



OPEN ACCESS

EDITED BY

Nicola Pedroni,
Polytechnic University of Turin, Italy

REVIEWED BY

Alberto Ghione,
Commissariat à l'Energie Atomique et aux
Energies Alternatives (CEA), France
Eric Aboud,
Lawrence Livermore National Laboratory
(DOE), United States

*CORRESPONDENCE

Ugur Merturek,
✉ mertureku@ornl.gov

RECEIVED 29 July 2025

ACCEPTED 02 October 2025

PUBLISHED 23 October 2025

CORRECTED 05 December 2025

CITATION

Sundaram A, Yin S, Merturek U and
Abdel-Khalik H (2025) An entropy-based
debiasing approach to quantifying experimental
coverage for novel applications of interest in the
nuclear community.
Front. Nucl. Eng. 4:1675308.
doi: 10.3389/fnuen.2025.1675308

COPYRIGHT

© 2025 This work is authored in part by Arvind
Sundaram, Shiming Yin, Ugur Merturek and
Hany Abdel-Khalik, © UT-Battelle LLC. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

An entropy-based debiasing approach to quantifying experimental coverage for novel applications of interest in the nuclear community

Arvind Sundaram¹, Shiming Yin¹, Ugur Merturek^{2*} and
Hany Abdel-Khalik¹

¹School of Nuclear Engineering, Purdue University, West Lafayette, IN, United States, ²Oak Ridge National Laboratory, Nuclear Energy and Fuel Cycle Division, Oak Ridge, TN, United States

This manuscript proposes a novel information-theoretic approach to the quantification of experimental relevance, i.e., coverage, to achieve optimal data assimilation results for nuclear engineering applications. Specifically, this work posits the need for a new metric, called coverage (q_C) of an application's quantity of interest, i.e., eigenvalue or power peaking for an advanced reactor concept, defined herein as the theoretically maximum achievable reduction in the quantity's uncertainty given measurements from a pool of experiments in a manner that is independent of the data assimilation procedure employed. Currently, reduction in a quantity's uncertainty is strongly biased by the underlying assumptions of the assimilation procedure to account for the under-determined nature of such problems and the similarity criterion employed to identify relevant experiments. To address this challenge, this work has developed a coverage metric, q_C , based on mutual information, which establishes a new conceptual framework for assessing coverage, one that is independent of the model parameters and responses degree of variations in both the experimental and application domains, i.e., linear vs non-linear, and their prior uncertainty distributions, i.e., Gaussian vs. non-Gaussian. The q_C is an entropic measure capable of addressing coverage for general nonlinear problems with non-Gaussian uncertainties and inclusive of the measurement uncertainties from multiple experiments. Numerical experiments from manufactured analytical problems as well as a set of benchmarks from the ICSBEP handbook are employed to demonstrate its theoretical and practical performance as compared to the c_k -based experiment selection methodology, commonly employed in the neutronic community. The manuscript then employs other well-known adaptations to existing data assimilation methodologies for nonlinear and non-Gaussian problems capable of achieving the coverage posited by q_C .

KEYWORDS

experimental coverage, similarity analysis, nuclear criticality analysis, uncertainty quantification, criticality safety, Bayesian data assimilation

1 Introduction

Data assimilation challenges in the nuclear community often require the prediction of novel application responses given physics-based models and a limited body of experimental data, such as eigenvalue predictions in advanced reactor designs, different sets of reactor conditions, etc. Various data assimilation procedures such as generalized linear least squares (GLLS) in TSURFER (Williams et al., 2011) commonly involve the adjustment of input parameters such as cross-sections and their associated covariance libraries using the available experimental data. This adjustment process is inherently underdetermined, with thousands of parameters informed by tens of experiments. The adjusted uncertainties are then propagated through the forward model of the application to obtain an estimate of the bias and uncertainty in the desired application response. Other procedures, such as MOCABA (Hofer et al., 2015), apply a direct mapping between the experiments and application through repeated Monte Carlo (MC) sampling and directly compute the bias and uncertainty without the need for adjustments. In general, these adjustment procedures are grounded in the Bayesian approach, which is ubiquitous in the nuclear community for its simplicity, scientific rigor, and amenability to verification, validation, and uncertainty quantification.

However, the Bayesian approach is based on the principle of incorporating all data and operates under the assumption that the correct adjustments will be obtained with infinite experimentation as the effects of less relevant experiments “cancel out.” This assumption is no longer valid when experimental data are scarce, as is common in the nuclear community, where experiments are cost-prohibitive and limited. In this case, the approach is susceptible to the error compensation phenomenon where the input parameters may be overcorrected based on a handful of experiments from a limited set assumed to be highly relevant, leading to overconfident and inaccurate predictions. Further, useful experimental data may be erroneously discarded as having low relevance, leading to underconfident predictions. In other words, a practical use of the Bayesian approach requires an assessment of the relevance of experimental data *a priori*: that is, an impact assessment on the application uncertainty—defined as *coverage* in this manuscript. The key question addressed here is, *How do we assess experimental relevance and quantify its coverage a priori?*

Previous work to assess coverage in the nuclear community include filtering out experiments based on their c_k -similarity (Broadhead et al., 2004), via the use of thresholds. For example, one may look for experiments with $c_k > 0.95$, or if such experiments with high similarity are not available, then a set of 5–10 experiments with $c_k > 0.9$ or 15–20 experiments with $c_k > 0.8$ (Broadhead et al., 2004). Beyond the nuclear community, practitioners may use metrics such as R^2 and its variants to quantify coverage after the adjustment procedure. However, this approach may confound the practitioner when performing diagnostics on their model. For instance, if a practitioner obtains a poor metric for the coverage, is it because of assumptions made by their data assimilation, i.e., adjustment procedure, such as linearity and Gaussianity, or is it the best-possible estimate due to incomplete information inherent in the experimental data? The consequences are markedly different; whereas the former informs the practitioner of inadequacies in the procedure, the latter informs them of inadequate relevant

experimental data and provides insight into what experiments may improve coverage.

This manuscript posits that an important distinction is to be made by the practitioner between experimental coverage and the adjustment procedure used by data assimilation. This is due to the various biases introduced by the adjustment procedure. For example, the adjustment of cross-sections given limited experimental data, e.g., critical eigenvalue, is highly ill-posed, with a large number of degrees of freedom that require coverage by the experimental data. Adopting a minimum-norm or other variant regularization criterion in such cases may appear logical for obtaining unique solution but is not resilient to the various sources of uncertainties and error compensation phenomenon that is unavoidable in practical problems, especially when the number of measurements is much smaller than the number of uncertainty sources. Examples of such uncertainties include epistemic uncertainties from lack of knowledge of nuclear modeling parameters and their prior uncertainties, aleatory uncertainties from inherent randomness, and physics modeling uncertainties from simplifying assumptions and numerical approximations. The chosen data assimilation procedure may erroneously adjust the underlying parameters in a manner that overcompensates some uncertainties for other uncertainties that are unaccounted for, as described earlier in the discussion on error compensation.

Mitigating these biases requires a full understanding of the underlying system and its various sources of uncertainties, that is, an accurate system-level high-fidelity physics model across all operating conditions, which is either expensive or infeasible. The use of machine-learning algorithms such as neural networks to approximate the physics model further brings additional assumptions and unknown biases, especially with limited data. The resulting biases are difficult to hedge for and their impact on inference in novel applications is unknown, reducing confidence in practitioners.

The proposed debiasing paradigm (Mertyurek and Abdel-Khalik, 2025) calls for a decoupling of the adjustment procedure from the question of experimental coverage through an entropic-based approach that directly maps relationships between the application response of interest and the experimental data based on known physics principles, sufficiently described by a first-principles model. Note that a first-principles model need not necessarily be high-fidelity and may even have relatively higher uncertainty in the input parameters (and corresponding outputs). In fact, a high-fidelity model may be over-tuned to a particular set of benchmark conditions and provide overconfident results when conditions change. It suffices that the model is physics-based and provides a faithful representation of the true relationships between the experimental data and the response(s) of interest. Capturing these relationships yields a coverage metric free of any potential biases introduced by the data assimilation adjustment procedure.

In this context, *coverage* is defined independently of any specific adjustment procedure and as a statistical limit on what can be inferred about the application of interest given the experimental data, akin to the concept of Shannon’s entropy in telecommunications (Shannon, 1948), where the channel capacity is calculated independently of the employed information error-correction encoding algorithm. When defined as such, it is a useful reference and a diagnostic tool to a practitioner in

assessing the efficacy of their chosen data adjustment procedure and the assumptions inherent in them, just as Shannon's channel capacity can be used to compare the efficacy of different error-correcting encoding algorithms. By defining coverage, the practitioner can assess the relevance of their limited body of experiments, select experiments that provide the most coverage, and quantify their anticipated impact on improving the estimates of the application of interest—that is, the anticipated reduction in uncertainty—prior to applying any adjustment procedure. Then, the practitioner can assess the performance of their procedure and perform diagnostics such as hyperparameter tuning, validating assumptions, *etc.*, until it approaches the precomputed statistical limit. Furthermore, they can optimize their experimental design to maximum its value to the application of interest.

To achieve these goals, this manuscript introduces the metric q_C for coverage quantification, as defined in Equation 1, borrowing from principles of information theory pioneered by C. Shannon in the 1940s in the field of digital communication (Shannon, 1948). Specifically, the concept of mutual information (Cover and Thomas, 1991), $F(\cdot)$, is adapted to a nuclear context as an assumption-free metric of quantifying coverage of the application response of interest, \mathbf{Y}_a , given the available body of experimental data, \mathbf{Y}_e , and associated underlying parameters, reference values, parameter uncertainties, measurements, measurement uncertainties, *etc.*, which are denoted by $\boldsymbol{\mu}$. For the rest of the manuscript, the subscripts a and e denote the application response(s) of interest and the experiment(s), respectively.

$$q_C(\mathbf{Y}_e, \mathbf{Y}_a) = 1 - e^{-F(\mathbf{Y}_a; \mathbf{Y}_e, \boldsymbol{\mu})} \quad (1)$$

$$F(\mathbf{Y}_a; \mathbf{Y}_e, \boldsymbol{\mu}) = H(\mathbf{Y}_a) - H(\mathbf{Y}_a | \mathbf{Y}_e, \boldsymbol{\mu}) \quad (2)$$

$$H(X) = - \int_{x \in X} p(x) \ln p(x) = \mathbb{E}[-\ln p(x)] \quad (3)$$

Previous work in the financial analysis community has presented mutual information as a global correlation coefficient for time-series data (Athe and Abdel-Khalik, 2014), which has been adapted to the nuclear community as a generalized similarity index (Dionisio et al., 2004). The scope of this manuscript is to build upon the theoretical underpinnings of the latter work, establish its ability to generalize existing indices such as c_k to nonlinear and non-Gaussian problems, and demonstrate its value in cases where multiple low- c_k experiments (typically discarded) may provide significant coverage.

