The ROMES method for reduced-order-model uncertainty quantification: application to data assimilation

M. Drohmann and K. Carlberg

Sandia National Laboratories

Workshop on Model Order Reduction and Data
Paris, France
January 6, 2014

Data assimilation by Bayesian inference

- Structural health monitoring

source: Holger Speckmann, Airbus

Given sensor data, what is the updated knowledge of material properties throughout the aircraft?

- Bayesian inference problem

$$\text{inputs } \boldsymbol{\mu} \rightarrow \boxed{\text{high-fidelity model}} \rightarrow \text{outputs } \mathbf{y}$$

Given measurements of the outputs, what is the posterior distribution of the inputs?

Bayesian inference

$$\text{inputs } \boldsymbol{\mu} \rightarrow \boxed{\text{high-fidelity model}} \rightarrow \text{outputs } \mathbf{y}$$

- Bayes' theorem

$$P(\boldsymbol{\mu}|\bar{\mathbf{y}}) = \frac{P(\bar{\mathbf{y}}|\boldsymbol{\mu})P(\boldsymbol{\mu})}{P(\bar{\mathbf{y}})}$$

  - measured outputs $\bar{\mathbf{y}} = \mathbf{y}(\boldsymbol{\mu}^\star) + \boldsymbol{\varepsilon}$ with noise $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$
  - posterior $P(\boldsymbol{\mu}|\bar{\mathbf{y}})$ is sought
  - prior $P(\boldsymbol{\mu})$ is given
  - normalizing factor $P(\bar{\mathbf{y}})$ is handled indirectly
  - likelihood $P(\bar{\mathbf{y}}|\boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{y}(\boldsymbol{\mu}), \sigma^2\mathbf{I})$ sampling requires **high-fidelity model evaluations**
- Objective: numerically sample the posterior distribution
  - \+ achievable in principle, e.g., by MCMC or importance sampling.
  - \- barrier: sampling requires high-fidelity forward solves

Reduced-order modeling and Bayesian inference

$$\text{inputs } \boldsymbol{\mu} \rightarrow \boxed{\text{reduced-order model}} \rightarrow \text{outputs } \mathbf{y}_{\text{red}}$$

- Replace the high-fidelity model with reduced-order model
  - measured outputs $\bar{\mathbf{y}} = \mathbf{y}(\boldsymbol{\mu}^\star) + \varepsilon \approx \mathbf{y}_{\text{red}}(\boldsymbol{\mu}^\star) + \varepsilon$
  - likelihood $P(\bar{\mathbf{y}}|\boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{y}_{\text{red}}(\boldsymbol{\mu}), \sigma^2\mathbf{I})$ sampling requires **reduced-order model evaluations**
  - sampling from the posterior becomes tractable
- **Problem**: neglects reduced-order-model errors

$$\bar{\mathbf{y}} = \mathbf{y}(\boldsymbol{\mu}^\star) + \varepsilon \tag{1}$$

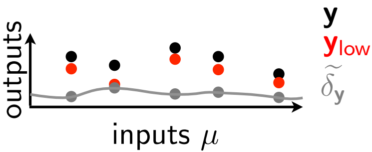
$$= \mathbf{y}_{\text{red}}(\boldsymbol{\mu}^\star) + \boxed{\delta_{\mathbf{y}}(\boldsymbol{\mu}^\star)} + \varepsilon \tag{2}$$

- "An interesting future research direction is the inclusion of estimates of reduced model error as an additional source of uncertainty in the Bayesian formulation." [Galbally et al., 2009]

