Interpolating between small- and large-g expansions using Bayesian model mixing

A. C. Semposki [1,*] R. J. Furnstahl [2,†] and D. R. Phillips [1,‡]

[1]*Department of Physics and Astronomy and Institute of Nuclear and Particle Physics, Ohio University, Athens, Ohio 45701, USA*

[2]*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*

(Received 14 June 2022; accepted 13 September 2022; published 20 October 2022)

Bayesian model mixing (BMM) is a statistical technique that can be used to combine models that are predictive in different input domains into a composite distribution that has improved predictive power over the entire input space. We explore the application of BMM to the mixing of two expansions of a function of a coupling constant g that are valid at small and large values of g respectively. This type of problem is quite common in nuclear physics, where physical properties are straightforwardly calculable in strong and weak interaction limits or at low and high densities or momentum transfers, but difficult to calculate in between. Interpolation between these limits is often accomplished by a suitable interpolating function, e.g., Padé approximants, but it is then unclear how to quantify the uncertainty of the interpolant. We address this problem in the simple context of the partition function of zero-dimensional ϕ^4 theory, for which the (asymptotic) expansion at small g and the (convergent) expansion at large g are both known. We consider three mixing methods: linear mixture BMM, localized bivariate BMM, and localized multivariate BMM with Gaussian processes. We find that employing a Gaussian process in the intermediate region between the two predictive models leads to the best results of the three methods. The methods and validation strategies we present here should be generalizable to other nuclear physics settings.

DOI: 10.1103/PhysRevC.106.044002

I. MOTIVATION AND GOALS

In complex systems, such as nuclei, there is typically more than one theoretical model that purports to describe a physical phenomenon. Among these models may be candidates that work well in disparate domains of the input space: energy, scattering angle, density, etc. In such a situation it is not sensible to select a single "best" model. How, instead, can we formulate a combined model that will be optimally predictive across the entire space?

A concrete realization of this situation is the combination of predictions from expansions about limiting cases, which provide theoretical control near those limits. Nuclear physics examples include low- and high-density limits (e.g., quark-hadron phase transitions in dense nuclear matter [1–4]); strong and weak interaction limits (e.g., expansions in positive and negative powers of the Fermi momentum times scattering length for cold atoms [5]); high- and low-temperature limits (e.g., Fermi-Dirac integrals [6]); and four-momentum transfer limits (e.g., the generalized GDH sum rule [7]).

We propose using Bayesian statistical methods, and the technique of Bayesian model mixing (BMM) in particular, to interpolate between such expansions. Bayesian methods are well suited to this problem since they enable the inclusion of prior information and provide principled uncertainty quantification (UQ). For a brief introduction to the basics of Bayesian statistics, see Appendix A.

In this work, we explore BMM strategies for interpolant construction in the case of the partition function of zero-dimensional ϕ^4 theory as a function of the coupling strength g [see Eq. (3)]. These series expansions—in g and $1/g$ respectively—provide excellent accuracy in their respective regions but, depending on the level of truncation, can exhibit a considerable gap for intermediate g where they both diverge (see Fig. 1).

A variety of resummation methods have been used to extend the domain of convergence of the small-g expansion of the zero-dimensional partition function of ϕ^4 theory, although only a few have incorporated the constraints from the large-g expansion; Ref. [5] discusses some of the possibilities. In Ref. [8], this zero-dimensional partition function was used to test the suitability of a generalized form of Padé approximants as an interpolant between the small-g and large-g expansions. The resulting "fractional powers of rational functions" perform well in that case, but there is no obvious principle that guides the choice of the "best" rational power and, relatedly, without access to the true partition function, the accuracy of different interpolants is not accessible in the region of the gap. The same criticism applies to the constrained extrapolations with order-dependent mappings introduced in Ref. [9].

Here we will apply Bayesian methods to obtain an interpolant that agrees with the true model not only at both ends of the input space where the power series are reliable, but also in that gap. A key objective is that the interpolant should come with a robust estimate of the uncertainty. A hallmark of a Bayesian approach is the use of probability distributions for

*as727414@ohio.edu
†furnstahl.1@osu.edu
‡phillid1@ohio.edu

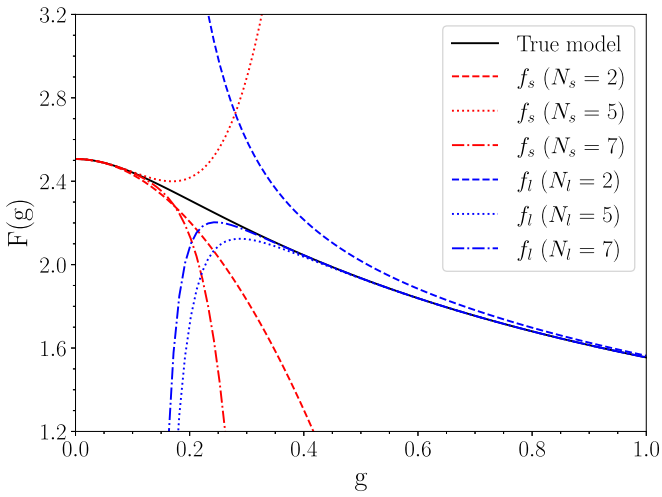

FIG. 1. The partition function in Eq. (3), denoted the true model, overlaid with two series expansions (red denoting the weak coupling expansion and blue the strong coupling expansion) truncated at various orders, here labeled N_s and N_l [see Eq. (4)].

(almost) every component of the problem. Particularly useful here will be the posterior predictive distribution, or PPD, which is the distribution of possible values that a function can take at a point where we do not have data. We will label the K models under consideration by $\mathcal{M}_k$ $(k = 1, \ldots, K)$. The kth model is specified by what it predicts for an observation y_i at a point x_i:

$$\mathcal{M}_k : y_i = f_k(x_i) + \varepsilon_{i,k}. \tag{1}$$

In our case the function f_k is a deterministic physics model depending on a single input x_i, or a Gaussian process used to improve the interpolation between models. The physics model will usually also depend on parameters θ_k, and those parameters would then have to be estimated by calibration to some observed data set $\mathbf{D}$, consisting of pairs $(x_1, y_1), \ldots, (x_n, y_n)$, but that is a complication we do not consider here.[1] The error term, $\varepsilon_{i,k}$, represents all uncertainties (systematic, statistical, computational) which may also enter at the point x_i. Since $\varepsilon_{i,k}$ is a random variable, its probability distribution has to be specified in order to specify the statistical model for the relationship between x_i and y_i. In our prototype, the f_k are (different) expansions of the integral in Eq. (3), with the x_i's a set of coupling strengths g, and $\varepsilon_{i,k}$ modeled from the truncation error at x_i for f_k.

The PPD for a new observation $(\tilde{x}, \tilde{y})$ for model k is denoted $p(\tilde{y}|\tilde{x}, \mathcal{M}_k)$. (Dependence on parameters θ_k, if present, would be integrated out—"marginalized"—here.) Our goal is a combined PPD, which is a weighted sum of the PPD for each model:

$$p(\tilde{y}|\tilde{x}) = \sum_{k=1}^{K} \hat{w}_k p(\tilde{y}|\tilde{x}, \mathcal{M}_k). \tag{2}$$

[1]We note that the BAND [10] software package `surmise` is capable of emulating and calibrating models and can be employed prior to using SAMBA to mix the models [11].

Conventional Bayesian model averaging (BMA) [12] and Bayesian stacking [13] are implementations of BMM for which the $\hat{w}_k$ are independent of x_i, but in general the weights $\hat{w}_k$ may depend on $\tilde{x}$ [14,15]. Such dependence clearly has to be present for expansions that are valid in different input domains.

The optimal ways to determine location-dependent weights in BMM for various applications is a topic of current research. The Bayesian Analysis of Nuclear Dynamics (BAND) project will produce a general software Framework for Bayesian analysis, with nuclear physics examples, and has a particular interest in the development and application of BMM methodologies. A general description of the BAND goals and BMM is given in the BAND Manifesto [16].

The present work develops a theoretical laboratory ("sandbox") that uses the example of the partition function in zero-dimensional ϕ^4 theory to explore some possible approaches to BMM that produce reliable interpolants with principled UQ. We begin this enterprise in Sec. II, where we present in detail the prototype case we have chosen, and provide a concise description of Bayesian methods in the context of our problem. Following this, in Sec. III, we discuss BMM based on linear mixtures of individual models' PPDs and present results of such an approach. In Sec. IV we apply an alternative method, that we dub "localized bivariate mixing," which uses the series expansions and their variances to produce a precision-weighted combination of the small-g and large-g expansions. As an improvement upon this method, we introduce a Gaussian process as a third model, and present results for "localized multivariate mixing" with that Gaussian process in Sec. V. A summary of lessons learned and the outlook for future applications of these techniques, as well as a short discussion of our publicly available SAndbox for Mixing using Bayesian Analysis (SAMBA) computational package, are contained in Sec. VI.

