

Bayesian parameter estimation for effective field theories

S. Wesolowski,^{1,*} N. Klco,^{2,3,†} R.J. Furnstahl,^{1,‡} D.R. Phillips,^{2,§} and A. Thapaliya^{2,¶}

¹*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*

²*Institute of Nuclear and Particle Physics and Department of Physics and Astronomy, Ohio University, Athens, OH 45701, USA*

³*Department of Physics, University of Washington, Seattle, WA 98195, USA*

(Dated: May 2, 2016)

We present procedures based on Bayesian statistics for estimating, from data, the parameters of effective field theories (EFTs). The extraction of low-energy constants (LECs) is guided by theoretical expectations in a quantifiable way through the specification of Bayesian priors. A prior for natural-sized LECs reduces the possibility of overfitting, and leads to a consistent accounting of different sources of uncertainty. A set of diagnostic tools are developed that analyze the fit and ensure that the priors do not bias the EFT parameter estimation. The procedures are illustrated using representative model problems, including the extraction of LECs for the nucleon mass expansion in SU(2) chiral perturbation theory from synthetic lattice data.

PACS numbers: 02.50.-r, 11.10.Ef, 21.45.-v, 21.60.-n

I. INTRODUCTION

Effective field theories (EFTs) describe physics in the presence of a separation of scales. They exploit known symmetries to construct general interactions at the lower momentum scale that do not depend on unresolved details of the physics from the higher scale [1–4]. A ratio (or, in the case of multiple scales, ratios) between the separated scales is formed, enabling a power-counting scheme that organizes the (renormalized) contributions of different operators to observables in an order-by-order expansion in the ratio(s) [4]. The effects of unresolved physics are parametrized by the coefficients of these operators, called low-energy constants (LECs). Estimating these LECs using experimental data or numerical simulations of the underlying theory is essential if the EFT is to be used to make quantitative predictions. But, since the EFT expansion is written in terms of a ratio of disparate scales, even in the absence of such data, we expect the LECs in the EFT should be of “natural” size, that is, of order unity when expressed using the appropriate scales.

In this paper we lay out a framework for EFT practitioners to obtain estimates of the LECs in their EFT while incorporating *all* the information at their disposal. A Bayesian formulation of the problem is ideal for this, as it facilitates folding theoretical expectations—such as that the LECs should be natural—into parameter estimation. Using only data (together with their statistical errors) to constrain the LECs can lead to distortions in the fits; this is already recognized by EFT practitioners who adapt their fitting procedures to include additional information (e.g., in fitting NN scattering data, see Sec. II A). The use of priors is thus very well motivated in the EFT

context. Even though it is still controversial in other contexts, such Bayesian reasoning has recently found several significant applications in EFTs for nuclear physics [5–10]. But priors can bias the parameter extraction if not used carefully. Validating the prior is a key element of robust parameter estimation.

The naturalness priors and Bayesian framework also allow us to examine the impact of higher-order terms in the EFT. Those terms can be incorporated into the fit, either through a known functional form or via the assumed EFT expansion for observables (see Ref. [8]), and then accounted for using the rule of marginalization. This allows the parameter estimation to account for all three sources of uncertainty in the extraction of LECs [11]:

1. Data uncertainties from experiment or numerical simulations.
2. Systematic errors from truncating the EFT.
3. Errors from methods used to compute observables.

These uncertainties are often intertwined and must be dealt with consistently, and each is nontrivial to estimate. For example, even when there is a large amount of high-quality data available, as in the case of NN scattering, that data must be carefully analyzed for consistency, as in the self-consistent database of Ref. [12]. But, while errors of the first type are often included in LEC extractions—and errors of the third type are sometimes also accounted for—it is far rarer for such extractions to consider the systematic errors from truncation of the EFT. Yet neglecting those means we fail to incorporate all the information we have about the EFT when determining the best values for LECs.

Using information on LEC naturalness also ameliorates the well-known phenomena of “overfitting” and “underfitting” within the EFT:

Overfitting. The parameters get fine-tuned to the data. A naturalness prior reduces the risk of this, since it restricts degenerate directions in the EFT

* wesolowski.14@osu.edu

† nk405210@ohio.edu

‡ furnstahl.1@osu.edu

§ phillips@phy.ohiou.edu

¶ at311509@ohio.edu

parameter space. Together with marginalization over higher orders, the prior additionally ensures that the LECs are not tuned to reproduce the data better than expected from the theoretical truncation error.

Underfitting. The model provides an inadequate representation of at least part of the data being fit (e.g., not high enough order). Marginalization over higher EFT orders ensures that the combined data and truncation uncertainties—which together determine the total uncertainty of the LECs—grows as the EFT parameter increases. Without this feature, blind application of least-squares fitting can lead to low-order LECs being overly influenced by data in an energy domain where the corresponding EFT expression has large errors.

When conducting EFT parameter estimation using data over a large range of the expansion parameter, it is possible to experience tendencies toward both overfitting and underfitting. In the past these biases have usually been decoupled in EFT fits by limiting the parameter range over which data are fitted. But, ideally, all available data would be used to inform results. A well-constructed prior protects against these biases while simultaneously relieving the scientist of the responsibility of justifying an arbitrarily chosen cutoff of data. We will show that our Bayesian EFT parameter estimation converges simply and intuitively, and that it does so even for data sets that include regions where the EFT converges poorly, or not at all.

This is achieved by computing so-called posterior probability distributions (pdfs) for LECs with all assumptions, such as naturalness, made explicit. We develop a system of diagnostics that ensure consistent, reproducible posteriors, with all assumptions manifested through the use of Bayesian priors. The diagnostics also allow tests of the sensitivity to particular prior prescriptions. Such a framework avoids both overfitting and underfitting in a statistically well-defined way, thereby maximizing the information from data that enters the LEC determination.

The probability distributions yield estimates of the parameters of the effective theory that are derived from both the data considered and underlying knowledge of the theory itself. Propagating the resulting uncertainties to the EFT's predictions displays the impact that uncertainties in the data used to fix the LECs have on those predictions. By combining those results with our prescription for truncation errors from Ref. [8] we can fully account for the uncertainties in the EFT (we do not address here the uncertainties from calculational methods, although their inclusion is well defined), thereby providing reliable guidance as to the power of new experimental or lattice data to improve theoretical predictions.

Following Refs. [11, 13], where a model problem was introduced to simulate the systematic behavior of an EFT expansion, we will explore representative examples to

demonstrate issues that occur in EFT parameter estimation. To simulate the extraction of EFT parameters from data, we contrive functions of a single variable, say, $g(x)$. This function is designed to have a Taylor series whose coefficients are order unity within a radius of convergence that is equal, or close, to 1:

$$g(x) = a_0 + a_1 x + a_2 x^2 + \dots; \quad |x| < 1. \quad (1)$$

The a_i s play the role of LECs in this example. They are extracted from synthetic data to which Gaussian noise is added to simulate experimental errors. In this case, the data are generated by choosing a set of N_d —typically evenly spaced— x points up to a maximum value of $x = x_{\max}$; given the j^{th} point x_j , the data point $d(x_j) \equiv d_j$ and error σ_j are

$$d_j = g(x_j)(1 + c\eta_j) \implies \sigma_j = c d_j, \quad (2)$$

where η_j is normally distributed with mean 0 and standard deviation 1. c is a specified relative error.

The goal of the model problems is then to reliably estimate as many EFT parameters as possible given certain synthetic data. We do this by choosing a truncation order k for the “theory” (which is really just a polynomial) and determining the posterior pdfs of the coefficients a_i in

$$g_{\text{th}}(x) \equiv \sum_{i=0}^k a_i x^i \quad (3)$$

from the data.¹ To ensure that the extraction is robust and reliable, one includes by a marginalization integral the contributions from a (hopefully finite) number of influential higher-order coefficients $\{a_{k+1}, \dots, a_{k_{\max}}\}$ up to some order k_{\max} . This accounts for the uncertainties in the posterior that arise as an artifact of finite-order truncation of the EFT, as presented in Ref. [8]. For the simple models here, this marginalization is the same as doing parameter estimation up to truncation order k_{\max} .

These model problems allow us to control the “natural” size of the coefficients being determined, the distribution and range of the “experimental” data, and its precision. We can explore pertinent EFT issues such as the presence of unnatural coefficients or improvements gained by adding new data, while avoiding the computational cost of computing observables from EFTs order-by-order.

However, such models do not include certain generic aspects of actual EFTs. An obvious extension of Eq. (3) is to a problem with two light scales, as occurs in chiral perturbation theory (χ PT), where the pion mass and

¹ The extension to an EFT with multiple coefficients at a given order and/or more than one expansion parameter because of multiple low-energy scales follows by generalizing the vector of coefficients (e.g., see Sec. V C).

the (assumed small) external momentum can be used to form two independent expansion parameters. Here we deal only with series in one variable; the extension to additional variables should be straightforward.

Another aspect of EFTs which is not reflected in Eq. (3) is that the coefficient of the term of $\mathcal{O}(x^i)$ will typically be a non-analytic function of x , $\mathcal{A}_i(x)$, of which the LEC a_i forms only the constant part. For the parameter-estimation problems being discussed here it is crucial that the non-analytic parts of \mathcal{A}_i —i.e. all but the LEC that appears in it—are predicted by the EFT. (Indeed, at certain orders in the EFT expansion \mathcal{A}_i contains no LEC and is completely predicted.) Thus, up to the order to which the EFT series has been computed, these non-analytic pieces can be removed from the parameter-estimation problem by simply subtracting them from the data. Of course, such a subtraction is only feasible for orders where the \mathcal{A}_i are known. But, at all orders in the EFT, the \mathcal{A}_i should be $\mathcal{O}(1)$ for x in the domain where the EFT applies. This allows us to deal with higher-order terms in the EFT series by imposing a naturalness prior on unknown higher-order coefficients—even in this general case where the coefficient is, in fact, a function—thereby facilitating marginalization over the higher-order terms.

Such non-analytic terms occur in one of our examples: the estimation of LECs for the two-flavor χ PT expansion of nucleon mass up to sixth chiral order. In this application, which was previously considered in Ref. [13], the LECs are linearly related to the data used for the extractions, just like our generic polynomial model. This example is difficult because it involves extracting a relatively large number of LECs from lattice data in a region where contributions at different orders are not easily distinguished.

In anticipation of later applications to chiral EFT for nuclei, we also examine the impact on our procedures of situations where the quantity for which we have data is a *nonlinear* function of the LECs a_i . In addition, we will need to account for higher-order contributions to the posterior when they have an unknown form or are impractical to compute. Our strategy is to build upon Ref. [8] and use the expected naturalness of the expansion coefficients $\{c_{k+1}, \dots, c_{k_{\max}}\}$ for the fit observables rather than the LECs themselves above the order k . This alternative approach is discussed in Sec. III A.

In Sec. II, we first briefly review alternative procedures used to supplement standard least-squares methods when fitting LECs for chiral EFT, and then outline our Bayesian framework for EFT parameter estimation. Sec. III gives an overview of the procedures and diagnostics for LEC estimation, illustrated with a particular model problem. A flowchart for using the diagnostic framework is presented in Sec. IV and the major steps are illustrated with examples using a different model. Additional case studies, including a nonlinear model and the χ PT mass expansion, are briefly explored in Sec. V. In Sec. VI we summarize and present plans for future anal-

yses with the diagnostic framework by the BUQEYE² collaboration.

II. STATISTICAL METHODS

A. Least-squares minimization

The underlying basis for extracting LECs is generally minimization of a least-squares objective function for a set of observables (e.g., see [14]):

$$\chi^2 = \sum_{i=1}^{N_d} \left(\frac{d_i - g_{\text{th}}(x_i)}{\sigma_i} \right)^2, \quad (4)$$

where x_i is the value of the EFT expansion parameter at which the observable was measured, N_d is the total number of data, d_i is the experimental measurement of the observable, $g_{\text{th}}(x_i)$ is the corresponding EFT prediction at x_i given a set of LECs (which we denote collectively as the vector \mathbf{a}), and σ_i is the uncertainty associated with the observable. A conventional least-squares minimization would take σ_i to be the experimental error, but, as already noted, in practice the procedure is modified to take account of additional elements, such as theoretical errors. Recent fits made of LECs for nucleon-nucleon scattering in chiral EFT in Ref. [15] and Ref. [16], which apply different regularization schemes, provide representative examples of the modifications.

Carlsson et al. add the different sources of uncertainty in quadrature for each observable [15],

$$\sigma^2 = \sigma_{\text{exp}}^2 + \sigma_{\text{numerical}}^2 + \sigma_{\text{method}}^2 + \sigma_{\text{model}}^2, \quad (5)$$

where σ_{exp} is the experimental uncertainty, $\sigma_{\text{numerical}}$ and σ_{method} are computational uncertainties (both included as the third type of uncertainty listed in the Introduction), and σ_{model} is the systematic EFT truncation error. The latter is estimated for scattering at momentum p as a constant times the expected expansion parameter raised to a power given by the first omitted order ($k+1$), namely $(p/\Lambda_b)^{k+1}$, with Λ_b the breakdown scale of chiral EFT. The constant is determined self-consistently by requiring that χ^2 including the full σ of Eq. (5) equals the number of degrees of freedom, as advocated in Ref. [17].

Epelbaum et al. [16] use $\sigma = \sigma_{\text{exp}}$ but address the danger of underfitting by limiting the region of fit data according to estimates of the domain of validity at each EFT order. They also augment the χ^2 function with penalty terms to enforce an expected relation between two LECs (Wigner symmetry) and to limit the allowed range of the D-state probability of the deuteron. These serve to restrict the parameter space for the least-squares fit, which limits overfitting. As part of the uncertainty

² Bayesian Uncertainty Quantification: Errors for Your EFT.

quantification (but not the fit itself), error estimates are made based on expectations for truncation errors. In Ref. [8], these estimates were shown to be justified semi-quantitatively by Bayesian arguments and a naturalness prescription.

The Bayesian approach we advocate for parameter estimation has a different structure to the procedures of Ref. [15, 16]; it is an interesting and relevant question whether those procedures can also be derived or motivated by a Bayesian framework under prescribed conditions. For example, in Ref. [11], it was noted (drawing on Stump et al. [18]) that when Gaussian priors are used one can account for missing higher-order contributions by a modified, augmented χ^2 . However, this does not take the form of adding experimental and theory errors in quadrature as in Eq. (5). We do not intend to pursue these connections further here. Rather we have sketched these examples to highlight that the use of *a priori* supplementary information is standard practice in state-of-the-art LEC optimization.

B. Bayesian parameter estimation

We generically use the notation $\text{pr}(x|y)$ to denote the pdf of x given that y is true. The Bayesian framework for parameter estimation determines the *posterior* $\text{pr}(\mathbf{a}|D, I)$, which is the joint probability distribution for the full set of LECs \mathbf{a} given the data D (including their errors) and any other information I . I includes the form of the EFT expansion and the order at which it is truncated, together with additional knowledge such as the expected naturalness of the LECs. We emphasize that the posterior distribution contains much greater content than just the “best-fit” values for the coefficients (which might, but need not necessarily, be identified as the vector \mathbf{a} that maximizes the posterior). When a Gaussian approximation to the posterior is appropriate, the full content may be effectively summarized by values for the maximum and a covariance matrix. But, for determining how the uncertainties propagate, it is generally necessary to start with the full posterior.

To make progress, we express the desired pdf $\text{pr}(\mathbf{a}|D, I)$ in terms of other pdfs that we can calculate or use to implement prior information, e.g., about naturalness. This is done using the basic rules of probabilistic inference, the sum and product rules, and their immediate consequences such as Bayes’ theorem and marginalization [19, 20]. It was demonstrated by Cox that this is the unique set of rules that follow from basic postulates of consistency [21–23].

The product rule dictates two ways to decompose a joint probability:

$$\text{pr}(x, y|I) = \text{pr}(x|y, I) \text{pr}(y|I) = \text{pr}(y|x, I) \text{pr}(x|I). \quad (6)$$

A simple rearrangement of the two decompositions yields Bayes’ theorem, which when applied to the posterior for

parameter estimation yields

$$\text{pr}(\mathbf{a}|D, I) = \frac{\text{pr}(D|\mathbf{a}, I)\text{pr}(\mathbf{a}|I)}{\text{pr}(D|I)}. \quad (7)$$

The pdf in the denominator, $\text{pr}(D|I)$, which is called the evidence for D , has no dependence on \mathbf{a} and is therefore determined here by normalization. (In this context the normalization follows from the sum rule for probabilities, but the unnormalized magnitude of the resulting integral is useful in Bayesian model selection, see Sec. II E below.) Rewriting Eq. (7) as a proportionality, we have

$$\text{pr}(\mathbf{a}|D, I) \propto \text{pr}(D|\mathbf{a}, I)\text{pr}(\mathbf{a}|I). \quad (8)$$

The posterior is therefore determined by the product of two probabilities, the *likelihood* $\text{pr}(D|\mathbf{a}, I)$ and the *prior* $\text{pr}(\mathbf{a}|I)$. The Bayesian procedure for estimating LECs \mathbf{a} proceeds by identifying a likelihood based on the available experimental data, assigning an appropriate prior, and finally analyzing the properties of the posterior for \mathbf{a} , including the dependence on the choice of prior. From the computed posterior, we can propagate the LEC uncertainties to the predictions of the EFT. We describe the practical implementation of this procedure in Sec. III. We pay particular attention to problems that can occur if the prior is not well chosen and develop a set of diagnostics by which such problems may be identified.

The unadorned least-squares minimization procedure with $\sigma_j = \sigma_{j,\text{exp}}$ follows from particular choices for the likelihood and prior. The pdf for N_d independent data $D = \{d_i\}$ given the LECs \mathbf{a} and relevant information I including the form and truncation order of the EFT is given by

$$\text{pr}(D|\mathbf{a}, I) = \prod_{j=1}^{N_d} \left(\frac{1}{\sqrt{2\pi}\sigma_{j,\text{exp}}} \right) e^{-\chi^2/2}, \quad (9)$$

with the χ^2 objective function defined in Eq. (4).³ Note that the theory itself, including the LECs and all other pertinent information, is on the right-hand side of the conditional here, i.e., it is given information. If we identify the “best-fit” parameters with the maximum of the posterior from Eq. (8) (maximum likelihood), then the Bayesian procedure is equivalent to least-squares minimization *if* we choose the prior $\text{pr}(\mathbf{a}|I)$ to be uniform (i.e., independent of \mathbf{a}) in regions where the likelihood is nonzero (or non-negligible). This equivalence does, though, blur the significant interpretative difference between $\text{pr}(D|\mathbf{a}, I)$ and $\text{pr}(\mathbf{a}|D, I)$. This difference in frequentist and Bayesian perspectives will be discussed in Sec. II D

³ The principle of maximum entropy (MaxEnt) [24] can be used to derive this likelihood in the Bayesian framework [19, 20]. In general we need a covariance matrix to account for correlated data.

One might think that a uniform prior is a natural, unbiased choice, but there are subtleties (e.g., in what variable it should be uniform and whether uniformity accounts for all available constraints) [20]. Furthermore, in the context of EFTs we know that the LECs should not have any unrestricted value, but instead can be expected to be distributed in some “natural” range around zero. Thus a uniform prior encodes information that is *not* expected, namely that parameters in any interval are equally likely. But this prior is the baseline for many analyses, and so we use it throughout as a benchmark. We will use the term “least-squares fit” interchangeably with maximizing the posterior of \mathbf{a} using a uniform prior. However, we stress again that in practice, especially in the EFT context, least-squares is often augmented with other criteria, such as those reviewed in Sec. II A.

C. Marginalization

The sum rule of probabilities says a pdf should be normalized, that is,

$$\int dx \text{pr}(x|I) = 1, \quad (10)$$

where the integration is over the appropriate range of x . The sum rule implies that for a joint pdf $\text{pr}(x, y|I)$ we find

$$\begin{aligned} \text{pr}(x|I) &= \int dy \text{pr}(x, y|I) \\ &= \int dy \text{pr}(x|y, I) \text{pr}(y|I), \end{aligned} \quad (11)$$

which is marginalization, with the product rule applied to obtain the second form.

This rule allows us to “integrate out” parameters from a joint distribution to obtain a pdf which depends only on the parameter(s) of interest. One can also “integrate in” parameters, expressing a pdf for some parameters in terms of integrals over joint distributions with other parameters. This is especially useful when the form of a pdf of interest is not known, but joint distributions with other parameters are known.

We use the rule of marginalization extensively in several different contexts.

1. Projection:

to quote LEC results and plot pdfs. We derive two-dimensional pdfs $\text{pr}(a_i, a_j|D, I)$ from the full joint posterior $\text{pr}(\mathbf{a}|D, I)$ by integrating over all the other parameters:

$$\begin{aligned} \text{pr}(a_i, a_j|D, I) \\ = \int \prod_{m=0, m \neq i, j}^k da_m \text{pr}(\mathbf{a}|D, I). \end{aligned} \quad (12)$$

Such marginalized posteriors for the model used in Refs. [11, 13] are shown in the plots below the diagonal in Fig. 1, where the x -axis scale is at the

bottom. (This model will be used in Sec. III to illustrate the general procedures and diagnostics.) The $\text{pr}(a_i, a_j|D, I)$ s are displayed as intensity plots, and so reveal the regions of high joint probability for pairs of LECs. Such projected posterior plots are therefore useful for diagnosing correlations between the LECs obtained from the analysis. Further details of their interpretation will be discussed in Sec. III B. Note that in this case, the posteriors are two-dimensional Gaussians, but in general they can be more complicated and even multi-modal.

One further marginalization step reduces the two-dimensional pdfs to one-dimensional pdfs for a particular LEC, i.e.

$$\text{pr}(a_i|D, I) = \int da_j \text{pr}(a_i, a_j|D, I). \quad (13)$$

These pdfs appear on the diagonal in Fig. 1. Here, the individual posteriors are simple Gaussians, but again, in general, they can be highly non-Gaussian.

2. **Higher orders:** to account for higher-order terms in the EFT expansion and avoid underfitting when extracting low-order LECs. Marginalization is a straightforward method to account for such higher-order terms whose coefficients may not be well-constrained by the data but can still be assumed to be natural. In particular, if we need to determine coefficients up to order k but wish to account for contributions up to order k_{\max} , we marginalize according to

$$\begin{aligned} \text{pr}(a_0, \dots, a_k|D, I) \\ = \int da_{k+1} \dots da_{k_{\max}} \text{pr}(\mathbf{a}|D, I). \end{aligned} \quad (14)$$

3. **Integrating in:** to describe pdfs of interest in terms of pdfs with known analytic forms. For example, the prior distribution $\text{pr}(\mathbf{a}|I)$ can be expressed in terms of a naturalness parameter, \bar{a}

$$\text{pr}(\mathbf{a}|I) = \int d\bar{a} \text{pr}(\mathbf{a}|\bar{a}, I) \text{pr}(\bar{a}|I), \quad (15)$$

where \bar{a} parameterizes the width of prior pdf for the LECs $\text{pr}(\mathbf{a}|\bar{a}, I)$. By introducing the marginalization integral over \bar{a} , the prior for the LECs is specified quantitatively by its width \bar{a} and can be calculated.

D. Interpretation

In Sec. II B we noted how the Bayesian procedure can yield the same maximum likelihood result obtained from a frequentist procedure. However, the interpretations are quite different. In the strict frequentist view, the only

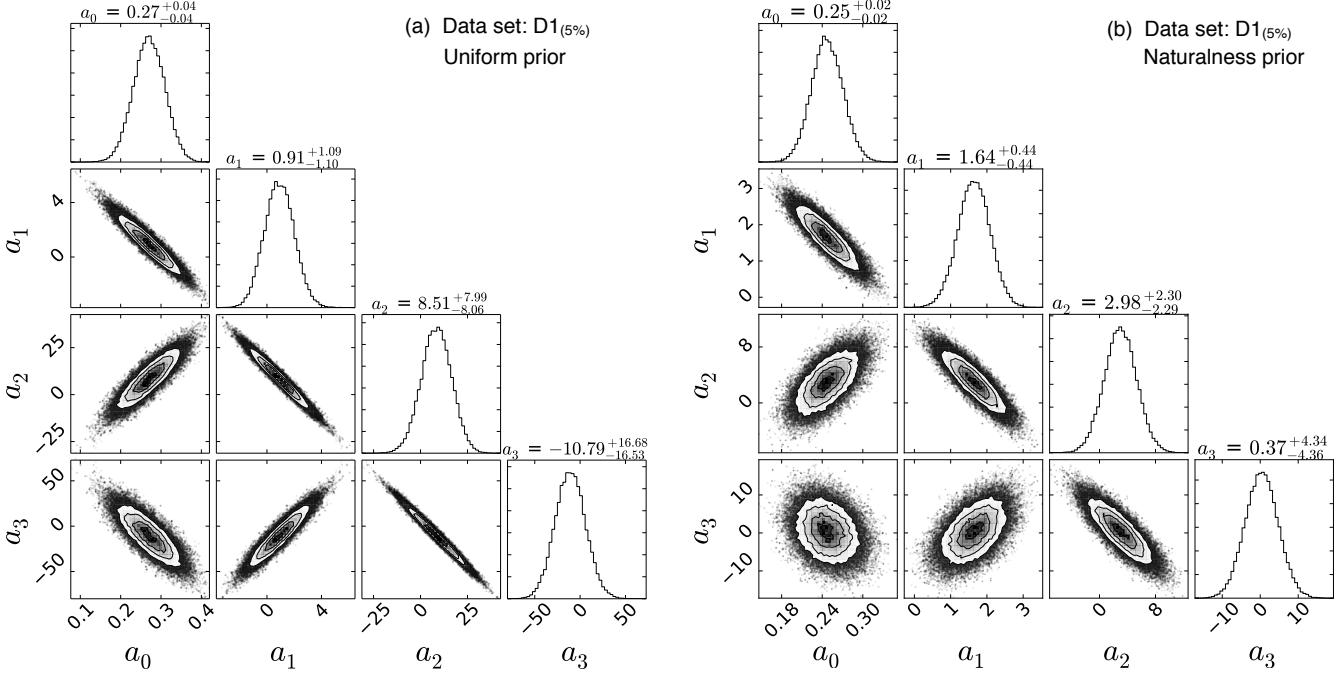


FIG. 1. Projected posterior plots [25] for Model D [the function defined in Eq. (23)] calculated at order $k = 3$, $k_{\max} = 3$ given data set $D1_{(5\%)}$, as described in Sec. III, with (a) a uniform prior, and (b) the Set C' prior with $\bar{a}_{\text{fix}} = 5$. The full posterior was calculated using Eq. (27). The plots on the diagonal are marginalized pdfs $\text{pr}(a_i|D, I)$ for the coefficient a_i on the x -axis, calculated according to Eq. (13). The lower triangle plots are two-dimensional pdfs $\text{pr}(a_i, a_j|D, I)$, calculated according to Eq. (12). The contours indicate 0.5, 1.0, 1.5, 2.0 sigma in a Gaussian approximation. The mean and 68% DoB interval are also given for each coefficient. Note the change of scale between (a) and (b).

valid interpretation of probability is as the relative frequency of an event out of a large ensemble of repeatable experiments [26]. Thus the outcomes of experiments are treated as random variables. Model parameters (e.g., the LECs), on the other hand, are fixed — they do not have a probability distribution, so the Bayesian posterior we seek has no meaning. The complete focus is on the likelihood. Thus, strictly speaking, a frequentist confidence interval (CI) exists only for data. It gives the probability that, for fixed parameters, a large set of experiments will yield results that fall in the interval $p\%$ of the time.

