International Journal for Uncertainty Quantification, 5 (5): 469–489 (2016)

RANDOM PREDICTOR MODELS FOR RIGOROUS UNCERTAINTY QUANTIFICATION

Luis G. Crespo, [*] ***Sean P. Kenny, & Daniel P. Giesy***

NASA Langley Research Center, MS 308, Hampton, Virginia 23681, USA

[*] Address all correspondence to Luis G. Crespo E-mail: Luis.G.Crespo@nasa.gov

Original Manuscript Submitted: 04/15/2015; Final Draft Received: 09/30/2015

This paper proposes techniques for constructing linear parametric models describing key features of the distribution of an output variable given input-output data. By contrast to standard models, which yield a single output value at each value of the input, random predictors models (RPMs) yield a random variable. The strategies proposed yield models in which the mean, the variance, and the range of the model's parameters, thus, of the random process describing the output, are rigorously prescribed. As such, these strategies encompass all RPMs conforming to the prescription of these metrics (e.g., random variables and probability boxes describing the model's parameters, and random processes describing the output). Strategies for calculating optimal RPMs by solving a sequence of optimization programs are developed. The RPMs are optimal in the sense that they yield the tightest output ranges containing all (or, depending on the formulation, most) of the observations. Extensions that enable eliminating the effects of outliers in the data set are developed.When the data-generating mechanism is stationary, the data are independent, and the optimization program(s) used to calculate the RPM is convex (or, when its solution coincides with the solution to an auxiliary convex program), the reliability of the prediction, which is the probability that a future observation would fall within the predicted output range, is bounded rigorously using Scenario Optimization Theory. This framework does not require making any *assumptions on the underlying structure of the data-generating mechanism.*

KEY WORDS: *uncertainty quantification, representation of uncertainty, stochastic response surface method, random process, parameter estimation, optimization*

1. INTRODUCTION

Metamodeling [1] refers to the process of creating a mathematical representation of a phenomenon based on input-output data. These models can be parametric (e.g., polynomial response surfaces, polynomial chaos expansions, bootstrapping techniques) or no-parametric (e.g., smoothing spline models, Kriging/Gaussian process models). In the parametric case, the analyst first prescribes the model's structure and then determines the value of the model's parameters such that a measure of the discrepancy between observations and predictions is minimized. This step is commonly referred to as model calibration or regression. Model-form uncertainty (i.e., uncertainty caused by the offset between the structure of the computational model and the structure of the data-generating mechanism), measurement noise, and numerical error often inhibit confidently prescribing a fixed constant value for such parameters. Consequently, it is preferable to prescribe a set of parameter values such that the collective prediction that results from evaluating the model at each set member accurately represents the ensemble of observations.

Several model calibration techniques are available in the literature. Most of them assume the structure

$$y = M(x, p) + \eta, \tag{1}$$

where $y \in \mathbb{R}^{n_y}$ is the *output*, M is a continuous function of its arguments, $x \in \mathbb{R}^{n_x}$ is the *input*, $p \in \mathbb{R}^{n_p}$ is a *parameter* or regression coefficient, and $\eta \in \mathbb{R}^{n_y}$ is a random variation caused by noise and measurement error.

2152–5080/15/$35.00

Traditionally, the realizations of the random error are assumed to be independent and identically distributed (IID) following a Normal distribution. A typical regression problem consists of estimating the value of p given the set of observations (x_i, y_i), for $i = 1, \ldots, N$, where $N > n_p$. A key assumption in this model structure is that measurement error is the only cause of discrepancy between the observations and the noise-free prediction (so there is no model-form uncertainty).

Parameter estimation is commonly carried out by solving for the parameter realization that minimizes the sum of squared errors between predictions and observations [2]. This approach yields the least squares (LS) parameter estimate $\hat{\mu}$. The precision of this estimate, which prescribes how much it can deviate from its "true value" within an epistemic framework (i.e., the true value of p is fixed and unknown), is often evaluated using confidence intervals. The calculation of confidence intervals [2] and prediction intervals requires a probabilistic description of p. In linear regression statistics, a prediction interval defines a range of values within which the output is likely to fall in given a specified value of the input. Linearly regressed data are often non-normally distributed. Normally distributed data are statistically independent of one another, whereas regressed data are dependent on x. The uncertainty represented by a prediction interval includes not only the uncertainties associated with the population mean and the new observations, but also with the uncertainty associated with the regression parameter $\hat{\mu}$. Because the uncertainties associated with the population mean and new observation are independent of the observations used to fit the model, the uncertainty estimates of these three sources are combined. In general, the calculation of confidence and prediction intervals often requires (i) assuming a distribution for η, (ii) the predicted y having a mathematically convenient dependency on p and η, or (iii) a nonlinear M being accurately represented by a linear approximation. As expected, the suitability of the resulting predictions depend tightly on the validity of such assumptions.

A common approach to model calibration is Bayesian inference. In Bayesian inference, the objective is to describe the model's parameters as a vector of possibly dependent random variables using Bayes' rule. The resulting vector, called the posterior, depends on an assumed prior random vector and the likelihood function, which in turn depends on the observations and on the structure of M. Whereas this approach does not make any limiting assumptions on the manner in which M depends on p, nor on the structure of the resulting posterior, it requires that the calibrated variables in p be epistemic. This vector might be comprised of physical epistemic uncertainties and hyperparameters of aleatory variables.[1] Note that the consideration of aleatory uncertainties requires assuming a structure for them so they can be parameterized in terms of nonphysical epistemic variables. The presence of aleatory and model-form uncertainty yields uncertainty characterizations that fail to describe the prediction error (i.e., the offset between the observations and the prediction resulting from a calibrated model). This deficiency can be mitigated by adding a fictitious discrepancy term to M [3]. This term, which can have a fixed epistemic or a fixed aleatory structure, is calibrated as if it were part of M. In spite of its high computational demands, and of the potentially high sensitivity of the posterior to the assumed prior, this method is commonly regarded as a benchmark.

Bayesian calibration is often applied to models M having a physics-based structure. In contrast, this paper yields data-based models having a linear parameter dependency. This structure enables a rigorous treatment. Extensions to models having an arbitrary structure, including physics-based models whose parameters are real numbers, are made in [4]. In this paper, we do not use a measurement error term such as η, nor do we make prior assumptions about a distribution of p. What is here called a random predictor model (RPM) has the general form $y = M(x, p)$, where p is a random vector, so the output y is a random process parameterized by x. We do not fully specify the distribution of p. Instead, we only seek to find an expected value, a covariance matrix, and, in some cases, a support set for p.

Making the prediction match the observations by adjusting the hyperparameters of a distribution of p is a long-standing approach used in reliability-based design optimization, moment matching algorithms, and backward propagation of variance [5–8]. In this paper, these hyperparameters are determined by solving optimization programs according to the input-output data available and a few design parameters chosen by the analyst. The role of these parameters is to limit the largest number of standard deviations that can separate the measured outputs from the mean function. The resulting description of p is chosen to be as tight as possible while satisfying this restriction. We further

[1] For instance, if q contains the physical parameters of the model M, where q_1 is epistemic and q_2 is aleatory, having a normal distribution with mean μ and standard deviation σ, the vector $p = [q_1, \mu, \sigma]^\top$ contains three epistemic variables, one physical and two nonphysical.

provide means of identifying outliers in the data set so that eliminating them from the modeling process can result in predictions having a narrower output range at the expense of a reduction in the reliability of the prediction. Such a reduction can be formally quantified using the scenario approach (see Section 5). This article extends the interval predictor models (IPMs) proposed in [9], for which p is prescribed as a bounded set, so p is prescribed as a random vector. The developments herein enable generating random descriptions of p, and thus of y, having an arbitrary structure. This structure can be a random vector (e.g., p can be Gaussian or a generalized beta), or a probability box (e.g., all random vectors having a fixed expected value, variance, and support set). As such, the resulting characterization of p is distribution free.

As in the Bayesian inference approach, the formulations proposed yield a probabilistic description of the model's parameters. In contrast to the Bayesian approach, however, the proposed methods do not require any prior description of the uncertainty in p, and the resulting models yield analytical characterizations for both the predicted output and the model's parameters. This paper focuses on computational models having a linear dependency on p and an arbitrary dependency on x. Furthermore, the support of the probability density function characterizing p will be prescribed as a hyper-rectangular set. The advantage of these sets is that each component of p can be selected arbitrarily in its interval independently of the choices made for any of the other parameters. As such, parameter interdependencies are avoided. This independence enables the calculation of RPMs whose parameters are independent random variables.

This paper is organized as follows. Section 2 describes the problem statement and main objectives of this article. Section 3 presents the mathematical framework for calculating IPMs. These models play an instrumental role in the calculation of some RPMs. Section 4 presents formulations for calculating RPMs having various levels of fidelity and insensitivity to outliers. The reliability of the these models is studied in Section 5. Finally, Section 6 presents a few concluding remarks.

2. PROBLEM STATEMENT

A data generating mechanism (DGM) is postulated to act on a vector of input variables, $x \in \mathbb{R}^{n_x}$, to produce an output, $y \in \mathbb{R}^{n_y}$. In the following, the focus will be on the single-output ($n_y = 1$), multi-input ($n_x \geq 1$) case. The dependency of the output on the input is entirely arbitrary. This covers the case in which y is a function of x (so there is one output value for each input value) and the case in which y is a random process parameterized by x (so there are infinitely many output values for each input value). Assume that N input-output pairs are obtained from a DGM, and denote by $\mathsf{z} = \{z_i\}$, with $z_i = (x_i, y_i)$ for $i = 1, \ldots, N$, the corresponding data sequence.

It is desired to build a mathematical model of the DGM based on z that will predict the output corresponding to an unobserved realization of the input. Let $X \subseteq \mathbb{R}^{n_x}$ a set of input variables, and $Y \subseteq \mathbb{R}^{n_y}$ be a set of outputs which might result from evaluating the model at elements of X. The presence of intrinsic variability, and parametric- and model-form uncertainty makes it unrealistic to build a model that will predict a single output for a fixed input. Instead, an IPM will predict an interval-valued function into which the output from an unobserved input is expected to fall, while an RPM will predict a random process matching key features of the data. Engineering judgment is used to select a computational model $y = M(x, p)$, where $p \in \mathbb{R}^{n_p}$ is a parameter vector. Instead of the standard practice of trying to match all the data as closely as possible with M evaluated at a single vector p of parameters, the thrust in this work is to restrict as much as possible a set in $\mathbb{R}^{n_p}$ from which p is chosen while, at the same time, having the property that each data point in z (except, possibly, for a few outliers neglected purposely by the analyst) can be fit exactly by the model evaluated at least one element of p in such a set.

One restriction to be considered is for p to belong to a set P. For a fixed value of the input x, the propagation of P through M yields an interval of output values. Thus these models are called interval predictor models. The objective here is to choose P to make the corresponding y intervals as small as possible and still allow each data point (x_i, y_i) to be modeled as $y_i = M(x_i, p)$ for some $p \in P$. The other form of restriction considered is to describe p as a random vector. For a fixed value of the input x, the propagation of this vector through M yields a random variable for the the outcome y. Various properties of $R_y(x)$, such as its moments and support set, are determined by those of p. The thrust here is to choose a random vector that leads to a prediction matching key features of the data.