II. FORMALISM

A. Test problem

Our prototype true model is a zero-dimensional ϕ^4 theory partition function [8]

$$F(g) = \int_{-\infty}^{\infty} dx\, e^{-\frac{x^2}{2} - g^2 x^4} = \frac{e^{\frac{1}{32g^2}}}{2\sqrt{2}g} K_{\frac{1}{4}}\left(\frac{1}{32g^2}\right), \tag{3}$$

with g the coupling constant of the theory, and $K_{\frac{1}{4}}$ the modified Bessel function of the second kind. This function can be expanded in either g or $1/g$ to obtain

$$F_s^{N_s}(g) = \sum_{k=0}^{N_s} s_k g^k, \quad F_l^{N_l}(g) = \frac{1}{\sqrt{g}} \sum_{k=0}^{N_l} l_k g^{-k}, \tag{4}$$

with coefficients

$$s_{2k} = \frac{\sqrt{2}\Gamma(2k + 1/2)}{k!}(-4)^k, \quad s_{2k+1} = 0, \tag{5}$$

and

$$l_k = \frac{\Gamma\left(\frac{k}{2} + \frac{1}{4}\right)}{2k!}\left(-\frac{1}{2}\right)^k. \tag{6}$$

The weak coupling expansion, $F_s^{N_s}(g)$, is an asymptotic series, while the strong coupling expansion, $F_l^{N_l}(g)$, is a convergent series. The former is a consequence of the factorial growth of the coefficients s_{2k} at large k; as for the latter, the coefficients l_k decrease factorially as $k \to \infty$.

To apply Bayesian methods to the construction of an interpolant that is based on these small- and large-coupling expansions it is necessary to think of each expansion as a probability distribution. We follow Refs. [17–19] and assign Gaussian probability distributions to the error made by truncating each series at a finite order. That is, we define the PPD for the small-g expansion at order N_s as

$$\mathcal{M}_{N_s} : F(g) \sim \mathcal{N}\big(F_s^{N_s}(g), \sigma_{N_s}^2(g)\big), \tag{7}$$

and that for the large-g expansion at order N_l as

$$\mathcal{M}_{N_l} : F(g) \sim \mathcal{N}\big(F_l^{N_l}(g), \sigma_{N_l}^2(g)\big). \tag{8}$$

Note that by assuming that the mode of the PPD is the value obtained at the truncation order we are assuming that the final result is equally likely to be above or below that truncated result.

B. Truncation error model

This leaves us with the task of defining the "1σ error," i.e., the standard deviation, associated with each PPD. The error in $F_s^{N_s}(g)$ is nominally of $O(g^{N_s+2})$ (note that the series contains only even powers of g) but the factorial behavior of the coefficients modifies the size of the error dramatically. We therefore build information on the asymptotics of the coefficient s_{2N_s+2} into our model for that error. Analyzing the first few terms of the series (see Appendix B) we might guess

$$\sigma_{N_s}(g) = \begin{cases} \Gamma(N_s+3)g^{N_s+2}\bar{c} & \text{if } N_s \text{ is even,} \\ \Gamma(N_s+2)g^{N_s+1}\bar{c} & \text{if } N_s \text{ is odd.} \end{cases} \tag{9}$$

The standard deviation takes this alternating form because the small-g series contains only even terms. $\bar{c}$ is a number of order one that is estimated by taking the (nonzero) terms in the expansion up to order N_s, rescaling them according to Eq. (9) so that the large-N_s behavior is accounted for, and computing their rms value. We deem Eq. (9) an "uninformative error model" for large N_s.

A closer look at the behavior of the series as N_s gets bigger leads to a more informed statistical model for the uncertainty at order N_s:

$$\sigma_{N_s}(g) = \begin{cases} \Gamma(N_s/2+1)(4g)^{N_s+2}\bar{c} & \text{if } N_s \text{ is even,} \\ \Gamma(N_s/2+1/2)(4g)^{N_s+1}\bar{c} & \text{if } N_s \text{ is odd.} \end{cases} \tag{10}$$

$\bar{c}$ is again estimated by taking the nonzero terms in the expansion, this time rescaling them according to Eq. (10), and computing their rms value. Further discussion of how we arrived at these two forms for the behavior of higher-order terms can be found in Appendix B.

Similarly the error in $F_l^{N_l}(g)$ is nominally of $O(g^{-N_l-3/2})$ [recall that $F(g)$ goes as $g^{-1/2}$ at large g] but we must first account for other factors that dramatically affect the convergence of the series. We therefore take

$$\sigma_{N_l}(g) = \frac{1}{\Gamma(N_l+2)}\frac{1}{g^{N_l+3/2}}\bar{d} \tag{11}$$

as an uninformative model of the error at large N_l and

$$\sigma_{N_l}(g) = \left(\frac{1}{4g}\right)^{N_l+3/2}\frac{1}{\Gamma(N_l/2+3/2)}\bar{d}, \tag{12}$$

as the informative model. $\bar{d}$ is to be estimated using the same approach taken for $\bar{c}$.

It is important to note that all sets of coefficients discussed are alternating, so an error model with mean zero is appropriate. However, the error model only builds in some aspects of the series' behavior: it assumes that the higher-order coefficients are randomly distributed around zero, which does not account for their alternating-sign pattern. Some consequences of this choice will later be manifest in Fig. 6.

III. LINEAR MIXTURE MODEL

A. Formulation

In some situations the true probability distribution is a linear mixture of two simpler probability distributions. For example, the height of adult males and the height of adult females[2] in the US are each approximately normally distributed [20], but the two probability distributions have different means and variances, and so the distribution of adult heights in the US is (at the same level of approximation) a linear mixture of the two:

$$p(\text{height of US adults}) = \alpha\, p(\text{height of US males}) + (1-\alpha)\, p(\text{height of US females}), \tag{13}$$

where α is the probability of drawing a male from the population.

This kind of linear mixture of probabilities defines a conceptually simple form of BMM; cf. Eq. (2). If the weights in Eq. (2) are chosen to be independent of the location in the input space and equal to the Bayesian evidence [the denominator in Eq. (A1)], what results is an example of BMA.

In this section we follow Coleman [21] and make the weights in Eq. (2) dependent on location in the input-space variable, g. Reference [21] applied this technique to mix two different models of heavy-ion collisions developed by the JETSCAPE Collaboration (although there the mixing protocol was also modified to account for correlations between observables across the input space).

We apply it to large-g and small-g models, $\mathcal{M}_{N_l}$ and $\mathcal{M}_{N_s}$. The PPD computed in this section therefore takes the form

$$p(F(g)|\boldsymbol{\theta}) \sim \alpha(g;\boldsymbol{\theta})F_s^{N_s}(g) + (1-\alpha(g;\boldsymbol{\theta}))F_l^{N_l}(g). \tag{14}$$

We take the variance of the two individual-model probability distributions to be the form of the informative error model [see Eqs. (10) and (12)]. The mixing function $\alpha(g;\boldsymbol{\theta})$ depends not just on the location in the input space, g, but also on parameters $\boldsymbol{\theta}$ defining that dependence on g. These parameters dictate how the mixing function "switches" from being 1 at $g = 0$, where $F_s^{N_s}$ dominates, to being 0 at $g = 1$, where $F_l^{N_l}$ dominates.

[2]"Male" and "female" here refer to the sex assigned at birth.

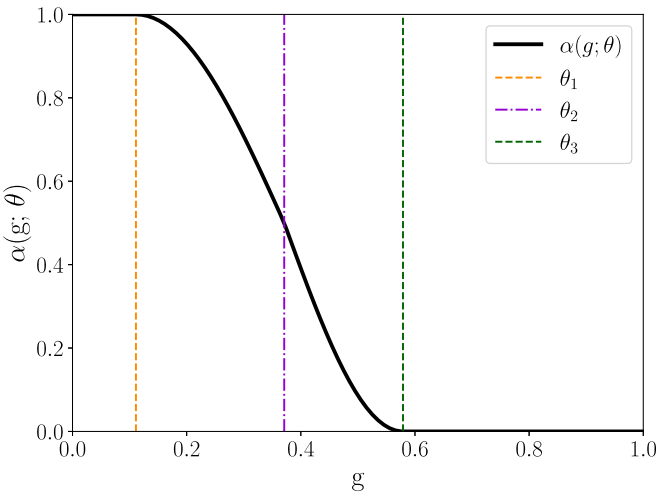

FIG. 2. The piecewise cosine mixing function, evaluated at the MAP values of $\boldsymbol{\theta}$, shown as the solid black curve, for the case $N_s = N_l = 2$. The dashed green, dashed orange, and dash-dotted purple lines represent the MAP values of the three parameters θ_1, θ_2, and θ_3, which give the location of transition points in the function [see Eq. (15)].

