SANDIA REPORT
SAND2021-6172
Printed Click to enter a date

Sandia National Laboratories

Gemma V&V/UQ/Credibility Activities – FY20 Progress

Casey Jelsema, John Red-Horse, Brian Rutherford, Gabriel Huerta, Aubrey Eckert

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

This report describes the credibility activities undertaken in support of Gemma code development in FY20, which include Verification & Validation (V&V), Uncertainty Quantification (UQ), and Credibility process application. The main goal of these activities is to establish capabilities and process frameworks that can be more broadly applied to new and more advanced problems as the Gemma code development effort matures. This will provide Gemma developers and analysts with the tools needed to generate credibility evidence in support of Gemma predictions for future use cases. The FY20 Gemma V&V/UQ/Credibility activities described in this report include experimental uncertainty analysis, the development and use of methods for optimal design of computer experiments, and the development of a framework for validation. These initial activities supported the development of broader credibility planning for Gemma that continued into FY21.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the work completed by Nevin Martin, formerly of Sandia National Laboratories, as a member of the Gemma Credibility team. The authors would also like to thank electromagnetics SMEs Brian Zinser, Salvatore Campione, Rob Pfeiffer, Joe Kotulski, and Adam Jones for their teamwork, engagement, and support of this work.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page left blank

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
CompSim	Computational Simulation
EM	Electromagnetics
PLOAS	Probability of loss of assured safety
QoI	Quantity of Interest
Q-Q	Quantile-Quantile Plot
SA	Sensitivity Analysis
SE	Shielding Effectiveness
UQ	Uncertainty Quantification
V&V	Verification & Validation

1. INTRODUCTION

This report describes the credibility activities that have taken place in parallel to Gemma code development in FY20. The credibility activities undertaken in support of Gemma code development include Verification & Validation (V&V), Uncertainty Quantification (UQ), and Credibility process application. The main goal of these activities is to establish capabilities and process frameworks that can be more broadly applied to new and more advanced problems as the Gemma code development effort matures. This will provide Gemma developers and analysts with the tools needed to generate credibility evidence in support of Gemma predictions for future use cases.

The credibility team supporting the Gemma code development effort is made up of a cross-discipline group of experts including V&V/UQ practitioners, Dakota experts, and statisticians. This team leverages the expertise and experience across these disciplines to develop strategies and solutions to support the Gemma code development effort. The efforts undertaken by this team include support of both experimental and computational simulation activities.

The credibility activities undertaken in support of Gemma development have been tied to an overall credibility process strategy that is structured around the credibility tools and methodologies that have been developed by department 1544. The computational simulation (CompSim) credibility process involves assembling and documenting evidence to ascertain and communicate the believability of predictions that are produced from computational simulations [1]. The components of the CompSim credibility process in which activities have been undertaken in support of the Gemma code development effort in FY20 include experimental activities to quantify experimental result uncertainty, the development of a methodology for determining the optimal design of computer experiments, and the development of a validation framework methodology and the application of this methodology to a problem of interest. Verification activities have also taken place in parallel to these activities and are described in a separate report [2]. Each of the credibility activities undertaken in FY20 is described in the sections that follow. These activities are tied to the development and execution of a larger Gemma V&V/UQ/Credibility plan that will be documented in FY21.

2. EXPERIMENTAL ACTIVITIES

2.1. Q-Killer Experimental Results & Analysis

An experiment "Q-killer" was planned and executed in order to quantify uncertainty in several quantities of interest (QoIs). As a test instrument this experiment used a long-slotted cylinder with an absorber material. The test was performed under several experimental conditions: (1) Using 1 mm absorber, 2 mm absorber, or 1 mm absorber with a D-Dot variation[1]; (2) Two modes, at 1.13 GHz and 1.23 GHz.

Prior to modeling, we first investigate the distribution of the results. Quantile-quantile (QQ) plots are commonly used to determine the suitability of a distribution for a given dataset. Given the small sample sizes, however, we pool the data in the following manner: For each treatment (combination of factor levels size and mode), we standardize the data by subtracting the mean and dividing by the standard deviation. If a normal distribution is satisfied, this will transform all of the data to a N(0,1) distribution, and the combined data would also represent a normal distribution.

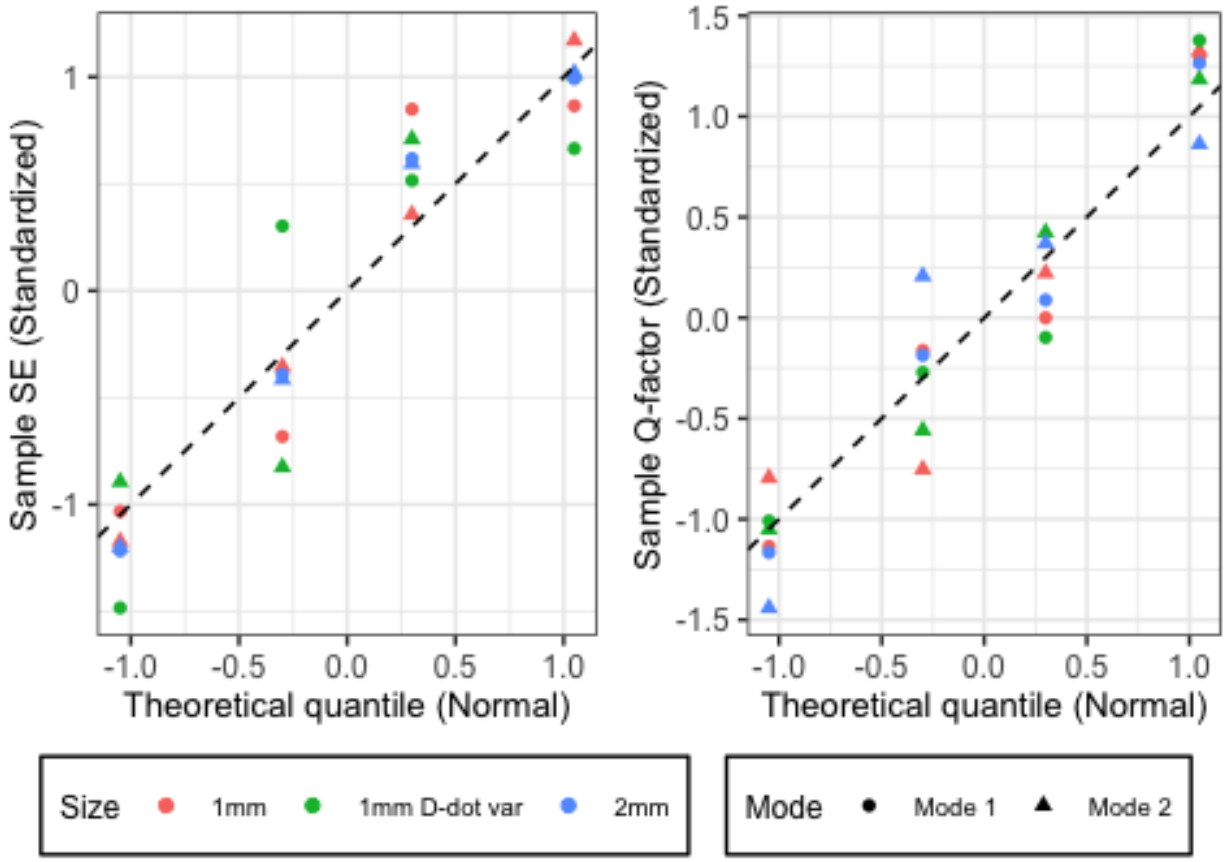

Figure 2-1 Quantile-quantile plots for SE and Q-factor.

In Figure 2-1 we provide the QQ plots for both shielding effectiveness (SE), which is a measure of the electric field in a cavity relative to the incident exterior field strength [5], and the quality factor

[1] D-Dot is the AD80 d-dot free field sensor. This sensor is connected to a hybrid coupler, and data is collected from both the sum and delta ports.

(Q-factor), which describes how good a resonator the cavity is at a given resonant frequency (a higher Q-factor will be associated with a sharper resonant peak in an SE-vs-frequency plot) [6]. The data for each is within the range of that which is expected from ordinary variation from a normal distribution. Hence, we will assume that a normal likelihood for the data is a reasonable approach.

To model the QoIs (SE and Q-factor) we construct a Bayesian hierarchical model. For explanation we refer to SE, but the same framework is applied to the Q-factor.
We specify a cell-means model, in which each treatment (a combination of size and mode) is modeled with a separate mean and separate standard deviation. There are 6 such combinations, which we denote with $i = 1, \dots, 6$. Then, letting $Y_{i,j}$ be the j^{th} measurement from the i^{th} treatment, we model $Y_{i,j}$ as below.

$$\begin{aligned} Y_{i,j}|\mu_i, \sigma_i^2 &\sim \mathrm{N}(\mu_i, \sigma_i^2) \\ \mu_i &\sim \mathrm{N}(\theta = 0, \tau^2 = 1e+05) \\ \sigma_i^2 &\sim \mathrm{InvGam}(0.001, 0.001) \end{aligned} \tag{2-1}$$

In this framework, the prior distribution for both the mean and the variance are considered "diffuse" or "non-informative," meaning the prior does not exert a strong influence on the results. This model is implemented through a Markov Chain Monte Carlo (MCMC) algorithm using the JAGS software. Running this analysis provides samples from the posterior distributions of μ_i and σ_i, that is, the distribution of these parameters after learning from the observed data. In addition, we can sample from the posterior predictive distribution of $Y_{i,j}$, which is analogous, but simulates the QoI after having updated the distributions of the parameters. These posterior and posterior predictive distributions inform about the behavior of the parameters or behavior of the QoI.

2.1.1. Results for SE

The posterior distributions of the mean and standard deviation of SE is provided in Figure 2-2, as well as numerical summaries in Table 2-1. These show a fairly striking difference between the 2 mm and 1 mm absorber thickness cases (with or without D-Dot variation), and between the two modes. Interestingly, there appears to be an interaction between the mode and the size because as, in comparison to mode 1, the 2 mm experiment has a smaller mean SE for mode 2, but both of the 1 mm experiments have a larger mean SE for mode 2. Additionally, the 1 mm without D-dot variation has a larger shift form mode 1 to mode 2 than the 1 mm with D-dot variation.

