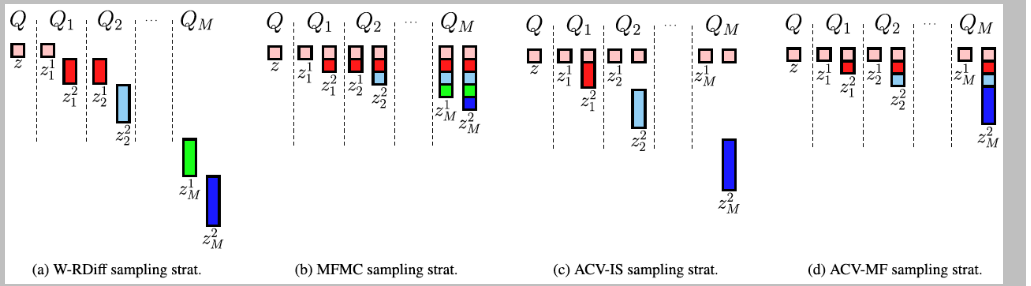
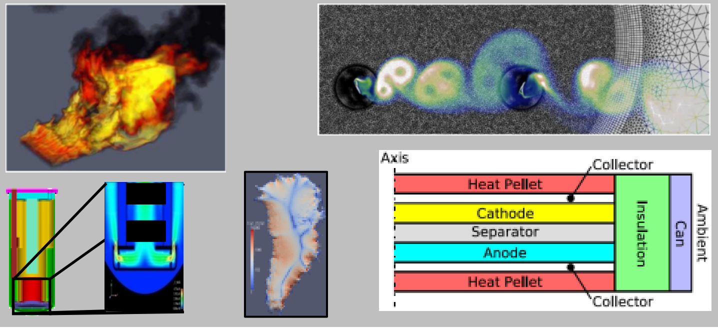


DAKOTA

Explore and predict with confidence



Recent Progress in Model Ensemble Configuration for Multifidelity UQ

Michael S. Eldred[1], John Jakeman[1], Alex Gorodetsky[2], Gianluca Geraci[1]

[1]*Optimization & Uncertainty Quantification Dept, Center for Computing Research, Sandia National Laboratories, Albuquerque NM*

[2]*Aerospace Engineering Department, University of Michigan, Ann Arbor MI*

Multilevel / Multifidelity Estimators based on MC Sampling

Sandia National Laboratories

Estimator	Type	Sample allocation
MLMC	1D: hierarchical, recursive	Analytic
MFMC	1D: hierarchical, recursive	Analytic, Numerical
MLMF MC	2D: HF,LF pair + resolutions	Analytic
ACV	Non-recursive / peer: all CV pairings target root	Numerical
Gen. ACV	Search over approx sets & DAGs (MFMC + ACV + intermediate)	Numerical
ML BLUE	Model groupings	Numerical
Group ACV	Relax BLUE independence, solve linearly constrained opt	Numerical

1D/2D hierarchical (MLMC, MFMC, MLMF MC)

non-hierarchical (ACV, Gen. ACV, ML BLUE, Group ACV)

Motivation: production deployments of ML/MF methods encounter a variety of challenges that can impede performance

- Accurate a priori / offline estimations of Covar[Q] are often impractical, and should rather be integrated and optimized
  → *iterated pilot approaches including relaxation*
- LF models often have parameters that trade accuracy vs. cost (set via SME judgment, but intuition often inaccurate in this context)
  → *hyper-parameter model tuning*
- Numerical solutions often suffer from multiple minima, desirable to drop model allocations while retaining conditioning
  → *robust numerical solves*: coarse global search, multi-start competed local from {global,analytic}, SDP
- For general model ensembles, the best approximation selections and CV pairings/groupings are not known a priori
  → *ensemble selection and configuration*

Each of these concerns can introduce additional iteration or expand the scale of an integrated optimization

Giles 2008, Peherstorfer et al. 2016, Geraci et al. 2017, Gorodetsky et al. 2020, Bomarito et al. 2022, Schaden and Ullmann 2020, Gorodetsky et al. 2024

Ensemble Configuration in Multifidelity Sampling

Model-tuning $\arg\min_{\theta} \left[\arg\min_{\mathbf{r},N} \frac{Var[Q]}{N}(1 - R^2(\theta, \mathbf{r})) \quad s.t. \quad N\left(w + \sum_{i=1}^{M} w_i(\theta) r_i\right) \leq C \right]$

Covariance refinement: ΔN_{shared} → Cov[Q]

Not swappable without N_{shared} re-eval.
Refinement is deferable:
tuning often projection-based

Ensemble selection: best subset (drop low value approx.)

CV selection (enumerate/optimize DAG pairings)

Separated for numerical reasons

$$\arg\min_{\mathbf{r},N} \frac{Var[Q]}{N}(1 - R^2(\theta, \mathbf{r})) \quad s.t. \quad N\left(w + \sum_{i=1}^{M} w_i(\theta) r_i\right) \leq C$$

Resource Allocation: cost ⟷ accuracy

Fixed covariance (GenACV w/ MFMC, ACV as special cases)

Variable covariance: converge on N_{shared}, tune w.r.t. θ

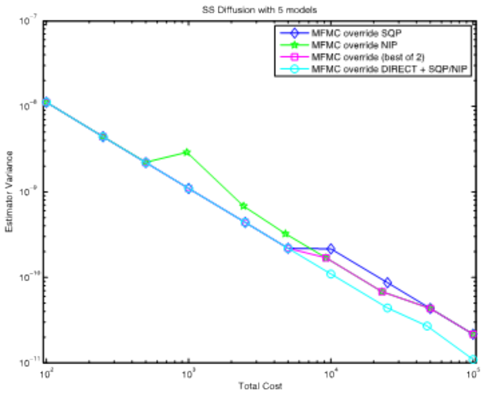
Review of previous work: Outer | Inner

Iterated Pilots → integrate pilot as *online cost;* optimize total

- Iterate shared $N^{(i)}$ for estimation of Covar[Q] across models

Initialize: select small shared pilot $N^{(0)}$ to under-shoot optimal profile
1) Sample all models
2) $N^{(i)}$ shared samples → $Cov_{LL}^{(i)}$, $Cov_{LH}^{(i)}$ → opt. solver → r^*, N^*
3) Compute one-sided ΔN for shared samples from $N^{(i)}$ to N^*
   A. Optional: apply under-relaxation factor γ
   B. If non-zero increment, advance (i) and return to 1)

- Avoid inefficiency (over-est.) or inaccuracy (under-est.).



Harden numerical solutions

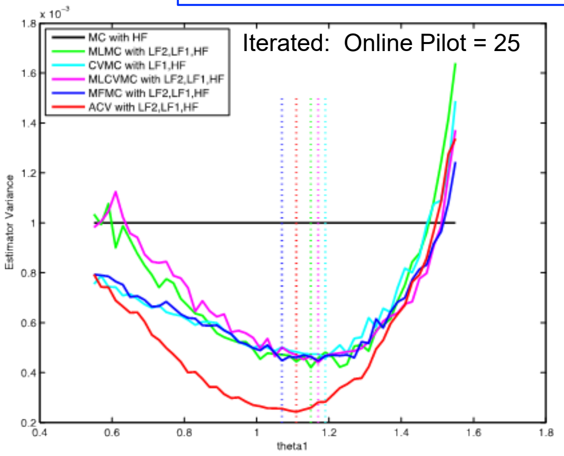
→ mitigate multi-modality
- Global search to identify promising regions:
  - SBGO, EA, EGO, DIRECT
- Competed NLP for local refinement:
  - SQP (via NPSOL), NIP (via OPT++)

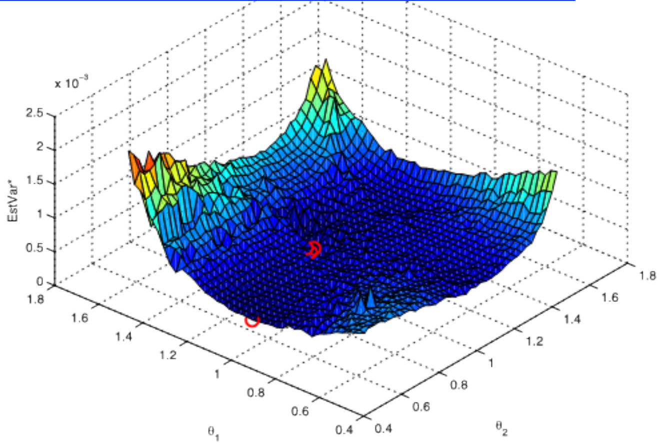
→ mitigate conditioning
- SDP to support indefinite solves from model removal

Hyper-Parameter Model Tuning

Tune approx to identify best accuracy vs. cost trade-off

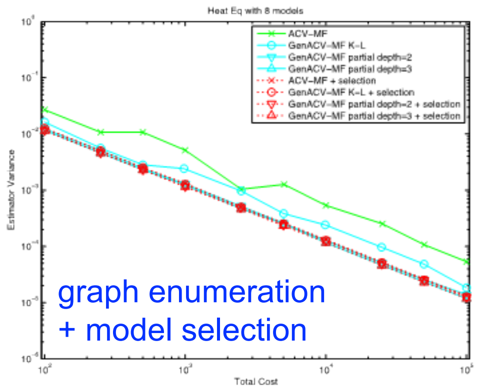
$$\arg\min_{\theta} \left[\arg\min_{\mathbf{r},N} \frac{Var[Q]}{N}(1 - R^2(\theta, \mathbf{r})) \quad s.t. \quad N\left(w + \sum_{i=1}^{M} w_i(\theta) r_i\right) \le C \right]$$





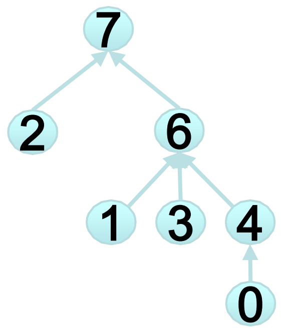
Ensemble selection / pairings:

Identify most performant approx: membership / relationship



Combinatorial growth in alternatives:
- For 1 set of 8 models, # DAG = 214,720
- With model selection, # DAG = 350,870

Mitigate using DAG depth throttles:
- K-L (ordered depth = 2): 22 or 800 w/ selection
- Depth = 2: 6,323 or 19,693 w/ selection
- Depth = 3: 48,260 or 105,870 w/ selection



Multilevel Best Linear Unbiased Estimator (ML BLUE)

Group-based estimator:

- Enumeration of model combinations via groups
- Allocate shared group samples (no profile in group)
- Group sample sets are independent

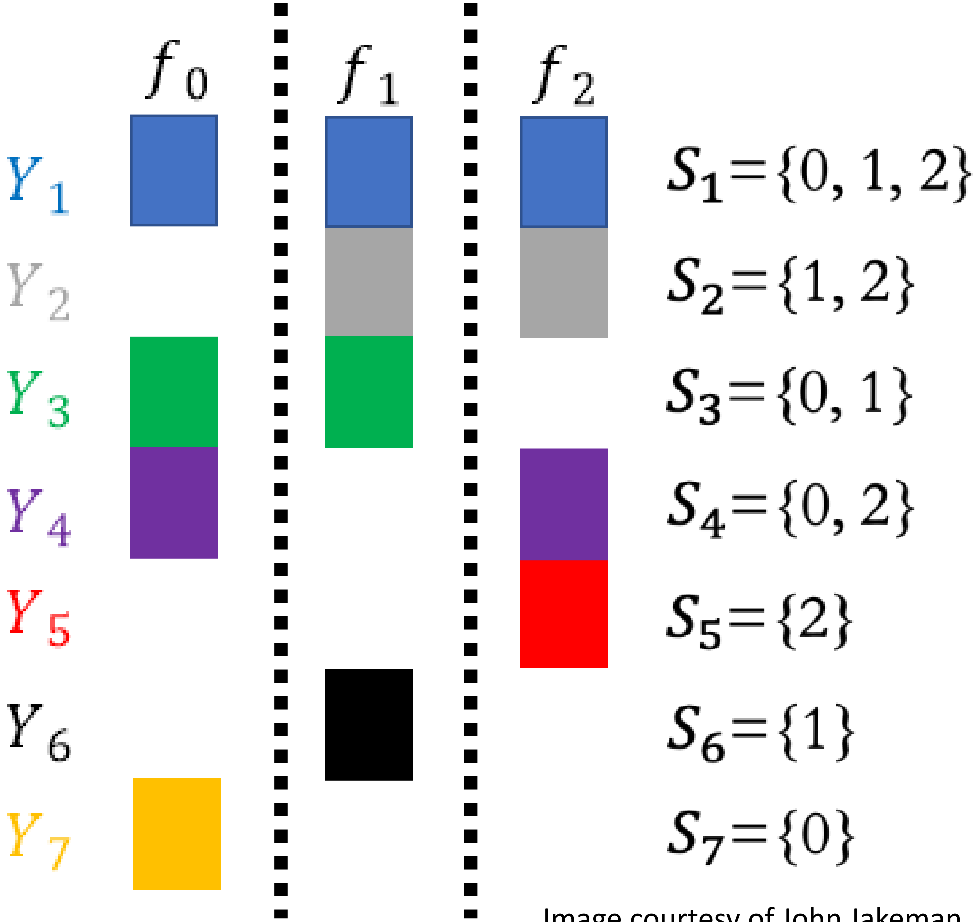


Image courtesy of John Jakeman

D Schaden, E Ullmann. On multilevel best linear unbiased estimators
SIAM/ASA J. Uncertainty Quantification 8 (2), 601 - 635

The ML BLUE estimator is $\Psi q^B = y$

$$\Psi = \sum_{k=1}^{K} m_k R_k^\top C_k^{-1} R_k \qquad y = \sum_{k=1}^{K} R_k^\top C_k^{-1} \sum_{i=1}^{m_k} Y_k^{(i)}$$

- For linear combination of the model means $q_\beta^B = \beta^\top q^B$, variance of ML BLUE is $\mathbb{V}[q_\beta^B] = \beta^\top \Psi^{-1} \beta$.
- Numerical soln for m minimizes this variance s.t. budget, where $\beta = [1, 0, 0, \ldots]^\top$ targets the HF mean.

Features of Dakota implementation

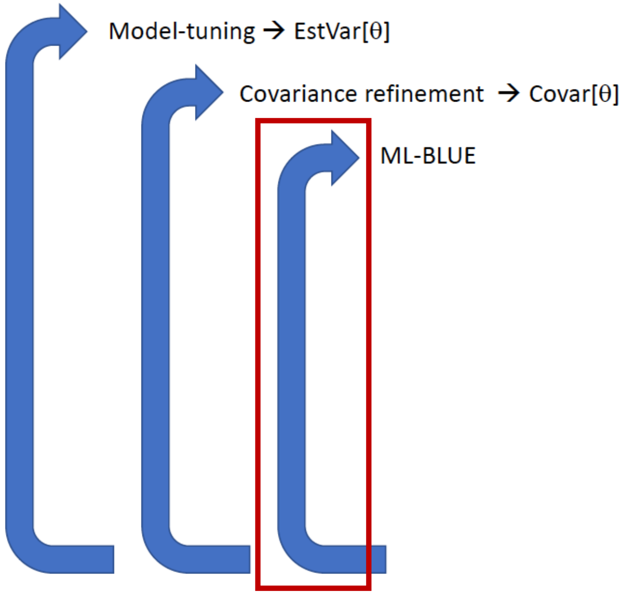
- Solution modes: {online,offline} x {perf projection,final stats}
- Shared vs. independent group pilot sampling
- Under-relaxation factors: fixed, recursive, sequence
- Group throttles: size (“SAOB,k”), MFMC (“FC,k”), common (MF, ML, pair CV)
- Online cost recovery, hyper-parameter control

Current limitations of Dakota implementation

- Conditioning mitigations limited to equilibration + iterative refinement
- Solvers limited to existing global + competed local (DIRECT + SQP/NIP)

Emerging extension: Group ACV https://arxiv.org/abs/2402.14736

Multilevel Best Linear Unbiased Estimator (ML BLUE)



Exploration of ensemble configuration with ML BLUE:

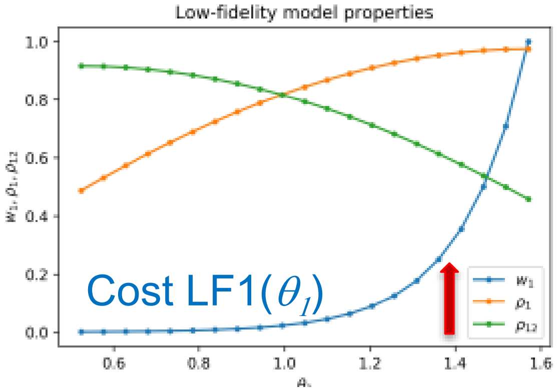
1. If matrix conditioning can be mitigated, ensemble config is embedded
2. Iterated covariance approaches take on new requirements → under-relaxation
   - Shared vs. independent pilot (if something interesting here)
   - Under-relaxation
3. Performance under model tuning → continuation of ACV/GenACV robustness trends?

ML BLUE performance on 3 standard test problems is observed to be linked to model and sample counts

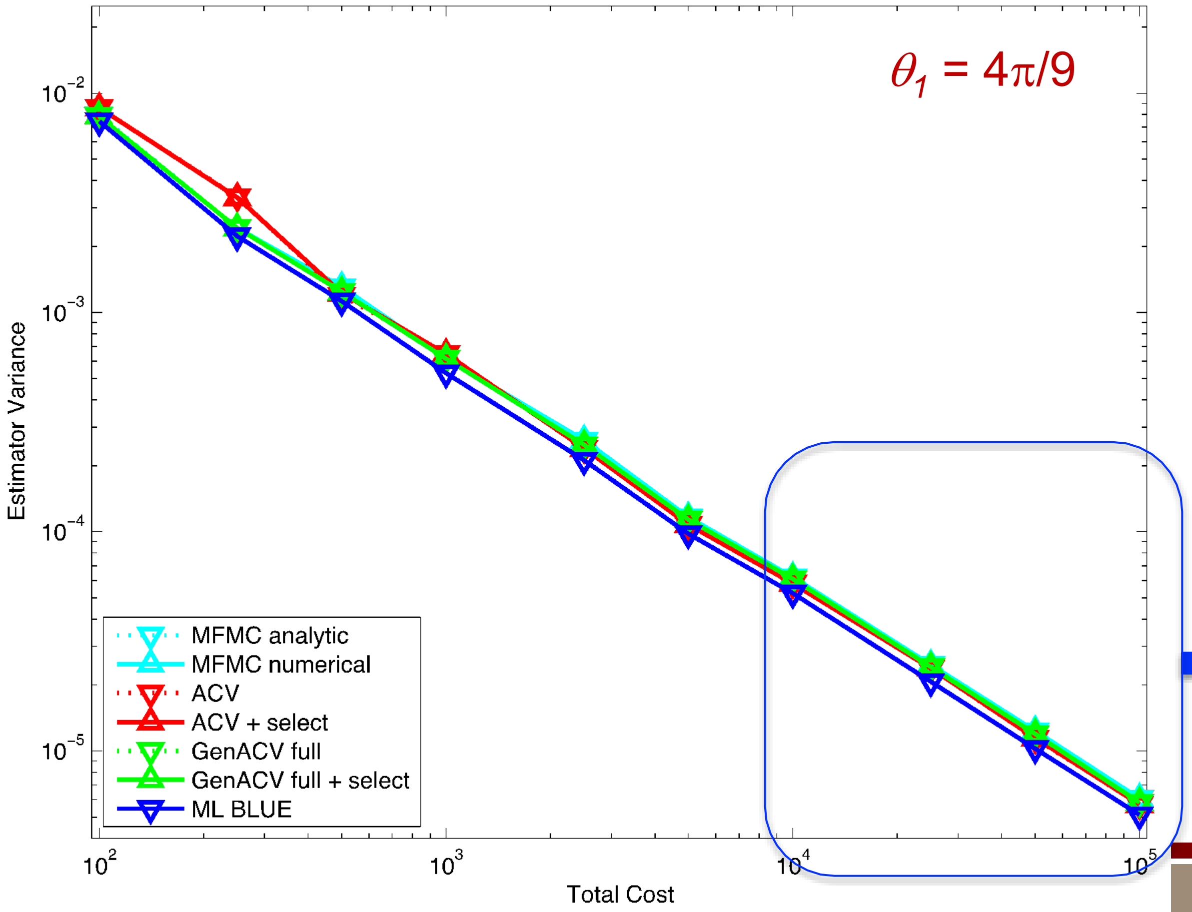
- *Tunable problem w/ 3 models with hyper-parameters:* excellent performance
- *Steady state diffusion w/ 5 resolutions*: conditioning is degrading for more aggressive sample distributions within large total budgets → group throttling required
- *Transient diffusion (heat eq) with 8 total models (2 MF x 4 RL):* solutions become unreliable without aggressive model pruning / group throttling

