

Application of Interval Predictor Models to Space Radiation Shielding

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

Ryan B. Norman, and Steve R. Blattnig,[‡]

NASA Langley Research Center, Hampton, VA, 23681, USA

This paper develops techniques for predicting the uncertainty range of an output variable given input-output data. These models are called Interval Predictor Models (IPM) because they yield an interval valued function of the input. This paper develops IPMs having a radial basis structure. This structure enables the formal description of (i) the uncertainty in the models parameters, (ii) the predicted output interval, and (iii) the probability that a future observation would fall in such an interval. In contrast to other metamodeling techniques, this probabilistic certificate of correctness does not require making any assumptions on the structure of the mechanism from which data are drawn. Optimization-based strategies for calculating IPMs having minimal spread while containing all the data are developed. Constraints for bounding the minimum interval spread over the continuum of inputs, regulating the IPMs variation/oscillation, and centering its spread about a target point, are used to prevent data overfitting. Furthermore, we develop an approach for using expert opinion during extrapolation. This metamodeling technique is illustrated using a radiation shielding application for space exploration. In this application, we use IPMs to describe the error incurred in predicting the flux of particles resulting from the interaction between a high-energy incident beam and a target.

Abbreviations

DGM Data Generating Mechanism
GCR Galactic Cosmic Rays
GP Gaussian Process
IPM Interval Predictor Model
LS Least Squares

*Corresponding Author, Luis.G.Crespo@nasa.gov. AIAA Associate Fellow.

[†]Aerospace Technologists, Dynamic Systems and Control Branch.

[‡]Aerospace Technologists, Durability, Damage Tolerance, & Reliability Branch.

MLE Maximum Likelihood Estimate
OP Optimization Program
RBF Radial Basis Function

I. Introduction

Metamodeling [19] 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 non-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 its 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 (DGM)), 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

$$y = M(x, p) + \eta, \quad (1)$$

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. 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 $i = 1, \dots, N$, 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.

Parameter estimation is commonly carried out by solving for the parameter realization that minimizes the sum of squared errors between predictions and observations [18]. This approach yields the *Least Squares* (LS) parameter estimate p_{LS} . 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 [18] and prediction intervals require 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 given a specified value of the input. The prediction error data for a linear regression 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 p_{LS} . 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 hyper-parameters of aleatory variables^a. Note that the consideration of aleatory uncertainties requires assuming a structure for them, so they can be parameterized in terms of non-physical 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 of a calibrated model). This deficiency can be mitigated by adding a fictitious discrepancy term to M [11]. 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 for metamodeling.

Another approach to metamodeling is Gaussian Process (GP) Modeling [17] or Kriging. A GP is a stochastic process in which the predicted output is prescribed by a normally distributed random variable whose mean and variance are functions of the input. Moreover, every finite collection of those random variables has a multivariate normal distribution. The distribution of a GP is the joint distribution of all those (infinitely many) random variables, and as such, it is a distribution over functions. The calculation of GPs requires prescribing a structure for a mean and a covariance function and solving for the value of its hyper-parameters. In spite of its high computational complexity, and its inability to accurately describe data having an input-dependent spread, this method is widely used due to its ability to yield broad predictions during extrapolation, e.g., predictions away from the data exhibit a larger spreads.

This paper focuses on IPMs with a linear parameter dependency and a radial basis structure. IPMs are metamodels that cast the spread in the data as interval uncertainty in the parameters of a computational model. Reference [6] presents the mathematical foundation of IPMs. This structure enables the formal description of the uncertainty in the models parameters, the predicted range of outputs, and the probability that a future observation would fall in such a range. This paper presents optimization-based strategies for calculating an IPM and constraints for bounding its spread, regulating its variation/oscillation, and centering its spread about a target point. Furthermore, we develop an approach for using expert opinion during extrapolation. A comparison of IPMs to other metamodeling techniques using numerical experiments is available in the companion paper [7].

^aFor 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 $p = [q_1, \mu, \sigma]^T$ contains three epistemic variables, one physical and two non-physical.

Uncertainty quantification in nuclear physics has attracted considerable attention recently [16, 4, 12, 13]. A component of the risk due to space exploration for astronauts originates from the natural space radiation environment. Space is filled with a high energy, low intensity background radiation environment called Galactic Cosmic Rays (GCR). Exposure to GCR for extended periods of time can lead to increased risk for cancer and cataracts, along with other health problems [9]. Material shielding can potentially reduce such a risk but cannot meet current requirements for long duration missions [8, 20]. To address the shortcomings of physics-based models and better understand the phenomenon, NASA has funded a set of measurements and an associated uncertainty quantification study. In this paper we use IPMs to describe the error incurred in predicting the flux of particles resulting from the interaction between a high-energy incident ion beam and a target.

II. Interval Predictor Models

A. Philosophy

A DGM is postulated to act on a vector of input/state 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. Assume that N input-output pairs are obtained from the DGM, and denote by $\mathbf{z} = \{z_i\}$, with $z_i = (x_i, y_i)$ for $i = 1, \dots, N$, the corresponding data sequence.

It is desired to build a mathematical model of the DGM based on \mathbf{z} which will predict the output corresponding to an unobserved realization of the input. Let $X \subseteq \mathbb{R}^{n_x}$ be 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 . 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 into which unobserved data is expected to fall. 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 fit all of 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 can be fit exactly by the model for at least one element of p in such a set. The restriction considered herein forces p to belong to a hyper-rectangle P . For a fixed value of the input x , the propagation of the set P through M yields an interval of output values y . Thus, these models are called interval predictor models. The thrust 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 $p \in P$.

In this setting, the two main problems of interest can be stated as follows. First, we want to find an empirical model (or rule) 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. These features, which are cast by the analyst as design requirements on the empirical model (e.g., we want all observed outputs to be within the predicted range), constitute the criteria driving the search for an optimal empirical model. Second, we want to quantify rigorously the probability that y_{N+1} be compliant with such requirements (e.g., the probability that y_{N+1} will be inside the predicted range). In this setting, the prediction targeted 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 . This will be pursued without making any assumptions on the underlying structure of the DGM^b.

B. Mathematical Structure

An IPM is simply a function that returns an interval as output. These models were proposed in the framework of differential inclusions and set-valued dynamical systems [1, 2], and further developed for the case in which prior knowledge on the DGM is available [14, 15]. In the context of this paper, an IPM is a rule that assigns to each instance vector $x \in X$ a corresponding outcome interval in Y . That is, an IPM is a set-valued map

$$I_y : x \rightarrow I_y(x) \subseteq Y, \quad (2)$$

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 $X \times Y$. A parametric IPM is obtained by associating to each $x \in X$ the set of outputs y corresponding to all parameter values p in the set P :

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

Attention will be limited to the case in which the output is a linear combination of Radial Basis Functions (RBF) of the input, p are the coefficients of such a combination, and the uncertainty set P is the bounded hyper-rectangle^c:

$$P = \{p : \underline{p} \leq p \leq \bar{p}\}. \quad (4)$$

The parameter points \underline{p} and \bar{p} are called the *Defining Vertices* of P . Hence, the corresponding IPM is given by

$$I_y(x, q, P) = \{y = p^\top \varphi(x; q), p \in P\}, \quad (5)$$

for $\varphi(x; q) = \{\varphi_1(x; q^1), \dots, \varphi_{n_p}(x; q^{n_p})\}$, where each φ_i is a non-negative-valued RBF and q^i is the part of q used by φ_i . This setting leads to

$$I_y(x, q, P) = [\underline{y}(x, q, \bar{p}, \underline{p}), \bar{y}(x, q, \bar{p}, \underline{p})], \quad (6)$$

where

$$\underline{y}(x, q, \bar{p}, \underline{p}) = (p_0 - m)^\top \varphi(x; q), \quad (7)$$

$$\bar{y}(x, q, \bar{p}, \underline{p}) = (p_0 + m)^\top \varphi(x; q), \quad (8)$$

with $p_0 = (\bar{p} + \underline{p})/2$, and $m = (\bar{p} - \underline{p})/2$. The functions \underline{y} and \bar{y} are, respectively, the lower and upper boundaries of the IPM. Each member of the family of infinitely many predictions that results from evaluating the model M at each realization $p \in P$ lies between them, and no tighter containing functions exist. Therefore, the boundaries of I_y are linear functions of \underline{p} and \bar{p} , and radial basis functions of the input. Note that the center of the interval depends

^bWe will only assume that the DGM is stationary and the observations in z are Independent and Identically Distributed (IID).

