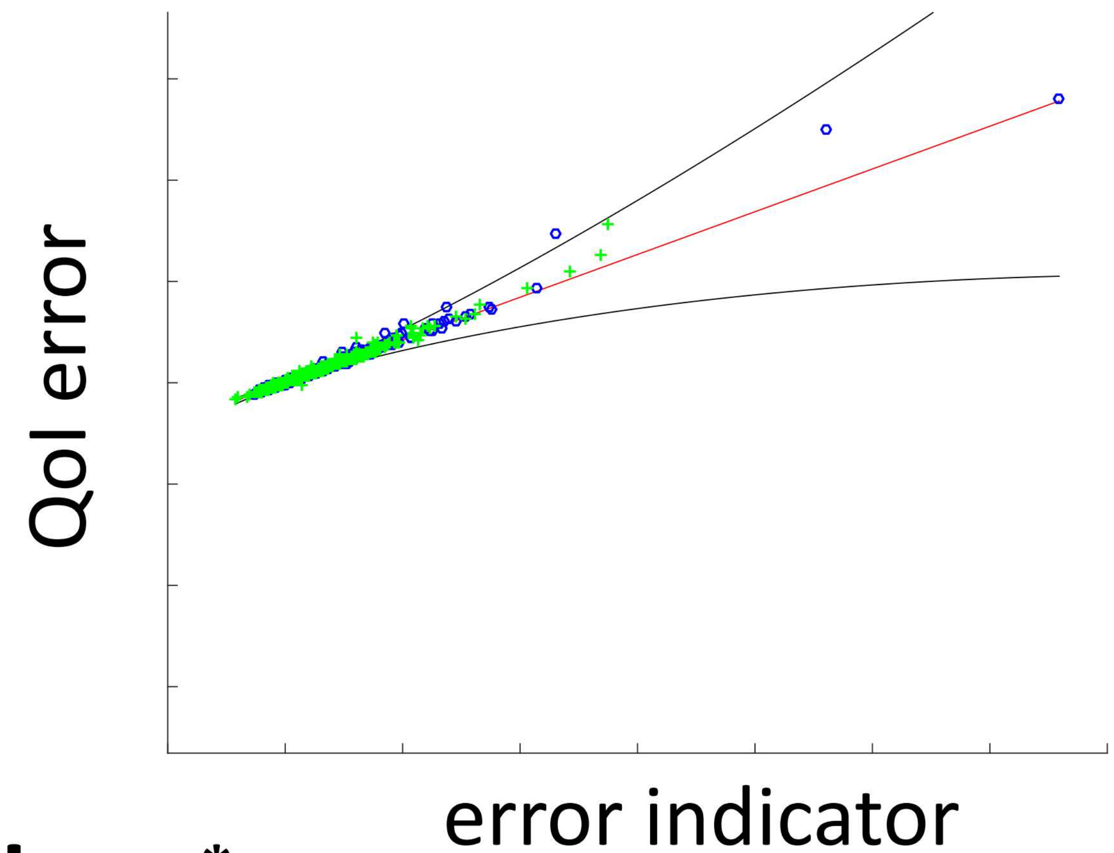
SAND2018-6942C

Machine-learning error models for quantifying the epistemic uncertainty in low-fidelity models

$$\underbrace{\tilde{\mathbf{q}}_{\mathsf{HFM}}(\boldsymbol{\mu})}_{\text{stochastic}} = \underbrace{\mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu})}_{\text{deterministic}} + \underbrace{\tilde{\delta}(\boldsymbol{\mu})}_{\text{stochastic}}$$



Kevin Carlberg[*]

Wayne Uy[^], Fei Lu[†], Matthias Morzfeld[Δ], Brian Freno[*]

[*]Sandia National Laboratories, [^]Cornell University,
[†]Johns Hopkins University, [Δ]University of Arizona

Uncertainty Quantification and Scientific Machine Learning

June 4, 2018

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & 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-NA0003525.

Surrogate modeling in UQ

inputs $\boldsymbol{\mu}$ → *high-fidelity model* → *outputs* $\mathbf{q}_{\mathrm{HFM}}$

- high-fidelity-model (HFM) noise model: $\mathbf{q}_{\mathrm{meas}} = \mathbf{q}_{\mathrm{HFM}}(\boldsymbol{\mu}) + \varepsilon$
- measurement noise ε has probability distribution $\pi_{\varepsilon}(\cdot)$
- HFM likelihood: $\pi_{\mathrm{HFM}}(\mathbf{q}_{\mathrm{meas}} \,|\, \boldsymbol{\mu}) = \pi_{\boldsymbol{\varepsilon}}(\mathbf{q}_{\mathrm{meas}} - \mathbf{q}_{\mathrm{HFM}}(\boldsymbol{\mu}))$

inputs $\boldsymbol{\mu}$ → *surrogate model* → *outputs* $\mathbf{q}_{\mathrm{surr}}$

- surrogate noise model: $\mathbf{q}_{\mathrm{meas}} = \mathbf{q}_{\mathrm{surr}}(\boldsymbol{\mu}) + \varepsilon$
  - inconsistent with HFM noise model
- surrogate likelihood: $\pi_{\mathrm{surr}}(\mathbf{q}_{\mathrm{meas}} \,|\, \boldsymbol{\mu}) = \pi_{\boldsymbol{\varepsilon}}(\mathbf{q}_{\mathrm{meas}} - \mathbf{q}_{\mathrm{surr}}(\boldsymbol{\mu}))$
  - inconsistent with HFM noise model

Surrogate modeling in UQ

$$\mathbf{q}_{\mathsf{HFM}}(\boldsymbol{\mu}) = \mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu}) + \boldsymbol{\delta}(\boldsymbol{\mu})$$

- HFM noise model: $\mathbf{q}_{\mathsf{meas}} = \mathbf{q}_{\mathsf{HFM}}(\boldsymbol{\mu}) + \boldsymbol{\varepsilon}$
  $= \mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu}) + \boldsymbol{\delta}(\boldsymbol{\mu}) + \boldsymbol{\varepsilon}$
- HFM likelihood: $\pi_{\mathsf{HFM}}(\mathbf{q}_{\mathsf{meas}} \mid \boldsymbol{\mu}) = \pi_{\varepsilon}(\mathbf{q}_{\mathsf{meas}} - \mathbf{q}_{\mathsf{HFM}}(\boldsymbol{\mu}))$
  $= \pi_{\varepsilon}(\mathbf{q}_{\mathsf{meas}} - \mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu}) - \boldsymbol{\delta}(\boldsymbol{\mu}))$

+ equivalent to HFM formulation

− not practical: the deterministic error $\boldsymbol{\delta}(\boldsymbol{\mu})$ is generally unknown

How can we account for the error $\boldsymbol{\delta}(\boldsymbol{\mu})$ in a manner that is consistent and practical?

Surrogate modeling in UQ: existing work

Simply replace high-fidelity with surrogate
[Xiu and Karniadakis, 2002; Marzouk and Najm, 2009; Frangos et al., 2010; Nguyen et al., 2010; Li, Marzouk, 2014; Cui et al., 2015]

+ straightforward, can optimize surrogate
- inconsistent with HFM noise model

Multifidelity model management [Peherstorfer, Willcox, Gunzberger, 2018]

- Forward UQ [Giles, 2008; Ng and Willcox, 2014; Narayan et al., 2014; Teckentrup et al., 2015; Peherstorfer et al., 2016]
- Inverse UQ [Christen and Fox, 2005; Efendiev et al., 2006; Cotter et al., 2013; Cui et al., 2013]

+ guaranteed convergence to high-fidelity UQ analysis
- expensive: many queries of high-fidelity model

Stochastic model of the surrogate error

- Surrogate itself is stochastic [Bilionis, Zabaras, 2012; Moustapha et al., 2016]
- Construct a stochastic error model
  [Ng and Eldred, 2012; Drohmann and C., 2015; Manzoni, Pagani, Lassila, 2016; Trehan, C., Durlofsky, 2017; Lu et al., 2017]

+ cheap: no queries of high-fidelity model
+ consistent with HFM noise model (quantify epistemic uncertainty)
- does not converge to high-fidelity UQ analysis

Surrogate modeling in UQ

$$\mathbf{q}_{\mathsf{HFM}}(\boldsymbol{\mu}) = \mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu}) + \boldsymbol{\delta}(\boldsymbol{\mu})$$

Idea: *stochastic model* $\tilde{\delta}(\boldsymbol{\mu})$ *for the error that models its uncertainty*

$$\underbrace{\tilde{\mathbf{q}}_{\mathsf{HFM}}(\boldsymbol{\mu})}_{\text{stochastic}} = \underbrace{\mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu})}_{\text{deterministic}} + \underbrace{\tilde{\boldsymbol{\delta}}(\boldsymbol{\mu})}_{\text{stochastic}}$$

- stochastic HFM noise model: $\mathbf{q}_{\mathsf{meas}} = \tilde{\mathbf{q}}_{\mathsf{HFM}}(\boldsymbol{\mu}) + \boldsymbol{\varepsilon}$
  $= \mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu}) + \tilde{\boldsymbol{\delta}}(\boldsymbol{\mu}) + \boldsymbol{\varepsilon}$
- stochastic HFM likelihood: $\pi_{\widetilde{\mathsf{HFM}}}(\mathbf{q}_{\mathsf{meas}} \mid \boldsymbol{\mu}) = \pi_{\boldsymbol{\varepsilon}+\tilde{\boldsymbol{\delta}}}(\mathbf{q}_{\mathsf{meas}} - \mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu}))$

\+ consistent with HFM noise model

\+ practical if the stochastic error model $\tilde{\delta}$ is computable

Questions

- What properties do we want in a stochastic error model?
- How does a stochastic error model affect Bayesian inference?
- What is the optimal stochastic error model?
- How can we construct a stochastic error model for reduced-order models?

Questions

- **What properties do we want in a stochastic error model?**
- How does a stochastic error model affect Bayesian inference?
- What is the optimal stochastic error model?
- How can we construct a stochastic error model for reduced-order models?

Stochastic error model

$$\underbrace{\tilde{\mathbf{q}}_{\mathsf{HFM}}(\boldsymbol{\mu})}_{\text{stochastic}} = \underbrace{\mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu})}_{\text{deterministic}} + \underbrace{\tilde{\delta}(\boldsymbol{\mu})}_{\text{stochastic}}$$

- HFM likelihood: $\pi_{\mathsf{HFM}}(\mathbf{q}_{\mathsf{meas}} \mid \boldsymbol{\mu}) = \pi_{\boldsymbol{\varepsilon}}(\mathbf{q}_{\mathsf{meas}} - \mathbf{q}_{\mathsf{HFM}}(\boldsymbol{\mu}))$
- stochastic HFM likelihood: $\pi_{\widetilde{\mathsf{HFM}}}(\mathbf{q}_{\mathsf{meas}} \mid \boldsymbol{\mu}) = \pi_{\boldsymbol{\varepsilon}+\tilde{\delta}}(\mathbf{q}_{\mathsf{meas}} - \mathbf{q}_{\mathsf{surr}}(\boldsymbol{\mu}))$

- Desired properties in stochastic error model $\tilde{\delta}(\boldsymbol{\mu})$
  1. cheaply computable: similar cost to evaluating the surrogate
  2. low variance: introduces as little epistemic uncertainty as possible

     $$\lim_{\mathrm{Var}(\tilde{\delta})\to 0} \pi_{\widetilde{\mathsf{HFM}}}(\mathbf{q}_{\mathsf{meas}} \mid \boldsymbol{\mu}) = \pi_{\mathsf{HFM}}(\mathbf{q}_{\mathsf{meas}} \mid \boldsymbol{\mu})$$

  3. generalizable: correctly models epistemic uncertainty in the error

Questions

- What properties do we want in a stochastic error model?
- **How does a stochastic error model affect Bayesian inference?**
- What is the optimal stochastic error model?
- How can we construct a stochastic error model for reduced-order models?

Stochastic error model

$$\underbrace{\tilde{\mathbf{q}}_{\text{HFM}}(\boldsymbol{\mu})}_{\text{stochastic}} = \underbrace{\mathbf{q}_{\text{surr}}(\boldsymbol{\mu})}_{\text{deterministic}} + \underbrace{\tilde{\delta}(\boldsymbol{\mu})}_{\text{stochastic}}$$

- Forward UQ: $\pi_{\text{HFM}}(\mathbf{q}_{\text{meas}} \mid \boldsymbol{\mu}) = \pi_{\varepsilon}(\mathbf{q}_{\text{meas}} - \mathbf{q}_{\text{HFM}}(\boldsymbol{\mu}))$

  $\pi_{\widetilde{\text{HFM}}}(\mathbf{q}_{\text{meas}} \mid \boldsymbol{\mu}) = \pi_{\varepsilon+\tilde{\delta}}(\mathbf{q}_{\text{meas}} - \mathbf{q}_{\text{surr}}(\boldsymbol{\mu}))$
- Inverse UQ: $\pi_{\text{post}}^{\text{HFM}}(\boldsymbol{\mu} \mid \mathbf{q}_{\text{meas}}) \propto \pi_{\text{prior}}(\boldsymbol{\mu})\pi_{\varepsilon}(\mathbf{q}_{\text{meas}} - \mathbf{q}_{\text{HFM}}(\boldsymbol{\mu}))$

  $\pi_{\text{post}}^{\widetilde{\text{HFM}}}(\boldsymbol{\mu} \mid \mathbf{q}_{\text{meas}}) \propto \pi_{\text{prior}}(\boldsymbol{\mu})\pi_{\varepsilon+\tilde{\delta}}(\mathbf{q}_{\text{meas}} - \mathbf{q}_{\text{surr}}(\boldsymbol{\mu}))$

Homoscedastic $\tilde{\delta}(\boldsymbol{\mu})$

- **Low-variance error model**: $\text{Var}(\tilde{\delta}(\boldsymbol{\mu}) + \varepsilon) \approx \text{Var}(\varepsilon)$
  - *Forward UQ:* measurements certain: $D_{\text{KL}}(\pi_{\text{HFM}} \,\|\, \pi_{\widetilde{\text{HFM}}})$ small
  - *Inverse UQ:* data informative: $D_{\text{KL}}(\pi_{\text{post}}^{\text{HFM}} \,\|\, \pi_{\text{post}}^{\widetilde{\text{HFM}}})$ small
- **High-variance error model:** $\text{Var}(\tilde{\delta}(\boldsymbol{\mu}) + \varepsilon) \gg \text{Var}(\varepsilon)$
  - *Forward UQ*: measurements uncertain: $D_{\text{KL}}(\pi_{\text{HFM}} \,\|\, \pi_{\widetilde{\text{HFM}}})$ large
  - *Inverse UQ*: data uninformative: $D_{\text{KL}}(\pi_{\text{prior}} \,\|\, \pi_{\text{post}}^{\widetilde{\text{HFM}}})$ small

Heteroscedastic $\tilde{\delta}(\boldsymbol{\mu})$

- measurements certain, data informative for $\boldsymbol{\mu}$ where surrogate trusted

Questions

- What properties do we want in a stochastic error model?
- How does a stochastic error model affect Bayesian inference?
- **What is the optimal stochastic error model?**
- How can we construct a stochastic error model for reduced-order models?