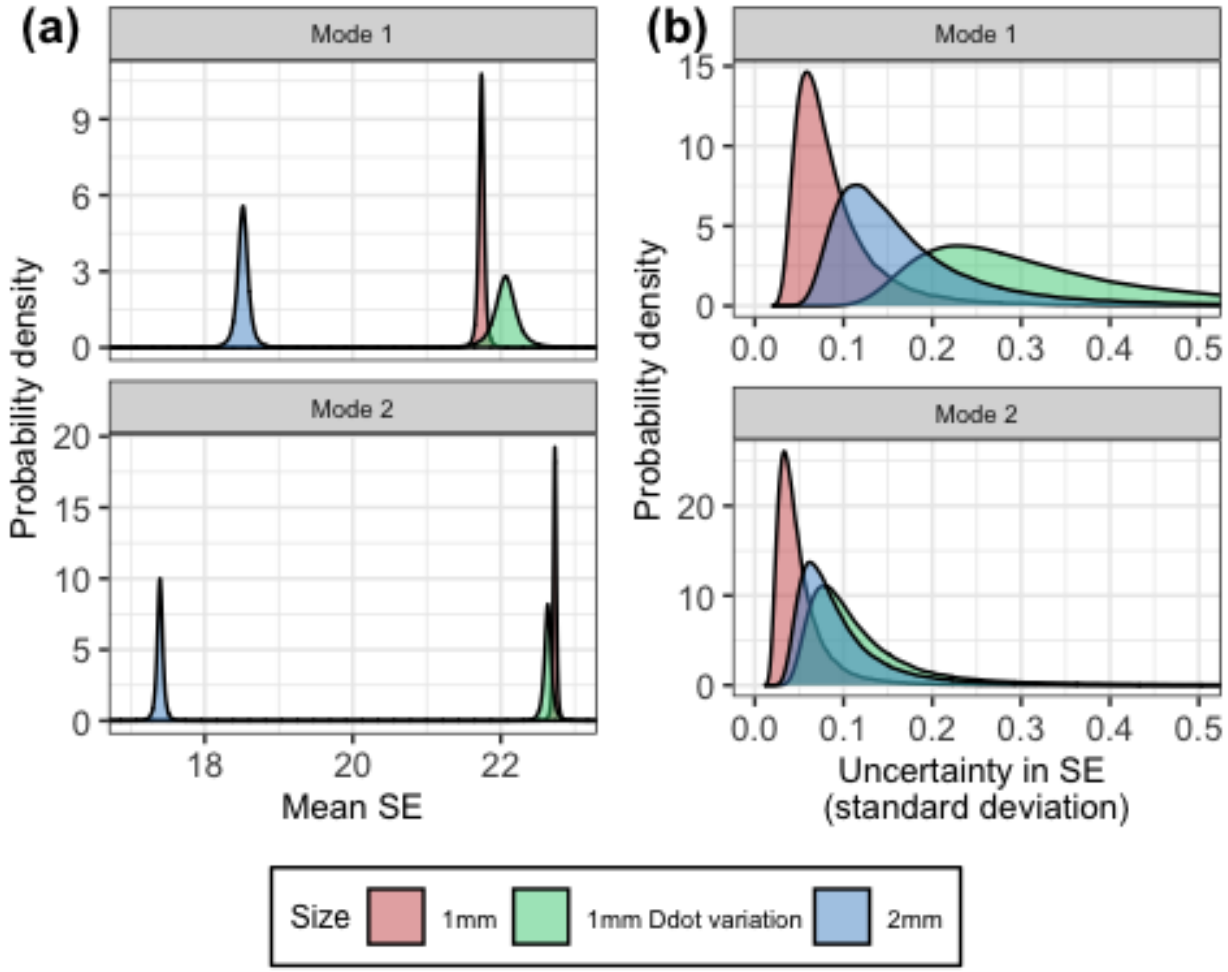

Figure 2-2: Posterior distribution of (a) mean SE and (b) uncertainty in SE.

There are also differences in the posterior distribution of the SE uncertainty (as measured by the standard deviation). In particular, the 1 mm absorber thickness without D-dot variation appears to minimize the uncertainty, though this is much less pronounced under Mode 2 setting.

Table 2-1: Summary statistics of posterior distribution of SE.

Parameter	Mode	Size	Mean	SD	+/- 1 SD Bound
SE Mean	Mode 1	1 mm	21.7404	0.0576	(21.6828, 21.7980)
		1 mm Ddot variation	22.0649	0.2209	(21.8440, 22.2858)
		2 mm	18.5163	0.1134	(18.4029, 18.6297)
	Mode 2	1 mm	22.7315	0.0336	(22.6979, 22.7651)
		1 mm Ddot variation	22.6378	0.0781	(22.5597, 22.7159)
		2 mm	17.3952	0.0626	(17.3326, 17.4578)
SE SD	Mode 1	1 mm	0.0929	0.0676	(0.0253, 0.1605)
		1 mm Ddot variation	0.3634	0.2629	(0.1005, 0.6263)
		2 mm	0.1804	0.1379	(0.0425, 0.3183)
	Mode 2	1 mm	0.0529	0.0412	(0.0117, 0.0940)
		1 mm Ddot variation	0.1242	0.0932	(0.0310, 0.2173)
		2 mm	0.1002	0.0824	(0.0178, 0.1826)

The posterior predictive distribution is shown in Figure 2-3, and summarized in Table 2-2. These show largely the same patterns, but do serve to illustrate the point that while the parameters may show substantial differences, there is more variability associated with individual values, and hence less distinction between the absorbers and modes (in particular, between the 1 mm and 1 mm with D-dot variation).

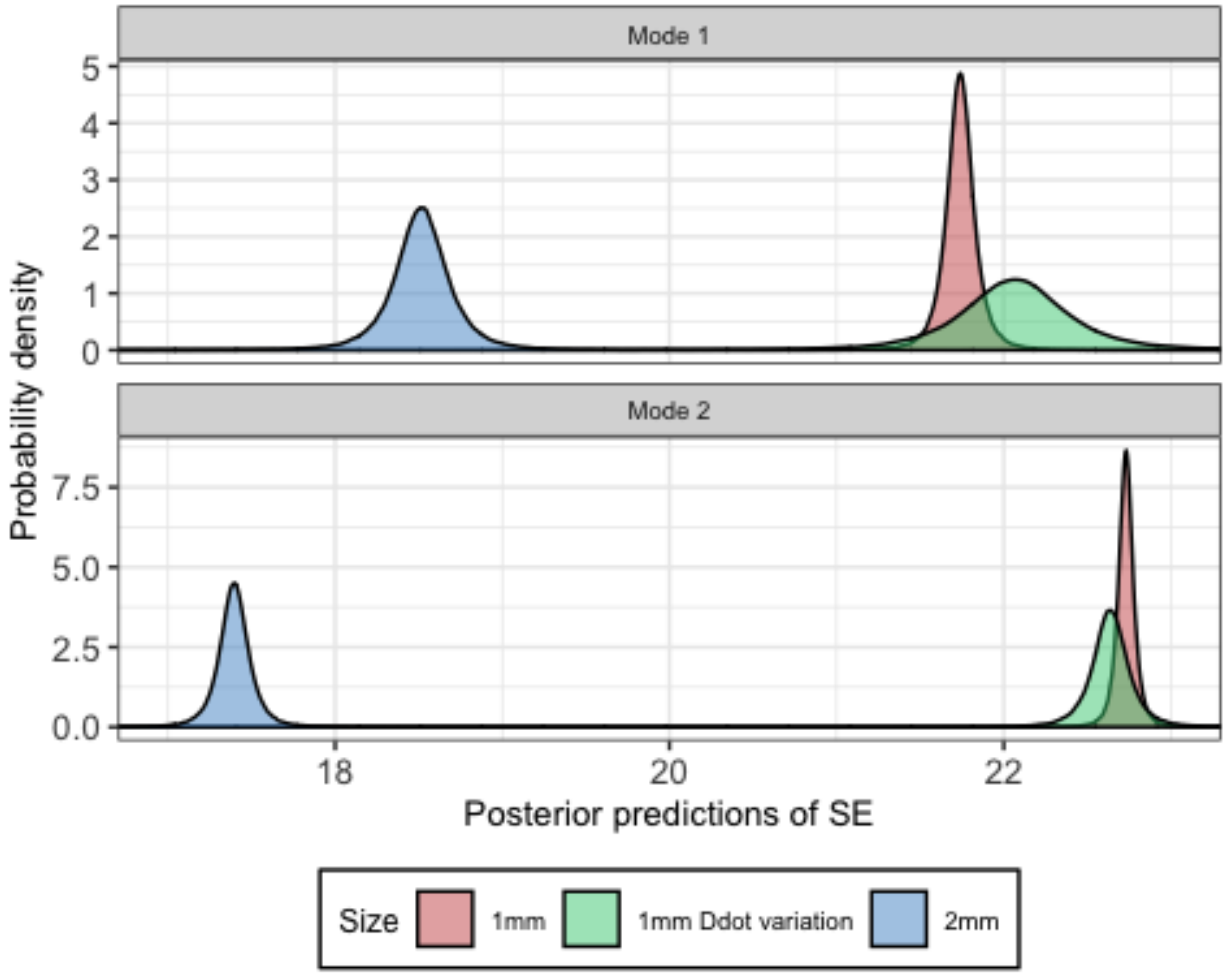

Figure 2-3: Posterior predictive distribution of SE.

Table 2-2: Quantiles of the posterior predictive distribution of SE.

Mode	Size	Percentile						
		0%	1%	5%	50%	95%	99%	100%
Mode 1	1 mm	16.947	21.400	21.564	21.740	21.917	22.082	25.719
	1 mm Ddot variation	-0.308	20.730	21.374	22.065	22.755	23.418	36.414
	2 mm	10.633	17.855	18.173	18.516	18.860	19.179	34.319
Mode 2	1 mm	16.154	22.536	22.632	22.732	22.832	22.924	27.803
	1 mm Ddot variation	17.036	22.181	22.402	22.638	22.875	23.097	30.803
	2 mm	5.814	17.026	17.205	17.396	17.586	17.764	27.050

2.1.2. *Results for Q-factor*

The posterior distribution of the mean Q-factor is provided in Figure 2-4, which again shows a distinct difference between the 2 mm vs 1 mm absorber thickness (with or without D-Dot variation), and between the two modes. Unlike with the SE, however, for the Q-factor the D-dot variations with the 1 mm absorber thickness did not appear to have an impact on the mean Q-factor; the posterior distributions for each are nearly coincident. Additionally, there does not appear to be any interaction with the mode, all three absorber thickness levels experienced a roughly similar reduction in Q-factor for Mode 2.

