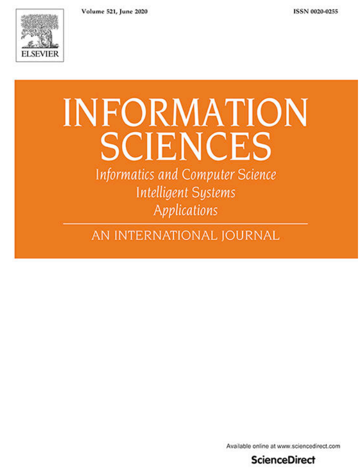
Journal Pre-proofs



A Performance Based Method for Information Acquisition in Engineering Design under Multi-Parameter Uncertainty

Zhenjun Ming, Anand Balu Nellippallil, Guoxin Wang, Yan Yan, Janet K. Allen, Farrokh Mistree

PII: S0020-0255(20)30944-0
DOI: https://doi.org/10.1016/j.ins.2020.09.034
Reference: INS 15873

To appear in: *Information Sciences*

Received Date: 2 January 2019
Revised Date: 4 August 2020
Accepted Date: 17 September 2020

Please cite this article as: Z. Ming, A. Balu Nellippallil, G. Wang, Y. Yan, J.K. Allen, F. Mistree, A Performance Based Method for Information Acquisition in Engineering Design under Multi-Parameter Uncertainty, *Information Sciences* (2020), doi: https://doi.org/10.1016/j.ins.2020.09.034

A Performance Based Method for Information Acquisition in Engineering Design under Multi-Parameter Uncertainty

Zhenjun Ming
Beijing Institute of Technology, Beijing, P.R. China
No. 5 Zhongguancun South Street, Haidian District, Beijing, China 100081
zhenjun.ming@bit.edu.cn

Anand Balu Nellippallil
Center for Advanced Vehicular Systems, Mississippi State University,
200 Research Blvd.
Starkville, MS 39759
Email: anand.nellippallil@msstate.edu

Guoxin Wang[1]
Beijing Institute of Technology, Beijing, P.R. China
No. 5 Zhongguancun South Street, Haidian District, Beijing, China 100081
wangguoxin@bit.edu.cn

Yan Yan
Beijing Institute of Technology, Beijing, P.R. China
No. 5 Zhongguancun South Street, Haidian District, Beijing, China 100081
yanyan331@bit.edu.cn

Janet K. Allen
The Systems Realization Laboratory
The University of Oklahoma, Norman, Oklahoma USA
Room 116-G, 202 W. Boyd Street, Norman, Oklahoma 73019
janet.allen@ou.edu

Farrokh Mistree
The Systems Realization Laboratory
The University of Oklahoma, Norman, Oklahoma USA
Felgar Hall, Room 306, 865 Asp Avenue, Norman, Oklahoma 73019
farrokh.mistree@ou.edu

[1] Corresponding Author: Vice Director of the Institute for Industrial Engineering and Associate Professor, School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China.

Abstract

Uncertainty pertaining to multiple parameters is a critical issue in designing complex systems. Whether or not to acquire more information to reduce uncertainty, and how to acquire information are the meta-level decisions to be made. Key challenges in making such decisions are that there are multiple information sources to choose from, and the cost of information as well as its effects on the overall design utility are different. To address these challenges, a performance-based stepwise information acquisition method is proposed. In the proposed method, the utility-based compromise Decision Support Problem construct is used to formulate design decisions to maximize the overall utility. For meta-level decisions, a performance index is developed for selecting the most appropriate information in each acquisition trial. The index is an integration of the improvement potential of the overall utility, the sensitivity of each ranged parameter, and the cost of the acquired information. Advantages of this proposed method are: 1) sensitivity-efficiency ensures that acquired information is invested on the critical parameters which avoids ineffective information acquisition; 2) cost-efficiency ensures that every acquisition is cost-efficient which avoids budget overruns. The efficacy of this method is demonstrated using the design of a hot rod rolling process. It is shown in the results that the performance-based method leads to an 8-45% larger drop of improvement potential compared to the random method.

Keywords: Uncertainty; Multi-parameter; Information Acquisition; Sensitivity; Value of Information

1. Frame of Reference

Designers often confront a variety of uncertainties in the design of systems and products [1-3]. In order to arrive at a better decision, designers collect information to reduce uncertainty associated with making a decision. However, gathering more information (e.g., by physical experiments or computer simulations) inevitably results in an increase in cost which could be in terms of human time, computational time, or delay. Accordingly, we suggest that engineers are confronted with resolving the following dilemma: make a decision under the current uncertainty or expending resources to reduce that uncertainty for a potentially better expected outcome. The trade-off between reduced uncertainty and the increased cost is a meta-level (or process-level) decision [4] that designers have to make in addition to decisions on the artifact or system that is being designed.

Simon [5] argues that humans are "Administrative man" making decisions under imperfect knowledge about nature and limited processing capability, instead of the theoretical "Economic man" who does not have such limitations. The idea of decision making under imperfect knowledge and searching for "satisficing" (good enough) solutions instead of optimal solutions is adopted in References [6, 7]. Howard [8] proposes the value-of-information construct for determining whether to get additional information for making a decision. As defined by Howard, value-of-information is the difference between the expected value of the objective for the alternative selected with the benefit of the information and that without. Value-of-information has been extensively studied in theoretical research fields including decision theory [9] and game theory [10].

In engineering design, various authors have written about the use of value-of-information construct in information acquisition for decision making under uncertainty. For example, Agogino and coauthors [11-13] use the expected value-of-information (EVI) for catalog selection problems wherein designers need to choose components from a catalog under significant uncertainty. Panchal and coauthors [4] propose an index called "improvement potential" to measure the value of information in simulation model refinement, and in the increasing of decision model fidelity for complex system design [14]. Based on the "improvement potential", Messer and coauthors [15] propose the "process performance indicator" for model selection under limited information in the context of integrated product and material design. In summary, existing value-of-information based approaches provide useful means for designers to evaluate both the benefit and the cost of additional information, so as to facilitate designers making decisions on whether to go ahead and make a decision based on the information at hand or delay the decision and expend further effort developing a better understanding of the problem. However, authors typically focus on a singular parameter under uncertainty and seldom touch upon multi-parametric uncertainty. For instance, the "improvement potential" index and the associated stepwise simulation model refinement method [4] work well for problems with a single uncertain parameter, but there is a lack of clarification or justification of how this method can guide the information acquisition process among multiple uncertain parameters. In the EVI based approach [11-13], even though it is mentioned that a vector of parameters are uncertain, but the process of how information is queried to reduce the uncertainty in multiple parameters is not specified in the catalog selection process.

In design of engineering systems, especially complex systems, designers usually face a situation where multiple parameters are concurrently uncertain, and they need to make meta-level decisions on information acquisition to reduce the uncertainty [4, 16]. Compared to a singular uncertain parameter, the challenge of information acquisition for multiple uncertain parameter is mainly embodied in two aspects: 1) multiple information sources are available for designers to choose from, and 2) the cost of information as well as its impact on the overall performance of design are different. Because resources are often limited during design, there is a need to acquire information strategically to reduce uncertainty of the parameters base on the value of the information, to make better meta-level decisions to improve the design. In order to address the need, we propose a performance-based approach for information acquisition accounting for design under multi-parametric uncertainty.

From the perspective of information sciences, the significance of this paper is anchored in that we recognize the importance of meta-level decisions in multi-parametric uncertainty problems. Key difference in meta-level decisions is that decisions makers not only need to focus on the formulation of the problem itself (given the uncertainty), but also need to decide whether or not to collect more information and what is the "best" way to collect information. Our contribution is that we provide an information acquisition method to guide decision makers to make these meta-level decisions. Using the method, decision makers can measure the performance of the information before it is acquired, considering the sensitivity of the uncertain parameters of the problem and the cost of the information to be acquired. This enables decision makers wisely spend the limited budget on the most valuable information.

The paper is organized as follows. Related literature and background are provided in Section 2. The method proposed for information acquisition accounting for design under multi-parametric uncertainty is presented in Section 3. A hot rod rolling process design example is presented in Sections 4 to demonstrate the use of the method. A generalized framework of the information acquisition method proposed in this paper is presented in Section 5. Closing thoughts are presented in Section 6.

2. Background and Foundation

2.1 Uncertainty in Engineering Design

Characterization of uncertainty in the modeling of physical systems is a critical topic in engineering design and analysis literature, many classifications and definitions have been proposed. For example, Isukapalli and coauthors [17] classify uncertainties as follows: a) "natural uncertainty or variability" which stands for inherent randomness or unpredictability of the physical system; b) "model uncertainty" which refers to approximations and simplifications in model formulation, and c) "data uncertainty" denoting incomplete knowledge of model parameters or inputs. Allen and coauthors [18] identify three types of uncertainties from a system function perspective: I) uncertainty in noise or environmental and other factors, II) uncertainty in design variables or control factors, and III) uncertainty introduced by modeling methods. Different uncertainty classifications reflect the difference in the views or computational representations of the problem. Zhai and coauthors [1] classify design uncertainties into randomness and fuzziness, and define the hybrid of these two types as twofold uncertainty. From a standpoint of whether the uncertainty can be reduced or not, uncertainty can be further categorized into "reducible" and "irreducible" [17, 19]. Epistemic uncertainty is usually a type of reducible uncertainty that can be diminished by improvements in measurements and/or model formulation and/or by increasing the accuracy or sample size of data. Aleatory uncertainty, on the other hand, is a type of irreducible uncertainty which is inherit in the physical system and can only be quantified in a statistical sense and cannot be reduced by gathering more information. Panchal and coauthors [4] point out that aleatory (irreducible) uncertainty is usually represented using probability distributions, while epistemic (reducible) uncertainty is generally modeled using intervals [20] or fuzzy sets [21].

Our purpose in this paper is to present an information acquisition method for reducing uncertainty in the expectation of making better design decisions, therefore we focus on the reducible uncertainty that can be reduced by acquiring more information. In order to visually present the uncertainty sources and identify the reducible uncertainty, the classification scheme proposed by Allen and coauthors [18] is used, see Figure 1. As shown in Figure 1, uncertainties may come from noise factors Y (type I), control factors X (type II), and system models $f(X,Y)$ (type III). Control factors are design variables which represent the specification of the system after the system is realized or implemented. At the design stage, because the system is yet to be built, a designer can hardly take action to reduce the variability in design variables. Therefore, uncertainty associated with the control factors is a type of irreducible uncertainty. Noise factors are uncontrollable parameters, which are usually given before design is performed and are used as constants during a design process. Uncertainty associated with noise factors can be categorized into variability and imprecision. Some noise factors are random parameters (Y_a in Figure 1) and represented using probability distributions, of which, as mentioned earlier, the uncertainty is not reducible. Some noise factors are ranged parameters (Y_e in Figure 1) and are represented using intervals, of which the uncertainty is reducible and thus is the focus of this paper, as shown in the small dashed rectangle in Figure 1. System model $f(X,Y)$ represents the mathematical relationship between the inputs X,Y and the outputs (response) Z of the system. Uncertainty associated with the system model is reducible by building models with higher fidelity. However, it is another research issue and out of the scope of this paper.

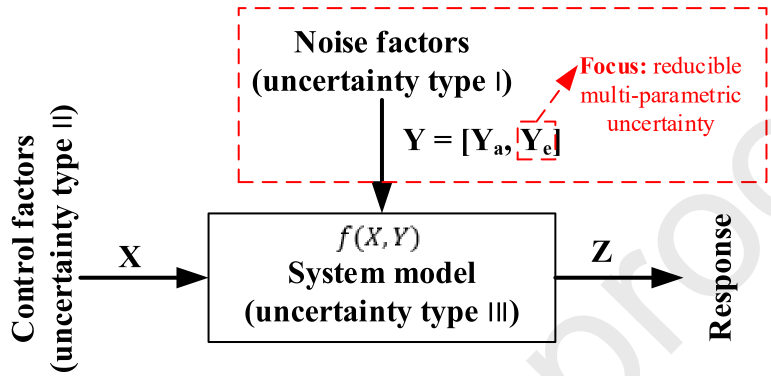


Figure 1. Three types of uncertainty sources [18] and the focus of this paper

2.2 Value-of-Information for Reducing Uncertainty

Lawrence [22] argues that the value-of-information for decision making can be measured at different stages in a decision making process: i) prior to consideration of incorporation of information; ii) after considering an information source but prior to receiving the information (*ex-ante* value); iii) after receiving additional information and making the decision, but before realization of the environmental state (*conditional* value); or iv) after addition of information and observing the outcome of the decision based on acquired information (*ex-post* value). In the engineering context, *ex-post* value can truly reflect the value-of-information for a decision based on the actual behavior of the system because the system is realized and the behavior can be measured exactly. However, during the design stage it is impossible for a designer to obtain the *ex-post* value because the system is yet to be built. What is critical for a designer is to predict the value of information before the information is received and the state occurs. Hence, we use the *ex-ante* value-of-information that captures the expected value by considering uncertainties in the system. The notion of *ex-ante* value-of-information for design under uncertainty is well recognized in literature [4, 12, 14, 15, 23]. Mathematically, the *ex-ante* value-of-information is calculated as:

$$v(x,y) = E_{x|y}\pi(x,a_y) - E_x\pi(x,a_0) \tag{1}$$

where $\pi(x,a)$ represents the payoff achieved by selecting an option a when the environment state is x. E_x $\pi(x)$ is the expected value of $\pi(x)$ and $E_{x|y}\pi(x)$ is the expected value of $\pi(x)$ given y. a_0 and a_y stand for the options selected by the decision maker in the absence and presence of information y.

