A Decision-Theoretic Framework for Adaptive Simulation Experiments

Kevin L Rotz[1], Terril N Hurst[2], Adam E Berryman[3], and David S Anderson[4]

Raytheon Technologies, Tucson, Arizona, 85756

A framework has been developed for simulation experiments that increases the efficiency of high-performance computing for building surrogate models to answer queries of interest for effective decision-making in the presence of quantified uncertainty. The framework applies Bayesian analysis to maximize the value-to-cost ratio of information obtained from simulation runs by accounting for the uncertainty that is present in queries, models, and simulation data. When using the framework to answer a precisely stated query, greater than 80 percent reduction has been observed in number of runs. Simple examples illustrate how using the framework can accomplish this.

1. Introduction

SEVERAL decades of rapid advances in computational science and engineering have resulted in increased reliance on modeling, simulation, and analysis (MSA) to make decisions when designing and testing complex physical systems. Pervasive MSA usage motivated increased focus on simulation verification and validation/ calibration to assure credibility of MSA results.[1] Once a simulation (shortened to "Sim" hereafter to distinguish from the *process* of simulation) has been verified, validated/calibrated, and deemed sufficiently "un-wrong" for its intended use, the next question becomes *how* to use the Sim to answer a specific question using available resources, including high-performance computing (HPC).[2] Answering this question is the primary focus of this paper, although Example 4 (Section 3) briefly explores implications of using an un-calibrated vs. a calibrated Sim.

Traditional design-of-experiments (DOE) methods provided a natural starting point for guiding Sim usage in MSA. However, since MSA queries routinely involve high-dimensional factor spaces and interactions, larger sample sizes than used in traditional DOE methods are often needed. This fact led to innovations for MSA such as Latin Hypercube Sampling (LHS) and Gaussian Process Models (GPMs).[3,4]

These innovations gave rise to new issues, options and tradeoffs. For example, with faster processors, a *fixed number of Sim runs* could be completed in less time, enabling more timely decisions; or, given a *fixed amount of elapsed time*, the number of Sim runs could be increased to reduce the statistical imprecision of parameter estimates, thereby reducing risk when making a decision.[5] This flexibility led to exploring *adaptive sampling* techniques. Rather than using what is called here *a priori* sampling, i.e. deciding beforehand where to make a large number of Sim runs and then analyzing the results, "adaptive sampling" means choosing where to make additional Sim runs after analyzing results that are based on a smaller initial, space-filling sample.

Given a mathematically precise query, the closed-loop framework illustrated in Figure 1 chooses where to sample on successive iterations. The framework schedules sampling predicated upon four interconnected models: (a) the *surrogate* model, e.g., a continuous correlated beta process,[6] which globally predicts the response using beta distributions; (b) the *value* model, e.g., mutual information,[7] for estimating the benefit of candidate runs for answering the query; (c) the *cost* model, to predict time needed to execute candidate runs, possibly from multi-fidelity Sim options; and (d) the *HPC* model, which represents aspects of the HPC grid to schedule Sim runs. Candidate Sim runs are chosen by maximizing value per cost. A Bayesian perspective is taken to formulate and update each of these models as results arrive to guide iterative run selection and HPC scheduling.

---

[1] Senior Systems Engineer II, Algorithms & Simulation Center, Bldg. 805 M/S C1.
[2] Senior Engineering Fellow, Systems Design & Analysis, Bldg. 805 M/S C1.
[3] Senior Principal Systems Engineer, Algorithms & Simulation Center, Bldg. 805 M/S C1.
[4] Systems Engineer II, Algorithms & Simulation Center, Bldg. 805 M/S C1.

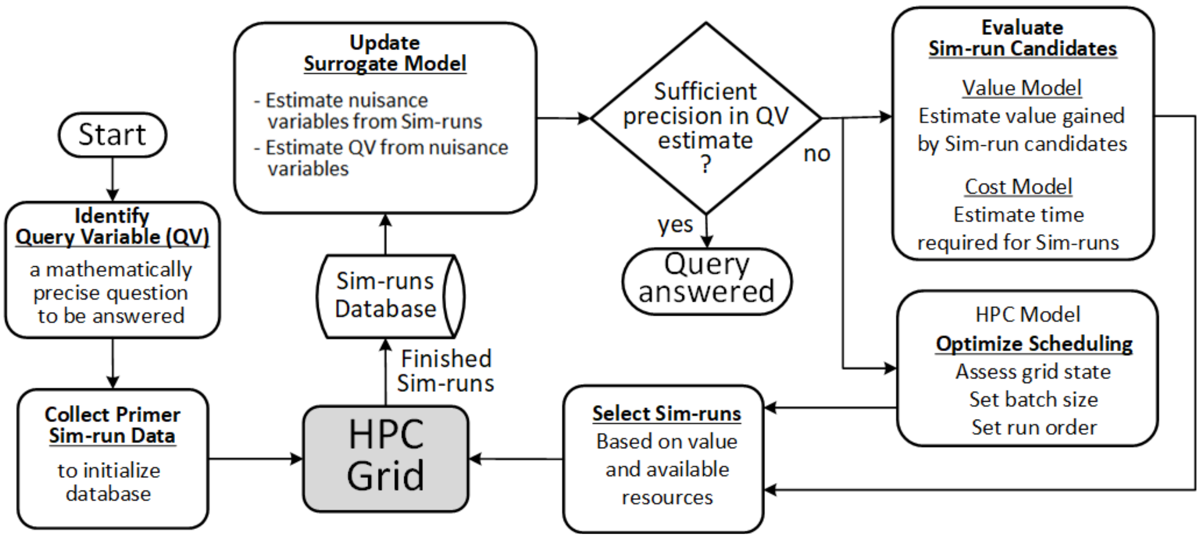


Figure 1: The ASHI loop.

The framework shown in Figure 1 is called ASHI, which stands for Adaptive Sampling for HPC Improvement. It is designed to specify Sim runs in an optimal value-to-cost manner for decision-making after answering a given query with quantified uncertainty/risk. Section 3 includes simple examples of how the ASHI loop is used.

The ASHI Query and Models

A simulation experiment supports decision-making by answering some question or query. We should not expect a simulation experiment to give a crisp answer to a vaguely stated query. The query must be precisely posed; it is defined as a random variable called a *query variable*, having the form $QV \sim p(\theta \mid y)$, where p is the posterior probability distribution of parameter θ given available data y.

The ASHI loop employs four models; examples of each are described briefly below and illustrated in Section 3.

1. The *Surrogate Model* parametrizes observed Sim output given known inputs. An example is the continuous correlated beta process model (CCBP), which estimates the proportions of successes and failures using beta-distributed uncertainty at every point in the Sim's factor space. The CCBP combines results using an exponentially decaying correlation function; CCBP output is used to predict the result of a candidate Sim run.

2. The *Value Model* quantifies the information gained for the QV by running a candidate Sim run, thereby estimating uncertainty in QV that *could* be reduced *after* choosing and observing results from the candidate Sim run. An example value model is mutual information.

3. The *Cost Model* estimates HPC resources required for a candidate Sim run, e.g., time or memory, based upon past runs. The generalized linear model is a form that is typically used. A given simulation experiment might have multiple Sim fidelity options that involve different costs. It may be desirable to balance value of information gained with the cost of a mixture of Sim runs from these multi-fidelity options.

4. The *HPC Model* accounts for current HPC workload and uses information from the query and the other ASHI models to schedule the next Sim runs and to obtain optimal value per cost.

Given appropriate instances of each of these models and results from selected Sim runs, the ASHI loop begins and ends with the query variable QV, having the form $p(\theta \mid y)$. The posterior distribution for QV should fully reflect the uncertainty arising from Sim modeling assumptions, Sim-run data y, and surrogate model parameters θ. Section 2 describes the relationship between the surrogate and value models being used to quantify QV uncertainty. Section 3 illustrates application of ASHI to simple examples; in each example, results are compared when using ASHI vs. using traditional *a priori* sampling and maximum-likelihood estimation. Section 4 summarizes results and concludes the paper.

2. Surrogate and Value Models

Since our interest is in the query variable QV, and since we can only run our Sim with specific inputs x to produce output Z_x, we choose candidate Sim runs for which the mutual information (MI) between predicted Z_x and QV should be high, using currently available data and the surrogate model. Mutual information is given by

$$I(QV; Z_x) = \sum_{QV} \sum_{Z_x} p(QV, Z_x) log \frac{p(QV, Z_x)}{p(QV)p(Z_x)}$$

Assuming that QV and Z_x are independent given some hidden parameter θ that is relevant to both the query and the Sim run, we can write the joint probability as

$$p(QV, Z_x) = \int p(QV, Z_x|\theta)p(\theta)d\theta = \int p(QV|\theta)p(Z_x|\theta)p(\theta)d\theta$$

Notice that using a point estimate of θ would result in $I = 0$ for all candidate Sim runs, since $\log(1) = 0$. Therefore, to estimate MI, we must employ a surrogate model that produces a full posterior distribution over θ. Common examples of surrogate models for doing this include Bayesian linear and generalized linear regression models, the continuous correlated beta process (CCBP), and Gaussian process regression.

Choice of surrogate model depends on the questions that one wants to answer. Bayesian regression using both linear and generalized linear models is often a good choice, and one can either sample coefficients using Markov chain Monte Carlo (MCMC) methods or estimate posteriors using Laplace approximation. In instances where there is high variation in the response surface, a non-parametric model such as Gaussian process regression is a powerful tool. Gaussian process regression can be useful for sensitivity analysis and optimization problems; the choice of a *parametric* regression function can unintentionally bias estimates of pointwise sensitivities or locations and number of local optima.

However, due to the growth of the covariance matrices involved, utilizing Gaussian processes with MCMC is limited to a small number of treatments (on the order of a thousand) before becoming impractical. In these situations, if posterior correlations between responses are not necessary to the query, one may resort to pseudo-Bayesian methods such as the continuous correlated Beta process (CCBP). Variational Bayesian methods, developed by Hensman *et al*, are also being investigated to overcome the small-data limitations of using MCMC.[8]

In summary, mutual information is a good value model for ASHI to choose Sim runs, and for simple queries and response surfaces, simple Bayesian parametric regression models are usually adequate. Responses involving many factors and localized response features are better served by non-parametric models. For example, the CCBP is a good non-parametric model when the query doesn't require considering pairwise posterior correlations between hidden variables θ, as is the case in Example 1 in Section 3. For queries that involve derivatives, pairwise correlations are crucial, as in Example 2. A Gaussian process-based surrogate model is then more appropriate.

3. Examples of Applying ASHI

ASHI has been applied to examples involving a diversity of query types, all explicitly accounting for uncertainty: (1) assessing compliance with a performance requirement, (2) sensitivity analysis of a system's response to input factors, and (3) engineering design optimization. Each example includes a brief description the simulation experiment's query and the Sim that was used, the experiment's input ("factor") space and surrogate model(s) chosen for inference, a description of *a priori* sampling for observing the system's response Z_x at each x-point ("treatment"), a comparison of maximum-likelihood vs. Bayesian inference results based on the *a priori* sample, and a description of ASHI sampling and the resultant Bayesian posterior distribution for QV. Example 4 illustrates how epistemic uncertainty from modeling assumptions can affect Sim-run choices and query results when using an un-calibrated vs. a calibrated Sim.

3.1 Compliance with a performance requirement. A 6-degree-of-freedom Sim was used in Example 1 to determine compliance with a system performance requirement, "The system shall perform with success probability $\pi_S \geq 0.85$ over at least 40% of the engagement space." The Sim was run at each treatment to estimate local π_i, the probability of guiding a missile to within specified proximity of a ground target for the i^{th} engagement, given specified values of five control-factor inputs: target range, target altitude, relative launcher-to-target altitude, launcher pitch, and launcher heading. For each control-factor treatment, uncertainty factors were also drawn randomly from Monte Carlo distributions representing the system's sensors, effectors, and environment.

For *a priori* sampling, 32 trials were made per treatment after drawing from the Sim's Monte Carlo factor distributions. Thus, each maximum-likelihood estimate for $i = 1$ to 1000 Latin Hypercube Sample (LHS) treatments took the form $\hat{\pi}_i = \#successes_i \div 32$. An empirical cumulative distribution function (CDF) of $\hat{\pi}_i$ was then constructed to represent performance across the system's full engagement space (Figure 2, left). The empirical CDF reflects the constant number of trials per treatment (33 possible values, i.e. for 0 to 32 successes). The 95% confidence interval (vertical dotted lines) for the 40th percentile's $\hat{\pi}_i$ quantile falls completely below 0.85, clearly indicating failure to comply with the performance requirement.

