



HAL
open science

Calibration of imprecise and inaccurate numerical models considering fidelity and robustness: a multi-objective optimization-based approach

Sez Atamturktur, Zhifeng Liu, Scott Cogan, Hsein Juang

► **To cite this version:**

Sez Atamturktur, Zhifeng Liu, Scott Cogan, Hsein Juang. Calibration of imprecise and inaccurate numerical models considering fidelity and robustness: a multi-objective optimization-based approach. *Structural and Multidisciplinary Optimization*, 2014, 51 (3), pp.659-671. 10.1007/s00158-014-1159-y . hal-02300231

HAL Id: hal-02300231

<https://hal.science/hal-02300231>

Submitted on 8 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Calibration of imprecise and inaccurate numerical models considering fidelity and robustness: a multi-objective optimization-based approach

Sez Atamturktur, Zhifeng Liu, Scott Cogan, Hsein Juang

Abstract Traditionally, model calibration is formulated as a single objective problem, where fidelity to measurements is maximized by adjusting model parameters. In such a formulation however, the model with best fidelity merely represents an optimum compromise between various forms of errors and uncertainties and thus, multiple calibrated models can be found to demonstrate comparable fidelity producing non-unique solutions. To alleviate this problem, the authors formulate model calibration as a multi-objective problem with two distinct objectives: fidelity and robustness. Herein, robustness is defined as the maximum allowable uncertainty in calibrating model parameters with which the model continues to yield acceptable agreement with measurements. The proposed approach is demonstrated through the calibration of a finite element model of a steel moment resisting frame.

Keywords Nondominated sorting genetic algorithm · Info-gap decision theory · Info-gap uncertainty model · Prediction looseness · Experiment-based model validation · Self-consistency

1 Introduction

Due to inevitably uncertain model parameters and incomplete representations of underlying physics, computer models only

provide an approximate representation of reality. The approximate nature of computer modeling has motivated the development of calibration techniques through which model parameters are tuned to achieve an improved agreement between model predictions and measurements. In the calibration of computer models, heuristic search algorithms developed for optimization problems are reported to be particularly well suited and thus, are widely implemented in many engineering fields (Nagai and Kadoya 1979; Bekey and Masri 1983; Luce and Cundy 1994; Franchini et al. 1998; Solomatine 1999; Mattsson et al. 2001; Farajpour and Atamturktur 2013).

Traditionally, in optimization-based calibration, fidelity to measurements is treated as the sole objective, in that a set of model parameter values is sought that maximizes the agreement between model predictions and experimentally-measured responses. In such treatments of model calibration however, compensations between various forms of uncertainties and errors inherent to the formulation and parameterization of the numerical model as well as the nature of the available experiments become inevitable. These compensations might be due to the limited spatial and frequency points, noisy data, model form errors, collinearity between parameters, etc. Thus, the obtained optimal fidelity model merely represents the best compromise between uncertainties and errors originating from diverse sources. These compensating effects leads to a family of plausible models, each displaying a *comparable* measure of fidelity to experiments, an issue commonly known as non-uniqueness (Berman 1995). What is more, the number of non-unique solutions tends to increase as measurement uncertainty increases or as the number of available measurements decreases. It is also important to note that these optimal and *a fortiori* suboptimal solutions may still not yield acceptable levels of fidelity for an intended use.

Among candidate models demonstrating comparable fidelity, a model could be considered preferable over another if it

S. Atamturktur (✉) · Z. Liu · H. Juang
Glenn Department of Civil Engineering, Clemson University,
Clemson, SC, USA
e-mail: sez@clemson.edu

S. Cogan
Department of Applied Mechanics, University of Franche-Comté,
Besançon, France

exhibits greater robustness to parameter uncertainty, meaning that the model predictions remain within a predefined threshold even when the uncertainties in the model parameters (or model form) are considered. Such robustness can be conveniently calculated using the framework of info-gap decision theory (Hemez and Ben-Haim 2004).

In this paper, the authors contend that not only the fidelity of the model predictions but also the robustness of this fidelity against uncertainty are important attributes to be considered during model calibration. Hence, the authors treat model calibration as a multi-objective optimization problem where both the model's fidelity to measurements and robustness to parameter uncertainty are treated as objectives. When these two objectives are conflicting, the multi-objective treatment of model calibration results in what is widely known as a Pareto front presenting multiple plausible solutions and demonstrating a trade-off relationship between the model's fidelity and robustness. Although multi-objective optimization can supply a clear picture of the trade-off relationship between fidelity and robustness, the decision maker may still be left to select one, best-performing model for a given intended use. Thus, an approach is suggested herein to select a model that yields a preferable compromise between the two objectives for a specific prediction. This paper formally presents the conceptual framework of multi-objective model calibration and demonstrates its application through a case study of a steel frame building with imprecisely known connection parameters and an incompletely known model form. The paper concludes with a discussion on incorporating the intended use of a numerical model into the optimization framework through the computation of self-consistency of predictions.

2 Robustness in model calibration

Let's consider a model predicting a system response, y_s is a function of model parameters, θ :

$$y_s(\theta) \quad (1)$$

Model calibration, in the most traditional sense, involves adjusting the values of θ in Eq. (1) to improve the agreement between model predictions and measurements, y_m . Optimal values for θ are then sought to minimize the difference between model predictions and measurements. Herein, the vector of parameter values that yields maximum fidelity (i.e., minimum disagreement between predictions and measurements) is indicated by $\tilde{\theta}_f$.

An alternative approach proposed by Ben-Haim et al. (1998), Hemez and Ben-Haim (2004), and Ben-Haim (2004,

2012) is to satisfy a fidelity threshold, R_c , and optimize the robustness of the model against uncertainty while delivering this threshold fidelity (Eq. (2)).

$$|y_s(\theta) - y_m| < R_c \quad (2)$$

Within this approach, robustness is defined as the maximum allowable uncertainty in the poorly known θ parameters such that model predictions agree with measurements no worse than an amount R_c . Herein, the vector of θ parameter values that yield maximum robustness is indicated by $\tilde{\theta}_r$.

The difference between these two approaches to model calibration, optimizing fidelity versus optimizing robustness, is illustrated in Fig. 1 for a model with a single model parameter. Here, the model output, y_s , is plotted as a function of the model parameter, θ . As seen in this figure, the parameter value that exhibits optimal fidelity, $\tilde{\theta}_f$, yields the smallest possible error between predictions and measurements denoted as R_f . This fidelity however is not assured due to the unavoidable presence of both aleatory and epistemic uncertainties. The former form of uncertainty reflects that the calibration parameters may have natural, irreducible variability. The latter reflects that in the context of model calibration, one cannot guarantee that $\tilde{\theta}_f$ represents the so-called *true* parameter value due to inevitable compensations between various forms of uncertainties and errors. Therefore, uncertainty remains in the fidelity-optimal parameter value, $\tilde{\theta}_f$ and as seen in Fig. 1, when this uncertainty is considered, with only a slight deviation from $\tilde{\theta}_f$ denoted as $\hat{\alpha}_f$ (indicating an absolute value), model predictions reach errors beyond acceptable levels defined by R_c .

On the other hand, the parameter value that exhibits optimal robustness to uncertainty, $\tilde{\theta}_r$ yields solutions that satisfy the fidelity threshold, R_c , even when large deviations from $\tilde{\theta}_r$ denoted as $\hat{\alpha}_r$ are considered ($\hat{\alpha}_r \gg \hat{\alpha}_f$). However, this robust-optimal parameter value, $\tilde{\theta}_r$ yields a larger error between the predictions and experiments denoted as R_r compared to the

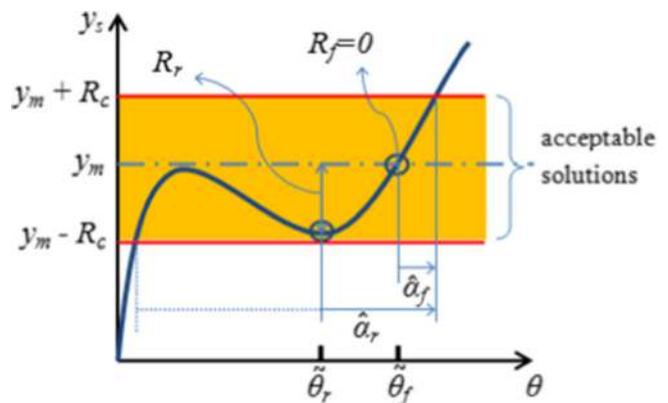


Fig. 1 Notional illustration of robustness and fidelity for a one-dimensional problem

fidelity-optimal parameter value ($R_f \gg R_r$). As evident in Fig. 1, the fidelity- and robust-optimal solutions would converge if the relationship between the θ parameter and the y_s output was linear. Hence, robust-optimal calibration exploits the nonlinear relationship between model inputs and outputs.

