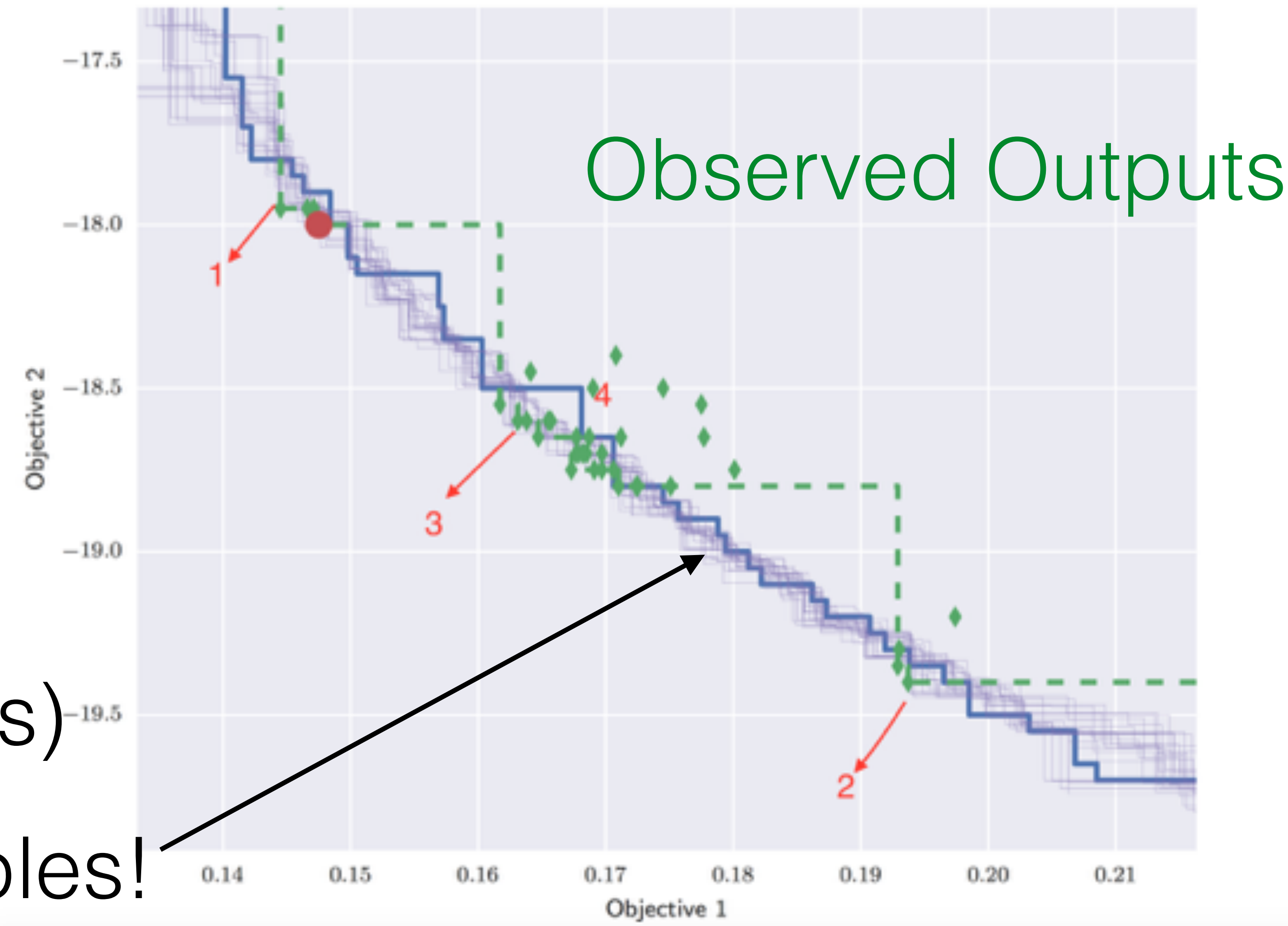
$$\min_{\mathbf{x}} \mathbf{E}[f_1(\mathbf{x})]$$

$$\min_{\mathbf{x}} \mathbf{E}[f_2(\mathbf{x})]$$

↓

Pareto Front
(set of non-dominated pts)