An intermediate method, dubbed "ASHI Lite," was completed for Example 1. 1000 *a priori* LHS treatments were run, but instead of doing 32 trials per treatment, the number for each varied between 8 and 100, based upon a *Beta*(*2,2*) conjugate prior distribution for Bernoulli trials. Each treatment's trials were increased from 8 until (a) the posterior probability of non-compliance (PNC) fell *below* 5%, or (b) it *exceeded* 95%, or (c) the *maximum* of 100 trials was reached. Figure 2 (right) shows typical credible intervals for eight treatments and denotes the number of trials and the posterior PNC for each one. PNC fell below 5% for Treatment 1 after 28 trials; after 8 trials, Treatment 2 had PNC > 95%; Treatments 4, 6, and 8 were too-close-to-call after 100 trials. The total number of Sim runs was reduced from 32,000 to 25,148, a 21% savings. 155 treatments remained unresolved after 100 trials; more trials would have resolved some of these, but it would not have changed the system's non-compliance.

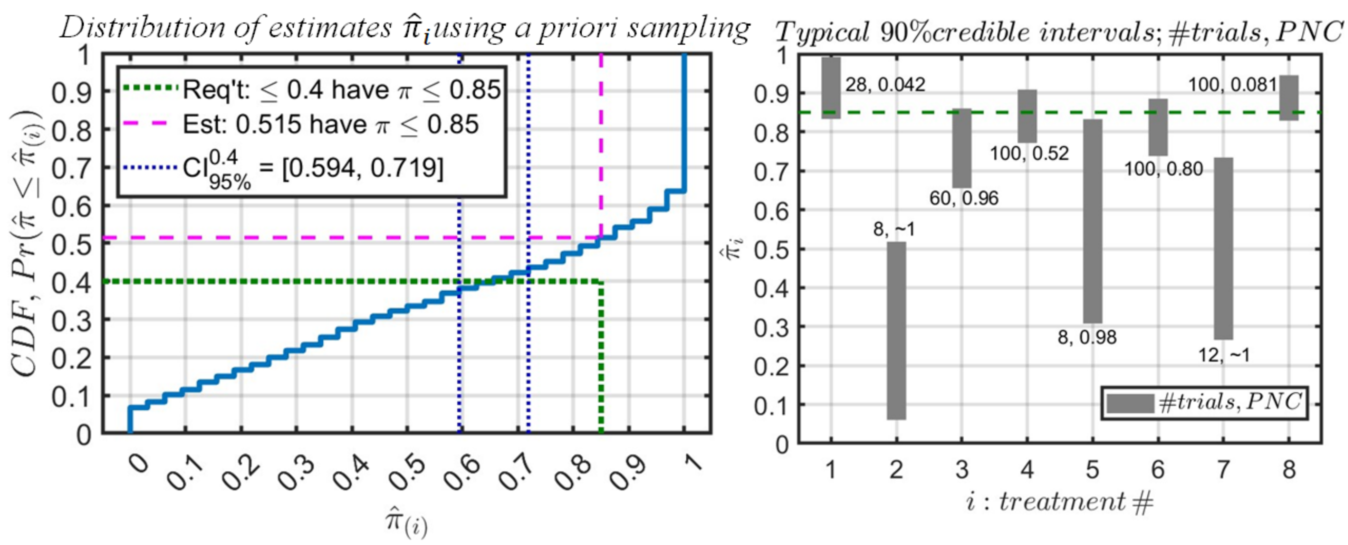


Figure 3: Example 1 results, using *a priori* sampling. *A CDF of maximum-likelihood estimates is shown on the left, and on the right are typical 90% credible intervals on treatments' posterior probability of non-compliance,* $PNC_i = 1 - p(\pi_i \mid y)$*, using ASHI Lite sampling. Labels for credible intervals show the number of trials needed and each treatment's PNC.*

To "prime the ASHI pump," sampling for Example 1 began with a *single* trial for each of the same 1000 LHS treatments selected during *a priori* sampling. Therefore, each treatment's outcome was either a "success" or a "fail," and the outcomes were used to construct a continuous correlated beta process (CCBP) spatial surrogate model, employing an exponential function to correlate outcomes from *individual* treatments across the control-factor space. Subsequent batches of 1000 were down-sampled from 10,000 LHS candidates after using the CCBP to rank value (mutual information). Finally, the system's compliance result was represented as a joint posterior distribution over the CCBP parameters.

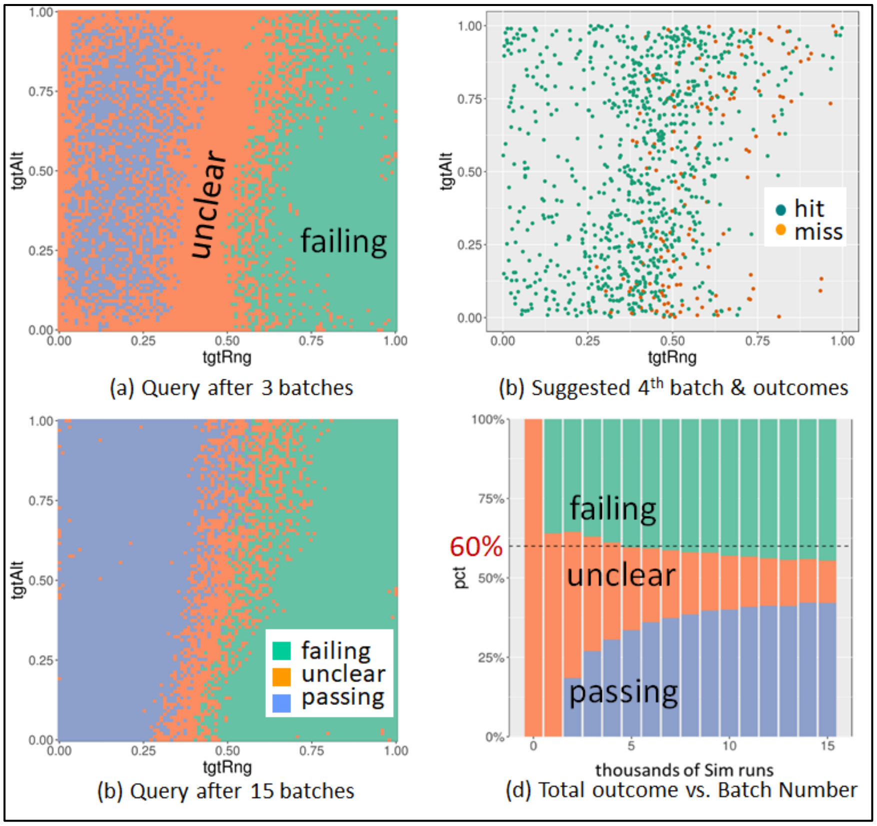


Figure 3: Example 1 ASHI-based compliance findings. *(a) After 3000 Sim runs, CCBP predictions were beginning to localize a comply/non-comply boundary; (b) suggested runs for Batch 4 are denser near this boundary; (c) the CCBP identified the compliance region after 15 batches, but (d) even after only 6 batches system non-compliance is clear. (Note that CCBP plots are 2-D subspace projections of the system's full joint posterior distribution.)*

As shown in Figure 3a, by using the CCBP surrogate model, the (non-)compliance boundary began to emerge fairly quickly—within 3000 runs. For too-close-to-call cases (boundary region), more data were taken, initially further from the compliance boundary but then closer to the boundary (Figure 3b). By 15,000 runs, the compliance boundary was established (Figure 3c); however, even after only 6000 runs, it was clear that less than 40% of the engagement space achieved $\pi \geq 0.85$ (see dotted line in Figure 3d), i.e. the system's performance requirement was not met. Thus, only 6000 runs were needed, vs. 32,000 *a priori* runs, to establish system non-compliance.

In summary, the same conclusion of system non-compliance was reached by *a priori* and ASHI sampling. However, to reach this conclusion, ASHI Lite reduced the cost from 32,000 to 25,148 runs—a savings of 21%, and ASHI required only 6000 runs—a savings of 81%. Some of these runs could have been used to further reduce uncertainty, although the query did not call for doing this.

3.2 Sensitivity analysis. Example 2 involved simulating operations to machine and inspect parts (Figure 4).[9] Part interarrival times were exponentially distributed with a mean of 3 minutes (assuming infinite capacity of first-in-first-out machining and inspection queues). After machining, each part was inspected and re-worked if found bad.

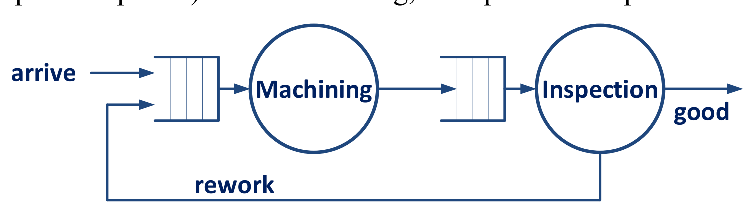
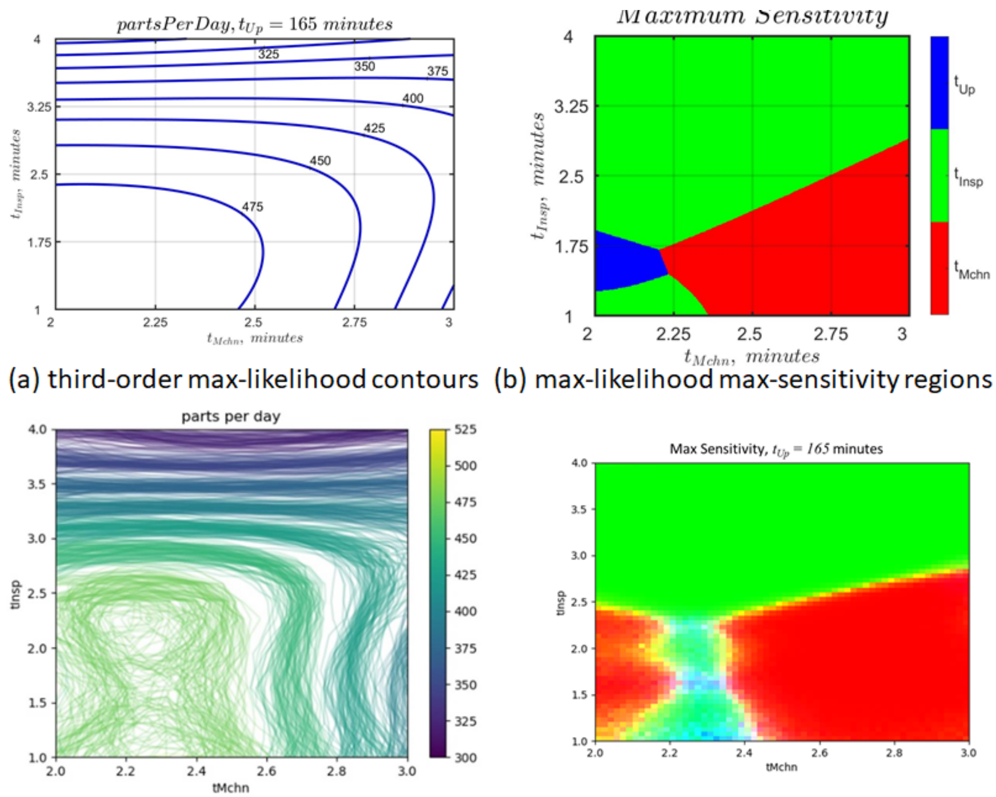


Figure 4: Simple machining/inspection operation.



(a) third-order max-likelihood contours (b) max-likelihood max-sensitivity regions

(c) Bayesian Gaussian process contours (d) Bayesian max-sensitivity regions

Figure 5: Example 2, maximum-likelihood vs. Bayesian surrogate model contours and max-sensitivity regions.

