

A critical study of the Monte Carlo replica method

Mark N. Costantini,^a Maeve Madigan,^b Luca Mantani,^a James M. Moore^a

^a*DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom*

^b*Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, D-69120, Heidelberg, Germany*

E-mail: mnc33@cam.ac.uk, madigan@thphys.uni-heidelberg.de,
luca.mantani@maths.cam.ac.uk, jmm232@cam.ac.uk

ABSTRACT: We present a detailed mathematical study of the Monte Carlo replica method as applied in the global fitting literature from the high-energy physics theory community. For the first time, we provide a rigorous derivation of the parameter distributions implied by the method, and show that, whilst they agree with Bayesian posteriors for linear models, they disagree otherwise. We proceed to numerically quantify the disagreement between the Monte Carlo replica method and the Bayesian method in the context of two phenomenologically relevant scenarios: fits of the SMEFT Wilson coefficients, and fits of PDFs (albeit in a toy scenario). In both scenarios, we find that uncertainty estimates of the quantities of interest are discrepant between the two approaches when non-linearity is relevant. Our findings motivate future investigation of Bayesian methodologies for global PDF fits, especially in the context of simultaneous determination of PDFs and SMEFT Wilson coefficients.

KEYWORDS: Bayesian Statistics, Global Fits, Parton Distributions, SMEFT

Contents

1	Introduction	1
2	The mathematics of the Monte Carlo replica method	3
2.1	Bayesian vs Monte Carlo inference	4
2.2	Calculation of the Monte Carlo posterior	6
2.3	Toy examples: linear, quadratic and circular theories	10
3	Applications in high energy physics	14
3.1	SMEFT fits	15
3.2	PDF fits	17
3.2.1	A toy PDF model	19
3.2.2	Results for DIS-only fits	22
3.2.3	Results for hadronic-only fits	22
3.2.4	Results for global fits	23
4	Conclusions	25
A	Additional analytic example: purely quadratic theory	27
B	Numerical challenges of the Monte Carlo replica method	29
C	Impact of training-validation splits	31

1 Introduction

Obtaining reliable interval estimates for model parameters is one of the most fundamental problems in statistics. The usual starting point is the observation of data and its modelling with a parametric theory. The objective is then to estimate a region in parameter space that has a statistical meaning. In textbook discussions, this is usually effected by obtaining *confidence regions* (in frequentist statistics) or *credible regions* (in Bayesian statistics).

This paper studies an inference methodology which is often compared to the Bayesian paradigm, namely the *Monte Carlo (MC) replica method*. This method has been deployed in various fits of the Wilson coefficients in the Standard Model Effective Field Theory (SMEFT) [1–5], and to fits of the parton distribution functions (PDFs) which parametrise hadron structure [6–16] (and, indeed, in simultaneous extractions of PDFs and SMEFT Wilson coefficients [17–22]).

The Monte Carlo replica method has a very intuitive basis, described in detail in Sect. 2. Given an observation drawn from some known distribution (almost exclusively a multivariate Gaussian in the high-energy physics literature), we can use it to simulate

further data samples. Each of these samples is subsequently used to estimate the parameters of a theoretical model employed for describing the data. Through the collection of estimates, one intends to propagate uncertainties from the data to the parameter space.

Despite this intuitive basis, it is important to note that - to the best of the authors' knowledge - little has been said in the literature about the mathematical foundations of the method, apart from the case of a linear model (see Ref. [23]). There are similarities between the Monte Carlo replica method and the *parametric bootstrap* described in the mathematics literature; however, Monte Carlo seems to be a singular case of the method corresponding to an instance where a single initial observation is used to simulate the underlying distribution.¹ The literature typically deals only with the limiting case where many initial observations of the same quantity are used to simulate the underlying distribution; further detail can be found in Ref. [24, 25].

On the other hand, there has been significant effort in the physics literature to numerically benchmark the Monte Carlo replica method against other frameworks. For example, in the context of PDF fitting, Ref. [26] compares the use of the Monte Carlo replica method with other methods of PDF error determination in the context of a toy fit. Only two PDF flavours are included, and they are fitted to a single deep inelastic scattering (DIS) dataset, where the corresponding theory predictions are linear in the PDFs. The authors find perfect agreement between the Monte Carlo replica method and the other standard methods of PDF error determination considered; however, as we shall explain in Section 2 and Section 3, this is to be expected when using only DIS data, and does not necessarily generalise to the inclusion of data from proton-proton collisions. Additionally, the benchmarking exercise between the Monte Carlo replica method and the *Hessian method* for estimating PDF uncertainties carried out in Ref. [27] suggests that there is good agreement between these two approaches. Further benchmarking will be possible given the advent of new Bayesian PDF fits, which have started appearing in the literature in recent years; see for example Ref. [28], in which the authors present a novel parton density determination code to analyse collider data within a Bayesian framework, and Ref. [29] in which Gaussian processes for the solution of the PDF fitting inverse problem are discussed.

In contrast, in the case of benchmarking Monte Carlo versus Bayesian SMEFT fits, significant tension was observed in Ref. [20]. Indeed, a careful discussion in App. E of this work highlighted that the issue was methodological by providing a complete calculation of the Monte Carlo ‘posterior distribution’ in a very simple toy case. To the authors’ knowledge, this is the first time in the literature that disagreement was explicitly shown and explained between the two approaches.

This work aims to clarify the Monte Carlo replica method in much more generality, on two fronts. Firstly, we provide the first rigorous mathematical derivation of the form of the ‘Monte Carlo posterior distribution’, and we show that it is typically inequivalent to the posterior distribution that would have been inferred from a Bayesian method. Secondly, we apply our results to further numerical benchmarks in the high-energy physics literature,

¹We are grateful for a discussion with the statistician Alastair Young, an expert in bootstrap methods, for clarifying this point.

namely global fits of the SMEFT Wilson coefficients and, separately, of the PDFs of the proton. In general, we find that interval estimates obtained from the Monte Carlo replica method do not agree with those produced by Bayesian methods.²

The structure of the paper is as follows. In Sect. 2, we introduce the Monte Carlo replica method and derive the ‘Monte Carlo posterior’ distribution for model parameters, as compared to the parameter distributions that would be implied by a Bayesian method. We proceed to give some toy examples of the comparison in analytically tractable cases, including the important example of a linear model, where the Monte Carlo replica method and Bayesian method coincide in a simple way. Subsequently, in Sect. 3, we compare the Monte Carlo replica method with a Bayesian method in the context of two phenomenologically relevant scenarios. We begin by discussing a top-sector fit of the SMEFT Wilson coefficients, displaying significant disagreement between the Wilson coefficient distributions implied by the Monte Carlo replica method and a Bayesian method. We continue by discussing fits of the PDFs of the proton using a simplified toy model for the PDFs based on linear interpolation. We begin with a fit only to deep inelastic scattering data, where the two inference methods coincide perfectly. However, we follow up with a discussion of a fit to hadronic-only (proton-proton) data, in which the PDFs enter theory predictions quadratically; here, we find significant underestimation of PDF uncertainties (at worst by around 90%) in the low- x region when the Monte Carlo replica method is used as compared to a Bayesian method. Finally, we perform a fit using the complete DIS plus hadronic dataset, still observing an underestimation of uncertainties (at worst by around 60%). Instead, good agreement is found in the mid to high- x region in all of the toy fits. We summarise the results of this work in Sect. 4.

2 The mathematics of the Monte Carlo replica method

In this section, the main mathematical results of our work are presented, accompanied by toy examples. The take-home message is that, in general, the Monte Carlo replica method does not reproduce Bayesian credible regions. In particular, it is quite difficult to assess a priori whether the outcome of the fit is affected by pathological behaviours or whether it approximates the Bayesian posterior distribution. The notable exception to this is the case of linear theories, in which the posterior distribution is Gaussian and well reproduced by the Monte Carlo methodology. Given this consideration, should the reader choose, they may proceed to Section 3, where the application of these results to significant examples in high-energy physics is presented, bypassing the technical details outlined here.

We begin in Sect. 2.1 by reminding the reader of a Bayesian framework for statistical inference, the paradigm with which the Monte Carlo replica method is most often compared, before introducing the Monte Carlo replica method itself. We explain how both methods are applied to obtain uncertainty estimates on theory parameters in the case that data are distributed according to a multivariate normal distribution; this is the case to which the

²This has one notable exception, as we shall show in Sect. 2, which is the case of linear models; it can be shown that Monte Carlo and Bayesian fits of a linear model coincide, provided that a sufficiently wide uniform prior is used for the Bayesian fit.

Monte Carlo replica method is almost universally applied within the high-energy physics literature. In Sect. 2.2, we proceed to give a detailed mathematical calculation of the parameter distributions implied by the Monte Carlo replica method, contrasting them with distributions obtained from the Bayesian method. In Sect. 2.3, we give a selection of toy examples, both analytic and numerical, displaying the behaviour in a range of settings.

2.1 Bayesian vs Monte Carlo inference

In traditional statistical inference, there are two standard paradigms for constructing uncertainty estimates on parameters, the *Bayesian* and *frequentist* frameworks. The Bayesian approach assumes that the parameters \mathbf{c} of interest are themselves random variables; a Bayesian $100\alpha\%$ *credible region* is then a region of parameter space within which the parameter has a probability α of lying, given some observed data. On the other hand, the frequentist approach assumes that the parameters \mathbf{c} are fixed; a frequentist $100\alpha\%$ *confidence region* is a region constructed from the observed data in such a manner to ensure that if the data and confidence region were regenerated 100 times, then 100α of the confidence intervals would cover the true parameter value. These methods sound extremely similar at face value, but they are not equivalent in general; agreement is only guaranteed in the large sample limit by the Bernstein-von Mises theorem [30].

The Bayesian approach is often used in global fits in the high-energy physics theory literature since it allows us to efficiently constrain a high-dimensional parameter space, while making use of the pre-processed information provided by experiments. This typically consists of a single ‘measurement’ of the data, usually a central value of a measured cross-section and some uncertainty estimate (or covariance matrix in the multi-dimensional case), see for example Refs. [1–3, 31–33]. Further, the Monte Carlo replica method itself is most frequently compared with the Bayesian paradigm; see for instance Refs. [1, 4, 16, 23], and particularly Ref. [26], where the Monte Carlo replica method is described as a ‘*frequentist approach to obtaining a Bayesian posterior*’. Therefore, in this work we shall choose to benchmark the Monte Carlo replica method exclusively against the Bayesian practice.

Bayesian method. In order to fix notation for the discussion of the Monte Carlo replica method in the sequel, let us describe the Bayesian approach in more detail for a specific setup. Let us suppose that experimental data comprising N_{dat} datapoints is distributed according to a multivariate normal distribution:³

$$\mathbf{d} \sim \mathcal{N}(\mathbf{t}(\mathbf{c}), \Sigma), \quad (2.1)$$

where Σ is an $N_{\text{dat}} \times N_{\text{dat}}$ experimental covariance matrix (of which we assume perfect knowledge), and $\mathbf{t} : \mathbb{R}^{N_{\text{param}}} \rightarrow \mathbb{R}^{N_{\text{dat}}}$ is a smooth theory prediction function, taking as

³In principle, other distributions could be considered, however, in the high-energy physics literature the multivariate normal is often used when applying the Monte Carlo replica method. Further, the Monte Carlo replica method requires that we can efficiently sample from the distribution from which the data is drawn, so a normal is a natural candidate with which to begin working. Although, it is worth noting that in NNPDF and SIMUnet a *truncated* multivariate normal is technically used, due to rejection of non-positive pseudodata [22, 34].

argument a vector of N_{param} unknown theory parameters $\mathbf{c} \in \mathbb{R}^{N_{\text{param}}}$. Given an observation \mathbf{d}_0 of the experimental data, our aim is to recover a reliable estimate of the region in which \mathbf{c} lies.

Since in Bayesian statistics, \mathbf{c} itself is assumed to be a random variable, it has some associated *prior probability density* $p(\mathbf{c})$ reflecting our knowledge of \mathbf{c} prior to the experimental observation. Bayes' theorem then tells us that after the observation, the probability density of \mathbf{c} given \mathbf{d}_0 is:

$$p(\mathbf{c}|\mathbf{d}_0) \propto p(\mathbf{c}) \cdot p(\mathbf{d}_0|\mathbf{c}) = p(\mathbf{c}) \exp\left(-\frac{1}{2}(\mathbf{d}_0 - \mathbf{t}(\mathbf{c}))^T \Sigma^{-1}(\mathbf{d}_0 - \mathbf{t}(\mathbf{c}))\right), \quad (2.2)$$

where we have inserted the probability density for a multivariate Gaussian $p(\mathbf{d}_0|\mathbf{c})$ according to Eq. (2.1). For convenience, the argument of the exponential in the posterior is usually written in terms of the χ^2 -*statistic* evaluated on the data,

$$\chi_{\mathbf{d}_0}^2(\mathbf{c}) := (\mathbf{d}_0 - \mathbf{t}(\mathbf{c}))^T \Sigma^{-1}(\mathbf{d}_0 - \mathbf{t}(\mathbf{c})). \quad (2.3)$$

The Bayesian posterior distribution is used to construct uncertainty estimates for the parameter \mathbf{c} ; in particular, a $100\alpha\%$ *credible region* is defined to be a region R such that:

$$N \int_R p(\mathbf{c}) \exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(\mathbf{c})\right) = \alpha, \quad (2.4)$$

where N is the appropriate normalisation constant in the proportionality relation Eq. (2.2). Such regions can be constructed efficiently using numerical algorithms, including Markov Chain Monte Carlo (MCMC) methods such as Nested Sampling (NS) [35–40].

The Monte Carlo replica method. In the Monte Carlo replica method, we begin by introducing the *pseudodata distribution* $\mathbf{d}_p \sim \mathcal{N}(\mathbf{d}_0, \Sigma)$, which is intended to approximate the actual distribution from which the measurement \mathbf{d}_0 was drawn. We then define the corresponding ‘*best-fit parameter*’ values to be those which minimise the χ^2 -statistic evaluated on the pseudodata:⁴

$$\mathbf{c}_p(\mathbf{d}_p) := \arg \min_{\mathbf{c}} \chi_{\mathbf{d}_p}^2(\mathbf{c}) = \arg \min_{\mathbf{c}} (\mathbf{d}_p - \mathbf{t}(\mathbf{c}))^T \Sigma^{-1}(\mathbf{d}_p - \mathbf{t}(\mathbf{c})), \quad (2.5)$$

essentially defining $\mathbf{c}_p(\mathbf{d}_p)$ as a function of a random variable \mathbf{d}_p . The distribution of the *Monte Carlo replicas* $\mathbf{c}_p(\mathbf{d}_p)$ is now usually interpreted in the same way as the Bayesian posterior [1, 12, 26]; we shall call this distribution the *Monte Carlo posterior* to make the comparison clear.

It is worth noting that, in practice, the determination of the minimiser is performed numerically; in particular, this means that the values obtained for $\mathbf{c}_p(\mathbf{d}_p)$ may depend on the numerical setup (e.g. choice of optimiser, initialisation, learning rate). This is also true for a Bayesian analysis when numerical methods are used. Ideally, we should strive

⁴This equation does not necessarily hold exactly in practice; often, numerical implementations of the Monte Carlo replica method also use a random training-validation split when finding the minimum, together with a cross-validation stopping. We further discuss this in App. C.

to reduce this ‘methodological uncertainty’ as much as possible, so that the uncertainty on the parameters is dominated by the information coming from the data; justification that we have indeed achieved this in our study is presented in App. B.

In the following section, we shall give a detailed calculation of the Monte Carlo posterior, and compare it to the Bayesian posterior.

2.2 Calculation of the Monte Carlo posterior

Naturally, it is important to ask whether the distribution of the Monte Carlo replicas $\mathbf{c}_p(\mathbf{d}_p)$ presented in Eq. (2.5) is comparable to the distribution obtained via the Bayesian approach in Eq. (2.2), since these methods are so often treated equivalently. In this section, we present a detailed calculation of the distribution of $\mathbf{c}_p(\mathbf{d}_p)$, using basic methods from probability theory, in order to facilitate this comparison.