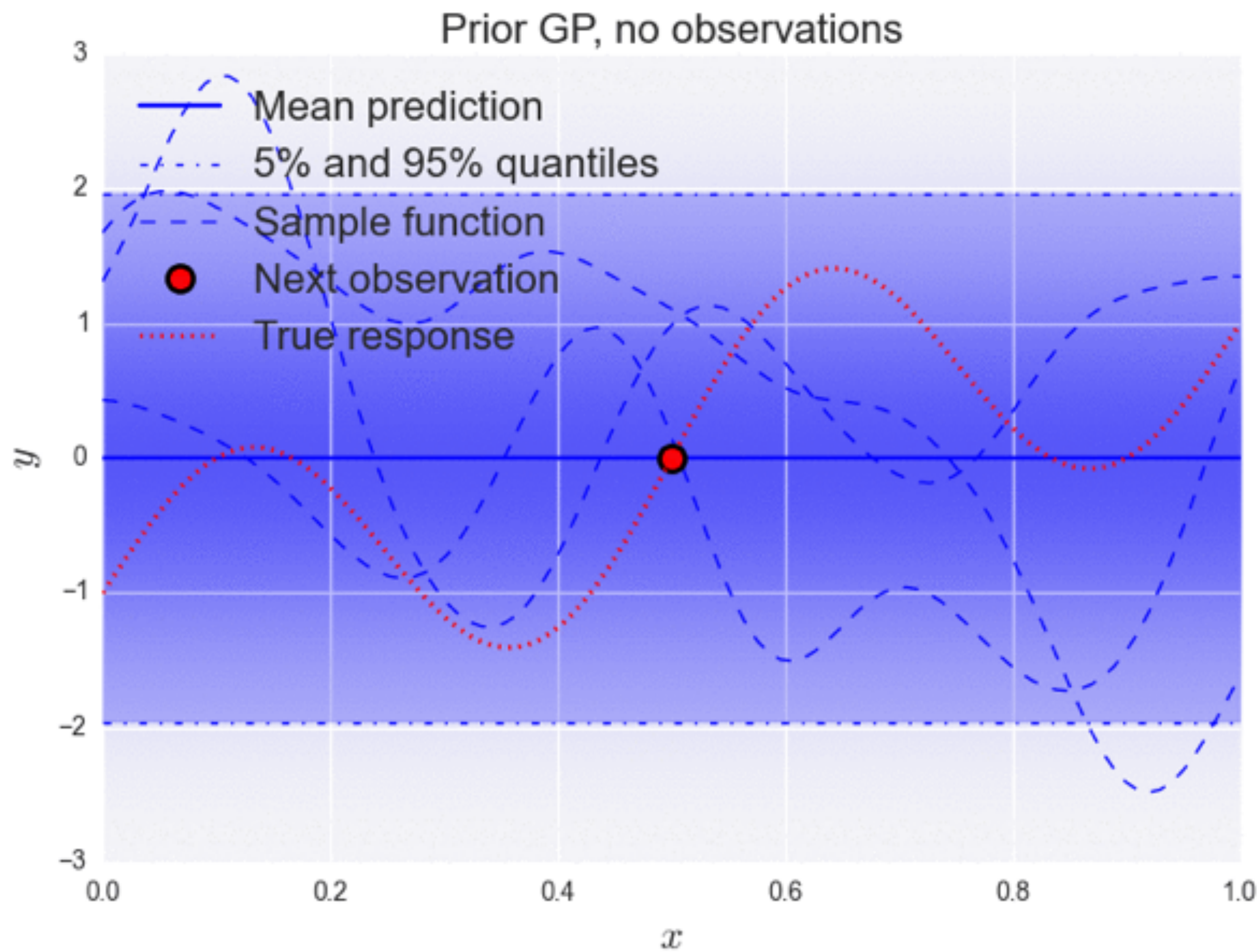
Pareto front samples!



Information Acquisition Decisions Under a Simulation Budget

PREDICTIVE
SCIENCE LABORATORY

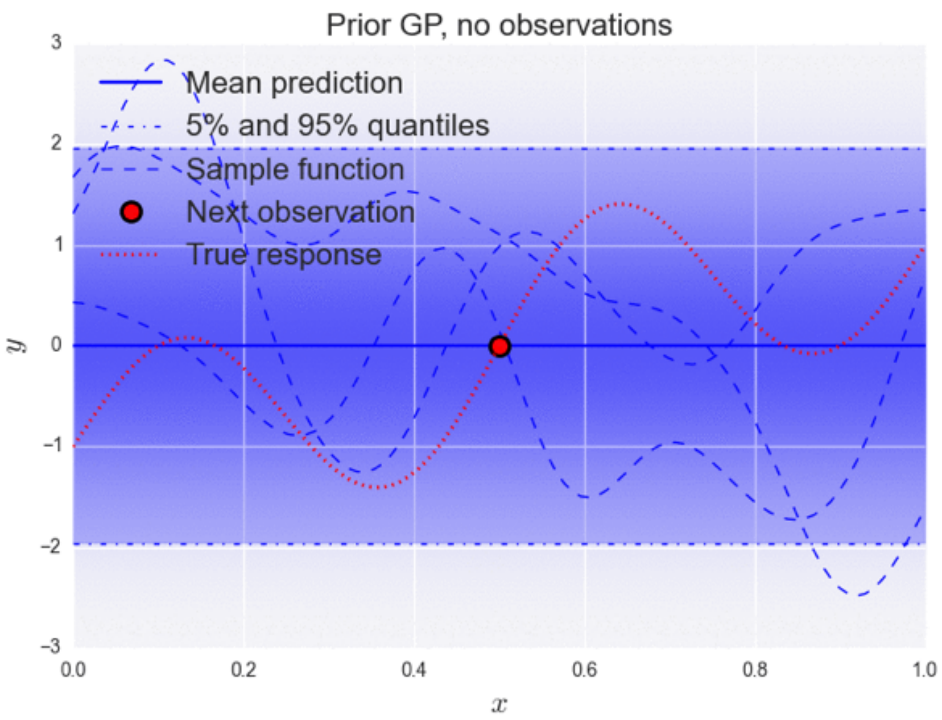
Adaptive selection: What is your goal?