Information theory defines the mutual information between two or more variables as the reduction in entropy, $H(\cdot)$, of one variable gained upon knowledge of the other(s) as defined in Equation 2. Here, entropy is a general metric characterizing the uncertainty of a variable defined in Equation 3, in a similar manner to how the mean and standard deviation characterizes the Gaussian distribution via the central limit theorem. The mutual information can then be interpreted as the average “information gain” obtained by reducing the entropy (i.e., uncertainty) in the application response of interest upon knowledge of the experimental data and associated parameters. It is also the theoretical upper limit (Carrara and Ernst, 2017) of inference for any procedure and is useful in diagnosing overfitting and/or underfitting.

The use of entropy enables the practitioner to work across a wide range of uncertainty distributions as opposed to using existing

methods that typically require Gaussianity. Under the principle of maximum entropy when the mean and covariance are known, the underlying parameters, e.g., cross-sections, are assumed to be Gaussian. Nevertheless, nonlinearities in experiment(s) and/or application(s) result in corresponding non-Gaussian priors for the responses, which hinders the quantification of coverage using traditional tools that only account for first-order sensitivities and assume Gaussianity of the experiment(s) and application. With the use of entropy, the practitioner can accurately construct reliable confidence intervals to improve parameter estimates for general non-Gaussian distributions.

The scope of this work is to provide a theoretical basis for q_C and mutual information, its value to practitioners, and demonstration of its effectiveness to tackle non-linear and non-Gaussian data assimilation problems, as well as a real use-case to identify high-value experiments on a set of benchmarks from the *International Criticality Safety Benchmark Evaluation Project* (ICSBEP) handbook. The handbook consists of a set of criticality safety benchmark experiments in a standardized format (Nuclear Energy Agency, 1995) describing the fissile material used (Plutonium, highly enriched Uranium, *etc.*), the physical form of the material (metal, compound, *etc.*), and the neutron spectrum (fast, thermal, *etc.*). The benchmarks contain the information to compute sensitivities and build representative calculational models as well as the experimental eigenvalues and their associated uncertainties. In the numerical experiment discussed in this manuscript, the sensitivities the experimental measurements are utilized. The rest of the manuscript is divided as follows: Section 2 provides a background on inference procedure, as well as existing coverage metrics and their shortcomings. Section 3 introduces the q_C metric, the intuition behind Equation 1, additional information on q_C computation, comparisons to GLLS-based metrics such as c_k , and applicability as a experimental selection tool. Section 4 outlines a numerical experiment using a dataset composed of linear and nonlinear models that highlight the necessity of a procedure-free coverage metric and the ability of q_C in quantifying coverage *a priori*. Then, the value of q_C to identify high-value but low c_k experiments is demonstrated on a set of 100 benchmarks from the ICSBEP handbook. Section 5 summarizes the results of the numerical experiment and discusses future avenues of research.

2 Background

The similarity coefficient, c_k , between an application and a single experiment has gained significant traction in the nuclear community in response to coverage quantification challenges. It is defined by the correlation coefficient in Equation 4 between the application and experiment sensitivities for a given experiment i , denoted by \mathbf{s}_a and \mathbf{s}_e^i , respectively, and is weighted by the prior covariance matrix $\boldsymbol{\Sigma}_z$ of the dependent variable \mathbf{z} (typically cross-sections).

$$c_k^i = \frac{\mathbf{s}_a^T \boldsymbol{\Sigma}_z \mathbf{s}_e^i}{\sqrt{\mathbf{s}_a^T \boldsymbol{\Sigma}_z \mathbf{s}_a} \sqrt{\mathbf{s}_e^i^T \boldsymbol{\Sigma}_z \mathbf{s}_e^i}} \quad (4)$$

Assuming strictly linear dependence of both the experiment and the application responses on the cross sections (both within and outside the operational range to account for changes in conditions),

Gaussian prior covariance of the cross-section, and zero measurement uncertainty, c_k can be interpreted as coverage of the application response of interest for a single experiment. Consequently, the approach suffers from a few key deficiencies, namely, the following.

Measurement Uncertainty: c_k ignores measurement uncertainty inherent to experimentation and therefore cannot be directly translated to coverage. For instance, experiments with high c_k and high measurement uncertainty may be less relevant than those with a relatively lower c_k but lower measurement uncertainty. In other words, it is not possible for a practitioner to answer the following question *a priori*: What coverage is achievable if I utilize experimental data with $c_k = 0.9$?

One-to-One: As a correlation coefficient, c_k can only assess an individual experiment and cannot account for synergistic relationships among multiple experiments in coverage. As demonstrated later in this manuscript, it is possible for multiple low c_k experiments below a user-defined threshold to provide greater coverage than a high c_k experiment, controlling for measurement uncertainty. Methodologies in (Broadhead et al., 2004; Kiedrowski and Ki, 2014) that combine c_k across multiple experiments are tied in with the GLLS procedure to obtain stable biases and do not enjoy the interpretation that the one-to-one variant enjoys by virtue of being a correlation coefficient.

Experimental Redundancy: c_k cannot diagnose experimental redundancy found in experimental benchmarks. Here, we imply redundant coverage, which may not necessarily be the same experiment performed multiple times. For example, two experiments may synergistically provide coverage for an application that makes a third experiment redundant; however, when viewed individually, the three experimental sensitivities may be dissimilar. This may lead to the selection of an experiment with high c_k (with the application) that is also heavily correlated with a pre-selected set of experiments, leading to almost no meaningful reduction in uncertainty (i.e., no additional coverage). This may even occur at the cost of a low c_k experiment that is independent and potentially offers greater coverage of the application. This phenomenon is better explained through the lens of subspace analysis presented in Sections 3.2, 5.2.

Nonlinearity and Non-Gaussianity: By virtue of being a correlation coefficient, c_k measures linear relationships and cannot generally assess relevance for nonlinear problems in the nuclear industry, such as eigenvalue dependence on H/X moderator-to-fuel ratio. The use of a threshold may disregard relevant experimental data that have a nonlinear but symmetric relationship, due to low c_k , e.g., quadratic around the reference. Further, the application and/or experimental sensitivities at the reference may not be characteristic of the function's behavior across the entire region of interest, resulting in low c_k at the reference, but potentially high c_k elsewhere.

Modifications to c_k have been previously suggested in the nuclear community to mitigate issues related to multiple experiments, experimental redundancy, and measurement uncertainty while retaining its interpretation as a metric (Broadhead et al., 2004; Kiedrowski, 2014) to assess experimental coverage. However, as a correlation coefficient, it relies on linear sensitivities and Gaussian uncertainties of the underlying parameters. As demonstrated in the numerical experiments later

in this manuscript, the approach may fail to accurately identify experimental relevance and quantify coverage when the experiments and/or application responses of interest exhibit nonlinearity.

The value of q_C in identifying potential useful experiments that have low one-to-one c_k values with applications is more apparent in the context of trending analysis (Broadhead et al., 2004) and WHISPER (Kiedrowski, 2014), where c_k is used as a threshold criterion to determine whether a system falls within the area of applicability of the experiment. In (Broadhead et al., 2004), it is observed that as experiments are added in decreasing order of c_k , the computed bias begins to significantly break away from previously computed values. The suggested approach is to increase the number of experiments selected to enable the bias estimate to stabilize and converge. On the other hand, WHISPER (Kiedrowski, 2014) utilizes the χ^2 -statistic computed following the adjustment and a threshold (default = 1.2) to accept the experiment or reject it as inconsistent. In this case, the method to accept/reject, i.e., to determine whether an experiment provides coverage, is tied to the GLLS procedure, and not independently determined.

Here, q_C may be utilized to assess the total value of a set of benchmark experiments independent of GLLS, by computing the mutual information of all the experiments with the application. Then, an iterative approach may be designed where the experiment with the highest q_C is initially selected, followed by the experiment that maximizes the q_C when added to the already-selected experiment. As mutual information is invariant to redundant information, the iterative procedure selects experiments one-by-one until the total q_C of the selected set is close to the value of the entire set, as defined by the tolerance criterion of the user. As demonstrated in the numerical sections on the manufactured analytical problems and the benchmarks from the ICSBEP handbook, this approach provides stable convergent behavior for GLLS, does not arbitrarily define a minimum number of experiments or c_k threshold like (Broadhead et al., 2004; Kiedrowski, 2014), and accounts for the synergy amongst the experiments not captured by the one-to-one c_k . By not discarding experiments solely based on c_k value, this approach also enables practitioners to extract the full value of their existing body of experiments and determining whether the captured coverage is sufficient before building new experimental benchmarks.

3 Methodology

The proposed q_C metric avoids the limitations of extant metrics through an entropic approach that places no assumptions on linearity, the probability distribution, and the specific adjustment procedure. Because this metric directly computes the mutual information, both linear and nonlinear dependencies between the application responses and the experimental data are accounted for. This approach also bypasses a critical computational limitation of computing individual entropies for some distributions. As critiqued by Jaynes in his work on the limiting density on discrete points (Jaynes, 1957), the entropy of continuous variables as defined by Shannon, $H(X) = -\int_{x \in X} p(x) \ln p(x) dx$, lacks several favorable properties of its discrete counterpart, especially the lack of invariance to the change of variables, as well as divergence if the

integral is approximated by Riemann summation with an increasing number of bins. However, since mutual information is the difference of entropies, the effect of divergence and change of variables are zeroed out (Jaynes, 1957). It also remains identical if computed using known analytical expressions for differential entropy, Jaynes' adaptation, or in the limit of Riemann summation, assuming the bins are kept identical for all variables. Another key advantage is that mutual information possesses the property of invariance to change of variables, redundant experiments, and irrelevant experiments, and it is typically computationally tractable for most problems of interest to the nuclear community using published algorithms such as k-nearest neighbors (Kraskov et al., 2003), mutual information neural estimation (Belghazi et al., 2018), etc. It also enables a practitioner to assign the value of a new redundant experiment as zero if it does not increase mutual information (and, by extension, q_C).

Mutual information for continuous variables as encountered in the nuclear community is bounded on $[0, \infty)$, inspiring the nonlinear transform in Equation 1 that bounds q_C on $[0, 1]$, where 0 indicates no coverage and 1 indicates perfect coverage by the experiment. As mutual information is an upper bound on coverage, the former case implies that no adjustment procedure can be used to glean value from the experimental data without additional assumptions/biases made by the practitioner; conversely, the latter case assures the practitioner that the application response is theoretically perfectly predictable given the experimental data. Being independent of the adjustment procedure, q_C informs the practitioner on "what" data to use for the analysis by providing an upper bound on coverage but not "how" to achieve the coverage; that is, it does not inform the practitioner of the specific adjustment procedure. To this end, the machine learning community has developed parameter-free methods and universal approximators such as neural networks to achieve close to the performance predicted by q_C , which may further be combined with Bayesian approaches for verification, validation, and uncertainty quantification.

