This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SAND2022-7326C

Sandia National Laboratories

Multifidelity uncertainty quantification for non-deterministic models

Bryan Reuter, Gianluca Geraci, Tim Wildey, Michael Eldred

June 8, ECCOMAS 2022

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

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.

SAND?????

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.

Non-deterministic models are relevant for many applications

Models with an uncontrollable source of stochasticity are ***non-deterministic***

Examples include:

- Turbulent flow
- Particle-in-cell methods for plasma simulation
- Radiation transport [Clements, 2021]
- Flows through random subsurfaces
- Cybersecurity [Geraci, 2021]
- Machine learning

Multifidelity uncertainty quantification

- Multifidelity (MF) approaches for uncertainty quantification (UQ) leverage information from multiple sources which vary in accuracy and cost
- Objective is to improve reliability of statistics of high-fidelity computational models
- Non-deterministic models are candidates for both high- and low-fidelity information sources
- The intrinsic stochasticity effectively weakens correlations between models
- Practitioners may try to drive stochastic noise contribution to zero (or as low as they can afford)

Goal

Explore the efficacy of common MFUQ approaches (MLMC, MFMC) for non-deterministic model sets

Preliminaries

Preliminaries

Consider the mapping $f: X \times \Omega \to \mathcal{D}$ which represents a computational model

$$y = f(x, \omega)$$

- $x \in X$ is the set of ***accessible*** input parameters
  - Explicitly specified for each model evaluation
- $\omega \in \Omega$ is the set of ***inaccessible*** random variables
  - The realization is not known, nor the underlying probability distribution
  - Represents the intrinsic stochasticity in the model

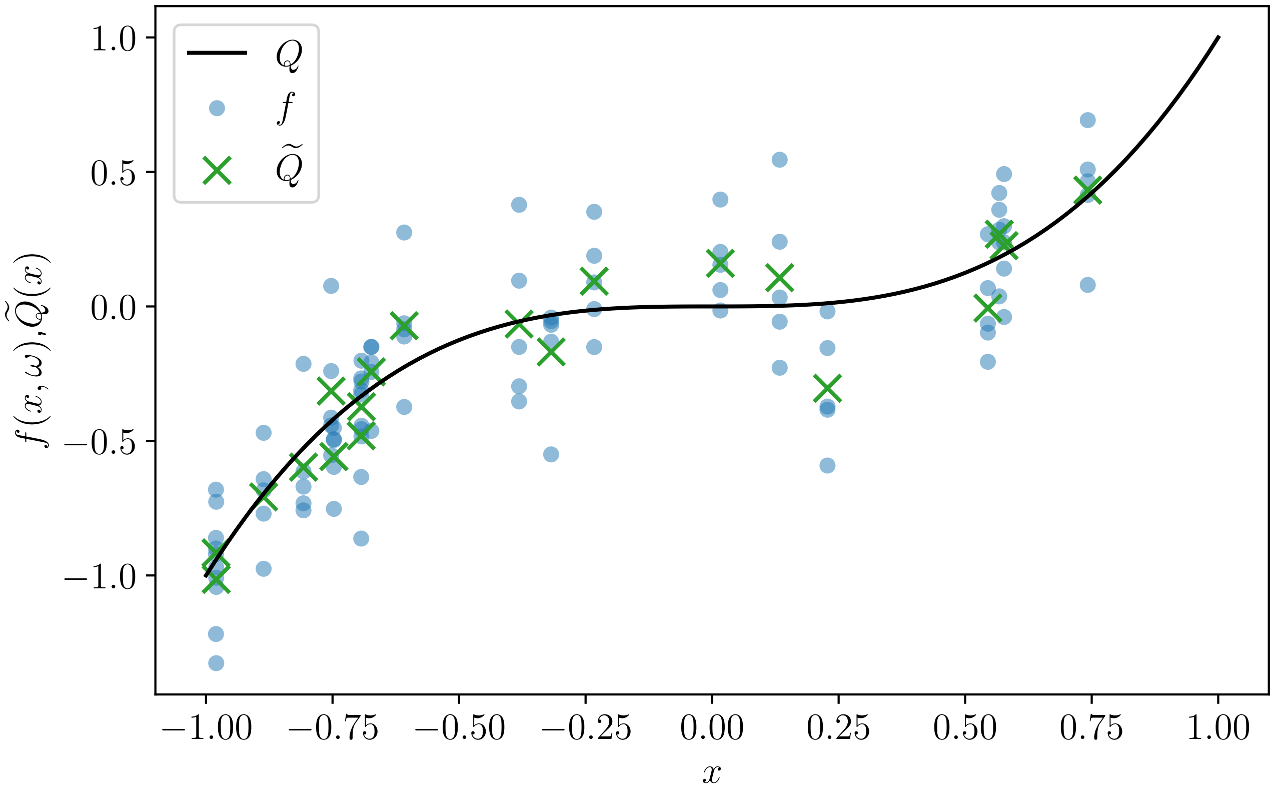
- Practically, for a given x, y is a random variable over Ω which we can characterize statistically through realizations (replicas) $\{f(x, \omega^{(j)})\}_j$

Preliminaries

We consider the quantity of interest (QoI) to be the expectation over ω:

$$Q(x) = \mathbb{E}_\omega[f] \approx \frac{1}{N_\omega} \sum_{j=1}^{N_\omega} f\left(x, \omega^{(j)}\right) = \tilde{Q}(x)$$

- We assume the model cost C is independent of the inputs
- The total cost of obtaining the noisy QoI is $N_\omega C$



