SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 9, No. 1, pp. 231–259

Representing Model Discrepancy in Bound-to-Bound Data Collaboration*

Wenyu Li†, Arun Hegde†, James Oreluk†, Andrew Packard†, and Michael Frenklach†

Abstract. We extend the existing methodology in bound-to-bound data collaboration (B2BDC), an optimization-based deterministic uncertainty quantification (UQ) framework, to explicitly take into account model discrepancy. The discrepancy is represented as a linear combination of finite basis functions, and the feasible set is constructed according to a collection of modified model-data constraints. Formulas for making predictions are also modified to include the model discrepancy function. Prior information about the model discrepancy can be added to the framework as additional constraints. Dataset consistency, a central feature of B2BDC, is generalized based on the extended framework.

Key words. model discrepancy, uncertainty quantification, bound-to-bound data collaboration, inverse problem

AMS subject classifications. 62P35, 68T37

DOI. 10.1137/19M1270185

1. Introduction.

During the past few decades, computational capabilities and data availability have seen substantial growth in many scientific and engineering fields. The growing demand for predictive models with quantifiable uncertainty has developed into an active research area, uncertainty quantification (UQ) [35]. Two principal objectives of UQ are inference of model parameters, also known as the inverse or calibration problem utilizing a set of known data (the *training* set), and model prediction outside such a set. Theories and methods have been developed from both statistical and deterministic perspectives. In the statistical perspective, specifically under a Bayesian inference framework [21], the prior distribution of model parameters is updated by the likelihood of observations through Bayes' theorem. The produced posterior distribution is utilized for model parameter inference and predictions. In the deterministic perspective, for example, using the bound-to-bound data collaboration (B2BDC) method [12, 18, 19, 45] employed in this paper, posterior ranges on model parameters and predictions are obtained through inequality-constrained optimization.

These two perspectives, while complementing each other in some aspects, address essentially the same problems and produce comparable outcomes. Both methodologies emphasize the critical role of identifying the source of uncertainty [17]. B2BDC shares conceptual similarities not only with Bayesian calibration methods but also Bayesian history matching [9] and methods under the set membership framework [8, 20, 33, 44]. Bayesian history matching retains a probabilistic interpretation of the data and defines a nonimplausible region in the parameter space that contains all acceptable parameter vectors. Although both methods

*Received by the editors June 24, 2019; accepted for publication (in revised form) October 14, 2020; published electronically February 11, 2021.
https://doi.org/10.1137/19M1270185
Funding: This work was supported by U.S. Department of Energy, National Nuclear Security Administration, under Award DE-NA0002375.

†Department of Mechanical Engineering, University of California at Berkeley, CA 94720-1740 USA (wenyuli@berkeley.edu, arun.hegde@berkeley.edu, jim.oreluk@berkeley.edu, apackard@berkeley.edu, frenklach@berkeley.edu).

seek a region containing valid parameter vectors, Bayesian history matching uses stochastic emulators and improves the quality of the emulators through iterative updates of the nonimplausible region, whereas B2BDC employs polynomial response surfaces and focuses primarily on evaluation of uncertainty in predictions [18, 19]. Similar to the setup in B2BDC, the set-membership framework in robust control describes uncertainty in prior information and data by constraints. Methods developed under this framework also formulate estimation/prediction of quantities of interest (QOIs) as regions defined by inequality constraints or solutions to optimization problems [33, 52]. Some methods employ simply shaped approximations of the complex regions formed therein, e.g., a bounding ellipsoid [14, 44] or a minimum-volume bounding parallelotope [8], whereas others approximate the target region with a bounding nonconvex geometry whose quality can be refined iteratively [7, 28, 38]. B2BDC uses polynomial surrogate models and handles the resulting nonconvexity through convex relaxation, leading to global guarantees on optimality [45]. A previous B2BDC analysis showed that approximation of the feasible set by a bounding ellipsoid or polytope may lead to overly conservative prediction results [42].

When analysis suggests disagreement among models and data, there are three possible causes: the model is correct and the data are flawed, the data are correct and the model is flawed, or both are flawed. In the present study, we focus on the second case, where we have more confidence in the data than in the model. The first case has been a subject of our past studies [12, 25], and the third case will be left as a challenge for future work. In the work of Kennedy and O'Hagan [30], experimental observations are assumed to be noisy measurement of the underlying true process which represents reality,

$$y = \mathcal{R}(s) + \epsilon, \tag{1.1}$$

where ϵ is the measurement noise, s are the scenario parameters, and $\mathcal{R}(s)$ represents reality. The scenario parameters, such as the initial temperature and pressure, are controllable properties known from the experimental setup and can vary from experiment to experiment. In any scientific endeavor, knowledge of the true process is an idealization; a model, considered as tentatively entertained [5], may have a systematic error in prediction. Kennedy and O'Hagan [30] suggest describing the uncertainty in the model form as an additive term, δ, referred to as model *inadequacy*, to the model output,

$$\mathcal{R}(s) = M(x^*, s) + \delta(s), \tag{1.2}$$

where x^* are the underlying true calibration parameters. The model parameters are uncertain parameters intrinsic to the model, $M(\cdot)$, and share common values across all experiments.

This approach of compensating for model discrepancy has received substantial interest. Some Bayesian frameworks (see, e.g., [2, 6, 26, 29, 31, 39, 40, 50, 53]) use a Gaussian process (GP) [43, 49] to represent $\delta(s)$ [6, 26, 39, 40, 50, 53], and others use a functional decomposition [29, 31], while all refer to model inadequacy as model discrepancy, model bias, model form uncertainty, model error, and model form error. Efforts have also been made to overcome the difficulty in identifying model discrepancy and model parameters individually, and to improve prediction performance at conditions different from the training data. For example, Brynjarsdóttir and O'Hagan [6] put constraints on the GP realization of model discrepancy

at specific conditions derived from domain knowledge. Plumlee [39] argued that the prior distribution of model discrepancy should be orthogonal to the gradient of the model under certain assumptions. Wang, Chen, and Tsui [53] estimated the model discrepancy and model parameters separately. Joseph and Melkote [29] constructed a statistical model of discrepancy in a sequential manner, limiting its contribution to the prediction. In Bayesian history matching, model discrepancy is taken into account by including a structured term in its statistical model of the relation between the data and the simulator's output [22].

Our objective here is to resolve disagreement among models and data in the deterministic setting of B2BDC using the perspective of (1.2). The optimization-based framework of B2BDC represents uncertainty by sets and has been successfully applied in several domains, including combustion science [12, 15, 19, 27, 41, 46] and engineering [37], atmospheric chemistry [47], quantum chemistry [10], and systems biology [11, 13, 55]. In the present work, we expand the B2BDC formalism by adding a deterministic model-form discrepancy function to the constraints derived using model and data, conceptually following Kennedy and O'Hagan [30]. We start with a brief recount of B2BDC in subsection 2.1, followed by a reformulation of the feasible set and prediction problems with model discrepancy in subsection 2.2. Application of the proposed methodology is presented in section 3 for two examples, a simple mass-spring-damper and a more realistic combustion system. A comparison of the proposed method and a Bayesian calibration method is also given in section 3. Further interpretation of model discrepancy as a general consistency measure is discussed in section 4. We conclude with summarizing comments in section 5.

2. Methodology.

Bound-to-bound data collaboration (B2BDC) is a deterministic optimization-based framework for systematically combining models and experimental data with quantified uncertainties [12, 18, 19, 45]. In this framework, the uncertainties in experimental data are specified by intervals $[L_e, U_e]$, where L_e and U_e are the lower and upper bounds assessed for the eth QOI. A prior uncertainty region of model parameters derived from domain knowledge is also given and denoted by $\mathcal{H}$. The collection of all provided information is referred to as a *dataset*.

2.1. B2BDC without model discrepancy.

The experimental data are utilized to carve out a smaller region in $\mathcal{H}$, referred to as the *feasible set* and denoted by $\mathcal{F}$,

$$\mathcal{F} = \{x \,|\, x \in \mathcal{H},\ L_e \le M(x, s_e) \le U_e,\ e = 1, 2, \ldots, N\}. \tag{2.1}$$

The constraints $L_e \le M(x, s_e) \le U_e$ are referred to as *model-data constraints* since they are derived by connecting model outputs with experimental bounds. The feasible set constitutes the posterior region of the model parameters that satisfy all of the model-data constraints listed in (2.1). In other words, the feasible set in the B2BDC framework is not computed explicitly but, rather, implicitly through constructing constrained optimization problems that ensure the feasibility constraints are satisfied. The dataset is referred to as *consistent* if its feasible set is nonempty and as *inconsistent* otherwise. The feasibility is determined by calculating a numerical measure termed the *scalar consistency measure* (SCM) [12]. The

quantity is defined by

$$
\begin{aligned}
C_{\mathrm{D}} \coloneqq \underset{x\in\mathcal{H}}{\text{maximize}} \quad & \gamma \\
\text{subject to} \quad & L_e + \gamma\left(\frac{U_e - L_e}{2}\right) \le M(x, s_e) \le U_e - \gamma\left(\frac{U_e - L_e}{2}\right), \\
& e = 1, 2, \ldots, N.
\end{aligned} \tag{2.2}
$$

The dataset is consistent if $C_{\mathrm{D}} \ge 0$, as its nonnegative value proves the existence of a parameter vector that satisfies all the constraints. The dataset is inconsistent when $C_{\mathrm{D}} < 0$, and its magnitude evaluates the degree of inconsistency. Sensitivity of the calculated C_{D} with respect to the experimental and parameter uncertainty bounds can also provide valuable information about the dataset [12]. For a consistent dataset, the predicted interval $[L_p, U_p]$ for an unmeasured QOI can be calculated by solving

$$
L_p = \min_{x\in\mathcal{F}} M(x, s_p) \tag{2.3}
$$

and

$$
U_p = \max_{x\in\mathcal{F}} M(x, s_p), \tag{2.4}
$$

where s_p are the corresponding scenario parameters.

An inconsistent dataset implies that the models, experimental data, and prior information are fundamentally incompatible with one another, making prediction essentially meaningless. In prior work, inconsistency was resolved by identifying likely offending experimental data. This was accomplished by computing sensitivities of the consistency measure [12] and/or minimal relaxations of the bounds to recover consistency [25]. In the present work, we focus on regaining dataset consistency and the ability to make predictions through scenario-dependent model corrections.

In practice, the optimization problems (2.2)–(2.4) are often challenging to solve numerically because the underlying model is not convex and is expensive to evaluate. The B2BDC approach is to bracket the solution to the optimization problems with two quantities, referred to as the inner bound and the outer bound, whose computation is more tractable [13, 45]. Computation of the outer bound, in particular, motivates the use of a quadratic/polynomial or rational quadratic surrogate model for the underlying model at each scenario condition, i.e., $M(x, s_e)$. The outer bound is then obtained by solving a semidefinite program (SDP) [45]. The inner bound is a local solution to the original optimization problem found by nonlinear constrained optimization solvers. In our B2BDC toolbox [32], the inner bound is calculated with the MATLAB function `fmincon` and the outer bound using the SDP solver SeDuMi [51] and the modeling system CVX for convex optimization [23, 24]. Considering that many QOIs depend mainly on a small fraction of the model parameters, a phenomenon termed effective sparsity [4], a subset of the model parameters that have a significant effect on $M(x, s_e)$, referred to as active parameters, is selected by a screening sensitivity analysis [16]. The coefficients of the surrogate model are computed by fitting simulations of $M(x, s_e)$ at selected points, referred to as design points, in the active parameter space. Development of

the active parameters results in a reduced computational cost of building and evaluating the surrogate models. In a typical B2BDC application, generation of surrogate models usually consumes the most CPU time (by running the simulator at design points). In addition to the savings gained by selecting active parameters for each QOI, the cost can be further reduced by running simulations at design points in parallel. Previous applications [46, 48] showed that the computation of inner and outer bounds took at most a few minutes on a desktop computer for datasets consisting of 55 and 102 model parameters along with 167 and 77 QOIs, respectively.