Simulation-based design (SBD) [24] is an emerging field of study in engineering design wherein the focus is on using computational models for supporting design decisions typically made by humans to ensure cost-effectiveness of a design process. In SBD, a computational model when exercised is a source of information that a designer can take into consideration in making decisions associated with efficacy of a design. Even though the actual system behavior is unknown at this stage, getting more information to increase the fidelity of computational models (e.g., modeling more effects) can improve the accuracy of prediction and approximate the actual system behavior. Therefore, refining simulation models is appropriate for reducing uncertainties. Panchal and coauthors [4] assume that the imprecision bounds (upper and lower bounds) of a simulation model are available, and propose the *improvement potential* (P_I) indices for measuring the value-of-information in model refinement.

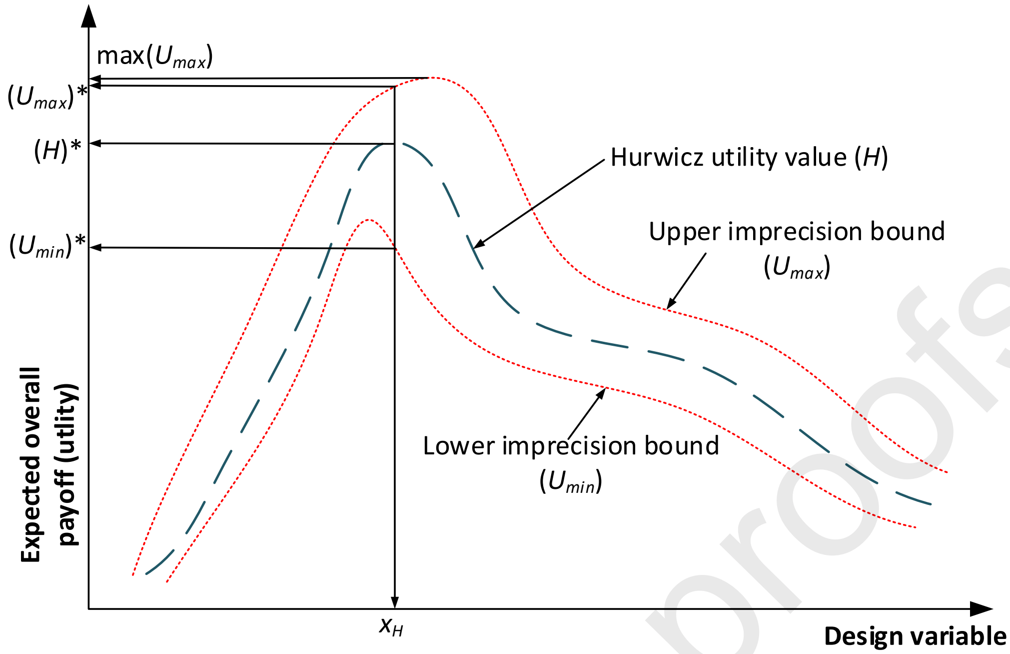


Figure 2. Decision making based on imprecision bounds [4]

The imprecision bounds of the simulation model represent the upper and lower bounds on the overall utility function (decision is made to maximize the overall utility), as shown in Figure 2. The actual payoff lies somewhere (unknown) in the belt confined by lower bound U_{min} and upper bound U_{max}. Even though the actual utility function is unknown, a designer can still make decisions based on some rules such as maximizing the lower bound (i.e., improving the worst case scenario) or upper bound (i.e., improving best possible scenario) on achievable payoff, or maximizing the weighted combination (i.e., Hurwicz criterion [25]) of payoff. When Hurwicz criterion is followed, the decision point should be the x_H that maximize the Hurwicz utility value H. The lower and upper bound on expected payoff at x_H are represented as $(U_{min})^*$ and $(U_{max})^*$ respectively. The maximum achievable payoff through the design space is max (U_{max}) which is calculated by maximizing the upper bound of the overall utility. Given that the imprecision bounds of the overall utility are available, the maximum possible value-of-information (namely, the maximum possible incremental payoff by reducing the range between the bounds) can be evaluated using the *improvement potential* (P_I) as follows:

$$P_I = \max(U_{max}) - (U_{min})^* \tag{2}$$

where $\max(U_{max})$ is the maximum expected payoff that can be achieved by any point in the design space and $(U_{min})^*$ is the lowest expected payoff value achieved by the selected point in the design space (before adding information).

In this paper, we assume that simulation models with different fidelities (characterized with different imprecision bounds and cost) are available as information sources for a designer to acquire in order to reduce the uncertainty associated with the parameters. The concept of *improvement potential* is extended to measure the value-of-information when there are multiple parameters are uncertain.

2.3 Sensitivity Analysis in Multiple Parameters under Uncertainty

In cases where multiple parameters are uncertain, the variation of the overall utility is a co-effect contributed by the parameters all-together. When a designer tries to acquire more information to reduce the uncertainty (range) of a parameter, it is important to know the effect of this action on the overall utility. This is important especially when a designer wants to spend a limited budget on acquiring information to reduce the uncertainty pertaining to the "most critical" parameters. Sensitivity Analysis (SA) is the study of how the uncertainty in the output of a mathematical model or system can be apportioned to different sources of uncertainty in its inputs [26]. It provides a way for a designer to understand the contribution to the overall variance in the output from each of the input parameters. In the SA literature, the method proposed by Sobol and the Fourier amplitude sensitivity test (FAST) method are two of the most popular methods. The method proposed by Sobol [27] is a variance-based sensitivity analysis method that decomposes the variance of the output into fractions which can be attributed to inputs or sets of inputs. The main advantage of the method proposed by Sobol is anchored in its capability of computing the "Total Sensitivity Index (S_{Ti})" defined as the sum of all the sensitivity indices involving that factor. The Sobol method has been used for sensitivity analysis in many applications, such as financial big data analysis [28] and petroleum engineering [29], etc. However, one drawback of the Sobol method lies in its computational efficiency – it relies on a large number of sample points in the input space [30]. The FAST method introduced by Cukier and coauthors [31, 32] is a computationally-efficient sensitivity analysis method that converts multidimensional input space into a single dimensional space by using a suitably defined search curve, which is given by

$$x_i(s) = G_i(sin\, w_i s) \tag{3}$$

where s is a scalar variable varying over the $-\infty < s < +\infty$, G_i are a transformation functions, and $\{w_i\}$ is a set of different frequencies to be properly selected associated with each factor. As s varies, all the input factors change simultaneously along a curve and systematically explores the input space. Each x_i oscillates periodically at its corresponding frequency w_i, and output Y ($Y = f(x_1,x_2,\ldots,x_n)$) shows different periodicities combined with the different frequencies w_i , whatever f is. If some x_i has a strong influence on the output, the oscillations of Y at frequency w_i shall be of high amplitude, which forms the basis for computing sensitivity. FAST has been regarded as one of the most elegant methods for SA [30]. The main advantage of FAST is its computational efficiency, however, it has limitations in computing the total effects (instead of the main effect). Saltelli and coauthors [30] proposed an extended FAST (EFAST) method that combines FAST's better efficiency and Sobol's capacity to compute total effects. The key idea of EFAST is assigning frequency w_i to the ith factor and a different frequency $w_{(i)'}$ to all the remaining factors. By evaluating the spectrum at $w_{(i)'}$ and its higher harmonics $pw_{(i)'}$, the partial variance $D_{(-i)}$ can be estimated, and the total effect (sensitivity S_i) of the ith factor is calculated by

$$S_i = D_{(Ti)} = D - D_{(-i)} \tag{4}$$

where the index $-i$ stands for "all but i." For detailed calculation of the variance D and $D_{(-i)}$, see [30]. Different transformation functions have been proposed [30, 31]. To make the samples distributed in a more general range [a_i, b_i], Lauret and coauthors [33] proposed a transformation function given by

$$x_i(s) = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{\pi} arcsin\,(sin\, w_i s) \tag{5}$$

In this paper, EFAST is used to evaluate the total effect on the overall utility contributed by each ranged parameter, to facilitate a designer making a decision on which parameter (or set of parameters) should be of first priority for reducing uncertainty when multiple parameters are uncertain.

3. Proposed Method

3.1 Modeling Design Decisions – the Utility-Based Compromise Decision Support Problem

In this paper, the utility-based compromise Decision Support Problem (cDSP) [7, 34] is used for modeling decisions in design. The cDSP construct is proposed by Mistree and coauthors [7] for formulating decisions that involve making a trade-off among multiple design objectives, it is a hybrid formulation based on mathematical programing and goal programing. Seepersad and coauthors [34] propose a utility-based cDSP wherein they replace the standard deviation function with a multi-attribute utility function. In this formulation, individual goals are formulated as single-attribute utility functions, and multiple goals are combined in the objective function using Archimedean weightings. The mathematical formulation of the utility-based cDSP is provided in Figure 3.

Given

An alternative to be improved through modification.
Assumptions used to model the domain of interest.
The system parameters.

n	number of system variables
$p+q$	number of system constraints
p	equality constraint
q	inequality constraints
m	number of system goals
$G_i(X)$	system constraint functions
$A_i(X)$	system goal functions
$u_i(A_i(X))$	utility function for each goal
$U(X)$	overall, multi-attribute utility function $= f[u_1(A_1(X)),\ldots,u_m(A_m(X))]$

Find

System variables

$X = X_1,\ldots,X_j \qquad j = 1,\ldots,n$

Deviation Variables

$d_i^-, d_i^+ \qquad i = 1,\ldots,m$

Satisfy

System constraints (linear, nonlinear)

$G_r(X) = 0 \qquad r = 1,\ldots,p$

$G_r(X) \geq 0 \qquad\qquad r = p + 1,\ldots,p + p$
System goals (linear, nonlinear)
$E[u_i(A_i(X))] + d_i^- - d_i^+ = 1$
Bounds
$X_j^{min} \leq X_j \leq X_j^{max} \qquad\qquad j = 1,\ldots,n$
$d_i^-, d_i^+ \geq 0; d_i^- \cdot d_i^+ = 0$
Minimize
Deviation Function: Additive Multi-Attribute Utility Function
$Z = 1 - E[U(X)] = \sum_{i=1}^{m} k_i(d_i^- + d_i^+)$

Figure 3. Mathematical formulation of utility-based cDSP [34]

3.2 Modeling Information Acquisition Decisions in Design Under Multi-Parametric Uncertainty

A problem formulated using the utility-based cDSP represents a decision to be made to maximize the utility embodied in the deviation function. Due to the lack of knowledge, a designer may have multiple uncertain parameters in the utility-based cDSP. These uncertain parameters are mainly from the coefficients of the system constraint functions $G_i(X)$ and system goal functions $A_i(X)$, and are represented using imprecision bounds, e.g., $p_k = [p_k^l, p_k^u]$, where p_k stands for a specific uncertain parameter and p_k^l, p_k^u stand for the lower and upper bounds of p_k respectively. Given multiple parameters are uncertain, as shown in Figure 4, the designer is facing meta-level decisions about acquiring information to reduce the uncertainty in the utility-based cDSP formulation. There are two types of meta-level decisions in this circumstance: one is whether to acquire more information, the other is which piece of information to pick from the information sources. As mentioned in Section 2.2.2, simulation models are assumed to be the information sources. Here, we further assume that a designer's budget for acquiring information is limited and model the meta-decisions as follows.