Preliminaries

In the multifidelity context, consider a set of models $\{f_\ell\}_\ell$ for $\ell = 1, \ldots, L$

- Following the multifidelity Monte-Carlo literature, let f_1 be the highest fidelity model
- Let $\{C_\ell\}_\ell$ be the costs of one realization of the models
- Let $\{N_{\omega,\ell}\}_\ell$ be the number of replicas for each model (these may be different from model to model)
- We assume number of replicas is independent of the location x

Mean estimators

Optimal strategy for single fidelity Monte-Carlo estimator of the mean is to use only one replica

$$\hat{Q}^{MC} = \frac{1}{N}\sum_{i=1}^{N}\tilde{Q}\left(x^{(i)}\right) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{N_\omega}\sum_{j=1}^{N_\omega} f\left(x^{(i)}, \omega^{(j)}\right)\right] \approx \mathbb{E}[Q(x)]$$

$$\mathbb{V}\left[\hat{Q}^{MC}\right] = \frac{1}{N}\mathbb{V}[\tilde{Q}]$$

By the law of total variance, it follows that

$$\mathbb{V}\left[\hat{Q}^{MC}\right] = \frac{1}{N}\left(\mathbb{V}[Q] + \frac{\mathbb{E}[\sigma_\omega^2]}{N_\omega}\right)$$

where $\mathbb{E}[\sigma_\omega^2] = \mathbb{E}\left[\mathbb{V}_\omega[f]\right]$

How do things change in a multifidelity setting?

- Increasing the number of replicas for a particular model will improve its effective correlations, but increase its effective cost
- For a given budget, this means improving correlations is at odds with exploring the uncertain parameter space
- For a given number of model replicas, the MLMC and MFMC algorithms can be used as normal
  - Model tuning exercise

Basic steps for non-deterministic model set

1. Estimate cost, noise, and noisy correlation between models with a pilot study
2. Estimate underlying “non-noisy” statistics
   - Variance deconvolution [Geraci, Olson 2022]
3. Decide on target
   - Total computational budget
   - Required statistical error
4. Solve optimization problem for a variety of estimators
   - How many samples for each model?
   - How many replicas per model?
5. Deploy optimal sampling strategy

Multilevel Monte-Carlo estimator for the mean

$$\tilde{Q}_\ell(x) \equiv \frac{1}{N_{\omega,\ell}} \sum_{j=1}^{N_{\omega,\ell}} f_\ell\left(x, \omega_\ell^{(j)}\right) \approx Q_\ell(x)$$

Define $\tilde{Y}_\ell(x) \equiv \tilde{Q}_\ell(x) - \tilde{Q}_{\ell+1}(x)$

$$\hat{Q}^{ML} = \sum_{\ell=1}^{L} \left\{ \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \tilde{Y}_\ell\left(x^{(i),\ell}\right) \right\}$$

$$= \sum_{\ell=1}^{L} \left\{ \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left[\frac{1}{N_{\omega,\ell}} \sum_{j=1}^{N_{\omega,\ell}} f_\ell\left(x^{(i),\ell}, \omega_\ell^{(j)}\right) - \frac{1}{N_{\omega,\ell+1}} \sum_{j=1}^{N_{\omega,\ell+1}} f_{\ell+1}\left(x^{(i),\ell}, \omega_{\ell+1}^{(j)}\right) \right] \right\}$$

$$\mathbb{V}[\hat{Q}^{ML}] = \sum_{\ell=1}^{L} \frac{\mathbb{V}[\tilde{Y}_\ell]}{N_\ell}$$

Unlike single fidelity, there exists a trade-off between replicas and UQ samples for MF methods

$$\mathbb{V}[\hat{Q}^{ML}] = \sum_{\ell=1}^{L} \frac{\mathbb{V}[\tilde{Y}_\ell]}{N_\ell}$$

Can show $\mathbb{V}[\tilde{Y}_\ell] = \mathbb{V}[Q_\ell - Q_{\ell+1}] + \frac{\mathbb{E}[\sigma^2_{\omega,\ell}]}{N_{\omega,\ell}} + \frac{\mathbb{E}[\sigma^2_{\omega,\ell+1}]}{N_{\omega,\ell+1}}$

$$\Rightarrow \mathbb{V}[\hat{Q}^{ML}] = \sum_{\ell=1}^{L} \frac{\mathbb{V}[Y_\ell]}{N_\ell} + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \frac{\mathbb{E}[\sigma^2_{\omega,\ell}]}{N_{\omega,\ell}} + \sum_{\ell=1}^{L-1} \frac{1}{N_\ell} \frac{\mathbb{E}[\sigma^2_{\omega,\ell+1}]}{N_{\omega,\ell+1}}$$

Consider two-model case:

$$\mathbb{V}[\hat{Q}^{ML}] = \frac{\mathbb{V}[Q_{HF} - Q_{LF}]}{N_{HF}} + \frac{\mathbb{V}[Q_{LF}]}{N_{LF}} + \frac{1}{N_{HF}} \frac{\mathbb{E}[\sigma^2_{\omega,HF}]}{N_{\omega,HF}} + \frac{1}{N_{HF}} \frac{\mathbb{E}[\sigma^2_{\omega,LF}]}{N_{\omega,LF}} + \frac{1}{N_{LF}} \frac{\mathbb{E}[\sigma^2_{\omega,LF}]}{N_{\omega,LF}}$$