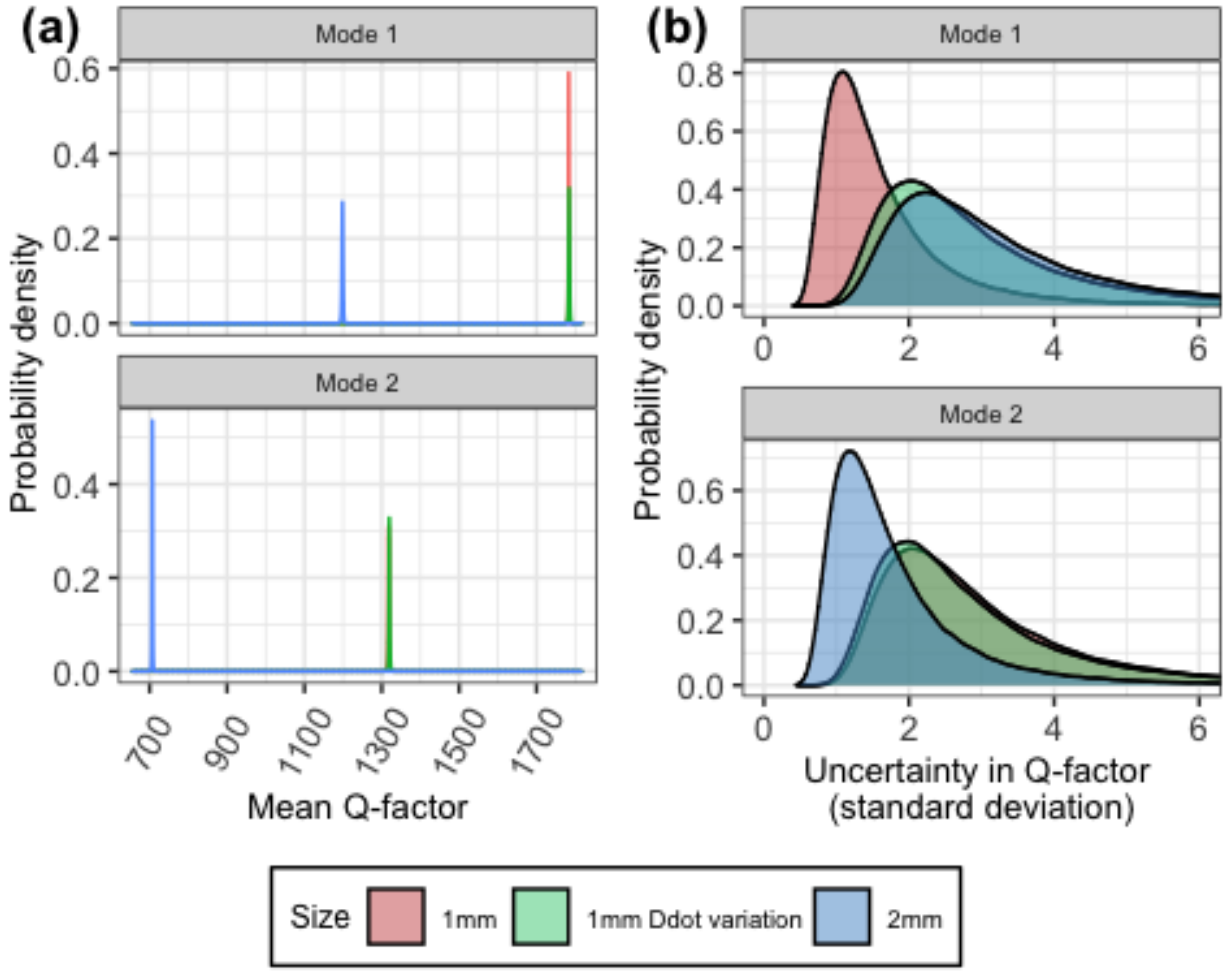

Figure 2-4: Posterior distribution of (a) mean SE and (b) uncertainty in Q-factor.

The standard deviation of the Q-factor exhibits an interesting pattern: the 1 mm with D-dot variation appears unchanged between the two modes, but the 1 mm and the 2 mm appear to swap: For Mode 1 the 1 mm thickness gives the smallest standard deviation, while for Mode 2 the 2 mm thickness appears to give the smallest (though neither of these would appear to be statistically significant).

Table 2-3: Summary statistics of posterior distribution of Q-factor.

Parameter	Mode	Size	Mean	SD	+/- 1 SD Bound
Q-factor Mean	Mode 1	1 mm	1783.2816	1.0485	(1782.2331, 1784.3301)
		1 mm D-dot variation	1784.1574	1.9619	(1782.1954, 1786.1193)
		2 mm	1198.4090	2.2179	(1196.1911, 1200.6269)
	Mode 2	1 mm	1317.7187	2.0261	(1315.6926, 1319.7448)
		1 mm D-dot variation	1319.4950	2.0849	(1317.4101, 1321.5799)
		2 mm	707.0319	1.1610	(705.8709, 708.1929)
Q-factor SD	Mode 1	1 mm	1.6970	1.2308	(0.4662, 2.9278)
		1 mm D-dot variation	3.1813	2.3169	(0.8644, 5.4982)
		2 mm	3.5327	2.6235	(0.9092, 6.1562)
	Mode 2	1 mm	3.2502	2.4056	(0.8446, 5.6557)
		1 mm D-dot variation	3.0896	2.4363	(0.6534, 5.5259)
		2 mm	1.8802	1.3844	(0.4958, 3.2646)

We provide a numeric summary of the posterior predictive distributions of the Q-factor in Table 2-4. Because of the large difference in mean Q-factor between the 2 mm thickness and the 1 mm thicknesses, and the relatively small uncertainties, a graphical representation does not provide much additional insight.

Table 2-4: Quantiles of the posterior predictive distribution of Q-factor.

Mode	Size	Percentile						
		0%	1%	5%	50%	95%	99%	100%
Mode 1	1 mm	1672.4	1777.1	1780.0	1783.3	1786.5	1789.4	1908.9
	1 mm D-dot variation	1606.0	1772.3	1778.1	1784.2	1790.2	1795.7	1894.1
	2 mm	1013.6	1185.3	1191.7	1198.4	1205.1	1211.1	1352.9
Mode 2	1 mm	1095.9	1305.7	1311.5	1317.7	1323.9	1329.6	1517.0
	1 mm D-dot variation	1026.6	1308.0	1313.6	1319.5	1325.3	1330.6	1475.2
	2 mm	592.0	700.1	703.4	707.0	710.6	713.9	771.5

3. UNCERTAINTY QUANTIFICATION & SENSITIVITY ANALYSIS

A major driver for the development of capabilities to support uncertainty quantification and sensitivity analysis for Gemma predictions is the desire to present prediction bounds for simulation results. An effort was undertaken to develop a methodology for selecting the optimal design of computer experiments, or simulation uncertainty runs, to develop uncertainty estimates. The development of this methodology is described in the following sections.

In addition to the methodology development described in Section 3.1, an effort was also undertaken to develop and pilot a driver to connect Gemma to Dakota. This effort will be documented along with broader credibility activity plans in FY21.

3.1. Optimal Design of Computer Experiments

3.1.1. General description

Selecting inputs for computer runs in a computational simulation project can be accomplished through the "Design of Computer Experiments". This process can be made substantially more efficient when a lower-fidelity code is available – in our case, we use a low-fidelity cavity coupling model to compute shielding effectiveness (SE) analytically for different values of three slot aperture parameters. This analytic model, described in [5], is based on conservation of electromagnetic energy and is referred to as a "power balance" method.

Getting at specific project objectives can often be accomplished by understanding specific quantities of interest (QoIs) with a reasonable level of precision. Examples of quantities of interest include the response maximum value, the inputs yielding the maximum response, the overall probability of failure, and the average response. When a lower-fidelity code is available, precision in the QoI can be approximated using that code, thus providing an estimate of the sampling requirements (number of computer runs needed) to achieve an acceptable level of precision when using the high-fidelity code.

It is often the case that a computer experimental design is accomplished in two phases. The first phase is used to get a rough approximation of the response over the entire input space and to get a reasonable initial estimate of the QoI. During the first phase a general sampling approach such as random sampling or Latin hypercube sampling (see [3]) can be employed to select input values for the low-fidelity computer runs. By examining results using different sample sizes, one can determine at what point a reasonable estimate of the QoI has been obtained. Generally, there is a point where further use of this sampling approach yields diminishing returns in establishing precision of the QoI estimates. The second phase is mainly focused on producing further computer runs on inputs that provide specific information more relevant to the QoI.

In FY20 we focused on both phases of this analysis but completed work mainly on the first phase. In this report, we give results for all accomplishments including some preliminary second phase results from a proposed approach to the design of computer experiments. Section 3.1.2 provides an overview of the response-modeling approach to computer experimental design. Section 3.1.3 shows how this approach is implemented using data generated using the power-balance model. We show results that constitute the first step in the design algorithm. More detail on the material in sections 3.1.2 and 3.1.3 is provided in [4]. Section 3.1.4 focuses on the first phase, getting reasonable precision on the initial estimate of the QoI.

3.1.2. Proposed Response-Modeling Methodology for Design of Computer Experiments

Over the past several decades, the use of computer codes supporting complex analyses in science and engineering has increased dramatically. In response, the statistics and engineering communities have developed a range of methods and tools for working with computer models in areas including prediction, uncertainty quantification, sensitivity analysis, parameter calibration and computer experimental design. Experimental design issues arise when the computer codes are complex and expensive to run. The challenge is to select the computer input levels that yield informative responses with limited computational effort. A good experimental design depends on the objectives of the analysis. The QoI is determined based on analysis objectives. Most QoIs either are functions of a computer response $y \in \mathbf{Y}$ or consist of a region or point in the k-dimensional computer input or design space $\mathbf{X}$. Examples of QoIs based on the response space include optimal response values, probability of a response below or above a specified threshold, mean or variances of the QoI, and quantiles of the response distribution. Examples of QoI in the input or design space include identification of inputs that yield an optimal response or identifying a failure region $F \subset \mathbf{X}$.