Case 1: Gaussian, homoscedastic error model

Proposition [C., Uy, Lu, Morzfeld, 2018]

If the following conditions hold:

1. $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$,
2. Error model is Gaussian and homoscedastic
$$\tilde{\boldsymbol{\delta}}(\boldsymbol{\mu}) \sim \mathcal{N}(\tilde{\mathbf{m}}(\boldsymbol{\mu}), \tilde{\sigma}^2 \mathbf{I}), \quad \forall \boldsymbol{\mu} \in \mathcal{D}$$
3. Error model is unbiased $\mathrm{E}[\tilde{\mathbf{m}}(\boldsymbol{\mu})] = \boldsymbol{\delta}(\boldsymbol{\mu}),\ \forall \boldsymbol{\mu} \in \mathcal{D}$,

then

$$\max_{\tilde{\sigma}^2} \mathrm{E}_{\text{prior}}[D_{\mathrm{KL}}(\pi_{\mathrm{HFM}} \,\|\, \pi_{\text{surr}}) - D_{\mathrm{KL}}(\pi_{\mathrm{HFM}} \,\|\, \pi_{\widetilde{\mathrm{HFM}}})]$$

is positive and is attained at

$$\tilde{\sigma}_\star^2 = \frac{1}{n_s} \mathrm{E}_{\text{prior}}[\|\mathbf{q}_{\mathrm{HFM}}(\boldsymbol{\mu}) - \mathbf{q}_{\text{surr}}(\boldsymbol{\mu}) - \tilde{\mathbf{m}}(\boldsymbol{\mu})\|^2]$$

\+ Suggests error-model noise should be computed as sample variance

$$\tilde{\sigma}^2 = \frac{1}{n_s} \mathrm{E}_{\text{data}}[\|\mathbf{q}_{\mathrm{HFM}}(\boldsymbol{\mu}) - \mathbf{q}_{\text{surr}}(\boldsymbol{\mu}) - \tilde{\mathbf{m}}(\boldsymbol{\mu})\|^2]$$

− Homoscedasticity may violate property 3

Case 2: Gaussian, heteroscedastic error model

Proposition [C., Uy, Lu, Morzfeld, 2018]

Let $\{\mathcal{D}_i\}_{i=1}^{M}$ be a non-overlapping partition of the parameter space $\mathcal{D}$.
If the following conditions hold:

1. $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$,
2. Error model is Gaussian and heteroscedastic such that
$$\tilde{\boldsymbol{\delta}}(\boldsymbol{\mu}) \sim \mathcal{N}(\tilde{\mathbf{m}}(\boldsymbol{\mu}), \tilde{\sigma}_i{}^2 \mathbf{I}), \quad \forall \boldsymbol{\mu} \in \mathcal{D}_i$$
3. Error model is unbiased $\mathrm{E}[\tilde{\mathbf{m}}(\boldsymbol{\mu})] = \boldsymbol{\delta}(\boldsymbol{\mu}), \ \forall \boldsymbol{\mu} \in \mathcal{D}$,

then
$$\max_{\tilde{\sigma}_i^2} \mathrm{E}_{\mathrm{prior}}[(D_{\mathrm{KL}}(\pi_{\mathrm{HFM}} \,\|\, \pi_{\mathrm{surr}}) - D_{\mathrm{KL}}(\pi_{\mathrm{HFM}} \,\|\, \pi_{\widetilde{\mathrm{HFM}}}))\, \mathbf{1}_{\mathcal{D}_i}(\boldsymbol{\mu})]$$
is positive and is attained at
$$\tilde{\sigma}_{i,\star}^2 = \frac{1}{n_s} \mathrm{E}_{\mathrm{prior}}[\|\mathbf{q}_{\mathrm{HFM}}(\boldsymbol{\mu}) - \mathbf{q}_{\mathrm{surr}}(\boldsymbol{\mu}) - \tilde{\mathbf{m}}(\boldsymbol{\mu})\|^2 \mid \mathcal{D}_i]$$

+ Error-model noise should be computed as sample variance over regions
$$\tilde{\sigma}_i^2 = \frac{1}{n_s} \mathrm{E}_{\mathrm{data}}[\|\mathbf{q}_{\mathrm{HFM}}(\boldsymbol{\mu}) - \mathbf{q}_{\mathrm{surr}}(\boldsymbol{\mu}) - \tilde{\mathbf{m}}(\boldsymbol{\mu})\|^2 \mid \mathcal{D}_i]$$
+ Heteroscedasticity can satisfy property 3

Questions

- What properties do we want in a stochastic error model?
- How does a stochastic error model affect Bayesian inference?
- What is the optimal stochastic error model?
- **How can we construct a stochastic error model for reduced-order models?**

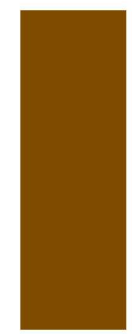
Reduced-order modeling

High-fidelity model

$$\mathbf{r}(\mathbf{x};\boldsymbol{\mu}) = 0, \quad \mathbf{q}_{\mathrm{HFM}}(\boldsymbol{\mu}) = \mathbf{q}(\mathbf{x}(\boldsymbol{\mu});\boldsymbol{\mu})$$

Reduced-order model

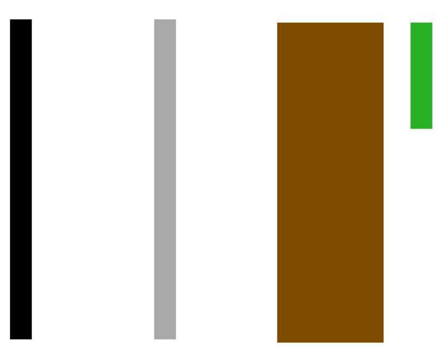
- **Offline:** construct low-dimensional basis $\boldsymbol{\Phi}$



- **Online:** reduce high-fidelity-model dimension

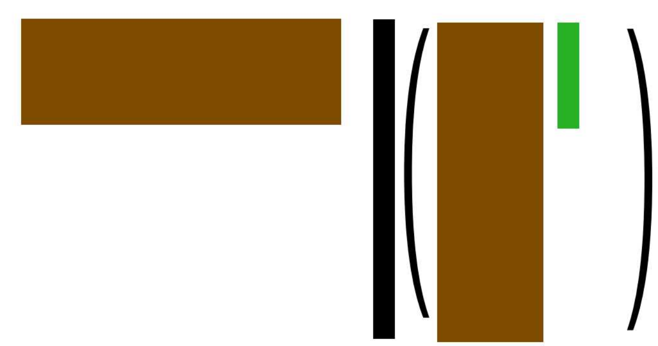
1. Reduce number of unknowns

$$\mathbf{x} \approx \tilde{\mathbf{x}} = \boldsymbol{\Phi}\hat{\mathbf{x}}$$



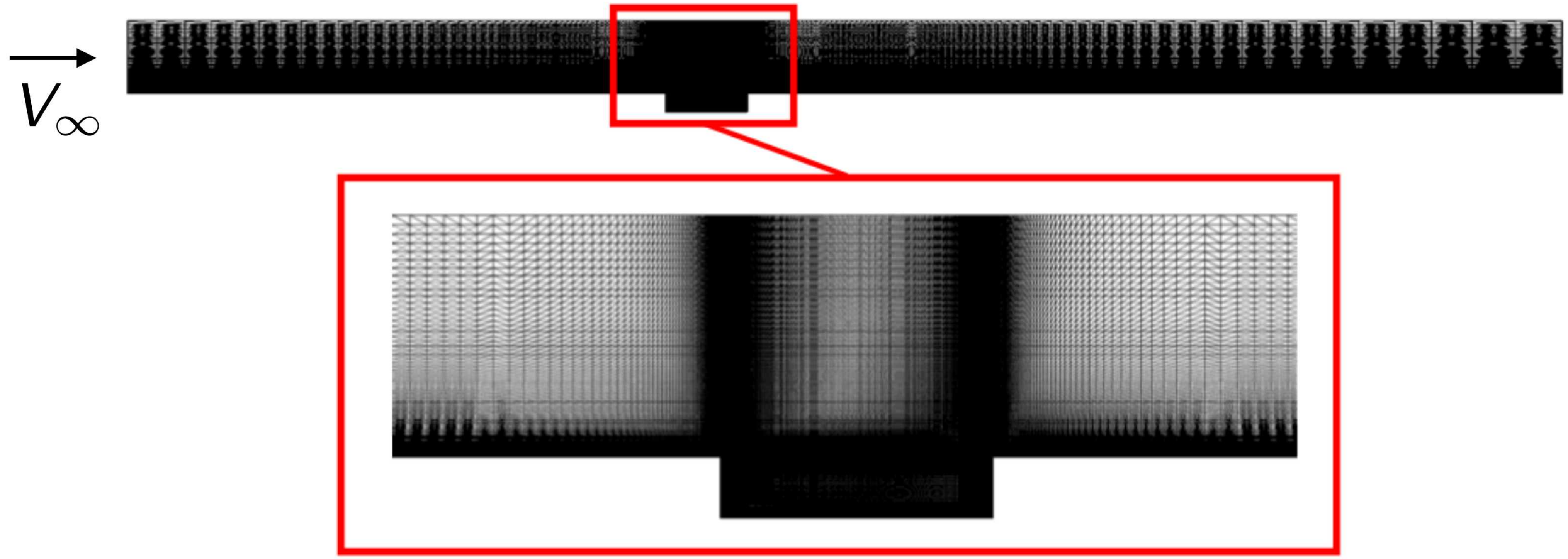
2. Reduce number of equations

$$\boldsymbol{\Phi}^T \mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu}) = \mathbf{0}$$



$$\boldsymbol{\Phi}^T \mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu}) = \mathbf{0}, \quad \mathbf{q}_{\mathrm{surr}}(\boldsymbol{\mu}) = \mathbf{q}(\boldsymbol{\Phi}\hat{\mathbf{x}}(\boldsymbol{\mu});\boldsymbol{\mu})$$

ROM demonstration: turbulent cavity



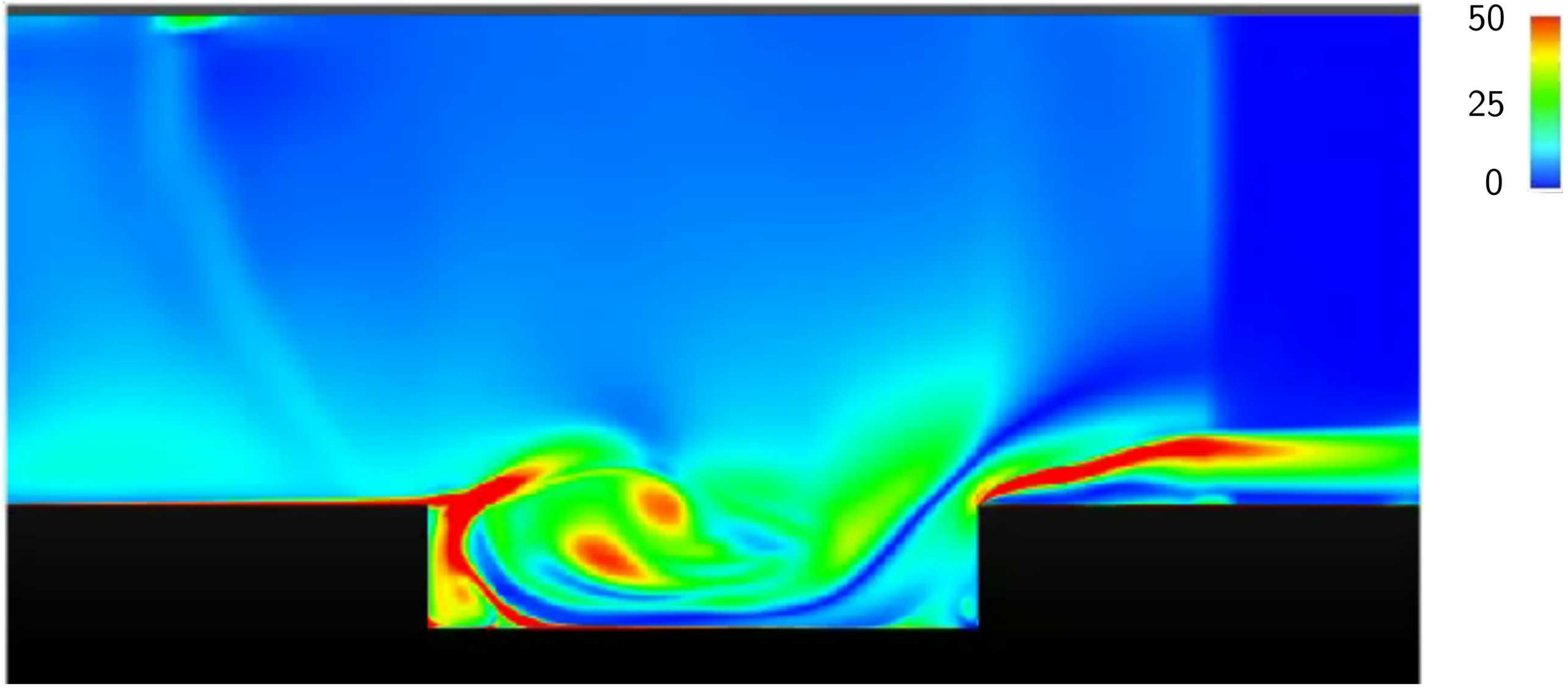
- Unsteady Navier–Stokes
- Re = 6.3 x 10^6
- $M_\infty = 0.6$