For a given number of replicas, classical MLMC solution applies (total budget fixed)

$$\mathbb{V}[\hat{Q}^{ML}] = \sum_{\ell=1}^{L} \frac{\mathbb{V}[\tilde{Y}_\ell]}{N_\ell}$$

$$N_\ell = \frac{C_{tot}}{S} \sqrt{\frac{\mathbb{V}[\tilde{Y}_\ell]}{\tilde{C}_\ell}}$$

$$S = \sum_{k=1}^{L} \sqrt{\mathbb{V}[\tilde{Y}_k]\tilde{C}_k}$$

$$\tilde{C}_\ell = N_{\omega,\ell} C_\ell + N_{\omega,\ell+1} C_{\ell+1}$$

Multifidelity Monte-Carlo

$$\hat{Q}^{MF} = \frac{1}{N_1}\sum_{i=1}^{N_1} \tilde{Q}_1\left(x^{(i),1}\right) + \sum_{\ell=2}^{L} \alpha_\ell \left\{ \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} \tilde{Q}_\ell\left(x^{(i),\ell}\right) - \frac{1}{N_{\ell-1}}\sum_{i=1}^{N_{\ell-1}} \tilde{Q}_{\ell-1}\left(x^{(i),\ell-1}\right) \right\}$$

$$= \frac{1}{N_1}\sum_{i=1}^{N_1} \frac{1}{N_{\omega,1}}\left[\sum_{j=1}^{N_{\omega,1}} f_1\left(x^{(i),1}, \omega_1^{(j)}\right)\right]$$

$$+ \sum_{\ell=2}^{L} \alpha_\ell \left\{ \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} \frac{1}{N_{\omega,\ell}}\left[\sum_{j=1}^{N_{\omega,\ell}} f_\ell\left(x^{(i),\ell}, \omega_\ell^{(j)}\right)\right] - \frac{1}{N_{\ell-1}}\sum_{i=1}^{N_{\ell-1}} \frac{1}{N_{\omega,\ell-1}}\left[\sum_{j=1}^{N_{\omega,\ell-1}} f_{\ell-1}\left(x^{(i),\ell-1}, \omega_{\ell-1}^{(j)}\right)\right]\right\}$$

Classical solution to optimization problem applies.

Hints at trade-off between improving correlations and increasing effective model cost.

$$\mathbb{V}[\hat{Q}^{MF}] = \frac{\mathbb{V}[\tilde{Q}_1]}{C_{tot}} \left\{ \sum_{\ell=1}^{L} \sqrt{\widetilde{W}_\ell \left(\tilde{\rho}_{1,\ell}^2 - \tilde{\rho}_{1,\ell+1}^2 \right)} \right\}^2$$

$$\tilde{\rho}_{1,\ell} = \frac{\mathbb{C}[\tilde{Q}_1, \tilde{Q}_\ell]}{\sqrt{\mathbb{V}[\tilde{Q}_1]\mathbb{V}[\tilde{Q}_\ell]}}$$

$$\widetilde{W}_\ell = N_{\omega,\ell} C_\ell$$

$$\mathbb{V}[\tilde{Q}_\ell] = \mathbb{V}[Q_\ell] + \frac{\mathbb{E}[\sigma_{\omega,\ell}^2]}{N_{\omega,\ell}}$$

Optimal number of high fidelity replicas is always one for MFMC

Can show $\frac{\partial \mathbb{V}[\hat{Q}^{MF}]}{\partial N_{\omega,1}} > 0$ always for non-trivial model collections

Outline of proof

1. Differentiate MFMC variance expression
2. A lot of algebra…
3. Recognize $\mathbb{C}[\tilde{Q}_1, \tilde{Q}_\ell] = \mathbb{C}[Q_1, Q_\ell]$ for $\ell \neq 1$
4. Use inequalities $\mathbb{C}[Q_1, Q_\ell] \leq \mathbb{V}[Q_1]\mathbb{V}[Q_\ell]$, $\mathbb{V}[\tilde{Q}_\ell] \geq \mathbb{V}[Q_\ell]$

Numerical examples

Roadmap

1. Explore optimal estimator variance vs. number of replicas for two model case (noisy polynomials)
2. Same as 1, but tweak noise levels
3. Extend polynomial test case to four models
4. Explore pilot reliability