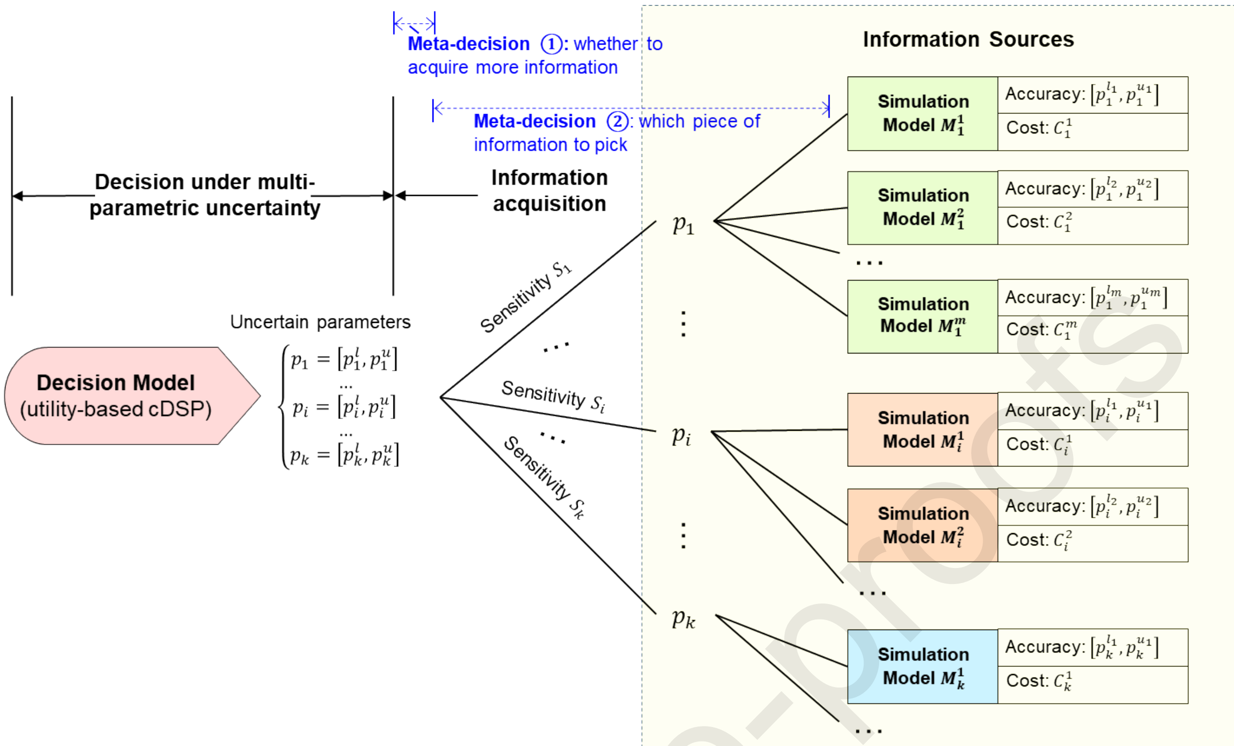


Figure 4. Information acquisition in design under multi-parametric uncertainty

Meta-decision ①: whether to acquire more information. In this meta-decision, the *improvement potential* (P_I) indices are evaluated to facilitate a designer understanding the value of acquiring more information. According to Equation 2, the evaluation of P_I is dependent on the imprecision bounds of the uncertain parameters and the overall utility function and is independent of the information sources. The two terms, $\max(U_{max})$ and $(U_{min})^*$, of Equation 2 are computed using the response surface method. For example, if the utility-based cDSP includes j design variables and k uncertain parameters, then within a $j+k$ dimensional space N^{j+k} sample points can be generated using an orthogonal grid with N (N can be different according to length of the range) elements per variable or parameter. Using the sample points, both $\max(U_{max})$ and $(U_{min})^*$ are calculated based on the overall utility function $U(X)$ as well as the constraint functions $G_i(X)$ (the points should be subject to the constraints). Given that the *improvement* *potential* P_I is known, the criterion below is used for the designer to make a decision on whether to acquire more information:

$$\begin{cases} P_I > P_T & acquire\ more\ information \\ P_I \leq P_T & stop\ acquiring \end{cases} \tag{6}$$

where P_T denotes a threshold value that is based on the designer's judgement according to the features of the design problem.

Meta-decision ②: which information to pick. This is a coupled meta-decision that includes two sub-decisions: 1) which parameter in the set of $\{p_1,p_2,\ldots,p_k\}$ to reduce uncertainty, and 2) which information source (simulation model) to choose for reducing the uncertainty pertaining to a parameter. The former mainly depends on the parameter sensitivities $S_i\ (i = 1,\ldots,k)$ which associates the *improvement potential* P_I to different parameters. The latter depends on accuracy (i.e., the new imprecision bounds after information acquisition) and the cost which measures the performance of the of the simulation models $M_i^j(i = 1,\ldots,k; j = 1,2,\ldots)$. It is a coupled decision because the sensitivity of a parameter also influences the overall performance of a simulation model that is used for reducing the uncertainty in the regarding parameter to improve the overall expected utility of the design. Therefore, the designer should concurrently consider sensitivity, accuracy, and cost in the decision about which piece of information (simulation model) to pick. Based on this, we propose a merit function for measuring the overall performance of the information sources given by

$$MP_i^j = K \cdot \frac{P_I \cdot S_i \cdot \left(1 - \frac{p_i^{u'} - p_i^{l'}}{p_i^u - p_i^l}\right)}{C_i^j} + \varepsilon \quad \left(i = 1,\ldots,k;\ j = 1,2,\ldots;\ C_i^j \leq B_{avai}\right) \tag{7}$$

where K is a scale coefficient to facilitate visualization, which keeps as a constant across information sources; P_I is the improvement potential which is calculated using Equation 2, see Section 2.2 for details; S_i is the sensitivity which is calculated using Equations 4 and 5, see Section 2.3 for details; $P_I \cdot S_i$ means the improvement potential attributed to parameter i; the term enclosed in the bracket means the capability of the information source for reducing the range of the parameter (the percentage of the reduced range); the term C_i^j means the cost of the information; and B_{avai} denotes the available budget for information acquisition; ε is used to capture the total error which is produced in the evaluation of P_I and S_i. Since P_I is not an estimated parameter (because both of its two terms, namely, $\max(U_{max})$ and $(U_{min})^*$, are not estimated), ε is mainly contributed by the evaluation of S_i which is estimated using sampling methods based on Equation 5. It is seen from Equation 7 that the new range $[p_i^{l'}, p_i^{u'}]$ should be narrower than the original range $[p_i^l, p_i^u]$, otherwise MP_i^j will be negative and the information source is useless. Since both S_i and C_i^j are factors of merit function MP_i^j, the meta-decision is a joint decision taking in account sensitivity- and cost-efficiency. The decision criterion for this meta-decision is to select an information source that maximizes MP_i^j. This is to ensure that every trial of information acquisition improves the expected overall utility as much as possible and that the budget for information acquisition is utilized wisely. Since the improvement potential and parameter sensitivity evolves as more information is given to the decision model, the evaluation of MP_i^j is an "online" rating process for the next candidate information sources so as to facilitate the selection of the most valuable information for the current decision. It should be noted that even though a performance index is not 100% accurate (with error ε), we can still use its mean value to indicate the merit of a specific information source.

3.3 Performance-Based Method for Stepwise Information Acquisition

In Figure 4 the situation of design under multi-parametric uncertainty and information acquisition with multiple information sources available are shown. We assume that a designer can only acquire information from only one source each time and propose a performance-based method for stepwise information acquisition shown in Figure 5. The method consists of two meta-decisions and the following seven steps.

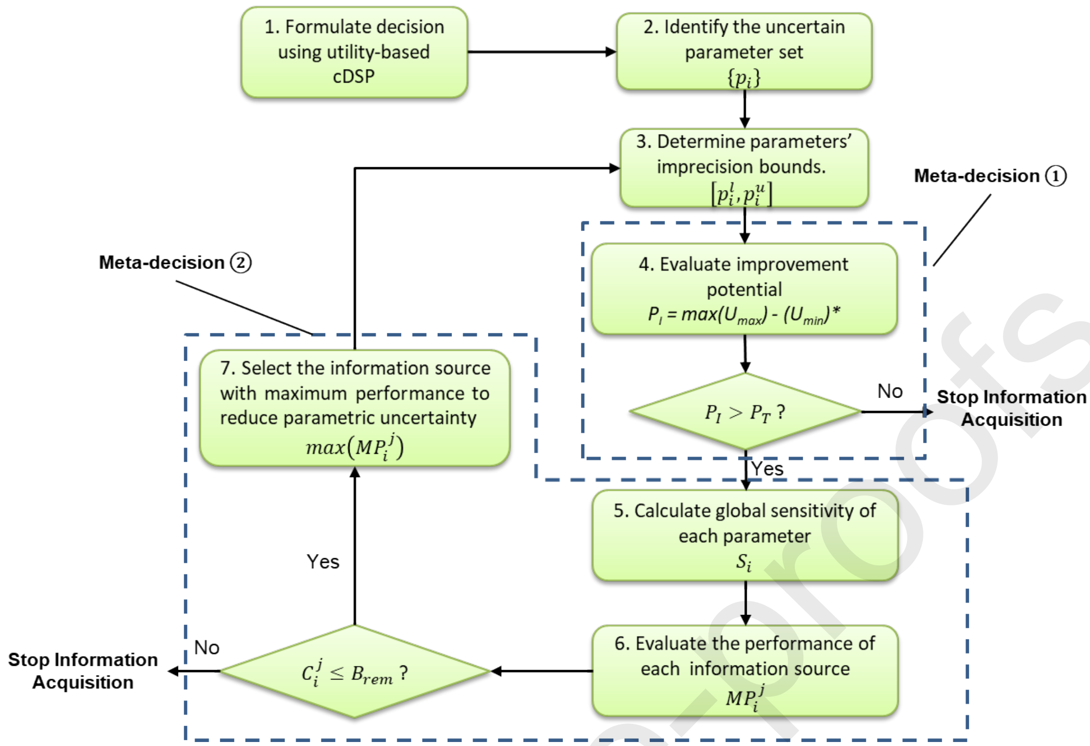


Figure 5. Performance-based method for stepwise information acquisition

Step 1. Formulate the design decision using the utility-based cDSP construct. Mathematic formulatation of the utility-based cDSP is shown in Figure 3. An instantiated example is presented in Section 4.

Step 2. Identify the uncertain parameter set. In this step, a designer identifies all the uncertain parameters in the utility-based cDSP formulation. As mentioned in Section 3.2, these parameters are mainly coefficients of the system goals and constraints.

Step 3. Determine the imprecision bounds of each parameter. These bounds are gatered to evaluate the *Improvement Potential* in Step 4.

Step 4. In this step, the *Improvement Potential* P_I is evaluated by Equation 2 using DOE method. $(U_{min})^*$ and max (U_{max}) of Equation 2 are computed using the response surface method, see [4] (Steps 4 and 5 of Figure 6 in [4]) for details. Both $(U_{min})^*$ and max (U_{max}) are taken as inputs for evaluating P_I. A desinger makes a decision after P_I is evaluated: if P_I is larger than the threshold then proceeds to Step 5, otherwise stops information acquisition. This is meta-decision ①.

Step 5. This is the beginning of meta-decision ②. In this step the global sensitivity of each parameter is calculated based on the EFAST method using the imprecision bounds gathered in Step 3 as the input. With sensitivity S_i, a designer will know what fraction of each parameter can contribute to the *Improvement Potential* P_I. Sensitivity analysis is performed before every information acquisition trial. To calculate the sensitivity, a designer needs to follow two sub-steps as follows.

5.a. Generation of samples of uncertain parameters within the specified imprecision bounds, using Equation 5 of the EFAST method, as defined in Section 2.3.

$$x_i(s) = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{\pi} arcsin\,(sin\,w_i s) \tag{5}$$

where b_i and a_i are the upper bound and lower bound of uncertain Parameter i, s is a variable over the -∞<s<+∞, w_i is a particular frequency, x_i is a generated sample point of Parameter i.

5.b. Calculation of the sensitivity (total effects) of each uncertain parameter using Equation 4, as defined in Section 2.3. The generated sample points x_i in Step 5.a are used as the input for calculating the variances in the overall design utility. For details of variances D and $D_{(-i)}$, see Section 2.3.

$$S_i = D_{(Ti)} = D - D_{(-i)} \tag{4}$$

Step 6. In this step, a designer evaluates the performance of each simulation model considering sensitivity of the parameter, accuracy and cost of the simulation model, and synthesizes them as MP_i^j using Equation 7. If the cost of the simulation model is under the remaining budget, the designer proceeds to Step 7, otherwise information acquisition stops.

Step 7. Among those simulation models of which the cost is under the remaining budget, a designer selects the one with maximum MP_i^j to reduce the uncertainty associsted to the corresponding parametr. From Step 3 to Step 7 is a iterative process. The uncertainty associated to the parameter set $\{p_i\}$ is systematically reduced until the *Improvement Potential* is under the threshold (which means the value of additional information is low) OR the available budget is less than the cost of further simulation models (which means the budget for information acuisition has run out).