2.2. B2BDC with model discrepancy. In practice, one may encounter an inconsistent dataset where the experimental data are more reliable than the model form (e.g., as in [36]). In such a case, resolution of the inconsistency may be sought by compensating the difference between the model and data through introduction of a discrepancy function. We adopt the definition of model discrepancy proposed by Kennedy and O'Hagan (1.2) and on that basis introduce a scenario-dependent *discrepancy function* $\delta(s)$ into the model-data constraints,

$$L_e \le M(x, s_e) + \delta(s_e) \le U_e, \quad e = 1, 2, \ldots, N. \tag{2.5}$$

We assume that the discrepancy takes the form of a linear combination of n basis functions, $\{\Phi_i\}_{i=1}^n$,

$$\delta(s) = \sum_{i=1}^{n} c_i \Phi_i(s), \tag{2.6}$$

where $\{c_i\}_{i=1}^n$ are unknown coefficients, and $n = 0$ refers hereafter to a zero discrepancy function (i.e., $\delta = 0$). Substitution of (2.6) into (2.5) results in the modified model-data constraints,

$$L_e \le M(x, s_e) + \sum_{i=1}^{n} c_i \Phi_i(s_e) \le U_e, \quad e = 1, 2, \ldots, N. \tag{2.7}$$

The linear form in (2.6) is motivated by the fact that the existing tools in B2BDC are directly applicable with (2.7) in the extended parameter space (x, c): while $\{\Phi_i\}_{i=1}^n$ can be any set of nonlinear functions, the modified model-data constraints are linear in $\{c_i\}_{i=1}^n$. Representing model discrepancy by a linear combination of basis functions has been done by others [29]. However, our method is different from those in that it does not set an objective to fit a particular model discrepancy function. Hence, orthogonality among basis functions, usually imposed to improve regression performance, is not required.

We define the joint feasible set in the extended parameter space of x and c by combining the prior uncertainty and modified model-data constraints,

$$\mathcal{F}_\delta = \left\{ (x, c) \,|\, x \in \mathcal{H},\, c \in \mathcal{H}_c,\, L_e \le M(x, s_e) + \sum_{i=1}^{n} c_i \Phi_i(s_e) \le U_e,\ e = 1, 2, \ldots, N \right\}, \tag{2.8}$$

where $\mathcal{H}_c$ represents the prior uncertainty region of the discrepancy-function coefficients. The feasibility of $\mathcal{F}_\delta$ can be calculated by modifying the model-data constraints accordingly in the

SCM formula in (2.2). The projection of $\mathcal{F}_\delta$ on the model parameter space is

$$\widetilde{\mathcal{F}} = \{x \,|\, \exists \hat{c} : (x, \hat{c}) \in \mathcal{F}_\delta\}, \tag{2.9}$$

which represents the set of feasible model parameters after including the discrepancy function. When the joint feasible set is not empty, prediction at an unmeasured scenario s_p can be obtained by solving the following modified versions of (2.3) and (2.4):

$$L_p = \min_{(x,c)\in\mathcal{F}_\delta} M(x, s_p) + \sum_{i=1}^{n} c_i \Phi_i(s_p) \tag{2.10}$$

and

$$U_p = \max_{(x,c)\in\mathcal{F}_\delta} M(x, s_p) + \sum_{i=1}^{n} c_i \Phi_i(s_p). \tag{2.11}$$

The proposed framework treats the data as the sum of model output and discrepancy function as implied by (1.2). The joint feasible set, defined over the extended parameter space (x, c), therefore represents only combinations of the model and discrepancy function that are consistent with the data. It is not possible to elucidate the model or the discrepancy function separately without further assumptions and/or additional information.

A challenge of the above approach is the choice of basis functions. A practitioner may simply choose a set with the least number of basis functions that resolves dataset inconsistency, following, for instance, the Akaike information criterion [1] of penalizing a larger number of parameters, and thus prevent overfitting. If extra insight about the behavior of the discrepancy function is provided by domain expert, various forms of discrepancy function can be investigated before making the final decision considering these extra factors in addition to the requirement that dataset consistency is recovered.

The developed framework with model discrepancy expressed using (2.6) has the general feature that, for a given dataset and a prediction QOI, the prediction interval becomes systematically wider if additional basis functions are included. To understand this, suppose two sets of basis functions are used in an analysis, with the second being a superset of the first. Let vector c represent the coefficients for the shared basis functions $\{\Phi_i\}_{i=1}^{n}$ and c' the coefficient vector for the additional basis functions $\{\Phi'_j\}_{j=1}^{n'}$, i.e.,

$$\begin{aligned} \delta^1(s) &= \sum_{i=1}^{n} c_i \Phi_i(s), \\ \delta^2(s) &= \delta^1(s) + \sum_{j=1}^{n'} c'_j \Phi'_j(s). \end{aligned} \tag{2.12}$$

The corresponding joint feasible sets formed by (2.8) are denoted by $\mathcal{F}_\delta^1$ and $\mathcal{F}_\delta^2$. Any feasible point $(x, c) \in \mathcal{F}_\delta^1$ is also feasible for $\mathcal{F}_\delta^2$ by setting c' to zero, i.e., $(x, c, c')|_{c'=0} \in \mathcal{F}_\delta^2$. Therefore, the posterior uncertainty interval of the QOI, predicted on $\mathcal{F}_\delta^1$, is always contained in that predicted on $\mathcal{F}_\delta^2$. The increased uncertainty in the prediction interval can depend on the

prediction QOI, the dataset, and the selected basis functions, as will be demonstrated in section 3.

Previous work (e.g., [6, 39]) has demonstrated the value of including prior knowledge of the model discrepancy function when applying statistical UQ methods. In B2BDC, this can be accomplished by incorporating additional constraints. For example, sign constraints on the discrepancy function, or its derivatives, can be enforced at specified scenario conditions by introducing linear inequalities in c. An example of forcing the model discrepancy function to be positive at selected scenarios is

$$\sum_{i=1}^{n} c_i \phi_i(s_j) > 0, \quad j = 1, 2, \ldots. \tag{2.13}$$

The effect of such constraints is automatically spread to predictions through augmenting the feasibility constraints in (2.10) and (2.11). Another example of constraining the magnitude of model discrepancy function is given in subsection 3.1.4.

The posterior uncertainty of the model discrepancy function at any specified scenario s_p can be calculated by solving the prediction problems (2.10) and (2.11) with the objective replaced by $\delta(s_p)$, i.e.,

$$L_\delta = \min_{(x,c)\in\mathcal{F}_\delta} \sum_{i=1}^{n} c_i \Phi_i(s_p) \tag{2.14}$$

and

$$U_\delta = \max_{(x,c)\in\mathcal{F}_\delta} \sum_{i=1}^{n} c_i \Phi_i(s_p). \tag{2.15}$$

Repeating this computation at various conditions in the scenario parameter space can identify regions where uncertainty in the model discrepancy function is large. In a similar manner, the posterior uncertainty can be calculated for each discrepancy-function coefficient c_i.

3. Numerical examples. Application of the extended B2BDC framework is demonstrated with two examples, an illustrative mass-spring-damper system and a realistic hydrogen combustion system. In addition, the proposed method is compared with a Bayesian calibration method via the simple machine example used by Brynjarsdóttir and O'Hagan [6]. In each example, we started with a postulated "true" model to represent the underlying true process. An inadequate model for the analysis was then created by omitting some parts of the true model, and a true calibration parameter value was selected. The developed framework was applied to the inadequate model, and the following results are reported and discussed for different choices of basis functions:

1. Dataset consistency.
2. If the dataset is consistent: (a) the predicted intervals at interpolated and extrapolated conditions; and (b) first, whether the true process values are contained in the predicted intervals, and second, whether the true calibration parameter x^* is in the feasible set.

Quadratic surrogate models were generated and used in the hydrogen combustion example, whereas the original models were used directly in the other two examples. Following our empirical rule that the number of design points should be at least twice that of surrogate model coefficients, 500 design points were used to generate each quadratic surrogate model, which have 21 coefficients. For the first two examples, the computation was conducted for two levels of experimental uncertainty to provide a more comprehensive characterization of the developed method. To clarify the nomenclature, δ^* and δ are used to represent the true model discrepancy defined in (1.2) and the linear combination (2.6), respectively. We also differentiate between *interpolated* and *extrapolated* predictions, where the former refers to s_p lying inside the training domain and the latter outside.

All the example scripts along with the general B2BDC software [32]—i.e., everything required to reproduce the results reported in this paper—can be found using the GitHub link https://github.com/B2BDC/. The numerical experiments were programmed in MATLAB and run on a desktop computer with 16 GB RAM and an Intel i7-6700K CPU.

3.1. One-dimensional mass-spring-damper system. The force, F, needed to extend or compress a spring by a small distance, z, is expressed using Hooke's law

$$F = -kz, \tag{3.1}$$

where k is a constant characteristic of the spring, its stiffness. We now consider a simple system: a spring attached to a ball on one end and affixed to a wall on the other; this is sketched in Figure 1.

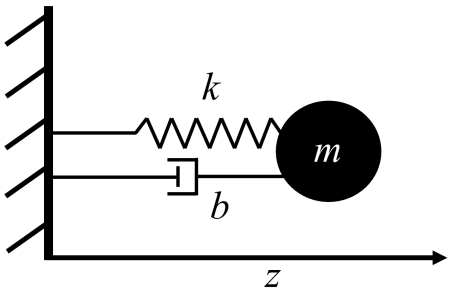

Figure 1. *Sketch plot of the mass-spring-damper system.*

The ball has a mass $m = 1$ and is placed initially at $z_0 = -1.5$ with an initial velocity $v_0 = 1$. In addition to the force exerted by the spring, a damping force proportional to the ball's velocity also affects the motion of the ball. Thus, the evolution of the ball's displacement is described by

$$\begin{aligned} &\frac{d^2z}{dt} = -kz - b\frac{dz}{dt}, \\ &z|_{t=0} = z_0 = -1.5, \\ &\left.\frac{dz}{dt}\right|_{t=0} = v_0 = 1, \end{aligned} \tag{3.2}$$

where b is the constant coefficient of the damping force, and its value was set to 0.05. For a given k, displacement evolution of system (3.2)—the "true" model in this example—is the solution to a second order, constant coefficient, ordinary differential equation and has the analytic form

$$z^*(k,t) = e^{-bt/2}\left[\frac{v_0 + 0.5bz_0}{\sqrt{k - b^2/4}}\sin\left(\sqrt{k - b^2/4}\,t\right) + z_0\cos\left(\sqrt{k - b^2/4}\,t\right)\right]. \tag{3.3}$$

The "inadequate" model was constructed by neglecting the damping force (i.e., $b = 0$), which results in the solution

$$z(k,t) = \frac{v_0}{\sqrt{k}}\sin\left(\sqrt{k}\,t\right) + z_0\cos(\sqrt{k}\,t). \tag{3.4}$$

In both the true and inadequate models, the stiffness k is the model parameter, and the time t is the scenario parameter. The true stiffness of the spring—the true calibration parameter value—was selected to be $k^* = 0.25$ with the prior uncertainty interval $\mathcal{H} = [0.2, 0.3]$. The real displacement was evaluated with $z^*(k^*, t)$. The displacements computed by the two models with $k = k^*$ and their difference, the model discrepancy (1.2), are demonstrated in Figure 2 for $t \in [0, 4]$.

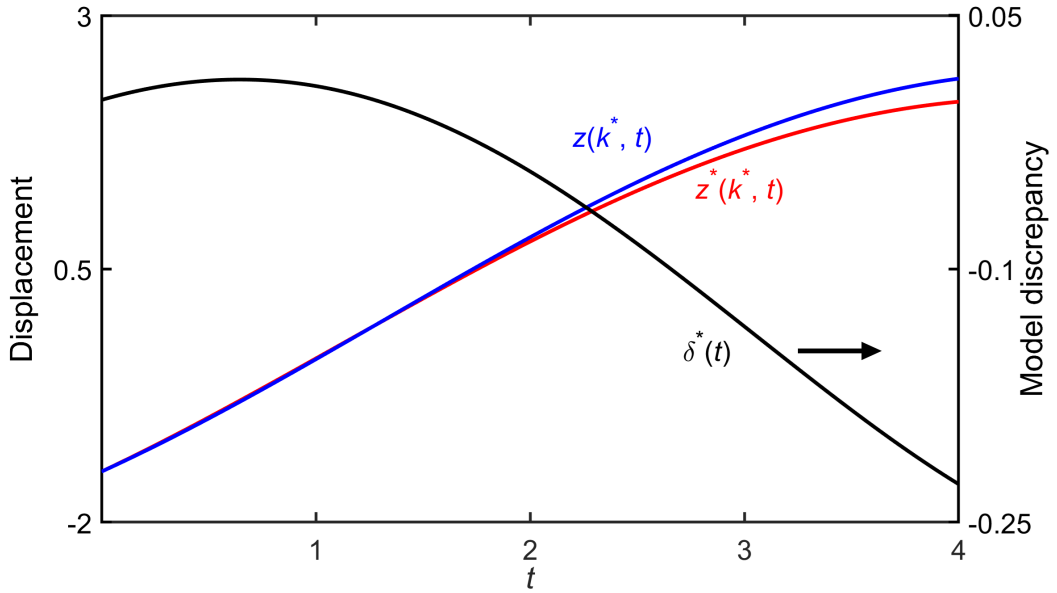

Figure 2. *The true model solution $z^*(k^*, t)$, the inadequate model solution $z(k^*, t)$ and the model discrepancy function $\delta^*(t) = z^*(k^*, t) - z(k^*, t)$.*

The QOIs for this example were chosen to be the displacements of the ball at specified times t_e. The dataset was composed of 20 of these QOIs in the scenario region $t \in [0, 3]$. For each QOI, an "observed" value was generated by adding uniform noise with a prescribed

maximum magnitude ϵ to the true process value,

$$z_e = z^*(k^*, t_e) + \epsilon\, u_e, \qquad u_e \sim \mathcal{U}(-1,\, 1), \quad e = 1, 2, \ldots, 20. \tag{3.5}$$

QOI uncertainty bounds were generated by setting $L_e = z_e - \epsilon$ and $U_e = z_e + \epsilon$. The present analysis was performed with ϵ values of 0.05 and 0.1. Three prediction QOIs were generated for t values of 1.5, 3.2, and 4. The first prediction case occurs at a scenario within the training-set domain of [0, 3] and is an interpolated prediction. The second and third cases occur at scenarios outside the training-set domain and are extrapolated predictions.