Two model example

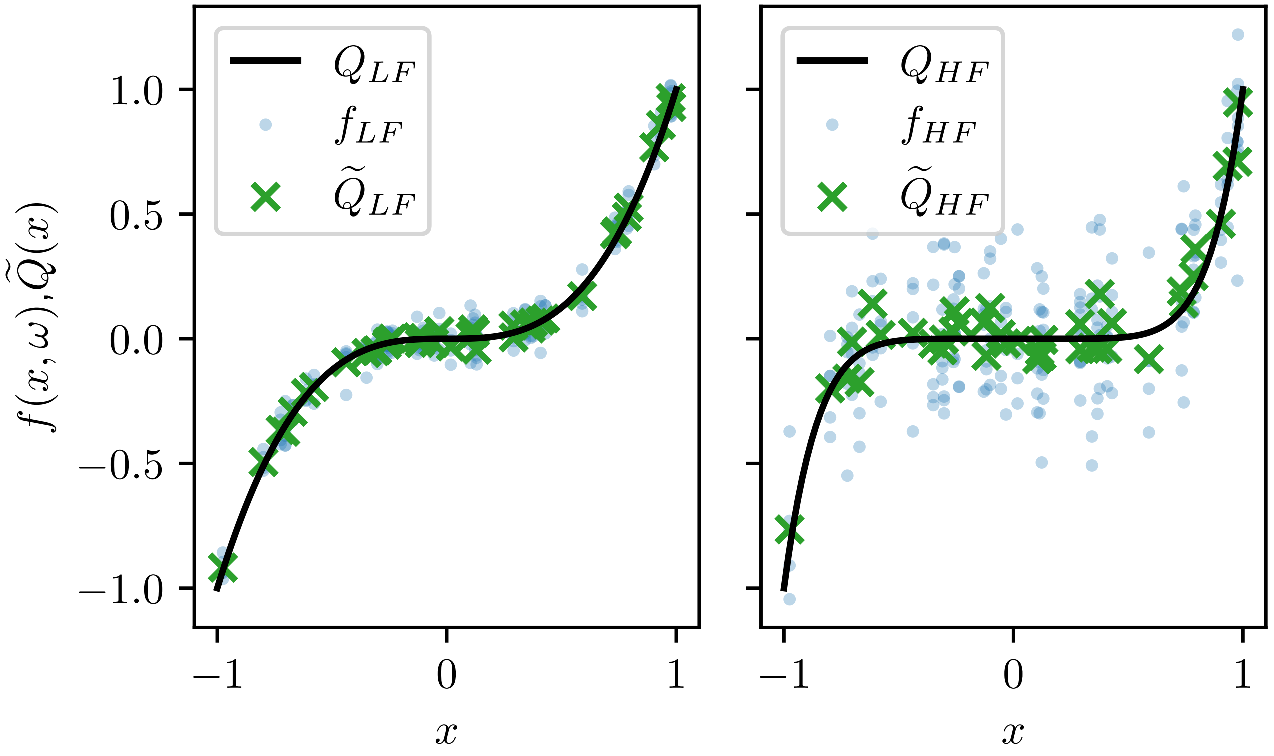
Consider a two model set of noisy monomials:

$$f_{HF} \equiv f_1 = x^7 + \omega_1 \qquad f_{LF} \equiv f_2 = x^3 + \omega_3$$

$$\text{with } x \sim U(-1,1) \qquad \omega_i \sim \mathcal{N}(0, \sigma_i)$$

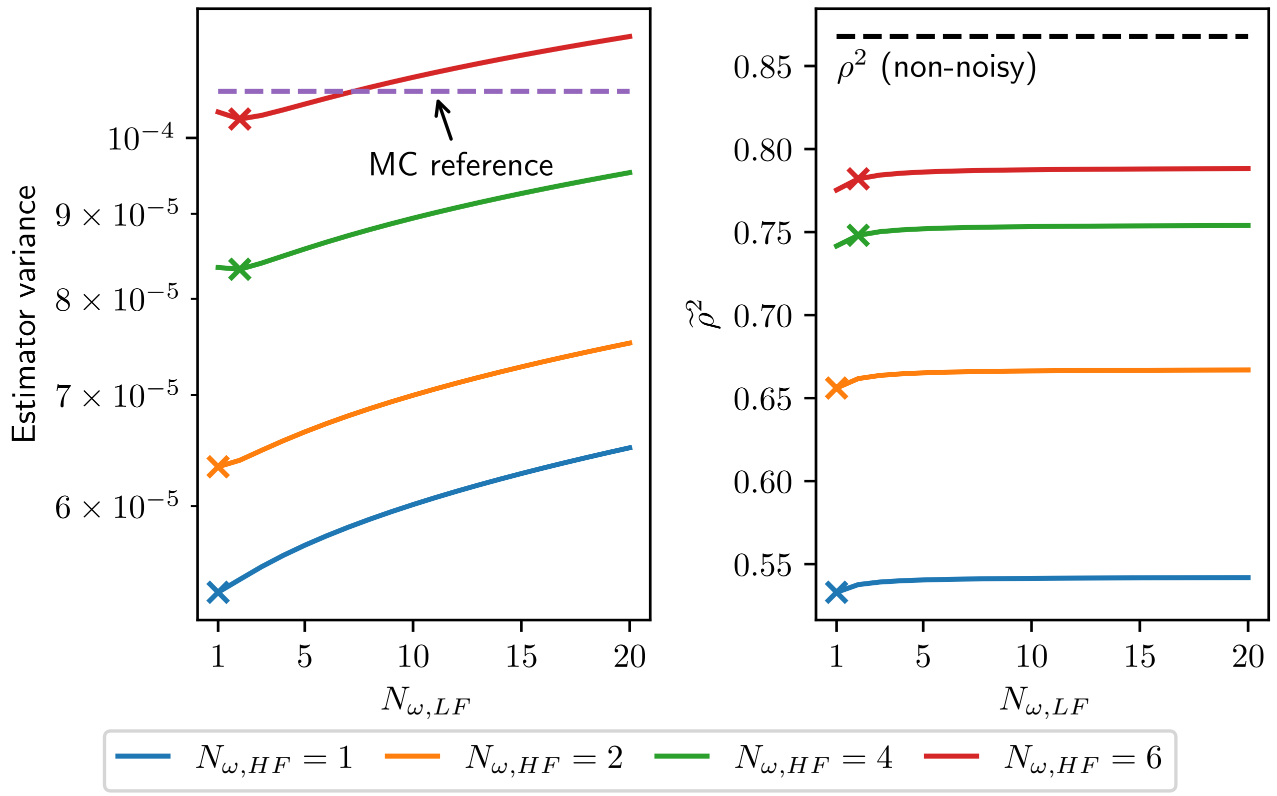
We can characterize two model sets by

1. The cost ratio: C_{LF}/C_{HF}
2. The “non-noisy” model correlation: ρ
3. The ”non-noisy” variances: $\mathbb{V}[Q_\ell]$
4. The stochastic noise levels: $\mathbb{E}[\sigma^2_{\omega,\ell}]$



Lower noise in LF model

No benefit to stochastic averaging, still performant despite fairly sub-optimal correlation.



