Moment-Matching Predictor Models with a Linear Staircase Structure

Luis G. Crespo[a*], Sean P. Kenny[b], and Daniel P. Giesy[c]

[a,b,c]Dynamic Systems and Controls Branch,
NASA Langley Research Center, Hampton, VA, 23681, USA

Abstract: This paper proposes techniques for constructing computational models describing the distribution of a continuous output variable given input-output data. These models are called Random Predictor Models (RPMs) because the predicted output corresponding to any given input is a random variable. We focus on RPMs having a linear parameter dependency, a bounded support set and prescribed functions for the first four moments. These constraints are realized by describing the model parameters as staircase random variables. The high versatility of such variables, and their low computational cost enable the efficient generation of possibly skewed and/or multimodal RPMs over an input-dependent interval. Optimization-based strategies for calculating RPMs using several optimality criteria are developed in this an a companion paper. These criteria include optimal moment-matching, presented herein, as well as minimal-dispersion and maximum-likelihood formulations. The computational demands of the first two approaches, which separate distribution-free and distribution-fixed steps, are kept low by not requiring the simulation of solution candidates during the search for the optimal RPMs. As an example we consider the estimation of the load being applied to a cantilever beam from deflection measurements.

Keywords: Predictor models, staircase variables, model calibration, moments.

1. INTRODUCTION

Metamodeling [1] is the process of creating a mathematical representation of a phenomenon based on input-output data. Metamodeling techniques can be parametric or non-parametric. In the parametric case, the functional form by which the output depends on the input is first prescribed using a model M, and then the parameters of such a model are characterized. This step is commonly referred to as to model calibration. The approach proposed below falls into this category.

Bayesian inference [2] is often used for model calibration. In Bayesian calibration, the objective is to describe the parameters of a model as a vector of possibly dependent random variables by 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 M. 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. Making the prediction match the observations by adjusting the hyper-parameters of a distribution is a long standing approach used in reliability-based design optimization, moment matching algorithms, and backward propagation of variance [3, 4, 5, 6]. In spite of its high computational demands, which entail simulating the predictor for each candidate combination of the calibrating variables, and of the potentially high sensitivity of the posterior to the assumed prior, this method is regarded as a benchmark in model calibration.

2. PROBLEM STATEMENT

A Data Generating Mechanism (DGM) is postulated to act on a vector of input variables, x, to produce an output, y. In this article, the focus will be on the single-output ($n_y = 1$) multi-input ($n_x \geq 1$) case.

* Corresponding author, Luis.G.Crespo@nasa.gov

The dependency of the output on the input is arbitrary. Assume that N Independent and Identically Distributed (IID) input-output pairs are obtained from a stationary DGM, and denote by $\mathbb{D} = \{x^{(i)}, y^{(i)}\}$, $i = 1, \ldots N$, the corresponding data sequence. The main objective of this article is to generate a model of the DGM based on $\mathbb{D}$. In a parametric model the output depends on both the input and the parameter through an equation. Denote by $y = M(x, p)$ such an equation, where $p \in \mathbb{R}^{n_p}$ are the model parameters. Instead of the standard practice of trying to match all the data as closely as possible with M evaluated at a single point p, the thrust in this work is to characterize p by either a bounded set P or by the joint PDF $f_p(p)$ supported in P. In both cases, the prescription of P must ensure that each data point in $\mathbb{D}$ can be fit exactly by the model evaluated at least one element of p in such a set. For a fixed value of the input x, and as long as P is a connected set and $M(x, p)$ is a continuous function (the only cases we will consider), the propagation of P through M yields an interval of output values. Thus, these models are called Interval Predictor Models (IPM). The desired IPM is a narrow interval of output values where unobserved data will likely fall. Conversely, for a fixed value of the input x, the propagation of $f_p(p)$ through M yields a random variable. Thus, these models are called RPMs. The desired RPM accurately describes the probability distribution governing the DGM.

Optimization-based strategies for calculating RPMs having a linear parameter dependency were developed. These optimality criteria include a moment-matching formulation, presented below, as well as a minimal-dispersion and maximum-likelihood formulations, presented in the companion paper [7]. The background supporting these developments is introduced next.

3. PRELIMINARIES

Consider the continuous random variable z with support set $\Delta_z = [z_L, z_U]$, Probability Density Function (PDF) $f_z : \Delta_z \subset \mathbb{R} \to \mathbb{R}^+$, and Cumulative Distribution Function (CDF) $F_z : \Delta_z \to [0, 1]$. Denote by m_r the r-th central moment of z, which is defined as

$$m_r = \int_{\Delta_z} (z - \mu)^r f_z(z) dz, \; r = 0, 1, 2, \ldots \tag{1}$$

where μ is the expected value of z. Note that $m_0 = 1$, $m_1 = 0$, m_2 is the variance, m_3 is the third-order central moment, and m_4 is the fourth-order central moment. Where reference is made to the r-th moment of a random variable, we assume that the corresponding integral in (1) converges for that distribution.