We employ parametric mixing, which means we must specify the functional form of α. In Ref. [21] Coleman employed the well-known logistic function (which is also referred to as a sigmoid function, and has been used as a simple activation function in recent machine learning applications) and the CDF (cumulative distribution function) of the normal distribution. However, these do not meet the requirements of our application. We require $\alpha(g;\boldsymbol{\theta})$ become strictly 0 before the value of g where the small-g power series diverges. Likewise, $\alpha(g;\boldsymbol{\theta})$ must remain constant at 1 for all g for which the large-g expansion diverges. We therefore adopted a third mixing function, henceforth referred to as the cosine mixing function, which is constructed piecewise so that it switches the large-g model on at a value θ_1 that is "late enough," and switches the small-g model off at a value θ_3 that is "early enough." Figure 2 provides a pictorial example of our function and its success in this constraint. The function is

$$\alpha(g;\boldsymbol{\theta}) = \begin{cases} 1, & g \leqslant \theta_1; \\ \frac{1}{2}\left[1+\cos\left(\frac{\pi}{2}\left(\frac{g-\theta_1}{\theta_2-\theta_1}\right)\right)\right], & \theta_1 < g \leqslant \theta_2; \\ \frac{1}{2}\left[1+\cos\left(\frac{\pi}{2}\left(1+\frac{g-\theta_2}{\theta_3-\theta_2}\right)\right)\right], & \theta_2 < g \leqslant \theta_3; \\ 0, & g > \theta_3. \end{cases} \tag{15}$$

The parameter θ_2 specifies the location in input space, i.e., the value of g, at which each model has a weight of 0.5 in the final PPD.

The mixing function thus has three parameters $\boldsymbol{\theta}$ that need to be estimated from data. Bayes' theorem, Eq. (A1), tells us that the posterior for the parameters is the product of the likelihood of that data and the prior distribution for the parameters. If we assume that we have $i = 1, \ldots, n$ evenly spaced data points in the gap and the error on these is $\sigma_{d_i}^2$, then we can use Bayes' theorem and the mixed-model definition [Eq. (14)] to write

$$p(\boldsymbol{\theta}|\mathbf{D}) = p(\boldsymbol{\theta}) \prod_{i=1}^{n} \left\{ \alpha(g_i;\boldsymbol{\theta}) \mathcal{N}\left(F_s^{N_s}(g_i), \sigma_{d_i}^2 + \sigma_{N_s}^2\right) + [1-\alpha(g_i;\boldsymbol{\theta})] \mathcal{N}\left(F_l^{N_l}(g_i), \sigma_{d_i}^2 + \sigma_{N_l}^2\right) \right\}. \tag{16}$$

Note that $\sigma_{N_s}^2$ and $\sigma_{N_l}^2$ correspond to the theory error model chosen (here the informative model). The prior distribution, $p(\boldsymbol{\theta})$, is defined as

$$\begin{aligned} p(\boldsymbol{\theta}) = {} & \mathcal{U}(\theta_1 \in [0, b]) \mathcal{N}(\theta_1; \mu_1, 0.05^2) \\ & \times \mathcal{U}(\theta_2 \in [\theta_1, b]) \mathcal{N}(\theta_2; \mu_2, 0.05^2) \\ & \times \mathcal{U}(\theta_3 \in [\theta_2, b]) \mathcal{N}(\theta_3; \mu_3, 0.05^2). \end{aligned} \tag{17}$$

This prior ensures that $\theta_1 < \theta_2 < \theta_3$. The value b is a cutoff that can be selected to indicate where we believe the end of the mixing region between the models is (heretofore and henceforth deemed "the gap"). This can be set to separate values in each uniform prior, depending on the user's knowledge of their system, or to the same value for all parameter priors. Note that this can indeed be part of the prior in this analysis, because we have access to information on where the gap between the expansions is based solely on the expansions themselves. Also, note that the values for the mean of the Gaussian priors [μ_1, μ_2, and μ_3 in Eq. (17) above] can be assigned depending on the estimated location of the gap.

We determine [Eq. (16)], and hence the posteriors of the three parameters $\boldsymbol{\theta}$, by employing Markov chain Monte Carlo (MCMC) sampling, implemented via the Python package `emcee` [22]. We assigned 20 walkers with 3000 steps each, for a total of 60 000 steps. We selected the first 200 steps of each walker as burn-in and removed them from consideration in the parameter chains.

We can write the PPD [Eq. (14)] as

$$p(\tilde{y}(g)|\boldsymbol{\theta}, \mathbf{D}) = \sum_{j=1}^{M} \alpha(g;\boldsymbol{\theta_j}) F_s^{N_s}(g) + (1-\alpha(g;\boldsymbol{\theta_j})) F_l^{N_l}(g), \tag{18}$$

where we sum over M values of the $\boldsymbol{\theta}$ vector from trace results in our MCMC samples. In this way we are not just including a maximum *a posteriori* (MAP) value of our estimated parameters, but using all of the posterior distribution information we gained from our MCMC calculation to determine them.

B. Results and discussion

Before discussing the results for the linear mixture model we point out that the need for data in the gap between these two expansions to calibrate the parameters of the mixing function could be a concern for applications to physical problems where obtaining these data would be costly, difficult, or both.

Even if data are available—as we have assumed they are—two obvious pitfalls of the linear mixture model in our test case can be seen in Figs. 3 ($N_s = N_l = 2$) and 4 ($N_s = N_l = 5$). The PPD mean connects to the two series expansions across the gap, but in both cases the discontinuities in the first derivatives of the mixed model at the left and right edges of the gap would not easily be explained by natural occurrence, unless

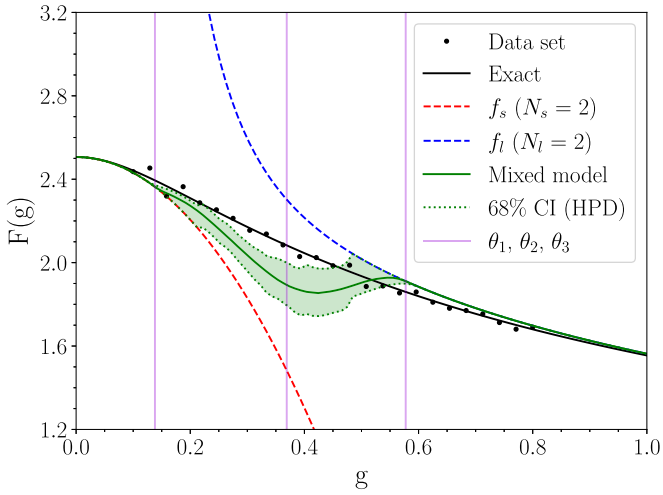

FIG. 3. The PPD median and 68% credibility interval for the linear mixture model [Eq. (18)] are shown in solid green and as a shaded green band, respectively, overlaid on the $N_s = 2$ (dashed red curve) and $N_l = 2$ (dashed blue curve) small-g and large-g expansions. The location in g of the three mixture parameters of the piecewise cosine mixture function, θ_1, θ_2, and θ_3, are indicated by the solid light purple lines. The true model curve is shown in solid black as a reference, with the black dots around it comprising the randomly distributed data set used in this analysis. This data has a 1% error from the true curve added to it.

one crossed a phase transition or other new physics entered at a specific value of g.

Second, even though the theory error of the small- and large-g expansions (in the form of the informative error model used in this analysis) is accounted for in our analysis, we see that in Fig. 3 the 68% credibility interval does not include the true model 68% of the time, indicating that this method is not accurately quantifying uncertainty in this toy problem. We did find cases where this linear mixture model produces a PPD with good empirical coverage properties, i.e., does not suffer from this defect, but we were only able to achieve that

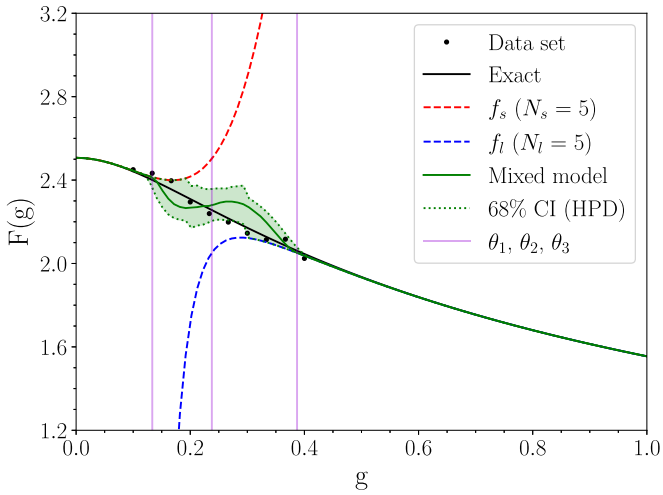

FIG. 4. The PPD median and 68% credibility interval, with all of the curves as for Fig. 3, but for the case $N_s = N_l = 5$.