3.1.1. Dataset consistency and QOI prediction. We first considered the ideal situation where the true model, given by (3.3), and the formulas in subsection 2.1 were used in the B2BDC calculations. The prediction results are displayed in Figure 3. With this setup, the dataset is consistent with k^* being feasible, and the predicted intervals contain the true process values for both tested ϵ's. The length of the predicted intervals at each prediction scenario is shorter for a smaller value of ϵ, indicating that more accurate measurements produce more accurate predictions.

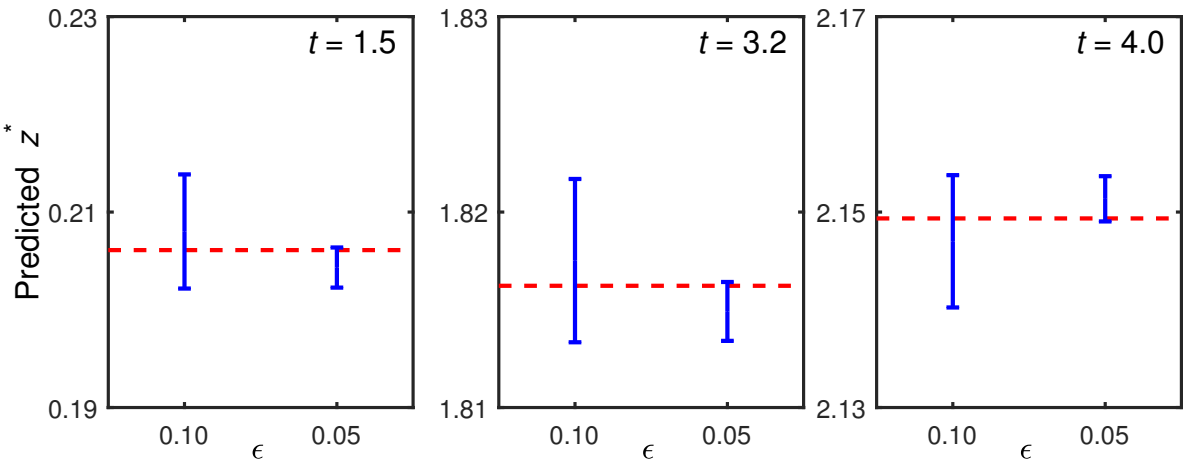

Figure 3. *Computed QOI prediction intervals for the mass-spring-damper example using the true model. The horizontal red dashed lines mark the displacement computed with the true model and true calibration parameter value, $z^*(k^*, t)$. The vertical blue solid lines designate the B2BDC predicted intervals, computed by solving optimization problems (2.3) and (2.4).*

We then moved to a more realistic situation where an inadequate model, given by (3.4), was examined through the modified B2BDC framework, described in subsection 2.2. Four different model discrepancy functions,

$$\delta(t) = \sum_{i=1}^{n} c_{i-1} t^{i-1}, \quad n = 1, 2, 3, 4, \tag{3.6}$$

were tested in addition to the case where $\delta = 0$. The discrepancy function is a polynomial in t of degree $n - 1$.

Table 1
Results of the dataset consistency analysis.

ϵ	n				
	0	1	2	3	4
0.05	inconsistent	inconsistent	inconsistent	$k^* \in \widetilde{\mathcal{F}}$	$k^* \in \widetilde{\mathcal{F}}$
0.10	inconsistent	$k^* \notin \widetilde{\mathcal{F}}$	$k^* \in \widetilde{\mathcal{F}}$	$k^* \in \widetilde{\mathcal{F}}$	$k^* \in \widetilde{\mathcal{F}}$

The outcome of the dataset consistency analysis is summarized in Table 1. Examination of these results shows that the dataset is inconsistent for both values of ϵ when $n = 0$, i.e., a zero model discrepancy function was used. For $\epsilon = 0.05$, a quadratic δ is required to obtain dataset consistency. In this case, k^* is also found to be feasible. For $\epsilon = 0.1$, a constant δ is enough to achieve consistency. However, k^* becomes feasible only after using a linear δ.

The predicted QOI intervals are displayed in Figure 4 for $t = 1.5$, 3.2, and 4. As expected, the prediction intervals with a higher order δ are wider for both ϵ values. In the cases where δ produced a consistent dataset for both ϵ values, a shorter prediction interval is observed with the smaller ϵ. For $\epsilon = 0.1$, the QOI interval predicted using a constant δ does not contain the true value at all time instances. With a linear δ, the predicted interval contains the true value at all three time instances. The predicted interval contains the true value for all tested times and for both values of ϵ with a quadratic and cubic δ.

3.1.2. Posterior bounds of model parameter and discrepancy-function coefficients. We now examine the posterior uncertainty bounds of model parameter k and discrepancy-function coefficients $\{c_i\}_{i=1}^n$ obtained for different polynomial orders of δ. These bounds are the one-dimensional projection of the joint feasible set $\mathcal{F}_\delta$ onto coordinate directions. The volume ratio of the joint feasible set and the multidimensional box, whose sides are the posterior projections of the parameters, was calculated as the fraction of 10^6 samples uniformly distributed in the box that lay in $\mathcal{F}_\delta$. The results are presented in Table 2. For comparison, the posterior interval of k obtained with the true model is also listed. The computed volume ratio results show that the joint feasible set becomes progressively smaller relative to the box as the dimension increases.

The B2BDC analysis with the true model resulted in a significantly narrower posterior uncertainty interval for model parameter k compared to its prior; the interval in this case, by design, contains the true calibration parameter value. With the inadequate model and a constant model discrepancy function ($n = 1$) at $\epsilon = 0.1$, an even narrower posterior interval was obtained; however, the true value, k^*, is completely missed. With a higher order δ, the posterior interval covers the same range as the prior. This outcome can be explained by considering two factors that affected the posterior interval of k.

First, the inadequate model has a different functional dependency on the model parameter k, resulting in the following problem-specific change of the posterior bounds: feasible k values for the true model can become infeasible for the inadequate model and vice versa. In the current example, this can be illustrated by comparing the displacement predicted using the true and inadequate models and its dependency on model parameter k, as demonstrated in Figure 5. Plotted in this figure are $z^*(k,t)$ and $z(k,t)$ computed for different values of k drawn from its prior interval along with the experimental bounds for the case where $\epsilon = 0.1$.

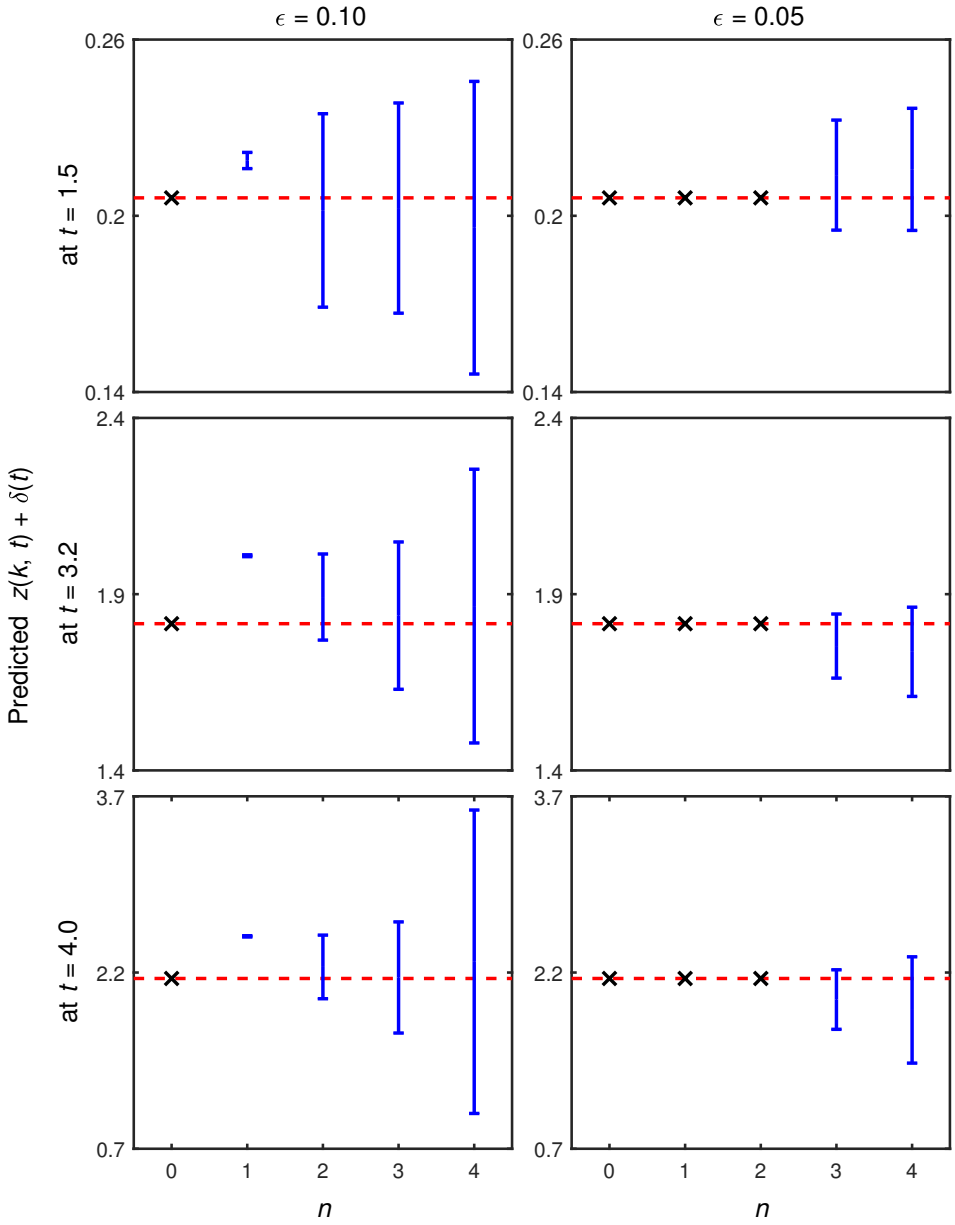

Figure 4. *Predicted QOI intervals at $t = 1.5, 3.2,$ and 4. The horizontal red dashed lines are the displacement derived with the true model and evaluated at the true model parameter value, $z^*(k^*, t)$. The vertical blue solid lines designate the B2BDC predicted intervals, computed by solving the optimization problems (2.10) and (2.11). The $\times$'s mark dataset inconsistency.*

For given t, larger k values produced larger displacements for both models. The resulting displacement bands (shown in cyan) cover similar vertical regions at smaller t values, but the band for the inadequate model gradually shifts upward with increasing magnitude at larger t values compared to that for the true model. For the last two observations, shown in the right

Table 2

Projection of the joint feasible set, computed with the inadequate model (3.4) and different model discrepancy functions (3.6), onto coordinate directions of model parameter k and discrepancy-function coefficients $\{c_i\}_{i=1}^n$; also shown is the computed volume ratio of the joint feasible set to the box made of the projected intervals. The symbol ∅ represents an empty posterior uncertainty due to dataset inconsistency.

n	k	c_0	c_1	c_2	c_3	Volume ratio
			$\epsilon = 0.05$			
	[0.24, 0.25]*					
0	∅					
1	∅	∅				
2	∅	∅	∅			
3	[0.20, 0.30]	[-0.02, 0.03]	[-0.08, 0.06]	[-0.05, 0.00]		5.2×10^{-3}
4	[0.20, 0.30]	[-0.02, 0.03]	[-0.12, 0.06]	[-0.13, 0.14]	[-0.05, 0.03]	2.3×10^{-4}
			$\epsilon = 0.10$			
	[0.24, 0.26]*					
0	∅					
1	[0.20, 0.21]	[0.00, 0.01]				5.2×10^{-1}
2	[0.20, 0.30]	[-0.01, 0.04]	[-0.11, 0.01]			4.9×10^{-2}
3	[0.20, 0.30]	[-0.02, 0.04]	[-0.13, 0.06]	[-0.05, 0.03]		1.2×10^{-2}
4	[0.20, 0.30]	[-0.04, 0.05]	[-0.19, 0.19]	[-0.31, 0.22]	[-0.07, 0.09]	2.0×10^{-4}

*Posterior uncertainty interval obtained with the true model.

inset plot, only a small portion of the band satisfies the QOI bounds. Note that this portion corresponds to smaller k values. However, predictions with these smaller k values invalidated at least one other QOI bound, motivating the use of δ to resolve inconsistency.