3.1 Equivalence for linear models with Gaussian uncertainties

The following sections relate q_C to well-known metrics such as C-similarity based on the GLLS methodology, which is well suited for linear and Gaussian problems. Under such conditions, Equation 1 simplifies to the reduction in uncertainty (as characterized by the standard deviation σ) of a single application response of interest, Y_a , given the body of experimental data, $\mathbf{Y}_e, \boldsymbol{\mu}$.

Assume a single application response of interest, represented by the random variable Y_a , and a set of experiments represented by the random vector, \mathbf{Y}_e , and parameters $\boldsymbol{\mu}$. If the underlying parameters are Gaussian and the applications and experiments are linear functions, then:

$$F(Y_a; \mathbf{Y}_e, \boldsymbol{\mu}) = H(Y_a) - H(Y_a | \mathbf{Y}_e, \boldsymbol{\mu})$$

Here, the entropy of a Gaussian variable $x \in X \sim \mathcal{N}(\bar{x}, \sigma_x)$ is derived using the probability distribution $p(\cdot)$ as

$$\begin{aligned} H(X) &= - \int_{x \in X} p(x) \ln p(x) \\ &= \mathbb{E}[-\ln p(x)] \\ &= \mathbb{E}\left[-\ln \frac{e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}}{\sqrt{2\pi\sigma_x^2}}\right] \\ &= \mathbb{E}\left[\ln \sqrt{2\pi\sigma_x^2} + \frac{(x-\bar{x})^2}{2\sigma_x^2}\right] \\ &= \ln \sigma_x \sqrt{2\pi} + \frac{\mathbb{E}[(x-\bar{x})^2]}{2\sigma_x^2} \\ &= \ln \sigma_x \sqrt{2\pi} + \frac{1}{2} = \ln \sigma_x \sqrt{2\pi e} \\ F(Y_a; \mathbf{Y}_e, \boldsymbol{\mu}) &= H(Y_a) - H(Y_a | \mathbf{Y}_e, \boldsymbol{\mu}) \\ &= \ln(\sigma_{a,prior} \sqrt{2\pi e}) - \ln(\sigma_{a,post} \sqrt{2\pi e}) \\ &= \ln \frac{\sigma_{a,prior}}{\sigma_{a,post}} \\ q_C^{LG} &= 1 - e^{-F(Y_a; \mathbf{Y}_e, \boldsymbol{\mu})} = 1 - \frac{\sigma_{a,post}}{\sigma_{a,prior}} \end{aligned} \quad (5)$$

As depicted in Equation 5, the simplified metric, q_C^{LG} , is readily interpretable to a practitioner in contrast to c_k . The value of a new experiment is clearly quantifiable as providing Δq_C^{LG} additional reduction in the uncertainty. For example, a $q_C^{LG} = 0.9$ implies that the posterior uncertainty is reducible to 1/10th the prior. If the addition of a new experiment increases q_C^{LG} to 0.95, then it may be understood that the experiment provides an additional 5% reduction in posterior uncertainty. On the other hand, the additional coverage characterized by an increase in c_k is both a function of measurement uncertainty and follows a more complex nonlinear relationship involving the ratio of variances. For example, assuming zero measurement uncertainty, $c_k^2 = 1 - \sigma_{a,post}^2 / \sigma_{a,prior}^2$ (i.e., R^2), which is not as readily interpretable as the simple linear relationship of q_C .

The expression can be simplified further when the linear model is known. Assume the parameters $\boldsymbol{\mu}$ consist of the measurement uncertainty $\boldsymbol{\Sigma}_m$ (typically diagonal and user-input), dependent variable \mathbf{z} with mean $\bar{\mathbf{z}}$ and prior covariance matrix $\boldsymbol{\Sigma}_z$ obtained from ENDF libraries in tools such as SCALE. The linear model describing the experiments and application response are represented using the sensitivity matrix \mathbf{S}_e and vector \mathbf{s}_a , respectively, where each column of \mathbf{S}_e corresponds to the sensitivity of a given experiment.

$$\begin{aligned} \mathbf{Y}_e &= \mathbf{S}_e^T \mathbf{z} + \boldsymbol{\epsilon}_m, \quad \boldsymbol{\epsilon}_m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_m) \\ Y_a &= \mathbf{s}_a^T \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\bar{\mathbf{z}}, \boldsymbol{\Sigma}_z) \end{aligned}$$

The prior application uncertainty is characterized by its standard deviation, $\sigma_{a,prior} = \sqrt{\mathbf{s}_a^T \boldsymbol{\Sigma}_z \mathbf{s}_a}$. For the linear model, a Bayesian update is the optimal procedure in an information-theoretic sense given an experimental measurement \mathbf{y}_m . The posterior covariance matrix, $\boldsymbol{\Sigma}_{z,post}$, is then estimated as follows, akin to the GLLS procedure:

$$\boldsymbol{\Sigma}_{z,post} = \boldsymbol{\Sigma}_z - \boldsymbol{\Sigma}_z \mathbf{S}_e (\mathbf{S}_e^T \boldsymbol{\Sigma}_z \mathbf{S}_e + \boldsymbol{\Sigma}_m)^{-1} \mathbf{S}_e^T \boldsymbol{\Sigma}_z$$

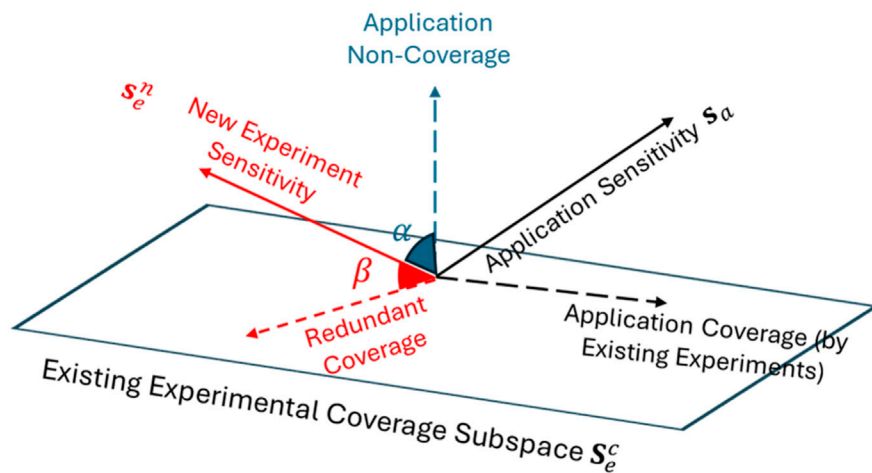


FIGURE 1
Graphical representation showing experimental coverage of application and the impact of adding a new experiment.

The adjustment covariance matrix is propagated through to the application, yielding the posterior standard deviation of the application response of interest, $\sigma_{a,post} = \sqrt{s_a^T \Sigma_{z,post} s_a}$. The proposed coverage metric, q_C , can be expressed solely in terms of the application and experimental sensitivities and the provided covariance information in Equation 6.

$$q_C^{LG} = 1 - \sqrt{\frac{s_a^T \Sigma_z (I - S_e (S_e^T \Sigma_z S_e + \Sigma_m)^{-1} S_e^T \Sigma_z) s_a}{s_a^T \Sigma_z s_a}} \quad (6)$$

3.2 Assessing the value of individual experiments

While q_C^{LG} in Equation 6 provides the total coverage of an experimental set, it is also capable of doing so in an incremental fashion by highlighting individual experiments that provide the highest incremental coverage. Suppose the set of available experiments are divided into S_e^c , a matrix containing the sensitivities of experiments already selected for Bayesian update and providing coverage c , and s_e^n , a new experimental sensitivity under consideration, differentiated by the superscript n . Initially, S_e^c may be an empty set. Without loss of generality, Equation 3 is better interpreted by first setting the prior covariance matrix of the dependent variable to identity and neglecting the measurement uncertainty—that is, by setting $\Sigma_z = I$, $\Sigma_m = 0$. The first selected experiment that provides the greatest coverage, in this case, is the experiment that aligns most closely to the application—specifically, the one with the highest cosine similarity, as expected.

For subsequent selection, $(I - S_e (S_e^T \Sigma_z S_e + \Sigma_m)^{-1} S_e^T \Sigma_z)$ in Equation 3 reduces to a projection operator onto the null space of the selected experiments in S_e^c . This is denoted as *non-coverage* in this manuscript, as depicted in Figure 1, where the null operator acting on the application sensitivity creates a residual vector (in dashed blue). This denotes the component of the application sensitivity not explained or accounted for by the selected experiment, contrasting with the component already accounted

for (in dashed black) and denoted as *coverage*. Similarly, the operator acting on the new experiment sensitivity also yields a residual not already accounted for by the existing body of experiments. Here, the component already covered by the existing body of experiments (in dashed red) is irrelevant or redundant. In the general case, Equation 3 reduces to Equation 7 below.

$$q_C = 1 - \sqrt{1 - \frac{\cos^2 \alpha}{\sin^2 \beta}} \quad (7)$$

Herein lies a key deviation from c_k -based methods. Whereas c_k is one-to-one, q_C considers the entire subspace spanned by the selected experiments in order to account for potential redundancies and/or synergies between experiments. The visual representation in Figure 1 shows that a new experiment with sensitivity s_e^n forms two angles: α with the application non-coverage vector and β with its (redundant) projection on the subspace spanned by S_e^c . Note that although the two angles appear complementary in Figure 1 indicating perfect coverage (or $q_C = 1$), this is because we have restricted it to a 2-D plane in three dimensions for visualization, and therefore perfect coverage is obtained as long as s_e^n has a component orthogonal the 2-D plane. In higher dimensions, the coverage subspace is a hyperplane, i.e., $\alpha + \beta$ may not necessarily equal 90° and therefore may not necessarily provide perfect coverage of $q_C = 1$.

Additionally, c_k considers only the relationship between s_e^n and s_a and may fail to account for redundancies ($\beta \sim 0^\circ$, $\alpha \sim 90^\circ$) where the new experiment adds minimal value. It may also miss cases of synergy (small α angles), where a traditionally low c_k experiment is capable of providing maximal coverage when combined with another experiment by jointly generating a subspace containing the application sensitivity. Through this iterative process, it is possible to select a subset of experiments that provide high coverage of a given application and converge quickly to the theoretical limit offered by the entire available body of experiments.