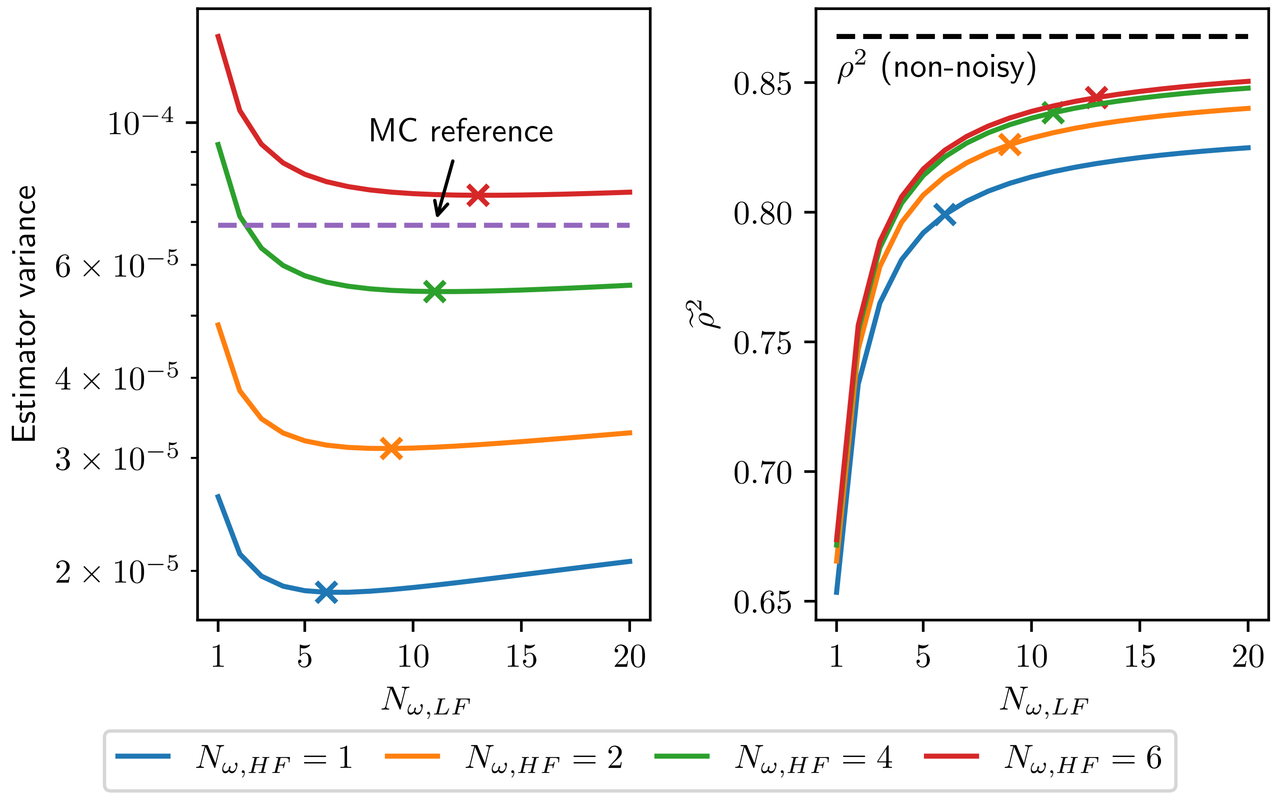
Fixed budget (1000 HF evals)

$$C_{LF}/C_{HF} = .001$$
$$\sigma_{HF} = .2$$
$$\sigma_{LF} = .05$$

MFMC result

Higher noise in LF model

Replication is beneficial, “noisy” correlation can be increased significantly for little cost.



Fixed budget (1000 HF evals)

$$C_{LF}/C_{HF} = .001$$
$$\sigma_{HF} = .05$$
$$\sigma_{LF} = .2$$

MFMC result

Many model example

Consider a four model set of noisy monomials:

$$f_1 = x^7 + \omega_1 \qquad f_2 = x^5 + \omega_2 \qquad f_3 = x^3 + \omega_3 \qquad f_4 = x + \omega_4$$

with $x \sim U(-1,1) \qquad \omega_i \sim \mathcal{N}(0, \sigma_i)$

We can characterize many model sets by

1. The cost ratios: C_ℓ / C_1
2. The "non-noisy" model correlations: $\rho_{1,\ell}$
3. The "non-noisy" variances: $\mathbb{V}[Q_\ell]$
4. The stochastic noise levels: $\mathbb{E}[\sigma^2_{\omega,\ell}]$

Here we have $\rho^2_{1,\ell} = \{1, .988, .932, .745\}$
and fix $C_\ell / C_1 = \{1, .1, .01, .001\}$
$\sigma_\ell = \{.2, .2, .2, .2\}$

Many model problem

- Optimal solution found by enumerating over potential model orderings/subsets and number of replicas
  - Minimal estimator variance for fixed budget
- Model set search space includes sets of size 2, 3, and 4
  - HF model (f_1) always included
- HF replicas fixed to be 1

Optimal solution for MFMC is to use models (1,3,4)