We begin by noting that $\mathbf{c}_p(\mathbf{d}_p)$ is a function of the random variable \mathbf{d}_p , which ensures that its distribution is given by the general formula:

$$p(\mathbf{c}) \propto \int d^d \mathbf{d}_p \delta(\mathbf{c} - \mathbf{c}_p(\mathbf{d}_p)) \exp\left(-\frac{1}{2}(\mathbf{d}_p - \mathbf{d}_0)^T \Sigma^{-1}(\mathbf{d}_p - \mathbf{d}_0)\right), \quad (2.6)$$

where an integral without limits denotes an integral over the entire space (in this case, $\mathbb{R}^{N_{\text{dat}}}$). Importantly, this formula assumes that $\mathbf{c}_p(\mathbf{d}_p)$ is a single-valued function of the pseudodata \mathbf{d}_p ; that is, there are not multiple equivalent minima of the χ^2 -statistic given in Eq. (2.5). This is not a deficiency of the mathematical approach; in the case that $\mathbf{c}_p(\mathbf{d}_p)$ is only discretely multi-valued, the posterior is simply upgraded to:

$$p(\mathbf{c}) \propto \int d^d \mathbf{d}_p \sum_{i=1}^{N_{\text{multi}}(\mathbf{d}_p)} \delta\left(\mathbf{c} - \mathbf{c}_p^{(i)}(\mathbf{d}_p)\right) \exp\left(-\frac{1}{2}(\mathbf{d}_p - \mathbf{d}_0)^T \Sigma^{-1}(\mathbf{d}_p - \mathbf{d}_0)\right), \quad (2.7)$$

where $\mathbf{c}_p^{(i)}(\mathbf{d}_p)$ is the i th branch of $\mathbf{c}_p(\mathbf{d}_p)$ arising from the pseudodata \mathbf{d}_p . In particular, we will need this case in one of the toy examples in the sequel. In the more serious case that $\mathbf{c}_p(\mathbf{d}_p)$ has some *continuum* of multiple values for a given piece of pseudodata, we say that the problem contains a *flat direction*. We discuss this in Appendix B. Therefore, for simplicity, we shall now proceed to assume that $N_{\text{branch}}(\mathbf{d}_p) \equiv 1$, safe in the knowledge that generalisation is possible. We now manipulate the integral Eq. (2.6) to present it in a form where it can be compared to the Bayesian posterior.

Coordinates on pseudodata space. We begin by introducing a new system of coordinates on the space of pseudodata, allowing us to reduce the integral to a convenient form. Currently, the pseudodata space is parametrised directly in terms of \mathbf{d}_p , but a natural alternative parametrisation arises for a pseudodata point \mathbf{d}_p in terms of its corresponding best-fit parameter values \mathbf{c}_p , and a vector $\boldsymbol{\lambda}$ describing the displacement of \mathbf{d}_p from the theory prediction of the best-fit parameter values, $\mathbf{t}(\mathbf{c}_p)$. These coordinates $(\mathbf{c}_p, \boldsymbol{\lambda})$ will allow us (in the generic case) to ‘absorb’ the delta functions in Eq. (2.6).

To construct such coordinates, we begin by noting that the function $\mathbf{c}_p(\mathbf{d}_p)$ must satisfy the system of equations:

$$\mathbf{0} = \left(\frac{\partial \mathbf{t}}{\partial \mathbf{c}}\right)^T (\mathbf{c}_p(\mathbf{d}_p)) \Sigma^{-1}(\mathbf{d}_p - \mathbf{t}(\mathbf{c}_p(\mathbf{d}_p))), \quad (2.8)$$

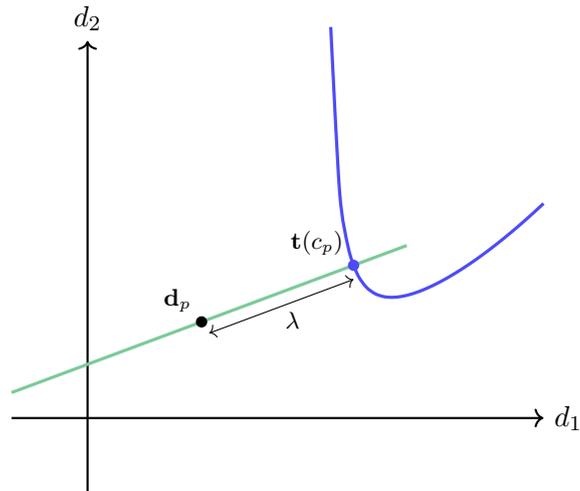


Figure 2.1. A visualisation of the construction of ‘natural’ coordinates on pseudodata space, in the case where $N_{\text{dat}} = 2$ and $N_{\text{param}} = 1$, and where $\Sigma = \sigma^2 I$ is an uncorrelated covariance matrix (so that ‘distance’ on pseudodata space coincides with Euclidean distance). In this case, the theory $\mathbf{t} : \mathbb{R}^2 \rightarrow \mathbb{R}$ may be viewed as constructing a parametric curve $\mathbf{t}(c)$ in the plane, shown in blue. The ‘natural’ coordinates on the point \mathbf{d}_p in pseudodata space are constructed in terms of the corresponding best-fit parameter values c_p , and the displacement $\mathbf{w} = M(c_p)\lambda$ from $\mathbf{t}(c_p)$ to \mathbf{d}_p , where in this case $\lambda \in \Lambda(c_p) \subseteq \mathbb{R}$; see the main text for details. The set $\Lambda(c_p)$ is indicated in the figure by the green solid line. It consists of all points whose closest point on the theory curve $\mathbf{t} : \mathbb{R} \rightarrow \mathbb{R}^2$ is $\mathbf{t}(c_p)$ (since $\Sigma = \sigma^2 I$ is Euclidean in this figure; in the general case, we mean ‘closest’ with respect to the distance measure induced from the inner product induced by Σ).

since $\mathbf{c}_p(\mathbf{d}_p)$ is a minimiser of the χ^2 -statistic evaluated on pseudodata, and hence is a stationary point of the χ^2 -statistic evaluated on pseudodata. In particular, this implies the relation:

$$\mathbf{d}_p = \mathbf{t}(\mathbf{c}_p(\mathbf{d}_p)) + \Sigma \mathbf{w}, \quad \text{where} \quad \mathbf{w} \in \text{Ker} \left(\left(\frac{\partial \mathbf{t}}{\partial \mathbf{c}} \right)^T (\mathbf{c}_p(\mathbf{d}_p)) \right). \quad (2.9)$$

That is, a *necessary* condition on pseudodata to lead to the best-fit parameter values \mathbf{c}_p is that it takes the form $\mathbf{d}_p = \mathbf{t}(\mathbf{c}_p) + \Sigma \mathbf{w}$, for some \mathbf{w} in the above kernel. Importantly though, it is *not* a *sufficient* condition; we have only shown that pseudodata $\mathbf{d}_p = \mathbf{t}(\mathbf{c}_p) + \Sigma \mathbf{w}$ for a given \mathbf{w} leads to \mathbf{c}_p being a *stationary* point of the χ^2 -statistic evaluated on \mathbf{d}_p - it could be a maximum or a saddle. Thus, in order to guarantee that pseudodata of the form Eq. (2.9) leads to $\mathbf{c}_p(\mathbf{d}_p)$ as a minimum, we must further restrict to an appropriate range of \mathbf{w} .

Now, rank-nullity implies that the dimension of the kernel is at least $N_{\text{dat}} - N_{\text{param}}$,

$$\dim \left(\text{Ker} \left(\left(\frac{\partial \mathbf{t}}{\partial \mathbf{c}} \right)^T (\mathbf{c}_p) \right) \right) \geq N_{\text{dat}} - N_{\text{param}}. \quad (2.10)$$

Let us write the dimension of the kernel as $N_{\perp}(\mathbf{c}_p)$ for short, denoting directions ‘orthogonal’ to the theory surface. Introducing a basis $\mathbf{w}_1, \dots, \mathbf{w}_{N_{\perp}(\mathbf{c}_p)}$ for the kernel, we can package

the basis as a single matrix:

$$M(\mathbf{c}_p) := \left(\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_{N_\perp(\mathbf{c}_p)} \right), \quad (2.11)$$

which allows us to express \mathbf{d}_p in the form:

$$\mathbf{d}_p = \mathbf{t}(\mathbf{c}_p(\mathbf{d}_p)) + \Sigma M(\mathbf{c}_p(\mathbf{d}_p))\boldsymbol{\lambda}, \quad (2.12)$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{N_\perp(\mathbf{c}_p(\mathbf{d}_p))}$ is a set of coordinates for the ‘orthogonal’ directions to the theory surface near $\mathbf{c}_p(\mathbf{d}_p)$. Recall, $\boldsymbol{\lambda}$ must be additionally restricted to an allowed range which results in this pseudodata giving $\mathbf{c}_p(\mathbf{d}_p)$ as a *minimiser* of the χ^2 -statistic; let us write this range as $\Lambda(\mathbf{c}_p(\mathbf{d}_p))$.

We would like to state that the ‘natural’ coordinates on pseudodata space are now $(\mathbf{c}_p, \boldsymbol{\lambda})$ indicating where on the theory surface we are (the \mathbf{c}_p coordinate), and then how far away we are in an orthogonal direction (the $\boldsymbol{\lambda}$ coordinate). An example of the construction of these coordinates is shown in Fig. 2.1. However, since $(\partial\mathbf{t}/\partial\mathbf{c})^T$ can have less than full rank, the number of these coordinates can exceed N_{dat} ; near points where this Jacobian matrix has less than full rank, some of the \mathbf{c}_p coordinates are redundant (see Fig. 2.2 for an example). Therefore, near any given $\mathbf{c} \in \mathbb{R}^{N_{\text{param}}}$ where the Jacobian matrix $(\partial\mathbf{t}/\partial\mathbf{c})^T(\mathbf{c})$ has rank $N_{\parallel}(\mathbf{c})$, we shall assume that we can introduce some smooth parametrisation:⁵

$$\mathbf{f} : \mathbb{R}^{N_{\parallel}(\mathbf{c})} \rightarrow \mathbb{R}^{N_{\text{param}}} \quad (2.13)$$

which removes this redundancy, allowing us to define a bijective correspondence between pseudodata \mathbf{d}_p and coordinates $(\mathbf{u}, \boldsymbol{\lambda})$ (at least locally in \mathbf{u}):

$$\mathbf{d}_p = \mathbf{t}(\mathbf{f}(\mathbf{u})) + \Sigma M(\mathbf{f}(\mathbf{u}))\boldsymbol{\lambda}. \quad (2.14)$$

In the case that $N_{\parallel}(\mathbf{c}) = N_{\text{param}}$, we are in the happy case where the Jacobian matrix has full rank, and we can simply take $\mathbf{f}(\mathbf{u}) = \mathbf{u}$, essentially using the coordinates $(\mathbf{c}_p, \boldsymbol{\lambda})$ as initially desired. Otherwise, this more complicated construction is needed.

Change of coordinates in Eq. (2.6). Let us now consider changing coordinates in Eq. (2.6) near \mathbf{c} to the ‘natural’ system $(\mathbf{u}, \boldsymbol{\lambda})$ that we defined above. The Jacobian of the transformation $\mathbf{d}_p \mapsto (\mathbf{u}, \boldsymbol{\lambda})$ is given by:

$$\left| \det \left(\frac{\partial\mathbf{t}}{\partial\mathbf{c}}(\mathbf{f}(\mathbf{u})) \frac{\partial\mathbf{f}}{\partial\mathbf{u}} + \frac{\partial(\Sigma M\boldsymbol{\lambda})}{\partial\mathbf{c}}(\mathbf{f}(\mathbf{u})) \frac{\partial\mathbf{f}}{\partial\mathbf{u}} \Big| \Sigma M(\mathbf{f}(\mathbf{u})) \right) \right|, \quad (2.15)$$

and hence the integral Eq. (2.6) reduces to:

$$\int d^{N_{\parallel}(\mathbf{c})}\mathbf{u} \delta(\mathbf{c} - \mathbf{f}(\mathbf{u})) \int_{\Lambda(\mathbf{c})} d^{N_\perp(\mathbf{c})}\boldsymbol{\lambda} \left| \det \left(\frac{\partial\mathbf{t}}{\partial\mathbf{c}}(\mathbf{f}(\mathbf{u})) \frac{\partial\mathbf{f}}{\partial\mathbf{u}} + \frac{\partial(\Sigma M\boldsymbol{\lambda})}{\partial\mathbf{c}}(\mathbf{f}(\mathbf{u})) \frac{\partial\mathbf{f}}{\partial\mathbf{u}} \Big| \Sigma M(\mathbf{f}(\mathbf{u})) \right) \right| \cdot \exp \left(-\frac{1}{2}(\mathbf{t}(\mathbf{c}) + \Sigma M(\mathbf{c})\boldsymbol{\lambda} - \mathbf{d}_0)^T \Sigma^{-1}(\mathbf{t}(\mathbf{c}) + \Sigma M(\mathbf{c})\boldsymbol{\lambda} - \mathbf{d}_0) \right). \quad (2.16)$$

⁵Whilst this assumption may seem reasonable, there exist examples of exceptionally pathological multi-variable functions which fail to have locally constant rank except at isolated points. The authors consider these unlikely to occur in a physical problem.

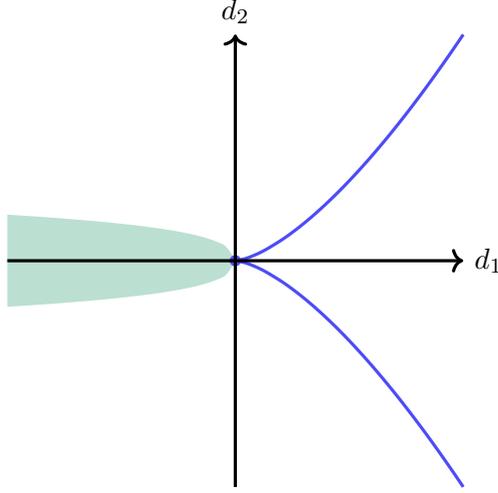


Figure 2.2. An example of a point on a theory surface $\mathbf{t}(c) = (c^2, c^3)^T$ (with Euclidean covariance $\Sigma = \sigma^2 I$) whose derivative does not have full rank at $c = 0$. In this case, the ‘natural’ coordinates (c, λ) constructed on the remainder of this space do not apply at this point; instead, the required coordinates are (λ_1, λ_2) with c neglected (the ‘local parametrisation’ function in this case is simply a constant, $f : \{\emptyset\} \rightarrow \mathbb{R}$ given by $f(\emptyset) = 0$). The space $\Lambda(0)$ is a two-dimensional subset of the left half-plane, indicated in green in the figure. At $c = 0$, there will be a delta function singularity in the Monte Carlo posterior, which is not present in the Bayesian posterior.

Expanding the exponent, this can be manipulated into the form:

$$\begin{aligned}
& \exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(\mathbf{c})\right) \\
& \cdot \int d^{N_{\parallel}(\mathbf{c})}\mathbf{u} \delta(\mathbf{c} - \mathbf{f}(\mathbf{u})) \int_{\Lambda(\mathbf{c})} d^{N_{\perp}(\mathbf{c})}\boldsymbol{\lambda} \left| \det\left(\frac{\partial \mathbf{t}}{\partial \mathbf{c}}(\mathbf{f}(\mathbf{u}))\frac{\partial \mathbf{f}}{\partial \mathbf{u}} + \frac{\partial(\Sigma M \boldsymbol{\lambda})}{\partial \mathbf{c}}(\mathbf{f}(\mathbf{u}))\frac{\partial \mathbf{f}}{\partial \mathbf{u}} \Big|_{\Sigma M(\mathbf{f}(\mathbf{u}))}\right) \right| \\
& \cdot \exp\left(-\frac{1}{2}\boldsymbol{\lambda}^T M(\mathbf{c})^T \Sigma M(\mathbf{c}) \boldsymbol{\lambda} + \boldsymbol{\lambda}^T M(\mathbf{c})^T (\mathbf{d}_0 - \mathbf{t}(\mathbf{c}))\right), \tag{2.17}
\end{aligned}$$

which allows comparison with the Bayesian posterior in Eq. (2.2). We see that the approaches are in general very different: the Bayesian posterior of Eq. (2.2) and the Monte Carlo posterior above are equivalent *if and only if* the prior in the Bayesian approach is chosen according to the complicated \mathbf{c} -dependent multiplicative factor featuring in the above equation.

In the case that $(\partial \mathbf{t} / \partial \mathbf{c})^T$ has full rank at \mathbf{c} , the general formula given in Eq. (2.17) significantly simplifies. We can take $\mathbf{f}(\mathbf{u}) = \mathbf{u}$, and $N_{\parallel}(\mathbf{c}) = N_{\text{param}}$, $N_{\perp}(\mathbf{c}) = N_{\text{dat}} - N_{\text{param}}$,

which allows us to absorb all of the delta functions in Eq. (2.17), reducing it to:

$$\begin{aligned} & \exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(\mathbf{c})\right) \int_{\Lambda(\mathbf{c})} d^{N_{\text{dat}}-N_{\text{param}}}\boldsymbol{\lambda} \left| \det\left(\frac{\partial\mathbf{t}}{\partial\mathbf{c}} + \frac{\partial(\Sigma M\boldsymbol{\lambda})}{\partial\mathbf{c}}\Big|_{\Sigma M(\mathbf{c})}\right) \right| \\ & \cdot \exp\left(-\frac{1}{2}\boldsymbol{\lambda}^T M(\mathbf{c})^T \Sigma M(\mathbf{c}) \boldsymbol{\lambda} + \boldsymbol{\lambda}^T M(\mathbf{c})^T (\mathbf{d}_0 - \mathbf{t}(\mathbf{c}))\right). \end{aligned} \quad (2.18)$$

Summary. We have shown that, in the general multivariable case for a general theory prediction, the Bayesian posterior and the Monte Carlo posterior coincide only for a very judicious choice of Bayesian prior (which seems unlikely to be motivated). Choosing such a prior, the resulting coincident posteriors have the following features:

- At points \mathbf{c} where the Jacobian matrix $(\partial\mathbf{t}/\partial\mathbf{c})^T$ has full-rank, the posterior is a ‘scaled’ version of the Bayesian posterior obtained under the assumption of a sufficiently wide uniform prior. Note that the overall scale factor is \mathbf{c} -dependent and may lead to shape differences. The integral factor that multiplies the Bayesian posterior depends on the size of the region $\Lambda(\mathbf{c})$ of the pseudodata which gives rise to the best-fit value \mathbf{c} in the Monte Carlo approach.
- At points \mathbf{c} where the Jacobian matrix $(\partial\mathbf{t}/\partial\mathbf{c})^T$ does not have full-rank, the posterior has delta function singularities. The strength of these singularities again corresponds to the size of the region $\Lambda(\mathbf{c})$ of the pseudodata which gives rise to the best-fit value \mathbf{c} in the Monte Carlo approach.