^cThroughout this paper, vector inequalities hold component-wise.

on the geometric center of the uncertainty box, p_0 ; not on its size. The spread of $I_y(x, P)$, which is the distance between the its boundaries, is

$$\delta_y(x, q, \underline{p}, \underline{p}) = 2m^\top \varphi(x; q). \quad (9)$$

Note that the spread depends on the semi-diagonal of P , m , and not on its geometric center p_0 . Further notice that a reduction in the volume of P might yield a larger spread, and that $\delta_y(x)$ exhibits the same asymptotic behavior of $M(x)$, e.g., If Gaussian RBFs are used, the spread converges to zero as the separation between x and the centers c increases.

The above equations fully specify an IPM given an uncertainty box P . Means to calculate P leading to optimal IPMs are presented next.

C. IPM Formulations

In the first formulation, we seek an IPM given by (4-5) where the limits of P are given by the solution to the following convex Optimization Program (OP).

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

$$\{\hat{p}, \hat{p}\} = \operatorname{argmin}_{u, v : u \leq v} \{E_x[\delta_y(x, q, v, u)] : \underline{y}(x_i, q, v, u) \leq y_i \leq \bar{y}(x_i, q, v, u), i = 1, \dots, N\}, \quad (10)$$

where $E_x[\cdot]$ is the expected value operator with respect to the input x , q is a constant vector set in advance, and (x_i, y_i) is an element of the data sequence \mathbf{z} .

Hence, we search for the set P that minimizes the expected interval spread such that all the observed outputs are in $I_y(x, q, P)$. When x is a random vector of analytically known distribution, the cost function in (10) can be calculated analytically or via numerical quadrature. Otherwise, the sample mean based on the data in \mathbf{z} should be used. The value of q can be prescribed using empirical arguments according to the concentration and spread of the data. The probability that a future observation will fall in $I_y(x, q, \hat{P})$ can be rigorously bound using the developments of Section II.E. A natural improvement to the formulation above results from simultaneously optimizing over the parameters q . This practice leads to the following OP.

Optimization Program 2 (OP2). *The value of q and the defining vertices of \hat{P} are given by*

$$\{\hat{q}, \hat{p}, \hat{p}\} = \operatorname{argmin}_{q, u, v : q \in Q, u \leq v} \{E_x[\delta_y(x, q, v, u)] : \underline{y}(x_i, q, v, u) \leq y_i \leq \bar{y}(x_i, q, v, u), i = 1, \dots, N\}, \quad (11)$$

where $E_x[\cdot]$ is the expected value operator with respect to the input x , Q is the parameter set corresponding to feasible RBFs, and (x_i, y_i) for $i = 1, \dots, N$ are the observations in \mathbf{z} .

Hence, we search for the parameter q and the set P that minimize the expected interval spread such that all the observed outputs are in $I_y(x, q, P)$. As before, the probability that a future observation will fall in $I_y(x, \hat{q}, \hat{P})$ can be rigorously bounded using the developments in Section II.E.

IPMs and uncertainty sets resulting from (11) are tighter than those resulting from (10). However, the OP in (11) is non-convex in general. This makes its computational demands

substantially greater than those of OP1. Note however that any feasible solution to (11) yields an IPM that encompasses all the data as required. As such, the only consequence of (11) converging to a non-global optimum is that the resulting IPM will have a suboptimal spread.

The developments that follow focus on Gaussian RBFs, for which

$$\varphi(x; q^i) = e^{-\left(\frac{x-c_i}{\sigma_i}\right)^2}, \quad (12)$$

for $i = 1, \dots, n_p$, where $c \in \mathbb{R}^{n_p}$ are the centers, $\sigma \in \mathbb{R}^{n_p}$ are the length-scale parameters, and $q = [c, \sigma]$. The same rationale can be applied to RBFs with other structures.

D. Over-Fitting and Under-Bounding

Figure 1 shows an IPM having a Gaussian RBF structure based on $N = 40$ observations and the OP1 formulation. The points shown as circles have the value of the centers c as the abscissas and the midpoint of the predicted interval as the ordinates. The LS prediction is shown as a dashed line. Note that all observations are within the IPM boundaries as intended, but the LS prediction is not. More importantly, note that the spread of the IPM converges to practically zero as x increases, thus, the IPM boundaries start interpolating/over-fitting the data. This feature, which is predicted by Equations (7-8), might occur at interior points of the input domain X . This potentially undesirable behavior can be controlled by using the constraints proposed below. In general, such constraints lead to larger uncertainty sets \hat{P} and IPMs with larger spreads. When the value of q is fixed these constraints are convex, thus, their inclusion into an OP preserves convexity.

1. Bounding the Minimum Spread

The value of \hat{P} along with the corresponding value of q might lead to IPMs having inadmissibly small spreads for some values of x in X . One correction for this is to introduce a constraint ensuring that

$$\delta_y(x, q, \bar{p}, \underline{p}) \geq \epsilon \text{ for all } x \in X. \quad (13)$$

The case in which the input is one dimensional, so $n_x = 1$, is considered first.

To implement this idea, we want to find the minimum spread $\delta_y(x)$ over the range of measured inputs. This domain, to be denoted as X_m , is often prescribed by the experimentalist before measuring data. In a one dimensional setting, we use $X_m = [\min(c), \max(c)]$. Since $\delta_y(x)$ is a smooth function of x , the minimum occurs either at an endpoint of the interval or at an interior point for which the derivate $\delta'_y(x)$ is equal to zero. One easily obtains an analytic expression for $\delta'_y(x)$, but the problem of solving for its zeros is analytically intractable. Numerical techniques could be used, but their integration into the search for an IPM with a minimum allowable spread could make the approach computationally prohibitive. Instead, we want to find a lower bound to the spread whose minimum can be evaluated relatively cheaply.

The strategy is to partition X_m into n_i sub-intervals, and generate a lower bound to each $\varphi(x; q_i)$ on each sub-interval by using an affine function. A lower bound to $\delta_y(x, q, \bar{p}, \underline{p})$ is obtained by replacing each component of $\varphi(x; q)$ in (9) with a linear combination of such functions. Because this combination is also affine, its minimum on the sub-interval occurs

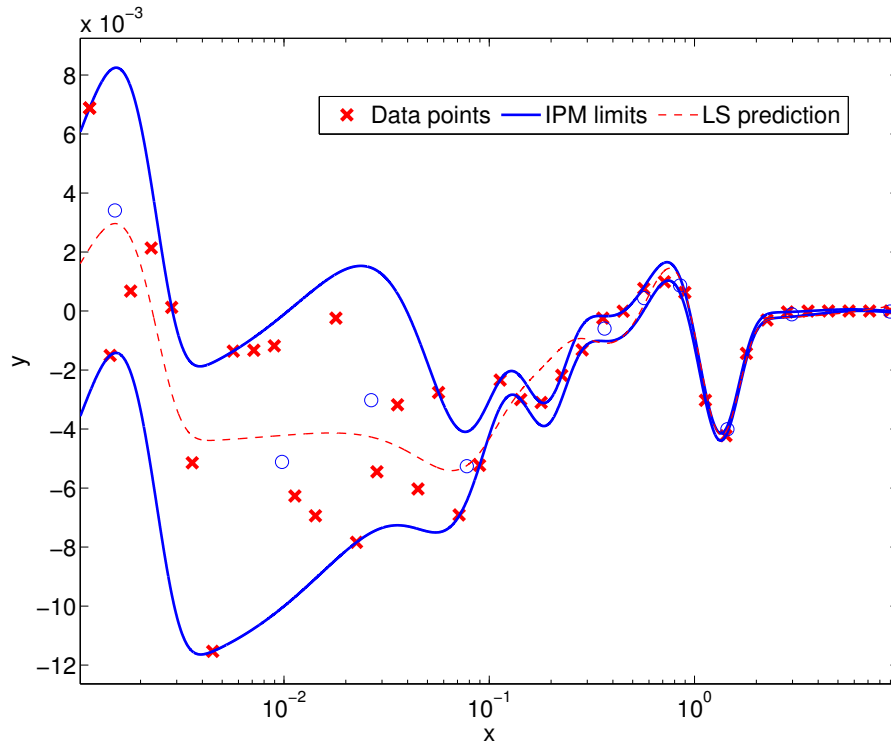


Figure 1. Unconstrained IPM.

at one of the endpoints. Hence, a lower bound to the minimum of $\delta_y(x, q, \bar{p}, \underline{p})$ on X_m is the minimum of all these sub-interval lower bounds.