In this setting the two main problems of interest can be stated as follows. First, we want to find an empirical model that, when evaluated at a new value x_{N+1} of the input, returns an informative prediction of the unobserved output

y_{N+1}. An informative prediction can be interpreted as a narrow interval that is consistent with salient features of the data comprising z. These features, which are prescribed by the analyst as design requirements (for example, we might want all observed outcomes to be less than 2 standard deviations from the mean function), are cast as inequality constraints in the optimization programs used to create the model. Second, we want to quantify the probability of y_{N+1} being compliant with such requirements. (In the previous example, we want to evaluate the probability that y_{N+1} is less than 2 standard deviations away from the mean function.) In this setting, the targeted prediction is a narrow output interval of high probability. Note that the second objective implies that the prediction must conform to the DGM for *any* value of N without having any knowledge about its underlying structure.[2]

3. INTERVAL PREDICTOR MODELS

This section introduces basic concepts from IPMs that are essential for the construction of RPMs. Additional information on IPMs and examples are available in [9]. An IPM is simply a mapping that assigns an output interval for each value of the input. In the context of this paper, an IPM assigns to each instance vector $x \in X$ a corresponding outcome interval in Y. That is, an IPM is a set-valued map,

$$I_y : x \to I_y(x) \subseteq Y, \tag{2}$$

where $I_y(x)$ is the prediction interval. Depending on context, the term IPM will refer to either the function I_y or its graph $\{(x, y) : x \in X, y \in I_y(x)\}$ in $X \times Y$. Let M be any functional acting on a vector x of inputs and a vector p of parameters to produce an output y; i.e., $y = M(x, p)$. A parametric IPM is obtained by associating to each $x \in X$ the set of outputs y corresponding to all values of p in P:

$$I_y(x, P) = \{y = M(x, p),\ p \in P\}. \tag{3}$$

$I_y(x, P)$ will be an interval as long as $M(x, p)$ is a continuous function of x and p, and P is a connected set. All instances of M and P considered in this paper satisfy these restrictions. Attention will be limited to the IPM given by

$$I_y(x, P) = \{y = p^\top \varphi(x),\ p \in P\}, \tag{4}$$

where $\varphi(x)$ is an arbitrary basis, and p is a member of the hyper-rectangular uncertainty set

$$P = \{p : \underline{p} \le p \le \overline{p}\}. \tag{5}$$

The parameter points $\underline{p}$ and $\overline{p}$ are called the defining vertices of P. This model structure enables describing the IPM as

$$I_y(x, P) = \left[\, \underline{y}(x, \overline{p}, \underline{p}),\ \ \overline{y}(x, \overline{p}, \underline{p}) \,\right], \tag{6}$$

where

$$\underline{y}(x, \overline{p}, \underline{p}) = \overline{p}^\top \left(\frac{\varphi(x) - |\varphi(x)|}{2} \right) + \underline{p}^\top \left(\frac{\varphi(x) + |\varphi(x)|}{2} \right), \tag{7}$$

$$\overline{y}(x, \overline{p}, \underline{p}) = \overline{p}^\top \left(\frac{\varphi(x) + |\varphi(x)|}{2} \right) + \underline{p}^\top \left(\frac{\varphi(x) - |\varphi(x)|}{2} \right). \tag{8}$$

The functions $\underline{y}$ and $\overline{y}$ are, respectively, the lower and upper boundaries of the IPM. Each member of the family of infinitely many functions that result from evaluating the model M at each realization $p \in P$ lies between them, and no tighter containing functions exist. Observe that the IPM boundaries are not members of such a family when $\varphi(x)$ changes sign. The IPM boundaries are linear functions of $\underline{p}$ and $\overline{p}$, and piecewise continuous functions of the input. As such, they will have derivative discontinuities on the hypersurfaces where $\varphi(x)$ changes sign. The spread of $I_y(x, P)$, which is the distance between the upper and lower boundaries, is

$$\delta_y(x, \overline{p}, \underline{p}) = (\overline{p} - \underline{p})^\top |\varphi(x)|. \tag{9}$$

[2]We will only assume that the DGM is stationary and the observations in z are IID.

The narrower the spread δ_y the more informative the IPM prediction. Note that the spread depends on the size of the uncertainty box P but is independent of its geometric center. Furthermore, notice that a reduction in the volume of P might yield a larger spread.

The particular case in which the basis is polynomial is considered next. A general representation of a multivariate polynomial basis is

$$\varphi(x) = \left[1, x^{i_2}, x^{i_3}, \ldots, x^{i_{n_p}}\right]^\top, \tag{10}$$

where $x = [x_1, \ldots, x_{n_x}]$ is the input, and the vector $i_j = [i_{j,1}, \ldots, i_{j,n_x}]$, with $i_j \neq i_k$ for $j \neq k$ having the exponents of the monomials. For a polynomial basis we have $\varphi(|x|) = |\varphi(x)|$, which further simplifies Eqs. (7)–(9).

The above equations fully specify an IPM given the uncertainty box P. A means to calculate P's leading to optimal IPMs is provided next.

3.1 Type-1 IPMs

In this formulation we seek an IPM given by Eqs. (4)–(9), where $P = \hat{P}$ is given by the solution to the following optimization program (OP):

Optimization Program 1 (OP1)**.** *The defining vertices of $\hat{P}$ are given by*

$$\left\{\hat{\underline{p}},\ \hat{\overline{p}}\right\} = \underset{u,\,v}{\operatorname{argmin}} \left\{\mathbb{E}_x[\delta_y(x,v,u)] : \underline{y}(x_i,v,u) \leq y_i \leq \overline{y}(x_i,v,u),\ u \leq v\right\}, \tag{11}$$

where $\mathbb{E}_x[\cdot]$ is the expected value operator with respect to the input x, and (x_i, y_i) for $i = 1, \ldots, N$ are the observations comprising z*.*

In this formulation we search for the uncertainty box P that minimizes the expected interval spread such that all the observed outputs are within the IPM. When x is a standard joint random vector, the cost function in (11) can be calculated analytically. Otherwise, the sample mean of δ_y based on the data in z should be used. The resulting IPM, which is calculated by solving the convex OP in (11), admits a rigorous reliability assessment (see Section 5). This assessment formally bounds the probability that a future observation will fall within $I_y(x, \hat{P})$.

The DGM is commonly approximated by the LS prediction, $y = \hat{\mu}^\top \varphi(x)$, where the LS parameter estimate $\hat{\mu}$ is given by

$$\hat{\mu} = (A^\top A)^{-1} A^\top [y_1, \ldots, y_N]^\top, \tag{12}$$

and $A_{i,j} = \varphi_j(x_i)$, for $i = 1, \ldots, N$ and $j = 1, \ldots, n_p$. The vector μ is the parameter value minimizing $\sum_{i=1}^{N} \left[y_i - p^\top \varphi(x_i)\right]^2$. The membership of $\hat{\mu}$ in $\hat{P}$ can be ensured by replacing the first constraint with $u \leq \hat{\mu} \leq v$ (i.e., $\hat{\mu} \in \hat{P}$), or adding the constraint $u + v = 2\hat{\mu}$ (i.e., $\hat{\mu}$ is the geometric center of $\hat{P}$). In general, the inclusion of these constraints leads to IPMs with larger expected spreads, with the equality constraint leading to the larger of the two. A formulation resulting from adding either of these two sets of constraints will be called Augmented OP1. Other types of IPMs are considered in [9].

4. RANDOM PREDICTOR MODELS

A RPM is a mapping that assigns to each input vector $x \in X$ a corresponding random variable in the output space Y. That is, a RPM is a random variable-valued map

$$R_y : x \to R_y(x) \subseteq Y, \tag{13}$$

where $R_y(x)$ is a random process having its support in Y. A parametric RPM is obtained by associating to each $x \in X$ the set of outputs y corresponding to all values of p described by a random vector with joint cumulative distribution function (CDF) $F_p(p)$ having P in (5) as its support set. As before, attention will be limited to the case where the output is linear in p. This leads to

$$R_y\left(x, F_p\right) = \{y = p^\top \varphi(x),\ p \sim F_p(p)\}. \tag{14}$$

Denote by $\mu \in \mathbb{R}^{n_p}$, $\nu \in \mathbb{R}^{n_p}$, and $c \in \mathbb{R}^{n_p(n_p-1)/2}$ the mean, variance, and correlation of p, respectively. The variance and correlation fully prescribe the covariance matrix $C(\nu, c) \in \mathbb{R}^{n_p \times n_p}$. It can be shown that any $F_p(p)$ supported in P must satisfy the consistency equations:

$$\underline{p} \leq \mu \leq \overline{p}, \tag{15}$$

$$0 \leq \nu \leq (\mu - \underline{p}) \odot (\overline{p} - \mu), \tag{16}$$

$$-1 \leq c \leq 1, \tag{17}$$

$$C(\nu, c) \succeq 0, \tag{18}$$

where the symbols $\odot$ and $\succeq$ denote the componentwise product of vectors and positive semidefiniteness, respectively.[3]

The random process $R_y(x, F_p)$ is fully prescribed by the model $y = p^\top \varphi(x)$ and $F_p(p)$. Naturally, key features of the prediction, such as statistical moments and its range, vary with x. In particular, the mean function is $\mu_y(x, \mu) = \mathbb{E}_p[y(x,p)] = \mu^\top \varphi(x)$, the variance function is

$$\nu_y(x, \nu, c) = \mathbb{E}_p\{[y(x,p) - \mu_y(x)]^2\} = \varphi(x)^\top C(\nu, c)\varphi(x), \tag{19}$$

and the interval-valued range or support function is given by (6). When the components of p are uncorrelated, Eq. (19) reduces to[4]

$$\nu_y(x, \nu) = \nu^\top \varphi^2(x). \tag{20}$$

A few metrics for characterizing $R_y(x)$ are introduced next. The σ surface, which connects all the outputs y that are τ standard deviations from the mean function, is defined as

$$s_\sigma(x, \mu, \tau, \nu, c) = \mu^\top \varphi(x) + \tau\sqrt{\nu_y(x, \nu, c)}, \tag{21}$$

where $\tau > 0$ corresponds to deviations above the mean and $\tau < 0$ to deviations below. The σ volume, defined as

$$v_\sigma(x, \mu, \tau, \nu, c) = \left[s_\sigma(x, \mu, -\tau, \nu, c),\ s_\sigma(x, \mu, \tau, \nu, c)\right], \tag{22}$$

is an interval-valued function that contains all the outputs y that are no more than τ standard deviations away from the mean function $\mu_y(x)$. For the value of τ to be feasible; i.e., for the σ surface to be within the support of $R_y(x, F_p)$, it must satisfy

$$\underline{y}(x, \overline{p}, \underline{p}) \leq s_\sigma(x, \mu, \tau, \nu, c) \leq \overline{y}(x, \overline{p}, \underline{p}). \tag{23}$$

Equation (23) ensures that the support of the process contains outcomes that are up to τ standard deviations from the mean function. Note that the range of τ values (i.e., range of standard deviations) satisfying these inequalities is a function of x.

The formulations that follow prescribe key features of F_p, thus of the random process $R_y(x, F_p)$, based on input-output data. As such they encompass all RPMs conforming to such features. Four types of RPMs are proposed. Type-1 RPMs prescribe the mean and variance of $R_y(x)$ when the entire data set is used. Type-2 RPMs prescribe the same statistics after eliminating the effects of a fixed percentage of the observations (i.e., outliers). Such observations are worst case in the sense that their removal tightens the σ volume the most. Type-3 and type-4 RPMs not only prescribe the mean and variance, but also the support P. Whereas type-3 RPMs emphasize the tightness of the range of y, type-4 RPMs emphasize the tightness of the σ volume. In contrast to type-1 and type-2 RPMs, which only require solving two OPs (one for μ and another one for ν), type-3 and type-4 RPMs require solving a sequence of three interdependent OPs (one for each μ, ν, and P). A summary of the main features of all four RPMs is provided in Table 1.