The random variables of interest will be constrained to have a bounded support set and given values for μ, m_2, m_3, and m_4. The bounded support constraint is $\Delta_z \subseteq \Omega_z$ where $\Omega_z = [\underline{z}, \overline{z}]$, with $\overline{z} \geq \underline{z}$, whereas the moment constraints are given by (1). The parameters of these constraints will be grouped into the variable $\theta_z \in \mathbb{R}^6$ given by

$$\theta_z = [\underline{z}, \overline{z}, \mu, m_2, m_3, m_4]. \tag{2}$$

Any random variable z having a support set contained by $[\underline{z}, \overline{z}]$ with moments μ, m_2, m_3, and m_4 must

[8] satisfy the feasibility conditions $g(\theta_z) \leq 0$, where[1]

$$g_1 = \underline{z} - \bar{z}, \tag{3}$$
$$g_2 = \underline{z} - \mu, \tag{4}$$
$$g_3 = \mu - \bar{z}, \tag{5}$$
$$g_4 = -m_2, \tag{6}$$
$$g_5 = m_2 - v \tag{7}$$
$$g_6 = m_2^2 - m_2(\mu - \underline{z})^2 - m_3(\mu - \underline{z}), \tag{8}$$
$$g_7 = m_3(\bar{z} - \mu) - m_2(\bar{z} - \mu)^2 + m_2^2, \tag{9}$$
$$g_8 = 4m_2^3 + m_3^2 - m_2^2(\bar{z} - \underline{z})^2, \tag{10}$$
$$g_9 = 6\sqrt{3}m_3 - (\bar{z} - \underline{z})^3, \tag{11}$$
$$g_{10} = -6\sqrt{3}m_3 - (\bar{z} - \underline{z})^3, \tag{12}$$
$$g_{11} = -m_4, \tag{13}$$
$$g_{12} = 12m_4 - (\bar{z} - \underline{z})^4, \tag{14}$$
$$g_{13} = (m_4 - vm_2 - um_3)(v - m_2) + (m_3 - um_2)^2, \tag{15}$$
$$g_{14} = m_3^2 + m_2^3 - m_4 m_2, \tag{16}$$

for $u = \bar{z} + \underline{z} - 2\mu$ and $v = (\mu - \underline{z})(\bar{z} - \mu)$. The realizations of θ satisfying the constraints in Equations (3-16) constitute the θ-feasible domain, Θ, defined as

$$\Theta = \{\theta : g(\theta) \leq 0\}. \tag{17}$$

A member of Θ will be called θ-feasible. Determining membership in Θ is a distribution-free assessment applicable to possibly infinitely many random variables satisfying the desired constraints.

A particular class of random variables that can realize most of Θ is proposed in [8]. This class is called *staircase* because the PDF of its members is piecewise constant over bins of equal width. Staircase variables, are calculated by solving the convex optimization program

$$\min_{\ell \geq 0} \{J(\theta, n_b) : A(\theta, n_b)\ell = b(\theta), \theta \in \Theta\}, \tag{18}$$

where J is the cost function used for optimization, n_b is the number of bins partitioning Ω_z, ℓ are the PDF values at the bin centers, and $A\ell = b$ are moment matching constraints. Staircase variables will be denoted as

$$z \sim S_z(\theta_z, n_b, J). \tag{19}$$

One of several cost functions J, each corresponding to different optimality criteria, can be chosen. The points $\theta \in \Theta$ for which a staircase variable with n_b bins exists constitutes the staircase feasible domain, $\mathbb{S}(n_b)$. As expected, $\mathbb{S}(n_b) \subset \Theta$. Additional details on staircase variables are available in [8].

4. INTERVAL PREDICTOR MODELS

This section presents the means to generate the support set of an RPM. This will be carried out by finding a baseline IPM using the same data sequence $\mathbb{D}$ that will be used to construct the RPM.

[1] Throughout this paper vector inequalities hold component-wise.

Additional details on IPMs are available in [9]. An IPM assigns to each instance vector $x \in X \subseteq \mathbb{R}^{n_x}$ a corresponding outcome interval in $Y \subseteq \mathbb{R}$. That is, an IPM is a set-valued map, $I_y : x \to I_y(x) \subseteq Y$, where $I_y(x)$ is the predicted 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 $X \times Y$. A non-parametric IPM is given by

$$I_y(x) = \left\{ \left[\underline{y}(x),\, \overline{y}(x)\right],\, \overline{y}(x) \geq \underline{y}(x) \right\}. \tag{20}$$

where the functions $\underline{y}(x)$ and $\overline{y}(x)$ are the lower and upper boundaries of the IPM respectively. A parametric IPM is obtained by associating to each $x \in X$ the set of outputs y that result from evaluating the function $y = M(x,p)$ at all values of p in the set P. Attention will be limited to the case in which the output depends linearly on p and arbitrarily on x, i.e.,

$$y = p^\top \varphi(x), \tag{21}$$

where $\varphi(x)$ is an arbitrary basis, and to uncertainty sets P having a hyper-rectangular shape, i.e.,

$$P = \{p : \underline{p} \leq p \leq \overline{p}\}. \tag{22}$$

The parameter points $\underline{p}$ and $\overline{p}$ are called the defining vertices of P. In this setting, a parametric IPM is given by

$$I_y(x,P) = \left[\, \underline{y}(x,\overline{p},\underline{p}),\ \overline{y}(x,\overline{p},\underline{p}) \,\right], \tag{23}$$

where

$$\underline{y}(x,\overline{p},\underline{p}) = \overline{p}^\top \left(\frac{\varphi(x) - |\varphi(x)|}{2} \right) + \underline{p}^\top \left(\frac{\varphi(x) + |\varphi(x)|}{2} \right), \tag{24}$$

$$\overline{y}(x,\overline{p},\underline{p}) = \overline{p}^\top \left(\frac{\varphi(x) + |\varphi(x)|}{2} \right) + \underline{p}^\top \left(\frac{\varphi(x) - |\varphi(x)|}{2} \right). \tag{25}$$