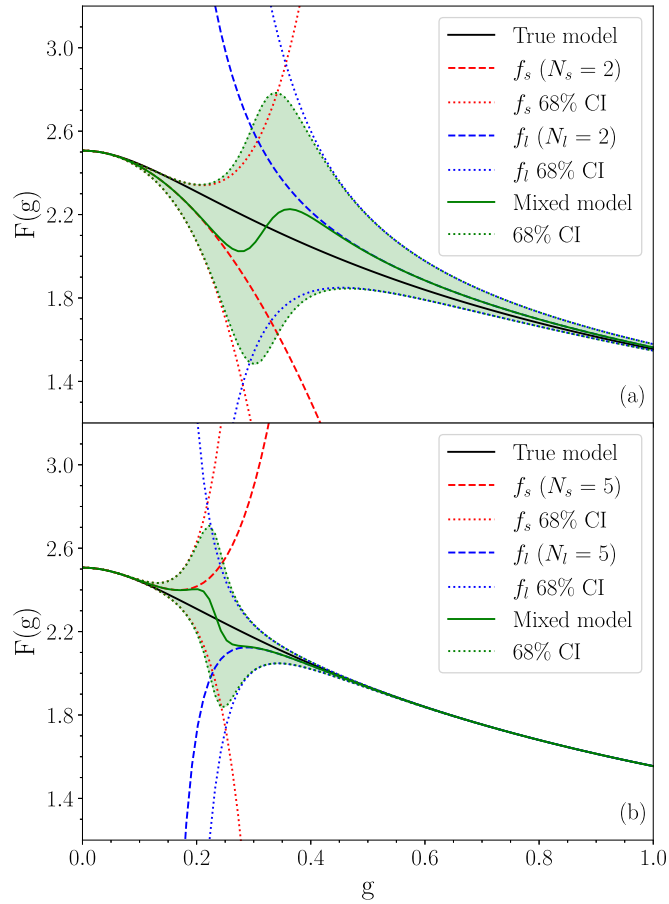

FIG. 5. The results of computing the bivariate model mixing are shown for (a) $N_s = 2$, $N_l = 2$ and (b) $N_s = 5$, $N_l = 5$, using the informative error model. The result from computing $f_\dagger$ [Eq. (20)] is shown in green, with the 68% credibility interval shown as the green shaded region.

for $N_s, N_l = 5$ (as in Fig. 4) as well as for some values that are much higher than would be attainable in a physics application.

There is a separate, but perhaps related, issue in choosing which mixture function we use by hand. If we are not choosing one that really represents the transition from one model's domain to the next, we are building in an implicit bias to the mixed model result, as the model will always adopt features of the mixing function chosen. In the next section, we construct another method that addresses this problem directly by eliminating the need for a mixing function to cross the gap.

IV. LOCALIZED BIVARIATE BAYESIAN MODEL MIXING

A. Formalism

One way to formulate a mixed model that incorporates the model uncertainty of the models being mixed is to use the simple method of combining Gaussian distributed random variables [16]. This can be written for any number of models K. Each model is weighted by its *precision*, the inverse of the variance:

$$f_\dagger = \frac{1}{Z_P}\sum_{k=1}^{K}\frac{1}{v_k}f_k, \quad Z_P \equiv \sum_{k=1}^{K}\frac{1}{v_k}. \tag{19}$$

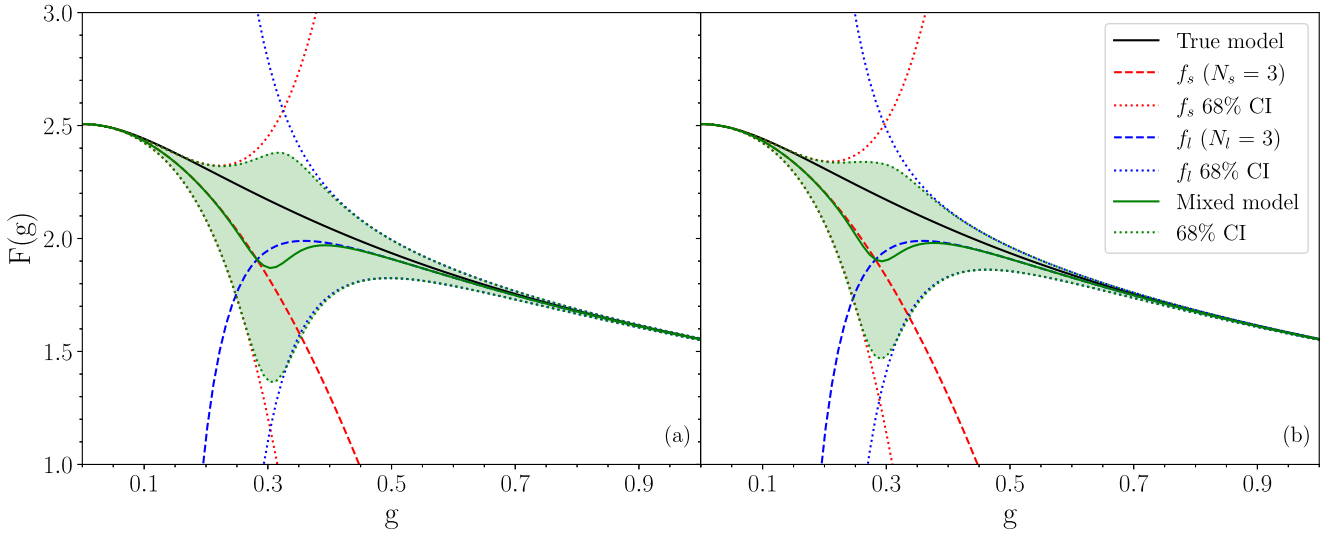

FIG. 6. A comparison between the uninformative error model (left panel) and the informative error model (right panel), shown side by side applied to the case $N_s = 3$, $N_l = 3$. The greater size of the uncertainty in the gap when the uninformative error model is used is evident in the left panel.

From the rule for a linear combination of independent normally distributed random variables $X_k \sim \mathcal{N}(\mu_k, v_k)$,

$$\sum_k a_k X_k \sim \mathcal{N}\left(\sum_k a_k \mu_k, \sum_k a_k^2 v_k\right), \tag{20}$$

the distribution of $f_\dagger$ is

$$f_\dagger \sim \mathcal{N}\left(Z_P^{-1} \sum_k \frac{1}{v_k} f_k, Z_P^{-1}\right). \tag{21}$$

The variance v_k is determined by the error models previously discussed in Sec. II B, thereby incorporating the theory uncertainty into this mixing method. This section focuses on the bivariate case of Eq. (21), where model 1 and model 2 are $F_s^{N_s}(g)$ and $F_l^{N_l}(g)$, respectively. We employ the so-called informative error model and several values of N_s and N_l (for more examples of N_s and N_l, see the Supplemental Material for this work [23]).

B. Results and discussion

The results of this method for two different cases of N_s, N_l are presented in Fig. 5. The bivariate mixed model completes the journey across the gap for both the lower orders [panel (a)] and the higher orders [panel (b)] in a continuous fashion, as desired. In both cases the true curve is also encompassed by the 68% interval of the PPD obtained through bivariate model mixing.

However, there are two negative aspects of the result of Fig. 5 which are quite striking. First, the shape of the mean of the PPD is not in line with our intuition for this problem. Second, the 68% interval is, if anything, too wide. Considering that F only varies between 2.5 and 1.8 between $g = 0$ and $g = 0.6$ an error bar of order 0.3, as is obtained even for $N_s = N_l = 5$, seems rather conservative.

In fact, as was discussed in Sec. IV A, the right variances to use in Eq. (20) to mix the two power series are nontrivial to determine. This adds to the challenges of this technique. There is a significant effect of the form of the variance on the results of the bivariate mixed model: the credibility interval changes by a noticeable amount depending on whether we employ the so-called uninformative or the informative formulas (see Fig. 6 and Sec. II B). Hence, there is a substantial amount of uncertainty—at least in the 68% interval of the interpolant's PPD—from not knowing the correct form of the truncation error. This affects the final result obtained from this two-model mixing in a way that cannot be ignored.

In fact, these "failures" all stem from the same root cause: the use of only the small-g and large-g power series of F to form the interpolant. If this is truly all the information one has then this is all one can say [24]. Given the errors on the small- and large-g expansions, behavior encompassed by the green band is possible. And nothing we have put into our bivariate model-mixing formulation precludes a sinusodial variation of the type seen in both panels of Fig. 5.

But such curvature of the mixed model mean is not a feature that is supported by any prior knowledge of our particular system. The intuitive expectation would be to see a smooth, linear path from one model to the other, and that is not being shown in our result from this simple two-model method. In the next section we incorporate this expectation by including in our mixing a nonparametric form for the function F in the gap between the two expansions.

V. BRIDGING THE GAP: LOCALIZED MULTIVARIATE BAYESIAN MODEL MIXING WITH GAUSSIAN PROCESSES

A. Formalism

To improve upon Sec. IV, we introduce a third model to interpolate across the gap in g in the form of a Gaussian process. A Gaussian process (GP) is fundamentally stochastic:

TABLE I. Results of the squared Mahalanobis distance for the Matérn 3/2 kernel and RBF kernel for all three training methods discussed. The mean of each training method per kernel is also calculated. From these data, it is evident that the Matérn 3/2 kernel is the most suitable for our problem.