The query response was mean system throughput, *partsPerDay*. The problem initially included 5 control factors: machining and inspection times were both gamma-distributed, having $2 \leq \mu_{tMchn} \leq 3$ and $1 \leq \mu_{tInsp} \leq 4$ minutes, respectively (rate parameter $\beta = 8$); bad parts were binomially distributed ($0.09 \leq \pi_{Bad} \leq 0.1$); and machine up-time and time-to-repair were exponentially distributed, $360 \leq \mu_{tUp} \leq 400$ minutes and $10 \leq \mu_{tRepair} \leq 20$ minutes, respectively. Latin Hypercube Sampling was used: 200 *a priori* treatments, each replicated 32 times. After factor-screening (i.e. fixing π_{Bad} at 0.10 and $\mu_{tRepair}$ at 10 minutes), a 3-factor, third-order ordinary least-squares surrogate model was constructed using the *a priori* sample.

Figures 5a shows third-order response surface contours vs. $\{t_{Mchn}, t_{Insp}\}$ at $t_{Up} = 165$ minutes, using *a priori* sampling. Figure 5b shows and max-sensitivity regions, identified for each point as the factor having the maximum absolute value component of the response surface's gradient. Figures 5c-d show Bayesian versions using a Gaussian Process Model and auto-relevance determination to fit the data.[10] Samples of *partsPerDay* contours drawn from the posterior are shown in Figure 5c. The posterior over *partsPerDay* induces a probability distribution over $p(Sens)$ and in turn $p(maxSens)$ at each point in the factor space. The value of a candidate at x is then the mutual information between the max sensitivity at x and the probable outcome of a new Sim-run at x.

Example 2 used mutual information between a Gaussian Process Model's predicted Sim response Z_x and query variable QV as a measure of candidate Sim-run value—see a snapshot in Figure 6 (left). Based on average entropy in the sensitivities, adaptive sampling rapidly beat *a priori* sampling (Figure 6, right). However, eventually, *a priori* sampling caught up and surpassed adaptive sampling, likely due to noise in the estimate of the mutual information when using MCMC. Since Monte Carlo samples were used to estimate $p(QV, Z_x)$, and since the mutual information between QV and Z_x became smaller as the data set became larger, the estimation of mutual information became swamped by statistical noise. More samples could have been drawn at the expense of additional downtime between batches. An alternative under investigation involves variational Bayesian methods to approximate the posterior distributions.[9]

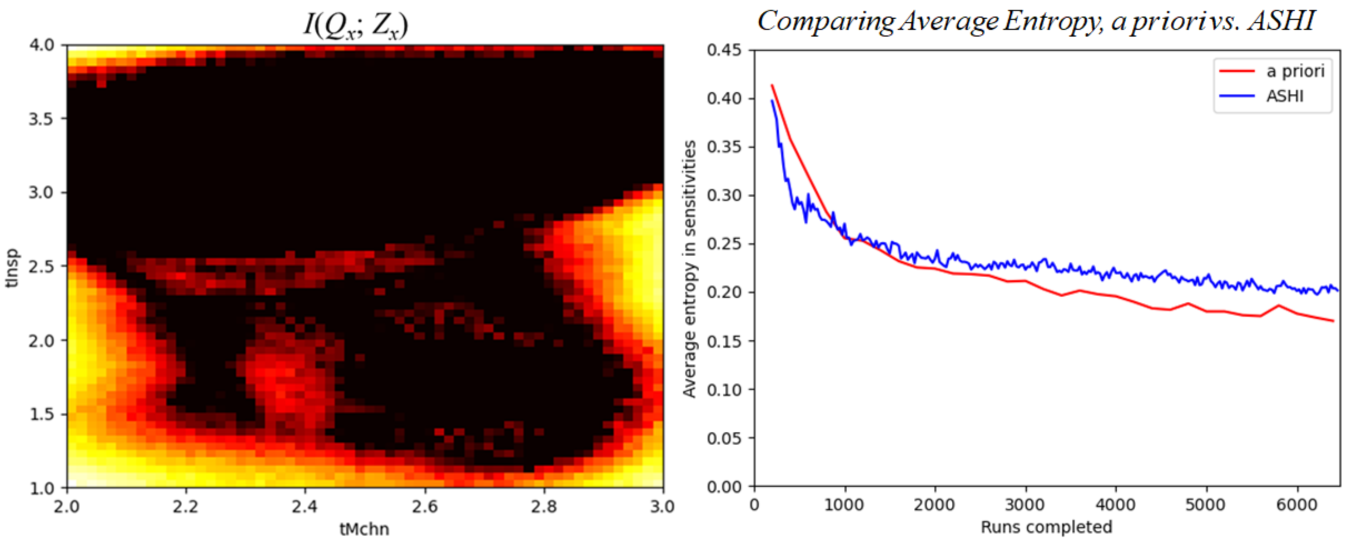


Figure 6. Mutual Information and Entropy, Example 2. *Value snapshot (left) and entropy comparison (right).*

3.3 Design optimization. Example 3 involved engineering design optimization using a 6-degree of freedom Sim. Four design variables for a guidance algorithm were of interest: two gain values and two gain-switching times-to-go. The objective was to find optimal design-variable values to maximize probability of successfully guiding to a target, denoted by $best\pi_S$. Another requirement was to achieve a 90% credible interval width of 0.10 for $best\pi_S$. For *a priori* sampling, 200 Latin Hypercube treatments, each replicated 64 times, were used; however, a useful surrogate model for design optimization was not created using this data, nor was a 0.10 credible interval achieved. Figure 7 shows Bayesian analysis results using the *a priori* data. The left plot compares 95% credible intervals using *a priori* sampling vs. ASHI, and the right plot compares posterior distributions on $best\pi_S$. While even after 12,800 *a priori* runs, the desired credible interval was not achieved, it took only 5850 ASHI runs to achieve 0.10, and the same width achieved with 12,800 *a priori* runs was reached after only 2350 ASHI runs.

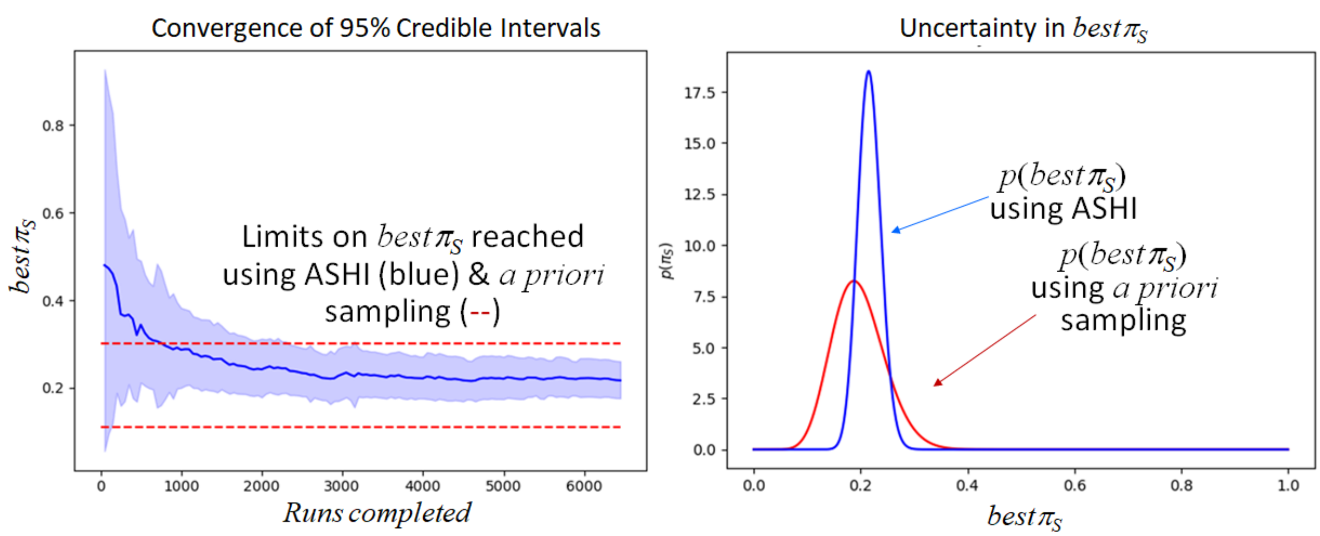
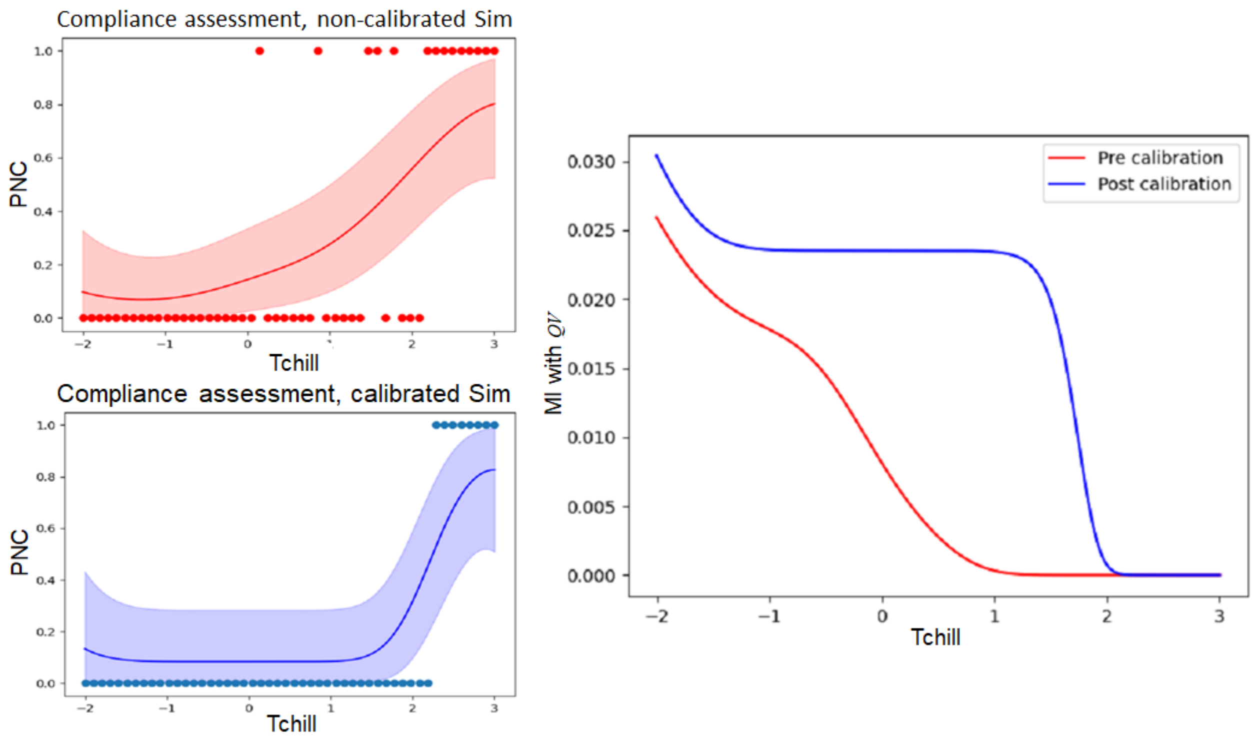


Figure 7: Example 3. Comparison of $best\pi_S$ uncertainty achieved with ASHI vs. *a priori* sampling.



(a) PNC before (top), after (bottom) calibration. (b) Mutual information, pre- vs. post-calibration.

Figure 8: Example 4. Comparison of (non-)calibrated Sim compliance estimates and mutual information for *coolMeat*.

3.4 VVUQ. Example 4 differs from the other three: of primary interest was how the value model differed between a non-calibrated and a calibrated Sim. The example, nicknamed *coolMeat*, was inspired by a food-safety study completed at the University of Nebraska.[11] The study involved cooling a cooked ham is cooled quickly to prevent re-growth of bacteria within the ham. Food safety requirements include maximum allowable cool-down times given chilling air temperature T_{chill}, i.e. the query was "Are we compliant at each given value of T_{chill}?" Two uncertainty factors were involved: the calibration factors, thermal conductivity k and convection heat transfer coefficient h. In the un-calibrated case, $\{k, h\}$ were drawn independently from uniform distributions. The calibrated case used correlated random draws from the joint posterior probability distribution of $\{k, h\}$ that was obtained using field data and Bayesian analysis.