Each member of the family of infinitely many predictions that results from evaluating the model M at each realization $p \in P$ lies between the IPM boundaries $\underline{y}(x,\overline{p},\underline{p})$ and $\overline{y}(x,\overline{p},\underline{p})$, and no tighter containing functions exist. The spread of $I_y(x,P)$, which is the distance between its upper and lower boundaries, is

$$\delta_y(x,\overline{p},\underline{p}) = (\overline{p} - \underline{p})^\top |\varphi(x)|. \tag{26}$$

The narrower δ_y the more informative the IPM. Several IPM types can be calculated within this framework. In this paper we seek IPMs given by (23), where the defining vertices of P are given by

$$\left\{\underline{\hat{p}}(c),\, \hat{\overline{p}}(c)\right\} = \underset{u,\,v:\,u\leq v}{\arg\min} \left\{ \mathbb{E}_x[\delta_y(x,v,u)] : \underline{y}\left(x^{(i)},v,u\right) \leq y^{(i)} \leq \overline{y}\left(x^{(i)},v,u\right),\, c(u,v) \leq 0,\, i = 1,\ldots N \right\}, \tag{27}$$

where $\mathbb{E}_x[\cdot]$ is the expected value operator with respect to x, and $c(u,v) \leq 0$ are a set of additional constraints that will be used to enforce additional attributes. This program is convex as long as $c(u,v) \leq 0$ is a convex set. The uncertainty box corresponding to this IPM is

$$\hat{P}(c) = \{p : \underline{\hat{p}}(c) \leq p \leq \hat{\overline{p}}(c)\}. \tag{28}$$

Scenario optimization [10, 9] theory enables making a formal, distribution-free assessment on the probability of unobserved data falling outside the predicted interval.

5. RANDOM PREDICTOR MODELS

An RPM is a mapping that assigns to each input vector $x \in X$ a corresponding random variable $R_y(x)$. A non-parametric RPM is the random variable-valued map given by

$$R_y(x) = \{f_{y(x)}(y),\ y(x) \in \Delta_y(x)\}, \tag{29}$$

where $f_{y(x)}$ is the PDF of y at $x \in X$ having the support set $\Delta_y(x) = [\underline{y}(x), \overline{y}(x)] \subseteq Y$. By contrast, 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 PDF $f_p(p)$ supported in Δ_p, each weighted according to its corresponding likelihood. Hence,

$$R_y(x, f_p) = \{y = M(x,p),\ p \sim f_p(p),\ p \in \Delta_p\}. \tag{30}$$

The RPMs developed hereafter assume the form in (21). This structure enables the analytical description of the moments of the output in terms of the moments of the parameters via

$$\mu_{y(x)} = \mathbb{E}_p[p]^\top \varphi(x), \tag{31}$$

$$\mathbb{E}_y\left[y^2\right] = \varphi^\top(x)\mathbb{E}_p\left[pp^\top\right]\varphi(x), \tag{32}$$

$$\mathbb{E}_y\left[y^3\right] = \varphi^\top(x)\mathbb{E}_p\left[pp^\top\varphi(x)p^\top\right]\varphi(x), \tag{33}$$

$$\mathbb{E}_y\left[y^4\right] = \varphi^\top(x)\mathbb{E}_p\left[pp^\top\varphi(x)\varphi(x)^\top pp^\top\right]\varphi(x). \tag{34}$$

The central moments of the output are given by

$$m_{2,y(x)} = \mathbb{E}_y\left[y^2\right] - \mu_{y(x)}^2, \tag{35}$$

$$m_{3,y(x)} = 2\mu_{y(x)}^3 - 3\mu_{y(x)}\mathbb{E}_y\left[y^2\right] + \mathbb{E}_y\left[y^3\right], \tag{36}$$

$$m_{4,y(x)} = \mathbb{E}_y\left[y^4\right] - 3\mu_{y(x)}^4 + 6\mu_{y(x)}^2\mathbb{E}_y\left[y^2\right] - 4\mu_{y(x)}\mathbb{E}_y\left[y^3\right]. \tag{37}$$

Hence, the moments of the output depend on $\varphi(x)$ and $\mu = \mathbb{E}_p[p]$, $m_2 = \mathbb{E}_p\left[(p-\mu)^2\right]$, $m_3 = \mathbb{E}_p\left[(p-\mu)^3\right]$ and $m_4 = \mathbb{E}_p\left[(p-\mu)^4\right]$. To simplify the notation, rewrite Equations (31), (35), (36) and (37) as

$$\mu_{y(x)} = h_\mu(\mu, x), \tag{38}$$

$$m_{2,y(x)} = h_{m_2}(\mu, m_2, x), \tag{39}$$

$$m_{3,y(x)} = h_{m_3}(\mu, m_2, m_3, x), \tag{40}$$

$$m_{4,y(x)} = h_{m_4}(\mu, m_2, m_3, m_4, x). \tag{41}$$

In contrast to non-parametric RPMs, assuming a particular functional dependency between the inputs and the output often renders RPMs with a suboptimal performance. The degradation in performance stems from the fact that the likelihood of the parameter realization $p \in \Delta_p$ via $f_p(p)$ corresponds to the likelihood of the function $y = p^\top\varphi(x)$ in the input-output space $X \times Y$. Hence, any assignment of probability via $f_p(p)$ will affect the probability distribution within the RPM throughout X. In fact, the set of parameter points in the intersection of P with a hyper-volume bounded by any pair of parallel hyper-planes both perpendicular to the vector $\varphi(x) \in \mathbb{R}^{n_p}$ is mapped by $y = p^\top\varphi(x)$ onto a subinterval of $\Delta_{y(x)} = I_y(x, P)$. This dependency on the input x precludes shaping the probability distribution of the RPM at a given input value x without affecting the probability distribution elsewhere.