Demo: selecting observations with maximum predictive uncertainty

PREDICTIVE
SCIENCE LABORATORY

Adaptive selection: What is your goal?



Equivalent to maximizing the **expected information gain** in the posterior of the hyper-parameters of the GP, if the posterior is well-approximated by a Gaussian (MacKay, 1991)

But that's not always what we want to do…

PREDICTIVE SCIENCE LABORATORY

Information Acquisition Decisions

Given

$$f(\cdot) \mid \mathcal{D}_n \sim p(f(\cdot) \mid \mathcal{D}_n)$$

What should our next observation be?

Depends on what we want to do…

PREDICTIVE
SCIENCE LABORATORY

Information Acquisition Decisions for Optimization of a Value Function

Assume that what we want to do is to solve

$$\mathbf{x}^* = \mathrm{argmax}_{\mathbf{x}} \mathbf{E}[f(\mathbf{x})]$$

where f(x) may be noisy. Bold **E** is expectation over this noise.

The *current expected optimal value* is:

$$\mathcal{V}_n = \mathbb{E}[\max_{\mathbf{x}} \mathbf{E}[f(\mathbf{x})] \mid \mathcal{D}_n],$$

The *pro-forma optimal value* is:

$$\mathcal{V}_{n+1}(\tilde{\mathbf{x}}) = \mathbb{E}[\max_{\mathbf{x}} \mathbf{E}[f(\mathbf{x})] \mid \tilde{\mathbf{x}}, \mathcal{D}_n],$$

Idea: Maximize the *marginal value of information*:

$$\mathbf{x}_{n+1} = \mathrm{argmax}_{\tilde{\mathbf{x}}} \{\mathcal{V}_{n+1}(\tilde{\mathbf{x}}) - \mathcal{V}_n\}$$