Goal: construct a statistical model of the reduced-order-model error

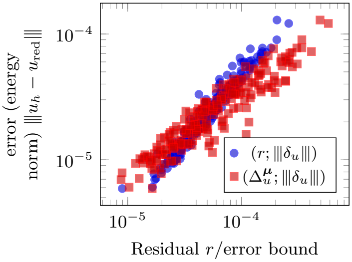
Strategies for ROM error quantification

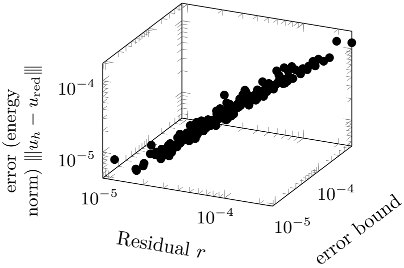
1. Rigorous error bounds
   - \+ independent of input-space dimension
   - \- not amenable to statistical analysis
   - \- often overestimate the error (i.e., high effectivity)
   - \- improving effectivity incurs larger costs [Huynh et al., 2010, Wirtz et al., 2012] or intrusive reformulation of discretization [Yano et al., 2012]
2. Multifidelity correction [Eldred et al., 2004]



   - surrogate model of low-fidelity error as a function of inputs
   - 'correct' low-fidelity outputs with surrogate
   - \+ amenable to statistics
   - \- curse of dimensionality
   - \- ROM errors often highly oscillatory in the input space [Ng and Eldred, 2012]

Our key observation

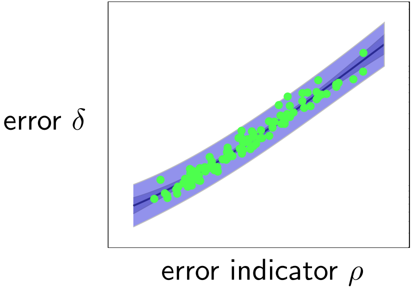




- Residual and error bound often correlate with the true error
- **Main idea**: construct a stochastic process that maps error indicators (not inputs $\boldsymbol{\mu}$!) to a distribution of the error
  - \+ independent of input-space dimension
  - \+ amenable to statistics

Reduced-order model error surrogates

ROMES



- Construct a stochastic process of the ROM error $\widetilde{\delta}(\boldsymbol{\rho})$
- Select a *small number* of error indicators $\boldsymbol{\rho} = \boldsymbol{\rho}(\boldsymbol{\mu})$ that are
  1. cheaply computable online, and
  2. lead to low variance of the stochastic process.
- First attempt: Gaussian process (GP) such that random variables $(\widetilde{\delta}(\boldsymbol{\rho}_1), \widetilde{\delta}(\boldsymbol{\rho}_2), \ldots)$ have joint Gaussian distribution

Gaussian process [Rasmussen and Williams, 2006]

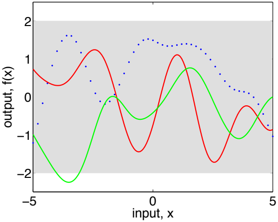
Definition (Gaussian process)

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

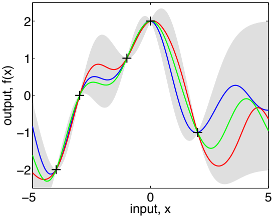
$$\widetilde{\delta}(\boldsymbol{\rho}) \sim \mathcal{GP}(m(\boldsymbol{\rho}), k(\boldsymbol{\rho}, \boldsymbol{\rho}'))$$

- mean function $m(\boldsymbol{\rho})$; covariance function $k(\boldsymbol{\rho}, \boldsymbol{\rho}')$
- given a training set $\{(\delta_i, \boldsymbol{\rho}_i)\}$, the mean and covariance functions can be inferred via Bayesian analysis
- consider two types of Gaussian processes
  1. kernel regression [Rasmussen and Williams, 2006]
  2. relevance vector machine (RVM) [Tipping, 2001]

GP #1: Kernel regression [Rasmussen and Williams, 2006]



(a) prior



(b) posterior

- **prior**: $\widetilde{\delta}(\underline{\boldsymbol{\rho}}) \sim \mathcal{N}(0, K\left(\underline{\boldsymbol{\rho}}, \underline{\boldsymbol{\rho}}\right) + \sigma^2 \mathbf{I})$
  - $k(\boldsymbol{\rho}_i, \boldsymbol{\rho}_j) = \exp \frac{\|\boldsymbol{\rho}_i - \boldsymbol{\rho}_j\|^2}{r^2}$ is a positive definite kernel
  - $\underline{\boldsymbol{\rho}} := \left[\underline{\boldsymbol{\rho}}_{\text{train}} \; \underline{\boldsymbol{\rho}}_{\text{predict}}\right]^T$
- **posterior**: $\widetilde{\delta}(\boldsymbol{\rho}_{\text{predict}}) \sim \mathcal{N}(m(\boldsymbol{\rho}_{\text{predict}}), \operatorname{cov}(\boldsymbol{\rho}_{\text{predict}}))$
- infer hyperparameters σ^2 and r^2