Note that despite the presence of a cosine (like c_k), the new experiment with the highest coverage may not necessarily be the one

with the highest cosine similarity ($\cos \alpha$) with the application non-coverage but must also balance against the redundancy ($\sin \beta$) from previous experiments. If the practitioner wishes to use c_k , they must re-project all the remaining experimental sensitivities onto the null operator defined above at each iteration, effectively ensuring $\sin \beta = 1$ for all remaining candidate experiments for selection.

When $\Sigma_z \neq \mathbf{I}$, the cosine similarity in the first experimental selection is replaced by c_k , and all experimental and application sensitivities are projected onto the subspace of the prior covariance. For problems of practical interest to the nuclear community, the effect of measurement uncertainty (i.e., $\Sigma_m \neq \mathbf{0}$) is a perturbation of the sensitivities in the system, which in turn has a “smoothing” effect on the coverage of each experiment. For linear (L), Gaussian (G), zero measurement uncertainty (Z), and a single experiment (S), there is no iterative selection process, and the q_C metric is related to c_k as indicated below in Equation 8. As it is a monotonic function, the highest c_k experiment provides the greatest coverage under LGZS conditions.

$$q_C^{LGZS} = 1 - \sqrt{\frac{\mathbf{s}_a^T \Sigma_z \mathbf{s}_a - \frac{(\mathbf{s}_a^T \Sigma_z \mathbf{s}_e)^2}{(\mathbf{s}_e^T \Sigma_z \mathbf{s}_e)}}{\mathbf{s}_a^T \Sigma_z \mathbf{s}_a}} = 1 - \sqrt{1 - \frac{(\mathbf{s}_a^T \Sigma_z \mathbf{s}_e)^2}{(\mathbf{s}_a^T \Sigma_z \mathbf{s}_a)(\mathbf{s}_e^T \Sigma_z \mathbf{s}_e)}} = 1 - \sqrt{1 - c_k^2} \quad (8)$$

3.3 Monte Carlo adaptation

MC sampling is a popular tool in the nuclear community that is used to circumvent the problem of computing sensitivities and computing vector–matrix products when the dependent variable has high dimensionality, such as cross-sections on the order of 10^{4-5} . Some algorithms, such as MOCABA (Hoefer et al., 2015), utilize this technique to directly predict integral functions of nuclear data and avoid adjusting the intermediate nuclear data itself. Although the original q_C formulation in Equation 1 is expressed directly in terms of the integral functions and can be directly computed using random samples of the application responses and experimental data, this section derives an equivalent formulation for q_C^{LG} .

Revisiting the linear model below, we first observe that the inner product between sensitivities can be computed via random sampling as

$$\mathbf{S}_e^T \Sigma_z \mathbf{s}_a = \mathbb{E}[(\mathbf{S}_e^T \mathbf{z})(\mathbf{s}_a^T \mathbf{z})] = \mathbb{E}[(\mathbf{S}_e^T \mathbf{z} + \boldsymbol{\epsilon}_m)(\mathbf{s}_a^T \mathbf{z})]$$

since $\mathbb{E}[(\boldsymbol{\epsilon}_m)(\mathbf{s}_a^T \mathbf{z})] = 0$.

$$\begin{aligned} \mathbf{Y}_e &= \mathbf{S}_e^T \mathbf{z} + \boldsymbol{\epsilon}_m, & \boldsymbol{\epsilon}_m &\sim \mathcal{N}(\mathbf{0}, \Sigma_m) \\ Y_a &= \mathbf{s}_a^T \mathbf{z}, & \mathbf{z} &\sim \mathcal{N}(\bar{\mathbf{z}}, \Sigma_z) \end{aligned}$$

Here, \mathbf{z} is sampled according to the prior covariance matrix Σ_z . Similarly, all inner products in Equation 6 may be identified and replaced. Given the presence of the expectation operator in both the numerator and denominator of Equation 6, it can be reduced to summation across all N samples of the responses. We introduce \mathbf{Y}_{Ne} and \mathbf{Y}_{Na} to describe the matrix and vector containing N samples of the experimental data and application responses, respectively, along each column with the mean subtracted. Note that computer models typically do not account for measurement uncertainty, and the user must simulate additional measurement uncertainty to avoid overestimating the coverage. We may compute q_C^{LG} using Monte Carlo using Equation 9 below.

$$\begin{aligned} q_C^{LG} &= 1 - \sqrt{\frac{\mathbf{s}_a^T \Sigma_z \mathbf{s}_a - \mathbf{s}_a^T \Sigma_z \mathbf{S}_e (\mathbf{S}_e^T \Sigma_z \mathbf{S}_e + \Sigma_m)^{-1} \mathbf{S}_e^T \Sigma_z \mathbf{s}_a}{\mathbf{s}_a^T \Sigma_z \mathbf{s}_a}} \\ &= 1 - \sqrt{1 - \frac{(\mathbf{Y}_{Na}^T \mathbf{Y}_{Ne})(\mathbf{Y}_{Ne}^T \mathbf{Y}_{Ne})^{-1} (\mathbf{Y}_{Ne}^T \mathbf{Y}_{Na})}{\mathbf{Y}_{Na}^T \mathbf{Y}_{Na}}} \end{aligned} \quad (9)$$

Following similar arguments presented in Section 3.2 and visualized in Figure 1, the methodology reduces to a minimum-norm least-squares procedure. It fits the simulated experiments (inflated by measurement uncertainty) to the application response of interest, assuming that the underlying parameter \mathbf{z} is sampled according to the prior covariance Σ_z . This approach achieves the same outcome as the GLLS procedure.

4 Numerical experiments

This section illustrates the value of q_C to the data assimilation community in (a) identifying synergistic experiments that greatly increase coverage currently discarded as having low relevance and (b) accurately capturing experimental coverage for nonlinear applications and/or experiments. Three numerical experiments are considered: (1) a purely linear problem, (2) a nonlinear problem, and (3) quantifying coverage among 100 randomly selected benchmarks from the ICSBEP handbook for a given application.

4.1 Linear model

Consider the following analytical problem composed of a set of four experiments and a single application response described by a linear model with the same notation as that given in the previous section. The four experimental outputs, which are denoted by the vector $\mathbf{Y}_e = [Y_e^1 Y_e^2 Y_e^3 Y_e^4]^T$, represent eigenvalues that are linear functions of the cross-sections, represented by the six-dimensional vector $\mathbf{z} = [z_1 z_2 z_3 z_4 z_5 z_6]^T$. The application eigenvalue, denoted by the scalar Y_a , is also a linear function of the cross-sections. The mean and prior covariance of the cross-sections are denoted by $\bar{\mathbf{z}}$ and Σ_z , respectively. The objective of this model is to identify the experiments that yield the greatest coverage of the application, and to predict the application bias and uncertainty at some unknown condition (e.g., hot full-power). Once the practitioner identifies the relevant experiments, they apply a Bayesian update to obtain a posterior estimate of the application bias and uncertainty.

$$\begin{aligned} \mathbf{Y}_e &= \begin{bmatrix} -0.0024 & 0.0072 & 0.0033 & -0.0033 & -0.0185 & 0.0143 \\ 0.0020 & -0.0030 & -0.0020 & 0.0036 & -0.0100 & 0.0223 \\ 0.0102 & 0.0054 & 0.0039 & -0.0173 & 0.0098 & 0.0089 \\ 0.0039 & 0.0065 & -0.0029 & -0.0058 & -0.0151 & 0.0172 \end{bmatrix} \\ &\quad \times \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \end{bmatrix} + \begin{bmatrix} 1.120 \\ 0.995 \\ 0.950 \\ 1.080 \end{bmatrix} + \boldsymbol{\epsilon}_m \\ Y_a &= -0.0020z_1 + 0.0024z_2 + 0.0009z_3 - 0.0004z_4 - 0.0084z_5 \\ &\quad + 0.0048z_6 + 1.14 \end{aligned}$$

TABLE 1 Numerical experiment 1 data sheet.

Experiment	c_k	Reference	Measurement (mean)
Experiment 1 (Y_e^1)	0.9598	0.9882	0.9792
Experiment 2 (Y_e^2)	0.9585	0.9885	0.9795
Experiment 3 (Y_e^3)	−0.2009	1.0082	1.0228
Experiment 4 (Y_e^4)	0.8627	0.9988	0.9937
Application (Y_a)	—	1.0815	—

$$\begin{aligned} \mathbf{z} &\sim \mathcal{N}(\bar{\mathbf{z}}, \Sigma_z), \boldsymbol{\epsilon}_m \sim \mathcal{N}(\mathbf{0}, 10^{-6} * \mathbf{I}) \\ \bar{\mathbf{z}} &= [7.2056 \ 7.3183 \ 0.8779 \ 10.1643 \ 9.4983 \ 2.7368] \\ \Sigma_z &= \begin{bmatrix} 0.2997 & 0.0728 & 0.0326 & -0.2604 & 0.2862 & -0.0670 \\ 0.0728 & 0.0992 & 0.0297 & -0.1685 & 0.0873 & 0.0308 \\ 0.0326 & 0.0297 & 0.0893 & -0.1098 & 0.0105 & 0.1196 \\ -0.2604 & -0.1685 & -0.1098 & 0.4998 & -0.2460 & -0.2558 \\ 0.2862 & 0.0873 & 0.0105 & -0.2460 & 0.4769 & -0.2630 \\ -0.0670 & 0.0308 & 0.1196 & -0.2558 & -0.2630 & 0.7697 \end{bmatrix} \end{aligned}$$

The experimenter performs highly accurate experiments to obtain measurements at these unknown conditions with zero-mean error, denoted by $\boldsymbol{\epsilon}_m$, and measurement uncertainty represented by a purely diagonal covariance matrix (i.e., it is proportional to the identity matrix \mathbf{I}). For brevity, constants are displayed to four digits of precision. The application and experimental eigenvalues are typically known at the reference conditions, such as cold conditions, fresh fuel, etc., at $\bar{\mathbf{z}}$ and tabulated in Table 1, along with the c_k values using first-order sensitivities.

4.2 Nonlinear model

This numerical experiment is designed to demonstrate the ability of q_C to capture coverage in nonlinear models in which nonlinearities may exist in both the experiments and applications. In this scenario, first-order sensitivities analyzed by metrics such as c_k may not necessarily capture all possible variations in the response of interest; likewise, using techniques that assume linearity of the experiments and applications, such as GLLS, may even provide under- or over-confident *posteriori* estimates of the predicted application bias and uncertainty. Consider the following nonlinear problem, of which all notation is consistent with the previous section. The dimensionality of the problem is reduced to three cross-sectional dimensions, two experiments, and a single application.

$$\begin{aligned} \mathbf{Y}_e &= \begin{bmatrix} Y_e^1 \\ Y_e^2 \end{bmatrix} = \begin{bmatrix} f(z_1, z_2, z_3) \\ g(z_1, z_2, z_3) \end{bmatrix} + \boldsymbol{\epsilon}_m \\ Y_a &= h(z_1, z_2, z_3) \\ \mathbf{z} &\sim \mathcal{N}(\bar{\mathbf{z}}, \Sigma_z), \boldsymbol{\epsilon}_m \sim \mathcal{N}(\mathbf{0}, 10^{-6} * \mathbf{I}) \\ \bar{\mathbf{z}} &= [6.86214 \ 8.07431 \ 4.18035] \\ \Sigma_z &= \begin{bmatrix} 0.05668 & 0.00443 & -0.01213 \\ 0.00443 & 0.06333 & 0.01029 \\ -0.01213 & 0.01029 & 0.00690 \end{bmatrix} \end{aligned}$$

TABLE 2 Numerical experiment 2 data sheet.