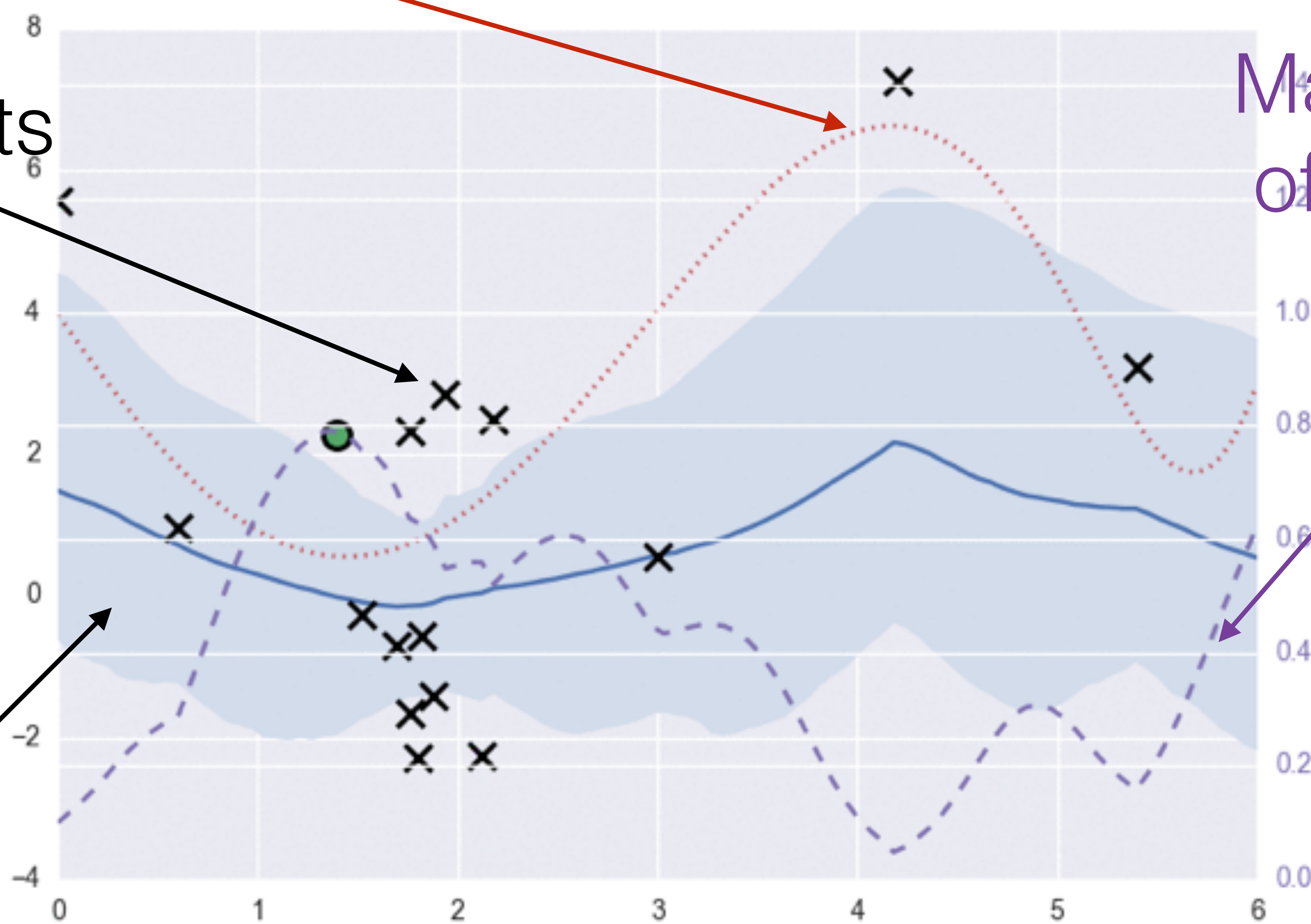
PREDICTIVE
SCIENCE LABORATORY

For special choices of approximating VoI…

- Expected Improvement (Jones, 1998)
- Knowledge Gradient (Frazier et al., 2008)

PREDICTIVE
SCIENCE LABORATORY

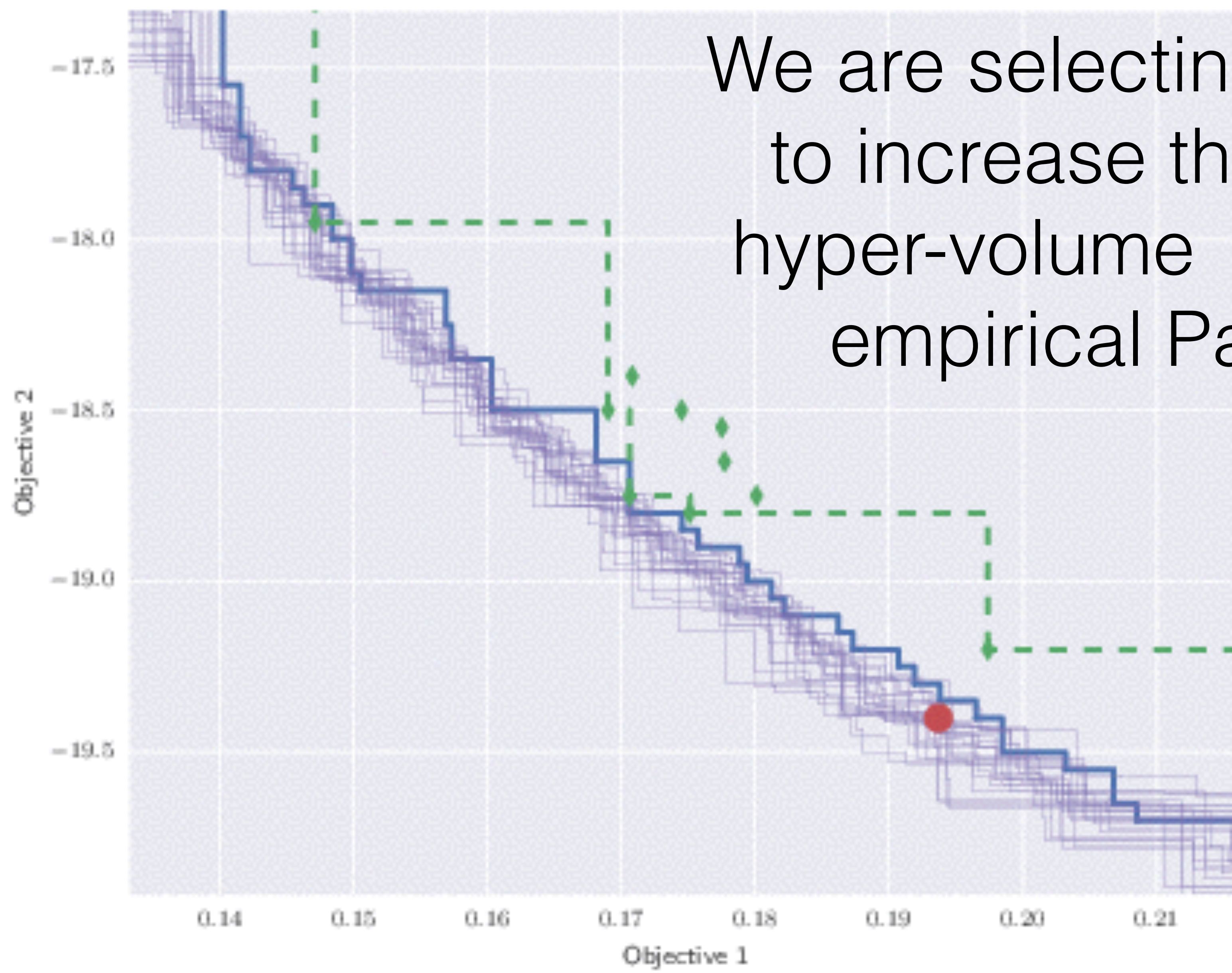
Example: 1D robust optimization



(Pandita and Bilionis, 2016)