		D^2_{MD} (Matérn 3/2) Training method			D^2_{MD} (RBF) Training method		
N_s	N_l	1	2	3	1	2	3
2	2	1.6	1.4	1.6	60	70	60
2	4	1.1	0.6	5.0	140	170	1500
3	3	0.5	0.3	3.0	65	110	7
4	4	1.8	4.3	5.6	66	29000	600
5	5	2.7	1.2	1.2	350	15	15
8	7	1.9	0.4	0.8	4200	210000	14
5	10	5.1	3.4	6.0	1100	140	2100
Mean		2.1	1.7	3.3	850	34000	610

a collection of random variables, any subset of which will possess a multivariate Gaussian distribution [19]. We can describe a GP with its mean function and positive semidefinite covariance function, or *kernel*. The kernel should reflect the physical and mathematical characteristics of our system. In this work, we focused on two of the many possible kernel forms: the RBF (radial basis function) kernel, and the Matérn 3/2 kernel. The former is described by

$$k(x, x') = \exp\left(-\frac{(x - x')^2}{2l^2}\right), \tag{22}$$

while the latter is given by

$$k(x, x') = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l}(x - x')\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}}{l}(x - x')\right), \tag{23}$$

where Γ is the gamma function, and K_ν is the modified Bessel function of the second kind. We employed the package `scikit learn` to compute our GP within the wrapper of SAMBA [25]. General draws from both of these kernels can be seen in Sec. 1.7.5 of the `scikit learn` documentation [26]. We chose the Matérn 3/2 kernel specifically, hence $\nu = 3/2$ in Eq. (23).

Employing an emulator typically involves three steps: setting up a kernel, choosing and training the emulator on a data set, and predicting at new points using the trained GP. Our toy model is trained using four points from training arrays generated across the input space. The four points chosen are determined by

(i) Generating an array of training points from both expansions.
(ii) Selecting two fixed points from the small-g expansion training array and the last fixed point in the large-g expansion training array.
(iii) Determining the fourth and final training point from the large-g expansion training array using what we will call a training method.

We also include the truncation error on each point of the training set. This is done as symmetric point-to-point uncertainty, fixed using the informative error model.

In our case, we applied three different training methods in the final step. Their results can be found in Table I. The first (referred to as method 1) fixes the fourth training point at $g = 0.6$ for every value of N_s and N_l investigated. The second (method 2) allows the training point to shift given the values of $F_s(g)$ and $F_l(g)$, locating the final training point where the value of $F_s(g)$ becomes much larger than the range of output values in this problem. The third (method 3) employs the theory error on each training array point. This method fixes the fourth training point at the location in the data set where the value of the theory error at that point changes by more than 5% of $F_l(g)$ at the next point.

With the GP in hand we then use Eq. (21) to mix it with $F_s^{N_s}(g)$ and $F_l^{N_l}(g)$ to obtain our mixed model. Results of this procedure are discussed in Sec. V C.

B. Diagnostics: Mahalanobis distance

We employ the (squared) Mahalanobis distance to check which kernel and training method are optimal for this toy model. The Mahalanobis distance includes correlated errors and maps the differences of the GP and the true solution across the input space to a single value. It is given by

$$D^2_{\mathrm{MD}} = (\boldsymbol{y} - \boldsymbol{m})^T K^{-1}(\boldsymbol{y} - \boldsymbol{m}), \tag{24}$$

where $\boldsymbol{y}$ is a vector of solutions (either of a draw from the GP being used or the true solution) and $\boldsymbol{m}$, K are the mean and covariance matrix from the GP, as previously noted [19].

To determine whether or not the calculated squared Mahalanobis distance is reasonable, we compare it to a reference distribution. To construct the reference distribution we take $\boldsymbol{y}$ to be numerous draws from a multivariate normal distribution constructed from the GP mean and covariance matrix. This reference distribution should converge to a χ^2 distribution with the degrees of freedom matching the number of points used to calculate D^2_{MD} [27]. We used samples from the emulator to construct the reference distribution and compared to a χ^2 with the appropriate degrees of freedom. This verifies

that the reference distribution is indeed the χ^2 distribution, so hereafter we compare our value of D^2_{MD} to the expected value given the χ^2.

Now we take $\boldsymbol{y}$ to be the vector representing the true solution at a series of points in g, as we want to compare our model to the true result for this toy problem. We pick three points in g to calculate the Mahalanobis distance, and ensure that they are sufficiently spaced that there are not too many points within one length scale of the GP. We obtain the results for $\mathrm{D}^2_{\mathrm{MD}}$ for each training method and kernel given in Table I. Seven pairs of truncation orders N_s, N_l were chosen to test each training method and kernel previously discussed. A suitable Mahalanobis distance is one that is close to the number of degrees of freedom of the χ^2 distribution we are comparing to. Examining the RBF kernel results, it is obvious that this kernel is not the correct choice for this problem. The computed D^2_{MD} is orders of magnitude too large in many cases there, regardless of training method. Because of this, we rule out using the RBF kernel entirely. The enormous squared Mahalanobis distance values given by this kernel seem to be due to its smoothness: the RBF kernel is so smooth that it breaks down and leads to K in Eq. (24) becoming asymptotically singular. Inverting this matrix then leads to these inflated D^2_{MD} numbers. This can be seen by invoking the singular value decomposition (SVD). This shows a drop in the magnitude of the eigenvalues by several orders, progressing from the first to the last. Often the final eigenvalue is on the order of 10^{-7} or smaller, because the RBF covariance matrix is near degenerate [28].

Examining the training method results under the Matérn 3/2 kernel, each of the calculated squared Mahalanobis distances is within the variance of the $\chi^2(N = 3)$ distribution. This implies that the method used does not have an overly strong correlation with the result of the interpolation such as to be detrimental if we chose, say, method 1 over method 2. In all of our final results we employed method 2, as it yields the smallest Mahalanobis distance of the three methods here.

C. Results and discussion

In Fig. 7, the mixed models have been plotted for $N_s, N_l = 2$ and $N_s, N_l = 5$ [panels (a) and (b), respectively]. These truncation orders are the same as those in Fig. 5, chosen for ease of direct comparison between the two figures. Adding the GP as an interpolant is shown to greatly lessen the curvature of the mixed model across the gap, as well as suppress the uncertainty in that region to obtain a more precise estimate of the true curve. The improvement in UQ is so significant that the error model chosen (discussed in Secs. II B and IV) no longer has a sizable effect on the results of the mixed model. This is excellent news in regards to generalizing this method: if the final mixed curve does not display a lot of sensitivity to the high-order behavior of the coefficients in the theory error model the interpolants that result will be much more robust.

A complementary verification of our success in estimating the uncertainty of the mixed model is evident in Table II, where the gap length (over g) is given for each truncation order pair, and the area in the uncertainty band is calculated via a simple implementation of Simpson's rule. The reduction

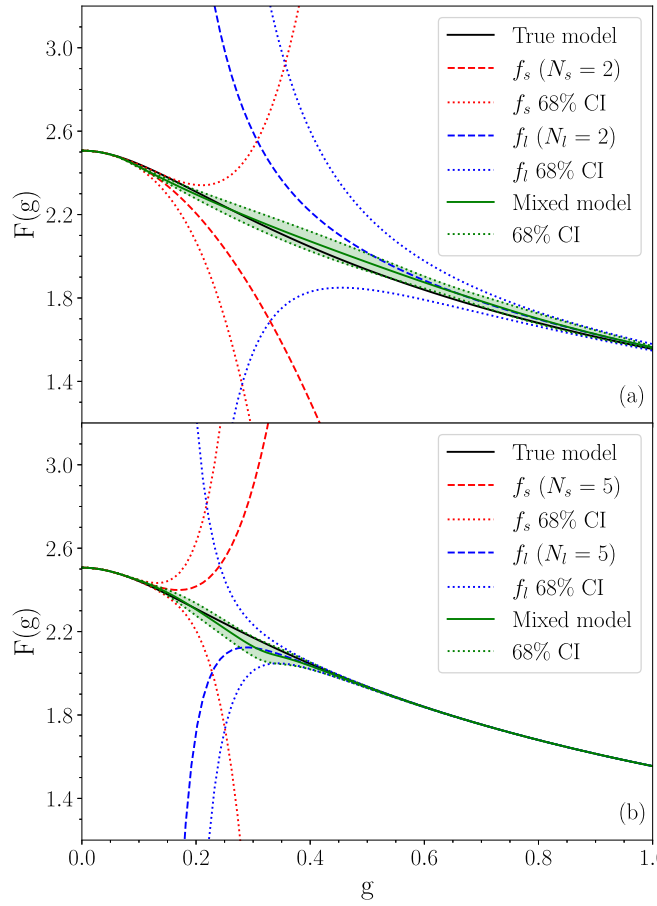

FIG. 7. The results of employing the Matérn 3/2 kernel in a GP to cross the gap between the two series expansions, truncated at (a) $N_s = 2, N_l = 2$ and (b) $N_s = 5, N_l = 5$. Compare this result to Fig. 5, which did not include the GP.

of the gap area as the truncation order increases and the width decreases supports our GP result as a good interpolant for this mixed model, because the size of the uncertainty band should decrease as one adds more terms (information) to the mixed model.

A direct comparison of the bivariate results to those of the multivariate model mixing using a GP for $N_s, N_l = 3$ can be examined in Fig. 8. In this case, the mixed model including the GP [panel (b)] is a very close approximation to the true

TABLE II. Calculated areas of each uncertainty band per truncation order in N_s and N_l. As expected, as both N_s and N_l increase, the uncertainty reduces.

