

Multifidelity uncertainty quantification and model validation of large-scale multidisciplinary systems

Giuseppe Cataldo,^{a,*} Elizabeth Qian^b and Jeremy Auclair^c

^aNASA Goddard Space Flight Center, Greenbelt, Maryland, United States

^bCalifornia Institute of Technology, Pasadena, California, United States

^cCentre d'Etudes Spatiales de la Biosphère, Toulouse, France

Abstract. A simulation-based framework for multifidelity uncertainty quantification is presented, which informs and guides the design process of complex, large-scale, multidisciplinary systems throughout their life cycle. In this framework, uncertainty in system models is identified, characterized, and propagated in an integrated manner through the analysis cycles needed to quantify the effects of uncertainty on the quantities of interest. This is part of the process to design systems and verify their compliance to performance requirements. Uncertainty quantification is performed through mean and variance estimators as well as global sensitivity analyses. These computational analyses are made tractable by the use of multifidelity methods, which leverage a variety of low-fidelity models to obtain speed-ups, while keeping the main high-fidelity model in the loop to guarantee convergence to the correct result. This framework was applied to the James Webb Space Telescope observatory integrated model used to calculate the wavefront error caused by thermal distortions. The framework proved to reduce the time required to perform global sensitivity analyses from more than 2 months to less than 2 days, while reducing the error in the final estimates of the quantities of interest, including model uncertainty factors. These technical performance improvements are crucial to the optimization of project resources such as schedule and budget and ultimately mission success. © 2022 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JATIS.8.3.038001](https://doi.org/10.1117/1.JATIS.8.3.038001)]

Keywords: uncertainty quantification; multifidelity; global sensitivity analysis; Sobol' indices; rank statistics.

Paper 21166G received Dec. 15, 2021; accepted for publication Jun. 30, 2022; published online Jul. 22, 2022.

1 Introduction

A rigorous simulation-based framework for multifidelity uncertainty quantification (UQ) is presented which informs and guides the design, verification, and validation processes of complex, large-scale, multidisciplinary systems throughout their life cycle. The term multifidelity refers to the use of multiple mathematical models, of varying fidelity (accuracy) and cost, to perform a computational analysis of uncertainties in a system. In such an approach, inexpensive low-fidelity (LF) models are used for computational acceleration, while an expensive high-fidelity (HF) model is used to ensure accuracy of the results and retain the physics of the phenomena the model is trying to represent. This multifidelity approach enables rigorous UQ and provides insightful results from a complex system model faster and with a better understanding of the model inputs and outputs than the traditional way of relying only on an expensive HF model. For large-scale projects or programs, the latter approach is often a cause of schedule and cost overruns, driven by the large amount of time spent designing and analyzing systems with an inefficient model at the detriment of other important activities, such as integration and testing. As a case study, the successful application of this framework to the James Webb Space Telescope (JWST)¹ is described in detail and highlights the benefits that this framework brings in terms of technical performance improvements and optimization of project resources such as schedule and budget.

*Address all correspondence to Giuseppe Cataldo, Giuseppe.Cataldo@NASA.gov

The development of complex systems relies on mathematical models to explore the trade space in the early phases of the system life cycle, optimize the system design in the more advanced phases, and simulate system performance to verify its ultimate compliance to mission requirements. Because a mathematical model is an imperfect representation of reality, it is necessary to subject the model to a validation process to confirm that it is an adequate representation of the physical system and is capable of predicting the system's behavior accurately with respect to the requirements within the domain of the intended application of the model.² This is accomplished through hardware testing and reconciliation of the test results with the model predictions.

A model's ability to perfectly reproduce reality is affected by several factors, including unknown model parameters, inadequacy of the model to represent certain phenomena, the fact that it may not be practical to run the model to observe the output for every configuration of interest, etc.³ Model results are therefore inherently affected by different sources of uncertainty that need to be understood and included in the analysis cycles to inform and guide the decision-making process in the design and management of a project or program. As a result, understanding the uncertainty and the factors that drive it is important to all players in the decision chain, as it affects project resources.

The framework discussed here represents a paradigm shift from the current state of the art and the modus operandi in several government and industrial settings, where model uncertainty factors (MUFs) and/or safety factors, often based solely on engineering judgment, are usually applied downstream of the mathematical models to the worst-case model outputs (Ref. 4, pp. 89–92). There are two drawbacks to this approach. First, as defined in Ref. 4, MUFs are a semiquantitative adjustment, either additive or multiplicative or both, that is made to the results of a simulation-based analysis to account for uncertainty. Here, semiquantitative means based on past experience or engineering expertise rather than data. Second, designing a system for worst-case scenarios is equivalent to developing the system for a sequence of events that are each individually unlikely to occur in the system lifetime and even more unlikely to occur together. As a consequence, the use of MUFs to the outputs of worst-case scenario analyses leads to overly conservative predictions and overdesigned systems. While sometimes overdesign may be the more efficient way to solve a design problem, uncertainty in design, especially in a large-scale mission, such as JWST, is expensive.

The significance of the research results presented here is twofold. First, in this multifidelity UQ framework, uncertainty in system models is identified, categorized, modeled, and propagated in an integrated manner throughout the analysis cycles needed to design systems and verify their compliance to key performance requirements. Such an integrated approach naturally leads to an increased knowledge of the system that is being designed and to an in-depth understanding of the uncertainty's impact on system performance, robustness, reliability, and safety.⁵ Gaining this knowledge is of paramount importance when considering that 85% of a project's total life cycle cost is locked in by the end of preliminary design.⁶ Therefore, this framework becomes a crucial instrument especially if applied early in the project life cycle, when there is larger freedom in the design and decisions can have a stronger impact in subsequent phases.

Second, this framework has the ability to provide methods and tools to robustly and efficiently design complex, large-scale, multidisciplinary systems where several forms of uncertainties are present. This approach thus improves project performance by reducing the analysis time and freeing up resources for activities, such as integration and testing, that can have a major influence on the system's ability to meet the requirements.

This multifidelity UQ framework finds applications in several fields, such as groundwater modeling,⁷ ocean dynamics,⁸ and remote sensing,⁹ and is proving crucial to a number of space missions,¹⁰ as the inclusion of such critical information, needed to support resource utilization decisions and risk mitigation strategies, will maximize the probability of mission success.

The paper is organized as follows. Section 2 provides the background and motivation behind this work. Section 3 describes the UQ framework and Sec. 4 specifically addresses its multifidelity implementation. Section 5 illustrates in more detail how to carry out a global sensitivity analysis (GSA) study to rank model parameters and prioritize research efforts for those parameters with the greatest uncertainty to reduce their impact on the model outputs. In addition, this section describes the first formulation of multifidelity GSA (MFGSA) based on rank statistics, which yields the greatest computational benefit. The James Webb Space Telescope is used

throughout as an example of a complex, large-scale, multidisciplinary system in which the described framework was employed to quantify uncertainty by overcoming the challenges imposed by the large computational cost of the models used. Section 6 summarizes the main findings and describes ongoing work and future perspectives.

2 Background and Motivation

The quality and reliability of a system's model depend on the model's adequacy to describe the system and its real-life behavior.¹¹ Model parameter values are often known only under ideal conditions or in a few limited laboratory test cases,^{12–15} which either do not necessarily apply to the specific system or mission being investigated or lie outside the domain of the intended application of the model.² In addition, test settings and boundary conditions, as well as the hardware configuration, lead engineers to introduce effective parameters to account for geometric complexity [e.g., in a multilayer insulation (MLI) blanket] and physical phenomena caused by hardware interactions (e.g., aluminum screws bolted into a titanium rod or the presence of test sensors, instrumentation, etc.), which are difficult to model. The exact values of these effective parameters and their variances are unknown and have to be estimated by the engineers based on their expertise or judgment. Finally, while new complex systems are most impacted by uncertainties, even later system copies suffer from this issue.

Simulation-based design under uncertainty addresses such shortcomings by incorporating uncertainty upstream in the design process and propagating its effects through the model to obtain an output where one can identify the dominant effects contributing to the system's performance or robustness degradation.

Modeling, analyzing, and validating systems that are built to satisfy increasingly stringent requirements become more and more challenging as such systems grow in complexity and scale.¹⁶ This leads to massive, computationally heavy models, characterized by a large number of parameters, easily exceeding the thousands. A multifidelity approach¹⁷ can combine computationally intensive HF models with computationally efficient LF models in a principled way, making simulation-based design under uncertainty possible even for large-scale, multidisciplinary systems. Multifidelity approaches have been used to enable UQ and optimization for a variety of applications, including aircraft design,^{18,19} structural analysis and optimization,^{20,21} and trajectory simulation for entry, descent, and landing.²²

Lastly, models developed for different systems often reside in separate platforms that require manual transfer of data from one model to another when performing integrated, end-to-end analyses. An example is a thermal-structural-optical analysis where each discipline model is developed in a specific software environment and requires a number of interfaces to pass the outputs of one model as inputs to the next model in the chain of analysis to compute the final figure of merit. In these cases, it is imperative to automate several of the analysis and UQ tasks that involve running models on different platforms a large number of times.

The work presented in this paper shows how to combine all of these aspects in the design process to maximize use of the available computational resources within a prescribed budget.

3 Uncertainty Quantification

Uncertainty is usually defined as either aleatory or epistemic. Aleatory uncertainty is representative of unknowns that differ each time the same experiment is run and can be quantified through random sampling techniques such as Monte Carlo (MC) or quasi-MC methods.²³ For this reason, it is also called statistical uncertainty and the evaluation of uncertainty by the statistical analysis of series of observations is termed a type A evaluation.²⁴ Epistemic (systematic) uncertainty is due to things one could know in principle but does not in practice. For example, epistemic uncertainty can arise due to a measurement that is not accurate, a model that neglects certain effects, or some data that have been deliberately hidden. Epistemic uncertainty can be reduced by gaining a better knowledge of the system, process, or mechanism under consideration. The evaluation of uncertainty by means other than the statistical analysis of series of observations is termed a type B evaluation (of uncertainty).²⁴ Methods commonly used for

epistemic uncertainties include probability bounds analysis,²⁵ fuzzy logic,²⁶ or evidence theory (i.e., the Dempster–Shafer theory,^{27,28} which is a generalization of the Bayesian theory of subjective probability).