Experiment	c_k	Reference	Measurement (mean)
Experiment 1 (Y_e^1)	−0.1679	1.0083	1.0174
Experiment 2 (Y_e^2)	0.9985	1.0015	1.0096
Application (Y_a)	—	1.0017	—

The objective remains the same: to identify the experiments that yield the greatest coverage of the application and to predict the application bias and uncertainty at some unknown condition (e.g., hot full-power). The full expressions include up to third-order and cross-interaction terms ($z_1 z_2, z_3 z_1^2$, etc.) and are provided in the Supplementary Material. Like the previous numerical experiment, the application and experimental eigenvalues are known at the reference conditions, and the c_k values, which are calculated using the first-order sensitivities at the reference point, are tabulated in Table 2.

4.3 ICSBEP benchmarks

In this numerical experiment, a randomly selected subset of 100 benchmarks from the ICSBEP handbook are chosen as a potential list of candidate experiments for GLLS, along with one application benchmark, namely, HEU-SOL-THERM-013–003. The 100 benchmarks have absolute c_k values across the entire spectrum of 0.0–1.0, with 80 benchmarks having $c_k > 0.8$, among which 19 benchmarks are highly similar with $c_k > 0.9$, and 9 benchmarks have $c_k > 0.97$. The highest c_k value in the subset is 0.986. The initial application uncertainty, as measured by the standard deviation, is approximately 0.00718 (718 pcm).

The objective of the numerical experiment is to demonstrate the ability of q_C to capture the theoretical coverage provided by the entire set of experiments, identify the experiment that provides the most coverage, and deploy the iterative selection to procedure to select relevant experiments. The GLLS procedure is then applied to compute the posterior bias based on the selected experiments. The approach is then compared to the c_k -based threshold and rejection approaches outlined in (Broadhead et al., 2004; Kiedrowski, 2014), where the benchmarks are ordered in decreasing order of c_k . We also demonstrate the marked break in the estimated bias with the inclusion of experiments in order of their c_k as observed in (Broadhead et al., 2004) where the c_k cutoff is lowered, whereas the effect is eliminated in the q_C -based approach.

5 Results and discussion

This section demonstrates the q_C metric in capturing coverage for both linear and nonlinear problems for which the response distribution may or may not be Gaussian. It also compares the performance of q_C to existing c_k -based methodologies used in the nuclear community and evaluates their respective impacts on data assimilation tools such as GLLS and machine learning.

TABLE 3 Computed q_C for numerical experiment 1.

Experiments	c_k values	q_C (MC)	q_C (Kraskov)	q_C^{LG}
Experiment 1	0.9618	0.7235	0.7235	0.7227
Experiments 1 and 2	0.9618, 0.9309	0.7235	0.7244	0.7228
Experiments 1 and 3	0.9618, -0.2323	0.9550	0.9554	0.9549
Experiments 1 and 4	0.9618, 0.8527	0.8646	0.8651	0.8643
Experiments 3 and 4	-0.2323, 0.8527	0.8835	0.8841	0.8834
Experiments 1, 2 and 3	0.9618, 0.9309, -0.2323	0.9557	0.9560	0.9555
Experiments 1, 2 and 4	0.9618, 0.9309, 0.8527	0.8842	0.8846	0.8839
All four experiments		0.9574	0.9577	0.9573

Bolded values indicate experiments that must be selected to provide coverage close to the theoretical maximum.

5.1 Linear model

In this section, we investigate experimental relevance in the linear problem in Section 4.1 via two approaches: c_k and q_C . The q_C -based methodology analyzes the four available experiments using the mutual information in Equation 1 and tabulated in Table 3. The MC methodology described in Section 3.3 and Equation 5 was deployed by simulating $N = 100,000$ samples of the application and experiments and corrupting the experiment simulation with measurement uncertainty. The analytical linear Gaussian expression in Equation 3 and a k-nearest-neighbor-based mutual information algorithm (Kraskov et al., 2003) are provided for comparison.

Computing q_C prior to applying the GLLS procedure with the application reveals the following:

- q_C reveals that experiments 1 and 3 provide the highest coverage for the application, and this method is capable of achieving ~95% reduction in the prior application uncertainty. This is counterintuitive to a pure c_k -based approach, which would nominally discard experiment 3 due to its low c_k value and would include only experiments 1 and 2.
- q_C reveals redundancy between experiments 1 and 2 despite their high c_k values, demonstrating that they provide very little additional value. For example, the addition of experiment 2 to experiment 1 adds a reduction of only 0.01% in uncertainty as measured by the q_C value.
- q_C reveals significant synergy between two mid/low- c_k experiments. The set containing experiments 3 and 4 provides 88.34% reduction, whereas a set containing experiments 1 and 2 (high c_k) provides a reduction of only 72.28%.

Once the experiments are selected and included in the relevant set by the practitioner, the GLLS procedure/Bayesian update as described in Section 3.1 is applied to obtain a posterior bias and uncertainty. The table below shows the GLLS results for the above cases with the reduction in ratio of standard deviations, indicating a perfect match with that predicted by q_C and validating Equation 2.

The results in Table 4 display some counterintuitive behavior, as noted above, primarily due to the inability of c_k to capture redundancies and synergies between experiments because it instead assesses experiments individually. It is unable to capture the value of an experiment *a priori* and may potentially discard high-value experiments that provide high coverage.

Furthermore, applying the GLLS procedure in descending order of c_k displays unstable behavior: drastic changes in bias are observed when low c_k experiments are included. The inclusion of experiments in the order of $1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \dots$, in descending order of c_k , yields unstable bias estimates ($-355 \rightarrow -355 \rightarrow -486 \rightarrow -505$) and a sharp reduction in uncertainty when low c_k experiments are added. This behavior is also reported in (Broadhead et al., 2004) and demonstrated in the ICSBEP benchmark results in Section 5.3.A practitioner, on the other hand, may anticipate and desire convergent behavior as more experiments are included and their relevance decreases. This outcome is achieved by ordering experiments by the additional coverage provided as described by the iterative selection process in Sections 2, 3.2, which quickly converges to a stable bias estimate with the inclusion of just two experiments (1 and 3).

Since the problem is linear, a subspace analysis of the sensitivities can also be performed by projecting the application sensitivities onto the experiments and computing the cosine between the application sensitivity and its projection, as visualized in Section 3.2. Since the prior covariance matrix $\Sigma_z \neq \mathbf{I}$, all application and experimental sensitivities are first projected onto the subspace of the prior covariance matrix and normalized. The cosine is calculated according to Equation 10 below.

$$\cos \theta = \frac{\mathbf{s}_a^T \Sigma_z \mathbf{S}_e (\mathbf{S}_e^T \Sigma_z \mathbf{S}_e)^{-1} \mathbf{S}_e^T \Sigma_z \mathbf{s}_a}{\sqrt{\mathbf{s}_a^T \Sigma_z \mathbf{s}_a} \sqrt{\mathbf{S}_e^T \Sigma_z \mathbf{S}_e (\mathbf{S}_e^T \Sigma_z \mathbf{S}_e)^{-1} \mathbf{S}_e^T \Sigma_z \mathbf{s}_a}} \quad (10)$$

As in Table 5, we see that the cosine expression reduces to the c_k value for a single application and a single experiment and can be extended to evaluate relevance between an application and multiple experiments. Note that this analysis does not account for the perturbation introduced by the measurement uncertainty, but it is assumed that the effect cancels out because it is identical across all four experiments. The highest coverage is achieved when

TABLE 4 GLLS results for numerical experiment 1.

Experiments included	Application bias	Reduction, $\sigma_{a,prior} = 913$ pcm	True solution
Experiment 1	-355 ± 253 pcm	0.7227	−506 pcm
Experiments 1 and 2	-355 ± 253 pcm	0.7228	
Experiments 1 and 3	-505 ± 41 pcm	0.9549	
Experiments 1 and 4	-473 ± 124 pcm	0.8643	
Experiments 3 and 4	-502 ± 106 pcm	0.8834	
Experiments 1, 2 and 3	-505 ± 41 pcm	0.9555	
Experiments 1, 2 and 4	-486 ± 106 pcm	0.8839	
All four experiments	-505 ± 39 pcm	0.9573	

Bolded values indicate experiments that must be selected to provide coverage close to the theoretical maximum.

TABLE 5 Linear subspace analysis for numerical experiment 1.

Experiments included	$\cos \theta$ between application sensitivity and its projection onto experiment sensitivities
Experiment 1	0.9618
Experiment 2	0.9309
Experiment 3	−0.2323
Experiment 4	0.8527
Experiments 1 and 2	0.9618
Experiments 1 and 3	1.0000
Experiments 1 and 4	0.9949
Experiments 3 and 4	0.9945

Bolded values indicate experiments that must be selected to provide coverage close to the theoretical maximum.

experiments 1 and 3 are included, with a cosine of 1 indicating perfect coverage. In other words, the application vector is fully contained in the subspace spanned by the sensitivities of experiments 1 and 3, and therefore the application can be predicted perfectly if the experimental measurements are known with zero measurement uncertainty. For nonlinear problems, the sensitivity at the reference is not sufficient to characterize all variations in the application, and other extensions (Bang et al., 2012) may be suitable to identify relevant experiments.

It is important to note that the experimental selection problem does not have an optimal substructure. In other words, the set of n experiments providing the highest coverage of a given application does not necessarily contain the set of $n - 1$ experiments that provide the highest coverage. As a counterexample, consider an application sensitivity $\mathbf{s}_a = [1\ 1\ 0]$ and a set of three experimental sensitivities, $\mathbf{s}_e^1 = [1\ 0\ 0]$, $\mathbf{s}_e^2 = [0\ 1\ 0]$, and $\mathbf{s}_e^3 = [1\ 1\ 0.01]$. The single experiment that provides the highest coverage is \mathbf{s}_e^3 ; however, it is not contained in the set of two experiments \mathbf{s}_e^1 and \mathbf{s}_e^2 that provide perfect coverage of the application. Building upon this counterexample, the iterative procedure of picking the experiment with highest coverage and appending to that set may not yield the optimal subset in the

TABLE 6 Computed q_C for numerical experiment 2.