In the subsequent section, we shall exhibit toy examples showcasing these phenomena, before extending to examples in the high-energy physics literature in Sect. 3.

2.3 Toy examples: linear, quadratic and circular theories

In this section, we present a selection of toy examples, demonstrating the ideas of the previous section more concretely. We give three examples: (i) a linear theory, where the Bayesian and Monte Carlo posteriors coincide in a simple way (they are coincident provided the Bayesian prior is a sufficiently wide uniform distribution); (ii) a quadratic theory of one datapoint, which exhibits the delta function behaviour discussed above; (iii) a circular theory, which does not exhibit the delta function behaviour but nonetheless demonstrates an unusual scaling behaviour because of the integral factors in Eq. (2.18).

A further analytic example is given in App. A, namely a calculation of the Monte Carlo posterior in the case of a purely quadratic theory of multiple datapoints.

Example 1 - Linear theory. The most basic example is the case of a linear theory, $\mathbf{t}(\mathbf{c}) = \mathbf{t}_0 + \mathbf{t}_{\text{lin}}\mathbf{c}$, where $\mathbf{t}_0 \in \mathbb{R}^{N_{\text{dat}}}$ and \mathbf{t}_{lin} is an $N_{\text{dat}} \times N_{\text{param}}$ matrix. In this case, the relevant Jacobian matrix is:

$$\left(\frac{\partial\mathbf{t}}{\partial\mathbf{c}}\right)^T = \mathbf{t}_{\text{lin}}^T, \quad (2.19)$$

which is independent of \mathbf{c} . In particular, the matrix $M(\mathbf{c})$ may be constructed independently of \mathbf{c} , obeying the property $M(\mathbf{c})^T \mathbf{t}_{\text{lin}} = 0$ by definition. Assuming that $\mathbf{t}_{\text{lin}}^T$ is of

full-rank (else we in fact get a flat direction), we may use Eq. (2.18) to calculate the Monte Carlo posterior:

$$\exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(\mathbf{c})\right) \int_{\Lambda(\mathbf{c})} d^{N_{\text{dat}}-N_{\text{param}}}\boldsymbol{\lambda} \left| \det\left(\mathbf{t}_{\text{lin}} \middle| \Sigma M\right) \right| \exp\left(-\frac{1}{2}\boldsymbol{\lambda}^T M^T \Sigma M \boldsymbol{\lambda} + \boldsymbol{\lambda}^T M^T (\mathbf{d}_0 - \mathbf{t}_0)\right). \quad (2.20)$$

Observe that the \mathbf{c} -dependence has entirely dropped out from the Monte Carlo ‘scale factor’, except for in the integration range $\Lambda(\mathbf{c})$. To compute this range, we must determine for which values $\boldsymbol{\lambda}$ the pseudodata $\mathbf{t}(\mathbf{c}) + \Sigma M \boldsymbol{\lambda}$ leads to \mathbf{c} minimising the χ^2 -statistic evaluated on this pseudodata. The χ^2 -statistic evaluated on this pseudodata is given by:

$$\begin{aligned} \chi^2(\mathbf{c}') &= (\mathbf{t}(\mathbf{c}') - \mathbf{t}(\mathbf{c}) - \Sigma M \boldsymbol{\lambda})^T \Sigma^{-1} (\mathbf{t}(\mathbf{c}') - \mathbf{t}(\mathbf{c}) - \Sigma M \boldsymbol{\lambda}) \\ &= (\mathbf{t}_{\text{lin}}(\mathbf{c}' - \mathbf{c}) - \Sigma M \boldsymbol{\lambda})^T \Sigma^{-1} (\mathbf{t}_{\text{lin}}(\mathbf{c}' - \mathbf{c}) - \Sigma M \boldsymbol{\lambda}) \\ &= (\mathbf{c}' - \mathbf{c})^T \mathbf{t}_{\text{lin}}^T \Sigma^{-1} \mathbf{t}_{\text{lin}} (\mathbf{c}' - \mathbf{c}) + \boldsymbol{\lambda}^T M^T \Sigma M \boldsymbol{\lambda}. \end{aligned} \quad (2.21)$$

This expression is minimised by \mathbf{c}' such that $\mathbf{t}_{\text{lin}}(\mathbf{c}' - \mathbf{c}) = \mathbf{0}$, since Σ^{-1} is positive definite. But assuming that \mathbf{t}_{lin} is full-rank, this has a unique solution $\mathbf{c}' = \mathbf{c}$. Thus $\Lambda(\mathbf{c}) = \mathbb{R}^{N_{\text{dat}}-N_{\text{param}}}$ for all values of \mathbf{c} , indicating that the Monte Carlo ‘scale factor’ is \mathbf{c} -independent.

It follows that the Monte Carlo posterior for this linear theory is proportional to:

$$\exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(\mathbf{c})\right), \quad (2.22)$$

which is simply the Bayesian posterior obtained from a sufficiently wide uniform prior. Thus we have shown: *for a linear theory, the Bayesian and Monte Carlo approaches coincide* (provided that the Bayesian approach uses a sufficiently wide uniform prior). This result will be very important in the phenomenological study of PDFs in Sect. 3.2.

Example 2 - Quadratic theory, one datapoint. Consider $t(c) = t_0 + t_{\text{lin}}c + t_{\text{quad}}c^2$, a quadratic theory in one parameter, with $t_{\text{quad}} > 0$. In this case, the Jacobian matrix is:

$$\frac{\partial t}{\partial c} = t_{\text{lin}} + 2ct_{\text{quad}}, \quad (2.23)$$

which has full rank unless $c = -t_{\text{lin}}/2t_{\text{quad}}$. Thus for all $c \neq -t_{\text{lin}}/2t_{\text{quad}}$, we can choose $M(c)$ empty, which seems to give the Monte Carlo posterior for $c \neq -t_{\text{lin}}/2t_{\text{quad}}$ as:

$$\left| \det\left(\frac{\partial t}{\partial c}\right) \right| \exp\left(-\frac{1}{2}\chi_{d_0}^2(c)\right) = |2ct_{\text{quad}} + t_{\text{lin}}| \exp\left(-\frac{1}{2}\chi_{d_0}^2(c)\right). \quad (2.24)$$

However, it is important to observe that in this case, the function $c_p(d_p)$ is in fact multi-valued; we can see this by writing the theory prediction as:

$$t(c) = t_{\text{quad}} \left(c + \frac{t_{\text{lin}}}{2t_{\text{quad}}} \right)^2 + t_0 - \frac{t_{\text{lin}}^2}{4t_{\text{quad}}}. \quad (2.25)$$

This shows that there is a symmetry in the parameter c about $c = -t_{\text{lin}}/2t_{\text{quad}}$; two equivalent values of c give the same prediction provided that $c \neq -t_{\text{lin}}/2t_{\text{quad}}$. Thus we must additionally include a factor of 2 in the Monte Carlo posterior accounting for the multi-valuedness; this follows from the more general form of the posterior given in Eq. (2.7).

In the remaining case, $c = -t_{\text{lin}}/2t_{\text{quad}}$, we take $M(-t_{\text{lin}}/2t_{\text{quad}}) = 1$, and parametrise the lower rank surface with $f : \{\emptyset\} \rightarrow \mathbb{R}$ by $f(\emptyset) = -t_{\text{lin}}/2t_{\text{quad}}$. Then, the Monte Carlo posterior about $c = -t_{\text{lin}}/2t_{\text{quad}}$ takes the form:

$$\exp\left(-\frac{1}{2}\chi_{d_0}^2(c)\right) \delta\left(c + \frac{t_{\text{lin}}}{2t_{\text{quad}}}\right) \int_{\Lambda(-t_{\text{lin}}/2t_{\text{quad}})} d\lambda \sigma^2 \exp\left(-\frac{1}{2}\sigma^2\lambda^2 + \lambda(d_0 - t(c))\right), \quad (2.26)$$

where σ^2 is the one-by-one covariance matrix. The set $\Lambda(-t_{\text{lin}}/2t_{\text{quad}})$ consists of all λ such that the pseudodata $t(-t_{\text{lin}}/2t_{\text{quad}}) + \sigma^2\lambda$ leads to $-t_{\text{lin}}/2t_{\text{quad}}$ as a minimiser of the χ^2 -statistic on this pseudodata. We note that the χ^2 -statistic on this pseudodata is given by:

$$\chi^2(c') = \frac{1}{\sigma^2} \left(t(c') - t\left(-\frac{t_{\text{lin}}}{t_{\text{quad}}}\right) - \sigma^2\lambda \right)^2 \quad (2.27)$$

$$= \frac{1}{\sigma^2} \left(t_{\text{quad}} \left(c' + \frac{t_{\text{lin}}}{2t_{\text{quad}}} \right)^2 - \sigma^2\lambda \right)^2, \quad (2.28)$$

using the form of the theory prediction given in Eq. (2.25). If $\lambda > 0$, then we can solve the quadratic in the bracket to obtain a value of c' different from $-t_{\text{lin}}/2t_{\text{quad}}$ which gives a χ^2 -statistic of zero to the pseudodata. Otherwise, if $\lambda \leq 0$, the minimiser is given by $c' = -t_{\text{lin}}/2t_{\text{quad}}$. Hence we see that $\Lambda(-t_{\text{lin}}/2t_{\text{quad}}) = (-\infty, 0]$.

Putting everything together, it follows that the complete Monte Carlo posterior is given by:

$$\exp\left(-\frac{1}{2}\chi_{d_0}^2(c)\right) \left[\delta\left(c + \frac{t_{\text{lin}}}{2t_{\text{quad}}}\right) \int_{-\infty}^0 d\lambda \sigma^2 \exp\left(-\frac{1}{2}\sigma^2\lambda^2 + \lambda(d_0 - t(c))\right) + 2|2ct_{\text{quad}} + t_{\text{lin}}| \right], \quad (2.29)$$

which (through an appropriate substitution in the integral) matches the formula given in Appendix E of Ref. [20], where a very brief discussion of the statistical validity of the Monte Carlo method was first presented.⁶

Importantly, this demonstrates a singular behaviour of the Monte Carlo posterior at the point $c = -t_{\text{lin}}/2t_{\text{quad}}$ as compared to the Bayesian posterior. Indeed, this can result in a significant bias in the central value and a significant underestimation of the uncertainties for the parameter c , since the posterior is highly concentrated around $c = -t_{\text{lin}}/2t_{\text{quad}}$. This phenomenon is showcased in the phenomenologically relevant cases of the SMEFT in Sect. 3.1 and of PDFs in Sect. 3.2.

⁶It should be noted that Eq. (E.11) in the reference Ref. [20] actually contains an *erratum*, namely the division by $|2ct_{\text{quad}} + t_{\text{lin}}|$ in the second term of the equation should actually be a multiplication, as presented here.

Example 3 - Circular theory. Consider a theory $\mathbf{t} : \mathbb{R} \rightarrow \mathbb{R}^2$ given by:

$$\mathbf{t}(c) = t_0 \begin{pmatrix} \cos(c) \\ \sin(c) \end{pmatrix}, \quad (2.30)$$

with Euclidean covariance matrix $\Sigma = \sigma^2 I$. In this case, the relevant Jacobian matrix is:

$$\frac{\partial \mathbf{t}}{\partial c} = t_0 \begin{pmatrix} -\sin(c) \\ \cos(c) \end{pmatrix}, \quad (2.31)$$

which is always of full-rank. We see that we may take $M(c) = \mathbf{t}(c)$, and hence the Monte Carlo posterior is given for all c by:

$$\exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(c)\right) \int_{\Lambda(c)} d\lambda t_0^2 \sigma^2 (1 + \lambda \sigma^2) \exp\left(-\frac{1}{2}t_0^2 \sigma^2 \lambda^2 + \lambda (\mathbf{t}(c) \cdot \mathbf{d}_0 - t_0^2)\right). \quad (2.32)$$

To determine the integration range $\Lambda(c)$, we can use some geometry. Since we are working with a Euclidean covariance matrix, $\Sigma = \sigma^2 I$, the Monte Carlo best-fit parameter according to a particular pseudodata value \mathbf{d}_p is simply the closest point on the theory surface to the pseudodata value \mathbf{d}_p . Now, the closest point on the theory surface to the pseudodata $\mathbf{t}(c) + \lambda \sigma^2 \mathbf{t}(c)$ is c (modulo 2π) for $\lambda \in [-1/\sigma^2, \infty)$, and $c + \pi$ (modulo 2π) otherwise; see Fig. 2.3 for a more detailed description. In particular, this implies that $\Lambda(c) = [-1/\sigma^2, \infty)$. This allows us to directly calculate the Monte Carlo posterior, via a change of variables:

$$u = t_0 \sigma \lambda - \frac{(\mathbf{t}(c) \cdot \mathbf{d}_0 - t_0^2)}{t_0 \sigma}. \quad (2.33)$$

This yields the final Monte Carlo posterior as:

$$\begin{aligned} & \exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(c)\right) \exp\left(\frac{(\mathbf{t}(c) \cdot \mathbf{d}_0 - t_0^2)^2}{2t_0^2 \sigma^2}\right) \int_{-\frac{\mathbf{t}(c) \cdot \mathbf{d}_0}{t_0 \sigma}}^{\infty} du \left(\sigma^2 u + \frac{\sigma \mathbf{t}(c) \cdot \mathbf{d}_0}{t_0}\right) \exp\left(-\frac{1}{2}u^2\right) \\ & \propto \exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(c)\right) \exp\left(\frac{(\mathbf{t}(c) \cdot \mathbf{d}_0 - t_0^2)^2}{2t_0^2 \sigma^2}\right) \left(\exp\left(-\frac{(\mathbf{t}(c) \cdot \mathbf{d}_0)^2}{2t_0^2 \sigma^2}\right) + \frac{\mathbf{t}(c) \cdot \mathbf{d}_0}{\sqrt{2}t_0 \sigma} \sqrt{\pi} \operatorname{erfc}\left(-\frac{\mathbf{t}(c) \cdot \mathbf{d}_0}{\sqrt{2}t_0 \sigma}\right)\right), \end{aligned} \quad (2.34)$$

where erfc is the complementary error function, defined by:

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} dt e^{-t^2}. \quad (2.35)$$

This formula can be simplified by absorbing the ‘Bayesian’ part of the posterior, yielding a Monte Carlo posterior which is proportional to the compact form:

$$1 + \frac{\mathbf{t}(c) \cdot \mathbf{d}_0}{\sqrt{2}t_0 \sigma} \sqrt{\pi} \exp\left(\frac{(\mathbf{t}(c) \cdot \mathbf{d}_0)^2}{2t_0^2 \sigma^2}\right) \operatorname{erfc}\left(-\frac{\mathbf{t}(c) \cdot \mathbf{d}_0}{\sqrt{2}t_0 \sigma}\right). \quad (2.36)$$

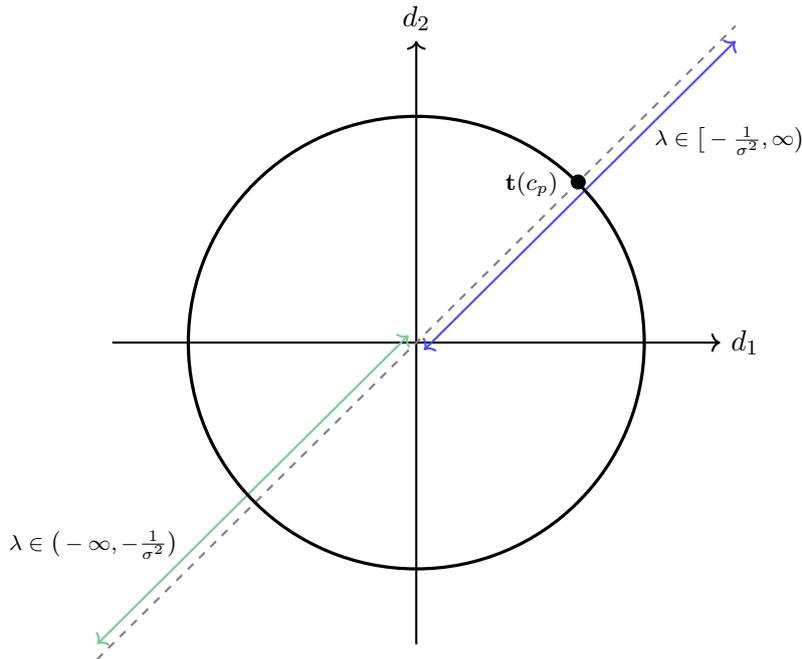


Figure 2.3. A depiction of the circular theory described in Example 3. The dashed grey line indicates the possible values of \mathbf{d}_p described by the range of λ shown. Pseudodata of the form $\mathbf{d}_p = \mathbf{t}(c_p) + \lambda\sigma^2\mathbf{t}(c_p) = (1 + \lambda\sigma^2)\mathbf{t}(c_p)$ has best-fit parameter value c_p if and only if $\lambda \in [-1/\sigma^2, \infty)$, the green portion of the dashed line indicated in the figure. On the other hand, pseudodata of this form has best-fit parameter value $c_p + \pi$ otherwise.

