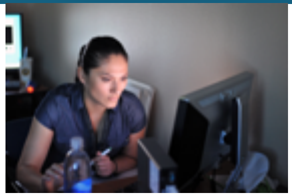
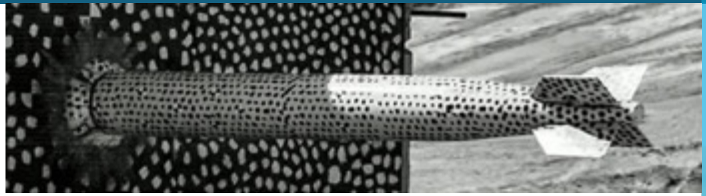
Embedded Model Error Quantification and Propagation in Physical Systems





PRESENTED BY

Khachik Sargsyan, 8351

CIS External Review, August 26-29, 2019

SAND 2019-9724 C

Algorithmic talk demonstrating broad application impact

- Overview
- **Model error**
  - Definition
  - Current state
  - What we do
- Applications
  - Climate land model (BER+ASCR)
  - Chemistry (BES)
  - Fusion science (FES+ASCR)
  - Thermodynamics (EERE)
  - ....
  - Turbulence modeling
- Summary/future

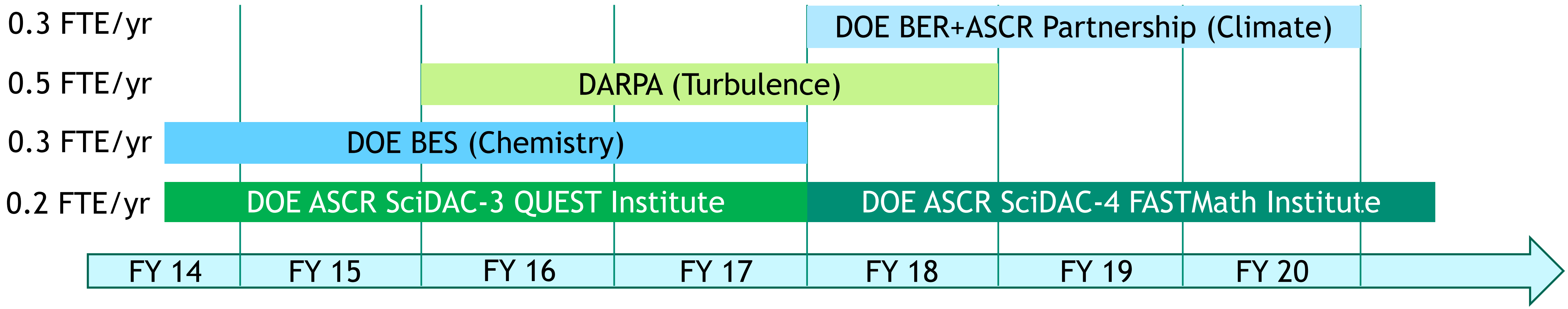
U.S. DEPARTMENT OF ENERGY
Office of Science

SciDAC
Scientific Discovery through Advanced Computing

DARPA

This work has been funded by DOE/DOD, including fundamental research and applications

model-error work incl. collaborators

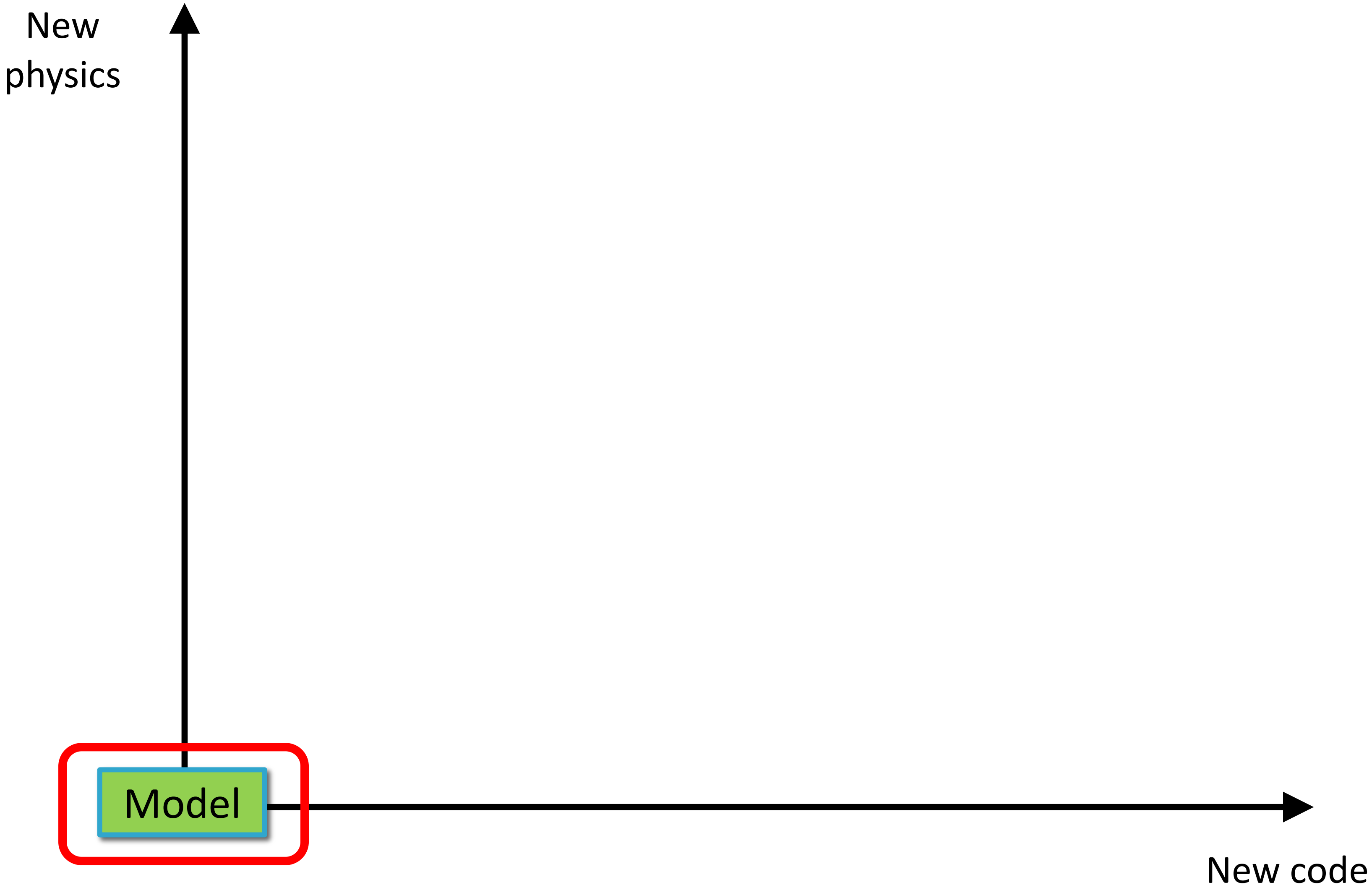


Habib Najm, Cosmin Safta, Tiernan Casey, James Oreluk, Bert Debusschere (SNL)

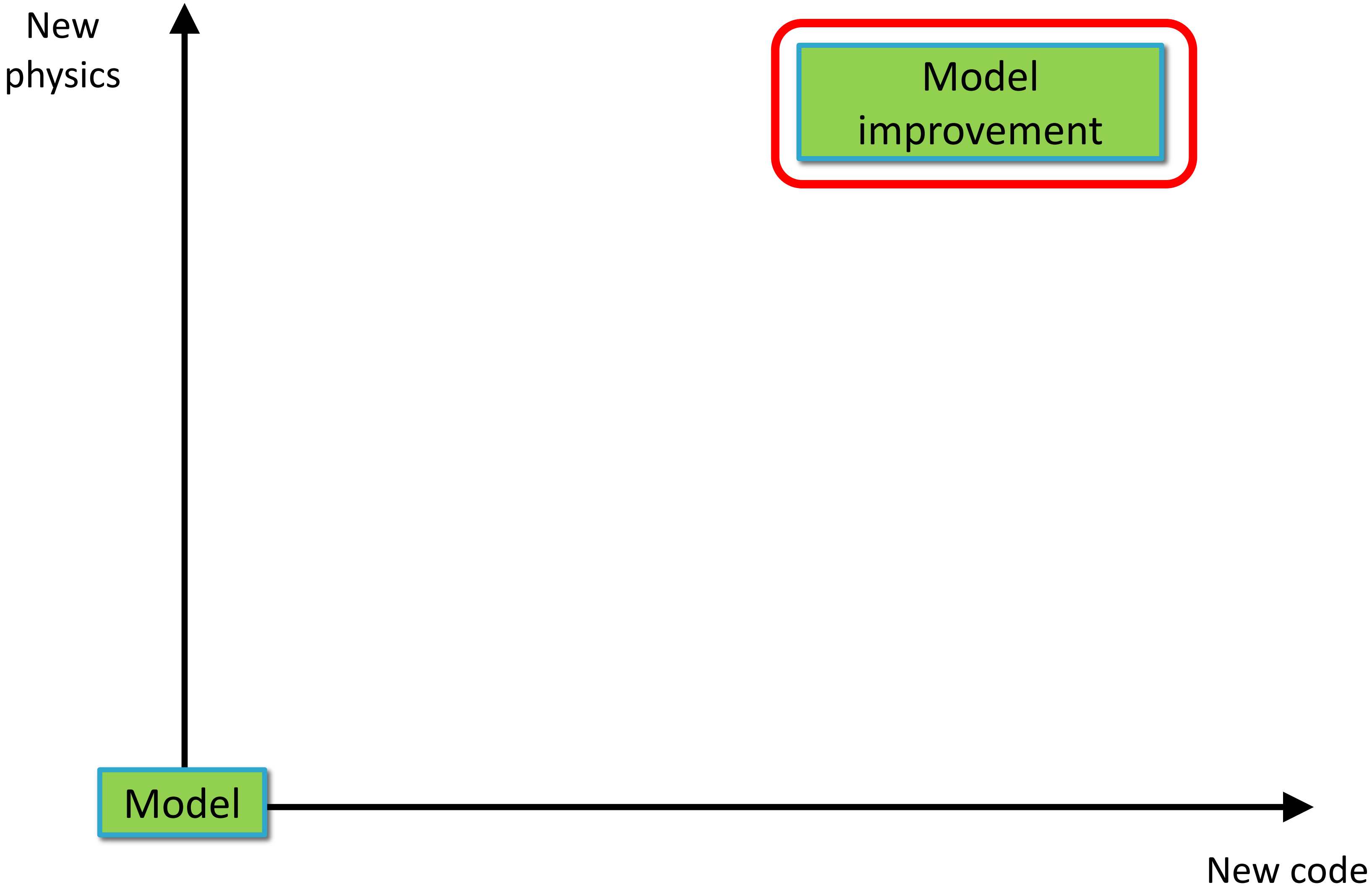
Daniel Ricciuto (ORNL), Jason Bender (LLNL)

Youssef Marzouk, Chi Feng (MIT), Roger Ghanem (USC), Xun Huan (UMichigan)

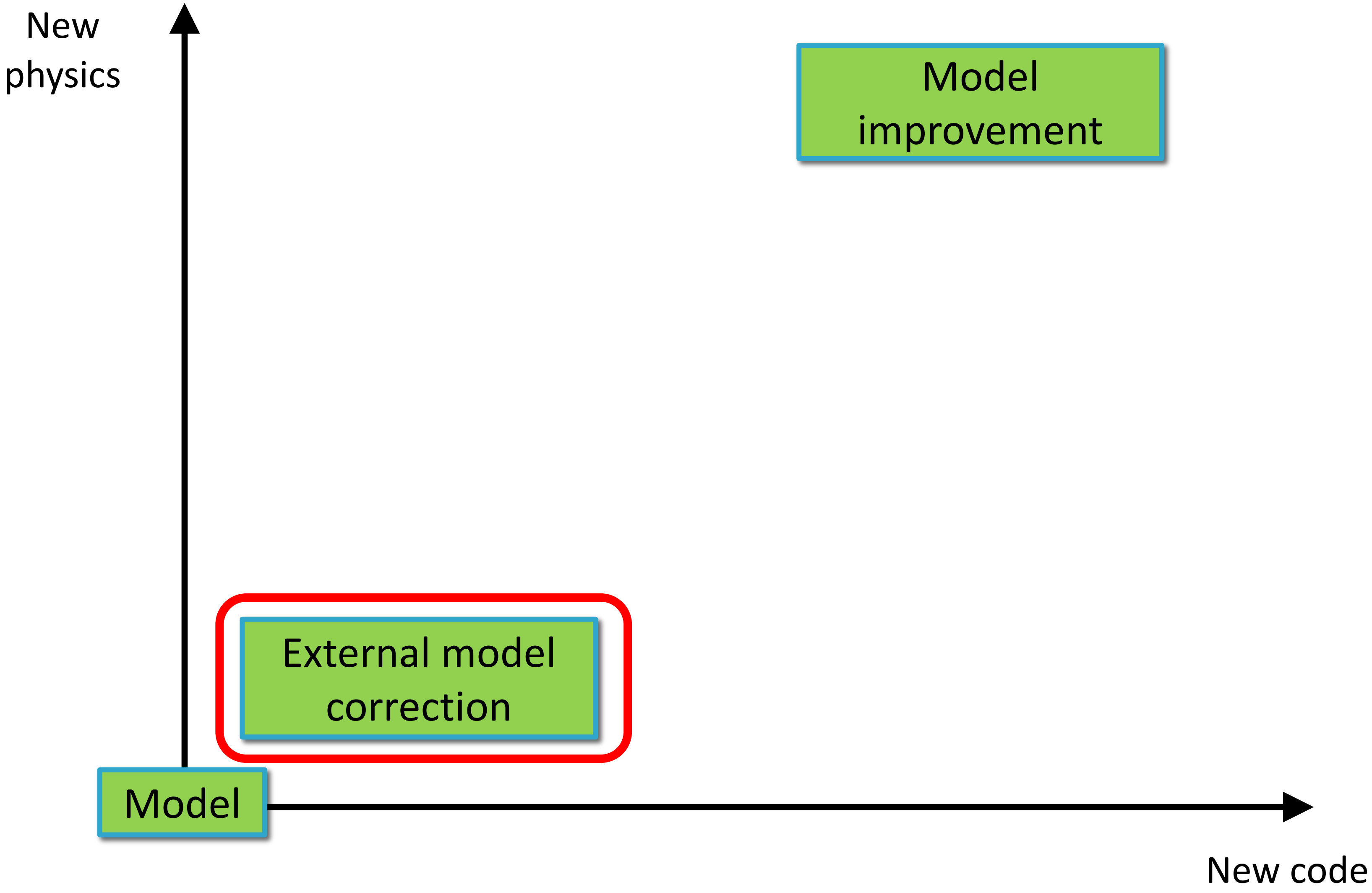
To improve computational models of physical phenomena, one targets improvements in the physics and the code



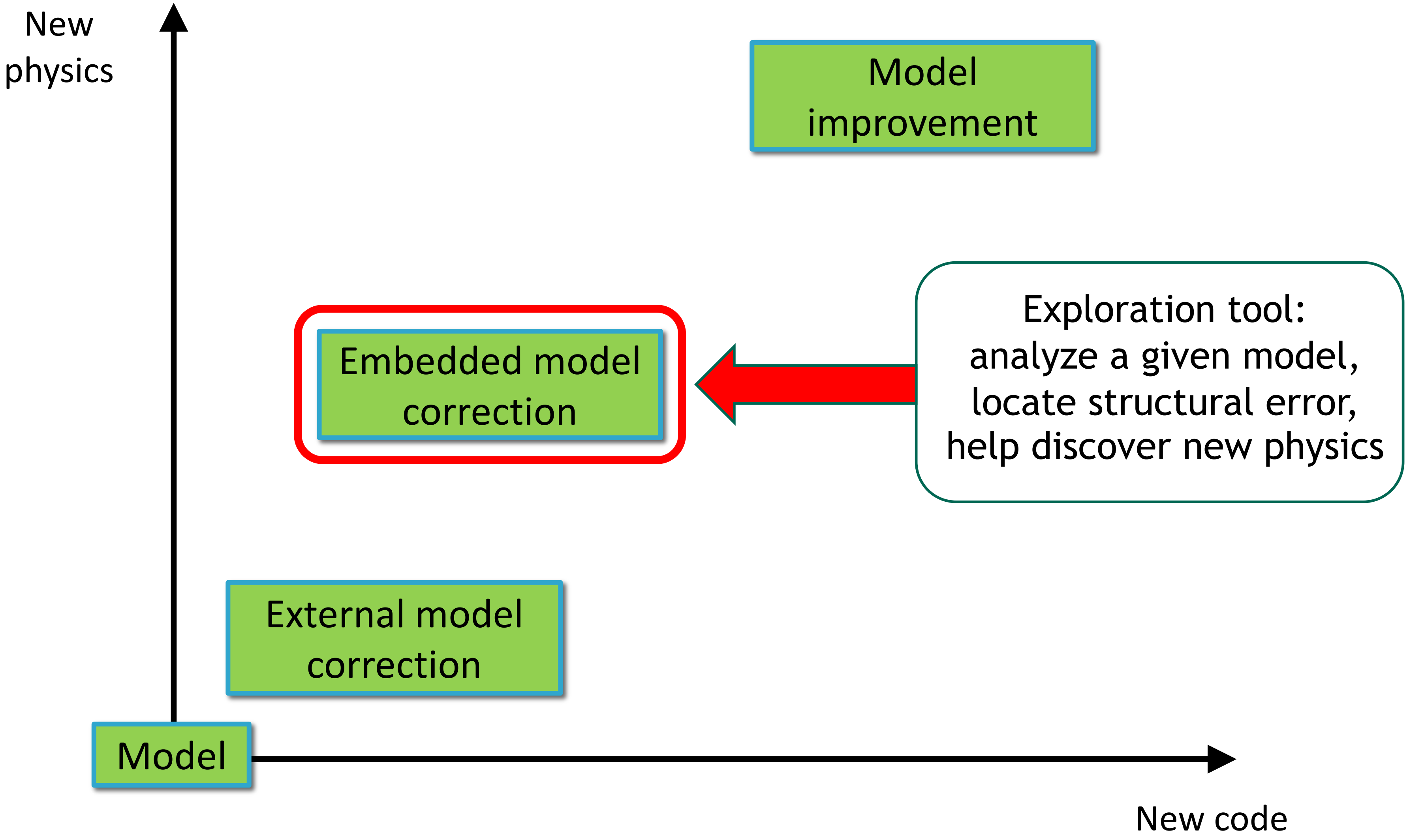
Ideal (but costly) situation: domain-specific ‘intrusion’ into the code



There is mature work in *statistical* model corrections, without physical interpretation



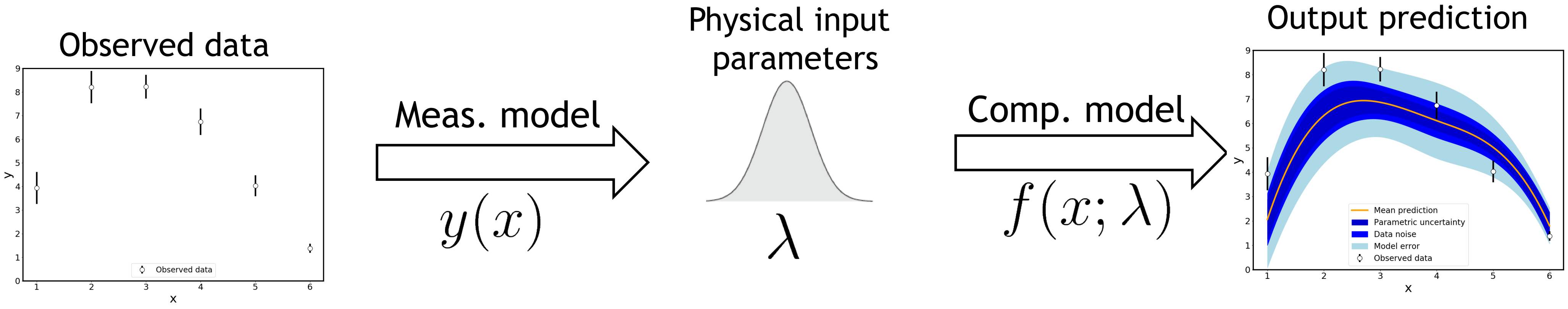
We develop approach for embedded statistical correction, that is driven by the physics but does not require code change (augments the existing code with stochasticity)