The response-modeling approach to computer experimental design can accommodate the general case where the objectives of the analysis are to estimate some QoI with as much precision as possible. In this approach, the computer response is modeled through an emulator (the "response-model") based on an ensemble of Gaussian process surfaces. The surfaces are generated probabilistically and conditioned on prior information. Figure 3-1 provides an example of three realizations of a response-model in two input dimensions. In actual application, the response-models consist of many more realizations, 20 or more in our applications. Note that for any set of input points $\mathbf{x}_1,..,\mathbf{x}_m \in \mathbf{X}$ the ensemble provides a discrete set of responses $[y_{i1}, \ldots, y_{im}]$ for each i, i = 1, 2, 3 in the case shown in Figure 3-2.

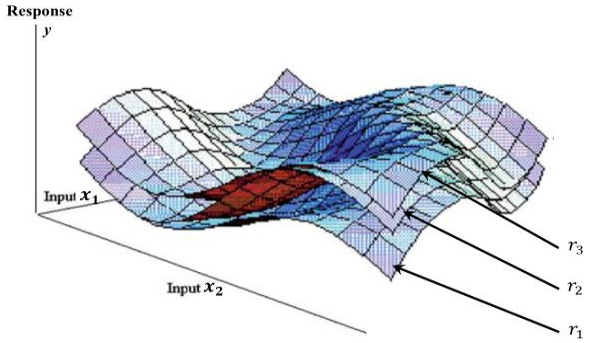

Figure 3-1. Illustration of a Response-Model.

Information contained in the response-models and their linkage to the QoI will impact the degree to which further experimentation at a point in the input space will prove informative. Three considerations that can influence the choice of an efficient experimental design are automatically accounted for through the response-modeling algorithm: the sensitivity of the QoI to the response at that point, the "closeness" in the input space to other computer experiments, and the relative likelihood of inputs at that point (where applicable – for some QoI, failure probability for example, the inputs will have an associated joint probability distribution).

A diagram illustrating the flow of the algorithm is given in Figure 3-2.

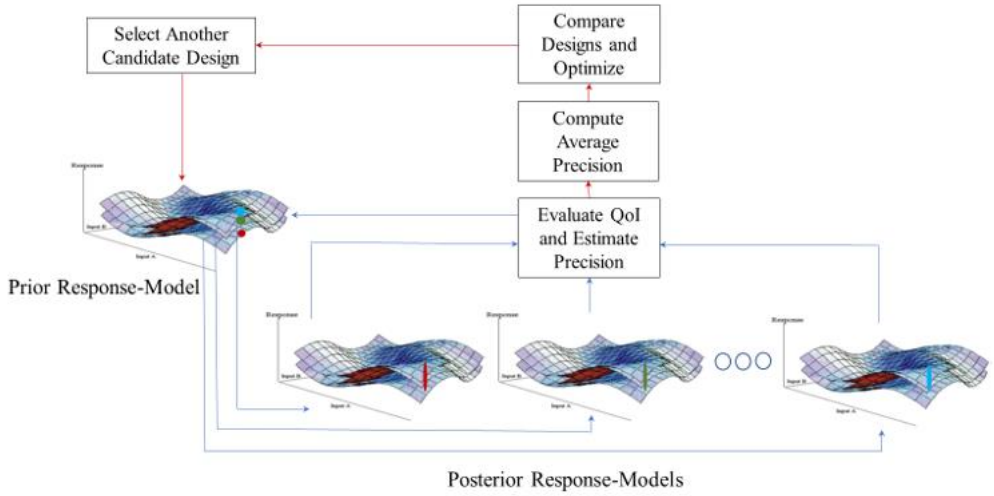

Figure 3-2. Flow Diagram for the Response-Modeling algorithm. The icons shown as multiple response surfaces represent the response-models.

The two processes indicated in the diagram address tasks that are common to most experimental design algorithms, namely, selection of candidate designs for consideration, and evaluation and comparison of the designs according to specific (problem-dependent) design criteria in search of an optimal design.

Response-models are used first to model prior information for parameters of the Gaussian processes used in their construction and to approximate uncertainty in the response for input locations where the computer code has not been run. The icon for the "prior response-model" is shown in the left of Figure 3-2. A single-point candidate design is indicated by the colored circles indicating different possible response values at that location calculated for different realizations. The "posterior response-models" are constructed iteratively to approximate the posterior response distribution. Note that there is one posterior response-model for each prior realization. Values in the prior response-model, at the design locations, are treated as part of the conditioning data for the corresponding posterior response-model. This is indicated by the coloring at the single design point location in the posterior response-models. The posterior response-models together provide the basis for estimating the QoIs and their expected precision and for evaluating and comparing candidate designs.

A summary outline of the methodology includes:

i) Candidate designs are selected.
ii) For each design, the prior response-model specifies probable response values at the design points through the realizations (one set of values per realization as shown in the colored circles for the single-point design).
iii) The set of values for each prior realization is used with other conditioning data for the posterior response-model corresponding to that realization (i.e. all realizations in that posterior response-model take on those same values at those design points).
iv) Each posterior response-model provides an estimate of the QoI and its precision.

v) The expected precision (averaged across posterior response-models) is computed and compared with other designs.
vi) The design yielding the best "relative expected precision" is selected.

Note that the process described above may be repeated in stages as part of an overall design strategy. Results from a set of computer experiments established at one stage become part of the conditioning data for the next stage.

3.1.3. Design of Experiments for Power Balance Code

In this section, we illustrate how the response-modeling algorithm could be used in the second phase of the experimental design process. Here we describe the design of computer experiment analyses related to the power balance code with 3 slot parameters as the main inputs: slot width, slot length and slot depth as described in [5]. For simplicity the 3D domain for the independent variables slot length, slot width and slot depth was limited to intervals where SE is monotonic. The parameters related to the size, shape and composition of the cylinder were set at fixed values. Because power balance code runs are so inexpensive, we can run it over a fine grid throughout the input space. Figure 3-3 illustrates the true (power balance calculated) relationship between SE and the three slot parameters calculated over an 11X11X11 grid. Yellow corresponds to higher values of SE while blue corresponds to lower values (note that SE, as defined here, is high when shielding is poor).

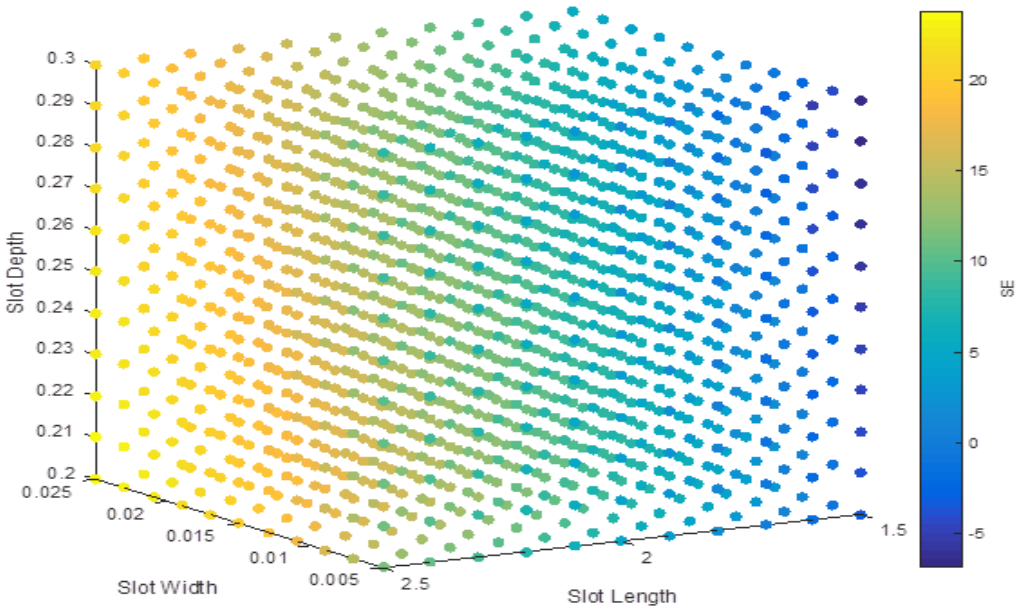

Figure 3-3. 3-D plot of shielding effectiveness (SE) in terms of the slot parameters for a high-resolution grid for the power balance code.

The first phase of the experimental design (see Section 3.1.1) was implemented to obtain a sample of size 40 of the three-dimensional slot parameter space. The sample was generated assuming normal distributions on the 3 inputs. Figure 3-4 shows the location of the Latin Hypercube samples colored by magnitude of SE.

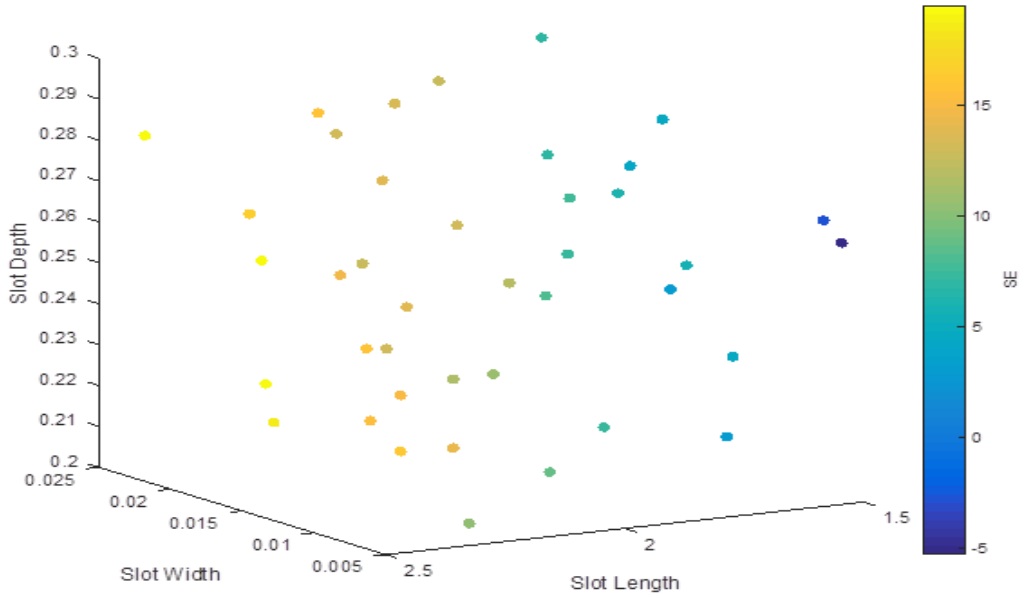

Figure 3-4. Location of 40 Latin-hypercube samples for the power balance code.