The presentation that follows focuses on the uncorrelated case. This case renders convex OPs that enable having a large number of observations. Extensions to the correlated case can easily be made. In the developments that follow, the performance of an RPM refers to the property evaluated by the cost function in the corresponding OP.

[3]The upper bound in (16) results from applying the expected value operator $\mathbb{E}_{p_i}[\cdot]$ to both sides of $p_i^2 \leq (\underline{p}_i + \overline{p}_i)p_i - \underline{p}_i\overline{p}_i$, which holds for all $p_i \in [\underline{p}_i, \overline{p}_i]$, and then using $\nu_i = \mathbb{E}_{p_i}[p_i^2] - \mu_i^2$ for $i = 1, \ldots, n_p$.

[4]When the correlation c is zero, the corresponding argument of any function depending on it will be dropped from the notation.

TABLE 1: Performance function J, number of constraints c, and decision variable s for all RPM types

	First OP	Second OP	Third OP
Type-1 RPM	$J = \sum[y_i - p^\top \varphi(x_i)]^2$ $c = 0$ $s = \mu$	$J = \mathbb{E}_x[\nu_y(x)]$ $c = 2N + n_p$ $s = \nu$	N/A
Type-2 RPM	$J = \sum[y_i - p^\top \varphi(x_i)]^2$ $c = 0$ $s = \mu$	$J = \mathbb{E}_x[\nu_y(x)]$ $c = n_p + 1$ $s = \nu$	N/A
Type-3 RPM	$J = \sum[y_i - p^\top \varphi(x_i)]^2$ $c = 0$ $s = \mu$	$J = \mathbb{E}_x[\delta_y(x)]$ $c = 2(N + n_p)$ $s = \{\overline{p}, \underline{p}\}$	$J = \mathbb{E}_x[\nu_y(x)]$ $c = 1 + 2n_p$ $s = \nu$
Type-4 RPM	$J = \sum[y_i - p^\top \varphi(x_i)]^2$ $c = 0$ $s = \mu$	$J = \mathbb{E}_x[\nu_y(x)]$ $c = n_p + 1$ $s = \nu$	$J = \mathbb{E}_x[\delta_y(x)]$ $c = 2N + 3n_p$ $s = \{\overline{p}, \underline{p}\}$

4.1 Type-1 RPMs

Type-1 RPMs prescribe the expected value and variance functions of $R_y(x, F_p)$ based on the entire data set in z. A type-1 RPM is given by Eq. (14), where F_p has a expected value[5] $\mu = \hat{\mu}$ given by (12) and a variance $\nu = \hat{\nu}$ given by the following OP:

Optimization Program 2 (OP2)**.** *For a given the mean* μ*, the variance* ν *is equal to*

$$\hat{\nu} = \underset{\nu}{\operatorname{argmin}} \left\{ \mathbb{E}_x[\nu_y(x, \nu)] : s_\sigma(x_i, \mu, -\tau_{\max}, \nu) \le y_i \le s_\sigma(x_i, \mu, \tau_{\max}, \nu) \text{ for } i = 1, \ldots, N,\ \nu \ge 0 \right\}, \tag{24}$$

where $\tau_{\max} > 0$ *is a parameter prescribed by the analyst, and* (x_i, y_i) *for* $i = 1, \ldots, N$ *are the observations in* z*.*

Hence, a Type 1-RPM minimizes the expected value of the output's variance such that all observations are no more than $\tau_{\max}$ standard deviations away from the mean function; i.e., all observations are within the σ-volume $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu})$, such that the sum of the squares of the prediction errors relative to the mean function is minimal.

The dependence of $\hat{\nu}$ on $\tau_{\max}$ is studied next. Equation (24), which is subject to $2N + n_p$ inequality constraints, is equivalent to the linear program

$$\hat{\nu} = \underset{\nu}{\operatorname{argmin}} \left\{ \nu^\top \mathbb{E}_x\left[\varphi^2(x)\right] : \tau_{\max}^2 \nu^\top \varphi^2(x_i) \ge \left[y_i - \mu^\top \varphi(x_i)\right]^2 \text{ for } i = 1, \ldots, N,\ \nu \ge 0 \right\}, \tag{25}$$

which is subject to $N + n_p$ constraints. The constraint set in (25) scales inversely with $\tau_{\max}^2$, so the scaled optimal objective value

$$\mathcal{I} = \tau_{\max}^2 \hat{\nu}^\top \mathbb{E}_x[\varphi^2(x)] \tag{26}$$

is invariant with respect to $\tau_{\max}$. It follows that the larger $\tau_{\max}$, the smaller $\|\hat{\nu}\|$, and the larger the number of standard deviations separating any given point (x, y) from the mean function. If $\hat{\nu}_1$ is the solution to (25) corresponding to $\tau_{\max,1}$, and $\hat{\nu}_2 = \alpha\hat{\nu}_1$ where $\alpha = (\tau_{\max,1}/\tau_{\max,2})^2$, then $\hat{\nu}_2$ is the solution to (25) corresponding to $\tau_{\max,2}$. This implies that $v_\sigma(x, \mu, \tau_{\max,1}, \hat{\nu}_1) = v_\sigma(x, \mu, \tau_{\max,2}, \hat{\nu}_2)$, and $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu})$ is independent of the choice of $\tau_{\max}$.

A type-1 RPM does not prescribe the support of p, and thus of $R_y(x, P)$. Any random vector satisfying the consistency equations (15)–(18) for $\mu = \hat{\mu}$ and $\nu = \hat{\nu}$ is a valid characterization of $F_p(p)$. Note that both type-1

[5]The selection of μ as $\hat{\mu}$ is arbitrary, and any other value can be used instead. This applies to all OPs derived hereafter.

IPMs and type-1 RPMs require solving a convex OP. As such, they can efficiently handle hundreds of thousands of data points, thus many more input dimensions than alternative metamodels. Since type-1 RPMs are calculated by solving a convex OP, they admit a rigorous reliability assessment. This assessment, presented in Section 5, bounds the probability that a future observation will fall outside $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu})$.

Example 1. Consider the DGM $y = x^2\cos(x) - \sin(3x)e^{-x^2} - \cos(x^2) + x(g-1)$, where x is uniformly distributed over $X = [-5.5, 5.5]$, and g is a standard normal distribution.[6] A total of $N = 150$ independent observations from the DGM were used to form the data sequence z. We assume that M is a linear combination of sixth-order polynomials so $\varphi(x) = [1, x, x^2, x^3, x^4, x^5, x^6]^\top$, and $n_p = 7$. In [9] we calculate several IPMs based on the same setup, for which the LS parameter estimate is $\hat{\mu} = [-0.8734, -1.1059, -0.9926, 0.0026, -0.0228, -0.0004, 0.0028]^\top$.

A type-1 RPM for $\tau_{\max} = 1$, to be referred to as RPM A, is shown in Fig. 1. This figure shows the observations ($\times$'s), the mean function $\mu_y(x)$ (solid line), as well as σ surfaces (green dashed-dotted lines) in increments of 0.5 standard deviations. Note that the observation near $(1, -15)$ limits the σ volume from below. The only significant variance[7] in $\hat{\nu}$ is $\hat{\nu}_1$, so the performance of RPM A is $\mathbb{E}_x[\nu_y] = \hat{\nu}_1 = 180.3824$.

4.1.1 Outliers

The presence of a few low-probability data points deviating considerably from the rest of the observations will make the σ volume and uncertainty set P much larger, diminishing the RPM's performance. Whereas the limits of $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu})$ might be driven by a few observations, the majority of them might be much closer to the mean function, e.g., for RPM A above, only nine observations are outside $v_\sigma(x, \mu, 0.5, \hat{\nu})$, whereas the remaining 141 observations are inside. The removal of such points from the data set will lead to narrower, more informative predictions at the expense of a reduced RPM's reliability. These observations, to be called *outliers* hereafter, can be identified using any one of several figures of merit. This paper will use

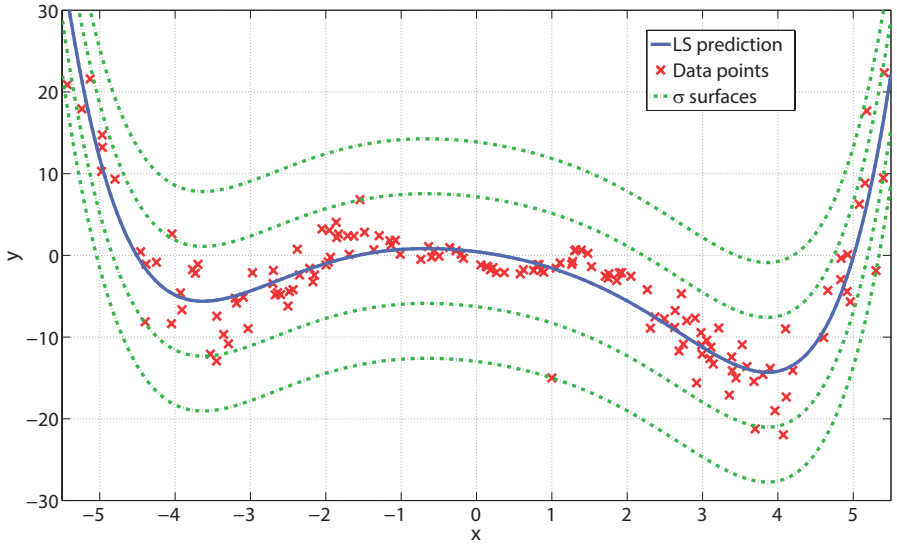

FIG. 1: RPM A: type-1 RPM for $\tau_{\max} = 1$.

[6]Note that no knowledge about DGM is required to calculate RPMs. This equation has been included solely for clarity in the presentation.

[7]For a given RPM, we might want to evaluate the contribution of individual terms in M to the resulting prediction. The term $\varphi_i(x)$ is insignificant when its contribution to the mean function, given by $\max_{x\in X}\{|\mu_i\varphi_i(x)|\}$, and its contribution to the variance, given by $\max_{x\in X}\{\nu_i\varphi_i^2(x)\}$, are sufficiently small. Terms satisfying both of these conditions, along with the conditions affecting $I_y(x, \hat{P})$, as explained in [9], can be removed from M without degrading the prediction.

$$\kappa_i(\mu, \nu, c) = \frac{\left(y_i - \mu^\top \varphi(x_i)\right)^2}{\nu_y(x_i, \nu)}. \tag{27}$$

The metric κ_i is a variance-normalized distance squared between the ith observed output and the mean function at the corresponding input. Outliers will be identified by determining the data points corresponding to the largest percentiles of the empirical CDF of κ, $F_{\kappa(\hat{\nu})}(\kappa)$, based on the N observations, i.e., (x_i, y_i) is an outlier if $F_{\kappa(\hat{\nu})}(\kappa_i) > \lambda$ where $0 \ll \lambda < 1$. Once the outliers are identified, they can be removed from the data sequence and a new type-1 RPM will be calculated. The resulting RPM will attain tighter predictions for a λ fraction of the observations in z, while the prediction for the remaining $1 - \lambda$ fraction might be considerably degraded. The outliers found by this procedure will be the same regardless of the value[8] of $\tau_{\max}$.