Note that this formula for the Monte Carlo posterior does not contain any singular behaviour; however, it still only agrees with the Bayesian approach if an extremely specific, unmotivated prior is chosen. If instead we choose a sufficiently wide uniform prior for the Bayesian method, we can readily compare the two posteriors through some plots to exhibit the behaviour. In Fig. 2.4 we show a comparison of the two posteriors in two cases: (i) when $|\mathbf{d}_0| \gg t_0$; (ii) when $|\mathbf{d}_0| \ll t_0$. In particular, we see that the Monte Carlo approach can both underestimate the parameter uncertainty that the Bayesian approach would imply, and overestimate this uncertainty. This is important, because it is not immediately obvious from the formula Eq. (2.18) that both possibilities can occur.

Further, this motivates the phenomenological study in the sequel; we emphasise the general point that *Eq. (2.18) is sufficiently complicated that the authors do not have a full understanding of its analytic behaviour*. Therefore, in realistic scenarios, our strategy in benchmarking the agreement of the Bayesian and Monte Carlo posteriors is to compute both and check.

3 Applications in high energy physics

In this section, we apply the mathematical discussion from Section 2 to some phenomenological examples in high-energy physics. In particular, we contrast the Bayesian and Monte Carlo posteriors in the case of fits of SMEFT Wilson coefficients in Sect. 3.1, before per-

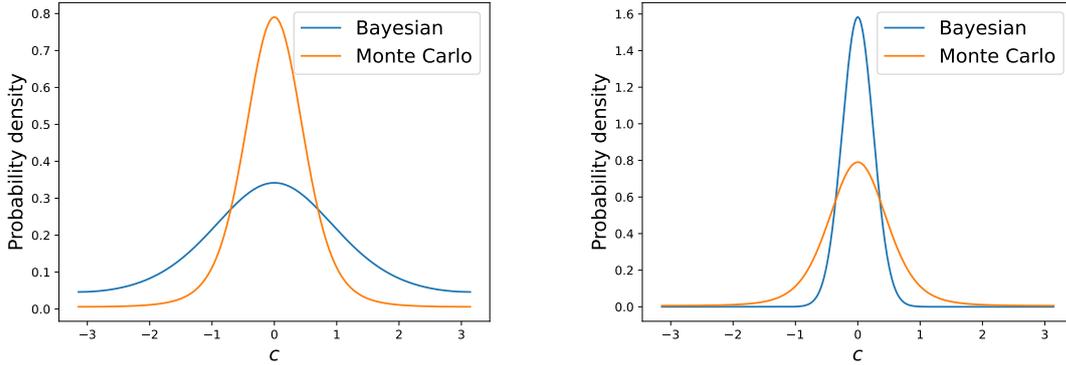


Figure 2.4. A comparison of the Monte Carlo and Bayesian posterior distributions for the circular theory described in Example 3. In the first panel, we show the distributions when $|\mathbf{d}_0| \gg t_0$ (in particular, with $\mathbf{d}_0 = (2, 0)$, $t_0 = 0.5$ and $\sigma = 1$); in this case, the Monte Carlo posterior is much more peaked compared to the Bayesian posterior. On the other hand, in the second panel we show the distributions when $|\mathbf{d}_0| \ll t_0$ (in particular, with $\mathbf{d}_0 = (2, 0)$, $t_0 = 8$ and $\sigma = 1$); in this case, the opposite behaviour is observed.

forming the same exercise in the context of uncertainty estimation for the PDFs of the proton in Sect. 3.2.

In summary, in Sect. 3.1 we show that current global SMEFT fits cannot reliably make use of the Monte Carlo replica method. We find that uncertainties on SMEFT Wilson coefficients produced by the Monte Carlo replica method are discrepant from those estimated by a Bayesian method. Similarly, in Sect. 3.2 we show that, in a heavily simplified model parametrisation, global PDF fits also suffer from the same issue. We demonstrate that when data from proton-proton collisions are included, the Monte Carlo replica method applied to our toy model leads to significantly smaller uncertainty estimates for the low- x PDFs, as compared to a Bayesian approach.

3.1 SMEFT fits

The *Standard Model Effective Field Theory* (SMEFT) treats the SM as a low-energy effective limit of an ultraviolet theory. As such, it extends the SM Lagrangian by a series of non-renormalisable operators built from the SM fields and respecting the SM symmetries:

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{d=5}^{\infty} \sum_{i=1}^{N_d} \frac{c_d^{(i)} \mathcal{O}_d^{(i)}}{\Lambda^{d-4}}, \quad (3.1)$$

where Λ is some characteristic energy scale of New Physics (i.e. the scale at which we expect the SM to break down, becoming unreliable as an effective theory), and the sum over d is a sum over the dimension of the operators. At each dimension d , there are N_d independent operators, $\mathcal{O}_d^{(i)}$, indexed by $i = 1, \dots, N_d$, with corresponding couplings $c_d^{(i)}$. The couplings $c_d^{(i)}$ are called *Wilson coefficients*, and parametrise deviations from the SM. See Ref. [41] for a review. At $d = 5$, neutrino measurements place strong constraints on $c_5^{(i)}$; however,

global fits of the dimension six operators have become a topic of interest in both theoretical and experimental literature in recent years, see for example Refs. [1–5, 20, 31, 42–57].

Cross-section predictions from the dimension-six SMEFT generically take the form:

$$\mathbf{t}(\mathbf{c}) = \mathbf{t}_{\text{SM}} + \mathbf{t}_{\text{lin}}\mathbf{c} + \mathbf{t}_{\text{quad}}(\mathbf{c} \otimes \mathbf{c}), \quad (3.2)$$

where \mathbf{t}_{SM} is a $N_{\text{dat}} \times 1$ vector of SM predictions, \mathbf{t}_{lin} is an $N_{\text{dat}} \times N_{\text{op}}$ matrix of linear SMEFT predictions (with N_{op} the number of SMEFT operators in the fit), and \mathbf{t}_{quad} is an $N_{\text{dat}} \times N_{\text{op}}^2$ matrix of quadratic SMEFT predictions. The use of the *Kronecker product* \otimes of two vectors allows us to express the quadratic predictions in a fully vectorial form.

Importantly, this is a non-linear theory, and as we saw in Example 2 of Sect. 2.3 this can lead to discrepancies between the Monte Carlo posterior and the Bayesian posterior (unless a specific highly non-trivial prior is chosen). The quadratic behaviour of the SMEFT theory predictions additionally implies that there may be delta function singularities as we saw in Example 2 of Sect. 2.3. We show that this is likely to be the case in the global fit carried out below.

In the remainder of this section, we will compare the Bayesian and Monte Carlo posterior distributions of the SMEFT Wilson coefficients constrained specifically by top sector observables. A comparison between the Bayesian and Monte Carlo posteriors was previously discussed in Appendix E of Ref. [20], for the simple case of one Wilson coefficient constrained by a single differential measurement. There it was found that the choice of methodology led to a marked difference in the posterior obtained: the Bayesian posterior was significantly wider than the Monte Carlo posterior, which exhibited the ‘spiked’ behaviour induced by the presence of a delta function as discussed in Sec. 2.2. In this section, we assess whether this discrepancy remains at the level of a global fit.

To do so, we perform an analysis of the 175 measurements used in the simultaneous PDF and SMEFT fit of Ref. [20], encompassing $t\bar{t}$, $t\bar{t} + X$, single top, single top + X , $t\bar{t}t\bar{t}$, $t\bar{t}b\bar{b}$ and top decay observables. These observables will be used to determine the posterior distributions of 25 Wilson coefficients of the dimension-6 SMEFT, the notation for which can be found in Table B.1 of [20]. Datapoints and theory predictions are taken directly from Ref. [20], and we include the quadratic effect of the dimension-6 coefficients on our SMEFT theory predictions. The SMEFT code [1] is used to compare the Nested Sampling and Monte Carlo replica methodologies due to the availability of both methodologies in the public code. Note that the SMEFT collaboration is aware of the limitations of the Monte Carlo replica method and has now deprecated the feature, stressing that it can be reliably used only for linear fits.

Fig. 3.1 shows the marginalised credible intervals at 68% and 95% on the SMEFT coefficients, resulting from the Nested Sampling and Monte Carlo replica methodologies. Firstly, we observe that the discrepancy between the methodologies, as observed in the 1-parameter fit of Ref. [20], remains at the level of the global fit. In a global fit of 25 Wilson coefficients to all available top datasets, visible discrepancies between the two approaches are found. In many of the four-fermion operators, the Monte Carlo replica method leads to a systematic shift downwards from the SM, for example c_{ut}^8 and $c_{qq}^{8,3}$, while the coefficients c_{tG} and c_{qt}^1 are similarly shifted upwards from the SM. In some cases the constraints found

by the Monte Carlo replica method are significantly wider than those obtained by Nested Sampling, for example c_{qd}^8 and c_{qq}^{11} , while c_{tZ} is found to be very well constrained by the Monte Carlo replica method compared to Nested Sampling.

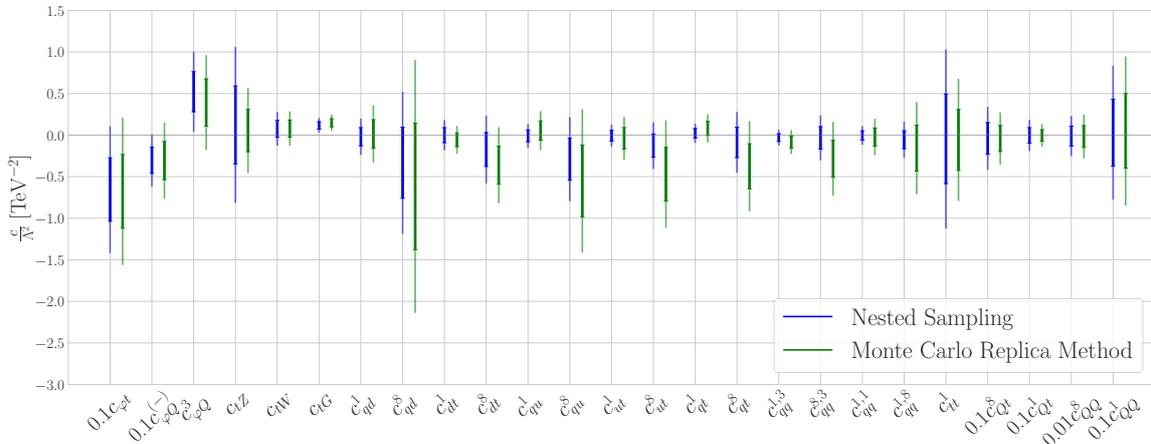


Figure 3.1. Marginalised constraints on the top sector of the SMEFT, comparing the Monte Carlo replica method and Nested Sampling.

The Bayesian and Monte Carlo posteriors, some of which are shown in Fig. 3.2, illuminate the source of the discrepancies found above. We observe that the narrow constraint on c_{tZ} obtained from the Monte Carlo replica method results from a spiked distribution, which showcases the delta function behaviour discussed in Sect. 2.2. Similarly spiked behaviour is observed in the posterior distributions of the four-fermion operators, in particular $c_{qq}^{1,1}$ and c_{qd}^8 , while the posterior distributions of c_{qt}^1 and c_{ut}^8 are found to be highly skewed and non-Gaussian relative to the Bayesian posteriors. Finally, we observe that for c_{tG} , both methodologies obtain Gaussian-like posterior distributions of similar widths; however the Monte Carlo replica method shifts the distribution in the positive direction, leading to a stronger pull from the SM than that obtained with Nested Sampling.

3.2 PDF fits

The parton distributions of the proton are essential ingredients in collider predictions whenever there are protons in the initial state. Roughly speaking, the PDFs represent the probability densities in x for the various constituents of the proton to participate in a collision, whilst carrying a fraction x of the parent proton’s momentum; the participating constituent then goes on to participate in a ‘hard reaction’ with other particle species in the collision.

In more detail, PDFs take the functional form $f_a(x, Q^2)$ where a denotes the flavour of the constituent, x denotes the momentum fraction, and Q^2 denotes a characteristic scale for the process, usually set to the energy of some particles in the collision.⁷ Predictions for the cross-section of a given collision are usually estimated in terms of a discretised version of the *factorisation theorems* for the relevant processes, evaluated on some discrete x -grid

⁷The scale is called the *factorisation scale*, and plays a similar role to the arbitrary scale introduced in renormalisation theory.

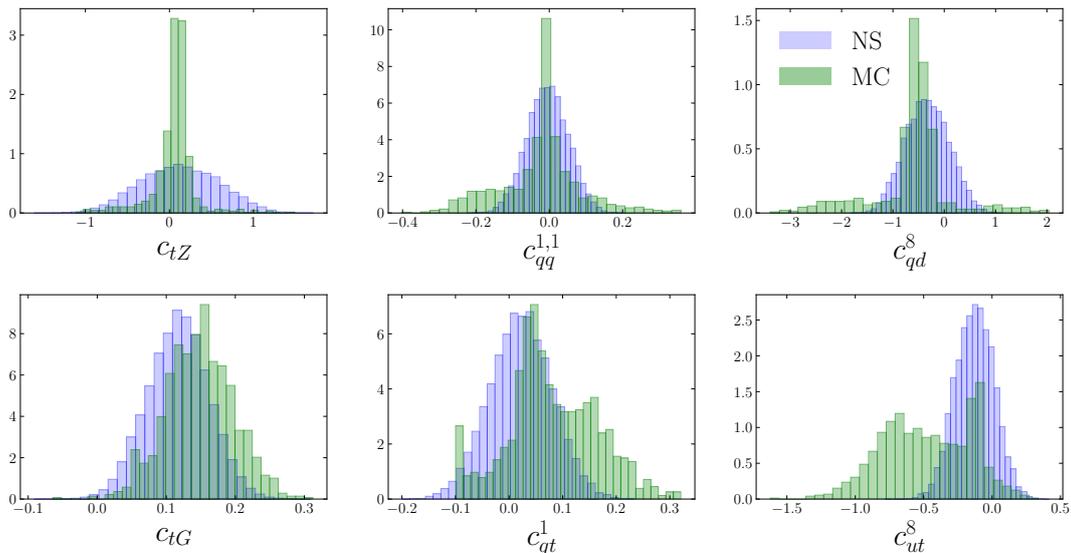


Figure 3.2. Selection of SMEFT posteriors showcasing significant differences in the results obtained by Nested Sampling and the Monte Carlo replica method.

$(x_1, \dots, x_{N_{\text{grid}}})$, taking the general form for the i th datapoint in a dataset:

$$t_i = \begin{cases} \sum_{a=1}^{N_{\text{flav}}} \sum_{\alpha=1}^{N_{\text{grid}}} \text{FK}_{i,a\alpha} f_a(x_\alpha, Q_0^2), & \text{if the } i\text{th point is deep-inelastic scattering data;} \\ \sum_{a,b=1}^{N_{\text{flav}}} \sum_{\alpha,\beta=1}^{N_{\text{grid}}} \text{FK}_{i,a\alpha b\beta} f_a(x_\alpha, Q_0^2) f_b(x_\beta, Q_0^2), & \text{if the } i\text{th point is hadronic data.} \end{cases} \quad (3.3)$$

The arrays $\text{FK}_{i,a\alpha}$ and $\text{FK}_{i,a\alpha b\beta}$ are called *fast-kernel tables* in the PDF-fitting parlance, and encode both the evolution of the PDFs from some fixed initial scale Q_0^2 (usually taken to be $Q_0 = 1.65$ GeV), together with the convolution of the evolved PDFs with a cross-section for the hard reaction.

Importantly, we observe that for deep-inelastic scattering data, the theory predictions are *linear* in the PDFs, but the contribution for hadronic (proton-proton) data is in fact *purely quadratic* in the PDFs. This will have significant consequences for the use of the Monte Carlo replica method for inferring the PDFs from collider data, as we shall discuss below.

In order to compare the Monte Carlo replica method and the Bayesian approach, we have developed a private code that allows both methodologies to be adopted when fitting PDFs. In particular, Bayesian PDFs are produced using the Nested Sampling algorithm `ultranest` [58–60], which is a well-established algorithm to sample from multimodal posterior distributions and infer estimates of parameter uncertainties. Because of this, it is an ideal framework to perform statistical inference in the presence of non-linearity.

Throughout the study we employ the publicly available FK tables and datasets provided by the NNPDF collaboration [12].