The multifidelity approach adopted in the framework presented here can be used to quantify the effects of both aleatory and epistemic uncertainties in the design cycle. For example, a finding that certain aleatory uncertainties contribute very little to the overall output uncertainty would allow these uncertainties to be neglected in future analyses, reducing the cost of these analyses. On the other hand, a finding that certain epistemic uncertainties contribute significantly to the variance of the output would indicate that more resources should be invested in reducing those uncertainties through, for example, further experiments. The analysis results presented in this work, however, focus on aleatory uncertainty, without loss of generality.

Figure 1 shows the general scheme adopted to quantify aleatory uncertainty in this framework. The study presented here does not account for model inadequacy, i.e., the uncertainty in the model itself due to missing or wrong physics and any other type of modeling error. The extension to consider model inadequacy based on the works in Refs. 3 and 29 is an area for future work.

In step 1, all possible sources of aleatory uncertainty are identified, such as, for example, material properties, geometry, orbital parameters, etc.

In step 2, these uncertainties are characterized probabilistically: each uncertain variable is assigned a probability distribution (defined by a probability density/mass function in the continuous/discrete cases, respectively). These probability distributions are estimated by engineers based on experimental data, literature reviews, and engineering expertise.

In step 3, the model input parameter uncertainties are propagated through the model itself to compute the model outputs, which will also be stochastic in nature. This is accomplished through MC-type simulations, whereby all the parameters are varied simultaneously, and their values are sampled from their respective probability density functions. Because MC simulations cannot be run for an infinite number of samples to observe the output for every configuration of interest, there will be some residual error that is quantified as sampling error.

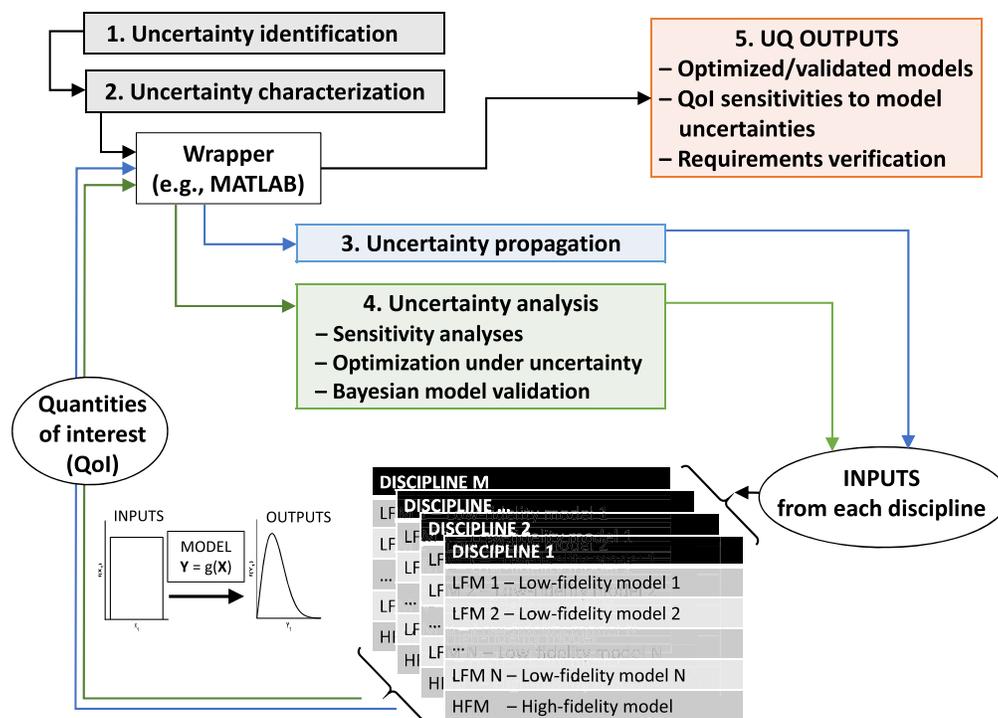


Fig. 1 UQ process. The sources of uncertainty are identified in step 1 and then characterized through probability distributions in step 2. By means of a wrapper, the uncertainty is propagated in step 3 through each discipline model to compute the Qoi. These are then used to perform uncertainty analysis in step 4. At the end of the process, in step 5, the UQ outputs are available for use.

In step 4, uncertainty analysis can be performed for all quantities of interest (QoIs) based on engineering needs. (Quantities of interest can also represent technical performance metrics. Here, the term “quantities of interest” is used to avoid loss of generality and reducing the scope of the methodologies presented in this work.) Some examples of uncertainty analysis are:

- Estimators: One could compute statistical properties such as the expected value (mean) and variance of the distributions of each model output, as well as MUFs.
- Sensitivity analysis: A GSA³⁰ can provide a list of parameters prioritized by their influence in terms of their contribution to the total variance of a specific QoI. This is important in the context of requirements verification and model validation. For example, cumulative distribution functions can be used to evaluate the probabilities that each QoI meets specific requirements within a specified level of confidence, as predicted by the model. If there is insufficient confidence that a requirement will be met, GSA can provide a list of parameters ranked by their contribution to the total variance on a specific QoI related to that particular requirement. Additional research could therefore be performed to reduce the uncertainty in these parameters and thus decrease their contribution to the model output uncertainty. This step is discussed in more detail in Sec. 5.
- Bayesian model validation: When experimental data are available, Bayes’ theorem can be used to update the model parameter prior distributions and thus increase confidence in the model.⁵
- Optimization under uncertainty: The results of step 3 can be used to optimize stochastic figures of merit that may be required for the design of a specific system, such as expectation and variance or a linear combination of the two, etc.³¹

Finally, step 5 gathers the outputs of the process in terms of optimized/validated models, QoI sensitivities to model uncertainties, and requirements verification results.

3.1 Application to JWST

During the course of its mission, JWST repeatedly changes its attitude as it slews from one science target to another. Because a typical observation lasts a few hours and the thermal time constant of JWST is on the order of days, there is insufficient time for JWST to reach thermal equilibrium before the next slew. As the thermal state changes during these maneuvers throughout the mission, the thermally induced deformations in the observatory structures also vary, causing perturbations in the wavefront error (WFE).³² In the work presented here, the hot-to-cold slew case is used as a benchmark to analyze the WFE of the telescope for an attitude change from +5 deg to −45 deg in pitch and from −5 deg to +5 deg in roll during nominal science operations. This change of state is referred to as HC23 (Fig. 2) and defines the allowed range of boresight pointing angles for the observatory relative to the Sun line, which must remain in the range 85 deg to 135 deg at all times to keep the telescope behind the sunshield. As a result of this hypothetical maneuver, the temperatures of several subsystems change, before stabilizing to their steady-state values. As an example, this work focuses on thermal distortions of the primary mirror (PM) backplane support structure, which are expected to dominate the WFE evolution.

The WFE is derived by first computing the temperature predictions, ΔT , through the JWST observatory thermal model. These predictions serve as input to the JWST observatory thermal distortion model, which provides the final WFE value. The thermal model is run in SINDA/G,³³ a finite-difference-based thermal software. Similar to the approach described in Ref. 34, the nominal ΔT values predicted by the thermal model are multiplied by some MUFs before mapping performed by the structural analysis team in preparation for input to the thermal distortion model, which is a finite-element model developed in NASTRAN[®] with millions of degrees of freedom.

For the thermal model, the subsystems of interest in this work are the telescope’s PM, primary mirror backplane assembly (PMBA), backplane support frame (BSF), integrated science instrument module (ISIM) bench, ISIM near-infrared spectrometer (NIRSpec) focal plane array (FPA) radiator, ISIM science instruments (SI) radiators and aft optics subsystem central baffle,

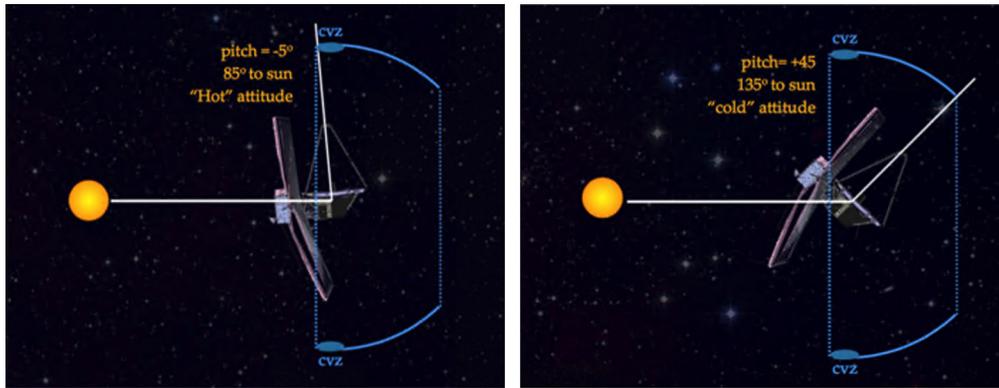


Fig. 2 The hot-to-cold slew maneuver consists of changing the telescope attitude from 85 deg to the Sun (hot) to 135 deg (cold), corresponding to a change in pitch angle from -5 deg to $+45$ deg. Because JWST operates in an ecliptic coordinate framework, two small continuous viewing zones can be seen, which are centered at each of the ecliptic poles. Their 5-deg radius is determined by the 85-deg solar exclusion zone, although any observation approaching the 85 deg limit will have additional limitations due to safety considerations.

for a total of $n = 7 \Delta T$ outputs. The thermal model input parameters identified for this study are materials' conductivities, emissivities, and effective emissivities (e.g., for an MLI blanket), for a total of $m = 194$ variables.

For the thermal distortion model, the output of interest is the WFE, whereas the input parameters identified for this study are the Young modulus (E), Poisson coefficient (ν), and coefficient of thermal expansion of different telescope structures, for a total of $m = 66$ variables.