Means to characterize $f_p(p)$ in (30) such that the resulting RPM accurately represents the data are presented next.

6. MOMENT-MATCHING RPMS

This formulation aims at minimizing the offset between the target moments estimated from the data, to be denoted as $\tilde{m}$, and those to be predicted by the RPM, to be denoted as m. Because these variables are functions of the input x, they will be denoted as $\tilde{m}_{y(x)}$ and $m_{y(x)}$ respectively. The algorithmic implementation of this approach requires solving a sequence of optimization programs. Each program entails searching for the combination of moments of p of a given order that minimizes the offset between the target and the prediction. Note that these searches for the optimum are carried out in a distribution-free setting. Once the optimal moments of all n_p parameters are solved for, we calculate a moment-matching staircase variable for each parameter in p. A target-matching RPM is obtained by using the resulting $f_p(p)$ in (30).

Means to prescribe the target functions given $\mathbb{D}$ are provided next. Denote by

$$\tilde{m}_{y(x)} = \left[\tilde{\mu}_{y(x)}, \tilde{m}_{2,y(x)}, \tilde{m}_{3,y(x)}, \tilde{m}_{4,y(x)}\right], \tag{42}$$

for $x \in X$, the target functions containing the mean $\tilde{\mu}_{y(x)}$, the variance $\tilde{m}_{2,y(x)}$, the third-order central moment $\tilde{m}_{3,y(x)}$, and the fourth-order central moment $\tilde{m}_{4,y(x)}$. Because not enough data (if any) is available to characterize the DGM at a fixed value of the input x, we will adopt the "sliding window" approach based on the weighted sample moments proposed in [11]. The approach proposed therein yields the target functions $\tilde{m}_{y(x)}$ in (42).

The predicted moments $m_{y(x)}$ to be solved for are given by

$$m_{y(x)} = \left[\mu_{y(x)}, m_{2,y(x)}, m_{3,y(x)}, m_{4,y(x)}\right]. \tag{43}$$

When $\mu_{y(x)} = \tilde{\mu}_{y(x)}$, $m_{2,y(x)} = \tilde{m}_{2,y(x)}$, $m_{3,y(x)} = \tilde{m}_{3,y(x)}$, and $m_{4,y(x)} = \tilde{m}_{4,y(x)}$ for all $x \in X$, a "perfect match" between the prediction and the target is attained. The nonparametric RPMs in [11] achieve a perfect match. This is often not the case when the RPM is parametric, which is the case considered herein. The predicted moments $m_{y(x)}$ depend on the support of p given in (22), and on the moments of p via Equations (38-41). The full set of parameters of independent variables will be grouped into the matrix $\theta_p \in \mathbb{R}^{n_p \times 6}$ given by

$$\theta_p = \left[\underline{p}, \overline{p}, \mu, m_2, m_3, m_4\right], \tag{44}$$

where $\underline{p}$ and $\overline{p}$ are the defining vertices of P, to be determined, and the moments μ, m_2, m_3 and m_4 are elements of $\mathbb{R}^{n_p}$. In this setting, we aim at shaping $m_{y(x)}$ in (43) by manipulating θ_p in (44) such that the offset between the predicted moment functions in $m_{y(x)}$ and the target moment functions in $\tilde{m}_{y(x)}$ is minimized. For θ_p to be a feasible parameter set for a random vector p, the constraints $g(\theta_p) \leq 0$ must be satisfied. By $g(\theta_p)$, we mean all $g(\theta_{p_i})$ where θ_{p_i} is row i of matrix θ_p for all $i = 1, \ldots, n_p$.

Means to solve this problem are provided next. To this end we will use g_{v}, with v being the subset of the index list $\{1\ldots,14\}$, as a short-hand notation for the corresponding components of the feasibility constraint functions $g(\theta)$ in (3-16). Furthermore, we define the uncertainty set

$$\Omega_p(\gamma, c) = \{p : \underline{p}(c,\gamma) \leq p \leq \overline{p}(c,\gamma)\}, \tag{45}$$

whose defining vertices are

$$\underline{p}(\gamma, c) = \underline{\hat{p}}(c) - \gamma(\hat{\overline{p}}(c) - \underline{\hat{p}}(c)), \tag{46}$$

$$\overline{p}(\gamma, c) = \hat{\overline{p}}(c) + \gamma(\hat{\overline{p}}(c) - \underline{\hat{p}}(c)), \tag{47}$$

for $\underline{\hat{p}}(c)$ and $\hat{\overline{p}}(c)$ given by (27). Note that both $I_y(x, \Omega_p(\gamma, c))$ for $\gamma > 0$ and $I_y(x, \Omega_p(0, \emptyset))$ enclose the data, but the spread of the latter is likely smaller. For a given set of target functions $\tilde{m}_{y(x)}$, θ_p is found by solving the following sequence of optimization programs.