We still need to specify how to partition X_m and which affine function to use as a lower bound for each basis function on each sub-interval. We first make some observations about the nature of the geometry of the graph of a Gaussian RBF. This curve has a unique maximum at c_i and exactly two inflection points at $c_i \pm \sigma_i/\sqrt{2}$. To the left of the left inflection point, the curve is monotone increasing and convex. Between the inflection points, it is concave. To the right of the right inflection point, it is monotone decreasing and convex. On any sub-interval which falls entirely within the concave portion, the chord joining the endpoints is a lower bound. On any sub-interval which falls entirely within one of the convex portions, any line segment which is tangent to the curve is a lower bound. We choose to use the tangent at the endpoint with the smaller basis function value. This ensures that the line segment also stays above the x -axis over the sub-interval.

This suggests that we choose a partition so that none of the sub-intervals cross over any of the inflection points. In addition, we will include all the centers in the partition. That way, the chords drawn within the concave region will closely bound the Gaussian basis function. Hence, we partition the interval X_m using all the points in the vector c of centers and any of the inflection points $c_i \pm \sigma_i/\sqrt{2}$ falling into X_m . Denote by x^j , with $j = 1, \dots, n_i + 1$, the nodes of the resulting partition. Any pair of consecutive nodes in this sequence defines the limits of a sub-interval.

Given this partition, we now need to find an affine lower bound for each Gaussian basis function on each sub-interval. This function is completely determined by its values at the two

sub-interval endpoints. Each node except the first and last participates in two intervals. The values it gets from the lower bounding affine functions on the left and the right might not be the same. Since we ultimately want a lower bound, we select the smaller of these values to use in what now becomes a piecewise continuous linear bound $l(x; q^i)$ for $\varphi(x; q^i)$. The values of $l(x; q^i)$ at each of the nodes depend only on q^i and not on m . When we form the linear combination of these bounds, one for each basis function, we get the piecewise linear function $2m^\top l(x; q)$, which in turn is a lower bound of $\delta_y(x, q)$. Numerical experiments indicate that this is bound is tight. A lower bound for the minimum spread on the interval X_m is the minimum value of this lower bound at any of its nodes. Consequently, the requirement in (13) is satisfied when the $n_i + 1$ convex constraints

$$2m^\top l(x^j; q) \geq \epsilon, \text{ for } j = 1, \dots, n_i + 1, \quad (14)$$

hold. This set of constraints is equivalent to the convex constraint

$$\min_{j=1, \dots, n_i+1} 2m^\top l(x^j; q) \geq \epsilon, \quad (15)$$

which correspond to the intersection of all n_i convex sets defined by (14).

The extension of this idea to a higher dimensional input space is presented next. The case where $n_x > 1$ presents problems absent in the $n_x = 1$ case. We have found a natural way to divide the 1-dimensional domain into sub-intervals, using geometrically significant features of the Gaussian basis functions – namely, the centers and inflection points. It is not so obvious how to make use of geometric properties of multi-dimensional Gaussian basis functions to generate a partitioning of a multi-dimensional hyper-rectangular domain.

For simplicity, we limit our attention to Gaussian basis functions which may have different length scale for each dimension but have no cross-correlation terms. Viewing the multi-dimensional hyper-rectangular domain as a Cartesian product of intervals, we partition each interval into sub-intervals and generate a partitioning of the hyper-rectangular domain into n_r sub-hyper-rectangles by taking, in every possible combination, the Cartesian product of a sub-interval generated by the partition of each coordinate interval. On each sub-hyper-rectangle, each Gaussian basis function is under-estimated by its minimum value over the sub-hyper-rectangle. This only requires knowledge of the values of the function at the vertices of such a sub-hyper-rectangle. Using this minimum value provides a lower bound for each Gaussian basis function on each sub-hyper-rectangle, although not one which bounds as closely as the one-dimensional bound using line segments. If, on the fixed sub-hyper-rectangle j , the minimum value for Gaussian basis function number i is l_i , then $2m^\top l_j$ is a lower bound for the spread. In this setting and analogous to (15), the requirement in (13) is satisfied when the convex constraint

$$\min_{j=1, \dots, n_r} 2m^\top l_j \geq \epsilon, \quad (16)$$

holds. Hence, the requirement in (13) is cast as a single constraint regardless of how fine the partition of X_m is. The choice of the individual coordinate partitions involves a trade-off. On the one hand, increasingly fine partitions produce sharper lower bounds. On the other hand, increasingly fine partitions increase n_r exponentially fast, thus the amount of computational effort required to evaluate (16).

Figure 2 shows the IPM resulting from adding the minimum spread constraints with $\epsilon = 2.3 \times 10^{-3}$ to the same formulation and data set used in Figure 1. Note that the spread

of the IPM exceeds the prescribed threshold over the entire X domain. Incidentally, the minimum spread constraint reveals an unwanted oscillation in the prediction. Means to regulate this oscillation are provided next.

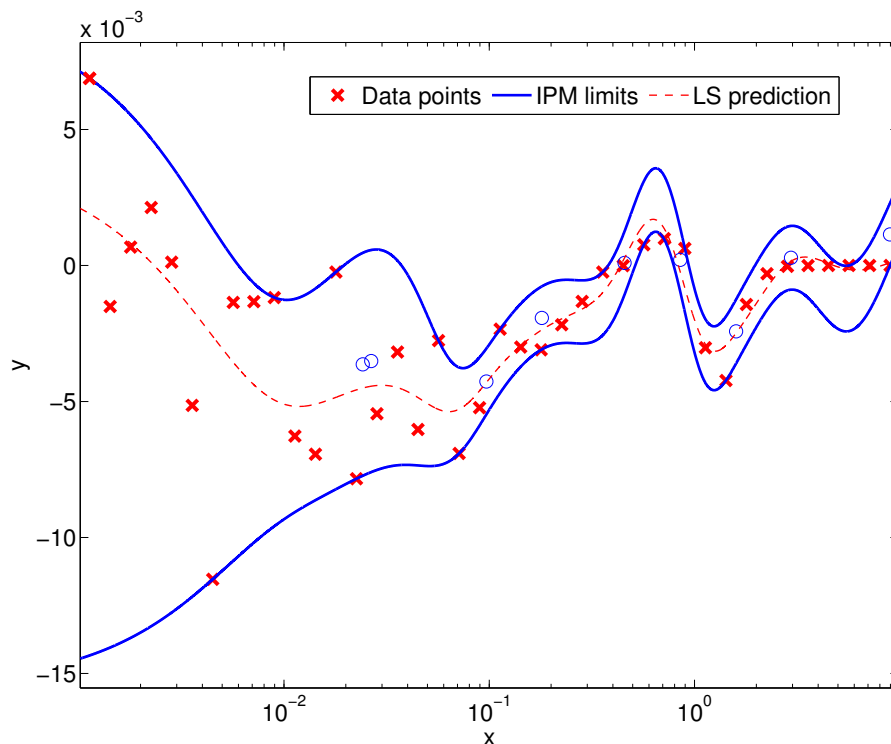


Figure 2. IPM with minimum spread constraints.

2. Regulating Oscillatory Behavior

The IPM at a fixed value of x is an interval that results from adding several sub-intervals, one for each term in the linear combination. Individual sub-intervals dominating this sum yield rapidly oscillating IPMs. This variation can be controlled by ensuring that no single sub-interval dominates the sum at the RBF centers. This can be enforced using the linear constraints:

$$(1 + \gamma)m_i\varphi(c_i; q^i) \leq \sum_{k=1, k \neq i}^{n_p} m_k\varphi(c_i; q^k), \text{ for } i = 1, \dots, n_p \quad (17)$$

where γ is a tuning parameter set by the analyst. The satisfaction of the i th constraint implies that the contribution of all but the i th the term to the spread at c_i exceeds the contribution of such a term by a factor proportional to γ .

Figure 3 shows the IPM resulting from adding (17) for $\gamma = 0.4$ to the same formulation and data set used in Figure 1. Note that the oscillation has been reduced. As before, the LS prediction is outside the IPMs at a few places. Larger values of γ yield more attenuation of the oscillation and larger local spreads.