The multi-objective model calibration methodology implemented herein can then be expressed as:

For a given R_c ,
 Find a population of $\tilde{\theta}$ to optimize : $R(\tilde{\theta})$ and $\hat{\alpha}(\tilde{\theta})$

$$(3)$$

where $\tilde{\theta}$ are the calibrated, best estimate model parameters to be used as the nominal values in the robustness analysis. Note that in the formulation given in Eq. (3), error between measurements and predictions at the best estimate parameter values, $R(\tilde{\theta})$ is minimized, while robustness, $\hat{\alpha}(\tilde{\theta})$ is maximized. Satisfying these two objectives will lead to multiple plausible solutions for $\tilde{\theta}$. Further discussion on this concept will be provided later.

The difference between fidelity-optimal and robust-optimal parameter calibration is further illustrated in Fig. 2a, which notionally plots prediction error, R against uncertainty in model parameters, α , where low prediction error indicates high fidelity. The fidelity-optimal model shown in Fig. 2 yields the smallest possible error when $\alpha=0$ indicated with R_f , but can only tolerate a small amount of uncertainty, before it fails to satisfy the fidelity threshold, R_c . On the other hand, the robust optimal model, which tolerates larger uncertainties, $\hat{\alpha}_r$ compared to the fidelity-optimal model yields higher nominal errors, R_r . Therefore, R_f and $\hat{\alpha}_r$ define the bounding limits for maximum fidelity (i.e., minimum error) and robustness obtained in the model predictions y_s for a given R_c . The most desirable outcome of course would be to reach $\hat{\alpha}_r$ and R_f simultaneously, what is widely known as the utopia point in multi-objective optimization. When objectives are conflicting as in this case, what is obtained instead is a family of competing solutions as shown in Fig. 2b. Once these competing solutions are obtained, a decision maker may still need to determine the preferred solution (Deb and Gupta 2010), i.e.

the model that strikes a balance between fidelity and robustness.

We must emphasize that herein $\hat{\alpha}_r$ is determined for a given R_c . In info-gap analysis, one need not select the fidelity threshold, R_c , a priori. Instead, the selection of R_c depends on the trade-off between model robustness and a given fidelity threshold (Hemez and Ben-Haim 2004). Thus, the model that minimizes the shaded area shown in Fig. 2c would yield high robustness considering all values of model fidelity bounded by the predefined fidelity threshold, R_c . Minimization of the shaded area is attractive as it strives to find a solution that is robust for not only a specific fidelity level defined by the threshold R_c but also for all fidelity levels that satisfy R_c . This approach will be implemented in our research and detailed in the remainder of the paper.

3 Background perspectives

In this section, approaches taken to achieve single objective solutions to model calibration, i.e. fidelity-optimal and robust-optimal solutions, will first be presented, followed by a discussion of the multi-objective optimization technique implemented herein.

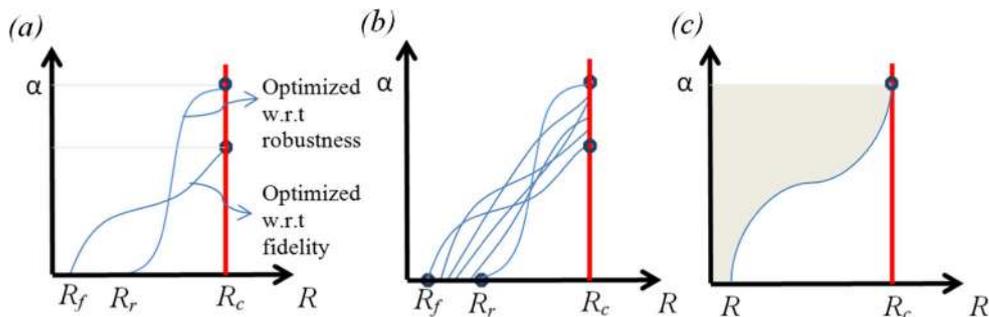
3.1 Calibration of imprecise and inaccurate numerical models

The model discussed in the previous section, $y_s(\theta)$, can be generalized to reflect control parameters, \mathbf{x} , that define the operational domain of the model, the domain within which the numerical model is executed in a predictive capacity:

$$y_s(\mathbf{x}, \theta) \quad (4)$$

Control parameters define the operational and environmental settings. The range assigned for the control parameters define the domain of applicability. It is within this domain the engineering system operates and the computer model of interest is used to make predictions. Hence, for model calibration, experiments must be conducted to explore the domain of applicability. There is a clear distinction between control parameters, \mathbf{x} and calibration parameters, θ in that the former

Fig. 2 (a) $\alpha-\hat{R}(\alpha)$ curves for fidelity optimal and robustness optimal model (b) $\alpha-\hat{R}(\alpha)$ curves for a family of alternative models (c) The shaded area is to be minimized while searching for preferable compromise between fidelity and robustness objectives (Note: Within this figure, α indicates the absolute value)



can be controlled by the experimentalist and is assumed to be known with certainty.

Provided that uncertainties in numerical solutions and experimental measurements are negligible, disagreements between the model predictions and experimental measurements can primarily be attributed to two factors: parameter uncertainty and model bias. In model calibration, parameter uncertainty is typically addressed by adjusting model parameter values to reduce the disagreement between experimental measurements and simulation either deterministically (Ma and Abdulhai 2001; Zhang et al. 2009) or stochastically (Campbell 2006; Higdon et al. 2004). Moreover, model bias, also known as model form error, arises due to simplified or omitted representation of physical phenomena during the development of the numerical model (Draper 1995). In model calibration, model bias is typically addressed by empirically training a mathematical error model to bias correct the predictions (Kennedy and O’Hagan 2001; Higdon et al. 2007; Atamturktur et al. 2011). Considering both parameter uncertainty and model bias, experimental measurement, y_m , can then be related to simulation result, y_s , through the following equation:

$$y_m(\mathbf{x}_i) = y_s(\mathbf{x}_i, \mathbf{t}) + \psi(\mathbf{x}_i) + \tau(\mathbf{x}_i), \quad i = 1, \dots, n \quad (5)$$

where n is the number of experiments, \mathbf{t} is the true but unknown vector of values for model parameters, $\boldsymbol{\theta}$, $\psi(\mathbf{x}_i)$ represents model bias, and $\tau(\mathbf{x}_i)$ represents experimental error, commonly defined by a zero-mean Gaussian random variable. In Eq. (5), the symbol \mathbf{t} drops out of y_m as \mathbf{t} becomes a constant (true) value during experiments. The implication of Eq. (5) is that the response of the physical system can be predicted more accurately by bias-correcting the inexact simulation model, provided that model bias can be estimated at settings where experiments are unavailable.

Consider a hypothetical model, y_s with one model parameter that is set at its best estimate value, $\tilde{\theta}$. Figure 3a illustrates the incompleteness of this hypothetical model by comparing the model predictions to five available physical experiments, where it can be clearly seen that this model is unable to match the experiments throughout the domain. One possible explanation for this disagreement might be the use of an incorrect value for θ (i.e. miscalibration of the model). In Fig. 3b, an ensemble of model predictions is obtained by selecting alternative values of θ from a range of possibilities. When this ensemble of predictions is compared to the five available physical experiments, it becomes evident that no θ value is correctly predicting *all* of the available five experiments. Thus, the numerical model is inadequate and possibly missing the necessary parameters to fully describe the physical response of interest. Figure 3c shows that this model incompleteness results in systematic bias in model predictions. If this model inadequacy is not taken into

account while calibrating the uncertain model parameters, parameters might be calibrated to mathematically viable but physically incorrect values that compensate for model bias. Draper (1995) emphasized that such compensating effects typically lead to over-confidence in model predictions.