The second factor, as discussed in subsection 2.2, is that inclusion of a higher order δ always results in a wider posterior interval. For a constant δ at $\epsilon = 0.1$, the posterior interval widened from the empty set (a zero δ) to an interval with finite length. With the constant δ, feasible (k, c_0) can be found with k limited to a very small region close to the prior lower bound. The red dashed curve in Figure 5 corresponds to the prediction with one of the feasible (k, c_0).

The posterior uncertainty intervals of $\{c_i\}_{i=1}^n$ also become systematically wider with a higher polynomial order of δ, as expected. The enlarged posterior uncertainty intervals associated with individual parameters k and $\{c_i\}_{i=1}^n$ are related to the phenomenon usually referred to in statistical literature (e.g., [6]) as confounding, manifesting itself in the presence of a strong correlation between model parameter(s) and model discrepancy despite their relatively wider marginal posterior distributions. We demonstrate this from a deterministic perspective by the plots shown in Figure 6 generated for the case of a linear δ at $\epsilon = 0.1$. The plots display the joint feasible set of k, c_0, and c_1 along with its two-dimensional projections. The three-dimensional plot clearly shows that the joint feasible set occupies only a small fraction of the enclosing cube. Inspection of the projections indicates that for fixed k, the uncertainty in c_0 and c_1 is reduced, on average, to 66% and 46% of their posterior ranges, respectively, and at fixed c_0, the uncertainty in c_1 is reduced, on average, to 26% of its posterior range.

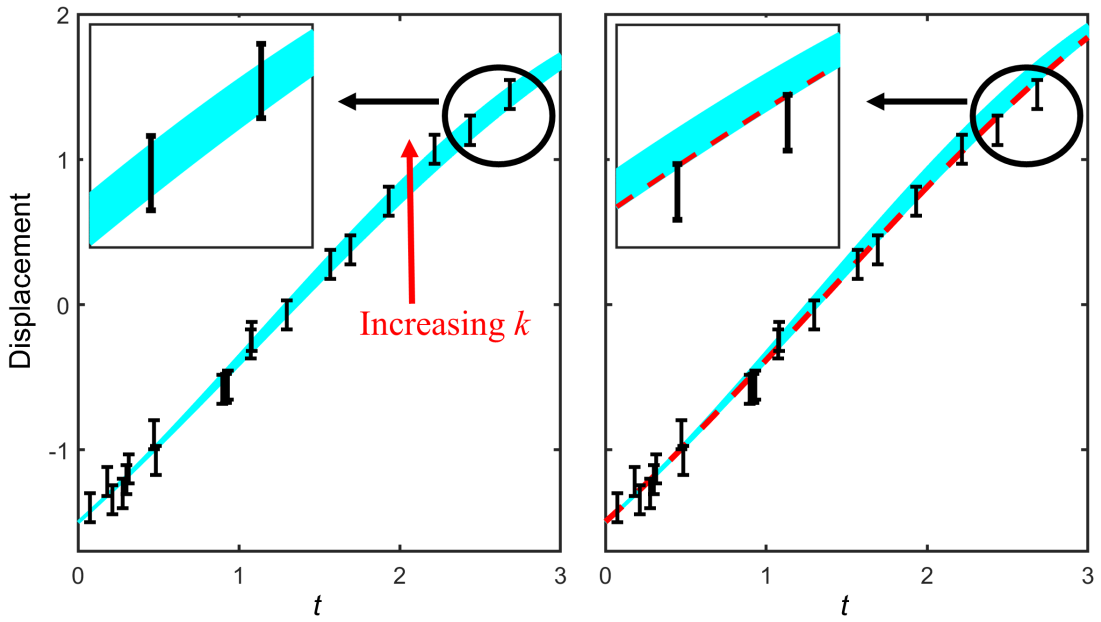

Figure 5. *Displacements computed with the true (left) and the inadequate (right) models for various k values drawn from its prior interval* [0.2, 0.3] *(cyan regions). The black vertical bars are observation QOI bounds. The red dashed line is one feasible realization of $z(k,t)+c_0$ with $k=0.2$ and $c_0=0.005$. The insets are zoomed on the last two observations for $t\in$* [2.3, 2.8].

3.1.3. Posterior uncertainty of model discrepancy. We now examine the upper and lower bounds of δ predicted at 1000 discrete time points, t_i, equally spaced in [0, 4]. The bounds were calculated by solving problems (2.14) and (2.15). This region is divided into the interpolation zone ($t\in$ [0, 3]), where data exists, and the extrapolation zone ($t\in$ [3, 4]) for comparison. The uncertainty bands are shown in Figure 7 for quadratic and cubic δ; they were generated by linearly interpolating adjacent upper and lower bounds.

Inspection of these results shows that the computed uncertainty bounds enclose δ^* in both the interpolation and extrapolation zones for both quadratic and cubic δ. The width of the predicted uncertainty band is effectively constrained within the interpolation zone. The predicted uncertainty band starts to widen toward the end of the interpolation zone and diverges rapidly in the extrapolation zone. The observed divergence is more dramatic for a cubic δ than for a quadratic δ. The uncertainty band for a fixed δ is, overall, narrower with a smaller ϵ in both the interpolation and extrapolation zones.

3.1.4. Additional constraints on model discrepancy. As discussed in subsection 2.2, constraints derived from domain knowledge about the model discrepancy can be included in the B2BDC calculations. We illustrate this feature with the following example.

Let us assume that although we introduced a discrepancy function, we would still like to rely on the model more than on the introduced correction when making predictions. This requirement can be attained by selecting among all feasible values of δ those that have their

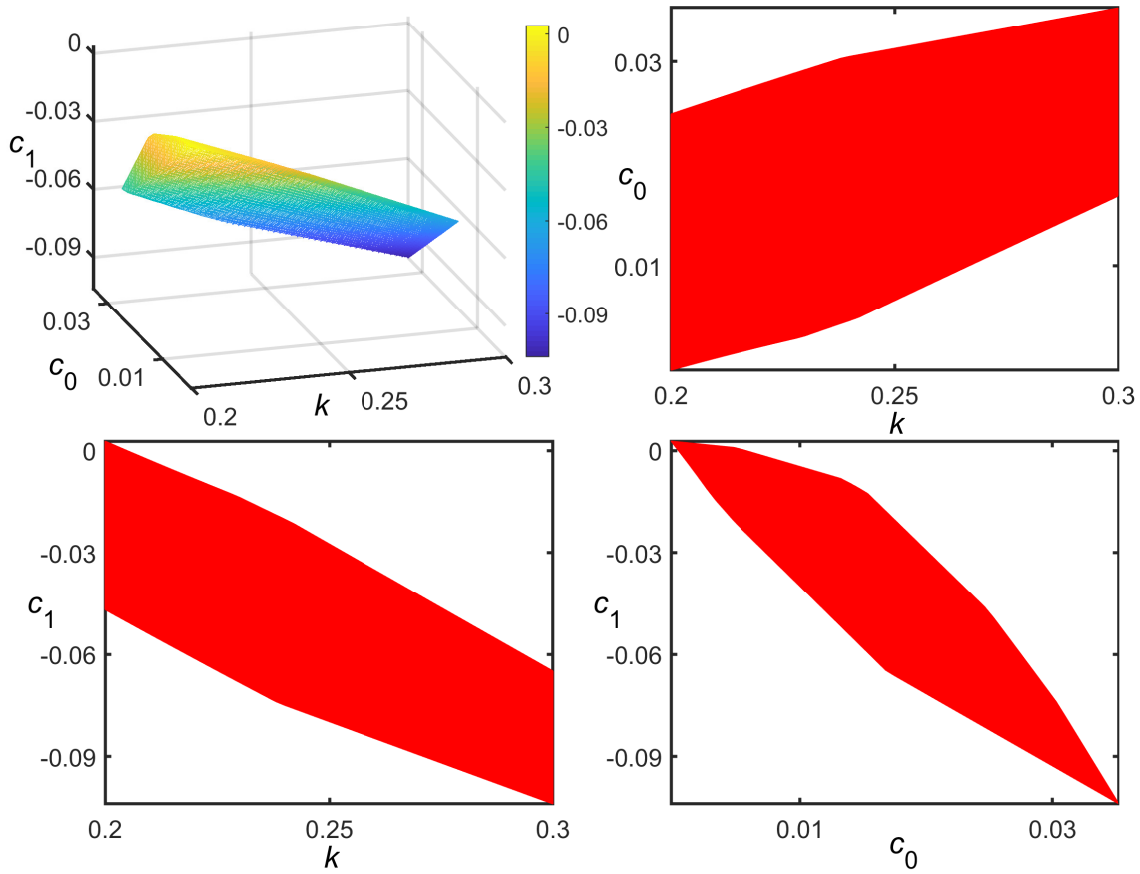

Figure 6. *Joint feasible set of k, c_0, and c_1 and its two-dimensional projections (colored in red), computed with linear δ and $\epsilon = 0.1$. The color bar in the upper-left figure color codes the values of c_1. The axes' limits of k, c_0, and c_1 were set to their calculated posterior uncertainty bounds.*

magnitude, averaged over data and prediction scenarios, below a prescribed threshold, α,

$$\frac{1}{N+1}\left(|\delta(t_p)| + \sum_{i=1}^{N}\left|\delta(t_i)\right|\right) \leq \alpha, \tag{3.7}$$

where N is the number of experimental QOIs. The constraint was added to the joint feasible set construction (2.8), and predictions were made with varying values of α. The results for $\epsilon = 0.1$ and cubic δ are shown in Figure 8. As expected, the predicted interval increases for larger α, eventually reaching the value obtained without (3.7) as this additional constraint becomes inactive.

3.2. A hydrogen combustion model. In this section we apply the formalism described above to a hydrogen combustion model: a homogeneous adiabatic H_2-air reaction system at constant volume. The evolution of the system states, i.e., species concentrations and temperature, is simulated numerically by solving a set of ordinary differential equations. The time