Figure 8a shows how the probability of non-compliance (PNC) varied with T_{chill} when using a non-calibrated vs. a calibrated Sim. By using the calibrated Sim, a crisper compliant/non-compliant transition region was obtained, and the hazard of M&S bias was reduced. Figure 8b shows the mutual information between PNC and a new Sim run, as would be used to recommend new Sim runs in the ASHI loop. The crucial point is how different mutual information is between a non-calibrated and a calibrated Sim; therefore, calibrating before sampling is clearly important, since mutual information will guide sampling towards different regions.

4. Summary and Conclusion

The four examples in Section 3 illustrate how sampling and inference vary between using ASHI and conventional, *a priori* sampling. Example 1 showed that ASHI saved 81% of the runs used in *a priori* sampling to decide compliance while also spatially describing the system compliance region. Example 2 illustrated how response contour uncertainty could be quantified and how regions of max-sensitivity factor could be more precisely identified. For Example 3, ASHI saved over 50% of runs vs. *a priori* sampling. Finally, Example 4 showed the difference in value model estimates between a non-calibrated and a calibrated Sim and how the compliance region differed—and hence why calibration is important to be done *before* deciding where to sample.

Query specification and surrogate model selection are key aspects in successfully using ASHI. Each example employed a different surrogate model. Important elements in choice of surrogate model include the specifics of the query and the amount of data being collected.

The ASHI loop is a significant contribution for improving the efficiency of simulation experiments. By employing Bayesian analysis and information-theoretic concepts, it enables quantifying uncertainty at each step of MSA, beginning with Sim calibration and ending with precisely representing query results. Additional use cases are being investigated, as well as exploring additional model types to employ within the ASHI loop.

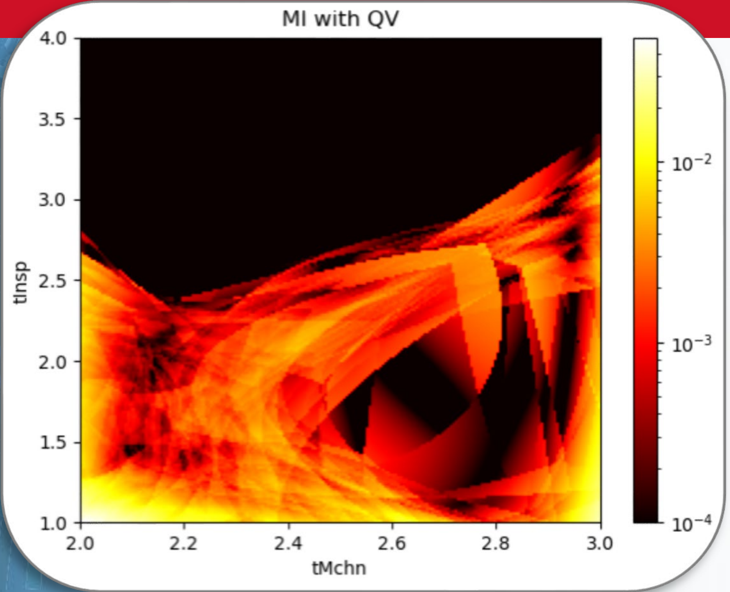
Acknowledgment

The authors wish to thank Dr. Jefferson Taft for his foundational contributions in architecting the ASHI loop.

References

1. Bayarri *et al*, "A Framework for Validation of Computer Models," Technical Report 162 (2005), National Institute of Statistical Sciences, www.niss.org.
2. Jones, J.S. A.I. Scherling, T.M. Allen, and T.N. Hurst, "Maximizing Effectiveness of a High-Performance Computing Grid by Using the DASE Protocol," 2019 AIAA Conference on Modeling & Simulation Technologies.
3. McKay, M.D., R. Beckman, and W. Conover, "A Comparison of Three Methods for Selecting Vales of Input Variables in the Analysis of Output From a Computer Code," *Technometrics* 21(1):239-245 (1979).
4. Rasmussen, C.E., "Gaussian Processes in Machine Learning," https://is.mpg.de/fileadmin/user_upload/files/publications/pdf2903.pdf
5. Berger, J.O., *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer (2010).
6. Goetschalckx, R., P. Poupart, J. Hoey, "Continuous Correlated Beta Processes," *Proceedings*, 22nd International Joint Conference on Artificial Intelligence, Vol. 22, No. 1, p. 1269 (2011).
7. Stone, J.V., *Information Theory: A Tutorial Introduction*, Sebtel Press (2015).
8. Hensman, J., N. Fusi, and N.D. Lawrence. "Gaussian processes for big data." arXiv preprint arXiv:1309.6835 (2013).
9. Law, A., *Simulation Modeling and Analysis*, 5th ed., McGraw-Hill (2015).
10. Titsias, M., "Variational learning of inducing variables in sparse Gaussian processes." *Artificial intelligence and statistics*, PMLR, 2009.
11. Amézquita, A., L. Wang, and C.L. Weller, "Finite Element Modeling and Experimental Validation of Cooling Rates of Large Ready-to-Eat Meat Products in Small Meat-Processing Facilities," https://digitalcommons.unl.edu/biosysengfacpub/97/?utm_source=digitalcommons.unl.edu%2Fbiosysengfacpub%2F97&utm_medium=PDF&utm_campaign=PDFCoverPages.

A Decision-Theoretic Framework for Adaptive Simulation Experiments



Kevin L Rotz, PhD
Terril N Hurst, PhD
Adam E Berryman
David S Anderson, PhD

Raytheon Missiles & Defense

This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

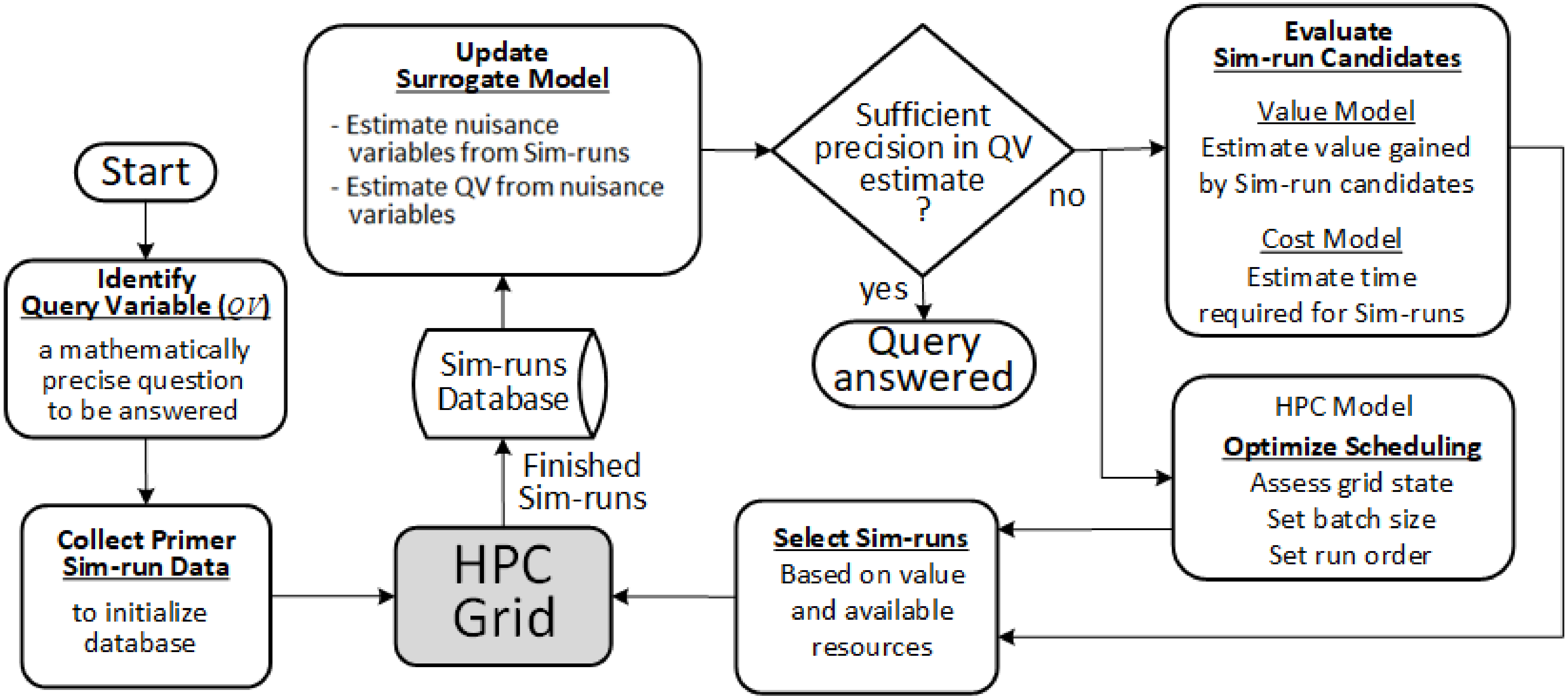
Copyright 2022, Raytheon Technologies. All rights reserved.

The rise of pervasive modeling, simulation, & analysis (MSA)

- Virtuous cycle enabled by decades of advances in computational science & engineering
- Performing simulation experiments is relatively easy; achieving credible MSA is harder. Trusting simulation output requires full *uncertainty quantification* (UQ) – see the VVUQ literature
- High-performance computing (HPC) grids enable UQ in high-dimensional factor spaces via
  - space-filling designs, e.g., Latin Hypercube Sampling (LHS)
  - surrogate models (SMs) for random functions, e.g., Gaussian processes
- These innovations introduce issues, options, and tradeoffs for MSA; e.g.,
  - a fixed number of Sim runs can be completed/repeated in a much shorter time, or
  - completing more Sim runs within a fixed time can reduce statistical (vs. practical?) imprecision/risk

Adaptive sampling is more efficient for UQ than *a priori* sampling

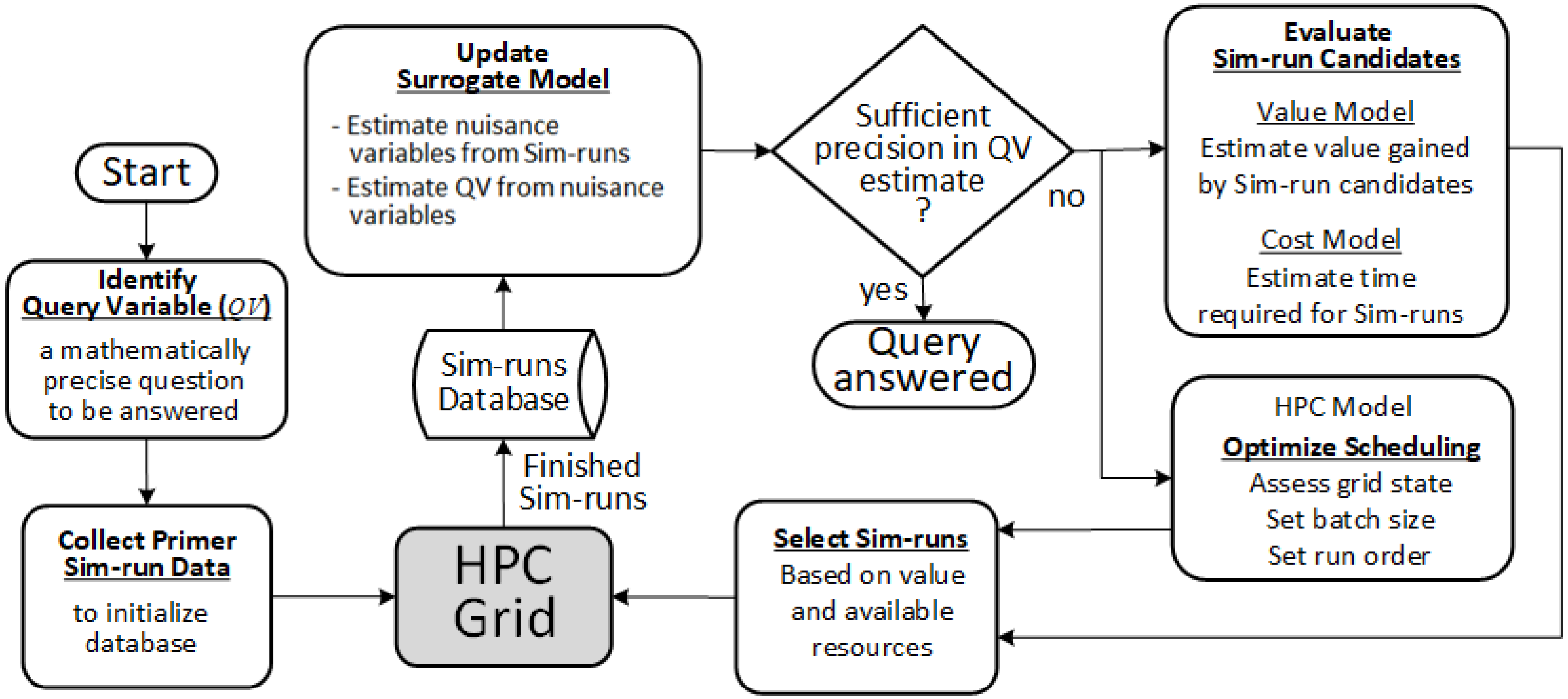
The ASHI Loop: Adaptive Sampling for HPC Improvement