At tested settings of control parameters, \mathbf{x}_i , model bias, $\psi(\mathbf{x}_i)$, can be calculated as the difference between $y_m(\mathbf{x}_i)$ and $y_s(\mathbf{x}_i, \tilde{\boldsymbol{\theta}})$. For untested settings, however, the model form error is unknown and an independent error model must be trained (Farajpour and Atamturktur 2013). This discrepancy model is an approximation of the true model bias, $\psi(\mathbf{x})$, obtained by fitting a purely mathematical model to the known values of model form error at the tested settings. Symbolically, the discrepancy model is expressed as, $\delta(\mathbf{x}, \boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ represents a vector of coefficients that describes the discrepancy model. These coefficients must be identified during model calibration. Note that the calibrated values of $\boldsymbol{\gamma}$ are dependent upon the calibrated, best-estimate parameter values of the simulation model, $\tilde{\boldsymbol{\theta}}$. With discrepancy model trained, best-estimate predictions, $\hat{\zeta}(\mathbf{x})$, can then be obtained by summing the calibrated model predictions with those of the trained discrepancy model:

$$\hat{\zeta}(\mathbf{x}) = y_s(\mathbf{x}, \tilde{\boldsymbol{\theta}}) + \delta(\mathbf{x}, \boldsymbol{\gamma}) \quad (6)$$

All too often, model bias is neglected during parameter calibration, possibly leading to questionable $\tilde{\boldsymbol{\theta}}$ values. It is however important to evaluate the role of model bias while calibrating the model parameters. Only when the model bias is determined sufficiently small, its omission can be justified. Later in this paper, we will revisit the role of model bias in model calibration.

3.2 Info-gap decision theory

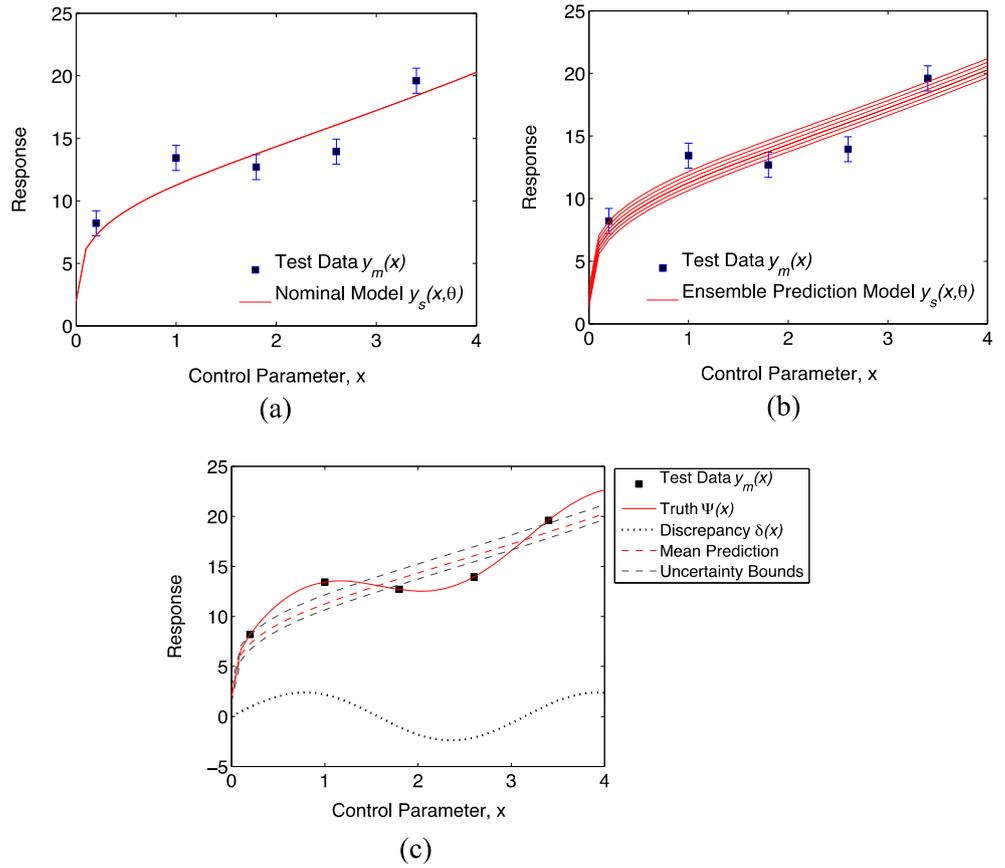
Herein, lack of knowledge regarding the precise values of the model parameters is represented with an info-gap uncertainty model (Ben-Haim 2006), specifically the envelope model. In this study, the uncertainty in parameters, α , also known as the horizon of uncertainty, is measured by the ratio of the absolute difference between the uncertain parameters, $\boldsymbol{\theta}$, and their nominal values, $\tilde{\boldsymbol{\theta}}$, to the nominal values, as given in Eq. (7).

$$U(\alpha, \tilde{\boldsymbol{\theta}}) = \left\{ \boldsymbol{\theta} : \left| \frac{\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}}{\tilde{\boldsymbol{\theta}}} \right| \leq \alpha \right\}, \alpha \geq 0 \quad (7)$$

Therefore, in the envelope info-gap model used herein, the horizon of uncertainty represents the uncertainty introduced to the nominal value as a fraction.

Note that in the context of the present work, $\tilde{\boldsymbol{\theta}}$ is the vector of best estimate values obtained through parameter calibration to reduce the disagreement between model predictions and measurements. As uncertainty is introduced to this best

Fig. 3 (a) Comparison between experiments and nominal model predictions (b) Ensemble of model predictions $y_s(x, \theta)$ obtained with $-1 \leq \theta \leq +1$ (c) Comparison of prediction statistics, “truth” function, and model form error



estimate value, agreement between model predictions and measurements may deteriorate. This deterioration is quantified by searching for the model with the worst predictive ability for a given horizon of uncertainty, α . The function that yields the worst case model predictions for increasing values of horizon of uncertainty can be calculated as follows:

$$\hat{R}(\alpha) = \max_{\theta \in U(\alpha, \tilde{\theta})} R(\theta) \quad (8)$$

where there will be a separate function for each calibrated $\tilde{\theta}$.

As shown in Fig. 4, $\hat{R}(\alpha)$ is a monotonic, non-decreasing function of α . As α increases, uncertain parameters are allowed to vary within a larger range, thereby encompassing a new worst-case value of $\hat{R}(\alpha)$. For a given threshold fidelity, R_c , robustness, $\hat{\alpha}$, consistent with the manner introduced earlier in this paper, can then be calculated as follows:

$$\hat{\alpha} = \max \left\{ \alpha : \hat{R}(\alpha) \leq R_c \right\} \quad (9)$$

Conceptually, $\hat{\alpha}$ is the maximum horizon to which the info-gap uncertainty model is allowed to expand before the $\hat{R}(\alpha)$ exceeds the critical performance value, R_c (see Fig. 4). The larger $\hat{\alpha}$ is, the larger the amount of uncertainty the model can tolerate and thus the greater the robustness. Hence, the horizon of uncertainty plotted against worst case model

prediction as shown in Fig. 4 (and shown earlier in Fig. 2) is referred to as robustness function. It should be clear that $\hat{R}(\alpha)$ supplies the inverse of the robustness function.