GP #2: Relevance vector machine [Tipping, 2001]

$$\widetilde{\delta}(\boldsymbol{\rho}) = \sum_{m=1}^{M} w_m \phi_m(\boldsymbol{\rho}) + \boldsymbol{\varepsilon}$$

- fixed basis functions ϕ_m (e.g., polynomials, radial-basis functions)
- stochastic coefficients w_m
- noise $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$
- **prior**: $w \sim \mathcal{N}(0, \mathrm{diag}(\alpha_i))$
- **posterior**: $w \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$ leads to posterior dist. of $\widetilde{\delta}(\boldsymbol{\rho})$
- infer hyperparameters σ^2 and α_i

ROMES Algorithm

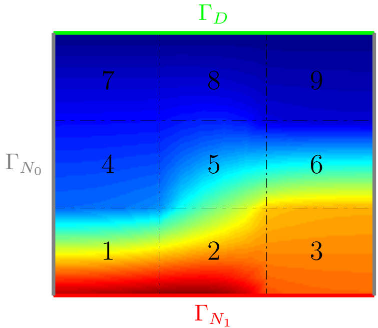
Offline

1. Populate ROMES database $\{(\delta(\boldsymbol{\mu}), \bar{\boldsymbol{\rho}}(\boldsymbol{\mu})) \mid \boldsymbol{\mu} \in \mathcal{D}_{\text{train}}\}$, where $\bar{\boldsymbol{\rho}}$ denotes candidate indicators.
2. Identify a few error indicators $\boldsymbol{\rho} \subset \bar{\boldsymbol{\rho}}$ that lead to low variance in the Gaussian process.
3. Construct the Gaussian process $\widetilde{\delta}(\boldsymbol{\rho}) \sim \mathcal{GP}(m(\boldsymbol{\rho}), k(\boldsymbol{\rho}, \boldsymbol{\rho}'))$ by Bayesian inference.

Online (for any $\boldsymbol{\mu}^\star \in \mathcal{D}$)

1. compute the ROM solution
2. compute error indicators $\boldsymbol{\rho}(\boldsymbol{\mu}^\star)$
3. obtain $\widetilde{\delta}(\boldsymbol{\rho}(\boldsymbol{\mu}^\star)) \sim \mathcal{N}(m(\boldsymbol{\rho}(\boldsymbol{\mu}^\star))), k(\boldsymbol{\rho}(\boldsymbol{\mu}^\star), \boldsymbol{\rho}(\boldsymbol{\mu}^\star))$
4. correct the ROM solution

Thermal block (Parametrically coercive and compliant, affine, linear, elliptic)

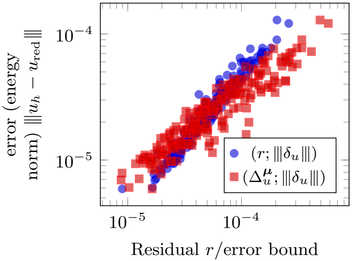


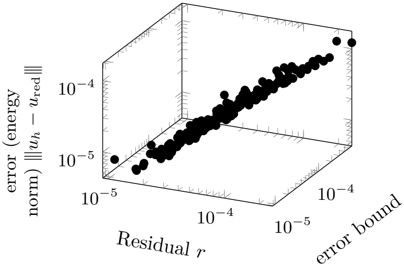
$$\triangle c(x;\boldsymbol{\mu})u(x;\boldsymbol{\mu}) = 0 \text{ in } \Omega \qquad u(\boldsymbol{\mu}) = 0 \text{ on } \Gamma_D$$