Spatial discretization
- 2nd-order finite volume
- DES turbulence model
- 1.2×10^6 degrees of freedom

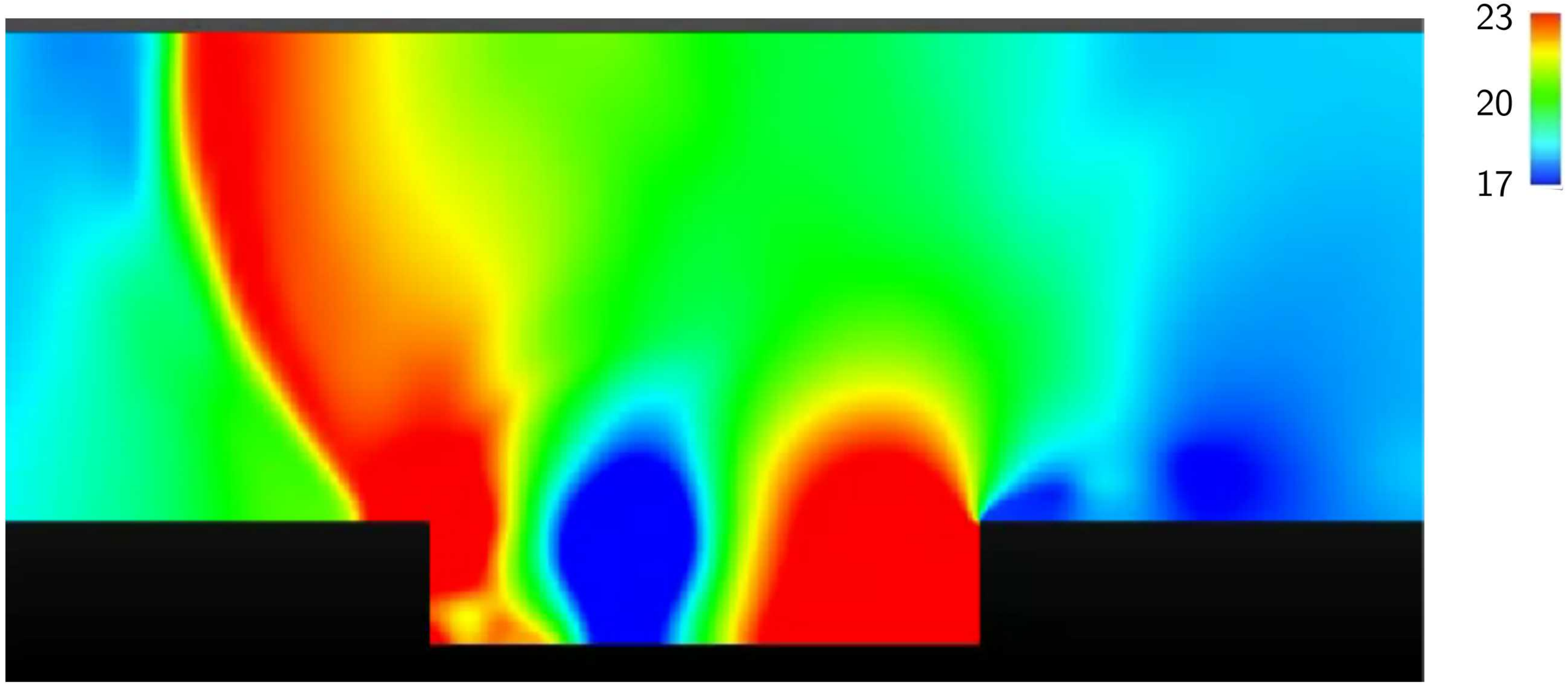
Temporal discretization
- 2nd-order BDF
- Verified time step $\Delta t = 1.5 \times 10^{-3}$
- 8.3×10^3 time instances

High-fidelity model solution

vorticity field

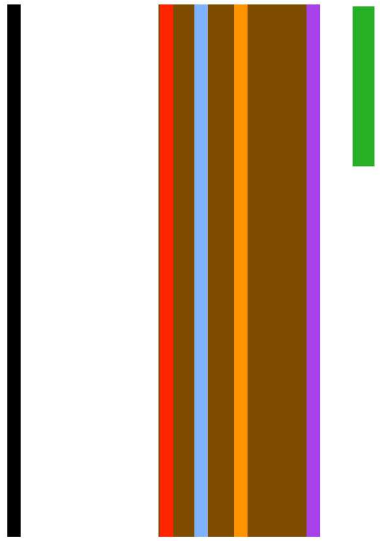


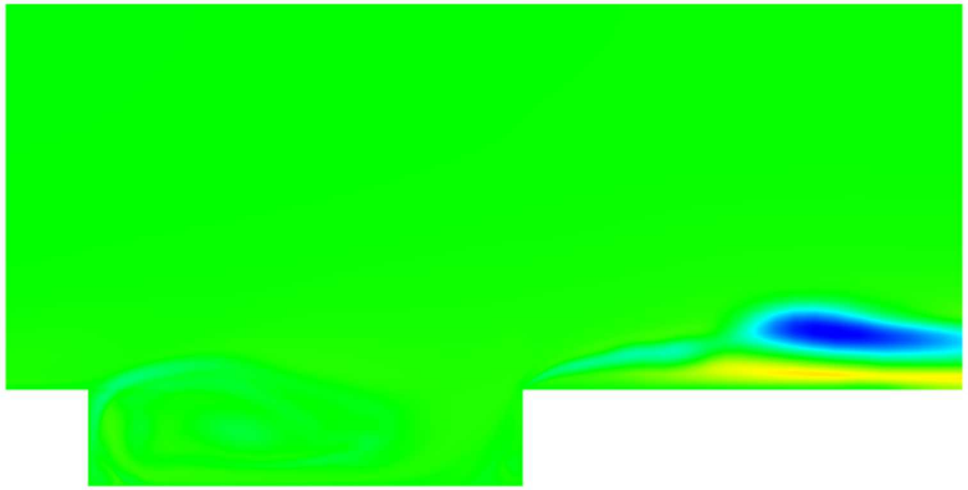
pressure field



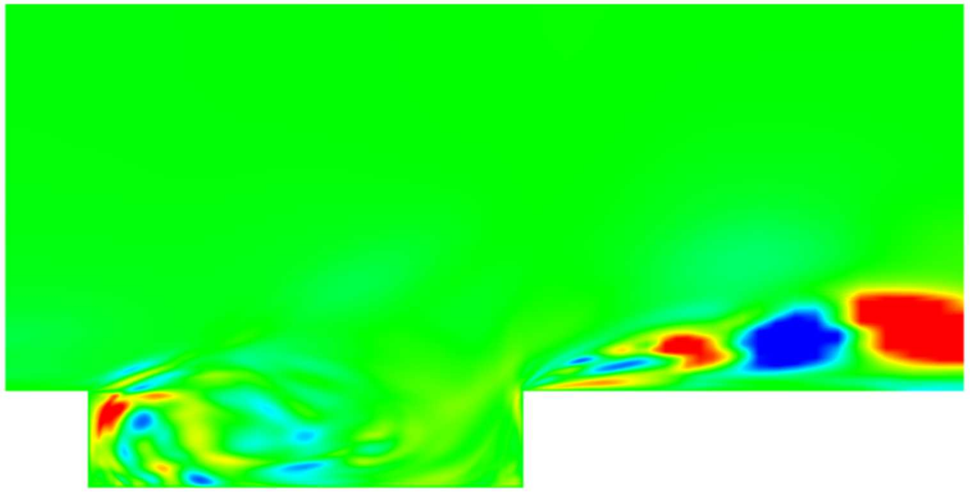
Principal components

$\mathbf{x}(t) \approx \boldsymbol{\Phi}\ \hat{\mathbf{x}}(t)$

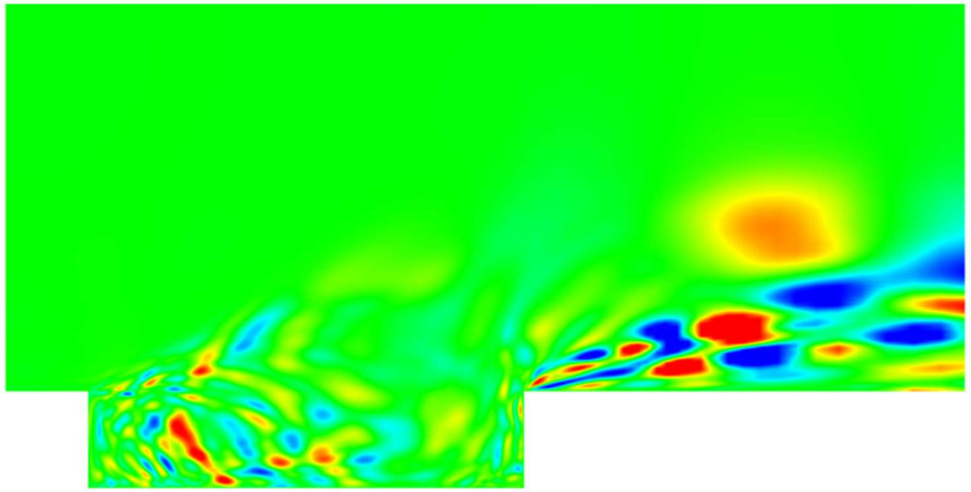




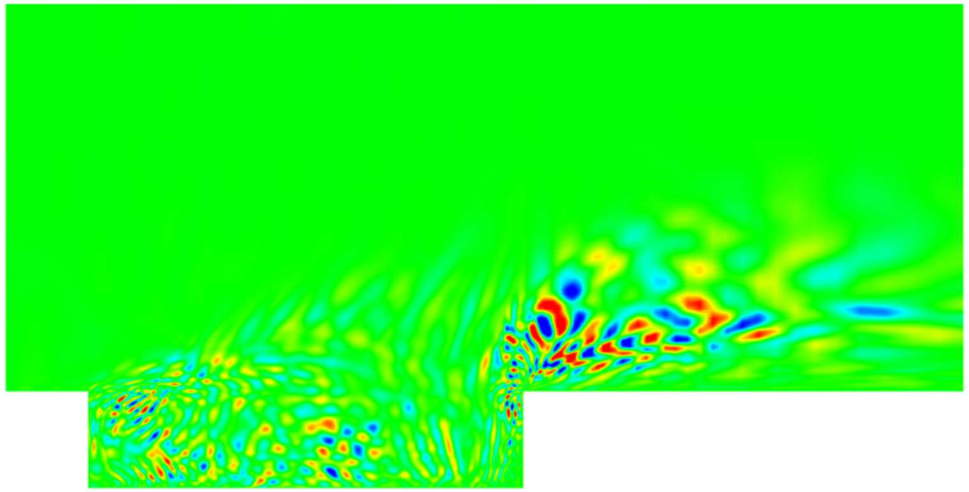
ϕ_1



ϕ_{21}



ϕ_{101}



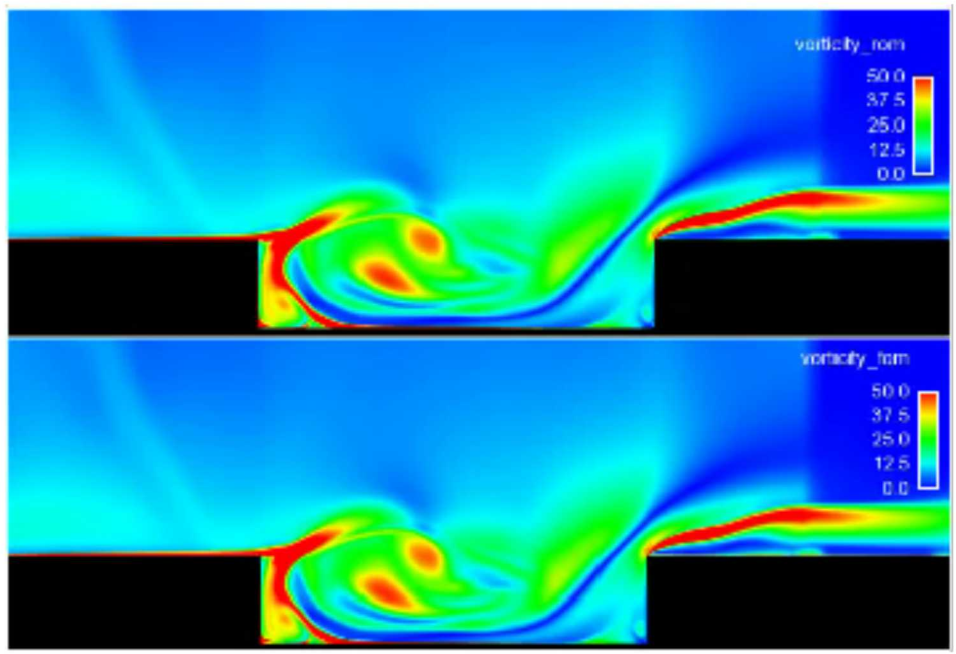
ϕ_{401}

Turbulent-cavity results [C., Barone, Antil, 2017]