Given a mathematically precise query, ASHI chooses where & when to make Sim runs

The query variable (QV) plays a central role in ASHI

- The ASHI query is represented as a random variable QV
- ASHI calculates a probability distribution $p(QV \mid y)$, where y is the data observed via Sim runs
- $p(QV \mid y)$ should fully reflect uncertainty arising from
  - Sim modeling assumptions
  - Sim-run data
  - Surrogate model parameters θ



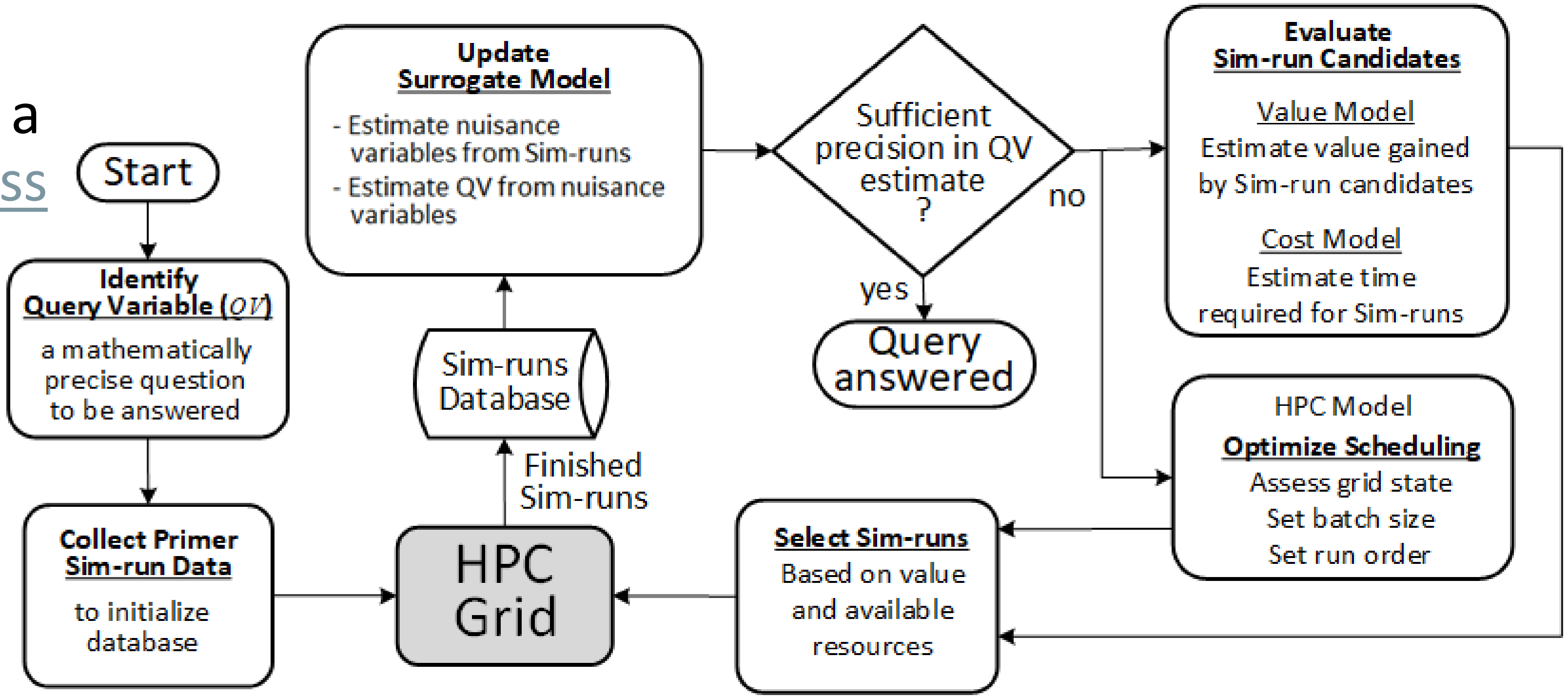
- ASHI checks whether desired QV precision is obtained before deciding to submit another batch of selected Sim runs to the HPC grid

The ASHI loop begins and ends with the query variable QV

ASHI employs four interconnected models

Raytheon Missiles & Defense

- The **Surrogate Model** (SM) – e.g., a continuous correlated beta process
- The **Information Model** – e.g., mutual information
- The **Cost Model** – e.g., run-time prediction, perhaps given multi-fidelity Sim options
- The **HPC Model** represents HPC grid aspects for optimally scheduling batches of Sim runs



ASHI uses Bayesian inference to formulate & update these models as Sim-run results arrive

Surrogate & information models for inferring QV

- Since our interest is in QV and we can only run our Sim with specific inputs x to produce a result Z_x, we choose those candidate Sim runs for which the *mutual information* (MI) between Z_x and QV is predicted to be high, using current data and the chosen SM:

$$I(QV;Z_x) = \sum_{QV}\sum_{Z_x} p(QV,Z_x)\log\frac{p(QV,Z_x)}{p(QV)p(Z_x)}$$

- We assume that QV and Z_x are independent given some parameter θ that is relevant to both the query and the Sim run:

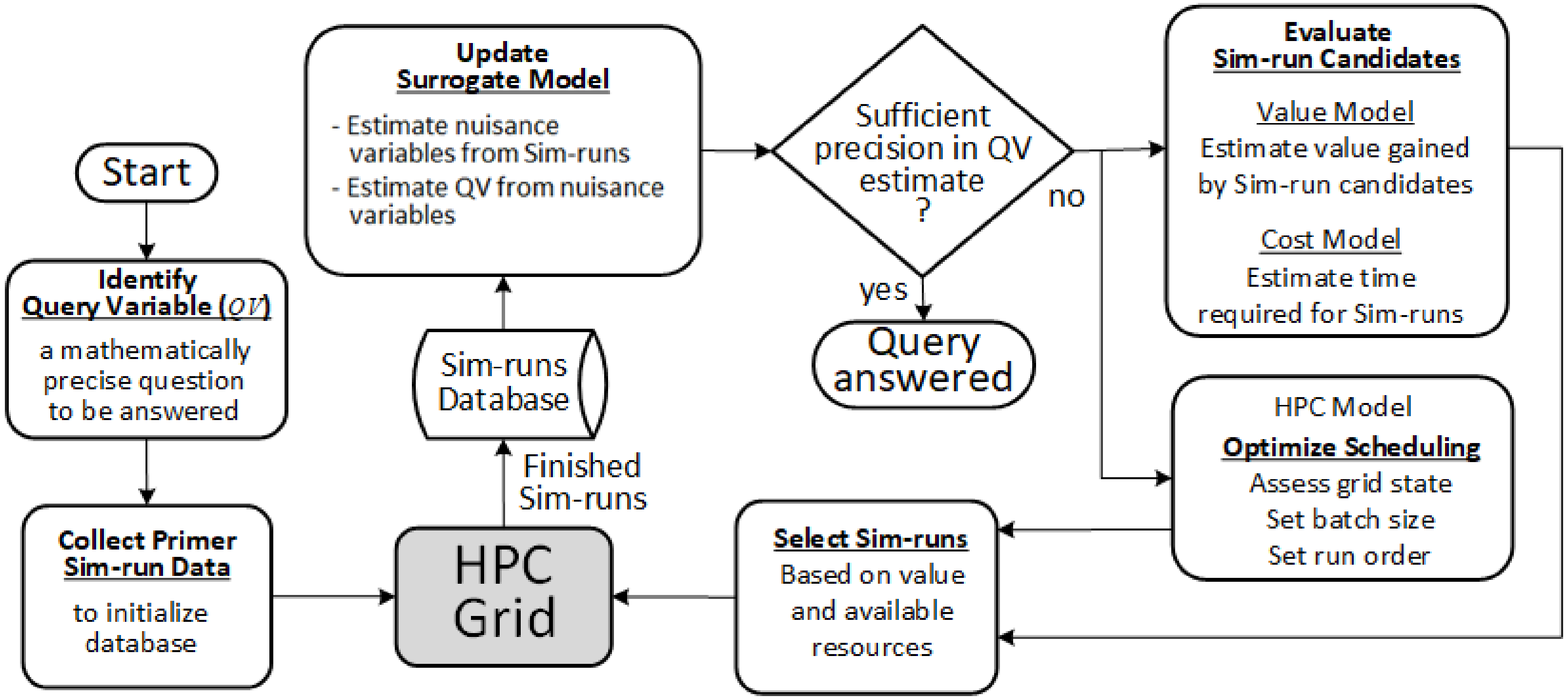
$$p(QV,Z_x) = \int p(QV,Z_x|\theta)p(\theta)d\theta = \int p(QV\,|\theta)p(Z_x|\theta)p(\theta)d\theta$$

- Using a point estimate of θ would result in MI = 0 for all candidate Sim runs ($\log(1) = 0$)
- Therefore, to facilitate estimating MI, we must employ surrogate models (SMs), $f(\theta, x)$, that produce a full posterior distribution over θ

Example SMs for estimating MI: Bayesian linear and generalized linear regression models, Continuous Correlated Beta Process, and Gaussian Process regression

The ASHI loop has been applied to several use cases, e.g.,

1. Compliance with a system performance requirement
2. Response sensitivity to control factors
3. Engineering design optimization



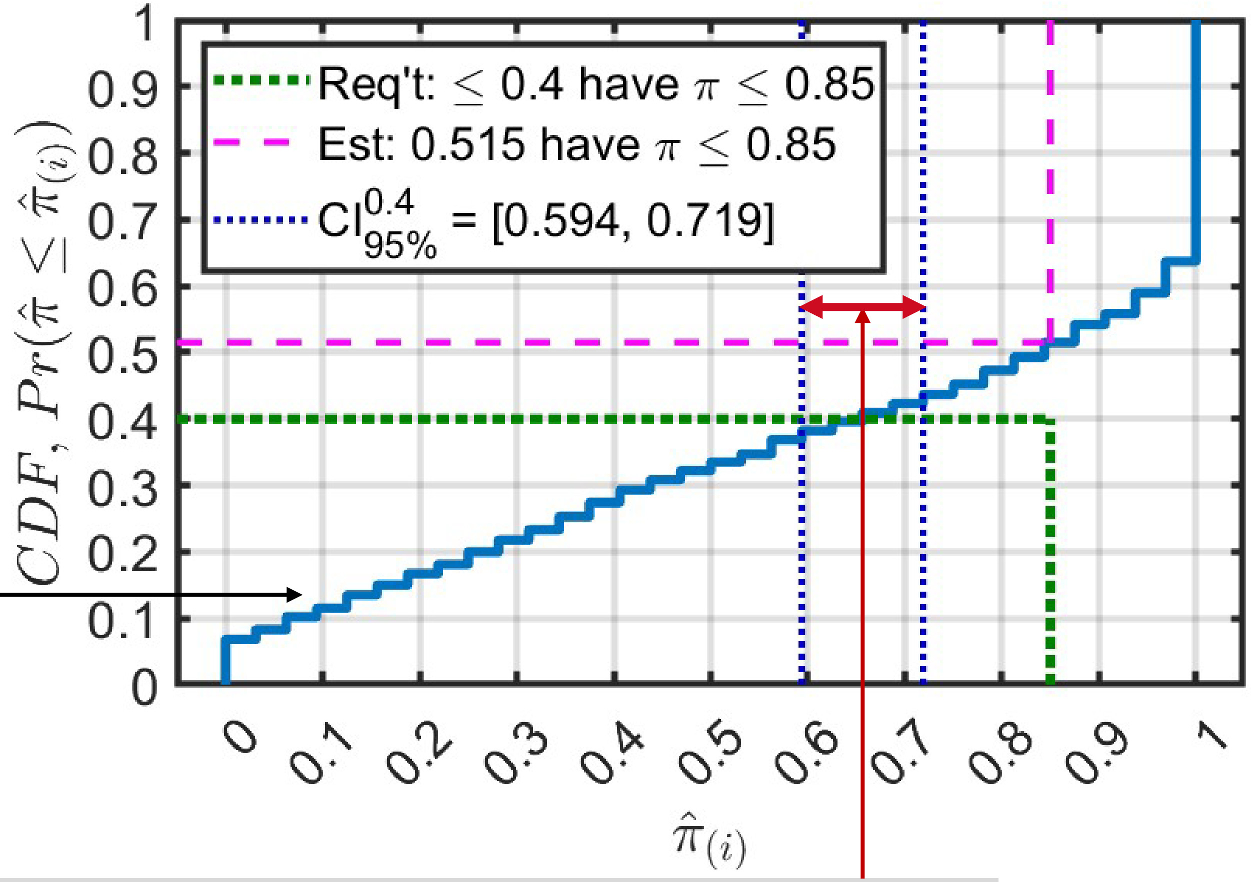
Details for these simple examples follow; for more, see the paper