Example 2. We now derive a type-1 RPM for $\tau_{\max} = 1$ after removing seven outliers from the data set. These outliers attain the largest values of κ. The resulting RPM, to be referred to as RPM B, is shown in Fig. 2. In this case there are seven observations outside $v_\sigma(x, \hat{\mu}, 1, \hat{\nu})$ by design (shown with circled cross symbols), and 114 within the $v_\sigma(x, \hat{\mu}, 0.5, \hat{\nu})$. The only sizable variances for RPM B are $\hat{\nu}_1 = 44.5139$ and $\hat{\nu}_2 = 0.5194$. The performance of RPM B, $\mathbb{E}_x[\nu_y] = 49.2469$, is 72.7% better than that of RPM A. The approach to eliminate the effects of outliers used above requires the identification and removal of observations from the data set and the calculation of two RPMs. Conversely, the approach described next achieves the same objective without identifying or removing outliers and requires calculating only a single RPM.

4.2 Type-2 RPMs

A formulation leading to an alternative RPM is presented next. In contrast to type-I RPMs, this approach searches for ν by using only a fixed percentage of the N observations available. The observations comprising the neglected set are worst case in the sense that their removal tightens the optimal σ volume the most. Whereas the outliers removed to construct RPM B are worst case for the value of $\hat{\nu}$ corresponding to RPM A only, those neglected in a type-2 RPM are worst case for the varying value of ν being considered during the optimization. This will be carried out without removing any point from the data sequence in advance.

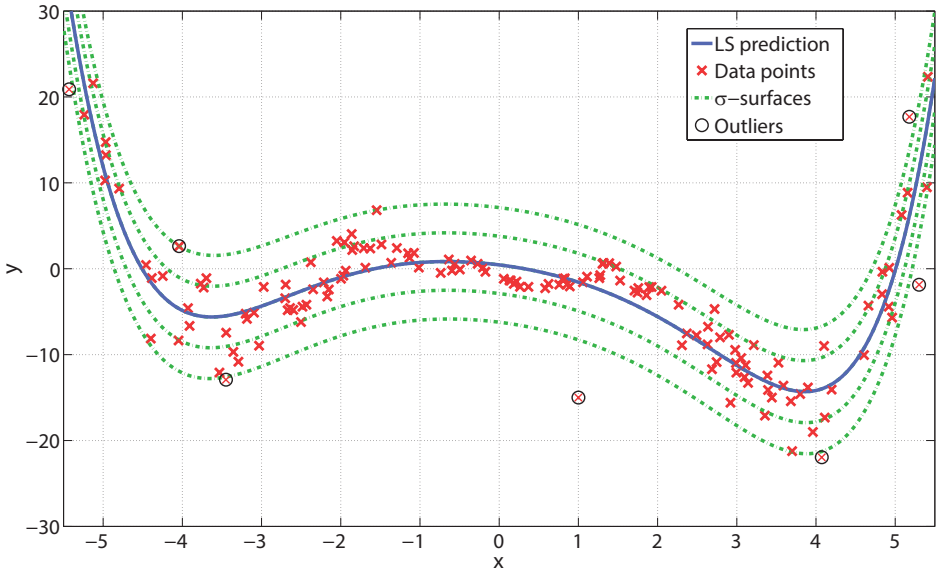

FIG. 2: RPM B: type-1 RPM after the removal of outliers.

[8]This is a consequence of the following observation. If $[\kappa_i, F_\kappa(\kappa_i)]$ are points on the optimal CDFs corresponding to $\tau_{\max,1}$, the points on the optimal CDF corresponding to $\tau_{\max,2}$ are $[\alpha\kappa_i, F_\kappa(\kappa_i)]$, where α was defined earlier.

In particular, a type-2 RPM is given by Eq. (14), where p has an expected value $\mu = \hat{\mu}$ given by (12) and a variance $\nu = \hat{\nu}$ given by the following OP:

Optimization Program 3 (OP3). *For a given the mean* μ*, the variance* ν *is equal to*

$$\hat{\nu} = \underset{\nu}{\operatorname{argmin}} \left\{ \mathbb{E}_x[\nu_y(x,\nu)] : F_{\kappa(\nu)}\left(\tau^2_{\max}\right) \geq \lambda,\ \nu \geq 0 \right\}, \tag{28}$$

where $\tau_{\max} > 0$ *is a parameter prescribed by the analyst,* $F_{\kappa(\nu)}$ *is the empirical CDF of* $\kappa(\nu)$ *in (27) based on the* N *observations in* z*, and* $0 < \lambda \leq 1$*, another parameter to be chosen by the analyst, is the proportion of observations to be contained by* $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu}(\lambda))$.

Hence, a Type-2 RPM minimizes the expected value of the output's variance such that a λ fraction of the observations are no more than $\tau_{\max}$ standard deviations apart from the mean function, such that the sum of the squares of the prediction errors relative to the mean function is minimal. The tightening of the prediction for such a fraction yields a σ volume $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu}(\lambda))$ that does not enclose the remaining $1 - \lambda$ fraction. This shows that (28) is a chance-constraint formulation [10], in which one is willing to accept the occurrence of unfavorable low-probability events (probability $1-\lambda$) for the sake of an improved performance for high-probability events (probability λ). As with type-1 RPMs, $\tau_{\max}$ is essentially a scaling factor.

OP3 is a nonconvex formulation. When $\lambda = 1$ the solution to OP3 and the solution to OP2, which is convex, are the same.[9] When $\lambda < 1$, a fixed number of observations (outliers) are neglected as the RPM is being calculated. Outliers can be easily identified by finding the data points for which $F_{\kappa(\hat{\nu})}\left(\kappa_i(\hat{\nu})\right) > \lambda$. The points violating this condition, which are the elements of z within $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu}(\lambda))$, constitute the sequence w. A type-1 RPM based on the data sequence w is equivalent to the type-2 RPM in (28) based on the data sequence z. This relationship enables performing a reliability assessment of type-2 RPMs. This assessment, presented in Section 5, formally bounds the probability that a future observation will fall outside $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu}(\lambda))$.

Example 3. We now derive a type-2 RPM for $\lambda = 143/150$ and $\tau_{\max} = 1$. As with RPM B, we want 143 observations to be less than one standard deviation away from the mean function. The resulting RPM, shown in Fig. 3, will be referred to as RPM C. Note that the process is more focused on the LS prediction than either RPM A or RPM B. The only sizable components of $\hat{\nu}$ are $\hat{\nu}_1 = 6.0124$, and $\hat{\nu}_2 = 3.2985$. Note that the outliers, falling outside

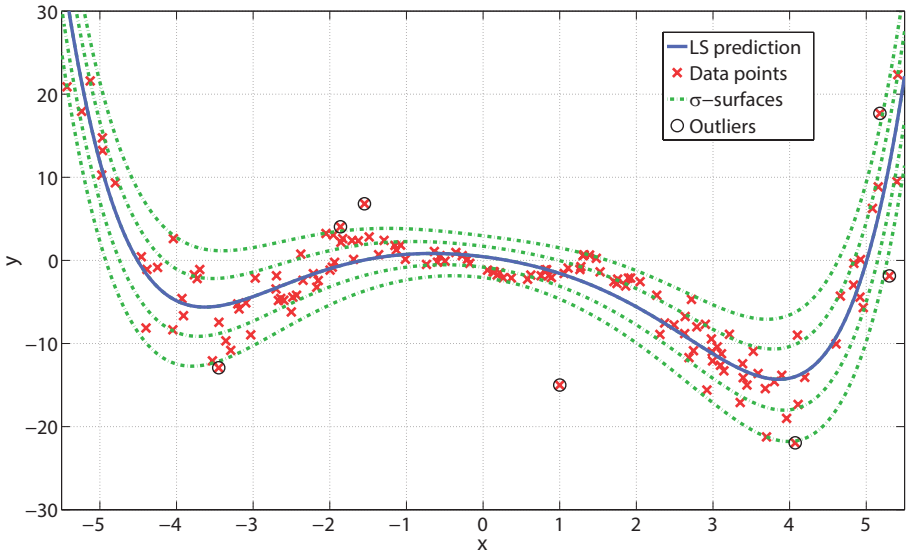

FIG. 3: RPM C: type-2 RPM for $\lambda = 143/150$ and $\tau_{\max} = 1$.

[9]Note that if $\mathbb{E}_x[\cdot]$ is calculated based on a sample mean, all N data points in z must be used to obtain equal solutions.

$v_\sigma[x, \hat{\mu}, 1, \hat{\nu}(\lambda)]$, differ from those of RPM B. The performance of RPM C, $\mathbb{E}_x[\nu_y] = 36.3341$, is 26% better than that of RPM B.

Figure 4 shows the empirical CDFs of $\omega = \kappa(\hat{\nu})\mathbb{E}_x[\nu_y(x, \hat{\nu})]$ for RPM A and RPM C. Note that the smaller the value of ω, the more concentrated the data about the mean function. The support of ω for RPM A is $[0, 180.4]$, whereas that for RPM B is $[0, 718.1]$. The upper limits of these intervals are shown as dotted lines. Hence, the concentration of the set of $N = 150$ data points about the mean is about four times higher for RPM A than that for RPM B. However, if we look at the quantile $\omega = F_\omega^{-1}(143/150)$, we note that the closest 143 data points to the mean for RPM B are more concentrated than those for RPM A. The range of ω corresponding to such points for RPM A is $[0, 48.32]$ whereas that for RPM B is $[0, 37.76]$. The upper limits of these intervals are shown as dashed lines. Therefore, when only 143 data points out of the 150 are considered, RPM B is about 30% better than RPM A. This illustrates that (28) is a chance-constraint formulation in which one is willing to accept a degraded performance (i.e., larger values of ω) for low-probability events (i.e., those occurring with probability $1 - \lambda$) for the sake of an improved performance (i.e., smaller values of ω) for high-probability events (i.e., those occurring with probability λ).

4.3 Type-3 RPMs

Type-3 RPMs prescribe the expected value, variance, and support of p, and thus of $R_y(x)$. In contrast to type-1 and type-2 RPMs, which require solving one OP for the mean and another one for the variance, a type-3-RPM requires solving a sequence of three OPs linked by the consistency equations (15)–(18). The additional OP is used to calculate the support P. The order of the sequence implies that the mean has priority over the support set, and the support set over the variance.