4. Example – Design of a Hot Rod Rolling System

In this section, the efficacy of the method proposed in Section 3 is established via a hot rod rolling process design problem which is part of a gear manufacturing process. The foundational problem is contributed by our industrial partner – the Tata Consultancy Services in India. From the raw material to the final gear product, the material goes through multiple manufacturing processes such as casting, rolling, cooling, forging, and machining. All these manufacturing processes are subject to different types of uncertainties. There is natural uncertainty that is inherent in the system due to the complex nature associated with each manufacturing process. These could be operating parameters like temperature ranges, pressure ranges, operating speeds, etc. or the randomness in material microstructure, phases and constituents that are present. From a simulation-based design perspective, the uncertainties present in the system may represent uncertainty in simulation model parameters, control factors, models themselves and uncertainty propagation as models interact or share information. In order to obtain the desired end properties of the gear produced, proper decisions need to be made about the process control parameters (set points) at each of these processes. Nellippallil and co-authors [35, 36] propose a goal-oriented, inverse design method to carry out the decision-based design exploration of these manufacturing processes to satisfy certain end performance requirements. Extension of their work to include Type I, II and III robust design by managing uncertainty across the manufacturing process chain is addressed in [37]. In this paper, we frame the boundary within this manufacturing process chain problem and focus on the rod (product) produced after the rolling and cooling processes. A description of the problem is presented in Section 4.1. Based on this problem, In Section 4.2 we investigate how a designer can cost-efficiently and sensitivity-efficiently perform information acquisition using the performance-based method when multiple (non-controllable) parameters are under uncertainty. To demonstrate the efficacy, the performance-based method is compared to two other acquisition strategies. In Section 4.3, we benchmark the performance-based method to a method in literature to demonstrate the superiority of the former. In Section 4.4, we summarize and discuss the results.

4.1 Problem Formulation Using the Utility-Based cDSP Construct

Step 1 of the performance-based information acquisition method is to formulate a utility-based cDSP. In the context of hot rod rolling process design context, this means formulating a utility-based cDSP to represent the design decision associated with the problem within the cooling stage and the end rod product requirements. The formulation is modified from [36, 38]. The relationships established in the utility-based cDSP are the end mechanical properties of the rod produced; yield strength (YS), tensile strength (TS), and hardness (HV), as a function of the microstructure variables, ferrite grain size (D_α) and phase fraction of ferrite (X_f) after rolling and cooling processes. Five uncertain parameters are identified for this manufacturing process chain problem. They include, pearlite interlamellar spacing (S_0), austenite to ferrite transformation temperature (T_{mf}), the concentrations of manganese, silicon and nitrogen ($[Mn]$, $[Si]$ and $[Ni]$). These parameters are considered uncertain as exact values for these parameters are difficult to measure and are subjected to change due to the complexity involved in the manufacturing process chain thereby affecting the design of the final rod produced. Bounds are defined to establish the upper and lower limits of the system variables. Constraints are defined to establish the maximum and minimum values of the mechanical properties, YS (yield strength), TS (tensile strength), HV (hardness) and toughness measured in terms of impact transition temperature (ITT). The target values for the goals are defined as YS_{Target}= 330 MPa, TS_{Target}= 750 MPa, HV_{Target}= 170. The design goal here is to achieve high values of mechanical properties close to the target by exploring the solution space of material microstructure given the uncertainty associated with the parameters. The utility-based cDSP reads as follows:

Given

1) End requirements identified for the rod rolling process
- Maximize Yield Strength (Goal)
- Maximize Tensile Strength (Goal)
- Maximize Hardness (Goal)
- Minimize ITT (Requirement in terms of constraint)
2) Well established empirical and theoretical correlations, RSMs and information flow from the end of cooling to the end product mechanical properties. Details provided in [35].
3) System variables and their ranges

Table 1. System variables and ranges for u-cDSP

Sr. No	System Variables	Ranges
1	X_1, ferrite grain size (D_α)	8-25 μm
2	X_2, the phase fraction of ferrite (X_f)	0.1-0.9

4) Uncertain parameters and their probable ranges

Table 2. Uncertain parameters and ranges for u-cDSP

Sr. No	Uncertain Parameters	Probable Ranges
1	Y_{e1}, the pearlite interlamellar spacing (S_0)	0.05-0.20 μm
2	Y_{e2}, manganese concentration after cooling ($[Mn]$)	0.6-2.0 %
3	Y_{e3}, the composition of Si ($[Si]$)	0.20-0.22 %
4	Y_{e4}, the composition of N ($[Ni]$)	0.0074-0.02 %
5	Y_{e5}, the austenite to ferrite transformation temperature (T_{mf})	570-700 °C

Find

System Variables

X_1, ferrite grain size (D_α)

X_2, the phase fraction of ferrite (X_f)

Deviation Variables

d_i^-, d_i^+, *i =1,2,3*

Satisfy

System Constraints

- Minimum yield strength constraint $YS \geq 220$ MPa
- Maximum yield strength constraint $YS \leq 330$ MPa
- Minimum tensile strength constraint $TS \geq 450$ MPa
- Maximum tensile strength constraint $TS \leq 750$ MPa
- Minimum hardness constraint $HV \geq 131$
- Maximum hardness constraint $HV \leq 170$
- Minimum ITT constraint $ITT \geq -100$℃
- Maximum ITT constraint $ITT \leq 100$℃

Utility based system goals for yield strength, tensile strength, and hardness

- U_{YS}: Maximize Yield Strength $1.2\left(\frac{YS(X_i)}{YS_{Target}}\right)^2 - 0.2\left(\frac{YS(X_i)}{YS_{Target}}\right) + d_1^- - d_1^+ = 1$
- U_{TS}: Maximize Tensile Strength $1.3\left(\frac{TS(X_i)}{TS_{Target}}\right)^2 - 0.3\left(\frac{TS(X_i)}{TS_{Target}}\right) + d_2^- - d_2^+ = 1$
- U_{HV}: Maximize Hardness $1.4\left(\frac{HV(X_i)}{HV_{Target}}\right)^2 - 0.4\left(\frac{HV(X_i)}{HV_{Target}}\right) + d_3^- - d_3^+ = 1$

Variable Bounds Defined in Table 1

Bounds on deviation variables

$$d_i^-, d_i^+ \geq 0 \text{ and } d_i^- * d_i^+ = 0, \text{i} = 1,2,3$$

Minimize

Deviation from weighted overall maximum utility:

$$Z = 1 - (0.34U_{YS} + 0.33U_{TS} + 0.33U_{HV})$$

Using the above utility-based cDSP, information acquisition to reduce the multi-parametric uncertainty in the formulation is performed in Section 4.2. In Section 4.2.1, multiple information sources for acquisition are presented. Section 4.2.2 is an baseline acquisition experiment using the performance-based method proposed in this paper. Section 4.2.3 is the comparison between the baseline experiment and two other.

4.2 Information Acquisition Regarding to the Uncertain Parameters

4.2.1 Pre-processing – Preparation of Information Sources

In Section 4.1, five parameters (S_0, Mn, Si, Ni, and T_{mf}) are identified as uncertain parameters and their associated imprecision bounds are given in Table 2. The variation of these parameters affects designers' selection of appropriate design variable values that maximize the overall utility as well as the utility achieved. In order to reduce the uncertainty, designers have an option to acquire more information about these parameters from information sources. Here, we assume that for each uncertain parameter there are five simulation models pertaining to it, as shown in Table 3. These simulation models form the sources for information acquisition. The functionality of these simulation models is to gradually narrow down the ranges of parameters so that they become more and more certain. For example, the initial range of parameter S_0 is [0.05, 0.2] μm, which can be reduced to [0.135, 0.14] μm after five information acquisition trials using simulation models $M11$- $M15$. It should be noted these simulation models must be used in a sequential manner, i.e., the fidelity from low to high. For example, $M12$ cannot be used before $M11$ because it is refined based on $M11$ by adding more fidelity to $M11$. This aligns with the concept in [4] that simulation models are gradually refined to become more and more accurate in prediction of system behavior. Refinement of simulation models inevitably result in cost, which is the price that designers must pay in information acquisition. To reflect the increasing of difficulty of improving the accuracy of simulation models by adding more fidelity, we use the exponential function below to estimate the cost of each model.

$$C = B \cdot (1 + r)^{\frac{1}{q} \cdot (1 - \frac{Range_{new}}{Range_{init.}})} \tag{8}$$

where B denotes the inherent cost of an uncertain parameter (in this case we assume that the inherent costs of the five parameters are \$3, \$20, \$3, \$30, and \$30 respectively), q denotes the base unit percentage of range reduction that incurs cost (q is set up based on designers' prior knowledge, in this case we assign the q values for the five parameters as $q(S_0) = 11\%$, $q(Mn) = 10\%$, $q(Si) = 11\%$, $q(Ni$

$) = 10\%$, and $q(T_{mf}) = 10\%$ to reflect different degrees of expensiveness in reducing their ranges), and r denotes the increasing rate which is assigned as 0.5 per range reduction of q. The cost of simulation model refinement will have an exponential growth as acquisition moves on. Using the exponential function, the cost of all the simulation models are calculated and given in the third column of Table 3.

Table 3. Information sources and associated cost for acquisition

Uncertain Parameters	Information Sources	Cost ($)
Pearlite interlamellar Spacing So (μm) Initial: [0.05,0.2]	$M11$: [0.07,0.19]	6.3
	$M12$: [0.09,0.18]	13.1
	$M13$: [0.11,0.16]	35.0
	$M14$: [0.13,0.15]	73.2
	$M15$: [0.135,0.14]	105.8
Manganese Concentration Mn (%) Initial: [0.6,2]	$M21$: [0.8,1.9]	47.7
	$M22$: [1,1.8]	113.7
	$M23$: [1.2,1.7]	271.0
	$M24$: [1.4,1.68]	512.6
	$M25$: [1.6,1.65]	997.8
Composition of Silicon Si (%) Initial: [0.2,0.22]	$M31$: [0.202,0.218]	6.3
	$M32$: [0.204,0.216]	13.1
	$M33$: [0.206,0.214]	27.4
	$M34$: [0.208,0.212]	57.3
	$M35$: [0.209,0.21]	99.5
Composition of Nickel Ni (%) Initial: [0.0074,0.02]	$M41$: [0.0089,0.0185]	78.8
	$M42$: [0.0104,0.017]	206.8
	$M43$: [0.0119,0.0155]	543.1
	$M44$: [0.0134,0.014]	1426.2
	$M45$: [0.0136,0.0138]	1622.1
Temperature T_{mf} (°C) Initial: [570,700]	$M51$: [583,687]	67.5
	$M52$: [596,674]	151.9
	$M53$: [609,661]	341.7
	$M54$: [622,648]	768.9
	$M55$: [632,638]	1434.7

Given the information sources and the associated cost presented in Table 3, there are many possible paths for designers to perform information acquisition in a stepwise manner. For example, the number of possible paths is up to 3125 (i.e., 5^5) if designers are given five total acquisition trails. In this paper, we set up the constraint for information acquisition so that the budget is $1000 and designers need to spend the money "wisely" to reduce uncertainty pertaining to the five parameters. In order to test the efficacy of the performance-based acquisition method proposed in this paper, in Section 4.2.2 we first implement this method in a baseline experiment, then in Section 4.2.3 we compare the baseline experiment with two controlled experiments that follow different information acquisition strategies, and finally the summary of the experiments and discussion are given in Section 4.3.

4.2.2 Baseline Experiment – Performance-Based Stepwise Information Acquisition

In the baseline experiment, designers follow a strategy of acquiring the information source with highest performance at each acquisition trial until the $1000 budget is used up (as shown in Figure 5). According to Equation 7, the performance of a specific information source is a function of the current improvement potential, sensitivity of the uncertain parameter, initial range of the uncertain parameter, the updated range of the uncertain parameter after acquiring the information, and the cost of the information. This function needs to be evaluated before each acquisition trial to facilitate designers selecting the information source with highest performance. Given the initial ranges of the five parameters presented in Table 2, the starting improvement potential P_I is calculated as 0.2152, and the sensitivities S of the five parameters (S_0, Mn, Si, Ni, and T_{mf}) are calculated as 0.0002, 0.2825, 0.0003, 0.4135, 0.2243 respectively. Using P_I and S as the input, the performances MP of the five alternatives (M11, M21, M31, M41, M51) available for the first acquisition trial are evaluated as 0.001 $\pm$ 0.0005, 0.27 $\pm$ 0.001, 0.002 $\pm$ 0.0005, 0.26 $\pm$ 0.001, and 0.14 $\pm$ 0.001 respectively. It is seen from the evaluation result that M21 has the top mean performance 0.27, therefore M21 is selected for the first information acquisition trial. After the first acquisition trial, the range of Mn is reduced from [0.6%, 2%] to [0.8%, 1.9%]. This new range will again be used to calculate the new improvement potential, new sensitivity, and to evaluate the performances of information sources for the next acquisition trial, then the second acquisition trial is carried out. The acquisition process iteratively moves on (as the flowchart shown in Figure 5) until accumulated cost reaches the $1000 budget (no improvement potential threshold is set to stop the process, unless the budget is used up). The process path of all the acquisition trials in the baseline experiment is captured in Figure 6, and the associated process data is presented in Table 4.