Example 1: Performance requirement compliance

Raytheon Missiles & Defense

- Query: Quantify the probability of complying with the requirement, "Probability of system success shall be $\pi_S \geq 0.85$ for at least 60% of the system's engagement space," defined within given values of 5 factors: target range, target altitude, relative launcher-to-target altitude, launcher pitch, and launcher heading

Using *a priori* sampling	# treatments	1000
	# MC replicates per treatment	32
	Max-Likelihood Surrogate Model	Summary Statistic
Using ASHI	$p(\theta \mid y)$ Surrogate Model	Continuous Correlated Beta



- Empirical CDF of maximum-likelihood estimates of $\hat{\pi}_{(i)}$ for 1000 *a priori* LHS treatments, each replicated with 32 Monte Carlo trials (ASHI results on later slides)

Non-compliance is clearly indicated at the 95% confidence level

"ASHI Lite" for PNC: Variable # trials for *a priori* treatments

- Estimated probability of non-compliance (PNC):

$$PNC_i = 1 - p(\pi_i \geq 0.85 \mid y, n), y = 0, 1, \dots, n$$

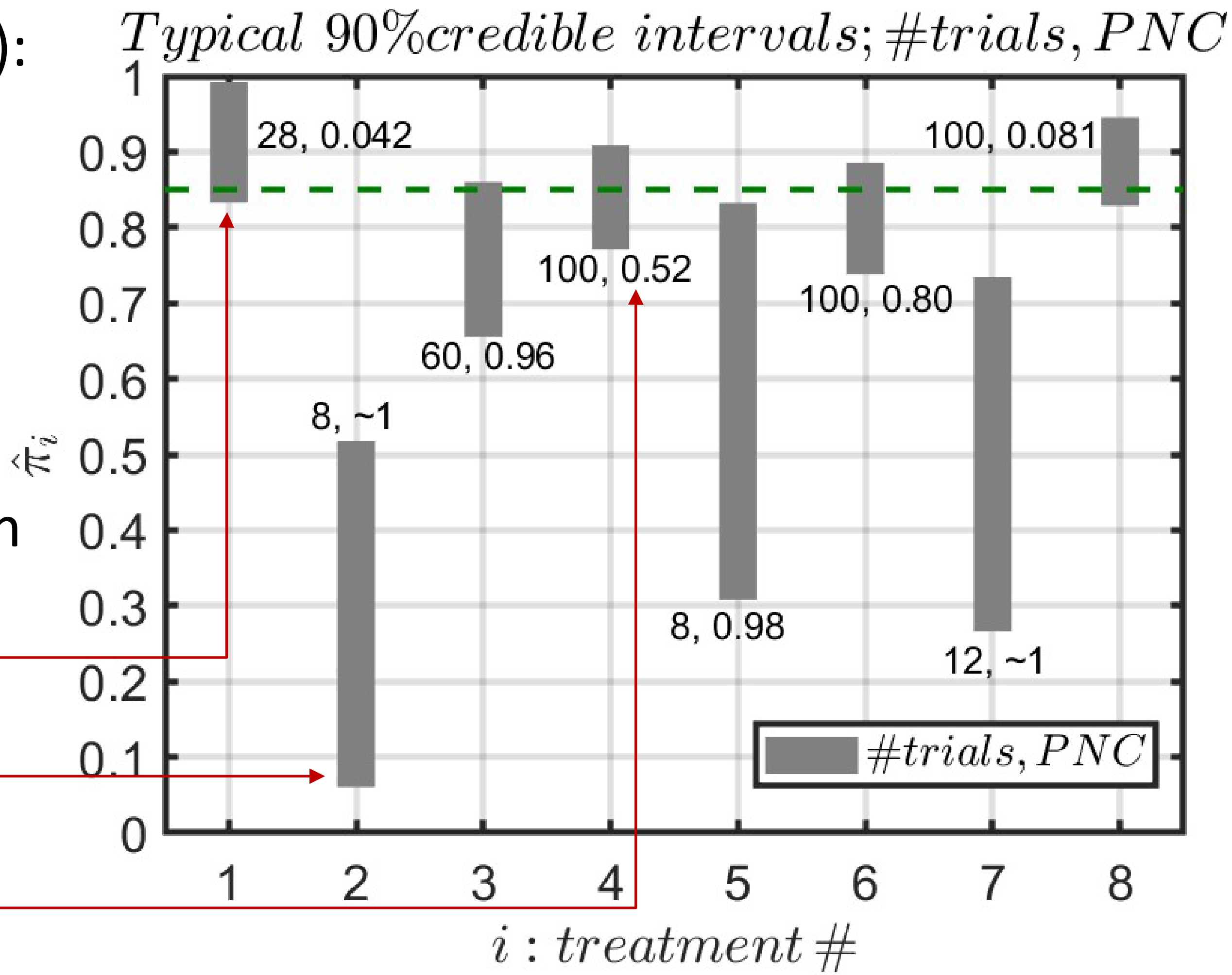
- Limited the number of treatment trials to

$$8 \leq n \leq 100$$

- Used a $Beta(2, 2)$ conjugate prior for binomial trials of each *a priori* treatment. Suspended trials of the i^{th} treatment after 8 runs and when reaching either limit,

$$PNC_i \leq 0.05$$
or
$$PNC_i \geq 0.95$$

- 155 of the 1000 LHS *a priori* treatments remained un-resolved after 100 trials



Could use some of the 6852 *a priori* runs saved to resolve more of un-resolved treatments

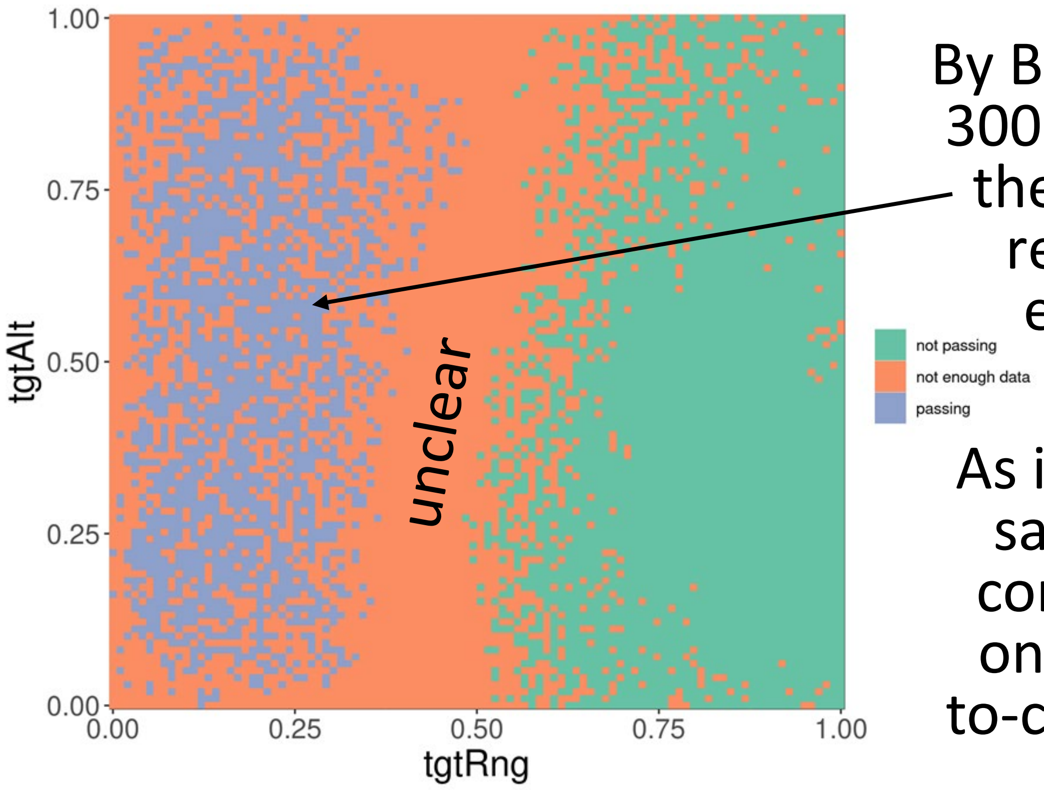
ASHI for Example 1 (requirement compliance)

Raytheon Missiles & Defense

- SM: The Continuous Correlated Beta Process (CCBP) uses a kernel-smoothing function $K(x_i, x)$ to model spatially correlated probability of system success π_S:

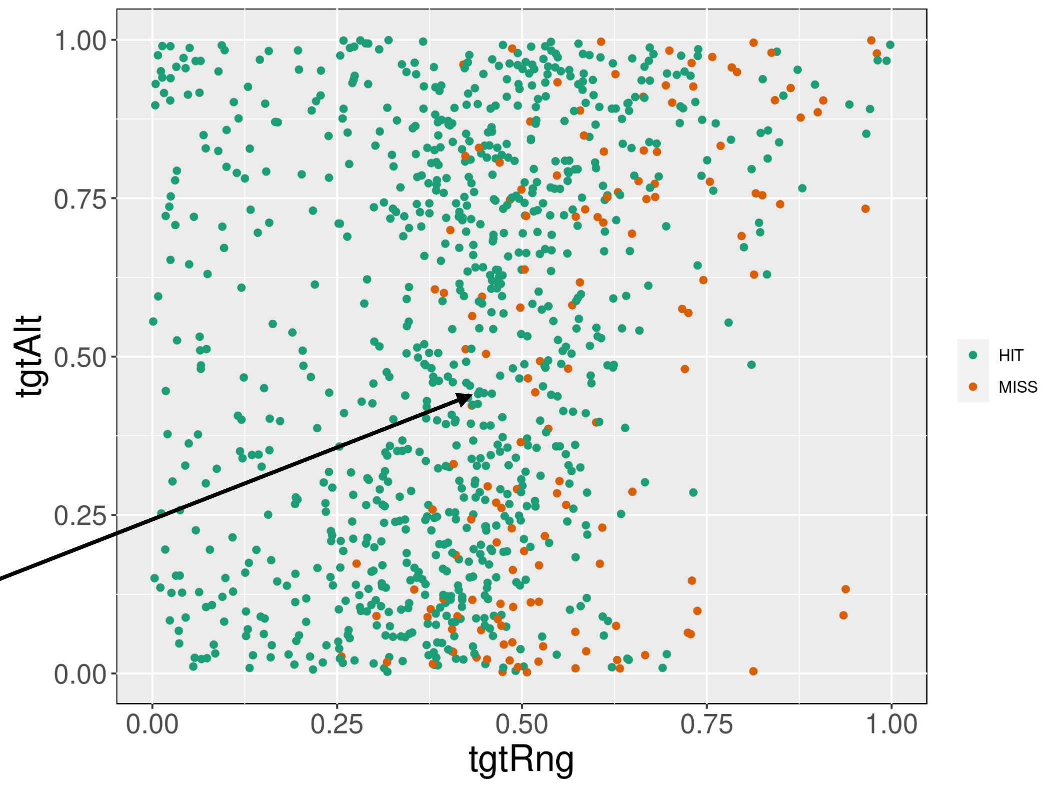
$$p\left(\pi_S(x) \middle| Z_{x_1} = z_1, Z_{x_2} = z_2, \cdots, Z_{x_n} = z_n\right) = Beta\left(\left[\alpha(x) + \sum_{i=1}^{n} \delta(z_i = 1) K(x_i, x)\right], \left[\beta(x) + \sum_{i=1}^{n} \delta(z_i = 0) K(x_i, x)\right]\right)$$