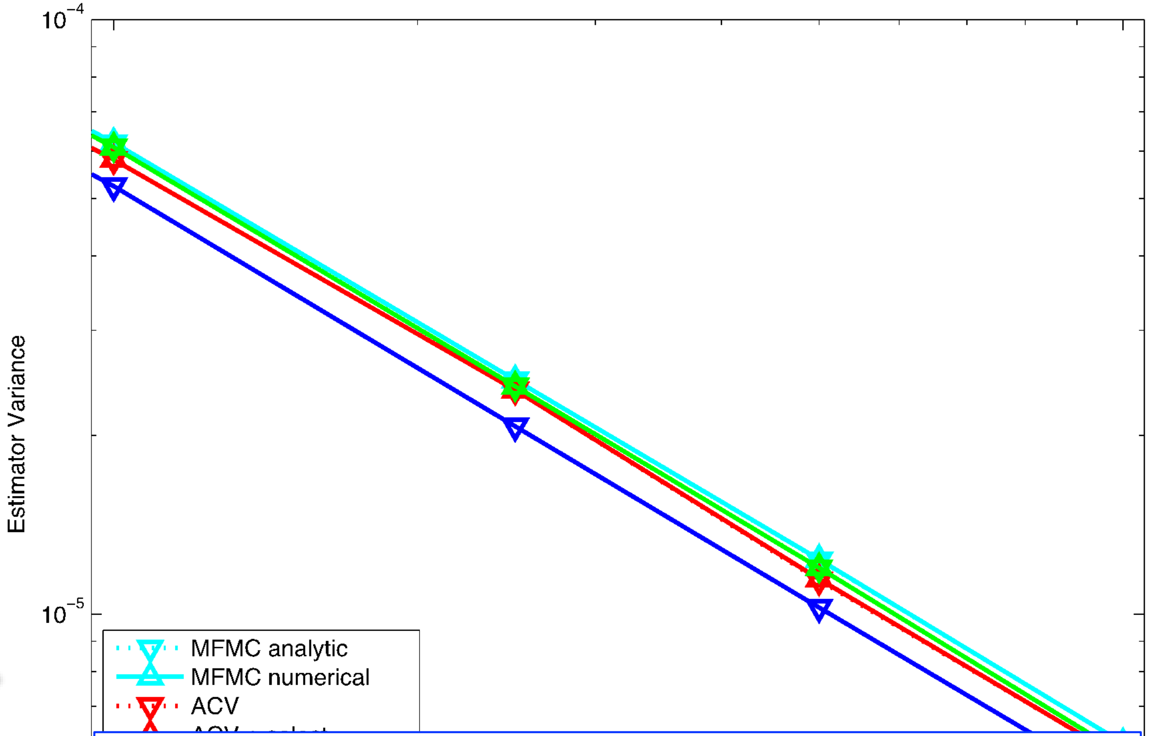
Tunable Model test problem (JCP 2020)

$$Q(\theta) = \sqrt{11}\ [\ \cos(\theta)\ x^5 + \sin(\theta)\ y^5\ ]$$
$$Q_1(\theta_1) = \sqrt{7}\ [\ \cos(\theta_1)\ x^3 + \sin(\theta_1)\ y^3\ ]$$
$$Q_2(\theta_2) = \sqrt{3}\ [\ \cos(\theta_2)\ x + \sin(\theta_2)\ y\ ]$$



Later we will tune hyper-params θ_1 and θ_2 *for fixed HF* $\theta = \pi/2$.
For this comparison, we fix $\theta_2 = \pi/6$ and $\theta_1 = 4\pi/9$.





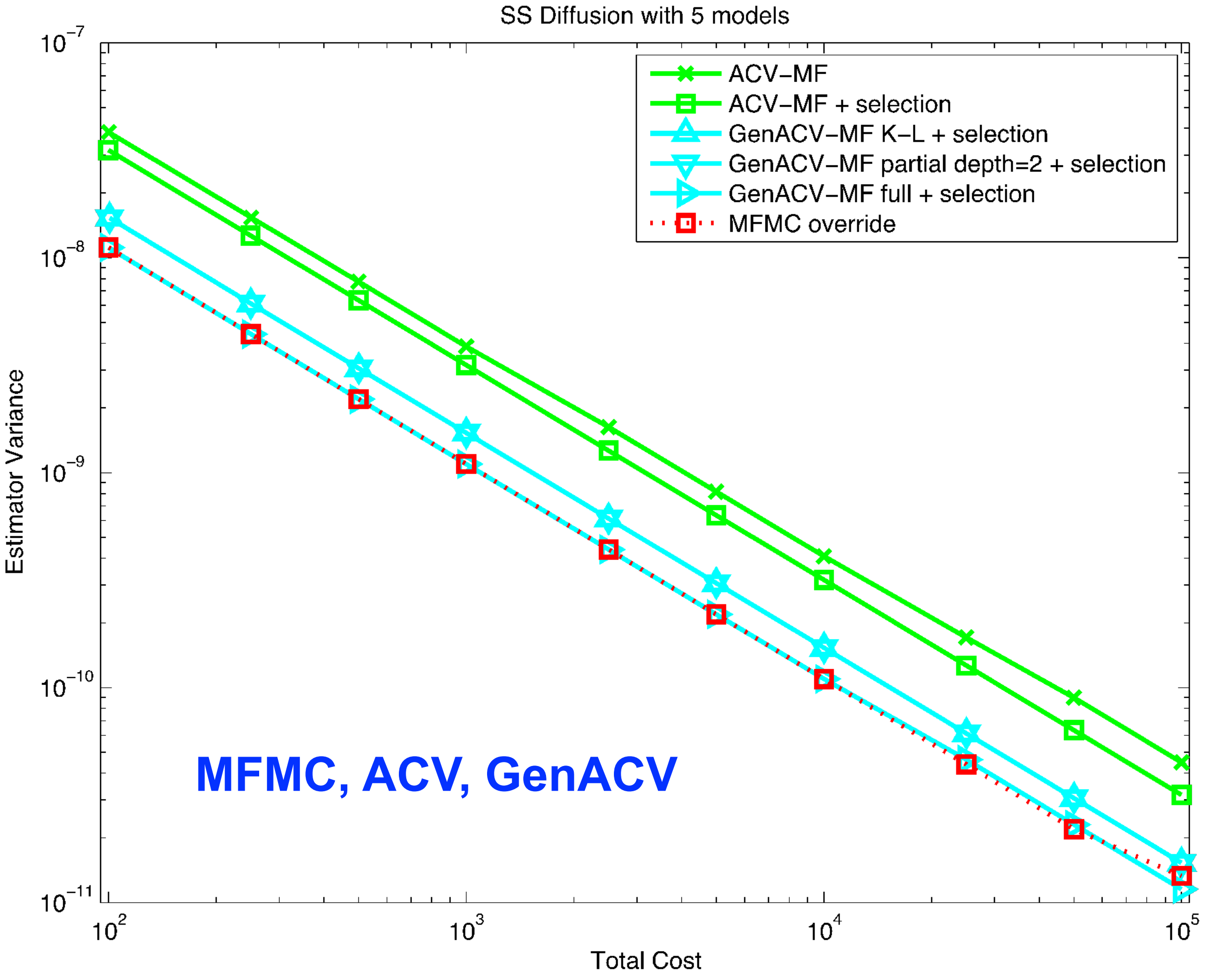
When conditioning issues are contained, ML BLUE has shown increments in performance relative to MFMC, ACV, GenACV