In order to improve or correct the model, we need to quantify model error

Prediction uncertainty budget

parametric uncertainty + data noise + model error

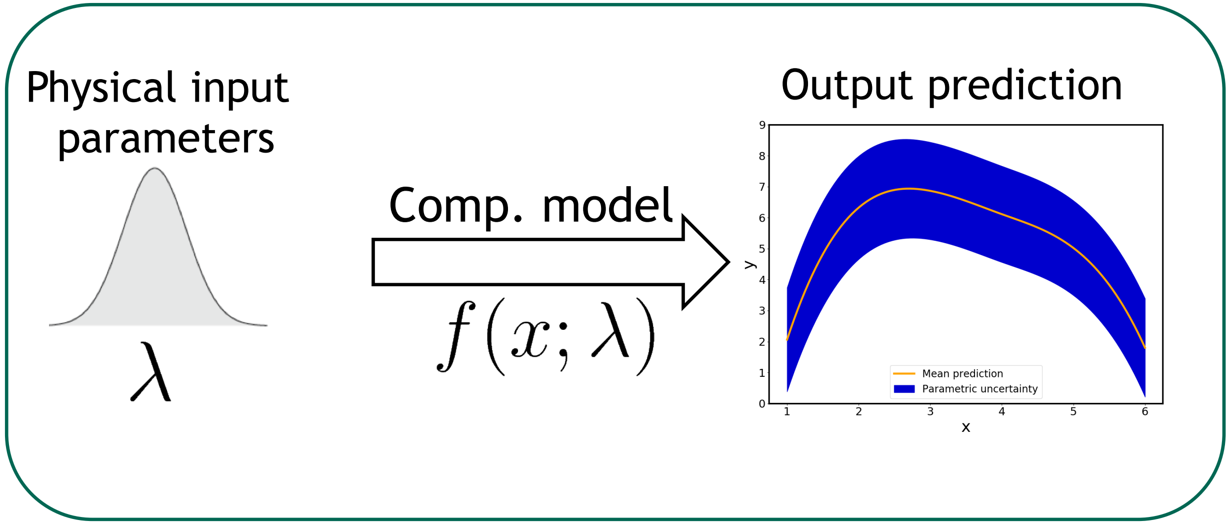


In UQ, we know how to incorporate parametric uncertainty

Prediction uncertainty budget

parametric uncertainty

Cosmin Safta's talk from 2018 CIS ERB: algorithmic advances to address high-dimensionality of λ and non-linearity of $f(x;\lambda)$

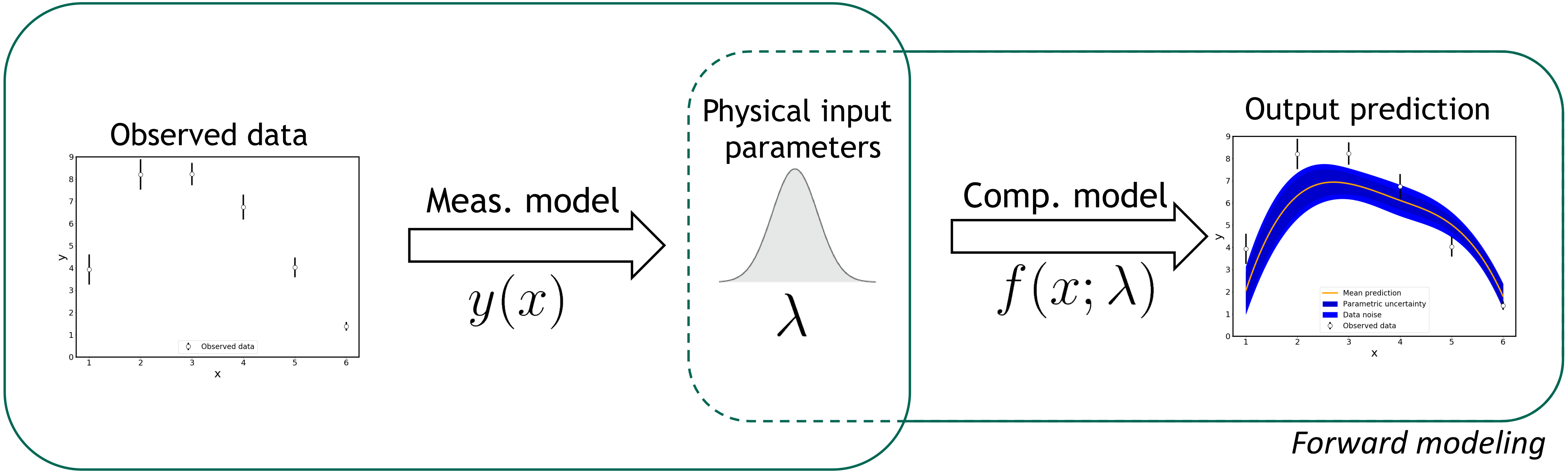


Forward modeling

In UQ, we know how to incorporate parametric uncertainty and data noise

Prediction uncertainty budget

parametric uncertainty + data noise

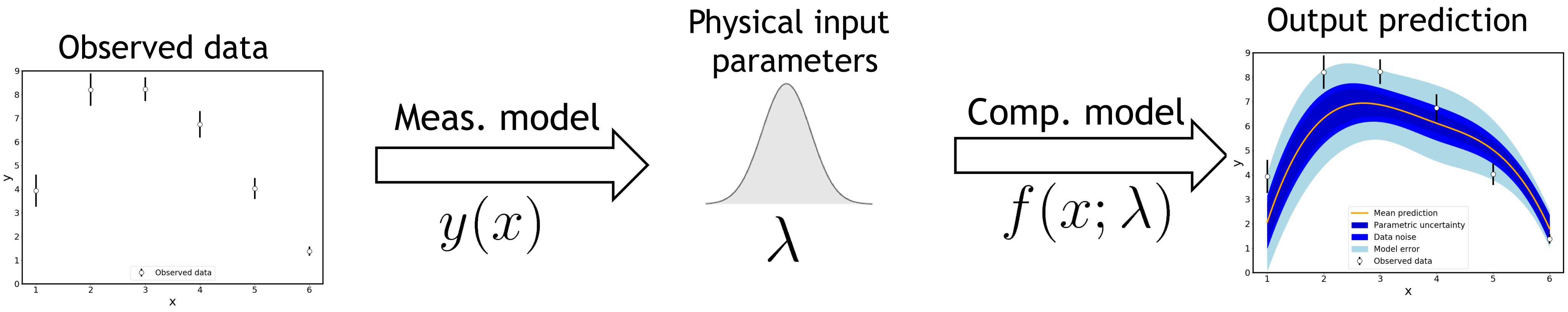


Inverse modeling

In order to improve or correct the model, we need to quantify model error

Prediction uncertainty budget

parametric uncertainty + data noise + model error



Model error is often the dominant component of predictive uncertainty

Ignoring model error leads to

- Biased parameter estimation
- Overconfident predictions

Data | Model | Data noise

$$y(x_i) = f(x_i; \lambda) + \epsilon_i$$

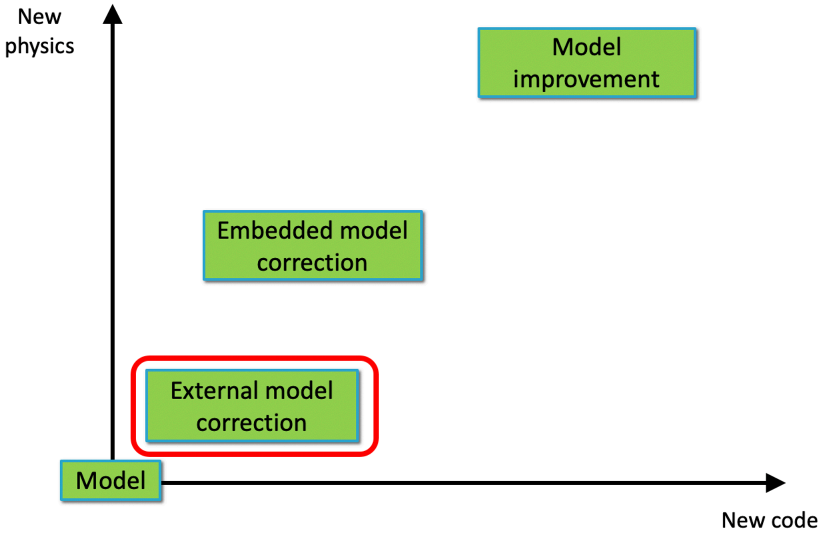
Representing and estimating model error is useful for

- Reliable computational predictions
- Model comparison, selection
- Scientific discovery and model improvement:
  - *"is it worth resolving details, or just parameterize empirically?"*
- Optimal resource allocation:
  - *"do I improve my model (e.g. high-res), or run more simulations?"*

Conventional approach from statistical literature: external correction to the model

Data — Model — Model error — Data noise

$$y(x_i) = \underbrace{f(x_i;\lambda) + \delta(x_i)}_{\text{Truth} \quad g(x_i)} + \epsilon_i$$



- Explicit Gaussian Process representation of model error [Kennedy-O'Hagan, 2001]
- See also [Higdon et. al, 2004], [Bayarri et. al, 2007]
- Usage: too many to cite
- Variants exist: multiplicative noise, non-linear maps etc.
- Implemented in Dakota – highly mature procedure

Conventional approach from statistical literature: external correction to the model

Data Model Model error Data noise

$$\underbrace{y(x_i)}_{\text{Data}} = \underbrace{f(x_i;\lambda) + \delta(x_i)}_{\text{Truth } g(x_i)} + \epsilon_i$$

DAKOTA

Explore and predict with confidence.

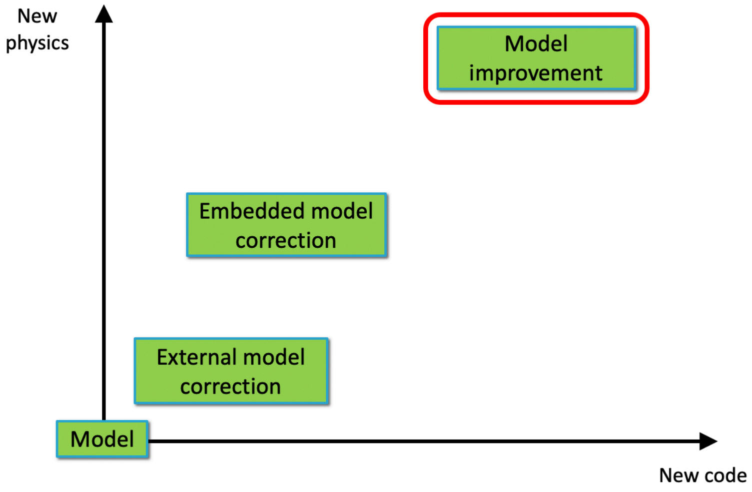
See Kathryn Maupin's poster

- Explicit Gaussian Process representation of model error [Kennedy-O'Hagan, 2001]
- See also [Higdon et. al, 2004], [Bayarri et. al, 2007]
- Usage: too many to cite
- Variants exist: multiplicative noise, non-linear maps etc.
- Implemented in Dakota – highly mature procedure

Embedded model error corrections have gained recent interest

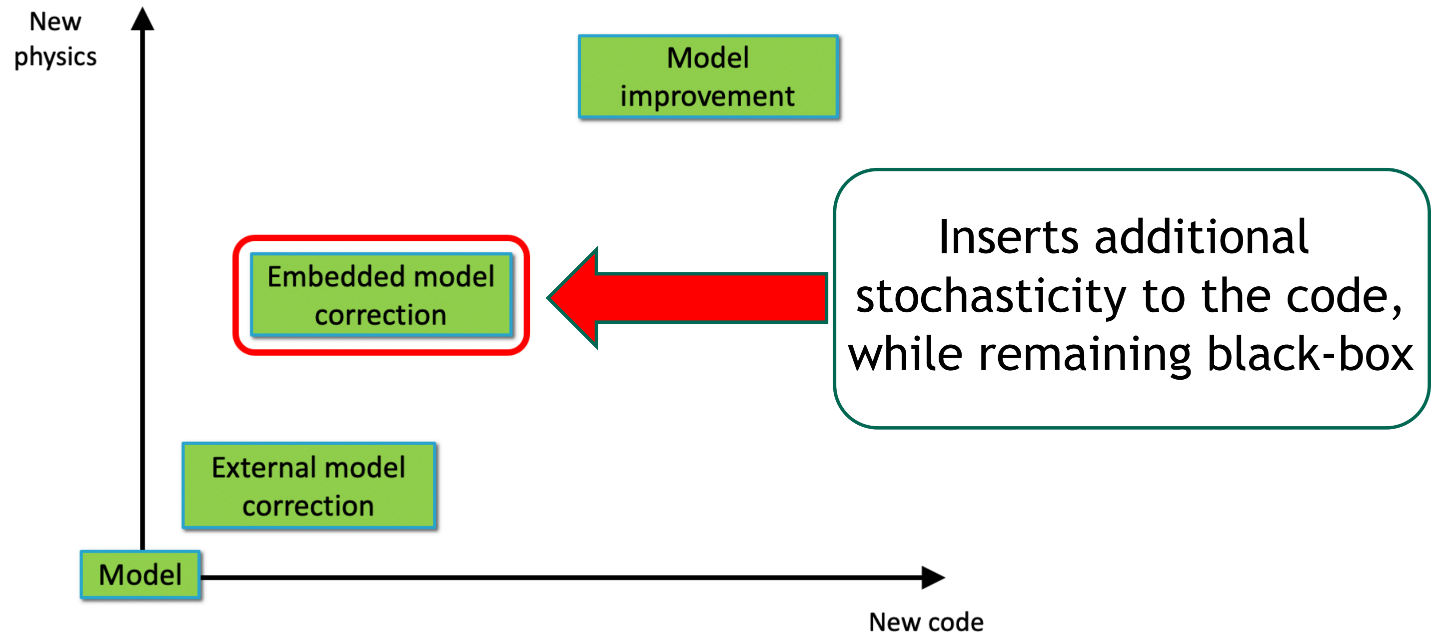
Domain scientists improve the models by problem-specific internal adjustments

- Validating extrapolative predictions [Oliver and Moser, 2011], [Oliver et. al, 2014]
- Engineering/statistical adjustment [Joseph and Melkote, 2009]
- Stochastic operator [Morrison et. al, 2016], [Sondak et. al, 2017]
- Additive corrections to submodels [Strong et. al, 2011]
- Field inversion and machine learning [Duraisamy et. al, 2015-]
- Turbulence anisotropy perturbation [Emory et. al, 2011]
- Random field correction [Brown and Atamturktur, 2016]
- Hierarchical mixture model [Feng, 2017]
- Parameter inflation [Pernot et. al, 2017]
- Hierarchical stochastic model [Wu et. al, 2017]
- Dynamic discrepancy [Bhat et. al., 2017]
- Hybrid correction [He and Xiu, 2016]



$$y(x_i) = \tilde{f}(x_i; \lambda, \delta(x_i)) + \epsilon_i$$

We have developed a non-intrusive way to embed model error within the model



Embed model error where model error (i.e. key modeling assumptions) happens:

- Specific submodel phenomenology
- Modified transport law, material property, turbulent constant

- Naturally preserves physics
- Disambiguates from data noise
- Allows prediction of non-observable QoIs
- Does not vanish with more data

We have developed a non-intrusive way to embed model error within the model

External model correction

$$y(x_i) = f(x_i;\lambda) + \delta(x_i) + \epsilon_i$$

Embedded model correction

Intrusive

change the code

$$y(x_i) = \tilde{f}(x_i;\lambda,\delta(x_i)) + \epsilon_i$$

Non-intrusive
(black-box)

$$y(x_i) = f(x_i;\lambda+\delta(x_i)) + \epsilon_i$$

change inputs to the code

Embed model error where model error (i.e. key modeling assumptions) happens:

- Specific submodel phenomenology
- Modified transport law, material property, turbulent constant

- Naturally preserves physics
- Disambiguates from data noise
- Allows prediction of non-observable QoIs
- Does not vanish with more data

Polynomial chaos (PC) expansion is employed as a model error representation

$$f\left(x;\lambda+\underbrace{\sum_k \alpha_k\Psi_k(\xi)}_{\text{PC}}\right) \overset{\text{NISP}}{\simeq} \sum_k f_k(x;\underbrace{\lambda}_{\text{physical}},\underbrace{\alpha}_{\text{model-error}})\Psi_k(\xi)$$

$$\rightarrow \mu(x;\lambda,\alpha)$$

$$\rightarrow \sigma^2(x;\lambda,\alpha)$$

Key technical ingredients:

- Non-intrusive spectral projection (NISP) for uncertainty propagation
- Simultaneous Bayesian inference of λ and α via Markov chain Monte Carlo
- Variance decomposition of prediction uncertainty

Prediction mean $\quad \mu(x)=\mathbb{E}_{\lambda,\alpha}[\mu(x;\lambda,\alpha)]$

Prediction variance $\quad \sigma^2(x)=\underbrace{\mathbb{V}_{\lambda,\alpha}[\mu(x;\lambda,\alpha)]}_{\text{Posterior}}+\underbrace{\mathbb{E}_{\lambda,\alpha}[\sigma^2(x;\lambda,\alpha)]}_{\text{Model error}}$

Polynomial chaos (PC) expansions are well-suited for uncertainty propagation and attribution

$$f\left(x;\lambda+\underbrace{\sum_k \alpha_k\Psi_k(\xi)}_{\text{PC}}\right) \overset{\text{NISP}}{\simeq} \sum_k f_k(\underbrace{x;\lambda}_{\text{physical}},\underbrace{\alpha}_{\text{model-error}})\Psi_k(\xi) \rightarrow \mu(x;\lambda,\alpha),\ \sigma^2(x;\lambda,\alpha)$$

Key technical ingredients:

- Non-intrusive spectral projection (NISP) for uncertainty propagation
- Simultaneous Bayesian inference of λ and α via Markov chain Monte Carlo
- Variance decomposition of prediction uncertainty

Prediction mean
$$\mu(x)=\mathbb{E}_{\lambda,\alpha}[\mu(x;\lambda,\alpha)]$$

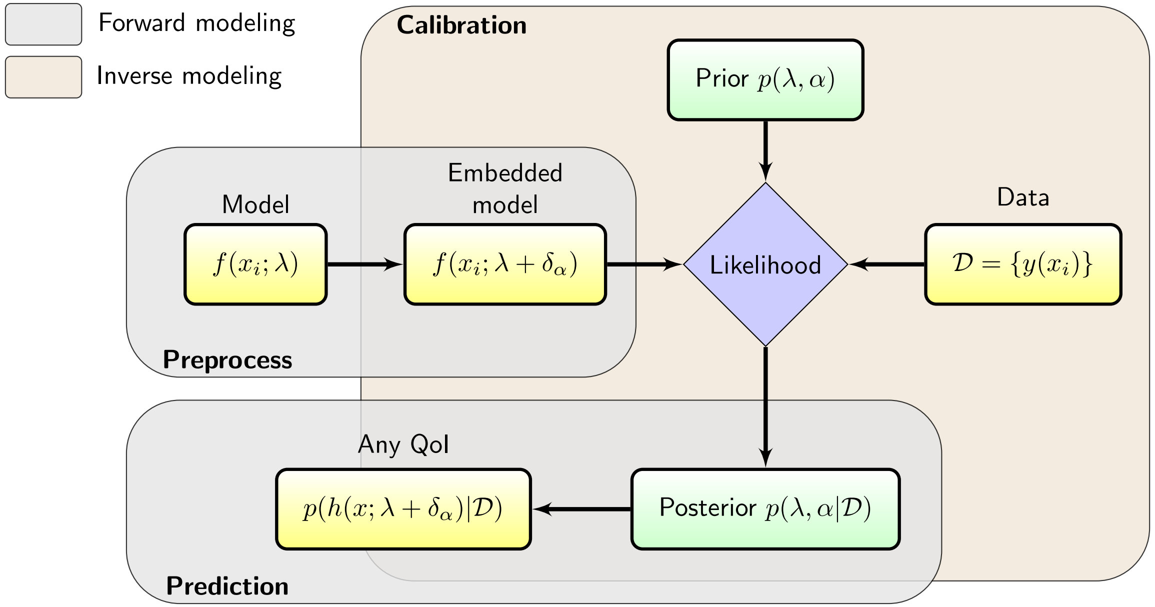
Prediction variance
$$\sigma^2(x)=\underbrace{\mathbb{V}_{\lambda,\alpha}[\mu(x;\lambda,\alpha)]}_{\text{Posterior}}+\underbrace{\mathbb{E}_{\lambda,\alpha}[\sigma^2(x;\lambda,\alpha)]}_{\text{Model error}}$$

Does not shrink with more data

Embedded, non-intrusive workflow is implemented in UQTk

FASTMATH UQTk

Funded by FASTMath SciDAC Institute, www.sandia.gov/uqtoolkit



Applications

- Earth System Land Model (BER)
- Chemical kinetic modeling (BES)
- Plasma surface interactions (FES)
- Turbulence modeling (DARPA)

Method: Sargsyan, Najm, Ghanem, IJCK (2015); Sargsyan, Huan, Najm, IJUQ (2019).