3.2.1 A toy PDF model

The PDFs, being functions, have infinitely many degrees of freedom, and hence cannot be determined completely by finite amounts of data; the problem is ill-posed. Therefore, PDF collaborations typically assume a specific functional form for the PDFs at the initial scale $Q^2 = Q_0^2$. Most collaborations, for example CTEQ [61, 62], JAM [13–15] and MSHT [63, 64], use a fixed functional form with $\mathcal{O}(30 - 40)$ parameters, while the NNPDF collaboration [12] uses a flexible Neural Network.

It remains an open question to what extent the Monte Carlo replica method as applied in a realistic PDF fit agrees or disagrees with a Bayesian method for uncertainty propagation. This paper does not intend to solve this problem, and we consider it important future work to address this issue (especially in the context of PDF-SMEFT interplay, as discussed in Ref. [20] and Sect. 3.1).

In this work we aim lower, and take a first step towards answering this question, by instead working with a simplified PDF model and artificially generated data.⁸ In the remainder of this section, we discuss the toy model (based on linear interpolation of the PDF grid), and describe the generation of the artificial data used in this study.

The toy model. In our toy model, to reduce dimensionality, we first suppose that the structure of the proton can be completely described in terms of three flavours in the evolution basis, Σ , g , V , the singlet, gluon and valence distributions respectively, where the singlet and the valence are defined in terms of the quark flavours as:

$$\Sigma = u + \bar{u} + d + \bar{d} + s + \bar{s} + 2c, \quad (3.4)$$

$$V = (u - \bar{u}) + (d - \bar{d}) + (s - \bar{s}). \quad (3.5)$$

Samples of these flavours taken from a realistic PDF fit (namely the recent NNPDF4.0 determination) are shown in Fig. 3.3. These three flavours have the following structure: (i) the singlet is a monotonically decreasing function; (ii) the gluon is peaked at $x \approx 10^{-2}$; (iii) the valence is peaked at $x \approx 0.2$.

Now, in standard PDF fits, a parametrisation (either polynomial-based or neural network) is assumed in order to model the structure of these flavours. However, both polynomial and neural network parametrisations are non-linear, so even in the DIS case where PDFs enter linearly into theory predictions, the results of Sect. 2.3 imply that we cannot expect exact agreement between the Monte Carlo replica method and the Bayesian method. Hence, in order to produce a useful benchmark exercise where agreement is demonstrated in a PDF fitting context, we introduce the following ‘linear interpolation’ parametrisation of the PDFs.

⁸To avoid confusion, we reserve the term *pseudodata* exclusively for discussion of Monte Carlo pseudodata. The term *artificial data* instead describes artificial central experimental values, \mathbf{d}_0 , in the discussion of the previous chapters.

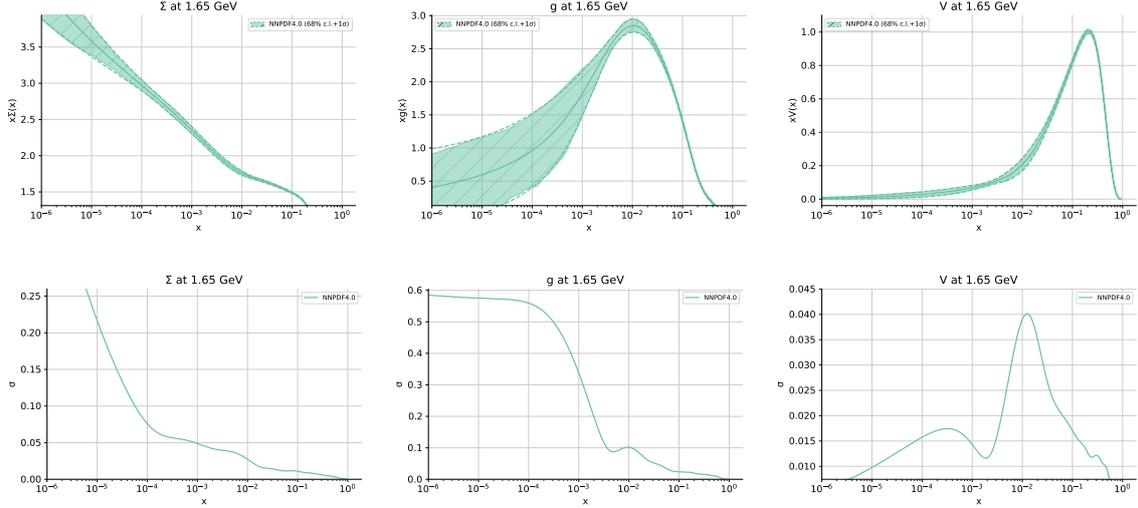


Figure 3.3. The singlet Σ , gluon g , and valence V distributions in the NNPDF4.0 fit; their respective percentage uncertainties (at 68%) are also shown in the panel below. The singlet is monotonically decreasing, whilst the gluon and valence distributions have peaks at around $x \approx 10^{-2}$ and $x \approx 0.2$ respectively.

In our parametrisation, for each flavour f , we select a ‘reduced x -grid’:

$$\left(x_1^f, \dots, x_{N_{\text{grid}}(f)}^f\right). \quad (3.6)$$

This reduced x -grid is now treated as an interpolation grid, from which each flavour of PDF is constructed. In particular, we define the PDF at the initial scale as a linear interpolant of its values on this reduced x -grid, via:

$$f(x, Q_0^2) = \begin{cases} f(x_1^f, Q_0^2) & \text{if } x \leq x_1^f; \\ \left(\frac{x_{i+1}^f - x}{x_{i+1}^f - x_i^f}\right) f(x_i^f, Q_0^2) + \left(\frac{x - x_i^f}{x_{i+1}^f - x_i^f}\right) f(x_{i+1}^f, Q_0^2) & \text{if } x \in [x_i, x_{i+1}], \text{ for } i = 1, \dots, N_{\text{grid}}(f); \\ f(x_{N_{\text{grid}}(f)}^f, Q_0^2) & \text{if } x > x_{N_{\text{grid}}(f)}^f. \end{cases} \quad (3.7)$$

In particular, the ‘parameters’ which describe this PDF model are simply the values of the PDF on the reduced x -grid; once these values $f(x_1^f, Q_0^2), \dots, f(x_{N_{\text{grid}}(f)}^f, Q_0^2)$ are specified, then the PDF itself is specified everywhere. We have chosen to extrapolate simply by taking the value of the PDF below the lowest grid point to always be equal to its value *at* the lowest grid point; similarly for the highest grid point.

This somewhat elementary approach has the following distinct advantages:

- Importantly for our purposes, the PDF parametrisation is *linear*; therefore, DIS predictions are truly linear predictions of the PDF parameters, and predictions for proton-proton collisions are truly quadratic predictions of the PDF parameters. In particular, this allows us to benchmark the Monte Carlo approach directly against a Bayesian approach in the DIS-only case.

- As we increase the number of grid points and flavours, the model becomes increasingly realistic. This implies that if we take the grid to be the FK-table grid in our parametrisation, and use all of the fitted flavours, the linear interpolation becomes completely immaterial, since the FK-table only ‘sees’ the PDF at the - now complete - interpolation grid. With the advent of increased computational power, making use of tools such as GPUs for instance, such a direct fit of the grid could be within reach in the next few years.

On a practical note, for this study we have selected grids such that as much of the ‘structure’ displayed in Fig. 3.3 is retained. We choose 12 grid points in each case, giving a total of 36 fitted parameters. The precise choice is:

$$(1.57 \times 10^{-4}, 3.62 \times 10^{-4}, 8.31 \times 10^{-4}, 1.90 \times 10^{-4}, 4.32 \times 10^{-3}, 9.70 \times 10^{-3}, \\ 2.11 \times 10^{-2}, 4.34 \times 10^{-2}, 8.23 \times 10^{-2}, 1.41 \times 10^{-1}, 2.20 \times 10^{-1}, 3.14 \times 10^{-1}) \quad \text{for } \Sigma; \quad (3.8)$$

$$(3.62 \times 10^{-4}, 5.49 \times 10^{-4}, 8.31 \times 10^{-4}, 1.90 \times 10^{-3}, 4.33 \times 10^{-3}, 9.70 \times 10^{-3}, \\ 2.11 \times 10^{-2}, 4.34 \times 10^{-2}, 8.23 \times 10^{-2}, 1.41 \times 10^{-1}, 2.20 \times 10^{-1}, 3.14 \times 10^{-1}) \quad \text{for } g; \quad (3.9)$$

$$(3.05 \times 10^{-2}, 4.34 \times 10^{-2}, 6.05 \times 10^{-2}, 8.23 \times 10^{-2}, 1.09 \times 10^{-1}, 1.78 \times 10^{-1}, \\ 2.65 \times 10^{-1}, 3.67 \times 10^{-1}, 4.80 \times 10^{-1}, 5.40 \times 10^{-1}, 6.01 \times 10^{-1}, 6.65 \times 10^{-1}) \quad \text{for } V. \quad (3.10)$$

Observe that we have populated the region near $x \approx 0.2$ for the valence with significantly more points, in order to model its peak structure in this region. For the singlet and gluon distributions, we reduce the number of points for $x < 10^{-4}$, where the uncertainty becomes larger, implying a lack of constraints from the dataset. In particular, because of the choice of the grid points, all of the parameters are constrained by the data and do not lie in the extrapolation region.

Generation of the artificial data. Due to the extremely simplified nature of the PDF model in this study, we choose to use purely artificial data throughout. Our artificial data is produced assuming that the true PDF law of Nature is a linear interpolant of the above form (thereby bypassing any modelling error), and random noise is added on top according to the experimental covariance matrix.

In more detail, we assume that the true law of Nature is NNPDF4.0 [12], and produce predictions for the dataset in question (either DIS only, proton-proton only, or the full dataset, explored in turn in the sequel) using this PDF set; write these predictions as \mathbf{t}_0 . The artificial data is then generated as a sample from the multivariate normal distribution:

$$\mathcal{N}(\mathbf{t}_0, \Sigma), \quad (3.11)$$

where Σ is the relevant \mathbf{t}_0 covariance matrix.⁹

⁹See Ref. [65] for details on the definition of the \mathbf{t}_0 covariance matrix.

3.2.2 Results for DIS-only fits

The vast majority of the data which enters a PDF fit is *deep inelastic scattering* data, which depends linearly on the PDFs.¹⁰ As indicated in Eq. (3.3), DIS predictions take the form:

$$\mathbf{t}(\mathbf{f}) = \text{FK} \cdot \mathbf{f},$$

where FK is the $N_{\text{dat}} \times (N_{\text{grid}} N_{\text{flav}})$ fast-kernel table for the prediction. Importantly for our purposes, this is a linear prediction, so Monte Carlo and Bayesian uncertainty estimates on PDFs in the grid parametrisation obtained from DIS data only must agree according to the general theory presented in Sect. 2.

Using a private code, we have benchmarked this agreement in the context of the model and artificial data described above, using the entire DIS dataset entering the NNPDF4.0 analysis [12], consisting of 2968 datapoints. We display the resulting PDFs in Fig. 3.4; as anticipated, agreement is exceptional between the Monte Carlo and Bayesian approaches.

The three PDF fits in Fig. 3.4 are obtained as follows. The Monte Carlo PDF is, as one might expect, produced using the Monte Carlo replica method. The Nested Sampling PDF is a Bayesian PDF produced using the Nested Sampling algorithm; a uniform prior is used on each of the grid points to ensure the agreement described in Sect. 2. Finally, the ‘analytic’ PDF is produced by sampling from the analytic posterior distribution that can be calculated for this linear problem; it is found to be:

$$\mathcal{N}\left((\text{FK})^T \Sigma^{-1} \text{FK}\right)^{-1} \text{FK}^T \Sigma^{-1} \mathbf{d}_0, (\text{FK}^T \Sigma^{-1} \text{FK})^{-1}\right). \quad (3.12)$$

This ‘analytic’ result allows us to test the efficacy of the numerical approaches presented by the Monte Carlo and Nested Sampling techniques.

3.2.3 Results for hadronic-only fits

In this section, we repeat the exercise of Sect. 3.2.2, but now using *hadronic-only* data instead of DIS-only data. In the context of PDF fitting, hadronic data refers to data coming from proton-proton collisions; as a result, the theory predictions for these processes take a quadratic form, as presented in Eq. (3.3). In particular, we can expect some singular behaviour of the Monte Carlo posterior, as presented in the toy examples of Sect. 2 and in the SMEFT fits of Sect. 3.1.

The settings remain precisely the same as in Sect. 3.2.2, except we now use the complete NNPDF4.0 hadronic dataset, with the exception of jet data.¹¹ This dataset comprises a total of 1027 points.

The result of Monte Carlo replica fits and Bayesian fits of these data are presented in Fig. 3.5. In this instance, good agreement between the Monte Carlo PDF and the Nested Sampling PDF is found in the mid to high- x region, whereas significant disagreement is observed at low- x ; in particular, the uncertainty on the Monte Carlo PDF is reduced

¹⁰Excluding, for example, ratio observables - the NMC measurement of F_2^d/F_2^p is an example [66].

¹¹The study is conducted by making use of the new theory pipeline of NNPDF and when producing results, jet data was not available. We note, however, that it has been recently implemented.

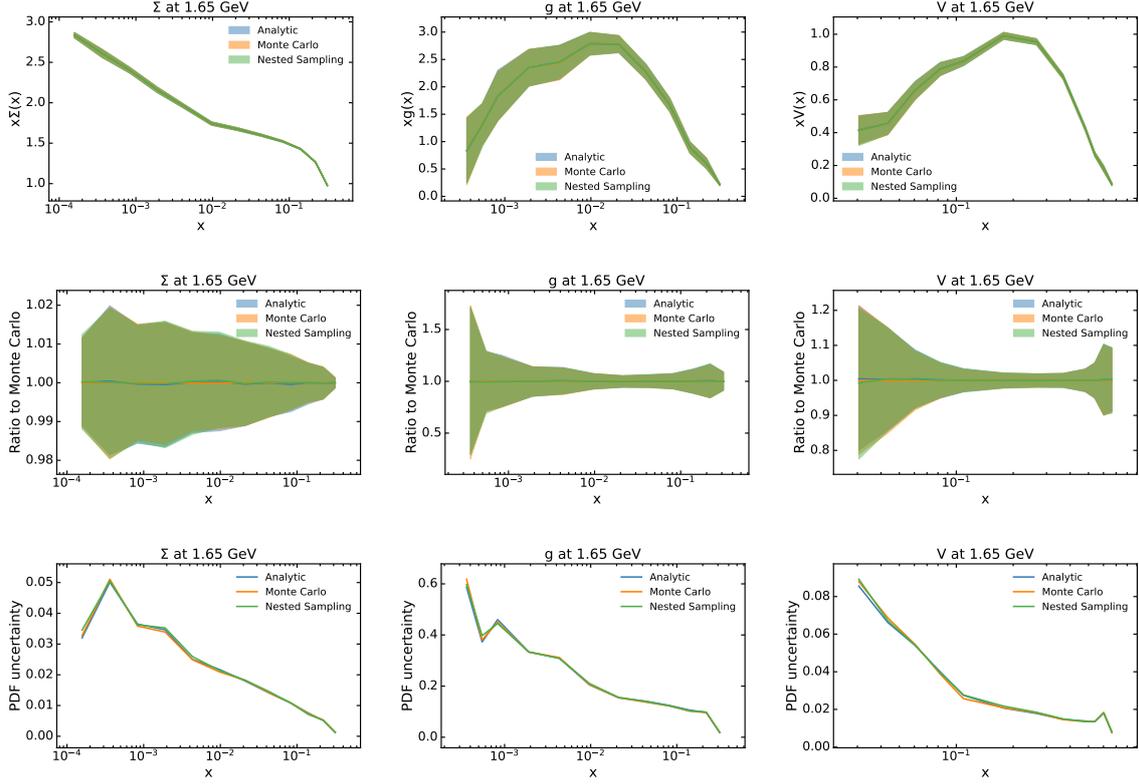


Figure 3.4. A comparison of three PDF fits to DIS-only data using the simplified model described above. The meaning of the Monte Carlo, Nested Sampling, and Analytic PDF sets is described in the text. The columns display the flavours Σ , g , V from left to right. The rows display the absolute values of the PDFs, the ratio of the PDFs to the Monte Carlo PDF, and the uncertainties on each of the PDF sets, from top to bottom. Agreement is outstanding between the three exercises.

compared to the Nested Sampling PDF for the singlet and gluon at lower values of x . The most striking reduction is for the gluon at $x = 5.49 \times 10^{-4}$, which undergoes a 90% reduction in uncertainty when going from the Bayesian PDF to the Monte Carlo PDF.

The valence distribution, on the other hand, agrees well between the Bayesian and Monte Carlo fits. However, due to the theoretical calculation being poorly understood in the non-linear case, we cannot say with confidence precisely why this behaviour occurs. A possible conjecture comes from the fact that the valence in this region is particularly well-constrained by these data, so that about its best-fit value V^* , we can write $V \approx V^* + \Delta V$ with ΔV particularly small. Then quadratic contributions V^2 can be linearised via $V^2 \approx (V^*)^2 + 2V^* \cdot \Delta V$, resulting in a Gaussian posterior around the best-fit value, where we expect good agreement between the methodologies.