In particular, a type-3 RPM is defined by Eq. (14), where $\mu = \hat{\mu}$ is given by (12), $P = \hat{P}$ is given by an augmented version of (11), and the variance $\nu = \hat{\nu}$ is the solution to the following OP:

Optimization Program 4 (OP4). *For a given mean μ and a given uncertainty set P with defining vertices $\overline{p}$ and $\underline{p}$, the variance ν is equal to*

$$\hat{\nu} = \underset{\nu}{\operatorname{argmin}} \left\{ \mathbb{E}_x[\nu_y(x, \nu)] : F_{\kappa(\nu)}\left(\tau^2_{\max}\right) \geq \lambda,\ 0 \leq \nu \leq \nu_{\max} \right\}, \tag{29}$$

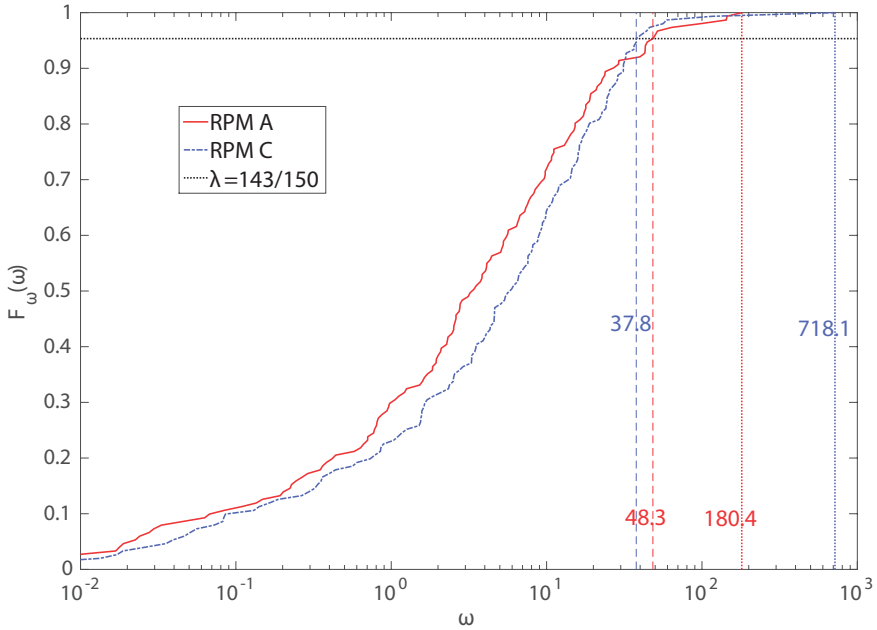

FIG. 4: Empirical CDFs of $\omega = \mathbb{E}_x[\nu_y]\kappa(\hat{\nu})$ for RPM A (red) and RPM C (blue).

where $\nu_{\max} = (\mu - \underline{p}) \odot (\overline{p} - \mu)$ *and* $F_{\kappa(\nu)}$ *is the empirical CDF of* $\kappa(\nu)$ *in (27) based on the* N *observations in* z*. The parameters* $\tau_{\max}$ *and* λ*, to be chosen by the analyst and defined earlier, must satisfy*

$$\tau_{\max} > \tau^*_{\max} = \max_{1 \le i \le N} \left\{ \frac{|y_i - \mu^\top \varphi(x_i)|}{\sqrt{\nu_y(x_i, \nu_{\max})}} \right\}, \tag{30}$$

and $0 < \lambda \le 1$.

Hence a type-3 RPM minimizes the expected value of the output's variance such that $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu}(\lambda))$ contains[10] $\lfloor \lambda N \rfloor$ observations, subject to (i) a P that minimizes the expected spread for a range $I_y(x, P)$ containing the full set of N observations, and (ii) a μ that minimizes the sum of the squares of the prediction errors relative to the mean function. Note that extreme observations from a set of N elements prescribe the support set $\hat{P}$ in OP1 according to δ_y, whereas extreme observations from a set of $\lfloor N\lambda \rfloor$ elements prescribe $\hat{\nu}$ in OP4 according to κ. The solution to (11) enters (29) via the upper bound on ν, $\nu_{\max}$. The constraint (30) ensures the feasible design space is nonempty. The ith component of the vector at the right-hand side of (30) is the absolute value of τ_i, where τ_i is the solution to $y_i = s_\sigma(x_i, \mu, \tau_i, \nu_{\max})$. Hence τ_i is the smallest number of standard deviations that can separate (x_i, y_i) from the mean function without letting ν exceed $\nu_{\max}$.

Whereas the augmented OP1 is convex, the inequality constraints in (29) make OP4 nonconvex. When $\lambda = 1$, such constraints are equivalent to the constraints in (25), so the solution to OP4 coincides with the solution to a convex OP. Therefore type-3 RPMs for the case in which $\lambda = 1$ can be found by solving a sequence of three convex OPs. When $\lambda < 1$, the constraints in (29) are equivalent to a subset of the constraints in (25). This subset is given by all the elements in z satisfying $F_{\kappa(\hat{\nu})}(\kappa_i) \le \lambda$. The $\lfloor N\lambda \rfloor$ observations satisfying this condition constitute the data sequence w. Therefore OP4, based on the data sequence z, renders the same empirical model as a convex OP based on the data sequence w. This is the basis used for bounding the reliability of type-3 RPMs. To this end (see Theorem 2), it is useful to determine if the containment condition $v_\sigma(x, \mu, \tau_{\max}, \nu) \subseteq I_y(x, P)$ holds for all $x \in X$, i.e., the range of $R_y(x, P)$ contains the σ volume corresponding to $\tau_{\max}$. This condition holds if and only if

$$(\overline{p} - \underline{p})^\top |\varphi(x)| - \left| (\overline{p} + \underline{p} - 2\mu)^\top \varphi(x) \right| - 2\tau_{\max} \sqrt{\varphi(x)^\top C(\nu, c) \varphi(x)} \ge 0, \tag{31}$$

$\forall x \in X$. Type-3 RPMs satisfying this semi-infinite constraint allow for a tighter reliability bound. Enforcing the containment condition by design requires incorporating (31) into (29).

Example 4. Two type-3 RPMs based on the same setup used earlier are derived next. Whereas the two RPMs differ in the value of λ used to calculate $\hat{\nu}$, both use the same support set $\hat{P}$. This set is calculated via an augmented OP1 so $\hat{\mu} \in \hat{P}$. This leads to $\underline{\hat{p}} = [-12.9837, -1.1488, -0.8339, 0.0013, -0.0379, -0.0001, 0.0032]^\top$, and $\hat{\overline{p}} = [7.2080, -1.1488, -0.8339, 0.0013, -0.0379, -0.0001, 0.0034]^\top$, and $\mathbb{E}_x[\delta_y] = 10.4942$. These values, in turn, yield an upper bound for ν where the only significant component is $\nu_{\max,1} = 90.8037$. The bound on $\tau_{\max}$ resulting from (30) yields $\tau^*_{\max} = 1.4094$. Thus we selected $\tau_{\max} = 1.5$.

A type-3 RPM for $\lambda = 1$ is calculated first. Therefore we require that all 150 observations be no more than 1.5 standard deviations from the mean function. The resulting RPM, to be referred to as RPM D and shown in Fig. 5, leads to a variance $\hat{\nu}$ for which the only significant term is $\hat{\nu}_1 = 80.1699$. The performance of RPM D is given by both $\mathbb{E}_x[\delta_y] = 10.4942$ and $\mathbb{E}_x[\nu_y] \approx \hat{\nu}_1$. Whereas the boundaries of $I_y(x, \hat{P})$ are shown as dashed black lines, σ surfaces separated by 0.5 units are shown as dashed-dotted green lines. Note that the augmented constraint yielded a mean function that deviates considerably from the center of $I_y(x, \hat{P})$. Furthernore, notice that the lower limit of the support coincides with the σ surface $s_\sigma(x, \hat{\mu}, -1.5, \hat{\nu})$ even though the functions have different functional forms. Conversely, the values of τ on (21) for which the corresponding σ surface coincide with $\overline{y}(x, \hat{\overline{p}}, \underline{\hat{p}})$ vary. Even though the portions of the σ-surfaces $s_\sigma(x, \hat{\mu}, \tau, \hat{\nu})$ spreading outside $I_y(x, \hat{P})$ are infeasible, e.g., almost the entire $s_\sigma(x, \hat{\mu}, -1.5, \hat{\nu})$, they are plotted for clarity. The feasible range of τ values at each value of x is given by (23). Because the majority of the observations are at the center of $I_y(x, \hat{P})$, neglecting a few outliers will considerably tighten the prediction.

[10]The floor operator $\lfloor \cdot \rfloor$ is the greatest integer less or equal than or eqval to its argument.

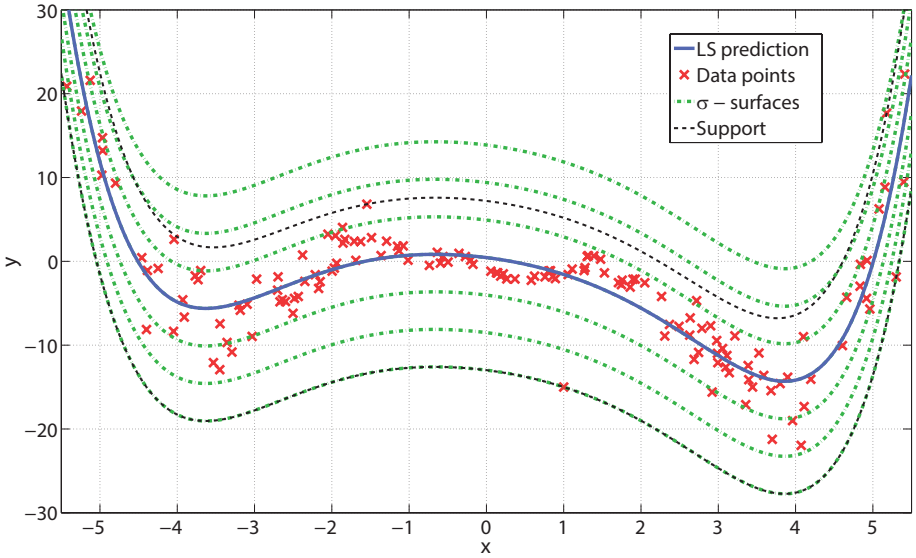

FIG. 5: RPM D: type-3 RPM for $\tau_{\max} = 1.5$ and $\lambda = 1$.

A type-3 RPM for $\lambda = 143/150$ is derived next. Therefore we require that 143 observations be no more than 1.5 standard deviations from the mean function. This model, to be referred to as RPM E, leads to a variance $\hat{\nu}$ for which $\hat{\nu}_1 = 22.2497$ is the only significant term. The performance of RPM E is given by $\mathbb{E}_x[\delta_y] = 10.4942$ as before, and by $\mathbb{E}_x[\nu_y] \approx \hat{\nu}_1$. In terms of the latter metric, RPM E is 3.6 times better than RPM D. Figure 6 shows σ surfaces corresponding to RPM E being 0.5 units apart. The same line conventions used before apply. A comparison between Figs. 5 and 6 indicates that RPM E yields a tighter probabilistic description for $100\lambda\%$ of the observations than RPM D. The containment condition in (31) is not satisfied by either RPM D or RPM E. This is reflected in Figs. 5 and 6, where $s_\sigma(x, \hat{\mu}, 1.5, \hat{\nu}) > \overline{y}(x, \hat{\overline{p}}, \hat{\underline{p}})$ for some x in X.