PREDICTIVE SCIENCE LABORATORY

Example: Multi-Objective Optimization



We are selecting simulations to increase the expected hyper-volume enclosed by empirical Pareto front

(Pandita and Bilionis, 2016)

PREDICTIVE
SCIENCE LABORATORY

Example: Phase Diagrams of Binary Alloys

The problem is:

$$\boldsymbol{\sigma}^*(\omega) = \arg\min_{\boldsymbol{\sigma}} G(\boldsymbol{\sigma}, \omega)$$

Thermodynamic parameters

Thermodynamic potential

Alloy configuration

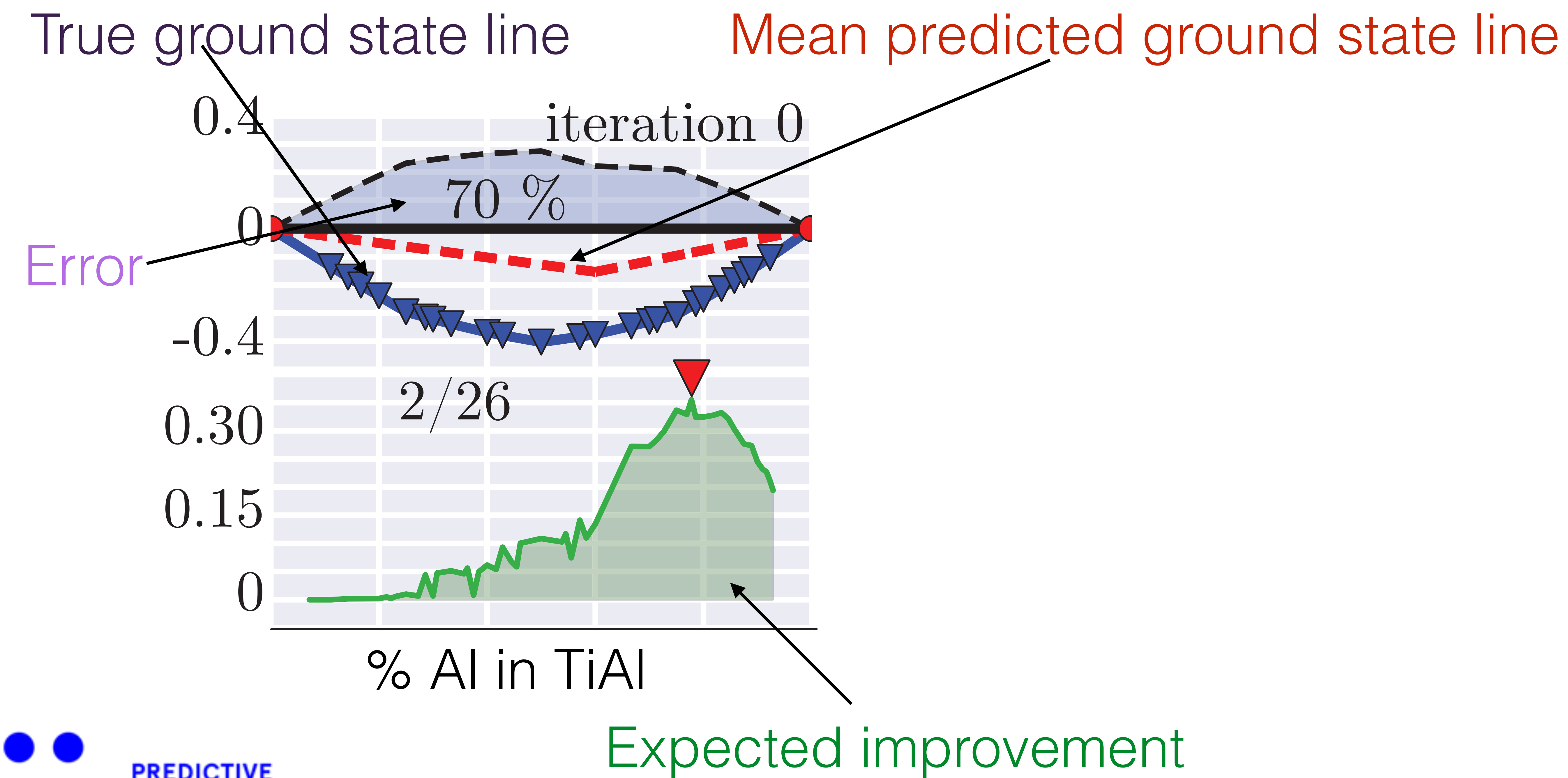
How?

- Put prior on G.
- Gather data.
- Compute posterior over G.
- Propose the simulation that maximizes expected improvement in G.