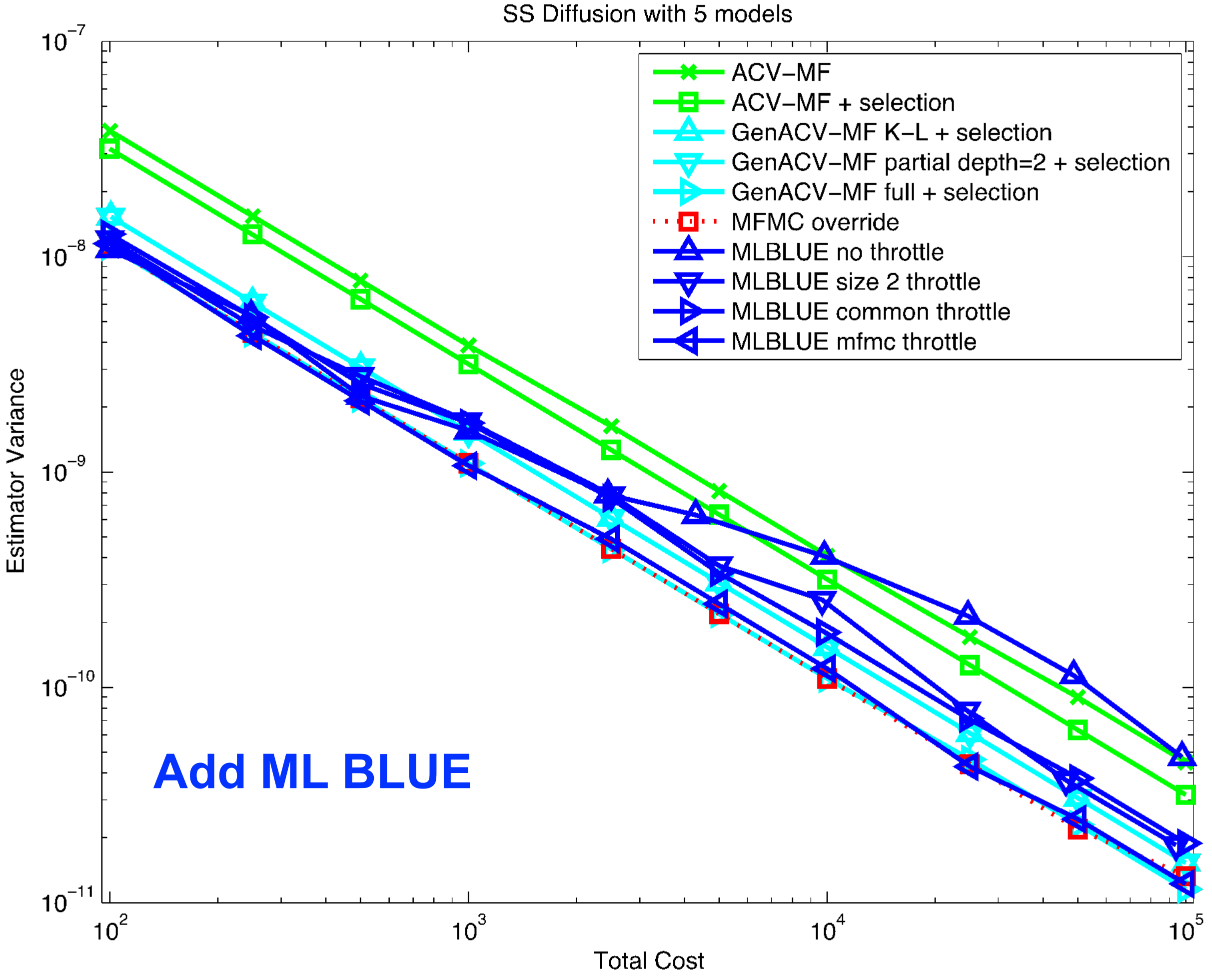
Steady state diffusion test problem

$$-\frac{d}{dx}\left[a(x,\xi)\frac{du}{dx}(x,\xi)\right] = 10, \quad (x,\xi) \in (0,1) \times I_\xi$$

$$u(0,\xi) = 0, \quad u(1,\xi) = 0.$$

Steady state 1D diffusion: 5 models in 1D hierarchy
resolutions = {4,8,16,32,64}, relative cost = {1,4,16,64,256}





ACV peer DAG is poor → GenACV recovers MFMC at full depth

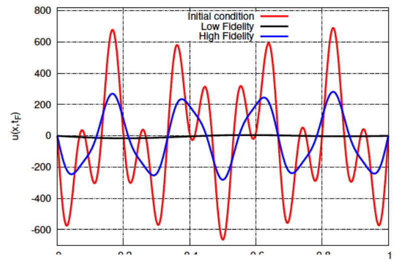
Throttles necessary to mitigate conditioning at higher sample counts

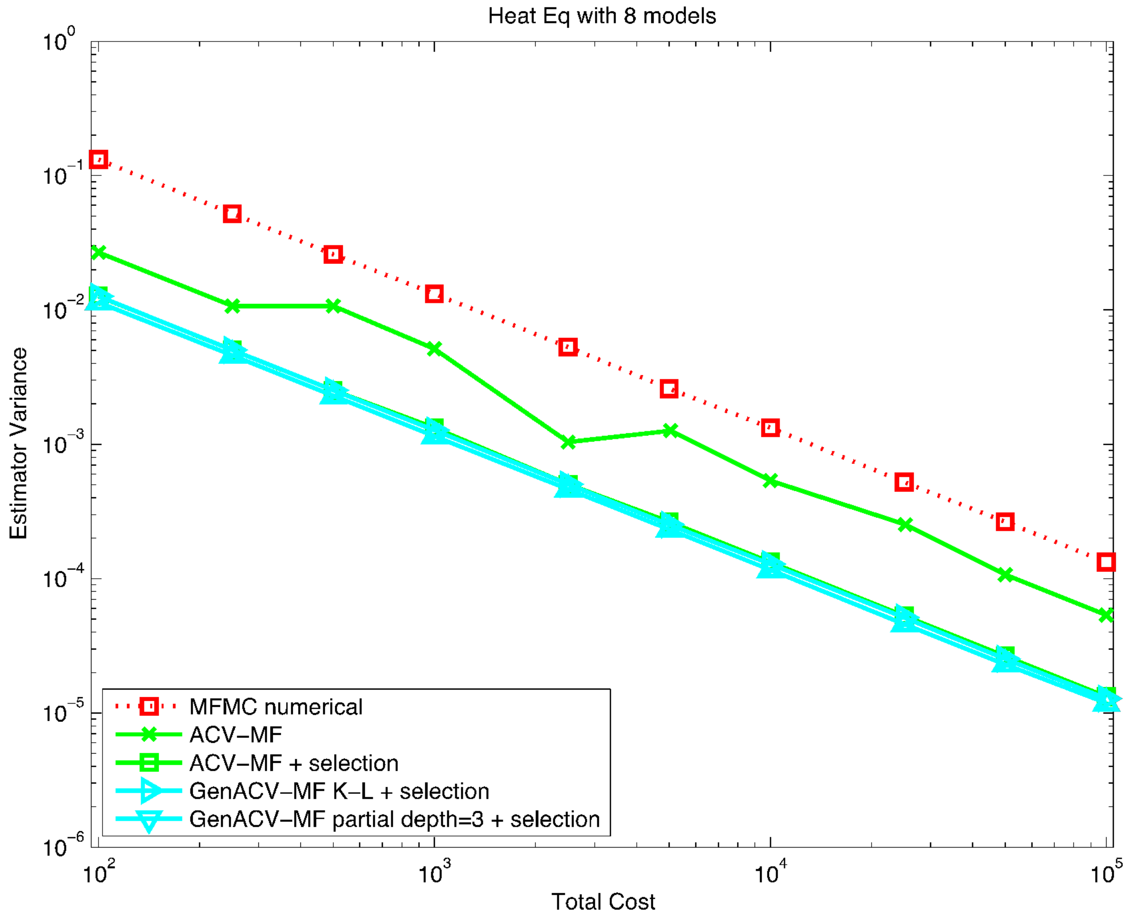
Transient diffusion test problem

1D transient diffusion ("heat equation")

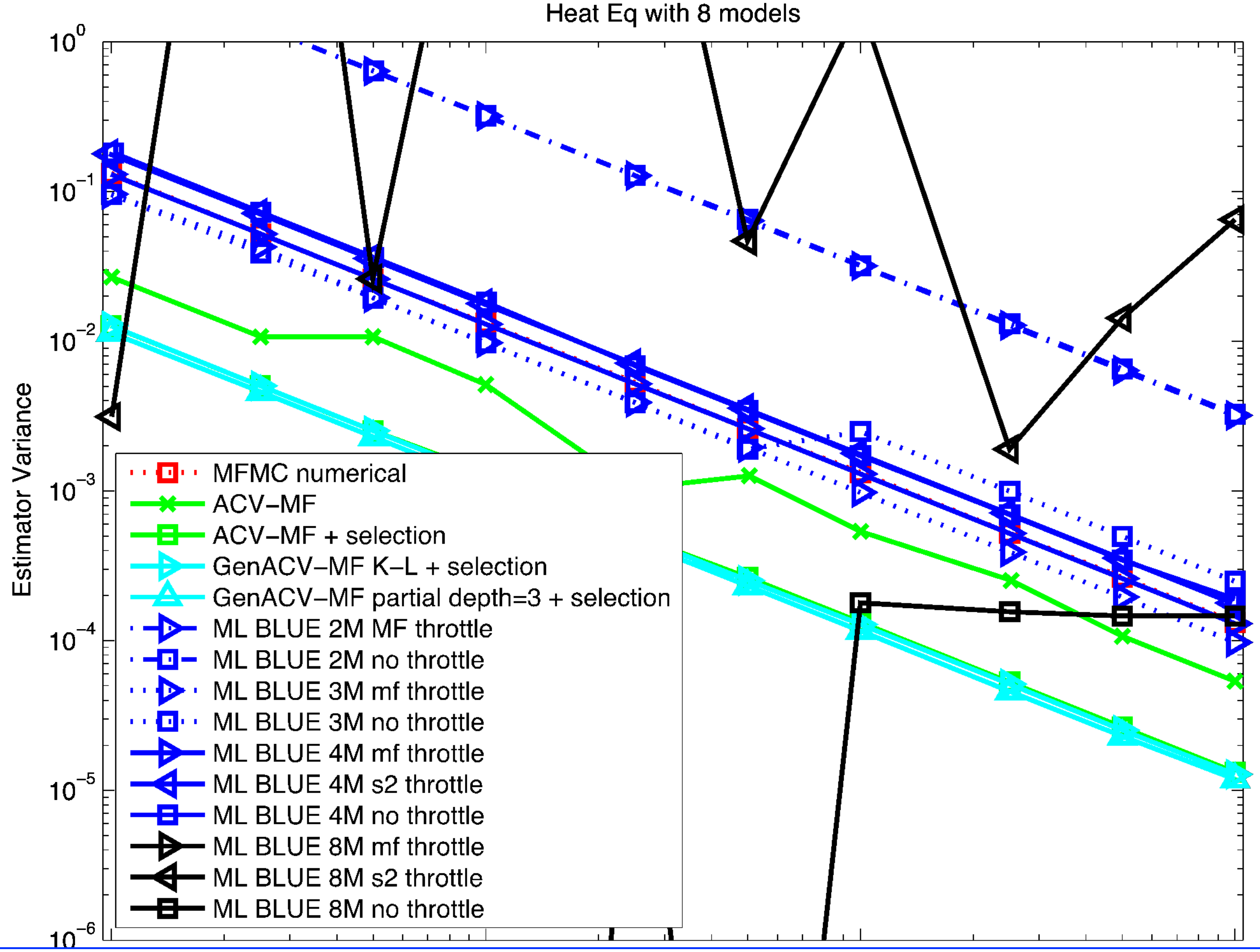
- 8 models in 2D hierarchy: multifidelity + multilevel
- Fourier solution modes = 3 LF, 21 HF
- Spatial coordinates = {5 15 30 60} LF, {30 60 100 200} HF