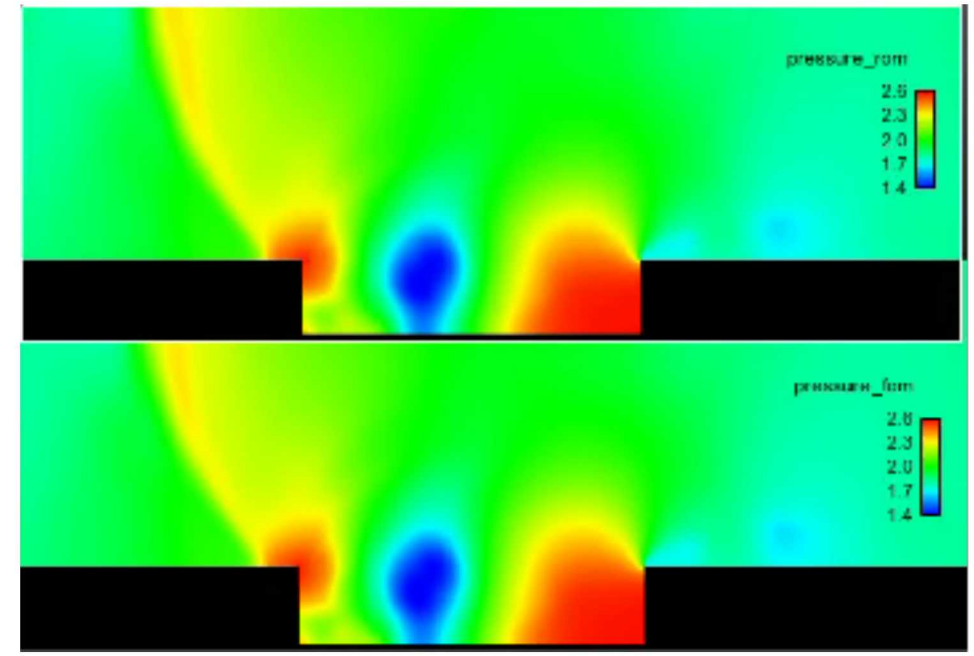
LSPG ROM
dim: 179
32 min, 2 cores

high-fidelity
dim: 1.2M
5 hours, 48 cores

vorticity field



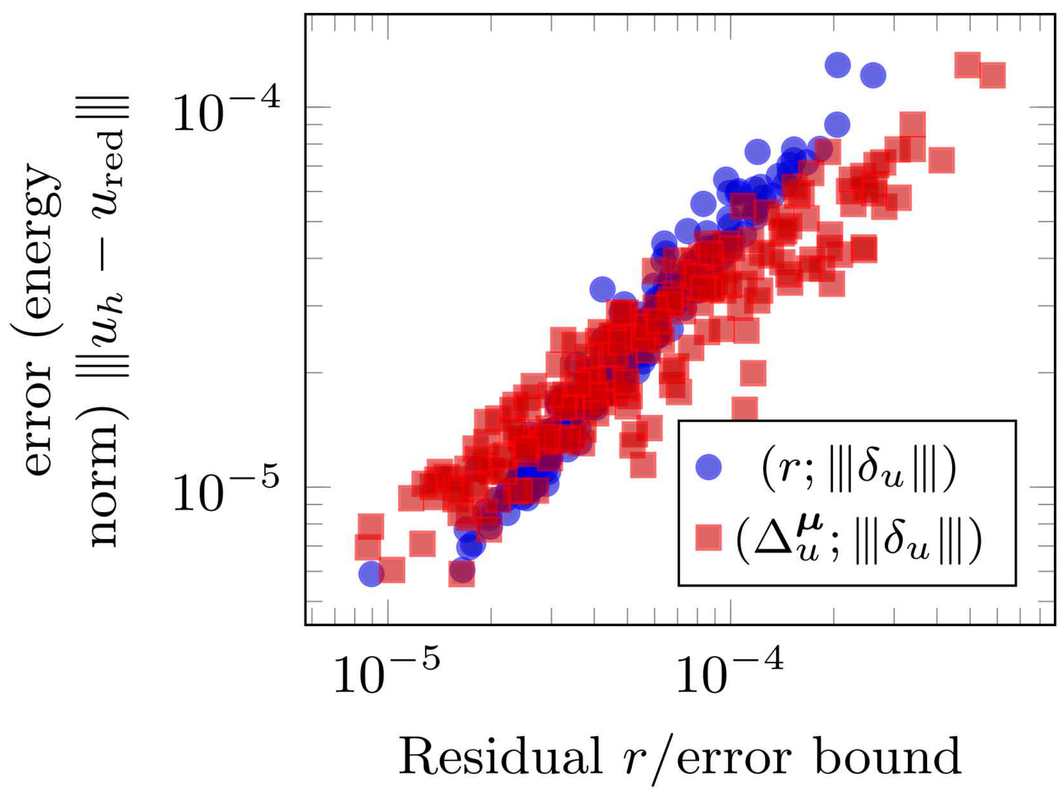
pressure field

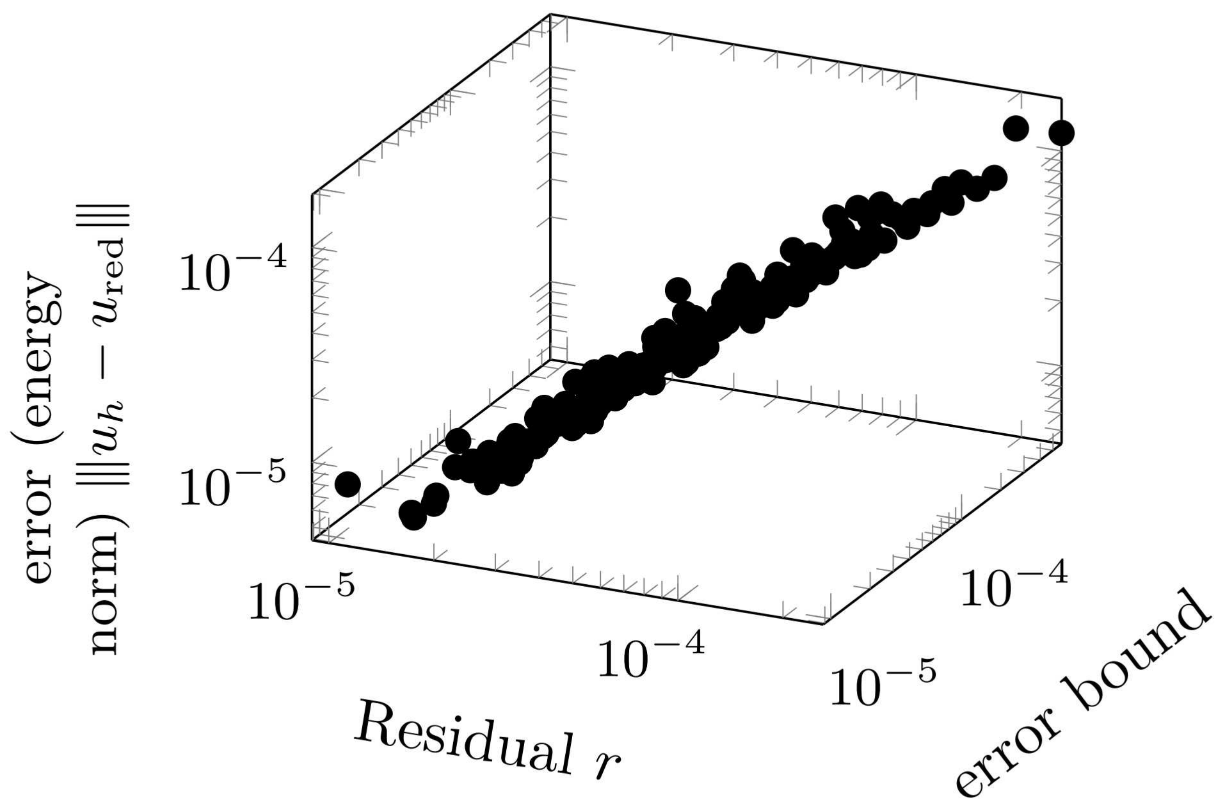


+ *<1% error*
+ *229X computational-cost reduction*

How can we construct an error model for the ROM?

Key insight



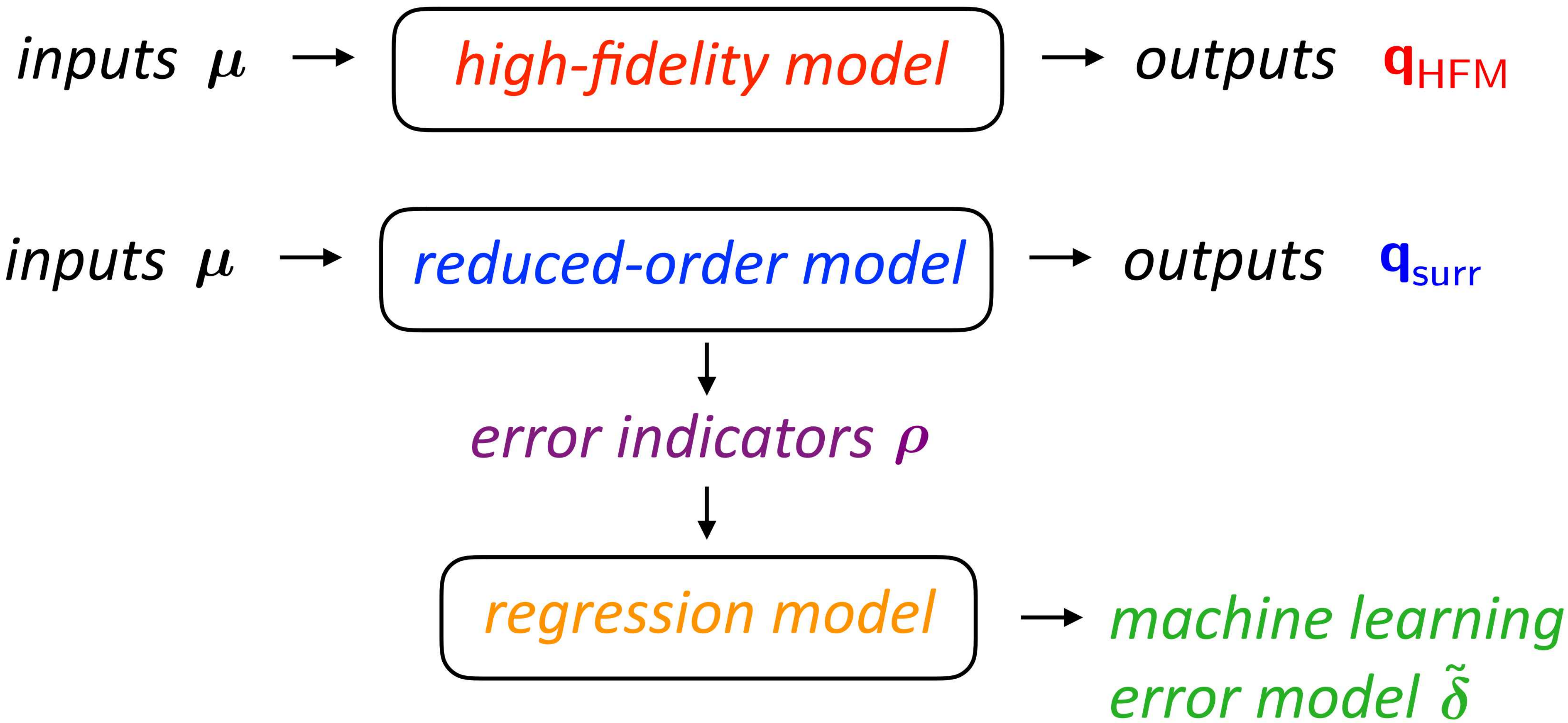


ROMs generate physics-based error indicators that are highly informative of the ROM error

Idea: *Apply machine learning regression to generate a mapping from error indicators to a distribution over the ROM error*

+ Can generate lower-variance error models than simply modeling $\mu \mapsto \delta(\mu)$
[Kennedy, O'Hagan, 2001; Ng, Eldred 2012]

Machine-learning error models



$$\underbrace{\tilde{\mathbf{q}}_{\mathrm{HFM}}(\boldsymbol{\mu})}_{\text{stochastic}} = \underbrace{\mathbf{q}_{\mathrm{surr}}(\boldsymbol{\mu})}_{\text{deterministic}} + \underbrace{\tilde{\delta}(\boldsymbol{\mu})}_{\text{stochastic}}$$

How to determine error indicators ρ and regression model?

Features and regression model

- Desired properties in stochastic error model $\tilde{\delta}(\boldsymbol{\mu})$
  1. cheaply computable, 2. low variance, 3. generalizable
- *Feature engineering*: select error indicators $\boldsymbol{\rho}$ to trade off:
  1. *Number of features*
     - ➔ Large number: costly, lower variance, high capacity regression
     - ➔ Small number: cheaper, higher variance, low capacity regression
  2. *Quality of features*
     - ➔ High quality: expensive, lower variance
     - ➔ Low quality: cheaper, higher variance
- *Regression model*:
  - ➔ High capacity: lower variance, more data to generalize
  - ➔ Low capacity: higher variance, less data to generalize

Method 1: Dual-weighted residual and Gaussian process regression
[Drohmann, C., 2015; Pagani, C., Manzoni, 2018]

Method 2: Large number of features and high-dimensional regression
[Trehan, C., Durlofsky, 2017; Freno, C., 2018]

Features and regression model

- Desired properties in stochastic error model $\tilde{\delta}(\boldsymbol{\mu})$
  1. cheaply computable, 2. low variance, 3. generalizable
- *Feature engineering*: select error indicators $\boldsymbol{\rho}$ to trade off:
  1. *Number of features*
     - Large number: costly, lower variance, high capacity regression
     - ➔ Small number: cheaper, higher variance, low capacity regression
  2. *Quality of features*
     - ➔ High quality: expensive, lower variance
     - Low quality: cheaper, higher variance
- *Regression model*:
  - High capacity: lower variance, more data to generalize
  - ➔ Low capacity: higher variance, less data to generalize

Method 1: Dual-weighted residual and Gaussian process regression
[Drohmann, C., 2015; Pagani, C., Manzoni, 2018]

Method 2: Large number of features and high-dimensional regression
[Trehan, C., Durlofsky, 2017; Freno, C., 2018]

Feature: dual-weighted residual [Drohmann, C., 2015]

- Approximate HFM output to first order

$$q_i(\mathbf{x}) \approx q_i(\boldsymbol{\Phi}\hat{\mathbf{x}}) + \frac{\partial q_i}{\partial \mathbf{x}}(\boldsymbol{\Phi}\hat{\mathbf{x}})(\mathbf{x} - \boldsymbol{\Phi}\hat{\mathbf{x}}) \quad (1)$$