Experiments	c_k values	q_C (MC)
Experiment 1	−0.2214	0.3795
Experiment 2	0.9985	0.5823
Both Experiments		0.9282

sense of maximizing coverage with n selected experiments. In the machine learning community, a similar problem exists in feature selection, where the best n features are not necessarily the n best features selected iteratively. Nevertheless, the problem of identifying the optimal subset of experiments by brute force is computationally intractable beyond a handful, and the iterative process mentioned above works reasonably well in practice and converges to a stable bias and uncertainty estimate on tests performed on larger datasets of 100 experimental evaluations from the ICSBEP Handbook as demonstrated in Section 5.3.

5.2 Nonlinear model

This section concerns experimental coverage for a nonlinear model given two experiments. As indicated in Table 2, a c_k -based approach would typically cause a practitioner to discard experiment 1 and assume that experiment 2 provides the highest coverage. We computed q_C in Table 6 below using MC sampling and the k -nearest-neighbors algorithm for mutual information.

The q_C results indicate that the two experiments provide significant coverage of the application individually and near-full coverage synergistically despite the wide variation in c_k values. Figure 2 below depicts the variation in the application response with respect to the two experiments, indicating that near-complete coverage is feasible. As a proof-of-concept, we utilized a three-layer feedforward neural network to compute an approximate fit of the function and visualize the surface of the predicted response on the 3D plot in Figure 2. It is observed that the application eigenvalues lie on the surface of the neural network-predicted response with low error (arising out of measurement uncertainty). Note that the neural net training itself is not the focus of the manuscript, and the fit is

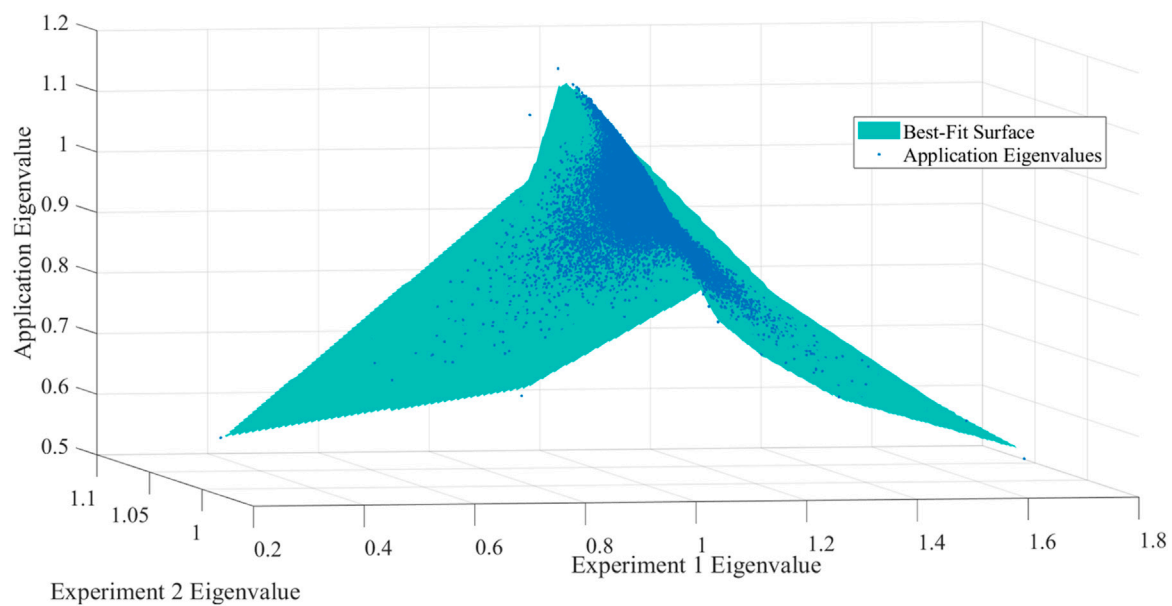


FIGURE 2
Plot depicting variation of application/experiments eigenvalues and the best-fit surface.

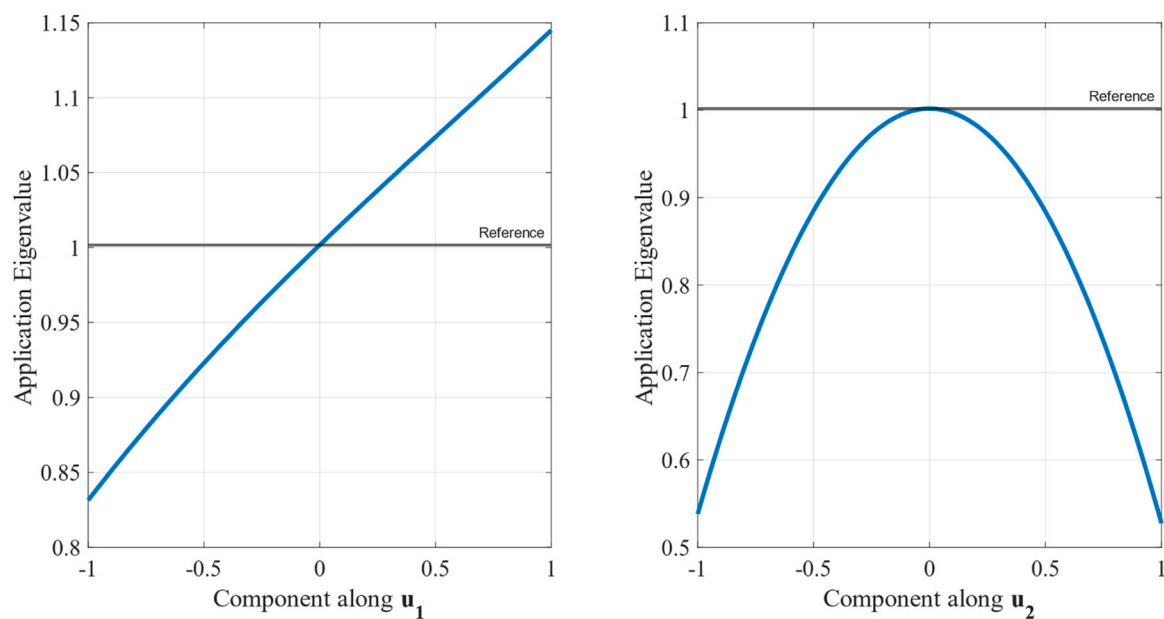


FIGURE 3
Application Variations plotted along its active subspace directions.

purely for illustrative purposes and to support the claim that the application is covered by the experiments.

Through active subspace techniques (Bang et al., 2012), we observe that the application response of interest primarily varies nonlinearly along two directions, which are orthogonalized into directions \mathbf{u}_1 and \mathbf{u}_2 and visualized in Figure 3. However, \mathbf{u}_2 vanishes at the reference point due to symmetry and is therefore not captured by the application sensitivity at the reference point (and therefore c_k which is taken at the

reference point). Repeating the analysis on the experiments indicates a single dominant direction of variation for each experiment despite the nonlinearity, indicating that the sensitivities at the reference point, \mathbf{s}_e^1 and \mathbf{s}_e^2 respectively, are sufficient to characterize the experiments over the entire range. The variation of the experiments along this direction is visualized in Figure 4.

Through the active subspace algorithm, the sensitivities are now amenable to the subspace analysis performed previously. Using

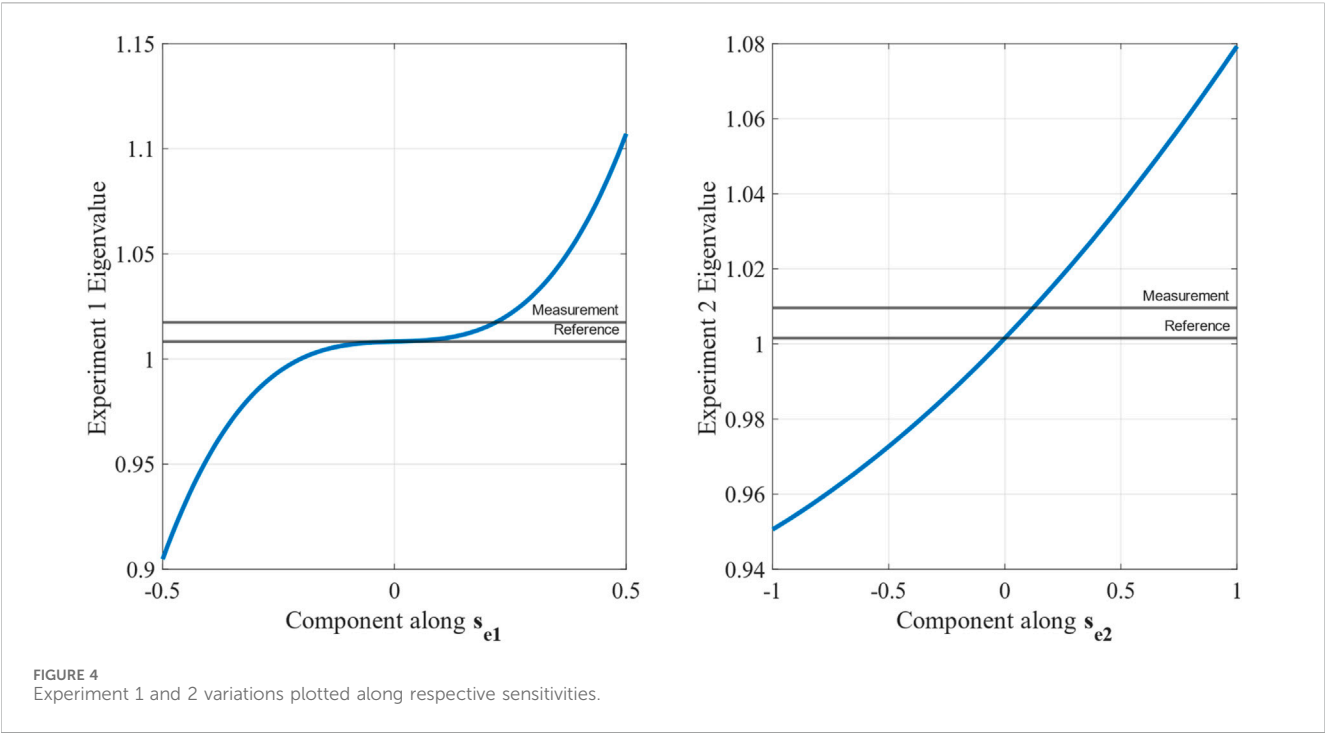


TABLE 7 Nonlinear subspace analysis for numerical experiment 2.