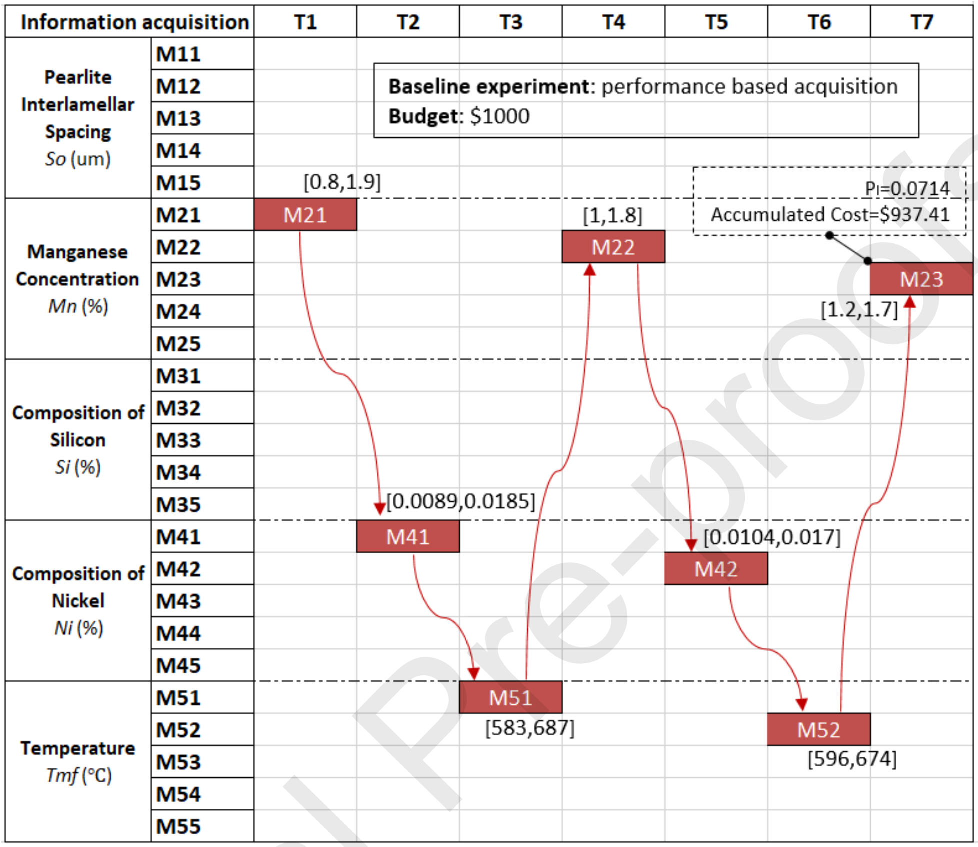


Figure 6. Baseline experiment using the performance-based method

Table 4. Process data the baseline experiment

Info. Acq.	Imp. Poten.	Sensitivity (total effects) S. - S_0	S. - M_n	S. - S_i	S. - N_i	S. - T_{mf}	Design Var. D_α	X_f	Selected Model	Perfo.	Accu. cost	Hurw. Util.
T1	0.2152	0.0002	0.2825	0.0003	0.4135	0.2243	21.22	0.81	M21	0.27 ± 0.01	47.68	0.8427
T2	0.2064	0.0002	0.1914	0.0003	0.4469	0.2771	21.22	0.81	M41	0.28 ± 0.01	126.46	0.8452
T3	0.1826	0.0003	0.2734	0.0004	0.3263	0.3710	17.44	0.81	M51	0.20 ± 0.01	193.96	0.8561
T4	0.1112	0.0014	0.2594	0.0005	0.3114	0.1717	8.00	0.63	M22	0.11 ± 0.005	307.64	0.9083
T5	0.1012	0.0016	0.1762	0.0005	0.4043	0.2156	8.00	0.63	M42	0.06 ± 0.005	514.49	0.9102

T6	0.0884	0.0029	0.2694	0.0012	0.3134	0.4068	8.00	0.63	M52	0.09 ± 0.005	666.37	0.9113
T7	0.0808	0.0032	0.2785	0.0010	0.3295	0.2363	8.00	0.63	M23	0.05 ± 0.005	937.41	0.9116
End	0.0714						8.00	0.63			937.41	0.9129

As shown in Figure 6 and Table 4, the information acquisition process goes through seven different trials: M21→M41→M51→M22→M42→M52→M23. In this process, it is seen that the improvement potential is reduced from the initial 0.2152 to 0.0714 which means a fairly low uncertainty in the obtained overall utility caused by the uncertainties pertaining to the input parameters. And by the end of Trial 7 the accumulated cost reaches 937.41, which is very closed to the budget limit, and designers cannot afford a further acquisition since cost of the next highest-performance information source is far beyond the remaining budget, and therefore the process is stopped. The seven acquisition trials are mainly distributed in Parameters Mn, Ni, and T_{mf}, no trial is distributed to Parameters S_0 and Si. This is because the sensitivities of S_0 and Si are closed to 0 (which means their variation has negligible effect on the overall utility); even though their cost is relatively low, it cannot compensate the low sensitivities. It is also seen in Figure 6 that acquisition trials alternated among Mn, Ni, and T_{mf} frequently. This is because the sensitivities of these parameters vary after each acquisition, which therefore affects the next acquisition. From the overall perspective, it is seen a decrease of improvement potential (P_I) and an increase of utility using Hurwicz criterion (H), as shown in Figure 7. During the first two acquisition trials, the variation of both P_I and H is small. However, through the third acquisition trial (i.e., the range of Temperature T_{mf} is reduced from the initial [570℃, 700℃] to [583℃, 687℃]), both P_I and H vary significantly – P_I decreases from 0.1826 to 0.1112 and H increases from 0.8561 to 0.9083. This rapid variation is due to the sensitivity of T_{mf} at the bounds of [570℃, 700℃] which is relatively high, and once being changed it would result in significant variation in the utility. After the third trial, both P_I and H go through a gentle variation until the acquisition process is stopped.

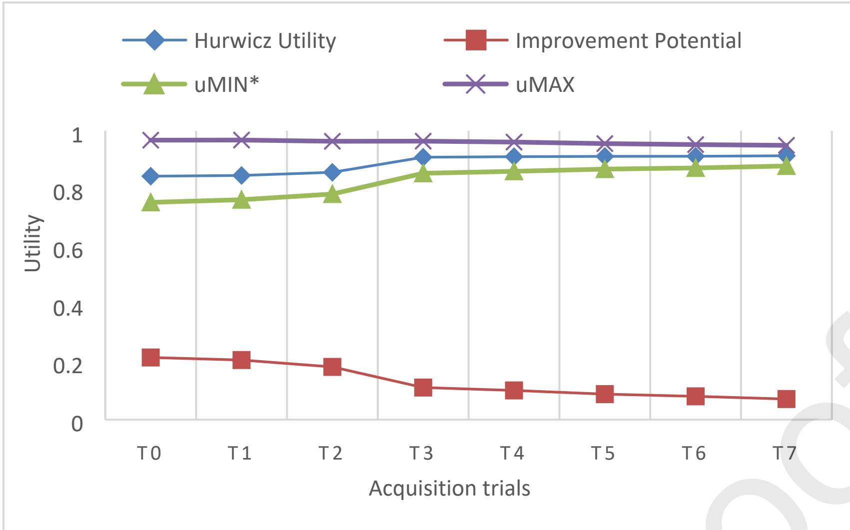


Figure 7. Variation of the Hurwicz utility and improvement potential after information acquisition trials

Figure 8 is an illustration of a designer's decision about the two design variables – D_α and X_f after the information acquisition process is stopped. The lower bounds on the utility for various values of the design variables (D_α and X_f) are plotted as U_{min}, whereas the upper bounds are plotted as U_{max}. The Hurwicz utility is plotted as H, which is maximum at $D_\alpha = 8\ um$ and $X_f = 0.63$. The maximum value of H is denoted as (H)*, which is equal to 0.9129. Some sample points of U_{min} at certain variable ranges (e.g., $X_f = 0.63, D_\alpha \sim [20, 25]$) are equal to 0. This is because solutions within these ranges violate the constraints and are treated as infeasible solutions with a utility value 0. From the plots of U_{max} and U_{min} the imprecision bounds of the overall utility value are very close to each other around the decision points ($D_\alpha = 8, X_f = 0.63$). The difference between them is measured by the improvement potential P_I, which is equal to 0.0714 – representing a low imprecision on the overall utility. Here, a conclusion can be drawn that the \$1000 budget is "wisely" spent to reduce the uncertainty of the overall utility to a relatively low level which is fairly safe for making a decision. If there is no limit on the budgets, uncertainty will keep decreasing and approaching 0 with infinite acquisition trials.

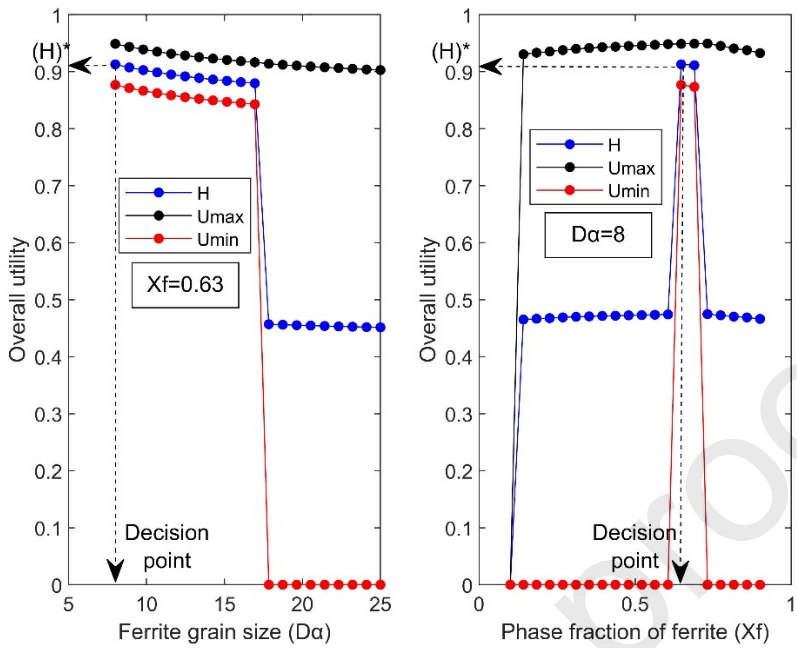


Figure 8. Decision made after 7 information acquisition trials

4.2.3 Comparison – Information Acquisition Using Other Strategies

In this section, we compare the performance-based information acquisition strategy with two other strategies, namely, the cost-only based acquisition strategy and sensitivity-only based strategy, to showcase the advantage of the information acquisition method proposed in this paper. The acquisition paths for the cost-only based strategy as well as the sensitivity-only based strategy are show in Figure 9. Both are subject to the same constraint as the baseline experiment in Section 4.2.2, namely, the budget for information acquisition is limited to $1000.