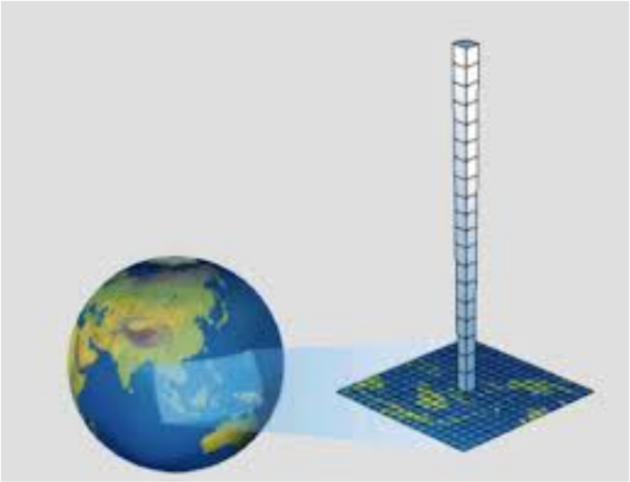
Applications: Huan et. al, AIAA J (2018); Hakim et. al, CTM (2018); Cekmer et. al, IJUQ (2018); Rizzi et. al, CMAME (2019).

Major application of focus: Earth System Land Model

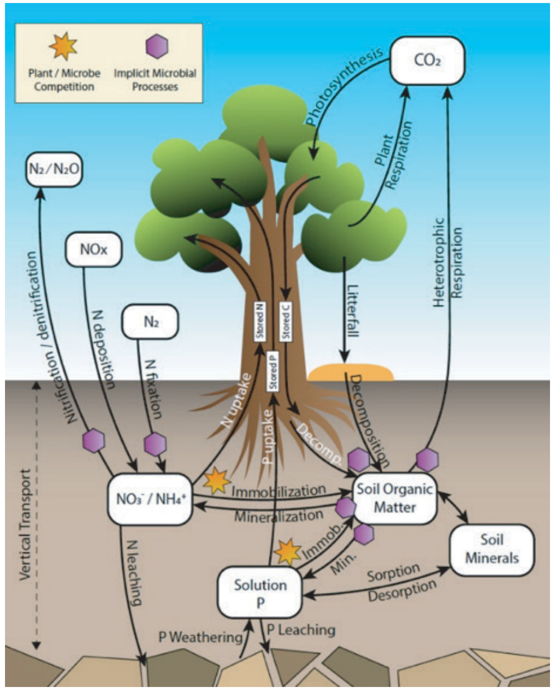
funded by **BER+ASCR**

- US DOE sponsored Earth system model
- Land, atmosphere, ocean, ice, human system components
- High-resolution, employ DOE leadership-class computing facilities

E3SM
Energy Exascale Earth System Model



NERSC National Energy Research Scientific Computing Center

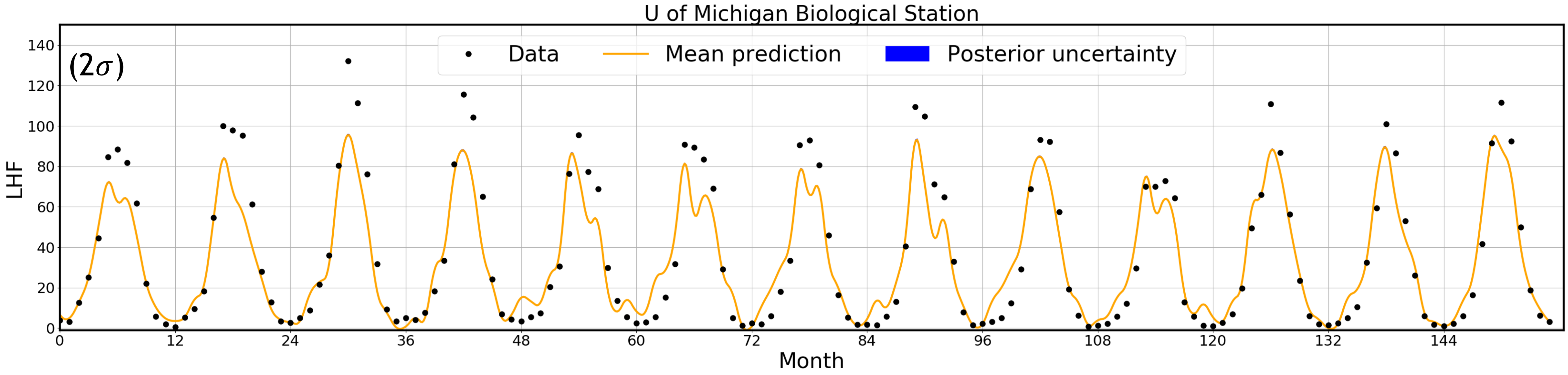


Serving as a UQ Lead for Land Model:
Direct UQ impact on several land components,
collaboration with multiple NLs

Oak Ridge National Laboratory · Berkeley Lab (Lawrence Berkeley National Laboratory) · Argonne National Laboratory · Pacific Northwest National Laboratory

Calibration without embedded model error

Land model calibration given FLUXNET observations

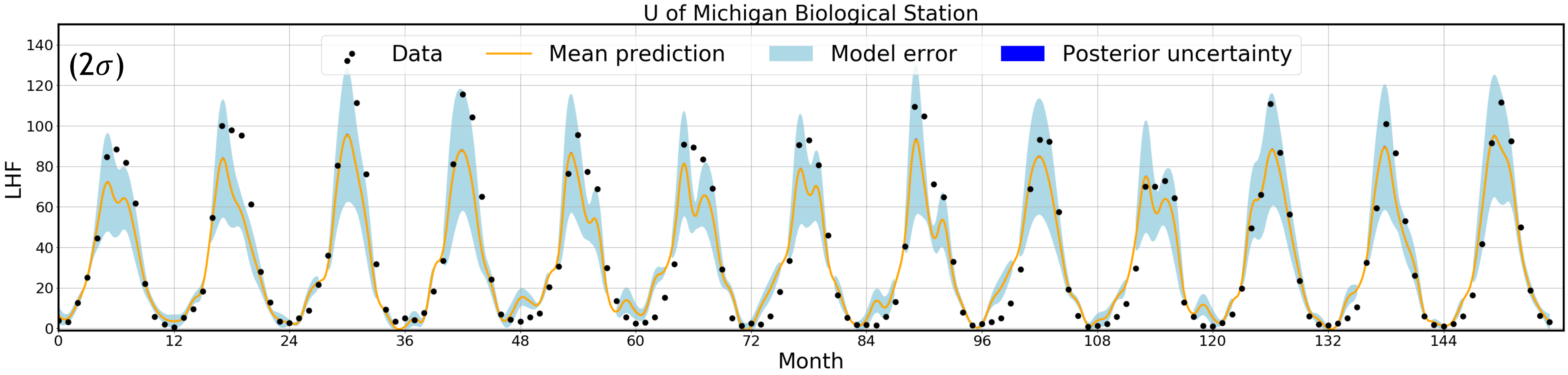


- Summer month peaks are not captured
- Posterior uncertainty negligible

LHF = Latent Heat Flux

Calibration with embedded model error

Land model calibration given FLUXNET observations

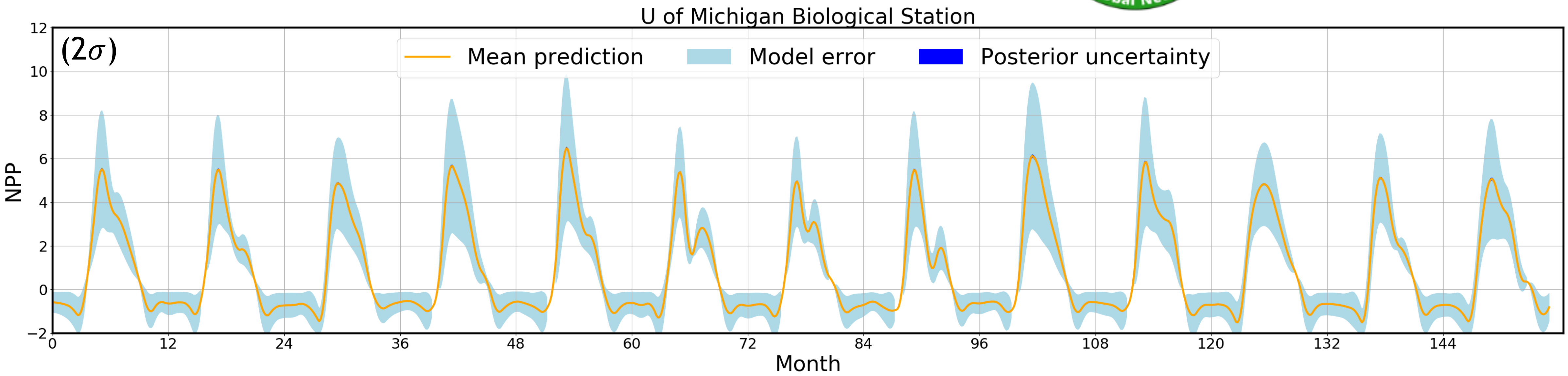


- Model error component dominates
- Captures model deficiency in summer months
- Indicates model improvement opportunities
- For further improvement: more intrusive embedding

LHF = Latent Heat Flux

Embedded model error allows prediction of unobservable QoIs

Land model calibration given FLUXNET observations

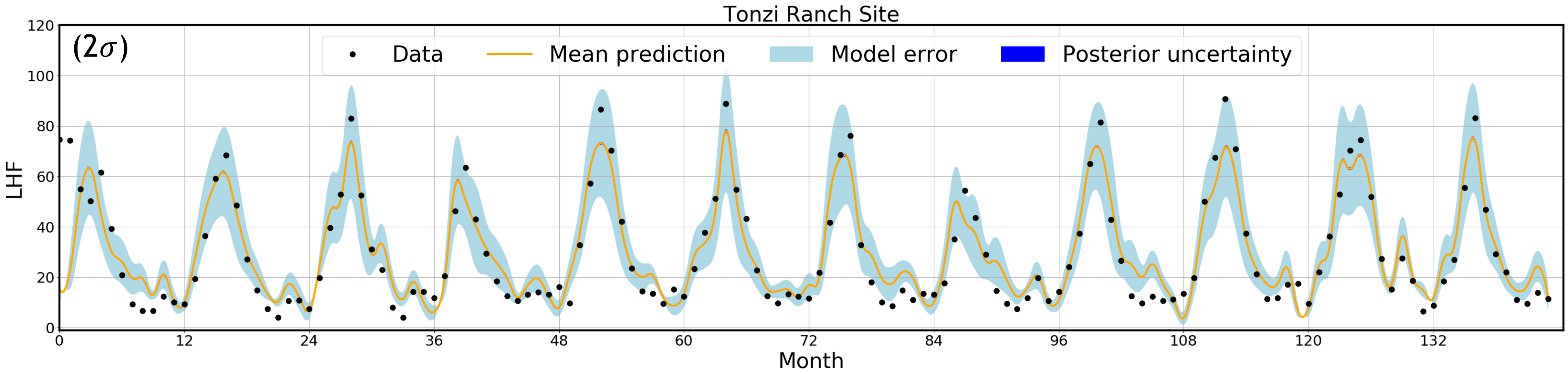


- Can be piped to human component or atmosphere model as a boundary condition

NPP = Net Primary Productivity

Embedded model error allows prediction at other FLUXNET sites

Land model calibration given FLUXNET observations



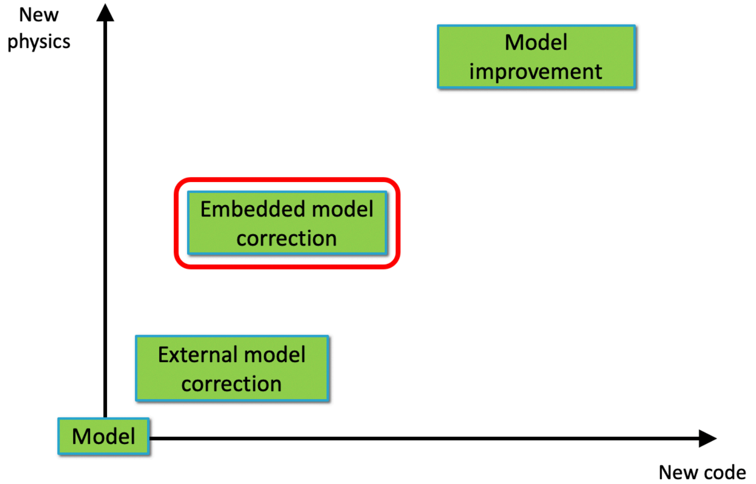
- Assumption: model goes wrong in a similar way

LHF = Latent Heat Flux

Unique capability for model exploration via embedded model error representation

Non-intrusive

$$y(x_i) = f(x_i; \lambda + \delta(x_i)) + \epsilon_i$$



- Allows meaningful extrapolation
- Respects physics
- Disambiguates model and data errors
- Predictive uncertainty attribution
  - surrogate errors
  - data noise
  - parametric uncertainty
  - **structural errors**

- Code available via UQTk (www.sandia.gov/uqtoolkit)
- Impacted many programs DOE/DOD/SNL
- Applied outside immediate group
- Provides alternative for external correction approach currently available in DAKOTA

Prediction variance = parametric uncertainty + data noise + model error

U.S. DEPARTMENT OF ENERGY | Office of Science

SciDAC Scientific Discovery through Advanced Computing

FASTMATH

Current/Future

- Method development under FASTMath and BER/ASCR SciDAC partnership
  - challenging Bayesian problem: approximate likelihood methods, informative priors
  - parameter *field* embedding: proven need for climate land model
  - probabilistic neural networks: many technical analogies; *see Ahmad Rushdi's poster*

Key References

- K. Sargsyan, H. Najm, R. Ghanem, "On the Statistical Calibration of Physical Models", Int. J. Chem. Kinetics, 47(4), 246-276, 2015, ~30 citations.

- K. Sargsyan, X. Huan, H. Najm. "Embedded Model Error Representation for Bayesian Model Calibration", arXiv:1801.06768, in press, Int. J. Uncert. Quant., 2019.

Contact Info: Khachik Sargsyan ksargsy@sandia.gov 925-294-4885 www.sandia.gov/~ksargsy

Additional slides

Model error can be defined in a variety of ways

In this work, model error is the difference between our model and the 'truth' model behind noisy data

$$\underbrace{f(x_i;\lambda)}_{\text{model}} \quad \text{vs} \quad \underbrace{y(x_i)}_{\text{data}} = \underbrace{g(x_i)}_{\text{truth}} + \underbrace{\epsilon_i}_{\text{noise}}$$

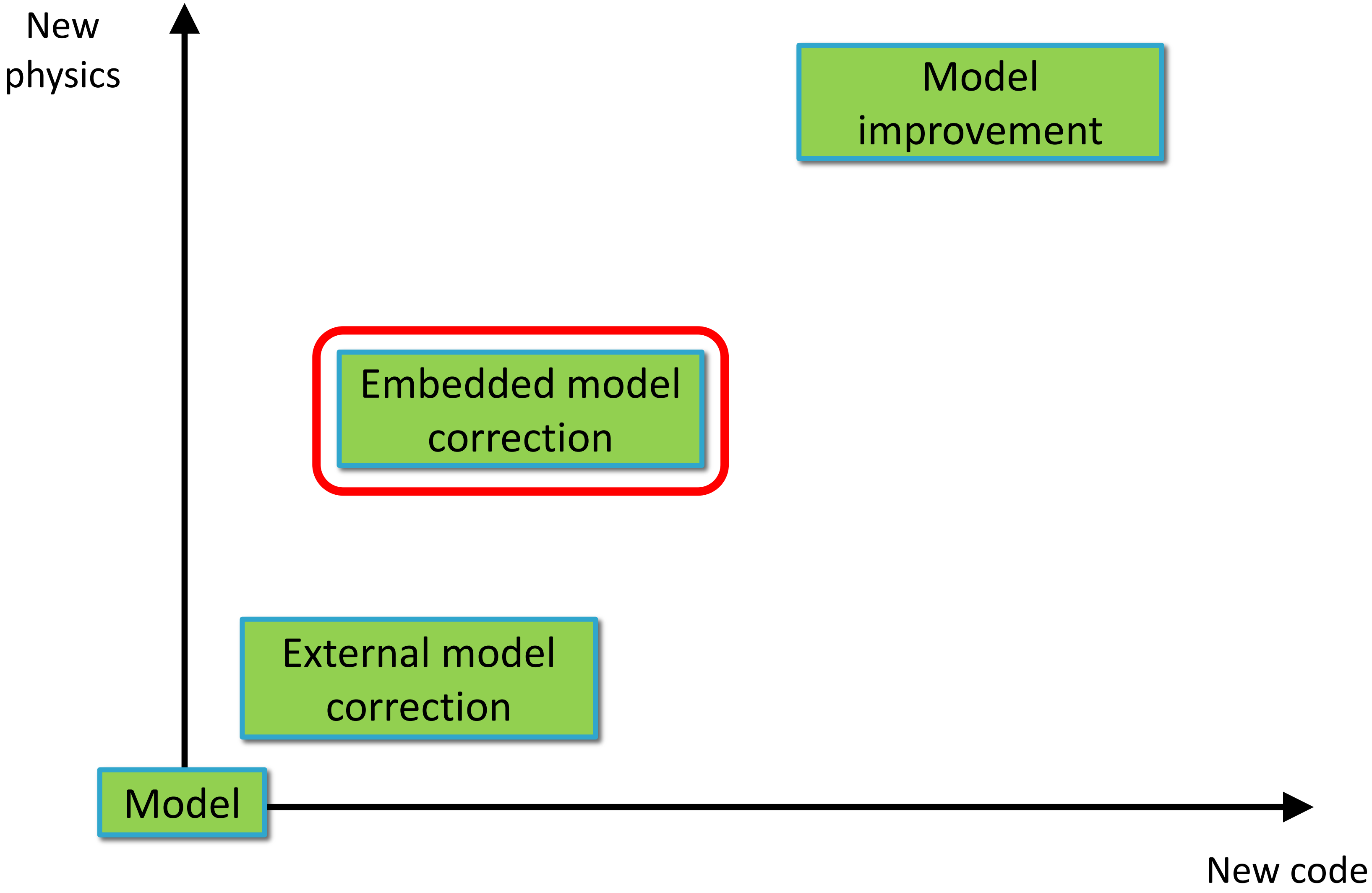
Model error is associated with

- Simplifying assumptions, parameterizations
- Mathematical formulation, theoretical framework

Very loaded concept
… otherwise called
(with altered meanings)

- model discrepancy
- model structural error
- model inadequacy
- model misspecification
- model form error
- model uncertainty

There is mature work in *statistical* model corrections, without physical interpretation



We develop approach for model structural error representation: embedded, but non-intrusive