Based on information gathered in the literature,^{12,13,15} experimental data collected over the years to build a JWST material property database, elicitation from subject matter experts, the parameters' probability density functions were found to be either uniform or normal. Wherever data were absent, uniform distributions were assumed, indicating a complete lack of knowledge regarding the distributions of those particular parameters, assuming equal probability for their values within acceptable bounds. All the uniform distributions used in this study have lower and upper bounds varying within $\pm 10\%$ and $\pm 5\%$ of the parameters' nominal values for the thermal and thermal distortion model, respectively. The normal distributions have a wider variability depending on the material.

The parametric uncertainties were propagated in the integrated model through MC simulations, whereby many samples were drawn from the parameter distributions to approximate the characteristics of the QoI distributions. The effects of radiative parameters such as materials' emissivities were accounted for after a model order reduction strategy was implemented, due to the high computational cost needed to vary each of these parameters in NASA's Thermal Synthesizer System (TSS)³⁵ and update the corresponding SINDA/G files. Each single TSS run required about 2 days and each of the 12 emissivity parameters was varied to its minimum and maximum values, for a total of 24 runs. This approach thus allowed defining an envelope for such parameters and their effects on the QoI distributions, implying that their contribution to the overall uncertainty could not be greater than this.

The QoIs are the mean and variance of the HC23 ΔT values for each of the subsystems mentioned above and the final WFE. In practice, estimators of these QoIs were computed with a finite number of samples, thus providing an approximation of the true mean and variance of the model outputs, as shown in Eq. (1). If the model is represented by the function $\mathbf{Y} = f(\mathbf{X})$, where $\mathbf{Y} = \{Y_1, X_Y, \dots, Y_n\}$ is a vector with n outputs and $\mathbf{X} = \{X_1, X_2, \dots, X_m\}$ is a vector with m input parameters, each with its own probability density function, the mean and variance estimators are, respectively, defined as follows:

$$\bar{\mathbf{Y}} = \frac{1}{N} \sum_{k=1}^N f(\mathbf{X}_k) \quad \text{and} \quad \sigma^2 = \frac{1}{N-1} \sum_{k=1}^N [f(\mathbf{X}_k) - \bar{\mathbf{Y}}]^2, \quad (1)$$

where \mathbf{X}_k is the k 'th independent realization of the input \mathbf{X} and N is the number of realizations (or samples). Here, \mathbf{Y} represents either the vector of ΔT s from the thermal model or the WFE from the thermal distortion model. The vector σ^2 is the vector of variances σ_j^2 for each of the j outputs.

The MC standard error, e_{MC} , relates the number of samples N with the confidence level P in the results: $e_{MC} = \sqrt{P(1-P)/N}$. In other words, to achieve a 1% error at the 1σ level ($P = 68\%$), $N = 2000$ samples are necessary. For the thermal model, a number of samples $N = 1500$ was adopted to achieve an $e_{MC} \approx 1.2\%$ at the 1σ level (i.e., 68% confidence level), while keeping the computational cost of running the JWST observatory thermal model within the prescribed budget. The computational budget was limited by a combination of factors, including simulation run time, number of available SINDA/G licenses, and project schedule requirements. The computation of the samples needed in Eq. (1) was parallelized on two Intel® Xeon® CPU E5- 2620 v2 @ 2.10-GHz processors and six cores, given the availability of six SINDA/G licenses. Each model run took ~ 4.5 h with a starting guess from a previously converged simulation.

For the thermal distortion model, a number of samples $N = 460$ was adopted ($e_{MC} \approx 2\%$ at the 1σ level), with each model run requiring ~ 55 min on average. The computation of the samples was performed on NASA's 129056-core "Discover" supercomputing cluster, an assembly of multiple Linux scalable units built upon commodity components capable of nearly 6.8 petaflops and parallelized by simultaneously running up to five simulations based on the number of available NASTRAN® license seats. Table 1 shows the mean and variance estimators for all of the outputs.

For each QoI, a MUF is defined as follows:

$$\text{MUF} = \frac{|\bar{Y}| + 2\sigma}{|Y_{\text{nom}}|}, \quad (2)$$

where Y_{nom} is the model output Y under nominal conditions, and \bar{Y} and σ are the MC estimators of the average Y and variance calculated according to Eq. (1) with the same samples previously drawn from the parameters' probability density functions for each QoI.

Figure 3 shows the MUFs for each thermal subsystem. The orange squares are associated with MUFs determined through a root-sum-square (RSS) approach as follows:

$$\text{MUF}_{\text{RSS}} = \frac{|Y_{\text{nom}}| + \sqrt{\sum_{i=1}^m (\Delta T_{\text{max},i})^2}}{|Y_{\text{nom}}|}. \quad (3)$$

Table 1 Estimators of the HC23 ΔT and WFE mean values and variances for the JWST subsystems of interest. In the MC simulations, a number of samples $N = 1500$ was employed for the thermal model and $N = 460$ for the thermal distortion model.

QoI	Mean	Variance
ΔT_{PM}	-0.1037 K	$1.4042 \times 10^{-6} \text{ K}^2$
ΔT_{PMBA}	-0.0951 K	$1.0858 \times 10^{-6} \text{ K}^2$
ΔT_{BSF}	-0.0782 K	$0.3430 \times 10^{-6} \text{ K}^2$
$\Delta T_{\text{ISIM Bench}}$	-0.0401 K	$0.4013 \times 10^{-6} \text{ K}^2$
$\Delta T_{\text{ISIM FPA Radiator}}$	-0.0011 K	$0.0731 \times 10^{-6} \text{ K}^2$
$\Delta T_{\text{ISIM SI Radiator}}$	-0.0077 K	$0.0744 \times 10^{-6} \text{ K}^2$
$\Delta T_{\text{Central Baffle}}$	-0.1916 K	$0.5607 \times 10^{-6} \text{ K}^2$
WFE	54.1599 nm	$1.6773 \times 10^{-4} \text{ nm}^2$

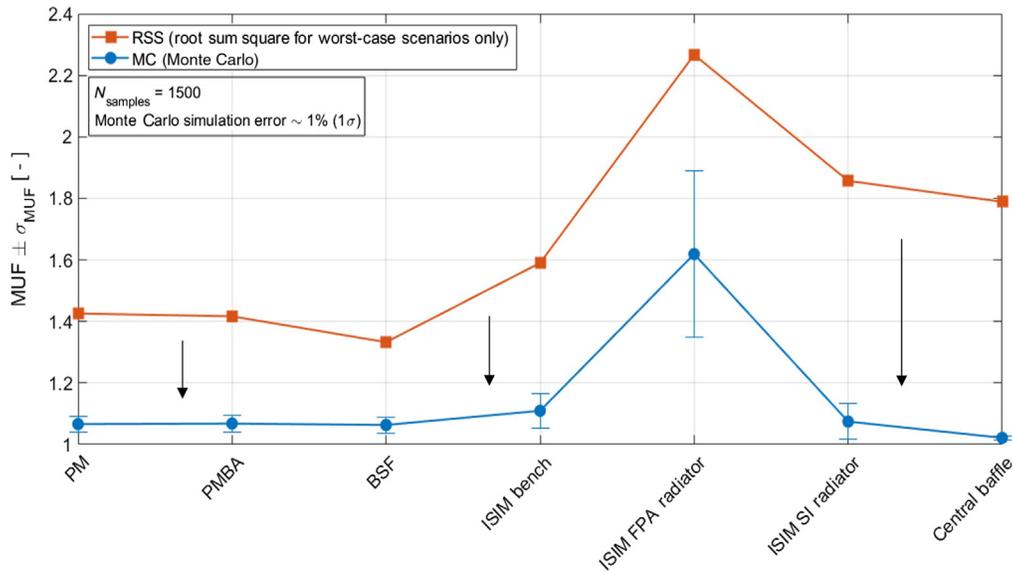


Fig. 3 MUFs and their standard deviations for the seven JWST thermal subsystems investigated. The blue circles identify the MUFs computed through the MC approach in Eq. (2), while the orange squares those computed through the RSS approach in Eq. (3). The RSS MUFs are larger than the MC MUFs because they account for the largest ΔT s resulting from changing each parameter to the minimum and maximum values of its distribution, thereby creating a worst-case-scenario envelope.

Here, each parameter, one at a time, is changed to the minimum and maximum values of its distribution (In the case of a normal distribution, the values at $\pm 6\sigma$ were considered as maximum and minimum, respectively, which extend to nearly the totality of the data) and the largest ΔT resulting from either change, ΔT_{\max} , is taken into the RSS. The RSS approach assumes that all the parameters are uncorrelated and does not account for any parameter interaction effects. In addition, because all the model parameters are pushed to their extreme values, the resulting MUFs can be seen as an envelope of worst-case scenarios. The blue circles, instead, represent the MUFs calculated through Eq. (2) with the probability density functions identified in this study. The nature of the MC approach also allows calculating the standard deviation of each MUF, thus providing a quantitative indication of their confidence or coverage range, which is not possible with an RSS approach. This example shows how easy it can be to overdesign systems when making worst-case scenario assumptions, as done with an RSS approach, and to underestimate available margins. Quantification of uncertainty eliminates these drawbacks by providing results based on data and by avoiding the improper use of project resources, including schedule and cost, caused by overly conservative assumptions and excessive reliance on subjective engineering judgment.

4 Multifidelity Uncertainty Quantification

In steps 3 and 4 of the UQ process in Fig. 1, the estimators of statistical properties obtained using a highly accurate, computationally expensive model are called HF estimators. The need to draw a large number of samples from this computationally expensive model can make such an analysis computationally prohibitive. The UQ analysis discussed in Sec. 3.1 took about 47 days to complete despite maximization of all the available computational resources. Orders-of-magnitude decreases in computational time can be achieved by the use of multifidelity Monte Carlo (MFMC)³⁶ estimators, which are explained below.

The MFMC method achieves computational speed-ups by allocating some portion of the available computational resources to cheap, LF model samples, while retaining some of the available computational resources for expensive, HF model samples. The result is a multifidelity estimator of the QoI that is unbiased with respect to the HF estimator: for a fixed cost,

the multifidelity estimator has lower variance than the HF estimator (more certainty in the multifidelity estimator); for a fixed estimator quality, the multifidelity estimator is cheaper to compute than the HF estimator.