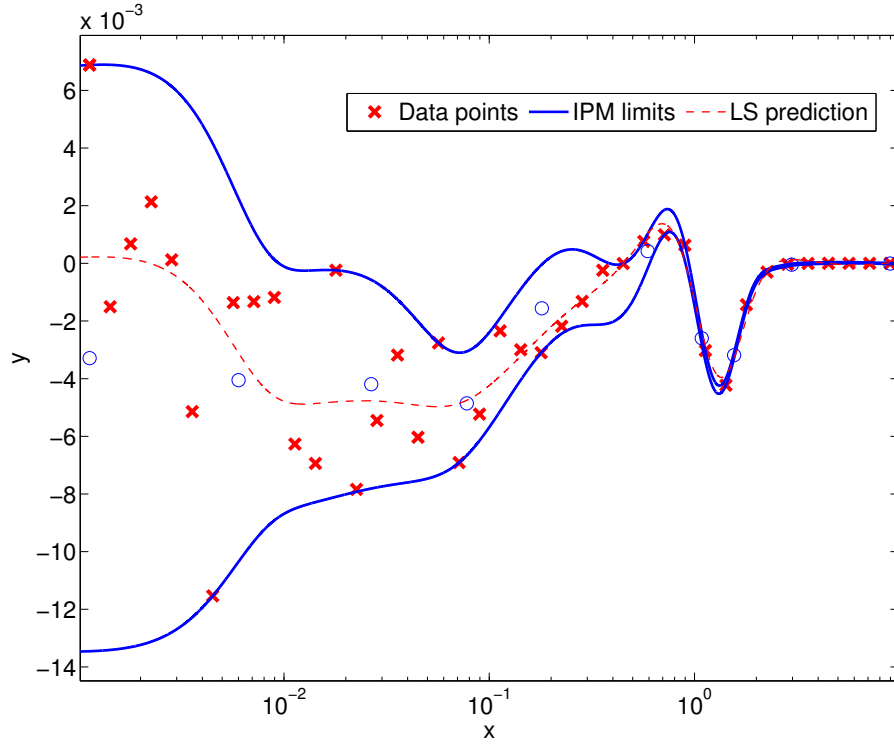


Figure 3. IPM with reduced oscillation constraint.

3. Using the Least Squares Prediction

The DGM is commonly approximated by the Least Square (LS) prediction. In the context of RBFs, this prediction is $y = p_{\text{LS}}^{\top} \varphi(x; q)$, where p_{LS} , the solution to

$$p_{\text{LS}}(q) = \underset{p}{\operatorname{argmin}} \sum_{i=1}^N (y_i - p^{\top} \varphi(x_i; q))^2, \quad (18)$$

is

$$p_{\text{LS}}(q) = (A^{\top} A)^{-1} A^{\top} [y_1, \dots, y_N]^{\top}, \quad (19)$$

where $A_{i,j} = \varphi_j(x_i; q)$, for $i = 1, \dots, N$ and $j = 1, \dots, n_p$.

The formulations for calculating IPMs above do not guarantee that $p_{\text{LS}} \in \hat{P}$. Whereas $y = p_{\text{LS}}^{\top} \varphi(x; q)$ describes the dominant trend of the data by weighting all data points equally, $I_y(x, q, \hat{P})$ describes their spread. As such, there is no reason to expect that $p_{\text{LS}} \in \hat{P}$ nor that $\underline{y}(x) \leq p_{\text{LS}}^{\top} \varphi(x; q) \leq \bar{y}(x)$ for all $x \in X$. The analyst might want to relate the LS and IPM predictions by ensuring that p_{LS} is a member of \hat{P} . For instance, p_{LS} might be thought as the best deterministic estimate of p , and a \hat{P} centered at such a point can be interpreted as the smallest set of perturbations required by the model to fit all observations. The membership of p_{LS} in \hat{P} can be ensured by replacing $u \leq v$ in (10-11) with

$$u \leq p_{\text{LS}}(q) \leq v, \quad (20)$$

to be called the LS inclusion constraint, or by using

$$u + v = 2p_{\text{LS}}(q), \quad (21)$$

to be called the P -centering constraint. The equality constraint ensures that the IPM prediction is centered at the LS prediction, so $p_{\text{LS}}^\top(q)\varphi(x; q) = (y(x, q, \hat{p}, \hat{p}) + \bar{y}(x, q, \hat{p}, \hat{p}))/2$. In general, the inclusion of either of these constraints leads to IPMs with larger spreads, with the P -centering constraint leading to the larger of the two.

Figure 4 shows the IPM resulting from adding the P -centering constraint to the same formulation and data set used in Figure 1. Note that the resulting IPM exhibits a larger spread in most of the range X and its midpoint function coincides with the LS prediction. Finally, Figure 5 shows the IPM resulting from using the minimum spread, the reduced

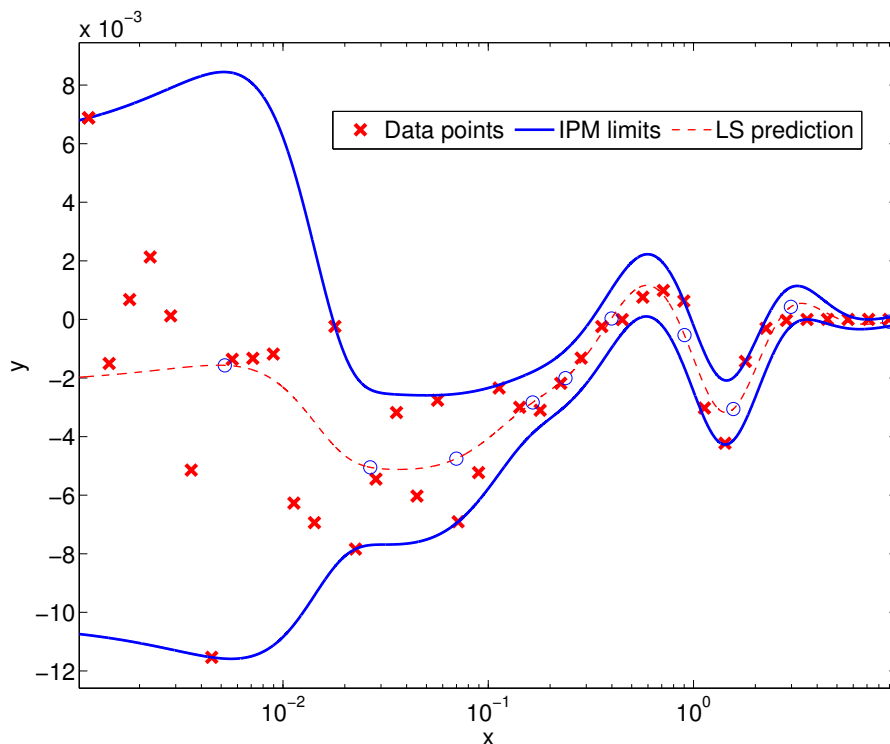


Figure 4. IPM with P -centering constraint.

oscillation and the P -centering constraints altogether. The resulting interval prediction exhibits the desired behavior within the data set X_m .

E. Reliability

Please refer to [6] for a detailed presentation on the reliability of IPMs. The developments that follow are included to make this paper self-contained. Denote by \mathbb{P} the *unknown* distribution of the DGM from which the data points (x_i, y_i) are obtained. No assumption is made on \mathbb{P} so that the functional form relating x and y can be arbitrary.