The sequential construction of a type-3 RPM, where the variance ν is solved for after solving for the support set P, restricts its probabilistic performance (i.e., the variance is calculated given an optimal P). This restriction manifests

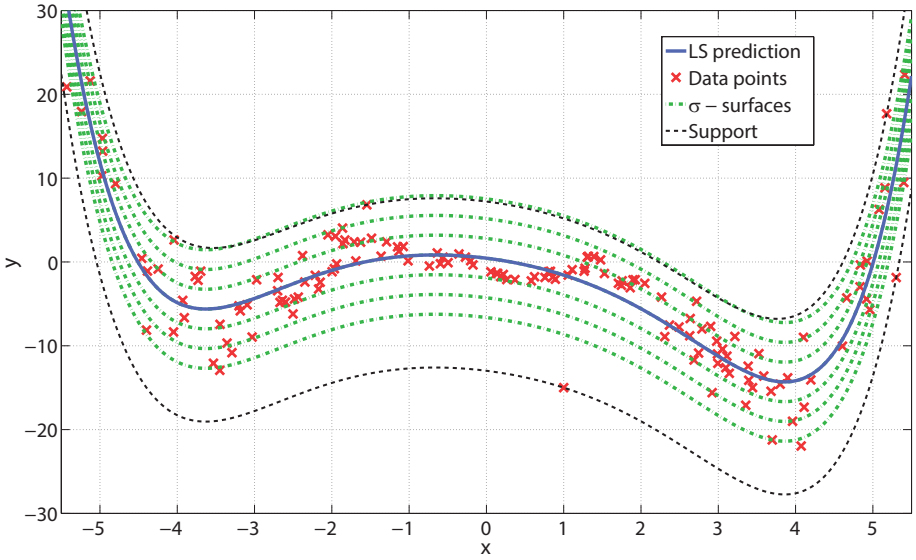

FIG. 6: RPM E: type-3 RPM for $\tau_{\max} = 1.5$ and $\lambda = 143/150$.

in the lower bound (30) to admissible values of $\tau_{\max}$. A sequential approach reversing the order and priorities of the model is presented next.

4.4 Type-4 RPMs

As with a type-3 RPM, a type-4 RPM prescribes the expected value, variance, and support set of F_p, and thus of $R_y(x)$, by solving three OPs. The first two OPs yield a type-2 RPM whereas the latter yields P. In contrast to type-3 RPMs, type-4 RPMs make the tightness of the σ volume more important than the spread of the output's range.

In particular, a type-4 RPM is given by Eq. (14), where the expected value $\mu = \hat{\mu}$ is given by (12), the variance $\nu = \hat{\nu}$ is given by (28), and $P = \hat{P}$ is given by the following OP:

Optimization Program 5 (OP5)**.** *For a given mean* μ *and a given variance* ν*, the defining vertices of* $\hat{P}$ *are given by*

$$\{\underline{\hat{p}}, \hat{\overline{p}}\} = \underset{u,\, v}{\operatorname{argmin}} \{\mathbb{E}_x[\delta_y(x,v,u)] : u \le \mu \le v,\ \underline{y}(x_i) \le y_i \le \overline{y}(x_i),\ i = 1,\ldots,N,\ \nu \le (\mu - u) \odot (v - \mu)\}. \quad (32)$$

Hence a type-4 RPM prescribes a P that minimizes the expected spread of the random process $R_y(x, F_p)$ such that the support set $I_y(x, P)$ contains all N observations subject to (i) a ν that minimizes the expected output variance for a σ volume $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu}(\lambda))$ containing $\lfloor \lambda N \rfloor$ observations, and (ii) a μ that minimizes the sum of the squares of the prediction errors relative to the mean function. Note that the solution to OP3 enters OP5 via the lower bound of the last constraint. Further notice that is further notice that OP3, used to calculate $\hat{\nu}$, is nonconvex, whereas OP5, used to calculate $\hat{P}$, is convex. This is the case even though the feasible design space associated with the bilinear constraints in (32) is nonconvex. The equivalence between OP3 and OP2, covered in Section 4.2, allows performing a reliability analysis of type-4 RPMs. This analysis bounds the probability that a future observation will fall outside both the σ-volume $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu}(\lambda))$ and the range $I_y(x, \hat{P})$. As before, the containment condition (31) plays a key role in the evaluation of such a bound.

Example 5. Next we derive two type-4 RPMs for $\tau_{\max} = 1$ and the same setup used earlier. The two RPMs differ in the value of λ used to calculate $\hat{\nu}$. Because $\tau_{\max} < \tau^*_{\max} = 1.4094$, there is no type-3 RPM that can tighten the σ-volume as much. This illustrates the limitations on the probabilistic performance resulting from type-3 RPMs.

The first RPM, referred to as RPM F, uses $\lambda = 1$. Hence we will require that all 150 observations be less than one standard deviation from the mean function. This setting led to RPM A in Example 1, so we have $\mathbb{E}_x[\nu_y] = \hat{\nu}_1 = 180.3824$. With $\hat{\nu}$ available, we then solve for $\hat{P}$ using (32). This leads to $\underline{\hat{p}} = [-12.9981, -1.1488, -0.8339, 0.0012, -0.0379, -0.0006, 0.0032]^\top$, and $\hat{\overline{p}} = [13.8920, -1.1488, -0.8339, 0.0012, -0.0379, 0.0001, 0.0032]^\top$ and $\mathbb{E}_x[\delta_y] = 13.4714$. Therefore, whereas the first and sixth component of p vary in a range, the other ones can be treated as fixed constants. The performance of RPM F, which is shown in Fig. 7, is given by both $\mathbb{E}_x[\delta_y] = 13.4714$ and $\mathbb{E}_x[\nu_y] = 180.3824$, which are 128% and 225% worse/larger than those of RPM D. Note that the containment condition holds for all $x \in X$. Both $v_\sigma(x, \hat{\mu}, \tau_{\max}, \hat{\nu}(\lambda))$ and $I_y(x, \hat{P})$ are centered about the LS prediction. This is not the case for other values of $\tau_{\max}$ (not shown). Because most of the observations are close to the mean function, it is natural to expect that neglecting a few outliers will considerably tighten the prediction.

We now calculate a type-4 RPM for $\lambda = 143/150$. This model, called RPM G, is shown in Fig. 8. The solution to OP3 led to RPM C for which $\mathbb{E}_x[\nu_y] = 36.3341$. With $\hat{\nu}$ available, we then solve for $\hat{P}$ using (32), which yields the vertices $\underline{\hat{p}} = [-10.2605, -3.8559, -0.8480, 0.0002, -0.0420, -0.0002, 0.0032]^\top$, and $\hat{\overline{p}} = [2.9670, 0.016, -0.8200, 0.0022, -0.0338, -0.0001, 0.0032]^\top$. Therefore, according to the spread of the output's range, the first, second, and fifth components of p are the dominant contributors. The performances of RPM G are $\mathbb{E}_x[\delta_y] = 12.3649$ and $\mathbb{E}_x[\nu_y] = 36.3341$. These values are 8% and 80% better/smaller than those of RPM F. The containment condition, which will be used to quantify the model's reliability, does not hold at $x = 0$ (not seen in Fig. 8). The support set of p_1 is not centered about its expected value of -0.8734 (see example 1). This causes a sizable offset between the mean function and the center of $I(x, \hat{P})$. This is further evidence that $v_\sigma(x, \hat{\mu}, 1, \hat{\nu}(143/150))$ contains most of the observations whereas outliers only affect $I_y(x, \hat{P})$, whose boundaries do not coincide with any σ surface.

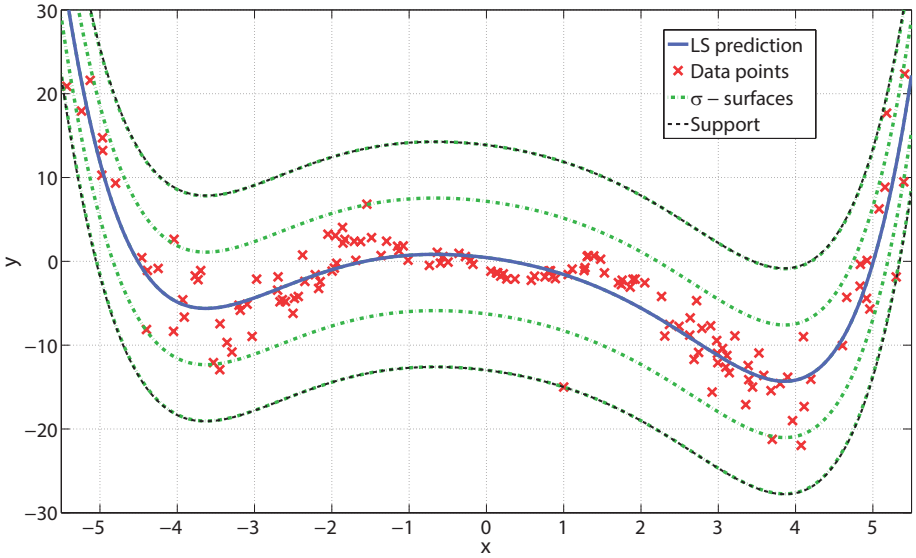

FIG. 7: RPM F: type-4 RPM for $\lambda = 1$ and $\tau_{\max} = 1$.

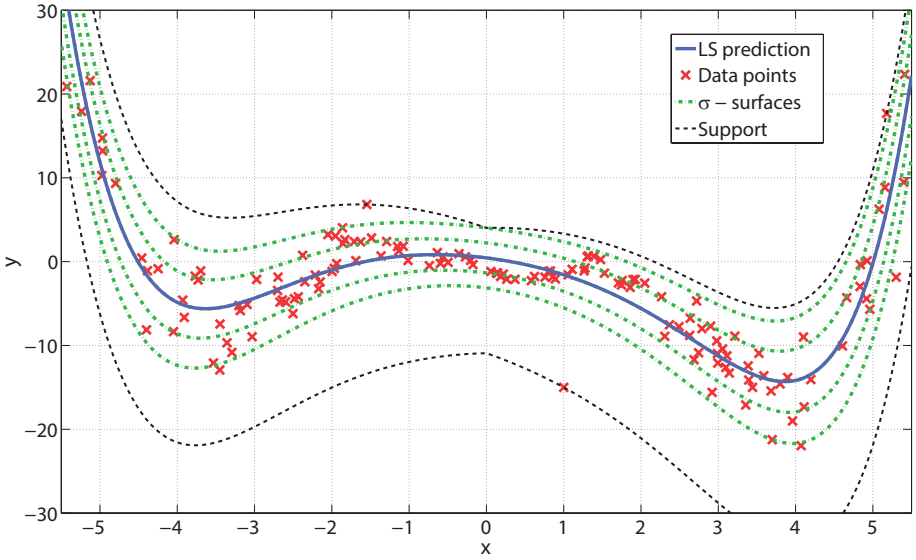

FIG. 8: RPM G: type-4 RPM for $\lambda = 143/150$ and $\tau_{\max} = 1$.

To better compare the probabilistic performance of type-3 RPMs against type-4 RPMs we make use of the invariant in Eq. (26). The comparison of RPM D with RPM F, for which $\lambda = 1$, yields $\mathcal{I}_D = \mathcal{I}_F = 180.38$. As such, changing the order of the OP sequence did not render any improvement. The comparison of RPM E with RPM G, for which $\lambda = 143/150$, yields $\mathcal{I}_E = 50.06$ and $\mathcal{I}_G = 36.33$, respectively. Therefore the prioritization of the variance over the support set improves the probabilistic performance by more than 27%. As expected, the improvement in probabilistic performance $\mathbb{E}_x[\nu_y]$ often causes a degradation of the nonprobabilistic performance $\mathbb{E}_x[\delta_y]$.

The comparison of F_κ for RPM F and RPM G (not shown) yields the same conclusions as drawn from Fig. 4. In this case RPM G is the better empirical model for $\lambda\%$ of the observations, whereas RPM F is the better model for the full data ensemble. As before, this illustrates the chance-constrained character of the formulation. As with type-2 RPMs, type-4 RPMs do not require prescribing the outliers in advance.