External model correction

Non-intrusive

Mature work on synthetic statistical correction

Embedded model correction

Non-intrusive

Exploration tool: analyze given model, locate structural error

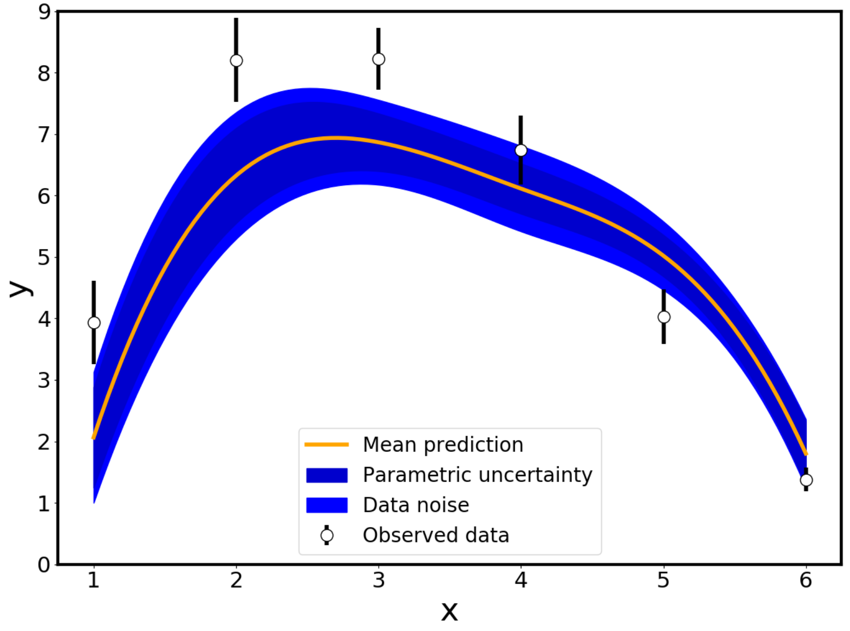
Embedded model correction

Intrusive

Improve/change the model, if you can

Non-intrusive = black-box

Prediction variance accounts for parametric uncertainty and data noise



Model Data

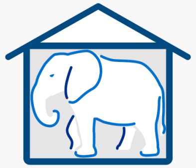
$$f(x; \lambda) \approx y(x)$$

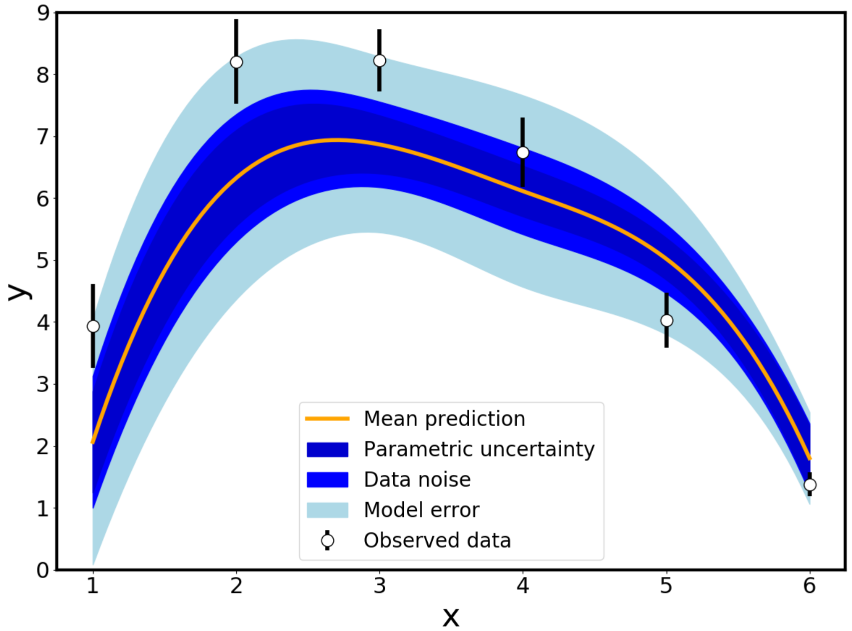
This is business as usual....

Posterior uncertainty (PU)

$$\underbrace{\text{parametric uncertainty} + \text{data noise}}$$

... assuming the model is perfect, i.e. represents the truth behind data

Elephant in the room:  **model *structural* error**



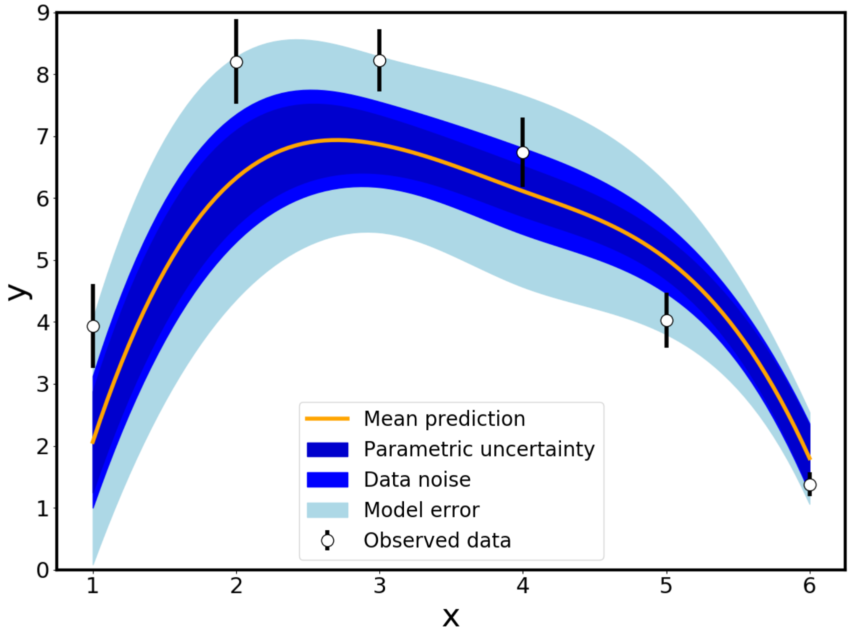
Model Data

$$f(x;\lambda) \approx y(x)$$

Uncertainty decomposition of model prediction needs to account for model error

$$\overbrace{\text{parametric uncertainty} + \text{data noise} + \underbrace{\text{model error}}_{\textbf{ME}}}^{\textbf{Posterior uncertainty (PU)}}$$

We want to incorporate model structural error in predictions



Model Data

$$f(x;\lambda) \approx y(x)$$

Uncertainty decomposition of model prediction needs to account for model error

Posterior uncertainty (PU)

$$\underbrace{\text{parametric uncertainty} + \text{data noise} + \text{model error}}_{\text{PU}}$$

$$\text{model error} = \text{ME}$$

External correction is not satisfactory for physical models

$$y_i = \underbrace{f(x_i;\lambda) + \delta(x_i)}_{\text{truth } g(x_i)} + \epsilon_i$$

- Explicit additive statistical model for model error [KOH, 2001]
- Potential violation of physical constraints
- Disambiguation of model error $\delta(x_i)$ and data error ϵ_i
- Yes, priors help: [Brynjarsdottir and O'Hagan, 2014], [Plumlee, 2017]
- Calibration of model error on measured observable does not impact the quality of model predictions on other QoIs
- Physical scientists are unlikely to augment their model with a statistical model error term on select outputs
  - Calibrated predictive model: $f(x;\lambda) + \delta(x)$ or $f(x;\lambda)$?
- Problem is highlighted in model-to-model calibration ($\epsilon_i = 0$)
  - no a priori knowledge of the statistical structure of $\delta(x)$

Calibrate $f(x; \lambda)$, given data $g(x)$

x are operating conditions, design parameters, various QoIs
λ are model parameters to be inferred/calibrated

$$g(x) = f(x; \lambda) + \epsilon$$

- Default: Ignore model errors:
  - Biased or overconfident physical parameters
  - Wrong model predictions

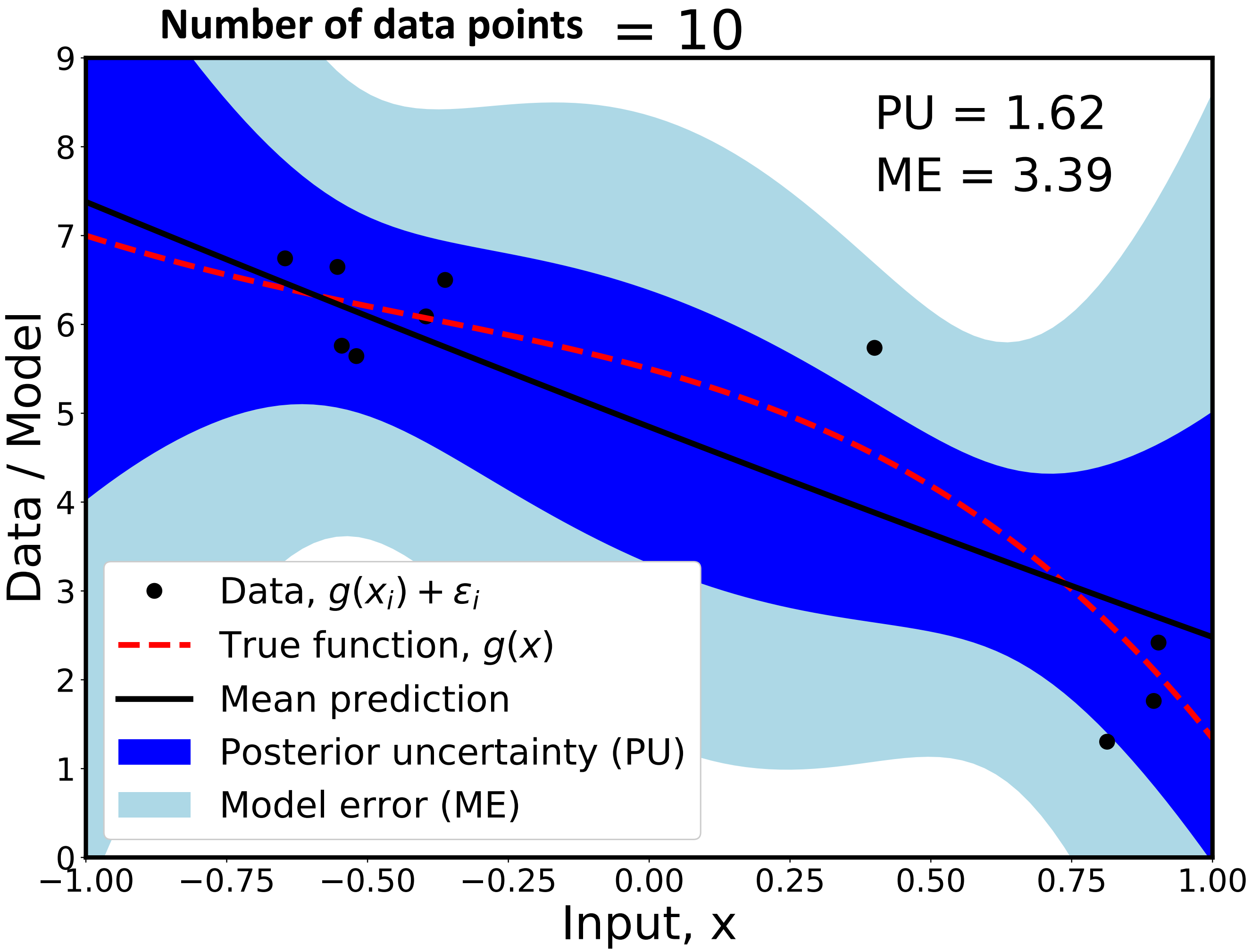
$$g(x) = f(x; \lambda) + \delta(x) + \epsilon$$

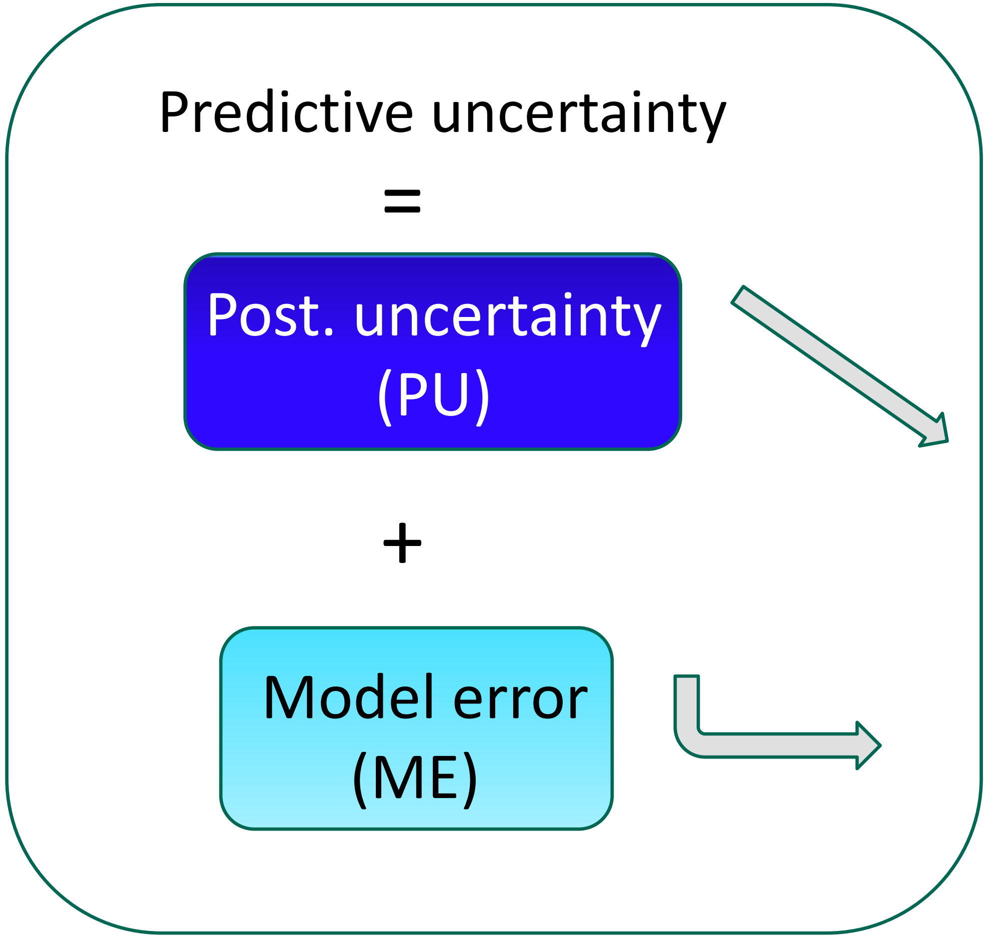
- Conventional: Correct for model errors:
  - Physical parameters are ok
  - Wrong model predictions (data-specific corrections)
  - Model and data errors mixed up

$$g(x) = f(x; \lambda + \delta(x)) + \epsilon$$

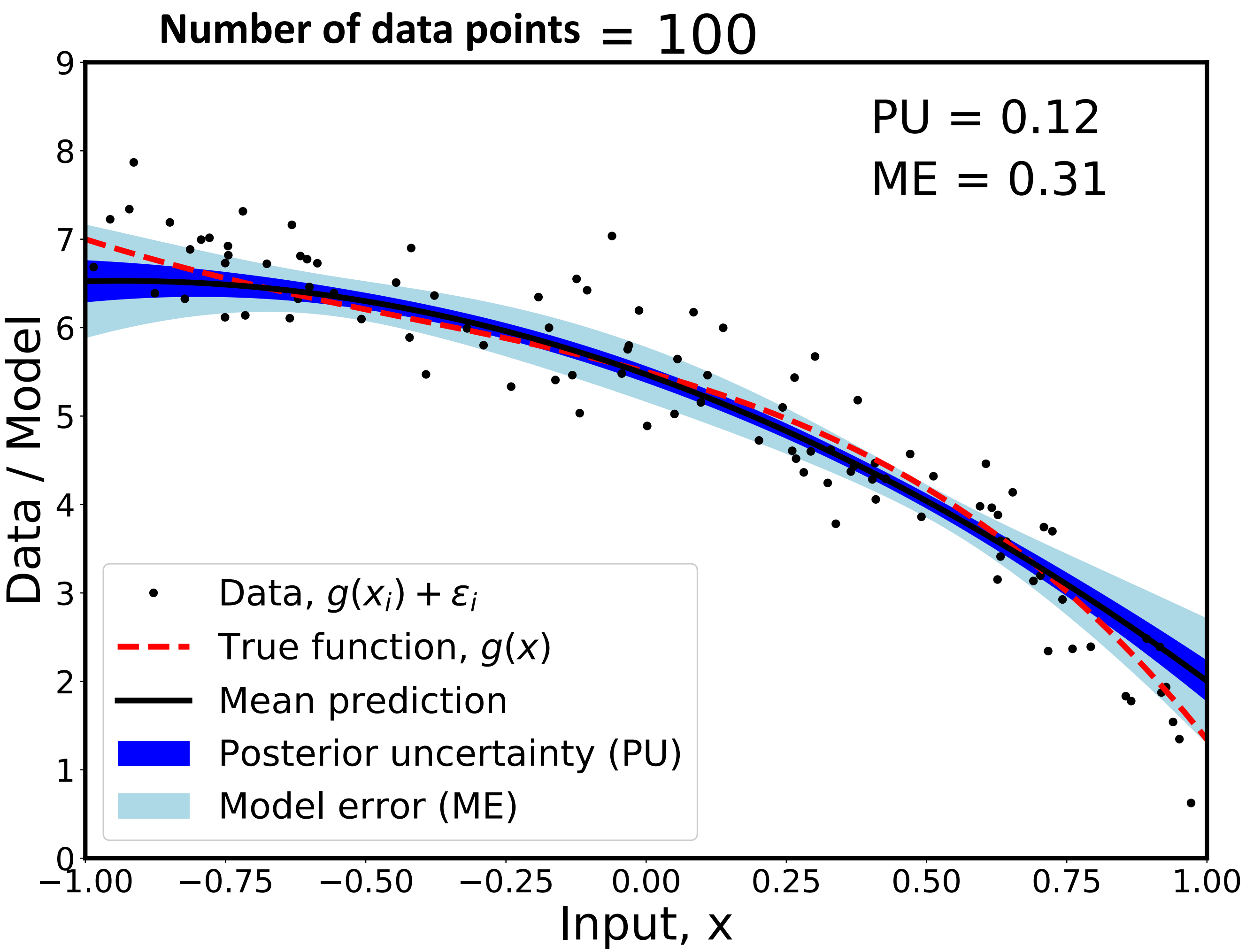
- What we do: Correct *inside* the model:
  - Embedded model error
  - Preserves model structure and physical constraints
  - Disambiguates model and data errors
  - Allows meaningful extrapolation

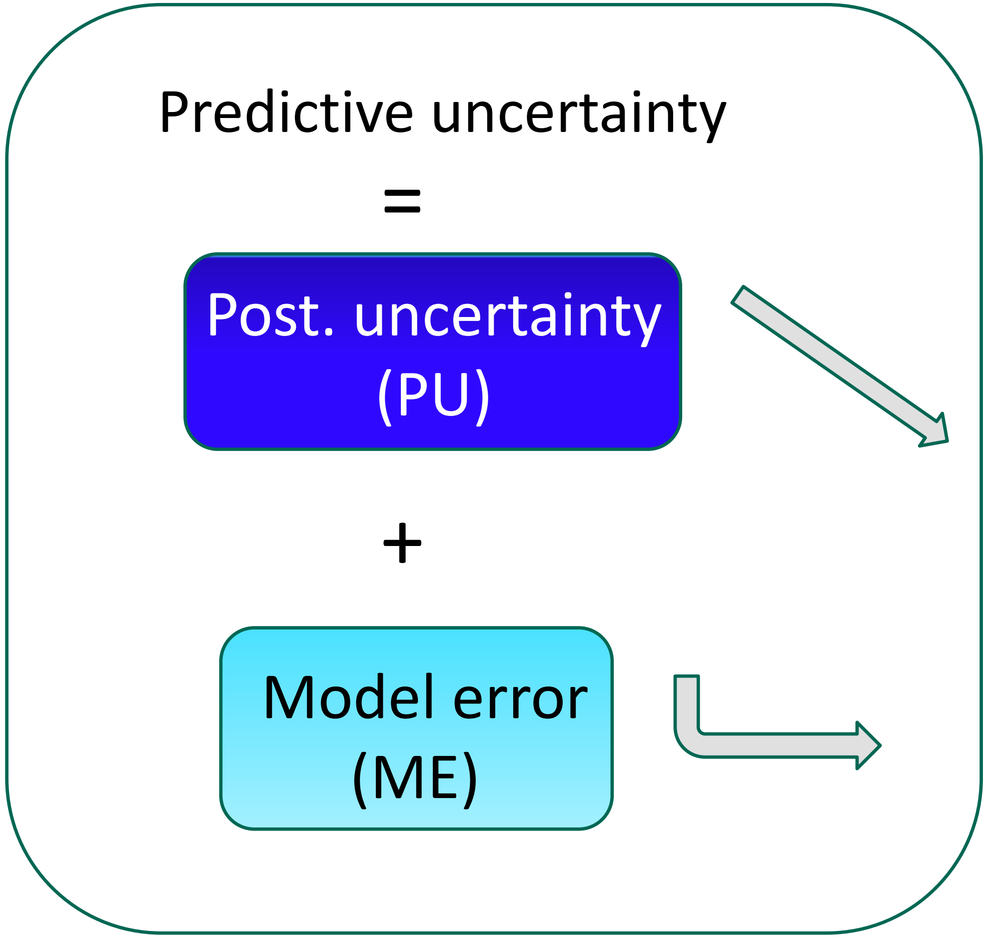
With more data points, posterior uncertainty (PU) goes down, but model error (ME) saturates to a limiting value