3.2.4 Results for global fits

Finally, in this section, we repeat the benchmarking study of the previous two sections for the global dataset, using both the DIS and hadronic data. Given the excellent agreement between the DIS-only Monte Carlo and Bayesian fits, and the poor agreement between the

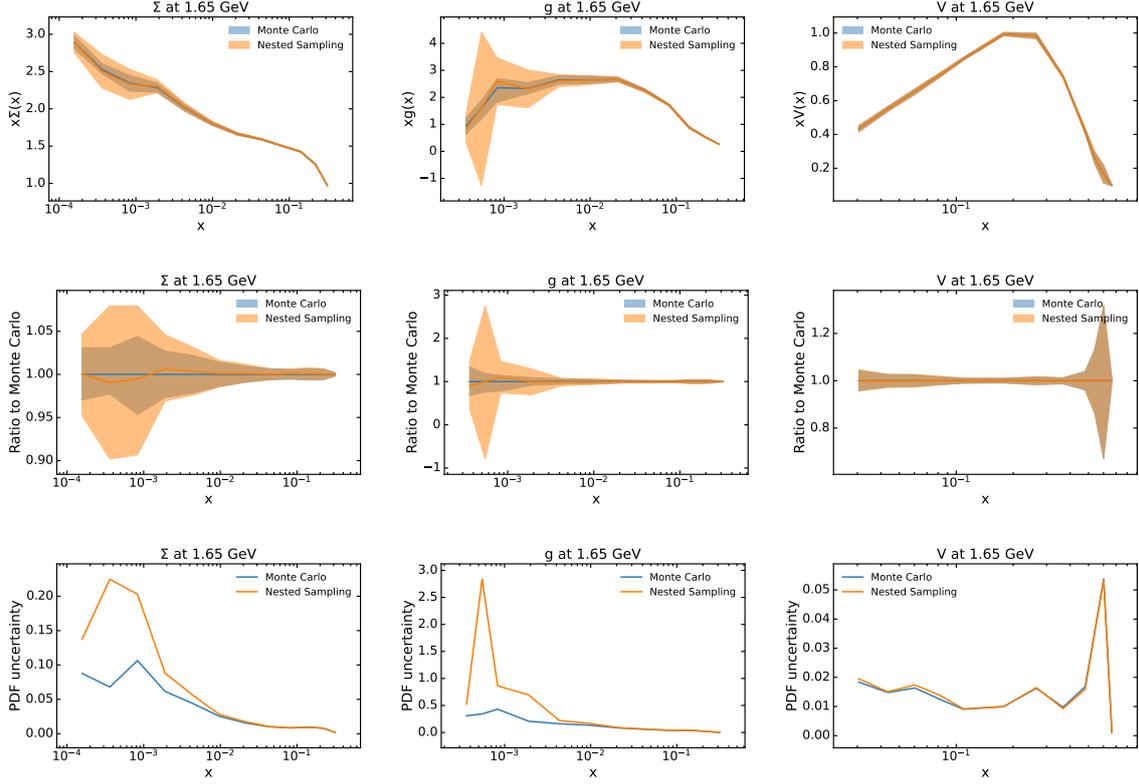


Figure 3.5. A comparison of three PDF fits to hadronic-only data using the simplified model described above. The meaning of the Monte Carlo and Nested Sampling PDFs is described in the text. The columns display the flavours Σ , g , V from left to right. The rows display the absolute values of the PDFs, the ratio of the PDFs to the Monte Carlo PDF, and the uncertainties on each of the PDF sets, from top to bottom. Significant disagreement is observed between the two exercises.

hadronic-only Monte Carlo and Bayesian fits, it is our general expectation that the result of the global fit will be some middle ground between the two cases.

The results of the Monte Carlo and Bayesian fits to the full dataset are presented in Fig. 3.6. As expected, we find that the Monte Carlo fit still shows some discrepancy from the Nested Sampling one in the low- x region, but the disagreement is less pronounced than in the purely hadronic case. In particular, the singlet PDF Σ shows a less significant reduction in uncertainties in the low- x region, of order 20% now, in going from the Nested Sampling to the Monte Carlo PDFs. On the other hand, the gluon PDF still shows a considerable discrepancy with the Bayesian result, displaying a conspicuous underestimation of PDF error bands; the worst discrepancy is around 60%.

In summary, in the broad dataset considered, the impact of the hadronic data on the full fit is already enough to cause the two methodologies to diverge. Whilst we have only demonstrated this discrepancy in the case of a simplified PDF model, we believe that the result motivates further efforts towards the implementation of Bayesian methods to quantify uncertainties in more realistic fits, especially if in the coming decade new and more accurate hadronic data from the LHC is included in global PDF fits.

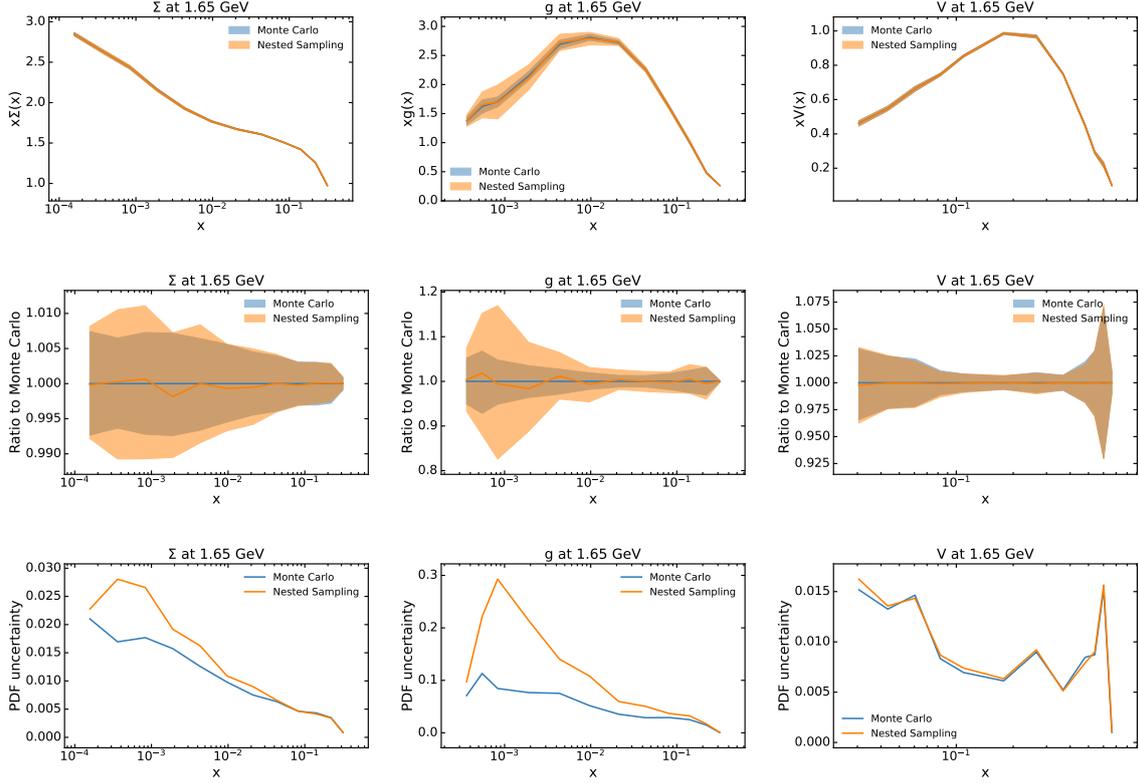


Figure 3.6. The same as figures Fig. 3.4 and Fig. 3.5, but now fitting the PDFs using the complete global dataset (combining the DIS and hadronic data from the previous cases). The level of disagreement between the Monte Carlo and Bayesian fits is reduced as compared to the hadronic-only case, but there is still significant tension between the two approaches.

4 Conclusions

For the first time, this work explored the mathematical foundations of the Monte Carlo replica method beyond linear models, before benchmarking the Monte Carlo replica method against a Bayesian approach in two global fit scenarios from the high-energy physics literature.

In Sect. 2, we produced an original calculation of the Monte Carlo ‘posterior’ distribution and showed it is only comparable to the Bayesian method in special cases (e.g. for a linear model, the two approaches agree if a sufficiently wide uniform prior is used for the Bayesian). The final result, Eq. (2.17), has some of its behaviour understood (particularly the delta function singularities that can appear in the posterior), but for the most part it remains intractably difficult to manipulate.

In Sect. 3.1, we provided a benchmark of the Bayesian method against the Monte Carlo replica method in the context of a fit of the SMEFT Wilson coefficients in the top sector. We find poor agreement between the implied parameter bounds, and as a result, we discourage the use of the Monte Carlo replica method in this arena of global fits. Many SMEFT fit collaborations [2, 31, 57] have already embraced a Bayesian perspective.

In Sect. 3.2, we perform the same benchmarking exercise between the two approaches in the context of a toy fit of PDFs to a global dataset. We found exceptionally good agreement between the Monte Carlo replica method and the Bayesian method when the DIS-only dataset was used (indeed, this must be the case according to the calculation given in Sect. 2). However, when the fit was instead performed on the hadronic-only dataset, we found that the PDF uncertainties implied by the Monte Carlo replica method, despite agreeing well at high- x , were significantly reduced compared to the Bayesian method at low- x . This behaviour persisted, although milder, when we performed a final fit using the global DIS plus hadronic dataset. Whilst the toy fit we performed was a simplification of a modern PDF fit, the results imply that caution is necessary when using the Monte Carlo replica method, especially as more hadronic data are added to the fits. On the other hand, the authors wish to emphasise the point that the level of disagreement, if any, in a realistic PDF fit remains unknown.

While conducting the PDF analysis, we implemented a framework to perform Bayesian inference using Nested Sampling. The developed tool will be featured in a forthcoming publication. We believe that further efforts should be dedicated towards the development of a fully Bayesian approach to PDF fitting, especially if fit simultaneously in combination with EFT coefficients. This work offers a compelling motivation for such a programme.

Acknowledgments

We are extremely grateful to Maria Ubiali for her support with this project. We would like to thank Manuel Morales Alvarado, Anke Biekötter, Luigi Del Debbio, Tommaso Giani, and Zahari Kassabov for useful discussions. We thank Richard Nickl for some advice at the early stages of the project. We thank Alastair Young for the discussion of bootstrap methods. We thank the NNPDF collaboration for comments on the manuscript and for making publicly available the data and the fast-kernel tables used in this study.

Mark N. Costantini, Luca Mantani, and James M. Moore are supported by the European Research Council under the European Union’s Horizon 2020 research and innovation Programme (grant agreement n.950246), and partially by the STFC consolidated grant ST/T000694/1 and ST/X000664/1. The work of Maeve Madigan is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 396021762 – TRR 257 Particle Physics Phenomenology after the Higgs Discovery and by the Alexander von Humboldt Foundation.

A Additional analytic example: purely quadratic theory

In this appendix, we give a further calculation of the Monte Carlo posterior for a toy example. Consider $\mathbf{t}(c) = \mathbf{t}_0 + \mathbf{t}_{\text{quad}}c^2$, a purely quadratic theory in one parameter. In this case, the Jacobian matrix is:

$$\left(\frac{\partial \mathbf{t}}{\partial c}\right)^T = 2c\mathbf{t}_{\text{quad}}^T. \quad (\text{A.1})$$

This is of full rank unless $c = 0$. The matrix $M(c)$ can be chosen independently of c to consist of $N_{\text{dat}} - 1$ columns which are orthogonal to the vector \mathbf{t}_{quad} , so that $M^T \mathbf{t}_{\text{quad}} = 0$. Further, the function $c_p(\mathbf{d}_p)$ is double-valued for $c \neq 0$, because of the symmetry $c \mapsto -c$ in the theory prediction. It follows that away from $c = 0$, the Monte Carlo posterior takes the form:

$$2 \exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(\mathbf{c})\right) \left| \det\left(2c\mathbf{t}_{\text{quad}} \middle| \Sigma M\right) \right| \int_{\Lambda(c)} d^{N_{\text{dat}}-1} \boldsymbol{\lambda} \exp\left(-\frac{1}{2}\boldsymbol{\lambda}^T M^T \Sigma M \boldsymbol{\lambda} + \boldsymbol{\lambda}^T M^T (\mathbf{d}_0 - \mathbf{t}_0)\right) \quad (\text{A.2})$$

The set $\Lambda(c)$ is the set of $\boldsymbol{\lambda}$ such that the pseudodata $\mathbf{t}(c) + \Sigma M \boldsymbol{\lambda}$ leads to c as a minimum of the χ^2 -statistic on this pseudodata. We note that the χ^2 on this pseudodata is given by:

$$\chi^2(c') = (\mathbf{t}(c') - \mathbf{t}(c) - \Sigma M \boldsymbol{\lambda})^T \Sigma^{-1} (\mathbf{t}(c') - \mathbf{t}(c) - \Sigma M \boldsymbol{\lambda}) \quad (\text{A.3})$$

$$= (\mathbf{t}_{\text{quad}}(c' - c)^2 - \Sigma M \boldsymbol{\lambda})^T \Sigma^{-1} (\mathbf{t}_{\text{quad}}(c' - c)^2 - \Sigma M \boldsymbol{\lambda}) \quad (\text{A.4})$$

$$= (c' - c)^4 \mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}} + \boldsymbol{\lambda}^T M^T \Sigma M \boldsymbol{\lambda}. \quad (\text{A.5})$$

Since Σ is positive definite, we have that $\mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}} > 0$ and $\boldsymbol{\lambda}^T M^T \Sigma M \boldsymbol{\lambda} > 0$, which shows that the minimum is at $c' = c$. Thus $\Lambda(c) = \mathbb{R}^{N_{\text{dat}}-1}$, the full range. Evaluating the integral then, we have that for $c \neq 0$ the Monte Carlo posterior takes the form:

$$2 \exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(\mathbf{c})\right) \left| \det\left(2c\mathbf{t}_{\text{quad}} \middle| \Sigma M\right) \right| \sqrt{\frac{(2\pi)^{N_{\text{dat}}-1}}{\det(M^T \Sigma M)}} \exp\left(\frac{1}{2}(\mathbf{d}_0 - \mathbf{t}_0)^T M (M^T \Sigma M)^{-1} M^T (\mathbf{d}_0 - \mathbf{t}_0)\right), \quad (\text{A.6})$$

using the standard formula for the Gaussian integral. We can simplify this with a couple of tricks. First, observe that by the Gram-Schmidt procedure, we may choose the columns of M to be orthogonal with respect to the inner product induced by Σ without loss of generality, so that:

$$M^T \Sigma M = I. \quad (\text{A.7})$$

This also allows us to compute the determinant factor. First, we note:

$$\det\left(2c\mathbf{t}_{\text{quad}} \middle| \Sigma M\right) = 2c \det(\Sigma) \det\left(\Sigma^{-1} \mathbf{t}_{\text{quad}} \middle| M\right). \quad (\text{A.8})$$

Further, $\Sigma^{-1} \mathbf{t}_{\text{quad}}$ and M are already orthogonal with respect to the inner product induced by Σ . Normalising $\Sigma^{-1} \mathbf{t}_{\text{quad}}$ with respect to this inner product, we have:

$$(\Sigma^{-1} \mathbf{t}_{\text{quad}})^T \Sigma (\Sigma^{-1} \mathbf{t}_{\text{quad}}) = \mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}}. \quad (\text{A.9})$$

It follows that:¹²

$$\det \left(\frac{\Sigma^{-1} \mathbf{t}_{\text{quad}}}{\mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}}} \middle| M \right) = \pm \sqrt{\det(\Sigma^{-1})}, \quad (\text{A.10})$$

and hence the Monte Carlo posterior simplifies to:

$$4\sqrt{(2\pi)^{N_{\text{dat}}-1} \det(\Sigma) \mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}} |c|} \exp \left(-\frac{1}{2} \chi_{\mathbf{d}_0}^2(\mathbf{c}) \right) \exp \left(\frac{1}{2} (\mathbf{d}_0 - \mathbf{t}_0)^T M M^T (\mathbf{d}_0 - \mathbf{t}_0) \right). \quad (\text{A.11})$$

We can also recognise $MM^T = M(M^T \Sigma M)^{-1} M^T$ as the projector onto the subspace orthogonal to $\mathbf{t}_{\text{quad}}^T \Sigma^{-1}$, which gives:

$$M(M^T \Sigma M)^{-1} M^T = \Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{t}_{\text{quad}} \mathbf{t}_{\text{quad}}^T \Sigma^{-1}}{\mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}}}. \quad (\text{A.12})$$

Hence the Monte Carlo posterior reduces to the (M -independent) simplified form:

$$4\sqrt{(2\pi)^{N_{\text{dat}}-1} \det(\Sigma) \mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}} |c|} \exp \left(-\frac{1}{2} \chi_{\mathbf{d}_0}^2(\mathbf{c}) \right) \cdot \exp \left(\frac{1}{2} (\mathbf{d}_0 - \mathbf{t}_0)^T \left(\Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{t}_{\text{quad}} \mathbf{t}_{\text{quad}}^T \Sigma^{-1}}{\mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}}} \right) (\mathbf{d}_0 - \mathbf{t}_0) \right). \quad (\text{A.13})$$

One can check that this reduces to the previous case with $t_{\text{lin}} = 0$ when we work with $N_{\text{dat}} = 1$.