Experiments included	$\cos \theta$ with u_1	$\cos \theta$ with u_2
Experiment 1	-0.2214	1.0000
Experiment 2	0.9985	-0.1679
Experiments 1 and 2	1.0000	1.0000

Equation 6, the cosine of each application sensitivity is computed with respect to the experiments and is tabulated in Table 7. It is observed that the combination of two experiments provides perfect coverage of both directions of variations of the applications: that is, the combined set of experiments spans the same subspace as that by u_1 and u_2 . This further supports the claim that the two experiments provide near-perfect coverage of the application when measurement uncertainty is also considered.

Next, we demonstrate that the naive GLLS methodology leads to incorrect adjustments due to experimental nonlinearity, incorrect estimates of the bias due to application nonlinearity, and mischaracterization of the *posteriori* distribution with the standard deviation due to non-Gaussianity (arising out of nonlinearity).

First, we observe that the experiment linearity assumption in GLLS causes a miscalculation of the adjustment itself. Intuitively, GLLS attempts to find a cross-section adjustment along the subspace spanned by the experimental sensitivities $S_e = [s_e^1 s_e^2]$ that minimizes the squared Euclidean distance from both the reference and the measurement. However, experimental nonlinearity may cause a significant overshoot or undershoot in the calculated adjustment depending on the concavity/convexity of the function, leading to an error in the bias estimate.

Second, we observe that the application linearity assumption in GLLS causes it to consider only the variation of the application given

by the sensitivity at the reference u_1 and thus neglect other variations (u_2). Therefore, it assumes experiment 1 is of low relevance, resulting in minimal adjustment to the covariance matrix. The opposite occurs when experiment 2 is included; the experimental relevance is overstated, and the predicted posterior uncertainty is highly overconfident. The application nonlinearity causes further error in the predicted bias as the procedure assumes linearity and computes the inner product of the adjustment vector with u_1 .

The issue of experimental nonlinearity can be corrected with a nonlinear adaptation to the GLLS procedure. Similar to GLLS, the adjustment is made along s_{e1} and s_{e2} to minimize the squared Euclidean distance from both the reference and the measurement, as visualized in Figure 4. However, the process may be performed iteratively in small steps with updated local gradients to account for the experimental nonlinearity as in (Sobes et al., 2016), or via nonlinear least-squares if the functional form of the experiment is known, or an inverse neural network model if it can be approximated. While a nonlinear iterative solver was utilized in this work, the specific method is left to the practitioner and outside the scope of the present work which focuses on the theoretically achievable coverage.

Regarding the issue of application nonlinearity, the bias is traditionally computed by considering only the reference sensitivity in GLLS. Instead, in the proposed adaptation, once an accurate adjustment is obtained from the above iterative procedure, it is input into the forward model of the application to provide an accurate estimate of the posterior bias. However, the nonlinearity of the application almost certainly makes the posterior distribution non-Gaussian even if the underlying parameters z are Gaussian. To tackle this challenge, we adopted an MC simulation approach by repeatedly sampling measurements from the provided measurement Y_e and simulated measurement uncertainty Σ_m ; we then repeated the above iterative procedure to obtain the posterior distribution of

TABLE 8 Comparison of methodologies for numerical experiment 2.

Methodology	Application bias	True solution
Naïve GLLS	995 ± 256 pcm	−1,016 pcm
Nonlinear Adaptation	−1011 ± 256 pcm	

\mathbf{z} , $p(\mathbf{z}|\mathbf{Y}_e)$ in the Bayesian sense. The posterior distribution is then input into the forward model to estimate the corresponding uncertainty of the application bias, obtain confidence intervals, etc. Note that both the posterior distribution of \mathbf{z} and the application uncertainty may be non-Gaussian and may not necessarily be fully characterized by the covariance matrix $\Sigma_{\mathbf{z}}$ or $\sigma_{a,post}$. For subsequent Bayesian updates, if desired, they may be approximated using known distributions or constructed using copulas to produce computationally tractable distributions.

Table 8 shows the predicted application bias with naïve GLLS, the proposed adaptation, and the trained neural network. As predicted, naïve GLLS provides an incorrect posterior estimate due to the experimental and application nonlinearity, which is resolved with the nonlinear adaptation. In this experiment, the posterior distribution can be approximated by a Gaussian, and the mean and standard deviation are provided in Table 8. Note that this may not be true in the general case, and it may be better suited to provide confidence intervals instead.

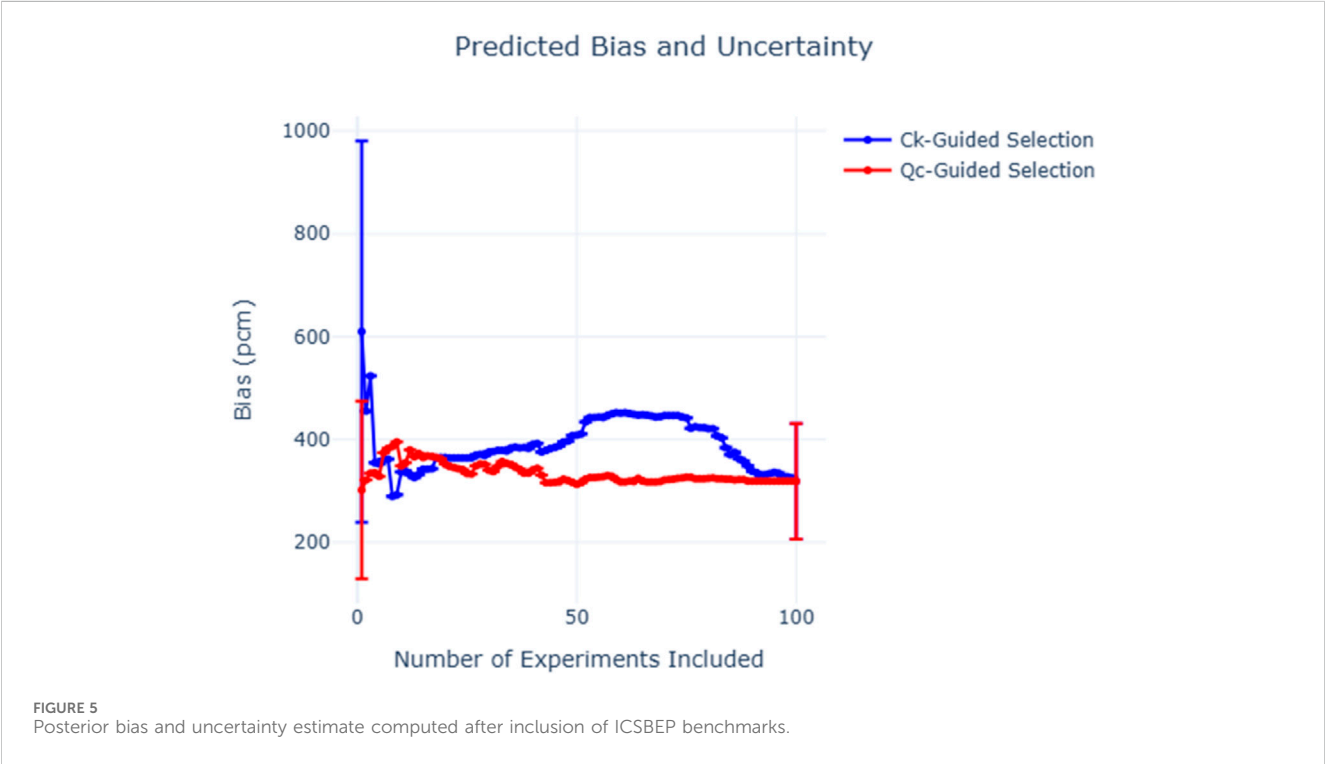
5.3 ICSBEP benchmark

As presented in Section 4.3, the selected benchmarks are ordered in decreasing order of their absolute c_k values with the GLLS procedure applied after each benchmark inclusion. The

procedure is then repeated using the iterative q_C procedure outlined in Section 2 and Section 3.2 – the first selected benchmark has the highest q_C , and subsequent benchmarks are selected in a manner that maximizes the total q_C of the selected set with the application. The posterior bias along with uncertainty (standard deviation) are provided in Figure 5 below. We also compute the total q_C of all 100 benchmarks to be 0.84, i.e., a theoretical 84% reduction in uncertainty achievable if all experiments are included.

We first note that the q_C -based approach accounts for measurement uncertainty of each benchmark in the set and identifies an experiment with $c_k = 0.978$ as providing the highest value to the chosen application, whereas the c_k -based approach selects the benchmark with $c_k = 0.986$. When the GLLS procedure is applied, the posterior uncertainty from the q_C approach with just one experiment reduces to 173 pcm from a prior of 718 pcm, a reduction of 76.07%. This is a marked improvement from the c_k -based approach which yields 371 pcm, i.e., only a 48.63% reduction in the prior uncertainty.

Figure 5 depicts significant breaks in the bias using the c_k based approach as the threshold is lowered to include experiments with c_k below 0.97, 0.86, 0.845, 0.814, 0.73, and 0.6. On the other hand, the q_C -based approach quickly stabilizes with the inclusion of ~25 experiments of varying c_k values, indicating that the remaining experiments add little value. A traditional c_k -based approach such as in (Broadhead et al., 2004) with a cutoff of $c_k = 0.9$ and 23 benchmarks yields a posterior bias of 364 ± 133 pcm, and $c_k = 0.8$ and 80 benchmarks yields a posterior bias of 421 ± 123 pcm. On the other hand, the q_C -based approach yields 343 ± 122 pcm and 325 ± 112 pcm respectively, compared to the inclusion of the full set of 100 benchmarks that yields a posterior bias of 319 ± 112 pcm with the GLLS procedure.



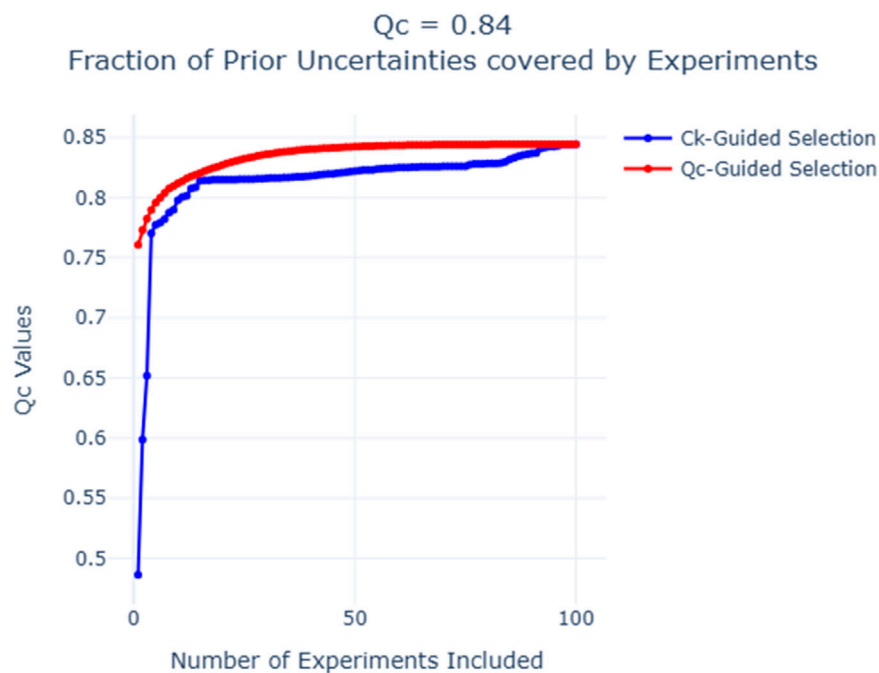


FIGURE 6
Comparison of the c_k and q_C -based selection procedures in capturing coverage.