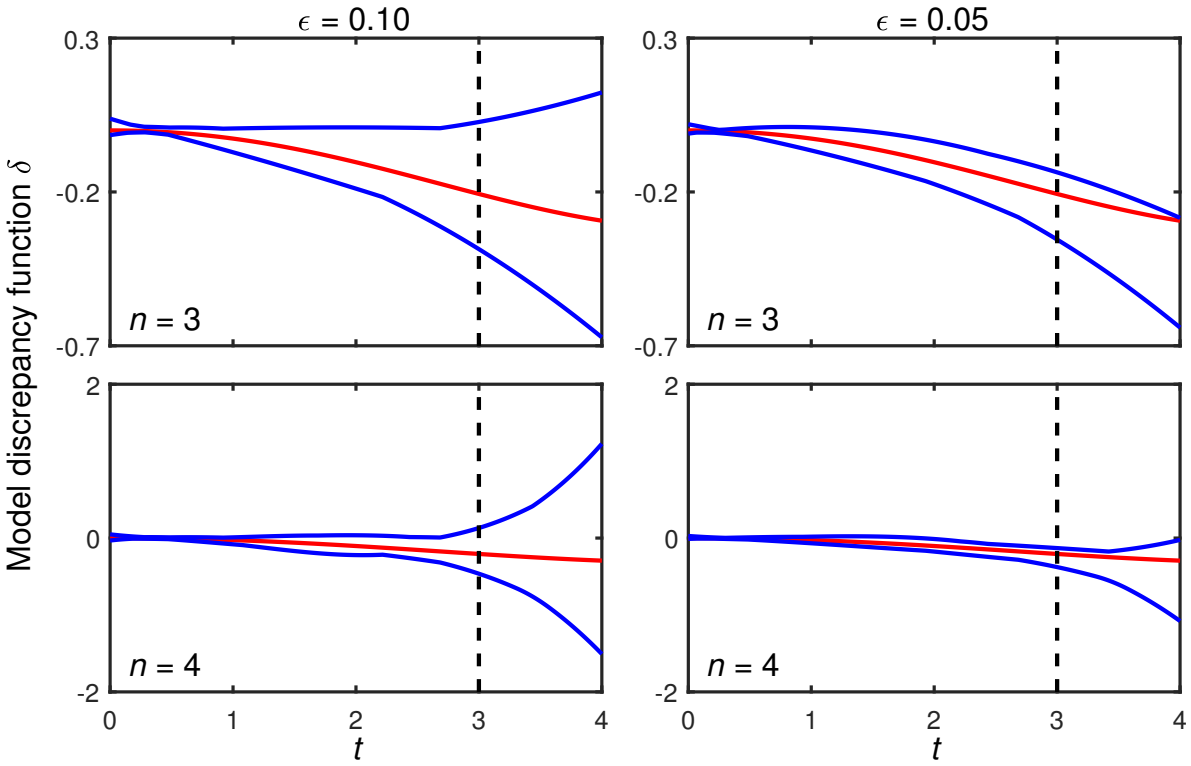

Figure 7. *Uncertainty bounds of quadratic and cubic δ (blue lines). The red line is the true model discrepancy δ^*.*

derivatives of species concentrations and temperature are calculated based on the specified chemical reaction mechanism and the energy equation. Simulations with detailed (21 reactions [56]) and reduced (5 reactions [54]) mechanisms, listed in Appendix A, were considered as the true and inadequate models, respectively. The model parameters, denoted by $\lambda \in \mathbb{R}^5$, are the logarithm of the multipliers associated with the five rate constants shared by both mechanisms, with their prior uncertainties taken from [56]. The true calibration parameter value was specified as $\lambda^* = \mathbf{0}$, where $\mathbf{0}$ is a vector of zeros.

The normalized scenario parameters, s_1, s_2, and s_3, were defined as

$$
\begin{aligned}
s_1 &= \frac{1000/T - 1000/T_{\text{center}}}{1000/T_{\text{low}} - 1000/T_{\text{high}}}, \\
&\quad T_{\text{center}} = 1370 \text{ K}, \ T_{\text{low}} = 1200 \text{ K}, \ T_{\text{high}} = 1600 \text{ K}, \\
s_2 &= \frac{\ln P - \ln P_{\text{center}}}{\ln P_{\text{high}} - \ln P_{\text{low}}}, \\
&\quad P_{\text{center}} = 3.2 \text{ atm}, \ P_{\text{low}} = 1 \text{ atm}, \ T_{\text{high}} = 10 \text{ atm}, \\
s_3 &= \frac{\phi - \phi_{\text{center}}}{\phi_{\text{high}} - \phi_{\text{low}}}, \\
&\quad \phi_{\text{center}} = 1, \ \phi_{\text{low}} = 0.75, \ \phi_{\text{high}} = 1.25,
\end{aligned}
\tag{3.8}
$$

where T, P, and ϕ are the initial temperature, initial pressure, and equivalence ratio of the mixture, respectively. In this example, equivalence ratio is the ratio of hydrogen to oxygen

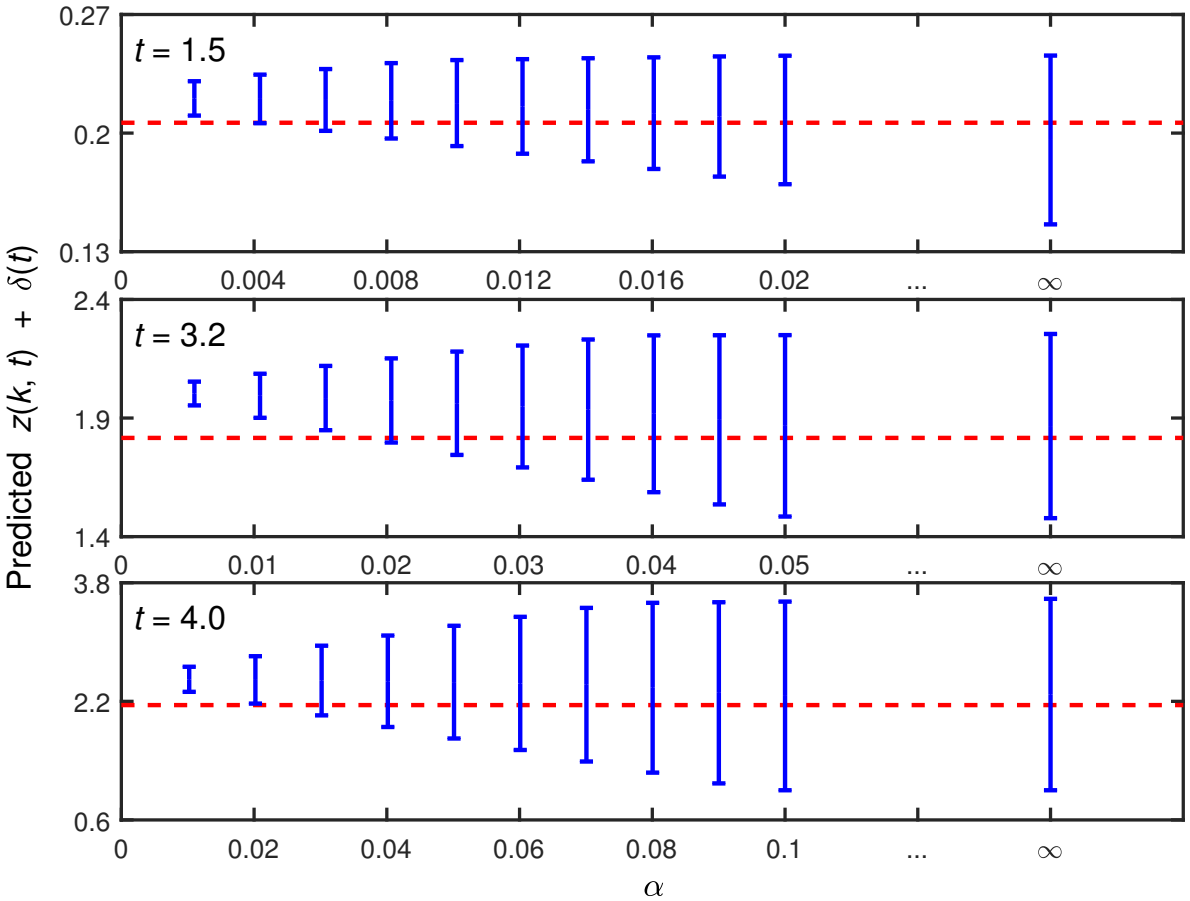

Figure 8. *Interpolation and extrapolation intervals computed by solving (2.10) and (2.11) with the extra constraint (3.7) for cubic δ at $\epsilon = 0.1$. The red dashed lines are the true prediction values.*

concentrations in the initial mixture to that in the stoichiometric mixture. The use of inverse temperature and logarithm of pressure for defining s_1 and s_2 is common in the combustion field; see e.g., [54].

A dataset was constructed using a second order orthogonal design [34] over the scenario region $[-1, 1]^3$. The scenario parameter values are listed in Table 3.

For each of the scenario conditions, the corresponding QOI was defined as the time when the hydrogen concentration drops to half of its initial value. This QOI was computed numerically from the simulated hydrogen concentration profile and denoted by $t^*_{1/2}(\lambda, T, P, \phi)$ and $t_{1/2}(\lambda, T, P, \phi)$ for the true and inadequate models, respectively. "Measured" QOIs, denoted by t_i, were generated by adding a relative noise to the true process values, specified as $t^*_{1/2}$ evaluated at the true calibration parameter value λ^*,

$$t_i = t^*_{1/2}(\lambda^*, T_i, P_i, \phi_i)(1 + \epsilon\, u_i), \qquad u_i \sim \mathcal{U}(-1, 1), \quad i = 1, 2, \ldots, 15. \tag{3.9}$$

The maximum noise magnitude, ϵ, was assigned values of 0.01 and 0.005. As before, the

Table 3
Design conditions for the training data.

Design index	s_1	T (K)	s_2	P (atm)	s_3	ϕ
1	1	1200	1	10	1	1.25
2	1	1200	1	10	-1	0.75
3	1	1200	-1	1	1	1.25
4	1	1200	-1	1	-1	0.75
5	-1	1600	1	10	1	1.25
6	-1	1600	1	10	-1	0.75
7	-1	1600	-1	1	1	1.25
8	-1	1600	-1	1	-1	0.75
9	0	1370	0	3.2	0	1
10	1.215	1170	0	3.2	0	1
11	-1.215	1660	0	3.2	0	1
12	0	1370	1.215	12.8	0	1
13	0	1370	-1.215	0.78	0	1
14	0	1370	0	3.2	1.215	1.3
15	0	1370	0	3.2	-1.215	0.7

Table 4
Tested model discrepancy functions.

Model discrepancy	Number of basis function n
No δ	0
$\delta = c_0$	1
$\delta = c_0 + \sum_{i=1}^{3} c_i s_i$	4
$\delta = c_0 + \sum_{i=1}^{3} c_i s_i + \sum_{i,j=1;\, i\le j}^{3} c_{i,j} s_i s_j$	10

uncertainty bounds were generated by computing $[(1-\epsilon)t_i, (1+\epsilon)t_i]$. The QOI computed with the inadequate model has no analytic solution, and a quadratic surrogate model S_i was generated for each QOI such that $S_i(\lambda) \approx \ln(t_{1/2}(\lambda, T_i, P_i, \phi_i))$. As in the previous example, we consider a polynomial model discrepancy function (Table 4) but now with the scenario parameters s_1, s_2, and s_3 defined by (3.8).

3.2.1. Dataset consistency and QOI prediction. Dataset consistency was calculated first, and the results are given in Table 5.

Inspection of these results shows that with $\epsilon = 0.01$, the dataset is inconsistent for both the zero δ and constant δ cases and becomes consistent when linear and quadratic δ are used. After ϵ was lowered to 0.005, linear δ was insufficient to keep the dataset consistent. For cases where the dataset is consistent, the distance between the true calibration parameter value and the feasible set, denoted by d_{λ^*} and defined as

$$d^2_{\lambda^*} = \min_{(\lambda,c)\in\mathcal{F}_\delta} (\lambda - \lambda^*)^T(\lambda - \lambda^*), \tag{3.10}$$

was calculated, and the results are also reported in Table 5. In all these cases, the true calibration parameter value is not in the feasible set. Its distance from the feasible set is larger when lower order δ or smaller ϵ were used.

For cases where the dataset is consistent, model predictions were computed at one interpolated and four extrapolated scenarios, which are specified in Table 6. The results are

Table 5

Results of dataset consistency and the distance between true model parameter value λ^* and the feasible set.

n	Dataset consistency	d_{λ^*}
	$\epsilon = 0.01$	
0	Inconsistent	—
1	Inconsistent	—
4	Consistent	0.167
10	Consistent	0.047
	$\epsilon = 0.005$	
0	Inconsistent	—
1	Inconsistent	—
4	Inconsistent	—
10	Consistent	0.072

Table 6

Scenario parameter values for model prediction.

Case index	Prediction	s_1	T (K)	s_2	P (atm)	s_3	ϕ
1	Interpolation	-0.6	1500	0.4	5	0	1
2	Extrapolation	-1.67	1800	0.4	5	0	1
3	Extrapolation	-1	1600	1.16	12	0	1
4	Extrapolation	-1	1600	0.4	5	1.6	1.4
5	Extrapolation	-1.67	1800	1.16	12	1.6	1.4

depicted in Figure 9.

Again, the lengths of predicted intervals are shorter with linear δ compared to quadratic δ. Similarly, smaller values of ϵ produced shorter predictions. At $\epsilon = 0.01$, the predicted interval with a linear δ contains the true value for cases 1, 2, and 3 but underpredicts the target for cases 4 and 5. With a quadratic δ, the predicted intervals contain the true values for all tested cases at both ϵ values.

3.2.2. Inference of model discrepancy. The projection of the feasible set on the parameter space of discrepancy-function coefficients, c's, describes not just one discrepancy function but the set of discrepancy functions that are consistent with the data. The following analysis with the linear δ and $\epsilon = 0.01$ shows an example of inferring model discrepancy from B2BDC calculations. The posterior uncertainty bounds of the discrepancy-function coefficients were calculated, and the results are given in Table 7. The volume ratio of the joint feasible set to the multidimensional box, specified similarly as in subsection 3.1, is 2.6×10^{-8} based on 10^9 samples.