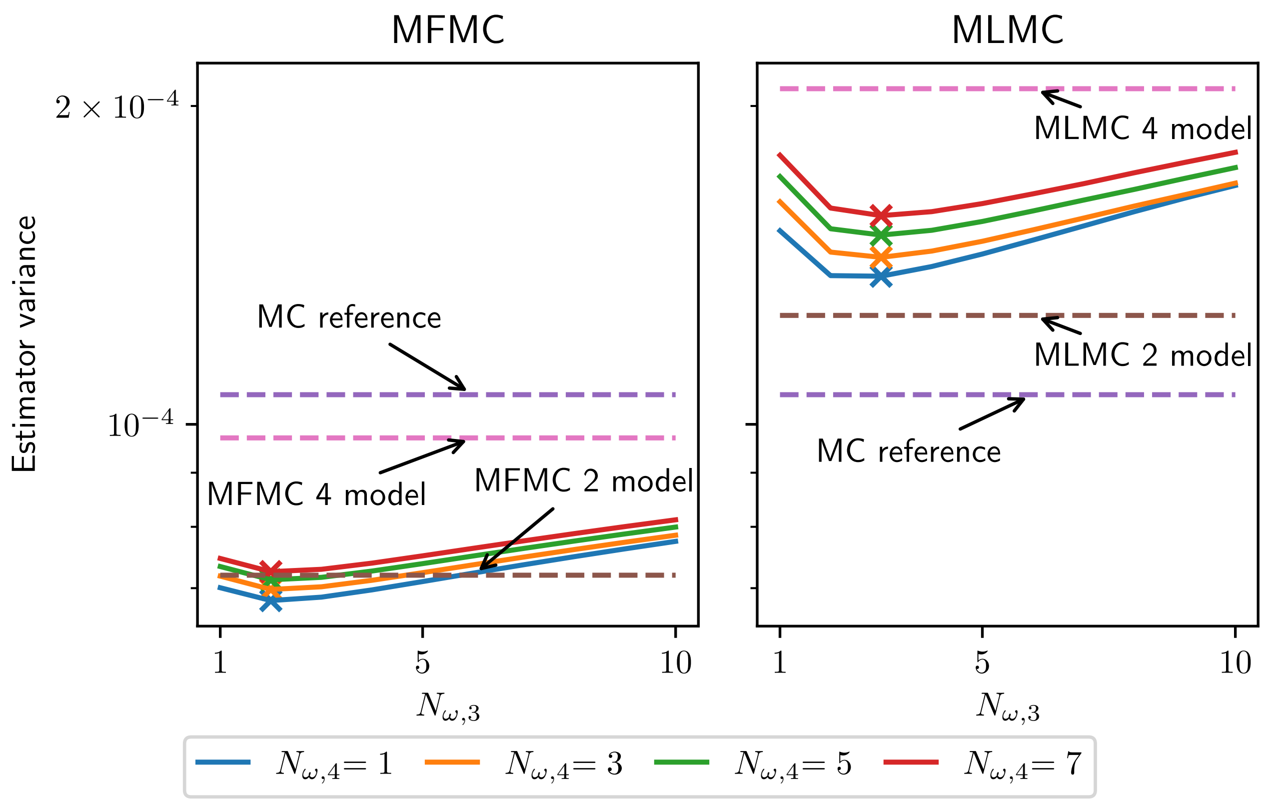
Optimal for MLMC is to use models (1,3)

Equal noise in all models (4 model set)

Using three models (1,3,4) is optimal for MFMC.

Despite relatively low final “noisy” correlations, MFMC beats MC.

MLMC cannot.