In the Bayesian interpretation, probabilities express the current state of knowledge about the parameters, which allows us to calculate Bayesian degree-of-belief (DoB) intervals for the true values of the parameters. These are established by integrations over posteriors for \mathbf{a} (or some subset) to determine the desired intervals. For example, a 68% DoB interval for a parameter a_i means that the particular input data and information lead to a probability of 68% that the true value of a_i is in that interval. The data are fixed while the parameters have a pdf.

The use of the word “belief” here is somewhat unfortunate, because it might suggest to some readers that results are subjective. And indeed, different priors can lead to different outcomes for the posterior pdfs—particularly in the case where data are sparse or of low quality. But,

the Bayesian framework is completely rigorous once the prior is specified. It permits a careful tracing of assumptions through to probabilistic consequences for parameters. In the case of EFT the assignment and interpretation of the prior should not be a source of controversy: it merely encodes *a priori* information about the parameter of interest. Furthermore, the prior can also represent knowledge from previous determinations or measurements. The use of Bayesian priors thus provides a quantifiable way to include guidance from both theory and previous experiments. It makes explicit the assumptions that are inherent in EFT parameter estimation, but are usually left largely implicit. We also mention in this context that, given specified pieces of “testable information” for the parameters, the Method of Maximum Entropy (MaxEnt) [24] can be used to derive the least informative priors based on the known information.

The interplay of likelihood and prior in limiting cases embodies desired outcomes that otherwise have to be imposed by hand. If the data are numerous and of good quality, a highly peaked likelihood will be the dominant contribution to the posterior if the prior is not too restrictive [20]. On the other hand, if the likelihood for a parameter is not well-constrained by the data, the prior will have the dominant effect. In the case of LEC parameter estimates, a naturalness prior can restrict the effect of parameters that are not well-constrained by data

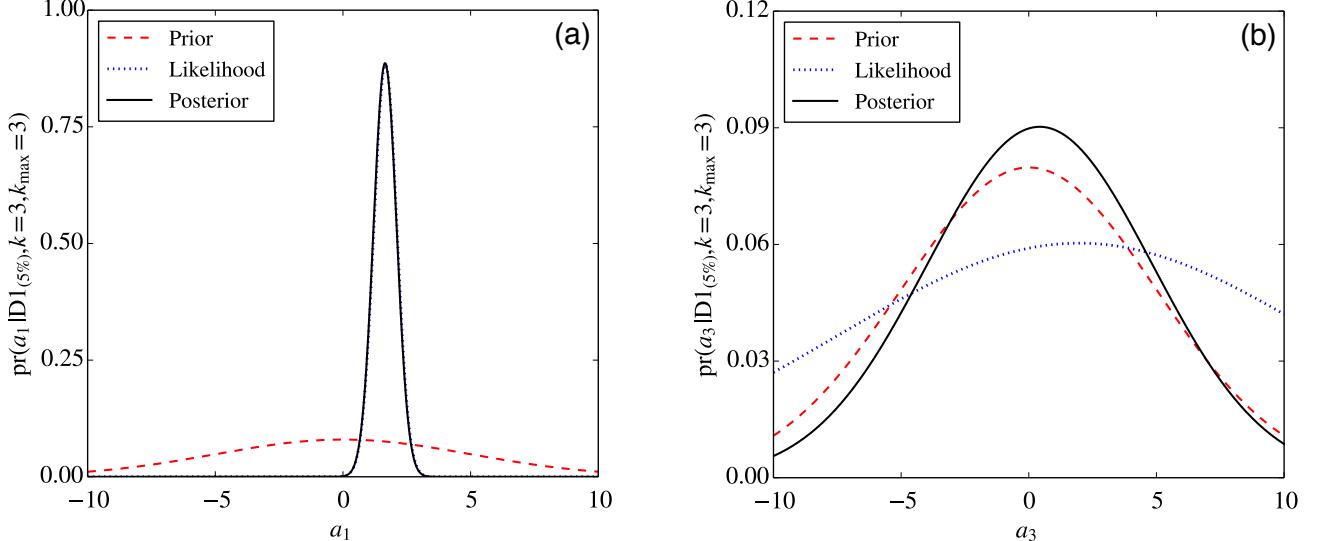


FIG. 2. (color online) Marginalized posteriors for two coefficients calculated at $k = 3$, $k_{\max} = 3$ given data set $D1_{(5\%)}$, which is described in Sec. III, along with the likelihood and prior pdfs in each case. For parameter a_1 in (a), the posterior is dominated by the likelihood (causing the two to overlap in the plot) and the prior is essentially uniform. For parameter a_3 in (b), the likelihood is widely spread and the posterior is mostly given by the prior.

by suppressing contributions from regions of parameter space driven by spurious correlations.

An example of a parameter that is well-determined by data is shown in Fig. 2(a), where the prior does not affect the posterior for that parameter because it is essentially uniform in the region where the likelihood has support. Figure 2(b) shows an example where the posterior largely reflects the input prior, since the parameter is not strongly localized by the slowly varying likelihood. We refer to this latter situation as the posterior “returning the prior”. These are marginalized posterior distributions from the same joint pdf $\text{pr}(\mathbf{a}|D, I)$, which means that even though the prior does not seem to play a role in Fig. 2(a), this is because it has restricted the domain of the parameter in Fig. 2(b). Thus the first parameter is insensitive, but only because the prior prevented other, higher-order, parameters from contributing in unnatural regions of parameter space. An example of a spurious correlation can be seen in Fig. 1(a), where a_0 is correlated with the ill-determined cubic coefficient a_3 . The amelioration of this problem is then seen in Fig. 1(b), where the correlation is almost completely removed when an appropriate naturalness prior is applied.

E. Bayes evidence

In this section we explain how Bayesian methods facilitate a quantitative approach to *model selection*: which model from a set of models is preferred by the data? Our presentation draws on material from Refs. [19, 20, 27].

We should first clarify what we mean by “model” in this context. After all, an EFT is said to be model-

independent because the theory is constructed in the most general form consistent with the symmetries of the underlying interaction. We use “model” here to mean any theoretical construct for computing an observable. Examples of quantitative Bayesian comparisons between EFT models could be the predictive power of EFTs with different degrees of freedom for the same data set, or an analysis of the increase in predictive power for observables using different orders in the same EFT.

For two models M_1 and M_2 and the same data set D , the relevant pdfs to compare are the posteriors for each model given the data, called the model evidence: $\text{pr}(M_1|D, I)$ and $\text{pr}(M_2|D, I)$. Note that there is no reference to a particular parameter set here; the comparison is not between two fits but between two models. We can use Bayes’ theorem to express the evidence ratio for the two models:

$$\frac{\text{pr}(M_2|D, I)}{\text{pr}(M_1|D, I)} = \frac{\text{pr}(D|M_2, I) \text{pr}(M_2|I)}{\text{pr}(D|M_1, I) \text{pr}(M_1|I)}, \quad (16)$$

where the common factor of $\text{pr}(D|I)$ cancels in the ratio. If neither model is *a priori* more likely, the ratio $\text{pr}(M_2|I)/\text{pr}(M_1|I) = 1$ so that the evidence ratio becomes equal to the *Bayes factor*

$$\frac{\text{pr}(M_2|D, I)}{\text{pr}(M_1|D, I)} \rightarrow \frac{\text{pr}(D|M_2, I)}{\text{pr}(D|M_1, I)}. \quad (17)$$

To express the right-hand-side in terms of the parameters of each model, we apply the marginalization rule so that

$$\frac{\text{pr}(D|M_2, I)}{\text{pr}(D|M_1, I)} = \frac{\int d\mathbf{a}_2 \text{pr}(D|\mathbf{a}_2, M_2, I) \text{pr}(\mathbf{a}_2|M_2, I)}{\int d\mathbf{a}_1 \text{pr}(D|\mathbf{a}_1, M_1, I) \text{pr}(\mathbf{a}_1|M_1, I)}. \quad (18)$$

Thus we integrate over the entire parameter space. The evidence ratio provides a quantitative diagnostic to assess the effectiveness of EFTs with different degrees of freedom (e.g., comparing χ EFT with and without explicit Δ -resonances [28]) or distinguishing between different power-counting schemes.

Here we focus on using the evidence ratio to assess whether adding another order to our EFT is favored by the given data. Thus we consider the case where $M_1 \rightarrow M_k$ and $M_2 \rightarrow M_{k+1}$ are the same EFT but evaluated at successive orders in the EFT expansion, and that neither is more likely *a priori* so that the evidence ratio for the two models is equal to the Bayes factor as in Eq. (17). (We will continue to use \mathbf{a}_1 and \mathbf{a}_2 to denote the corresponding vectors of coefficients.)

For simplicity, let M_{k+1} have one additional parameter a' and assume that the priors for the parameters factor into individual priors, so

$$\begin{aligned} \text{pr}(\mathbf{a}_2|M_{k+1}, I) &\equiv \text{pr}(\mathbf{a}_1, a'|M_{k+1}, I) \\ &= \text{pr}(\mathbf{a}_1|M_{k+1}, I) \text{pr}(a'|M_{k+1}, I). \end{aligned} \quad (19)$$

If the values of a' that contribute to the integrand in the numerator of Eq. (18) are determined predominantly by the likelihood (for the values of \mathbf{a}_1 where the integrand is large), and that likelihood is sharply peaked, then we can approximate the impact of the prior by taking it to be roughly constant with an effective width $\Delta a'$, so $\text{pr}(a'|M_{k+1}, I) \approx 1/\Delta a'$. We then approximate the integral over a' as the value at the likelihood peak \hat{a}' times the width of that peak, $\delta a'$. This yields the Bayes factor

$$\begin{aligned} \frac{\text{pr}(D|M_{k+1}, I)}{\text{pr}(D|M_k, I)} &\approx \frac{\delta a'}{\Delta a'} \int d\mathbf{a}_1 \text{pr}(D|\mathbf{a}_1, \hat{a}', M_{k+1}, I) \\ &\quad \times \text{pr}(\mathbf{a}_1|M_{k+1}, I) \\ &\quad \times \left[\int d\mathbf{a}_1 \text{pr}(D|\mathbf{a}_1, M_k, I) \text{pr}(\mathbf{a}_1|M_k, I) \right]^{-1}. \end{aligned} \quad (20)$$

The ratio of \mathbf{a}_1 integrals represents the gain in the likelihood from having this extra parameter with value \hat{a}' . But the ratio is also multiplied by the “Occam factor” or “Occam penalty” $\delta a'/\Delta a'$, which is the factor by which the parameter space collapses when the data are taken into account (this is large when the parameters can range over a large domain compared to the restriction imposed by the data).

Equation (20) tells us that we will favor the more complicated model only if the likelihood gain is larger than this factor. We are assuming that M_k is contained within M_{k+1} , i.e., M_k is M_{k+1} with $a' = 0$, which means this likelihood gain is always greater than or equal to one. In a situation such as Fig. 2(a), we have a peaked likelihood compared to the prior pdf, which yields a substantial Occam penalty. However, the likelihood gain here is much greater than that penalty, because it involves the likelihood peak at $\hat{a}' > 0$ in comparison to the likelihood for $a' = 0$, and the latter is very small. That means that the

evidence ratio for the addition of a' is $\gg 1$, and the inclusion of this parameter is therefore highly favored. This leads us to expect that if we calculate the evidence for the data $\text{pr}(D|M_k, I)$ as a function of the order k (which we will henceforth call the “evidence” for brevity) that function will increase as long as additional parameters are leading to a marked decrease in the χ^2 . However, if the likelihood function at $a' = 0$ is not well below the peak value (e.g., its width is greater than in this example, or the peak is shifted toward zero), than the Occam penalty can dominate and disfavor an additional parameter, which is an implementation of Occam’s razor. In such a situation the evidence will decrease from order k to order $k + 1$. We would then expect the evidence function $\text{pr}(D|M_k, I)$ to show a peak.

A different behavior can be expected if the comparison of likelihood and prior as in Fig. 2(b) holds, in which case the analysis of Eq. (18) is inverted. That is, we now say that the dependence on a' in the likelihood can be replaced by a constant \hat{a}' and the likelihood pulled out of the integral of a' :

$$\begin{aligned} \frac{\text{pr}(D|M_{k+1}, I)}{\text{pr}(D|M_k, I)} &\approx \int d\mathbf{a}_1 \text{pr}(D|\mathbf{a}_1, \hat{a}', M_{k+1}, I) \\ &\quad \times \text{pr}(\mathbf{a}_1|M_{k+1}, I) \int da' \text{pr}(a'|M_{k+1}, I) \\ &\quad \times \left[\int d\mathbf{a}_1 \text{pr}(D|\mathbf{a}_1, M_k, I) \text{pr}(\mathbf{a}_1|M_k, I) \right]^{-1}. \end{aligned} \quad (21)$$

The integral over a' is now just a normalization integral, equal to one. We are dominated by the prior, so $\hat{a}' \approx 0$ and the Bayes ratio is one, because M_{k+1} with the last coefficient equal to zero is simply M_k . The same argument goes through for each higher value of k , meaning that we have *saturation* rather than a peak for $\text{pr}(D|M_{k+1}, I)$ as a function of the order k [27]. In summary, the naturalness prior cuts down the “wasted” parameter space that might be ruled out by the data and which leads to an Occam penalty. This means that it limits the “phase space” of an EFT, which is therefore a simpler model (in the model selection sense) than the same functional form with unconstrained or only weakly constrained LECs.

F. Sampling with MCMC

While in some situations it is possible to evaluate posterior pdfs analytically, in many cases we must resort to numerical methods. Although many of the integrals we confront in this present work do not require its use, we employ Markov Chain Monte Carlo (MCMC) sampling methods to obtain posterior pdfs, since they are easily generalized to cases with more complicated probability distributions.

The MCMC algorithm generates N samples $\{a_j\}$ according to the posterior probability distribution $\text{pr}(\mathbf{a}|D, I)$. Expectation integrals may then be performed

using those samples:

$$\langle f(\mathbf{a}) \rangle = \int d\mathbf{a} \text{pr}(\mathbf{a}|D, I) f(\mathbf{a}) \approx \frac{1}{N} \sum_{j=1}^N f(\mathbf{a}_j). \quad (22)$$

The result accounts for all correlations between the parameters. The propagation of those correlations to the function f can be simplified if the posterior is well approximated as a correlated Gaussian; in that case the correlation matrix can be employed for this purpose. But Eq. (22) achieves this task independent of whether $\text{pr}(\mathbf{a}|D, I)$ is Gaussian or not. Marginalization integrals over parameters [see, e.g., Eq. (12)] can also be performed trivially by retaining only samples in the parameters of interest. The samples $\{\mathbf{a}_j\}$ are constructed via MCMC, in a particular implementation called emcee [29]. Implementation details can be found in Appendix A.

III. DIAGNOSTICS AND PROCEDURES FOR PARAMETER ESTIMATION

A. Set up of a test case

In this section we present a suite of diagnostics and associated procedures for EFT parameter estimation. For continuity, we use for illustration a model problem explored in previous work [11, 13], which is to use a given data set to extract as many coefficients in a particular function's Taylor expansion as possible. For the function we choose, in this section [13]:

$$g(x) = \left(\frac{1}{2} + \tan\left(\frac{\pi}{2}x\right) \right)^2 = 0.25 + 1.57x + 2.47x^2 + 1.29x^3 + \dots \quad (23)$$

The function is designed to have coefficients that are $\mathcal{O}(1)$ up to about tenth order, and the singularity at $x = 1$ means this is where the simulated EFT expansion breaks down. We follow the previous work and consider 10 equally spaced data points covering the range $0 < x \leq 1/\pi$, each with a data error of 5% [thus $c = 0.05$ in Eq. (2)]. The data set is enumerated in Appendix C of Ref. [13] and is also available as supplementary material to this paper. To conform to later usage, we refer to this as data set D1_(5%), where “D” indicates the function defined by Eq. (23), the “1” labels a choice of x points where the function was sampled, and the subscript indicates the data error. Other models and the nucleon mass expansion, with varied data ranges, precision, and numbers of points, will be considered in Secs. IV and V.

Now that we have data for a simulated EFT, the first step in our analysis is to select prior pdfs for the coefficients. Some possible sets of priors encoding naturalness

assumptions are listed in Table I. Prior Sets A, B, and C were used in Ref. [8]. The assumption for the coefficients themselves (the a_i s) in Sets A and B correspond to a flat distribution bounded by a maximum, \bar{a} . For Set A, this naturalness parameter \bar{a} is itself subject to a scale-invariant uniform (Jeffreys) prior [13] between the limits $\bar{a}_<$ and $\bar{a}_>$, while Set B has the naturalness parameter \bar{a} following a log-normal distribution of width σ . Set C assumes a Gaussian prior distribution for the a_i s, with the width \bar{a} distributed according to the Jeffreys prior. In general, arguments can be made to support any of these assumptions, which means we need to carefully check the sensitivity of our results to different choices.

For most of our examples, we choose a simpler variation of Set C, labeled Set C', in which we adopt the Gaussian prior on the coefficients $\mathbf{a} \equiv \{a_0, \dots, a_k\}$,

$$\text{pr}(\mathbf{a}|\bar{a}, I) = \left(\frac{1}{\sqrt{2\pi}\bar{a}} \right)^{k+1} \exp\left(-\frac{\mathbf{a}^2}{2\bar{a}^2}\right), \quad (24)$$

where I will include the order of the model and the marginalization order, and fix the value of the naturalness parameter at \bar{a}_{fix} :

$$\text{pr}(\bar{a}) = \delta(\bar{a} - \bar{a}_{\text{fix}}), \quad (25)$$

so that the marginalization over \bar{a} as in Eq. (15) picks out this value. The impact of the choice of a particular \bar{a}_{fix} can be anticipated by consulting the \bar{a} posterior plot described below.

We seek the posterior for the parameters \mathbf{a} for the model EFT expansion up to order k , but we will marginalize over coefficients as in Sec. II C, up to order k_{max} where we define the coefficients $\mathbf{a}_{\text{marg}} \equiv \{a_{k+1}, \dots, a_{k_{\text{max}}}\}$; we designate this posterior as $\text{pr}(\mathbf{a}|D, k, k_{\text{max}})$. We can calculate this using Eqs. (14)–(15), and Bayes' theorem:

TABLE I. Examples of prior pdfs encoding various naturalness assumptions.

set	$\text{pr}(a_i \bar{a})$	$\text{pr}(\bar{a})$
A	$\frac{1}{2\bar{a}} \theta(\bar{a} - a_i)$	$\frac{1}{\ln \bar{a}_>/\bar{a}_<} \frac{1}{\bar{a}} \theta(\bar{a} - \bar{a}_<) \theta(\bar{a}_> - \bar{a})$
A'	$\frac{1}{2\bar{a}} \theta(\bar{a} - a_i)$	$\delta(\bar{a} - \bar{a}_{\text{fix}})$
B	$\frac{1}{2\bar{a}} \theta(\bar{a} - a_i)$	$\frac{1}{\sqrt{2\pi}\bar{a}\sigma} e^{-(\log \bar{a})^2/2\sigma^2}$
C	$\frac{1}{\sqrt{2\pi}\bar{a}} e^{-a_i^2/2\bar{a}^2}$	$\frac{1}{\ln \bar{a}_>/\bar{a}_<} \frac{1}{\bar{a}} \theta(\bar{a} - \bar{a}_<) \theta(\bar{a}_> - \bar{a})$
C'	$\frac{1}{\sqrt{2\pi}\bar{a}} e^{-a_i^2/2\bar{a}^2}$	$\delta(\bar{a} - \bar{a}_{\text{fix}})$

TABLE II. Diagnostic tools for parameter estimation. Figure numbers in Secs. II and III are given for each type of plot. Additional examples of these figures can be found in Secs. IV and V.

Name	Fig.	Description and uses
projected posterior plot	1	Matrix of subplots with posteriors from marginalizing $\text{pr}(\mathbf{a} D, I)$ over all but one a_i (diagonal) or all but a pair a_i, a_j (lower triangle). Compare different priors; visualize correlations; identify when posterior \approx prior.
\bar{a} posterior	6	$\text{pr}(\bar{a} D, k, k_{\max})$. Show weighting of \bar{a} values when marginalized; identify appropriate marginalization range for \bar{a} ; signals unnatural coefficients.
\bar{a} relaxation plot	7	Evolution of marginalized posterior for single coefficients for full range of \bar{a} . Check whether \bar{a} is too restrictive (rapid changes with \bar{a}); identify regions insensitive to \bar{a} to help identify marginalization range; show transition to least-squares result.
evidence vs. k	8	$\text{pr}(D k, \bar{a}) \propto \text{pr}(k D, \bar{a})$ (if uniform prior on k). Show transition from dominance by likelihood [e.g., Fig. 2(a)] to dominance by prior [e.g., Fig. 2(b)].
x_{\max} plot	9	Evolution of marginalized posterior for single coefficients, $\text{pr}(a_i D, I)$, as data range specified by x_{\max} is increased. Check for stability with data range.
multi-set plot	10(a)	Evolution of marginalized posterior for single coefficients using multiple equal-sized sets of data, each over the same range.
accumulation plot	10(b)	Comparison of marginalized posteriors for single coefficients as data are accumulated from combining multiple equal-sized sets of data.
residual plot	12	Log-log plot of residuals versus expansion parameter x . Check for power-law behavior at different orders. Test for data or theory error dominance in different regions of x .

$$\begin{aligned} \text{pr}(\mathbf{a}|D, k, k_{\max}) &= \int da_{k+1} \cdots da_{k_{\max}} \text{pr}(\mathbf{a}, a_{k+1}, \dots, a_{k_{\max}}|D, k, k_{\max}) \\ &= \int d\mathbf{a}_{\text{marg}} \frac{\text{pr}(D|\mathbf{a}, \mathbf{a}_{\text{marg}}, k, k_{\max}) \text{pr}(\mathbf{a}, \mathbf{a}_{\text{marg}}|k, k_{\max})}{\text{pr}(D|k, k_{\max})} \\ &= \int d\mathbf{a}_{\text{marg}} \int d\bar{a} \frac{\text{pr}(D|\mathbf{a}, \mathbf{a}_{\text{marg}}, k, k_{\max}) \text{pr}(\mathbf{a}, \mathbf{a}_{\text{marg}}|\bar{a}, k, k_{\max}) \text{pr}(\bar{a})}{\text{pr}(D|k, k_{\max})}, \end{aligned} \quad (26)$$

where we have assumed that the prior for the naturalness parameter is independent of the truncation and marginalization orders. Substituting the prior from Eq. (25), we have

$$\text{pr}(\mathbf{a}|D, k, k_{\max}) = \int d\mathbf{a}_{\text{marg}} \frac{\text{pr}(D|\mathbf{a}, \mathbf{a}_{\text{marg}}, k, k_{\max}) \text{pr}(\mathbf{a}, \mathbf{a}_{\text{marg}}|\bar{a}_{\text{fix}}, k, k_{\max})}{\text{pr}(D|k, k_{\max})}, \quad (27)$$

The likelihood $\text{pr}(D|\mathbf{a}, \mathbf{a}_{\text{marg}}, k, k_{\max})$ for this problem is simply the least-squares likelihood from Eq. (9) calculated with *all* the coefficients $\{\mathbf{a}, \mathbf{a}_{\text{marg}}\}$ up to order k_{\max} . For Set C' the joint prior $\text{pr}(\mathbf{a}, \mathbf{a}_{\text{marg}}|\bar{a}_{\text{fix}}, k, k_{\max})$ is simply the product of Gaussian priors with width \bar{a}_{fix} in Eq. (24) for all the coefficients up to order k_{\max} .

Note on k and k_{\max} . The posterior pdf in Eq. (26) is the pdf for the coefficients \mathbf{a} up to some order k , having accounted for the higher-order coefficients $\{a_{k+1}, \dots, a_{k_{\max}}\}$. Had we computed the posterior for *all* the coefficients up to order k_{\max} and then performed the marginalization integral to generate the projected posterior plot up to order k , the results would be the same. Thus in the examples in this paper, where the posterior has the form of Eq. (26), the parameter estimation will be completely controlled by k_{\max} . The order $k \leq k_{\max}$ controls the actual calculation of observables, for example as in Fig. 12.

Alternative accounting of higher-order effects. As noted in Sec. I, explicitly calculating the observables appearing in the likelihood up to order k_{\max} may not be feasible—yet we still need to account for these contributions. To do so, we build on Ref. [8] and use the expected naturalness of the expansion coefficients for observables at higher order. That is, the contribution to an observable beyond order k is given by the truncation error (which may also have an overall scale [8])

$$\Delta_k(x) = \sum_{i=k+1}^{k_{\max}} c_i x^i, \quad (28)$$

and we introduce a naturalness prior for the c_i s. For the linear model observables introduced in Eq. (3), the c_i s here are just the a_i s. The generalization of Eq. (26) where we marginalize over $\mathbf{c}_{\text{marg}} \equiv \{c_{k+1}, \dots, c_{k_{\max}}\}$ is

$$\text{pr}(\mathbf{a}|D, k, k_{\max}) = \int d\mathbf{c}_{\text{marg}} \frac{\text{pr}(D|\mathbf{a}, \mathbf{c}_{\text{marg}}, k, k_{\max}) \text{pr}(\mathbf{a}|k, k_{\max}) \text{pr}(\mathbf{c}_{\text{marg}}|k, k_{\max})}{\text{pr}(D|k, k_{\max})}, \quad (29)$$

where we have assumed statistical independence of the coefficients \mathbf{a} from each other and from higher-order corrections, so that the prior pdf $\text{pr}(\mathbf{a}|\mathbf{c}_{\text{marg}}, k, k_{\max}) = \text{pr}(\mathbf{a}|k, k_{\max})$. The posterior now depends on where the line is drawn (at k) between higher-order corrections and the LECs we want to compute. Thus, in contrast to the case just described for Eq. (26), it is possible to have posteriors for \mathbf{a} up to order k that are not controlled just by k_{\max} . The two prior pdfs for the \mathbf{c}_{marg} and the \mathbf{a} coefficients will now encode our naturalness assumptions. Reference [8] contains an extensive analysis of different naturalness priors for observable coefficients. The likelihood in Eq. (29) will not in general be a simple least-squares likelihood, because the systematic error in the calculation given in Eq. (28) cannot simply be added in quadrature to the experimental error [11, 18]. This alternative accounting will not be used in this paper, but will be explored in future work.

When we are not marginalizing over higher orders and $k_{\max} = k$, Eq. (27) can be computed analytically as an augmented χ^2 , which was derived by Schindler and Phillips in Ref. [13]. When $k_{\max} > k$, the marginalization over the higher-order coefficients can be understood as accounting for the correlated systematic errors from leaving out higher-order terms. The posterior in this case is then a *modified*, augmented χ^2 [11] which was derived in Refs. [13, 18]. As a benchmark, we compare results obtained for estimates of \mathbf{a} using prior Set C' from Eq.(27) with the results of least-squares fits (equivalent to choosing a uniform prior instead) using our test data set D1(5%).

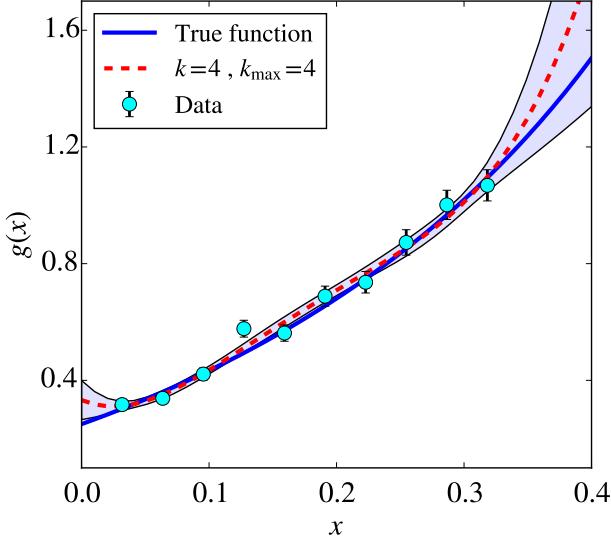


FIG. 3. (color online) Comparison of data set D1(5%), the underlying function for Model D from Eq. (23), and a least-squares prediction calculated at order $k = 4$, $k_{\max} = 4$ from that data set. The error bands represent 1- σ (68% DoBs).