Denote by f_1 the HF model, with cost (usually given by model run time) w_1 . Denote by f_2, f_3, \dots, f_l a set of $l - 1$ LF models, with costs w_2, w_3, \dots, w_l , respectively. LF models map the same inputs to the same quantity of interest as the HF model but sacrifice some amount of accuracy to gain speed. Some examples of LF models include projection-based reduced models,³⁷⁻⁴⁰ data-fit interpolation, and regression models,^{41,42} including machine-learning-based methods, such as support vector machines,^{43,44} and other simplified models^{45,46} that rely on early-stopping criteria, coarse-grid approximations, or simplified physics assumptions. The multifidelity estimator for the QoI mean is given as follows for two models (the HF model f_1 and one LF model f_2):

$$\bar{\mathbf{Y}}_{\text{MF}} = \frac{1}{N_1} \sum_{i=1}^{N_1} f_1(\mathbf{X}_i) + \alpha_2 \left(\frac{1}{N_2} \sum_{i=1}^{N_2} f_2(\mathbf{X}_i) - \frac{1}{N_1} \sum_{i=1}^{N_1} f_2(\mathbf{X}_i) \right), \quad (4)$$

where α_2 is scaling constant³⁶ that stretches f_2 to the same scale as f_1 , N_1 is the number of samples drawn from the expensive model f_1 , and N_2 is the number of samples drawn from the cheaper LF model N_2 , and $N_2 > N_1$. By re-arranging Eq. (4) as follows:

$$\bar{\mathbf{Y}}_{\text{MF}} = \alpha_2 \frac{1}{N_2} \sum_{i=1}^{N_2} f_2(\mathbf{X}_i) + \left(\frac{1}{N_1} \sum_{i=1}^{N_1} f_1(\mathbf{X}_i) - \alpha_2 \frac{1}{N_1} \sum_{i=1}^{N_1} f_2(\mathbf{X}_i) \right), \quad (5)$$

it can be seen that the multifidelity estimator can be viewed as a high sample estimate of the LF mean, plus an additive correction based on a few samples of the HF model. While here we include the formula for only two models (the HF model and one LF model), the extension to multiple LF models is straightforward, involving adding the term $\alpha_l \left(\frac{1}{N_l} \sum_{i=1}^{N_l} f_l(\mathbf{X}_i) - \frac{1}{N_1} \sum_{i=1}^{N_1} f_l(\mathbf{X}_i) \right)$ to Eq. (4) for each additional LF model $l = 3, 4$, etc.

The multifidelity estimator is unbiased, i.e., the expected value of the multifidelity mean estimator $\bar{\mathbf{Y}}_{\text{MF}}$ in Eq. (4) is equal to the expected value of the HF mean estimator $\bar{\mathbf{Y}}$ in Eq. (1), $\mathbb{E}[\bar{\mathbf{Y}}_{\text{MF}}] = \mathbb{E}[\bar{\mathbf{Y}}]$.³⁶ Remarkably, this unbiasedness property holds even when the LF model(s) is (are) biased, due to the terms in parentheses in Eq. (4) canceling when the expectation of the whole equation is taken. The multifidelity estimator is also more certain than the HF estimator with the same cost, i.e., the variance of the multifidelity mean estimator is less than the variance of the HF mean estimator, $\text{Var}[\bar{\mathbf{Y}}_{\text{MF}}] < \text{Var}[\bar{\mathbf{Y}}]$, when certain conditions are met.³⁶ roughly speaking, these conditions require that the loss in fidelity of the LF model be compensated for by a sufficient reduction in computational cost.

The work in Ref. 36 provides a model selection algorithm that ensures that the set of models used in the multifidelity estimate satisfy the conditions for the multifidelity estimator to outperform the HF estimator. This allows selection from a rich diversity of models of varying approximation quality and cost, based on different aspects of the system physics. Reference 36 also provides an algorithm to optimally allocate a fixed computational budget (usually given in CPU time) among the different models, that is, the algorithm outputs the number of samples N_1, N_2 , etc. that should be given to each model, which yield the multifidelity estimator with lowest variance. These algorithms are used in the subsequent numerical results for model selection and the distribution of samples among the different models available.

The MFMC approach described above for estimating the QoI mean can be extended to estimate QoI variances and/or other statistics by replacing each of the mean estimators in Eq. (4) with the appropriate estimator for the variance or other statistic of interest.⁴⁷ This approach is used in Sec. 5.1 for multifidelity estimation of sensitivity indices.

4.1 Application to JWST

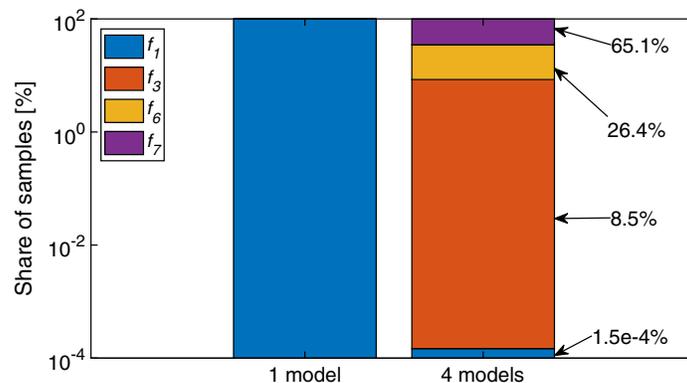
Table 2 shows the LF models created from the JWST HF observatory thermal model, f_1 , how well they correlate with it, and their run time.

Table 2 List of models employed in the JWST MFMC analysis with their correlation coefficients and run times. The correlation coefficients are a measure of how well each LF model correlates with the HF model.

	Model	Correlation coefficient, ρ	Run time, t (min)
f_1	HF	1	270
f_2	Early-stop criteria	0.0103	45
f_3	Polynomial interpolation (first degree)	0.3826	0.00008
f_4	Polynomial interpolation (second degree)	0.9999	0.00663
f_5	Polynomial interpolation (third degree)	0.9999	0.39511
f_6	Low-rank canonical approximations	0.7606	0.00055
f_7	Kriging (Gaussian process)	0.8814	0.00165
f_8	Support vector machine for regression	0.0635	0.00020

Seven different LF models were established as follows. Model f_2 was created by setting the convergence criteria of the HF model to less stringent values to decrease the total run time from 4.5 h to 45 min, while agreeing with the HF model predictions within 5%. Models f_3 , f_4 , and f_5 were developed by interpolation between sampled points of f_1 of first-, second-, and third-order degree polynomials, respectively. Three other models were derived with UQLab,⁴⁸ an open-source MATLAB toolbox for UQ. Model f_6 is a canonical low-rank approximation (LRA),⁴⁹⁻⁵¹ which UQLab identifies by performing a threefold cross-validation with 1500 training points and 1500 validation points from the HF model to select a model with rank 3 and Legendre polynomial bases of degree 2. Model f_7 is a Kriging approximation with constant trend and anisotropic ellipsoidal Matérn 5/2 correlation function.⁵²⁻⁵⁴ The hyperparameters of the model were determined through a maximum likelihood estimation method, which employed a hybrid genetic optimization algorithm⁵⁵ and 1200 training points. Model f_8 is a regression-based model with support vector machines (SVR),^{43,44} which uses a linear ϵ -insensitive loss function with an isotropic Matérn 5/2 kernel. The hyperparameters of the model were determined through a hybrid covariance matrix adaptation-evolution strategy (HCMAES) optimization algorithm⁵⁶ to minimize the span estimate of the leave-one-out (LOO) error with 300 training points.

For the JWST observatory thermal model, the MFMC model selection algorithm³⁶ selected three of the LF models shown in Table 2, automatically eliminating those that do not meet the constraint on the ratios of the costs and correlation coefficients as described earlier in Sec. 4 and Ref. 36. The share of samples for each model is shown in Fig. 4. Most of the samples are

**Fig. 4** Sample sharing. More than half of the samples is allocated to f_3 , which is the fastest among the models used, despite its low correlation with f_1 . Its lower accuracy is compensated by f_6 and f_7 , which have a slightly longer run time but higher correlation with f_1 . This is kept in the loop enough to increase the accuracy of the results without worsening the computational burden.

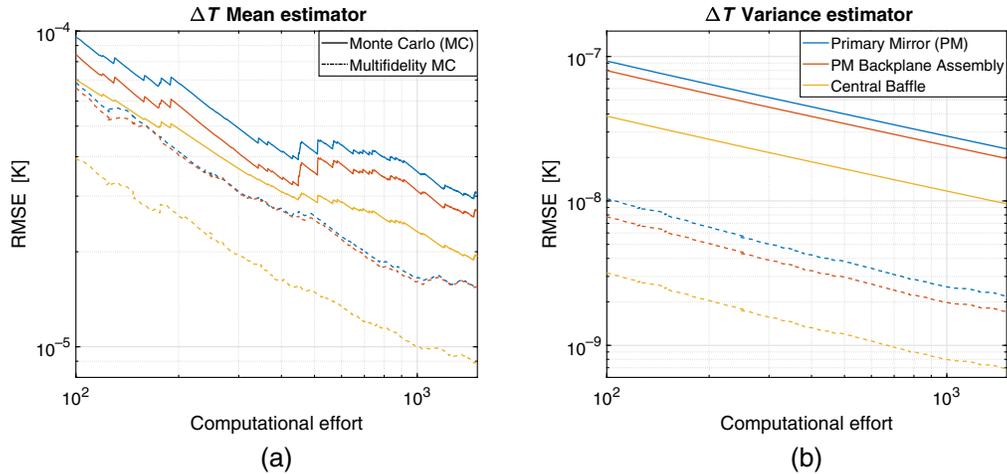


Fig. 5 (a) Mean and (b) variance estimators for the ΔT s of three of the seven JWST thermal subsystems analyzed. The colors identify each system. The solid lines refer to the MC approach using only the HF model, the dashed lines refer to the MFMC approach.

allocated to f_3 , which is the fastest among all models, despite its low correlation with f_1 . This is compensated by f_6 and f_7 , which are characterized by a slightly longer run time but higher correlation. f_1 is kept in the loop enough to increase the accuracy of the results without worsening the computational burden.