N_s	N_l	Gap length (g)	Gap area
2	2	0.92	0.05
2	4	0.60	0.02
3	3	0.92	0.04
4	4	0.26	0.03
5	5	0.21	0.02
8	7	0.13	0.01
5	10	0.11	0.003

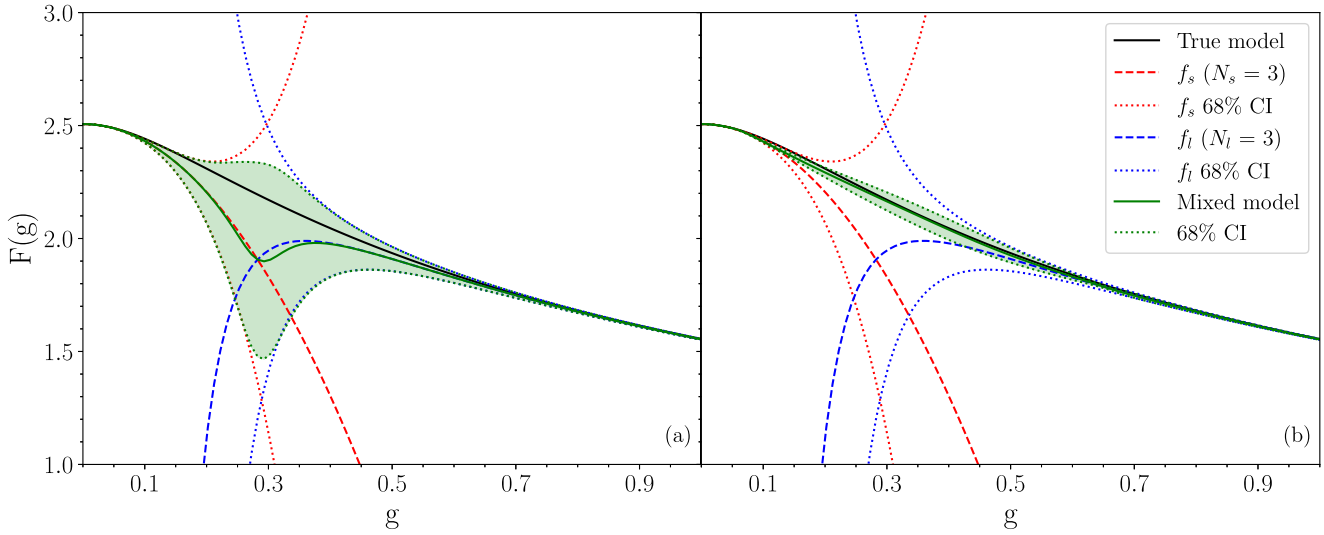

FIG. 8. A comparison of the results of the method from Sec. IV and that of Sec. V, applied to the $N_s = 3$ and $N_l = 3$ cases. The reduction in the credibility interval is evident in the latter method, as is the lessening of the curvature in the gap between the expansions.

curve, a stark contrast to the mixed model in panel (a), where the mixed curve follows the individual models much more closely than the true curve. The effect of the two expansions both approaching $-\infty$ is nearly nonexistent when the GP is employed for this case, showing that this method has greatly improved upon Sec. IV in multiple ways. Both the reduction of the overly conservative error bar in panel (a) and the improvement of the mean curve are promising results. These further our argument that this method, suitably generalized, will be a great asset to anyone wishing to model mix across an input space such as this.

D. Weights

As previously discussed, within BMM it is possible to construct weights that are dependent upon the location in the input space considered; for our toy problem this means the weights are a function of g. We explicitly plot them for each BMM method used for the cases of $N_s = N_l = 2$ (Fig. 9) and $N_s = N_l = 5$ (Fig. 10). These weights are plotted as α and $(1 - \alpha)$ functions in the linear mixture model [panel (a) in both figures], and normalized precisions in the case of the bivariate BMM and trivariate BMM [panels (b) and (c) in both figures].

Panels (a) in these two figures both look like we would expect, given our knowledge of $\alpha(g;\boldsymbol{\theta})$ from Sec. III: the red curve dominates until some crossing point where the blue takes over as the dominant weight, indicating that the most influential model on the mixed model turns slowly from the $F_s^{N_s}(g)$ (weighted by the red curve) to $F_l^{N_l}(g)$ (weighted by the blue curve). The same can be seen in panels (b) in Figs. 9 and 10 where the two curves also progress in the same fashion as in panels (a), showing the crossover from $F_s^{N_s}(g)$ dominance at low g to $F_l^{N_l}(g)$ dominance at large g. But, the crossover in the bivariate case is much sharper than that in the linear mixture model case.

Panels (c) highlight the substantial influence of the GP in the trivariate mixed model result across the gap: in the center of the gap for both Figs. 9 and 10, the GP reaches a weight of nearly 1.0, indicating it is exerting the most control over the result of BMM there. The decline of the GP curve is rather slow on the large-g side compared to the sharp rise on the low-g side of the input space, which could be related to the gap slowly closing on the large-g end in comparison to its somewhat quicker opening on the small-g end [see Fig. 7(a), where this is evident]. The location in the input space of the crossover from GP to $F_l^{N_l}(g)$ for the trivariate mixed model compared to when $F_l^{N_l}(g)$ takes over when we mix only two

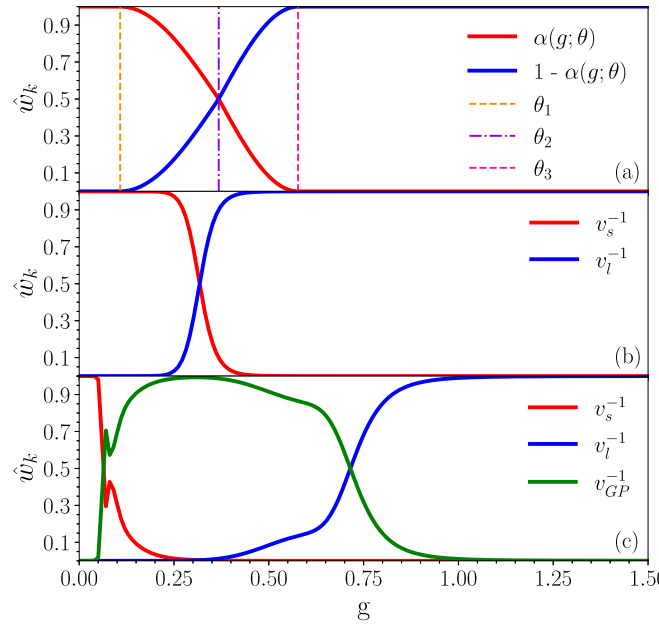

FIG. 9. The normalized weights of each model ($N_s, N_l = 2$, shown in typical red and blue) computed across the input space g, for each method: (a) linear mixture model, (b) bivariate BMM, and (c) trivariate BMM including the GP. Note the green in panel (c) is the weights curve for the GP interpolant.

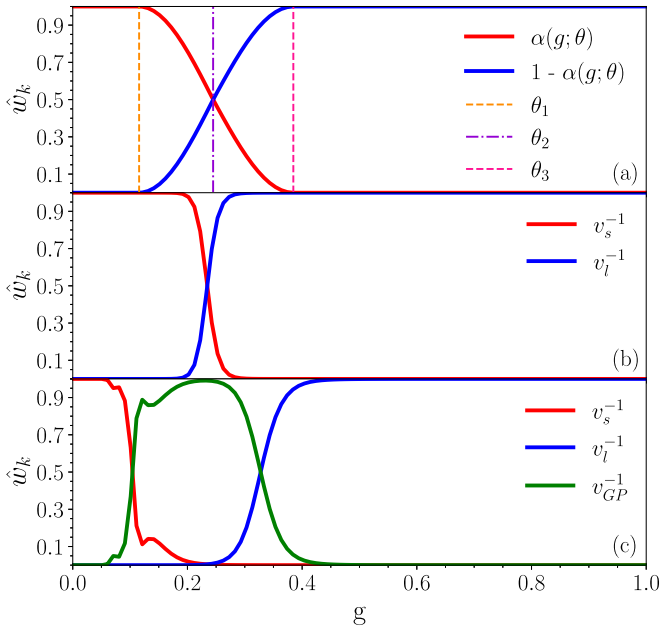

FIG. 10. The normalized weights of each model as a function of the input space g, as in Fig. 9, though now for $N_s, N_l = 5$. Note the decrease in the region in which the GP dominates in this figure, in contrast to the case in Fig. 9.

models varies with the orders considered. In the $N_s = N_l = 5$ case it is similar to the linear mixture model but for the $N_s = N_l = 2$ case the GP dominates well beyond where the gap closes for two models.