The reliability of the arbitrary IPM \mathcal{I} , $r(\mathcal{I})$, is defined as

$$r(\mathcal{I}) = \text{Prob}_{\mathbb{P}} \left[(x, y) \in I_y \left(x, q, \hat{P} \right) \right], \quad (22)$$

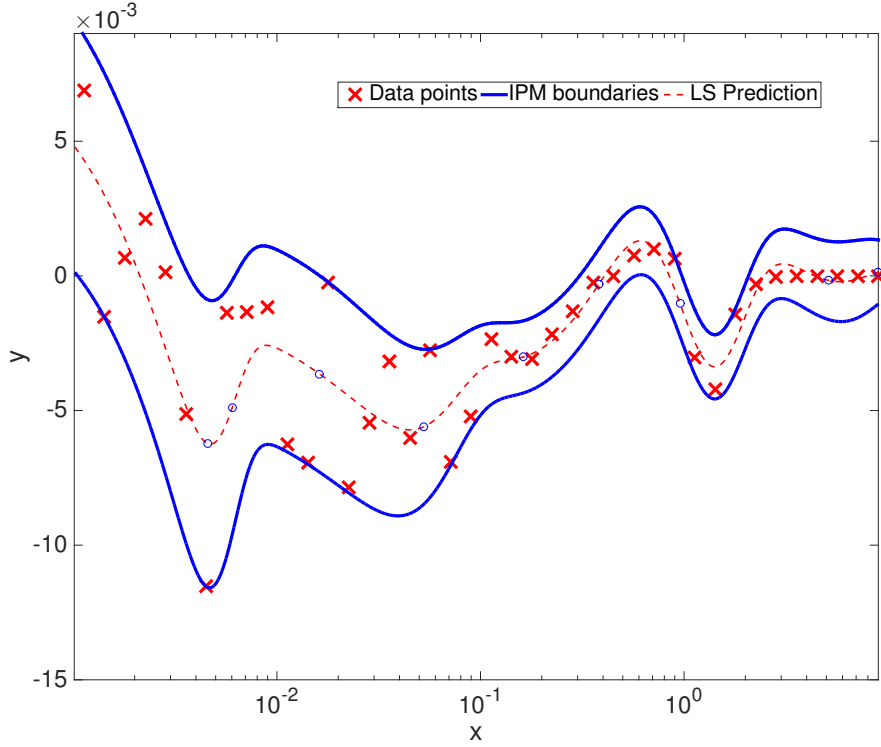


Figure 5. Fully constrained IPM.

where $\text{Prob}_{\mathbb{P}}[\cdot]$ is the probability operator based on \mathbb{P} . Hence, $r(\mathcal{I})$ is the probability that the unobserved input-output pair (x, y) will fall into the predicted interval. $r(\mathcal{I})$ can be bounded using the following Theorem, based on the Scenario Approach of Campi [5] et al.

Theorem 1. *Let \mathbf{z} be a data sequence of N independent elements resulting from a stationary DGM. Suppose the model \mathcal{I} is calculated by solving a convex constrained OP having a unique solution based on \mathbf{z} . Then, for any confidence parameter $\beta \in (0, 1)$ and any reliability parameter $\epsilon \in (0, 1)$ which satisfy*

$$\sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1 - \epsilon)^{N-i} \leq \beta, \quad (23)$$

then

$$\text{Prob}_{\mathbb{P}^N} [r(\mathcal{I}) \geq 1 - \epsilon] > 1 - \beta. \quad (24)$$

This theorem provides an assessment on unobserved data. The theorem states that the reliability of \mathcal{I} is no worse than $1 - \epsilon$ with probability greater than $1 - \beta$. As for the probability $1 - \beta$, one should note that \mathcal{I} is a random IPM that depends on N observations of \mathbb{P} . Therefore, its reliability can be greater or equal than $1 - \epsilon$ for some random observations but not for others, and β refers to the probability $\mathbb{P}^N = \mathbb{P} \times \dots \times \mathbb{P}$ of observing a bad set of N 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. The confidence probability $1 - \beta$ is key for obtaining results that are guaranteed independently of the DGM.

It is worth noting that the confidence parameter can be made very small such that it loses any practical significance, thus $r(\mathcal{I}) \geq 1 - \epsilon$. This can be done without letting N be too large because β vanishes exponentially fast with N . Note that assessing the reliability of the model does not require knowing \mathbb{P} . Equation (24) is a fundamental relation linking the amount of information available (represented by the number of observations N), 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 .

The convexity of the OP1 enables the direct application of Theorem 1 to the resulting IPMs. This is not the case for IPMs derived using OP2 since such a formulation is non-convex. However, the reliability of such models can be established by using the *Principle of Equivalence* [6]. This principle is based on finding an auxiliary convex OP that will result in the very same IPM found by solving the non-convex formulation. If this is attained, the reliability of the empirical 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 IPMs derived using OP2 by noticing that OP1 with $q = \hat{q}$ is the corresponding auxiliary problem.

Note that the inclusion of the constraints above will likely increase the actual reliability of the IPM (since the corresponding spread increases), but will not affect its analytical lower bound $1 - \epsilon$.

Using the data sequence \mathbf{z} to derive both p_{LS} and $I_y(x, q, \hat{P})$ having constraint (21) 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 [5]. 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 \mathbf{z} changes p_{LS} , thus, all the constraints in (10) and (11). Hence, strictly speaking, the usage of constraint (21) makes all such $2N$ constraints support constraints. 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. This unwanted dependency is eliminated by choosing a value for p_{LS} that is independent from the data used to build the IPM. This, for instance, can be attained by partitioning the data set into two subsets, using one to calculate p_{LS} via (18), and using the other one to calculate $I_y(x, q, \hat{P})$ via (10) or (11). The partition of the data into subsets is a common practice in model calibration, testing and validation.

F. Extrapolation

The developments above enable the construction of data-based metamodels. The input domain associated with the data, denoted as X_m , is usually chosen by the experimentalist before data is drawn from the DGM. The use of a metamodel outside X_m will be called extrapolation. Whereas the prediction within X_m can be objectively assessed by comparing it with the data, there is no basis to expect any data-based metamodel to properly represent the DGM during extrapolation. Conjectures about the behavior of the DGM outside X_m are often used to justify extrapolation, say, because of expert opinion, assumed continuity of the moments, or the understanding of the underlying physics driving the DGM. The goodness of such a prediction will only depend on the correctness of such assumptions. Note that

the existence of data gaps, say, because our inability to take measurements outside X_m , is equivalent to having a DGM based on the conditional probabilistic cloud $\{\mathbb{P}|x \in X_m\}$. Using a metamodel based on data extracted from $\{\mathbb{P}|x \in X_m\}$ to make predictions on a DGM based on \mathbb{P} over a larger input domain might lead to significant error. Naturally, inferences and predictions based on any data-based metamodel techniques must conform to the measured DGM to be valid. This applies to the upper reliability bound of IPMs, which is only applicable to future observations falling within X_m .

The predicted interval corresponding to any data-based metamodeling technique should contain as many observations as the confidence level dictates as tightly as possible. This often implies having an interval spread that, as the data do, expands and contracts with the value of the input. The tighter the interval the smaller the conservatism and the more informative the prediction. On the basis of the arguments above, the prediction resulting from a data-based metamodel during extrapolation is unreliable. To counteract for this deficiency and to avoid the usage of unsubstantiated narrow intervals, predictions outside X_m should be conservative. This is one of the main appeals of Gaussian Process (GP) models, whose predicted variance grows rapidly^d with the separation from X_m . This behavior is a safeguard against using the model beyond the intended domain of application.

A technique for modifying the behavior of a data-based IPM outside X_m is presented next. Reasons to change the prediction during extrapolation include the desire to incorporate expert opinion, to make the prediction satisfy physics-based considerations, and to make the prediction more conservative. The latter modification can be used to indirectly restrict and communicate the input-domain where the metamodel's prediction is sufficiently accurate.

Consider the IPM given by

$$I_y(x) = I_{\text{data-based}}(x) + I_{\text{belief-based}}(x), \quad (25)$$

where $I_{\text{data-based}}(x)$ is calculated based on one of the OPs above, and $I_{\text{belief-based}}(x)$ is a subjective IPM built according to the following considerations: (i) $I_{\text{belief-based}}(x) = [0, 0]$ when $x \in X_m$ (so the prediction within X_m is data-driven), (ii) the spread of $I_{\text{belief-based}}(x)$ depends on the separation between x and the set X_m (so we can account for the effects of having to generate a prediction away from the data), (iii) the spread of $I_{\text{belief-based}}(x)$ in the n_x input dimensions is proportional to $m_x \in \mathbb{R}^{n_x}$, where m_x is set in advance by the metamodeler (so we can account for the different units and features of individual components of x).

Means to calculate the separation between the point x and the set X_m are provided next. This separation will be given by the semi-diagonal of the largest hyper-rectangle with center x and orientation m_x not intersecting X_m . In what follows, we assume that X_m is a convex set^e. Convexity enables evaluating the separation between x and X_m , thus, $I_{\text{belief-based}}(x)$

^dUnfortunately, the tuning of a GP model alters the prediction inside and outside X_m in a coupled and non-intuitive manner. Whereas the prediction within X_m converges to a process with a constant variance (so it is unsuitable for representing DGMs having input-dependent spreads), the predicted variance during extrapolation grows unbounded as desired. Tuning this growth however, which strongly depends on the assumed covariance structure and the corresponding prior, is a difficult under-parameterized task coupled to the prediction in X_m .