The response-modeling algorithm (see Section 3.1.2) is used to approximate the relationship between shielding effectiveness (SE) and the slot width, slot length and slot depth parameters. Figure 3-5 illustrates the response-model estimated relationship based on the 40 samples shown in Figure 3-4. Note that compared to the true values shown in Figure 3-3, the SE response is similar with respect to the slot width and slot length parameters but no longer monotonic with respect to the slot depth parameter.

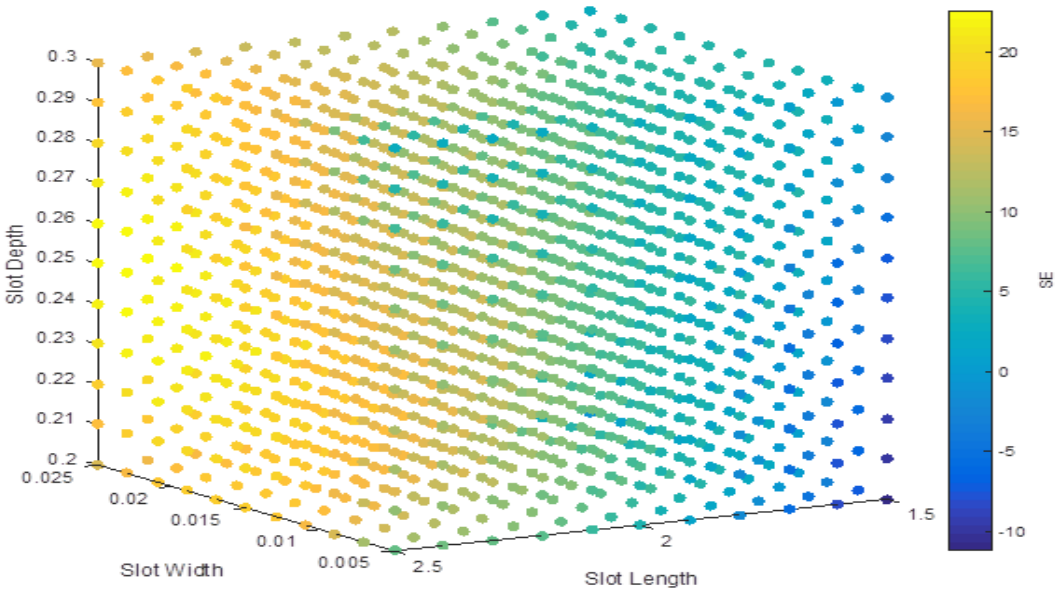

Figure 3-5. Estimated SE values calculated using the response-modeling algorithm based on the 40 Latin-hypercube samples shown in Figure 3-4.

The response-modeling algorithm finds an optimal design based on expected precision for specific QoI. For illustration, here, we show initial results for two QoI: (a) the 0.9 quantile and (b) the estimation of a "failure region", which is defined as all the point in input space for which values of SE are above 15.5. For each analysis, the expected QoI precision is assessed as a function of the 3 slot parameters to determine what computer experiment should be run next. Figure 3-8 maps the expected uncertainty for the estimation of the 0.9 quantile of SE, while Figure 3-9 shows the corresponding mapping of expected uncertainty for the estimation of the failure region. In both cases (Figure 3-8 and Figure 3-9) the more preferable design points in input space are those that align with smaller values of expected uncertainty. These results agree with our intuition. In Figure 3-6, the preferable design points (indicated by the dark blue) are close to the region where Figure 3-3 tells us that the maximum value is. In Figure 3-7 the preferable design points are further back toward the center of the input space at locations closer to the border of the failure region.

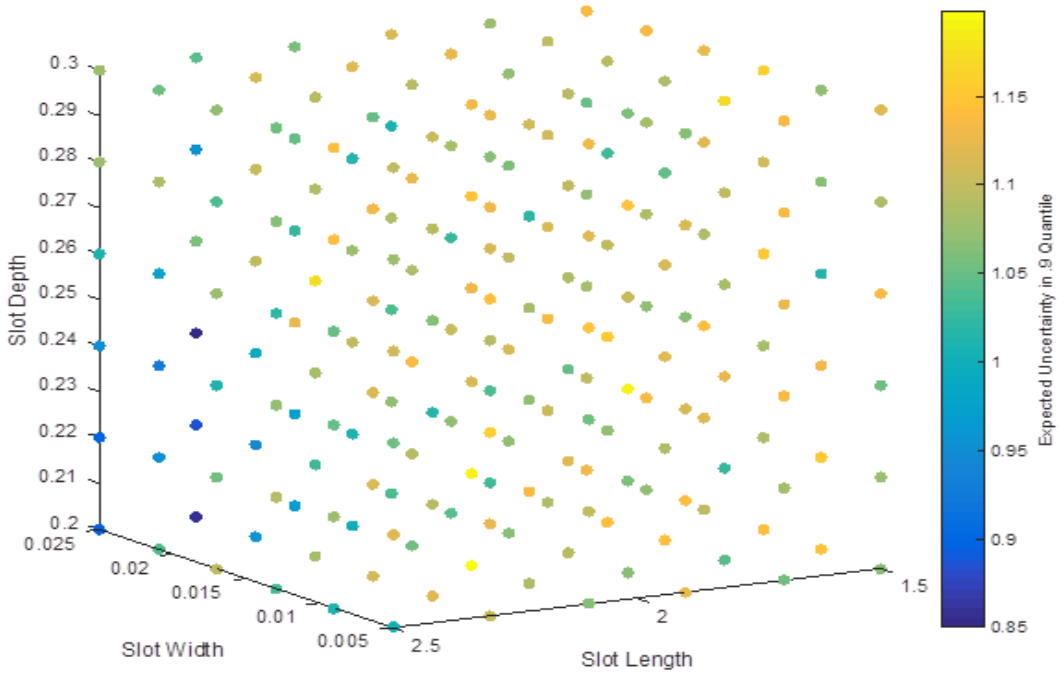

Figure 3-6. Expected uncertainty in 0.9 quantile. Lower values of expected uncertainty provide more preferable design points.

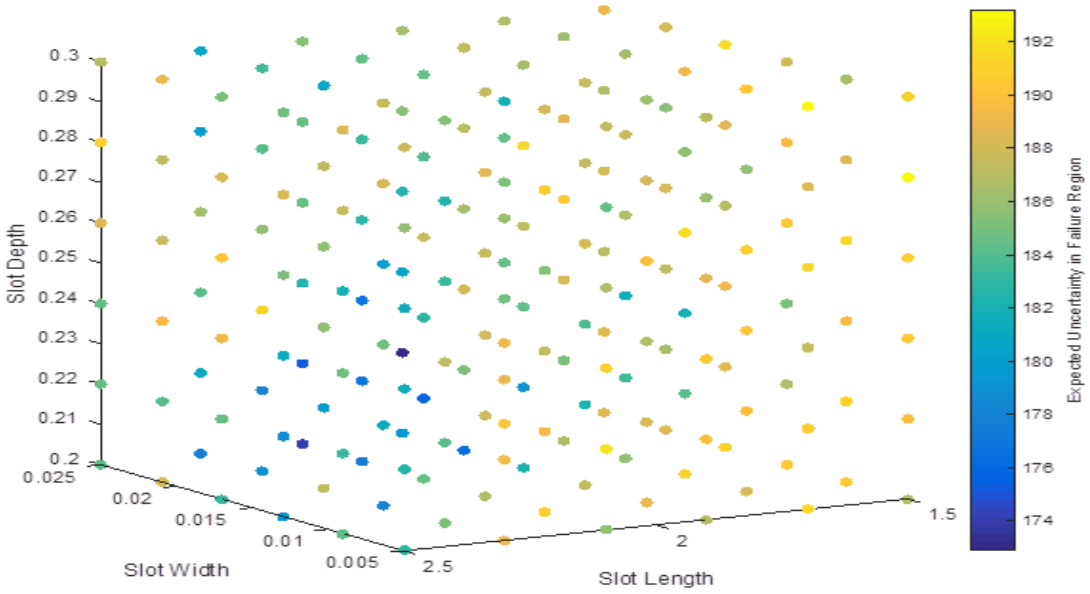

Figure 3-7. Expected uncertainty of failure, where failure is defined as values of SE greater than 15.5. Lower values of expected uncertainty provide more preferable design points.

The next step in the algorithm would be to select the more promising points and run the code using these inputs. One should see a substantial reduction in the uncertainty of the corresponding QoI.

3.1.4. Diminishing returns. Analyses of precision in QoI as a function of sample size

In this section, we address the first phase in computer experimental design as discussed in Section 3.1.1. By examining results using different sample sizes, one can determine at what point a reasonable estimate of the QoI has been obtained. Generally, there is a point where further use of this sampling approach yields diminishing returns in establishing the precision of the QoI estimates. We performed this analysis for several different QoI. We describe the procedure for the QoI failure-probability but we include results for the other QoI as well.

Consider the analysis for failure probabilities, assuming a "failure" is any situation where SE is greater than 15.5 and assuming the slot parameter inputs have a normal distribution. In Figure 3-8, we plot the estimated precision for failure probability for a range of sample sizes.