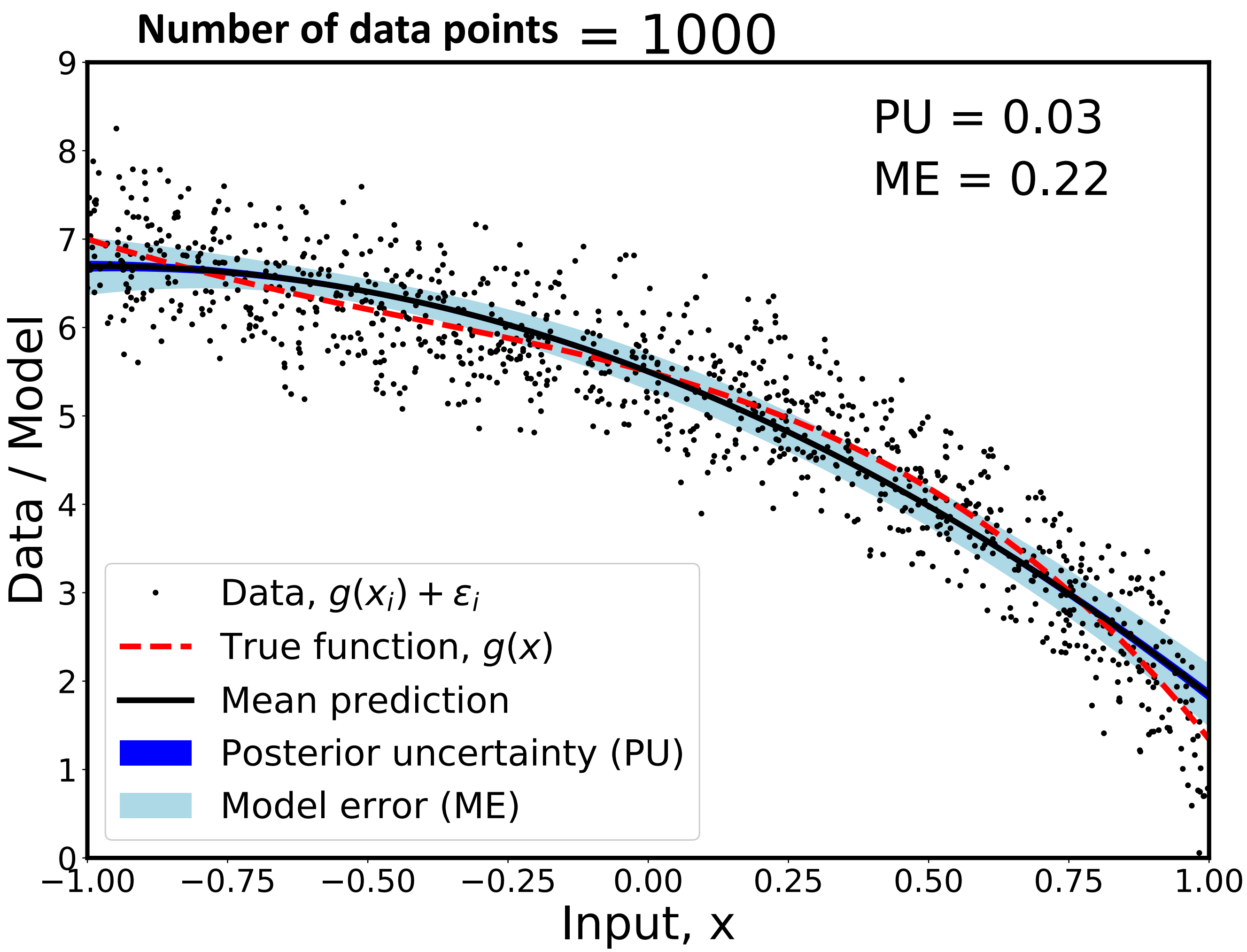


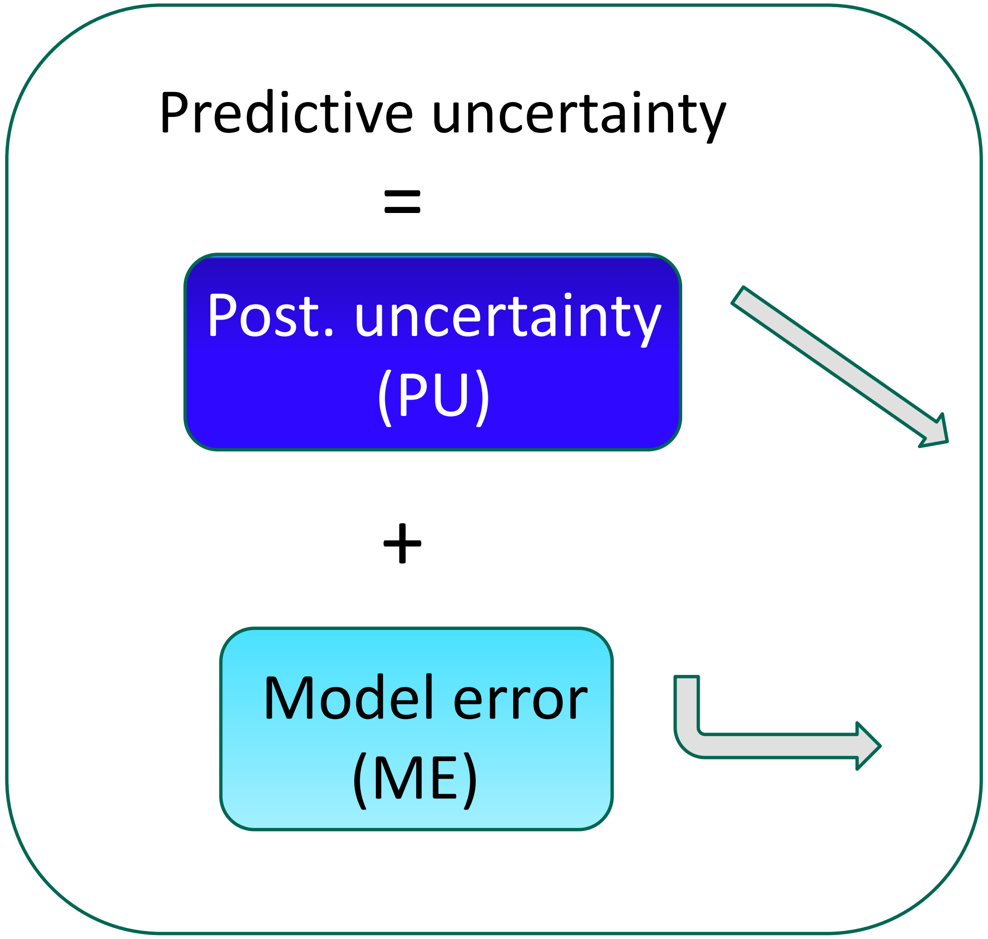
With more data points, posterior uncertainty (PU) goes down, but model error (ME) saturates to a limiting value



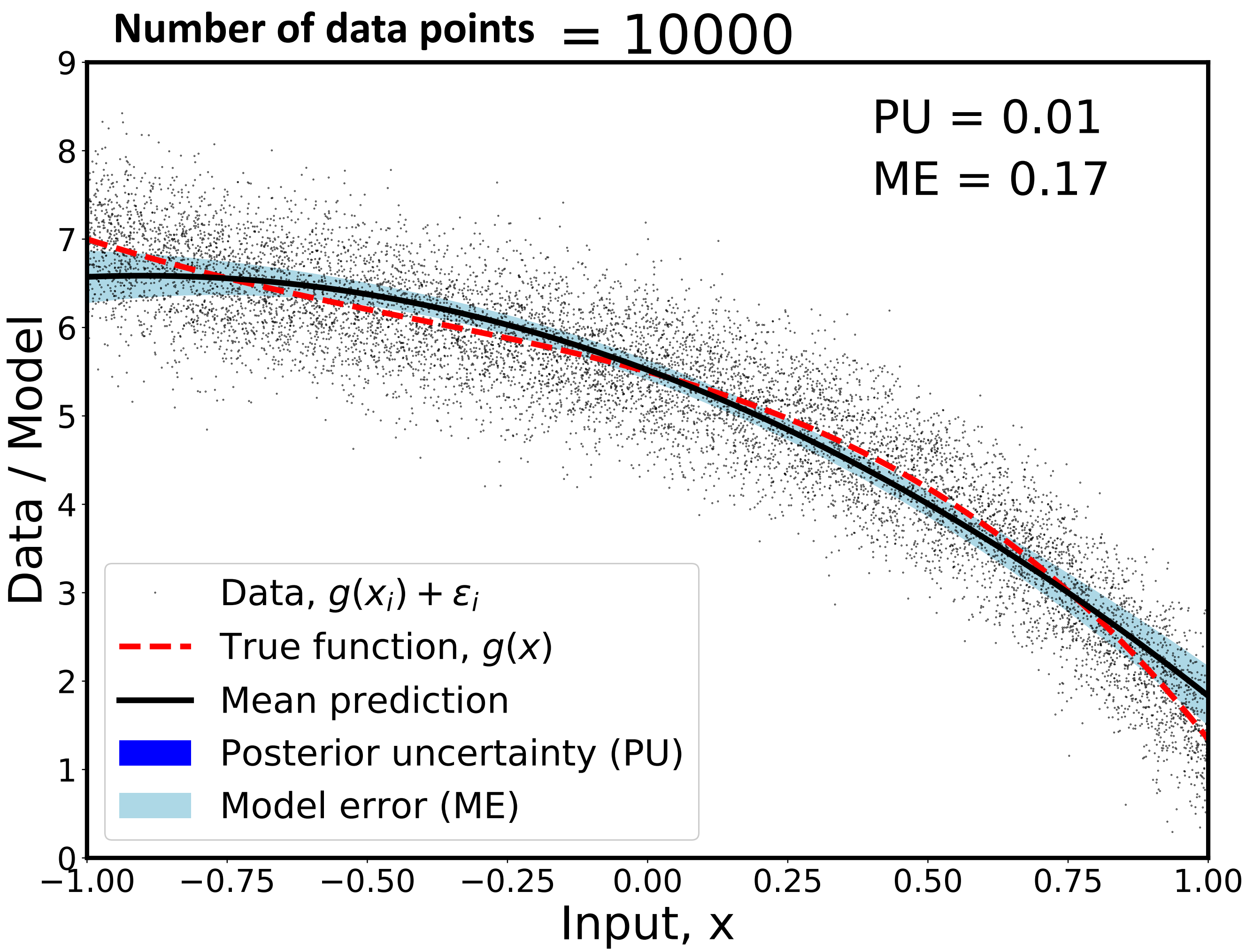


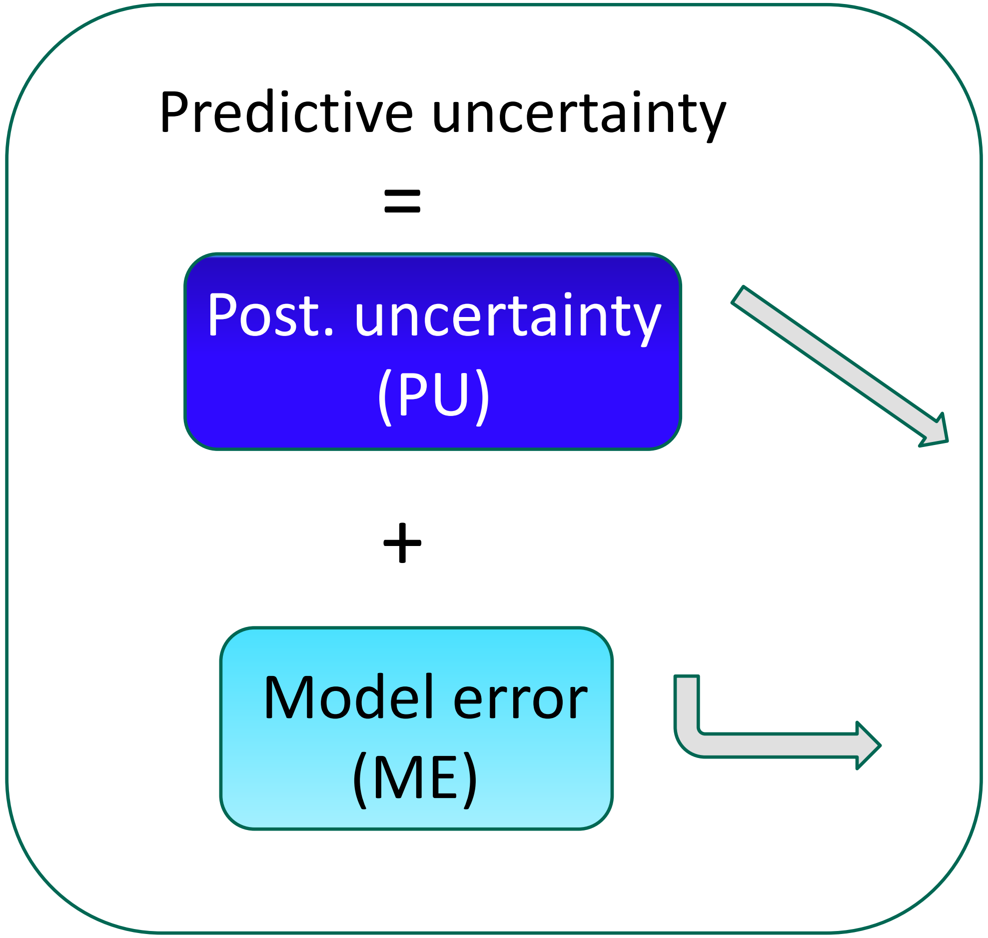
With more data points, posterior uncertainty (PU) goes down, but model error (ME) saturates to a limiting value



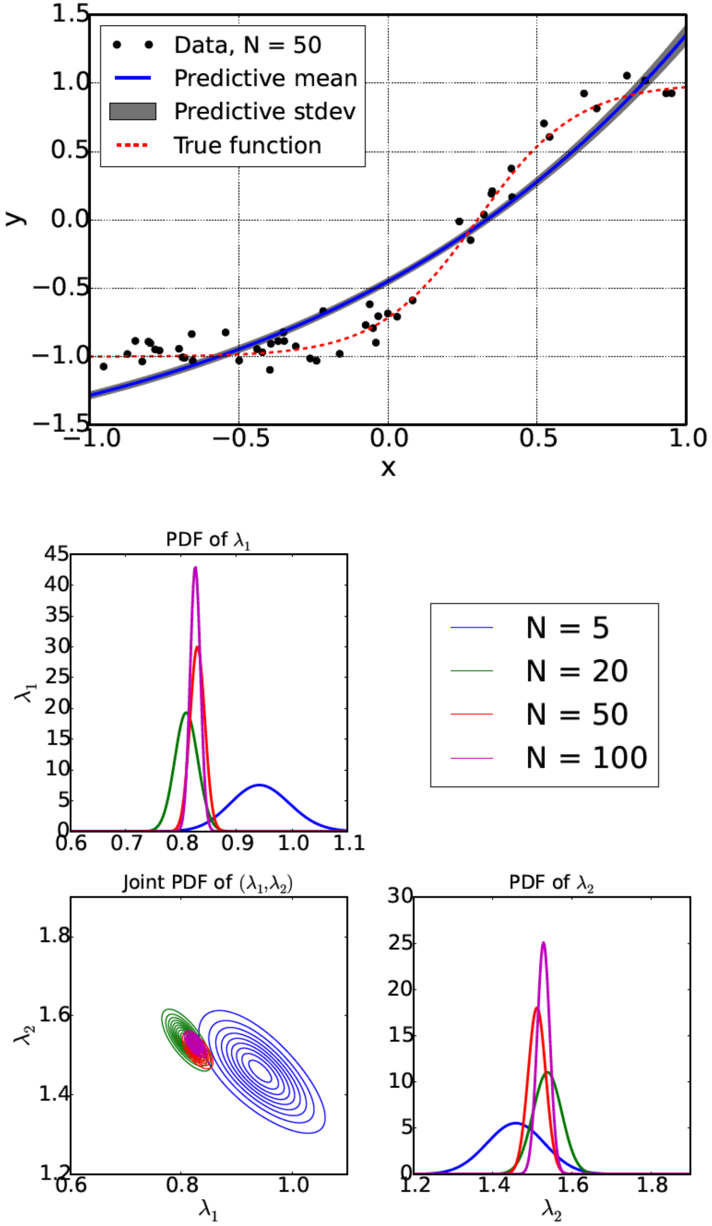


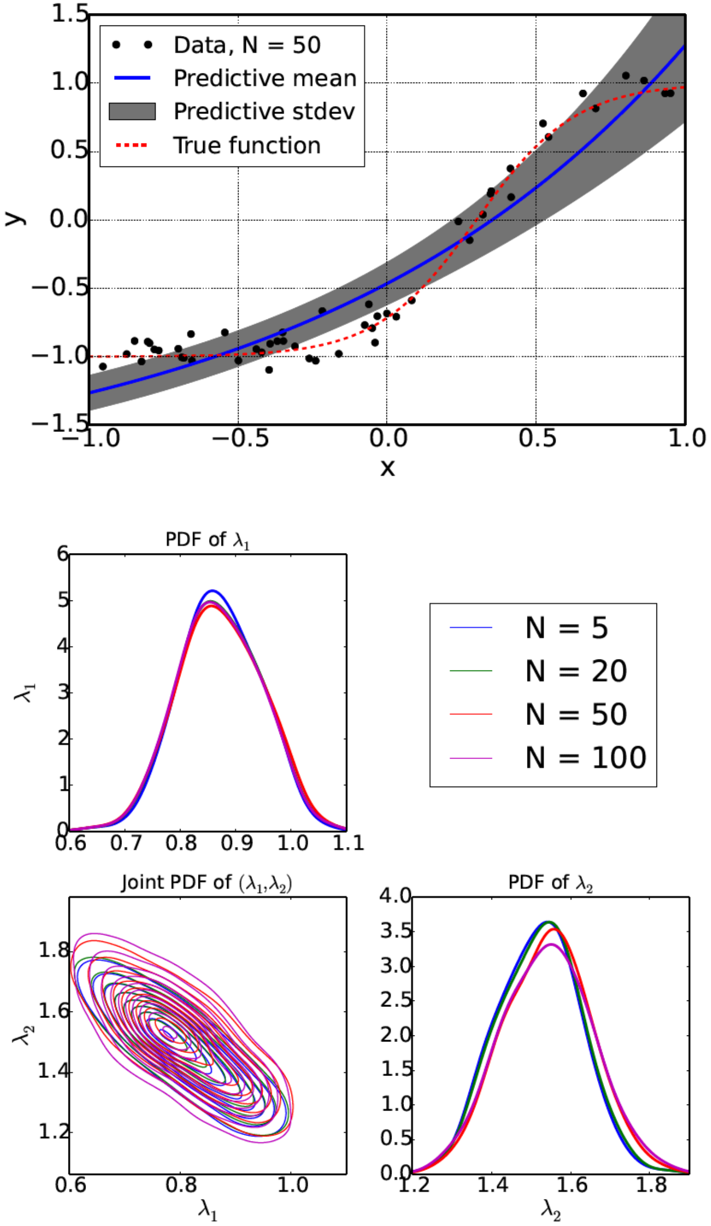
Irrespective of data amount, model error captures the discrepancy between model and the truth