$$\begin{cases} \dfrac{\partial u(x,\xi,t)}{\partial t} - \alpha(\xi)\dfrac{\partial^2 u(x,\xi,t)}{\partial x^2} = 0, \quad \alpha > 0,\ x \in [0,L] = \Omega \subset \mathbb{R} \\ u(x,\xi,0) = u_0(x,\xi), \quad t \in [0,t_F] \quad \text{and} \quad \xi \in \Xi \subset \mathbb{R}^d \\ u(x,\xi,t)|_{\partial\Omega} = 0 \\ u_0(x,\xi) = \mathcal{G}(\xi)\mathcal{F}_1(x) + \mathcal{I}(\xi)\mathcal{F}_2(x) \end{cases}$$



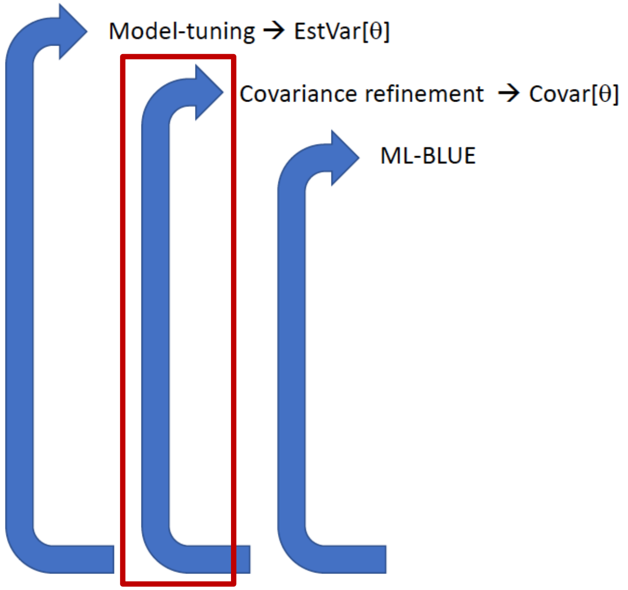


More complex hierarchy benefits significantly from DAG search



Work in progress (truncating eigenvalues from Ψ, SDP formulations)
Expected to unify inner configurations

Multilevel Best Linear Unbiased Estimator (ML BLUE)



Exploration of ensemble configuration with ML BLUE:

1. If matrix conditioning can be mitigated, ensemble config is embedded
2. Iterated covariance approaches take on new requirements
   - Under-relaxation becomes more important due to ordering of budget allocation
   - Shared vs. independent pilot: both can be iterated, w/ greater budget freedom in the former
3. Performance under model tuning → continuation of ACV/GenACV robustness trends?

Iterated ML BLUE

Initialize: select a small initial pilot sample expected to under-shoot the optimal profile
define group a as the group containing *all* models

1) Sample all models for *(i)* group a only (reuse covariances), *(ii)* all $k = 1,\ldots,K$ groups independently
2) Shared covariance iteration (option *(i)* only)
   A. m_a samples → $C_k^{\text{(shared)}}$ → Ψ, y → opt. solver $\min_N \beta^T\Psi^{-1}\beta$ → $\boldsymbol{m}$
   B. Compute one-sided Δm_a and under-relax step (full budget will not be expended, even unrelaxed)
   C. If $\Delta m_a = 0$, stop shared iteration; else perform Δm_a and return to 2A
3) Compute $\Delta \boldsymbol{m}$ for every group (only Δm_a is zero for *(i)*), under-relax step, and perform sample increments
4) Independent iteration (both options *(i)* and *(ii)*)
   A. Group samples → $C_k^{\text{(independent)}}$ → Psi, y → opt. solver → $\boldsymbol{m}$
   B. Compute one-sided $\Delta \boldsymbol{m}$ and under-relax steps (full budget expended on 1st iter. unless step is relaxed)
   C. If $\Delta \boldsymbol{m} = 0$, stop independent iteration; else perform $\Delta \boldsymbol{m}$ and return to 3A

Finalize: solve for μ for statistics of interest (currently moments 1 through 4 for each QoI)

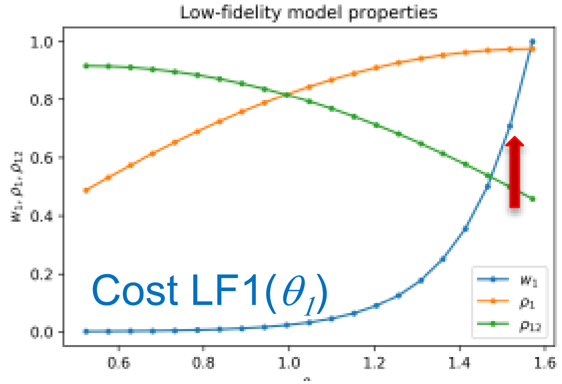
Iterated MFMC / ACV / GenACV converge on group a covariances as in step 2) above, prior to any LF increments