Table 7

Projection of the joint feasible set, computed with linear δ at $\epsilon = 0.01$, onto coordinate directions of each discrepancy-function coefficient.

Coefficient	Posterior uncertainty bounds
c_0	[-0.139, 0.112]
c_1	[-0.018, 0.005]
c_2	[-0.015, -0.006]
c_3	[-0.042, -0.013]

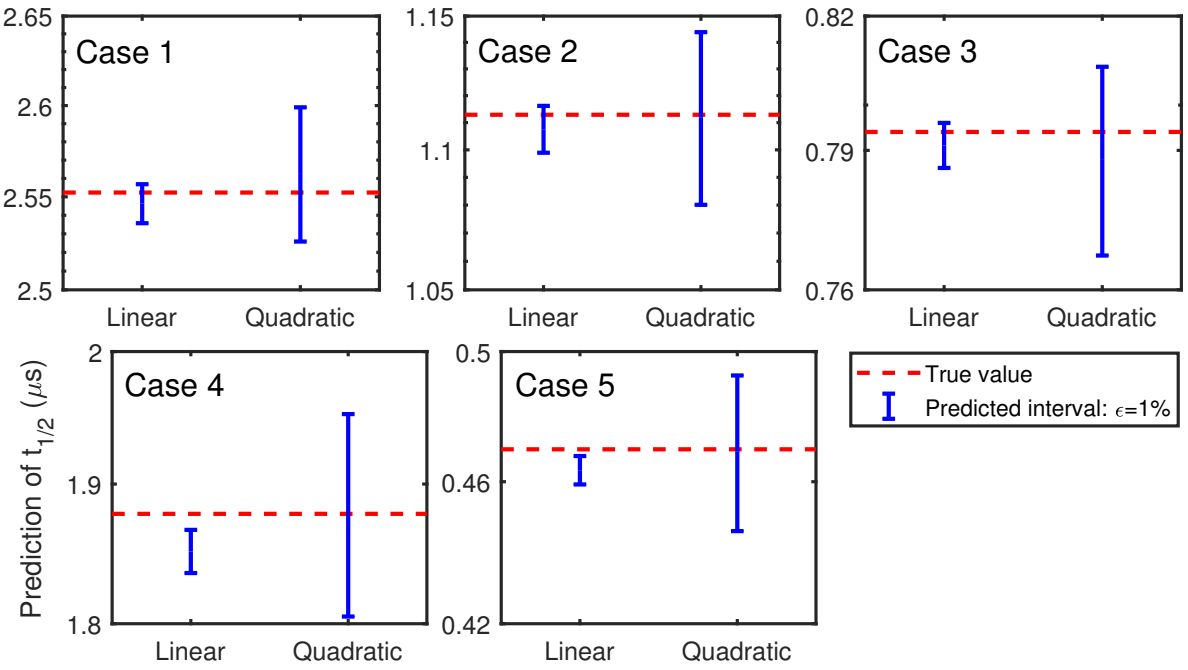

Figure 9. *QOI prediction intervals for the five cases in Table 6. The black dashed lines are the true QOI values. The predicted QOI intervals are drawn as blue ($\epsilon = 0.01$) and red ($\epsilon = 0.005$) vertical lines. The red $\times$'s mark dataset inconsistency.*

The results show that all feasible c_2 and c_3 are negative since the calculated posterior upper bounds are negative for these two coefficients. For the linear δ, the coefficients are also the partial derivatives of the discrepancy function with respect to the scenario parameters, i.e.,

$$c_i = \frac{\partial \delta}{\partial s_i}. \tag{3.11}$$

All feasible δ's are therefore smaller at larger s_2 or s_3 values when other scenario parameters are fixed.

The predicted interval of δ, i.e., $[L_\delta, U_\delta]$ from (2.14) and (2.15), was then calculated in the s_2-s_3 (P-ϕ) space at three fixed s_1 (T) values. The computed intervals were examined to determine the sign of feasible δ's at each specified scenario, and the results are shown in Figure 10.

Similar patterns are observed for the three tested temperature values: except for a lower left triangle region where both pressure and equivalence ratio are relatively small, all feasible δ's are negative. As a result of dataset consistency, predictions made at the grey-region scenarios always add a negative correction to the model output, suggesting that the inadequate model systematically overpredicts the QOI. By the result that c_2 and c_3 are negative in the feasible set, we see that the overprediction is likely to be stronger at larger s_2 and s_3 values, i.e., at higher pressures and equivalence ratios.

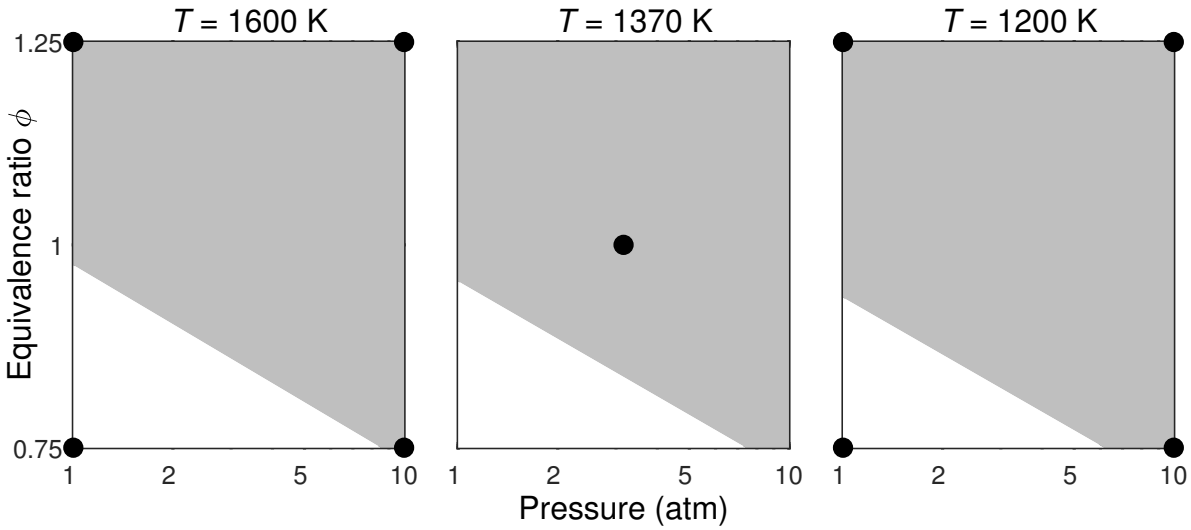

Figure 10. *The sign of model discrepancy function in the scenario region* $(s_2, s_3) \in [-1, 1]^2$ *for temperature values* 1600, 1370, *and* 1200K. *The grey region represents scenarios where* $U_\delta < 0$, *and the white one indicates where the interval* $[L_\delta, U_\delta]$ *contains* 0. *The black points are design scenarios in Table 3.*

3.3. Comparison with Bayesian calibration. In this section we compare the proposed method and the Bayesian calibration method using the simple machine example of Brynjarsdóttir and O'Hagan [6]. The simple machine simulates the amount of work delivered as a function of the amount of input effort and the efficiency of the machine. For this example, the scenario parameter is the amount of input effort, denoted by s, and the calibration parameter is the efficiency of the machine, denoted by θ. The true model is defined as

$$\eta^*(\theta^*, s) = \frac{\theta^* s}{1 + s/20}, \tag{3.12}$$

where η is the delivered work, and $\theta^* = 0.65$ is the true calibration parameter value. The inadequate model is given as

$$\eta(\theta, s) = \theta s. \tag{3.13}$$

"Measurements" were generated, similarly to Brynjarsdóttir and O'Hagan [6], by adding a zero-mean Gaussian noise to η^* at multiple scenario-parameter values s_i, which were equidistantly distributed within the interval $s \in [0.2, 4]$. The Gaussian noise was set with a standard deviation of $\sigma = 0.01$. For specifying prior uncertainty bounds, as required for application of the B2BDC framework, we generated the lower and upper bounds for each measurement by adding and subtracting 2σ intervals to the "measured" value. To ensure the uncertainty interval contains the underlying true value, we truncated the Gaussian noise by 2σ in the process; i.e., if a generated random noise had a magnitude greater than 2σ, it was abandoned, and a new one was drawn from the preset Gaussian distribution. We considered the same three measurement sets investigated by Brynjarsdóttir and O'Hagan [6] with 11, 31, and 61 measurement points. For each of these measurement sets, again following Brynjarsdóttir and

O'Hagan [6], we computed the posterior uncertainty of θ, an interpolated prediction at $s = 1.5$, and an extrapolated prediction at $s = 6.0$.

For the B2BDC analysis, the prior uncertainty region of θ was assumed to be [0.6, 0.7]. We performed the analysis using a set of polynomial discrepancy functions

$$\delta(s) = \sum_{i=1}^{n} c_{i-1} s^{i-1}, \quad n = 4, 5, 6. \tag{3.14}$$

The inadequate model in this example is a linear monomial in the scenario parameter s, and therefore the coefficient c_1 in the discrepancy function (3.14) is confounded with θ. To avoid the confounding, we assumed that the linear part is captured by the inadequate model and force $c_1 = 0$ in the discrepancy function.

Brynjarsdóttir and O'Hagan [6] introduced the following constraints that the discrepancy function must satisfy:

$$\begin{aligned} &\delta(0) = 0, \\ &\delta'(0) = 0, \\ &\delta(s) \le 0 \quad \forall s \in \mathcal{S}, \\ &\delta'(s) \le 0 \quad \forall s \in \mathcal{S}, \end{aligned} \tag{3.15}$$

where δ' is the derivative of the discrepancy function, and $\mathcal{S}$ is a nonempty set. The latter two constraints in (3.15) are referred to here as *semi-infinite* constraints, as embedding them in an optimization problem, such as (2.14) and (2.15), leads to a semi-infinite program [3]. In [3], the last two inequality constraints, which need to be satisfied for an infinite number of s values, were not implemented directly to maintain the computational convenience of a Gaussian process. Instead, these two inequalities were replaced by enforcing $\delta'(s) \le 0$ at $s = 0.5$ and $s = 1.5$, resulting in the following set of constraints:

$$\begin{aligned} &\delta(0) = 0, \\ &\delta'(0) = 0, \\ &\delta'(0.5) \le 0, \\ &\delta'(1.5) \le 0. \end{aligned} \tag{3.16}$$

Since these constraints are linear in the discrepancy-function coefficients, they will be referred to hereafter as *linear* constraints on δ to differentiate them from the semi-infinite constraints of (3.15).

The B2BDC computation was performed using the inadequate model with the δ functions (3.14) for three cases: without any constraints, with linear constraints (3.16), and with semi-infinite constraints (3.15) applied for $s \in [0, 4]$. The last case was implemented using the MATLAB function `fseminf`, which is designed for handling this type of semi-infinite constraint.

When no model discrepancy function was used, the dataset was inconsistent for all three measurement sets. Compared to the very narrow credible interval that does not cover θ^* reported by Brynjarsdóttir and O'Hagan [6], the B2BDC analysis flags a disagreement among the model and the data by identifying the dataset inconsistency.

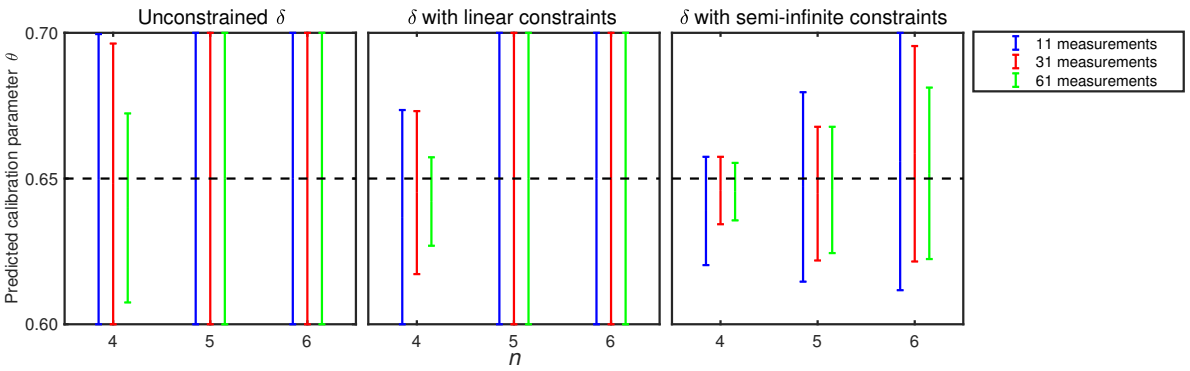

Figure 11. *Predicted intervals for the calibration parameter θ without constraints on δ, with linear constraints on δ, and with semi-infinite constraints on δ. The horizontal black dashed lines indicate the true calibration parameter value, θ^*. The vertical blue, red, and green solid lines designate the B2BDC predicted intervals for* 11, 31, *and* 61 *measurement sets.*

For the case when a model discrepancy function was included in the B2BDC analysis, Figure 11 shows the posterior uncertainty interval of the calibration parameter θ, computed with/without additional constraints on the discrepancy function.