In the next subsections we analyze and interpret $\text{pr}(\mathbf{a}|D, k, k_{\max})$ using the set of diagnostics summarized in Table II. Since the posterior is simple to compute for this linear model problem and prior Set C', we can easily check computations using MCMC sampling against the analytic results. Marginalization in more complicated situations such as nonlinear models will be discussed in

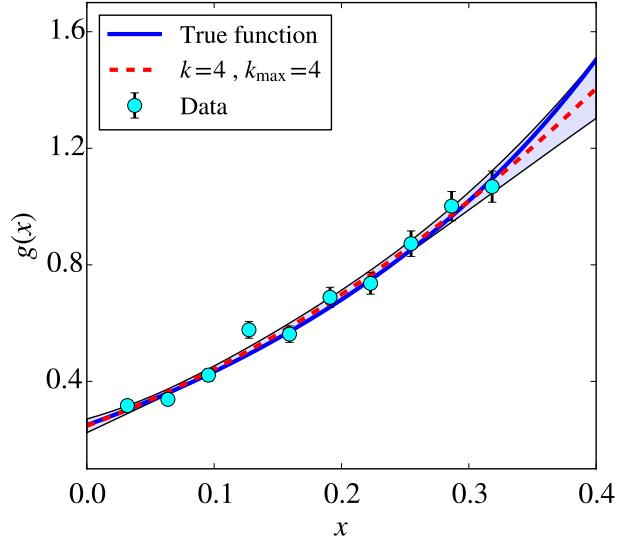


FIG. 4. (color online) Comparison of data set D1(5%), the underlying function for Model D from Eq. (23), and a Bayesian prediction calculated at order $k = 4$, $k_{\max} = 4$ using prior Set C' with $\bar{a}_{\text{fix}} = 5$, from that data set. The error bands represent 1- σ (68% DoBs).

Sec. V. For Model D and both priors in Fig. 1 we are sampling from a posterior which is a $(k+1)$ -dimensional Gaussian parametrized by the LEC means and covariance matrix. We therefore use the LEC means and 1- σ error bands (which are equivalent to 68% DoB intervals since the posterior is Gaussian) in the subsequent diagnostic plots, but remind the reader that in general the posterior will not be Gaussian, and that many distributions will need to be treated more carefully before resorting to a covariance analysis.

B. Coefficient estimates and correlations

The marginalized posterior pdfs computed for $k = 3$, $k_{\max} = 3$, are shown in Figs. 1(a) and 1(b) respectively for the uniform prior and for the prior Set C' with $\bar{a}_{\text{fix}} = 5$. Projected posterior plots as in Fig. 1 help to visualize the correlations among the extracted param-

TABLE III. Coefficient estimates from sampling of $\text{pr}(\mathbf{a}|D_{1(5\%)}, k, k_{\max})$ given the expansion from Eq. (3) at different orders (these results are controlled by k_{\max} only, see Sec. III A). The left side of the table is for a uniform prior, which is equivalent to a least-squares fit, and includes the χ^2/dof values. The right side of the table is using prior Set C' from Table I with $\bar{a}_{\text{fix}} = 5$, and includes the evidence $\text{pr}(D_{1(5\%)}|k, k_{\max})$. For both priors the posterior pdf is a multi-dimensional Gaussian.

		Uniform prior			Gaussian prior				
k	k_{\max}	χ^2/dof	a_0	a_1	a_2	Evidence	a_0	a_1	a_2
0	0	67	0.48 ± 0.01			~ 0	0.48 ± 0.01		
1	1	2.2	0.20 ± 0.01	2.6 ± 0.1		6.0×10^2	0.20 ± 0.01	2.6 ± 0.1	
2	2	1.6	0.25 ± 0.02	1.6 ± 0.4	3.3 ± 1.3	3.3×10^3	0.25 ± 0.02	1.6 ± 0.4	3.1 ± 1
2	3	1.9	0.27 ± 0.04	1.0 ± 1	8.1 ± 8.0	2.9×10^3	0.25 ± 0.02	1.7 ± 0.5	3.0 ± 2
2	4	2.0	0.33 ± 0.07	-1.9 ± 3	45 ± 30	2.8×10^3	0.25 ± 0.02	1.7 ± 0.5	3.0 ± 2
2	5	1.4	0.57 ± 0.1	-15 ± 7	280 ± 100	2.8×10^3	0.25 ± 0.02	1.7 ± 0.5	3.0 ± 2
2	6	1.9	0.59 ± 0.3	-16 ± 20	310 ± 400	2.8×10^3	0.25 ± 0.02	1.7 ± 0.5	3.0 ± 2
True values		0.25	1.57	2.47			0.25	1.57	2.47

ters, and are particularly useful for comparing the effects of different priors (including a uniform prior). For example, it is evident that for a_3 in Fig. 1(b) the marginalized one-dimensional posterior (on the diagonal) is simply returning the Gaussian prior of width \bar{a}_{fix} [the case illustrated in Fig. 2(b)]; in Fig. 1(a) the same posterior, while still Gaussian, is very wide. This leads to overfitting in the uniform-prior case, as a_3 is able to play off against other coefficients to push the maximum of the likelihood to unnatural regions of parameter space.⁴ Note also that the parameters a_0 and a_3 are strongly correlated without the prior but become largely uncorrelated with the prior. This trend of decoupling low-order coefficients from poorly determined high-order ones via application of a prior continues as k is increased. In the case of the uniform prior, the overfitting becomes more pronounced as k is increased.

Figure 3 shows the results of a least-squares prediction where overfitting occurs at $k = k_{\max} = 4$. The $1-\sigma$ error bands indicate that the leading behavior is poorly estimated, and that the fit is fine-tuned to reproduce the data where the error band is smaller. Figure 4 shows the same prediction made using Bayesian parameter estimates with the Gaussian naturalness prior. The true leading-order behavior (manifested at low x) and the prediction for x above the last data point are much better reproduced, indicating that overfitting is avoided.

Table III shows the central values and 68% intervals obtained for the three leading coefficients, a_0 , a_1 , and a_2 from both the least-squares posterior pdf (left side) and the posterior pdf corresponding to the naturalness prior (right side). Results are shown for fits at different k_{\max} up to $k_{\max} = 6$. The parameter estimates are obtained by MCMC integration and are consistent with the exact

results in Tables 1 and 4 in Ref. [13]. The details of the exact calculations for these simple priors can be found in Ref. [13]. When computing the parameter estimates with MCMC, the error bars for the estimates were obtained from the ensemble of samples (see Sec. II F). These results for the parameter estimates with their error bars are presented with a precision that is a conservative estimate of the error of the MCMC method itself; by increasing the number of samples in these calculations, the estimates will approach the exact ones in Ref. [13]. We note that Table 4 of Ref. [13] contains an error: the logarithm of the posterior at the maximum parameter values quoted there for the Bayesian fit should be shifted down by one row. This mistake is corrected in the calculation that produced Table III.

Our results confirm the finding of Ref. [13]: as the number of parameters is increased, the least-squares estimates of the leading coefficients degrade. As seen in Fig. 3 for the $k = k_{\max} = 4$ prediction, overfitting occurs as the coefficients become correlated in order to reproduce the data. This leads to large errors outside the fit region: compare the quality of the predictions there in Figs. 3 and 4.

The left side of the table uses the χ^2/dof to indicate the quality of the fit, while the right side gives the evidence $\text{pr}(D|k, k_{\max})$ to show the relative extent to which the model describes the data (see Sec. III D). The evidence here is controlled by k_{\max} , as discussed in Sec. III D, and as we increase k_{\max} , the naturalness prior prevents fine-tuning by restricting the correlations between the leading coefficients and higher-order ones as we saw in the lower left of Fig. 1(b). For any value of k_{\max} , the first two coefficients ($k = 1$) are reliably extracted, with the posterior indicating when a coefficient is not well-determined by the data.

In Fig. 5, we show a corresponding projected posterior plot to Fig. 1(b) but using prior Set A' (see Table I) with $\bar{a}_{\text{fix}} = 5$. The marginalized posteriors on the diagonal for the higher-order coefficients become quite skewed and highly non-Gaussian; this would significantly com-

⁴ Indeed, we generically find problems with the least-squares extraction of a a_k in a k th-order fit. Presumably this is because a_{k+1} , with which a_k is highly anti-correlated, is artificially forced to zero in such a fit.

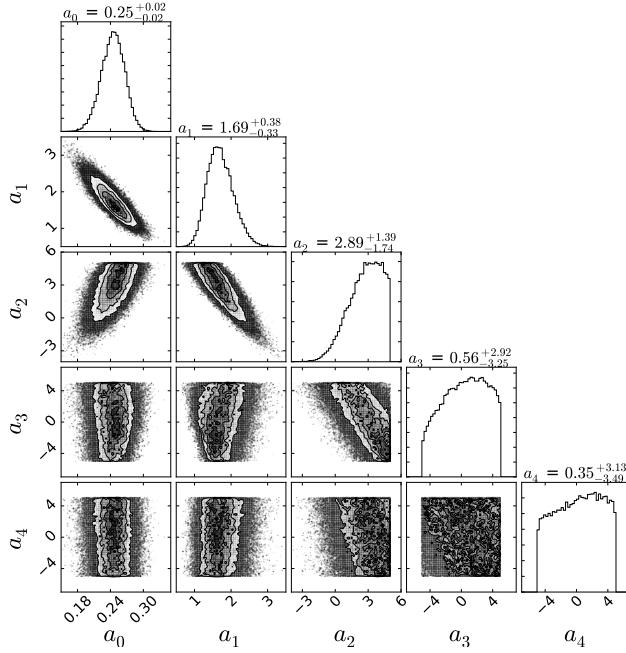


FIG. 5. Projected posterior plot (see Fig. 1) calculated at order $k = 4$, $k_{\max} = 4$ given data set $D1_{(5\%)}$ using Prior A' with $\bar{a}_{\text{fix}} = 5$.

plicate the determination of DoB intervals. However, the posteriors for these coefficients are primarily returning the prior—which in this case is flat. For the lowest-

This quantity can be expressed using marginalization and Bayes' theorem as

$$\text{pr}(\bar{a}|D, k, k_{\max}) = \frac{1}{\text{pr}(D|k, k_{\max})} \int d\mathbf{a} \int d\mathbf{a}_{\text{marg}} \text{pr}(D|\mathbf{a}, \mathbf{a}_{\text{marg}}, k, k_{\max}) \text{pr}(\mathbf{a}, \mathbf{a}_{\text{marg}}|\bar{a}, k, k_{\max}) \text{pr}(\bar{a}). \quad (30)$$

Note that this posterior can be obtained for different prior assumptions.

The posterior $\text{pr}(\bar{a}|D1_{(5\%)}, k = 3, k_{\max} = 3)$ is shown as a representative example in Fig. 6, in the case where the prior is Set C. The plot shows the region most likely for \bar{a} , which in this case implies natural coefficients.

Another way to check the sensitivity of results to the naturalness parameter is to examine the parameter estimates when \bar{a} is fixed to various values \bar{a}_{fix} . Hence we now employ prior Set C' and vary \bar{a}_{fix} from a small value to one large enough that the prior is effectively uniform. We use such results to create an \bar{a} relaxation plot (Fig. 7), which shows the projected posterior mean and width of different parameters as a function of a fixed value \bar{a}_{fix} for \bar{a} . \bar{a}_{fix} sets the width of the allowed region for the EFT parameters: Fig. 7 shows that too-small values of \bar{a} (e.g., of order 0.1) will bias the posterior severely, while large values relax to the least-squares (uniform prior) result.

Such “ \bar{a} relaxation” plots should be interpreted with care though, because a marginalization over \bar{a} when computing the coefficient posterior will weight different re-

order coefficients, the posteriors are Gaussian-like and their modes and widths agree with those from Fig. 1(b) and the corresponding values listed in Table III up to $k = 1$ when $k_{\max} = 4$. We recover about the same two-parameter correlation plot for a_0-a_1 as before, while those involving higher-order coefficients reflect details of the input prior. (Note that if we chose a distributed prior for $\text{pr}(\bar{a})$ and marginalized, these distributions would be significantly smoothed; using a delta function represents an extreme example.) Thus the meaningful results for this model are insensitive to the choice of prior; it is only necessary that the prior restrict the range of higher-order coefficients consistent with the expectations of naturalness.

C. Prior diagnostics

Our test cases make use of a fixed value of \bar{a} rather than marginalizing over a finite-width distribution for this naturalness parameter. This simplification is justified by using diagnostics for the prior that explore the sensitivity to \bar{a} .

In general one could marginalize over a range of \bar{a} values, and the posterior for \bar{a} , $\text{pr}(\bar{a}|D, k, k_{\max})$, when used in conjunction with the \bar{a} relaxation plot discussed next, is then useful for identification of an appropriate marginalization range for \bar{a} . The \bar{a} posterior will also signal the presence of unnatural coefficients (see Sec. V A).

gions of \bar{a} in the integration according to the posterior $\text{pr}(\bar{a}|D, k, k_{\max})$. This can be seen by using the rule of marginalization to express the coefficient posterior from Eq. (27) in an alternate form:

$$\begin{aligned} \text{pr}(\mathbf{a}|D, k, k_{\max}) &= \\ &\int d\bar{a} \text{pr}(\mathbf{a}|\bar{a}, D, k, k_{\max}) \text{pr}(\bar{a}|D, k, k_{\max}), \end{aligned} \quad (31)$$

where the first term in the integrand is the posterior for the coefficients given a specific \bar{a} and the second term is the posterior for \bar{a} that was calculated in Eq. (30). The result of marginalizing over \bar{a} therefore cannot directly be read from an \bar{a} relaxation plot. Figure 7 simply shows how the results change for fixed values of $\bar{a} = \bar{a}_{\text{fix}}$.

Ideally one finds (and marginalizes over) a slowly varying region in \bar{a} that is consistent with naturalness expectations; there should be little sensitivity to the endpoints of this region. There is sometimes a plateau in the \bar{a}_{fix} dependence, but not always. If there is a plateau, and

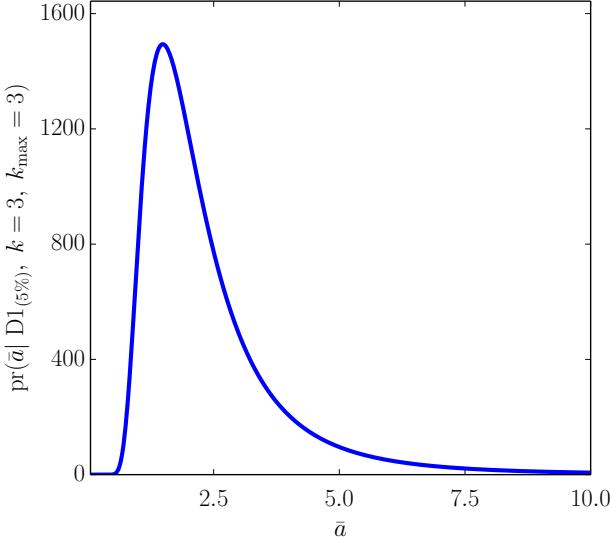


FIG. 6. The posterior pdf $\text{pr}(\bar{a}|D, k, k_{\max})$ calculated at $k = 3$, $k_{\max} = 3$ using prior Set C from Table I with $\bar{a}_< = 0.05$ and $\bar{a}_> = 20$, given data set $D1_{(5\%)}$.

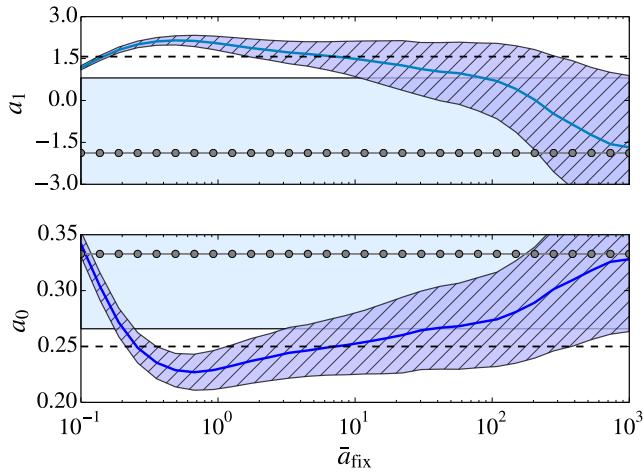


FIG. 7. (color online) Bayesian coefficient estimates calculated at $k = 1$, $k_{\max} = 4$ (solid lines with darker hatched error bands) as a function of \bar{a}_{fix} using prior Set C' given $D1_{(5\%)}$. The constant line with circles with lighter solid error bands is the least-squares estimate, which is independent of \bar{a}_{fix} . The error bands represent 68% DoBs ($1-\sigma$ errors).

coefficient estimates are largely independent of \bar{a}_{fix} in the \bar{a} region where $\text{pr}(\bar{a}|D, k, k_{\max})$ is significant, then marginalization will be equivalent to fixing \bar{a} at an \bar{a}_{fix} in this range. Taking a value for \bar{a}_{fix} somewhat above the peak region for \bar{a} in the \bar{a} posterior then provides a choice which is not overly restrictive. In the present case, we conclude that the model in question will not be sensitive to details of how \bar{a} is marginalized in a wide region, and for simplicity we fix $\bar{a}_{\text{fix}} = 5$.

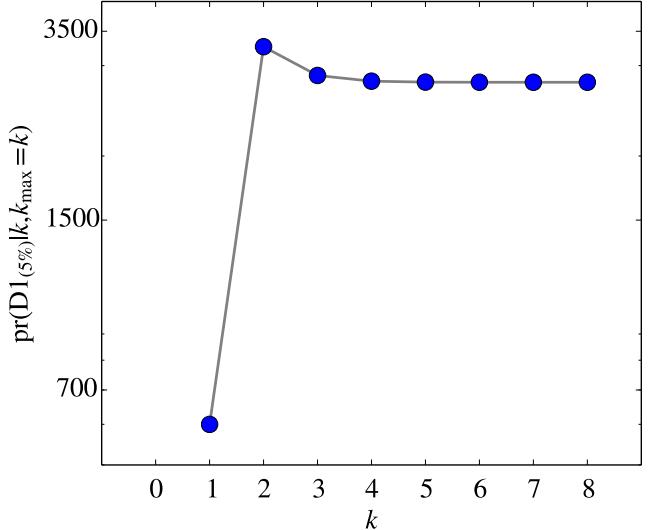


FIG. 8. (color online) Evidence $\text{pr}(D1_{(5\%)}|k, k_{\max} = k)$ for several values of k using prior Set C' with $\bar{a}_{\text{fix}} = 5$. (The evidence is not shown for $k = 0$ since it is nearly zero).

D. Model quality: Evidence

We discussed the evidence, which we define as $\text{pr}(D|k, k_{\max})$, in general terms in Sec. II E. In cases where the high-order coefficients are constrained by the naturalness prior, as in the situation reflected in Eq. (21), we expect saturation behavior of the evidence as parameters that are not well-constrained by the likelihood are added to the model. The evidence for a model at order k marginalized over coefficients up to order k_{\max} is

$$\begin{aligned} \text{pr}(D|k, k_{\max}) &= \int d\bar{a} \int d\mathbf{a} \int d\mathbf{a}_{\text{marg}} \\ &\times \text{pr}(D|\mathbf{a}, \mathbf{a}_{\text{marg}}, k, k_{\max}) \\ &\times \text{pr}(\mathbf{a}, \mathbf{a}_{\text{marg}}|\bar{a}, k, k_{\max}) \text{pr}(\bar{a}), \end{aligned} \quad (32)$$

where we have used the rule of marginalization from II C, the fact that the prior for the naturalness parameter is independent of the truncation and marginalization orders.

Just as in the discussion of the impact of k and k_{\max} on the posterior of \mathbf{a} in Sec. III A, our evidence calculations are also controlled solely by k_{\max} . For any of our models calculated at k with $k \leq k_{\max}$, Eq. (32) collapses to an integral over all the coefficients up to k_{\max} :

$$\begin{aligned} \text{pr}(D|k \leq k_{\max}, k_{\max}) &= \int d\bar{a} \int d\mathbf{a} \\ &\times \text{pr}(D|\mathbf{a}, k = k_{\max}, k_{\max}) \\ &\times \text{pr}(\mathbf{a}|\bar{a}, k = k_{\max}, k_{\max}) \text{pr}(\bar{a}), \end{aligned} \quad (33)$$

i.e., the value of k is irrelevant so long as it is less than or equal to k_{\max} since all the higher-order coefficients up to k_{\max} must also be integrated. This result also applies for models which are nonlinear in the coefficients. For this reason we compute only the evidence calculated at

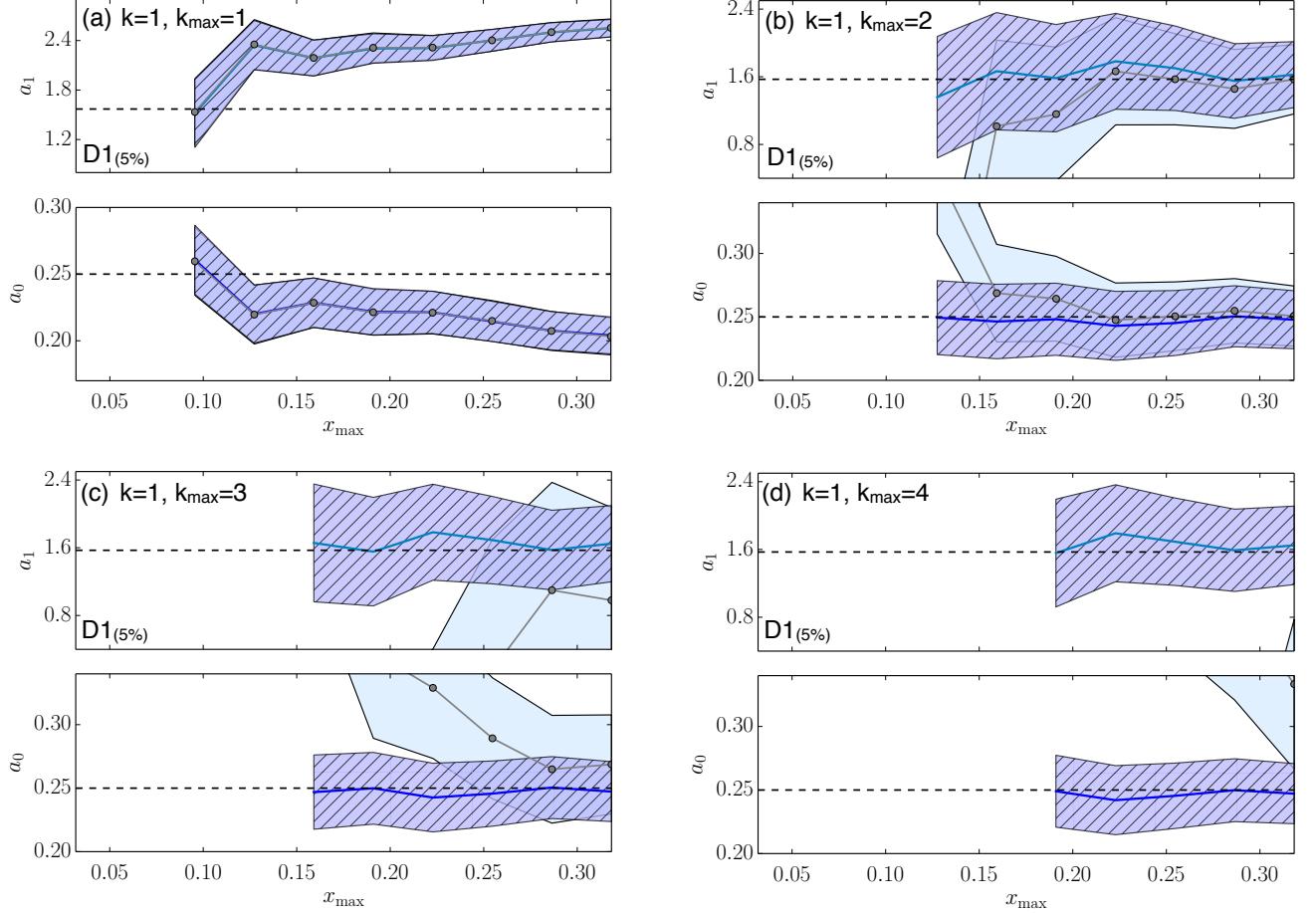


FIG. 9. (color online) Bayesian coefficient estimates as data from data set $D1_{(5\%)}$ are sequentially added at the high- x end. The largest x -value in the set is denoted as x_{\max} . The lines with darker hatched error bands represent estimates using prior Set C' with $\bar{a}_{\text{fix}} = 5$, and the line with circles with lighter solid error bands represents the least-squares estimates. The error bands represent 68% DoBs ($1-\sigma$ errors), which coincide in (a).

$k, k_{\max} = k$, e.g., as in Fig. 8. When the likelihood in Eq. (32) as a function of \mathbf{a}_{marg} is more complicated, these integrals will not necessarily collapse so simply and the evidence may be affected separately by k and k_{\max} , as discussed in Sec. III A.

Comparing the evidence of models as parameters are added (or as we marginalize over more high-order coefficients) provides an ideal order-by-order comparison for the suitability of a model to describe the data [19, 20]. Therefore when we perform parameter estimation in our Bayesian framework, we also compute the evidence for each model to quantitatively decide how many terms can be extracted from the data, and how many higher-order terms to marginalize.

Figure 8 shows the evidence $\text{pr}(D1_{(5\%)}|k, k_{\max} = k)$ for $k = 1$ to $k = 8$ using prior Set C' with $\bar{a}_{\text{fix}} = 5$. The evidence is also tabulated for each of these orders in Table III. Comparing different orders, we see that the Bayes ratio $\text{pr}(D1_{(5\%)}|k = 2)/\text{pr}(D1_{(5\%)}|k = 1) \approx 5$, implying that the quadratic model is somewhat more fa-

voritable than the linear one.⁵ Comparing higher orders, the Bayes ratio for the $k = 3$ to $k = 2$ case is about 1, indicating that they are equally favorable. This we define as evidence saturation, based on the discussion in Sec. II E, where going to higher order in the model does not improve the description of the data. When doing the parameter estimation for our examples, if $k + 1$ coefficients are expected to be determined by the data, the order $k_{\max} > k$ should not affect these $k + 1$ estimates; any coefficients above order k will essentially return the prior. However, verifying that this is true is an important part of the parameter estimation procedure. As emphasized in Ref. [30], simply truncating the fit at order k amounts to taking a delta-function prior for all coefficients of higher-order terms, since it assumes they are precisely zero.

⁵ We follow the empirical scale in Table 1 of Ref. [27], for which the thresholds for weak, moderate, and strong evidence favoring one model over another are ratios of 3, 12, and 150, respectively.