- Used CCBP and MI to down-sample 1000 runs from 10,000 candidate LHS runs per batch



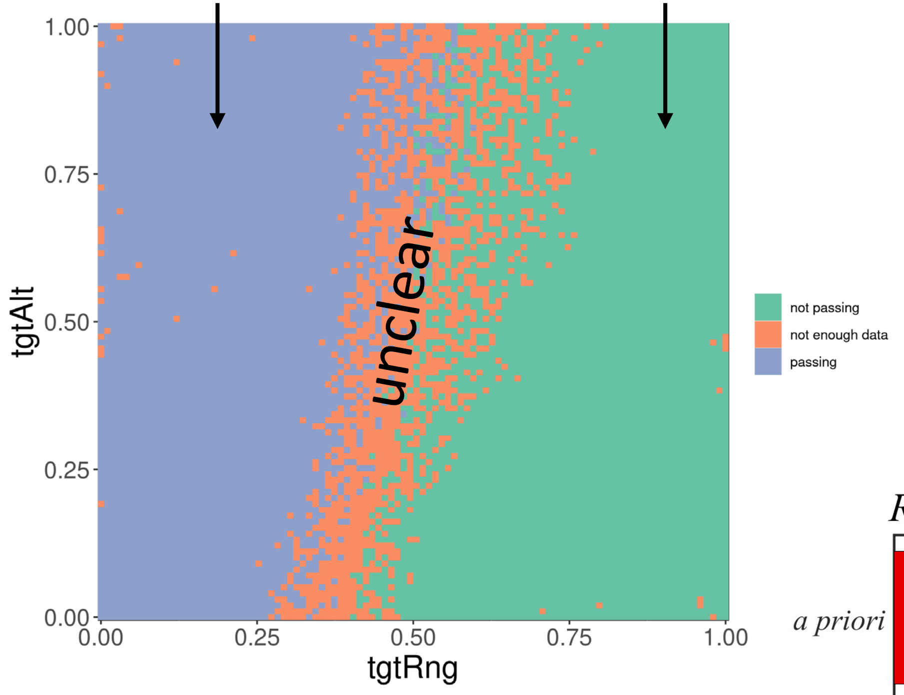
By Batch 3 (after 3000 Sim runs), the "passing" region was emerging

As in ASHI Lite, samples are concentrated on too-close-to-call interface

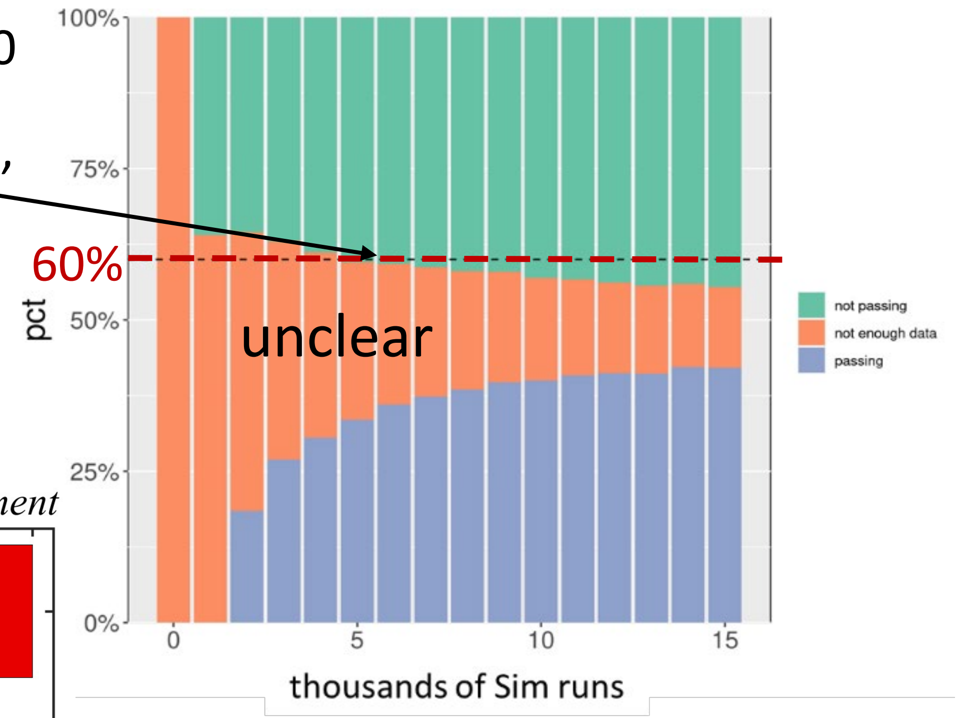


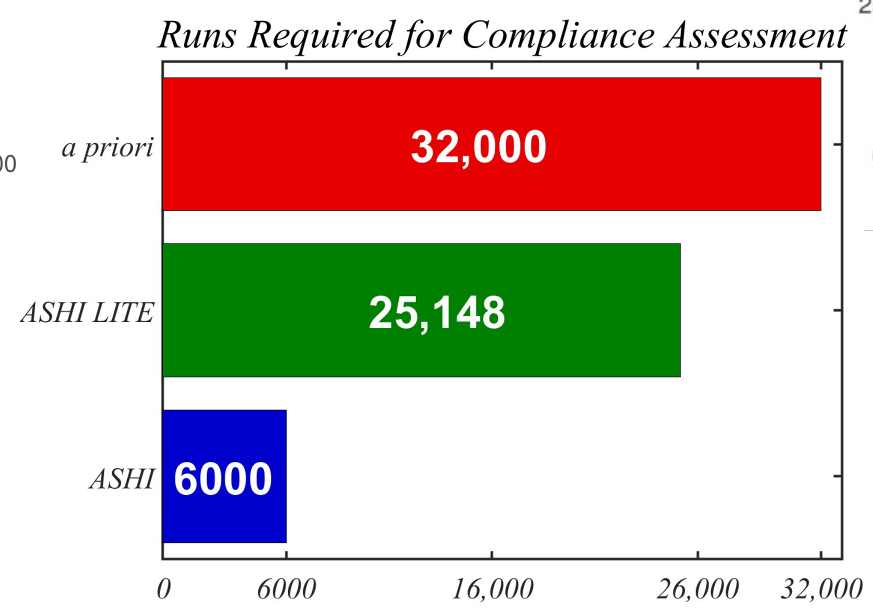
ASHI results, Example 1 (requirement compliance)

Main regions were known by Batch 15 (after 15,000 Sim runs)

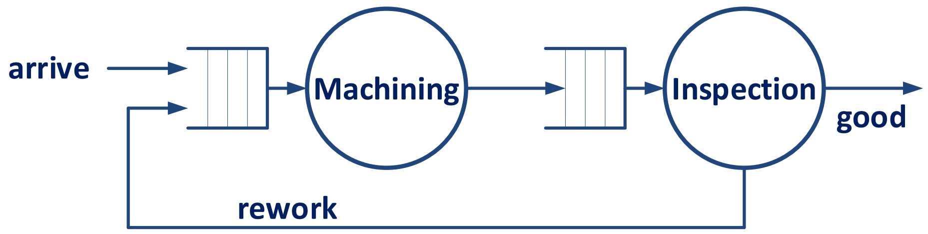


By Batch 6 (after 6000 Sim runs), non-compliance was clear, i.e. more than 40% were not passing

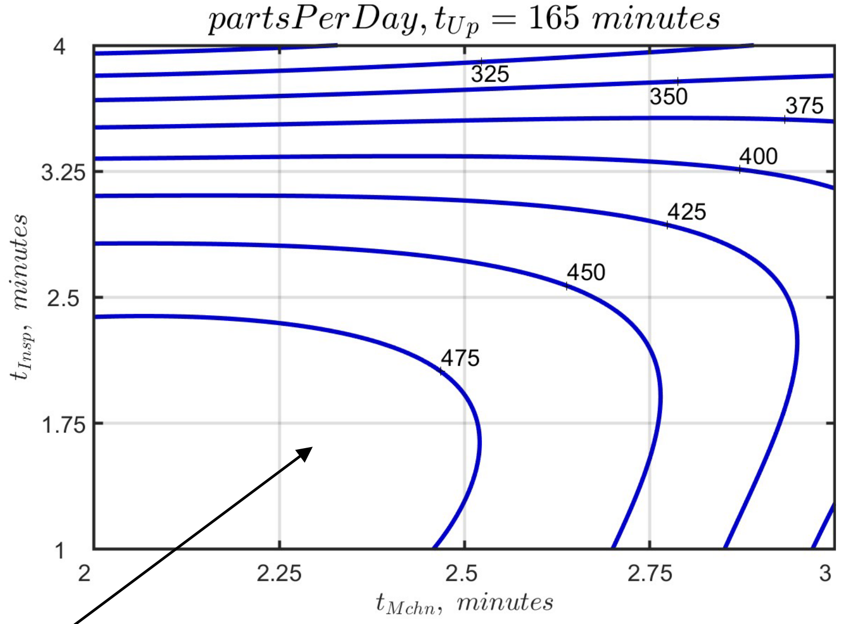




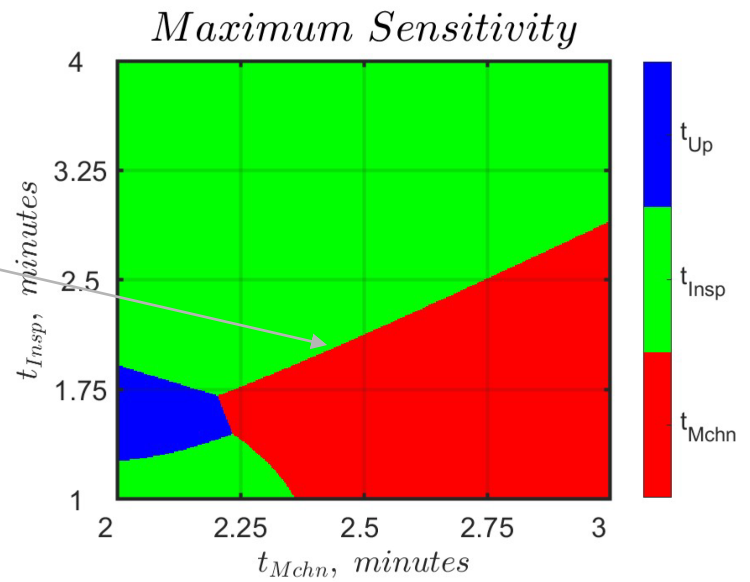
Example 2: Sensitivity analysis for factory throughput



Using *a priori* sampling	# treatments	200
	# MC replicates per treatment	32
	Max-Likelihood Surrogate Model	OLS Regression
Using ASHI	$p(\theta\|y)$ Surrogate Model	Bayesian Poisson Regression