In Fig. 5, the root mean square error (RMSE) of the MFMC approach is compared to that of a regular MC UQ analysis to show how, in general, for a given RMSE, multifidelity estimates converge in less time and, for a given computational budget, achieve a lower RMSE.

5 Multifidelity Global Sensitivity Analysis

GSA seeks to quantify the sensitivity of a model's QoI to its individual inputs. In particular, variance-based GSA^{57–59} divides the total variance of the model QoI into percentages that can be attributed to the j 'th input alone or the interaction between the i 'th and j 'th inputs, etc.⁶⁰ Two types of variance-based sensitivity indices are of particular interest to engineers: the Sobol' main index, $S_{m,i}$, and the Sobol' total index, $S_{t,i}$.⁵⁷ Both Sobol' indices lie in the interval $[0,1]$. The main index measures the fraction of the total QoI variance that can be attributed to the i 'th input alone, whereas the total index measures the fraction of the total QoI variance that can be attributed to the i 'th input as well as the interactions of the i 'th input with other inputs. Inputs with larger sensitivity indices contribute more to QoI uncertainty. Computing Sobol' indices for all inputs thus allows inputs to be ranked in terms of importance, enabling engineers to drop unimportant inputs for simplified analyses or to prioritize further characterization of important inputs to reduce the overall QoI uncertainty.

Let $D = \text{Var}(Y)$ denote the total variance of a model output. For vector outputs, the GSA procedure is simply repeated for each element of the vector output. The Sobol' main effect sensitivity index $S_{m,i}$ of the i 'th input parameter is defined as follows:

$$S_{m,i} = \frac{D_i}{D} = \frac{\text{Var}(\mathbb{E}[Y|X_i])}{\text{Var}(Y)}, \quad i = 1, \dots, m. \quad (6)$$

The numerator in Eq. (6) is the variance over the distribution of X_i of the conditional mean of the output Y conditioned on the input X_i . $S_{m,i}$ measures the contribution of input X_i to the output variance, without accounting for any interactions with other input parameters. For some intuition, note that if $f(\mathbf{X}) = f(X_i)$, i.e., if X_i is the only input that contributes to the output, then $\mathbb{E}[Y|X_i] = f(X_i) = f(\mathbf{X})$, so in this case $D_i = D$ and the Sobol' main index would be $S_{m,i} = 1$, indicating 100% of the output variance is attributable to input X_i . On the other hand, if X_i does not have any influence on Y , then $\mathbb{E}[Y|X_i] = \mathbb{E}[Y]$, so $D_i = \text{Var}(\mathbb{E}[Y]) = 0$.

The Sobol' total effect sensitivity index of factor X_i is defined as follows:

$$S_{t,i} = 1 - \frac{\text{Var}(\mathbb{E}[Y|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m])}{\text{Var}(Y)}, \quad i = 1, \dots, m. \quad (7)$$

The numerator in Eq. (7) is the variance of the conditional mean, where the conditioning is on all inputs except the i 'th input. $S_{t,i}$ measures the total contribution of input X_i on the output vector's variance, accounting for the influence of X_i alone as well as its interactions with the other input parameters.

Sobol' indices are usually computed via MC sampling methods using what are known as "fixing methods" or the "pick-freeze" approach. These terms refer to the way that the conditional expectations in Eqs. (6) and (7) are estimated. Such methods generate pairs of input samples by first generating one set of N input samples, and then for each input sample, the input component(s) on which the expectation is conditioned is "fixed" or "frozen" and all other input components are resampled to generate a second associated sample. The products of the outputs associated with these pairs of inputs are used to estimate the variances of the conditional means in Eqs. (6) and (7). Several different pick-freeze estimators for these indices have been proposed in the literature,^{58–63} each of which requires $\mathcal{O}(mN)$ evaluations of the model f to estimate indices for m different model inputs, where the number of sample pairs N required depends on the desired confidence level in the estimator. Recall, for example, that to achieve an MC standard error $e_{\text{MC}} \approx 1\%$ at the 1σ level, $N = 2000$ samples are necessary. With a large number of parameters, m , and a computationally expensive model, the cost of performing GSA via computing Sobol' indices becomes prohibitive. This is a challenge in the analysis of any complex system.

The MFMC approach described in Sec. 4 is used in Ref. 47 to accelerate the estimation of Sobol' sensitivity indices by replacing the mean estimators in Eq. (4) with the pick-freeze estimators of Sobol' sensitivity indices. This MFGSA approach is fundamentally different from some other approaches in the literature which also use the term "multifidelity,"^{64,65} where Sobol' indices are computed for an LF model whose accuracy is increased using HF model evaluations. In such approaches, the resulting estimators are biased, whereas the MFGSA approach of Ref. 47 yields conditional expectation estimators which are unbiased with respect to the HF model (regardless of bias introduced by LF models!). The next section describes how this MFGSA approach is used for a GSA analysis of the JWST observatory thermal model.

5.1 Application to JWST

A GSA analysis previously performed on the JWST observatory thermal model is described in Ref. 66. Here, the results were shown for the Sobol' main effect sensitivity indices of $m = 5$ parameters (out of 194) through regular MC integration.⁶⁷ To achieve an $e_{\text{MC}} \approx 1\%$ at the 1σ level, $N = 2000$ samples were required for a total of $(m + 1) \times N = 12,000$ simulations. Using only LF model f_2 , which warrants a 5% agreement with HF model f_1 in terms of nominal values and runs in about 45 min (as opposed to the 4.5 h required by f_1), and parallelizing the simulations on 6 cores, 12 days of computational time per Sobol' index were needed, for a total of ~ 60 days. While this analysis showed which of the five parameters of interest needed to be addressed by further research, it also showed the limiting factors of an MC-based GSA in terms of quality of results and run time when dealing with expensive models.

The MFGSA approach of Ref. 47 overcomes these limitations by leveraging the smaller cost of LF models that are kept in the loop with the HF model for accuracy. An MFGSA was performed on the JWST HF observatory thermal distortion model, g_1 , which takes the temperature values output by the thermal model discussed in Sec. 3.1 to perform a thermal distortion analysis, with the WFE being the final QoI.

Four different LF models were created as follows. A linearized model g_2 was built by evaluating the first derivatives of the model output with respect to each input and expanding the model output in a Taylor series truncated at the first order. This is analogous to performing a local sensitivity analysis. Three other models were derived with UQLab. Model g_3 is a canonical LRA, which UQLab identifies by performing a threefold cross-validation with 100 training points and 360 validation points from g_1 to select a model with rank 1 and Hermite and Legendre

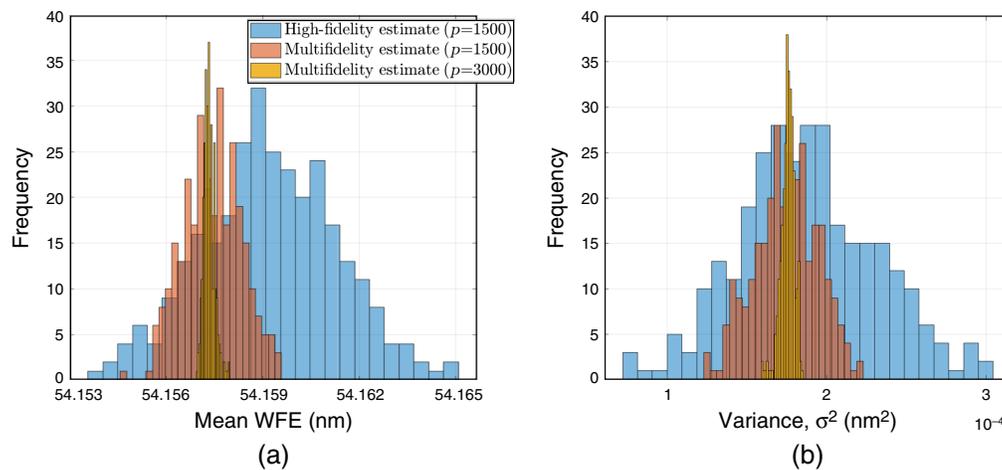
Table 3 Correlation coefficients, run time, and number of samples per model for a computational budget of $p = 1500$.

	Model	Correlation coefficient, ρ	Run time, t (min)	GSA, N_k	MFGSA, N_k
g_1	HF	1	110	12	12
g_2	Linearized HF	0.8874	4.21×10^{-4}	—	3422
g_3	LRAs	0.8552	6.10×10^{-6}	—	96,661
g_4	Kriging	0.9924	6.23×10^{-6}	—	—
g_5	Support vector machine for regression	0.8899	1.41×10^{-6}	—	—

polynomial bases of degree 1. Model g_4 is a Kriging approximation with linear trend and anisotropic ellipsoidal Matérn 5/2 correlation function. The hyperparameters of the model were determined through a covariance matrix adaptation-evolution strategy optimization algorithm⁵⁶ to perform a 70-fold cross-validation. Model g_5 is a regression-based SVR model which uses a linear ε -insensitive loss function with anisotropic ellipsoidal Gaussian kernel. The hyperparameters of the model were determined through a Broyden–Fletcher–Goldfarb–Shanno optimization algorithm^{68–71} to minimize the span estimate of the LOO error.

Table 3 shows the LF models created and selected by the MFGSA algorithm, their correlation to the HF model, their run time, and the number of samples that were allocated to each model as determined by the algorithm with a budget $p = 1500$. For the HF MC estimators, only 12 HF samples are affordable with this computational budget. For the MFGSA estimators, the leftover budget (not enough for another HF sample) is used to generate 3422 samples from LF model g_2 and 96,661 samples from LF model g_3 . For this particular combination of budget, model costs, and model fidelities, the model allocation algorithm uses the leftover budget after 12 HF evaluations for the very low-cost, LF evaluations, so both the multifidelity and single-fidelity MC approach use the same number of HF evaluations. However, this is not always the case: in other circumstances it is common to trade off a few HF evaluations in exchange for more LF evaluations.