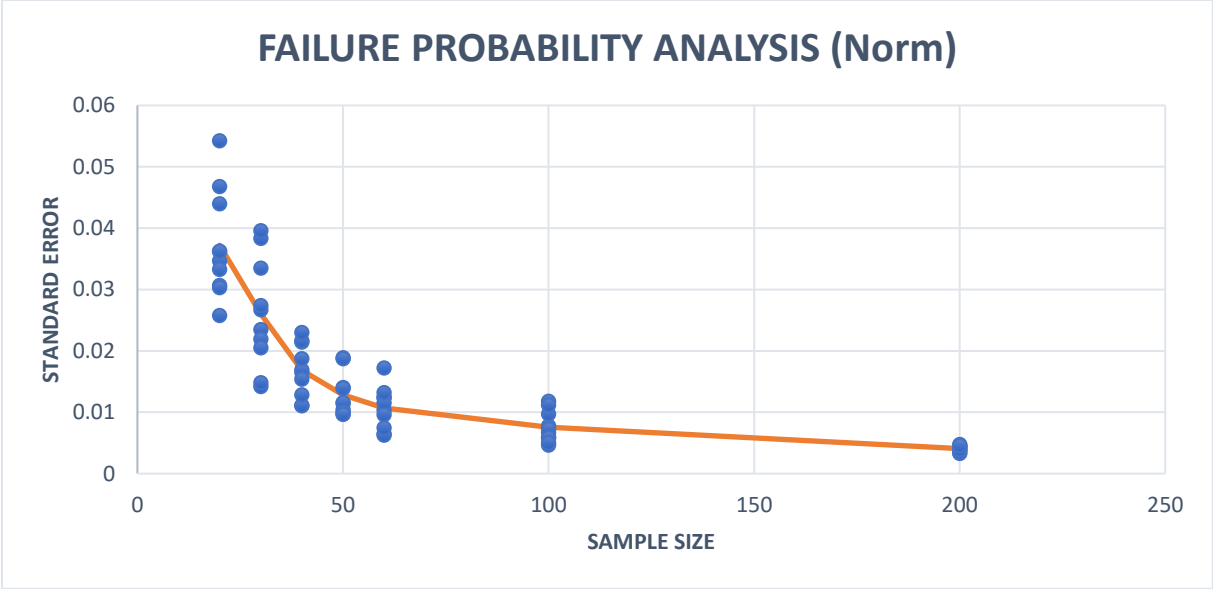

Figure 3-8. Standard error for failure probability as a function of sample size. Each dot represents an estimate of uncertainty at that sample size. The orange curve connects the mean values.

Each dot represents the response-modeling estimate of uncertainty using a response-model conditioned on a specific sample. The uncertainty estimate is just the standard deviation of the QoI estimate across the realizations of the response-model. For each sample size N, 10 different sets of N sample points were taken. Each dot represents the value of standard error for an N-point sample set. The orange curve connects the mean standard error values corresponding to each sample size. If one were to assess the required sample size based on this information, one could consider that most of the uncertainty has been resolved with a sample size of around 60 observations. Beyond a sample size value of 60, the decrements in standard error are more marginal.

An alternative approach to estimating the required sample size that might be appropriate using this power-balance model would be to use repeated sampling at each of several sample sizes, fit a Gaussian process model to each sample and use it to calculate the QoI response. A reasonable sample size could be determined as the sample size where the standard deviation of these predicted QoI responses became acceptable. This approach, however, would not be possible with a higher fidelity code which may take hours to run rather than seconds.

Figure 3-9 through Figure 3-12 present the results of standard error analyses for other QoIs similar to the analysis shown in Figure 3-8 for failure-probability. These other QoIs are the maximum SE value, mean SE value, 0.95 SE quantile and analysis of the failure region. One thing we expected in this analysis was to find that the recommended sample sizes would depend on the QoI. This appears to be the case. For the mean and 0.95 quantile analyses, a sample size of about 60 resolved most of the uncertainty in terms of standard error and the benefits of increasing the sample size were negligible. For the case of the maximum value and identifying the failure region, the required sample size to resolve most of the uncertainty would be higher, on the order of 100.

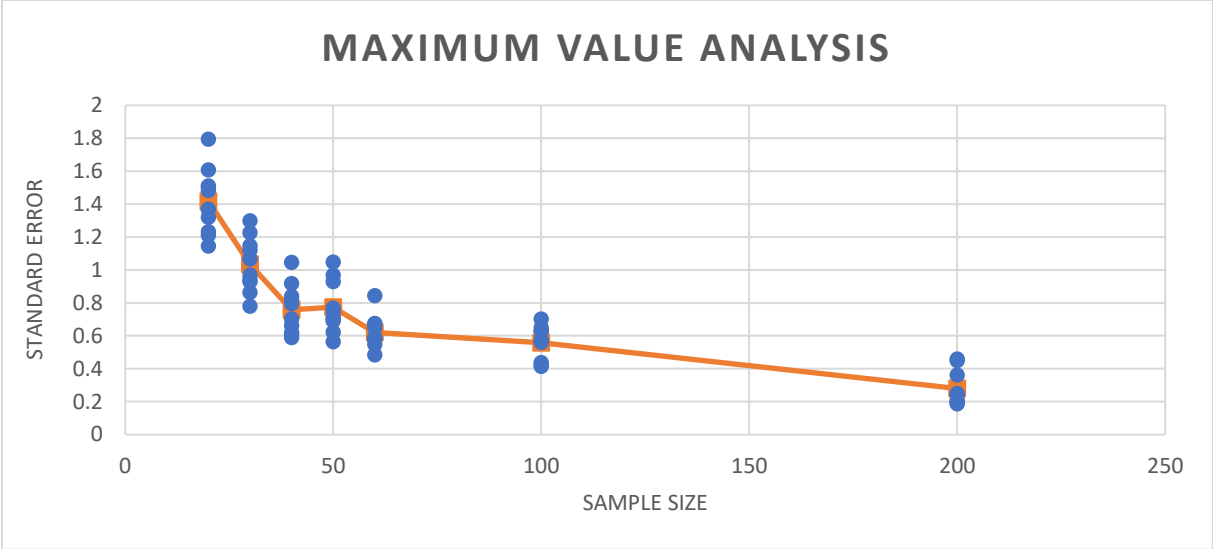

Figure 3-9. Standard error for estimated maximum value of SE as a function of sample size. Each dot represents an estimate of uncertainty at that sample size. The orange curve connects the mean values.

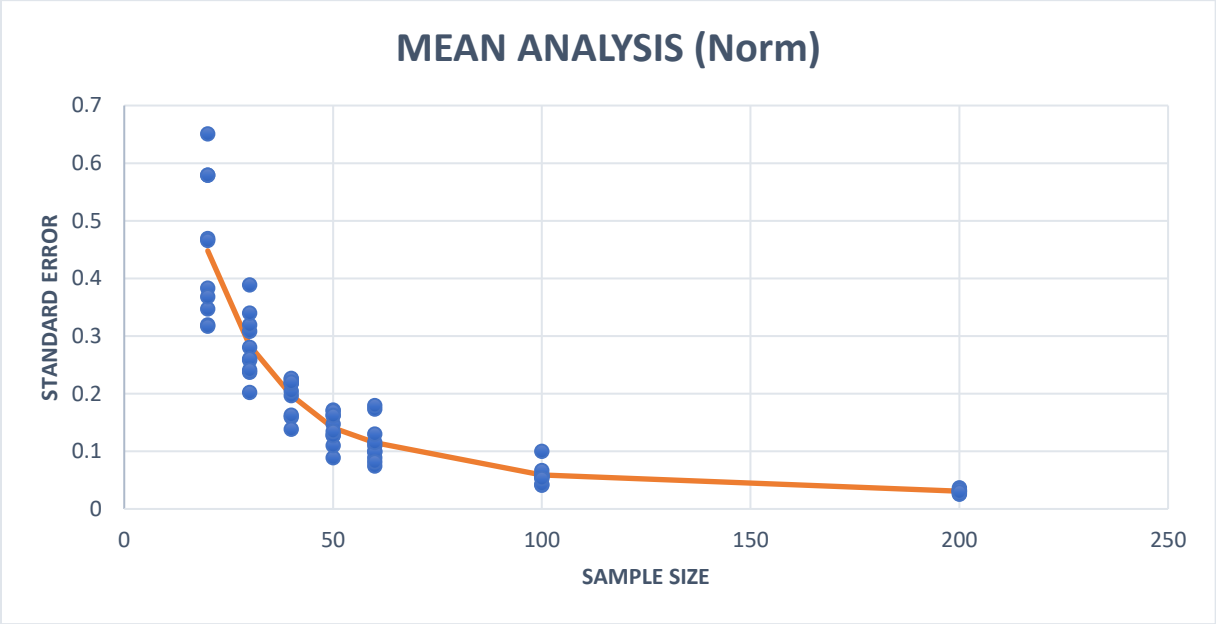

Figure 3-10. Standard error for the estimated mean value of SE as a function of sample size. Each dot represents an estimate of uncertainty at that sample size. The orange curve connects the mean values.

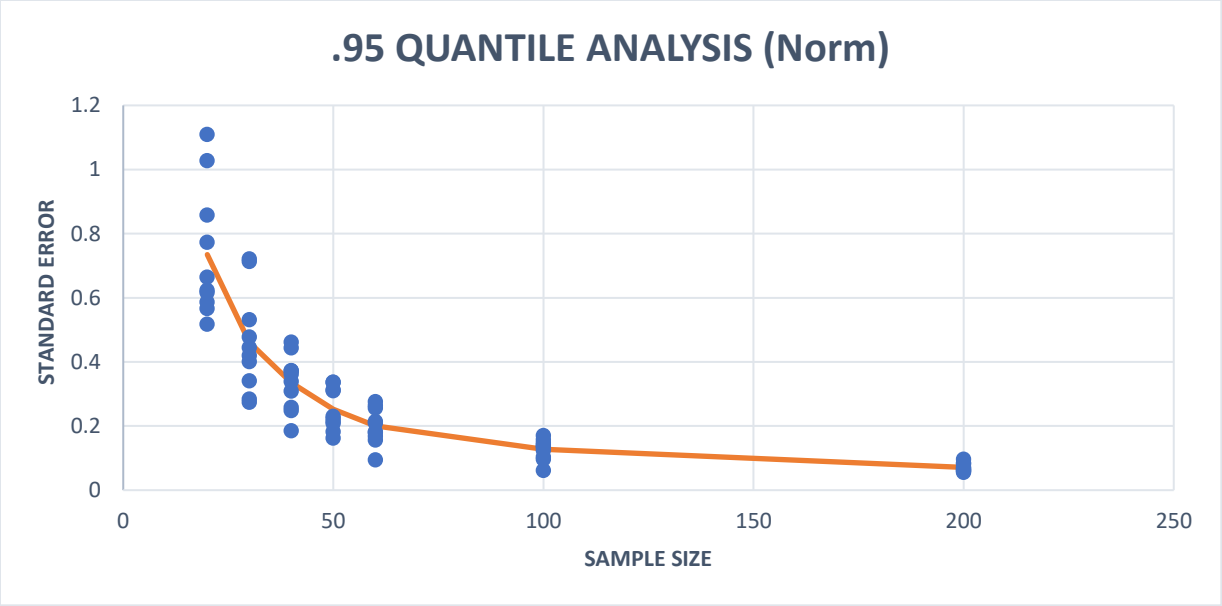

Figure 3-11. Standard error for the 0.95 quantile of SE as a function of sample size. Each dot represents an estimate of uncertainty at that sample size. The orange curve connects the mean values.