PREDICTIVE SCIENCE LABORATORY

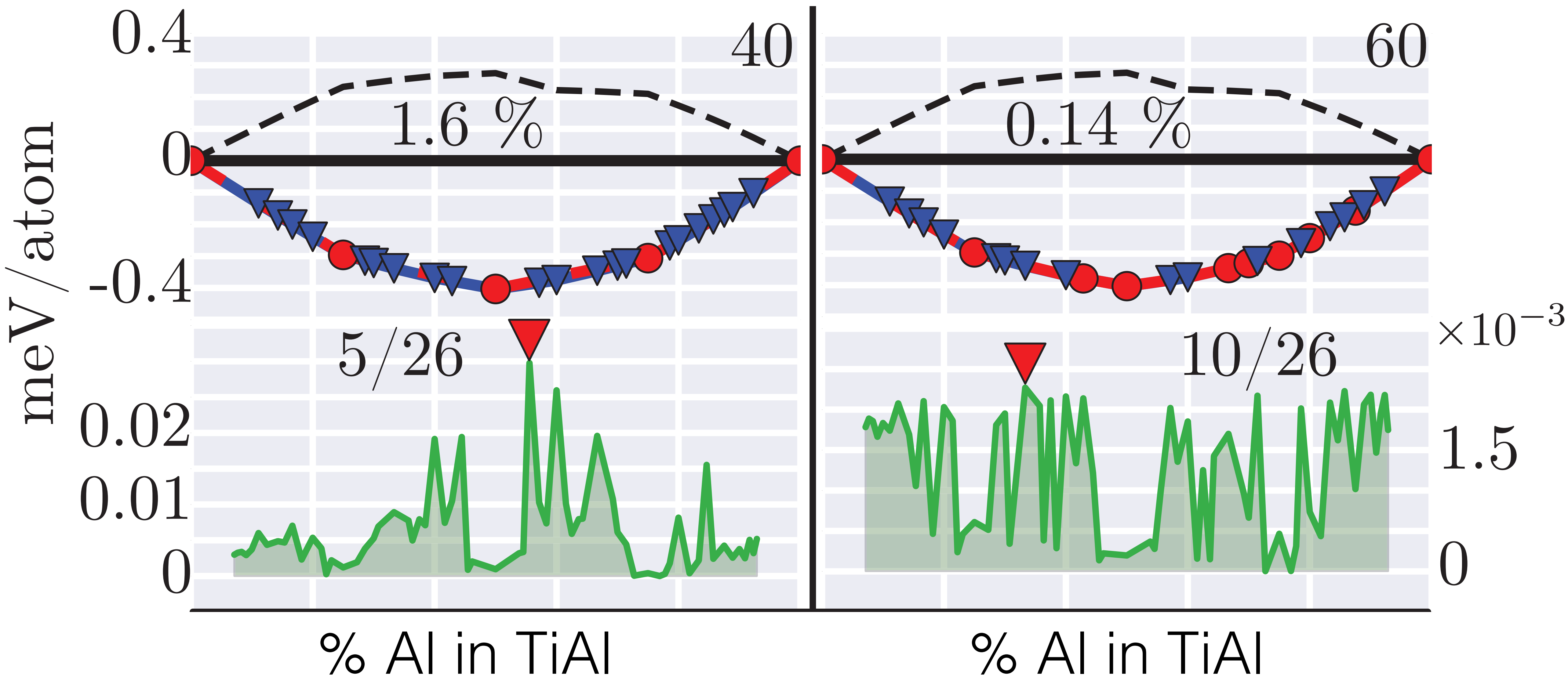
Kristensen, Bilionis, and Zabaras, 2016 (submitted to JCP)