In Fig. 6, the mean and variance estimates of the WFE are shown for three different cases, highlighting the benefits of using a multifidelity approach. In blue is a histogram of the mean and variance as determined by an MC simulation with only the HF model with the prescribed budget of $p = 1500$. The histogram in orange represents the multifidelity estimates for the same budget, and the histogram in yellow represents the multifidelity estimates with an increased budget ($p = 3000$).

**Fig. 6** Comparison of (a) the WFE mean and (b) variance estimators for UQ performed with only the HF model (blue), and the HF model and two LF models (orange) for the same budget ($p = 1500$). In yellow are the same multifidelity estimates with an increased budget ($p = 3000$).

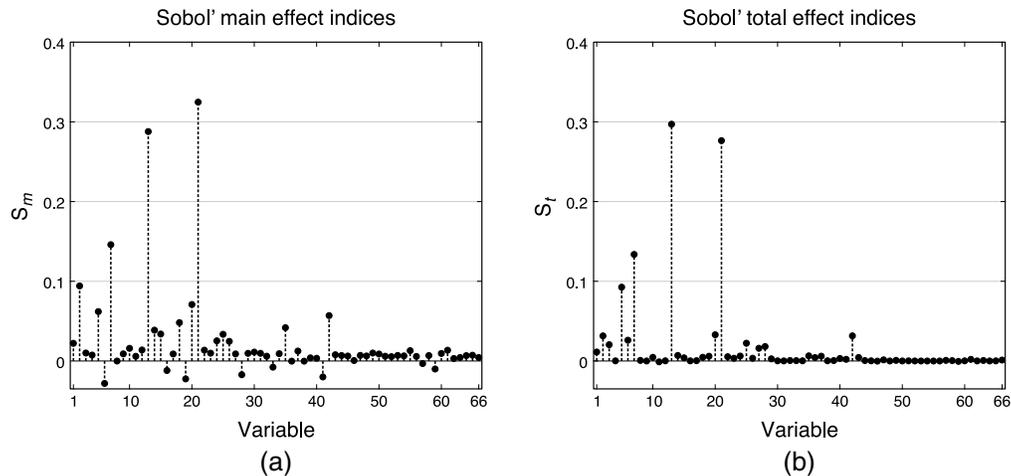


Fig. 7 Sobol' indices: (a) main effect and (b) total effect, as determined by MFGSA with a budget $p = 1500$.

indicating a reduction in the spread of both estimators. The histogram in yellow shows an additional improvement when the budget is increased to $p = 3000$ with a number of samples equal to 24, 6230, and 194,733 for the same three models used above.

The Sobol' main and total effect sensitivity indices for $p = 1500$ are shown in Fig. 7. Both sets of indices point to parameters #13 and #21 as the ones carrying about 1/3 of the variance contribution to the output, followed by #7 and #5 with about 1/10. The main effect indices appear to have some residual variability, which would be reduced with a higher number of simulations to improve convergence. Such residual is much less noticeable in the total indices. Despite this, other parameters that appear to contribute to the WFE variance are #20 and #42 at a <5% level.

In terms of overall performance, MFGSA brought several improvements. First, it enabled computing both the main and total effect sensitivity indices for all the 66 parameters of interest within the prescribed budget, without having to leave any parameters out, as was the case for the JWST observatory thermal model discussed in Sec. 3.1. Second, accuracy guarantees on the final results were retained through the HF model, thus removing that additional source of uncertainty that was introduced in the JWST observatory thermal model when performing GSA with only one LF model.

The main drawback was related to the ability to draw HF samples from the NASTRAN[®] model, which could only be run on a supercomputer without a direct interface with the MFGSA software in MATLAB. This part of the process could not be automated and required saving the HF samples offline. Based on the values shown in Table 3, the minimum number of HF simulations needed was $(m + 2) \times N = (66 + 2) \times 12 = 816$. Because the NASTRAN[®] model could only simulate one condition at a time, the HC23 change of state of the telescope (see Sec. 3.1) required two simulations per sample, bringing the total number to 1632 and an equivalent computational time of about 62 days. This number is close to the 60 days needed to perform the GSA described in Sec. 3.1 for only five parameters, which represents a major improvement. Nonetheless, collecting samples from an expensive HF model continues to represent the main hurdle to making even a tool such as MFGSA practical, especially when this type of information is needed over shorter timescales.

5.2 Rank Statistics-Based MFGSA

This section describes a powerful new multifidelity estimator for global sensitivity indices that unites the multifidelity framework with a recently proposed Sobol' main effect index estimator based on rank statistics. The new approach requires $\mathcal{O}(N)$ model evaluations and is more efficient than the pick-freeze approach described in Sec. 5.1, which requires $\mathcal{O}(mN)$ evaluations. While the pick-freeze methods have long been the state of the art, they have two major

disadvantages. First, they rely on a particular experimental design (fixing some input components and resampling others) that may be unavailable in practice. Second, the number of model calls to estimate all first-order Sobol' indices grows linearly with the number of input parameters m , as discussed in previous sections.

These two limitations were recently addressed and eliminated by the work of Gamboa et al.,⁷² which is able to embed Chatterjee's method⁷³ in the GSA framework, resulting in a dramatic reduction of the number of samples needed to estimate Sobol' indices of any order. Chatterjee's method consists of defining a coefficient of correlation between parameters based on rank statistics. In statistics, ranking is the data transformation in which numerical or ordinal values are replaced by their rank when the data are sorted. For example, the following observed numerical data: 5.3, 1.8, 9.2, 6.4, would be ranked in ascending order as 2, 1, 4, 3, respectively. Despite being simple in nature, like other classical coefficients (e.g., Pearson's, Spearman's or Kendall's correlation), Chatterjee's coefficient is able to estimate the degree of dependence between the parameters of interest while retaining a simple asymptotic theory under the assumption of independence. Gamboa generalizes Chatterjee's method, requiring only one set of N samples to compute Sobol' indices for all m inputs.

In this framework, called rank statistics-based GSA (RSB GSA), the Sobol' main effect sensitivity index of variable X_i for output Y is defined as follows:

$$S_{m,i}^G = \frac{\frac{1}{N} \sum_{k=1}^N Y_k Y_{R_i(k)} - \bar{Y}^2}{\text{Var}(Y)}, \quad i = 1, \dots, m, \quad (8)$$

where \bar{Y} is the sample mean of Y , $\text{Var}(Y)$ is the sample variance of Y , and $R_i(k)$ defines a reordering (permutation) of the output Y samples based on sorting the i 'th input components X_i in ascending order, as follows: let $\phi_i(k)$ denote the rank of $X_{i,k}$ in the sample $\{X_{i,1}, X_{i,2}, \dots, X_{i,N}\}$. That is, $\kappa = \phi_i(k)$ denotes the place of $X_{i,k}$ in the set after rearranging the inputs $\{X_{i,k'}\}_{k'=1}^N$ in ascending order. Let $k = \phi_i^{-1}(\kappa)$ denote the inverse of this operation, i.e., $\phi_i^{-1}(\kappa)$ gives us the original index of the κ 'th smallest X_i in the original ordering of the set. Then, $R_i(k)$ is defined as follows:

$$R_i(k) = \begin{cases} \phi_i^{-1}(\phi_i(k) + 1) & \text{if } \phi_i(k) + 1 \leq N, \\ \phi_i^{-1}(1) & \text{if } \phi_i(k) = N. \end{cases} \quad (9)$$

From Eq. (8) it can be seen that a total of N model runs is required to compute all Sobol' main effect sensitivity indices, since the same Y samples are reordered according to Eq. (9) for every parameter X_i . The number of runs, therefore, no longer depends on the number of parameters, m . This represents a considerable reduction from the GSA/MFGSA budgets, which grow linearly with m . Equation (8), however, usually requires N to be very large (of the order of 10^4 – 10^5) to achieve convergence.⁷² Reducing the number of runs to N may therefore still not be sufficient to making this approach useful when dealing with expensive models. Algorithm 1 combines the RSB global sensitivity estimators with the MFGSA approach of Ref. 47 in a new approach that enables GSA of multidisciplinary, large-scale systems at even further reduced cost. This new approach is called RSB MFGSA.

The results from Algorithm 1 are shown in Fig. 8. In gray are the indices obtained with an RSB GSA performed with HF model g_1 and all the 90 available samples, which were not enough to achieve convergence. In black are the indices obtained with the RSB MFGSA, which are similar to those in Fig. 7(a) but were obtained with a remarkably lower effort. With a budget $p = 47$, the number of samples (shown at the top right-hand side of Fig. 8) required by the RSB MFGSA approach is similar to the number of samples required by MFGSA with a budget $p = 3000$ (see Sec. 5.1 and Fig. 6).

Table 4 summarizes all the computational methodologies developed over the years to evaluate Sobol' indices and a comparison of the required computational budgets for the JWST case study. This comparison is based on the observation that the RSB MFGSA approach of Fig. 8 with a budget of $p = 47$ leads to index estimates that are qualitatively similar to the estimates of the MFGSA of Fig. 7 with a budget of $p = 1500$. The MFGSA approach is therefore approximately 42× more expensive than the RSB MFGSA approach. To estimate the budget

Algorithm 1 Rank statistics-based multifidelity global sensitivity analysis (RSB MFGSA).