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

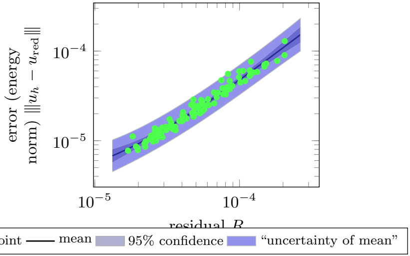
- Inputs $\boldsymbol{\mu} \in [0.1, 10]^9$ define diffusivity c in subdomains
- Output $\mathbf{y}(\boldsymbol{\mu}) = \int_{\Gamma_{N_1}} u(\boldsymbol{\mu})\mathrm{d}x$ is compliant
- ROM constructed via RB–Greedy [Patera and Rozza, 2006]
- Consider two errors: 1) energy norm of state-space error $|||u(\boldsymbol{\mu}) - u_{\mathrm{red}}(\boldsymbol{\mu})|||$, 2) output bias $\mathbf{y}(\boldsymbol{\mu}) - \mathbf{y}_{\mathrm{red}}(\boldsymbol{\mu})$

Error #1: energy norm $|||u(\boldsymbol{\mu}) - u_{\text{red}}(\boldsymbol{\mu})|||$



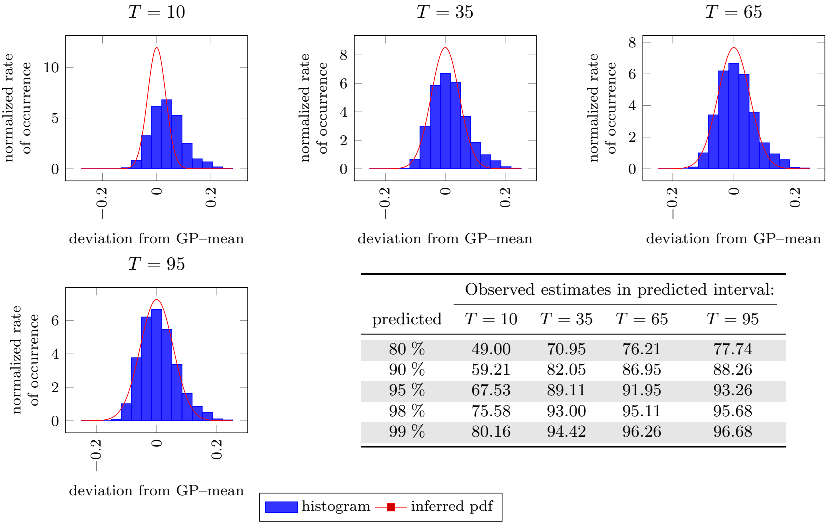


- Residual and error bound correlate with error



- ROMES (Kernel, residual indicator) in log-log space promising

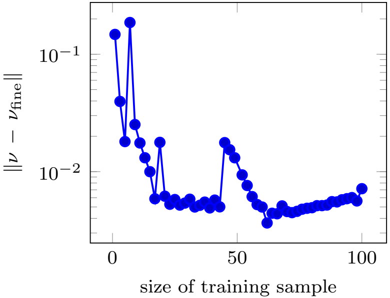
Gaussian-process assumptions verified (Kernel, residual indicator)



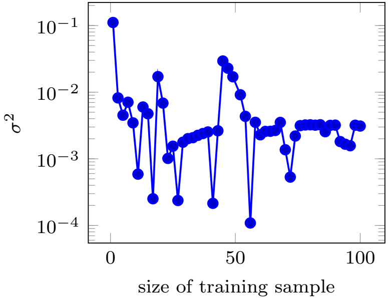
predicted	Observed estimates in predicted interval: $T=10$	$T=35$	$T=65$	$T=95$
80 %	49.00	70.95	76.21	77.74
90 %	59.21	82.05	86.95	88.26
95 %	67.53	89.11	91.95	93.26
98 %	75.58	93.00	95.11	95.68
99 %	80.16	94.42	96.26	96.68

GP variables converge (Kernel GP, residual indicator)