Inspection of Figure 11 reveals that the three tested discrepancy functions are able to resolve the dataset inconsistency for the three measurement sets with or without additional constraints imposed on δ. The prediction interval contains the true calibration parameter value for all tested cases. With an unconstrained δ, the predicted uncertainty interval of θ is the same as the prior uncertainty interval, except for the 11, 31, and 61 measurement sets with a cubic δ ($n = 4$). With the linear constraints on δ, the prediction interval shrinks only for the cubic δ compared to the unconstrained case. The accuracy in the prediction is significantly improved when the semi-infinite constraints are enforced: the prediction interval is reduced compared to the prior uncertainty interval for all tested cases.

The computed B2BDC predictions at interpolated ($s = 1.5$) and extrapolated ($s = 6.0$) conditions are given in Figures 12 and 13. The prediction interval contains the true value at interpolated and extrapolated conditions for all tested cases. The length of the prediction interval reduces if more measurements are included in the computation and if additional constraints on δ are forced. The uncertainty in the prediction interval is much larger at the extrapolated condition than at the interpolated condition. The uncertainty in the prediction interval is also larger with a higher degree polynomial, especially at the extrapolated condition. The reduction in the prediction interval length is more substantial if the semi-infinite constraints are used compared to the linear constraints.

Comparing the present B2BDC results with those of the Bayesian calibration method presented by Brynjarsdóttir and O'Hagan [6], we can make the following observations. The two methods, having different mathematical ways for representing model discrepancy, are able to produce an improved UQ result once the inadequate model is compensated with a modification term. The change in the length of the B2BDC prediction interval or in the Bayesian credible interval is less significant when the number of measurements grows from 31 to 61 compared to that from 11 to 31. For this specific example, the B2BDC prediction intervals with the tested

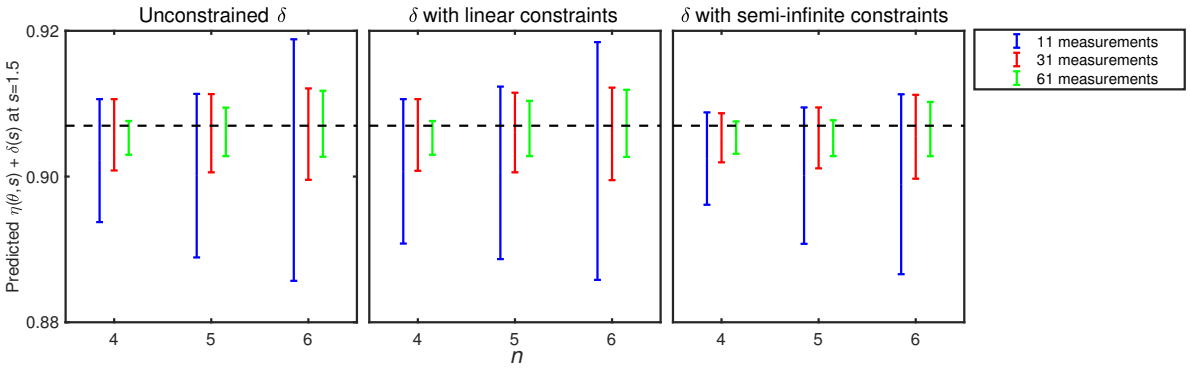

Figure 12. *Predicted intervals at* $s = 1.5$ *for* $\eta + \delta$ *without constraints on* δ*, with linear constraints on* δ*, and with semi-infinite constraints on* δ*. The horizontal black dashed lines are the true work delivered,* $\eta^*(\theta^*, 1.5)$*. The vertical blue, red, and green solid lines designate the B2BDC predicted intervals for* 11*,* 31*, and* 61 *measurement sets, computed by solving the optimization problems* (2.10) *and* (2.11)*.*

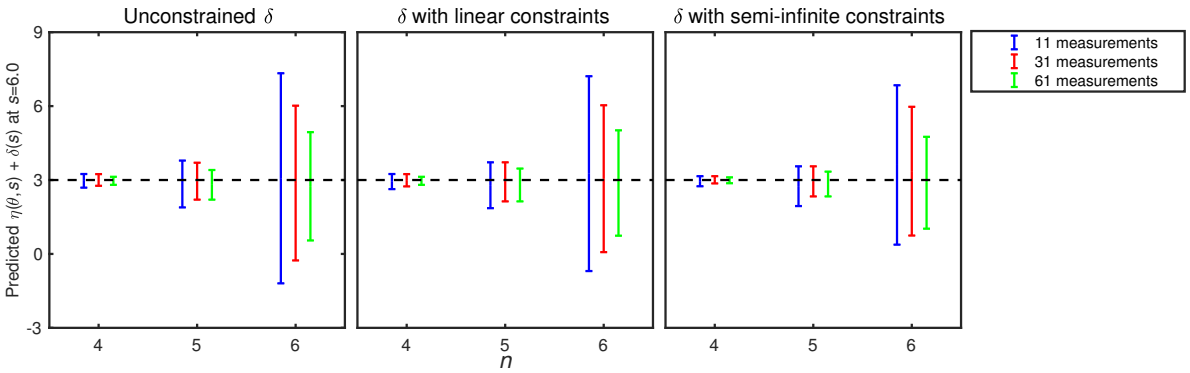

Figure 13. *Predicted intervals at* $s = 6.0$ *for* $\eta + \delta$ *without constraints on* δ*, with linear constraints on* δ*, and with semi-infinite constraints on* δ*. The horizontal black dashed lines are the true work delivered,* $\eta^*(\theta^*, 6.0)$*. The vertical blue, red, and green solid lines designate the B2BDC predicted intervals for* 11*,* 31*, and* 61 *measurement sets, computed by solving the optimization problems* (2.10) *and* (2.11)*.*

discrepancy functions contain the true value of the calibration parameter and of the QOI for both the interpolation and extrapolation predictions. This may be attributed to the fact that the true discrepancy function, $\eta^*(\theta^*, s) - \eta(\theta^*, s)$, can be well represented by the selected polynomial basis functions. Both methods can take into account prior knowledge of the model discrepancy. The present example demonstrates one of the benefits of the optimization-based infrastructure of B2BDC, namely, its natural ability to accommodate a variety of constraints.

4. Discrepancy as a consistency measure. The above examples and discussion primarily focus on the impact of model discrepancy on prediction. The inclusion of discrepancy in B2BDC also provides the opportunity to calculate a more general consistency measure. For

a given collection of basis functions $\{\Phi_i\}_{i=1}^n$, we may define a new consistency measure as

$$\begin{aligned} \min_{x,c} \quad & f(c) \\ \text{subject to} \quad & L_e \le M_e(x) + \sum_{i=1}^{n} c_i \Phi_i(s_e) \le U_e, \\ & x \in \mathcal{H}, \\ & c \in \mathcal{H}_c, \\ & e = 1, \dots, N, \end{aligned} \tag{4.1}$$

where the objective $f(\cdot)$ is a function of only the coefficients $\{c_i\}_{i=1}^n$ and reflects the "complexity" of the discrepancy function. In essence, (4.1) asks the following question: what is the least complex discrepancy function required to recover dataset consistency? Different choices of $f(\cdot)$ and $\Phi_i(\cdot)$ produce different consistency measures. For example, defining the complexity and discrepancy functions as

$$\begin{aligned} f(c) &= \sum_{i=1}^{N} |c_i|, \\ \Phi_i(s) &= -\mathbb{1}_{\{s_i\}}(s) = \begin{cases} -1, & s = s_i, \\ 0 & \text{otherwise}, \end{cases} \end{aligned} \tag{4.2}$$

where $n = N$, and the $\{s_i\}_{i=1}^N$ are the dataset scenarios. Note that with this choice of discrepancy, the eth model-data constraint in (4.1) becomes

$$L_e \le M_e(x) - c_e \le U_e. \tag{4.3}$$

This is exactly a version of the vector consistency measure presented in earlier work [25, eq. (4.4)]. Other choices of f, such as the integral over $\mathcal{H}_c$ of the squared discrepancy (or its squared derivatives, should each Φ_i be differentiable), can also be handled in the B2BDC framework.

Consistency measures formulated in this fashion gauge the disagreement between models and observations based on the "simplest" (or least complex) discrepancy required to render a dataset consistent. One potential application of this type of consistency measure is for model comparison. For a fixed set of basis functions $\{\Phi_i\}_{i=1}^n$, multiple models can be compared by evaluating (4.1). This generalized consistency analysis with model discrepancy is currently being investigated and will be discussed in future work.

5. Conclusions.

We examined the inclusion of model discrepancy as a linear combination of finite basis functions in B2BDC. The existing B2BDC framework was extended by reformulating the feasible set to include both model parameters and discrepancy-function coefficients; the prediction formulas were adjusted accordingly. Dataset consistency can be effectively recovered with the developed framework by increasing the complexity of the model discrepancy function used. The developed method offers a flexible construction of discrepancy function structure through the selection of basis functions; prior information on model discrepancy

can be included naturally in the optimization problems as additional constraints. The confounding effect between model parameters, and model discrepancy function in the posterior uncertainty, presented and discussed in statistical methodologies (e.g., [6]), was demonstrated from the deterministic perspective.

Appendix A. Detailed and reduced hydrogen mechanisms. The detailed and reduced mechanisms used in the present work are listed in Table 8. The reduced mechanism consists of the five reactions, which are indicated by bold font and check marks.

Table 8

Detailed and reduced H_2-O_2 reaction sets and associated parameters of the rate coefficients, $AT^n e^{-E/RT}$, in the units of cm^3, mol, s, cal, K (from [56]).

Reduced		Reactions	A	n	E
✓	1	**$H + O_2 = O + OH$**	2.65×10^{16}	-0.6707	17041
✓	2	**$O + H_2 = H + OH$**	3.87×10^{4}	2.7	6260
✓	3	**$OH + H_2 = H + H_2O$**	2.16×10^{8}	1.51	3430
	4	$OH + OH = O + H_2O$	3.57×10^{4}	2.4	-2110
	5[a]	$H + H + M = H_2 + M$	1.00×10^{18}	-1.0	0
		$H + H + H_2 = H_2 + H_2$	9.00×10^{16}	-0.6	0
		$H + H + H_2O = H_2 + H_2O$	6.00×10^{19}	-1.25	0
	6[b]	$O + O + M = O_2 + M$	1.20×10^{17}	-1.0	0
	7[c]	$O + H + M = OH + M$	4.71×10^{18}	-1.0	0
	8[d]	$H + OH + M = H_2O + M$	2.20×10^{22}	-2.0	0
✓	9[e]	**$H + O_2 + M = HO_2 + M$**	5.75×10^{19}	-1.4	0
		$H + O_2 = HO_2$	4.65×10^{12}	0.44	0
	10	$H + HO_2 = O + H_2O$	3.97×10^{12}	0.0	671
✓	11	**$H + HO_2 = H_2 + O_2$**	2.99×10^{6}	2.12	-1172
	12	$H + HO_2 = OH + OH$	8.40×10^{13}	0.0	635
	13	$O + HO_2 = OH + O_2$	2.00×10^{13}	0.0	0
	14	$OH + HO_2 = H_2O + O_2$	2.89×10^{13}	0.0	-497
	15	$HO_2 + HO_2 = H_2O_2 + O_2$	1.30×10^{11}	0.0	-1630
		$HO_2 + HO_2 = H_2O_2 + O_2$	4.20×10^{14}	0.0	12000
	16[f]	$OH + OH + M = H_2O_2 + M$	1.46×10^{11}	0.868	-8548
		$OH + OH = H_2O_2$	8.71×10^{9}	0.869	-2191
	17	$H + H_2O_2 = H_2O + OH$	1.00×10^{13}	0.0	3600
	18	$H + H_2O_2 = HO_2 + H_2$	1.21×10^{7}	2.0	5200
	19	$O + H_2O_2 = HO_2 + OH$	9.63×10^{6}	2.0	4000
	20	$OH + H_2O_2 = H_2O + HO_2$	1.74×10^{12}	0.0	318
		$OH + H_2O_2 = H_2O + HO_2$	7.59×10^{13}	0.0	7272
	21[c]	$O + OH + M = HO_2 + M$	8.00×10^{15}	0.0	0

[a]Collision efficiency: Ar = 0.63.
[b]Collision efficiencies: H_2 = 2.4, H_2O = 15.4, Ar = 0.83.
[c]Collision efficiencies: H_2 = 2, H_2O = 12, Ar = 0.7.
[d]Collision efficiencies: H_2 = 0.73, H_2O = 3.65, Ar = 0.38.
[e]Collision efficiencies: H_2O = 12, Ar = 0.53. Troe parameters: $a = 0.5$, $T^{***} = 10^{-30}$, $T^{*} = 10^{30}$, $T^{**} = 10^{100}$.
[f]Collision efficiencies: H_2 = 2, H_2O = 6, Ar = 0.67. Troe parameters: $a = 1.0$, $T^{***} = 10^{-30}$, $T^{*} = 10^{30}$, $T^{**} = 10^{30}$.