Solving for $\hat{\mu}$:

$$\hat{\mu} = \arg\min_{\mu} \left\{ \sum_{i=1}^{N} \left(\tilde{\mu}_{y(x^{(i)})} - h_{\mu}\left(\mu, x^{(i)}\right)\right)^2 \right\}. \tag{48}$$

Solving for $\hat{m}_2$: Let $\underline{p}$ and $\overline{p}$ be the defining vertices of $\Omega_p(\gamma_0, c_1)$, where $\gamma_0 \geq 0$ and $c_1 = g_{\mathsf{a}}|_{\mu=\hat{\mu}}$ for $\mathsf{a} = \{2,3\}$. The notation $g_{\mathsf{a}}|_{\mu=\hat{\mu}}$ means that the variable μ in the functions g_{a} is held fixed at $\hat{\mu}$. $\hat{\mu}$ is a feasible mean for p with support set $\Omega_p(\gamma_0, c_1)$. The optimal variance is

$$\hat{m}_2 = \arg\min_{m_2} \left\{ \sum_{i=1}^{N} \left(\tilde{m}_{2,y(x^{(i)})} - h_{m_2}\left(\hat{\mu}, m_{2,}, x^{(i)}\right)\right)^2 : c_2(m_2) \leq 0 \right\}. \tag{49}$$

where $c_2 = g_{\mathsf{b}}|_{\underline{z}=\underline{p},\, \overline{z}=\overline{p},\, \mu=\hat{\mu}}$ and $\mathsf{b} = \{4,5\}$.

Solving for $\hat{m}_3$: Let $\underline{p}$ and $\overline{p}$ be the defining vertices of $\Omega_p(\gamma_1, c_3)$ for $c_3 = g_{\mathsf{a}\cup\mathsf{b}}|_{\mu=\hat{\mu},\, m_2=\hat{m}_2}$. The optimal third moment is

$$\hat{m}_3 = \arg\min_{m_3} \left\{ \sum_{i=1}^{N} \left(\tilde{m}_{3,y(x^{(i)})} - h_{m_3}\left(\hat{\mu}, \hat{m}_2, m_3, x^{(i)}\right)\right)^2 : c_4(m_3) \leq 0 \right\}, \tag{50}$$

where $c_4 = g_{\mathsf{c}}|_{\underline{z}=\underline{p},\, \overline{z}=\overline{p},\, \mu=\hat{\mu},\, m_2=\hat{m}_2}$ and $\mathsf{c} = \{6,7,8,9,10\}$.

Solving for $\hat{m}_4$: Finally, let $\underline{p}$ and $\overline{p}$ be the defining vertices of $\Omega_p(\gamma_2, c_5)$ for $c_5 = g_{\mathsf{a}\cup\mathsf{b}\cup\mathsf{c}}|_{\mu=\hat{\mu},\, m_2=\hat{m}_2,\, m_3=\hat{m}_3}$. The optimal fourth moment is

$$\hat{m}_4 = \arg\min_{m_4} \left\{ \sum_{i=1}^{N} \left(\tilde{m}_{4,y(x^{(i)})} - h_{m_4}\left(\hat{\mu}, \hat{m}_3, m_4, x^{(i)}\right)\right)^2 : c_6(m_4) \leq 0 \right\}. \tag{51}$$

where $c_6 = g_{\mathsf{d}}|_{\underline{z}=\underline{p},\, \overline{z}=\overline{p},\, \mu=\hat{\mu},\, m_2=\hat{m}_2,\, m_3=\hat{m}_3}$ and $\mathsf{d} = \{11,12,13,14\}$.

Hence, at each program in the sequence we search for the moments of p of a given order that minimizes the offset between the predicted and the target moment functions of such an order. The solution of any program in the sequence becomes a parameter of the programs that follow. Furthermore, the values of $\underline{p}$ and $\overline{p}$ at any step of the sequence require calculating an IPM subject to constraints $c_i \leq 0$ that make the previously found optimal moments feasible. Note that these optimization programs are carried out in a distribution-free setting. By choosing values of the γs greater than zero the feasibility constraints depending on $\underline{p}$ and $\overline{p}$ are relaxed, and a better matching is attained.

A data-enclosing IPM of minimal spread consistent with the optimal moments is obtained by solving (27) subject to the full probabilistic constraint $c \leq 0$ where $c = g|_{\mu=\hat{\mu},\, m_2=\hat{m}_2,\, m_3=\hat{m}_3,\, m_4=\hat{m}_4}$. The defining vertices of the parameter box of this IPM will be denoted as $\hat{\underline{p}}$ and $\hat{\overline{p}}$ hereafter. Equation (44) evaluated at this set of optimal moments yields

$$\hat{\theta}_p = \left[\, \hat{\underline{p}}, \hat{\overline{p}}, \hat{\mu}, \hat{m}_2, \hat{m}_3, \hat{m}_4 \,\right]. \tag{52}$$

Equations (48-51) evaluated at the optimal moments yields the predicted moment functions

$$\hat{\mu}_{y(x)} = h_{\mu}(\hat{\mu}, x), \tag{53}$$

$$\hat{m}_{2,y(x)} = h_{m_2}(\hat{\mu}, \hat{m}_2, x), \tag{54}$$

$$\hat{m}_{3,y(x)} = h_{m_3}(\hat{\mu}, \hat{m}_2, \hat{m}_3, x), \tag{55}$$

$$\hat{m}_{4,y(x)} = h_{m_4}(\hat{\mu}, \hat{m}_2, \hat{m}_3, \hat{m}_4, x). \tag{56}$$