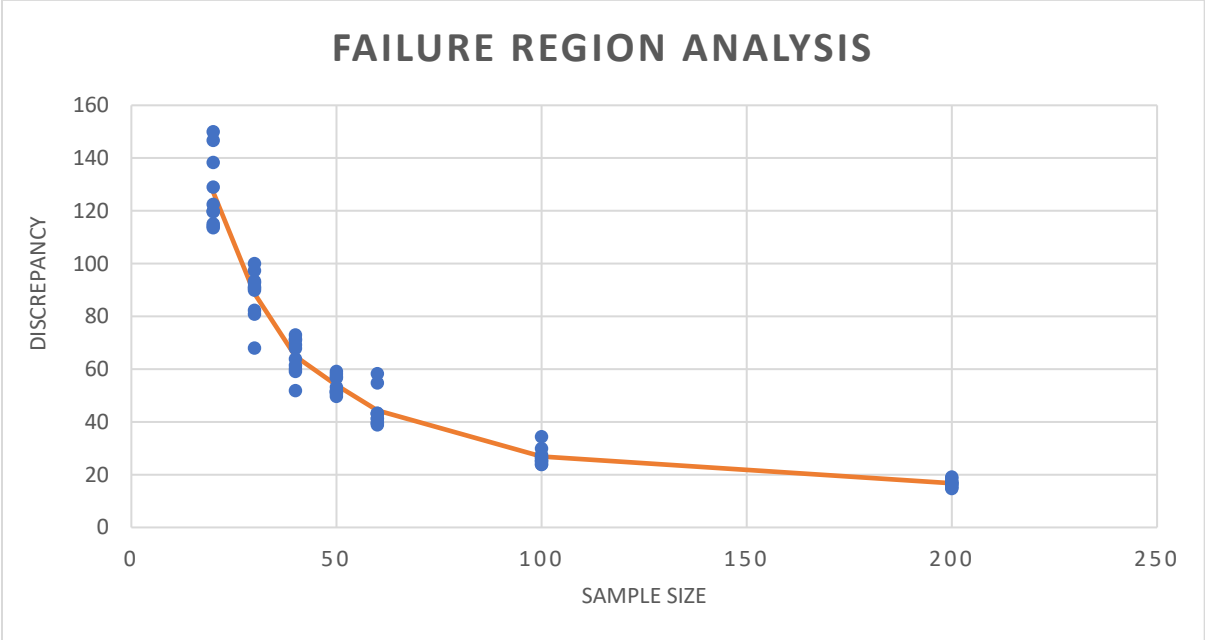

Figure 3-12. Discrepancy measure for the identification of the failure region as a function of sample size. Each dot represents an estimate of uncertainty at that sample size. The orange curve connects the mean values. Note that the identification of the failure region does not yield a point estimate and consequently standard errors are not available. See [4] for a description of the discrepancy measure.

4. VALIDATION FRAMEWORK DEVELOPMENT

An approximate uncertainty quantification method for model validation is described, the results of which can used to estimate bounds in the CompSim predictions. The uncertainty bounds accompany nominal CompSim predictions to provide confidence in the CompSim results for use in applications. The approach leverages information from existing experimental data by bounding discrepancies between corresponding CompSim and experimental data quantities of interest (QoIs) using an inverse analysis. These enveloping bounds take the form of QoI probabilistic estimates for standard deviation and yield best-fit estimates for system input parameter variances.

Linear system input-parameter-to-QoI response surface approximations are used to provide efficient functional maps representing a parametric design space. All of the design-space approximations are anchored at the QoI values computed at the nominal input parameter values and are assumed to represent the QoI mean values. These approximations are used in the estimation of the system input parameter variances, thus circumventing the need to make assumptions on input parameter probability distributions.

Finally, the estimated input parameter variances can be propagated into a large number of prediction scenarios, using similarly constructed linear response surface approximations, to estimate prediction QoI variances.

4.1. Motivation and Introduction

Physics-based modeling and CompSim have evolved incredibly in recent years. Despite this, due to any number of factors, discrepancy is routinely observed when comparing CompSim results modeling an experiment to data from the physical experiment itself. This motivates the need for representing CompSim results in the form of presumptive bounds rather than a single fixed value.

The process of establishing uncertainty aspects is a sub-type of model validation in that CompSim/experiment discrepancies are used to assess not only the suitability of our CompSim models, but also to inform downstream uncertainty models and produce predictions based on them. In typical CompSim applications there are some significant constraints that impact decisions on how best to estimate at these bounds. Physics model form error, say as a closed functional form, is always unknown and cannot be separated from other sources of discrepancy without a large time and cost budget, making this an important constraint. In addition, in many cases there is an extremely limited amount of data to assess, a priori, uncertainty in system input parameters. Finally, large physics-based CompSim models in the face of limited computational resources, namely machine time and number of processors available, mean that the number of CompSim function evaluations that could be performed was constrained.

With these limitations in mind, any QoI discrepancies are assumed to be accounted for by propagated system input parameter uncertainty. Also, for this uncertainty a low-order probabilistic metric—variance or standard deviation—is used to express QoI bounds.

4.2. Technical Approach

The foundation of the approach is the assumption that model uncertainty can be addressed through the uncertain system input parameters, $\mathbf{P}$, as this input uncertainty is reflected in the system QoI, $q = q(\mathbf{P})$. Typical examples of such parameters include, say, material properties or properties of a physical entity, such as a gap length or width. These parameters are modeled as probabilistic random variables.

Nominal values for the input parameters, which are assumed to have been acquired during prior simulation model development and **model calibration** activities, are assumed to represent their mean values, $\mu_{\mathbf{P}}$. A first estimate for the variance, $\sigma_{\mathbf{P}}^2$, of such an input random variable, or its square root, the standard deviation, $\sigma_{\mathbf{P}}$, is also made based on whatever additional information arises during the model development phase of a project.

Naturally, the effect of uncertainty in input parameters propagates into a response QoI, q. Mean values, μ_q, for the various system response QoIs that are considered in the project, such as EM wave characteristic, are assumed to equal those QoIs computed at these nominal parameter values, that is, $\mu_q = q(\mu_{\mathbf{P}})$. This relationship is exact for linear systems, but is an approximation otherwise.

QoI variances, σ_q^2, are adjusted to ensure that all discrepancies are a distance of no more than one QoI standard deviation from the nominal QoI. The constraints that these adjustments impose are used as criteria for making modifications to the input parameter variances. This process is a form of **model validation** in the sense that not only is it possible to make assessments on the suitability of the CompSim models, but underlying uncertainty is also captured in a manner that readily lends itself to prediction.

To reduce the computational burden in the propagation procedure, input to output function mappings for the input parameters are developed using linear response surface approximations. A byproduct of the process of generating these linear approximations is that a local **sensitivity analysis** is implicitly performed.

In the final phase, **prediction**, the now-modified uncertain input parameter variance estimates calculated during the model validation process are used in a number of prediction scenarios. In these, identical linear approximation procedures are employed using new CompSim QoI data local to each desired prediction point in the application domain. A visual summary of the entire technical approach is depicted in Figure 4-1.

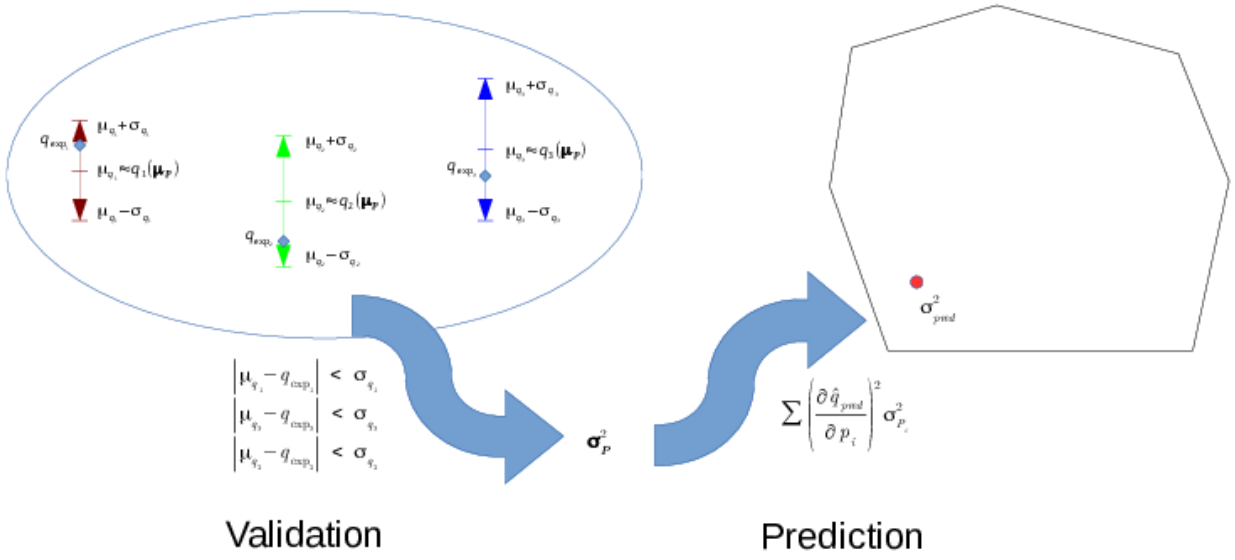

Figure 4-1. Overview of the Margin Estimation Process

4.2.1. Linear Response Surface Approximation

The circumstances underlying a typical CompSim project dictate the use of a linear response surface approximation, $\hat{q}$, of the general input parameter, $\mathbf{P}$, to QoI, q, design space functional maps, $q \approx \hat{q} = L(\mathbf{P})$. Thus,

$$\hat{q}(\mathbf{P}) = q(\mu_{\mathbf{P}}) + \sum_{i=1}^{n} \left.\frac{\partial q}{\partial p_i}\right|_{\mu_{\mathbf{P}}} (P_i - \mu_{P_i}) \tag{4-1}$$

where P_i and μ_{P_i} are understood to represent the i^{th} components of $\mathbf{P}$ and $\mu_{\mathbf{P}}$, respectively. Also, in Eq (4-1) the approximation

$$\partial q / \partial p_i|_{\mu_{\mathbf{P}}} \approx [\, q(\mu_{\mathbf{P}} + \Delta p_i) - q(\mu_{\mathbf{P}}) \,]/\Delta p_i \tag{4-2}$$

is used if one is calculating $L(\,\cdot\,)$ using a one-sided finite difference. Alternative formulations based on central difference and, in rare circumstances, linear fits via least squares can also be employed; in these cases the calculated partial derivatives in Eq (4-1) are performed in the natural way.