REFERENCES

[1] H. Akaike, *A new look at the statistical model identification*, in Selected Papers of Hirotugu Akaike, Springer, 1974, pp. 215–222.

[2] M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu, *A framework for validation of computer models*, Technometrics, 49 (2007), pp. 138–154.

[3] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Ser. Oper. Res., Springer, 2000.

[4] G. Box and R. D. Meyer, *Some new ideas in the analysis of screening designs*, J. Res. Nat. Bur. Stand., 90 (1985), pp. 495–500.

[5] G. E. Box and W. G. Hunter, *The experimental study of physical mechanisms*, Technometrics, 7 (1965), pp. 23–42.

[6] J. Brynjarsdóttir and A. O'Hagan, *Learning about physical parameters: The importance of model discrepancy*, Inverse Problems, 30 (2014), 114007.

[7] J.-P. Calliess, S. Roberts, C. Rasmussen, and J. Maciejowski, *Nonlinear set membership regression with adaptive hyper-parameter estimation for online learning and control*, in Proceedings of the 2018 European Control Conference (ECC), IEEE, 2018, pp. 1–6.

[8] L. Chisci, A. Garulli, and G. Zappa, *Recursive state bounding by parallelotopes*, Automatica, 32 (1996), pp. 1049–1055.

[9] P. S. Craig, M. Goldstein, A. H. Seheult, and J. A. Smith, *Pressure matching for hydrocarbon reservoirs: A case study in the use of Bayes linear strategies for large computer experiments*, in Case Studies in Bayesian Statistics, Springer, 1997, pp. 37–93.

[10] D. E. Edwards, D. Y. Zubarev, A. Packard, W. A. Lester, Jr., and M. Frenklach, *Interval prediction of molecular properties in parametrized quantum chemistry*, Phys. Rev. Lett., 112 (2014), p. 253003, https://doi.org/10.1103/PhysRevLett.112.253003.

[11] R. Feeley, M. Frenklach, M. Onsum, T. Russi, A. Arkin, and A. Packard, *Model discrimination using data collaboration*, J. Phys. Chem. A, 110 (2006), pp. 6803–6813, https://doi.org/10.1021/jp056309s.

[12] R. Feeley, P. Seiler, A. Packard, and M. Frenklach, *Consistency of a reaction dataset*, J. Phys. Chem. A, 108 (2004), pp. 9573–9583.

[13] R. P. Feeley, *Fighting the Curse of Dimensionality: A Method for Model Validation and Uncertainty Propagation for Complex Simulation Models*, Ph.D. thesis, University of California, Berkeley, 2008.

[14] E. Fogel and Y.-F. Huang, *On the value of information in system identification–bounded noise case*, Automatica, 18 (1982), pp. 229–238.

[15] M. Frenklach, *Transforming data into knowledge—Process Informatics for combustion chemistry*, Proc. Combust. Inst., 31 (2007), pp. 125–140, https://doi.org/10.1016/j.proci.2006.08.121.

[16] M. Frenklach, A. Packard, and R. Feeley, *Optimization of reaction models with solution mapping*, in Modeling of Chemical Reactions, R. W. Carr, ed., Elsevier, 2007, pp. 243–291.

[17] M. Frenklach, A. Packard, G. Garcia-Donato, R. Paulo, and J. Sacks, *Comparison of statistical and deterministic frameworks of uncertainty quantification*, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 875–901, https://doi.org/10.1137/15M1019131.

[18] M. Frenklach, A. Packard, and P. Seiler, *Prediction uncertainty from models and data*, in Proceedings of the 2002 American Control Conference, Vol. 5, IEEE, 2002, pp. 4135–4140.

[19] M. Frenklach, A. Packard, P. Seiler, and R. Feeley, *Collaborative data processing in developing predictive models of complex reaction systems*, Int. J. Chem. Kinet., 36 (2004), pp. 57–66.

[20] A. Garulli, A. Vicino, and G. Zappa, *Conditional central algorithms for worst case set-membership identification and filtering*, IEEE Trans. Automat. Control, 45 (2000), pp. 14–23.

[21] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, Chapman and Hall/CRC, 2013.

[22] M. Goldstein and J. Rougier, *Reified Bayesian modelling and inference for physical systems*, J. Statist. Plann. Inference, 139 (2009), pp. 1221–1239.

[23] M. Grant and S. Boyd, *Graph implementations for nonsmooth convex programs*, in Recent Advances in Learning and Control, V. Blondel, S. Boyd, and H. Kimura, eds., Lecture Notes in Control and Inform. Sci. 371, Springer, 2008, pp. 95–110, https://doi.org/10.1007/978-1-84800-155-8_7.

[24] M. Grant and S. Boyd, *CVX: MATLAB Software for Disciplined Convex Programming, Version* 2.1, http://cvxr.com/cvx, Mar. 2014.
[25] A. Hegde, W. Li, J. Oreluk, A. Packard, and M. Frenklach, *Consistency analysis for massively inconsistent datasets in bound-to-bound data collaboration*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 429–456, https://doi.org/10.1137/16M1110005.
[26] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne, *Combining field data and computer simulations for calibration and prediction*, SIAM J. Sci. Comput., 26 (2004), pp. 448–466, https://doi.org/10.1137/S1064827503426693.
[27] S. Iavarone, J. Oreluk, S. T. Smith, A. Hegde, W. Li, A. Packard, M. Frenklach, P. J. Smith, F. Contino, and A. Parente, *Application of bound-to-bound data collaboration approach for development and uncertainty quantification of a reduced char combustion model*, Fuel, 232 (2018), pp. 769–779.
[28] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Interval analysis*, in Applied Interval Analysis, Springer, 2001, pp. 11–43.
[29] V. R. Joseph and S. N. Melkote, *Statistical adjustments to engineering models*, J. Qual. Technol., 41 (2009), pp. 362–375.
[30] M. C. Kennedy and A. O'Hagan, *Bayesian calibration of computer models*, J. R. Stat. Soc. Ser. B Stat. Methodol., 63 (2001), pp. 425–464.
[31] W. Kleiber, S. R. Sain, M. J. Heaton, M. Wiltberger, C. S. Reese, and D. Bingham, *Parameter tuning for a multi-fidelity dynamical model of the magnetosphere*, Ann. Appl. Stat., 7 (2013), pp. 1286–1310.
[32] W. Li, A. Hegde, J. Oreluk, A. Packard, and M. Frenklach, *B2BDC Toolbox*, https://github.com/B2BDC/B2BDC/, 2017.
[33] M. Milanese and A. Vicino, *Optimal estimation theory for dynamic systems with set membership uncertainty: An overview*, Automatica, 27 (1991), pp. 997–1009.
[34] R. Myers, D. Montgomery, and C. Anderson-Cook, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley Series in Probability and Statistics, Wiley, 2009, https://books.google.fi/books?id=89oznEFHF_MC.
[35] W. L. Oberkampf and C. J. Roy, *Verification and Validation in Scientific Computing*, Cambridge University Press, 2010.
[36] J. Oreluk, Z. Liu, A. Hegde, W. Li, A. Packard, M. Frenklach, and D. Zubarev, *Diagnostics of data-driven models: Uncertainty quantification of pm7 semi-empirical quantum chemical method*, Sci. Rep., 8 (2018), 13248.
[37] J. Pedel, J. N. Thornock, and P. J. Smith, *Ignition of co-axial turbulent diffusion oxy-coal jet flames: Experiments and simulations collaboration*, Combust. Flame, 160 (2013), pp. 1112–1128, https://doi.org/10.1016/j.combustflame.2013.01.022.
[38] N. D. Perić, R. Paulen, M. E. Villanueva, and B. Chachuat, *Set-membership nonlinear regression approach to parameter estimation*, J. Process. Control, 70 (2018), pp. 80–95.
[39] M. Plumlee, *Bayesian calibration of inexact computer models*, J. Amer. Statist. Assoc., 112 (2017), pp. 1274–1285.
[40] P. Z. Qian and C. J. Wu, *Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments*, Technometrics, 50 (2008), pp. 192–204.
[41] T. Russi, A. Packard, R. Feeley, and M. Frenklach, *Sensitivity analysis of uncertainty in model prediction*, J. Phys. Chem. A, 112 (2008), pp. 2579–2588, https://doi.org/10.1021/jp076861c.
[42] T. Russi, A. Packard, and M. Frenklach, *Uncertainty quantification: Making predictions of complex reaction systems reliable*, Chem. Phys. Lett., 499 (2010), pp. 1–8.
[43] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, *Design and analysis of computer experiments*, Statist. Sci., 4 (1989), pp. 409–435.
[44] F. Schweppe, *Recursive state estimation: Unknown but bounded errors and system inputs*, IEEE Trans. Automat. Control, 13 (1968), pp. 22–28.
[45] P. Seiler, M. Frenklach, A. Packard, and R. Feeley, *Numerical approaches for collaborative data processing*, Optim. Eng., 7 (2006), pp. 459–478.
[46] N. A. Slavinskaya, M. Abbasi, J. H. Starcke, R. Whitside, A. Mirzayeva, U. Riedel, W. Li, J. Oreluk, A. Hegde, A. Packard, M. Frenklach, G. Gerasimov, and O. Shatalov, *De-*

velopment of an uncertainty quantification predictive chemical reaction model for syngas combustion, Energy Fuels, 31 (2017), pp. 2274–2297, https://doi.org/10.1021/acs.energyfuels.6b02319.

[47] G. P. Smith, M. Frenklach, R. Feeley, A. Packard, and P. Seiler, *A system analysis approach for atmospheric observations and models: Mesospheric HO_x dilemma*, J. Geophys. Res., 111 (2006), https://doi.org/10.1029/2005JD006846.

[48] G. P. Smith, D. M. Golden, M. Frenklach, N. W. Moriarty, B. Eiteneer, M. Goldenberg, C. T. Bowman, R. K. Hanson, S. Song, J. W. C. Gardiner, V. V. Lissianski, and Z. Qin, *GRI-Mech* 3.0, http://combustion.berkeley.edu/gri-mech/.

[49] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Ser. Statist., Springer, 1999.

[50] C. B. Storlie, W. A. Lane, E. M. Ryan, J. R. Gattiker, and D. M. Higdon, *Calibration of computational models with categorical parameters and correlated outputs via Bayesian smoothing spline anova*, J. Amer. Statist. Assoc., 110 (2015), pp. 68–82.

[51] J. F. Sturm, *Using SeDuMi* 1.02*, a MATLAB toolbox for optimization over symmetric cones*, Optim. Methods Softw., 11 (1999), pp. 625–653.

[52] E. Walter and H. Piet-Lahanier, *Estimation of parameter bounds from bounded-error data: A survey*, Math. Comput. Simul., 32 (1990), pp. 449–468.

[53] S. Wang, W. Chen, and K.-L. Tsui, *Bayesian validation of computer models*, Technometrics, 51 (2009), pp. 439–451.

[54] F. A. Williams, *Detailed and reduced chemistry for hydrogen autoignition*, J. Loss Prevent. Proc. Ind., 21 (2008), pp. 131–135.

[55] T.-M. Yi, M. Fazel, X. Liu, T. Otitoju, J. Goncalves, A. Papachristodolou, S. Prajna, and J. Doyle, *Application of robust model validation using SOSTOOLS to the study of G-protein signaling in yeast*, in Proceedings of the International Conference on Foundations of Systems Biology in Engineering (FOSBE), 2005, pp. 133–136, https://orbilu.uni.lu/handle/10993/20437.

[56] X. You, A. Packard, and M. Frenklach, *Process informatics tools for predictive modeling: Hydrogen combustion*, Int. J. Chem. Kinet., 44 (2012), pp. 101–116.