4.5 Discussion

There are infinitely many CDFs with F_p matching the requirements on the mean, variance, and support set resulting from the above formulations. One way to fully characterize characterize F_p given the features of a type-1 or type-2 RPM, is to assume that p is a vector of uncorrelated normal random variables. For a type-3 and type-4 RPM, this can be attained by assuming that F_p is an uncorrelated generalized beta random vector. The prescription of uncertainty as a probability box eliminates the need for such assumptions. The probability boxes prescribed in [11] account for all possible random vectors conforming to such restrictions.

The formulations above assume that the parameters in p are uncorrelated. Preliminary experiments enabling c to take on nonzero values led to improved probabilistic performances. This practice requires making c an additional decision variable in Eqs. (24), (28), and (29), and making the consistency conditions (17) and (18) additional inequality constraints. The reliability assessment of such RPMs, however, remains elusive.

4.6 Model Selection

A few comments regarding the use of the above formulations is in order. Note that the boundaries of type-1 IPMs and the limits of type-1 RPMs are driven by extreme observations, possibly having a small chance of occurrence. As such the resulting prediction is wider and thus less informative than those resulting from the other formulations. Type-2 RPMs tighten the prediction by neglecting extreme observations. If such observations fall within the long probability tails of the DGM, the resulting prediction is considerably better than that of type 1 RPMs. In type-3 and type-4 RPMs, extreme observations prescribe the support of the process, whereas only a fraction λ of them prescribes the variance. This fraction is chosen such that the informative character of the probabilistic prediction, which is where the bulk of the probability lies, is improved. Type-4 RPMs attain a better probabilistic performance than type-3 RPMs, whereas type-3 RPMs are better suited to describe the output's range.

A simple case study comparing alternative metamodeling techniques is presented next. Figure 9 shows the σ volume corresponding to $\tau_{\max} = 1$ that results from (i) a Gaussian process (GP) model, (ii) a prediction based on the confidence intervals (COI) for the coefficients of the linear regression, (iii) a type-2 RPM for $\lambda = 143/150$, (iv)

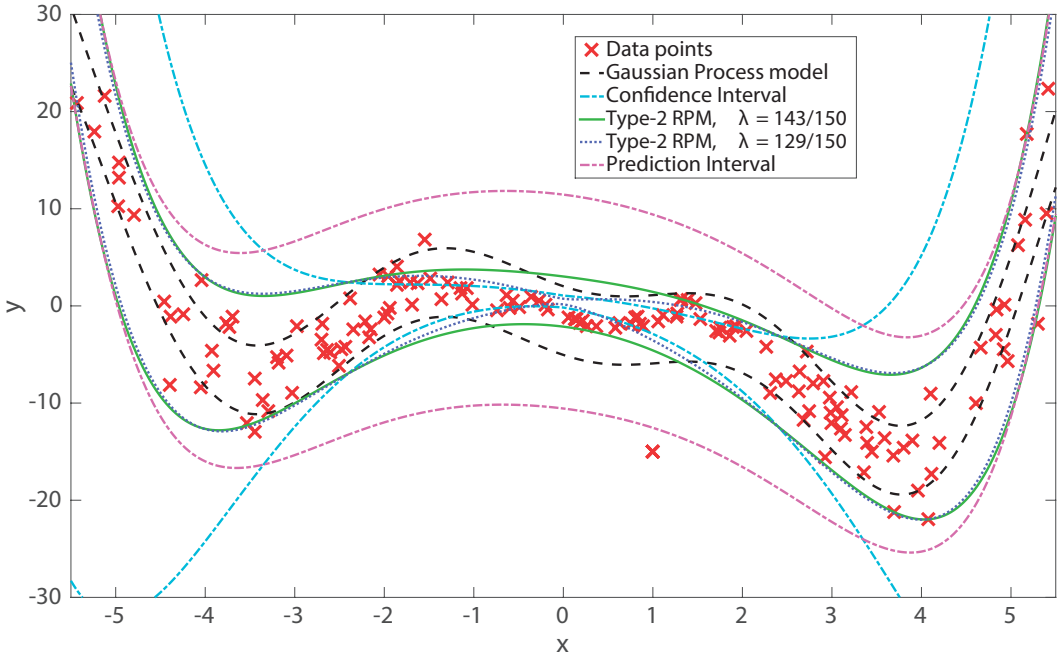

FIG. 9: σ-volume v_σ for $\tau_{\max} = 1$ resulting from a few metamodeling techniques.

a type-2 RPM for $\lambda = 129/150$, and (v) a prediction interval (PI).[11] All parametric techniques will use the same polynomial structure used in previous examples. The GP model uses a zero-mean function and the square exponential covariance function. Note that the σ volume corresponding to the GP, which implicitly assumes the structure in (1), excludes 21 data points, whereas the COI leaves 29 observations out. The first RPM led to RPM C (see example 3), while the second RPM was built to exclude the same number of outliers as the GP model. The RPMs attain the desired number of outliers by design, without making any assumption on the DGM, whereas the other methods require that the assumptions on the DGM (e.g., the DGM being a Gaussian process, the measurement/prediction error being normally distributed) be true in order to yield an accurate prediction. Note that the RPMs adjust more tightly to the spread of the data (i.e., it contracts and expands where needed), and the COI is excessively varying, whereas the GP and PI have a fairly constant width where the data is present. In particular, the variance of the width of the σ volume for the GP, COI, the type-2 RPMs, and the PI are 0.02, 1101, 35.46, 38.51, and 0.33, respectively. Regarding GP models, the largest aleatory spread of the DGM reached at some x in X prescribes the global predicted variance throughout X. The calculation of an RPM requires solving a sequence of OPs, which for the convex case can be done very efficiently for a large number of decision variables and constraints, i.e., on the order of 10^5 using standard optimization algorithms. This enables considering problems with many more data points, such as might be wanted in a high-dimensional input problem, than alternative approaches. For instance, GP models are restricted to a few thousand points before becoming computationally intractable. Numerical experiments performed by the authors indicate that the computational complexity of IPMs is about 2 orders of magnitude less than that of GP models. Whereas solving for a GP became numerically intractable for the setup listed above (it was still running after 36 h), an IPM was solved in about 3 min (using a desktop computer with modest hardware capabilities and standard software).

5. RELIABILITY

This section presents a framework for bounding the reliability of the predictor models proposed above. The reliability of an arbitrary model $\mathcal{E}$, $r(\mathcal{E})$, is the probability that a future observation will be within the predicted interval-valued function(s). The developments that follow are based on Scenario Optimization Theory [12–15]. Denote by $\mathbb{P}$ the *unknown* distribution of the DGM from which the points of the data sequence z are obtained. $\mathbb{P}$ can be interpreted as a probabilistic cloud in the $X \times Y$ space. The case in which y is a deterministic function of x only is a particular case where $\mathbb{P}$ is concentrated over the function. A general $\mathbb{P}$ leads to y being an arbitrary random process of x. No assumption is made on the underlying structure of $\mathbb{P}$. The following theorem enables bounding a model's reliability whenever the OP used for its calculation is convex [15].

Theorem 1. *Let z be a data sequence of N independent elements resulting from a stationary DGM. Suppose the model $\mathcal{E}$ is calculated by solving a convex constrained OP having a unique solution based on z. Furthermore, assume that k observations out of the N available have been discarded when calculating $\mathcal{E}$. Assume $k < N - d$, where d is the number of optimization variables used to calculate $\mathcal{E}$. Then, for any confidence parameter $\beta \in (0, 1)$ and any reliability parameter $\epsilon \in (0, 1)$ which satisfy*

$$\binom{k+d-1}{k} \sum_{i=0}^{k+d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i} \leq \beta, \tag{33}$$

then

$$\mathrm{Prob}_{\mathbb{P}^N} [r(\mathcal{E}) \geq 1 - \epsilon] > 1 - \beta. \tag{34}$$

This theorem provides an assessment of unobserved data. The theorem states that the reliability of $\mathcal{E}$ is no worse than $1 - \epsilon$ with probability greater than $1 - \beta$. As for the probability $1 - \beta$, one should note that $\mathcal{E}$ is a random model by virtue of the randomness in $\mathbb{P}$ prescribing z. Therefore its reliability can be greater than or equal to $1 - \epsilon$ for some random observations but not for others, and β refers to the probability $\mathbb{P}^N = \mathbb{P} \times \cdots \times \mathbb{P}$ of observing a bad set of N

[11] The COI and the PI are both based on a confidence level of $100(1 - 0.6827)\%$, which corresponds to $\tau_{\max} = 1$ for a Gaussian output.

samples such that the reliability of the model is less than $1-\epsilon$. Parameter ϵ is referred to as the reliability parameter, while β is the confidence parameter. It is worth noting that the confidence parameter can be made small enough that it loses any practical significance and $r(\mathcal{E}) \geq 1-\epsilon$. This can be done without letting N be too large, because β vanishes exponentially with N. Note that assessing the reliability of the model does not require making any assumptions on $\mathbb{P}$, and that the result is not asymptotic, i.e., it is valid for any finite value of N.

Equation (34) is a fundamental relation linking the amount of information available (represented by the number of observations N and the number of discarded data points k), the complexity of the model (represented by the number of decision variables d of the OP), and the probabilistic levels of reliability ϵ and confidence β. Inequality (33) should be interpreted as a relationship among five different variables, ϵ, β, N, k, and d. We can solve for optimal values of any of these variables depending upon the needs of the application.

5.1 Reliability of Type-1 IPMs

The reliability of type-1 IPMs, to be denoted as $\mathcal{I}$, is defined as

$$r(\mathcal{I}) = \text{Prob}_{\mathbb{P}}\left[(x,y) \in I_y\left(x, \hat{P}\right)\right]. \tag{35}$$

The convexity of the OP1 enables the direct application of Theorem 1.

5.2 Reliability of Type-1 and Type-2 RPMs

Denote by $\mathcal{R}$ any type-1 or type-2 RPM. The reliability of $\mathcal{R}$ is defined as

$$r\left(\mathcal{R}\right) = \text{Prob}_{\mathbb{P}}\left\{(x,y) \in v_\sigma\left(x, \mu, \tau_{\max}, \hat{\nu}(\lambda)\right)\right\}. \tag{36}$$

The convexity of OP2 enables the direct application of Theorem 1 to type-1 RPMs. This includes the cases in which none ($k=0$) and some ($k>0$) of the observations are removed from the data set. In contrast to OP2, OP3 is nonconvex. This opens the possibility of (28) having multiple optima. The RPMs corresponding to each local optima will likely be different. Because type-2 RPMs are calculated by solving a nonconvex program, Theorem 1 cannot be applied directly. However, the reliability of such models can be established by using the *principle of equivalence* (PE). This principle is based on identifying an auxiliary convex formulation that will result in the very same empirical model found by solving the nonconvex formulation. If this is attained, the reliability of the model, which is independent of the means used to calculate it, can be rigorously evaluated via the auxiliary formulation. This approach can be applied to type-2 RPMs. In particular, the solution to OP3 using the original data sequence z for a given value of λ is equivalent[12] to the solution of OP2, which is a convex OP, with the data sequence w. Because only the $N-k^*$ elements in w, where

$$k^* = \lfloor N(1-\lambda) \rfloor, \tag{37}$$

are required by the auxiliary program, the reliability of type-2 RPMs is given by (34) with $k=k^*$ in (33). These k^* observations satisfy $F_{\kappa(\hat{\nu})}(\kappa) > \lambda$.