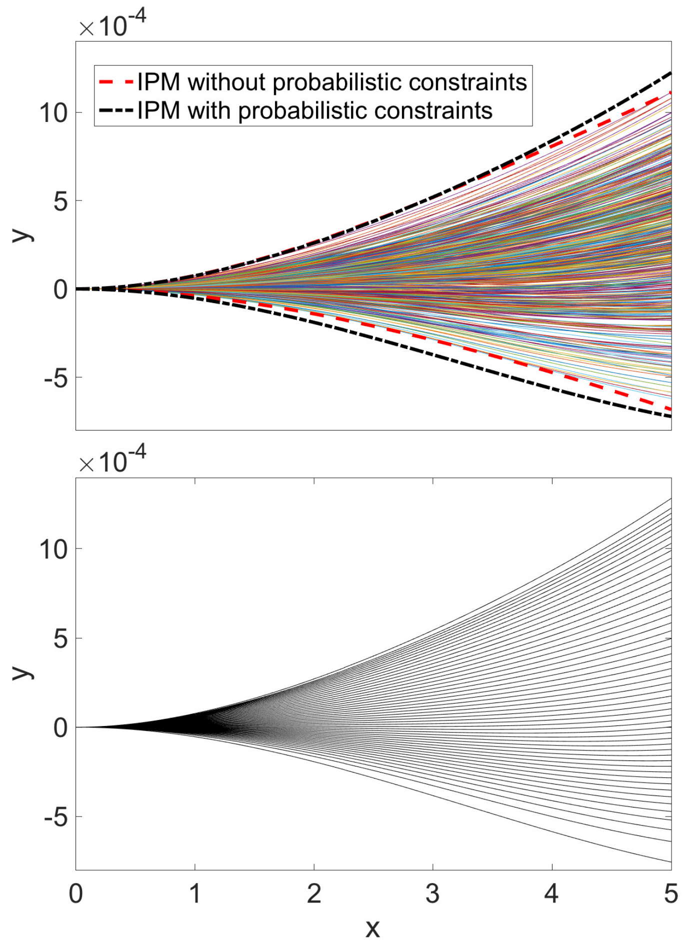

Figure 1: Top: Data ensemble of beam deflections (solid lines), IPM without probabilistic constraints (red dashed line) and IPM with probabilistic constraints (black dashed-dotted line). Bottom: Two-percentiles of the moment-matching RPM.

Any RPM for which $f_p(p)$ in (30) realizes $\hat{\mu}$, $\hat{m}_2$, $\hat{m}_3$, $\hat{m}_4$ yields these predicted moment functions. Such moments along with $\underline{\hat{p}}$ and $\overline{\hat{p}}$ define a probability-box possibly comprised of infinitely many random variables. Each member of this family realizes the optimal moments for the given bound on the support set. In the developments that follow we seek a particular member of this family. For the sake of simplicity we will further assume that the components of p are independent. The more general case of the parameters in p being dependent can be addressed using copulas. In particular, we will pursue a moment-matching RPM having the staircase structure given in (19), such that

$$p_i \sim S_{p_i}\left(\hat{\theta}_{p_i}, n_b, J\right) \text{ for } i = 1, \ldots n_p, \tag{57}$$

where $\hat{\theta}_{p_i}$ is the i-th row of $\hat{\theta}_p$. This form enables the efficient calculation of $f_p(p)$. The resulting PDFs, which depend on the selected cost J, can take on arbitrary shapes including highly skewed and multimodal distributions. This flexibility can be reduced by either restricting the feasible space in (48-51) or by choosing a suitable J.

Example: Next we consider the problem of characterizing the unknown loading applied to a cantilever beam from displacement measurements. $N = 500$ measurements, each consisting of a vector of locations on the beam where the sensors are placed, x, and the corresponding deflection vector y, are available. Therefore, an input-output pair $\{x^{(i)}, y^{(i)}\} \in \mathbb{D}$ for some $i = 1, \ldots, 500$ is given by two vectors of equal length. The top of Figure (1) shows the data ensemble. Note that the displacement at $x = 0$, where the beam is clamped, is zero for all cases.

Euler beam theory justifies the selection of the parametric model in (21), where p represents the applied load. This is a consequence of the principle of superposition which states that on a linear elastic structure, the combined effect of several loads acting simultaneously is equal to the algebraic sum of the effects of each load acting individually. The basis of the model $\varphi(x)$ is chosen to be the beam deflection corresponding to standard loading conditions of magnitude one. Such conditions include point forces and moment forces at fixed input values, as well as distributed loading conditions having a uniform or triangular form over the full extension of the beam. The basis terms corresponding to such loading conditions are

$$\varphi_{\text{force}}(x) = \begin{cases} \frac{x^2}{6EI}(3a - x) & \text{if } 0 \leq x \leq a, \\ \frac{a^2}{6EI}(3x - a) & \text{if } x \geq a \end{cases} \tag{58}$$

$$\varphi_{\text{moment}}(x) = \begin{cases} \frac{x^2}{2EI} & \text{if } 0 \leq x \leq a, \\ \frac{a}{2EI}(2x - a) & \text{if } x \geq a \end{cases} \tag{59}$$

$$\varphi(x)_{\text{uniform}} = \frac{x^2}{24EI}(x^2 + 6L^2 - 4Lx), \tag{60}$$

$$\varphi(x)_{\text{triangular increasing}} = \frac{x^3}{120EIL}(20L^3 - 10L^2x + x^3), \tag{61}$$

$$\varphi(x)_{\text{triangular decreasing}} = \frac{x^2}{120EIL}(10L^3 - 10L^2x + 5Lx^2 - x^3), \tag{62}$$

where E is the Young modulus, I is the area moment of inertia, L is the length of the beam, and $a \leq L$ is the x value at which the load is applied. Note that the selection of this parametric model is supported by physics-based arguments governing the DGM.