- Approximate HFM residual to first order

$$\mathbf{0} = \mathbf{r}(\mathbf{x}) \approx \mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}}) + \frac{\partial \mathbf{r}}{\partial \mathbf{x}}(\boldsymbol{\Phi}\hat{\mathbf{x}})(\mathbf{x} - \boldsymbol{\Phi}\hat{\mathbf{x}})$$

- Solve for the error

$$\mathbf{x} - \boldsymbol{\Phi}\hat{\mathbf{x}} \approx -\left[\frac{\partial \mathbf{r}}{\partial \mathbf{x}}(\boldsymbol{\Phi}\hat{\mathbf{x}})\right]^{-1} \mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}}) \quad (2)$$

- Substitute (2) in (1)

$$q_i(\mathbf{x}) - q_i(\boldsymbol{\Phi}\hat{\mathbf{x}}) \approx \mathbf{y}_i{}^T \mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}})$$

with the dual solution $\mathbf{y}_i$ satisfying

$$\frac{\partial \mathbf{r}}{\partial \mathbf{x}}(\boldsymbol{\Phi}\hat{\mathbf{x}})^T \mathbf{y}_i = \frac{\partial q_i}{\partial \mathbf{x}}(\boldsymbol{\Phi}\hat{\mathbf{x}})^T$$

Feature: dual-weighted residual [Drohmann, C., 2015]

$$q_i(\mathbf{x}) - q_i(\boldsymbol{\Phi}\hat{\mathbf{x}}) \approx {\mathbf{y}_i}^T \mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}})$$

$$\frac{\partial \mathbf{r}}{\partial \mathbf{x}}(\boldsymbol{\Phi}\hat{\mathbf{x}})^T \mathbf{y}_i = \frac{\partial q_i}{\partial \mathbf{x}}(\boldsymbol{\Phi}\hat{\mathbf{x}})^T$$

- Want to avoid HFM-scale solves, so approximate dual as

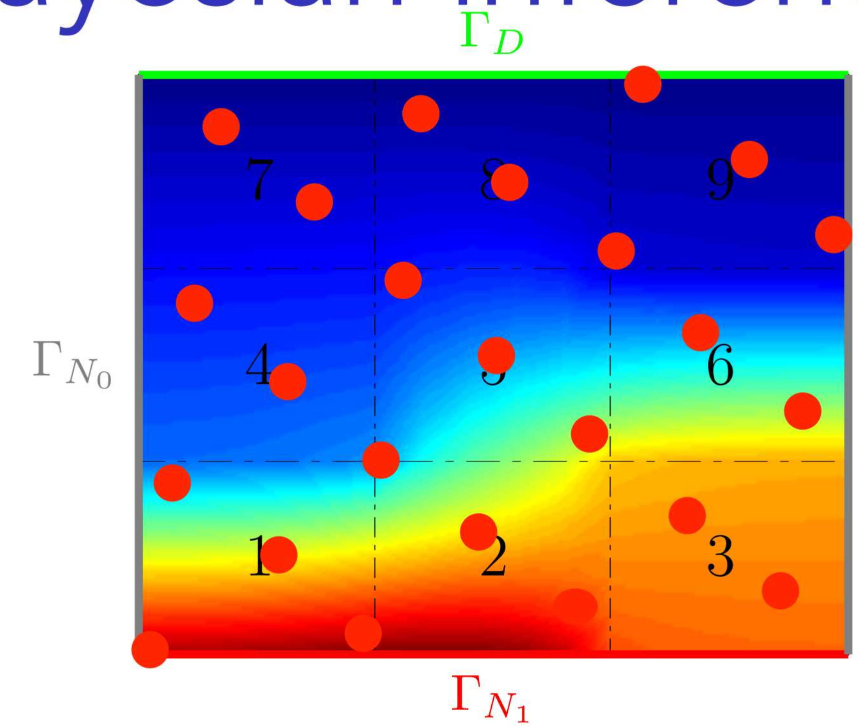
$$\mathbf{y}_i \approx \tilde{\mathbf{y}}_i = \boldsymbol{\Phi}_i \hat{\mathbf{y}}_i$$

and construct ROM for the dual

$${\boldsymbol{\Phi}_i}^T \frac{\partial \mathbf{r}}{\partial \mathbf{x}}(\boldsymbol{\Phi}\hat{\mathbf{x}})^T \boldsymbol{\Phi}_i \hat{\mathbf{y}}_i = {\boldsymbol{\Phi}_i}^T \frac{\partial q_i}{\partial \mathbf{x}}(\boldsymbol{\Phi}\hat{\mathbf{x}})^T$$

- **One feature**: $q_i(\mathbf{x}) - q_i(\boldsymbol{\Phi}\hat{\mathbf{x}}) \approx \rho_i = {\tilde{\mathbf{y}}_i}^T \mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}}) = {\hat{\mathbf{y}}_i}^T {\boldsymbol{\Phi}_i}^T \mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}})$
- **Regression model**: Gaussian process [Rasmussen, Williams, 2006]

Application: Bayesian inference



$$\triangle c(x;\boldsymbol{\mu})u(x;\boldsymbol{\mu}) = 0 \text{ in } \Omega \qquad \boldsymbol{x}(\boldsymbol{\mu}) = 0 \text{ on } \Gamma_D$$

$$\nabla c(\boldsymbol{\mu})\boldsymbol{x}(\boldsymbol{\mu}) \cdot n = 0 \text{ on } \Gamma_{N_0} \qquad \nabla c(\boldsymbol{\mu})\boldsymbol{x}(\boldsymbol{\mu}) \cdot n = 1 \text{ on } \Gamma_{N_1}$$

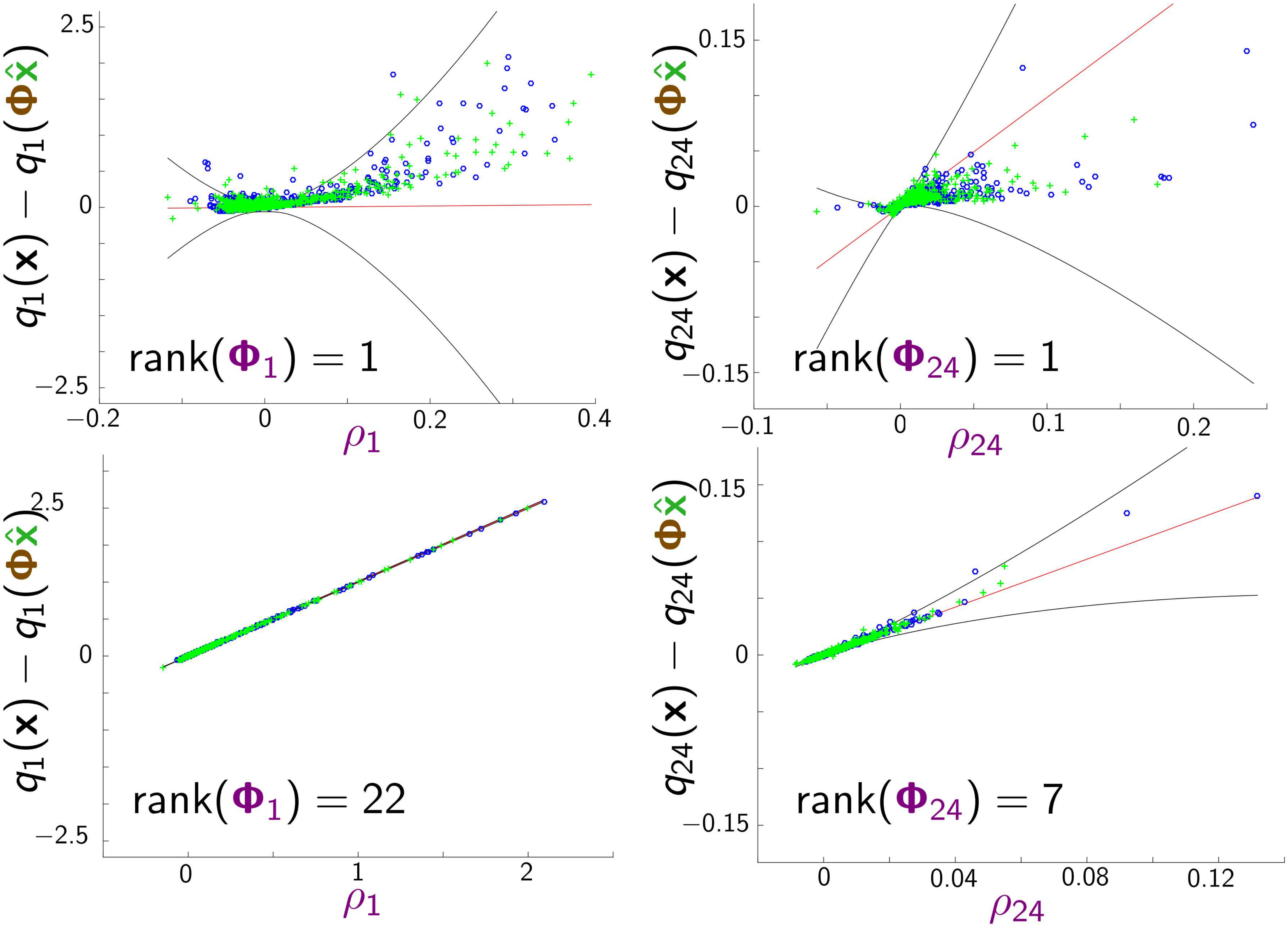
- Inputs $\boldsymbol{\mu} \in [0.1, 10]^9$ define diffusivity in *c* in subdomains
- Outputs $\mathbf{q}$ are 24 measured temperatures
- ROM constructed via RB-Greedy [Patera and Rozza, 2006]
- $\pi_{\text{prior}}(\boldsymbol{\mu})$: Gaussian with variance 0.1
- $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1 \times 10^{-3})$
- Posterior sampling: 1×10^5 samples w/ implicit sampling [Tu et al., 2013]

Machine learning error models

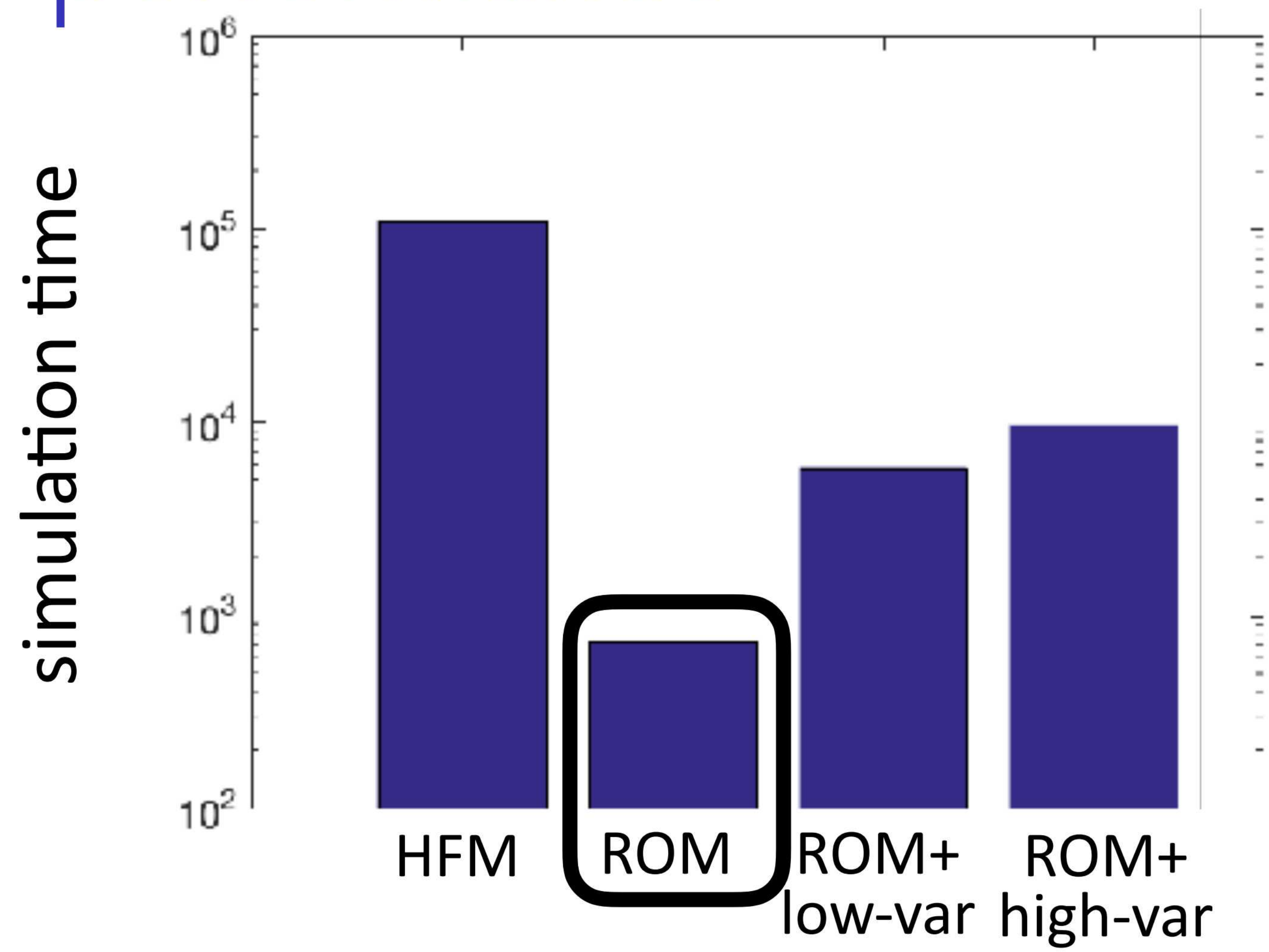
$$\tilde{\delta}_i(\boldsymbol{\mu}) \sim \mathcal{N}(\beta\rho_i(\boldsymbol{\mu}), \alpha_1 + \alpha_2|\rho_i(\boldsymbol{\mu})|^{\alpha_3})$$