-
- 1: Define budget ρ .
 - 2: Compute statistics for all $l = 1, \dots, L$ models: mean, variance, correlation, etc.
 - 3: Get optimal number of evaluations m_l and weights α_l for each model given ρ .
 - 4: Draw $N_L = \max\{N_l\}$ realizations $X_{i,1}, \dots, X_{i,N_L}$ from every input variable.
 - 5: Compute all N_1 evaluations of the HF model for every output of interest, Y .
 - 6: For every input variable, apply Gamboa's method to the HF model:
 - Sort $X_{i,1}, \dots, X_{i,N_1}$ in ascending order and get ordering index vector ϕ .
 - Sort ϕ in ascending order and get ordering index vector $\phi(k)$.
 - Compute the argument of ϕ^{-1} as in Eq. (9).
 - Compute $R(k)$ by reordering ϕ with the indices determined in the previous step for every variable.
 - Get output corresponding to the new rank, $Y_{R_i(k)}$.
 - Initialize S_m^G as in Eq. (8).
 - 7: **for** $l = 2:L$ **do**
 - 8: Compute all N_l evaluations of each LF model for every output of interest.
 - 9: Repeat Gamboa's method (i.e., step 6) with both N_l and N_{l-1} samples.
 - 10: Estimate corrections as in Eq. (8) with $N = N_l$ and $N = N_{l-1}$.
 - 11: Define $\Delta S_m^G = S_m^G(N = m_l) - S_m^G(N = m_{l-1})$.
 - 12: Add correction to existing estimate: $S_m^G \leftarrow S_m^G + \alpha_l \Delta S_m^G$.
 - 13: **end for**
-

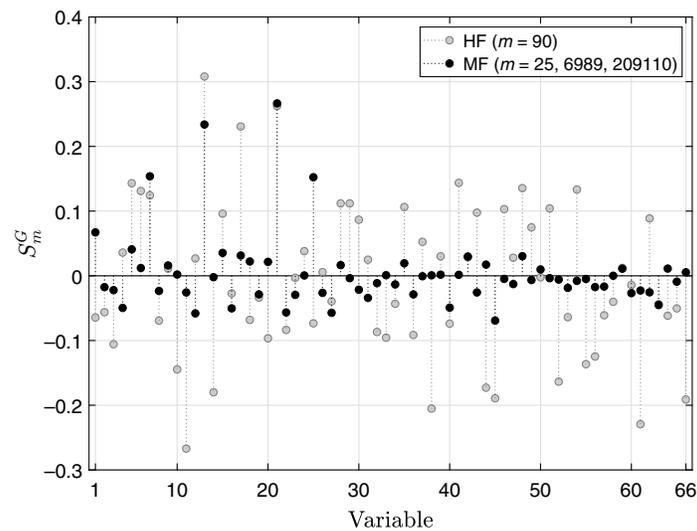


Fig. 8 Sobol' main effect sensitivity indices as derived from the Gamboa approach, i.e., Eq. (8), with all the 90 HF model samples available (gray) and from the multifidelity version developed in this work with the addition of two LF models and only 25 HF model samples (black).

Table 4 Review of Sobol' indices computation methodologies and comparison of computational budgets required to obtain Sobol' index estimates of similar quality to the multifidelity estimates in Figs. 7 and 8. Note that the computational budget is given in CPU hours. The analyses may take less "real-world" time if, for example, some model samples can be run in parallel, or more "real-world" time if, for example, the user must wait in a queue for a software license to become available.

Method	Authors	Budget (CPU hours)
GSA	Sobol', ⁵⁷ Cannavò ⁶⁷	$p = 6000$
MFGSA	Qian et al. ⁴⁷	$p = 1500$
RSB GSA	Gamboa et al. ⁷²	$p = 188$
RSB MFGSA	Cataldo et al.—this paper	$p = 47$

required for the HF -only estimators to converge to a similar rate, we note that the variances of the HF -only variance estimators are approximately $4\times$ larger than the variances of the multifidelity variance estimators for both the GSA and the RSB GSA approaches, respectively. Since MC variance scales inversely with the number of samples, we estimate that the HF -only versions of the estimators will require a $4\times$ larger budget to achieve similar levels of convergence to the multifidelity estimates.

Note that these budgets are for the JWST case study considered in this work, and that the specific budgets required for GSA will depend on the problem at hand, the method used, and the desired level of convergence. However, the comparison in Table 4 demonstrates that for this complex multidisciplinary system with 66 input parameters, the required budget decreases due to the introduction of both multifidelity and RSB approaches, making the combination of the two a powerful tool for the computation of Sobol' indices of expensive models and a large number of parameters. Note also that these budgets are what is required for a single set of Sobol' index estimates: in practical settings, the user may wish to repeat the estimates multiple times to understand the randomness in the estimates.

6 Conclusions

A framework to quantify the effects of parametric uncertainty in mathematical models that are used to design complex, large-scale, multidisciplinary systems was presented along with simulation procedures that address the difficulties of dealing with computationally expensive models and a number of parameters exceeding 60. Multifidelity techniques have been shown to reduce the execution time from months to days by leveraging different types of LF models to accelerate the UQ tasks (requiring expensive MC simulations), while keeping the HF model in the loop to maintain accuracy guarantees on the final result.

This framework was successfully applied to the James Webb Space Telescope's observatory integrated model, whose thermal module outputs temperature values that become the input to the structural module, which in turn provides the WFE of the PM caused by thermal distortions. The uncertainties of both modules' input parameters were identified and characterized in terms of probability density functions and propagated through the model. Uncertainty analysis was then performed by computing mean and variance estimators and MUFs for seven telescope subsystems. Global sensitivity analyses were carried out to identify the parameters that could cause the greatest reduction in the model output variance by targeting them with experiments aimed at reducing their variance.

The computational cost of all these tasks, which required exploring the design space with a number of samples sufficient to achieve at least an MC standard error $\leq 1\%$ at the 1σ level, was shown to be reduced by adopting multifidelity techniques. These were applied to the two above-mentioned models, which were characterized by 194 and 66 input parameters and an average run

time of ~ 4.5 and 1 h, respectively. When coupled with rank statistics, multifidelity techniques made global sensitivity analyses efficient and practical to use in the context of designing multidisciplinary, large-scale systems by reducing the number of runs of the HF model by orders of magnitude from a few thousand to 10 to 25. This reduced the time required for such an analysis from over 2 months to only 2 days. In addition, for a given computational budget, multifidelity techniques enabled achieving lower errors on the final estimates.

The approach described in this work allowed identifying the main drivers for subsequent research activities aimed at reducing their variance and better understanding their impact on system performance. More in general, this approach has demonstrated the ability to guide and inform the decision-making process with an increased knowledge of the system under consideration. This is most crucial in the early phases of a project life cycle, before most of the cost is locked in. However, it is applicable throughout with benefits spanning from a more effective use of computational resources, faster way of obtaining critical knowledge on the system's main drivers, the ability to keep the project on schedule and its budget on track, and the ability to invest more in activities such as integration and testing that can uncover problems before launching the system in its operational environment.

Future work attempts to include model inadequacy as a way to quantify the uncertainty of the model itself and enable a Bayesian-based model validation that is able to account for the impacts of model inadequacy on the system's response.

Acknowledgments

The material was based upon work supported by the National Aeronautics and Space Administration (NASA) under Award No. 80GSFC21M0002. The authors gratefully acknowledge the support provided by Enzo Kim in the development of the rank statistics-based GSA code, Christopher P. May in running the required TSS simulations, and Gary E. Mosier, Malcolm B. Nieder, and S. Harvey Moseley for their technical expertise and advice. GC also thanks the University of Maryland, Baltimore County for administering his appointment at the NASA Goddard Space Flight Center. The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

Code, Data, and Materials Availability

MATLAB code that implements the MFUQ method of Ref. 36 is available in a Github repository at <https://github.com/pehersto/mfmc>, whereas the one implementing the MFGSA strategy of Ref. 47 and the RSB MFGSA approach described in this work can be found at <https://github.com/elizqian/mfgsa>.

References

1. J. P. Gardner et al., "The James Webb Space Telescope," *Space Sci. Rev.* **123**(4), 485–606 (2006).
2. S. Blattnig et al., "Towards a credibility assessment of models and simulations," in *49th AIAA/ASME/ASCE/AHS/ASC Struct., Struct. Dyn., and Mater. Conf., 16th AIAA/ASME/AHS Adapt. Struct. Conf., 10th AIAA Non-Deterministic Approaches Conf., 9th AIAA Gossamer Spacecraft Forum, 4th AIAA Multidisciplinary Design Optimization Specialists Conf.*, p. 2156 (2008).
3. M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *J. R. Stat. Soc. Series B Stat. Methodol.* **63**(3), 425–464 (2001).
4. "Rules for the design, development, verification, and operation of flight systems," GSFC-STD-1000G, NASA Goddard Space Flight Center, Greenbelt, Maryland (2016).
5. NASA, "Standard for models and simulations," NASA-STD-7009A (2016).
6. K. Stout, "Bayesian-based simulation model validation for spacecraft thermal systems," PhD thesis, Massachusetts Institute of Technology (2015).