Online pilot integration – under-relaxation for small pilots

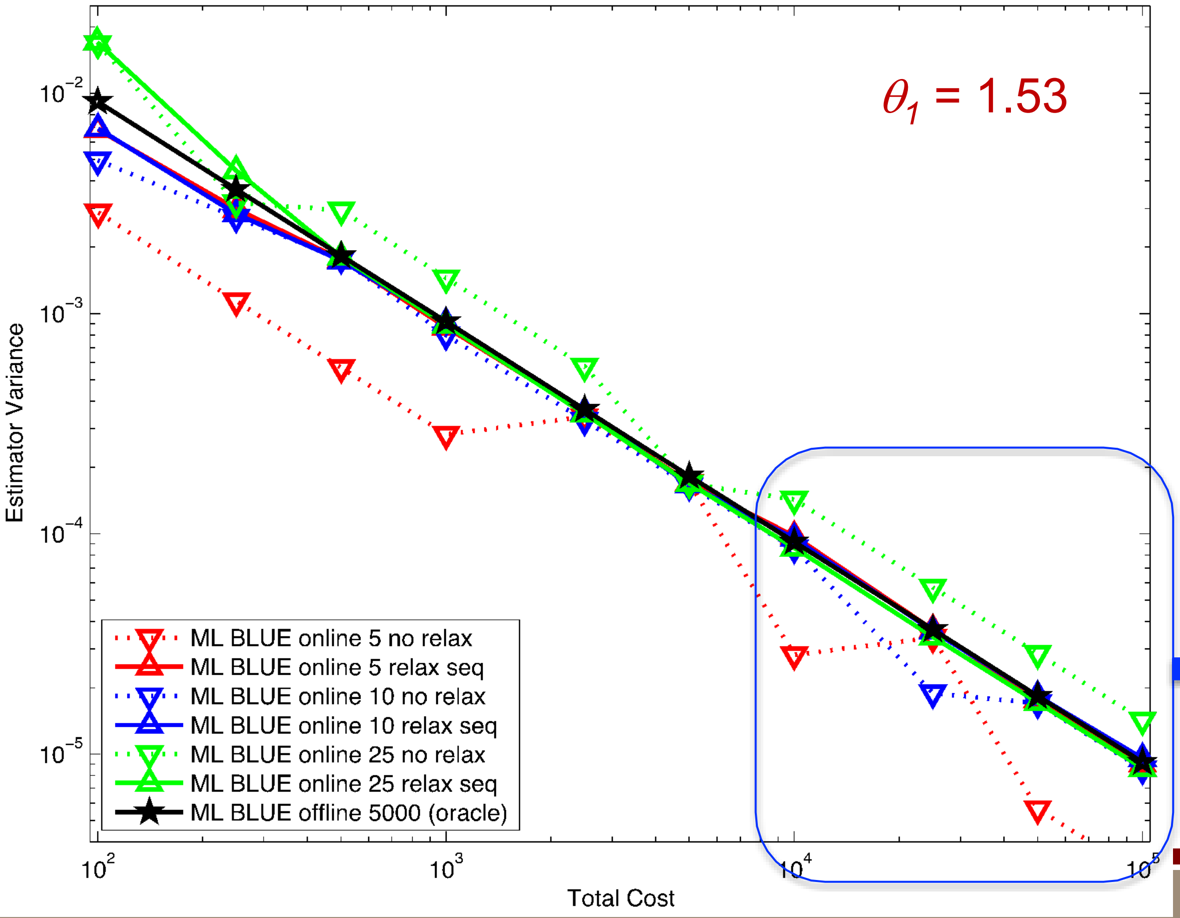
Sandia National Laboratories

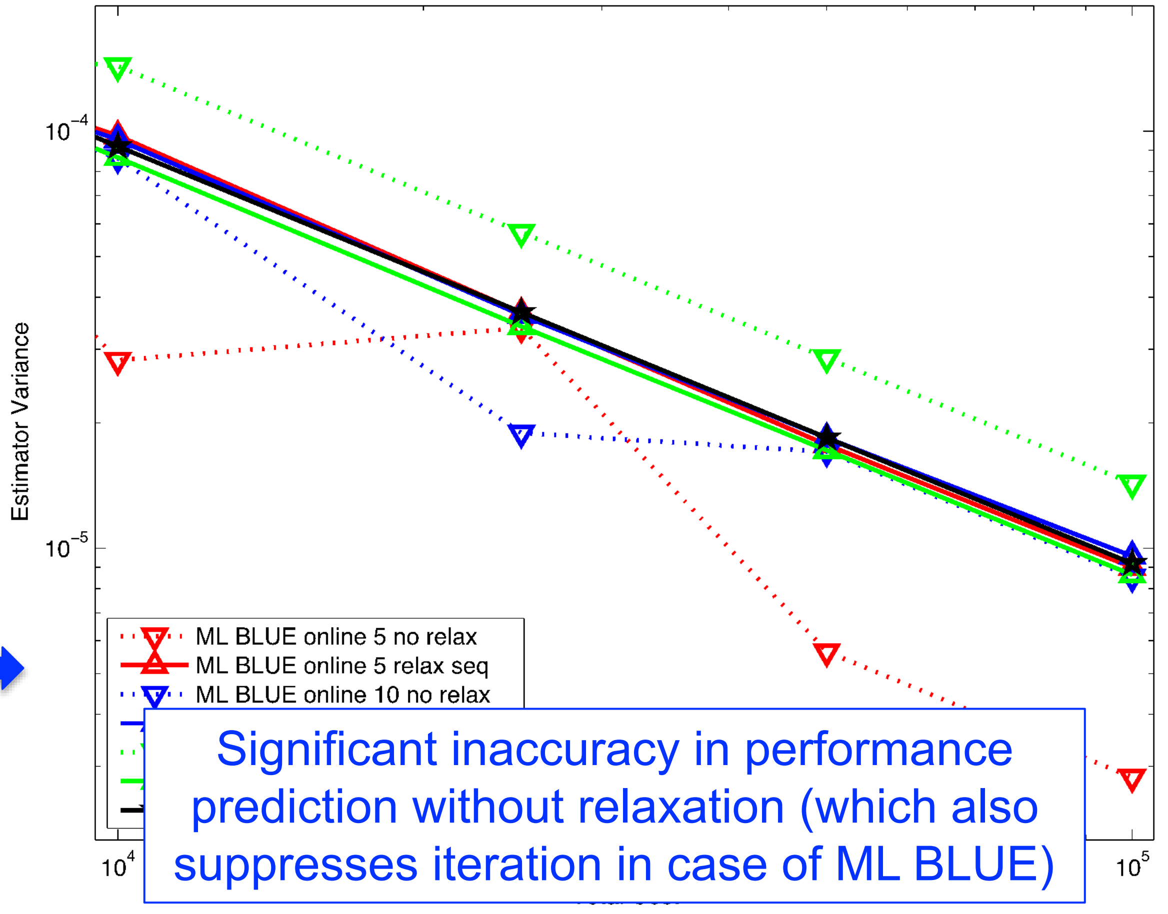
"Tunable Model" Definitions (JCP 2020)

$$Q(\theta) = \sqrt{11}\ [\ \cos(\theta)\ x^5 + \sin(\theta)\ y^5\ ]$$
$$Q_1(\theta_1) = \sqrt{7}\ [\ \cos(\theta_1)\ x^3 + \sin(\theta_1)\ y^3\ ]$$
$$Q_2(\theta_2) = \sqrt{3}\ [\ \cos(\theta_2)\ x + \sin(\theta_2)\ y\ ]$$



Later we will tune hyper-params θ_1 and θ_2 *for fixed HF* θ = π/2.
For this study, we fix θ_2 = π/6 and
θ_1 = 1.53 (MF similar to HF)
Under-relax sequence = {.5, .8, 1.}





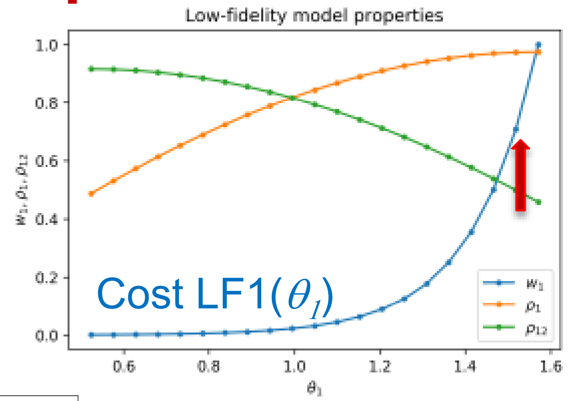
Significant inaccuracy in performance prediction without relaxation (which also suppresses iteration in case of ML BLUE)

Online pilot integration – effect of pilot over-estimation

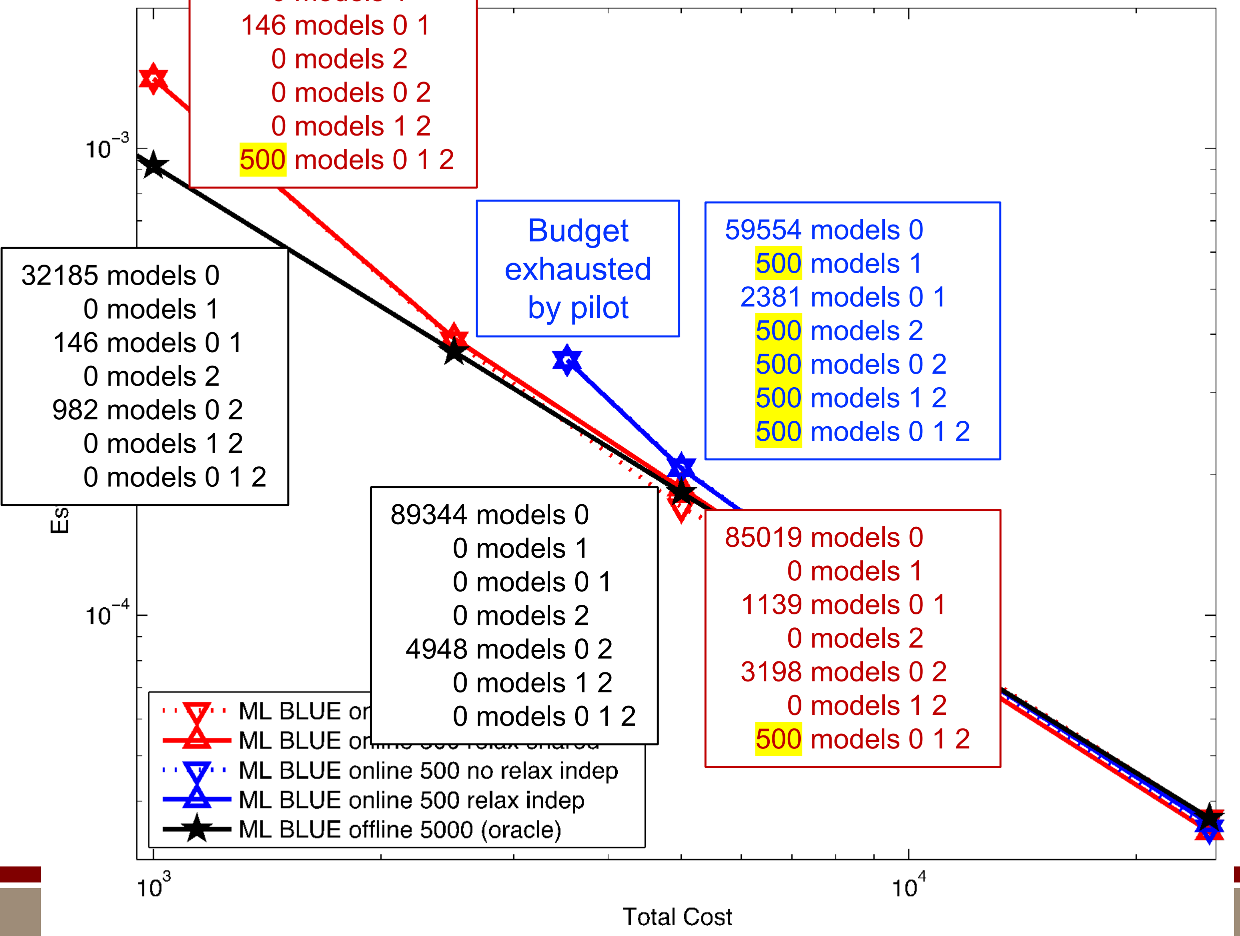
Sandia National Laboratories

"Tunable Model" Definitions (JCP 2020)

$$Q(\theta) = \sqrt{11}\ [\ \cos(\theta)\ x^5 + \sin(\theta)\ y^5\ ]$$
$$Q_1(\theta_1) = \sqrt{7}\ [\ \cos(\theta_1)\ x^3 + \sin(\theta_1)\ y^3\ ]$$
$$Q_2(\theta_2) = \sqrt{3}\ [\ \cos(\theta_2)\ x + \sin(\theta_2)\ y\ ]$$



Later we will tune hyper-params θ_1 and θ_2 *for fixed HF* $\theta = \pi/2$.
For this study, we fix $\theta_2 = \pi/6$ and study $\theta_1 = 4\pi/9$ and $\theta_1 = 1.53$.



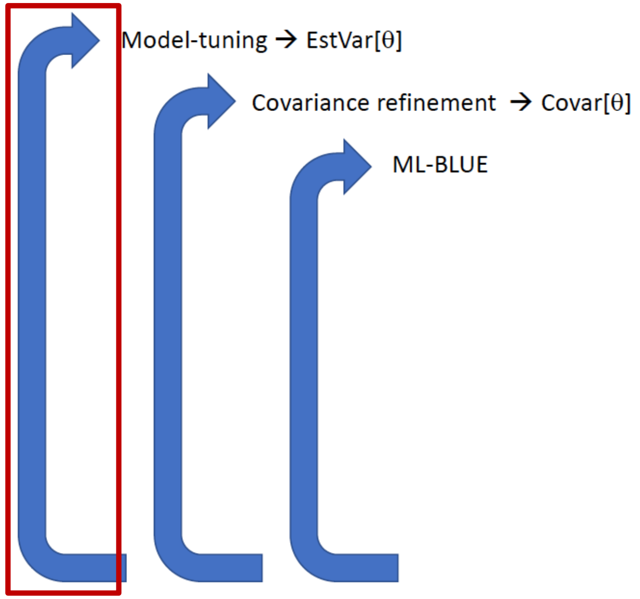
Config: θ_1 = 1.53, 3 models, 7 groups, online pilot size = 500

Online iteration + relaxation not a concern for resolved covariances from larger pilot. Issue is rather the *inefficiency of this pilot relative to optimal online allocation*.