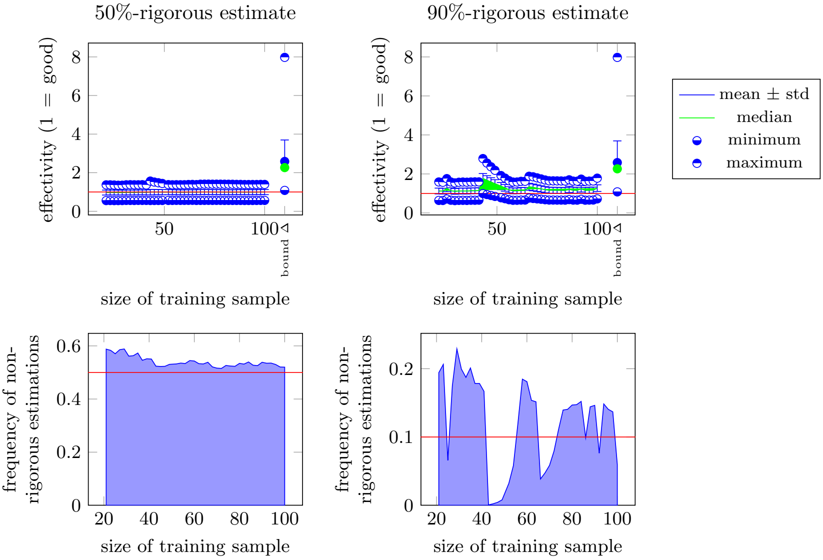
(i) convergence of mean m



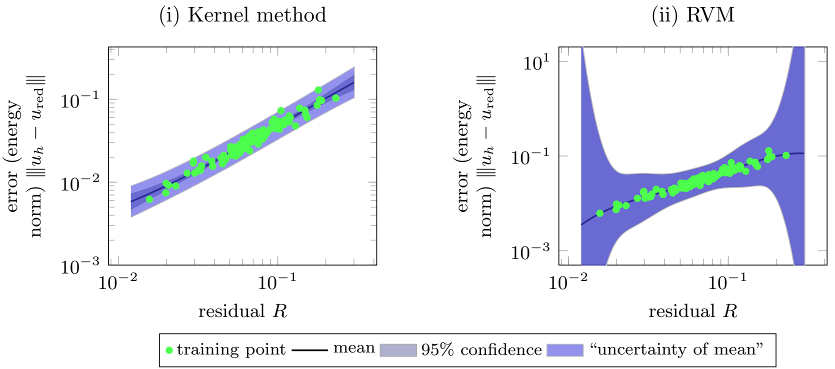
(ii) convergence of variance σ^2



Can achieve 'statistical rigor' (Kernel GP, residual indicator)

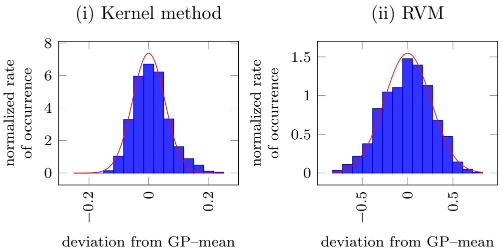


ROMES: Kernel and RVM GP comparison



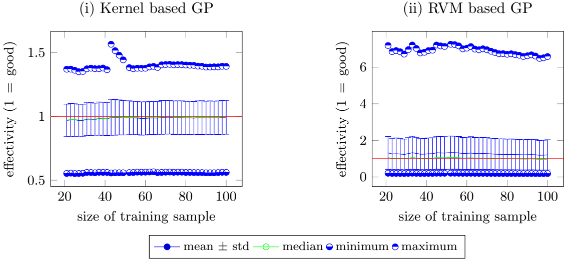
- RVM (Legendre-polynomial basis functions): significant uncertainty in mean's high-order polynomial coefficients