Back to toy example

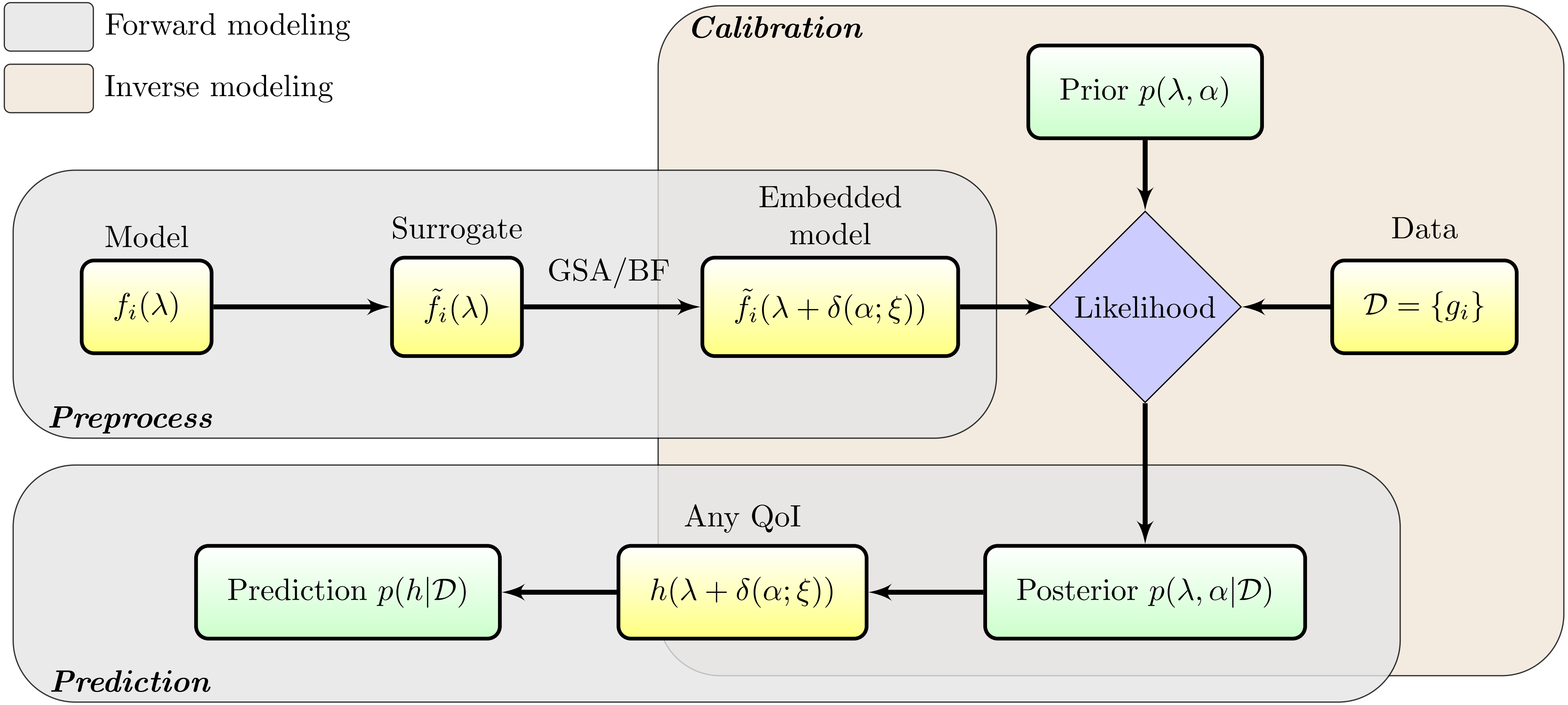




Predictive uncertainty captures model error

Stable prediction of “physical” parameters of the exponential function

Workflow


Forward modeling
Inverse modeling
Calibration
Prior $p(\lambda, \alpha)$
Embedded
model
Model
Surrogate
Data
GSA/BF
$f_i(\lambda)$
$\tilde{f}_i(\lambda)$
$\tilde{f}_i(\lambda + \delta(\alpha; \xi))$
Likelihood
$\mathcal{D} = \{g_i\}$
Preprocess
Any QoI
Prediction $p(h|\mathcal{D})$
$h(\lambda + \delta(\alpha; \xi))$
Posterior $p(\lambda, \alpha|\mathcal{D})$
Prediction

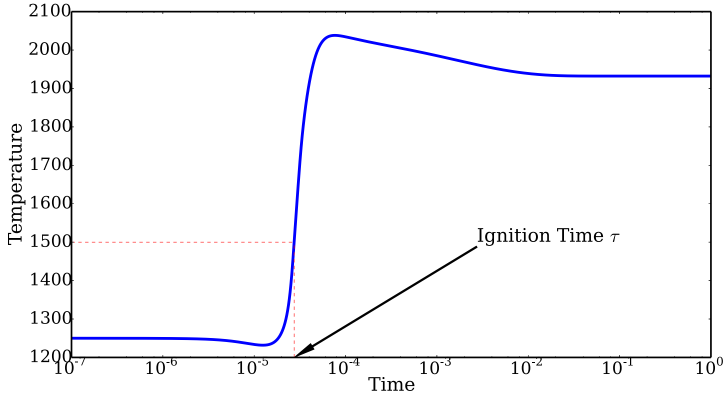
Application 2: Chemistry

funded by **BES**

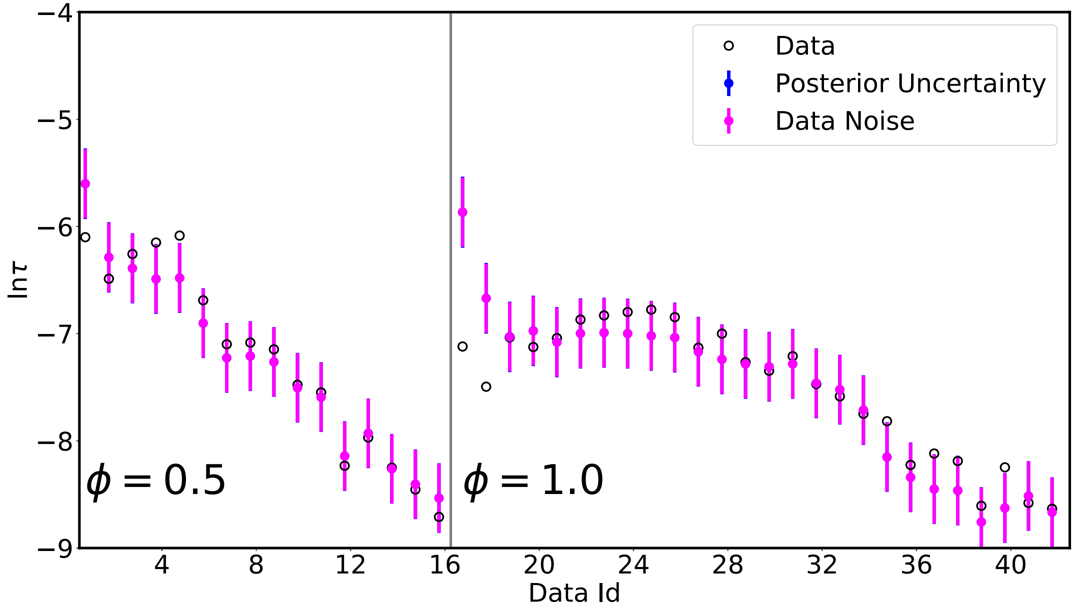
- A sandbox for developing the method
- Calibrating a simple 2-step reaction mechanism

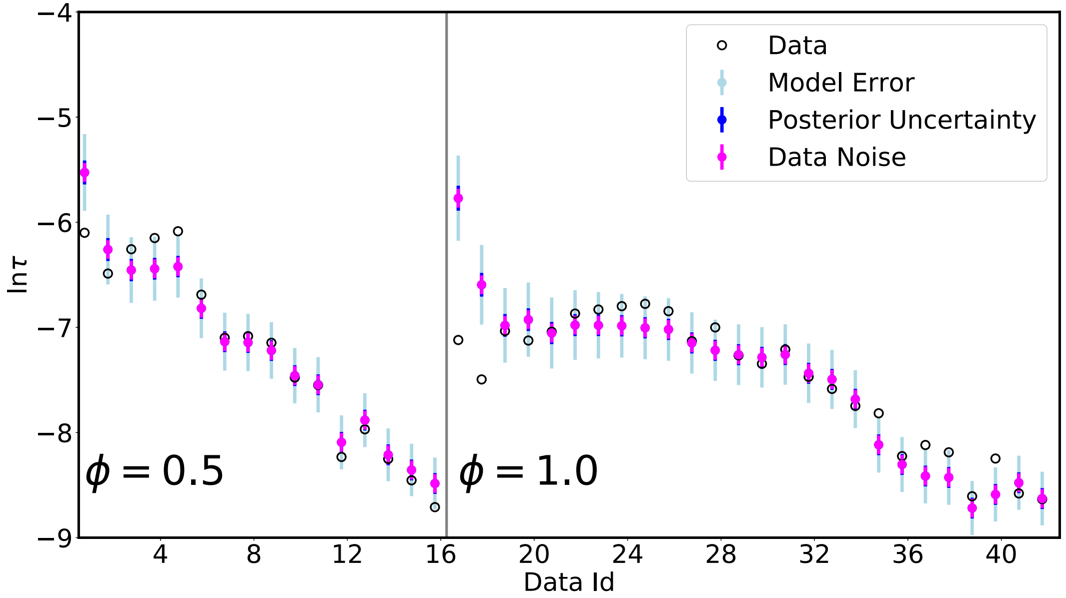
given high-fi model or experimental data

QoI: log-ignition time



Without model error,
all the discrepancy
is attributed to data noise



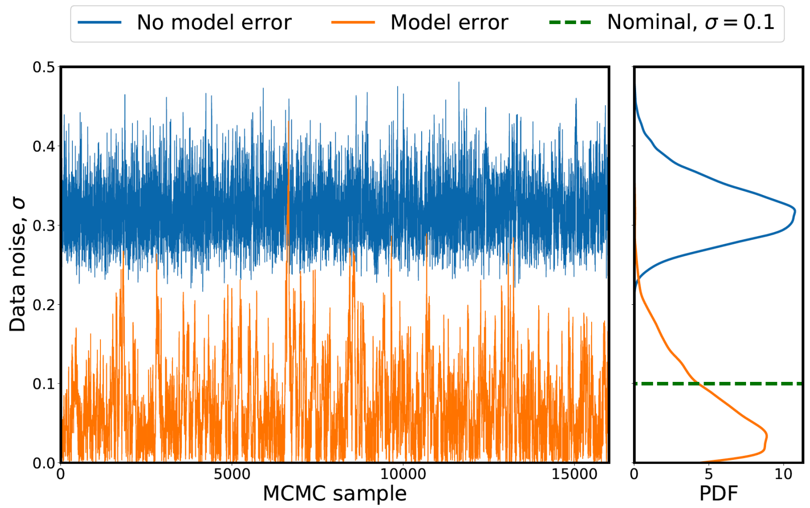


Sargsyan, Najm, Ghanem, IJCK (2015); Sargsyan, Huan, Najm, IJUQ (2019); Hakim et. al, CTM (2018)

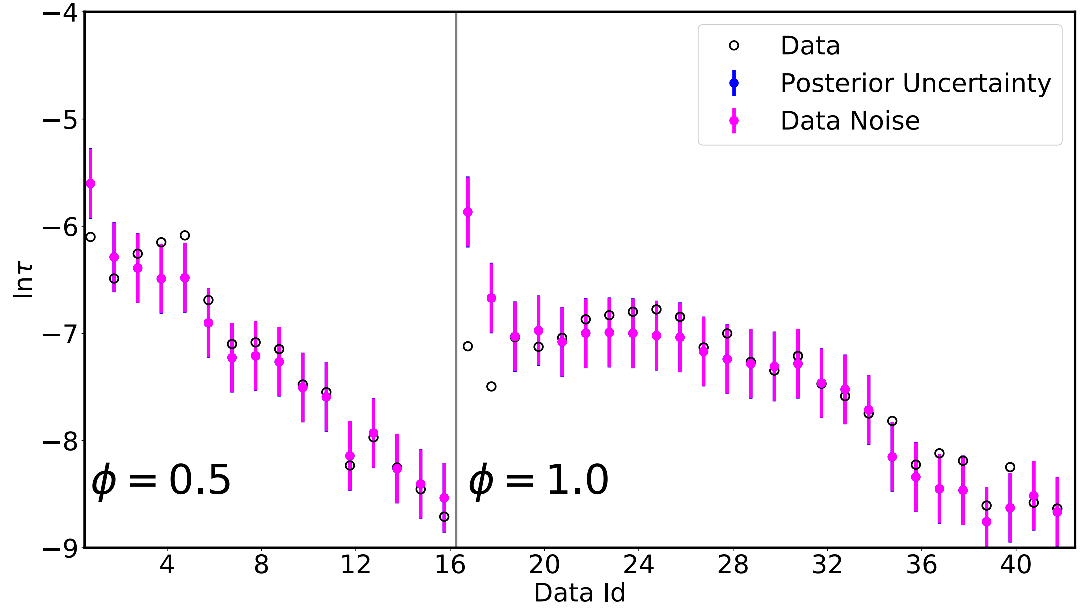
Application 2: Chemistry

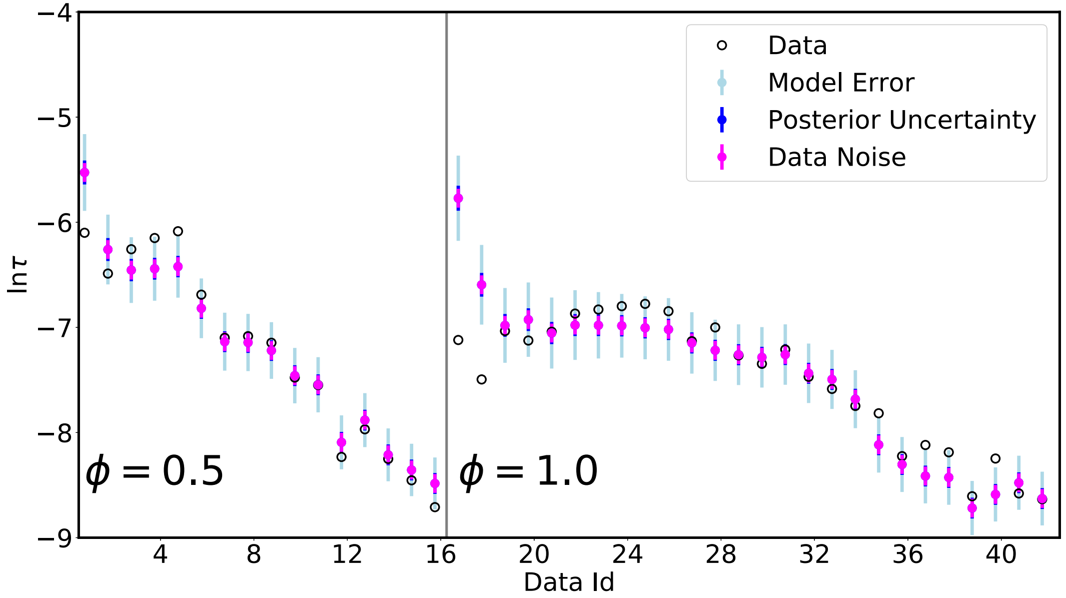
funded by **BES**

- A sandbox for developing the method
- Calibrating a simple 2-step reaction mechanism given high-fi model or experimental data



Without model error all the discrepancy is attributed to data noise



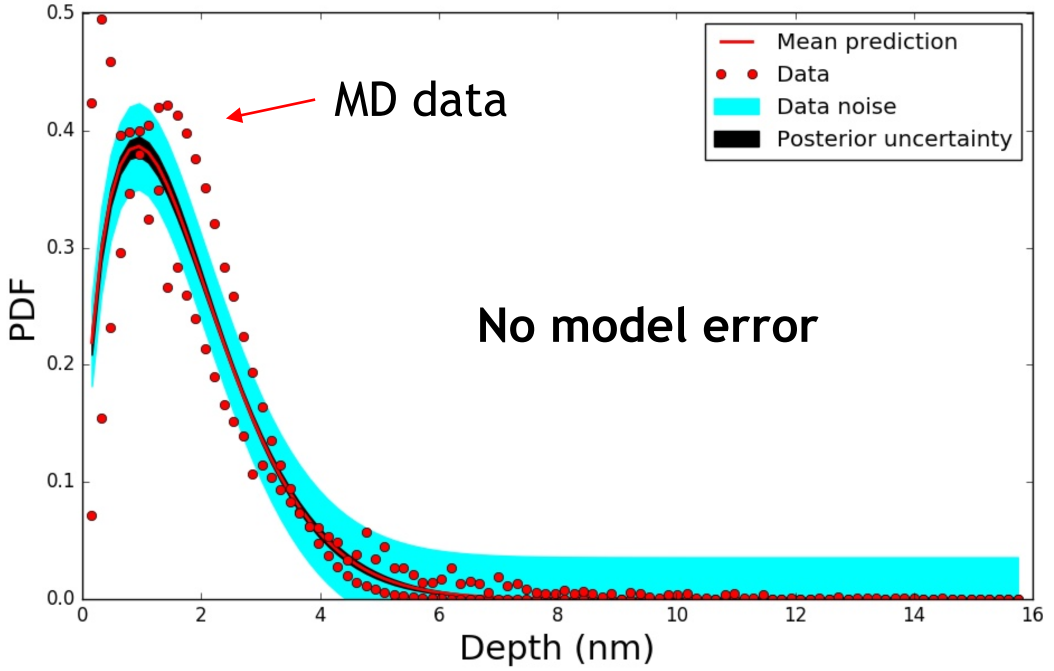


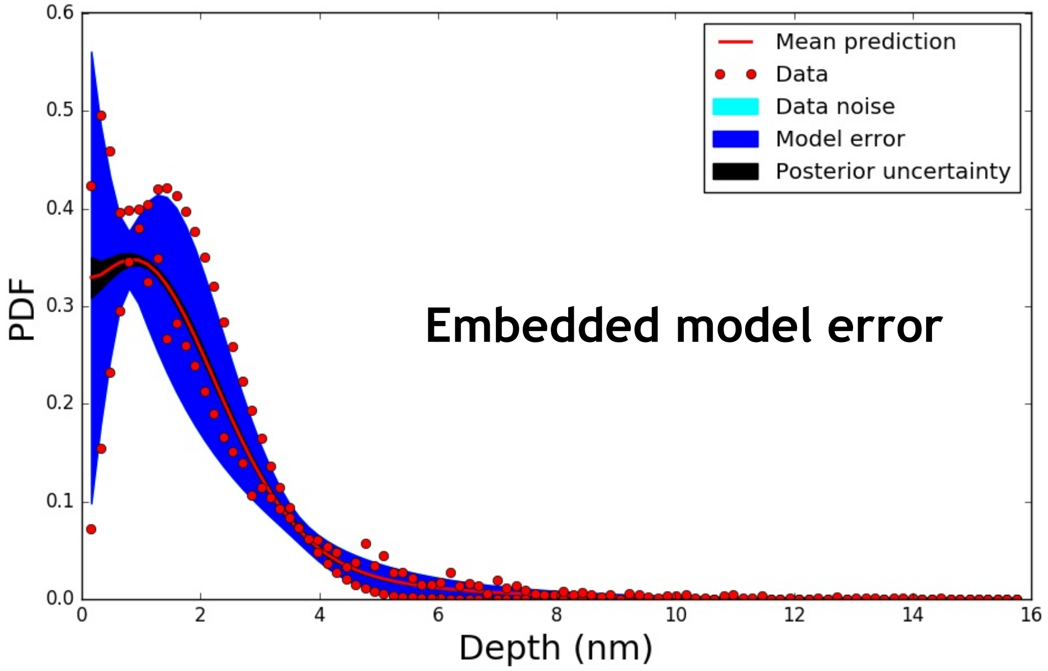
Sargsyan, Najm, Ghanem, IJCK (2015); Sargsyan, Huan, Najm, IJUQ (2019); Hakim et. al, CTM (2018)

Application 3: Plasma Surface Interaction funded by **FES+ASCR**

- Multi-institution partnership, but direct collaboration with ORNL and UTK
- Constructing uncertain input profiles for tungsten depth to propagate through Xolotl (PSI code)

Ignoring model error wrongly attributes uncertainty to data

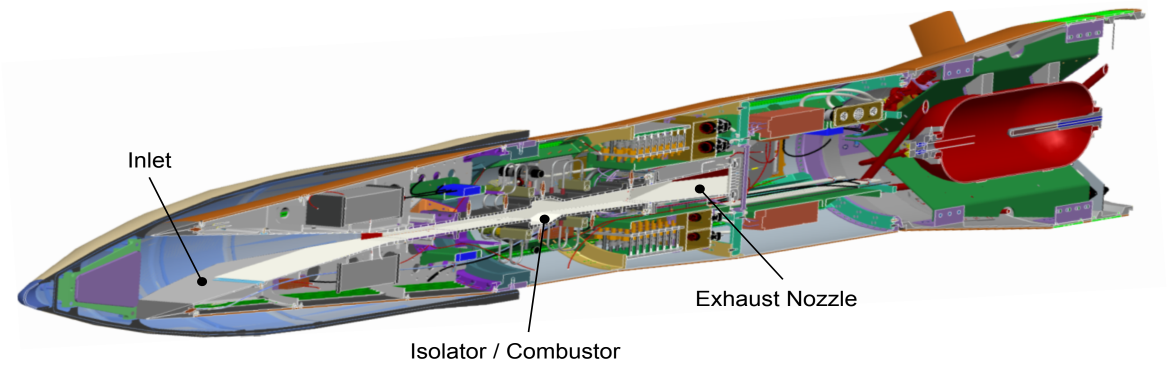


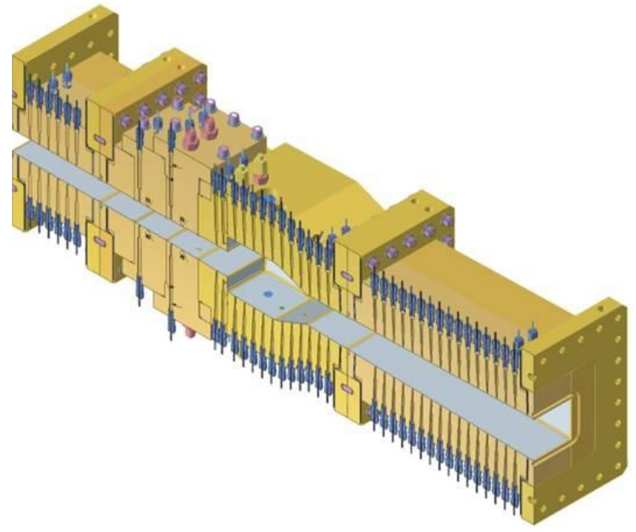


O. Cekmer, *K. Sargsyan*, S. Blondel, H. Najm, D. Bernholdt,, B.D. Wirth, "Uncertainty quantification for incident helium flux in plasma-exposed tungsten", *Int. J. Uncertainty Quantification*, Vol. 8, No. 5, p.429–446, 2018.