On the other hand, about $c = 0$, we can parametrise the theory via $f : \{\emptyset\} \rightarrow \mathbb{R}$ with $f(\emptyset) = 0$. Further, the matrix $M(0)$ can simply be taken to be the $N_{\text{dat}} \times N_{\text{dat}}$ identity matrix. The Monte Carlo posterior is then given by:

$$\exp \left(-\frac{1}{2} \chi_{\mathbf{d}_0}^2(c) \right) \delta(c) \det(\Sigma) \int_{\Lambda(0)} d^{N_{\text{dat}}} \boldsymbol{\lambda} \exp \left(-\frac{1}{2} \boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda} + \boldsymbol{\lambda}^T (\mathbf{d}_0 - \mathbf{t}_0) \right). \quad (\text{A.14})$$

To compute the set $\Lambda(0)$, we must find all $\boldsymbol{\lambda}$ such that the pseudodata $\mathbf{t}(0) + \Sigma \boldsymbol{\lambda}$ gives rise to the best-fit value of the parameter c . The χ^2 -statistic evaluated on this pseudodata is given by:

$$\chi^2(c') = (\mathbf{t}(c') - \mathbf{t}(0) - \Sigma \boldsymbol{\lambda})^T \Sigma^{-1} (\mathbf{t}(c') - \mathbf{t}(0) - \Sigma \boldsymbol{\lambda}) \quad (\text{A.15})$$

$$= (\mathbf{t}_{\text{quad}} c'^2 - \Sigma \boldsymbol{\lambda})^T \Sigma^{-1} (\mathbf{t}_{\text{quad}} c'^2 - \Sigma \boldsymbol{\lambda}) \quad (\text{A.16})$$

$$= c'^4 \mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}} - 2c'^2 \mathbf{t}_{\text{quad}}^T \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda} \quad (\text{A.17})$$

$$= \mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}} \left(c'^2 - \frac{\mathbf{t}_{\text{quad}}^T \boldsymbol{\lambda}}{\mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}}} \right)^2 - \frac{(\mathbf{t}_{\text{quad}}^T \boldsymbol{\lambda})^2}{\mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}}} + \boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda}. \quad (\text{A.18})$$

¹²Note that if A is a square matrix with orthonormal columns with respect to the inner product induced by Σ , we have $1 = \det(A^T \Sigma A) = \det(\Sigma) \det(A)^2$, which implies $\det(A) = \pm \sqrt{\det(\Sigma^{-1})}$.

It follows that if $\mathbf{t}_{\text{quad}}^T \boldsymbol{\lambda} \leq 0$, we have that $c' = 0$ is the unique minimiser of the χ^2 -statistic evaluated on the relevant pseudodata; otherwise, we get different roots. Hence $\Lambda(0) = \{\boldsymbol{\lambda} : \mathbf{t}_{\text{quad}}^T \boldsymbol{\lambda} \leq 0\}$, which yields the Monte Carlo posterior about $c = 0$:

$$\exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(c)\right) \delta(c) \det(\Sigma) \int_{\mathbf{t}_{\text{quad}}^T \boldsymbol{\lambda} \leq 0} d^{N_{\text{dat}}} \boldsymbol{\lambda} \exp\left(-\frac{1}{2}\boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda} + \boldsymbol{\lambda}^T (\mathbf{d}_0 - \mathbf{t}_0)\right). \quad (\text{A.19})$$

Overall then, the Monte Carlo posterior takes the form:

$$\begin{aligned} & \exp\left(-\frac{1}{2}\chi_{\mathbf{d}_0}^2(c)\right) \left[\delta(c) \det(\Sigma) \int_{\mathbf{t}_{\text{quad}}^T \boldsymbol{\lambda} \leq 0} d^{N_{\text{dat}}} \boldsymbol{\lambda} \exp\left(-\frac{1}{2}\boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda} + \boldsymbol{\lambda}^T (\mathbf{d}_0 - \mathbf{t}_0)\right) \right. \\ & \left. + 4\sqrt{(2\pi)^{N_{\text{dat}}-1} \det(\Sigma) \mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}}} |c| \exp\left(\frac{1}{2}(\mathbf{d}_0 - \mathbf{t}_0)^T \left(\Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{t}_{\text{quad}} \mathbf{t}_{\text{quad}}^T \Sigma^{-1}}{\mathbf{t}_{\text{quad}}^T \Sigma^{-1} \mathbf{t}_{\text{quad}}}\right) (\mathbf{d}_0 - \mathbf{t}_0)\right) \right] \end{aligned} \quad (\text{A.20})$$

B Numerical challenges of the Monte Carlo replica method

In this section, we discuss some technical challenges related to the MC methodology. As an initial disclaimer, these are not insurmountable issues and with enough fine-tuning and understanding of the problem at hand, they can be overcome. However, during the course of this and previous studies [20, 22], we often had to deal with unexpected difficulties which hindered the results of the fits.

The origin of the problem is that while our objective is to obtain a sample from the posterior distribution of the model parameters, this is achieved not through a sampling algorithm, but through an optimisation procedure which relies on gradient descent. For this reason, the presence of multiple local minima and of degeneracies, that is, continuous surfaces all of which have equal minimal χ^2 -statistic, can impact the result of the fit in ways that could be difficult to spot. These problems are instead well tackled by dedicated posterior sampling algorithms like Nested Sampling or MCMC, where multimodal distributions can be mapped efficiently and reliably.

The first and most basic issue has to do with making sure that the global minimum was actually reached. In our experience, this can be problematic in two distinct cases: when the posterior has a multimodal character and in the presence of strong correlations. Both of these can ruin a successful optimisation because in the first case, one could get trapped in a local minima while in the second, convergence to the true minimum can be very slow (since, in presence of strong correlations, the gradient can be very small).

The methodological uncertainty problem is well known by NNPDF, which has designed a suite of closure tests specifically to estimate the level of success of the neural network in finding the global minimum. In a level 0 closure test, multiple replicas are generated with the same pseudodata and all fit with the same model, the only difference being a distinct initialisation. In the ideal case, all of the replicas would find the same minimum and the resulting PDF set would have no uncertainty. However, this is not what it is observed by

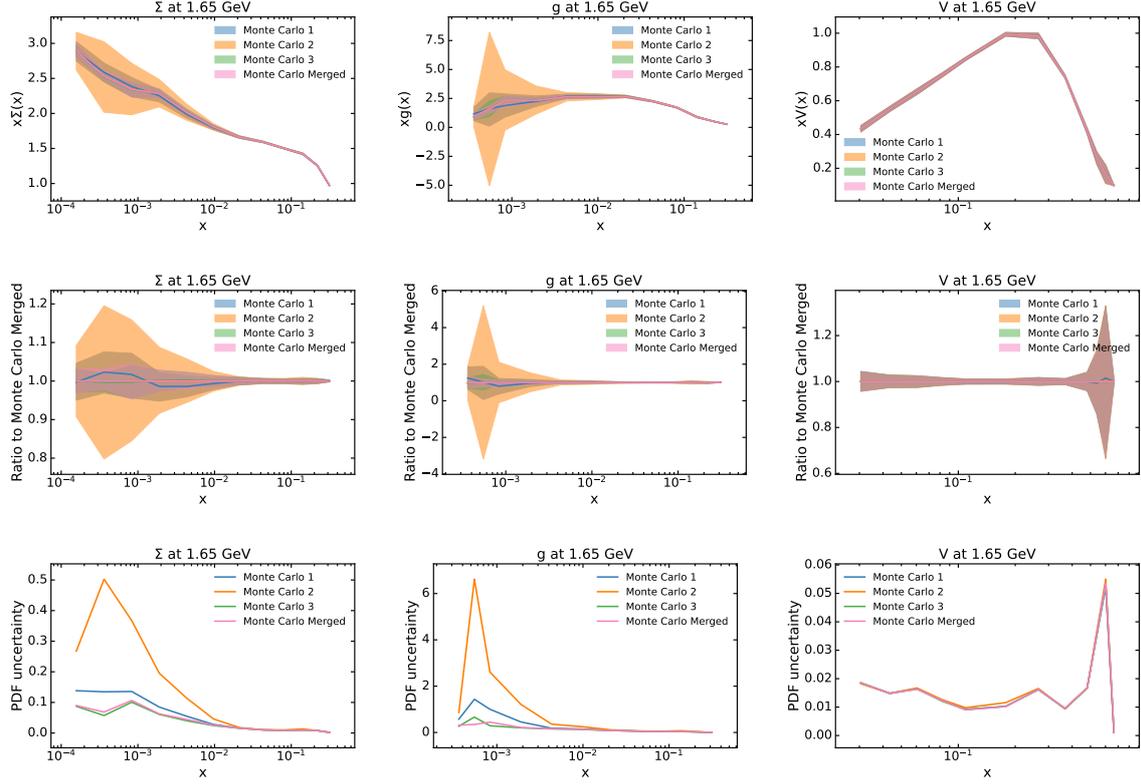


Figure B.1. Plots showing the “merging” procedure used to obtain the final result of the hadronic fit in Sect. 3.2.3. The figures show that fits ran with different initialisations found different results, discovering local minima. The merging procedure adopted selects the best replica in terms of the training χ^2 from all the fits, making the result more robust and increasing the confidence that the global minimum has been reached.

the collaboration and a methodological uncertainty is found, indicating a dependence on the initialisation of the parameters when performing the fit.

In our studies, we have repeatedly found similar problems and at times, the strong dependence on the initialisation has been surprising. In particular, the presence of degeneracies can be quite disruptive and without further analysis difficult to discover. As a matter of fact, while in a posterior sampling algorithm the flat directions are discovered automatically, in a Monte Carlo replica setting they can disguise as constrained directions simply because of a poor initialisation.

Specifically, in order to obtain the final result presented in Sec. 3.2.3, we ran multiple Monte Carlo replica fits, changing only the initialisation of the model, and we ‘merged’ them in order to obtain the fitted replicas with the lowest training χ^2 . By doing so, the result of the fit becomes more and more robust, since we are able to reject replicas that are stuck in local minima, even if they apparently had a good χ^2 . In Fig. B.1, we show an example of this from the hadronic PDF fit where we merged 7 different fits (3 of them shown in the plots). It can clearly be seen that the final merged result is considerably different from some of the preliminary fits that had been obtained.

These are common problems in machine learning and plenty of techniques have been designed to confront them. However, it must be noted that the objective is to estimate with high precision physical quantities and not get a ball-park estimate. Making sure that these are reliable is of paramount importance, especially for the years ahead. The transparency and ease of use of the posterior sampling methods go in this direction and allow for a much more controlled fit environment.

C Impact of training-validation splits

In one of the footnotes in Sect. 2.1, we briefly mentioned that the Monte Carlo replica method is often applied using a ‘training-validation split’, together with a cross-validation stopping, when the best-fit parameters are found for a given piece of pseudodata. In more detail, given a piece of pseudodata \mathbf{d}_p generated from the simulated distribution $\mathcal{N}(\mathbf{d}_0, \Sigma)$, we additionally generate a random subset $S \subseteq \{1, \dots, N_{\text{dat}}\}$ of ‘training indices’ as some fraction f_{train} of the full set of data indices. The best-fit values of the parameters are then given by some form of gradient descent, starting at some random initial value of the parameters (the choice of the initialisation is a point discussed in App. B), such that the χ^2 -statistic to the pseudodata *evaluated only on the training subset* $\{(d_p)_i : i \in S\}$ is minimised, while the χ^2 -statistic to the pseudodata *evaluated only on the validation subset* $\{(d_p)_i : i \notin S\}$ is also still decreasing. At the point where the χ^2 -statistic to the pseudodata evaluated on the validation subset starts increasing, the minimisation is halted.

This point is discussed briefly in the SMEFIT code paper, Ref. [1]; in particular, they state that a training-validation split with a cross-validation stopping is not required for the Monte Carlo replica method in the context of the SMEFT fits they conduct. A direct quote from their paper is the following: ‘*We note that as opposed to the PDF fit case no cross-validation is required here, since overlearning is not possible for a discrete parameter space, where the best-fit value coincides with the absolute maximum of the likelihood.*’ In our paper we work entirely with discrete PDFs evaluated on a grid of 36 points, so we consider the SMEFIT reasoning adequate to justify the lack of the inclusion of a training-validation split in our study.

On the other hand, for completeness, we have also computed all of the Monte Carlo PDF results using training-validation splits (different replica by replica), with a training fraction $f_{\text{train}} = 0.75$, and cross-validation stopping. In Fig. C.1, we display a comparison of the result of the DIS-only Monte Carlo PDF fit using the parametrisation and data discussed in Sect. 3.2.2 with the same Monte Carlo fit conducted with a training-validation split. We observe that the uncertainties are systematically inflated as a result of including the training-validation split; for the linear problem, this is well understood, because each replica effectively sees less data, so we expect larger uncertainties as a result. In particular, we have learned that the use of a training-validation split for the linear problem artificially inflates uncertainties, as compared with the analytic solution we presented in Sect. 3.2.2, with which Monte Carlo without a training-validation split agrees perfectly.

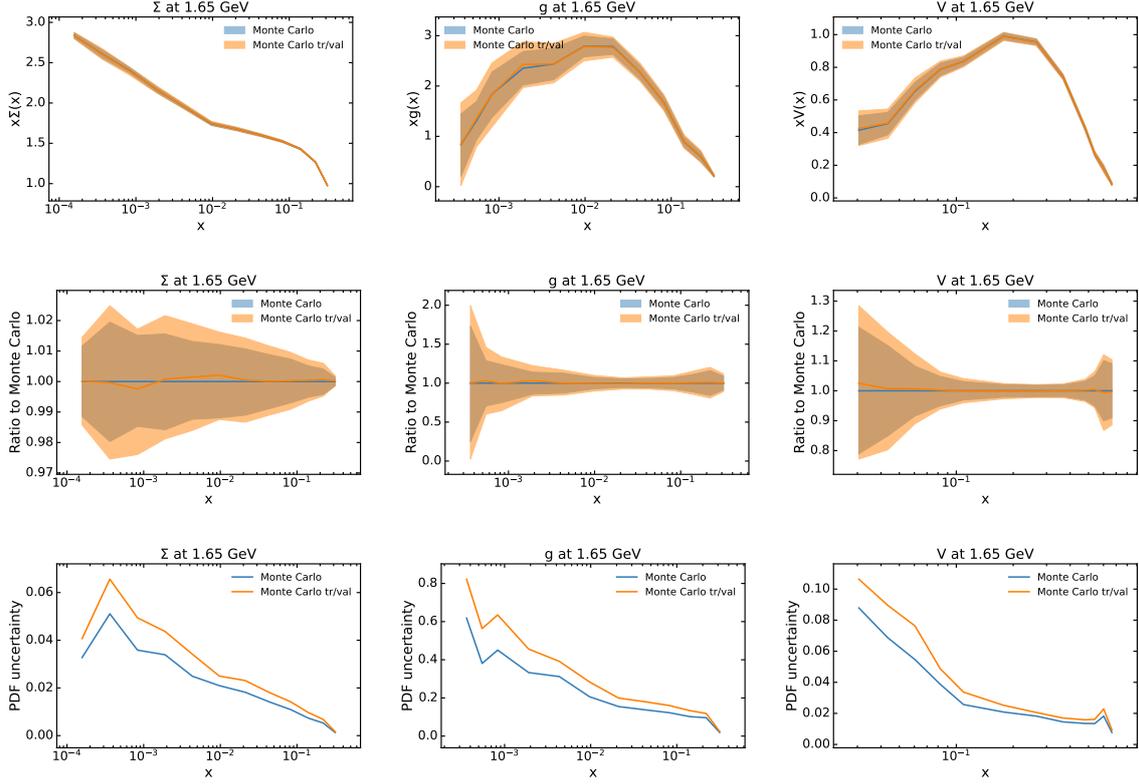


Figure C.1. A comparison of Monte Carlo PDF fits to DIS-only data using the complete dataset when minimising to the pseudodata, versus using a random training-validation split for each pseudodata replica, and applying cross-validation. As we saw in Sect. 3.2.2, the Monte Carlo result using the complete dataset agrees perfectly with both a numerical and analytic Bayesian approach, as expected. On the other hand, when using random training-validation splits and applying cross-validation, the uncertainties in the Monte Carlo approach are systematically *overestimated*.

On the other hand, in Fig. C.2, we display a comparison of the result of the hadronic-only Monte Carlo PDF fit using the parametrisation and data discussed in Sect. 3.2.3 with the same Monte Carlo fit conducted with a training-validation split. This time, we have no theoretical control of the Monte Carlo replica method in either case, as discussed in the main text. Somewhat surprisingly, our intuition fails us, and we instead see that the uncertainties are not systematically inflated, but seem to be ‘reshuffled’ between the flavours instead. Further, the central value is also distorted by the use of a training-validation split. This emphasises the point that we have little control over the Monte Carlo replica method in the non-linear case, and it is difficult to predict the result of a fit prior to performing it.