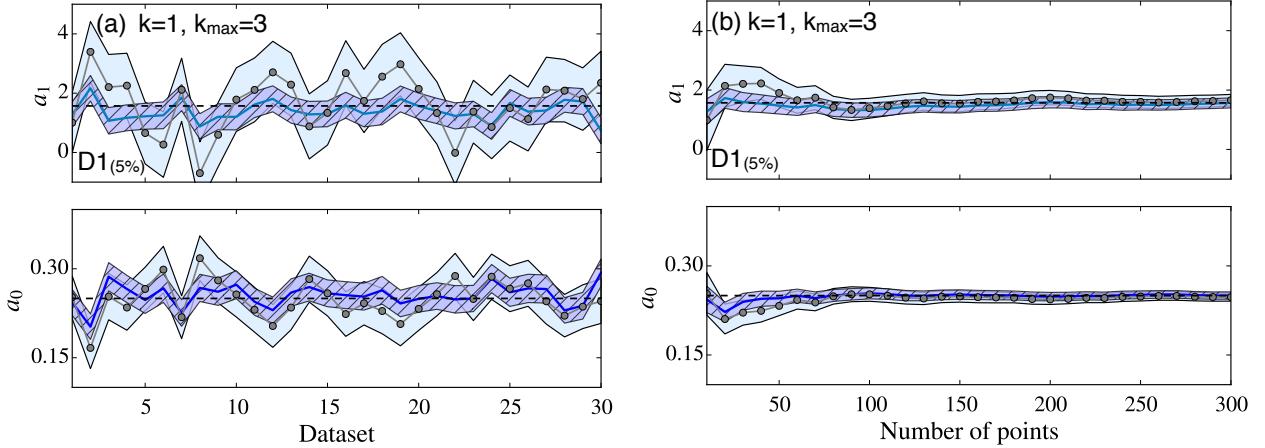


FIG. 10. Multi-set (a) and accumulation plots (b) calculated at $k = 1$, $k_{\max} = 3$. The shaded regions denote 68% error bands for the uniform (line with circles with lighter solid band) and naturalness prior (line with darker hatched band). The data sets used in (a) are 30 samples on the $D1_{(5\%)}$ mesh of 10 points. The same data are accumulated set by set to generate (b). In each case the prior was Set C' with $\bar{a}_{\text{fix}} = 5$.

E. x_{\max} plots

Variable x_{\max} plots show the projected maxima and widths of the parameters (which, for this problem, are the means and $1-\sigma$ errors) as a function of the maximum parameter value to which data are fit. In this way we examine the coefficients as we sequentially add data sampled at larger values of x , where higher-order contributions become increasingly important. The effect on the computed posterior can be seen from Eq. (4), where the data at different values of x form the residuals of the likelihood in Eq. (9), which goes into the coefficient posterior calculation. Plotting the evolution of the parameters as a function of x_{\max} shows the influence of each datum on the stability of the posterior in each dimension. Projected posterior plots such as those in Fig. 1 would be necessary to further analyze the correlations between orders at each value of x_{\max} .

A series of such plots are shown in Fig. 9 for data set $D1_{(5\%)}$ at four different fixed orders. The data set can be seen in Figs. 3 and 4 where we plot predictions from parameter estimates using different priors on the coefficients. We plot the $k = 1$ parameter results as the order of the marginalization k_{\max} is increased, starting in Fig. 9(a) with $k = 1$ (no higher-order coefficients) and going up to $k_{\max} = 4$ in Fig. 9(d). We plot the estimates with the naturalness prior as solid lines with error bands and compare these results with the least-squares results, plotted as lines with circles.

Some observations:

- The $k_{\max} = 1$ plot illustrates underfitting: the linear model only works well for the most infrared data (here the smallest x_{\max} point only) and significantly deviates from the true result elsewhere. Consequently, the result for larger x_{\max} shows no

stability. The prior for a_0 and a_1 is seen to be irrelevant here, and the least-squares results are the same as the estimates with the naturalness prior.

- The $k_{\max} = 2$ plot with no prior shows *overfitting* for the lowest values of x_{\max} , as there are too many terms available for the fit data, given the size of the (simulated) experimental error. As x_{\max} increases, the least-squares result becomes reliable. With the naturalness prior included, there is stability for a_0 and a_1 over the entire range. This marginalization over a_2 is the key feature for stability with x_{\max} .

- The $k_{\max} = 3$ plot shows that the uniform-prior results are off the scale for much of the displayed x_{\max} range, in strong contrast to the result with the naturalness prior, which gives stable central values and $1-\sigma$ errors. These naturalness-prior results are similar to the corresponding results when $k_{\max} = 2$. These patterns continue—for both the uniform-prior and naturalness-prior results—when $k_{\max} = 4$ [shown in Fig. 9(d)] and beyond (not shown).

This example illustrates the utility of using x_{\max} plots to check for instability with respect to the data range, which can signal underfitting and/or overfitting.

The stability with x_{\max} indicates that all high- x (UV) effects have been sufficiently accounted for, and that adding more high- x data does not improve the estimates. The k value at which these particular extractions become stable with respect to x_{\max} corresponds to that at which the evidence saturates. In this example, stability with x_{\max} begins when $k_{\max} = 2$, which is where the evidence saturates in Fig. 8. In fact, we will see in Sec. V B that x_{\max} plots can show stability when the evidence is past the saturation peak. This emphasizes the need to examine all the diagnostics before reaching a conclusion on the validity of the parameter estimation.

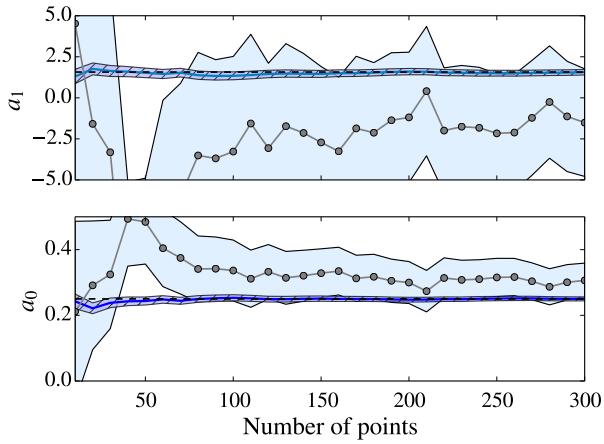


FIG. 11. Accumulation plot for $D1_{(5\%)}$ at $k = 1$, $k_{\max} = 6$. Same implementation as Fig. 10(b) but with $k_{\max} = 6$.

F. Multi-set and accumulation

Multi-set [Fig. 10(a)] and accumulation [Fig. 10(b)] plots are useful when there is enough data to subdivide it into a collection of smaller but (roughly) equivalent data sets. Multi-set plots provide a visualization of how fluctuations in the data affect the parameters. Figure 10(a) illustrates the fluctuations of the parameters a_0 and a_1 , estimated with the naturalness prior, as compared to the least-squares estimates, for the case $k_{\max} = 3$. For 30 data sets sampled on the same grid with the same random error as $D1_{(5\%)}$, we compare the maxima and width of the projected posteriors for each coefficient. This plot illustrates how large the fluctuations in the coefficients are as the data fluctuates. The added restriction from the naturalness prior reduces both the spread in the calculated central values and the size of the coefficients' error, compared to the result using the uniform prior.

Accumulation plots illustrate the utility of the prior when few data are available. Reading Fig. 10(b) from left-to-right shows that when there is less data available, the naturalness prior increases the precision of the estimates. Once there is enough data the uniform prior and naturalness prior results do not differ since the likelihood is the dominant component of the posterior. However, the danger with a uniform prior is that overfitting can degrade the parameter estimation, even when there is a large amount of data, see Fig. 11. Even with a large data set the uniform prior on included coefficients—and δ -function prior on omitted coefficients—means care must be taken in choosing k_{\max} , so that neither overfitting or underfitting takes place. Bayesian parameter estimation avoids this delicate selection, because the estimates of the low-order coefficients are stable with respect to k_{\max} .

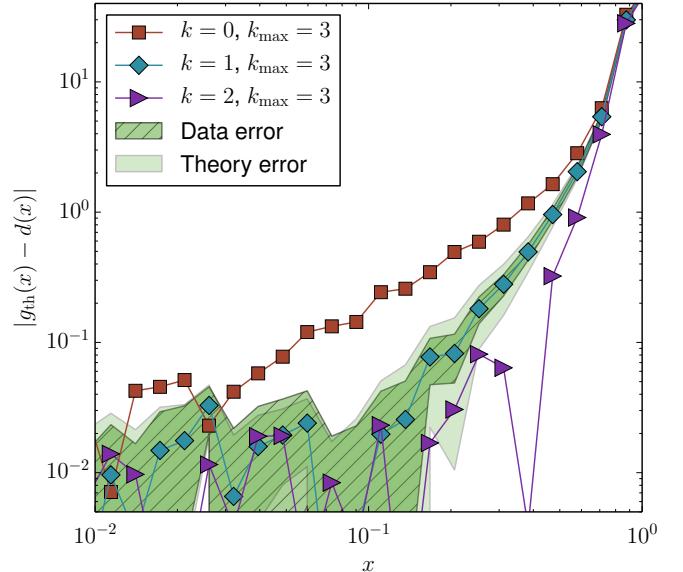


FIG. 12. Plot of residuals for predictions of Model D at leading order ($k = 0$), next-to-leading order ($k = 1$), and next-to-next-to-leading order ($k = 2$), all with $k_{\max} = 3$, given $D1_{(5\%)}$. The 68% error bands from both theory and data are shown on the $k = 1$ prediction as a representative example.

G. Residual plots

For sufficiently small values of the expansion parameter x , the truncation errors for an EFT calculated to order k should scale like the first omitted term (e.g., like x^{k+1} if all orders are present). This can be tested by a log-log plot of the absolute values of the residuals (difference of prediction and data) as a function of x . Plots of these type were introduced to analyze EFT behavior in Ref. [31] and are commonly called “Lepage plots”. A successful realization of the EFT should ideally reveal a clear signal of power-law behavior, with the slope given by the order in the expansion, in an extended region of x . This signal can be masked by data errors, particularly at low x , and by still higher-order truncation errors as the breakdown scale is neared. A Lepage plot can manifest the different regions, help to disentangle regulator artifacts from errors due to truncating the EFT Hamiltonian, and, in principle, approximately determine the breakdown scale [11].

We show an example of such a plot for Model D using $D1_{(5\%)}$ in Fig. 12. The three sets of points are residuals at orders $k = 0$, 1, and 2, from a parameter estimation at $k_{\max} = 3$. Data and theoretical error bands have been added for the $k = 1$ residuals to help identify the crossover between the region where data errors mask the signal of the first omitted term and where the anticipated power-law behavior should be seen. In this example, the crossover occurs at successively higher values of x , with at best a very narrow region of power-law behavior for $k = 2$.

In practice data are frequently sparse and too noisy

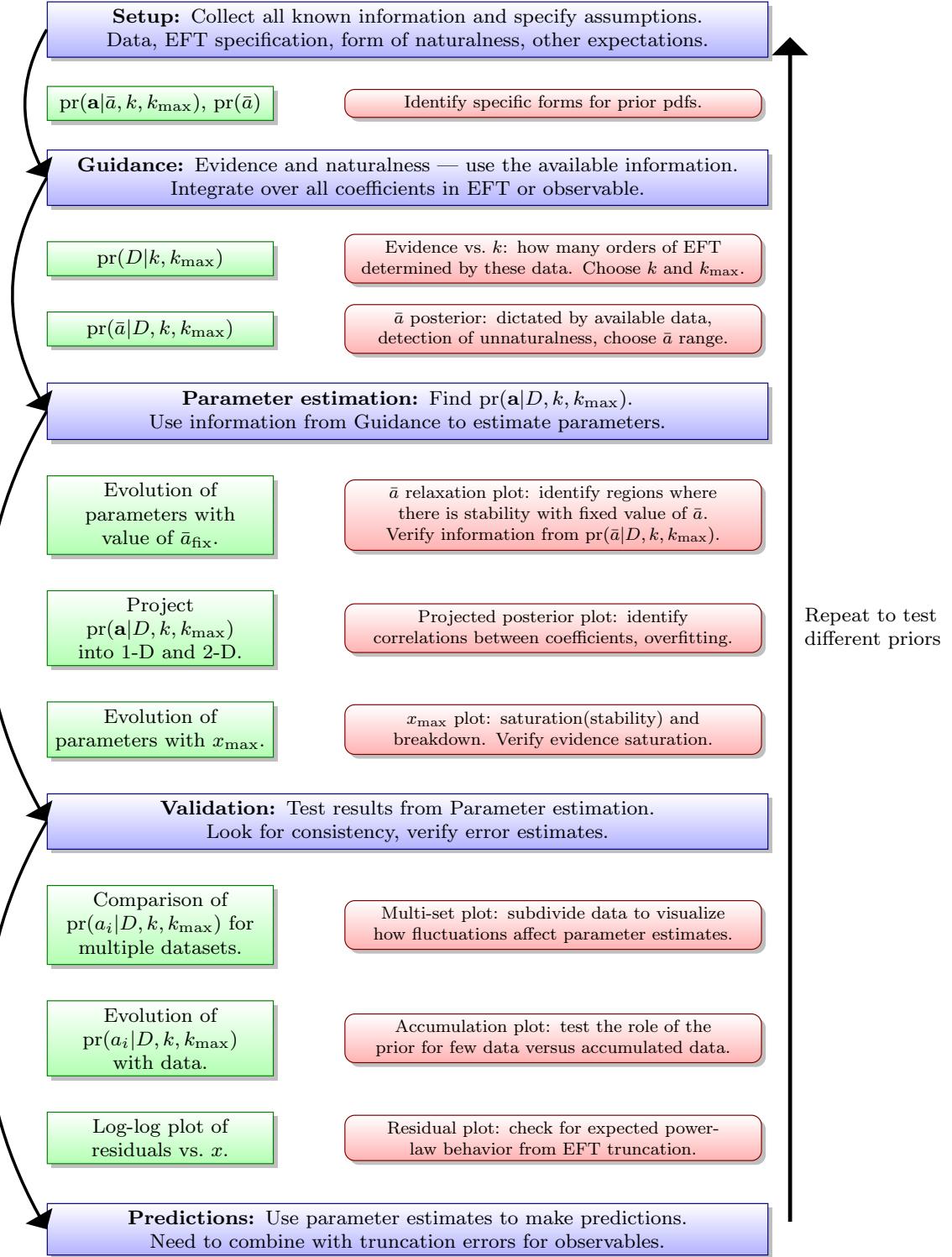


FIG. 13. (color online) Flowchart for parameter estimation applying diagnostic tools from Table II.

to robustly discern from Lepage plots whether power-law exponents are in accordance with EFT expectations. A different type of error plot that avoids this problem compares theory residuals from different values of an EFT regulator scale (i.e., cutoff) rather than between theory and data [32–34]. Examples and further discussion of residual plots are given in Ref. [11] and we do not consider them further in the present work.

IV. THE PARAMETER ESTIMATION PROCESS

A. Overview

In Fig. 13, we present a possible flowchart for the full parameter estimation process, which orders and builds on the diagnostic tools described in the last section. The process starts with the Setup of the model and specification of all available information and theoretical expectations (including the form of the priors but not their widths), and ends with Predictions of observables from the fit parameters (or, more precisely, from the posteriors for the parameters). The intermediate steps are divided into Guidance, Parameter estimation, and Validation. In this section we consider each of these in turn, not exhaustively, but to highlight how individual diagnostics can offer different insights into parameter estimation for EFTs. We emphasize that while we have for clarity described the process as a forward flow, in practice one would backtrack if later diagnostics do not support earlier conclusions.

We choose another convenient model for these explorations:

$$g(x) = \frac{\beta^2}{(\beta + x)^2}, \quad (34)$$

with fixed $\beta = 1.3$. The Taylor expansion, which identifies the coefficients we seek to estimate, is

$$g(x) = 1 - 1.54x + 1.78x^2 - 1.82x^3 + 1.75x^4 + \dots. \quad (35)$$

Note that the pole is at negative x so that the radius of convergence is “hidden” in the data. This also means that the coefficients in the Taylor expansion have alternating sign. The magnitudes of the coefficients are ~ 1 up until $\mathcal{O}(x^{10})$, where they begin decreasing. We consider a variety of data sets, of varying precision, and with data over different ranges in x and with different numbers of points. These sets are enumerated in Table IV (and available in files in the supplementary material). We add a subscript to the label to indicate the percent relative error. Figures 14 and 15 show the underlying model function from Eq. (34), along with one data set $H0_{(1\%)}$. Both fits agree quite well in the region where there is data. But, outside that interval, the uniform-prior predictions in Fig. 14 show the consequences of overfitting. The result is not in agreement with the underlying function already a disappointingly small distance above the

fit region. And the problems are not just at high x : the overfitting reduces precision at low x too. The Bayesian result in Fig. 15 does not suffer from these problems.

TABLE IV. Model H data set labels for sampling grid ranges and number of points. The breakdown scale is $x = 1.3$.

Label	# of pts.	Grid	Spacing
H0	10	$0.01 \leq x \leq 0.1$	linear
H1	10	$0.05 \leq x \leq 0.5$	linear
H2	15	$0.05 \leq x \leq 0.75$	linear
H3	10	$0.05 \leq x \leq 0.5$	quadratic
H4	15	$0.10 \leq x \leq 1.5$	linear

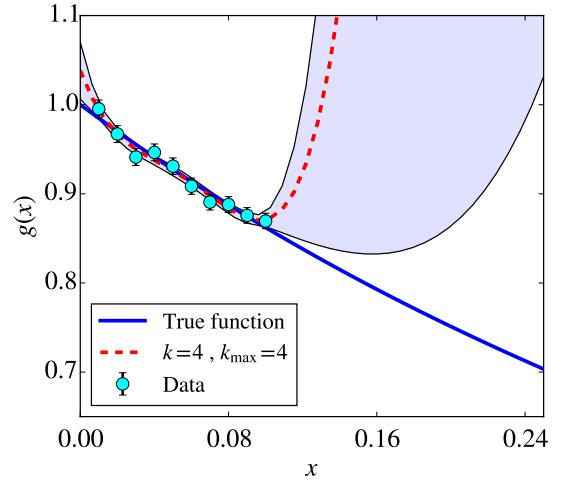


FIG. 14. (color online) Comparison of data set $H0_{(1\%)}$ (corresponding to the first row of Table IV), the underlying function for Model H from Eq. (34), and a least-squares prediction calculated at order $k = 4$, $k_{\max} = 4$ from that data set. The error bands represent 1- σ (68% DoBs).

The data sets in Table IV are representative of different situations that might be encountered in EFT parameter estimation. $H0_{(1\%)}$ has very small errors and is sampled on a mesh where the expansion parameter is very small. $H1_{(5\%)}$ is sampled over a small range of x with a fairly large data error. $H2_{(5\%)}$ expands upon the $H1_{(5\%)}$ mesh by adding an additional 5 ultraviolet (UV) measurements to the mesh to probe the improvement of leading-order extractions when UV data are added. Finally, $H3_{(1\%)}$ is an accurate data set measured on a quadratically spaced mesh (in the same range as $H1_{(5\%)}$). This both simulates an application to EFT expansions with even powers only and results in a data set with several high-precision measurements at small x . $H4_{(5\%)}$ is used in Sec. V B as a case study of an EFT fit beyond its breakdown scale.

Tables V, VI, VII, and VIII each show the results from the Parameter estimation stage using the data sets enumerated above, which are generated at random and are not selected to be “typical” in any way. As a consequence, the impact of fluctuations is manifested. With-

TABLE V. Coefficient estimates from sampling of $\text{pr}(\mathbf{a}|H_{0(1\%)}, k, k_{\max})$ given the expansion from Eq. (3) (these results are controlled by k_{\max} only, see Sec. III A). The left side of the table is for a uniform prior, which is equivalent to a least-squares fit, and includes the χ^2/dof values. The right side of the table is using prior Set C' from Table I with $\bar{a}_{\text{fix}} = 5$, and includes the evidence. For both priors the posterior pdf is a multi-dimensional Gaussian.

		Uniform prior			Gaussian prior				
k	k_{\max}	χ^2/dof	a_0	a_1	a_2	Evidence	a_0	a_1	a_2
0	0	20	0.92 ± 0.00			~ 0	0.92 ± 0.00		
1	1	0.90	0.99 ± 0.01	-1.3 ± 0.1		7.2×10^9	0.99 ± 0.01	-1.3 ± 0.1	
2	2	0.64	1.0 ± 0.01	-2.1 ± 0.5	6.7 ± 4	1.0×10^{10}	1.0 ± 0.01	-1.8 ± 0.4	4.0 ± 3
2	3	0.74	1.0 ± 0.02	-2.2 ± 1	9.8 ± 30	1.0×10^{10}	1.0 ± 0.01	-1.8 ± 0.4	4.0 ± 3
2	4	0.67	1.0 ± 0.03	-5.6 ± 4	130 ± 100	1.0×10^{10}	1.0 ± 0.01	-1.8 ± 0.4	4.0 ± 3
2	5	0.54	1.1 ± 0.1	-14 ± 9	580 ± 400	1.0×10^{10}	1.0 ± 0.01	-1.8 ± 0.4	4.0 ± 3
2	6	0.69	1.1 ± 0.1	-8.6 ± 20	190 ± 2000	1.0×10^{10}	1.0 ± 0.01	-1.8 ± 0.4	4.0 ± 3
True values		1.0	-1.54	1.78			1.0	-1.54	1.78

TABLE VI. Same as Table V except sampling from $\text{pr}(\mathbf{a}|H_{1(5\%)}, k, k_{\max})$.

		Uniform prior			Gaussian prior				
k	k_{\max}	χ^2/dof	a_0	a_1	a_2	Evidence	a_0	a_1	a_2
0	0	13	0.65 ± 0.01			~ 0	0.65 ± 0.01		
1	1	2.3	0.89 ± 0.03	-0.74 ± 0.1		1.4×10^2	0.89 ± 0.03	-0.74 ± 0.08	
2	2	2.3	0.95 ± 0.05	-1.3 ± 0.4	0.97 ± 0.6	6.2×10^1	0.95 ± 0.05	-1.3 ± 0.4	0.95 ± 1
2	3	2.5	1.0 ± 0.08	-2.1 ± 1	4.4 ± 4	3.9×10^1	0.96 ± 0.06	-1.5 ± 0.7	1.8 ± 3
2	4	2.9	0.92 ± 0.1	-0.04 ± 3	-10 ± 20	3.3×10^1	0.97 ± 0.06	-1.6 ± 0.7	1.8 ± 3
2	5	3.5	0.75 ± 0.2	5.5 ± 7	-66 ± 70	3.0×10^1	0.97 ± 0.06	-1.6 ± 0.7	1.9 ± 3
2	6	4.0	0.13 ± 0.5	29 ± 20	-380 ± 200	3.0×10^1	0.97 ± 0.06	-1.6 ± 0.7	1.7 ± 3
True values		1.0	-1.54	1.78			1.0	-1.54	1.78

TABLE VII. Same as Table V except sampling from $\text{pr}(\mathbf{a}|H_{2(5\%)}, k, k_{\max})$.

		Uniform prior			Gaussian prior				
k	k_{\max}	χ^2/dof	a_0	a_1	a_2	Evidence	a_0	a_1	a_2
0	0	29	0.53 ± 0.01			~ 0	0.53 ± 0.01		
1	1	2.4	0.89 ± 0.02	-0.70 ± 0.04		1.7×10^5	0.89 ± 0.02	-0.70 ± 0.04	
2	2	1.2	1.0 ± 0.04	-1.4 ± 0.2	0.77 ± 0.2	2.7×10^7	1.0 ± 0.04	-1.4 ± 0.2	0.77 ± 0.2
2	3	1.3	1.0 ± 0.1	-1.7 ± 0.5	1.6 ± 1	6.0×10^6	1.0 ± 0.05	-1.6 ± 0.5	1.4 ± 1
2	4	1.3	1.1 ± 0.1	-2.6 ± 1	5.9 ± 5	2.6×10^6	1.0 ± 0.06	-1.7 ± 0.6	1.6 ± 2
2	5	1.3	1.0 ± 0.1	0.63 ± 3	-17 ± 20	1.6×10^6	1.0 ± 0.06	-1.8 ± 0.7	2.0 ± 2
2	6	1.3	1.1 ± 0.2	-3.5 ± 5	23 ± 50	1.2×10^6	1.1 ± 0.06	-1.9 ± 0.7	2.2 ± 2
True values		1.0	-1.54	1.78			1.0	-1.54	1.78

TABLE VIII. Same as Table V except sampling from $\text{pr}(\mathbf{a}|H_{3(1\%)}, k, k_{\max})$.

		Uniform prior			Gaussian prior				
k	k_{\max}	χ^2/dof	a_0	a_1	a_2	Evidence	a_0	a_1	a_2
0	0	380	0.68 ± 0.00			~ 0	0.68 ± 0.00		
1	1	4.4	0.94 ± 0.00	-0.87 ± 0.01		8.8×10^3	0.94 ± 0.00	-0.87 ± 0.01	
2	2	0.88	0.98 ± 0.01	-1.23 ± 0.1	0.64 ± 0.1	3.6×10^8	0.98 ± 0.01	-1.2 ± 0.07	0.64 ± 0.12
2	3	0.73	0.96 ± 0.02	-0.96 ± 0.2	-0.50 ± 0.9	1.7×10^8	0.96 ± 0.01	-0.96 ± 0.2	-0.48 ± 0.8
2	4	0.62	0.94 ± 0.03	-0.35 ± 0.6	-4.8 ± 4	1.1×10^8	0.96 ± 0.02	-0.94 ± 0.3	-0.66 ± 1
2	5	0.58	0.90 ± 0.05	0.80 ± 2	-17 ± 10	8.7×10^7	0.96 ± 0.02	-0.91 ± 0.3	-0.87 ± 2
2	6	0.78	0.90 ± 0.1	0.70 ± 4	-15 ± 50	8.0×10^7	0.96 ± 0.02	-0.88 ± 0.3	-0.99 ± 2
True values		1.0	-1.54	1.78			1.0	-1.54	1.78

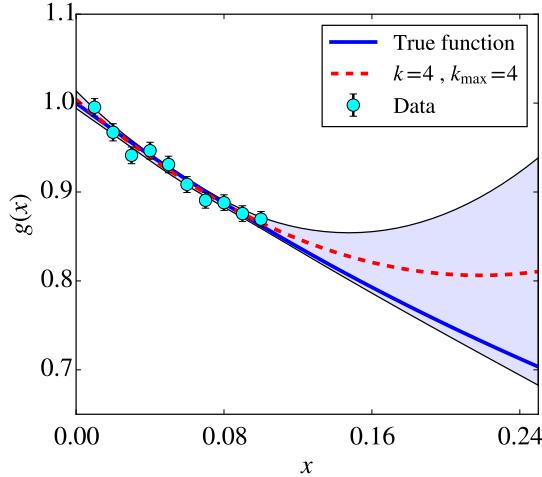


FIG. 15. (color online) Comparison of data set $H0_{(1\%)}$ (corresponding to the first row of Table IV), the underlying function for Model H from Eq. (34), and a Bayesian prediction calculated at order $k = 4$, $k_{\max} = 4$ using prior Set C' with $\bar{a}_{\text{fix}} = 5$, from that data set. The error bands represent 1- σ (68% DoBs).

out accounting for the errors, one might jump to false conclusions by naive comparisons to the true values. We see two examples immediately. First, the estimates for a_1 in Table VI with the Gaussian prior are much closer to the true values than those in Table V. But, the latter's tighter 68% DoB limits say it is the more precise prediction; it is just chance that the extraction from the H1 data set results in a central value that is close as the DoB limits are quite wide. Second, in Table VIII, the estimates with a Gaussian prior differ from the true values well beyond the quoted error bars. But, as documented below, this particular data set happened to have 2σ fluctuations in the data points, which leads to estimates for which the true value is outside the 68% DoB interval. This should not be a surprise; indeed, it is expected one-third of the time! The occurrence of a large fluctuation in $H3_{(1\%)}$ provides a testing ground for our multi-set analysis (see Sec. III F); this will reveal the impact of fluctuations on parameter estimates.