7. J. Roskam, *Airplane Design Part VIII: Airplane Cost Estimation Design Development and Manufacturing and Operating*, Darcorporation (1990).
8. D. Higdon, H. Lee, and C. Holloman, "Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems," in *Bayesian Statistics*, J. M. Bernardo et al., Eds., Vol. 7, pp. 181–197, Oxford University Press, Oxford, United Kingdom (2003).
9. I. W. McKeague et al., "Statistical inversion of South Atlantic circulation in an abyssal neutral density layer," *J. Mar. Res.* **63**, 683–704 (2005).
10. H. Haario et al., "Markov chain Monte Carlo methods for high dimensional inversion in remote sensing," *J. R. Stat. Soc. Series B Stat. Methodol.* **66**, 591–608 (2004).
11. NASA, "NASA Technology Roadmap," (2015).
12. J. Brynjarsdóttir and A. Kennedy, "Learning about physical parameters: the importance of model discrepancy," *Inverse Probl.* **30**(11), 114007 (2014).
13. E. R. Canavan and J. G. Tuttle, "Thermal conductivity and specific heat measurements of candidate structural materials for the JWST optical bench," *AIP Conf. Proc.* **824**(1), 233–240 (2006).
14. J. G. Tuttle et al., "Thermal properties of double-aluminized Kapton at low temperatures," *AIP Conf. Proc.* **986**(1), 34–41 (2008).
15. J. G. Tuttle and E. R. Canavan, "Recent NASA/GSFC cryogenic measurements of the total hemispheric emissivity of black surface preparations," *IOP Conf. Ser. Mater. Sci. Eng.* **102**(1), 012015 (2015).
16. D. Gilmore, Ed., *Spacecraft Thermal Control Handbook, 2nd ed., Vol. 2: Cryogenics*, The Aerospace Press, El Segundo, California (2002).
17. G. Cataldo et al., "Model-based thermal system design optimization for the James Webb Space Telescope," *J. Astron. Telesc. Instrum. Syst.* **3**(4), 044002 (2017).
18. B. Peherstorfer, K. E. Willcox, and M. Gunzburger, "Survey of multifidelity methods in uncertainty propagation, inference, and optimization," *SIAM Rev.* **60**(3), 550–591 (2018).
19. D. Allaire and K. Willcox, "A mathematical and computational framework for multifidelity design and analysis with computer models," *Int. J. Uncertain. Quantif.* **4**(1), 1–20 (2014).
20. A. Chaudhuri et al., "Multifidelity optimization under uncertainty for a tailless aircraft," in *AIAA Non-Deterministic Approaches Conf.*, p. 1658 (2018).
21. A. March, K. Willcox, and Q. Wang, "Gradient-based multifidelity optimisation for aircraft design using Bayesian model calibration," *Aeronaut. J.* **115**(1174), 729–738 (2011).
22. H. Xu and Z. Liu, "Control variate multifidelity estimators for the variance and sensitivity analysis of mesostructure–structure systems," *ASCE-ASME J. Risk Uncertain. Eng. Syst. B: Mech. Eng.* **5**(2), 020907 (2019).
23. J. Warner et al., "Multi-model Monte Carlo estimators for trajectory simulation," in *AIAA SciTech 2021 Forum*, p. 0761 (2021).
24. R. E. Caflisch, "Monte Carlo and quasi-Monte Carlo methods," *Acta Numer.* **7**, 1–49 (1998).
25. B. N. Taylor and C. E. Kuyatt, "Guidelines for evaluating and expressing the uncertainty of NIST measurement results," NIST Technical Note 1297, 1994 Edition (1994).
26. W. T. Tucson and S. Ferguson, "Probability bounds analysis in environmental risk assessments," Applied Biomathematics Technical Report (2003).
27. L. Zadeh, "Fuzzy sets," *Inf. Control* **8**(3), 338–353 (1965).
28. A. P. Dempster, "A generalization of Bayesian inference," *J. R. Stat. Soc. Series B (Methodol.)* **30**(2), 205–247 (1968).
29. G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press (1976).
30. Y. Ling, J. Mullins, and S. Mahadevan, "Selection of model discrepancy priors in Bayesian calibration," *J. Comp. Phys.* **276**, 665–680 (2014).
31. A. Saltelli, "Sensitivity analysis for importance assessment," *Risk Anal.* **22**(3), 579–590 (2002).
32. N. V. Sahinidis, "Optimization under uncertainty: state-of-the-art and opportunities," *Comput. Chem. Eng.* **28**(6), 971–983 (2004).
33. J. Gersh-Range and M. D. Perrin, "Improving active space telescope wavefront control using predictive thermal modeling," *J. Astron. Telesc. Instrum.* **1**(1), 1–18 (2014).
34. MSC Software Team, "SINDA 2013 user's guide," Technical Report, MSC Software (2013).

35. T. D. Panczak et al., "Thermal synthesizer system: an integrated approach to spacecraft thermal analysis," SAE Technical Paper, SAE International, pp. 1–19 (1991).
36. B. Peherstorfer, K. Willcox, and M. Gunzburger, "Optimal model management for multifidelity Monte Carlo estimation," *SIAM J. Sci. Comput.* **38**(5), A3163–A3194 (2016).
37. L. Sirovich, "Turbulence and the dynamics of coherent structures. III. Dynamics and scaling," *Q. Appl. Math.* **45**(3), 583–590 (1987).
38. G. Rozza, D. B. P. Huynh, and A. T. Patera, "Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations," *Arch. Comput. Methods Eng.* **15**(3), 229–275 (2008).
39. S. Gugercin, A. C. Antoulas, and C. Beattie, "H₂ model reduction for large-scale linear dynamical systems," *SIAM J. Matrix Anal. Appl.* **30**(2), 609–638 (2008).
40. P. Benner, S. Gugercin, and K. Willcox, "A survey of projection-based model reduction methods for parametric dynamical systems," *SIAM Rev.* **57**(4), 483–531 (2015).
41. A. I. Forrester, A. Sobester, and A. J. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*, John Wiley & Sons (2008).
42. A. I. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Prog. Aerosp. Sci.* **45**(1–3), 50–79 (2009).
43. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**, 273–297 (1995).
44. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011).
45. A. J. Majda and B. Gershgorin, "Quantifying uncertainty in climate change science through empirical information theory," *Proc. Natl. Acad. Sci. U. S. A.* **107**(34), 14958–14963 (2010).
46. L. W. T. Ng and K. E. Willcox, "Monte Carlo information-reuse approach to aircraft conceptual design optimization under uncertainty," *J. Aircr.* **53**(2), 427–438 (2016).
47. E. Qian et al., "Multifidelity Monte Carlo estimation of variance and sensitivity indices," *SIAM/ASA J. Uncertain. Quantif.* **6**(2), 683–706 (2018).
48. S. Marelli and B. Sudret, *UQLab: A Framework for Uncertainty Quantification in Matlab*, pp. 2554–2563, American Society of Civil Engineers (2014).
49. I. Markovsky and K. Uusevich, *Low Rank Approximation*, Vol. **139**, Springer (2012).
50. M. Chevreuril et al., "A least-squares method for sparse low rank approximation of multivariate functions," *SIAM/ASA J. Uncertain. Quantif.* **3**(1), 897–921 (2015).
51. K. Konakli and B. Sudret, "Polynomial meta-models with canonical low-rank approximations: numerical insights and comparison to sparse polynomial chaos expansions," *J. Comput. Phys.* **321**, 1144–1169 (2016).
52. T. J. Santner et al., *The Design and Analysis of Computer Experiments*, Vol. 1, Springer (2003).
53. C. Rasmussen and C. Williams, "Gaussian processes for machine learning (Internet ed.). Adaptive computation and machine learning," (2006).
54. V. Dubourg, "Adaptive surrogate models for reliability analysis and reliability-based design optimization," PhD thesis, Université Blaise Pascal-Clermont-Ferrand II (2011).
55. D. E. Goldberg, "Genetic algorithms in search, optimization, and machine learning," (1989).
56. N. Hansen, "The CMA evolution strategy: a tutorial," arXiv:1604.00772 X (2016).
57. I. M. Sobol, "Sensitivity estimates for nonlinear mathematical models," *Math. Model. Comput. Exp.* **1**(4), 407–414 (1993).
58. T. Homma and A. Saltelli, "Importance measures in global sensitivity analysis of nonlinear models," *Reliab. Eng. Syst. Saf.* **52**(1), 1–17 (1996).
59. A. B. Owen, "Variance components and generalized Sobol' indices," *SIAM/ASA J. Uncertain. Quantif.* **1**(1), 19–41 (2013).
60. A. Saltelli et al., *Global Sensitivity Analysis: The Primer*, John Wiley & Sons (2008).
61. A. Saltelli and J. Marivoet, "Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques," *Reliab. Eng. Syst. Saf.* **28**(2), 229–253 (1990).
62. A. Saltelli, "Making best use of model evaluations to compute sensitivity indices," *Comput. Phys. Commun.* **145**(2), 280–297 (2002).
63. A. Janon et al., "Asymptotic normality and efficiency of two Sobol' index estimators," *ESAIM: Probab. Stat.* **18**, 342–364 (2014).

64. L. Le Gratiet, C. Cannamela, and B. Iooss, "A Bayesian approach for global sensitivity analysis of (multifidelity) computer codes," *SIAM/ASA J. Uncertain. Quantif.* **2**(1), 336–363 (2014).
65. P. S. Palar et al., "Global sensitivity analysis via multi-fidelity polynomial chaos expansion," *Reliab. Eng. Syst. Saf.* **170**, 175–190 (2018).
66. G. Cataldo, "Integrated modeling under uncertainty for the James Webb Space Telescope," *Proc. SPIE* **10705**, 269–278 (2018).
67. F. Cannavò, "Sensitivity analysis for volcanic source modeling quality assessment and model selection," *Comput. Geosci.* **44**, 52–59 (2012).
68. C. G. Broyden, "The convergence of a class of double-rank minimization algorithms 1. General considerations," *IMA J. Appl. Math.* **6**, 76–90 (1970).
69. R. Fletcher, "A new approach to variable metric algorithms," *Comput. J.* **13**, 317–322 (1970).
70. D. Goldfarb, "A family of variable-metric methods derived by variational means," *Math. Comput.* **24**(109), 23–26 (1970).
71. D. F. Shanno, "Conditioning of quasi-newton methods for function minimization," *Math. Comput.* **24**(111), 647–656 (1970).
72. F. Gamboa et al., "Global sensitivity analysis: a new generation of mighty estimators based on rank statistics," *Bernoulli* hal-02474902v4 (2021).
73. S. Chatterjee, "A new coefficient of correlation," *J. Am. Stat. Assoc.* **116**(536), 1–21 (2020).

Giuseppe Cataldo is the Head of Backward Planetary Protection and the Containment Assurance Modeling & UQ Lead for the Mars Sample Return (MSR) Capture, Containment and Return System (CCRS). He leads research efforts in multidisciplinary design optimization and multifidelity uncertainty quantification for complex, large-scale systems. He received his PhD in aeronautics and astronautics from Massachusetts Institute of Technology (MIT) and his BS and MS degrees in aerospace engineering from the Polytechnic Institutes of Milan/Turin and ISAE-SUPAERO. He is a member of SPIE.

Elizabeth Qian is a von Karman instructor in the Department of Computing + Mathematical Sciences at Caltech. Her research focuses on making computational methods used in engineering decision-making efficient and scalable. She is interested in model reduction and scientific machine learning for physical systems, and in multifidelity formulations for uncertainty quantification and optimization. She received her SB and SM degrees in aerospace engineering and her PhD in aerospace computational engineering from MIT.

Jeremy Auclair is an Earth observation engineer working in Toulouse, France. He received his degree from ISAE-SUPAERO in 2020 and started working on Earth observation sciences in 2021. He studies the impact of different agricultural practices on land surface albedo and carbon stock and its contribution to climate change through satellite imagery. He worked as a research assistant at NASA Goddard in 2019 during a 6-month internship.