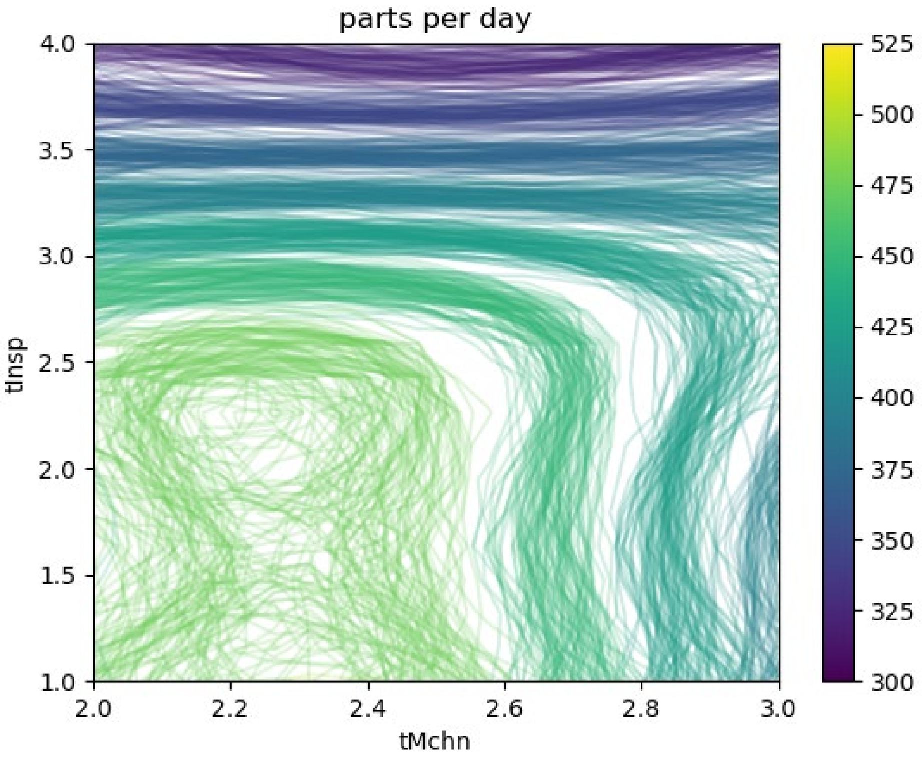
- Query: Find the factor to which throughput is the most sensitive: machining, inspection, and up-times (t_{Mchn}, t_{Insp}, t_{Up})
- Fitted a 3rd-order max-likelihood least-squares model to results from 200 *a priori* LHS treatments, each replicated 32 times
- Max-sensitivity region colors denote which of the 3 factors has the highest impact on system throughput; note crisp boundaries

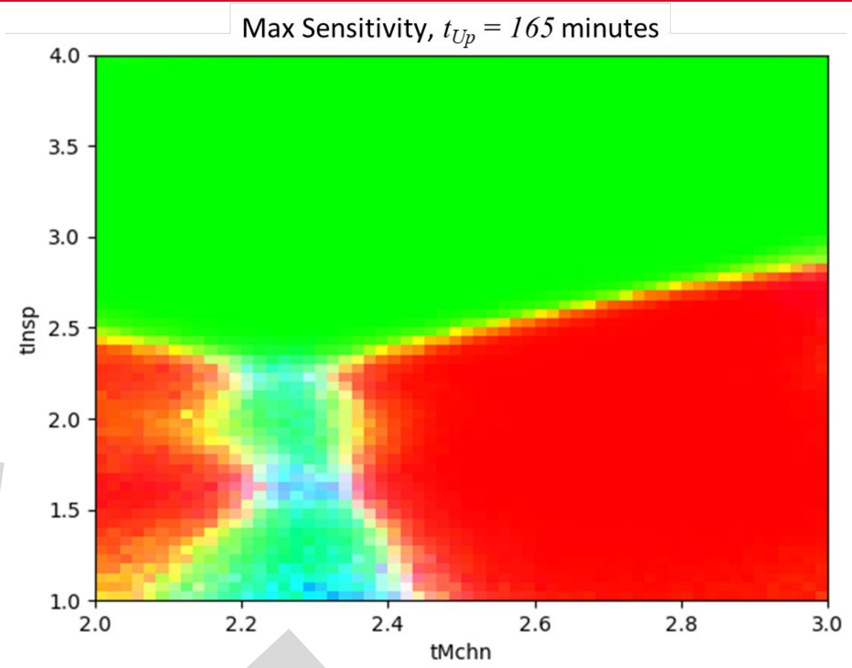


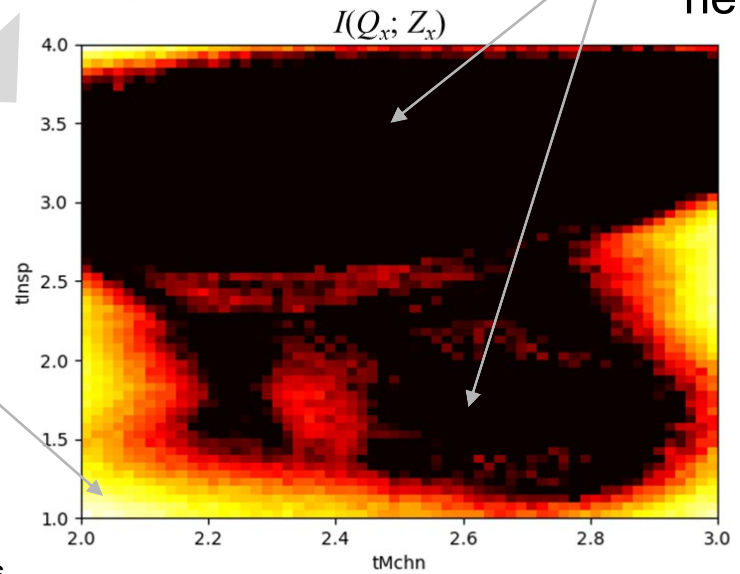
Adaptive sampling should choose to sample where we have high uncertainty in max sensitivity

Example 2: Mutual Information after 24, 25-run ASHI batches

- SM: Gaussian Process regression
- Sensitivity to factor i: $S_i = |\frac{\partial P}{\partial x_i}|$ (P = parts per day)
- Maximum sensitivity: $\mathrm{argmax}_i\, S_i$







Not worth sampling here

High MI value of sampling where we're the most confused about max sensitivity

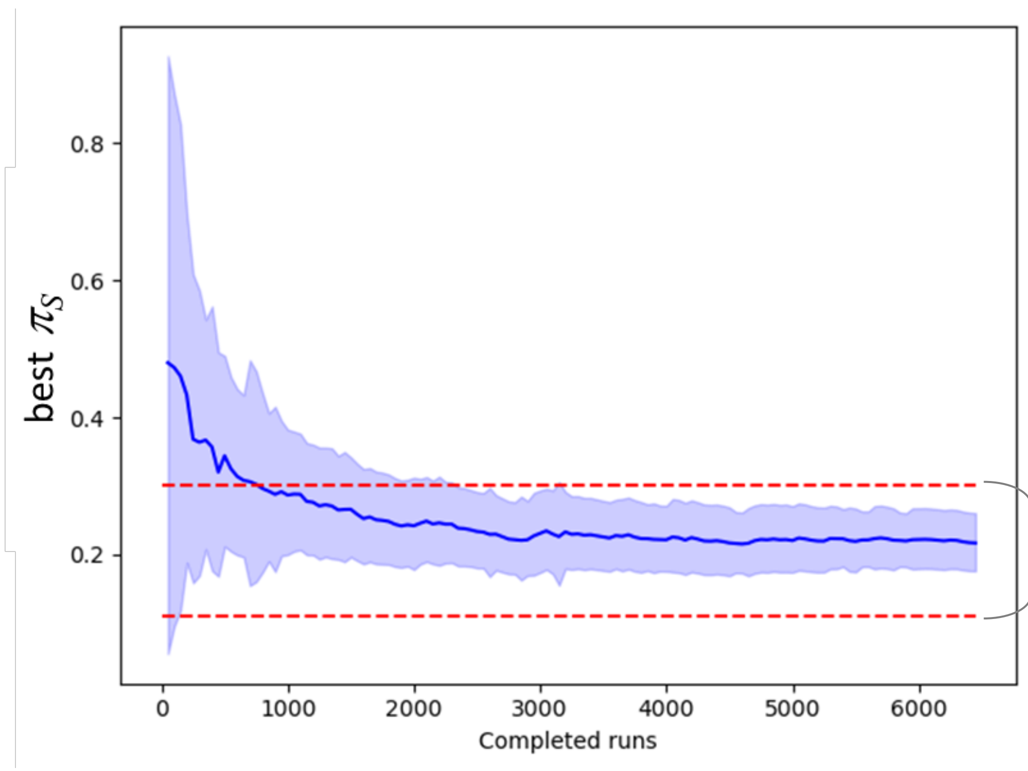
Example 3: Engineering design optimization

- Used a 6-degree-of-freedom (6DoF) Sim
- Query: “What values of guidance loop variables maximize success π_S?” Four design variables: initial gain g_1, initial time-to-go switching t_1, final gain g_2, and final time-to-go switching t_2
- Stopping criterion: At the optimal guidance settings, width of 95% credible interval for best π_S must be less than 0.10
- Unable to achieve a useful max-likelihood surrogate model using 200 *a priori* LHS treatments replicated 64 times – figure shows Bayesian analysis of *a priori* data, considering treatments as independent (not using SM)

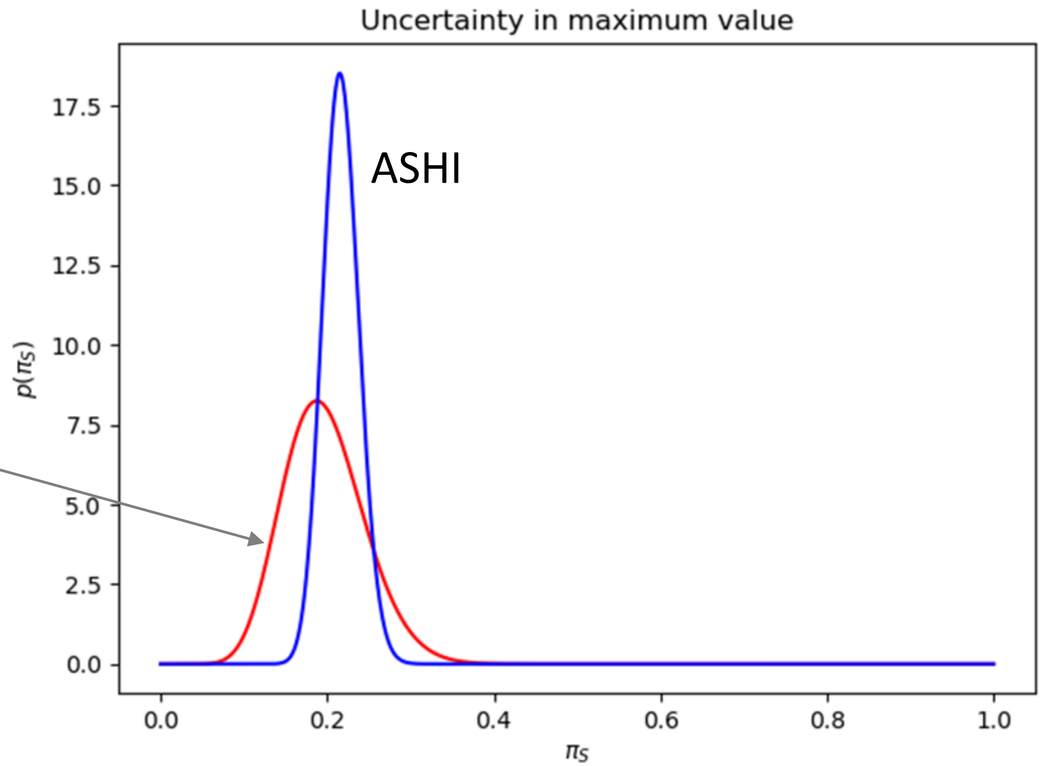
Using *a priori* sampling	# treatments	200
	# MC replicates per treatment	64
	Max-Likelihood Surrogate Model	Logistic regression not usable
Using ASHI	$p(\theta \mid y)$ Surrogate Model	CCBP

Did not meet 0.10 stopping criterion with 12,800 *a priori* runs

ASHI results, Example 3 (engineering design optimization)



$p(\pi_S)$
and limits on
best π_S reached
using *a priori*
sampling



Using localized MI value in search of argmax π_S rapidly decreased uncertainty in max π_S

- ASHI achieved desired 95% credible interval width of 0.10 for best π_S after 5850 runs, vs. not being achieved after 12,800 *a priori* runs
- ASHI achieved same uncertainty in 2350 runs as 12,800 *a priori* runs

Summary

- For credible MSA, the experiment's query must be precisely posed and uncertainty quantified
  - Surrogate models play a key role in quantifying uncertainty in the query results
  - ASHI efficiently answers the query using both a surrogate model and mutual information to choose where to sample given current Sim-run results
- The benefits of using the ASHI loop for adaptive sampling vary depending on the query
  1. Compliance example: 80% savings in runs
  2. Sensitivity Analysis example: Higher certainty in system's max-sensitivity regions
  3. Engineering Design Optimization example:
     - Saved > 50% runs to achieve the desired uncertainty level in best performance
     - Saved > 80% runs to achieve a level of uncertainty similar to *a priori* sampling

The ASHI loop is a major step forward for MSA that enables efficient sampling for effective, risk-informed decision-making