^eWhen data gaps are present in the input domain, we can represent X_m as a group of non-intersecting convex sets, calculate the separation to all of those sets, and chose the separation corresponding to the smallest $\|s\|$.

efficiently. Without loss of generality, the set X_m is represented as

$$X_m = \{x : g(x) < 0\}, \quad (26)$$

where $g \in \mathbb{R}$. For instance, if X_m is a sphere of radius R centered at x_c , we have $g(x) = \|x_c - x\|_2 - R$, whereas for a hyper-rectangle with semi-diagonal $m > 0$ centered at x_c we have $g(x) = \|x_c - x\|_\infty^m - 1$, where $\|a\|_\infty^m = \max_{i=1, \dots, n_x} a_i/m_i$.

Denote by $s \in \mathbb{R}^{n_x}$ the separation between x and X_m , where $s_i > 0$ is the separation in the i th input direction, and by $m_x > 0$ the desired rate of growth in the spread. The separation s is given by

$$s(x, m_x) = \begin{cases} 0 & \text{if } g(x) < 0, \\ \hat{\alpha} m_x & \text{otherwise,} \end{cases} \quad (27)$$

where

$$\{\hat{\alpha}, \hat{x}_w\} = \underset{\alpha, x_w}{\operatorname{argmin}} \{ \alpha : -\alpha m_x \leq x_w - x \leq \alpha m_x, g(x_w) \leq 0, \alpha \leq 0 \}. \quad (28)$$

In this setting, a hyper-rectangle with center $x \notin X_m$ and semi-diagonal s touches the boundary of X_m at \hat{x}_w .

A model structure for $I_{\text{belief-based}}(x)$ is given by

$$I_{\text{belief-based}}(x) = -u(s)I_{\text{data-based}}(x) + [v(x) - w(s), v(x) + w(s)], \quad (29)$$

where the functions $u \in \mathbb{R}$, $v \in \mathbb{R}$ and $w \in \mathbb{R}$ are prescribed by the analyst. The function u is a non-decreasing function of all the components of s having the range $[0, \mu]$ and satisfying the conditions:

$$u(0) = 0, \quad (30)$$

$$\lim_{\|s\| \rightarrow \infty} u(s) = \mu \leq 1. \quad (31)$$

The function $u(s)$ aims at canceling out the data-based IPM in (25) as the separation between x and X_m increases, e.g., $u(s) = \mu - e^{-s^\top s}$. Note that $u(s) = 0$ attains the desired objective when $I_{\text{data-based}}(x)$ uses a radial basis. The function $v(x)$ is assumed to be any continuous function satisfying

$$v(x) = 0 \text{ for all } x \in X_m. \quad (32)$$

This function aims at describing the midpoint of $I_y(x)$ as the value of s increases, e.g., $v(x) = s^\top s$. Finally, the function $w(s)$ is a non-negative function satisfying the conditions:

$$w(0) = 0, \quad (33)$$

$$\lim_{\|s\| \rightarrow \infty} w(s) = k \geq 0, \quad (34)$$

where k is the asymptotic value of the spread of $I_y(x)$. The function $w(s)$ describes the spread of $I_y(x)$ as a function of the separation s , e.g., $w(s) = k - e^{-s^\top s}$. The IPM resulting from substituting (29) into (25) is equal to the data-based IPM for all x in X_m , and converges

to $[v(x) - w(s), v(x) + w(s)]$ as the separation of x from X_m increases. In this setting, s is an interpolating variable blending the data-based and belief-based IPMs.

Figure 6 shows a family of IPMs having the structure in (25) for the same $I_{\text{data-based}}(x)$. $I_{\text{data-based}}(x)$, shown in blue, is based on the observations in \mathbf{z} that fall within $X_m = \{x : x < 0.1\}$. As such, the separation is given by $s(x) = \max\{0, x - 0.1\}$. Note that this IPM only contains the data in X_m . It is desired to change the IPM during extrapolation according to two main features. For this example lets assume that physics-based arguments require for y to converge to zero for a sufficiently large x . Furthermore, we also want to avoid reaching this asymptotic behavior at values of x that are not large enough so the prediction is not overly tight. Members of a family of belief-based IPMs having $u(s) = 0$, $v(x) = 0$, and $w(x) = ase^{-bs^2}$ for $a, b > 0$ are shown. Each member of the family corresponds to a different $\{a, b\}$ combination. The function $w(x)$ ensures that the spread during extrapolation grows (so the desired conservatism is attained) before decreasing towards zero (so the desired asymptotic behavior is attained). It is up to the analyst to shape $w(x)$, thus $I_{\text{belief-based}}(x)$ and $I_y(x)$, by searching for suitable values of a and b . As the figure shows, none of the resulting IPMs contain all the data (not that the analyst would know this in practice). Whereas the reliability upper bound for $I_{\text{data-based}}(x)$ and $I(x)$ holds for data in X_m , that is not the case during extrapolation. To the best of the authors' knowledge both of these extrapolation

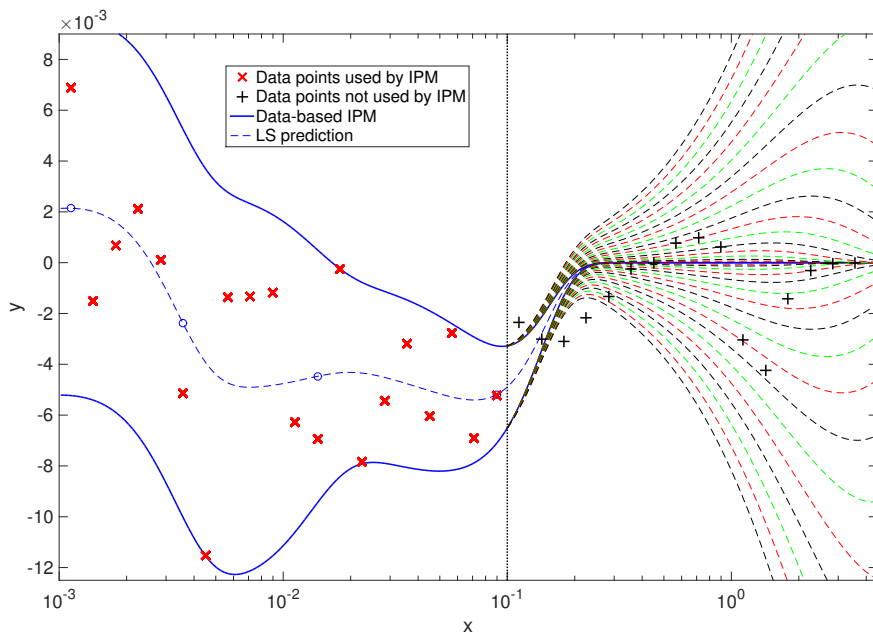


Figure 6. Extrapolating IPM.

features can not be attained by a GP model.

III. Application to Radiation Shielding

A component of the risk due to space exploration for astronauts originates from the natural space radiation environment. Space is filled with a high energy, low intensity background radiation environment called Galactic Cosmic Rays (GCR). GCR consist of fully ionized

nuclei, including all elements which are naturally occurring, traveling at near the speed of light. Exposure to GCR for extended periods of time can lead to increased risk for cancer and cataracts, along with other health problems [9]. Material shielding can reduce the radiation risk to astronauts but cannot meet current requirements for long duration missions [8, 20]. The number of safe days in space varies based on assumptions and models used, but calculations suggest that the Permissible Exposure Limits will be exceeded for a Mars mission using current technology [8]. The validation of these thick radiation shielding exposure calculations has been limited and better experimental data is needed to quantify the model uncertainties.

To address the lack of experimental data and the need for a better understanding of model uncertainty, NASA has funded a set of measurements and an associated project to quantify model uncertainty using those measurements. The uncertainty in the models compared to the laboratory experiments will be assessed and this uncertainty will then be extrapolated to the space environment. In the laboratory, a beam of ions is delivered to the shielding material at a given energy. The energy and its spread are well defined. The incident beam encounters a material target and is altered through interaction with the target. Detectors placed around the target material will measure the number, type, and energy of ions and particles produced through the interactions. Different beam ions, beam energies, target material types and thicknesses are the independent variables being tested. The differential spectral flux of the detected particles will then be compared to physics-based model predictions in order to generate uncertainty estimates for this type of model.