- Highlighted constraints force solns away from Oracle
  - For fixed pilot size, increasing budget reduces effect
  - Independent pilots amplify inefficiency of over-est.
- Effects not as severe for flexible numerical estimators: solns are fairly resilient and find near-optimal alternatives

Iterated + relaxed avoids inefficiency from pilot over-est. and inaccuracy from under-est.

Multilevel Best Linear Unbiased Estimator (ML BLUE)



Exploration of ensemble configuration with ML BLUE:

1. If matrix conditioning can be mitigated, ensemble config is embedded
2. Iterated covariance approaches take on new requirements → under-relaxation
   - Shared vs. independent pilot (if something interesting here)
   - Under-relaxation
3. Performance under model tuning → continuation of ACV/GenACV robustness trends?

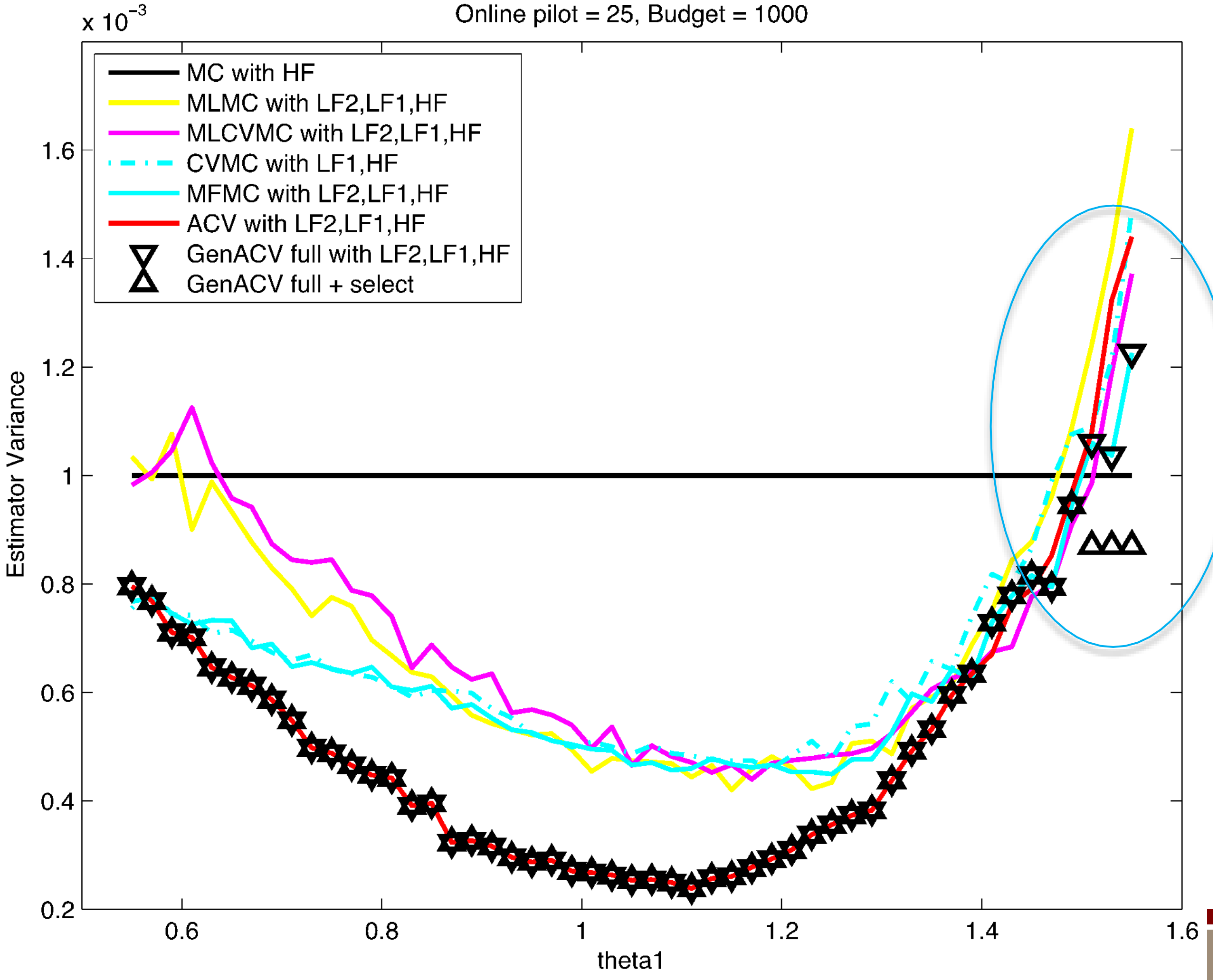
Bi-level tuning of hyper-parameters θ

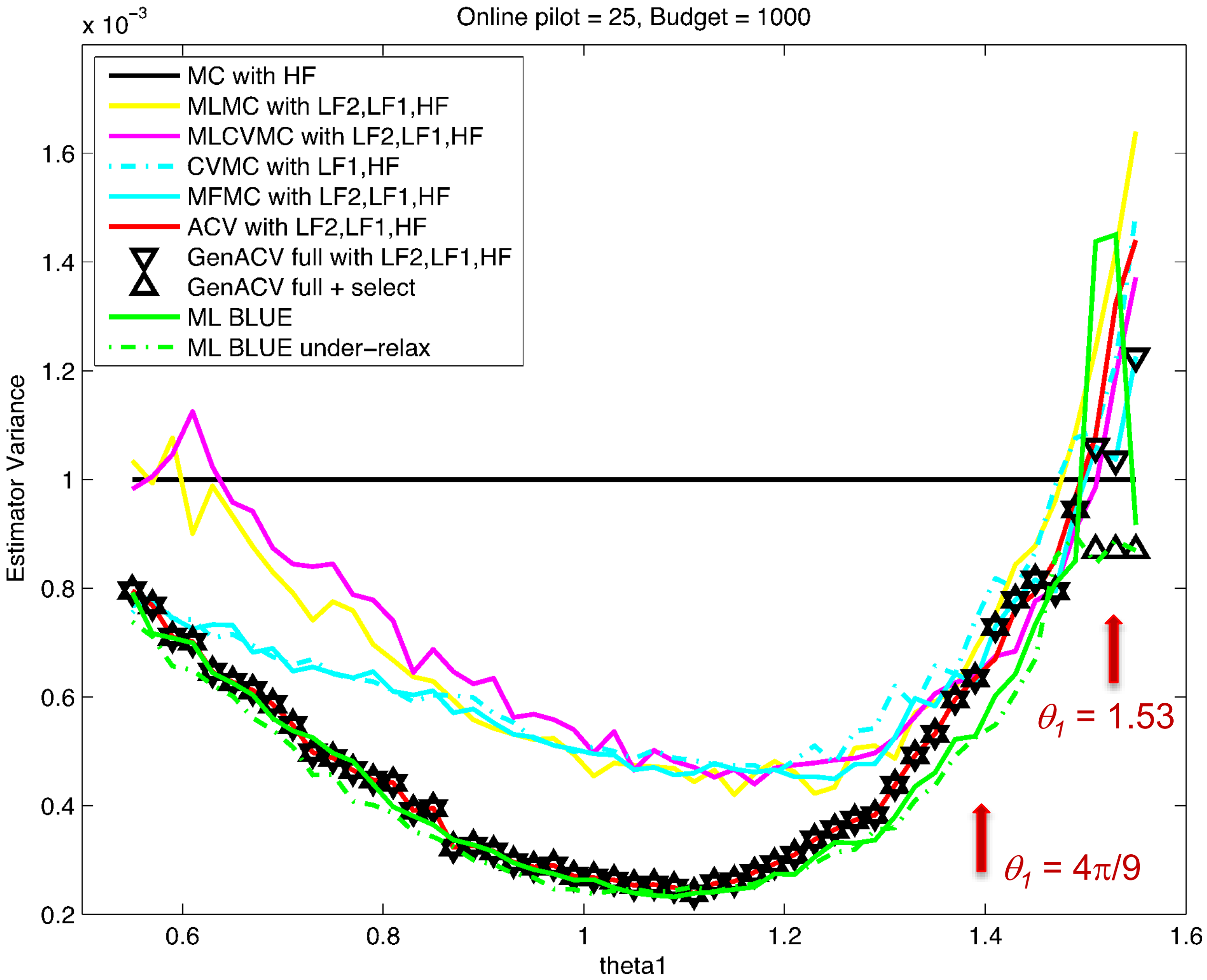
$$\arg\min_{\theta} \left[\arg\min_{\mathbf{r},N} \frac{Var[Q]}{N}(1 - R^2(\theta, \mathbf{r})) \quad s.t. \quad N\left(w + \sum_{i=1}^{M} w_i(\theta) r_i\right) \leq C \right]$$

Putting it all together: model tuning, iterated + under-relaxed, ensemble configuration

$$
\begin{aligned}
Q(\theta) &= \sqrt{11}\ [\ \cos(\theta)\ x^5 + \sin(\theta)\ y^5\ ] \\
Q_1(\theta_1) &= \sqrt{7}\ [\ \cos(\theta_1)\ x^3 + \sin(\theta_1)\ y^3\ ] \\
Q_2(\theta_2) &= \sqrt{3}\ [\ \cos(\theta_2)\ x + \sin(\theta_2)\ y\ ]
\end{aligned}
$$

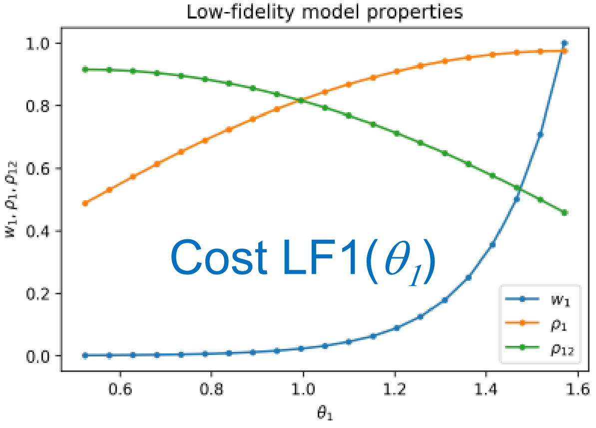
Start with parameter sweep for $\pi/6 < \theta_1 < \pi/2$ for mid-fidelity with high / low hyper-parameters fixed at $\theta = \pi/2$, $\theta_2 = \pi/6$.