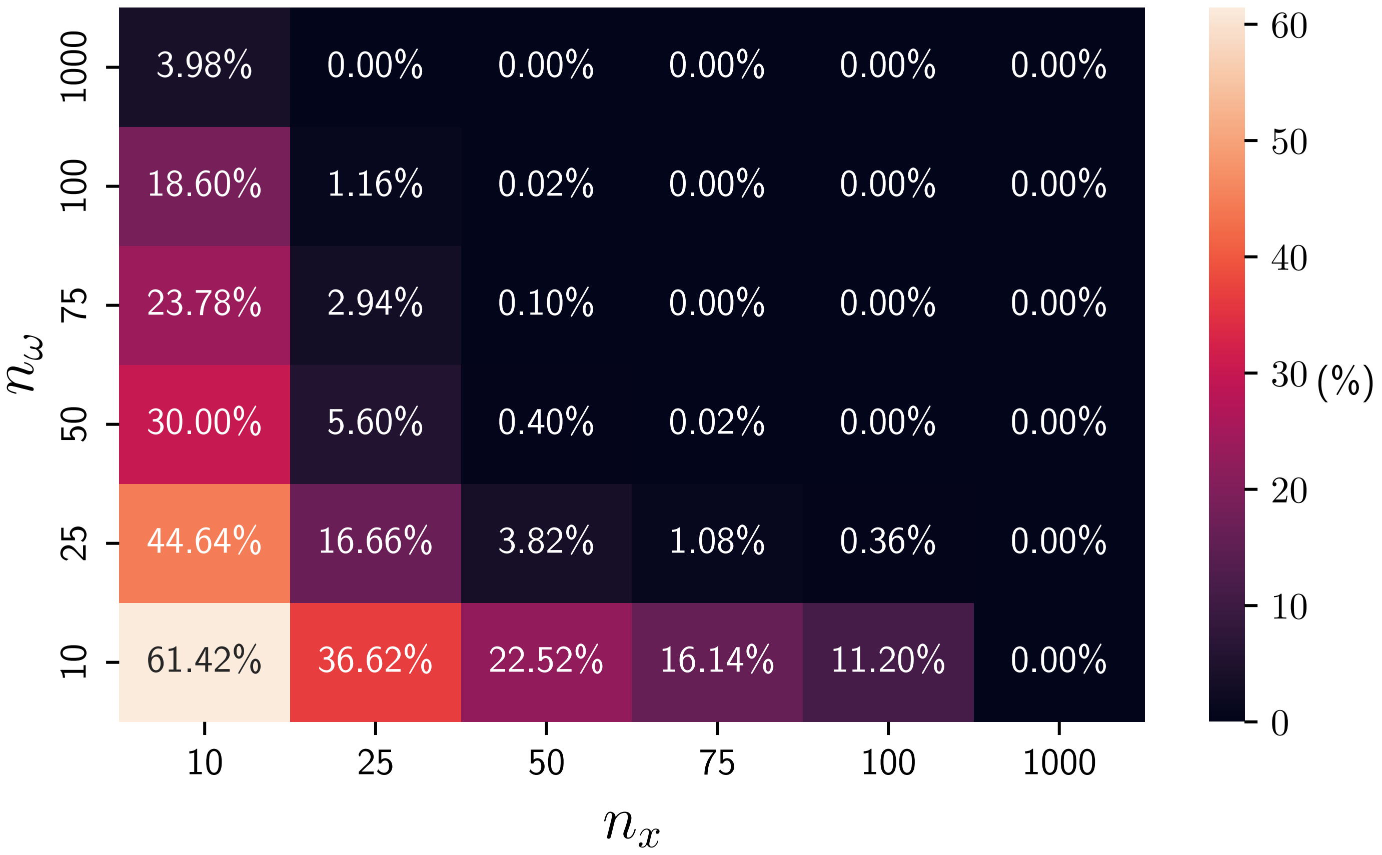


Fixed budget (1000 HF evals)

$$C_\ell / C_1 = \{1, .1, .01, .001\}$$
$$\sigma_\ell = \{.2, .2, .2, .2\}$$

$N_{\omega,1} = 1$ is fixed

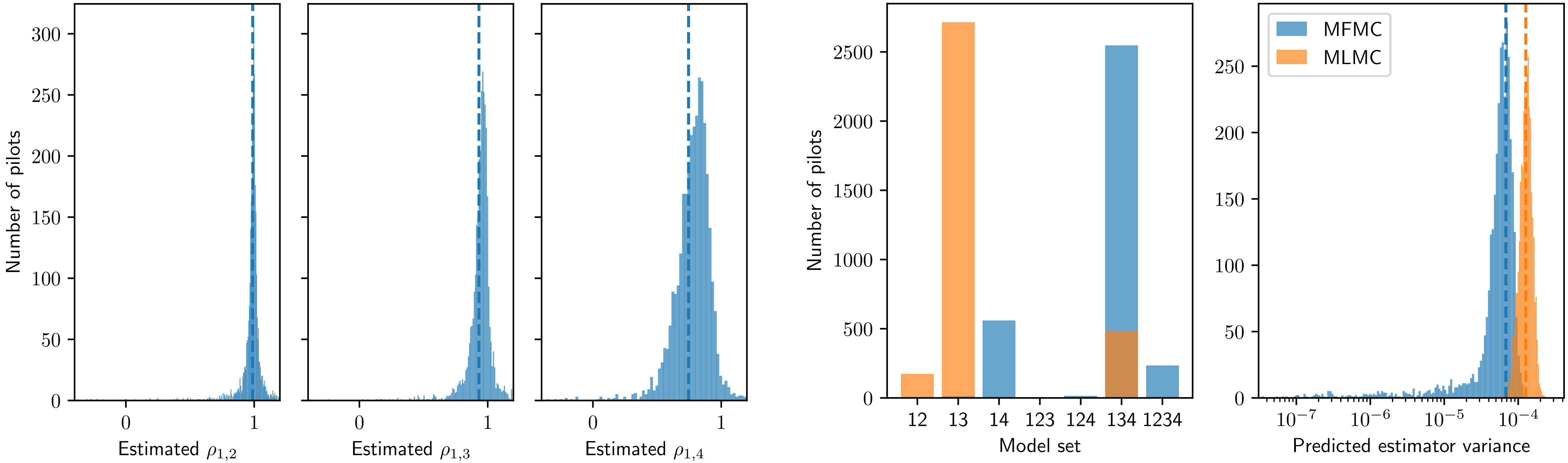
Under-resolved pilots can return negative parametric variances or non-realizable correlations