Application 4: Turbulent Flow

funded by **DARPA**





Advance the state of the art in UQ methods and software, in Large Eddy Simulation (LES) of a laboratory scale Scramjet combustor NASA Langley Hypersonic International Flight Research and Experimentation (HIFiRE) configuration

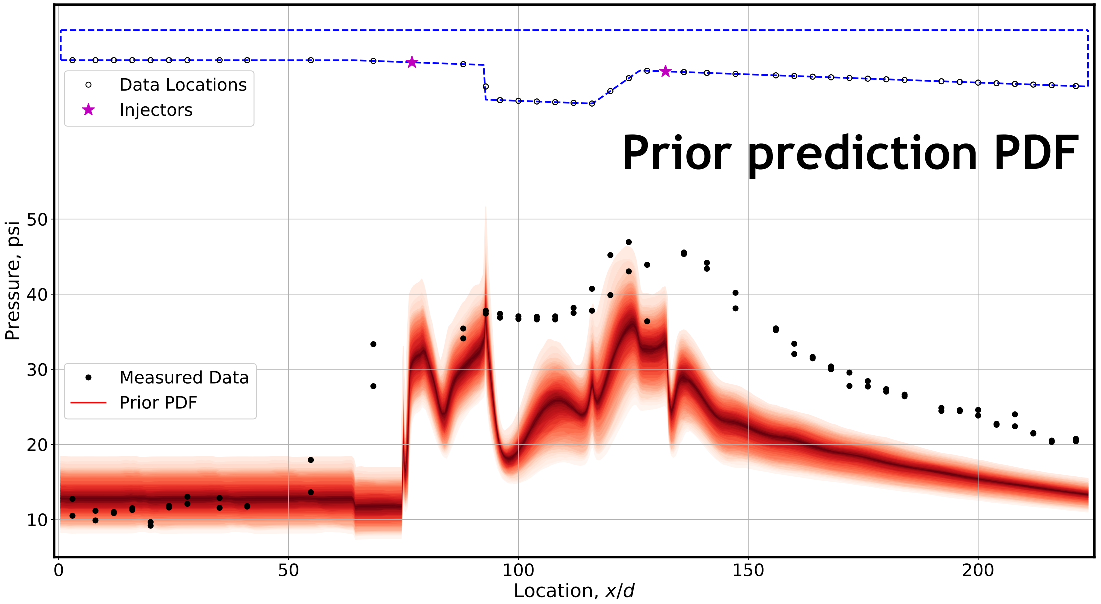
Application 4: Turbulent Flow

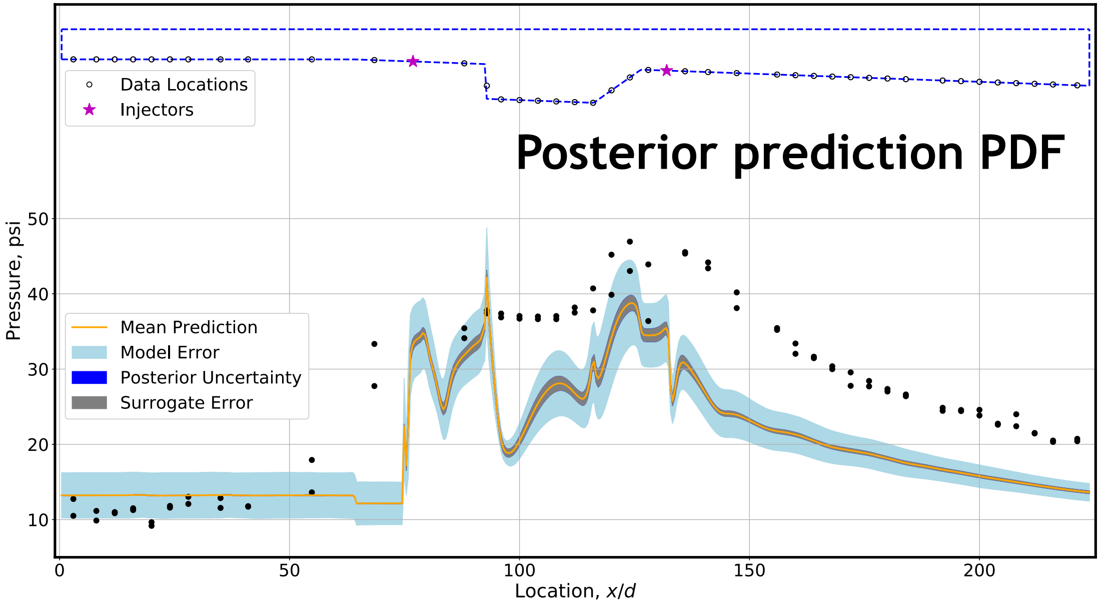
funded by **DARPA**

- Major UQ challenges for turbulent flow (LES)
  - Nonlinear dynamics
  - Large number of uncertain parameters
  - **LES model structural error**
  - Optimize design under uncertainty

Experimental data obtained from NASA

Model error is the main contributor of the predictive variance





X. Huan, C. Safta, *K. Sargsyan*, G. Geraci, Michael S. Eldred, Zachary P. Vane, G. Lacaze, Joseph C. Oefelein, Habib N. Najm, "Global Sensitivity Analysis and Estimation of Model Error, toward Uncertainty Quantification in Scramjet Computations", *AIAA Journal*, Vol. 56, No. 3, p.1170–1184, 2018

Applications (other)

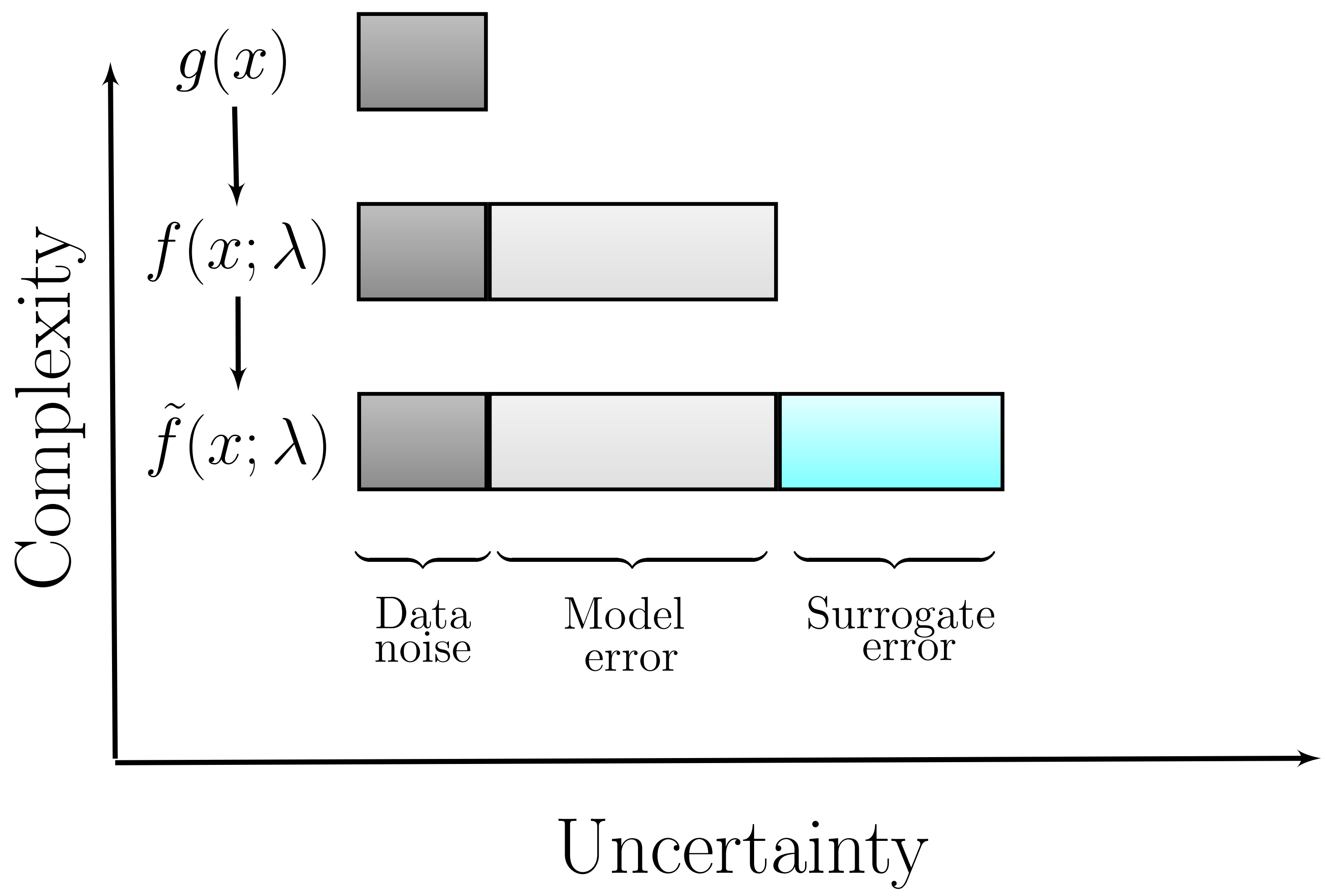
Thermodynamics (DOE EERE), work by Bert Debusschere

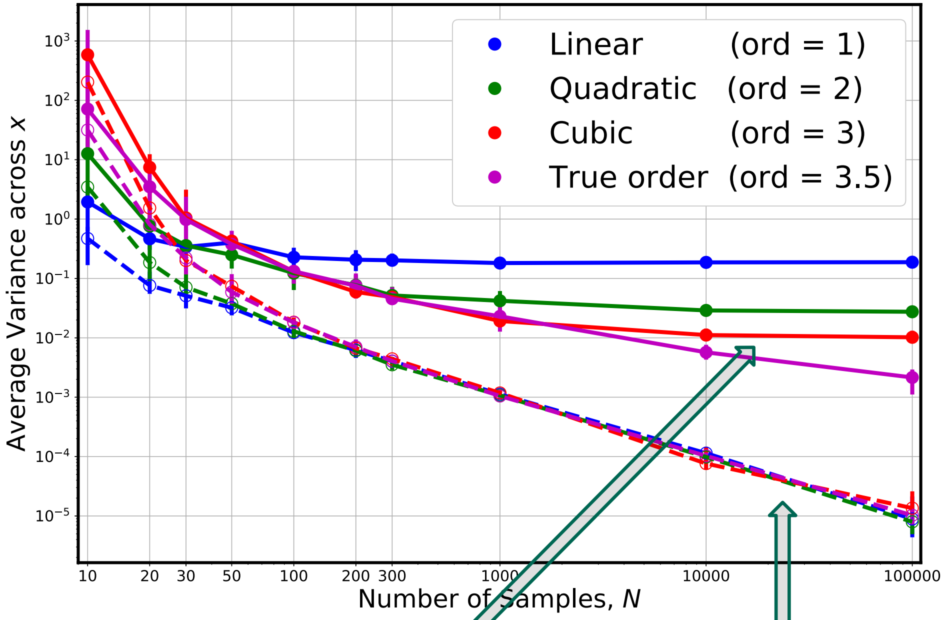
- The UQTk model error tools have been used to study the impact of using lower fidelity models for the thermodynamics of redox active materials that are used in water splitting methods for Hydrogen generation.
- This work is part of the HydroGEN Advanced Water Splitting Materials project (https://www.h2awsm.org/). HydroGEN is funded by DOE’s Fuel Cell Technologies Office in the Office of Energy Efficiency and Renewable Energy (EERE).

Flood Modeling (NSF/UMichigan), collaboration/advising within DOE SCGSR program

Materials Science (DOE/LDRD): work by Reese Jones, Francesco Rizzi

Uncertainty Budget

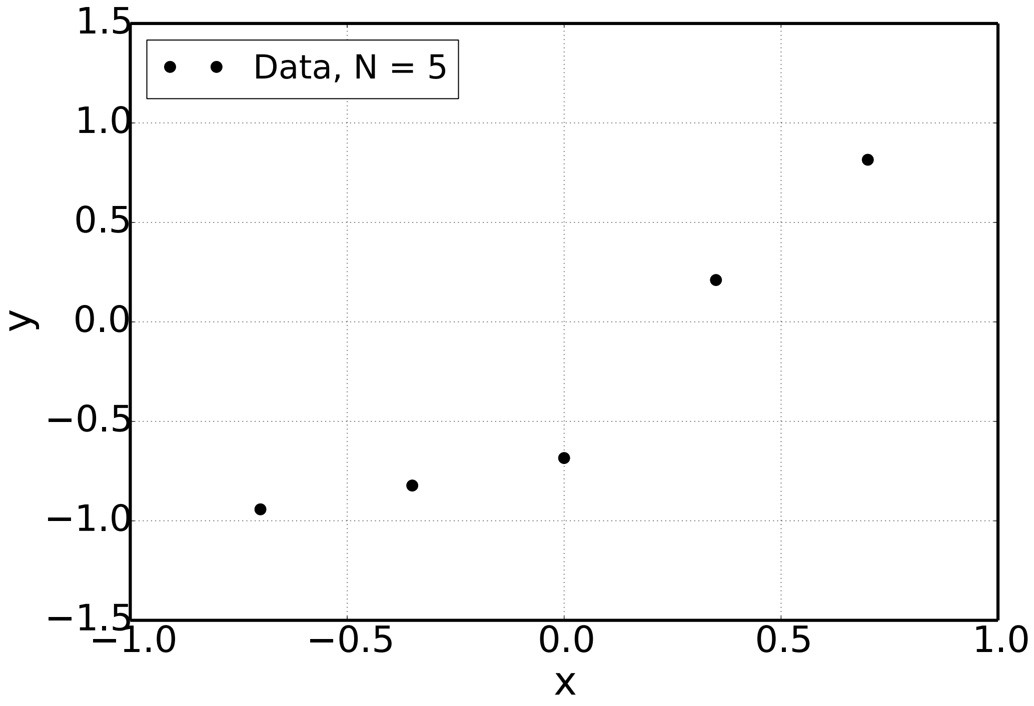




Model error (ME)

Post. uncertainty (PU)