low quality
high variance
cheap

high quality
low variance
costly

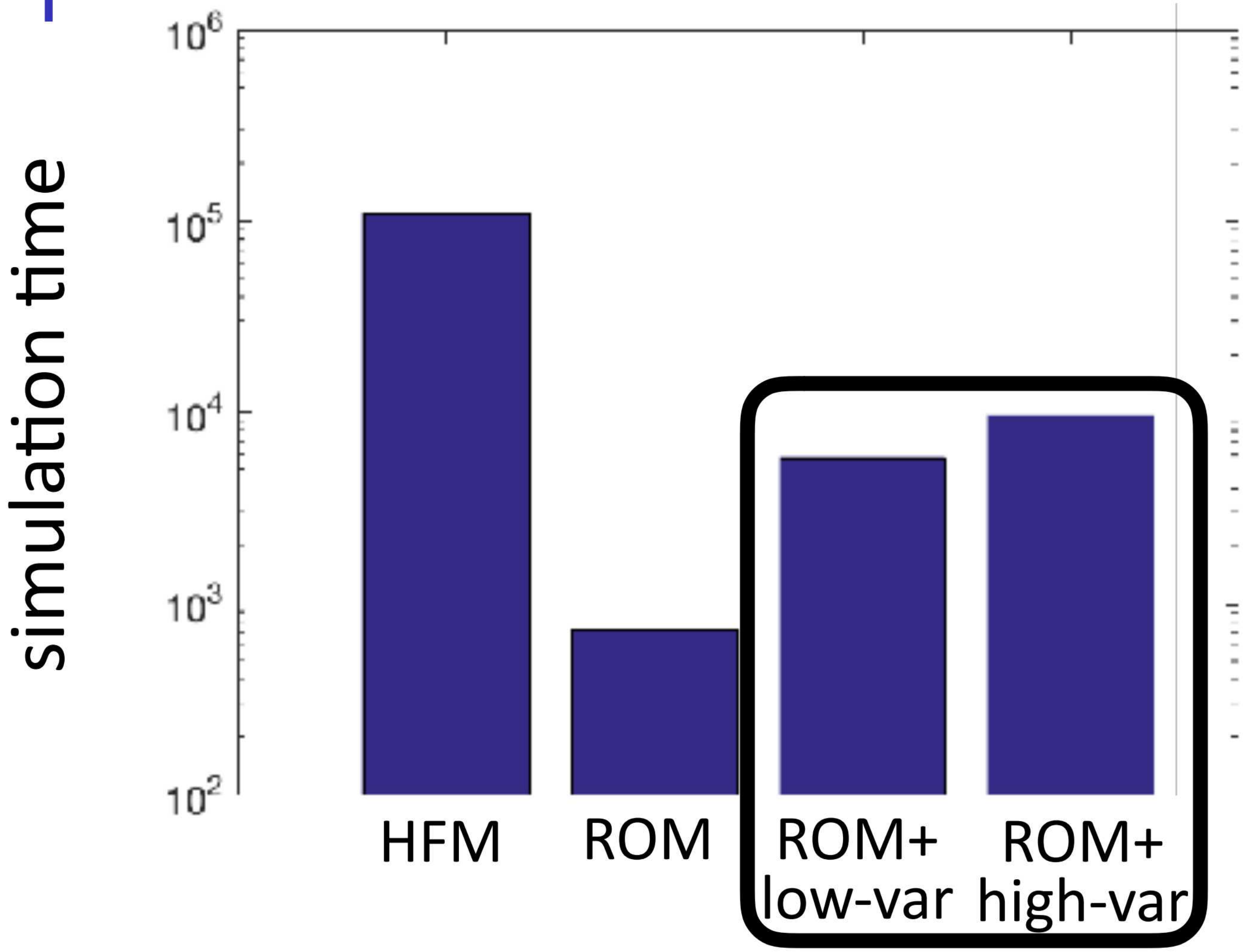


Wall-time performance



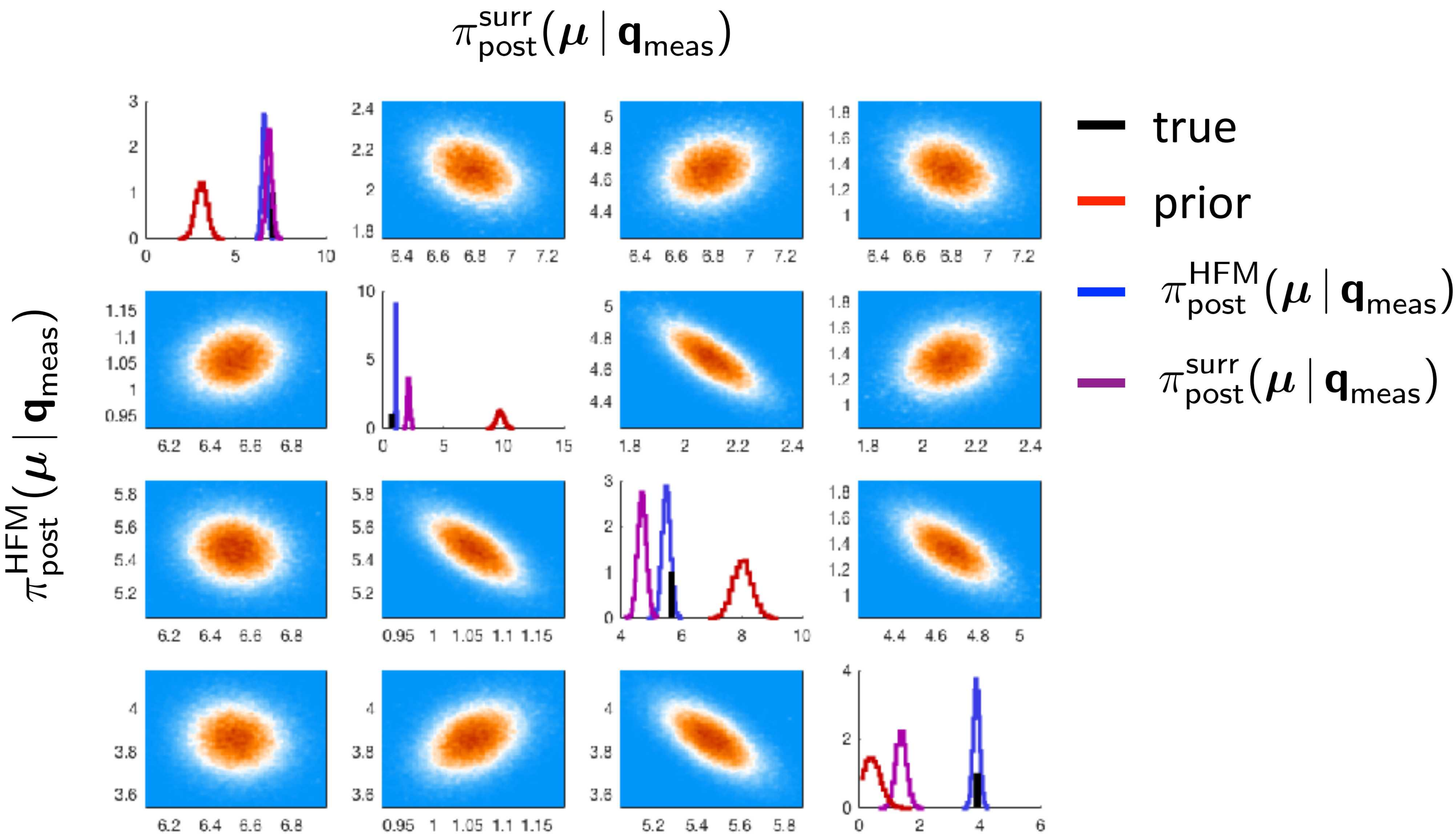
- ROM:
  + cheapest
  - inconsistent formulation

Wall-time performance



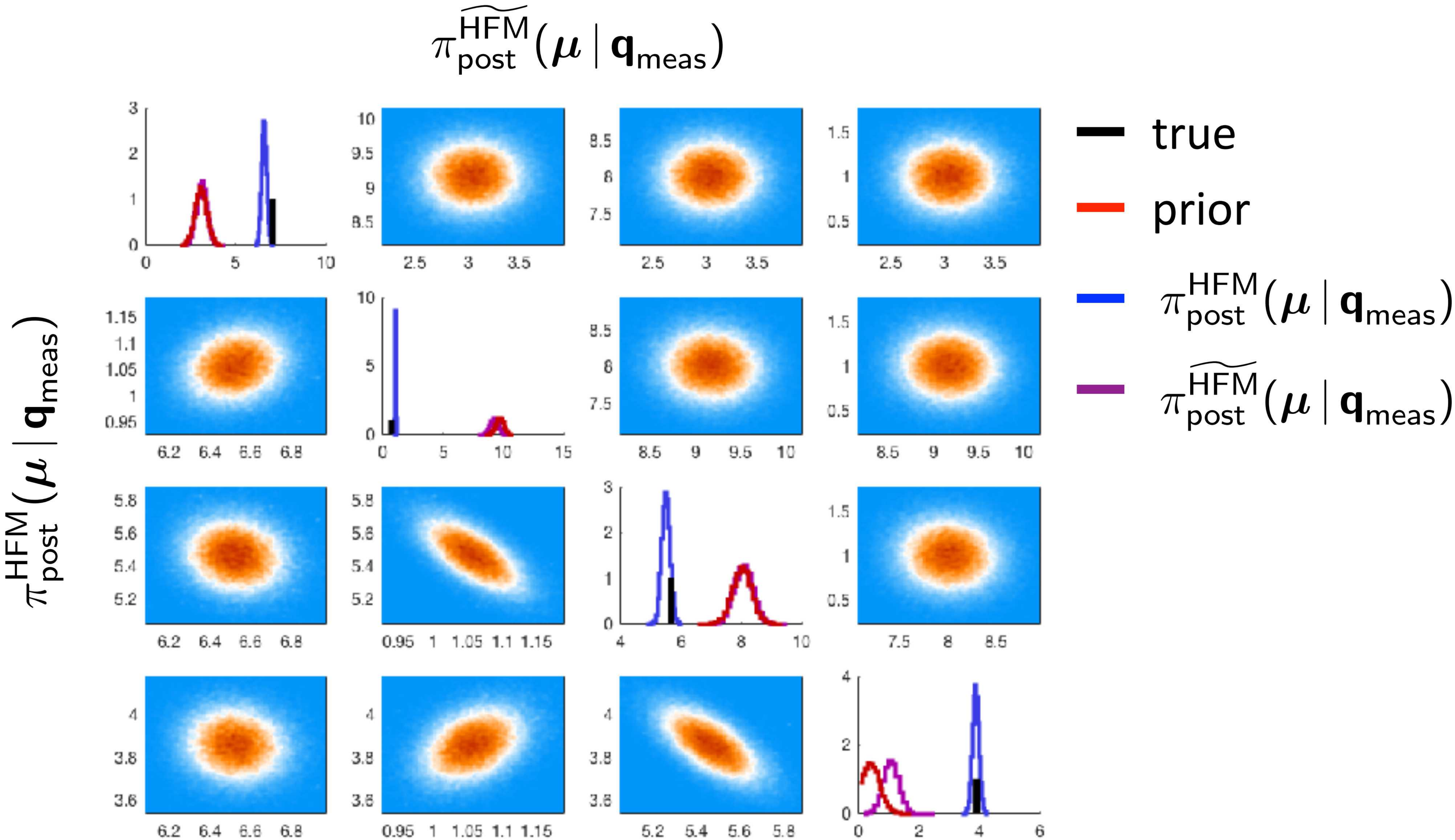
- ROM:
  - + cheapest
  - - inconsistent formulation
- ROM + error models:
  - + cheaper than HFM
  - - more expensive than ROM
  - + consistent formulation

Posteriors: ROM



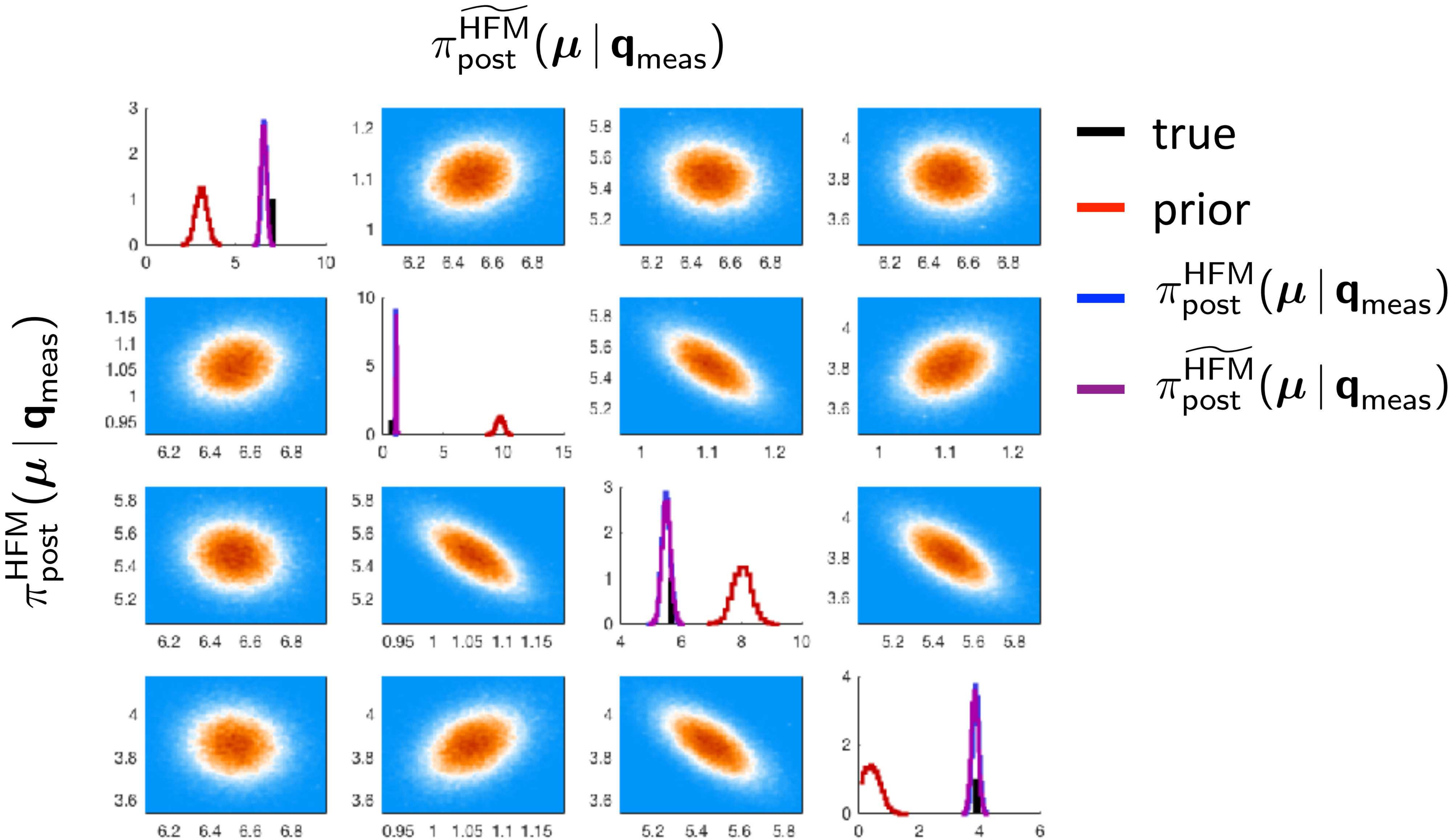
+ HFM posterior: close to **true parameters**
- ROM posterior: far from prior and **true parameters**