B. Guidance: Exploring the prior, k , and k_{\max}

The goal of the Guidance stage of our parameter estimation process is to take a particular data set and determine two aspects of the associated fit: the number of EFT parameters that can be reliably extracted and the range of naturalness parameters that should be used in that extraction. (Note: these procedures can be adapted to priors for other information.) The evidence $\text{pr}(D|k, k_{\max})$ illuminates the information content of the data with respect to the order of our fit. It is important to do the parameter estimation in the k -region where the evidence has saturated, or else the estimates of

those coefficients that can be determined will not be stable with respect to the order of marginalization k_{\max} . In Sec. III D it was discussed that for these model problems when $k \leq k_{\max}$, the evidence is controlled exclusively by k_{\max} . Therefore when we calculate the evidence, we do so for $k = k_{\max}$ as we did in Fig. 8. For k_{\max} s not in the saturation region the coefficients are underfit—even in the presence of a naturalness prior. Meanwhile, the pertinent values of \bar{a} are indicated by the posterior $\text{pr}(\bar{a}|D, k, k_{\max})$, which shows the data's content with respect to the naturalness of the coefficients.

The Guidance diagnostics are interrelated and should be compared for different prior assumptions—in this case we will compare Set C (or C') and Set A as an example. As confirmed below, for Model H we can fix \bar{a} to a reasonable value of \bar{a}_{fix} and the parameter estimates are largely insensitive to this choice. More detailed comparisons would be needed if results were more sensitive to the choice of prior. In the following we use the results in Tables V, VI, VII, and VIII to evaluate what we learn from the Guidance stage.

Small error, very small range. We first consider $H0_{(1\%)}$, a very accurate data set sampled at small x values, and examine what the evidence and \bar{a} posterior plots tell us in advance of parameter estimation. The \bar{a} posterior in Fig. 16 uses prior Set C. It shows that the most likely value for \bar{a} given this information is about 1, and that choosing $\bar{a}_{\text{fix}} = 5$ will not be overly restrictive. The evidence plot for Set C' in Fig. 17 shows saturation at $k = 2$, with no real improvement from $k = 1$. This suggests that the limited range of this data set in x means that we cannot extract information past the quadratic order (and that order will have limited information), and that we should marginalize over parameters $a_{i>2}$. This conclusion is the same for different priors, and the evidence saturation behavior is compared for Set A and Set C' in Fig. 17.

The expectations developed in the Guidance stage for $H0_{(1\%)}$ are verified by the parameter estimates in Table V. In particular, the results on the right side of the table using a Gaussian prior are consistent with determining successive orders with decreasing accuracy. The last coefficient determined by the data is a_2 , which has a large error, consistent with the evidence saturation. Higher-order coefficients simply return the prior.

It is instructive to compare to the results on the left side of the table, which are extracted using a uniform prior, i.e., assuming no prior knowledge of coefficient size. One might think that a high-quality data set with many points near $x = 0$ should be dominated by the lowest-order terms, but these results show that the effects of fine-tuning are severe. In this case Table V shows that we have underfitting at $k_{\max} = 1$, but, by the time we use $k_{\max} = 2$ for our fit, a_2 and a_1 have both acquired large (in absolute terms) central values. This overfitting only gets more marked as k_{\max} increases. In general the uncertainties allow a_0 and a_1 to be within the 68% (or, at worst 95%) DoB interval (once $k_{\max} > 1$) but the 68%

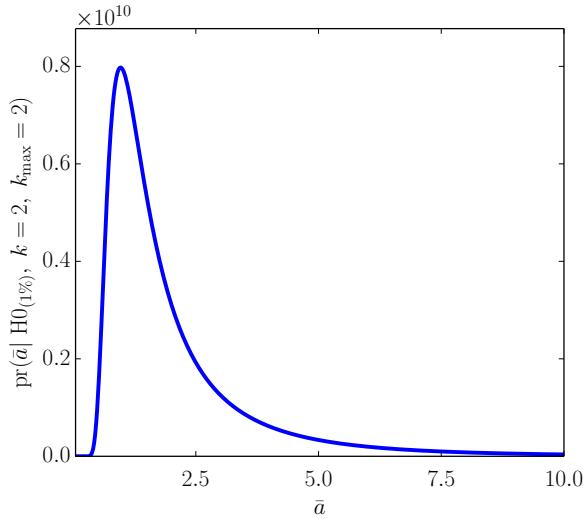


FIG. 16. The posterior pdf $\text{pr}(\bar{a}|D, k, k_{\max})$ calculated at $k = 2$, $k_{\max} = 2$ assuming prior Set C from Table I with $\bar{a}_< = 0.05$ and $\bar{a}_> = 20$, given data set $H0_{(1\%)}$.

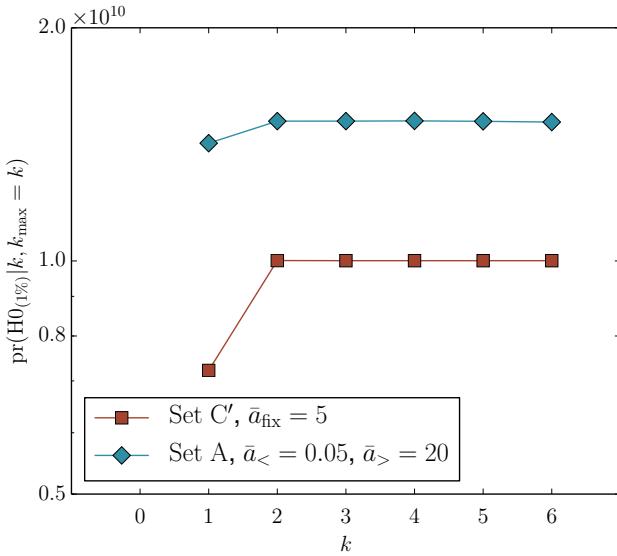


FIG. 17. Evidence $\text{pr}(H0_{(1\%)}|k, k_{\max} = k)$ using different prior assumptions for several values of k with $k_{\max} = k$. (The evidence for $k = 0$ is not shown for either prior because it is nearly zero). Note that only ratios of the evidence at different k for the same data and priors are significant.

interval is so wide in the case of a_1 that this statement is of little practical use. This is disappointing: one might have expected better from this low- x (infrared) data. A projected posterior plot verifies that it occurs because a_1 and a_2 are highly anti-correlated, and there is nothing in this data set to pin down the value of a_2 .

At higher orders, plotting the uniform prior prediction, e.g., $k = k_{\max} = 4$ in Fig. 14, the fine-tuning (overfitting) problem is manifest. Similar results using the uniform

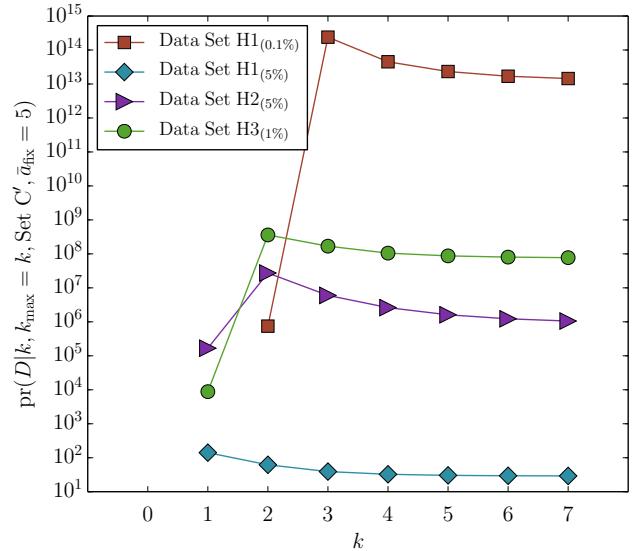


FIG. 18. Data set evidence $\text{pr}(D|k, k_{\max} = k)$ for several values of k using prior Set C' with $\bar{a}_{\text{fix}} = 5$ in each case. (The evidence is not shown for all calculations at $k = 0$ because they are nearly zero, and also the evidence for data set $H1_{(0.1\%)}$ at $k = 1$ for the same reason).

prior for our other data sets are shown in Tables VI, VII, and VIII, where generic underfitting/overfitting is also present. These issues make it clear that not including known information results in worse predictions, which can manifest in subtle ways. For example, the a_0 results for the uniform prior in Table V are not too far off from the true value but are in fact influenced by fine-tuning. The remaining significant error bar on a_1 shows that data in this narrow x window does not have a good lever arm to accurately determine the slope of the underlying function at $x = 0$, a feature that is quantified by the evidence.

Large error, small range. In the case of $H1_{(5\%)}$, we might expect better constraints on a_2 from data that covers a wider but still small range of x than $H0_{(1\%)}$. Fig. 18 shows a slight peak at $k = 1$ and saturation at higher k for $H1_{(5\%)}$. The interaction of the prior and likelihood is not as clean as in Eq. (21) when $k = 1$. By using only the $k = 1$ order EFT where the evidence peaks rather than the saturation region, we are not including prior information that the EFT has higher-order coefficients—this results in a poor extraction, as is evident on the right of Table VI. Estimates should be made in the saturation region, and the coefficients extracted at lower orders become stable with respect to k_{\max} because we appropriately account for higher-order effects. The uniform-prior results for $H1_{(5\%)}$ in Table VI show that the $k_{\max} = 1$ result is not very good, but that in fact $k_{\max} = 1$ and $k_{\max} = 2$ have the same values for the χ^2/dof . Hence the lowest χ^2 or the evidence peak are not reliable diagnostics of good parameter estimates—only when the evidence has saturated do the values of coefficients determined by the data cease to change with

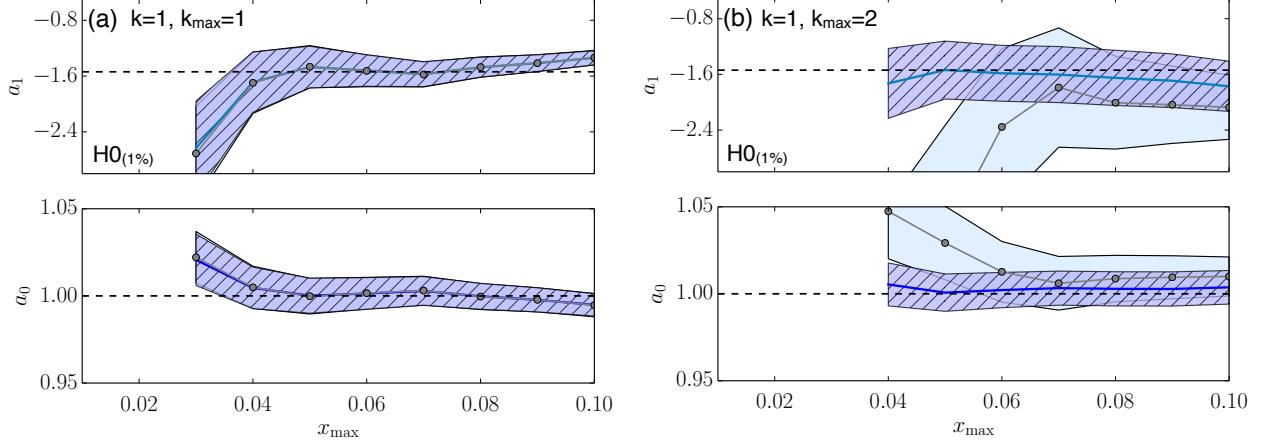


FIG. 19. (color online) Bayesian coefficient estimates as data from data set $H0_{(1\%)}$ are sequentially added at the high- x end. The largest x -value in the set is denoted as x_{\max} . The lines with darker hatched error bands represent estimates using prior Set C' with $\bar{a}_{\text{fix}} = 5$, and the line with circles with lighter solid error bands represents the least-squares estimates. The error bands represent 68% DoBs ($1-\sigma$ errors), which coincide in (a).

increasing k_{\max} .

For comparison, we also calculate the evidence for data set $H1_{(0.1\%)}$, which has 50 times smaller relative error than $H1_{(5\%)}$. For this highly precise data sampled on the $H1$ mesh from Table IV, the evidence saturates at $k = 3$ in Fig. 18, indicating that there are limitations from where the data are sampled even if it is very accurate and precise. There is still a small peak in the evidence as there was in the case of 5% error.

Large error, larger range. $H2_{(5\%)}$ has the same density of points as $H1_{(5\%)}$ but has 5 extra points up to $x = 0.75$. Does the addition of 5 more UV points improve the lower-order estimates by constraining the higher-order ones better? Yes: Fig. 18 shows that the evidence with $k = 1$ is now several orders of magnitude lower than the evidence in the saturation region near $k = 3$. The evidence is highest at $k = 2$ but the estimates in Table VII using a naturalness prior are not stable with respect to changing k_{\max} until the evidence saturates.

Small error, small range. Lastly, we examine $H3_{(1\%)}$, which samples the same region of x as $H1_{(5\%)}$, but has more data points at small x and has higher precision. The uniform-prior results are shown on the left side of Table VIII, and we see the generic overfitting/underfitting. Figure 18 shows saturation at $k = 2$ for this set. Compared to the evidence for $H1_{(5\%)}$, the decisiveness of the evidence at $k = 2$ to $k = 1$ is better. Turning to parameter estimation results using prior Set C' on the right of Table VIII, the extraction stabilizes as we enter the saturation region but the results do not seem consistent with the true values. This is not a failure, however; this particular $H3_{(1\%)}$ data set is one of the 32% of data sets that will not produce LECs in the 68% band. The a_0 extraction is driven by the lowest- x data point so that it is less than the true value, $a_0 = 1$. We will discuss this data set further when considering

multi-set plots below.

Summary. A uniform prior (least-squares method) is limited to the information in the data sets themselves. Whenever lower-order coefficients are correlated with higher-order ones, there will be underfitting of coefficients associated with the last included order—and overfitting of lower-order coefficients. While high-precision data can ameliorate the situation for a few coefficients, the potential of such a data set is still under-realized in a least-squares analysis. More generally, such difficulties cannot be avoided by choosing a particular x domain for the fit.

The Bayesian procedure is much more effective but is still constrained by the limitations of the data set, and these limitations are quantifiable using the evidence. Using a wider region, but with less precise data, decreases the accuracy of the a_0 determination markedly. (a_0 is determined with essentially the same accuracy from $H1_{(5\%)}$ and $H2_{(5\%)}$.) The decrease in the accuracy of a_1 is not as bad in $H1_{(5\%)}$ and $H2_{(5\%)}$ as $H0_{(1\%)}$, because some of the loss in data precision is made up by the increased lever arm. All three data sets suggest that only a precise data set, over a wide range of x , will be able to give unambiguous information on a_2 . Perhaps most notably, in each case, $H0_{(1\%)}$, $H1_{(5\%)}$, and $H2_{(5\%)}$, both the central values and 68% DoB intervals stabilize with k_{\max} once $k_{\max} \geq 3$. Those stabilized results are consistent with the underlying values of all of a_0 , a_1 , and a_2 .

C. Parameter estimation: x_{\max} analyses

For the Parameter estimation step in the process, the principal diagnostic is always the projected posterior plot, which illuminates the correlations and shows features such as which parameters are dominated by the

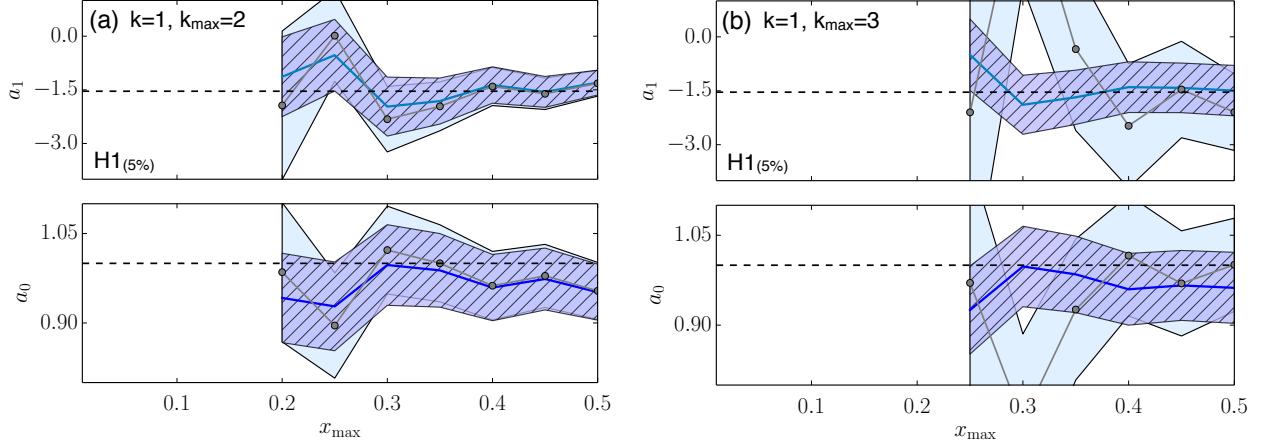


FIG. 20. (color online) x_{\max} plots for data set $H1_{(5\%)}$ with same description as Fig. 19.

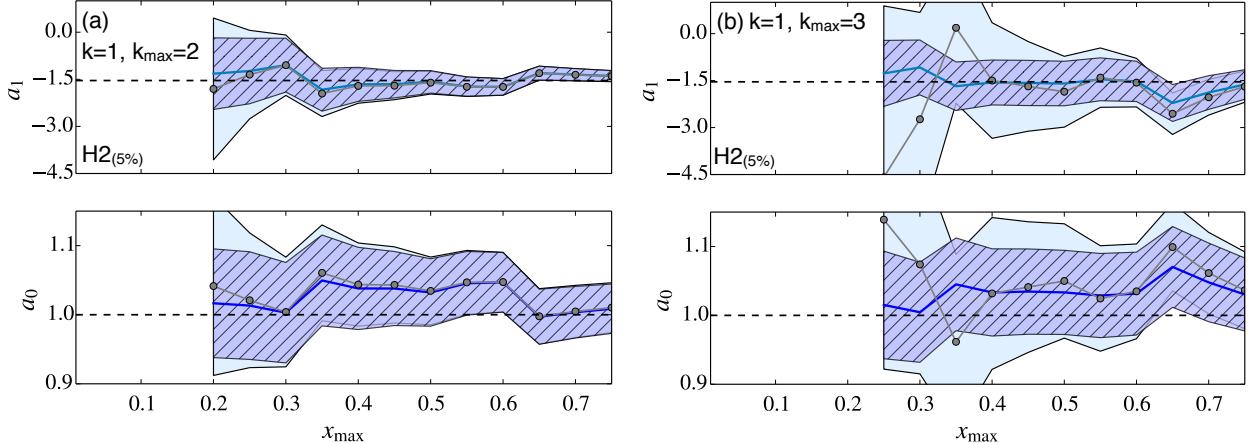


FIG. 21. (color online) x_{\max} plots for data set $H2_{(5\%)}$ with same description as Fig. 19.

prior. We have already examined examples of projected posterior plots and associated tables of parameter means and standard deviations, so here we focus instead on the added utility of x_{\max} plots. These track the evolution of parameter estimates as the data set is built up sequentially with more and more high- x (UV) data, thus making it clear that choosing ranges of data to fit certain orders in the EFT may omit important information.

Small error, very small range. We saw in Fig. 17 that evidence saturation for $H0_{(1\%)}$ occurs at $k = 2$, which is slightly favored over $k = 1$. But, Table V showed that, while the posterior contains some information on a_2 , it is largely undetermined. Therefore we explore the $k = 1$ results (the first two coefficients), marginalizing over higher coefficients for $k_{\max} > 1$. Figure 19(a) shows the $k = 1$, $k_{\max} = 1$ x_{\max} plot. Here the uniform and prior Set C' parameter estimates overlap and are not stable with increasing x_{\max} , indicating underfitting. When we marginalize to $k_{\max} = 2$, where the evidence saturated, in Fig. 19(b), the naturalness prior parameter es-

timates are stable with x_{\max} . The uniform-prior results are not and overfitting occurs, especially for small x_{\max} . Once $x_{\max} \approx 0.07$, the uniform prior error bars encompass the true value but are significantly larger than the naturalness results. As k_{\max} is increased further into the saturation region, the parameter estimates with the naturalness prior continue to be stable with x_{\max} (not shown).

Large error, different ranges. Turning to $H1_{(5\%)}$ and $H2_{(5\%)}$, which are sampled at larger ranges of x with larger errors, we explore the x_{\max} behavior at k_{\max} values near the evidence saturation. Figures 20(a) and 20(b) show the results at $k = 1$ with $k_{\max} = 2, 3$ respectively for $H1_{(5\%)}$. In light of the evidence in Fig. 18, $k_{\max} = 3$ is approximately in the saturation region while $k_{\max} = 2$ does not account for enough higher orders. At $k_{\max} = 2$ the estimates at the largest x_{\max} overlap with the uniform-prior results, while the corresponding estimate for $k_{\max} = 3$ is slightly better. The results for $k_{\max} = 3$ are also fairly stable with x_{\max} , consistent with

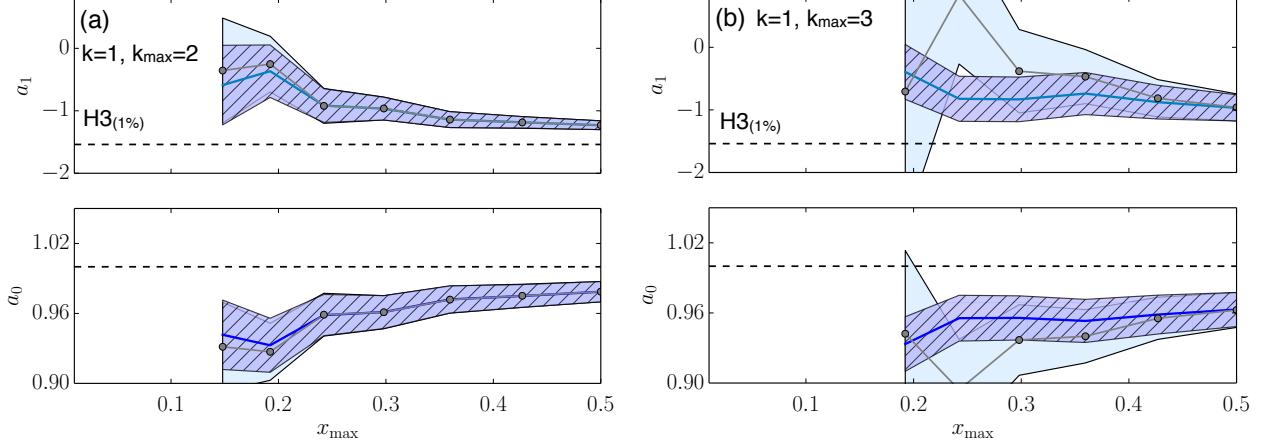


FIG. 22. (color online) x_{\max} plots for data set $H3_{(1\%)}$ with same description as Fig. 19.

the evidence saturation.

For $H2_{(5\%)}$, with extra UV data, we show the $k = 1$, $k_{\max} = 2, 3$ results in Figs. 21(a) and 21(b), which again explores the transition region near evidence saturation seen in Fig. 18. As we saw for $H1_{(5\%)}$, the estimates are not stable with x_{\max} until $k_{\max} = 3$ when the evidence truly saturates. In particular the final estimates with the largest x_{\max} in Fig. 21(a) significantly underestimate the error in a_1 .

Small error, small range. The evidence for $H3_{(1\%)}$ in Fig. 18 clearly saturates. We compare both $k_{\max} = 2, 3$, just showing the first two coefficients ($k = 1$), but not $k_{\max} = 1$ because it is in the underfitting regime. Surprisingly perhaps from the perspective of the evidence, $k_{\max} = 2$ in Fig. 22(a) still gives underfitting and for no x_{\max} is the estimate consistent with the true parameter to the $1-\sigma$ level. Even for low x_{\max} , underfitting still occurs. But the essential feature here is the absence of stability with respect to x_{\max} , *not* that the 68% DoBs miss the true parameter values. In contrast, for $k_{\max} = 3$, shown in Fig. 22(b), there is stability with x_{\max} , even as the uniform-prior results begin overfitting. But, even here, the 68% DoB intervals miss the true value. Such “occasional misses” are, though, consistent with a statistical interpretation of these error bands.

Summary. Using x_{\max} plots rather than a fixed range of data provides valuable insight into the parameter extraction. They establish the region of x that provides stable predictions, help to identify where in x new information on coefficients stops, and confirm the evidence saturation from the Guidance.

D. Validation: Multi-set analysis

Finally, for the Validation step we focus on how fluctuations in data affect our results as well as how the errors in the available data limit parameter estimation.

We posit that some larger set of measurements for our model is available, which we divide into smaller subsets and use to perform the parameter estimation to see the effects of statistical fluctuations (multi-set plots outlined in Sec. III F). We also take these data sets and build them up sequentially to see how much data at that error level is needed to produce precise estimates (accumulation plots outlined in Sec. III F).

Small error, very small range. Figure 23(a) shows a multi-set plot for $H0_{(1\%)}$ at $k = 1$ with $k_{\max} = 3$, which is well into the evidence saturation region. As overfitting occurs with the uniform prior, data fluctuations lead to fluctuations in the coefficients with large errors. The results using the naturalness prior have a smaller spread about the central value and smaller error bands. As k_{\max} is increased, the overfitting and coefficient fluctuations in the uniform prior case become even more severe, while the results with the naturalness prior are insensitive to increasing k_{\max} . Figure 23(b) shows how the estimates improve as more data are accumulated (in sets of 10 sampled the same way as $H0_{(1\%)}$). Even with 300 points, the uniform prior results are not as precise as the naturalness prior results. The advantage of extra prior information is more in evidence in the top panel with a_1 , where the uniform prior error bands are much larger.

Large error, small range. The multi-set plot for $H1_{(5\%)}$ is shown in Fig. 24(a), this time for $k = 1$ with $k_{\max} = 4$, which is in the saturation region. Compared to $H0_{(1\%)}$, the fluctuations of estimates are much larger, due to the larger data error and sparser mesh, while the uniform-prior results are fluctuating even more due to overfitting. The accumulated results for the uniform prior in Fig. 24(b) are not consistent at the $1-\sigma$ level until about 150 points—and even there they still have errors that are larger due to different input information. The naturalness prior results are consistent with the true value even for only 10 points, but the error in the estimates does not decrease much as more data are added. Any further increase in precision is limited by the range

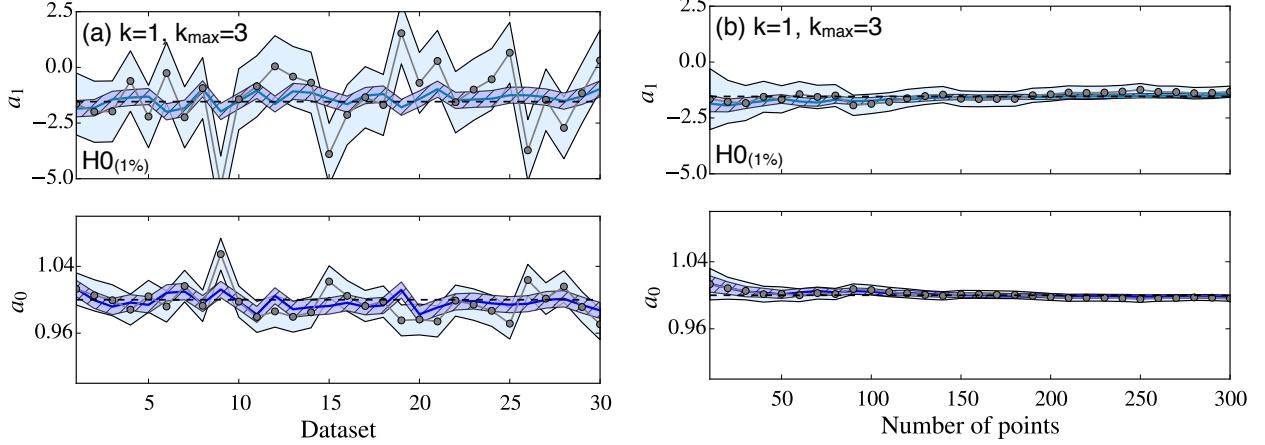


FIG. 23. Multi-set (a) and accumulation plots (b) calculated at $k = 1$, $k_{\max} = 3$. The shaded regions denote 68% error bands for the uniform (line with circles with lighter solid band) and naturalness prior (line with darker hatched band). The data sets used in (a) are 30 samples on the $H0_{(1\%)}$ mesh from Table IV. The same data are accumulated set by set to generate (b). In each case the prior was Set C' with $\bar{a}_{\text{fix}} = 5$.