E. Comparison with the FPR method

In Ref. [8] a method involving fractional powers of rational functions (FPR method) was used to interpolate between these two limits. The FPR method postulates a form

$$F(g) = g^a \left(\frac{1 + c_1 g + c_2 g^2 + \cdots + c_p g^p}{1 + d_1 g + d_2 g^2 + \cdots + d_q g^q} \right)^\alpha \tag{25}$$

that is designed to mimic a Padé approximant, but be more flexible since it allows an arbitrary leading power at small g and quite a lot of flexibility as regards the choice of the fractional power α to which the Padé will be raised. If we have N_s (N_l) coefficients from the small (large)-g expansion then we need [8]

$$p = \frac{1}{2}\left(N_l + N_s + 1 - \frac{(a-b)}{\alpha} \right),$$
$$q = \frac{1}{2}\left(N_l + N_s + 1 + \frac{(a-b)}{\alpha} \right). \tag{26}$$

The p and q coefficients c_k and d_k are then adjusted to satisfy the constraints imposed by the need to reproduce the first N_s terms of the small-g expansion of F and the first N_l terms of the large-g expansion of F. As is clear from Eq. (25), the coefficient a serves to ensure the small-g behavior is correctly reproduced. In our specific toy problem, we possess $a = 0$ and $b = -1/2$.

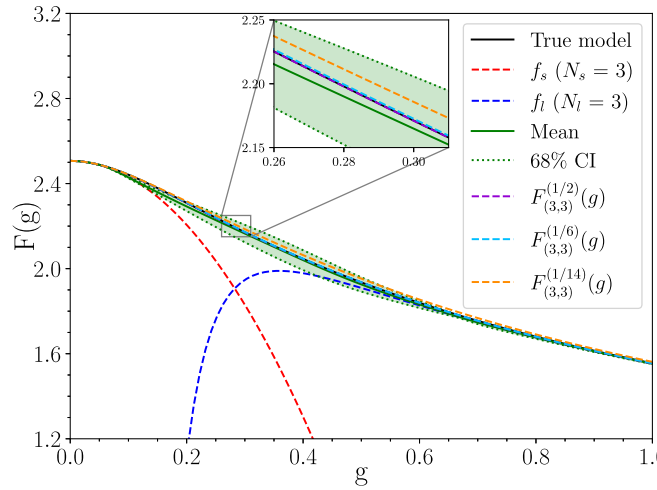

FIG. 11. Overlay of the FPR method curves for different values of α in the $N_s, N_l = 3$ case with the mixed model results using the GP. *Inset*: A closeup of the central mixing region, where it is easier to note the deviation of the FPR results from the true solution as the value of α decreases.

In Fig. 11 we compare the results of Honda's FPR code [8] with our multivariate mixed model for $N_s, N_l = 3$. Honda chooses three different α values, and as α decreases the accuracy of the FPR method decreases, as shown in the inset. Our GP mixed model mean is not as close to the true result as the FPR results for $\alpha = 1/2$ and $1/6$. Nevertheless, the FPR method does not include any UQ, which is arguably the most important development of this work, making our method a compelling option for researchers wishing to include principled UQ in their results. Furthermore, while the FPR interpolation method applies only to series expansions, our approach can be adapted to many other situations.

VI. SUMMARY AND OUTLOOK

In this work we have explored how to apply Bayesian model mixing to a simple test problem, interpolating expansions in the coupling constant g of a zero-dimensional ϕ^4 theory partition function (see Sec. II A). Rather than choose the single "best" model, as in Bayesian model selection, or an optimal global combination of predictions, as in BMA, we seek to vary the combination of models as a function of input location. The goal is to optimize the predictive power across the full input domain and quantify the prediction's uncertainty. Introducing BMM with a transition region is more flexible than treating this as a model selection problem. (Model selection can be considered a special case of BMM, for which the weights of each model are either zero or one.)

The basic output of BMM for our example and for more general cases is a combined PPD, see Eq. (2), which is a probability distribution for the output at a new input point based on an appropriate weighted sum of individual PPDs. To be clear, the prediction is not simply the most probable outcome, although the distribution may be summarized by MAP values and the variance (or covariances in general). The challenge is to how to best determine these input-dependent weights.

A general feature of this task is the need to incorporate error estimates from theory or learned from data. In our toy model, we exploit analytic solutions for the full integral and for small g and large g expansions. Analytic knowledge of these expansions enables credible error estimates; in practice we simulate different levels of credibility with "uninformative" and "informative" error models, as in Fig. 6 for example. By changing the order of each expansion, we can adjust the gap between their individual regions of highest validity as a function of g. Thus this problem serves as a prototype for combining predictive models with different patterns of coverage across the full input domain of interest.

We have explored a sequence of BMM strategies. In Sec. III we considered linear mixtures of PPDs, following Coleman [21], using the informative error model and parametric mixing with a switching model (see Fig. 2). We identified pitfalls with this approach when it is necessary to "mind the gap"; in particular, when the gap is large, this approach does not yield a statistically consistent result for the intermediate region. The linear mixture model did produce a PPD with good empirical coverage properties for high-order expansions. Perhaps unsurprisingly, it is easier to bridge the gap when the gap is small.

While reaching such high orders is unlikely to be realistic for physics applications, linear mixing may be an appropriate procedure when the domains of applicability overlap significantly. Many discussions about what to do in the region between the two expansions become moot if the expansions share a region in which both are convergent. Competing EFT expansions provides an example, such as those for pionless and chiral effective field theory (EFT) in the few-nucleon problem.

In Sec. IV we considered localized bivariate mixing, with each model weighted by its precision (inverse variance) to incorporate the theory uncertainty. Here we found good but rather conservative empirical coverage and a nonintuitive mean prediction for our model problem: where we expected a more-or-less straight joining of the limiting expansions the mean curves were significantly curved. These are not, however, failures because they correctly reflect the limited prior information about the behavior of the function in the gap. Since we used only the small-g and large-g expansions to inform the mixture, sinusoidal behavior of the mean is not ruled out, and is seen in Fig. 5. The small- and large-g expansions themselves do not preclude a rapid change in the region where neither is valid.

Taking care not to overconstrain intermediate behavior may be relevant for the application to bridging the nuclear matter equation of state between saturation density and QCD "asymptopia." At the low density end, chiral EFT provides an expansion with theoretical truncation errors, while at the high density end one can anchor the result with a QCD calculation, as in Ref. [29]. Because the question of an intermediate phase transition is open, we would want to minimize prior constraints.

Ambiguities as regards the behavior in the intermediate regime will disappear in applications where competing expansions overlap, as already pointed out for linear mixing. Even if expansions do not overlap we may have prior information that the observable being studied is smooth in the intermediate region. In Sec. V we introduced a model to bridge the gap in the toy model with prior expectations of smoothness, implemented as localized multivariate mixing using a GP. Not unexpectedly, the details of the GP, such as the choice of the kernel, matter (Matérn 3/2 is much favored over RBF), but the use of a Mahalanobis diagnostic to check on the kernel and training method shows reasonable insensitivity to the choice of the latter. The final results are very good. If we did not seek an uncertainty estimate, the FPR results [8] might be favored for highest precision in this particular application. But, for the more likely extensions to nuclear physics problems, our multivariate model mixing is a compelling approach.

When using this approach to construct the mixed model we did not take into account correlations between the truncation error in each of the power series across different values of g. Such correlations do not affect the mixing at a particular value of g. Extending the multivariate model mixing approach employed here to obtain the covariance matrix of the mixed model that emerges from combining K models, each of which has a nondiagonal covariance matrix, is an interesting topic for future investigation.

We expect that our explorations could be extended to many different phenomena in nuclear physics. For example, another candidate for BMM is to bridge the gap between the situations in nuclear data where R-matrix theory dominates and where Hauser-Feshbach statistical theory is the principal choice. This and other applications will require extending the BMM considered here to multidimensional inputs, as has been done in the case of Bayesian model averaging (e.g., [30]). The SAMBA computational package will enable new practitioners to learn the basics and to explore strategies in a simple setting. Not only can users try the toy model detailed in this work, but they will be able to include their own data and try these model mixing methods for themselves in the next release. This will include adoption of other BMM strategies, such as using Bayesian additive regression trees (BART) [31,32].

ACKNOWLEDGMENTS

We thank Özge Sürer, Jordan Melendez, Moses Chan, Daniel Odell, John Yannotty, Matt Pratola, Tom Santner, and Pablo Giuliani for helpful discussion on many sections of this work, and suggestions with regards to the structure of SAMBA. We thank Frederi Viens for supplying names for the methods of Secs. IV and V and Kostas Kravvaris for suggesting an interesting application of this approach. This work was supported by the National Science Foundation Awards No. PHY-1913069 and PHY-2209442 (R.J.F.), CSSI program Award No. OAC-2004601 (A.C.S., R.J.F., D.R.P.), and the U.S. Department of Energy under Contract No. DE-FG02-93ER40756 (D.R.P.).

APPENDIX A: BAYESIAN BACKGROUND

We provide a simple description of the mechanics of Bayesian statistics (see [16] and [33] for a more comprehensive exposition). Bayes' theorem relates two different

conditional probability distributions:

$$p(\boldsymbol{\theta}|\mathbf{D}) = \frac{p(\mathbf{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{D})}, \tag{A1}$$

i.e., it relates the likelihood determined from data $\mathbf{D}$ and the prior information on parameters $\boldsymbol{\theta}$ to the posterior of $\boldsymbol{\theta}$. The evidence term in the denominator of Eq. (A1) serves as the normalization of this relation and can be dropped if one is only interested in working within a single model.

The prior probability distribution in Eq. (A1), $p(\boldsymbol{\theta})$, is a specific representation of any prior knowledge we may have about the parameters we are seeking. It is imperative that we include any physical information known about the system in this quantity when applying Bayesian methods to real physics problems; this is the way we can include physical constraints in the statistical framework.