Pilot study failures

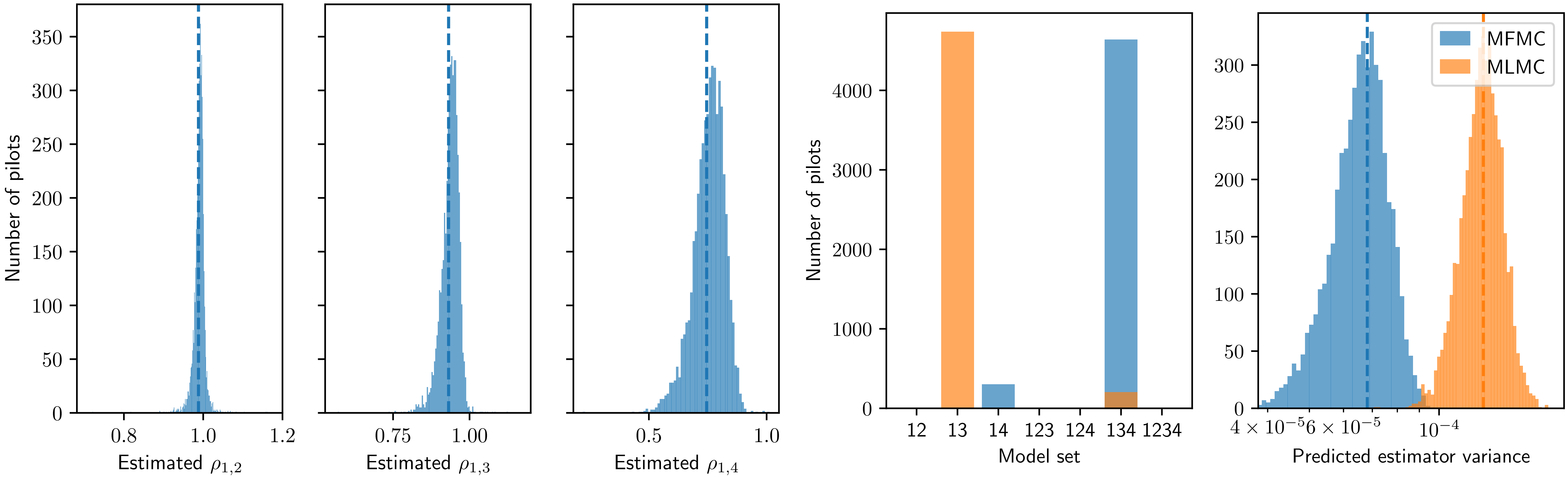
$$\sigma_\ell = \{.2, .2, .2, .2\}$$

Small pilot with few replicas can lead to unrealistic correlations, incorrect model selection, and errors in variance prediction



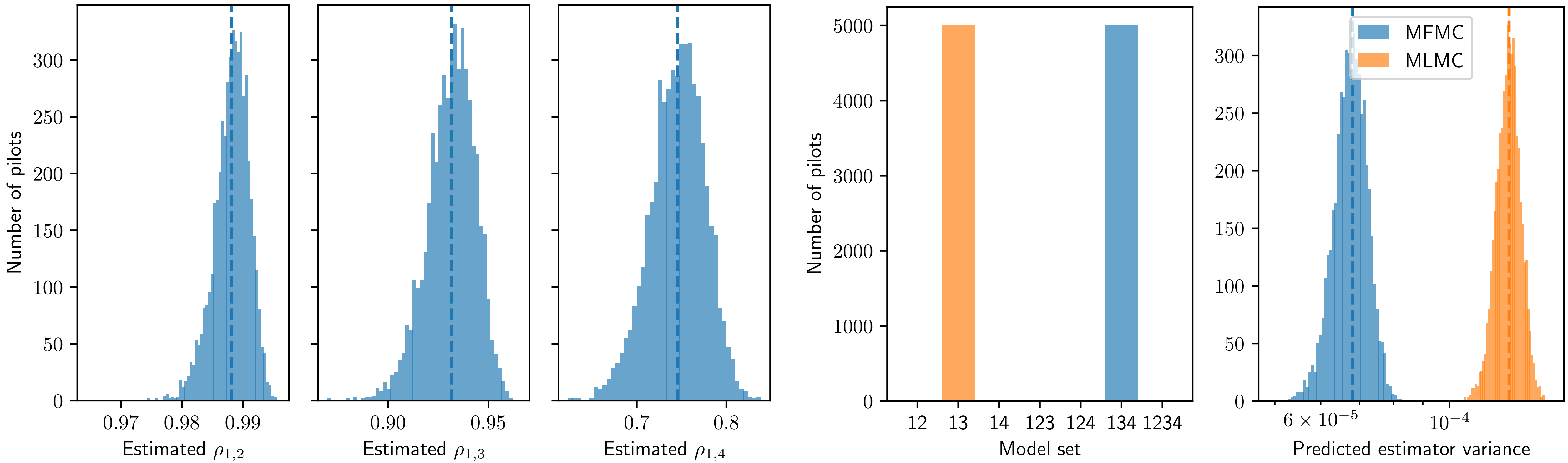
$$n_x = 10 \quad n_\omega = 10$$

Statistics improve with more samples and replicas



$n_x = 25 \quad n_\omega = 25$

Convergence to optimal model set, but pilot cost quite large



$$n_x = 100 \quad n_\omega = 100$$

Conclusions and future work

- Multifidelity approaches can provide variance reduction for mean estimation of QoIs from non-deterministic model sets
- This is a specific example of model tuning for multifidelity UQ
- Some practical concerns of note
  - estimating statistics from pilot
  - model ordering and the cost/correlation constraints of the analytical MFMC solution
  - the minimum number of replicas may be constrained by other analyses
- Extension to ACV approaches [Gorodetsky, 2020]
- Computation of higher order statistics/tail probabilities
- PDE-based examples

References

[Geraci, 2021]: Geraci, et al. “Multifidelity UQ Sampling for Stochastic Simulations.” *USNCCM* conference presentation (2021)

[Clements, 2021]: Clements, et al. “A Variance Deconvolution Approach to Sampling Uncertainty Quantification for Monte Carlo Radiation Transport Solvers.” *CSRI Summer Proceedings 2021* (2021)

[Gorodetsky, 2020]: Gorodetsky, Alex, et al. "A generalized approximate control variate framework for multifidelity uncertainty quantification." *Journal of Computational Physics* (2020)

[Peherstorfer, 2016]: Peherstorfer, Benjamin, et al. "Optimal model management for multifidelity Monte Carlo estimation." *SIAM Journal on Scientific Computing* (2016)

[Giles, 2008]: Giles, Michael "Multilevel Monte Carlo path simulation." *Operations Research* (2008)

[Oliver, 2021]: Oliver, Todd, et al. "Extensions to Multifidelity Monte Carlo Methods for Simulations of Chaotic Systems." *arXiv preprint arXiv:2106.13844* (2021)