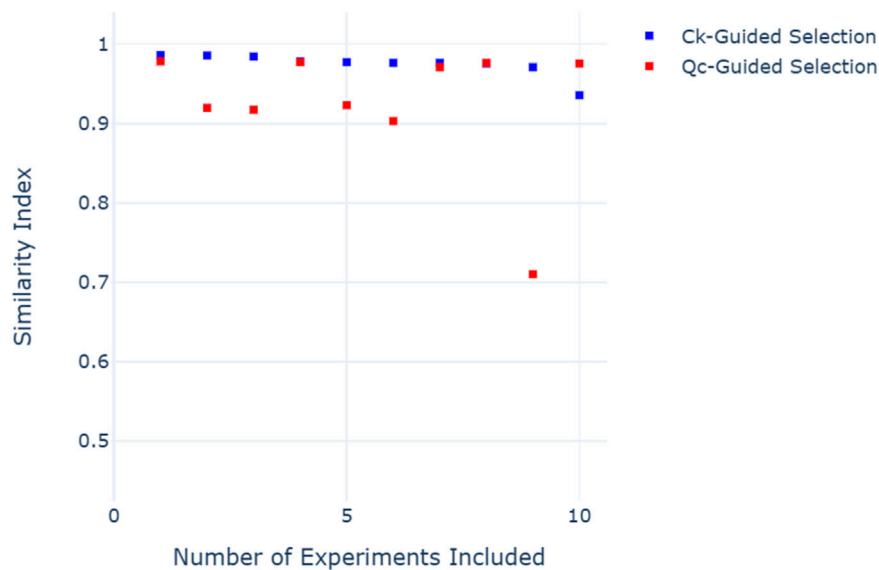


FIGURE 7
 c_k values of selected experiments between the two approaches.

The quicker convergence is also supported by Figure 6 which computes the q_C value after the inclusion of only a subset of experiments from the set of 100 benchmarks. For comparison, the number of benchmarks required by the c_k -threshold approach to achieve the same q_C is depicted. For instance, the c_k -based approach requires four benchmarks to achieve the same coverage as just one benchmark with the q_C approach. The q_C -based approach also converges faster, achieving $q_C = 0.83$ (within 0.01 of

the theoretical maximum) with 23 benchmarks, while the c_k -based approach requires 85 benchmarks and a c_k threshold of 0.71.

Upon closer inspection of the first ten benchmarks selected by the q_C approach as depicted in Figure 7, we notice that the approach prefers the selection of benchmarks with c_k values as low as 0.7 and 0.9 over some of the benchmarks with $c_k > 0.97$, indicating some degree of redundancy among the high c_k experiments. In fact, when expanding the selection to the first 23 benchmarks, the

q_C -based approach selects 7 benchmarks with $c_k \in (0.47, 0.82)$, benchmarks that would otherwise be discarded with a threshold-based approach.

While this study pertains to a set of 100 benchmarks, the ability to extract value out of low c_k experiments is expected to be valuable for applications where the number of available benchmarks is limited and/or there is a lack of sufficiently high c_k experiments. In such cases, q_C advises the practitioner on the most information that can be extracted, which can then be used to inform future experiments to improve their estimates.

6 Conclusion

The manuscript introduces q_C , a novel coverage metric to capture and quantify experimental relevance in nuclear datasets independent of the specific data assimilation procedure used. Unlike existing metrics such as c_k that are better suited for single-experiment, linear, and Gaussian problems, q_C works across a wide range of problems; here, a linear and a nonlinear analytical problem are demonstrated as well as a real case with a set of 100 benchmarks from the ICSBEP handbook. We demonstrated that c_k , as a one-to-one metric, is not capable of identifying experimental redundancies and synergies in providing coverage of an application response of interest and may potentially discard high-value experiments as irrelevant while overstating the value of highly similar but low-value redundant experiments. q_C , on the other hand, captures coverage for not only a single experiment and linear and Gaussian problems, but also for nonlinear problems with non-Gaussian distributions and the coverage across multiple experiments.

The scope of this manuscript is confined to the theoretical underpinnings of the approach, its relationship to existing coverage quantification metrics, the applicability of q_C to nonlinear and non-Gaussian problems, and its value in identifying valuable low- c_k experiments. As q_C provides theoretical coverage, it is also useful as a diagnostic tool to improve data assimilation algorithms, as demonstrated using a nonlinear adaptation of the GLLS procedure. The authors have applied the methodology successfully to various reactor physics applications. For instance, power and void histories of spent fuel samples were inferred based on q_C analysis of nuclide concentrations from destructive assay measurements, leading to significant improvements in predicted isotopic concentrations due to corrected operational histories (Yin et al., 2024; Islam et al., 2024; Yin et al., 2025). We are currently investigating another use case involving targeted nuclear data covariance adjustments guided by q_C contributions from selected critical experiments. Future work will be focused on machine-learning applications for nuclear by combining neural network-based adjustment procedures with Bayesian uncertainty quantification to address nonlinearities and non-Gaussianity, where q_C is used as a diagnostic tool.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

AS: Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft. SY: Data curation, Investigation, Validation, Writing – review and editing. UM: Conceptualization, Project administration, Resources, Supervision, Validation, Visualization, Writing – review and editing. HA-K: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported in part by the U.S. Department of Energy's (DOE) National Nuclear Security Administration, Office of Defense Nuclear Nonproliferation Research and Development (NA-22). The authors also acknowledge the DOE/NRC Collaboration for Criticality Safety Support for Commercial-Scale HALEU for Fuel Cycles and Transportation (DNCSH) initiative for their support and collaboration.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Correction note

A correction has been made to this article. Details can be found at: [10.3389/fnuen.2025.1736634](https://doi.org/10.3389/fnuen.2025.1736634).

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Licenses and Permissions

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government

purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-publicaccess-plan>). All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

References

- Athe, P., and Abdel-Khalik, H. S. (2014). Mutual information: a generalization of similarity indices. *Trans. Am. Nucl. Soc.* 111, 751–754.
- Bang, Y., Abdel-Khalik, H. S., and Hite, J. M. (2012). Hybrid reduced order modeling applied to nonlinear models. *Int. J. Numer. Meth Eng.* 91 (9), 929–949. doi:10.1002/nme.4298
- Belghazi, M. I., Baratin, A., Rajeshwar, S., Ozair, S., Bengio, Y., Courville, A., et al. (2018). "MINE: mutual information neural estimation," in *Proceedings of the 35th international conference on machine learning*. Editors J. Dy and A. Krause (Stockholm, Sweden: PMLR), 531–540.
- Broadhead, B. L., Rearden, B. T., Hopper, C. M., Wagschal, J. J., and Parks, C. V. (2004). Sensitivity- and uncertainty-based criticality safety validation techniques. *Nucl. Sci. Eng.* 146 (3), 340–366. doi:10.13182/nse03-2
- Carrara, N., and Ernst, J. (2017). *On the upper limit of separability*. arXiv: 1708.09449 [hep-ex].
- Cover, T. M., and Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley and Sons.
- Dionisio, A., Menezes, R., and Mendes, D. A. (2004). Mutual information: a measure of dependency for nonlinear time series. *Phys. A* 344, 326–329. doi:10.1016/j.physa.2004.06.144
- Hoefer, A., Buss, O., Hennebach, M., Schmid, M., and Porsch, D. (2015). MOCABA: a general monte carlo-bayes procedure for improved predictions of integral functions of nuclear data. *Ann. Nucl. Energy*. 77, 514–521. doi:10.1016/j.anucene.2014.11.038
- Islam, T., Yin, S., Merturyek, U., and Abdel-Khalik, H. S. (2024). Estimation of void fraction history using post-irradiation fuel isotopics. *Trans. Am. Nucl. Soc.* 131 (1), 219–222. doi:10.13182/T131-46224
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Phys. Rev.* 106, 620–630. doi:10.1103/physrev.106.620
- Kiedrowski, B. C. (2014). *Methodology for sensitivity and uncertainty-based criticality safety validation*. Los Alamos (NM): Los Alamos National Laboratory. Report No.: LA-UR-14-23202.
- Kraskov, A., Stoegebauer, H., and Grassberger, P. (2003). Estimating mutual information. *Phys. Rev. E* 69 (6), 16. doi:10.1103/PhysRevE.69.066138
- Merturyek, U., and Abdel-Khalik, H. S. (2025). *Physics-guided analytical model validation*. [US Patent]. US12,367,408.
- Nuclear Energy Agency (1995). *Document content and format guide for the international criticality safety benchmark evaluation project (ICSBEP)*. Paris: OECD Nuclear Energy Agency. Report No.: NEA/NSC/DOC(95)03.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27 (4), 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x
- Sobes, V., Leal, L., Arbanas, G., and Forget, B. (2016). Resonance parameter adjustment based on integral experiments. *Nucl. Sci. Eng.* 183 (3), 347–355. doi:10.13182/nse15-50
- Williams, M. L., Broadhead, B. L., Jessee, M. A., Wagschal, J. J., and Lefebvre, R. A. (2011). *TSURFER: an adjustment code to determine biases and uncertainties in nuclear system responses by consolidating differential data and benchmark integral experiments*. Oak Ridge (TN): Oak Ridge National Laboratory. Report No.: ORNL/TM-2005/39, Version 6.1.
- Yin, S., Islam, T., Merturyek, U., and Abdel-Khalik, H. S. (2024). Potential burnup indicator identification based on power history decomposition. *Trans. Am. Nucl. Soc.* 131 (1), 276–279.
- Yin, S., Islam, T., Merturyek, U., Procop, G., and Abdel-Khalik, H. S. (2025). "EDIM local power history and burnup inference based on destructive assay data," in *Proceedings of the international conference on mathematics and computational methods applied to nuclear science and engineering*, 2176–2185.