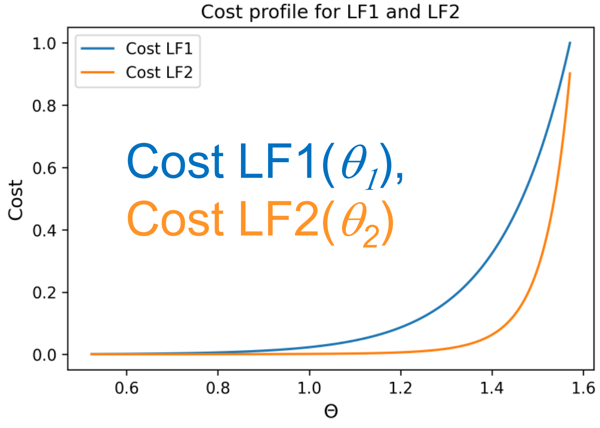




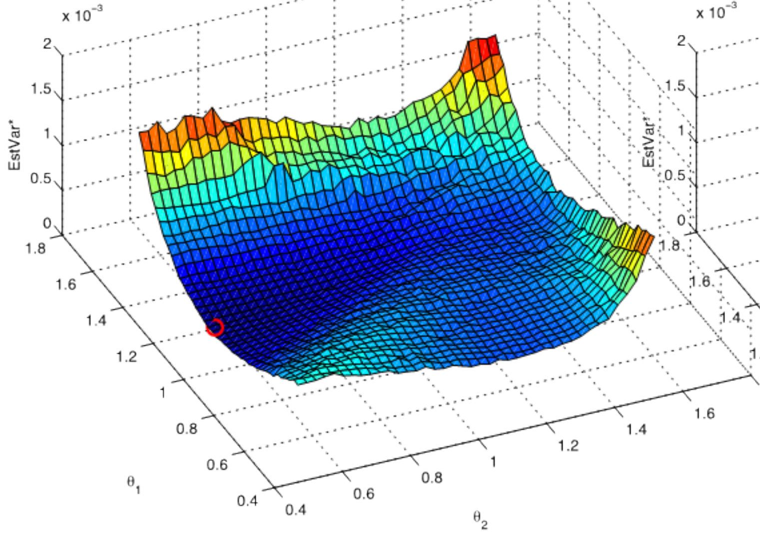
Putting it all together: model tuning, iterated + under-relaxed, ensemble configuration

$$
\begin{aligned}
Q(\theta) &= \sqrt{11}\ [\ \cos(\theta)\ x^5 + \sin(\theta)\ y^5\ ] \\
Q_1(\theta_1) &= \sqrt{7}\ [\ \cos(\theta_1)\ x^3 + \sin(\theta_1)\ y^3\ ] \\
Q_2(\theta_2) &= \sqrt{3}\ [\ \cos(\theta_2)\ x + \sin(\theta_2)\ y\ ]
\end{aligned}
$$

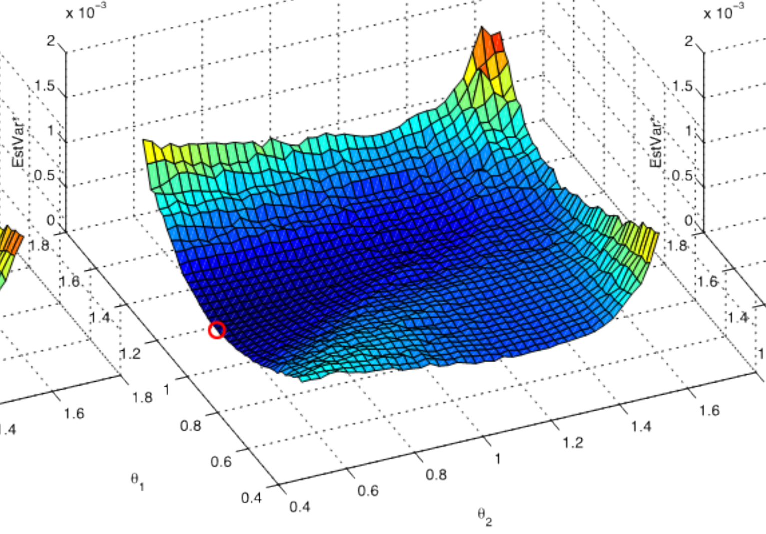




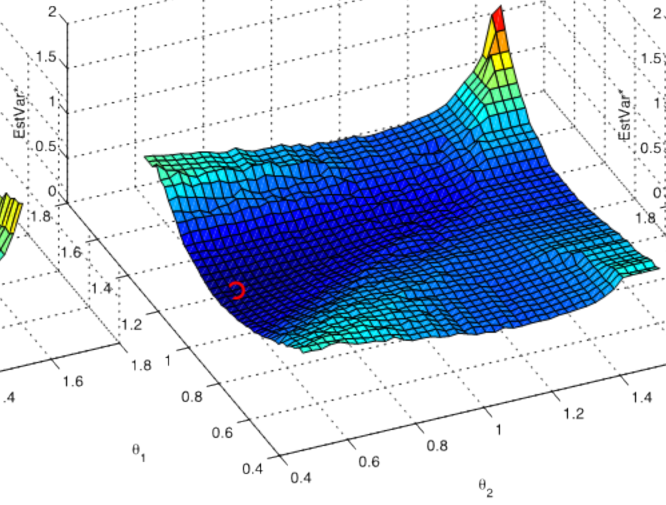
ACV iterated



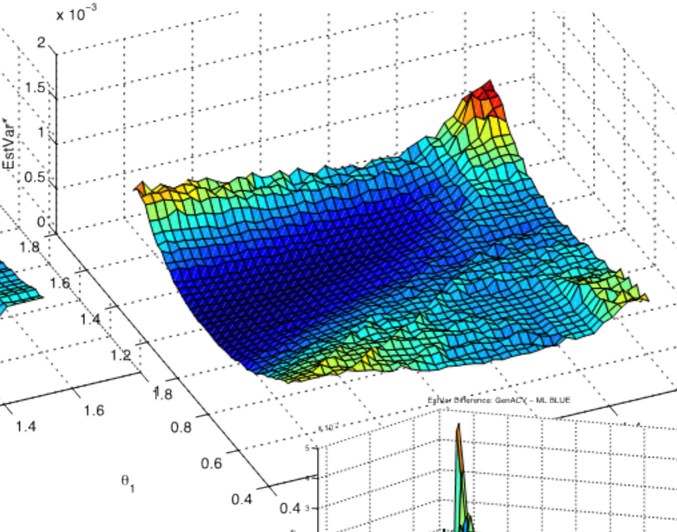
GenACV iterated, full DAG recur

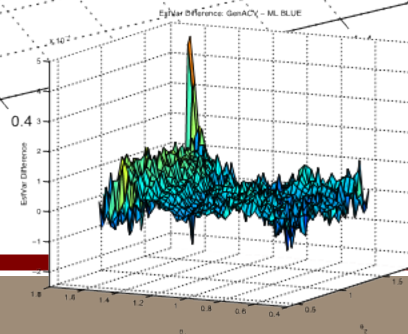


GenACV iterated, full DAG recur, model select



ML BLUE iterated, relaxed





Whether for efficiency or robustness, ensemble + DAG / group flexibility amplifies effect of model tuning

Summary Observations

Sandia National Laboratories

Production deployments of multifidelity methods encounter a variety of challenges

- Accurate offline estimations of Covar[Q] should be integrated / optimized → Outer: ***Iterated online pilots***
- LF models often have parameters that trade accuracy vs. cost → Outer: ***Hyper-parameter model tuning***
- Numerical solutions are not always reliable w/ local solvers → Inner: ***Multistart/multisolver, global/local***
- Best selections/pairings/groupings often unknown a priori → Inner: ***Model ensemble configuration***

Multilevel Best Linear Unbiased Estimator (ML BLUE)

- Allocations per enumerated group show significant promise
- Recent exploration has subjected ML BLUE to the same practical considerations as other estimators in Dakota
  - Solution modes, shared vs. independent pilots, under-relaxation, group enumeration throttles

Explore model configuration aspects of ML BLUE and compare to existing estimators

- Inner loop (covariance fixed)
  - Estimator variance at least on par with GenACV, with potential for more direct & efficient solutions
    - Avoids need to enumerate model subsets and DAGs, instead enumerating group memberships w/i integrated solve
  - Numerical conditioning requires effective mitigation
- Outer loop (converge/tune covariance):
  - Iterated approaches are again effective, but must now rely on under-relaxation due to ordering of budget allocation
  - Model tuning enhances efficiency/robustness, reinforcing the link between hyper-parameter utility and estimator flexibility

Next steps

- Work in progress
  - Numerics: eigenvalue truncation in Ψ, Semidefinite programming
  - Group ACV approaches relax independence, allowing sample reuse (https://arxiv.org/abs/2402.14736)