3.3 Multi-objective optimization

The approach proposed in the previous section for model calibration naturally leads to a multi-objective optimization problem with two distinct objectives: improving the fidelity and robustness of the model predictions. When these two objectives conflict, a single solution that optimizes both of these objectives would not exist in the solution space. However, a set of solutions, referred to as a Pareto front, that are better than all other solutions can be obtained. In Fig. 5, a

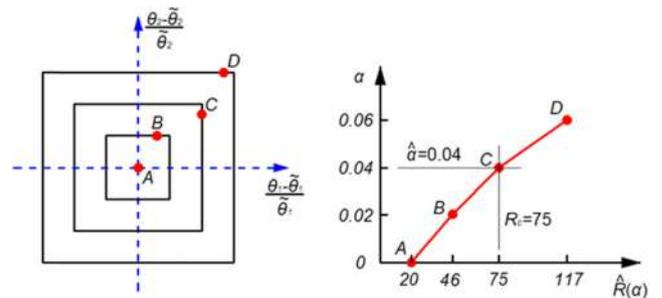
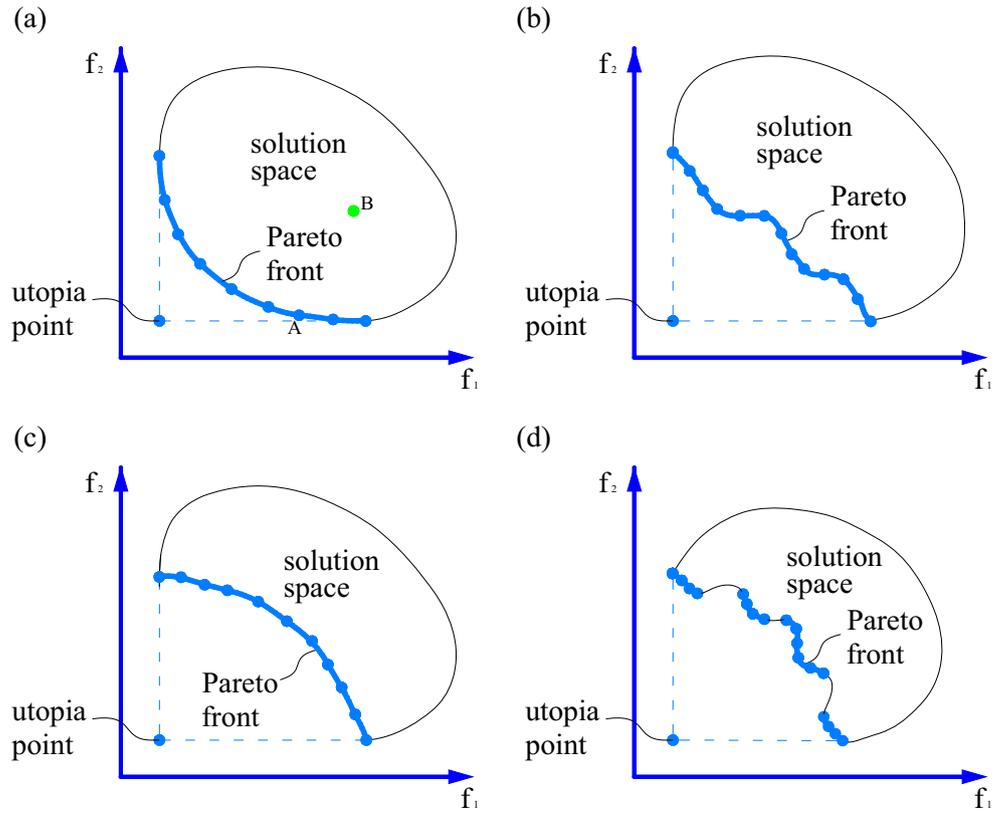


Fig. 4 Illustration of Info-gap robustness

Fig. 5 Plausible forms of Pareto fronts that can be obtained in a bi-objective space



few representative Pareto fronts are shown for two-objective optimization problems. The shape of the Pareto front visually displays the level of compromise needed from one criteria to improve other and hence, reflects the nature and complexity of the tradeoff between multiple objectives of model calibration. Of course, if the fidelity and robustness objectives do not conflict for a particular problem then the optimum solutions for both objectives would merge at a single point in the solution space.

In this study, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) is implemented because of its proven capability in identifying the Pareto front efficiently while maintaining solution diversity (Deb et al. 2002; Cagkan et al. 2006). NSGA-II is a variant of the Genetic Algorithm, with the essential part of non-dominated sorting. In the context of NSGA-II, the Pareto front presents a set of solutions that are not dominated by all other designs. The domination relation is defined as follows (Marler and Arora 2004): B is dominated by A if A is not inferior to B in all objectives, and A is superior to B in at least one objective (Fig. 5). As for a two-objective case, B is dominated by A if only one of the following criterions is satisfied: 1) A is better than B in both objectives or 2) A is better than B in one objective, but A equals B in the second objective. If one solution is not dominated by any other solution, it belongs to the Pareto front.

In NSGA-II, the Genetic Algorithm operators, such as selection, recombination and mutation are used to guide the

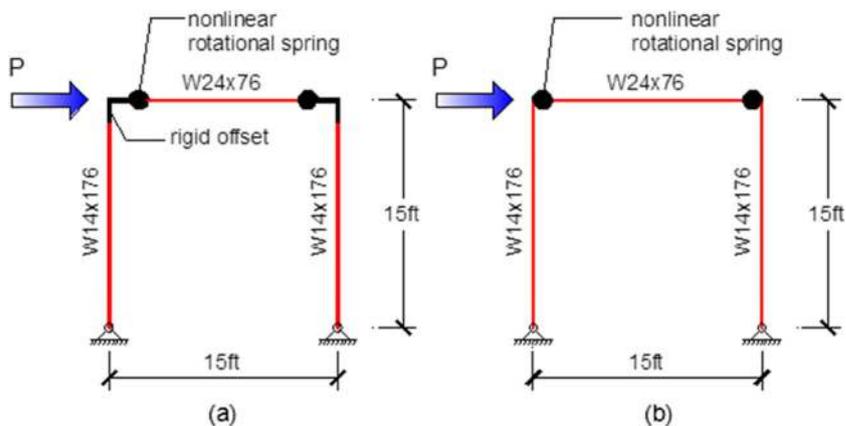
evolution of populations (i.e. group of solutions) from generation to generation until the solution reaches an optimum state. In our study, a population size of 50 is used for each generation with the Pareto front acquired after 100 generations. The intermediate method is selected as a crossover operator with a crossover ratio of 1.2; the Gaussian method is selected as a mutation operator with scale and shrink coefficients set to 0.1 and 0.5, respectively (Lobo et al. 2007).

Being a population-based optimization method, NSGA-II operates on a group of solutions rather than one solution at a time thus efficiently yielding the Pareto front in a single run. Of course, many other multi-objective optimization methods can also be implemented to derive the Pareto front, a comprehensive survey of which is provided by Marler and Arora (2004).

4 Case study

The proposed model calibration methodology is demonstrated on the plastic deformation of a steel portal frame with pinned supports shown in Fig. 6 using the object-oriented, open source software framework, The Open System for Earthquake Engineering Simulation (OpenSees). Beam plasticity is represented using an elastic beam with nonlinear rotational springs at both ends. The nonlinear rotational

Fig. 6 (a) Hypothetical accurate model with precise IK model parameter values used to generate synthetic experimental data (b) Inaccurate model missing rigid offset with unknown parameter values



springs are modeled using the Ibarra Krawinkler (IK) model (Ibarra et al. 2005) (Fig. 6).

The IK model used to represent the nonlinear rotational spring shown in Fig. 7 requires the definition of five parameters (Lignos and Krawinkler 2009):

- (1) K_i , the initial stiffness of the linear segment;
- (2) M_y , the moment at which the beam begins to yield;
- (3) β_1 , the strain hardening ratio, a ratio of the pre-capping stiffness to K_i ;
- (4) β_c , the post-capping hardening ratio, a ratio of the post-capping stiffness to K_i ;
- (5) M_c/M_y , which is the ratio of the moment at the capping point (M_c) to yielding moment.

Figure 6a shows the hypothetical ‘accurate’ model that considers a rigid offset and uses the hypothetical ‘precise’ values of the five IK model parameters (listed in Table 1). Using this model, synthetic experimental data is generated through push-over analysis where the load P is applied in 25 steps until 10 % drift is reached. Applied force at each increment of 0.4 % drift is recorded to obtain drift-force curves as shown in Fig. 8.