An initial set of simulated measurement results has been created, so that the effort to identify the uncertainty metrics could begin before the beam measurements are completed. To simulate the experiment and model results, the FLUKA model (FLUKA) [3, 10] was used. FLUKA is a fully integrated Monte Carlo simulation package for describing the interaction and transport of particles and nuclei in matter. FLUKA has many applications in particle physics, high energy experimental physics and engineering, shielding, detector and telescope design, cosmic ray studies, dosimetry, medical physics, radiobiology. In the context of this study, beams of ^{12}C , ^{14}N , and ^{28}Si ions at 500 MeV/n, 1000 MeV/n, and 1500 MeV/n kinetic energies incident on a 15 cm thick aluminum target were modeled and the differential flux of particles and ions produced were used to simulate the experimental data. For the simulated model results, beam ions with a slightly higher mass were used to ensure there was a difference between the simulated experiment and the model predictions while all other parameters were kept the same.

A. Example

Denote by E_i the energy of the ions of the incident beam and by E_o the energy of an ion of a particular type leaving the target. FLUKA is used to predict the flux of output ions within a prescribed sub-interval of energy E_o . These predictions are compared to experimental observations over the same sub-intervals to obtain a prediction error of the flux, ΔF . IPMs in which E_o is the input and ΔF is the output are calculated below. There is a different IPM for each combination of incident ion/particle type and energy, target thickness and type, and output ion/particle type. In the examples that follow we assume that the values of ΔF are scalar quantities.

The E_o domain from which measurements are taken is partitioned using a logarithmic

scale. As such, we will assume a probability distribution for x that makes every subinterval in the partition equally likely. This distribution is used to calculate the cost function in OP1 and OP2. Furthermore, we assume an expansion with $n_p = 10$ RBF terms, and use the minimum spread-, reduced-oscillation- and P -centering-constraints. All the IPMs presented below are data-based, so the developments in Section II.F are not used.

Figure 7 shows $N = 40$ data points corresponding to a carbon projectile with $E_i = 1\text{GeV/nucleon}$, an aluminum target, and a proton output particle. The vertical lines are the limits of the subintervals. The OP1 formulation for a suboptimal value for q leads to IPM A, whose boundaries are shown as black lines. As before, the LS prediction is showed as a dashed line whereas the RBF centers are shown as circles. Note that IPM A contains all the data and its limits pass through four data points, one at the upper limit and three at the lower limit. The OP2 formulation leads to IPM B, which is shown in blue. The improved tightness of IPM B relative to IPM A is apparent. This is reflected in the value of the cost, which is $E_x[\delta_y(x; q)] = 7.007 \times 10^{-3}$ for IPM A and $E_x[\delta_y(x; \hat{q})] = 4.855 \times 10^{-3}$ for IPM B. Note that the centers of IPM B are different from those of IPM A (used as initial condition in the search for \hat{q} in (11)), and the boundaries of IPM B pass through nine data points, five at the upper limit and four at the lower limit. The increased number of points on the limits of the IPM is an indication of the tightening of the prediction. Since both IPMs are based on the same number of observations, and on the same number of RBF terms, they admit the same reliability upper bound. For $N = 40$, $d = 10$, $\beta = 10^{-4}$, Theorem 1 leads to $\epsilon = 0.825$. Therefore, we expect future observations to fall outside the predicted range with a probability that won't exceed 0.825. This large value is a consequence of the small sample size. Note that whereas the same upper bound applies to both IPMs, the actual reliability of IPM B will likely be smaller than that of IPM A by virtue of its improved tightness. The closed-form expressions for the IPM boundaries, along with the reliability bound, enable rigorous uncertainty quantification for FLUKA.

Figure 8 shows IPMs corresponding to a carbon projectile for three values of E_i , an aluminum target, and a Helium output particle. In all three cases we use the same number of RBFs. The same line conventions used previously apply. Note the high sensitivity of the IPM to the value of E_i throughout the E_o range. This range is not the same for all three cases. Dashed vertical lines are used to denote the data point with largest E_o value. The IPMs corresponding to the red and blue lines are extrapolating beyond that value. This is highlighted by showing the IPM boundaries as dashed-dotted lines during extrapolation. As predicted by (9), the spread of IPM having a radial basis converges to zero as the separation between the test and the measured inputs increases. This is consistent with the nature of the physical quantity being predicted. Bases having terms of global influence, such as a polynomial basis, will not exhibit this behavior.

A two-dimensional IPM having the input $x = \{E_i, E_o\}$ can be readily calculated by combining the data used to construct the three one-dimensional IPMs, and using OP1 or OP2. The reliability of the resulting IPM can be bounded using Theorem 1. Alternatively, the three one-dimensional IPMs in Figure 8 can be combined to make predictions at other E_i values. This can be achieved by interpolating the set P and the parameter q based on the three points available. Any interpolation scheme that yields feasible values for these variables is viable, i.e., if z is the interpolating variable, it is required that $q(z) \in Q$ and $m(z) > 0$. Figure 9 shows the resulting IPM. As expected, all observations (not shown) are within the IPM boundaries, and sections at $E_i = 0.5 \text{ GeV/n}$, $E_i = 1\text{GeV/n}$, and $E_i = 1.5\text{GeV/n}$, are

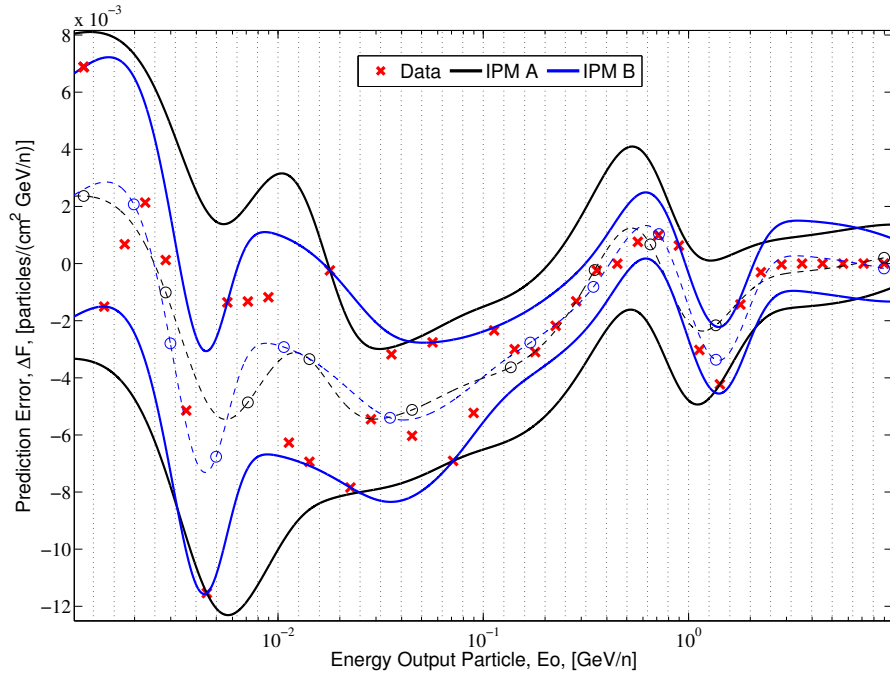


Figure 7. IPMs based on OP1 (black) and OP2 (blue) for $^{12}\text{C}(1 \text{ GeV/n}) + \text{Al} \rightarrow \text{Proton}$