References

- [1] T. Giani, G. Magni, and J. Rojo, *SMEFiT: a flexible toolbox for global interpretations of particle physics data with effective field theories*, [arXiv:2302.06660](https://arxiv.org/abs/2302.06660).

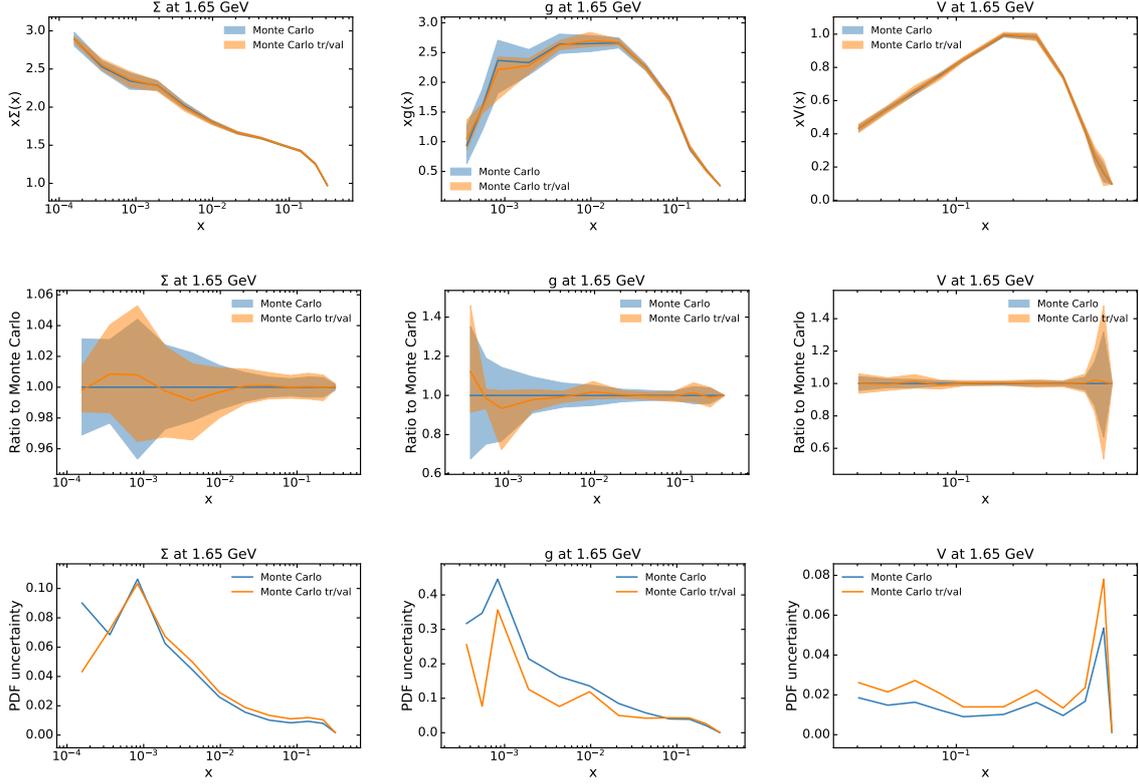


Figure C.2. The same as Fig. C.1, but for fits to hadronic-only data. In this case, the uncertainties seem to be reshuffled amongst the flavours, and the central value appears to be distorted by the use of training-validation splits and cross-validation stopping.

- [2] **SMEFiT** Collaboration, J. J. Ethier, G. Magni, F. Maltoni, L. Mantani, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang, *Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC*, JHEP **11** (2021) 089, [[arXiv:2105.00006](#)].
- [3] J. J. Ethier, R. Gomez-Ambrosio, G. Magni, and J. Rojo, *SMEFT analysis of vector boson scattering and diboson data from the LHC Run II*, Eur. Phys. J. C **81** (2021), no. 6 560, [[arXiv:2101.03180](#)].
- [4] N. P. Hartland, F. Maltoni, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang, *A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector*, JHEP **04** (2019) 100, [[arXiv:1901.05965](#)].
- [5] A. Biekötter, T. Corbett, and T. Plehn, *The Gauge-Higgs Legacy of the LHC Run II*, SciPost Phys. **6** (2019), no. 6 064, [[arXiv:1812.07587](#)].
- [6] **NNPDF** Collaboration, R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali, *A Determination of parton distributions with faithful uncertainty estimation*, Nucl. Phys. B **809** (2009) 1–63, [[arXiv:0808.1231](#)]. [Erratum: Nucl.Phys.B 816, 293 (2009)].
- [7] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *A first unbiased global NLO determination of parton distributions and their uncertainties*, Nucl. Phys. B **838** (2010) 136–206, [[arXiv:1002.4407](#)].

- [8] **NNPDF** Collaboration, R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO*, Nucl. Phys. B **855** (2012) 153–221, [[arXiv:1107.2652](#)].
- [9] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244–289, [[arXiv:1207.1303](#)].
- [10] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, [[arXiv:1410.8849](#)].
- [11] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, Eur. Phys. J. C **77** (2017), no. 10 663, [[arXiv:1706.00428](#)].
- [12] **NNPDF** Collaboration, R. D. Ball et al., *The path to proton structure at 1% accuracy*, Eur. Phys. J. C **82** (2022), no. 5 428, [[arXiv:2109.02653](#)].
- [13] **JAM** Collaboration, C. Cocuzza, A. Metz, and N. Sato, *Simultaneous Extraction of Unpolarized PDFs and Nuclear Effects*, SciPost Phys. Proc. **8** (2022) 098.
- [14] **Jefferson Lab Angular Momentum (JAM)** Collaboration, N. T. Hunt-Smith, W. Melnitchouk, N. Sato, A. W. Thomas, X. G. Wang, and M. J. White, *Global QCD analysis and dark photons*, JHEP **09** (2023) 096, [[arXiv:2302.11126](#)].
- [15] N. T. Hunt-Smith, C. Cocuzza, W. Melnitchouk, N. Sato, A. W. Thomas, and M. J. White, *On the resolution of the sign of gluon polarization in the proton*, [arXiv:2403.08117](#).
- [16] **NNPDF** Collaboration, R. D. Ball, J. Cruz-Martinez, L. Del Debbio, S. Forte, Z. Kassabov, E. R. Nocera, J. Rojo, R. Stegeman, and M. Ubiali, *Response to "Parton distributions need representative sampling"*, [arXiv:2211.12961](#).
- [17] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo, and M. Ubiali, *Can New Physics hide inside the proton?*, Phys. Rev. Lett. **123** (2019), no. 13 132001, [[arXiv:1905.05215](#)].
- [18] A. Greljo, S. Iranipour, Z. Kassabov, M. Madigan, J. Moore, J. Rojo, M. Ubiali, and C. Voisey, *Parton distributions in the SMEFT from high-energy Drell-Yan tails*, JHEP **07** (2021) 122, [[arXiv:2104.02723](#)].
- [19] S. Iranipour and M. Ubiali, *A new generation of simultaneous fits to LHC data using deep learning*, JHEP **05** (2022) 032, [[arXiv:2201.07240](#)].
- [20] Z. Kassabov, M. Madigan, L. Mantani, J. Moore, M. Morales Alvarado, J. Rojo, and M. Ubiali, *The top quark legacy of the LHC Run II for PDF and SMEFT analyses*, JHEP **05** (2023) 205, [[arXiv:2303.06159](#)].
- [21] E. Hammou, Z. Kassabov, M. Madigan, M. L. Mangano, L. Mantani, J. Moore, M. M. Alvarado, and M. Ubiali, *Hide and seek: how PDFs can conceal new physics*, JHEP **11** (2023) 090, [[arXiv:2307.10370](#)].
- [22] M. N. Costantini, E. Hammou, Z. Kassabov, M. Madigan, L. Mantani, M. Morales Alvarado, J. M. Moore, and M. Ubiali, *SIMUnet: an open-source tool for simultaneous global fits of EFT Wilson coefficients and PDFs*, [arXiv:2402.03308](#).
- [23] L. Del Debbio, T. Giani, and M. Wilson, *Bayesian approach to inverse problems: an application to NNPDF closure testing*, Eur. Phys. J. C **82** (2022), no. 4 330, [[arXiv:2111.05787](#)].
- [24] P. Hall, *The bootstrap and edgeworth expansion*, 1992.

- [25] J. Shao and D. Tu, *The jackknife and bootstrap*, 1996.
- [26] N. T. Hunt-Smith, A. Accardi, W. Melnitchouk, N. Sato, A. W. Thomas, and M. J. White, *Determination of uncertainties in parton densities*, Phys. Rev. D **106** (2022), no. 3 036003, [[arXiv:2206.10782](#)].
- [27] R. D. Ball et al., *Benchmarking of parton distributions and their uncertainties*, in HERA and the LHC: 4th Workshop on the Implications of HERA for LHC Physics, pp. 53–73, 3, 2009.
- [28] F. Capel, R. Aggarwal, M. Botje, A. Caldwell, O. Schulz, and A. Verbytskyi, *PartonDensity.jl: a novel parton density determination code*, [arXiv:2401.17729](#).
- [29] A. Candido, L. Del Debbio, T. Giani, and G. Petrillo, *Bayesian Inference with Gaussian Processes for the Determination of Parton Distribution Functions*, [arXiv:2404.07573](#).
- [30] A. W. v. d. Vaart, *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.
- [31] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, *Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory*, JHEP **04** (2021) 279, [[arXiv:2012.02779](#)].
- [32] N. Castro, J. Erdmann, C. Grunwald, K. Kröninger, and N.-A. Rosien, *EFTfitter—A tool for interpreting measurements in the context of effective field theories*, Eur. Phys. J. C **76** (2016), no. 8 432, [[arXiv:1605.05585](#)].
- [33] J. De Blas et al., *HEPfit: a code for the combination of indirect and direct constraints on high energy physics models*, Eur. Phys. J. C **80** (2020), no. 5 456, [[arXiv:1910.14012](#)].
- [34] **NNPDF** Collaboration, R. D. Ball et al., *An open-source machine learning framework for global analyses of parton distributions*, Eur. Phys. J. C **81** (2021), no. 10 958, [[arXiv:2109.02671](#)].
- [35] R. M. Neal, *Probabilistic inference using markov chain monte carlo methods*, 2011.
- [36] C. J. Geyer, *Practical markov chain monte carlo*, Statistical Science **7** (1992), no. 4 473–483.
- [37] G. Ashton et al., *Nested sampling for physical scientists*, Nature **2** (2022) [[arXiv:2205.15570](#)].
- [38] J. Skilling, *Nested Sampling*, AIP Conference Proceedings **735** (11, 2004) 395–405, [https://pubs.aip.org/aip/acp/article-pdf/735/1/395/11702789/395_1_online.pdf].
- [39] J. Skilling, *Nested sampling for general Bayesian computation*, Bayesian Analysis **1** (2006), no. 4 833 – 859.
- [40] F. Feroz, M. P. Hobson, and M. Bridges, *MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics*, Mon. Not. Roy. Astron. Soc. **398** (2009) 1601–1614, [[arXiv:0809.3437](#)].
- [41] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, Phys. Rept. **793** (2019) 1–98, [[arXiv:1706.08945](#)].
- [42] Z. Han and W. Skiba, *Effective theory analysis of precision electroweak data*, Phys. Rev. D **71** (2005) 075009, [[hep-ph/0412166](#)].
- [43] E. da Silva Almeida, A. Alves, N. Rosa Agostinho, O. J. P. Éboli, and M. C. Gonzalez-Garcia, *Electroweak Sector Under Scrutiny: A Combined Analysis of LHC and Electroweak Precision Data*, Phys. Rev. D **99** (2019), no. 3 033001, [[arXiv:1812.01009](#)].

- [44] J. Ellis, V. Sanz, and T. You, *The Effective Standard Model after LHC Run I*, JHEP **03** (2015) 157, [[arXiv:1410.7703](#)].
- [45] E. d. S. Almeida, A. Alves, O. J. P. Éboli, and M. C. Gonzalez-Garcia, *Electroweak legacy of the LHC run II*, Phys. Rev. D **105** (2022), no. 1 013006, [[arXiv:2108.04828](#)].
- [46] A. Biekötter, T. Corbett, and T. Plehn, *The Gauge-Higgs Legacy of the LHC Run II*, SciPost Phys. **6** (2019), no. 6 064, [[arXiv:1812.07587](#)].
- [47] S. Kraml, T. Q. Loc, D. T. Nhung, and L. D. Ninh, *Constraining new physics from Higgs measurements with Lilith: update to LHC Run 2 results*, SciPost Phys. **7** (2019), no. 4 052, [[arXiv:1908.03952](#)].
- [48] J. Ellis, C. W. Murphy, V. Sanz, and T. You, *Updated Global SMEFT Fit to Higgs, Diboson and Electroweak Data*, JHEP **06** (2018) 146, [[arXiv:1803.03252](#)].
- [49] T. Corbett, O. J. P. Eboli, J. Gonzalez-Fraile, and M. C. Gonzalez-Garcia, *Robust Determination of the Higgs Couplings: Power to the Data*, Phys. Rev. D **87** (2013) 015022, [[arXiv:1211.4580](#)].
- [50] A. Buckley, C. Englert, J. Ferrando, D. J. Miller, L. Moore, M. Russell, and C. D. White, *Constraining top quark effective theory in the LHC Run II era*, JHEP **04** (2016) 015, [[arXiv:1512.03360](#)].
- [51] I. Brivio, S. Bruggisser, F. Maltoni, R. Moutafis, T. Plehn, E. Vryonidou, S. Westhoff, and C. Zhang, *O new physics, where art thou? A global search in the top sector*, JHEP **02** (2020) 131, [[arXiv:1910.03606](#)].
- [52] S. Bißmann, J. Erdmann, C. Grunwald, G. Hiller, and K. Kröninger, *Constraining top-quark couplings combining top-quark and B decay observables*, Eur. Phys. J. C **80** (2020), no. 2 136, [[arXiv:1909.13632](#)].
- [53] N. Elmer, M. Madigan, T. Plehn, and N. Schmal, *Staying on Top of SMEFT-Likelihood Analyses*, [arXiv:2312.12502](#).
- [54] L. Allwicher, D. A. Faroughy, F. Jaffredo, O. Sumensari, and F. Wilsch, *Drell-Yan tails beyond the Standard Model*, JHEP **03** (2023) 064, [[arXiv:2207.10714](#)].
- [55] R. Boughezal, Y. Huang, and F. Petriello, *Exploring the SMEFT at dimension eight with Drell-Yan transverse momentum measurements*, Phys. Rev. D **106** (2022), no. 3 036020, [[arXiv:2207.01703](#)].
- [56] R. Bartocci, A. Biekötter, and T. Hurth, *A global analysis of the SMEFT under the minimal MFV assumption*, [arXiv:2311.04963](#).
- [57] I. Brivio, S. Bruggisser, N. Elmer, E. Geoffray, M. Luchmann, and T. Plehn, *To Profile or To Marginalize – A SMEFT Case Study*, SciPost Phys. **16** (2024) 035, [[arXiv:2208.08454](#)].
- [58] J. Buchner, *A statistical test for nested sampling algorithms*, Statistics and Computing **26** (Sept., 2014) 383–392.
- [59] J. Buchner, *Collaborative nested sampling: Big data versus complex physical models*, Publications of the Astronomical Society of the Pacific **131** (Aug., 2019) 108005.
- [60] J. Buchner, *UltraneST – a robust, general purpose bayesian inference engine*, 2021.
- [61] **CTEQ-TEA** Collaboration, I. Sitiwaldi, K. Xie, A. Ablat, S. Dulat, T.-J. Hou, and C. . P. Yuan, *Precision studies of the post-CT18 LHC Drell-Yan data in the CTEQ-TEA global analysis*, Phys. Rev. D **108** (2023), no. 3 034030, [[arXiv:2305.10733](#)].

- [62] **CTEQ-TEA** Collaboration, K. Xie, T. J. Hobbs, T.-J. Hou, C. Schmidt, M. Yan, and C. P. Yuan, *Photon PDF within the CT18 global analysis*, Phys. Rev. D **105** (2022), no. 5 054006, [[arXiv:2106.10299](#)].
- [63] R. Thorne, S. Bailey, T. Cridge, L. Harland-Lang, and A. D. Martin, *The MSHT20 Parton Distribution Functions*, SciPost Phys. Proc. **8** (2022) 018.
- [64] T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, *QED parton distribution functions in the MSHT20 fit*, Eur. Phys. J. C **82** (2022), no. 1 90, [[arXiv:2111.05357](#)].
- [65] **NNPDF** Collaboration, R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties*, JHEP **05** (2010) 075, [[arXiv:0912.2276](#)].
- [66] **New Muon** Collaboration, M. Arneodo et al., *Accurate measurement of $F_2(d) / F_2(p)$ and $R^{*d} - R^{*p}$* , Nucl. Phys. B **487** (1997) 3–26, [[hep-ex/9611022](#)].