ROMES: Kernel and RVM GP comparison



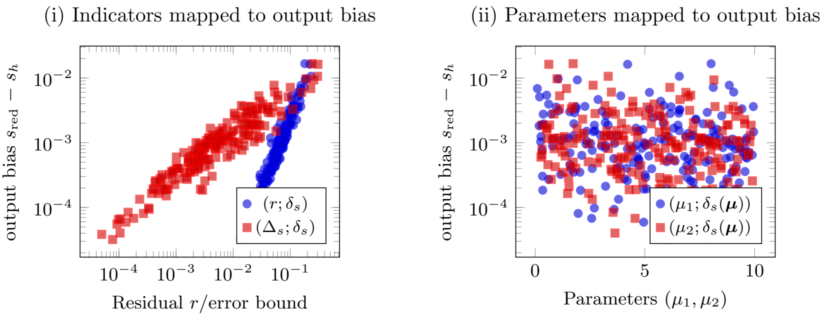
c	(i)	(ii)
80 %	77.7	77.9
90 %	88.1	88.7
95 %	92.9	94.7
98 %	95.6	97.4
99 %	96.6	99.2

- GP structure and confidence intervals verified in both cases



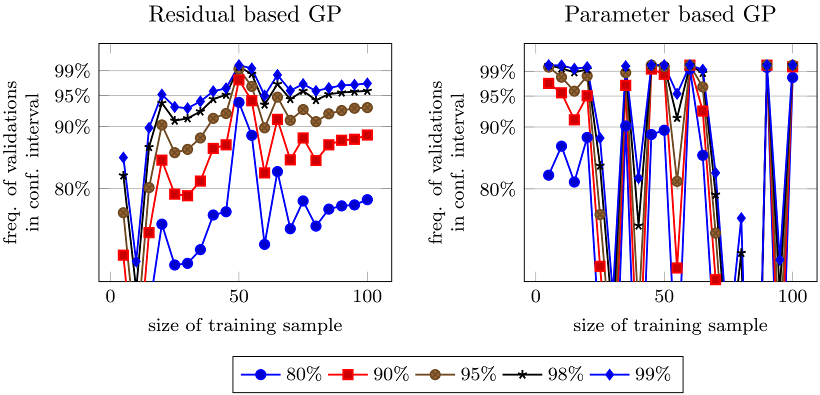
- Kernel GP produces better effectivity

Error #2: output bias $\mathbf{y}(\boldsymbol{\mu}) - \mathbf{y}_{\text{red}}(\boldsymbol{\mu})$



+ Residual and error bound are good indicators (ROMES)
- Inputs are poor indicators (Multifidelity correction)

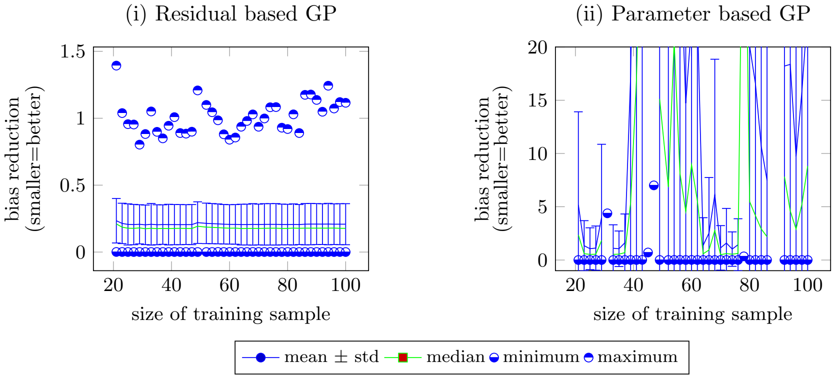
Gaussian-process assumption verification



+ ROMES (Kernel GP, residual indicator) confidence intervals converge.
- Multifidelity correction (Kernel GP, input indicator) confidence intervals do not converge.