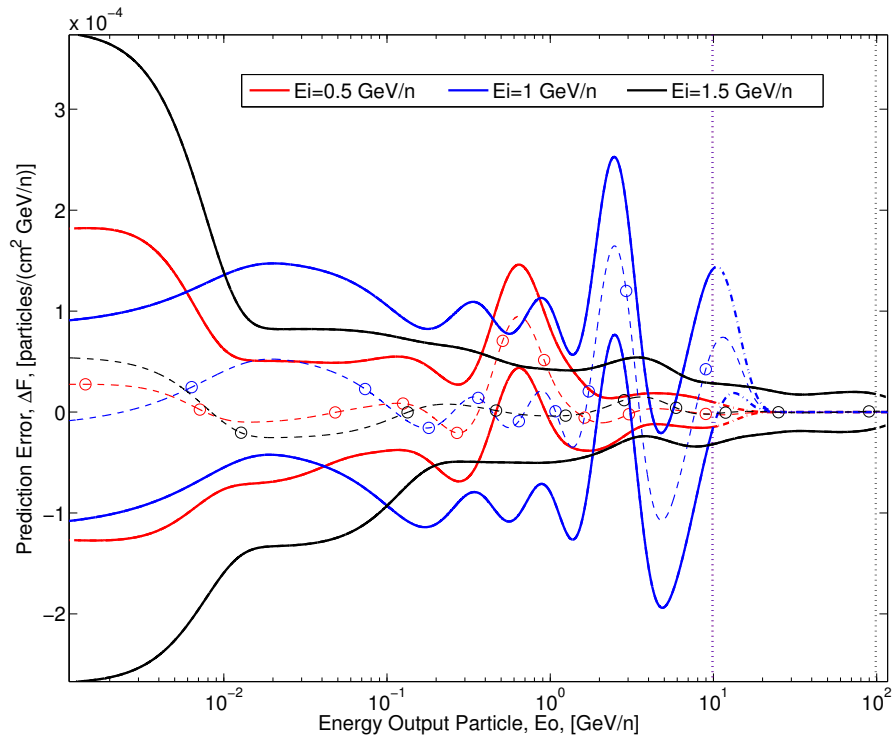


Figure 8. IPMs for $x = E_o$ and $^{12}\text{C}(E_i) + \text{Al} \rightarrow {}^3\text{He}(E_o)$, for three values of E_i .

equal to the IPMs in Figure 8. The reliability bound of each of the three one-dimensional

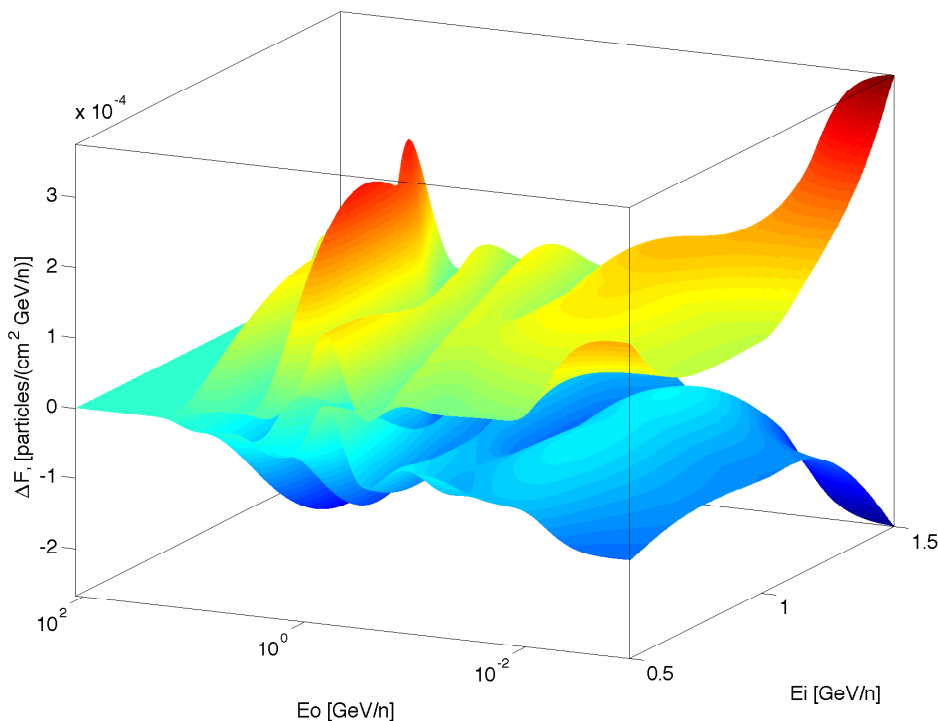


Figure 9. IPM for $x = \{E_i, E_o\}$ and $^{12}\text{C}(E_i) + \text{Al} \rightarrow ^3\text{He}(E_o)$

IPMs can be calculated using Theorem 1. However, that is not the case for any other value of E_i , nor for the two-dimensional IPM resulting from interpolation.

IV. Conclusions

This paper focuses on the construction and shaping of IPMs having a Gaussian radial basis. The local nature of this basis enables the accurate description of data generating mechanisms having an input-dependent spread with minimal modeling effort. This advantage, however, might yield predictions that overfit the data or for which the prediction is overly narrow. This is always the case during extrapolation. Means to counteract for this deficiency by constraining the search for an optimal IPM are developed herein. This meta-modeling technique is illustrated using a radiation shielding application. In this application we use IPMs to describe the error incurred in predicting the flux of particles resulting from the interaction between a high-energy incident beam and a target.

References

- ¹ J. P. Aubin and A. Cellina. *Differential Inclusions*. Springer-Verlag, Berlin, Germany, 1984.

- ² J. P. Aubin, J. Lygeros, M. Quincampoix, S. Sastry, and N. Seube. Impulse differential inclusions: a viability approach to hybrid systems. *IEEE Transactions on Automatic Control*, 47(1):2–20, 2002.
- ³ G. Battistoni, S. Muraro, P. R. Sala, F. Cerutti, A. Ferrari, S. Roesler, A. Fasso, and J. Ranft. The fluka code: Description and benchmarking. In *Proceedings of the Hadronic Shower Simulation Workshop, AIP Conference Proceeding 896*, pages 31–49, Fermilab 6-8 September 2006.
- ⁴ D. A. Brown, M. Herman, S. Hoblit, E. A. McCutchan, G. P. Nobre, B. Pritychenko, and A. A. Sonzogni. Uncertainty quantification in the nuclear data program. *Journal of Physics G: nuclear and particle physics*, 42(034020):1–26, 2015.
- ⁵ M. C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.
- ⁶ 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*, pages 1–20, 2016. Accepted.
- ⁷ L. G. Crespo, S. P. Kenny, D. P. Giesy, R. B. Norman, and Steve R. Blattnig. A comparison of metamodeling techniques using numerical experiments. In *AIAA Scitech 2016*, pages 1–35, San Diego, CA, USA, January 2016. Accepted.
- ⁸ FA Cucinotta, M-HY Kim, LJ Chappell, and JL Huff. How safe is safe enough? radiation risk for a human mission to mars. *PLoS ONE*, 8:e74988, 2013.
- ⁹ Marco Durante and Francis A. Cucinotta. Physical basis of radiation protection in space travel. *Rev. Mod. Phys.*, 83:1245–1281, Nov 2011.
- ¹⁰ A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft. FLUKA: a multi-particle transport code. Technical Report CERN-2005-10, CERN, 2005.
- ¹¹ M.C. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society B*, 63(3):425–464, 2001.
- ¹² A. E. Lovell and F. M. Nunes. Systematic uncertainties in direct reaction theories. *Journal of Physics G: nuclear and particle physics*, 42(034014):1–13, 2015.
- ¹³ J. D. McDonell, N. Schunck, D. Hidgon, J. Sarich, S. M. Wild, and W. Nazarewicz. Uncertainty quantification for nuclear density functional theory and information content of new measurements. *Physical review letters*, 114(1):122501 1–6, 2015.
- ¹⁴ M. Milanese and C. Novara. Set-membership identification of nonlinear systems. *Automatica*, 40(6):957–975, 2004.
- ¹⁵ M. Milanese and C. Novara. Set-membership prediction of nonlinear time systems. *IEEE Transactions on Automatic Control*, 50(11):1655–1669, 2005.

- ¹⁶ J. Piekarewicz, Wei-Chia Chen, and F. J. Fattoyev. Information and statistics: a new paradigm in theoretical nuclear physics. *Journal of Physics G: nuclear and particle physics*, 42(034018):1–22, 2015.
- ¹⁷ Carl. E. Rasmussen and Christopher K. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- ¹⁸ G. A. Seber and C. J. Wild. *Nonlinear Regression*. JohnWiley & Sons, Hoboken, New Jersey, USA, 2003.
- ¹⁹ T. Simpson, J. Peplinski, P. Koch, and J. Allen. Metamodels for computer-based engineering design: survey and recommendations. *Engineering with Computers*, 17(1):129–150, 2001.
- ²⁰ T. Slaba, C. Mertens, and S. Blattnig. Radiation shielding optimization on mars. *NASA Technical Paper 217965*, 2013.