Posteriors: ROM + high-variance error model



+ ROM + high-var error model posterior: close to prior

Posteriors: ROM + low-variance error model



+ ROM + low-var error model posterior: close to HFM posterior

Feature engineering

- Desired properties in stochastic error model $\tilde{\delta}(\boldsymbol{\mu})$
  1. cheaply computable, 2. low variance, 3. generalizable
- *Feature engineering*: select error indicators $\boldsymbol{\rho}$ to trade off:
  1. *Number of features*
     - ➔ Large number: costly, lower variance, high capacity regression
     - Small number: cheaper, higher variance, low capacity regression
  2. *Quality of features*
     - High quality: expensive, lower variance
     - ➔ Low quality: cheaper, higher variance
- *Regression model*:
  - ➔ High capacity (low bias, high variance): more data to generalize
  - Low capacity (high bias, low variance): less data to generalize

Method 1: Dual-weighted residual and Gaussian process regression
[Drohmann, C., 2015; Pagani, C., Manzoni, 2018]

Method 2: Large number of features and high-dimensional regression
[Trehan, C., Durlofsky, 2017; Freno, C., 2018]

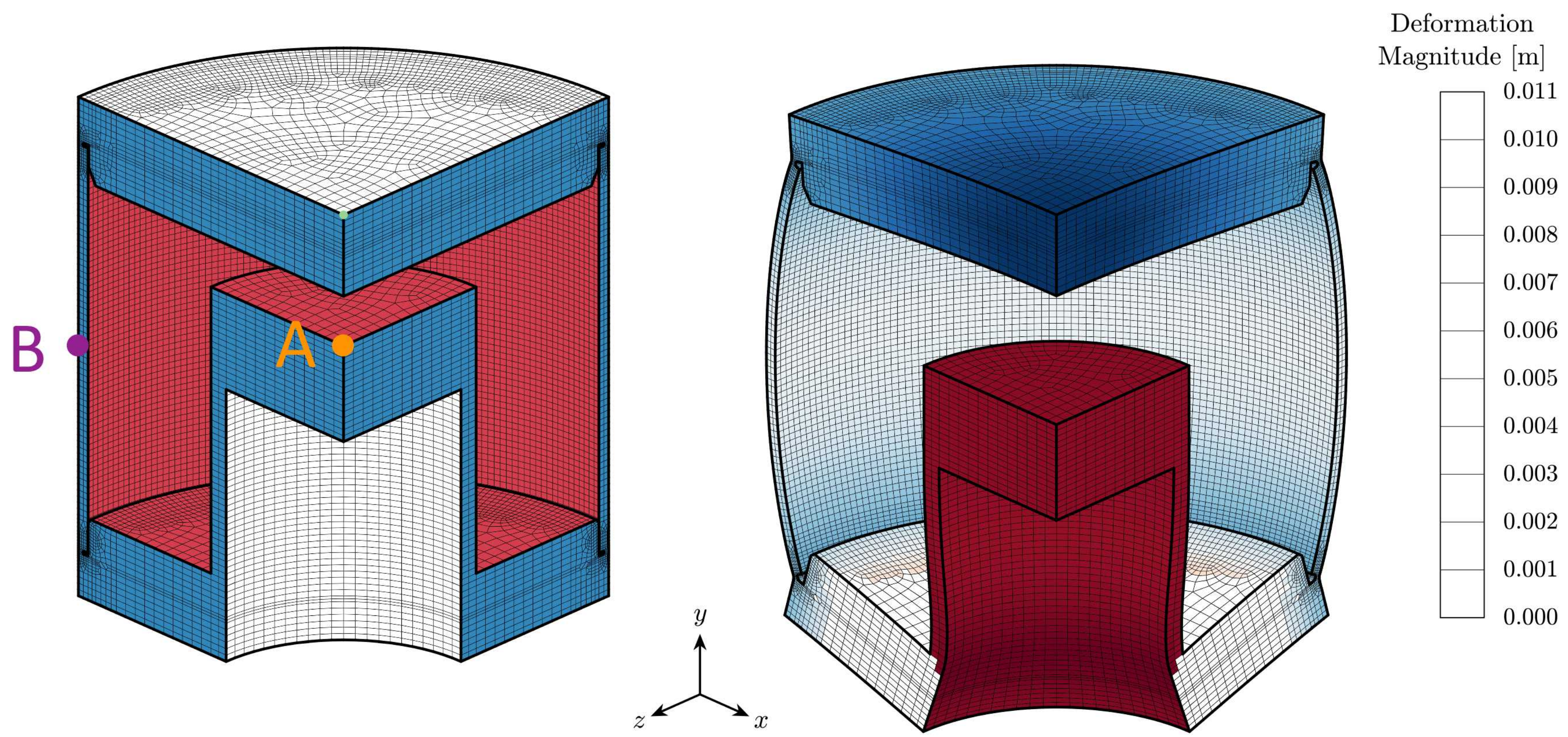
Features: residual samples [Freno, C., 2018]

- **Dual-weighted residual**:

$$q_i(\mathbf{x};\boldsymbol{\mu}) - q_i(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu}) \approx \mathbf{y}_i(\boldsymbol{\mu})^T\mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu}) = \sum_{j=1}^{N} y_{ij}(\boldsymbol{\mu}) r_j(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu})$$

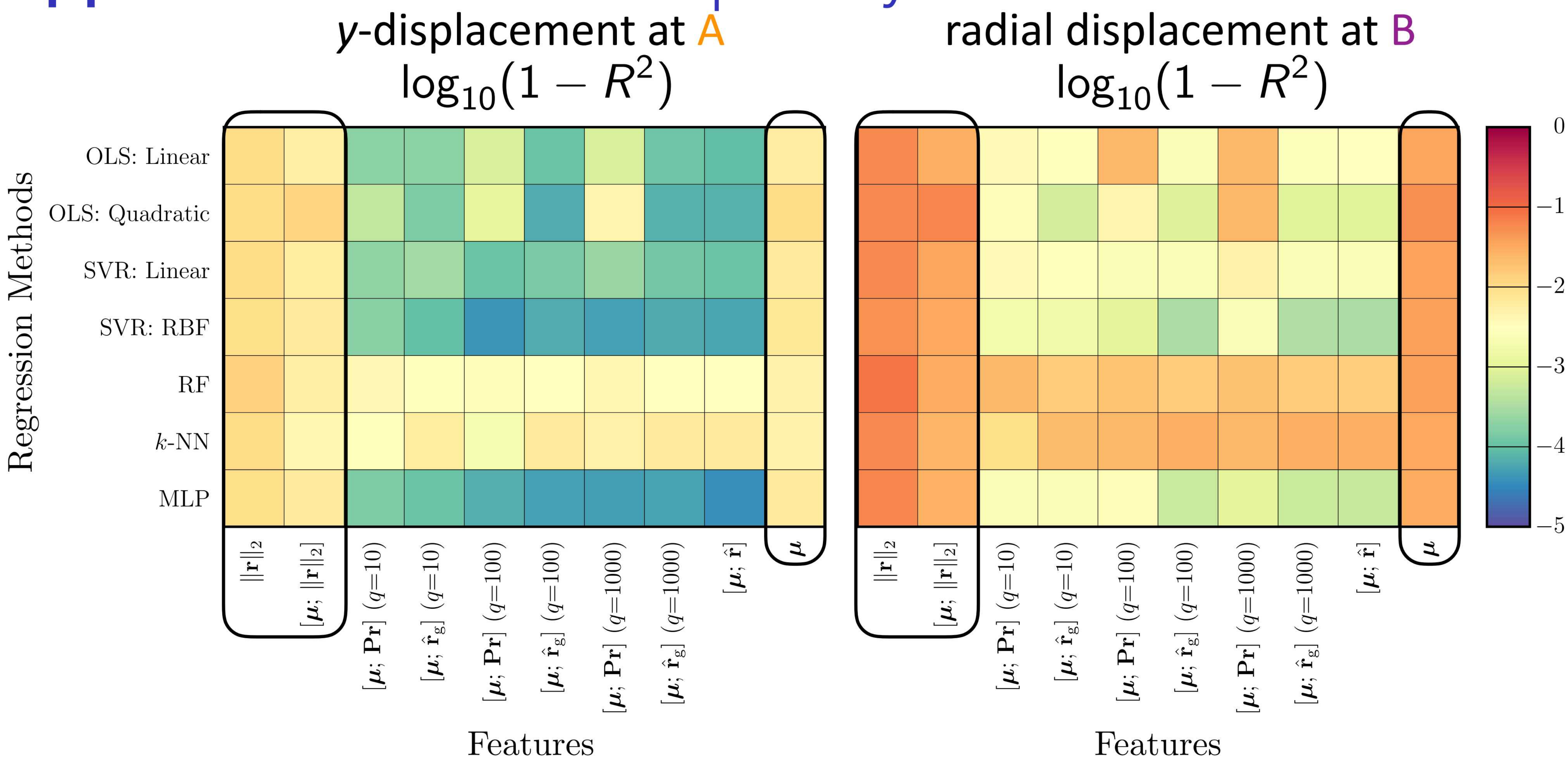
  + high quality
  - costly and less practical: requires dual solve
  - *Note*: this is parameter-dependent weighted sum of residual elements
- **Candidate features:**
  - parameters $\boldsymbol{\mu}$
    - low quality, cheap
  - residual norm $\|\mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu})\|_2$
    - small number, low quality, costly
  - residual $\mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu})$
    - large number, low quality, costly
  - residual samples $\mathbf{P}\mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu})$
    + moderate number, cheap
    - low quality
  - residual PCA $\hat{\mathbf{r}} := \boldsymbol{\Phi}_{\mathbf{r}}^T\mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu})$
    + moderate number, high-quality
    - costly
  - gappy PCA $\hat{\mathbf{r}}_g := (\mathbf{P}\boldsymbol{\Phi}_{\mathbf{r}})^+\mathbf{P}\mathbf{r}(\boldsymbol{\Phi}\hat{\mathbf{x}};\boldsymbol{\mu})$
    + moderate number, high-quality, cheap
- **Candidate regression methods:**
  - SVR, random forests, *k*-nearest neighbors, multilayer perceptron

Application: Predictive capability assessment



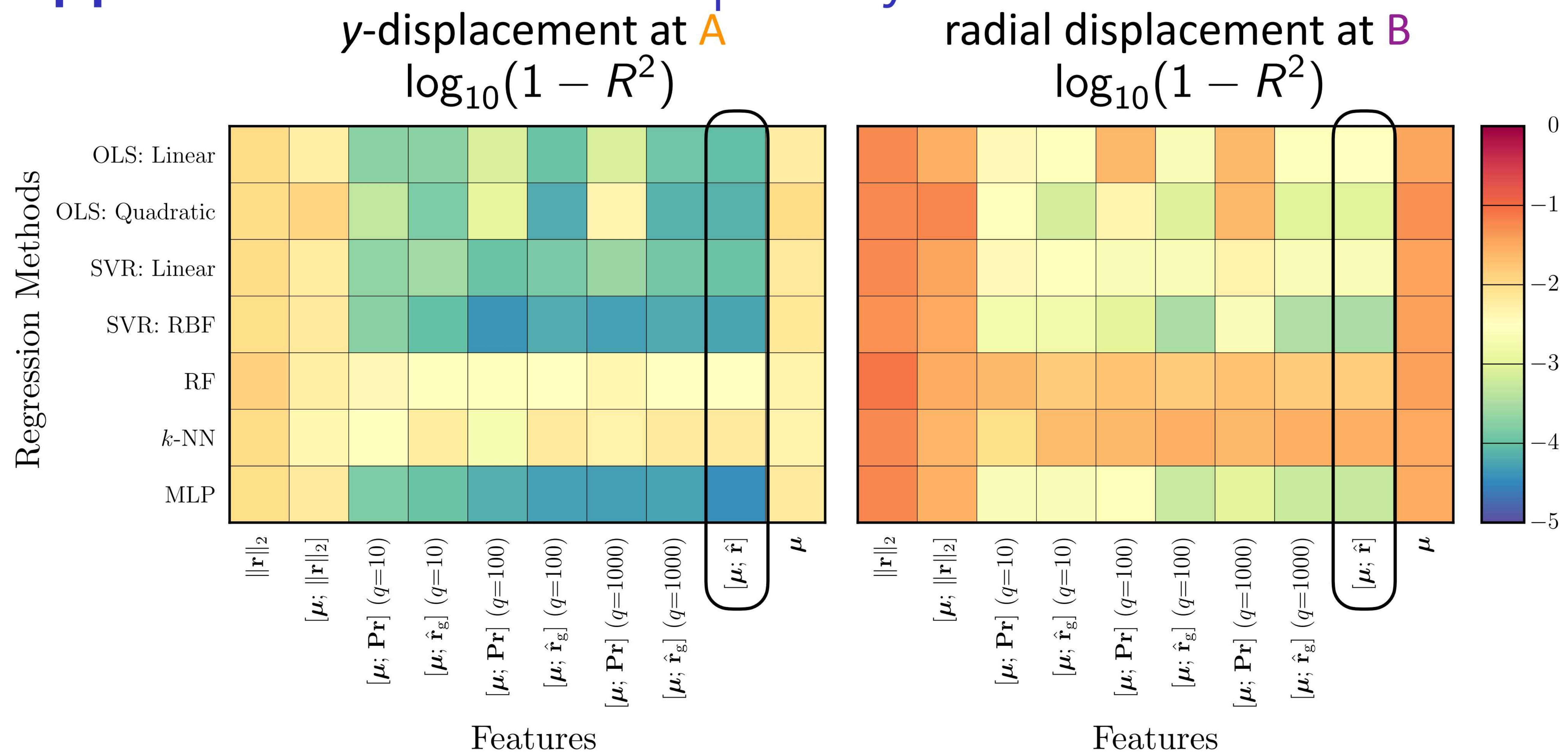
- *high-fidelity model dimension:* 2.8×10^5
- *reduced-order model dimension:* 6
- *inputs* $\boldsymbol{\mu}$*:* elastic modulus, Poisson ratio, applied pressure
- *quantities of interest: y*-displacement at A, radial displacement at B
- 150 training examples, 150 testing examples