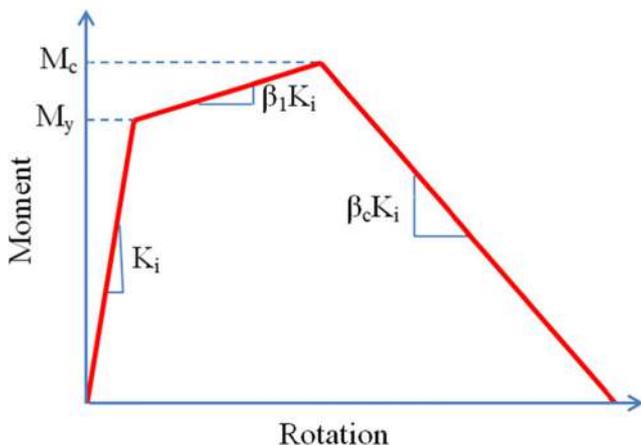


Fig. 7 Ibarra Krawinkler deterioration model (adopted from Lignos and Krawinkler 2011)

Two different scenarios are evaluated. In the first scenario, the model developer is assumed to be completely knowledgeable about the physical nature of the problem in that an exact model form is assumed to be available. In the second scenario, the model developer, is assumed to neglect the rigid-offset necessary to model the panel zone (resulting in model bias). In either scenario, the model developer is assumed to be unaware of the precise values of IK model parameters but knowledgeable about their plausible ranges (resulting in parameter uncertainty). The ranges for these parameters are taken from those suggested by Lignos and Krawinkler (2011) (see Table 1). Hence, the problem in hand is one of calibrating the poorly-known model parameters (i.e. IK connection parameters) and when necessary, bias correcting the model prediction to account for model incompleteness (i.e. lack of rigid offset to model panel zone).

In the following sections, the proposed multi-objective optimization-based methodology to model calibration is demonstrated on three different schemes considering varying knowledge levels of the model developer:

- 1) An accurate model with imprecise parameters, where only model parameters are calibrated (i.e., model developer has knowledge of the correct model form);
- 2) An inaccurate model with imprecise parameters, where model parameters are calibrated and a discrepancy model is trained (i.e. model developer does not have the knowledge of the correct model form, but is aware of the model’s inadequacy);

Table 1 Precise values and plausible ranges for the five IK model parameters

	K_i (kip*in/rad)	β_1	β_c	M_y (kip*in)	M_c/M_y
Exact value	2000000	0.03	-0.05	12000	1.2
Lower bound	1800000	0.02	-0.08	10000	1
Upper bound	2200000	0.04	-0.03	14000	1.4

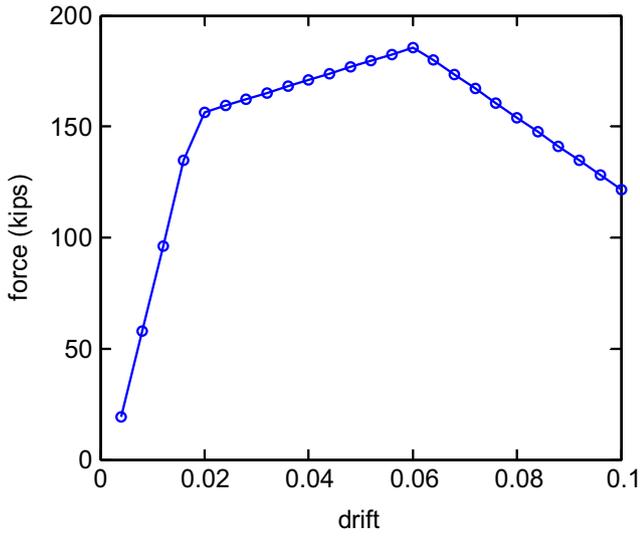


Fig. 8 Drift-force curve obtained using the precise and accurate (i.e. true) model, also referred to as synthetic experiments

- 3) An inaccurate model with imprecise parameters, where only model parameters are calibrated and discrepancy is neglected (i.e. model developed does not have the knowledge of the correct model form, and is not aware of the model's inadequacy).

For all three schemes, NSGA-II is employed to solve the proposed multi-objective calibration problem simultaneously considering fidelity and robustness of the predictions. A total of 25 measurements for the drift at varying load levels are assumed available for model calibration (recall Fig. 8).

4.1 Scheme 1: exact model form with poorly known parameters

The proposed calibration methodology is first deployed on an accurate model without the need to consider prediction bias. Therefore, $R(\theta)$ is defined as follows:

$$R(\theta) = \sqrt{\sum_{i=1}^k (y_m(x_i) - y_s(x_i, \theta))^2} \quad i = 1, \dots, n \quad (10)$$

where y_s denotes the plastic deformation predictions of the OpenSees model of the steel portal frame with pinned supports, and n denotes the number of experiments. Here, an accurate model i.e., the model with panel zone is considered. For each set of calibrated model parameter values, $\hat{R}(\alpha)$, is calculated considering fractional uncertainty in parameter values between 0 and 0.1 in steps of 0.01. For a threshold fidelity value (R_c of 12 kips), info-gap robustness, $\hat{\alpha}$, is then determined as the α value of the intersection point of the $\alpha - \hat{R}(\alpha)$ curve with $\hat{R}(\alpha) = R_c$. Multi-objective optimization as previously presented in Eq. (3) is solved to optimize both fidelity and robustness of model predictions. Figure 9 presents

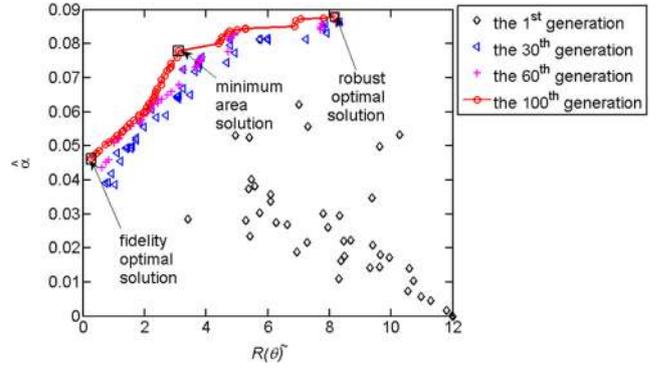


Fig. 9 Distribution of robustness and fidelity of candidate models at selected generations for Scheme 1

the evolution of the fidelity and robustness of model predictions obtained through NSGA-II generations.

As evident in Fig. 9, solutions of the first generation perform poorly in both $\hat{\alpha}$ and $R(\hat{\theta})$, however solutions evolve from generation to generation and ultimately reach the converged Pareto front yielding models with improved robustness and fidelity. Within the Pareto front solutions, the tradeoff relationship between $\hat{\alpha}$ and $R(\hat{\theta})$ can be clearly observed in Fig. 9, where the model with larger $\hat{\alpha}$ exhibits a greater $R(\hat{\theta})$, and the model with smaller $R(\hat{\theta})$ exhibits a lesser $\hat{\alpha}$.

Evaluating the two extremes, the fidelity-optimum solution has an $R(\hat{\theta})$ of 0.2552 and an $\hat{\alpha}$ of 0.0457, while the robustness-optimum solution has an $R(\hat{\theta})$ of 7.804 and an $\hat{\alpha}$ of 0.0889. As shown, though the fidelity-optimal model yields the smallest nominal errors, it tolerates the smallest amount of uncertainty before the fidelity threshold, R_c , is violated. While the robust-optimal model can tolerate the maximum amount of uncertainties, it yields the largest nominal errors.

In Fig. 10, the fidelity- and robust-optimum solutions are visually compared against all other Pareto front solutions by overlapping the $\alpha - \hat{R}(\alpha)$ curves. As discussed earlier, among all Pareto front solutions finding an alternative model that strikes a compromise between the two objectives is possible

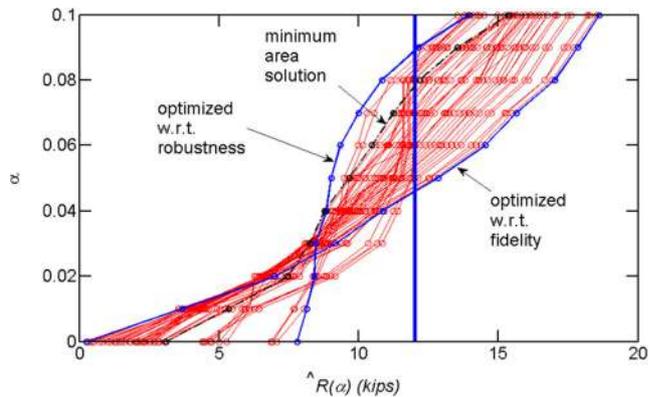


Fig. 10 $\alpha - \hat{R}(\alpha)$ curves for obtained Pareto front solutions for Scheme 1

by minimizing the area above the $\alpha\text{-}\widehat{R}(\alpha)$ curve, referred henceforth as area-optimum model. This area-optimum solution, shown in Fig. 10, has an $R(\widetilde{\theta})$ of 3.105 (a roughly 60 % reduction in $R(\widetilde{\theta})$ compared to the robust-optimal model), an $\widehat{\alpha}$ of 0.078 (a roughly 70 % gain in $\widehat{\alpha}$ compared to the fidelity-optimal model).