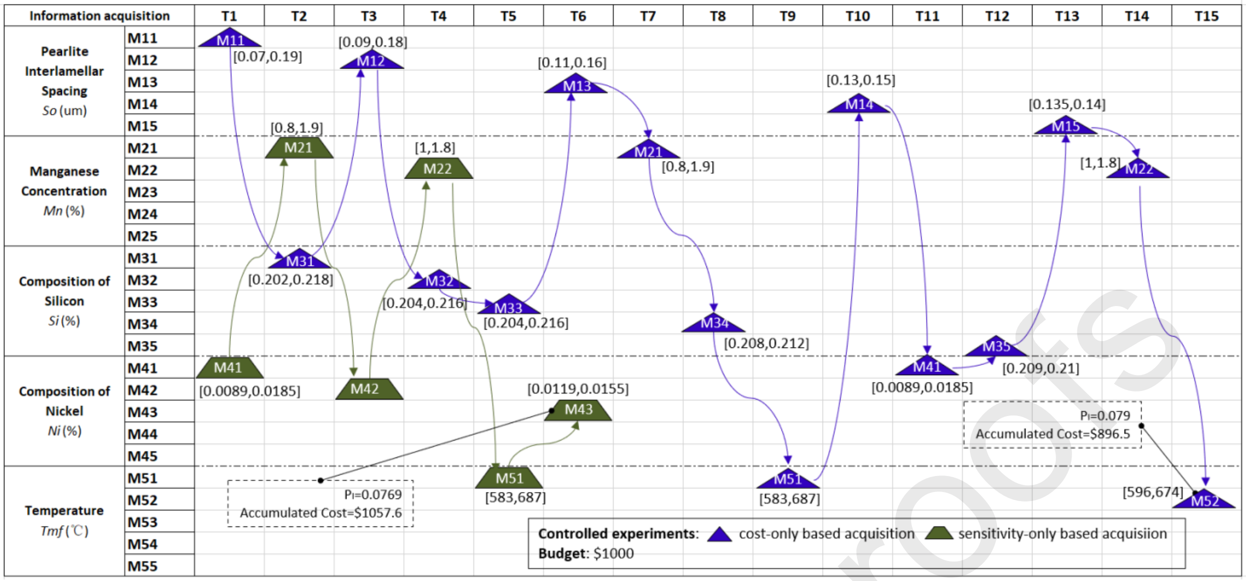


Figure 9. Information acquisition following the cost-only and sensitivity-only strategies

Cost-only Based Information Acquisition. This strategy means that designers always use the information source with minimum cost at any acquisition trial, regardless of the effect the acquired information can have on the overall utility. It represents a prudent attitude in spending money on information acquisition. For example, with caution, designers at the first trial select M11 for $6.3 which is the minimum cost out of the five alternatives. Driven by this strategy, the acquisition process goes through 15 different trials and is stopped when the remaining budget is not able to support a further trial. The path is: M11→M31→M12→M32→M33→M13→M21→M34→M51→M14→M41→M35→M15→M22→M52. Along this path the acquisition trials are distributed to all the five uncertain parameters, and the trials evolve in a way that the cost of acquisition increases sequentially form low to high. In Figure 10, the improvement potential values for all acquisition trials following the cost-only based strategy are plotted (blue dots) and are compared with those following the performance-based strategy (red dots). It is observed that cost-only based acquisition results in a slower decrease of improvement potential than the performance-based acquisition: it only takes the latter 7 trails to decrease improvement potential from 0.2152 to 0.0714 and the process is stopped; however the former only drops to 0.189 after the 8 trials, and it takes the former another 7 trails to decrease to 0.079. This is because in the previous 8 trials of the cost-only based strategy, all the money is spent on acquiring information for those uncertain parameters (So and Si) having little impact on the overall utility. Here, the conclusion is that performance-based acquisition is more efficient than the cost-only based acquisition in reducing the uncertainty pertaining to the overall utility. This is further verified by a bar chart for accumulated cost shown in Figure 11. In performance-based acquisition, the accumulated cost increases very fast and reaches the budget limit in 7 trials. Each acquisition trial, although more expensive than the cost-only based acquisition, is invested in the critical parameters. However, with regard to the cost-only based acquisition, most of the trails are invested for the inconsequential parameters and therefore requires more trials to reach the same uncertainty level as the performance-based acquisition.

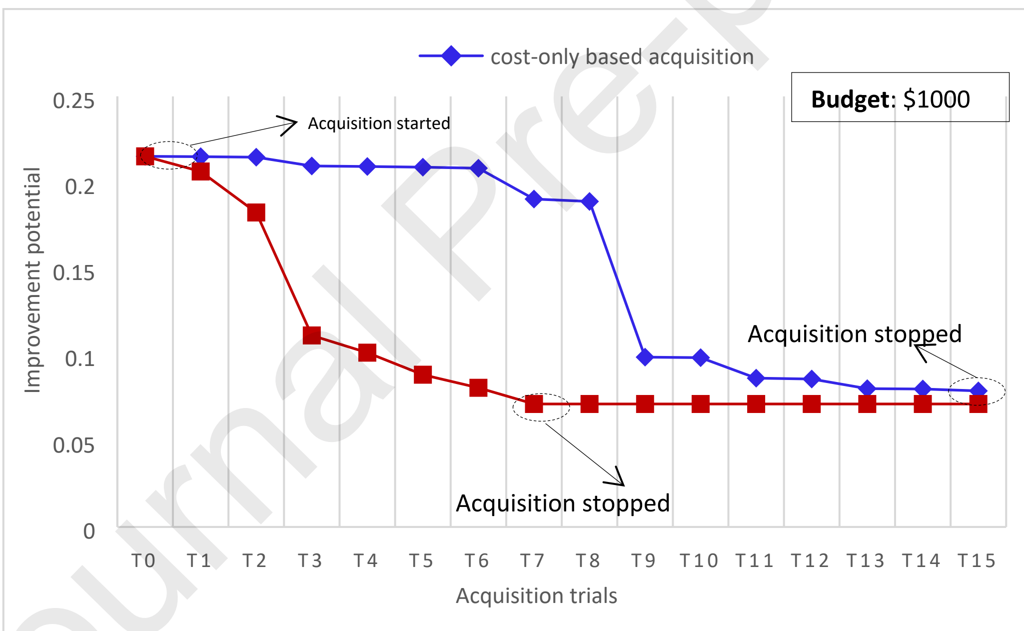


Figure 10. Improvement potential variations in cost-only based acquisition and performance-based acquisition

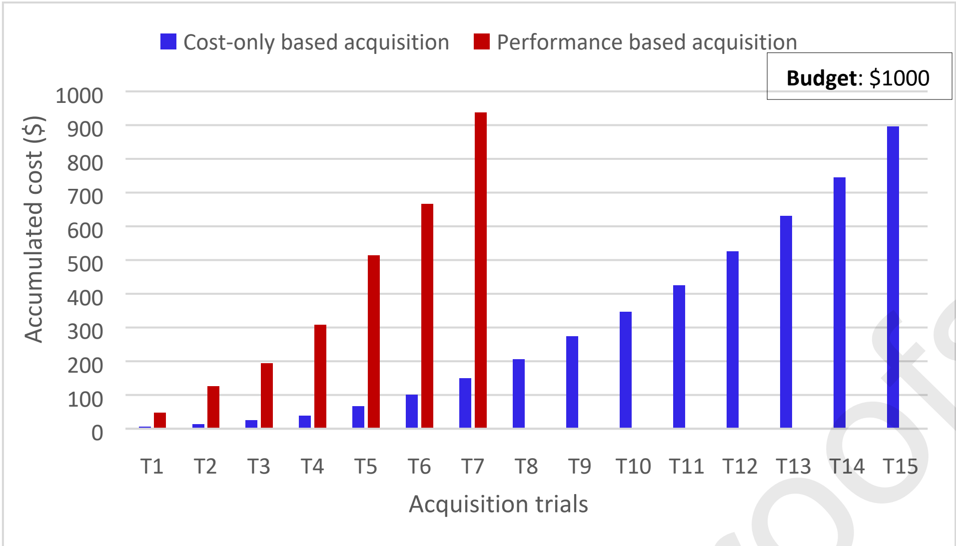


Figure 11. Accumulated cost variations in cost-only based acquisition and performance-based acquisition

Sensitivity-only Based Information Acquisition. This strategy means that designers always acquire the information source pertaining to the uncertain parameter with the greatest sensitivity, regardless of how much the information will cost. Compared to the cost-only strategy, this strategy stands for a greedy attitude in information acquisition because designers want every trial to have the highest impact on the overall utility. Driven by this strategy, the acquisition path is: M41→M21→M42→M22→M51→M43. The process is stopped after the 6th trial when the budget is overrun. It is observed in Figure 9 that all the trials are distributed to only three parameters (Mn, Ni, and T_{mf}) which have high sensitivities. This is the same as the baseline experiment where the performance-based strategy is followed. The improvement potential values for the 6 acquisition trials following the sensitivity-only based strategy are plotted in Figure 12 and are compared with those following the performance-based strategy. It is observed that both strategies have similar tendencies in the decreasing of improvement potential – the improvement potential values of both strategies drops to the same level (0.0884) by the 5th trial. The difference is that in the next acquisition (i.e., Trial 6), the sensitivity-only based acquisition process is stopped because of budget overrun, while the performance-based acquisition process is stopped after two more trials. This is verified by a bar chart of accumulated cost in Figure 13. It is observed that there is a dramatic increase of accumulated cost of the sensitivity-only strategy by the 6th trial which results in budget overrun. By the comparison, it is concluded that even though the sensitivity-only based acquisition strategy is efficient in reducing the uncertainty pertaining to the overall utility, it is likely to result in budget overrun because cost is never considered in information acquisition. Performance based acquisition, by taking both sensitivity and cost into account, is relatively safe against budget overrun especially in the late stages of acquisition when the cost increases dramatically.

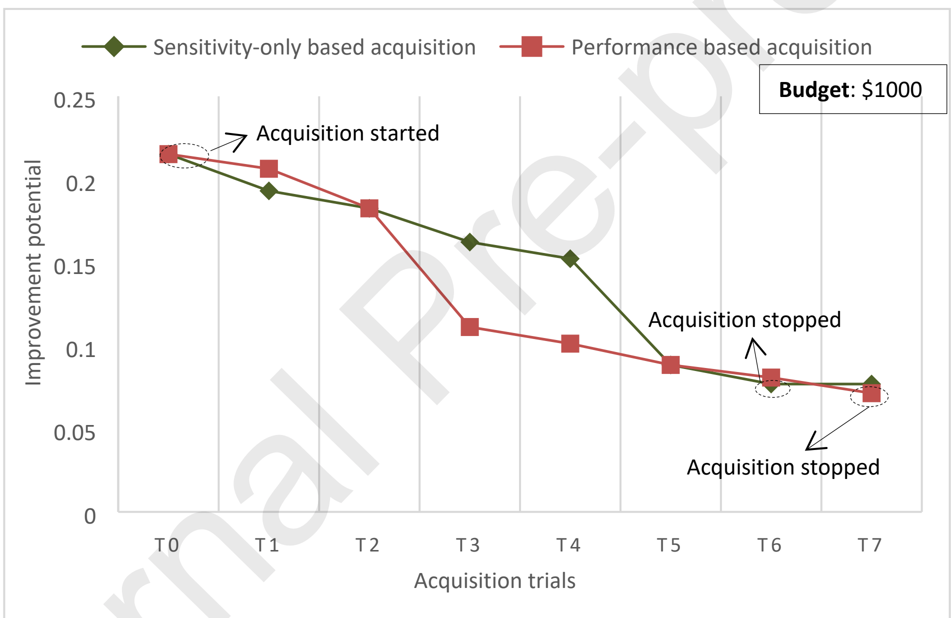


Figure 12. Improvement potential variations in sensitivity-only based acquisition and performance-based acquisition

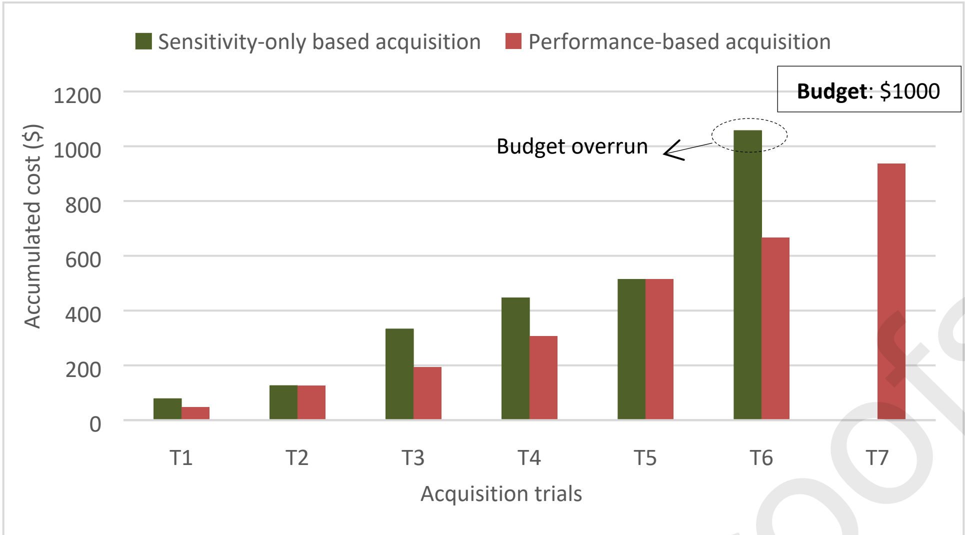


Figure 13. Accumulated cost variations in sensitivity -only based acquisition and performance-based acquisition

4.3 Benchmarking against Methods in Literature

The performance-based information acquisition method is benchmarked against a random method used in Reference [4]. In Reference [4], the focus is on addressing the issue of whether or not to acquire more information when parameters are uncertain. Since evaluation of the performance of multiple sources is not provided in Reference [4] as the guidance for the information acquisition process, designers must randomly select a parameter to reduce its range by acquiring information from the associated simulation models (sources), then randomly switch to another parameter and perform the same task until the improvement potential reaches a low level. We apply this random method to the hot rod rolling process design problem and compare its efficacy in information acquisition with the performance-based method, given that the budgets for both are the same $1000. The comparison results are shown in Table 5. Within the $1000 budget, there are 15 possible paths for information acquisition based on the random method. For example, Path 1 stands for that designers randomly choose Parameter S_o out of the five parameters and use the associated simulation models for acquisition, then after 5 trials (M11-M15) when the associated simulation models cannot be further refined they randomly switch to another parameter *Mn,* and finally stop at simulation model M23 when the remaining budge cannot support another trail. Therefore, the path of P1 is M11→ M12→ M13→ M14→ M15→ M21→ M22→ M23. After Path P1 the improvement potential has dropped from 0.2152 to 0.1698, a percentage of 21.1%. We enumerate all the 15 paths using the random method and obtain an average improvement potential drop of 38.46%, which is about 28% smaller than the drop using the performance-based method. All the improvement potential drops after the 15 acquisition paths are plotted in Figure 14. We observe the maximum drop (58.83%) is achieved by Path P4, and the minimum drop (21.1%) is achieved by Path P1, both are smaller than the drop of 66.82% achieve by the performance-based method. Therefore, we conclude that the performance-based information acquisition method results in a 8-45% larger drop of improvement potential than the random method for the same budget.