Six loading conditions were used to assemble $\varphi(x)$ for given values of E, I, L and a's. With $\varphi(x)$ in hand we then solved for the minimum-spread IPM in (27) with $c = \emptyset$ (no probabilistic constraints). In this setting, the uncertainty box $\hat{P}$ resulting from the IPM calculation prescribes each force and moment as a bounded interval. Figure (1) shows the boundaries of the resulting IPM as red dashed lines. This IPM tightly encloses the data as desired.

The empirical moment functions, shown in Figure (2) as solid lines, were estimated from the data ensemble. These functions were then used to characterize the moments of p by solving (48-51), which in turn, yield the predicted moments for the RPM. These moments, which closely approximate the empirical moments throughout X, are also shown in Figure (2). With the bound to the support set and the optimal moments of p in hand, we can then use (57) to calculate the distribution of the loads. Figure (3) shows the resulting staircase variables for a cost function J that minimizes the degree of non-unimodality in the distributions. Note that the shape of the PDFs, which include multimodal and highly skewed forms, fall outside the spectrum of PDFs described by standard random variables.

Finally, the bottom of Figure (1) shows the two-percentiles of the RPM. This figure was obtained by simulating the RPM and post-processing the predicted responses. Note that the predictor captures well not only the range of deflections but also the input-dependent skewness of the distribution. The

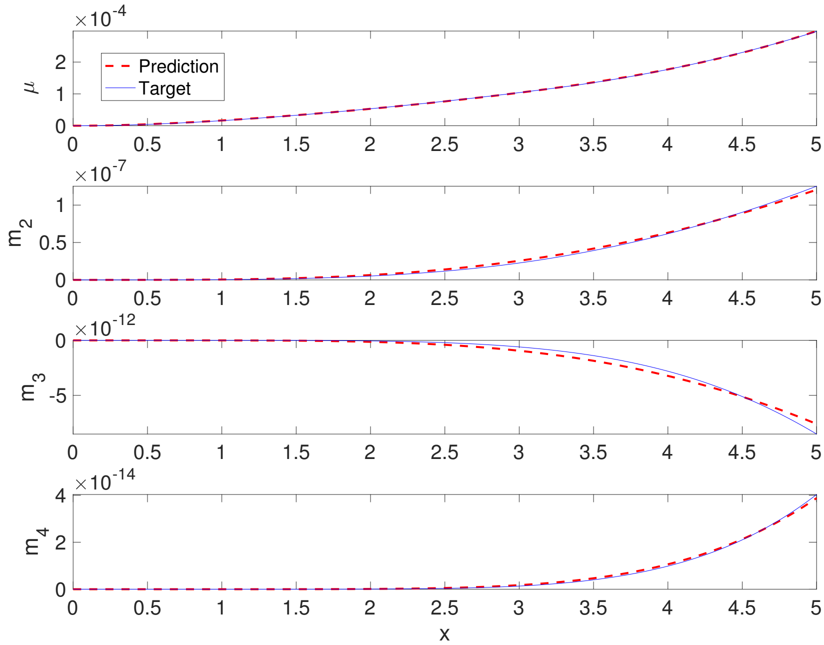

Figure 2: Empirical (solid) and predicted moment functions associated with the RPM (dashed).

boundaries of an IPM constrained by the optimal moments are superimposed at the top. Note that the tightness of the IPM is slightly degraded by the inclusion of such constraints.

This example demonstrates how to infer the randomness in the loading conditions from measurements. Even though Euler beam theory justifies the selection of (21), the analyst must choose which terms to include in $\varphi(x)$. Whereas the exclusion of important terms will yield to model-form uncertainty, thus mischaracterizations of the loads, the inclusion of unnecessary terms will rapidly increase the computational expense of the approach. Such an expense is proportional to order of the moment being matched. Note that an RPM that closely matches the empirical moments does not necessarily match the empirical percentiles. This is particularly valid when there is significant model-form uncertainty.

7. DISCUSSION

IPMs provide the means to describe the support set of a DGM, whereas RPMs provide the means to approximate the manner in which the data is distributed within such a set. Scenario optimization theory enables making a formal, distribution-free assessment on the probability of unobserved data falling outside the predicted interval [9]. By contrast, RPMs enable making a subjective assessment on the manner in which the data are distributed inside their range. Both IPMs and RPMs admit a functional interpretation: they describe an ensemble of infinitely many input-output functions. Only RPMs however, weight each member function according to its likelihood.