The calibrated parameter values, $\widehat{\alpha}$, $R(\widetilde{\theta})$ and area above the $\alpha\text{-}\widehat{R}(\alpha)$ curve are listed in Table 2 for the three optimality conditions (i.e. fidelity-, robustness- and area-optimum). Although the model developer is assumed in this scheme to be perfectly knowledgeable of the model form (which is rarely the case in practical applications), model calibration fails to yield the ‘true’ values for the parameters due to the unavoidable compensations between five calibration parameters. It must be noted that the fidelity-optimum model yields the highest errors (up to 25 %) in the five calibrated IK parameters compared to the robust-optimum and area-optimum solutions.

4.2 Scheme 2: inexact model form with poorly known parameters (with bias-correction)

For the second scheme, the model developer is assumed to disregard the panel zones thereby introducing inaccuracy, or model bias. However, the model developer is assumed to be aware of the presence of such bias. Hence, to account for the resulting model incompleteness, a discrepancy model is trained to bias correct the inaccurate model. Here, $R(\theta)$ is defined as follows:

$$R(\theta) = \sqrt{\sum_{i=1}^k \left(y_m(\mathbf{x}_i) - y_s(\mathbf{x}_i, \theta) - \delta(\mathbf{x}_i, \gamma | \theta) \right)^2} \quad i = 1, \dots, n \quad (11)$$

where y_s denotes the model prediction based on the inaccurate model, and n denotes the number of experiments.

Fidelity is inversely related to the prediction error, which is defined as the difference between measurements, y_m and calibrated and bias corrected predictions, $\widehat{\zeta}(\mathbf{x})$ (see Eq. (6)). While training the discrepancy model for bias correction, the procedure described in Farajpour and Atamturktur (2013) is followed by gradually increasing the order of the polynomial emulator and observing the calibration (at the training points) and generalization (at other than the training points) errors. Accordingly, a 7th order polynomial model is trained to approximate the discrepancy bias. Consistent with the previous scheme, an R_c level of 12 kips is used. The evolution of the NSGA-II algorithm obtained for this scheme is shown in Fig. 11, and the $\alpha\text{-}\widehat{R}(\alpha)$ curves for the Pareto front solutions are plotted in Fig. 12. The calibrated parameter values, $\widehat{\alpha}$, $R(\widetilde{\theta})$ and area above the $\alpha\text{-}\widehat{R}(\alpha)$ curve are listed in Table 3 for the three optimality conditions (i.e. fidelity-, robustness- and area-optimum).

Consistent with the observations of Scheme 1, Fig. 12 demonstrates that the fidelity-optimum model yields the smallest nominal errors, but can tolerate the smallest amount of uncertainty before R_c is violated. The robust-optimal model, which can tolerate the maximum amount of uncertainty, yields the largest nominal errors. Hence, calibrating the model parameters for only one objective leads to a calibrated model that performs worst in the other objective. However, the area-optimum model strikes a balance between $R(\widetilde{\theta})$ and $\widehat{\alpha}$ in that a 31 % gain in fidelity is achieved at the expense of only a 6 % loss in robustness. Once again the compensating effects result in calibrated model parameter values deviating from their hypothetical true values (up to 15 % for the robust- and area-optimum models). It is important to note that no optimality criterion offers a remedy to fully prevent the compensation between calibration parameters as well as the emulator trained to represent model bias.

Table 2 Properties of the calibrated models for Scheme 1 evaluated for fidelity optimum, robustness optimum and area optimum solutions

	Ki (kip*in/ rad)	β_1	β_c	My (kip*in)	Mc/My	$\widehat{\alpha}$	$R(\widetilde{\theta})$	Area (kips)
Exact value	2000000	0.03	-0.05	12000.00	1.20	-	-	-
	Fidelity optimum							
Calibrated value	2047470	0.032	-0.039	12283	1.261	0.0457	0.2552	0.293
Percent difference	2.35 %	6.45 %	-24.72 %	2.33 %	4.96 %	-	-	-
	Robustness optimum							
Calibrated value	1988312	0.0293	-0.045	12086	1.202	0.0889	7.804	0.272
Percent difference	-0.59 %	-2.36 %	-10.53 %	0.71 %	0.17 %	-	-	-
	Area optimum							
Calibrated value	1992300	0.0297	-0.042	12181	1.233	0.078	3.105	0.251
Percent difference	-0.39 %	-1.01 %	-17.39 %	1.50 %	2.71 %	-	-	-

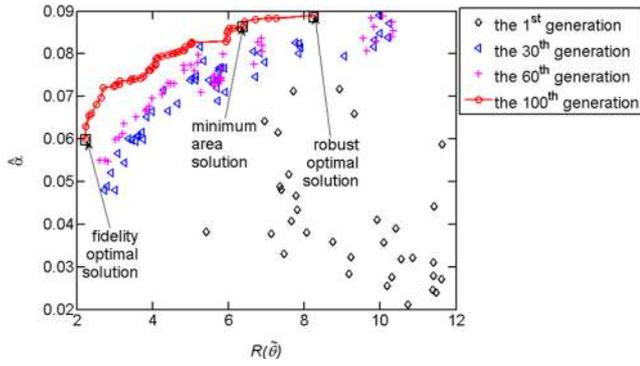


Fig. 11 Distribution of robustness and fidelity at selected generations for Scheme 2

4.3 Scheme 3: inexact model form with poorly known parameters (without bias-correction)

As in the second scheme, the model developer in the third scheme is assumed to neglect the panel zones resulting in model bias. Unlike the previous scheme however, the model developer is assumed to be unaware of model incompleteness. Hence, here a discrepancy model is not trained, meaning that the calibrated model parameters, θ , are likely to compensate for the model bias. $R(\theta)$ is then defined as follows:

$$R(\theta) = \sqrt{\sum_{i=1}^k (y_m(\mathbf{x}_i) - y_s(\mathbf{x}_i, \theta))^2} \quad i = 1, \dots, n \quad (12)$$

where y_s denotes the model prediction based on the inaccurate model and n denotes the number of experiments.

Since the model predictions are not bias corrected in this scheme, the disagreement between model predictions and measurements is significantly higher compared to the first two schemes in that no solution can be found that satisfies the fidelity threshold of 12 kips (R_c value used in schemes 1 and 2). Therefore, an R_c value of 150 kips is defined in this scheme with the objective being to obtain a number of plausible solutions similar to the first two schemes evaluated

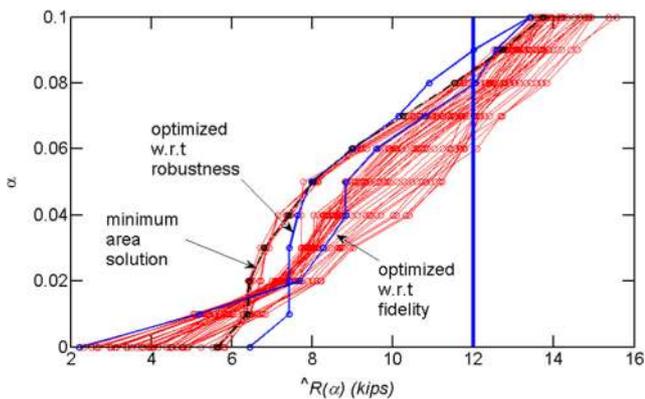


Fig. 12 $\alpha-\hat{R}(\alpha)$ curves for obtained Pareto front solutions for Scheme 2

herein. As before, the evolution process of $\hat{\alpha}$ and $R(\tilde{\theta})$ is shown in Fig. 13. The $\alpha-\hat{R}(\alpha)$ curves for the Pareto front solutions are plotted in Fig. 14. The calibration parameters $\hat{\alpha}$, $R(\tilde{\theta})$ and area above the $\alpha-\hat{R}(\alpha)$ curve are listed in Table 4 for fidelity-optimum, robustness-optimum and area-optimum solutions.