Table 5. Information acquisition paths of the random and the performance-based methods

	Acquisition No.	**Path**	**Covered Parameters**	**Imp. Poten. Drop**
Random method of ref. [4]	P1	M11→ M12→ M13→ M14→ M15→ M21→ M22→ M23	So, Mn	21.10%
	P2	M11→ M12→ M13→ M14→ M15→ M31→ M32→ M33→ M34→ M35→ M21→ M22	So, Si, Mn	35.83%
	P3	M11→ M12→ M13→ M14→ M15→ M31→ M32→ M33→ M34→ M35→ M41→ M42	So, Si, Ni	34.67%
	P4	M11→ M12→ M13→ M14→ M15→ M31→ M32→ M33→ M34→ M35→ M51→ M52	So, Si, Tmf	58.83%
	P5	M11→ M12→ M13→ M14→ M15→ M41→ M42	So, Ni	20.17%
	P6	M11→ M12→ M13→ M14→ M15→ M51→ M52	So, Tmf	58.32%
	P7	M21→ M22→ M23→ M24	Mn	17.80%
	P8	M31→ M32→ M33→ M34→ M35→M11→ M12→ M13→ M14→ M15→ M21→ M22	Si, So, Mn	35.83%
	P9	M31→ M32→ M33→ M34→ M35→M11→ M12→ M13→ M14→ M15→ M41→ M42	Si, So, Ni	34.67%
	P10	M31→ M32→ M33→ M34→ M35→M11→ M12→ M13→ M14→ M15→ M51→ M52	Si, So, Tmf	58.83%
	P11	M31→ M32→ M33→ M34→ M35→M21→ M22→ M23	Si, Mn	30.90%
	P12	M31→ M32→ M33→ M34→ M35→M41→ M42	Si, Ni	33.41%
	P13	M31→ M32→ M33→ M34→ M35→M51→ M52→ M53	Si, Tmf	41.59%
	P14	M41→ M42→ M43	Ni	25.56%
	P15	M51→ M52→ M53	Tmf	40.99%
			Average Imp. Poten.	*38.46%*
Performance-based method proposed in this paper (PT)		M21→M41→M51→M22→M42→M52→M23	Mn, Ni, Tmf	66.82%

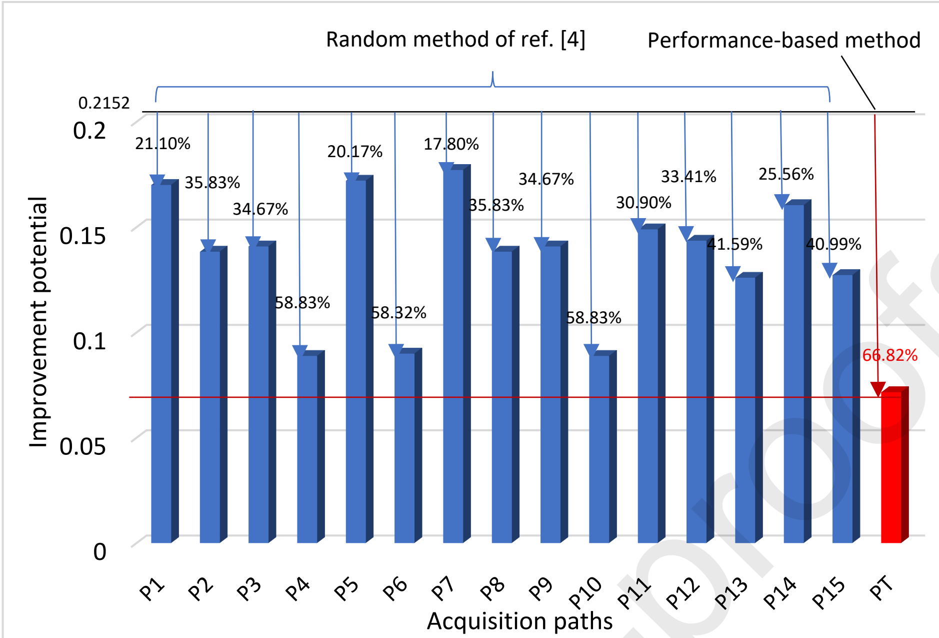


Figure 14. Improvement potential drops after information acquisition based on the random and the performance-based methods

4.4 Discussion of the Results

By the hot rod rolling process design problem, the sensitivity-efficiency and cost-efficiency of the proposed performance based stepwise information acquisition method for design decision making under multiple parametric interval uncertainty is established. First, we use the utility based cDSP construct to formulate the design decision to be made. In the formulation, 5 uncertain parameters (S_0, Mn, Si, Ni, and T_{mf}) and their initial imprecision bounds are identified. Second, as a baseline experiment, we apply the performance-based method to acquire information and gradually reduce the ranges of these 5 uncertain parameters. It is shown in the results that after 7 trials the given acquisition budget is used up, and the improvement potential P_I decreases to a low level which indicates that it is quite safe to make a decision on the two design variables (D_α and X_f). Third, under the same conditions, we conduct another two experiments following two different acquisition strategies – cost-only based acquisition and sensitivity-only acquisition and compare the results with those in the baseline experiment. From the comparison, it is seen that the performance-based method has some advantages over the other two strategies, which is embodied in:

- **Sensitivity-efficiency**. Given a certain budget, it takes less information acquisition trials of the performance-based method to reduce the impacts on the overall utility due to the uncertain parameters. Because parameter sensitivity is considered in the merit function, the performance-based method ensures that each acquisition trial is invested on a critical parameter which guarantees that the acquired information is valuable. This avoids many feckless acquisition trials (e.g., the cost-only strategy) and directs the acquisition quickly to the state with low improvement potential. Therefore, it is a sensitivity-efficient way for information acquisition under multi-parameter uncertainty.
- **Cost-efficiency**. Given a certain level of uncertainty (improvement potential P_I) to reduce to, it takes less resources (money) of the performance-based method. Cost of the information is considered in the merit function of the performance-based acquisition method, which ensures that each acquisition trial is cost-efficient. Cost-efficiency of information acquisition is very important in the later stages of the acquisition process, where the precision of some parameters is already very high and collecting more information to further reduce their ranges will result in a dramatic increase of cost which easily leads to budget overruns. Therefore, the performance-based method is also a cost-efficient method for information acquisition.

Finally, the performance-based method is benchmarked against a random method proposed in Reference [4]. It is observed that the former leads to an 8-45% larger drop of improvement potential than the latter for the same $1000 budget. This is due to the fact that the performance-based method takes in account both sensitivity- and cost-efficiency, which guides the acquisition process towards the most valuable information sources to reach a larger drop of improvement potential, instead of randomly (or blindly) selecting the parameters and investing in ineffective sources.

In the hot rod rolling process design problem, we use 1) simulation models as the sources for designers to acquire information to reduce the ranges of the 5 uncertain parameters, and 2) an exponential function to measure the cost of simulation model with different fidelities. However, this does not mean that the performance-based method is confined to the use of simulation models and exponential functions. Because Equation 7 is problem-independent, the performance-based method can also be applied to problems with other information sources (e.g., physical experiments, experts' knowledge, and technical reports from consulting companies, and any other sources that can narrow down the parameter ranges) and other cost measures (e.g., the linear and polynomial functions that capture the increasing behavior of the cost). The $1000 budget is used as an example to show a threshold of the information cost. In information acquisition context such as conducting physical experiments using high-precision equipment or running high-fidelity simulation models on high-performance computers, the cost can be as large as millions of dollars or more. Therefore, careful evaluation of the potential information sources before using them to make a decision is critically important.

5. Generalization of the Information Acquisition Framework

In this section, we extend the performance-based information acquisition method to a more generic framework so as to facilitate solving multi-parametric uncertainty problems across domains. The generalized framework is shown in Figure 15. The problem is simplified as a mathematical model with two inputs, namely, X and P, and one output Y. X is a vector representing control factors or variables, P is a vector representing uncontrollable factors or parameters, and Y is a function of both X and P represented as $Y = f(X,P)$. The goal is to choose "right" values of X for maximizing Y given that P is uncertain. The uncertainty in P is characterized by a set of ranges with lower bounds and upper bounds. This is a very common problem encountered in many engineering domains such as design, manufacturing, and control, etc. Given this problem, the decision maker can perform information acquisition using the performance-based method to reduce uncertainty in P so as to make safer decisions on choosing X for the maximization of Y. The process are summarized as three key steps as follows.

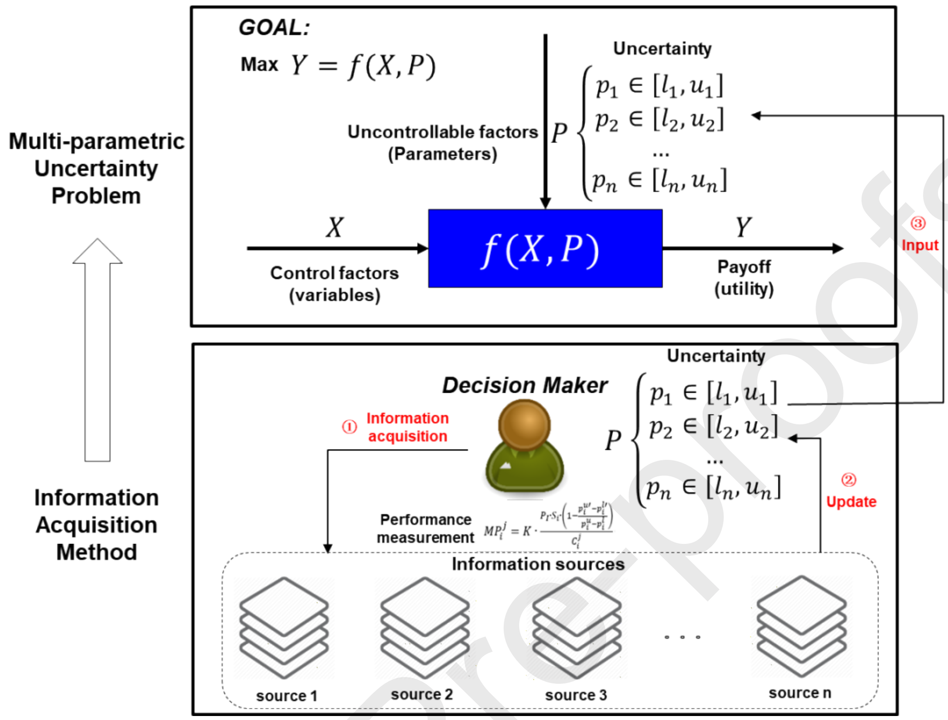


Figure 15. A generalized framework of performance-based information acquisition for multi-parametric uncertainty problems

- **Step ①**: Information acquisition. In order to reduce the uncertainty in P, the decision maker acquires information from multiple information sources (i.e., Source 1, 2, …, n) considering the sensitivity of the uncertain parameters (i.e., p_1,p_2,…,p_n) in P and the cost of the information. The information acquisition process is facilitated by a performance-based step-wise method introduced in Section 3.3. The core is the equation (Equation 7) that measures the performance of the information to be acquired by taking sensitivity and cost into account. For details of the method, see Section 3.3.
- **Step ②**: Updating the uncertainty in P. The acquired information is used to update the knowledge (accuracy) of the uncertain parameters in P. With the acquired information, the ranges of Parameters p_1,p_2,…,p_n is reduced.