Building these approximations at each point in application space encompasses all of the CompSim burden. Note that the analysis that aided in the determination of which linear approximation method to use itself constitutes a form of local sensitivity analysis. As final notes on the linear approximations consider that

- The uncertain input parameters in Eq (4-1), $P_i \in \mathbf{P}$, are assumed to be statistically uncorrelated.
- While q has been assumed to be scalar in Eq (4-1), it would be trivial to generalize it to the vector case.

4.2.2. Response Variance Approximation

The response QoI variance approximation follows from Eq (4-1),

$$\begin{aligned} \sigma_q^2 &= E\left[(q(\mathbf{P}) - \mu_q)^2\right] \\ &\approx E[(\hat{q}(\mathbf{P}) - \hat{q}(\mu_{\mathbf{P}}))^2] \qquad (4\text{-}3) \\ &\approx \sum_i \left(\frac{\partial \hat{q}}{\partial p_i}\right)^2 \Bigg|_{\mu_{\mathbf{P}}} \sigma_{P_i}^2 \qquad (4\text{-}4) \end{aligned}$$

where $\sigma_{P_i}^2$ is the variance of the i^{th} input parameter in the vector, $\mathbf{P}$. Also, in Eq (4-3) an approximation to the mean of q, $\mu_q = q(\mu_{\mathbf{P}})$, has been introduced; and, finally, all comments and assumptions regarding Eq (4-1) apply to Eq (4-4).

4.2.3. Model Validation

For model validation, integrating the discrepancy between CompSim QoIs and corresponding experimental QoI data, Eq (4-4), along with the constraint

$$|\mu_{q_i} \pm \delta_{Q_i} - q_{exp_i} \pm \varepsilon_{Q_{exp_i}}| \leq \sigma_{q_i} = \sigma_{q_i}(\sigma_{\mathbf{P}}) \tag{4-5}$$

yields a set of inequalities used to adjust the individual parameter variances, $\sigma_{P_i}^2 \in \sigma_{\mathbf{P}}^2$. In Eq (4-5), δ_{Q_i} and $\varepsilon_{Q_{exp_i}}$ represent suitably chosen bounds estimates representing numerical and experimental errors, respectively.

Note that Eq (4-5) represents one inequality for each QoI in a given validation scenario and that the collective must be satisfied for all i simultaneously. This collective is depicted for the case where no numerical or experimental errors are present in Figure 4-2.

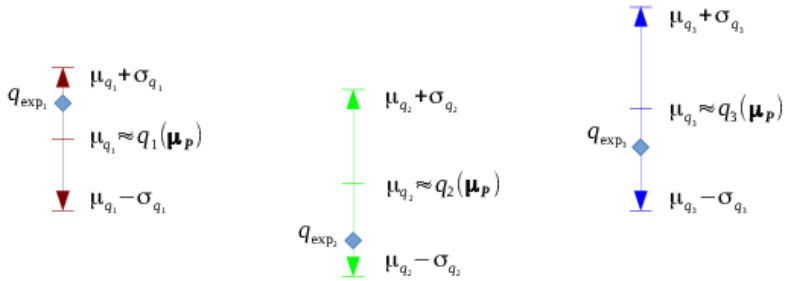

Figure 4-2. CompSim/Experiment Comparisons with Margins

where one can observe that not only do CompSim mean values vary point-to-point for each scenario/QoI, but the variances do as well; thus uniform QoI adjustments are often non-optimal. The collective inequality set to be satisfied is depicted in the Validation portion of Figure 4-1.

This validation process was carried out for a simple example of three QoIs for a single mode, TM010, of the shielding effectiveness response vs frequency for the Higgins Cylinder [7]. The Sandia EM code Gemma was used to provide the simulation QoI response data using a method-of-moments approach. Results revealed an issue with one of the QoIs, the 50% Q, a measure of the width of the SE signal about frequency at which the peak response is attained. This has yet to be resolved. The other two QoIs, the value of the peak response itself and the value of the frequency at which this peak is attained, were retained in the proposed bounding procedure. This reduced-order-validation QoI response set required only a small adjustment to the assumed input parameter standard deviations, specifically, an increase of 15%.

4.2.4. Prediction

Margin predictions can next be performed using the fixed, validation-derived $\sigma_{P_i}^2 \in \sigma_{\mathbf{P}}^2$ values and Eq (4-4) which require a new set of CompSim QoI predictions. The margin estimates can be represented by the resulting QoI standard deviations. This aspect of the procedure is depicted for one prediction point in the application space on the right side of Figure 4-1.

4.3. Concluding Remarks

In this section, we provided an overview of the technical process that is followed to estimate confidence bounds associated with CompSim predictions. We also provided motivation for specific key technical aspects of our approach and provided an individual discussion on each of these key aspects. These discussions included a justification for using a low-order uncertainty quantification approach, and the strategies for encapsulating all observed simulation-to-experiment discrepancies through system input parameter uncertainty and for limiting the computational burden through the use of low-order physics-based model approximations. A simple validation exercise using the Higgins Cylinder was described that resulted in the identification of a potential problem with the experimentally measured Q for the first, or lowest frequency, resonant mode. This mode is identified as TM010 for the EM shielding effectiveness response. A minor adjustment to the assumed standard deviations for the uncertain input parameters was required to ensure that the experiment QoI data were bounded by the nominal analysis result plus or minus a standard deviation for the remaining QoIs considered.

Finally, a visual depiction of a low-order uncertainty quantification process was provided in its entirety from validation to prediction for a typical prediction point. A procedure could be devised in which the results of the proposed validation study are used to propagate quite general probabilistic information, such as estimated probability of loss of assured safety (PLOAS), with results that are expressed as intervals with statistical tolerance limits. Such an approach also mitigates the problem of assuming too much information in assumed probability models characterized with too little data.

5. CONCLUSION

The V&V/UQ/Credibility process activities described in this report represent the beginning of a broader set of credibility activities supporting the code development efforts for the Gemma electromagnetics code. These efforts will be documented in a V&V/UQ/Credibility plan for Gemma in FY21. The overall goal of these activities is to provide Gemma developers and analysts with the tools needed to generate credibility evidence in support of Gemma predictions for future use cases. As the code development and application plans for Gemma continue to evolve, the planning and execution of credibility activities for Gemma will need to be adapted to meet changing needs and expectations.

Several lessons learned can be captured from the work described in this report and will be applied to the planning and execution of credibility activities for Gemma moving forward. First, it is critically important for V&V/UQ/Credibility activity planning and execution to include EM SMEs and Gemma code developers. This involvement is important to ensure that V&V/UQ/Credibility experts are made aware of important considerations in the physics, understand the experimental data that they are processing, and are able to connect credibility activities to larger planning efforts and requirements. A lack of communication and understanding on these aspects of the analysis can lead to an increase in analysis timelines and rework. As a result of this lesson learned, EM SMEs and Gemma code developers will be included as authors for the FY21 Gemma credibility plan, and strategies for communication will be put in place to communicate and document important aspects of upcoming analyses. As a result of the FY20 efforts, a stronger teaming across V&V/UQ/Credibility partners and EM SMEs/Gemma code developers has already been developed.

Second, it is clear that there is a wide variety of V&V/UQ/Credibility needs and stakeholders for both Gemma and EM analyses in general. Competing priorities and analysis needs can make it difficult to develop a clear plan for execution of credibility activities. Future planning will involve a prioritization of activities and stakeholders to streamline execution and ensure that credibility activities leverage V&V/UQ/Credibility capabilities and expertise efficiently.

Through evolving planning and execution of V&V/UQ/Credibility activities for Gemma, credibility evidence and the capabilities to generate this evidence will be developed and made available to Gemma code developers and analysts. This will support the development and delivery of the credibility evidence needed to support the use of Gemma across an applicable range of computational simulation problems.

REFERENCES

[1] Red-Horse, J., Dowding, K., Mullins, J., Schroeder, B., Black, A., & Eckert, A. V&V/UQ Credibility Assessment Communication Template Set. Sandia Presentation – SAND2020-4839O.

[2] Krueger, A., Hamel, J., Matula, N., & Freno, B. (2020) An Overview of Gemma FY20 Verification Activities. Sandia Report – SAND2020-11229.

[3] McKay, M.D., Conover, W.J. and Beckman, R. J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code", Technometrics, Vol. 21, No. 2, 239-245.

[4] Rutherford, B. (2020). "A Response-Modeling Approach to Optimal Design of Computer Experiments", White Paper, Sandia National Laboratories.

[5] Campione, S., Stephens, J.A. Martin, N., Eckert, A., Warne, L.K., Huerta, G., Pfeiffer, R.A. and Jones, A. (2019). "Developing uncertainty quantification strategies in electromagnetic problems involving highly resonant cavities", submitted for publication, Sandia National Laboratories. *Submitted for Publication.*

[6] Harrington, R. (1961). Time-Harmonic Electromagnetic Fields. McGraw-Hill.

[7] Higgins, M. & Charley, D. (2007). Electromagnetic Radiation (EMR)Coupling to Complex Systems: Aperture Coupling into Canonical Cavities in Reverberant and Anechoic Environments and Model Validation. Sandia Report – SAND2007-7931.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Salvatore Campione	1352	sncampi@sandia.gov
Adam Jones	1352	adajone@sandia.gov
Joe Kotulski	1352	jdkotul@sandia.gov
William Langston	1352	wllangs@sandia.gov
Robert Pfeiffer	1352	rapfeif@sandia.gov
Brian Zinser	1352	bzinser@sandia.gov
Adam Stephens	1463	jasteph@sandia.gov
Neil Matula	1541	nmatula@sandia.gov
Aubrey Eckert	1544	acecker@sandia.gov
Brian Freno	1544	bafreno@sandia.gov
Aaron Krueger	1544	amkrueg@sandia.gov
Gabriel Huerta	6673	jghuert@sandia.gov
Casey Jelsema	6674	cmjelse@sandia.gov
Brian Rutherford	6674	bmruthe@sandia.gov
Technical Library	01977	sanddocs@sandia.gov

This page left blank

This page left blank

Sandia National Laboratories

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.