5.3 Reliability of Type-3 and Type 4 RPMs

Denote by $\hat{\mathcal{R}}$ any type-3 or type-4 RPM. The reliability of $\hat{\mathcal{R}}$ is defined as

$$r(\hat{\mathcal{R}}) = \text{Prob}_{\mathbb{P}}\left\{(x,y) \in I_y\left(x, \hat{P}\right) \cap v_\sigma\left(x, \mu, \tau_{\max}, \hat{\nu}(\lambda)\right)\right\}. \tag{38}$$

The following theorem provides the means to bound $r(\hat{\mathcal{R}})$:

[12]When $\mathbb{E}_x[\delta_y]$ is evaluated by the sample mean, equivalence is attained by using w to evaluate the constraints and z to evaluate the cost function.

Theorem 2. *Let $\hat{\mathcal{R}}$ be a type-3 or type-4 RPM based on the data sequence* z *of N independent elements obtained from a stationary DGM. When the containment condition (31) holds, we have*

$$\text{Prob}_{\mathbb{P}^N}\left[r(\hat{\mathcal{R}}) \geq 1-\epsilon\right] > 1-\beta, \tag{39}$$

where ϵ and β are given by (33), with $d = n_p$ and $k = k^*$. Otherwise, $\epsilon = \epsilon_1 + \epsilon_2$ and $\beta = \beta_1 + \beta_2$, where ϵ_1 is given by

$$\sum_{i=0}^{2n_p-1} \binom{N}{i} \epsilon_1^i (1-\epsilon_1)^{N-i} \leq \beta_1, \tag{40}$$

whereas ϵ_2 and β_2 are given by (33) for $d = n_p$ and $k = k^*$.

Proof. When the containment condition holds, the two events defining the reliability are dependent and $r(\hat{\mathcal{R}}) = \text{Prob}_{\mathbb{P}}[(x,y) \in v_\sigma]$. In this case the reliability is given by Theorem 1 after applying the PE to the nonconvex OPs (29) or (28) to type-3 and type-4 RPMs, respectively. In both cases $k = k^*$ as defined in (37), and $d = n_p$. When Eq. (31) does not hold, use the bound $r(\hat{\mathcal{R}}) \geq \text{Prob}_{\mathbb{P}}[(x,y) \in I_y(x,\hat{P})] + \text{Prob}_{\mathbb{P}}\{(x,y) \in v_\sigma(x,\mu,\tau_{\max},\hat{\nu}(\lambda))\} - 1$. This bound is generally loose, so the actual model's reliability is probably larger. Each of the two events will be considered separately. Because the event $(x,y) \in I_y(x,\hat{P})$ is enforced by solving the convex OP in (11) or (32) with N observations, we use Theorem 1 for $d = 2n_p$ and $k = 0$ to calculate ϵ_1. Conversely, the event $(x,y) \in v_\sigma(x,\mu,\tau_{\max},\hat{\nu}(\lambda))$ is enforced by solving the nonconvex OPs in (29) for a type-3 RPM and (28) for a type-4 RPM. The PE for an auxiliary convex OP with $k = k^*$ and $d = n_p$ leads to ϵ_2. Theorem 2 results from substituting these expressions into Theorem 1. □

Although the RPMs corresponding to different local minima will likely be different, they admit the same reliability upper bound (i.e., the auxiliary problems use the same values for N, k, and d; thus for a given β they will lead to the same value of ϵ). Hence having different sets of k outliers might lead to RPMs with different performance values for the same reliability upper bound. The actual reliability of the model, however, will likely be different.

Example 6. The reliability of RPM D and E, which are type-3 RPMs, is considered first. Since neither model satisfies the containment condition (31), the reliability of each event must be added. Whereas the first event in (38), for which $N = 150$, $k = 0$, and $d = 14$, yields $1-\epsilon_1 = 0.8452$ with confidence $1-\beta_1 = 0.99$, the second event, for which $N = 150$, $k = 0$, and $d = 7$, leads to $1-\epsilon_2 = 0.9058$ with the same confidence. Therefore the reliability of RPM D is no less than $1-\epsilon_1-\epsilon_2 = 0.7510$ with confidence $1-\beta_1-\beta_2 = 0.98$. In the case of RPM E we have the same value for ϵ_1 as that for RPM D, whereas for the second event, for which $N = 150$, $k = 7$, and $d = 7$, leads to $1-\epsilon_2 = 0.7738$ with confidence $1-\beta_2$. Therefore the reliability of RPM D is no less than $1-\epsilon_1-\epsilon_2 = 0.6190$ with confidence $1-\beta = 0.98$. Hence discarding seven outliers improved performance by 74% at the expense of a reduction in reliability of 17.6%. Finally, we will evaluate the reliability of RPM F and G, which are type-4 RPMs. Recall that the containment condition holds for RPM F but not for RPM G. The reliability of RPM F, for which $N = 150$, $k = 0$, and $d = 7$, is no less than $1-\epsilon = 0.8032$ with confidence $1-\beta = 0.99$. In the case of RPM G, the first event in (38), for which $N = 150$, $k = k^* = 7$, and $d = 7$, leads to $1-\epsilon_1 = 0.7682$ with confidence $1-\beta_1 = 0.995$, whereas the second event, for which $N = 150$, $k = 0$, and $d = 14$, leads to $1-\epsilon_2 = 0.8372$ with confidence $1-\beta_2 = 0.995$. Therefore the reliability of RPM G is no less than $1-\epsilon_1-\epsilon_2 = 1-\epsilon = 0.6054$ with confidence $1-\beta = 0.99$. The values for β_1 and β_2 chosen make β for RPM F and RPM G equal, so their reliability can be compared. The reduction of 21.19% in the reliability of RPM G relative to that of RPM F is affected by the conservatism in Theorem 2. This illustrates the benefits of satisfying the containment condition. This example illustrates the typical trade-off between performance and reliability. These figures of merit should be traded off until the desired balance is reached. This balance can be reached by changing the number of observations N, of outliers via λ, or by changing the model's structure via n_p, which prescribes d.

5.4 Using the LS Parameter Estimate as the Mean Parameter

The selection $\mu = \hat{\mu}$ made above is subjective and solely based on engineering grounds. Unfortunately, using the data sequence z to derive both the mean μ and the volume $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu})$ violates the independence assumption of Theorem 1. The reason for this violation is linked to the concept of support constraints and how they are used in the corresponding proof [15]. A *support constraint* is defined as a constraint whose removal from the OP changes the optimum. The rationale supporting Theorem 1 makes use of d being the largest number of supporting constraints a convex OP admits. Removing an observation from z changes $\hat{\mu}$ and thus all the constraints in (24), (28), and (29) that depend on $\hat{\mu}$. Hence, strictly speaking, choosing $\mu = \hat{\mu}$ makes all such constraints support constraints, e.g., there are $2N$ supporting constraints for type-1 RPMs. This unwanted dependence is expected to be minor for moderately large values of N as the LS parameter estimate approaches its asymptotic value and becomes practically insensitive to additional data. As such we expect the theory to be "robust" and maintain its validity when μ and $v_\sigma(x, \mu, \tau_{\max}, \hat{\nu})$ are based on the same data. This unwanted dependency is eliminated by choosing a value for μ that is independent from the data used to build the RPM. This, for instance, can be attained by partitioning the data set into two subsets, using one to calculate $\hat{\mu}$ via (12) and using the other one to calculate $v_\sigma(x, \hat{\mu}, \tau_{\max}, \hat{\nu})$ via (24), (28), or (29). Alternatively, we can make μ an additional design variable in (24), (28), and (29). This practice not only eliminates the unwanted dependency among constraints, but also yields RPMs having an improved performance (a tighter σ volume).

6. CONCLUSIONS

This paper proposes techniques for constructing linear parametric models describing key features of the distribution of an output variable given input-output data. This structure enables a rigorous characterization of the uncertainty in the model's parameters, of key features of the prediction, and of the reliability of the resulting metamodel. Because such features conform to all possible probabilistic models for p, the resulting characterization of both the uncertainty and the predicted output are distribution-free. A few types of models exhibiting various degrees of insensitivity to outliers are developed. The differences between RPMs and standard metamodels are both conceptual and practical. First and foremost among them is the ability to formally evaluate the reliability of the resulting metamodel without having to make any assumptions on the structure of the underlying data generating mechanism. This is a substantive advantage over alternative techniques. Furthermore, the calculation of the proposed RPMs requires solving a sequence of optimization programs, which for the convex case can be done very efficiently for a large number of design variables and constraints, i.e., on the order of 10^5 using standard optimization algorithms. This enables considering problems with many more data points, thus input dimensions, than alternative approaches, e.g., Gaussian Process models become numerically intractable after a few thousand data points.

REFERENCES

1. Simpson, T., Peplinski, J., Koch, P., and Allen, J., Metamodels for computer-based engineering design: Survey and recommendations, *Eng. Comput.*, 17(1):129–150, 2001.
2. Seber, G. A. and Wild, C. J., *Nonlinear Regression*, John Wiley and Sons, Hoboken, NJ, USA, 2003.
3. Kennedy, M. and O'Hagan, A., Bayesian calibration of computer models, *J. R. Stat. Soc. B*, 63(3):425–464, 2001.
4. Crespo, L. G., Giesy, D. P., and Kenny, S. P., Calibration of predictor models using multiple validation experiments, AIAA-2015-0659, in *The 18th AIAA Non-Deterministic Approaches Conference*, Kissimmec, FL, Jan. 5–9, 2015.
5. Allen, M. and Maute, K., Reliability-based design optimization of aeroelastic structures, *Struct. Multidiscip. Optim.*, 27:228–242, 2004.
6. Eldred, M. S., Agarwal, H., Perez, V., Wojtkiewicz, S., and Renaud, J., Investigation of reliability method formulations in DAKOTA/UQ, *Struct. Infrastruct. Eng.*, 3(3):199–213, 2007.
7. Swiler, L. P., Adams, B., and Eldred, M., Model calibration under uncertainty: Matching distribution information, AIAA-2008-5944, in *The 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Victoria, British Columbia, Canada, Sept. 10–12, 2008.

8. McAndrew, C. C., *Compact Modeling: Principles, Techniques, and Applications*, Chap. 16, Springer, New York, 2010.
9. Crespo, L. G., Kenny, S. P., and Giesy, D. P., Interval predictor models with a linear parameter dependency, *ASME J. Verif. Valid. Uncertainty Quantification*, pp. 1–20, 2016.
10. Charnes, A., Cooper, W. W., and Symonds, G. H., Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil, *J. Instit. Operat. Res. Manag. Sci.*, 4(3):235–263, 1958.
11. Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D. S., and Sentz, K., Constructing probability boxes and Dempster-Shafer structures, Tech. Rep. SAND2002-4015, Sandia National Laboratories, 2003.
12. Calafiore, G. and Campi, M. C., The scenario approach to robust control design, *IEEE Trans. Automatic Control*, 51(1):742–753, 2006.
13. Campi, M. C. and Garatti, S., The exact feasibility of randomized solutions of uncertain convex programs, *SIAM J. Optim.*, 19(3):1211–1230, 2008.
14. Campi, M., Calafiore, G., and Garatti, S., Interval predictor models: Identification and reliability, *Automatica*, 45(2):382–392, 2009.
15. Campi, M. C. and Garatti, S., A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality, *J. Optim. Theory Appl.*, 148(1):257–280, 2011.