Compared to the first two schemes, the calibration parameters deviate significantly more from the exact values. This is expected since the best estimate predictions are not bias-corrected resulting in the model error being lumped into parameter errors. This observation is true for all optimality conditions, where the deviations from true values are as high as 50 % for the robust- and area-optimum model. A close look at the calibrated model parameters reveals that the fidelity- and robust-optimal models reach two significantly different sets of parameters. The area-optimal model presents yet another alternative solution that can be said to outperform the robust-optimal model with the exception of K_i . However, none of the optimality conditions entirely prevent calibration parameters from compensating for each other as well as for the inherent bias in model predictions. This observation demonstrates the fundamental flaw in the prevailing engineering paradigm of selecting a single, optimum model in model calibration.

5 Intended use of the model

The proposed multi-objective optimization-based model calibration approach can be extended to incorporate other objectives, such as the intended use of the model. Computer models are typically developed with the intent to predict a response (either the same as or different from the calibration experiments) at untested settings or configurations. Hence, experiments are rarely available to determine the fidelity or robustness of the model specifically for its intended use. While evaluating the performance of a model to fulfill its intended use in cases where experiments are unavailable, the *self-consistency* of model predictions supplies an alternative (Ben-Haim and Hemez 2012). The term self-consistency implies the tightness (or lack of variability) of the model predictions when the uncertainties in model parameters are considered.

For demonstration purposes, the model calibration problem is configured as a three-objective optimization with fidelity, robustness and self-consistency as objectives. The aim herein is to maximize the self-consistency of the model predictions for the intended use of the model. The model developed to analyze the portal frame, shown earlier in Fig. 6, is utilized and the intended use is assumed to be the prediction of the spring rotation (IK model). Self-consistency is calculated considering the predictions of the model at untested settings using the calibrated parameter values, $\tilde{\theta}$, and the associated

Table 3 Properties of the calibrated models for Scheme 2 evaluated for fidelity optimum, robustness optimum and area optimum solutions

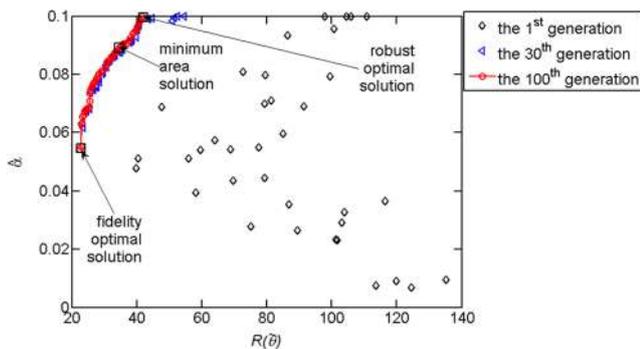
	Ki (kip*in/ rad)	β_1	β_c	My (kip*in)	Mc/My	$\hat{\alpha}$	$R(\hat{\theta})$	Area (kips)
Exact value	2000000	0.03	-0.05	12000	1.20	-	-	-
Fidelity optimum								
Calibrated value	2047470	0.03	-0.045	12176	1.232	0.06	2.243	0.201
Percent difference	2.35 %	0.00 %	-10.53 %	1.46 %	2.63 %	-	-	-
Robustness optimum								
Calibrated value	1977420	0.0291	-0.043	12047	1.176	0.089	8.263	0.191
Percent difference	-1.14 %	-3.05 %	-15.05 %	0.39 %	-2.02 %	-	-	-
Area optimum								
Calibrated value	1994300	0.0298	-0.043	12082	1.212	0.087	6.375	0.124
Percent difference	-0.29 %	-0.67 %	-15.05 %	0.68 %	1.00 %	-	-	-

robustness, $\hat{\alpha}$. The NSGA-II algorithm is employed with 200 individuals in the population and is executed for 200 generations.

The trade-offs between self-consistency and the fidelity of the model for a predefined desirable robustness level are shown in Fig. 15. In this figure, there are three distinct zones. In the first zone, a small but rapid improvement in self-consistency can be obtained by incurring an insignificant loss in fidelity. In the second zone, an insignificant gain in self-consistency can be had at the expense of a large loss in fidelity. Finally, a third zone exists in which a significant gain in self-consistency can be achieved with a minimal loss of fidelity. For instance, if the decision maker decides to stay in the first zone, among models with identical robustness and almost identical fidelity, it would be logical to justify the selection of the model with the highest self-consistency. Of course, model calibration must be repeated if the intended use of the model changes.

6 Conclusions

The non-unique solutions that are commonly obtained through optimization-based model calibration, where different combinations of calibration parameters reproduce experimental data

**Fig. 13** Distribution of robustness and fidelity at selected generations for Scheme 3

with similar fidelity, clearly demonstrate that model fidelity is an incomplete metric for parameter calibration. The findings obtained through the case study structure reveal that optimizing calibration parameters solely to maximize fidelity to measurements would yield an arbitrary solution from a family of plausible sets of parameter values. Hence, in model calibration, the compensation between the parameters is an inevitable issue even with the absence of bias in the model predictions. This finding is demonstrated in Scheme 1 where the calibrated parameters of the exact model converge to values that are different than the true values. When bias is presented in the model predictions, which is represented in Schemes 2 and 3, the compensating effects are reduced when an independent error model is trained to consider the bias (as evident in Scheme 2) compared to the case when model incompleteness is neglected (as observed in Scheme 3). The findings of this paper demonstrate that the prevailing engineering paradigm of selecting a single, fidelity-optimal model in calibration is fundamentally flawed due to compensating effects, especially when the model incompleteness is neglected.

An alternative approach to calibrating a model with respect to its fidelity is calibrating it with respect to its robustness. In this approach, a model that satisfies a certain fidelity threshold

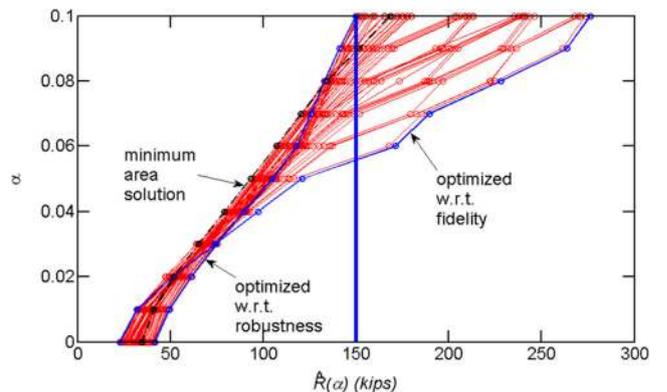
**Fig. 14** $\alpha-\hat{R}(\alpha)$ curves for obtained Pareto front solutions for Scheme 3

Table 4 Properties of the calibrated models for Scheme 3 evaluated for fidelity optimum, robustness optimum and area optimum solutions

	Ki (kip*in/ rad)	β_1	β_c	My (kip*in)	Mc/My	$\hat{\alpha}$	$R(\hat{\theta})$	Area (kips)
Exact value	2000000	0.03	-0.05	12000	1.20	-	-	-
Fidelity optimum								
Calibrated value	2199360	0.0343	-0.051	13124	1.186	0.0558	22.59	2.0533
Percent difference	9.49 %	13.37 %	1.98 %	8.95 %	-1.17 %	-	-	-
Robustness optimum								
Calibrated value	2035200	0.04	-0.03	13973	1.0702	0.1	41.69	1.3086
Percent difference	1.74 %	28.57 %	-50.00 %	15.19 %	-11.44 %	-	-	-
Area optimum								
Calibrated value	2190690	0.0398	-0.03	12892	1.167	0.0891	34.71	0.7237
Percent difference	9.10 %	28.08 %	-50.00 %	7.17 %	-2.79 %	-	-	-

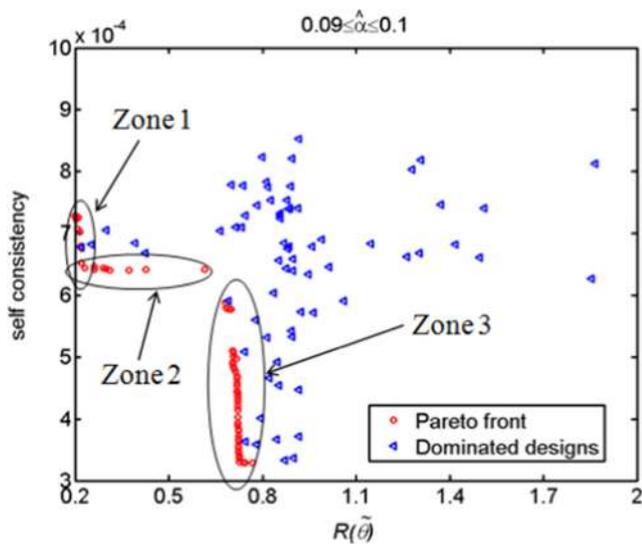
and maximizes the robustness against uncertainties in model parameters is sought. In this study, such robust-optimal solutions are observed to yield the most truthful model parameters for the case with no inherent bias. However, the presence of model bias is observed to increase the severity of compensating effects in robustness-based calibration where obtained solutions deviate significantly from their true parameter values. In cases where model incompleteness is remedied through bias correction, deviations from true parameter values are seen to be as high as 15 %; however when model incompleteness is neglected altogether during model calibration, the deviations reach as high as 50 %. Moreover, the results obtained in this study demonstrate that the selected robust-optimal model is strictly dependent upon the predefined fidelity threshold, R_c , and is not guaranteed to be optimal for all fidelity levels that satisfy the threshold requirement ($R < R_c$).