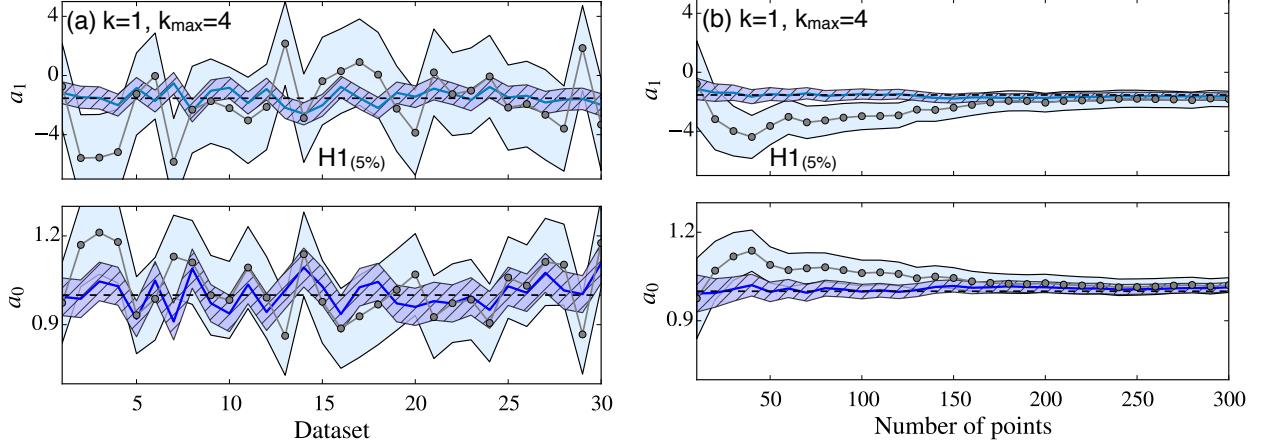


FIG. 24. Multi-set (a) and accumulation plots (b) calculated at $k = 1$, $k_{\max} = 4$ for data set $H1_{(5\%)}$. The description is otherwise the same as Fig. 23.

of the data and the fluctuations in it.

Large error, larger range. The multi-set and accumulated results for $H2_{(5\%)}$ are shown in Figs. 25(a) and 25(b) for $k = 1$, $k_{\max} = 4$. Compared to $H1_{(5\%)}$, the fluctuations between data sets in Fig. 25(a) are not as large for either prior, but the naturalness prior reduces the spread. The accumulated results in Fig. 25(b) show that the prior improves the results until there are about 75 data points available and the results with different priors become the same. However, for larger values of k_{\max} the uniform prior result will not converge to the Bayesian one as well as it does for $k_{\max} = 4$, as we saw for Model D in Fig. 11.

Small error, small range. The multi-set plot in Fig. 26(a) for the first two parameters given $H3_{(1\%)}$, marginalized to $k_{\max} = 4$, show tight error bands com-

pared to the previous results using 5% error data sets, as would be expected because of the smaller 1% relative error in the data. The estimates in Fig. 26(a) for the naturalness prior fluctuate statistically about the true value. Before, the results in the right-hand-side of Table VIII showed that the 68% interval missed the true parameter values, and a plot of results from that estimation in Fig. 27 at $k = 2$, $k_{\max} = 2$ (in the saturation region) shows a resulting prediction from that data set. Fig. 26(a) confirms that this was merely a result of the random error in that data set. This is important to keep in mind: the available physical data could lead to this type of estimate in an EFT. Thus statistically meaningful error bars are important so that resulting predictions have a true significance based on the data. The accumulated results in Fig. 26(b) show that even with just

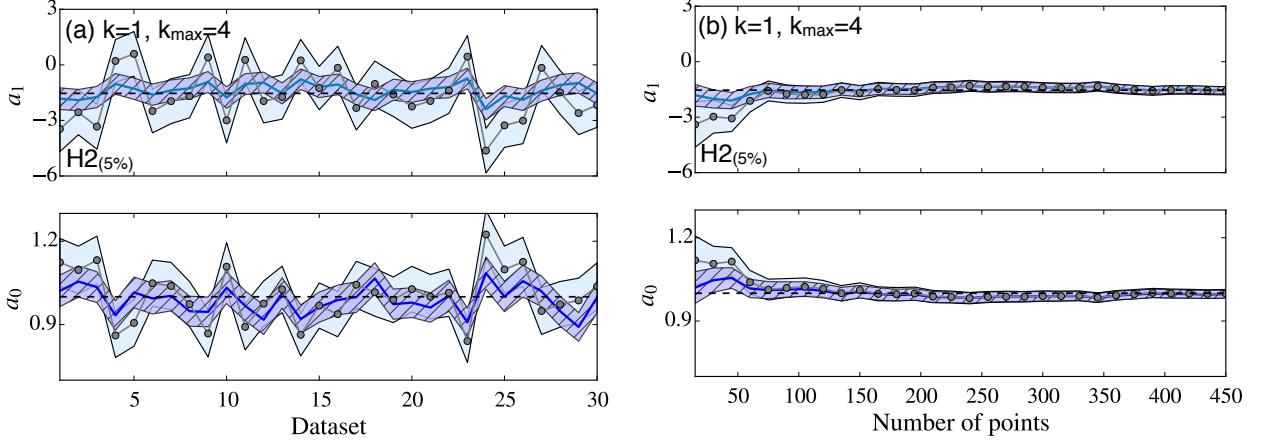


FIG. 25. Multi-set (a) and accumulation plots (b) calculated at $k = 1$, $k_{\max} = 4$ for data set $H2_{(5\%)}$. The description is otherwise the same as Fig. 23.

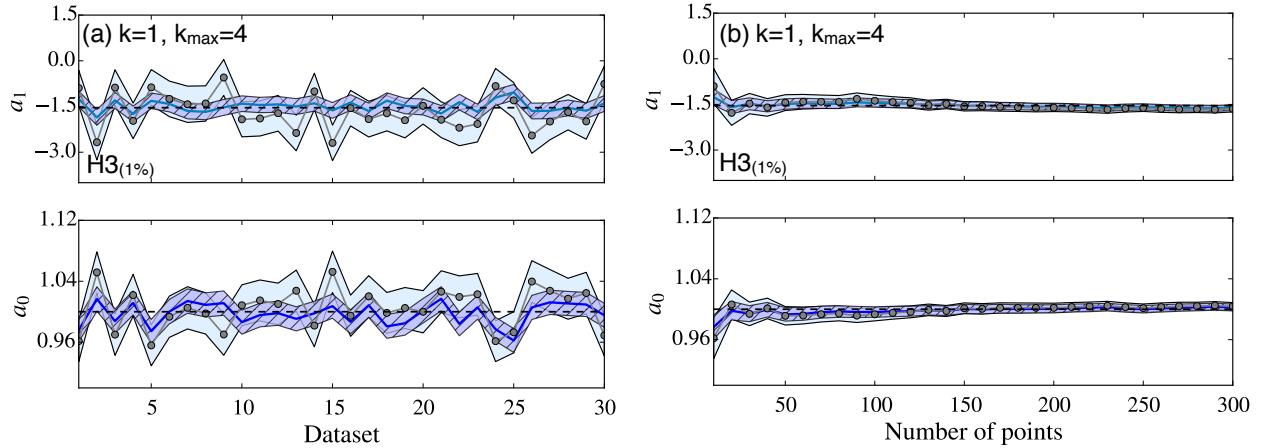


FIG. 26. Multi-set (a) and accumulation plots (b) calculated at $k = 1$, $k_{\max} = 4$ for data set $H3_{(1\%)}$. The description is otherwise the same as Fig. 23.

two data sets at this value of k_{\max} , the naturalness and uniform-prior results become the same.

Summary. The multi-set plots let us compare the impact of fluctuations on results using different prior information. In some cases the accumulated results can provide a sanity check that sufficient data will overwhelm the prior. The accumulated results also illustrate that data errors can severely limit what can be extracted from even a large amount of data when overfitting occurs. Using prior naturalness information about the coefficients can help even when many data are available.

E. Predictions

After the Validation stage of Fig. 13, the posteriors for the LECs are available to make predictions of observables, with consistent propagation of data and theory

uncertainties. Additional full loops in the flowchart may be appropriate to explore the impact on these predictions of different priors. In addition, we emphasize that truncation errors for predicted observables must be included for a full uncertainty quantification. We will not consider the Predictions stage further here, but reserve it for applications to actual EFTs in future work.

It should be evident that the Bayesian framework we advocate is far from automatic; there are many pitfalls along the way, but this is the nature of the problem. The simple case studies considered here illustrate some general truths about parameter estimation for EFTs: the size of data errors and the range of available data make a big difference; fluctuations happen; only so much information is available and the Bayesian evidence is necessary to quantify it. We have given examples of how the multiple tools we have outlined can be used together to optimize the extraction of the LECs and the uncer-

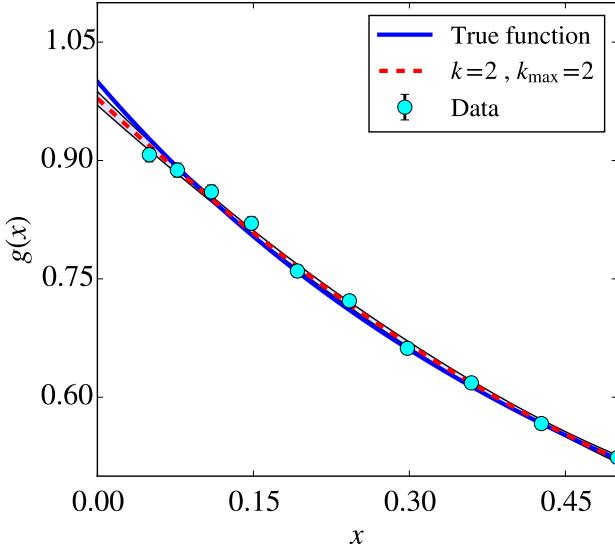


FIG. 27. Comparison of data set $H3_{(1\%)}$ (corresponding to the fourth row of Table IV), the underlying function for Model H from Eq. (34), and a Bayesian prediction using prior Set C' with $\bar{a}_{\text{fix}} = 5$ calculated at order $k = 2$, $k_{\max} = 2$ from that data set. The error bands represent 68% DoBs, which in this case are $1-\sigma$ bands in the Gaussian approximation.

tainty quantification. But we recognize the additional burden on the practitioner, which can introduce a significant and potentially prohibitive computational load. An important topic for further study is to what degree one can lighten this load while still working within the Bayes framework to provide robust statistical error bars (e.g., as in Ref. [35]). We caution that taking too many shortcuts may lose information from the data that could improve your LECs.

V. ADDITIONAL CASE STUDIES

In this section we consider an additional set of parameter estimation case studies, which highlight some particular difficulties that may be encountered in EFT applications. Our philosophy is that validating the procedures and testing the diagnostics with model problems (which at the least means analyzing the EFT with synthetic data), where we can explore a full range of issues in fitting LECs with known answers, is essential before turning to optimizations from real data. The following examples serve as prototypes for behavior we expect to see in “real-world” cases.

A. Blind test: detecting unnaturalness

We performed several different blind tests of the procedure of Sec. IV to verify that already knowing the answer has not been causing confirmation bias in our analyses.

For example, one might worry that our use of a naturalness prior by construction could guarantee that we will always find natural LECs unless we were already aware of unnatural coefficients. The corresponding blind test, described below, included an unnatural coefficient at a particular order, but the only information provided to the tester was a set of data points and a specification of the basis functions (the model EFT). The procedure set up in Sec. IV was applied to estimate the parameters of the model problem.

The first signals of unnaturalness should be detected in the Guidance stage, where the pdf of the naturalness parameter for the given prior choice is explored in the posterior pdf of \bar{a} : $\text{pr}(\bar{a}|D, k, k_{\max})$. During the Parameter estimation stage, an observation of unnaturalness should be confirmed by the evolution of the parameters with fixed \bar{a}_{fix} (using prior Set C' or A'), that is, the estimates will vary rapidly for expected natural values of \bar{a}_{fix} , but will stabilize once \bar{a}_{fix} is larger.

The case study we present here is a simple modification of Model D from Sec. III. In particular, an unnatural coefficient was added to the model of Eq. (23):

$$g(x) = \left(\frac{1}{2} + \tan\left(\frac{\pi}{2}x\right) \right)^2 + 20x^3, \quad (36)$$

which therefore has the same Taylor series as in Eq. (23) except for an unnatural coefficient at third order in x , altering the convergence rate—but not the convergence radius—of the expansion. Note that this is a somewhat extreme example because the unnaturalness is very large and occurs at a high order in the expansion. We refer to this as Model D̃. We consider a data set $\tilde{D}1_{(5\%)}$ that consists of 10 points sampled in the range $0 < x \leq 1/\pi$.

Beginning with the Guidance stage, we compute the evidence of $\tilde{D}1_{(5\%)}$ at several different orders and compute the posterior for \bar{a} . Looking at the evidence on the right-hand side of Table IX for Set C', we see that saturation occurs at $k_{\max} = 3$, but that the ratio to $k_{\max} = 2$ is not much greater than unity. Therefore, we expect that the first three coefficients— a_0 , a_1 , and a_2 —should be determined. It is likely that the unnatural coefficient's value can *not* be obtained with any real precision, making this a bastard of a blind test.

The posterior for \bar{a} using prior Set C is shown in Fig. 28 for $k_{\max} = 4$. This posterior indicates that the most probable values for \bar{a} lie in a wide range peaked around $\bar{a} = 5$, with a long tail extending to much larger values. This is a clear signature of unnaturalness. Choosing $\bar{a} < 5$ will likely cause the parameter estimates to be distorted by overly restricting the prior. We therefore choose $\bar{a}_{\text{fix}} = 15$ in the extractions of Table IX to avoid restricting the parameter space too much.

Unnaturalness is further confirmed during the Parameter estimation stage. Figure 29 shows how the first three coefficients, which should be rather well constrained by the data, change as a function of \bar{a}_{fix} using prior Set C'. Without the guidance of the dashed lines indicating the

TABLE IX. Coefficient estimates from sampling of $\text{pr}(\mathbf{a}|\tilde{\mathcal{D}}1_{(5\%)}, k, k_{\max})$ given the expansion from Eq. (3) (these results are controlled by k_{\max} only, see Sec. III A). The left side of the table is for a uniform prior, which is equivalent to a least-squares fit, and includes the χ^2/dof values. The right side of the table is using prior Set C' from Table I with $\bar{a}_{\text{fix}} = 15$, and includes the evidence. For both priors the posterior pdf is a multi-dimensional Gaussian.

		Uniform prior			Gaussian prior				
k	k_{\max}	χ^2/dof	a_0	a_1	a_2	Evidence	a_0	a_1	a_2
0	0	100	0.47 ± 0.01			~ 0	0.47 ± 0.01		
1	1	9.9	0.13 ± 0.01	3.8 ± 0.1		7.5×10^{-13}	0.13 ± 0.01	3.8 ± 0.1	
2	2	2.2	0.28 ± 0.02	0.30 ± 0.5	13 ± 2	5.1×10^0	0.28 ± 0.02	0.35 ± 0.4	13 ± 2
2	3	2.4	0.25 ± 0.04	1.3 ± 1	4.1 ± 9	5.4×10^0	0.27 ± 0.03	0.81 ± 0.7	8.6 ± 5
2	4	2.4	0.35 ± 0.1	-3.1 ± 3	63 ± 40	5.4×10^0	0.26 ± 0.03	0.88 ± 0.8	8.2 ± 5
2	5	2.9	0.36 ± 0.1	-3.7 ± 8	75 ± 100	5.4×10^0	0.26 ± 0.03	0.92 ± 0.8	8.1 ± 5
2	6	0.73	-0.54 ± 0.3	55 ± 20	-1300 ± 400	5.4×10^0	0.26 ± 0.03	0.88 ± 0.8	8.3 ± 5
True values		0.25	1.57	2.47			0.25	1.57	2.47

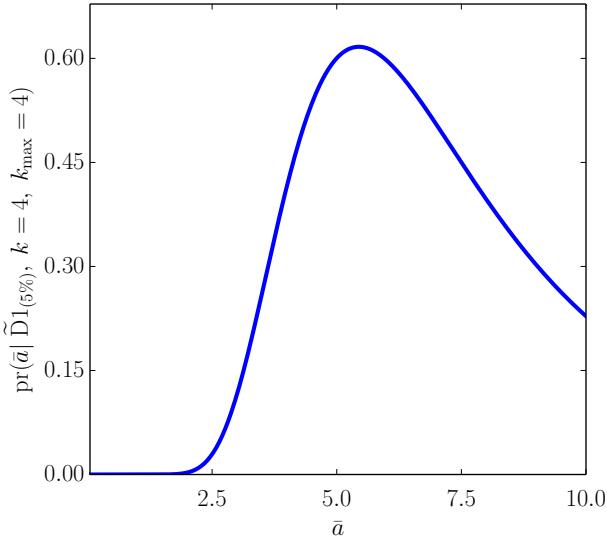


FIG. 28. The posterior pdf $\text{pr}(\bar{a}|D, k, k_{\max})$ assuming prior Set C from Table I with $\bar{a}_< = 0.05$ and $\bar{a}_> = 20$, given data set $\tilde{\mathcal{D}}1_{(5\%)}$ calculated at $k = 4$, $k_{\max} = 4$.

true values of the underlying expansion, there is a strong indication that a_2 is about 10, while the first two coefficients are approximately natural. The evolution plateaus between \bar{a}_{fix} of about 5 to 100, after which it rapidly transitions to the least-squares result.

It is also difficult to see where the unnaturalness is present in the expansion. In fact, if we did not marginalize over a_3 and a_4 with $k_{\max} = 4$, the results for these coefficients corresponding to that row in Table IX are $a_3 = 8.02 \pm 12$ and $a_4 = 5.32 \pm 14$. The unnaturalness is present in the estimate of a_3 , which is, however, rather poorly determined. The correlations inherent in a (naive) polynomial fit result in the unnaturalness being shared between a_3 and its neighbors. I.e., the presence of unnaturalness can distort the fit result of other coefficients which are natural. In this example it is—absent high-

quality data in an appropriate region of x —difficult to disentangle the specific order at which the one unnatural coefficient occurs.

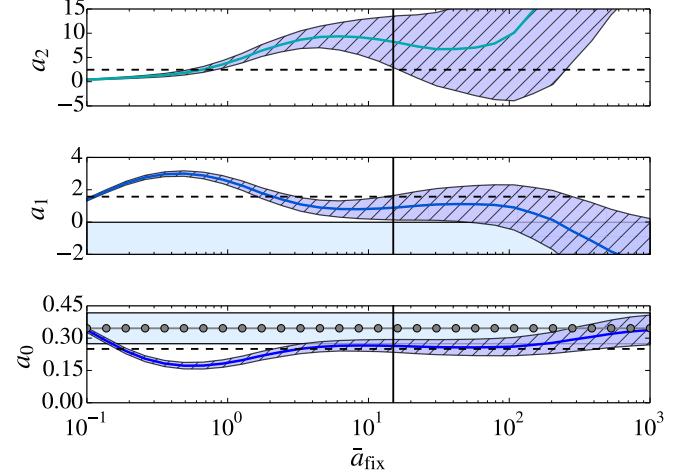


FIG. 29. (color online) Bayesian coefficient estimates (lines with darker hatched error bands) calculated at $k = 2$, $k_{\max} = 4$ as a function of \bar{a}_{fix} using prior Set C' given $\tilde{\mathcal{D}}1_{(5\%)}$. The constant line with circles (when visible on the chosen y-scale) with lighter solid error bands is the least-squares estimate, which is independent of \bar{a}_{fix} . The error bands represent 68% DoBs ($1-\sigma$ errors). The vertical lines in each panel at $\bar{a}_{\text{fix}} = 15$ indicate the value chosen for Set C' in the analysis.

In spite of this, unnaturalness in the EFT fit can be identified using the procedure of Sec. IV: there are clear signals in the Guidance and Parameter estimation stages. Disentangling at which order the unnatural LEC appears in the expansion is difficult, and the presence of an unnatural LEC can distort LECs which are in fact natural. With more information, such as knowing which orders have natural and unnatural contributions, we could update the prior pdfs for \mathbf{a} accordingly.

B. Data past breakdown of theory

An EFT expansion parameter will typically take the form of a ratio of scales that becomes one at the breakdown scale, where dynamics not explicitly included in the EFT appear. The analogous theory breakdown for our models is at the radius of convergence of the expansion. We explore estimating the parameters of Model H from Sec. IV using a data set sampled past the theory breakdown. In this case we have data set H4(5%), which is described in Table IV. We have now three points sampled at $x \geq 1.3$ where the expansion does not converge. Since the pole for Model H is at negative x , the function does not exhibit singular behavior at $x = 1.3$ and so the breakdown is “hidden”.

Our aim is to determine signatures of the EFT breakdown using this data set, i.e., determine whether the procedure of Sec. IV is sensitive to it. We compute the evidence and \bar{a} posteriors to complete the Guidance stage. The \bar{a} posteriors do not indicate unnaturalness for any order, and so we compute the evidence for H4(5%) using $\bar{a}_{\text{fix}} = 5$.

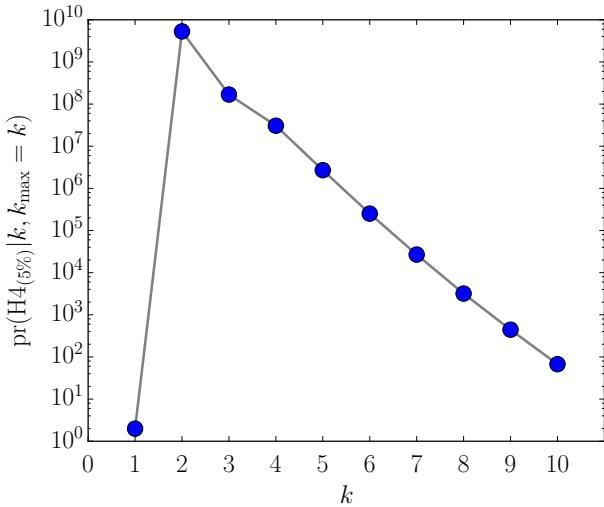


FIG. 30. Evidence $\text{pr}(\text{H4}_{(5\%)}|k, k_{\text{max}} = k)$ using prior Set C' with $\bar{a}_{\text{fix}} = 5$ for several values of k with $k_{\text{max}} = k$. (The evidence for $k = 0$ is not shown since it is nearly zero).

Figure 30 shows the evidence values. Compared to the saturation behavior using data sets sampled at smaller x values, the behavior of the evidence is strikingly different. The evidence decreases exponentially with increasing order of the expansion, and the pattern continues without saturating past the highest order shown in Fig. 30. We turn to the Parameter estimation stage to further examine the estimates with increasing k_{max} . Table X shows the parameter estimates for the coefficients up to $k_{\text{max}} = 6$. Although the results for a_0 and a_1 are fairly stable, the results for a_2 and beyond are not stable as we saw in other test cases. Examining the projected posterior plot in Fig. 31 at $k = k_{\text{max}} = 4$ gives us a clear picture of the interaction of the likelihood with

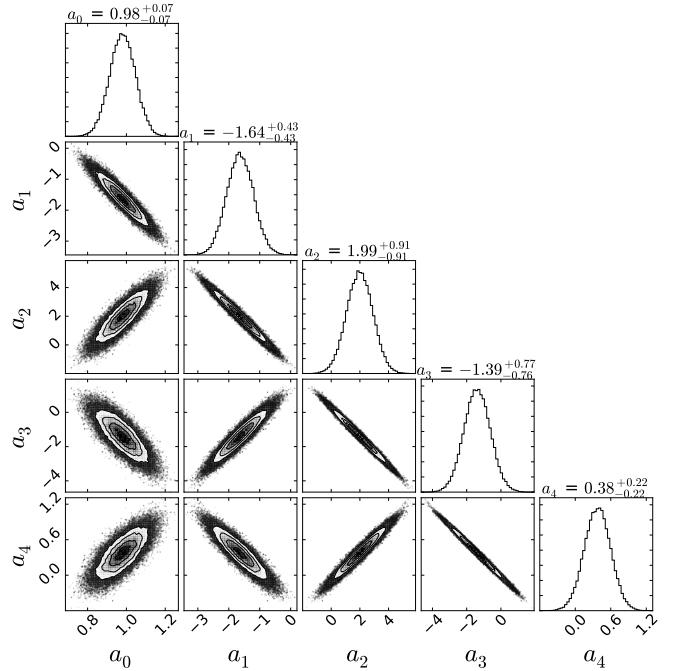


FIG. 31. Projected posterior plot (see Fig. 1) calculated at order $k = 4$, $k_{\text{max}} = 4$ given data set H4(5%) using prior Set C' with $\bar{a}_{\text{fix}} = 5$.

the prior in this case. We see tight correlations between coefficients remain in spite of the naturalness prior, unlike before where higher-order coefficients became almost uncorrelated with leading-order coefficients.

We continue with the Parameter estimation stage, examining the x_{max} plots for H4(5%), which compare the prior Set C' results with the uniform-prior results. Figure 32(a) shows the x_{max} plot for $k = k_{\text{max}} = 1$, and we continue up to $k = 1$, $k_{\text{max}} = 4$ in Fig. 32(d). In Fig. 32(b), which corresponds to the evidence peak in Fig. 30, the results are not stable with respect to x_{max} , and without relying on the dashed lines of the true values, no conclusive statement can be made. However, once we go to order $k_{\text{max}} = 4$ in Fig. 32(d), the results are fairly stable with respect to x_{max} , but a_0 and a_1 remain significantly correlated with the higher-order coefficients.

When we have data over a large range in the expansion parameter, the correlations between lower-order and higher-order coefficients may not be washed out by the naturalness prior. Since we have data at values of x where terms at different orders have similar contributions, the higher-order coefficients are not uniquely determined, resulting in large correlations and delicate cancellations. The likelihood should be further analyzed using singular value decomposition (SVD) to see the effects of the data. The example here shows that with enough low- x data, the leading behavior can still be extracted, but that further analysis is necessary to find the best way to utilize the available data.