Example: Phase Diagrams of Binary Alloys



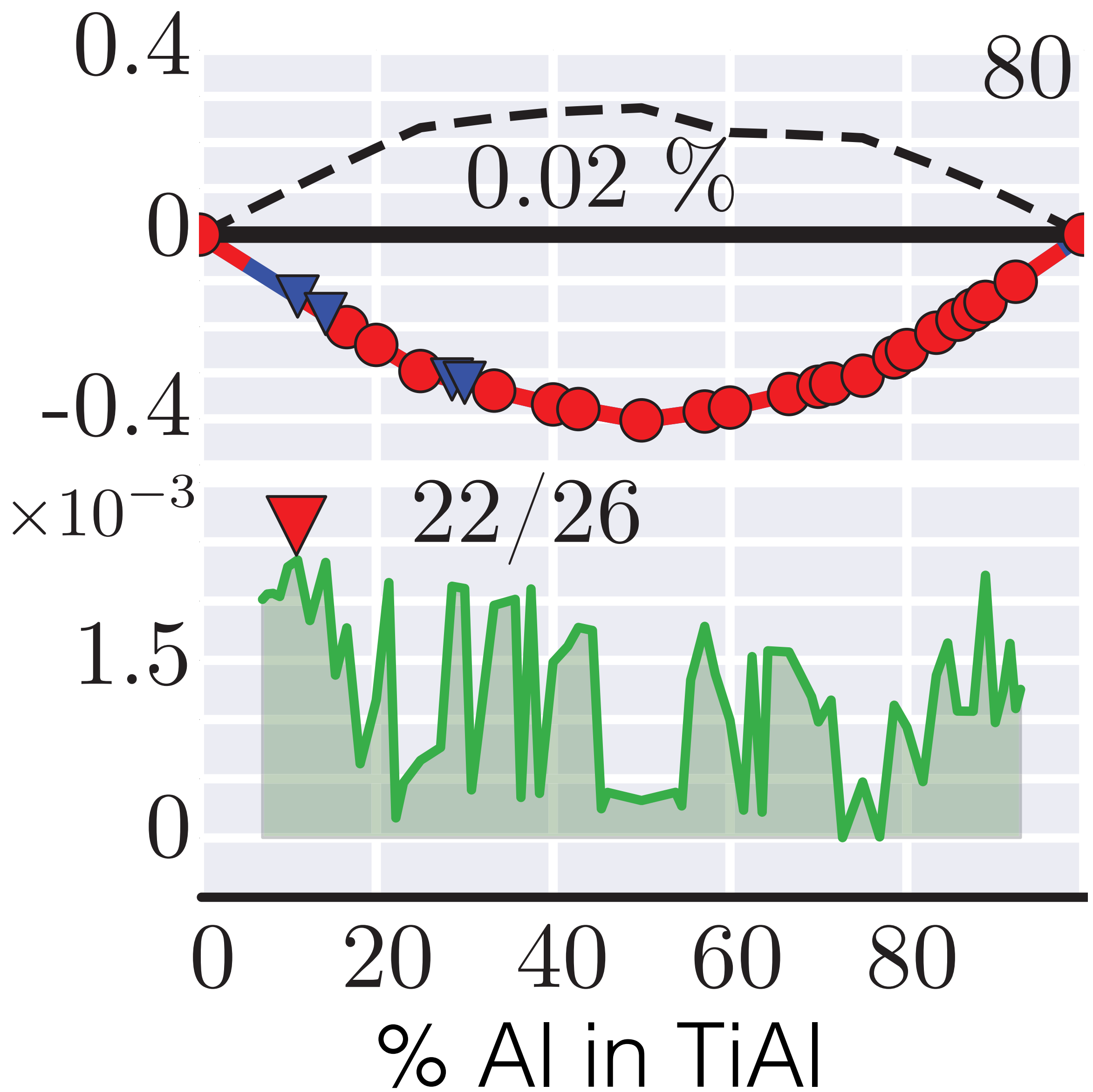
PREDICTIVE
SCIENCE LABORATORY

Example: Phase Diagrams of Binary Alloys



PREDICTIVE SCIENCE LABORATORY

Example: Phase Diagrams of Binary Alloys

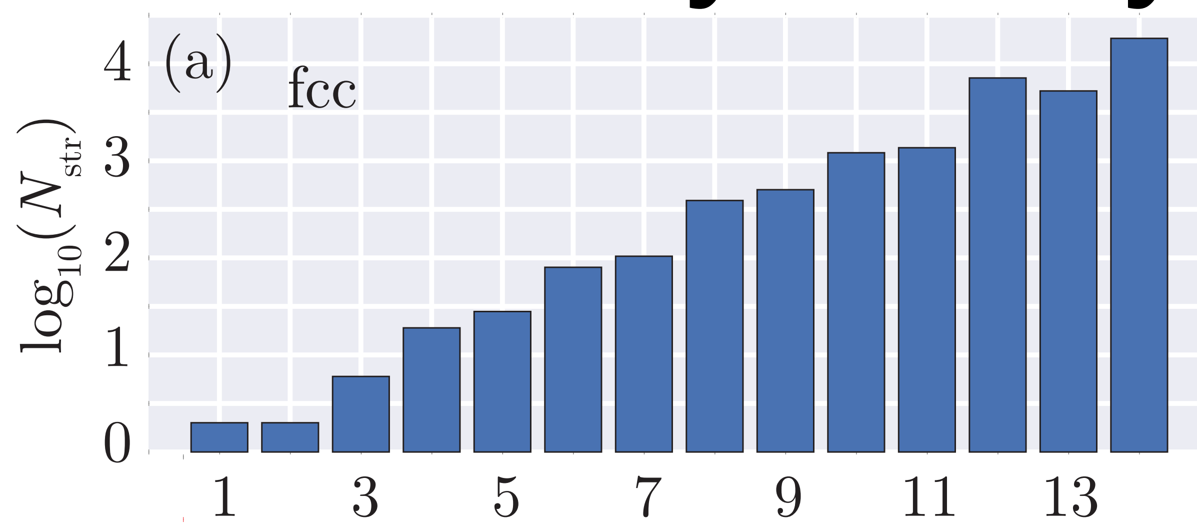


3

PREDICTIVE
SCIENCE LABORATORY

Example: Phase Diagrams of Binary Alloys

Number of structures



Unit cell size

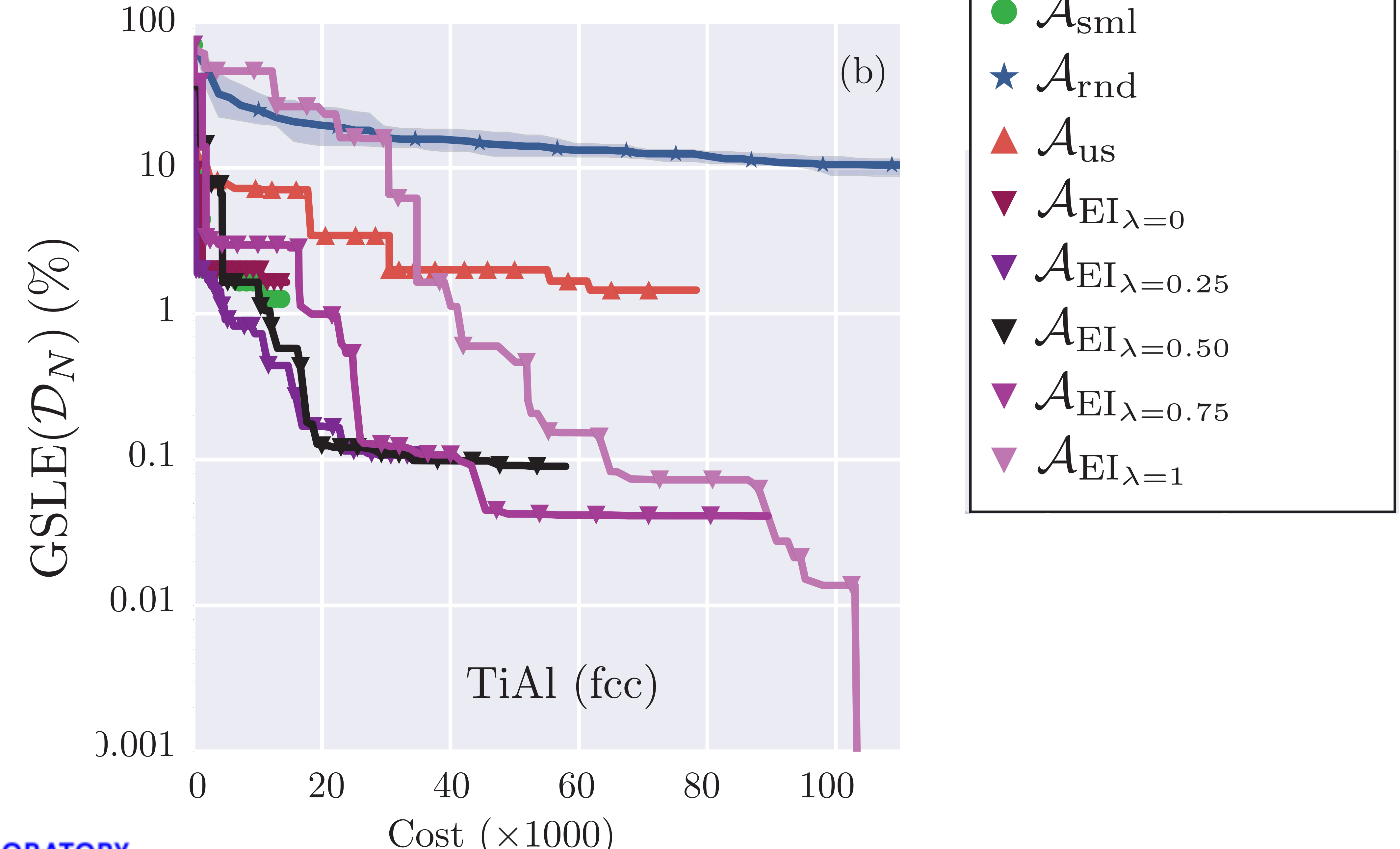
Big cell simulations have more structures and cost more to simulate…

$$\text{EI}_\lambda(\tilde{\boldsymbol{\sigma}}, \omega) \equiv \lambda \text{EI}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega}) - (1-\lambda)\text{C}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega})$$

PREDICTIVE SCIENCE LABORATORY

Example: Phase Diagrams of Binary Alloys

$$\mathrm{EI}_\lambda(\tilde{\boldsymbol{\sigma}},\omega) \equiv \lambda \mathrm{EI}(\tilde{\boldsymbol{\sigma}},\tilde{\omega}) - (1-\lambda)\mathrm{C}(\tilde{\boldsymbol{\sigma}},\tilde{\omega})$$



PREDICTIVE SCIENCE LABORATORY

Open Questions

- Can you find information acquisition functions that are good for:
  - uncertainty propagation
  - model calibration
- Yes, if you can pose these problems as optimization problems? Possible, but another talk.

PREDICTIVE
SCIENCE LABORATORY

Thank you!

PREDICTIVE
SCIENCE LABORATORY