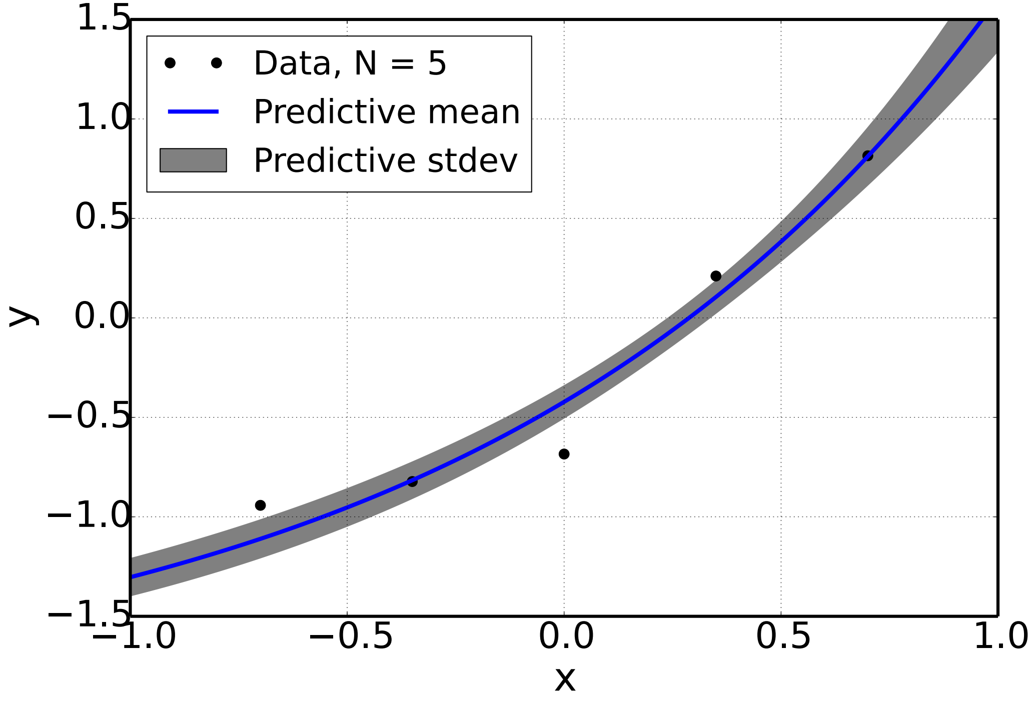
Wrong model leads to biased estimation



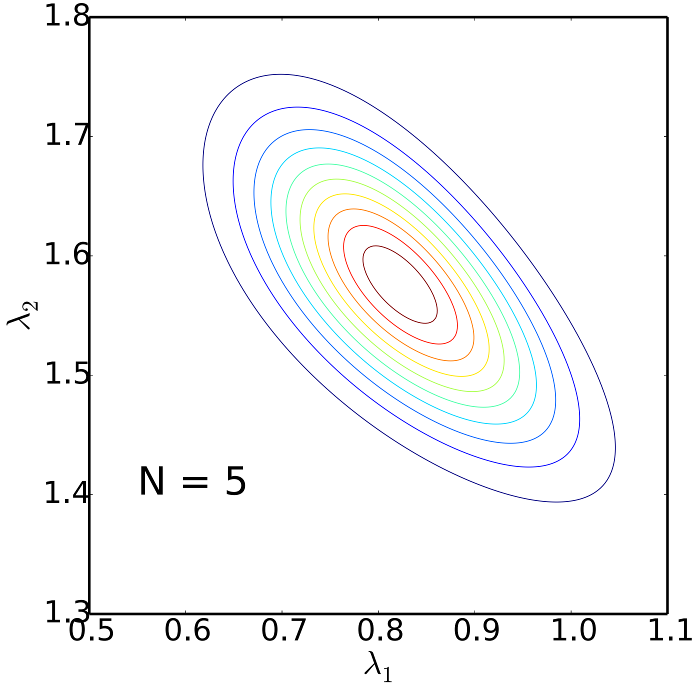
Given noisy data

Wrong model leads to biased estimation

Model prediction vs data



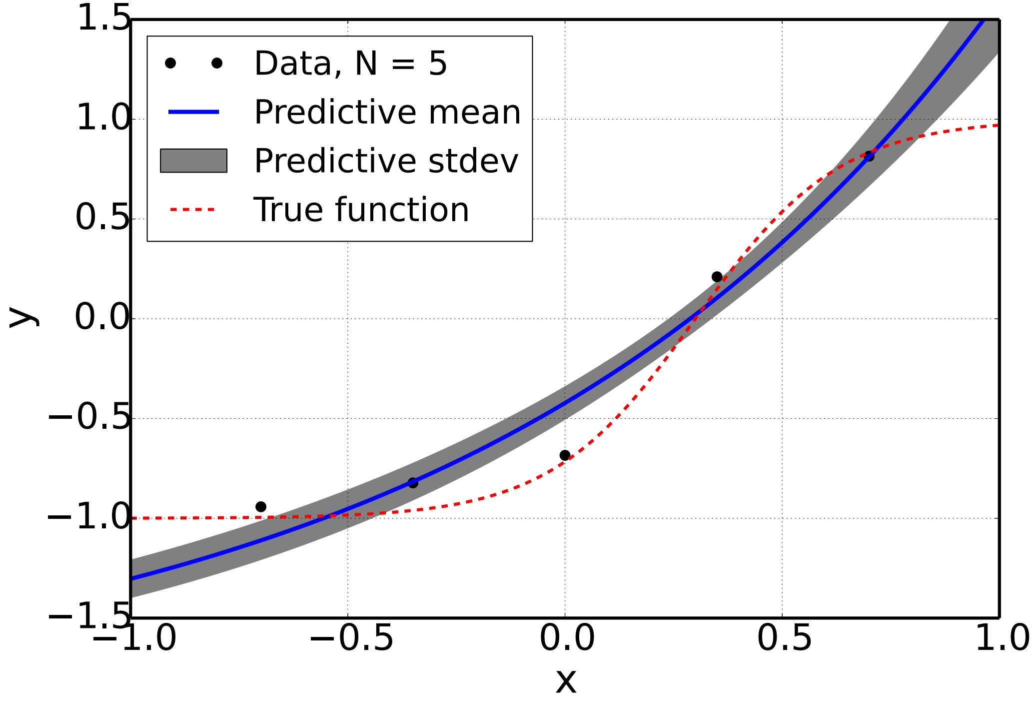
Posterior PDF of exponential parameters



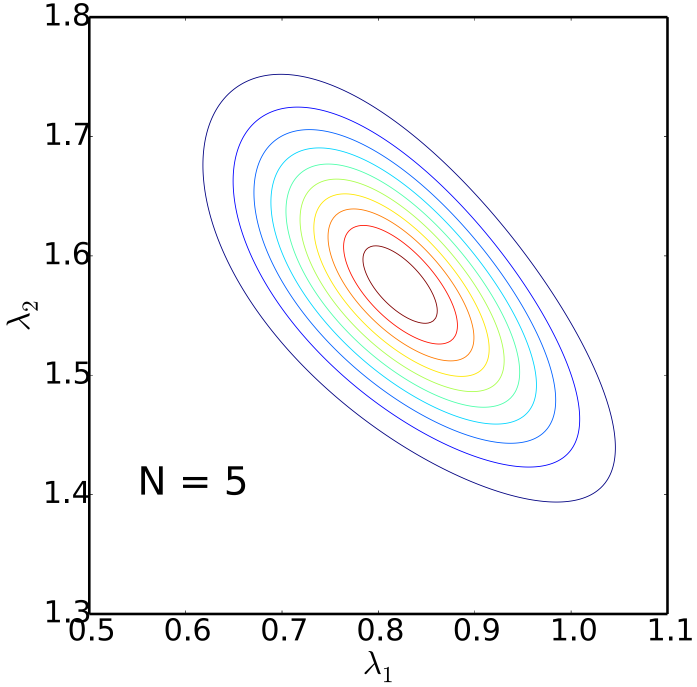
Calibrate an exponential model

Wrong model leads to biased estimation

Model prediction vs data



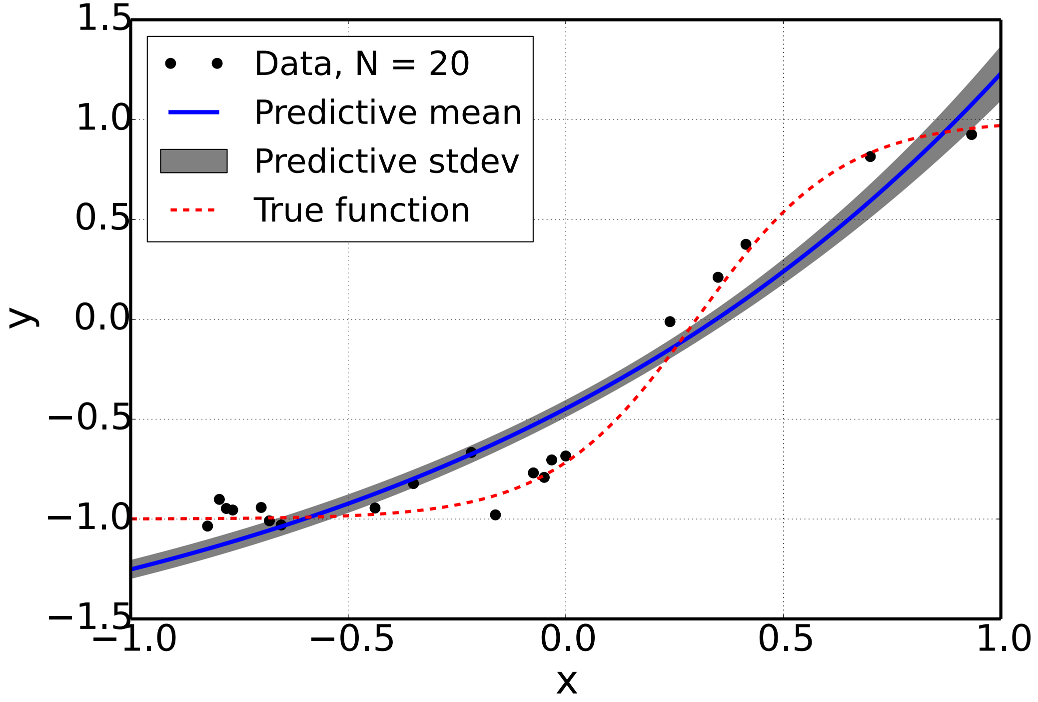
Posterior PDF of exponential parameters



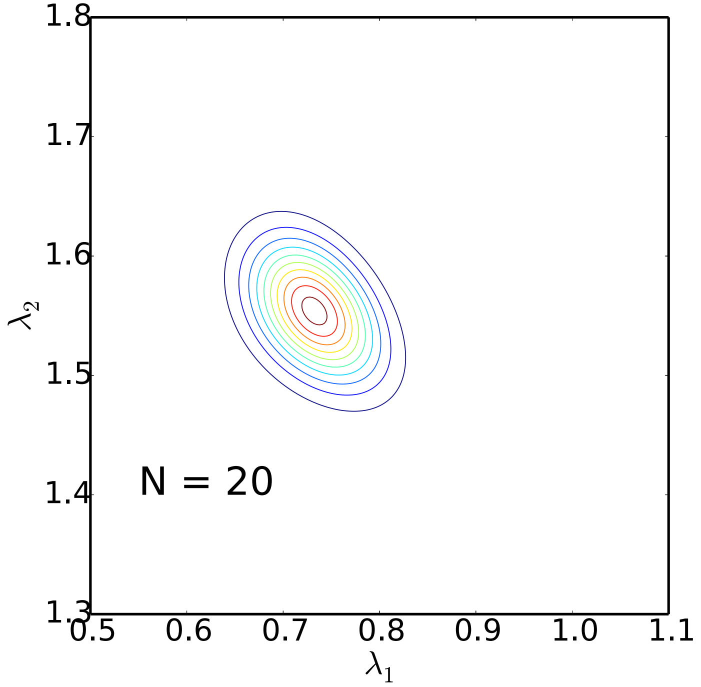
Calibrate an exponential model, but data comes from a different function (there is model error!)

Wrong model leads to biased estimation

Model prediction vs data



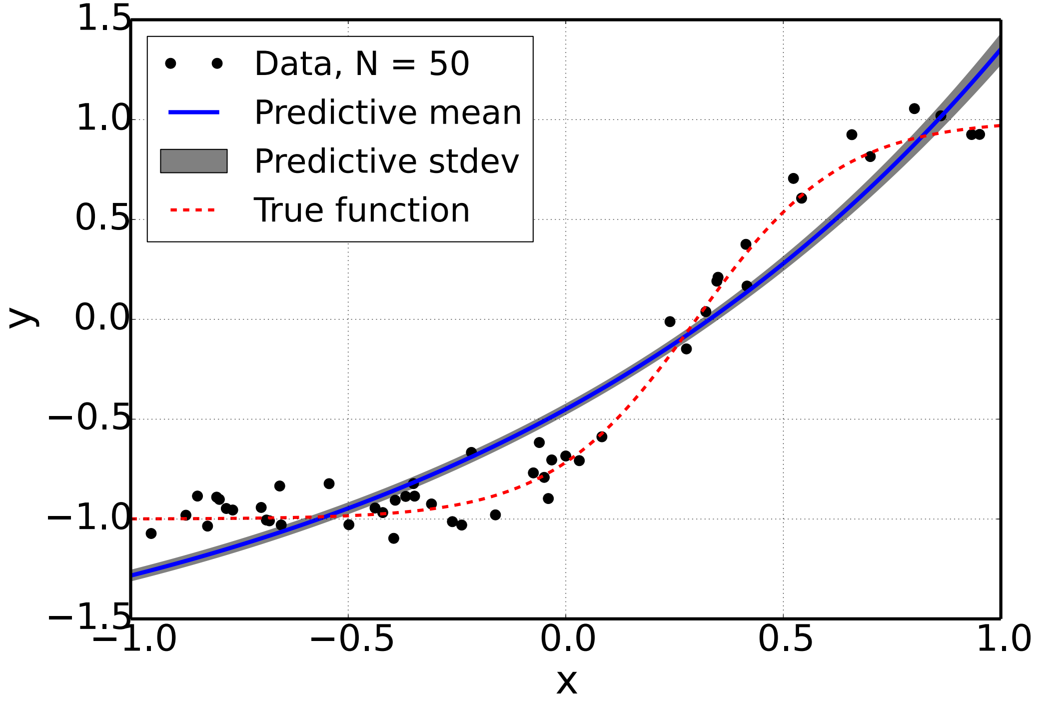
Posterior PDF of **model** parameters



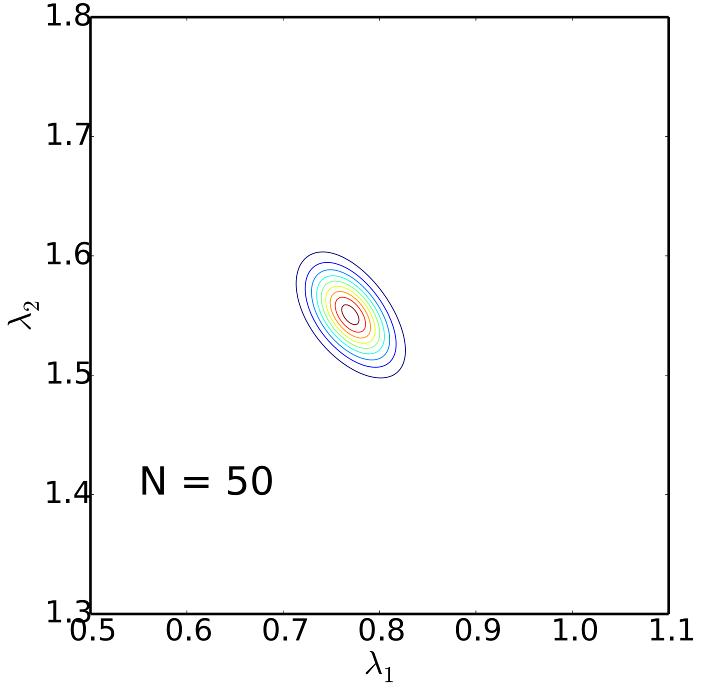
Collecting more data: become increasingly sure about the wrong values of parameters

Wrong model leads to biased estimation

Model prediction vs data

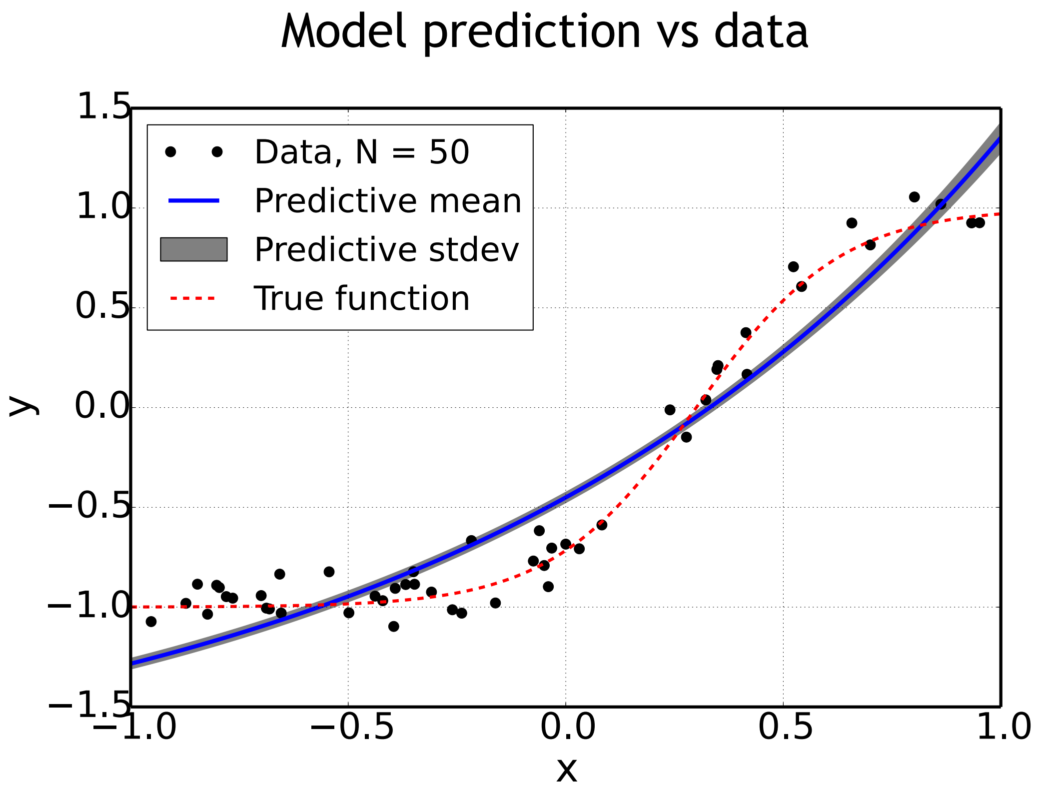


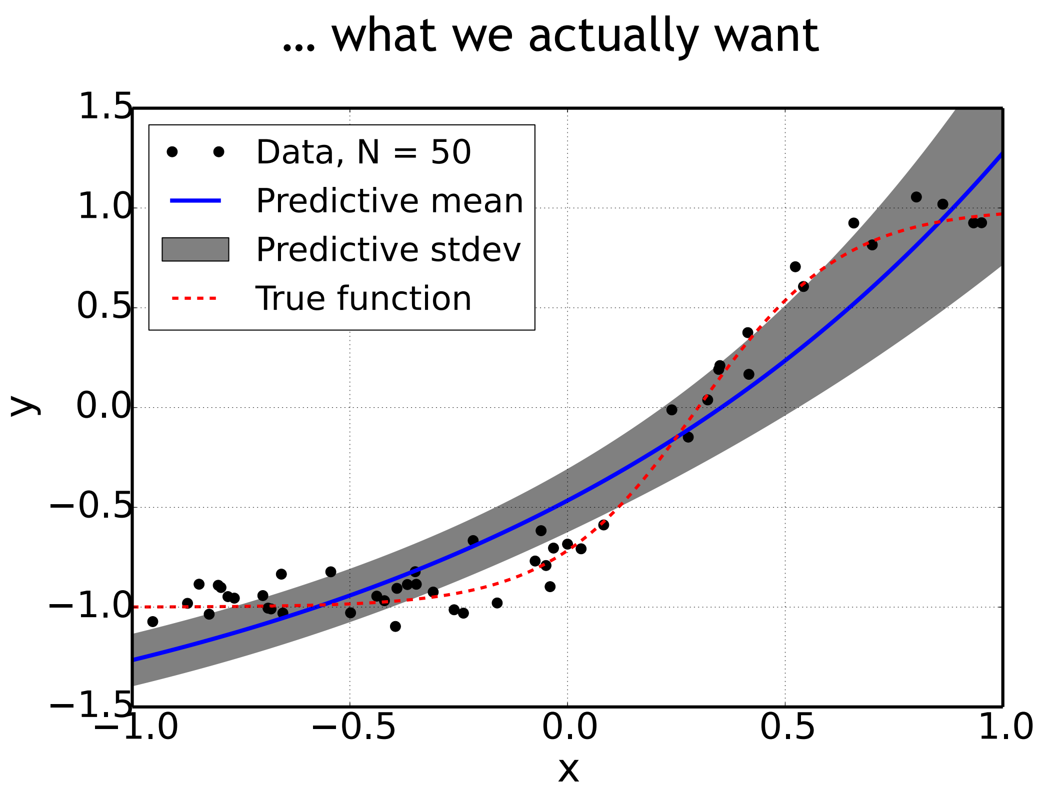
Posterior PDF of exponential parameters



Collecting more data: become increasingly sure about the wrong values of parameters

Wrong model leads to biased estimation





Predictive uncertainties **do not** capture model-data discrepancy

Predictive uncertainties capture model-data discrepancy

Embedded, non-intrusive workflow is implemented in UQTk

Funded by FASTMath SciDAC Institute, www.sandia.gov/uqtoolkit

- Model error correction inside the model, parameterized by ***polynomial chaos (PC)***
- Simultaneous ***Bayesian inference*** of physical parameters and model error parameters
- Calibrated uncertain prediction with model error component

PC

$$y(x_i) = f(x_i; \lambda + \delta(x_i)) + \epsilon_i$$

Bayesian inference

Applications

- Earth System Land Model (BER)
- Chemical kinetic modeling (BES)
- Plasma surface interactions (FES)
- Turbulence modeling (DARPA)

Method: Sargsyan, Najm, Ghanem, IJCK (2015); Sargsyan, Huan, Najm, IJUQ (2019).

Applications: Huan et. al, AIAA J (2018); Hakim et. al, CTM (2018); Cekmer et. al, IJUQ (2018); Rizzi et. al, CMAME (2019).

Unique capability for model exploration via **embedded** model error representation

Non-intrusive

$$y(x_i) = f(x_i; \lambda + \delta(x_i)) + \epsilon_i$$

- Allows meaningful extrapolation
- Respects physics
- Disambiguates model and data errors
- Predictive uncertainty attribution
  - surrogate errors
  - data noise
  - parametric uncertainty
  - **structural errors**

Intrusive

$$y(x_i) = \tilde{f}(x_i; \lambda, \delta(x_i)) + \epsilon_i$$

- For best impact, always look under the hood
  - currently, we do this for the land model

- Code available via UQTk (www.sandia.gov/uqtoolkit)
- Impacted many programs DOE/DOD/SNL
- Applied outside immediate group
- Provides alternative for external correction approach currently available in DAKOTA

Prediction variance = parametric uncertainty + data noise + model error