The Pareto front solutions obtained in this study demonstrate that calibrating a model strictly for one objective

(fidelity or robustness) results in the worst performance in the other objective. Furthermore, the two extreme cases (fidelity-optimum versus robust-optimum) may yield solutions that may greatly sacrifice one objective to marginally improve the other. Thus, it might be necessary to search for a compromise between these two approaches. In that regard, the authors suggest one such approach that aims to maximize the robustness for any fidelity threshold (lower than R_c). These area-optimal solutions are observed to find a preferable trade-off between fidelity and robustness for all three schemes. Other definitions of suitable compromise between fidelity and robustness can also be defined based on the specifics of the problem.

Note that none of the three optimality criterion fully remedy the risk of converging to a set of model parameters that best compensate for each other (and/or for model bias) to minimize a certain objective and only additional dedicated tests would be able to reduce this effect. Consequently, the authors contend that the evaluation of the calibrated model parameter values for all Pareto front solutions is necessary to determine the severity of such compensating effects. The lack of a mechanism to make a clear distinction between family of plausible solution can hence be listed as the major weakness of the multi-objective calibration approach. Future research will focus on investigating the possible clustering of the calibrated model parameter values of Pareto front solutions.

The multi-objective calibration approach can be extended to incorporate additional objectives that may be important for a specific problem. In this paper, this versatility is demonstrated focusing on the self-consistency of the model predictions at settings where experiments are not available. Here, self-consistency is defined to be inversely proportional to the variability of the model predictions obtained with parameter values sampled from a convex domain with a dimension of alpha. With this implementation, calibration parameters that minimize the error between model predictions and experiments (i.e. at tested settings), maximize the robustness of the

**Fig. 15** Configuring model calibration specific to the intended use of the model in a three-objective optimization problem

model predictions to uncertainty in input parameters (i.e. at tested settings) and minimize the spread of the model predictions (i.e. at untested settings).

The drawback of the multi-objective calibration is its high computational demands. The case-study problem implemented herein included a model that was amenable for obtaining a large number of computer runs quickly. The multi-objective calibration still took approximately 50 h on a laptop with i7-3632 QM cpu @ 2.2 GHz for all three schemes with two objective optimization. The multi-objective calibration considering three objectives (as discussed in the section 5) took approximately 3 weeks to complete on the same laptop. Hence, if multi-objective calibration is to be applied for more computationally demanding problems, implementation of a fast-running emulator (also known as surrogate model or meta-model) would be crucial.

References

- Atamturktur S, Hemez F, Williams B, Tome C, Unal C (2011) A forecasting metric for predictive modeling. *Comput Struct* 89(23–24): 2377–2387
- Bekey GA, Masri SF (1983) Random search techniques for optimisation of nonlinear systems with many parameters. *Math Comput Simul* 25:210–213
- Ben-Haim Y (2004) Uncertainty, probability and information-gaps. *Reliab Eng Syst Saf* 85:249–266
- Ben-Haim Y (2006) Info-gap decision theory: decisions under severe uncertainty, 2nd edn. Academic, Oxford
- Ben-Haim Y (2012) Modeling and design of a Hertzian contact: an info-gap approach. *J Strain Anal Eng Des* 47(3):153–162
- Ben-Haim Y, Hemez FM (2012) Robustness, fidelity and prediction-looseness of models. *Proc R Soc A* 468:227–244
- Ben-Haim Y, Cogan S, Sanseigne L (1998) Usability of mathematical models in mechanical decision processes. *Mech Syst Signal Process* 12:121–134
- Berman A (1995) Multiple acceptable solutions in structural model improvement. *AIAA* 33(5):924–927
- Cagkan E, Selin C, Andy D (2006) Multiobjective optimization and evolutionary algorithms for the application mapping problem in multiprocessor system-on-chip design. *IEEE Trans Evol Comput* 10(3):358–374
- Campbell K (2006) Statistical calibration of computer simulations. *Reliab Eng Syst Saf* 91(10–11):1358–1363
- Deb K, Gupta S (2010) Understanding knee points in bicriteria problems and their implications as preferred solution principles. Kanpur Genetic Algorithms Laboratory (KanGAL), KanGAL Report Number 2010005
- Deb K, Pratap A, Agarwal S (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 6(2):182–197
- Draper D (1995) Assessment and propagation of model uncertainty. *J R Stat Soc Ser B* 57(1):45–97
- Farajpour I, Atamturktur S (2013) Error and uncertainty analysis of inexact and imprecise computer models. *J Comput Civ Eng* 27(4): 407–418
- Franchini M, Galeati G, Berra S (1998) Global optimization techniques for the calibration of conceptual rainfall–runoff models. *Hydrol Sci J* 43(3):443–458
- Hemez FM, Ben-Haim Y (2004) Info-gap robustness for the correlation of tests and simulations of a nonlinear transient. *Mech Syst Signal Process* 18:1443–1467
- Higdon D, Kennedy M, Cavendish JC, Cafo J, Ryne RD (2004) Combining field data and computer simulations for calibration and prediction. *SIAM J Sci Comput* 26(2):448–466
- Higdon D, Gattiker J, Williams B, Rightley M (2007) Computer model validation using high-dimensional outputs. In: Bernardo J, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM, West M (eds) *Bayesian statistics*, vol. 8. Oxford University Press, Oxford
- Ibarra LF, Medina RA, Krawinkler H (2005) Hysteretic models that incorporate strength and stiffness deterioration. *Earthquake Eng Struct Dyn* 34:1489–1511
- Kennedy MC, O’Hagan A (2001) Bayesian calibration of computer models. *J R Stat Soc Ser B* 63(3):425–464
- Lignos DG, Krawinkler H (2009) Sidesway collapse of deteriorating structural systems under seismic excitations. Rep. No. TB 172, The John A. Blume Earthquake Engineering Center, Stanford Univ., Stanford, CA
- Lignos DG, Krawinkler H (2011) Deterioration modeling of steel beams and columns in support to collapse prediction of steel moment frames. *ASCE J Struct Eng* 137(11):1291–1302
- Lobo FG, Lima CF, Michalewicz Z (2007) *Parameter setting in evolutionary algorithms*, Springer Publishing Company, ISBN: 3540694315 9783540694311
- Luce CH, Cundy TW (1994) Parameter identification for a runoff model for forest roads. *Water Resour Res* 30(4):1057–1069
- Ma T, Abdulhai B (2001) Genetic algorithm-based combinatorial parametric optimization for the calibration of microscopic traffic simulation models. *Intelligent Transportation Systems Conf. Proc, IEEE, New Brunswick, NJ*, 848–853
- Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for engineering. *Struct Multidiscip Optim* 26(6):369–395
- Mattsson H, Klisinski M, Axelsson K (2001) Optimization routine for identification of model parameters in soil plasticity. *Int J Numer Anal Methods Geomech* 25:435–472
- Nagai A, Kadoya M (1979) Optimization techniques for parameter identification of runoff models. *Annual Report of Disaster Prevention Research Institute in Kyoto Univ.*, 22:(B-2)
- Solomatine DP (1999) Random search methods in model calibration and pipe network design. In: Savic D, Walters G (eds) *Water industry systems. modelling and optimization applications*, vol. 2. Res. Stud, Baldock, pp 317–332
- Zhang X, Srinivasan R, Zhao K, Liew MV (2009) Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. *Hydrol Process* 23(3):430–441