IPMs are driven by the extreme, outer-most data points of the data ensemble, which often are low-probability events occurring in the long tails of a distribution. As such, the bulk of the data is inconsequential to the IPM. RPMs mend this deficiency by describing probabilistic features of the DGM within this interval. This higher fidelity description comes with a price: IPMs require solving a single convex optimization program, whereas parametric staircase RPMs require solving one or several

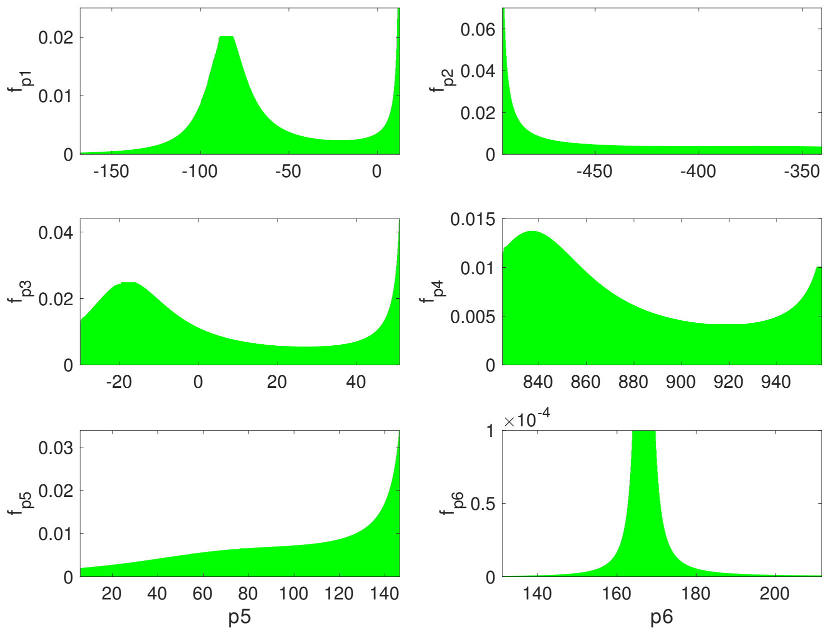

Figure 3: Identified loads.

polynomial optimization programs. Moment-matching RPMs are designed to match functions extracted from the data.

In the examples above, the superset of the support, $\Omega_{y(x)}$, was kept fixed during the calculation of the moments. The prescription of this set has substantive implications on the resulting distribution. Overly tight sets along with extreme moments often yield PDFs with sharp spikes. This is a consequence of over-restricting Θ by using exceedingly small values for $\overline{z}-\underline{z}$. Larger values of $\overline{z}-\underline{z}$ increase the size of the available set of feasible moments, thus the potential for better probabilistic predictions. Hence, the analyst should choose between making either the tightness of the support of the RPM or its probability allocation a priority. These conflicting objectives should be traded-off until the desired balanced is attained.

Note that the selection of the parameters of $\varphi(x)$ affect both the tightness of the support and the predicted distribution. Choosing parameter values that suit one of them might incidentally degrade the performance of the other one. This prescription should be made in accordance with the data in hand and the intended use of the model. Furthermore, RPMs in which a single term in $\varphi(x)$ dominates an input-output region cannot closely track target moments varying therein.

The versatility of a predictor is intrinsically linked to that of the assumed distribution structure for its parameters. As such, staircase variables are particularly suitable for describing complex DGMs. However, all this freedom might render predictors having undesirable spikes and spurious multimodal distributions. This can be avoided by further restricting the feasible space Θ or by relaxing some of the staircase constraints [8].

References

[1] T. Simpson, J. Peplinski, P. Koch, and J. Allen, "Metamodels for computer-based engineering design: survey and recommendations," *Engineering with Computers*, vol. 17, no. 1, pp. 129–150,

2001.

[2] M. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society B*, vol. 63, no. 3, pp. 425–464, 2001.

[3] M. Allen and K. Maute, "Reliability-based design optimization of aeroelastic structures," *Structural and Multidisciplinary Optimization*, vol. 27, pp. 228–242, 2004.

[4] M. S. Eldred, H. Agarwal, V. Perez, S. Wojtkiewicz, and J. Renaud, "Investigation of reliability method formulations in DAKOTA/UQ," *Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance*, vol. 3, no. 3, pp. 199–213, 2007.

[5] L. P. Swiler, B. Adams, and M. Eldred, "Model calibration under uncertainty: Matching distribution information, AIAA-2008-5944," in *AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Victoria, British Columbia, Canada*, September 2008.

[6] C. C. McAndrew, *Compact Modeling: Principles, Techniques, and Applications: Chapter 16 in Statistical Modeling using Backward Propagation of Variance*. New York: Springer, 2010.

[7] L. G. Crespo, D. P. Giesy, and S. P. Kenny, "Minimal-dispersion and maximal-likelihood predictors with a linear parameter dependency," in *Probabilistic Safety Assessment and Management conference, PSAM 14, Los Angeles, CA, USA*, September 2018.

[8] L. G. Crespo, D. P. Giesy, and S. P. Kenny, "On the calculation and shaping of staircase random variables," in *ESREL 2017, Portoroz, Slovenia*, June 2017.

[9] L. G. Crespo, S. P. Kenny, and D. P. Giesy, "Interval predictor models with a linear parameter dependency," *ASME Journal of verification, validation and uncertainty quantification*, vol. 1, no. 2, pp. 1–10, 2016.

[10] M. C. Campi and S. Garatti, "A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality," *Journal of optimization theory and applications*, vol. 148, no. 1, pp. 257–280, 2011.

[11] L. G. Crespo, D. P. Giesy, and S. P. Kenny, "Random predictor models with a nonparametric staircase structure," in *ESREL 2017, Portoroz, Slovenia*, June 2017.