Application: Predictive capability assessment



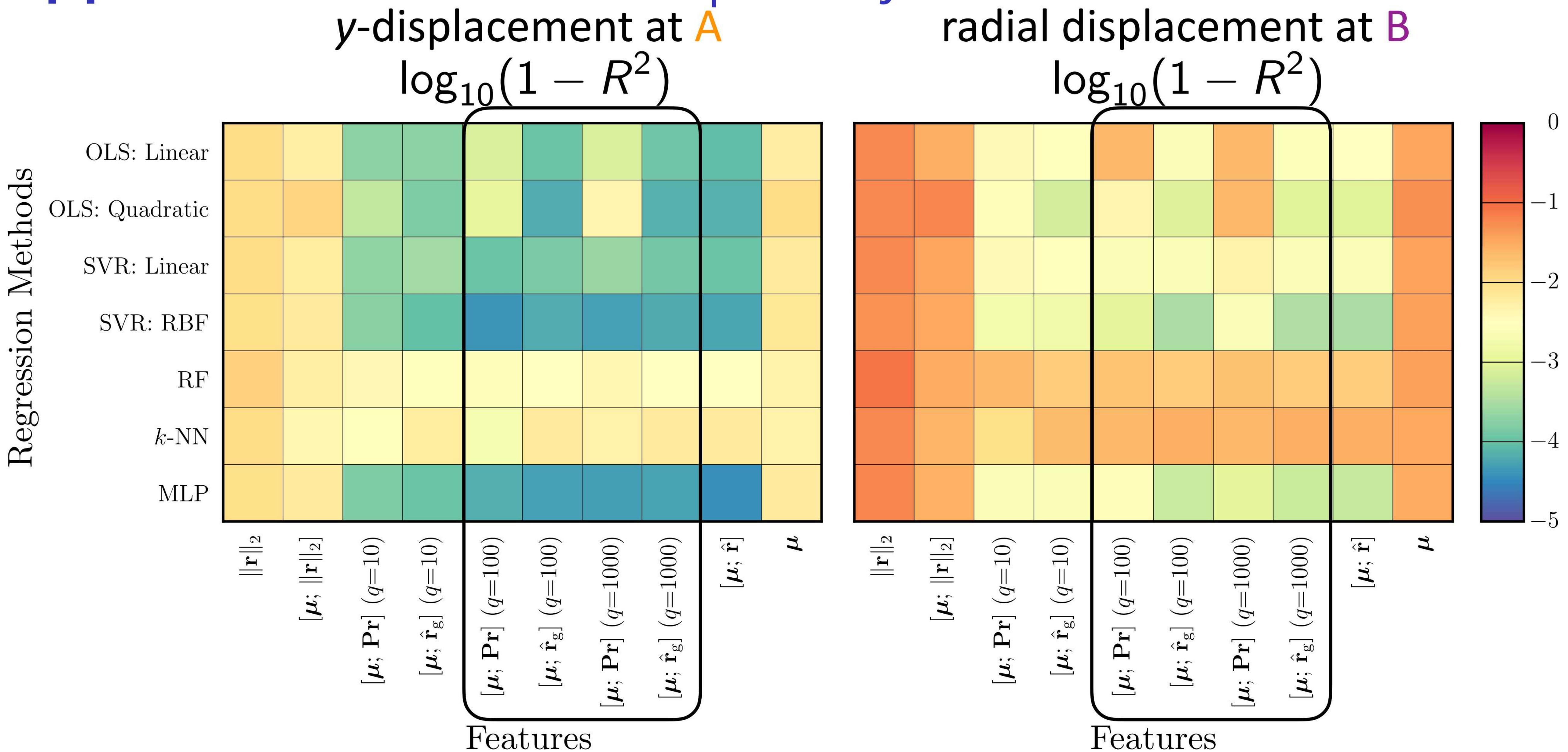
- parameters: large variance
- small number of low-quality features: large variance

Application: Predictive capability assessment



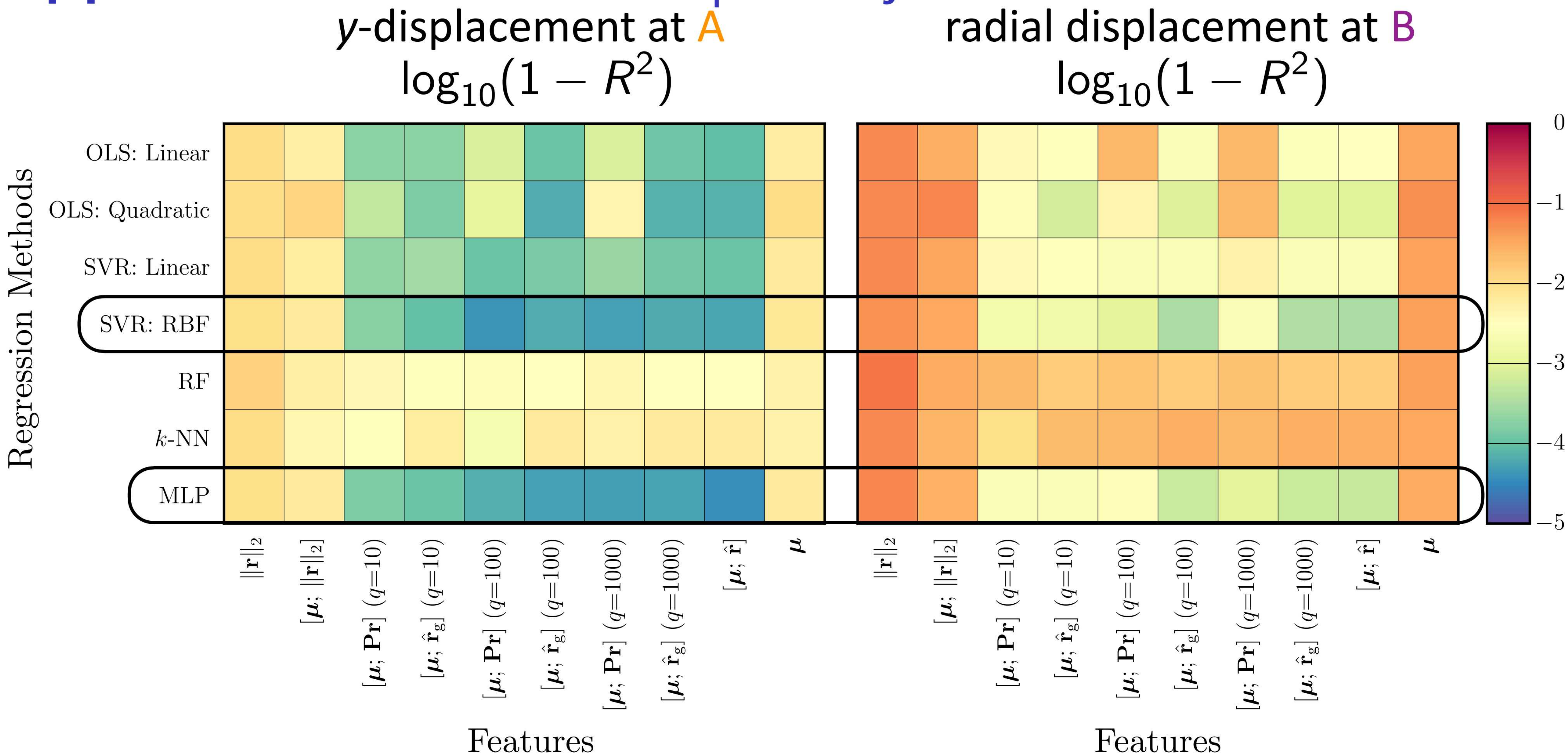
- parameters: large variance
- small number of low-quality features: large variance
- residual PCA: lowest variance overall but costly

Application: Predictive capability assessment



- parameters: large variance
- small number of low-quality features: large variance
- residual PCA: lowest variance overall but costly
- gappy PCA: nearly as low variance, but much cheaper

Application: Predictive capability assessment



- parameters: large variance
- small number of low-quality features: large variance
- residual PCA: lowest variance overall but costly
- + gappy PCA: nearly as low variance, but much cheaper
- + Multilayer perceptron and SVR: yield lowest-variance models

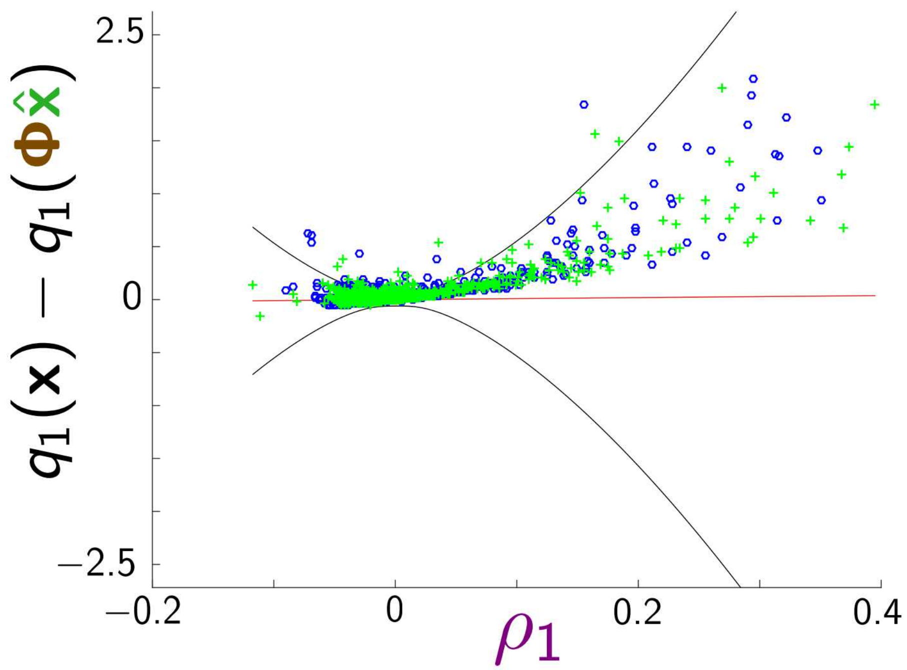
Summary

- What properties do we want in a stochastic error model?
  - *Cheaply computable, low variance, validated*
- What is the optimal stochastic error model?
  - *Error-model variance should be the sample variance over regions*
- How does a stochastic error model affect Bayesian inference?
  - *High-variance: close to prior*
  - *Low-variance: close to HFM posterior*
- How can we construct a stochastic error model for reduced-order models?
  - *Machine-learning error models based on error-indicator features*

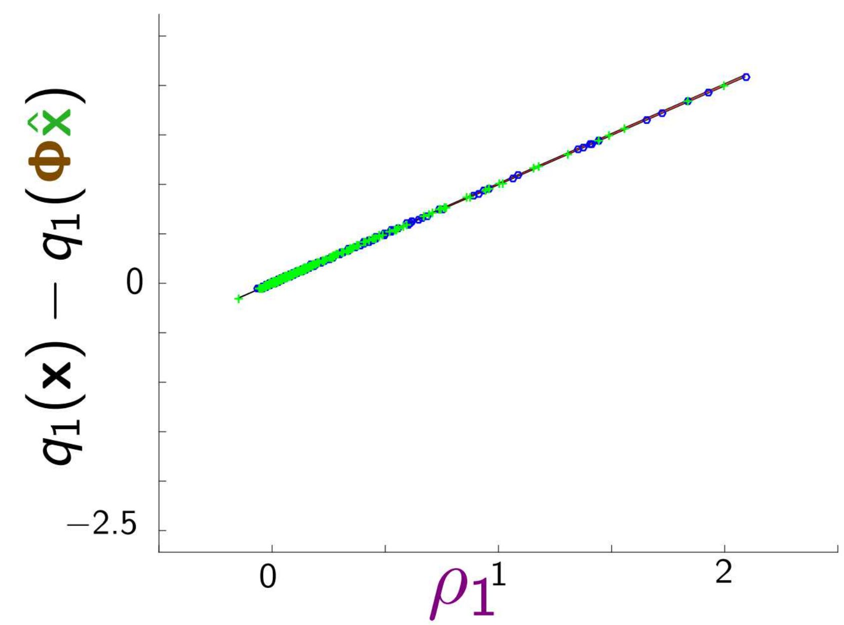
Questions?

References

- C, Uy, Lu, Morzfeld. Rigorous integration of reduced-order models in Bayesian inference via stochastic error models, *in preparation*, 2018.
- Freno, C. Machine-learning error models for approximate solutions to parameterized systems of nonlinear equations, *in preparation*, 2018.
- Trehan, C, and Durlofsky. Error modeling for surrogates of dynamical systems using machine learning. International Journal for Numerical Methods in Engineering, 112(12):1801–1827, 2017.
- Drohmann and C. The ROMES method for statistical modeling of reduced- order-model error. SIAM/ASA Journal on Uncertainty Quantification, 3(1):116–145, 2015.



high-variance error model



low-variance error model

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.