TABLE X. Coefficient estimates from sampling of $\text{pr}(\mathbf{a}|\text{H4}_{(5\%)}, k, k_{\max})$ given the expansion from Eq. (3), and the evidence at each order (these results are controlled by k_{\max} only, see Sec. III A). The prior used in these estimates was Set C' from Table I with $\bar{a}_{\text{fix}} = 5$.

k	k_{\max}	Evidence	a_0	a_1	a_2	a_3	a_4
0	0	~ 0	0.31 ± 0.00				
1	1	1.5×10^1	0.69 ± 0.01	-0.34 ± 0.01			
2	2	8.8×10^{10}	0.86 ± 0.03	-0.78 ± 0.06	0.23 ± 0.03		
3	3	1.1×10^{10}	0.90 ± 0.04	-0.98 ± 0.2	0.49 ± 0.2	-0.10 ± 0.1	
4	4	2.1×10^9	0.98 ± 0.07	-1.6 ± 0.4	2.0 ± 0.9	-1.4 ± 0.8	0.38 ± 0.2
4	5	5.2×10^8	0.97 ± 0.08	-1.5 ± 0.7	1.5 ± 2	-0.59 ± 3	-0.15 ± 2
4	6	1.4×10^8	0.97 ± 0.08	-1.5 ± 0.7	1.5 ± 2	-0.31 ± 3	-0.70 ± 3
True values			1.0	-1.54	1.78	-1.82	1.75

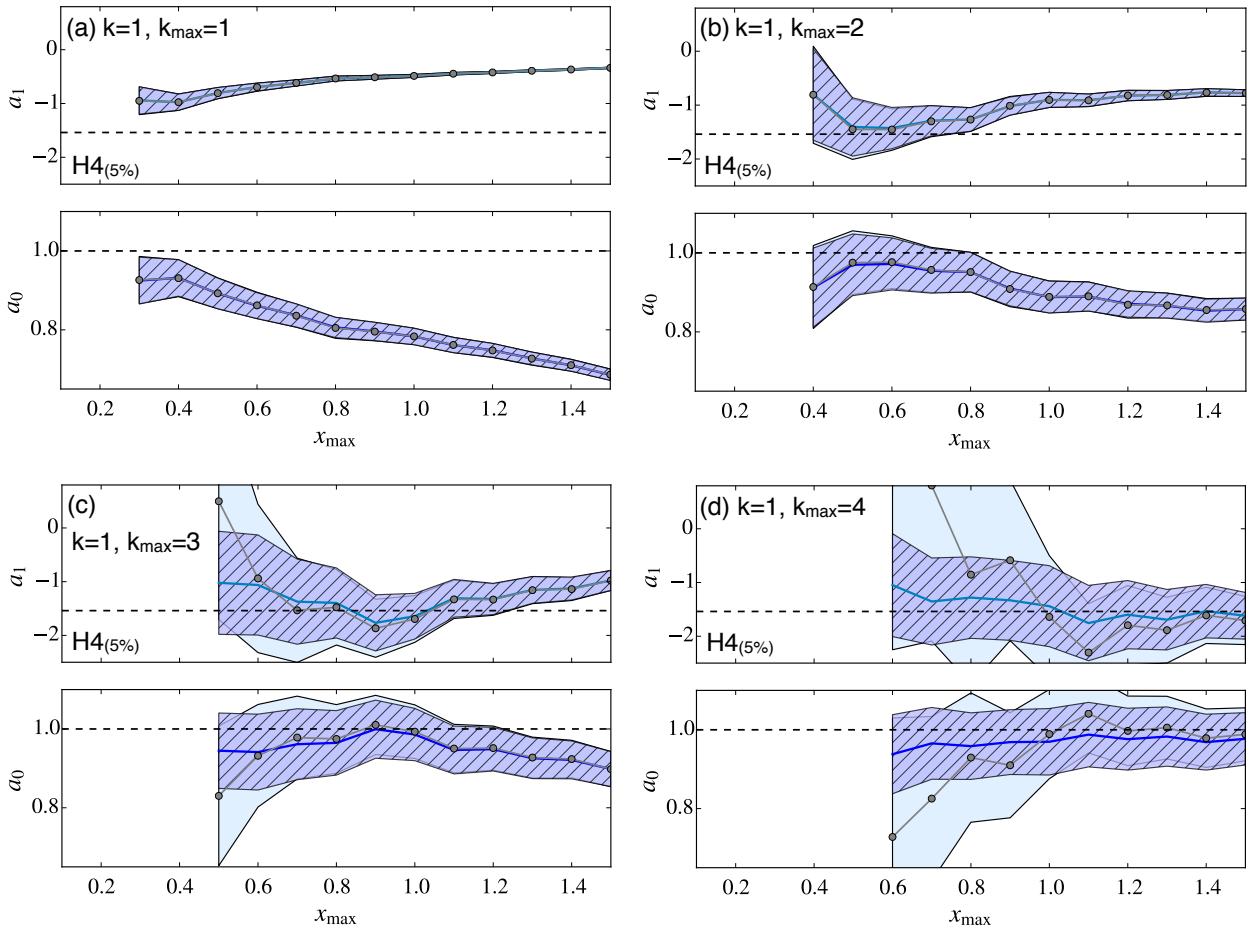


FIG. 32. Bayesian coefficient estimates as data from data set $\text{H4}_{(5\%)}$ are sequentially added at the high- x end. The largest x -value in the set is denoted as x_{\max} . The solid lines with darker hatched error bands represent estimates using prior Set C' with $\bar{a}_{\text{fix}} = 5$, and the lines with circles with lighter solid error bands represent the least-squares estimates. The error bands are 68% DoBs (1- σ errors), which coincide in (a) and (b).

C. Nucleon mass in χ PT

Lattice calculations of observables near the physical pion mass are computationally expensive. Therefore, there is a need for extrapolations to the physical point,

and chiral perturbation theory (χ PT) should be an ideal tool for extrapolations [36]. However, the lack of lattice data near (or below) the physical point means that extracting the coefficients in an expansion about zero pion mass can be difficult. It often turns out that we need to

disentangle contributions of different orders in the χ PT expansion at a pion mass where the expansion parameter is not particularly small. In this section we confront such difficulties, by applying our generic EFT tools for parameter estimation to chiral extrapolations from synthetic lattice data for the nucleon mass as a function of the pion mass, m .

At a technical level these χ PT expansions differ from the model problems described above since there can be several LECs at a single order, due to the possibility of different non-analytic functions multiplying the same overall power of m . Orders can also be skipped, e.g., there is no term $\propto m$ in the χ PT expansion of the nucleon mass. Therefore in this section we no longer refer to k and k_{\max} , replacing them by p and p_{\max} , which indicate χ PT orders, rather than numbers of LECs.

Following previous work in Ref. [13], we fit the LECs of the expansion for the nucleon mass from SU(2) χ PT. We are particularly interested in the first two low-order coefficients. The leading coefficient is the nucleon mass in the chiral limit, and the term at chiral order $p = 2$ is related to the pion-nucleon sigma term [37].

The nucleon mass expansion in terms of the intrinsic scale Λ is given by [13]:

$$\begin{aligned} \frac{M_N(m)}{\Lambda} = & \frac{M_0}{\Lambda} + \tilde{k}_1 \left(\frac{m}{\Lambda} \right)^2 + \tilde{k}_2 \left(\frac{m}{\Lambda} \right)^3 \\ & + \tilde{k}_3 \left(\frac{m}{\Lambda} \right)^4 \log \left(\frac{m}{\mu} \right) + \tilde{k}_4 \left(\frac{m}{\Lambda} \right)^4 \\ & + \tilde{k}_5 \left(\frac{m}{\Lambda} \right)^5 \log \left(\frac{m}{\mu} \right) + \tilde{k}_6 \left(\frac{m}{\Lambda} \right)^5 \\ & + \tilde{k}_7 \left(\frac{m}{\Lambda} \right)^6 \log \left(\frac{m}{\mu} \right)^2 + \tilde{k}_8 \left(\frac{m}{\Lambda} \right)^6 \log \left(\frac{m}{\mu} \right) \\ & + \tilde{k}_9 \left(\frac{m}{\Lambda} \right)^6 + \dots, \end{aligned} \quad (37)$$

where m is the pion mass, μ is the renormalization scale for the loop contributions, and the expansion is up to sixth order in χ PT. Since the expansion is about zero pion mass, ideally we would determine the free parameters from data sampled at small quark masses. We explore synthetic lattice data sampled at various ranges of m , investigating the feasibility of chiral extrapolations.

We can define scale-invariant LECs in this expansion by expressing the basis functions in terms of m/Λ , absorbing the μ dependence into the coefficients. If we wish to take into account the contributions at that order using marginalization, it is sufficient to account for the dominant power m^n . However, we do not make this approximation here, instead accounting for all terms in the chiral expansion up to sixth order, as written in Eq. (37).

Following Ref. [13], we generate synthetic data by taking $M_0 = 880$ MeV and computing values for the LECs of Eq. (37) at the renormalization scale $\mu = M_0$ with $\Lambda = 500$ MeV. These constants are given by (with $\tilde{k}_0 \equiv M_0/\Lambda$

and the relevant dimensions of $[{\rm GeV}]^{-n}$) [13, 38]

$$\begin{aligned} \tilde{k}_0 &= 1.76, & \tilde{k}_1 &= 1.92, & \tilde{k}_2 &= -1.41, & \tilde{k}_3 &= 0.81, \\ \tilde{k}_4 &= 1.03, & \tilde{k}_5 &= 2.97, & \tilde{k}_6 &= 4.41, & & \\ \tilde{k}_7 &= 0.4, & \tilde{k}_8 &= 0.31, & \tilde{k}_9 &= -3.13. & & \end{aligned} \quad (38)$$

At this value of Λ these LECs are approximately natural, and henceforth we use it in our calculations.

With these constants chosen, we add to the chiral expansion a model-dependent term that turns on slowly and becomes strong around $\Lambda = 500$ MeV which simultaneously turns off the expansion. The synthetic data are generated from the function [13]

$$M(m) = M_N(m) \left(1 - g \left(\frac{m}{\Lambda} \right) \right) + M_{\text{model}}(m) g \left(\frac{m}{\Lambda} \right). \quad (39)$$

This model-dependent term has the form

$$M_{\text{model}}(m) = \alpha + \beta m, \quad (40)$$

where the values $\alpha = 1$ GeV and $\beta = 1$ are selected following Ref. [13]. The function that controls the dominance of each term in Eq. (39) is [13]

$$g(x) = \frac{2}{\pi} \arctan(x^8). \quad (41)$$

Figure 33 shows a plot of the function of Eq. (39) from which synthetic data are sampled compared to the chosen chiral expansion alone and a data set $MN2_{(1.5\%)}$.

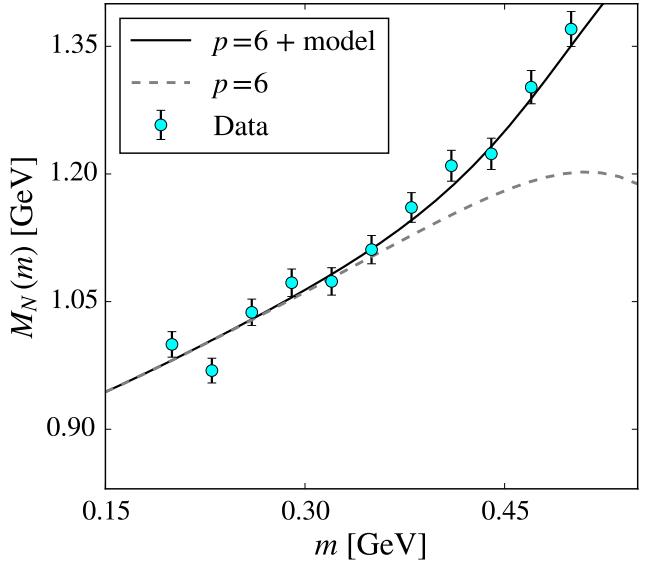


FIG. 33. (color online) Comparison of data set $MN2_{(1.5\%)}$ (corresponding to the third row of Table XI, the underlying expansion to order $p = 6$ from Eq. (37), and the function of Eq. (39), from which the data set was sampled).

We consider three data sets to represent different scenarios for lattice simulations. The pion mass meshes at which each was sampled are enumerated in Table XI.

TABLE XI. Nucleon mass data set labels for sampling grid ranges and number of points, where m is the pion mass.

Label	# of pts.	Grid	Spacing
MN0	11	$0.05 \leq m \leq 0.15$ GeV	linear
MN1	11	$0.14 \leq m \leq 0.44$ GeV	linear
MN2	11	$0.2 \leq m \leq 0.5$ GeV	linear

We first examine two realistic cases of synthetic data, $\text{MN2}_{(1.5\%)}$ and $\text{MN1}_{(1.5\%)}$, which are sampled at and above the physical pion mass with a 1.5% relative error. After seeing the limitations of data that simulates the current limitations of lattice calculations we explore an idealized set, $\text{MN0}_{(1.5\%)}$, where the sampling begins at $m = 50$ MeV.

We first analyze data sets that emulate currently available lattice calculations. $\text{MN1}_{(1.5\%)}$ is sampled with the lowest m -value near the physical point $m = 140$ MeV up to $m = 440$ MeV, where there is noticeable contribution to the generated data from the model-dependent part. In the Guidance stage, we consider the evidence in Table XII, which peaks at $p_{\max} = 2$ but decreases by two orders of magnitude from $p_{\max} = 2$ to $p_{\max} = 6$. There is no indication of unnaturalness in the \bar{a} posteriors at any order.

We proceed with the Parameter estimation, looking for stability with increasing orders of the extraction in accordance with the evidence calculations from the Guidance stage. Even though the evidence peaks at $p_{\max} = 2$, the parameter estimates in Table XII do not stabilize at that order. In fact, they do not stabilize until $p_{\max} = 4$. The result for \tilde{k}_0 is well-determined and consistent with the true value at the $1-\sigma$ level, but \tilde{k}_1 is rather poorly determined, although the estimate is consistent with the true value to within the quoted error estimates. With the data currently available to practitioners, lattice extrapolations will likely be difficult past the leading order.

We also examine the results of parameter estimation from $\text{MN2}_{(1.5\%)}$. A plot of this data set is shown in Fig. 33 compared to the underlying nucleon mass function used to generate the data and the chiral expansion up to sixth order. The data are sampled up to a region where the model-dependent term becomes significant. Table XIII shows the evidence and parameter estimates using prior Set C' compared to the uniform-prior results. The evidence shows no pleasing saturation behavior for this data set, perhaps indicating that we are sampling too near the breakdown of the expansion, as in Sec. V B. However, there is no indication of unnaturalness in the \bar{a} posterior at any order. It appears that the results are not stabilizing as the order is increased, but comparing the x_{\max} plots of Fig. 34(a) and Fig. 34(b) at $p_{\max} = 4$ and $p_{\max} = 5$ respectively, the final results look consistent and stable with x_{\max} . The coefficients are reproduced at the $1-\sigma$ level with rather large errors. Even extracting the leading coefficient in this case is very difficult, with the next-order coefficient mostly undeter-

mined. As in Sec. V B, it is not possible to extract conclusive results where the expansion cannot describe the data well.

We finally examine $\text{MN0}_{(1.5\%)}$, which is an idealized set sampled at significantly smaller pion masses than the previous two sets. We first compute the evidence and naturalness parameter posteriors to complete the Guidance stage for $\text{MN0}_{(1.5\%)}$. Figure 35 shows the posterior for \bar{a} assuming prior Set C at chiral order $p = 2$, $p_{\max} = 2$. We see no indications of unnaturalness in this data set, and again choose $\bar{a}_{\text{fix}} = 5$ for prior Set C'. Computing the evidence using this choice, the results in Table XIV show that the evidence reaches saturation at about $p_{\max} = 2$, where the values peak slightly, and then levels off at higher orders. The expectation is that only two coefficients will be determined by this data. Even in the unrealistic case where we have high-quality data below the physical pion mass, we do not expect to determine more than two LECs in the chiral expansion.

Turning to the Parameter estimation, the results in Table XIV show that the data set $\text{MN0}_{(1.5\%)}$ leads to an underestimate of the parameter \tilde{k}_0 compared to the true value, while \tilde{k}_1 has a better central value than was obtained from $\text{MN2}_{(1.5\%)}$, albeit still with a sizeable uncertainty. The underestimation of \tilde{k}_0 is another example of the effect of fluctuations in the data. The sizeable error on \tilde{k}_1 indicates that, although the coefficient is fixed by the data, the precision we can obtain is quite limited. At order $p_{\max} = 2$, which was the saturation order from the Guidance stage, the naturalness-prior parameter estimate for \tilde{k}_1 is spuriously precise—likely a result of underfitting of $\text{MN2}_{(1.5\%)}$ at this order. This illustrates the importance of going well into the region of p_{\max} where the evidence flattens out, in order to eliminate the possibility of underfitting. Examining the results using a uniform prior in Table XIV, we see very similar parameter estimates as for the naturalness prior at $p = p_{\max} = 2$, but any attempt to check the stability of that result by going to higher chiral order leads to extreme overfitting: by the time we reach $p_{\max} = 6$, where there are ten total coefficients, least squares is producing central values and 68% intervals of order 10^6 or 10^7 . The additional information imparted to the fit by the naturalness prior is essential to meaningful parameter estimation once we go to $p_{\max} = 3$. And going to $p_{\max} = 3$ is necessary in order to achieve stability of the extraction with chiral order.

The x_{\max} plots in Figs. 37(b) and 37(a) also show that stability with respect to the pion-mass interval over which data is fit is only achieved at order $p_{\max} = 3$ —in accordance with the inference from Table XIV that there is underfitting at $p_{\max} = 2$. The \bar{a} relaxation plot in Fig. 36 confirms a naturalness parameter \bar{a}_{fix} between 1 and about 5; this is consistent with the guidance from Fig. 35.

Finally, the Validation stage gives some indication of the effects of fluctuations in the data set, where we compare many independent samplings on the $\text{MN0}_{(1.5\%)}$ grid in the multi-set plot of Fig. 38(a), where we choose

TABLE XII. Coefficient estimates from sampling of $\text{pr}(\tilde{\mathbf{k}}|\text{MN1}_{(1.5\%)}, p, p_{\max})$ given the nucleon mass expansion from Eq. (37) (these results are controlled by p_{\max} only, see Sec. III A). The left side of the table is for a uniform prior, which is equivalent to a least-squares fit, and includes the χ^2/dof values. The right side of the table is using prior Set C' from Table I with $\bar{a}_{\text{fix}} = 5$, and includes the evidence. For both priors the posterior pdf is a multi-dimensional Gaussian.

		Uniform prior			Gaussian prior				
p	p_{\max}	χ^2/dof	\tilde{k}_0	\tilde{k}_1	\tilde{k}_2	Evidence	\tilde{k}_0	\tilde{k}_1	\tilde{k}_2
0	0	35	2.1 ± 0.01			~ 0	2.1 ± 0.01		
2	2	0.85	1.8 ± 0.02	0.82 ± 0.04		4.1×10^5	1.8 ± 0.0	0.82 ± 0.04	
3	3	0.96	1.8 ± 0.04	0.82 ± 0.4	0.002 ± 0.4	3.0×10^4	1.8 ± 0.0	0.82 ± 0.3	0.00 ± 0.4
3	4	0.91	1.6 ± 0.3	7.7 ± 10	-27 ± 60	1.4×10^4	1.8 ± 0.1	2.1 ± 1	-1.2 ± 4
3	5	0.98	5.1 ± 4	-490 ± 600	5800 ± 6000	6.9×10^3	1.8 ± 0.1	1.9 ± 2	-0.97 ± 4
3	6	0.53	-2000 ± 1000	$8.8 \times 10^5 \pm 5 \times 10^5$	$-2.4 \times 10^7 \pm 1 \times 10^7$	3.9×10^3	1.8 ± 0.1	1.8 ± 2	-0.83 ± 4
True values		1.76	1.92	-1.41			1.76	1.92	-1.41

TABLE XIII. Same as Table XII except sampling from $\text{pr}(\tilde{\mathbf{k}}|\text{MN2}_{(1.5\%)}, p, p_{\max})$.

		Uniform prior			Gaussian prior				
p	p_{\max}	χ^2/dof	\tilde{k}_0	\tilde{k}_1	\tilde{k}_2	Evidence	\tilde{k}_0	\tilde{k}_1	\tilde{k}_2
0	0	51	2.2 ± 0.01			~ 0	2.2 ± 0.01		
2	2	1.4	1.8 ± 0.02	0.89 ± 0.04		1.4×10^4	1.8 ± 0.02	0.89 ± 0.04	
3	3	1.4	1.9 ± 0.1	0.49 ± 0.4	0.37 ± 0.3	1.8×10^3	1.9 ± 0.1	0.47 ± 0.4	0.38 ± 0.3
3	4	1.7	2.6 ± 0.7	-21 ± 20	89 ± 80	3.1×10^2	1.8 ± 0.1	0.98 ± 2	0.10 ± 4
3	5	1.2	37 ± 20	-2700 ± 1000	$2.5 \times 10^4 \pm 1 \times 10^4$	1.2×10^2	1.8 ± 0.1	0.61 ± 2	0.49 ± 4
3	6	0.20	450 ± 8000	$-1.3 \times 10^5 \pm 2 \times 10^6$	$3.1 \times 10^6 \pm 5 \times 10^7$	6.0×10^1	1.9 ± 0.2	0.14 ± 2	0.87 ± 4
True values		1.76	1.92	-1.41			1.76	1.92	-1.41

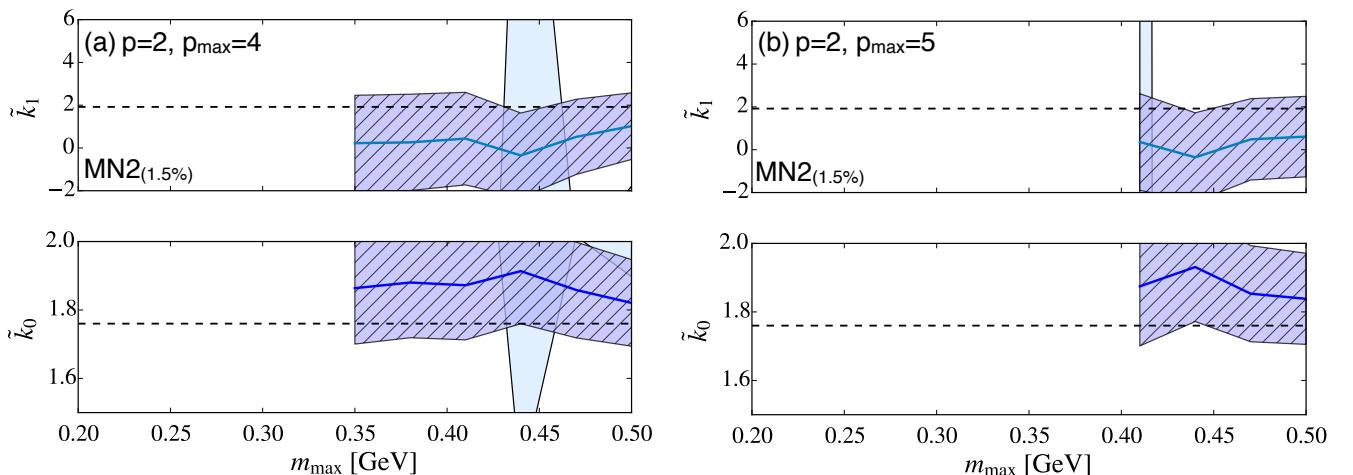


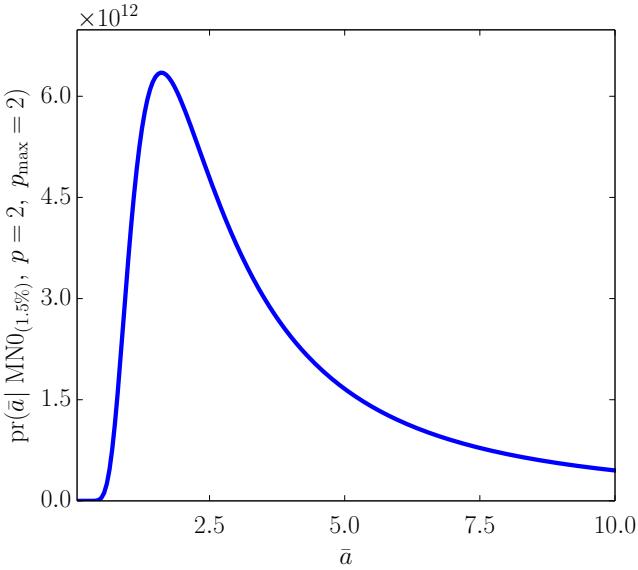
FIG. 34. (color online) Bayesian coefficient estimates as data from data set $\text{MN2}_{(1.5\%)}$ are sequentially added at the high- m end. The largest m -value in the set is denoted as m_{\max} . The solid lines with darker hatched error bands represent estimates using prior Set C' with $\bar{a}_{\text{fix}} = 5$, and the line with circles (off-scale in these panels) with lighter solid error bands represents the least-squares estimates. The error bands represent 68% DoBs ($1-\sigma$ errors).

$p_{\max} = 4$. The fluctuations in the results using the uniform prior are severe, and the results using prior Set C' fluctuate much less. Comparing the results for the first two coefficients at $p_{\max} = 4$ with these 30 independent data sets to the previous estimates with our single data set $\text{MN0}_{(1.5\%)}$, we conclude that the naturalness prior

result for \tilde{k}_0 in Table XIV is consistent with the fluctuations seen in Fig. 38(a). At the same order we examine the accumulation plot with these data sets in Fig. 38(b). The results for \tilde{k}_0 are consistent to within $1-\sigma$ when we have about 50 data points, but the error reduces significantly once we have more than 300 points. The error in

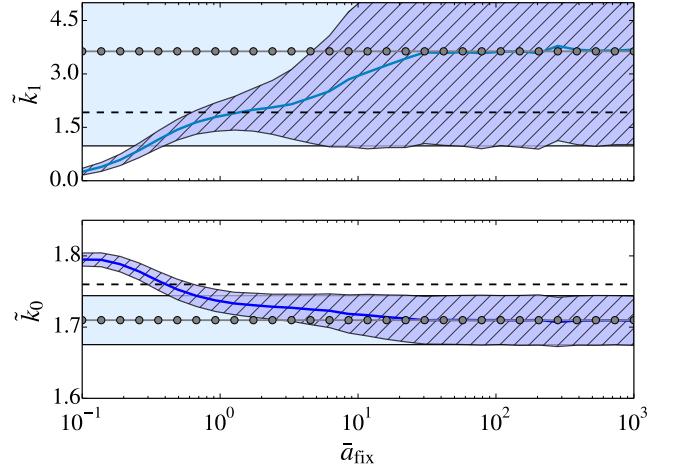
TABLE XIV. Same as Table XII except sampling from $\text{pr}(\tilde{\mathbf{k}}|\text{MN}0_{(1.5\%)}, p, p_{\max})$.

		Uniform prior			Gaussian prior				
p	p_{\max}	χ^2/dof	\tilde{k}_0	\tilde{k}_1	\tilde{k}_2	Evidence	\tilde{k}_0	\tilde{k}_1	\tilde{k}_2
0	0	5.4	1.82 ± 0.01			0.015	1.82 ± 0.01		
2	2	1.1	1.73 ± 0.02	2.2 ± 0.33		4.0×10^6	1.73 ± 0.02	2.2 ± 0.3	
3	3	1.2	1.71 ± 0.03	3.7 ± 2.7	-4.7 ± 8	3.3×10^6	1.72 ± 0.02	2.4 ± 1.4	-0.72 ± 4
3	4	1.3	2.1 ± 0.3	-140 ± 110	2000 ± 2000	3.1×10^6	1.72 ± 0.02	2.5 ± 1.4	-0.66 ± 4
3	5	1.9	1.5 ± 5	450 ± 5000	$-16000 \pm 2 \times 10^5$	3.1×10^6	1.72 ± 0.02	2.5 ± 1.4	-0.60 ± 4
3	6	0.25	1500 ± 1000	$-4.9 \times 10^6 \pm 5 \times 10^6$	$3.7 \times 10^7 \pm 4 \times 10^7$	3.1×10^6	1.72 ± 0.02	2.5 ± 1.4	-0.61 ± 4
True values		1.76	1.92	-1.41			1.76	1.92	-1.41

FIG. 35. (color online) The posterior pdf $\text{pr}(\bar{a}|D, p, p_{\max})$ calculated at $p = 2, p_{\max} = 2$ using prior Set C from Table I with $\bar{a}_{<} = 0.05$ and $\bar{a}_{>} = 20$, given data set $\text{MN}0_{(1.5\%)}$.

\tilde{k}_1 reduces slightly once we have over 300 data, but at this level of error in the data it is still difficult to extract \tilde{k}_1 with small uncertainty. The uniform-prior results for \tilde{k}_1 in Fig. 38(b) are much less precise than the results using naturalness information, even though \tilde{k}_0 is well determined by these data and only a uniform prior.