A commonly used method to determine the posterior density distributions of parameters $\boldsymbol{\theta}$ is Markov Chain Monte Carlo (MCMC) sampling. This technique often utilizes a Metropolis-Hastings algorithm to map out the posterior of each parameter given the likelihood and prior probability density functions (PDFs). We employed this technique in Sec. III.

APPENDIX B: TRUNCATION ERROR MODEL DEVELOPMENT, OR HOW TO BE A GOOD GUESSER, OR HOW TO BE A POOR GUESSER AT FIRST AND THEN GET BETTER

We take two models f_1 and f_2, with variances v_1 and v_2, and form the PDF for the mixed model, $f_\dagger$, according to

$$f_\dagger \sim \mathcal{N}\left(\frac{v_1 f_2 + v_2 f_1}{v_1 + v_2}, \frac{v_1 v_2}{v_1 + v_2}\right), \tag{B1}$$

as detailed in Sec. 3.4 of Ref. [16]. In our implementation the variances are defined by the error model associated with omitted terms.

Uninformative error model. For $F_s(g)$, the expansion in positive powers of g, we have coefficients that behave as

$$\begin{aligned}\frac{s_n}{n!} = \{&2.5066, -3.7599, 5.4833, -6.0316,\\ &5.2507, -3.7688, 2.2984, -1.2178,\\ &0.5702, -0.2391, 0.09081, -0.03150,\\ &0.01006, -0.002975, 0.0008193;\\ n = 2, 4, 6, \ldots, 30\},\end{aligned} \tag{B2}$$

with even more rapid decrease thereafter. All odd powers of g have vanishing coefficients. Taking the first-omitted term approximation to the part of the expansion omitted at order $2n$ we can adopt

$$\sigma_{N_s} = \bar{c}(N_s + 2)! g^{N_s+2} \tag{B3}$$

for N_s even and

$$\sigma_{N_s} = \bar{c}(N_s + 1)! g^{N_s+1} \tag{B4}$$

if N_s is odd. Here the coefficient $\bar{c}$ is estimated by taking the root-mean-square value of coefficients up to the order of truncation N_s. This will have a tendency to overestimate the next coefficient, but it provides a conservative error bar.

For $F_l(g)$, the expansion in negative powers of g, we have coefficients that behave as

$$\begin{aligned}l_m \times m! = \{&1.813, -0.3064, 0.1133, -0.05744,\\ &0.03541, -0.02513, 0.01992, -0.01728,\\ &0.01618, -0.01620, 0.01719, -0.01923,\\ &0.02257, -0.02765, 0.03526;\\ m = 0, \ldots, 14\},\end{aligned} \tag{B5}$$

with eventual rapid increase. Here we once again take the first-omitted term approximation to v_2 to estimate the error due to truncation at order N_l as

$$\sigma_{N_l} = \bar{d}\frac{1}{(N_l + 1)!}\frac{1}{g^{N_l+3/2}}. \tag{B6}$$

Here, $\bar{d}$ is estimated by taking the root-mean-square value of coefficients from order 2 to the mth order.[3] There might be some tendency to underestimate the next coefficient if we are unlucky as to where we stop. But, up to order $m = 14$, this seems to provide a reasonable estimate of the size of the next term.

Informative error model. Let us suppose we are better guessers, as regards the asymptotic behavior of the coefficients. Then we might form

$$\begin{aligned}\frac{s_{2n}}{(n-1)!4^{2n}} = \{&-0.469993, 0.514055, -0.530119,\\ &0.538402, -0.543449, 0.546846,\\ &-0.549287, 0.551126, -0.552562,\\ &0.553713;\\ n = 2, 4, \ldots, 20\},\end{aligned} \tag{B7}$$

which is remarkably stable out to $n = 60$. Similarly, we might look at

$$\begin{aligned}l_m \times \Gamma(m/2 + 1) \times 4^m = \{&1.8128, -1.086, 0.9064,\\ &-0.8145, 0.7553, -0.7127,\\ &0.6798, -0.6533, 0.6312,\\ &-0.6125, 0.5962, -0.5818;\\ n = 0, \ldots, 12\},\end{aligned} \tag{B8}$$

which has only a very slow decrease, reaching 0.3968 by $n = 50$. Based on these behaviors we formulate error model 2, as defined by Eqs. (10) and (12).

[3]The first two coefficients shift $\bar{d}$ to a value much higher than any of the succeeding coefficients would indicate. This, in turn, widens the interval on the large-g expansion much more than we would wish to observe, so in this model we neglect them.

[1] N. K. Glendenning, Phys. Rev. D **46**, 1274 (1992).
[2] N. K. Glendenning, Phys. Rep. **342**, 393 (2001).
[3] L. McLerran and S. Reddy, Phys. Rev. Lett. **122**, 122701 (2019).
[4] S. Han, M. A. A. Mamun, S. Lalit, C. Constantinou, and M. Prakash, Phys. Rev. D **100**, 103022 (2019).
[5] C. Wellenhofer, D. R. Phillips, and A. Schwenk, Phys. Status Solidi B **258**, 2000554 (2021).
[6] S. M. Johns, P. J. Ellis, and J. M. Lattimer, Astrophys. J. **473**, 1020 (1996).
[7] X. Zheng *et al.* (CLAS Collaboration), Nat. Phys. **17**, 736 (2021).
[8] M. Honda, J. High Energy Phys. 12 (2014) 19.
[9] C. Wellenhofer, D. R. Phillips, and A. Schwenk, Phys. Rev. Res. **2**, 043372 (2020).
[10] M. Y.-H. Chan, R. J. DeBoer, R. J. Furnstahl, D. Liyanage, F. M. Nunes, D. Odell, D. R. Phillips, M. Plumlee, A. C. Semposki, Ö Sürer, and S. M. Wild, BANDFramework: An Open-Source Framework for Bayesian Analysis of Nuclear Dynamics, Version 0.2.0 (2022), https://github.com/bandframework/bandframework.
[11] M. Plumlee, Ö Sürer, and S. M. Wild, *Surmise Users Manual*, Version 0.1.0 (NAISE, Evanston, IL, 2021).
[12] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, Stat. Sci. **14**, 382 (1999).
[13] Y. Yao, A. Vehtari, D. Simpson, and A. Gelman, Bayesian Anal. **13**, 917 (2018).
[14] Y. Yao, G. Pirš, A. Vehtari, and A. Gelman, Bayesian Anal. (2021).
[15] C. Fernandez and P. J. Green, J. R. Stat. Soc B **64**, 805 (2002).
[16] D. R. Phillips, R. J. Furnstahl, U. Heinz, T. Maiti, W. Nazarewicz, F. M. Nunes, M. Plumlee, M. T. Pratola, S. Pratt, F. G. Viens, and S. M. Wild, J. Phys. G: Nucl. Part. Phys. **48**, 072001 (2021).
[17] M. Cacciari and N. Houdeau, J. High Energy Phys. 09 (2011) 39.
[18] R. J. Furnstahl, N. Klco, D. R. Phillips, and S. Wesolowski, Phys. Rev. C **92**, 024005 (2015).
[19] J. A. Melendez, R. J. Furnstahl, D. R. Phillips, M. T. Pratola, and S. Wesolowski, Phys. Rev. C **100**, 044001 (2019).
[20] J. Ellenberg, *How Not to be Wrong: The Power of Mathematical Thinking*, 8th ed. (Penguin, London, 2014).
[21] J. R. Coleman, Topics in Bayesian Computer Model Emulation and Calibration, with Applications to High-Energy Particle Collisions, Ph.D. thesis, Duke University, 2019.
[22] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publ. Astron. Soc. Pac. **125**, 306 (2013).
[23] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevC.106.044002 for additional PPD results of other choices of N_s and N_l for each BMM method described in the paper.
[24] L. Wittgenstein, *Tractatus Logico-Philosophicus* (Routledge, London, 2001), Proposition 7.
[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011).
[26] scikit-learn developers, scikit-learn User Guide Sec. 1.7.5. Kernels for Gaussian Processes, https://scikit-learn.org/stable/modules/gaussian_process.html#kernels-for-gaussian-processes, 2007–2022, accessed 2022-05-19.
[27] L. S. Bastos and A. O'Hagan, Technometrics **51**, 425 (2009).
[28] R. Ababou, A. C. Bagtzoglou, and E. F. Wood, Math. Geol. **26**, 99 (1994).
[29] S. Huth, C. Wellenhofer, and A. Schwenk, Phys. Rev. C **103**, 025803 (2021).
[30] V. Kejzlar, L. Neufcourt, W. Nazarewicz, and P.-G. Reinhard, J. Phys. G: Nucl. Part. Phys. **47**, 094001 (2020).
[31] H. A. Chipman, E. I. George, and R. E. McCulloch, Ann. Appl. Stat. **4**, 266 (2010).
[32] J. Hill, A. Linero, and J. Murray, Annu. Rev. Stat. Appl. **7**, 251 (2020).
[33] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Phys. Rev. C **98**, 034318 (2018).