Bias improvement

$$\text{bias reduction} = \frac{E\left(\mathbf{y}_{\mathrm{red}}(\boldsymbol{\mu}) + \widetilde{\delta}_{\mathbf{y}}(\boldsymbol{\mu}) - \mathbf{y}(\boldsymbol{\mu})\right)}{\mathbf{y}_{\mathrm{red}}(\boldsymbol{\mu}) - \mathbf{y}(\boldsymbol{\mu})}$$



+ ROMES reduces bias by roughly an order of magnitude
- Multifidelity correction often *increases* the bias

Conclusions

- ROMES
  - combines existing ROM error indicators with supervised machine learning to statistical quantify ROM error
  - relies on identifying error indicators that yield low variance
  - 'statistical rigor' achievable
  - outperforms multifidelity correction (inputs are poor error indicators)
  - highlights strength of reduced-order models for data assimilation: other surrogates (likely) do not have such powerful error indicators
- Future work
  - apply to nonlinear, time-dependent problems
  - incorporate in likelihood function

$$\bar{\mathbf{y}} = \mathbf{y}_{\text{red}}(\boldsymbol{\mu}^\star) + \boxed{\delta_{\mathbf{y}}(\boldsymbol{\mu}^\star)} + \varepsilon$$

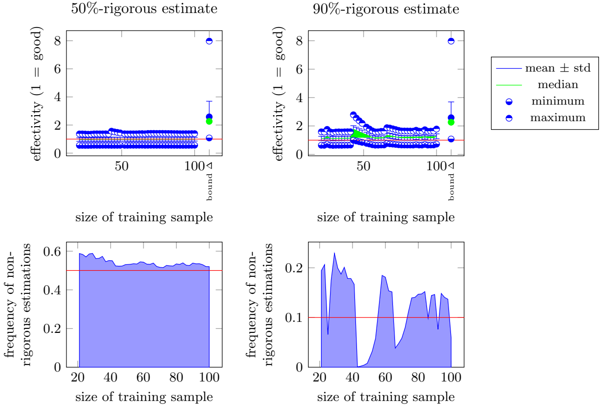
    where $\delta_{\mathbf{y}}$ and ε may have different distributions
  - develop error indicators for this purpose
  - automated selection of indicators and Gaussian process

Acknowledgments

- Khachik Sargsyan: helpful discussions
- This research was supported in part by an appointment to the Sandia National Laboratories Truman Fellowship in National Security Science and Engineering, sponsored by Sandia Corporation (a wholly owned subsidiary of Lockheed Martin Corporation) as Operator of Sandia National Laboratories under its U.S. Department of Energy Contract No. DE-AC04-94AL85000.

Questions?

- M. Drohmann & K. Carlberg, "The ROMES method for reduced-order-model uncertainty quantification," *in preparation*.



Bibliography I

Eldred, M. S., Giunta, A. A., Collis, S. S., Alexandrov, N. A., and Lewis, R. M. (2004).
Second-order corrections for surrogate-based optimization with model hierarchies.
In *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, NY*, number AIAA Paper 2004-4457.

Galbally, D., Fidkowski, K., Willcox, K., and Ghattas, O. (2009).
Non-linear model reduction for uncertainty quantification in large-scale inverse problems.
International Journal for Numerical Methods in Engineering, 81(12):1581–1608.

Bibliography II

Huynh, D. B. P., Knezevic, D. J., Chen, Y., Hesthaven, J. S., and Patera, A. T. (2010).
A natural-norm successive constraint method for inf-sup lower bounds.
Comput. Methods Appl. Mech. Engrg., 199:1963–1975.

Ng, L. and Eldred, M. S. (2012).
Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation.
In *AIAA 2012-1852*.

Patera, A. T. and Rozza, G. (2006).
Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations.
MIT.

Bibliography III

Rasmussen, C. and Williams, C. (2006).
Gaussian Processes for Machine Learning.
MIT Press.

Tipping, M. E. (2001).
Sparse bayesian learning and the relevance vector machine.
The Journal of Machine Learning Research, 1:211–244.

Wirtz, D., Sorensen, D. C., and Haasdonk, B. (2012).
A-posteriori error estimation for DEIM reduced nonlinear dynamical systems.
Preprint Series, Stuttgart Research Centre for Simulation Technology.

Bibliography IV

Yano, M., Patera, A. T., and Urban, K. (2012).
A space-time certified reduced-basis method for Burgers' equation.
Math. Mod. Meth. Appl. S., submitted.