In summary, extrapolating the nucleon mass in the chiral limit from lattice data is a difficult problem due to the lack of data at low pion masses, the level of error in available lattice data, and the contributions of several coefficients at the same order. Attempting to estimate LECs of the chiral expansion of the nucleon mass past chiral order $p_{\max} = 2$ will be very difficult for these reasons. Constraints on terms non-analytic in the quark mass will certainly assist in the extraction, but extracting \tilde{k}_0 and \tilde{k}_1 is still problematic. The use of the Bayesian parameter estimation procedure laid out in this work will assist practitioners in obtaining chiral extrapolations with credible uncertainty estimates, which can be improved as new

FIG. 36. (color online) Bayesian coefficient estimates (lines with darker hatched error bands) calculated at $p = 2, p_{\max} = 3$ as a function of \bar{a}_{fix} using prior Set C' given $\text{MN}0_{(1.5\%)}$. The constant line with circles with lighter solid error bands is the least-squares estimate, which is independent of \bar{a}_{fix} . The error bands represent 68% DoBs ($1-\sigma$ errors).

data becomes available.

D. Nonlinear models

The models considered so far are linear in the coefficients that represent LECs. However, in many EFTs of interest, such as chiral EFT, the observables are nonlinear functions of the LECs and can be more complicated and/or expensive to compute. Here we give a brief example to show the effectiveness of our procedures for cases in which the observable is nonlinear in the coefficients. We will address in future work cases where calculating an observable to order k_{\max} is not practical, and how to circumvent this issue using the naturalness of the expansion of the observable used in the fit, as discussed in Sec. III A. This will be relevant in applications to real EFTs such as chiral EFT. We will now briefly discuss one model example to demonstrate the utility of the diagnostics to observables nonlinear in the parameters.

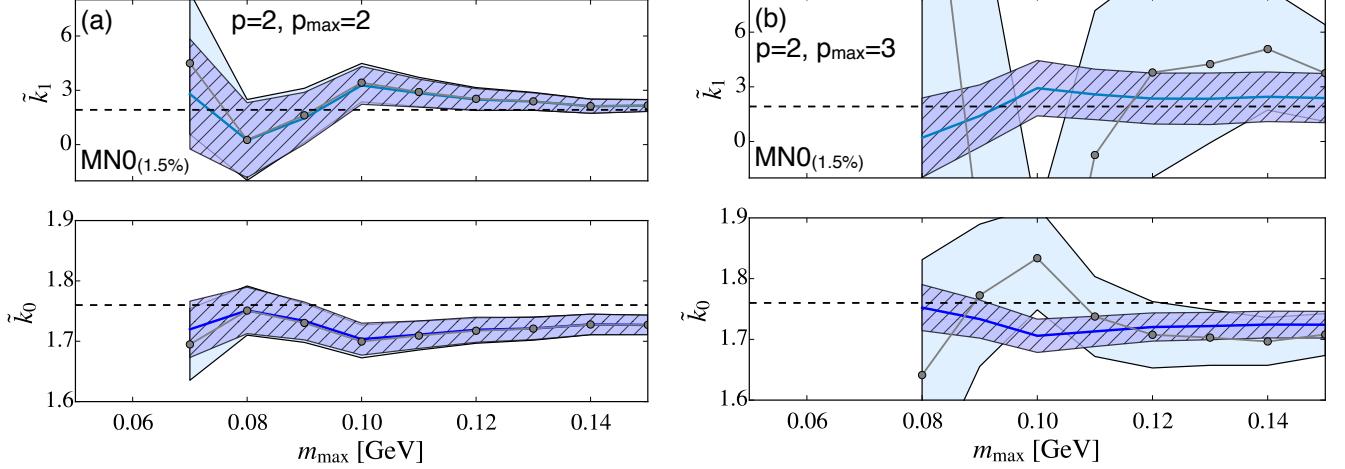


FIG. 37. (color online) Bayesian coefficient estimates as data from data set $MN0_{(1.5\%)}$ are sequentially added at the high- m end. The largest m -value in the set is denoted as m_{\max} . The lines with darker hatched error bands represent estimates using prior Set C' with $\bar{a}_{\text{fix}} = 5$, and the line with circles with lighter solid error bands represents the least-squares estimates. The error bands represent 68% DoBs ($1-\sigma$ errors).

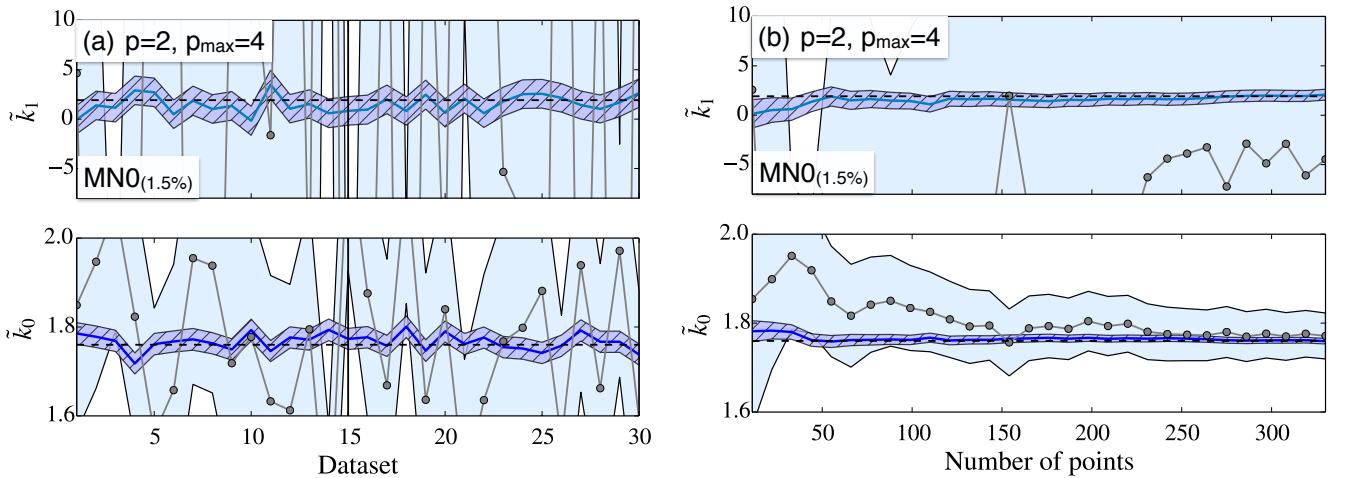


FIG. 38. (color online) Multi-set (a) and accumulation plots (b) calculated at $p = 2$, $p_{\max} = 4$. The shaded regions denote 68% error bands for the uniform (line with circles with lighter solid band) and naturalness prior (line with darker hatched band). The data sets used in (a) are 30 samples on the $MN0_{(1.5\%)}$ mesh from Table XI. The same data are accumulated set by set to generate (b). The naturalness prior used was Set C' with $\bar{a}_{\text{fix}} = 5$.

We define a model observable (called Model T) for this example by

$$T[g(x)] = \frac{1}{1 + g(x)}, \quad (42)$$

where $g(x)$ can represent any of the linear expansions we have considered. We will consider the model linear observable $g(x)$ from Sec. IV, Model H defined in Eq. (34). This model will be referred to by Model T[H]. The theoretical prediction for the observable is

$$T[g_{\text{th}}(x)] = \frac{1}{1 + g_{\text{th}}(x)}, \quad (43)$$

where $g_{\text{th}}(x)$ is defined in Eq. (3), and this observable can be calculated to order k using the vector of coefficients \mathbf{a} up to that order. Using Eq. (26), we can compute the posterior for the coefficients \mathbf{a} with a simple modification to the likelihood. Since the observable is no longer linear in \mathbf{a} , the likelihood becomes significantly more complicated, and the repeated computations necessary for sampling the posterior or computing the evidence can become cumbersome.

We will consider a single data set $T[H]0_{(1\%)}$ in this section, which is sampled on a linear mesh in the range $0.01 \leq x \leq 0.1$, representing a high-quality data set sampled at small expansion parameter values, similar to the

TABLE XV. Coefficient estimates from sampling of $\text{pr}(\mathbf{a}|T[H]0_{(1\%)}, k, k_{\max})$ given the expansion from Eq. (3) (these results are controlled by k_{\max} only, see Sec. III A). The left side of the table is for a uniform prior, which is equivalent to a least-squares fit, and includes the χ^2/dof values. The right side of the table is using prior Set C' from Table I with $\bar{a}_{\text{fix}} = 5$, and includes the evidence. For both priors the posterior pdf is well-approximated by a multi-dimensional Gaussian.

		Uniform prior				Gaussian prior			
k	k_{\max}	χ^2/dof	a_0	a_1	a_2	Evidence	a_0	a_1	a_2
0	0	5.6	0.92 ± 0.01			9.1×10^4	0.92 ± 0.01		
1	1	1.5	0.99 ± 0.01	-1.3 ± 0.21		6.9×10^{11}	0.99 ± 0.01	-1.3 ± 0.2	
2	2	1.5	1.0 ± 0.02	-2.6 ± 0.93	11 ± 8	7.4×10^{11}	0.99 ± 0.02	-1.6 ± 0.5	2.9 ± 4
2	3	1.6	1.0 ± 0.04	-4.6 ± 2.8	57 ± 60	7.4×10^{11}	0.99 ± 0.02	-1.6 ± 0.5	2.8 ± 4
2	4	1.2	1.1 ± 0.1	-18 ± 7.3	540 ± 300	7.4×10^{11}	0.99 ± 0.02	-1.6 ± 0.5	2.9 ± 4
2	5	1.2	1.0 ± 0.1	1.3 ± 18	-460 ± 900	7.4×10^{11}	0.99 ± 0.02	-1.6 ± 0.5	2.9 ± 4
2	6	1.5	1.1 ± 0.3	-22 ± 47	1100 ± 3000	7.4×10^{11}	0.99 ± 0.02	-1.6 ± 0.5	3.0 ± 4
True values		1.0	-1.54	1.78			1.0	-1.54	1.78

analysis of $H0_{(1\%)}$ for the linear Model H in Sec. IV. We expect to reproduce the leading behavior of the underlying expansion well using these data. Table XV shows results up to $k = 2$ for the marginalized posteriors at different k_{\max} s, comparing results using a uniform prior and Set C'. We see that, as in the linear cases, overfitting occurs when naturalness information is not included, starting as early as $k_{\max} = 2$ in the left column of Table XV. We will now demonstrate some of the diagnostics from the flowchart in Fig. 13.

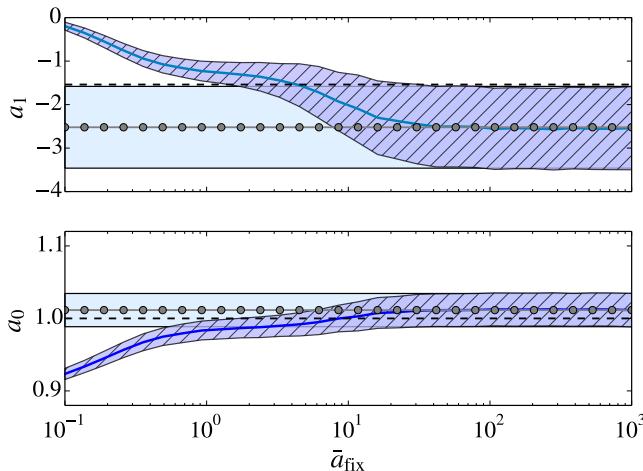


FIG. 39. (color online) Bayesian coefficient estimates (lines with darker hatched error bands) calculated at $k = 1$, $k_{\max} = 2$ as a function of \bar{a}_{fix} using prior Set C' given $T[H]0_{(1\%)}$. The constant line with circles with lighter solid error bands is the least-squares estimate, which is independent of \bar{a}_{fix} . The error bands represent 68% DoBs ($1-\sigma$ errors).

The evidence for different values of k_{\max} is shown in Table XV, and we see that the evidence saturates around $k_{\max} = 2$, while the ratio of the evidence at $k_{\max} = 2$ to $k_{\max} = 1$ is only about 1.1, indicating that no more information is gained for a_2 from this data set. Using prior Set C' with $\bar{a}_{\text{fix}} = 5$ results in excellent agreement

of a_0 and a_1 with their true values, but with a_2 not well-constrained. The results for the posterior maxima in the right side of Table XV confirm the guidance from the evidence that extractions past first order will not contain much information.

In terms of choosing a prior, Fig. 39 demonstrates that at $k_{\max} = 2$, the first two coefficients are consistent with the true values for \bar{a}_{fix} between about 2 and 10. We could also compute the \bar{a} posterior to analyze naturalness information contained in the data. Using the simplified prior Set C' with $\bar{a}_{\text{fix}} = 5$ therefore produces consistent results in this case.

The inclusion of additional information in the parameter estimation is beneficial, as it was in Sec. IV. In this nonlinear problem, the procedure results in the best possible parameter estimates with the available information. Even at $k_{\max} = 2$ overfitting occurs when naturalness information is not included, and the multi-set plot of Fig. 40(a) shows the benefit of the naturalness prior—and that the consequences of statistical fluctuations are still unavoidable, even with the high-quality data set $T[H]0_{(1\%)}$. The accumulation plot in Fig. 40(b) shows that the prior is beneficial even for a large data set until about 10 data sets are available, at which point the uniform prior and naturalness prior results become the same.

The procedures detailed in Fig. 13 are thus directly applicable to parameter estimation in cases when the available data are related nonlinearly to the parameters. We have demonstrated in a simple test case the success of the diagnostics with respect to the type of observable used. Of course, the limitations identified in this section and in the test cases of Sec. IV will still be present in nonlinear applications, but the major theme persists: the best parameter estimates are obtained by using all available information about the problem at hand.

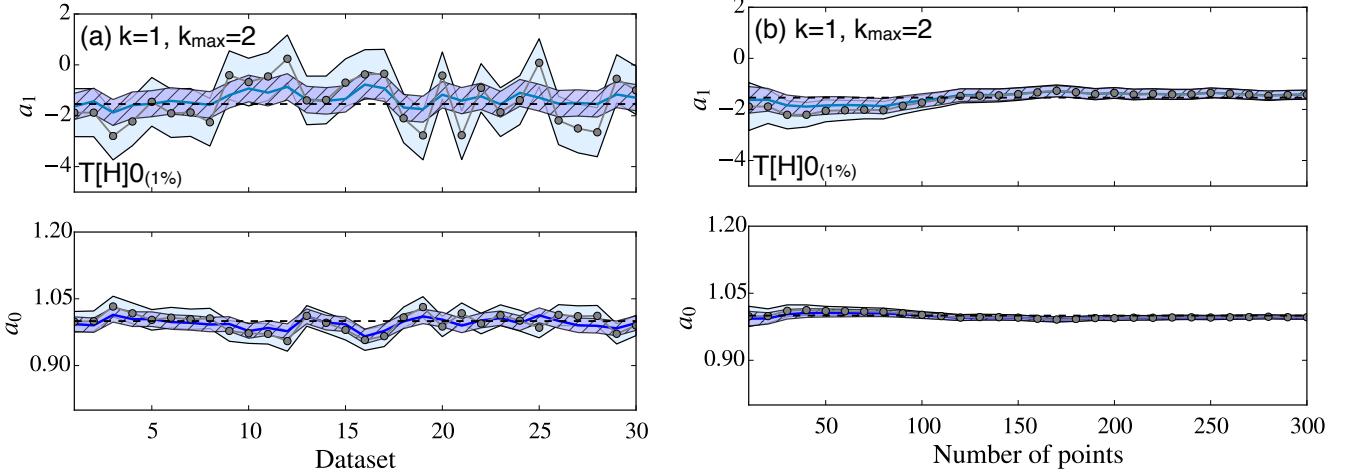


FIG. 40. (color online) Multi-set (a) and accumulation plots (b) calculated at $k = 1$, $k_{\max} = 2$. The shaded regions denote 68% error bands for the uniform (line with circles with lighter solid band) and naturalness prior (line with darker hatched band). The data sets used in (a) are 30 samples on the $T[H]0_{(1\%)}$ mesh which is the same as the $H0_{(1\%)}$ mesh from Table IV. The same data are accumulated set by set to generate (b). In each case the prior was Set C' with $\bar{a}_{\text{fix}} = 5$

VI. SUMMARY AND OUTLOOK

In this paper, we present a detailed framework for EFT parameter estimation using Bayesian statistics, which naturally combines statistical and systematic uncertainties. Fitting the LECs in an EFT poses a basic dilemma: we want to use more data to suppress the impact of data errors and avoid overfitting, but the accuracy decreases for larger values of the EFT expansion parameters, so using too much can lead to underfitting. By marginalizing over omitted higher-order terms using naturalness priors, we are able to use *all* of the data, without having to decide a particular value of the EFT parameter at which to break off the fit. In particular, a prior on naturalness suppresses overfitting by limiting how much LECs can play off each other. This in turn means we find stability with respect to expansion order and amount of data.

This success is not achieved automatically, however; there are multiple ways the parameter estimation can go wrong. By studying simple models reflecting the possible behaviors of an EFT, we have identified pitfalls and ways to address them. We have developed a suite of diagnostic tools to guide and analyze the fit of LECs. These diagnostics:

- use the Bayes evidence ratio to find how many LECs can be determined by the data;
- identify potential sensitivity to the choice of prior;
- trace the way the prior eliminates spurious correlations between undetermined coefficients;
- indicate the region in the EFT parameter where a particular LEC is determined;

- show the impact that fluctuations in the data have on the parameter estimation.

They are summarized in Table II and a possible flowchart for applying them in parameter estimation is given in Fig. 13. Multiple examples in Sec. III, IV, and V illustrate how to use these tools. These examples are not exhaustive but highlight some important features of Bayesian uncertainty quantification for EFT parameter estimation.

Our goal has been to make this framework adaptable to a wide range of EFT applications. We plan to test the framework in applications such as fitting LECs for the nucleon mass chiral expansion with genuine lattice data and for chiral nucleon-nucleon and three-nucleon forces. This will require us to address practical challenges such as the computational cost of MCMC evaluations, which may require compromises in our procedures. It will also be of interest to investigate whether augmentations of basic least-squares fitting procedures commonly used by practitioners to address the special features of EFTs can be justified within a Bayesian framework.

While we have focused here on parameter estimation, we believe that the Bayesian framework will enable practitioners to consistently achieve all the general goals of UQ for EFT calculations:

- reflect all sources of uncertainty in an EFT prediction using a likelihood or prior for each;
- compare theory predictions and experimental results statistically, with error bands as Bayesian degree-of-belief (DoB) intervals;
- distinguish uncertainties from IR (long-range) vs. UV (short-range) physics by avoiding overfitting;

- provide guidance on how to better extract LECs by exploiting how Bayes’ theorem propagates new info (e.g., will an additional or better measurement or lattice calculation help and by how much?);
- test whether EFT is working as advertised — do our predictions exhibit the anticipated systematic improvement? — by analyzing the trends of DoB intervals and applying model selection.

Investigations are in progress to extend and validate our diagnostics and procedures for each of these goals.

ACKNOWLEDGMENTS

We thank H. Griesshammer, D. Higdon, J. McGovern, E. Lawrence, A. Peter, M. Schindler, A. Steiner, and S. Wild for useful discussions. Useful feedback on the manuscript was provided by A. Dyhdalo, M. Heinz, S. Koenig, S. More, R. Perry, and C. Plumberg. This work was supported in part by the National Science Foundation under Grant No. PHY-1306250, the U.S. Department of Energy under grant DE-FG02-93ER40756, the NUCLEI SciDAC Collaboration under DOE Grant DE-SC0008533, and the Seattle Chapter of the Achievement Rewards for College Scientists Foundation.

Appendix A: Details on implementation of Markov Chain Monte Carlo via emcee

MCMC sampling methods are particularly suited to efficiently sample pdfs with large parameter spaces [29], and have been applied widely in many physical applications, such as astrophysics [27, 39]. We apply a particular MCMC sampling algorithm called emcee [29], which is a Python implementation of Goodman and Weare’s affine-invariant ensemble sampler for MCMC. The goal of the algorithm is to improve over traditional Metropolis-Hastings (M-H) methods [40] and modifications that require significant burn-in phases with the price of many computationally expensive evaluations of the pdf. Reference [41] contains summaries of several methods for posterior simulation, including the aforementioned M-H methods and Gibbs sampling.

The emcee code uses an ensemble of several walkers to explore the parameter space. Each starts in a user-specified part of the parameter space and is allowed to diffuse, exploring different features of the distribution in parallel. The number of walkers needed to sufficiently explore the parameter space generally depends on the problem of interest, but Ref. [29] suggests using hundreds of walkers. For the simple model problems in this work we find it is sufficient to use a number of walkers equal to twice the number of parameters, and to be safe we use four times the number of parameters.

The dangers that occur during optimization can also occur with MCMC walkers— they can find unlikely parts

of the parameter space and become “stuck”. Assessment of the convergence of the walkers is crucial to diagnose such problems. Sometimes walkers that are not converging can be identified graphically, but several useful quantitative methods exist to diagnose the convergence. Foreman-Mackey et al. advocate using the autocorrelation time, which quantifies the number of evaluations needed to obtain independent samples of the distribution [29]. When the autocorrelation time is large, the number of evaluations needed by the MCMC algorithm to achieve convergence will be large.

They also outline some proxies for assessing convergence [29]. One proxy is the fraction of proposed steps that were accepted in the sampling, called the acceptance fraction a_f , which can be used to diagnose walkers that did not converge during the sampling. We use this simple diagnostic to ensure that walkers converged in our MCMC sampling throughout this work, using the rejection criteria $a_f < 0.2$ for each separate walker. Low acceptance fractions are a sign that the sampler is stuck in an area of low probability, which often occurs in multimodal distributions. Some methods to deal with multimodality include clustering algorithms [39] and parallel-tempering ensemble MCMC, which is implemented in emcee.

-
- [1] D. B. Kaplan, (1995), arXiv:nucl-th/9506035 [nucl-th].
- [2] D. R. Phillips, Czech. J. Phys. **52**, B49 (2002), arXiv:nucl-th/0203040 [nucl-th].
- [3] C. P. Burgess, Ann. Rev. Nucl. Part. Sci. **57**, 329 (2007), hep-th/0701053.
- [4] E. Epelbaum, (2010), arXiv:1001.3229 [nucl-th].
- [5] S. Descotes-Genon, N. H. Fuchs, L. Girlanda, and J. Stern, Eur. Phys. J. C **34**, 201 (2004), arXiv:hep-ph/0311120 [hep-ph].
- [6] T. Ledwig, J. Nieves, A. Pich, E. Ruiz Arriola, and J. Ruiz de Elvira, Phys. Rev. D **90**, 114020 (2014), arXiv:1407.3750 [hep-ph].
- [7] X. Zhang, K. M. Nollett, and D. R. Phillips, Phys. Lett. B **751**, 535 (2015), arXiv:1507.07239 [nucl-th].
- [8] R. J. Furnstahl, N. Klco, D. R. Phillips, and S. Wesolowski, Phys. Rev. C **92**, 024005 (2015), arXiv:1506.01343 [nucl-th].
- [9] E. A. C. Pérez and T. Papenbrock, Phys. Rev. C **92**, 064309 (2015), arXiv:1510.02401 [nucl-th].
- [10] H. W. Griesshammer, J. A. McGovern, and D. R. Phillips, (2015), arXiv:1511.01952 [nucl-th].
- [11] R. J. Furnstahl, D. R. Phillips, and S. Wesolowski, J. Phys. G **42**, 034028 (2015), arXiv:1407.0657 [nucl-th].
- [12] R. Navarro Pérez, J. E. Amaro, and E. Ruiz Arriola, Phys. Rev. C **88**, 024002 (2013), [Erratum: Phys. Rev. C **88**, 069902 (2013)], arXiv:1304.0895 [nucl-th].
- [13] M. R. Schindler and D. R. Phillips, Annals Phys. **324**, 682 (2009), arXiv:0808.3643 [hep-ph].
- [14] A. Ekström, G. Baardsen, C. Forssén, G. Hagen, M. Hjorth-Jensen, et al., Phys. Rev. Lett. **110**, 192502 (2013), arXiv:1303.4674 [nucl-th].
- [15] B. Carlsson, A. Ekström, C. Forssén, D. F. Strömborg, O. Lilja, et al., Phys. Rev. X **6**, 011019 (2015), arXiv:1506.02466 [nucl-th].
- [16] E. Epelbaum, H. Krebs, and U. G. Meißner, Eur. Phys. J. A **51**, 53 (2015), arXiv:1412.0142 [nucl-th].
- [17] J. Dobaczewski, W. Nazarewicz, and P.-G. Reinhard, J. Phys. G **41**, 074001 (2014), arXiv:1402.4657 [nucl-th].
- [18] D. Stump et al., Phys. Rev. D **65**, 014012 (2001), arXiv:hep-ph/0101051 [hep-ph].
- [19] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge University Press, 2005).
- [20] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial* (Oxford University Press, 2006).
- [21] R. T. Cox, American Journal of Physics **14**, 1 (1946).
- [22] R. T. Cox, *The Algebra of Probable Inference* (Johns Hopkins University Press, 1961).
- [23] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).
- [24] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
- [25] D. Foreman-Mackey, A. Price-Whelan, G. Ryan, Emily, M. Smith, K. Barbary, D. W. Hogg, and B. J. Brewer, “corner.py v1.0.0,” (2014).
- [26] K. Olive et al. (Particle Data Group), Chin. Phys. C **38**, 090001 (2014).
- [27] R. Trotta, Contemp. Phys. **49**, 71 (2008), arXiv:0803.4089 [astro-ph].
- [28] E. Epelbaum, H. Krebs, and U.-G. Meißner, Nucl. Phys. A **806**, 65 (2008), arXiv:0712.1969 [nucl-th].
- [29] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publications of the Astronomical Society of the Pacific **125**, pp. 306 (2013).
- [30] G. Lepage, B. Clark, C. Davies, K. Hornbostel, P. Mackenzie, et al., Nucl. Phys. Proc. Suppl. **106**, 12 (2002), arXiv:hep-lat/0110175 [hep-lat].
- [31] G. P. Lepage, (1997), arXiv:nucl-th/9706029 [nucl-th].
- [32] H. W. Griesshammer, in *8th International Workshop on Chiral Dynamics (CD 2015) Pisa, Italy, June 29-July 3, 2015* (2015) arXiv:1511.00490 [nucl-th].
- [33] H. W. Griesshammer, Nucl. Phys. A **744**, 192 (2004), arXiv:nucl-th/0404073 [nucl-th].
- [34] P. F. Bedaque, G. Rupak, H. W. Griessammer, and H.-W. Hammer, Nucl. Phys. A **714**, 589 (2003), arXiv:nucl-th/0207034 [nucl-th].
- [35] D. Higdon, J. D. McDonnell, N. Schunck, J. Sarich, and S. M. Wild, J. Phys. G **42**, 034009 (2015), arXiv:1407.3017 [physics.data-an].
- [36] C. Bernard, in *8th International Workshop on Chiral Dynamics (CD 2015) Pisa, Italy, June 29-July 3, 2015* (2015) arXiv:1510.02180 [hep-ph].
- [37] G. Bali et al., Nuclear Physics B **866**, 1 (2013).
- [38] M. R. Schindler and D. R. Phillips, Annals Phys. **324**, 2051 (2009).
- [39] F. Hou, J. Goodman, D. W. Hogg, J. Ware, and C. Schwab, Astrophys. J. **745**, 198 (2012), arXiv:1104.2612 [astro-ph.IM].
- [40] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Journal of Chemical Physics **21**, 1087 (1953).
- [41] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (Chapman and Hall/CRC, 2003).