- **Step ③**: Inputting the updated P to the mathematical model of the problem and judging whether more information is needed. The uncertain parameters with reduced ranges are input to the mathematical model (which is formulated using the utility-based cDSP construct, see Section 3.1) to calculate the payoff (or utility) as well as the *improvement potential* (P_I, see Section 2.2). Based on the current *improvement potential,* the decision maker judges whether or not more information is needed. If it is needed, the whole process is repeated from Step ① to Step ③.

Since the multi-parametric uncertainty problem is simplified as a mathematical model and the information acquisition method is summarized as three key steps, we believe it is easy to duplicate in different domains. In this paper, we assume that the information sources (i.e., simulation models) are not encrypted and designers have no problem when acquiring information from them, and our focus is on finding the "optimal" sequence or path for designers to acquire information from multiple sources and gradually reduce the uncertainty in their design. In situations where the information sources are encrypted, one of the possible solutions for information acquisition is to combine the performance-based method with homomorphic encryption based on learning with errors [39]. Future research opportunities lie in the consideration of information security when acquiring information in design under multi-parametric uncertainty.

6. Closing Remarks

Resolving multi-parameter uncertainty is a problem that designers often confront as they design complex engineering systems. This problem embodies a dilemma: On the one hand, acquiring more information to reduce the uncertainty is an option to potentially improve the outcome of the decisions to be made. On the other hand, information acquisition may result in cost, and some acquired information may have very little effect on the outcome of the decision. Given the existence of this dilemma, designers must make meta-level decisions to determine whether or not to acquire additional information and how to acquire information. The difficulty in making such meta-level decisions in the context of multiple parameter uncertainty is anchored in the fact that multiple information sources may be available to choose from, and the cost of the information as well as its impact on the overall utility of design are different. In this paper, we propose a performance based stepwise information acquisition method to address this difficulty. What is new in this paper includes the following:

i. We use the utility-based compromise Decision Support Problem (u-cDSP) construct for formulating the decisions on the design variables to maximize the overall utility of design. The u-cDSP construct provides a structured way for integrating multiple goals in one single utility function, which forms the foundation for measuring the value (utility) of information under multi-parameter uncertainty.

ii. We develop a performance-based index (Equation 7) to facilitate choosing the critical (sensitive) and cost-efficient information to reduce the parameter ranges at every acquisition trial, and iteratively reduce uncertainty of the overall utility to a level that can improve the possibility of making a good decision.

iii. We develop a step-wise information acquisition procedure using the proposed performance-based index. The procedure facilitates designers efficiently acquiring the most valuable information to reduce uncertainty given a limited budget.

iv. We demonstrate the efficacy of the method using a hot rod rolling process design problem. In the example problem, we compare the performance-based information acquisition strategy with two different strategies – cost-only based information acquisition and sensitivity-only information acquisition, and highlight the sensitivity- and cost-efficiency of the proposed method.

v. We benchmark the proposed method against the random method in literature. It is shown that the performance-based information acquisition method leads to an 8-45% larger drop of improvement potential than the random method with the same budget.

Information acquisition is an important means for dealing with the reducible uncertainties in engineering design. In this paper, we draw a boundary of the problem to focus on the reducible uncertainty pertaining to the uncontrollable factors Y of the system model $f(X,Y)$, as shown in Figure 1. Future research opportunities are anchored in developing information acquisition methods with respect to the uncertainty associated with the mathematical relationship f.

Acknowledgement

Zhenjun Ming acknowledges the support from National Natural Science Foundation of China (51805033), China Postdoctoral Science Foundation (2018M631351), and Beijing Institute of Technology Research Fund Program for Young Scholars (3030011182037). Guoxin Wang and Yan Yan acknowledges the support from National Natural Science Foundation of China (51375049) and National Ministries Foundation (JCKY2016602B007). The authors thank TRDDC, Tata Consultancy Services, Pune for supporting this work (Grant No. 105-373200). Janet K. Allen and Farrokh Mistree gratefully acknowledge funding from the John and Mary Moore Chair and L.A. Comp Chair at the University of Oklahoma, respectively. This paper is an outcome of the International Systems Realization Partnership between the Institute for Industrial Engineering @ The Beijing Institute of Technology, the Systems Realization Laboratory @ The University of Oklahoma and the Design Engineering Laboratory @ Purdue University.

References

[1] H. Zhai, J. Zhang, Equilibrium reliability measure for structural design under twofold uncertainty, Information Sciences, 477 (2019) 466-489.
[2] B. Zahiri, P. Jula, R. Tavakkoli-Moghaddam, Design of a pharmaceutical supply chain network under uncertainty considering perishability and substitutability of products, Information Sciences, 423 (2018) 257-283.
[3] Y. Feng, Z. Hong, G. Tian, Z. Li, J. Tan, H. Hu, Environmentally friendly MCDM of reliability-based product optimisation combining DEMATEL-based ANP, interval uncertainty and Vlse Kriterijumska Optimizacija Kompromisno Resenje (VIKOR), Information Sciences, 442-443 (2018) 128-144.
[4] J.H. Panchal, C.J.J. Paredis, J.K. Allen, F. Mistree, A value-of-information based approach to simulation model refinement, Engineering Optimization, 40 (2008) 223-251.

[5] H.A. Simon, Administrative behavior: a study of decision-making processes in administrative organization, Macmillan Co., New York, 1947.
[6] E. Galvan, C. Hsiao, S. Vermillion, R. Malak, A Parallel Approach for Computing the Expected Value of Gathering Information, SAE International Journal of Materials and Manufacturing, 8 (2015) 271-282.
[7] F. Mistree, O.F. Hughes, B.A. Bras, The Compromise Decision Support Problem and the Adaptive Linear Programming Algorithm, Structural Optimization: Status and Promise, (Kamat, M. P., Ed.), AIAA, Washington, DC, (1993) 247-286.
[8] R.A. Howard, Information Value Theory, IEEE Transactions on Systems Science and Cybernetics, 2 (1966) 22-26.
[9] M. Tavana, F.J. Santos-Arteaga, D. Di Caprio, The value of information as a verification and regret-preventing mechanism in algorithmic search environments, Information Sciences, 448-449 (2018) 187-214.
[10] C.F. Martin, Y.C. Ho, Value of information in the Polya urn process, Information Sciences, 147 (2002) 65-90.
[11] S. Bradley, A. Agogino, W. Wood, Intelligent engineering component catalogs, in: Artificial intelligence in design'94, Springer, 1994, pp. 641-658.
[12] S.R. Bradley, A.M. Agogino, An Intelligent Real Time Design Methodology for Component Selection: An Approach to Managing Uncertainty, J Mech Design, 116 (1994) 980-988.
[13] W.H. Wood, A.M. Agogino, Decision-based conceptual design: modeling and navigating heterogeneous design spaces, J Mech Design, 127 (2005) 2-11.
[14] J.H. Panchal, C.J.J. Paredis, J.K. Allen, F. Mistree, Managing Design-Process Complexity: A Value-of-Information Based Approach for Scale and Decision Decoupling, J Comput Inf Sci Eng, 9 (2009) 021005-021005-021012.
[15] M. Messer, J.H. Panchal, V. Krishnamurthy, B. Klein, P.D. Yoder, J.K. Allen, F. Mistree, Model Selection Under Limited Information Using a Value-of-Information-Based Indicator, J Mech Design, 132 (2010) 121008-121008-121013.
[16] M. Messer, J.H. Panchal, J.K. Allen, F. Mistree, Model refinement decisions using the process performance indicator, Engineering Optimization, 43 (2011) 741-762.
[17] S. Isukapalli, A. Roy, P. Georgopoulos, Stochastic response surface methods (SRSMs) for uncertainty propagation: application to environmental and biological systems, Risk analysis, 18 (1998) 351-363.
[18] J.K. Allen, C. Seepersad, H. Choi, F. Mistree, Robust Design for Multiscale and Multidisciplinary Applications, J Mech Design, 128 (2006) 832-843.
[19] E. Nikolaidis, D.M. Ghiocel, S. Singhal, Types of uncertainty in design decision making, in: Engineering Design Reliability Handbook, CRC Press, 2004, pp. 137-156.
[20] A.C. Ward, T. Lozano-Perez, W.P. Seering, Extending the constraint propagation of intervals, Ai Edam, 4 (1990) 47-54.
[21] E.K. Antonsson, K.N. Otto, Imprecision in engineering design, Journal of Vibration and Acoustics, 117 (1995) 25-32.
[22] D.B. Lawrence, The economic value-of-information, NewYork: Springer, 1999.
[23] J.M. Ling, J.M. Aughenbaugh, C.J. Paredis, Managing the Collection of Information Under Uncertainty Using Information Economics, J Mech Design, 128 (2006) 980-990.
[24] W. Mefteh, Simulation-Based Design: Overview about related works, Mathematics and Computers in Simulation, 152 (2018) 81-97.
[25] K. Arrow, L. Hurwicz, An optimality criterion for decision-making under ignorance, In: C.F. Carter and J.L. Ford, eds. Uncertainty and expectation in economics: essays in honour of G.L.S. Shackle. Oxford: Blackwell, (1972) 1–11.
[26] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global sensitivity analysis: the primer, John Wiley & Sons, 2008.

[27] I.M. Sobol, Sensitivity Analysis for Nonlinear Mathematical Models, Mathematical Modeling & Computational Experiment, 1 (1993) 407-414.
[28] J. Ruan, Z. Yan, B. Dong, Q. Zheng, B. Qian, Identifying suspicious groups of affiliated-transaction-based tax evasion in big data, Information Sciences, 477 (2019) 508-532.
[29] E. Poisson, S. Brunel-Muguet, F. Kauffmann, J. Trouverie, J.C. Avice, A. Mollier, Sensitivity analyses for improving sulfur management strategies in winter oilseed rape, PLoS One, 13 (2018) 20.
[30] A. Saltelli, S. Tarantola, K.-S. Chan, A quantitative model-independent method for global sensitivity analysis of model output, Technometrics, 41 (1999) 39-56.
[31] R. Cukier, C. Fortuin, K.E. Shuler, A. Petschek, J. Schaibly, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory, The Journal of chemical physics, 59 (1973) 3873-3878.
[32] R. Cukier, H. Levine, K. Shuler, Nonlinear sensitivity analysis of multiparameter model systems, Journal of computational physics, 26 (1978) 1-42.
[33] P. Lauret, E. Fock, T.A. Mara, A node pruning algorithm based on a Fourier amplitude sensitivity test method, IEEE transactions on neural networks, 17 (2006) 273-293.
[34] C.C. Seepersad, F. Mistree, J.K. Allen, Designing evolving families of products using the utility-based compromise decision support problem, International journal of mass customisation, 1 (2005) 37-64.
[35] A.B. Nellippallil, V. Rangaraj, B.P. Gautham, A.K. Singh, J.K. Allen, F. Mistree, An Inverse, Decision-Based Design Method for Integrated Design Exploration of Materials, Products, and Manufacturing Processes, J Mech Design, 140 (2018) 111403-111403-111417.
[36] A.B. Nellippallil, K.N. Song, C.-H. Goh, P. Zagade, B.P. Gautham, J.K. Allen, F. Mistree, A Goal-Oriented, Sequential, Inverse Design Method for the Horizontal Integration of a Multistage Hot Rod Rolling System, J Mech Design, 139 (2017) 031403-031403-031416.
[37] A.B. Nellippallil, P. Mohan, J.K. Allen, F. Mistree, Robust Concept Exploration of Materials, Products and Associated Manufacturing Processes, in: ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, 2018, pp. V02BT03A010-V002BT003A010.
[38] Z. Ming, A.B. Nellippallil, Y. Yan, G. Wang, C.H. Goh, J.K. Allen, F. Mistree, PDSIDES—A Knowledge-Based Platform for Decision Support in the Design of Engineering Systems, J Comput Inf Sci Eng, 18 (2018) 041001-041001-041014.
[39] C. Gentry, A. Sahai, B. Waters, Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based, in: Annual Cryptology Conference, Springer, 2013, pp. 75-92.

Zhenjun Ming: Conceptualization, Methodology, Software, Validation, Formal analysi, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization

Anand Balu Nellippallil: Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing

Guoxin Wang: Supervision, Project administration, Funding acquisition

Yan Yan: Supervision, Project administration, Funding acquisition

Janet K. Allen: Writing - Review & Editing, Supervision, Funding acquisition

Farrokh Mistree: Writing - Review & Editing, Supervision, Funding acquisition

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: