Published for SISSA by Springer

Received: *March 27, 2025*
Revised: *June 20, 2025*
Accepted: *July 15, 2025*
Published: *August 13, 2025*

Beyond scale variations: perturbative theory uncertainties from nuisance parameters

Frank J. Tackmann

Deutsches Elektronen-Synchrotron DESY,
Notkestr. 85, 22607 Hamburg, Germany

E-mail: frank.tackmann@desy.de

Abstract: We develop a new approach to estimate the uncertainty due to missing higher orders in perturbative predictions (the perturbative "theory uncertainty"), which overcomes many inherent limitations of the currently prevalent methods based on varying unphysical renormalization scales. In our approach, the true underlying sources of the theory uncertainty, namely the missing higher-order terms, are identified and parameterized in terms of mutually independent theory nuisance parameters (TNPs). The TNPs are true parameters of the calculation, i.e., they have a well-defined true value that is not or only imprecisely known. This approach affords the theory uncertainty all benefits of a truly parametric uncertainty: it provides correct correlations and allows for consistent error propagation and combination. Furthermore, the TNPs can be profiled in fits, allowing the data to reduce the theory uncertainties. On the theory side, it allows maximally exploiting all available higher-order information to reduce the theory uncertainty, such as partial higher-order results or any nontrivial knowledge of the higher-order or all-order structure.

We first discuss the method in general as it can be applied across the board of perturbative calculations, including the various choices it requires and corresponding strategies for making them. As a concrete application, we then discuss the resummed transverse momentum (q_T) spectrum in Drell-Yan production, and how TNP-based uncertainties can correctly capture the correlations across the q_T spectrum and between Z and W production. This application is the basis of the theory model enabling the recent precise measurement of the W-boson mass by the CMS experiment. In a forthcoming paper, we use it to study the theory uncertainties in extracting the strong coupling constant α_s from the Z q_T spectrum.

Keywords: Electroweak Precision Physics, Higher-Order Perturbative Calculations, Resummation, Specific QCD Phenomenology

ArXiv ePrint: 2411.18606

https://doi.org/10.1007/JHEP08(2025)098

JHEP08(2025)098

Contents

1 Introduction

The interpretation of precision measurements requires equally precise theoretical predictions. Just as for experimental measurements, theoretical predictions are ultimately only as useful

JHEP08(2025)098

as their uncertainties are meaningful. We are specifically interested in theory predictions based on perturbation theory and their uncertainty due to missing higher-order corrections, which we will refer to as the perturbative "theory uncertainty". For a theory uncertainty to be meaningful it must realistically reflect our degree of knowledge. This not only means that it has a realistic size but also that it provides correct correlations and allows for some form of statistical interpretation.

The prevalent traditional approach for estimating perturbative theory uncertainties based on scale variations is neither particularly reliable nor does it provide correlations let alone a meaningful statistical interpretation. These limitations are in principle well known. They have been a long-standing bottleneck in our ability to interpret experimental measurements using theoretical predictions, which is only becoming more severe as experimental measurements become ever more precise. The approach put forward in this paper allows us to address this issue by equipping perturbative predictions with meaningful theory uncertainties.

When designing measurement and interpretation strategies we optimize the total uncertainty budget, and the theory uncertainty is part of this budget. Currently, this optimization often gets skewed toward reducing as much as possible the impact of unreliable theory uncertainties. This inevitably leads to sacrificing experimental precision. Reliable, meaningful theory uncertainties make such sacrifices unnecessary and thus allow improving the overall uncertainty budget beyond just the immediate effect of improving the theory prediction itself. They can also enable entirely new measurement strategies that would otherwise be unfeasible. An example is the recent precision measurement of the W-boson mass by CMS [1]. Thus, meaningful theory uncertainties greatly facilitate our ability to fully exploit the potential of existing and future precision measurements.

The above requirements for a meaningful theory uncertainty are elaborated on in section 2.1. The key points are: first, the theory uncertainty is a property of the current prediction that should reflect its intrinsic precision. In particular, it is *not* meant or defined to be the unknown difference to the all-order result (or some formally more accurate result standing in for the all-order one). Second, "theory correlations" — the correlations in the theory uncertainties of different predictions — are required as soon as more than a single theory prediction is used at a time. An important example is considering a differential spectrum, as each of its bins has a priori its own theory prediction. The bin-by-bin correlations are essential when one is interested in shape effects, since a shape uncertainty is basically a statement about how the uncertainties at different points in the spectrum are correlated. Theory correlations are thus critical if one wants to distinguish the shape effect induced by a to-be-determined parameter of interest from that caused by theory uncertainties. Third, a theory prediction simply cannot be used for interpreting experimental measurements without any quantitative statistical meaning for its uncertainty.

The limitations of scale variations are discussed in more detail in section 2.3. In short, their lack of correlations basically stems from the fact that their variation cannot be interpreted like that of a normal parameter whose uncertainty is being propagated. Hence, they are notoriously unreliable for estimating shape uncertainties, which unfortunately is exactly what they are often used for (primarily due to the lack of alternatives). This is becoming a severe limitation in many precision studies. Presently, to be on the safe side we like to avoid attaching

JHEP08(2025)098

any statistical meaning to theory uncertainties derived from scale variations. However, this is not helpful at all. It just skirts the issue and puts the burden on the users of our predictions since they are now forced to attach some ad hoc quantitative statistical meaning to them. This state of affairs is clearly unsatisfactory and frankly speaking rather embarrassing.

Some frequentist statistical models attached to theory uncertainties are discussed for example in refs. [2, 3]. There have been various proposals to obtain theory uncertainty estimates with a more meaningful statistical interpretation via a Bayesian model [4–7], or series acceleration [8], or based on a set of reference processes [9]. These methods go in the right direction by trying to more directly estimate the size of missing higher-order corrections and by more explicitly exposing the assumptions made. However, like scale variations they base the uncertainty estimate on the known lower-order terms without parameterizing the actual underlying source of uncertainty and as a result share many of the limitations of scale variations. They have a similar level of arbitrariness and reliability, and in particular they also lack theory correlations.

A theory uncertainty is a form of systematic (epistemic) uncertainty and as such we cannot hope to render it as robust as a purely statistical (aleatoric) uncertainty. However, the same requirements to be meaningful are shared by experimental systematic uncertainties. Our approach thus treats theory uncertainties following the same logic that is routinely applied for experimental systematic uncertainties to cast them into parametric uncertainties. This is the key to render them meaningful and is discussed in section 2.2 and section 3. In a nutshell, we identify the underlying sources of uncertainty, namely the relevant missing perturbative ingredients, and parameterize them in terms of unknown parameters, which are the "theory nuisance parameters" (TNPs). Predictions for different cross sections that depend on the same perturbative ingredient will share a common TNP and the associated uncertainty will be 100% correlated among them. The TNPs have true values, which can in principle be determined from a higher-order calculation, but which are a priori unknown (or treated as such). Without explicit knowledge of their true value, we can use auxiliary information at our disposal to constrain their allowed ranges. The TNPs are then explicitly varied or floated in fits within their allowed ranges to account for the theory uncertainties and propagate them with correct correlations to subsequent interpretations.

Whilst constraining the TNPs based on auxiliary theoretical information necessarily involves making some educated choices, this can be thought of as an imagined auxiliary measurement. Furthermore, depending on the context, such theoretical constraints can be supplemented or even replaced by constraining the TNPs with real auxiliary measurements or in situ in the interpretation of the nominal measurement itself. As a result, the TNP-based theory uncertainties admit an analogous statistical treatment and interpretation as experimental systematics based on nuisance parameters constrained by (real or imagined) auxiliary control measurements (see e.g. refs. [10, 11]). Finally, even if individual TNPs may not necessarily have a precisely known probability distribution, since the total theory uncertainty will typically arise as the combination of a number of TNPs, the central-limit theorem ensures that it will be asymptotically Gaussian distributed.

Another key advantage of our approach is that it overcomes the paradigm of only being able to systematically improve theory predictions in large discrete steps based on completely

JHEP08(2025)098

known formal orders. The desire to utilize available higher-order information for actual phenomenological benefit, i.e. to reduce theory uncertainties, without having to wait until the formally complete next order eventually becomes available is more than obvious. In fact, likely sooner than later this is going to become an actual requirement for making progress, because as we push to higher and higher orders, formally complete orders for final predictions are increasingly difficult to achieve and might eventually become out of reach. For this reason, more and more predictions are appearing that are formally "approximate" in some way ranging from very unjustified to very well justified. The underlying issue is of course that at present we lack meaningful theory uncertainties, so the primary guiding principle are formally complete orders.

We believe that in the long run an essential benefit of our approach will be to allow our community to break out of this rigid paradigm. Meaningful theory uncertainties are by construction a much better judge of the actual precision than the formal accuracy. Our approach naturally allows for predictions that are formally incomplete in a fully justified, systematic, and formally consistent manner. It ultimately allows for an (almost) continuous integration of newly available higher-order results into final theory predictions, taking full and immediate advantage of them for reducing theory uncertainties and thereby for maximal and immediate phenomenological impact. Moreover, our approach makes it very transparent which missing perturbative ingredients are causing the largest uncertainties at any given stage, allowing one to anticipate already beforehand the impact of explicitly calculating a certain higher-order ingredient. This can greatly help to guide efforts and to provide clear and tangible justification for allocating resources.

The approach of this paper was first advocated in ref. [12], and has already been used since in several instances [13–15]. In these cases, the TNPs serve to estimate the uncertainty due to still missing ingredients at the nominal, approximate working order. The application of our approach to the resummed transverse momentum (q_T) spectrum of W and Z bosons produced in hadronic collisions as discussed in section 6 forms the basis of the theoretical modelling that has enabled a high-precision measurement of the W-boson mass by the CMS experiment [1]. In a forthcoming paper [16], we use it to study the expected theory uncertainties in the extraction of the strong coupling constant α_s from the Z p_T spectrum. A promising first application of our approach to a variety of fixed-order single-differential distributions has been carried out in ref. [17].

At a basic level, it is of course not a new idea to estimate a missing higher-order coefficient and the uncertainty caused by it. For example, in the past resummed calculations at N^{3}LL and beyond (see e.g. refs. [18–23]) have used Padé approximations for varying the 4-loop cusp anomalous dimension and other 3-loop ingredients missing at the time. In high-precision QED and electroweak calculations, scale variations are less prevalent than for QCD calculations, and theory uncertainties are more commonly estimated by explicit, more-or-less ad hoc estimates of the expected size of missing higher-order terms (see e.g. ref. [24]) including attempts to constrain them from measurements (see e.g. ref. [25]). The methods of refs. [4–8] amount to modelling the size of missing terms based on the size of the known terms.

While the main focus of our discussion is on perturbative predictions in QCD, our approach in principle applies to any other systematic, truncated expansion and its truncation

JHEP08(2025)098

uncertainty, such as the power expansions performed in effective field theories. For example, a similar strategy can be followed to account for the truncation uncertainty in the SMEFT, see e.g. refs. [26–28].

This paper is organized as follows. As already mentioned, in section 2 we discuss general aspects of perturbative theory uncertainties. Section 3 gives a general discussion of the approach of theory nuisance parameters and is intended for all audiences. Section 3.1 gives a general overview of the approach, while the remaining subsections discuss several specific aspects. Readers interested in an executive 5-page summary of our approach can just read sections 2.2 and 3.1. Section 4 provides a guide for how to derive suitable parameterizations in terms of TNPs. It is intended for readers who wish to implement TNP-based uncertainties into their predictions, providing principles and strategies to follow as well as several examples for illustration. In section 5 we then focus on TNPs for scalar series in QCD and discuss how we can obtain robust theory constraints on them based on our theoretical knowledge and available information from existing higher-order calculations. In section 6, we present an explicit full-fledged example application of our approach for the case of q_T resummation for $pp \to Z/\gamma^*$ and $pp \to W$ production. We conclude in section 7.

2 Perturbative theory uncertainties

In section 2.1 we elaborate on the criteria for meaningful theory uncertainties. Readers who find them self-evident or are happy to accept them can skip this subsection. In section 2.2 we derive our basic approach to estimate theory uncertainties as parametric uncertainties. In section 2.3 we discuss the limitations of scale variations and why uncertainties derived from them cannot be regarded as parametric uncertainties.

2.1 Philosophy

As mentioned in the introduction, for a theory (or really any) uncertainty to be meaningful, it must

1. have a size that reflects our level of knowledge,
2. provide correct correlations, and
3. allow for some form of statistical interpretation.

Before elaborating further on these criteria, we stress that despite the title of this subsection, having meaningful theory uncertainties is not just a philosophical or academic issue — quite the opposite. As discussed in the introduction, it has important implications for interpreting experimental measurements.

2.1.1 Size and statistical interpretation

A theory uncertainty is a systematic uncertainty, and as such will always require some element of human judgement. Nevertheless, like for any systematic uncertainty, its size must reflect our level of knowledge or lack thereof. With faithfully estimated theory uncertainties, the precision of a perturbative prediction should be judged primarily by its uncertainty and not

JHEP08(2025)098

so much by its formal perturbative accuracy. Of course, for a given quantity, we expect a formally higher-order prediction to be more precise than a formally lower-order one. The key point is that this should be the *outcome* of the uncertainty estimation procedure rather than an *input* to it. This essentially precludes estimating the theory uncertainty (solely) based on the size of the last known perturbative correction.

To see this, consider the experimental analog of two measurements A and B of the same quantity, where B is more precise than A due to increased statistics or improved systematics or both. These "formal" improvements may make us more confident in measurement B, but in the end what really counts is their respective uncertainty. Assuming both have faithfully estimated uncertainties, we expect B's uncertainty to be smaller than A's. For simplicity, imagine that B's uncertainty is so much smaller than A's (and uncorrelated) that only A's uncertainty matters for comparing A and B. Consider the case that A does not agree with B: since B is deemed to be more reliable (formally "better"), we would conclude that A's uncertainty was underestimated, i.e., in this case we can invalidate A's uncertainty. Crucially, the reverse conclusion is not allowed: if A does agree with B within its uncertainty, this *does not* validate A's uncertainty, i.e., we *cannot* conclude that A's uncertainty is not underestimated. If that was allowed, then taken to its logical conclusion, if A's central value would agree perfectly with B, we would have to conclude that A has a vanishingly small uncertainty, which is clearly nonsense.

The above discussion applies identically when A is a lower-order and B a higher-order calculation of the same quantity. For A to agree with B within its uncertainty is only a necessary but not sufficient condition for A's uncertainty to be correctly estimated. In particular, we *cannot* estimate the uncertainty of A by comparing its central value to B. In other words, the difference in central values, i.e. the true size of the higher-order correction, is at best a (rough) *lower limit* on A's uncertainty.

Unfortunately, this is exactly what happens frequently in perturbative predictions: we are mistakenly led to think of the theory uncertainty as the difference of our result to the all-order result (or a formally more accurate higher-order one). This inevitably leads to the conclusion that we fundamentally cannot know the theory uncertainty because we will never know the true all-order result. Or perhaps less dramatically, that we will only really know the uncertainty at the present perturbative order once we have calculated the next order(s). The experimental analog would be to say that we can never know the uncertainty of a measurement because we will never know the true value in nature.

To summarize the above discussion: the theory uncertainty is not defined as the difference to the all-order (or the next-order) result. Instead, it must be a property of our present result reflecting its intrinsic precision. When estimating it, we are meant to estimate a possible range that contains the all-order result. Of course, we cannot estimate this range with absolute certainty. We can only hope to be able to estimate a range that contains the all-order result with some probability or some level of confidence. This leads us to the third criterium above, which basically means that we must have some way to quantify this probability or level of confidence. Otherwise, we cannot actually utilize the prediction for an interpretation, because to do so one must be able to interpret its uncertainty in terms of some quantitative statistical meaning.

JHEP08(2025)098

ρ	99.5%	98%	95.5%	87.5%
$\delta_{f/g}/\delta$	0.1	0.2	0.3	0.5

Table 1. Reduction of the relative uncertainty in the ratio f/g for different correlations ρ, see text for details.

2.1.2 Correlations

In practice theory predictions are almost never utilized in isolation but almost always in combination with one another, at which point the correlation in their uncertainties becomes relevant. This is the case whenever one considers more than a single process or phase-space region. Consider the following prototype of a data-driven method,

$$f = [g]_{\text{measured}} \times \left[\frac{f}{g}\right]_{\text{theory}}, \tag{2.1}$$

where a desired quantity f (the "signal" region/process) is obtained from a precisely measured quantity g (the "control" region/process) by multiplying it with their ratio predicted from theory. Loosely speaking, if f and g are different but closely related, their perturbative corrections should be very similar and largely cancel in the ratio, such that eq. (2.1) yields an improved description of f compared to its direct prediction from theory alone. More precisely, the theory uncertainties of f and g should be strongly correlated in order to cancel in the ratio. This cancellation is often implicitly assumed or relied on, but in reality it is very sensitive to the exact correlation.

To appreciate this, consider f and g having relative uncertainty δ with correlation ρ. The relative uncertainty of their ratio, $\delta_{f/g}$, as a function of ρ is given by

$$\delta_{f/g} = \delta\sqrt{2(1-\rho)}\,. \tag{2.2}$$

We are interested in the limit of strong correlation and large cancellation, i.e., ρ close to 1. In this limit, $\delta_{f/g}$ is very sensitive to the precise value of ρ, as illustrated in tables 1, because the square root becomes infinitely steep for $\rho \to 1$. For example, $\delta_{f/g}$ is 10 times smaller than δ for $\rho = 99.5\%$, while already for $\rho = 98\%$ it doubles in size to only 5 times smaller than δ. The same correlation information is required whenever one performs a simultaneous interpretation of two or more measurements. A prime example is the interpretation of a differential spectrum, which requires bin-by-bin (or point-by-point) theory correlations, as already discussed in the introduction. The specific example of the transverse-momentum spectrum of W and Z bosons at the LHC is discussed in section 6. The importance of theory correlations in the context of modern machine learning methods was also stressed e.g. in ref. [29].

It is important to keep in mind that different quantities we want to predict, such as cross sections for different processes or at different points in phase space, do not by themselves have a notion of being correlated to each other. A priori, they are only more or less related to each other by the extent to which their theory descriptions depend on common ingredients. What is correlated is the uncertainty in their prediction due to the limited knowledge of those common ingredients. If two quantities share a common source of uncertainty, the impact

JHEP08(2025)098

of that uncertainty on both is 100% correlated between them, and this is fundamentally the only way a correlation can arise.

The simplest example is a common input parameter. Its imprecise knowledge represents a common source of uncertainty and its resulting uncertainty in all quantities that depend on it is 100% correlated. When expressed as a covariance matrix, it yields a 100% correlated covariance matrix for all quantities. A standard way to evaluate the correlated impact on all quantities is to use a common nuisance parameter, which can be explicitly varied or floated in a fit and whose variation is equivalent to varying the input parameter itself.

When several quantities depend on multiple independent sources of uncertainty, the final correlation depends on the relative impact of the various 100% correlated contributions from each source. Expressed with covariance matrices, the total covariance matrix is given by a sum of several 100% correlated ones, which is in general not 100% correlated any longer. Of course, different (nuisance) parameters can themselves have (partially) correlated uncertainties, which can be propagated using standard error propagation.

More generally, the standard procedure to treat experimental systematic uncertainties is to cast them into parametric uncertainties by parameterizing the underlying source or effect in terms of one or more nuisance parameters, which have true but a priori unknown values. Their values are then constrained by auxiliary (real or imagined) control measurements. The resulting best-estimate but imprecise values of the nuisance parameters are then propagated to the nominal measurement and its interpretation. Without any auxiliary constraint on a nuisance parameter, its uncertainty is a priori infinite and its value will only be constrained by the nominal measurement itself, reducing the power of the measurement for constraining the parameters of interest.

To correctly quantify and account for theory correlations we simply follow the same procedure: we identify and parameterize the common and mutually independent sources of theory uncertainty and treat them respectively as 100% correlated and uncorrelated among all quantities of interest. This is exactly what the theory nuisance parameters will do.

2.2 Parametric perturbative uncertainties

Consider the formal perturbative expansion of a quantity f in a small parameter α,

$$f(\alpha) = f_0 + f_1\,\alpha + f_2\,\alpha^2 + f_3\,\alpha^3 + \mathcal{O}(\alpha^4)\,. \tag{2.3}$$

By calculating the values of the first few coefficients of the series, we obtain a prediction for f at leading order (LO), next-to-leading order (NLO), next-to-next-to-leading order (NNLO),

$$\begin{aligned} \text{LO:} \quad & f(\alpha) = \hat{f}_0\,, \\ \text{NLO:} \quad & f(\alpha) = \hat{f}_0 + \hat{f}_1\,\alpha\,, \\ \text{NNLO:} \quad & f(\alpha) = \hat{f}_0 + \hat{f}_1\,\alpha + \hat{f}_2\,\alpha^2\,, \end{aligned} \tag{2.4}$$

and so on. We always denote the true value of a quantity by a hat, so $\hat{f}_n$ are the true values of the series coefficient f_n. When applying perturbation theory in this way, we always work under the general assumption that the series in eq. (2.3) converges.[1]

[1]When α is a coupling constant, it is well know that the series coefficients f_n can grow factorially, $f_n \sim n!$, which for sufficiently high n overcomes the power suppression by α^n, so the series is only asymptotic. In

JHEP08(2025)098

The predictions for $f(\alpha)$ in eq. (2.4) are not exact but approximations of its all-order result. The theory uncertainty we consider here is due to this intrinsic inexactness.[2] Its fundamental sources are the higher-order terms in eq. (2.3) that are missing in eq. (2.4). Our assumption of convergence implies that the predictions get increasingly better, i.e. more precise, by including more and more terms in the series. This is equivalent to saying that the dominant source of theory uncertainty for the prediction at a given order is the next missing term, i.e, that the sum of all missing higher orders is dominated by the first missing one. (One might then consider treating the second missing one as the "error on the error" [3, 30].)

Strictly speaking, the actual source of uncertainty is not so much the missing term as a whole; it is rather the unknown *series coefficient* f_n, as we do know the exact power α^n it comes with. At NNLO, if we knew $\hat{f}_3$, we could add the next term $\hat{f}_3\,\alpha^3$ to increase the precision. Hence, very strictly speaking, what is unknown is not the series coefficient per se but rather its *true value* $\hat{f}_n$. We do know f_n in the sense that we know its exact definition, we know it has a well-defined true value, and we know how to calculate it in principle (even if we may not have the means to compute it in practice). Importantly, these distinctions are not just semantics, but are relevant in what follows.[3]

Let us stress another important logical distinction: the unknown $\hat{f}_n$ is *not* the theory uncertainty itself (as discussed in section 2.1.1); it is only the *source* of the uncertainty. The theory uncertainty is the impact on the prediction of not knowing $\hat{f}_n$, which also depends for example on the size of α^n. Therefore, to estimate the theory uncertainty we do not need a precise estimate of $\hat{f}_n$. We rather need to quantify our limited (or lack of) knowledge of f_n. We will discuss further how to do so in section 3. For now, it is sufficient to think of f_n as an unknown (or imprecisely known) parameter (not necessarily a scalar) which is going to be varied in some way. To estimate the theory uncertainty we have to propagate this variation into the prediction. For this purpose, f_n has to actually appear in our prediction, which means we have to include the next term that contains the dependence on f_n. For

practice, by using perturbation theory to obtain predictions we implicitly assume (and confirm empirically) that the series is still converging at the orders we are working, i.e., that the asymptotic behaviour only becomes relevant at much higher orders than we are working at. This can fail when the series is affected by (leading) renormalons, which essentially spoil the convergence of the series already at low orders. This can be remedied by working in an appropriate renormalon-free scheme in which the nonconverging pieces of the series are absorbed into the definitions of suitable (nonperturbative) parameters. Therefore, our general assumption is that f is expanded in a suitable perturbative scheme that is free of (leading) renormalons, such that the factorial growth of the coefficients does not yet affect the convergence of the series.

[2]To be crystal clear, it is not the uncertainty due to the imprecisely known value of α or any other input parameter.

[3]We can draw the following contrast for illustration: it could be the case that we do not know the structure of the series itself but only how to obtain f in some well-defined limit $\alpha \to 0$. In this case, the missing higher-order terms as a whole are the source of the theory uncertainty, which clearly makes it more difficult to estimate. An example would be a theory where we only know the free theory but not the interacting one. A phenomenologically important example is the kinematic limit in which parton showers are formulated, where we do not even in principle know the structure of the expansion around this limit. Similarly, resummed predictions are performed in a kinematic power expansion for which we know the leading-power limit, but we do not necessarily know the structure of the associated power series (although there has been a lot of progress in recent years to better understand it).

JHEP08(2025)098

example, the NLO and NNLO predictions in eq. (2.4) turn into

$$\begin{aligned} \mathrm{N}^{1+1}\mathrm{LO}: &\qquad f(\alpha, f_2) = \hat{f}_0 + \hat{f}_1\,\alpha + f_2\,\alpha^2\,, \\ \mathrm{N}^{1+2}\mathrm{LO}: &\qquad f(\alpha, f_2, f_3) = \hat{f}_0 + \hat{f}_1\,\alpha + f_2\,\alpha^2 + f_3\,\alpha^3\,, \\ \mathrm{N}^{2+1}\mathrm{LO}: &\qquad f(\alpha, f_3) = \hat{f}_0 + \hat{f}_1\,\alpha + \hat{f}_2\,\alpha^2 + f_3\,\alpha^3\,. \end{aligned} \tag{2.5}$$

The notation $\mathrm{N}^{m+k}\mathrm{LO}$ is meant to indicate that in addition to the first m fully known terms we include k further terms with unknown coefficients for estimating the theory uncertainty.[4]

We have now derived the essence of our approach: the missing series coefficients are the sources of the theory uncertainty. They are well-defined parameters of the perturbative series with a true but unknown value, and we simply treat them accordingly: we include them in the prediction and vary them to account for the theory uncertainty they cause. In this way, the theory uncertainty becomes a truly parametric uncertainty, which is the basis for making it meaningful. Note also that its source is actually different at each order, which also implies that the theory correlations depend on the order of the prediction.

As discussed further in section 3, in reality, the series coefficients have internal structure (e.g. color, partonic channels, etc.). They can also be functions of additional parameters (e.g. quark masses) as well as kinematic variables. Hence, instead of varying them directly, it will be more convenient to parameterize them as $f_n(\theta_n)$ in terms of one or more theory nuisance parameters θ_n that are the unknown parameters to be varied.

The actual range of variation for f_n (really the θ_n) is something we have to decide, which we also discuss further in section 3. By default it will be a sufficiently large range covering the generic, natural size of f_n without knowing the true value $\hat{f}_n$ or *as if* we had no knowledge of $\hat{f}_n$. In addition (or instead) we can also constrain f_n (really the θ_n) from experimental measurements.

If we are able to obtain a more precise estimate of $\hat{f}_n$, that is of course even better. We can include this information to reduce the theory uncertainty due to f_n. At this point, however, we have to remember the uncertainty due to f_{n+1}, which so far we were only able to neglect because it was subdominant compared to f_n. It has to be included now as soon as it becomes relevant compared to the reduced uncertainty from f_n. In this way, a lower-order prediction can gradually turn into a higher-order one. For example, when f_2 is still unknown, we would start at $\mathrm{N}^{1+1}\mathrm{LO}$. As f_2 becomes better known, e.g., due to partial or approximate calculations and/or experimental constraints, we switch at some point to $\mathrm{N}^{1+2}\mathrm{LO}$, which eventually turns into $\mathrm{N}^{2+1}\mathrm{LO}$ when $\hat{f}_2$ has been fully calculated. Our approach thus allows continuously improving theory predictions instead of being tied to large discrete steps from demanding complete formal orders.

2.3 Limitations of scale variations

A well-known limitation of scale variations is that they only have information from the known lower-order terms but no information about the genuine higher-order terms or structure, which makes them not very reliable and prone to underestimation due to accidentally small lower-order terms or due to important new structures appearing at higher order (e.g. new

[4] This notation implies a small departure from conventional wisdom in that $1+1 \neq 2$ and $1+2 \neq 2+1$.

JHEP08(2025)098

partonic channels or new functional dependences on kinematic invariants). Since the amount of variation is largely arbitrary, one also runs the risk of overestimating the uncertainties, which is of course also undesirable.

Even if with sufficient experience and appropriate care one is able to mitigate these dangers of misestimation, scale variations suffer from an even more severe and fundamental limitation: the scales that are being varied are unphysical: they are not actual parameters of the calculation that have a true but only imprecisely known value. There can easily be no value of the scale that actually captures the higher-order result. This immediately tells us that it makes very little sense to try and constrain them from data. Since the scales have no notion of a true value or an uncertainty that is being propagated, their variation also cannot be interpreted as such. This implies that they are fundamentally incapable of correctly determining theory correlations.

There is of course a more fundamental reason for the scales to appear, i.e. the renormalization of the theory, which however has a priori nothing to do with theory uncertainties. To all orders, the calculation does not depend on the scales. By truncating the perturbative series at a finite order, a residual scale dependence remains, i.e., it is an artifact of the calculation. Since this residual dependence must be cancelled by the truncated higher-order terms, it can be exploited to get a sense for the potential size of those missing higher-order terms, but no more than that.

We can capture the essence of the scale-variation approach and expose its limitations already at the level of the generic expansion in eq. (2.3). The key point is that the series coefficients f_n depend on the perturbative scheme by which we mean the exact way of performing the perturbative expansion. In our case here, it corresponds to the exact choice of the expansion parameter α. We can define a new scheme by introducing a new expansion parameter $\tilde{\alpha}$ that differs from α by higher-order terms,

$$\tilde{\alpha}(\alpha) = \alpha\big[1 + b_0\,\alpha + b_1\,\alpha^2 + b_2\,\alpha^3 + \mathcal{O}(\alpha^4)\big]\,. \tag{2.6}$$

The new scheme is uniquely determined by the coefficients b_k appearing in eq. (2.6). Since α and $\tilde{\alpha}$ are the same at lowest order, $\tilde{\alpha} = \alpha + \mathcal{O}(\alpha^2)$, they are equally good expansion parameters as long as we choose $b_k \sim \mathcal{O}(1)$. Apart from this condition, we can choose the b_k freely, so eq. (2.6) actually represents infinitely many possible expansion parameters.

To be concrete, for QCD scale variations we have $\alpha \equiv \alpha_s(\mu_0)$ and $\tilde{\alpha} \equiv \alpha_s(\mu)$, where μ_0 is the central scale and μ is the varied scale. Expanding $\alpha_s(\mu)$ in terms of $\alpha_s(\mu_0)$, we can easily determine the explicit b_k in eq. (2.6) that are implied by scale variations,

$$\begin{aligned}
b_0 &= \frac{\beta_0}{2\pi}\ln\frac{\mu_0}{\mu} && = 0.85\,L\,,\\
b_1 &= \frac{\beta_0^2}{4\pi^2}\ln^2\frac{\mu_0}{\mu} + \frac{\beta_1}{8\pi^2}\ln\frac{\mu_0}{\mu} && = 0.72\,L^2 + 0.34\,L\,,\\
b_2 &= \frac{\beta_0^3}{8\pi^3}\ln^3\frac{\mu_0}{\mu} + \frac{5\beta_0\beta_1}{32\pi^2}\ln^2\frac{\mu_0}{\mu} + \frac{\beta_2}{32\pi^3}\ln\frac{\mu_0}{\mu} && = 0.61\,L^3 + 0.72\,L^2 + 0.13\,L\,.
\end{aligned} \tag{2.7}$$

They are $(k+1)$th-order polynomials in $\ln(\mu_0/\mu)$, and β_k are the $(k+1)$-loop QCD β function coefficients governing the μ dependence of $\alpha_s(\mu)$. In the second expressions we used

JHEP08(2025)098

$n_f = 5$ and defined $L \equiv \ln(\mu_0/\mu)/\ln 2$ to give explicit numerical results for illustration. By convention, we vary μ by a factor of two around μ_0 so L varies between ± 1. Note that scale variations do not actually provide us with the freedom to choose all b_k freely. Instead, they are all determined by choosing a single value for μ or equivalently L.

Continuing our discussion, we now have two (or really infinitely many) equally good ways to perform the perturbative expansion for f, using either α or $\tilde{\alpha}$,

$$\begin{aligned} f(\alpha) &= f_0 + f_1\,\alpha + f_2\,\alpha^2 + f_3\,\alpha^3 + \mathcal{O}(\alpha^4)\,, \\ \tilde{f}(\tilde{\alpha}) &= \tilde{f}_0 + \tilde{f}_1\,\tilde{\alpha} + \tilde{f}_2\,\tilde{\alpha}^2 + \tilde{f}_3\,\tilde{\alpha}^3 + \mathcal{O}(\tilde{\alpha}^4)\,. \end{aligned} \tag{2.8}$$

Since they are expansions of the same f, to all orders they are identical: $f(\alpha) = \tilde{f}(\tilde{\alpha}) = f$. Plugging eq. (2.6) back into $\tilde{f}(\tilde{\alpha})$ and demanding that $f(\alpha) = \tilde{f}(\tilde{\alpha}(\alpha))$ at each order in α, we can easily derive the scheme translation that relates the $\tilde{f}_n$ to the original f_n,

$$\tilde{f}_0 = f_0\,, \qquad \tilde{f}_1 = f_1\,, \qquad \tilde{f}_2 = f_2 - b_0 f_1\,, \qquad \tilde{f}_3 = f_3 - 2b_0(f_2 - b_0 f_1) - b_1 f_1\,. \tag{2.9}$$

Hence, the scheme choice essentially amounts to shuffling around terms between orders in the series.

Although $f(\alpha) = \tilde{f}(\tilde{\alpha})$ to all orders, when we truncate $\tilde{f}(\tilde{\alpha})$ at a finite order to obtain predictions in our new scheme, they will differ by higher-order terms from our original predictions truncating $f(\alpha)$ in eq. (2.4). For example, up to NNLO we have

$$\begin{aligned} \text{LO:}\quad & \tilde{f}(\tilde{\alpha}) = \hat{\tilde{f}}_0 && = \hat{f}_0\,, \\ \text{NLO:}\quad & \tilde{f}(\tilde{\alpha}) = \hat{\tilde{f}}_0 + \hat{\tilde{f}}_1\,\tilde{\alpha} && = \hat{f}_0 + \hat{f}_1\,\alpha + b_0\hat{f}_1\,\alpha^2 + b_1\hat{f}_1\,\alpha^3 + \mathcal{O}(\alpha^4)\,, \\ \text{NNLO:}\quad & \tilde{f}(\tilde{\alpha}) = \hat{\tilde{f}}_0 + \hat{\tilde{f}}_1\,\tilde{\alpha} + \hat{\tilde{f}}_2\,\tilde{\alpha}^2 && = \hat{f}_0 + \hat{f}_1\,\alpha + \hat{f}_2\,\alpha^2 + \big[2b_0(\hat{f}_2 - b_0\hat{f}_1) + b_1\hat{f}_1\big]\alpha^3 + \mathcal{O}(\alpha^4). \end{aligned} \tag{2.10}$$

In the second expressions we used eqs. (2.6) and (2.9) to rewrite the prediction in terms of the original $\hat{f}_n$ and α to explicitly expose the differences highlighted in red. In general, the $\mathrm{N}^n\mathrm{LO}$ predictions in the two schemes agree up to $\mathcal{O}(\alpha^n)$ but differ by $\mathcal{O}(\alpha^{n+1})$ and higher terms (except for the LO predictions, which happen to agree exactly because there is no scheme dependence yet at this order).

In the scale-variation approach, this higher-order scheme dependence is now exploited by taking the difference between the two schemes as an estimate Δf of the theory uncertainty,

$$\begin{aligned} \text{LO:}\quad & \Delta f(\alpha) = 0\,, \\ \text{NLO:}\quad & \Delta f(\alpha) = b_0\hat{f}_1\,\alpha^2 + b_1\hat{f}_1\,\alpha^3 + \mathcal{O}(\alpha^4)\,, \\ \text{NNLO:}\quad & \Delta f(\alpha) = \big[2b_0(\hat{f}_2 - b_0\hat{f}_1) + b_1\hat{f}_1\big]\alpha^3 + \mathcal{O}(\alpha^4)\,. \end{aligned} \tag{2.11}$$

The limitations of the scale-variation approach should be clear from the above discussion. They are fundamentally caused by the fact that the scalar parameter L (or μ) that is being varied is not a true parameter of the prediction, i.e., it has no notion of having a true value $\hat{L}$ that reproduces the true missing $\hat{f}_n$. This is because the coefficients of α^n in eq. (2.11) are in general not a valid parameterization of the missing higher-order coefficients f_n. As soon as the f_n have some nontrivial internal structure, they will not just be given by fixed linear

JHEP08(2025)098

combinations of lower-order coefficients.[5] The fact that $\Delta f(\alpha)$ is proportional to the true values of the lower-order coefficients causes the common pitfall of underestimation already mentioned at the beginning of this subsection. For example at NLO, if $\hat{f}_1$ happens to be smaller than its natural size, or lacks relevant internal structures of f_2, $b_0\hat{f}_1$ will underestimate the natural size of f_2 and thus the uncertainty due to it. This is made even more severe by the fact that we practically always use the same conventional value for b_0 regardless of the actual properties of $\hat{f}_1$ and f_2.

Besides these dangers of misestimation, as L is not a true parameter of the prediction, its variation fundamentally cannot yield a meaningful theory uncertainty to begin with. That is, it cannot imply correlations or be constrained by measurements, and the resulting uncertainty estimates do not admit a meaningful statistical interpretation.

These limitations of scale variations have been known for a long time. A common method to alleviate the possible underestimation is to perform a variety of scale variations. The individual differences are then combined into a total uncertainty estimate Δf by taking their envelope because the different variations just probe the potential size of the same missing higher-order terms in different ways and are not individually meaningful. To mitigate the lack of correlations, the best we can do is to impose a context-specific correlation model on the total Δf. Dedicated correlation models have been discussed in the context of both fixed-order predictions (see e.g. refs. [31–38]) as well as resummed predictions (see e.g. refs. [32, 39–44]). Deciding whether or how to correlate or uncorrelate scale variations for different predictions also just amounts to choosing some ad hoc correlation model. While such correlation models can be theoretically motivated, they are still ad hoc assumptions, so they are only bandaids and do not cure the underlying problem.

In practice, the scale-variation based uncertainties are often propagated by introducing ad hoc nuisance parameters θ_f by writing the predictions at a given order as $f + \theta_f \Delta f$ with $\theta_f = 0 \pm 1$. Although this may be done to simplify the technical implementation, we cannot stress enough that doing so obviously does not turn Δf automagically into an actual parametric uncertainty. Such ad hoc nuisance parameters are not genuine nuisance parameters and must not be treated or misinterpreted as such. In particular, despite the fact that this has become a common mispractice, they *may not* be profiled.

3 Theory nuisance parameters

This section gives a general and largely self-contained discussion of theory nuisance parameters (TNPs) and TNP-based theory predictions and uncertainties. It is intended for all audiences. Readers should have read section 2.2, but not necessarily other subsections of section 2.

Section 3.1 gives an introduction and general overview of the TNP approach picking up where we left off in section 2.2. It is a prerequisite for the subsequent subsections and the rest of the paper. In the subsequent subsections we further discuss several aspects of

[5] The attentive reader might have noticed that in the special case where the f_n are pure scalars, we would in principle have enough degrees of freedom to correctly reproduce each $\hat{f}_n$ if we were to choose each b_k separately. However, apart from the fact that scale variations do not actually provide this freedom, as we will see in sections 3 and 4, the cases where we could parameterize f_n correctly as a single scalar are rare. Furthermore, this essentially precludes accounting for any correlations.

JHEP08(2025)098

the TNP approach. They are largely independent of each other, so readers not concerned with any one of these aspects can freely skip the respective subsection: section 3.2 discusses some implementation aspects. Section 3.3 discusses constraining the TNPs based on theory considerations and measurements. Section 3.4 discusses how scale and perturbative scheme choices figure into our approach.

3.1 General overview

We consider the expansion of a quantity f in the small parameter α,

$$f(\alpha) = f_0 + f_1\,\alpha + f_2\,\alpha^2 + f_3\,\alpha^3 + \mathcal{O}(\alpha^4)\,. \tag{3.1}$$

As before, we use a hat to distinguish a parameter (f_n, θ_n, ...) from its true value ($\hat{f}_n$, $\hat{\theta}_n$, ...). To obtain a perturbative prediction for f at order $\mathrm{N}^{m+k}\mathrm{LO}$ in our approach, we include the true values of the first m series coefficients and in addition the next k terms whose coefficients are (considered to be) unknown parameters, which account for the (dominant) theory uncertainty. For example,

$$\begin{aligned}
\mathrm{N}^{1+1}\mathrm{LO}: &\qquad f(\alpha,\theta_2) = \hat{f}_0 + \hat{f}_1\,\alpha + f_2(\theta_2)\,\alpha^2\,,\\
\mathrm{N}^{1+2}\mathrm{LO}: &\qquad f(\alpha,\theta_2,\theta_3) = \hat{f}_0 + \hat{f}_1\,\alpha + f_2(\theta_2)\,\alpha^2 + f_3(\theta_3)\,\alpha^3\,,\\
\mathrm{N}^{2+1}\mathrm{LO}: &\qquad f(\alpha,\theta_3) = \hat{f}_0 + \hat{f}_1\,\alpha + \hat{f}_2\,\alpha^2 + f_3(\theta_3)\,\alpha^3\,.
\end{aligned} \tag{3.2}$$

In addition to eq. (2.5), we have now parameterized the unknown series coefficients $f_n(\theta_n)$ in terms of theory nuisance parameters θ_n for $n = m+1, \ldots, m+k$.

In general, f_n has a nontrivial internal structure involving various discrete and continuous variables. In principle, some of this structure needs to be accounted for in the "TNP parameterization" $f_n(\theta_n)$, which therefore depends in general on multiple TNPs $\theta_{n,i}$. For example, when f_n depends on different flavor or color channels, we might need a $\theta_{n,i}$ for each. When f_n depends on a continuous kinematic variable, we might need to parameterize this dependence in terms of several $\theta_{n,i}$. The required number of TNPs thus depends on how f_n's internal structure is parameterized. For notational simplicity we always let $\theta_n \equiv \{\theta_{n,i}\}$ stand for the full set of $\theta_{n,i}$.

Different quantities can depend on a common $\theta_{n,i}$ when their coefficients internally depend on the same perturbative ingredient parameterized by $\theta_{n,i}$. Some obvious examples are universal objects in QCD which appear in many places, such as the beta function, splitting functions, or the cusp anomalous dimension. In this case, a given $\theta_{n,i}$ is always varied simultaneously everywhere it appears and the resulting uncertainty is treated as 100% correlated. This is how theory correlations among different quantities are correctly accounted for. In fact, as we will discuss in section 4, which parts of the internal structure we need to parameterize is directly determined by the theory correlations we need to account for. On the other hand, different $\theta_{n,i}$ should a priori be mutually independent and correspond to independent sources of uncertainties. They can then be varied independently and their resulting uncertainties can be treated as a priori uncorrelated.

An essential requirement on the TNP parameterization is that it must be able to reproduce the coefficient's true value $\hat{f}_n$. That is, the TNPs must have true values $\hat{\theta}_n \equiv \{\hat{\theta}_{n,i}\}$

JHEP08(2025)098

corresponding to $\hat{f}_n$,

$$\hat{f}_n = f_n(\hat{\theta}_n)\,. \tag{3.3}$$

This is what makes the TNPs themselves true parameters of the perturbative series, and what allows us to obtain meaningful constraints on their (a priori unknown) values. That is, as for any physical parameter whose true value is unknown, we can obtain a best estimate for the θ_n with some uncertainty, which we denote as

$$\theta_n = u_n \pm \Delta u_n\,. \tag{3.4}$$

This estimate could come from theory considerations or experimental measurements or both. It should be accompanied with a quantitative statistical interpretation of the uncertainty, which may be more or less rigorous depending on where it comes from. Statistically speaking, we want to treat eq. (3.4) as coming from a real or imagined auxiliary measurement, as for any other systematic uncertainty. Normally, eq. (3.4) will only provide a loose constraint for the θ_n to have their natural size but not a precise estimate of $\hat{\theta}_n$. To emphasize this point, we mostly talk about *constraints* on the θ_n rather than *estimates* of them. When we have several constraints for the same $\theta_{n,i}$, we combine them appropriately. The central theory prediction is then obtained by setting the θ_n to their central value u_n, while the theory uncertainty is evaluated by appropriately propagating or incorporating the uncertainties Δu_n, including their statistical interpretation, into the final results. In this way, TNPs provide us with parametric, meaningful theory uncertainties. They can (and should) always be propagated, combined, and interpreted like standard parameter uncertainties.

To summarize, there are two main steps to obtain a theory prediction with TNP-based uncertainties:

1. Derive an appropriate TNP parameterization $f_n(\theta_n)$ that satisfies all requirements for all quantities of interest and implement it into the predictions.

2. Obtain suitable auxiliary constraints on all relevant TNPs θ_n and propagate them into the final results.

It is important to separate these two steps both logically and practically, because they depend on different levels of approximations and assumptions.

The TNP parameterization in step 1 is determined by the internal structure of f_n, which is what it is and not really debatable. As we will see in section 4, all choices we can make here are based on external requirements and can be framed as making approximations that are systematically improvable if needed. Hence, the theory uncertainty and correlation structure encoded by a given TNP parameterization is always correct to some formal accuracy. Deriving it requires expert domain knowledge. It must thus be provided as part of the prediction and cannot be left to users. This also means that we cannot provide a generic parameterization that is going to work in all cases. Instead, in section 4 we discuss the general principles and strategies for constructing suitable parameterizations, and in section 6 we discuss a full-fledged example application.

On the other hand, in step 2 we can debate to what extent a specific constraint (theoretical or experimental) is deemed sufficiently reliable or not and informed users can choose to include

JHEP08(2025)098

it or not based on their preferences or requirements. We will see in section 5 that it is indeed possible to obtain robust theory constraints. Furthermore, users can choose their preferred method of propagating the TNP uncertainties. One could either vary the TNPs explicitly or derive a theory covariance matrix for all predictions by performing a standard Gaussian error propagation. When fitting to data one could repeat the fit for each variation (sometimes called scanning or offset method), or use the derived theory covariance matrix, or profile the TNPs as genuine nuisance parameters with eq. (3.4) imposed as an auxiliary constraint. The option to profile the TNPs is of course a key advantage, and where their name comes from, as it directly constrains the TNPs by the data. We will come back to this in section 3.3.

3.2 Approximate implementation

In practice, to upgrade an existing $\mathrm{N}^m\mathrm{LO}$ prediction to a full-fledged $\mathrm{N}^{m+k}\mathrm{LO}$ prediction with TNP uncertainties, one has to implement the correct structure of the next k orders in terms of the parameterized $f_n(\theta_n)$. Depending on the complexity of the prediction and parameterization this can be a challenging task in itself. Therefore, as an approximation to the $\mathrm{N}^{m+1}\mathrm{LO}$ implementation one can also consider using the structure of the existing $\mathrm{N}^m\mathrm{LO}$ prediction and absorb the uncertainty term into the highest known order, for example,

$$\begin{aligned} \mathrm{N}^{1+0}\mathrm{LO}: &\qquad f(\alpha,\theta_2) = \hat{f}_0 + \left[\hat{f}_1 + \alpha_0 f_2(\theta_2)\right]\alpha\,, \\ \mathrm{N}^{2+0}\mathrm{LO}: &\qquad f(\alpha,\theta_3) = \hat{f}_0 + \hat{f}_1\,\alpha + \left[\hat{f}_2 + \alpha_0 f_3(\theta_3)\right]\alpha^2\,. \end{aligned} \tag{3.5}$$

Here, α_0 denotes a fixed value of α, which is not part of the formal series structure, e.g., it is does not participate in counting orders of α. In extension to our notation, we denote this as $\mathrm{N}^{m+0}\mathrm{LO}$.[6]

This approximation makes little difference for our simple illustrative series, but it can make more of a difference for real-life series. For example, it might require approximating or dropping parts of the internal structure of $f_n(\theta_n)$ that cannot be absorbed into the existing structure of f_{n-1}. Furthermore, when the full series involves a product of several individual series (as e.g. in resummed predictions), one correctly accounts for all $\mathcal{O}(\alpha^{m+1})$ cross terms of lower-order pieces at $\mathrm{N}^{m+1}\mathrm{LO}$, while they are neglected at $\mathrm{N}^{m+0}\mathrm{LO}$. So whilst this approximation still provides parametric uncertainties, the impact of the parameters is only approximately correct because one effectively uses the $\mathcal{O}(\alpha^{m+1})$ uncertainty parameters with the lower $\mathcal{O}(\alpha^m)$ uncertainty structure. We might expect this to have only a limited effect on the overall size of the theory uncertainty, while it might have a bigger effect on the theory correlations. We generally recommend using the $\mathrm{N}^{m+1}\mathrm{LO}$ prediction. If this is unfeasible for practical reasons, one can still resort to $\mathrm{N}^{m+0}\mathrm{LO}$, but one should ideally check with available orders how much of a difference this approximation makes.

As discussed at the end of section 2.2, when the $\theta_{n,i}$ become strongly constrained, we have to include at some point the $\theta_{n+1,i}$, which means upgrading the prediction from $\mathrm{N}^{m+1}\mathrm{LO}$ to $\mathrm{N}^{m+2}\mathrm{LO}$. A convenient way to test if this is already warranted or not is to include the $\theta_{n+1,i}$ in this approximate fashion, i.e, approximate $\mathrm{N}^{m+2}\mathrm{LO}$ by $\mathrm{N}^{m+1+0}\mathrm{LO}$. Another possible scenario is a mixed case where some $\theta_{n,i}$ are well estimated or exactly known such that their

[6] This approximation thus comes at the minor cost of further breaking basic arithmetic: $m + 0 \neq m$.

JHEP08(2025)098

corresponding $\theta_{n+1,i}$ should be included, while most others are still largely unconstrained. In this case, it would be premature to upgrade to N^{m+2}LO but one can already include the few required $\theta_{n+1,i}$ approximately.

3.3 Constraining the TNPs

As already mentioned, since the TNPs are proper parameters with true values, it is perfectly consistent to profile them in fits to data, in stark contrast to scale-variation based approaches. We discuss several aspects of this in section 3.3.2 below.

Nevertheless, we still need a theory uncertainty estimate for the "pre-fit" theory predictions, i.e., before confronting them with experimental measurements. This is obviously necessary for any theoretical studies where we do not (yet) fit to data. Even when fitting to data, it might be unfeasible or undesirable to always constrain all TNPs entirely from data alone. Another reason is to be able to judge or test whether the data constrains some $\theta_{n,i}$ too strongly. Therefore, we need some constraint on the TNPs based on theory considerations, which we briefly discuss next and in much more detail in section 5.

3.3.1 Theory constraints

As our default theory constraint, without any further information, we will have $u_n = 0$ and Δu_n given by the "natural size" of θ_n, by which we mean we would generically expect $|\hat{\theta}_n| \lesssim \Delta u_n$. To be more concrete, if we knew with 68% confidence level that $|\hat{\theta}_n| \leq N_n$ we would take $\Delta u_n = N_n$. The default theory constraint thus requires us to estimate the natural size of θ_n and then allows θ_n to vary within it. Without loss of generality, we assume that θ_n is normalized to have a natural size of $\mathcal{O}(1)$, i.e., such that we generically expect $|\hat{\theta}_n| \lesssim 1$ and thus $\Delta u_n \simeq 1$.

Estimating the natural size of θ_n is basically equivalent to estimating the natural size of f_n, which is usually possible by considering its known higher-order structure. For example, just pulling out the known leading color and loop factors is usually sufficient to normalize θ_n to have $\mathcal{O}(1)$ natural size. As this estimate directly determines the eventual size of the theory uncertainty we would of course like to narrow it down better than just a generic $\mathcal{O}(1)$ factor, ideally to within a factor of 2 or better. This can then be tested extensively on many known series coefficients. As we will see in section 5, doing so we are able to obtain a (almost surprisingly) robust estimate for Δu_n, well within a factor of 2, and with a well-defined statistical interpretation.

In some cases we have further theoretical information that relates a priori independent structures in f_n (or different coefficients or quantities), which have to satisfy certain relations. An example are consistency relations between different anomalous dimensions, which they must satisfy exactly. We have different options to incorporate such information. If the relations are exact and simple enough, one option is to solve them explicitly and eliminate some $\theta_{n,i}$ by expressing them in terms of others. This amounts to incorporating the constraint directly at the level of the parameterization. Otherwise, especially in cases with inexact relations, we can keep all a priori independent $\theta_{n,i}$ and account for each relation by imposing a corresponding auxiliary theory constraint, which can then lead to nontrivial a posteriori correlations between some $\theta_{n,i}$.

JHEP08(2025)098

3.3.2 Measurement constraints

Theory-based constraints, unless they are exact constraints, inevitably involve some theoretical prejudice in the size of the uncertainty. (They can also induce a potential bias due to scheme and parameterization dependences, as discussed in sections 3.4 and 4.4.) However, when the theory predictions are used to interpret experimental measurements, which is when the theory uncertainties arguably matter the most, the TNPs can be constrained by the measurements themselves by including them in the fit as actual nuisance parameters. Hence, we have the choice to avoid (or at least minimize) the dependence on some undesired theoretical prejudice by not imposing (or reducing) some theory-based constraint and thereby rely more on the measurements. Of course, this comes at the expense of some experimental sensitivity. A key advantage of our approach is that it actually gives us this choice. Thus, profiling the TNPs in fits to data has many important benefits:

- It allows constraining the theory uncertainties by data.
- It avoids or reduces the susceptibility to possible theory prejudice or biases.
- It allows taking into account possibly important correlations between the TNPs and the parameters of interest.

The last point is because by profiling the TNPs we let the fit decide between moving a parameter of interest vs. moving the theory predictions.

One might be worried that when the TNPs are constrained by the data, they also absorb the effects of all yet higher-order corrections that have not been included in the theory uncertainty estimate, or more generally, the effect of any type of missing contribution or deficiency in our description. However, this problem is always there: any such effect is *always* collectively absorbed into all fitted parameters (both nuisance parameters and parameters of interest). The inclusion of TNPs in the fit does not make this any worse. In fact, it is likely to *reduce* this problem as far as missing theory contributions are concerned, because it is not unlikely that they are structurally similar to the theory uncertainty terms we now include. This means, they get absorbed more likely into the fitted TNPs than into the parameters of interest, thus reducing the contamination of the parameters of interest, which is exactly what we want.

We should of course not blindly let the TNPs get misused for unintended purposes. Formally, any unaccounted theory effect is really just an unaccounted source of theory uncertainty. By neglecting it we assume that it is small enough to be neglected against other uncertainties, which is equivalent to accepting that it will be effectively absorbed somewhere (hopefully mostly into the TNPs). However, this is exactly equivalent to the conditions under which we are allowed to neglect f_{n+1} compared to f_n as discussed in section 2.2. The same discussion obviously applies to any other source of theory uncertainty as well. In particular, if with sufficiently precise data we want to actually determine θ_n, then we effectively elevate it to a parameter of interest. We then have to at least include θ_{n+1} to account for the remaining leading theory uncertainty, as discussed at the end of section 2.2, and more generally also any other source of theory uncertainty of similar size.

JHEP08(2025)098

3.4 Scheme dependence

In our approach, we still have to choose a specific scale or scheme to perform the perturbative expansion. For our purposes, the perturbative scheme includes all choices of renormalization and factorization schemes as well as the choices of all associated scales we have to make. One might wonder how the dependence on this scheme figures into our approach now. In general, the scheme dependence is not a problem. The scheme just has to be well defined so we can translate from one scheme to another, and the scheme dependence has to be treated consistently.

We already discussed the scheme dependence at the level of our example series in section 2.3. To briefly recap, by choosing different expansion parameters α or $\tilde{\alpha}$, we have different ways to perform the perturbative expansion for the same quantity f,

$$\begin{aligned} f(\alpha) &= f_0 + f_1\,\alpha + f_2\,\alpha^2 + f_3\,\alpha^3 + \mathcal{O}(\alpha^4)\,, \\ \tilde{f}(\tilde{\alpha}) &= \tilde{f}_0 + \tilde{f}_1\,\tilde{\alpha} + \tilde{f}_2\,\tilde{\alpha}^2 + \tilde{f}_3\,\tilde{\alpha}^3 + \mathcal{O}(\tilde{\alpha}^4)\,. \end{aligned} \tag{3.6}$$

To all orders, the two series give identical results, $f(\alpha) = \tilde{f}(\tilde{\alpha}) = f$, but at any truncated order they differ by higher-order terms [see eq. (2.10)]. The two schemes are uniquely defined relative to each other by the relation between α and $\tilde{\alpha}$,

$$\tilde{\alpha}(\alpha) = \alpha\big[1 + b_0\,\alpha + b_1\,\alpha^2 + b_2\,\alpha^3 + \mathcal{O}(\alpha^4)\big]\,, \tag{3.7}$$

from which the relation between the series coefficients f_n and $\tilde{f}_n$ follows [see eq. (2.9)],

$$\tilde{f}_0 = f_0\,, \qquad \tilde{f}_1 = f_1\,, \qquad \tilde{f}_2 = f_2 - b_0 f_1\,, \qquad \tilde{f}_3 = f_3 - 2b_0(f_2 - b_0 f_1) - b_1 f_1\,. \tag{3.8}$$

To discuss the scheme dependence or ambiguity in the context of our TNP-based predictions, it is important to distinguish two places where the scheme choice enters: first, the scheme dependence of f_n is inherited by θ_n. We thus pick a common "reference scheme" in which the θ_n are defined via the TNP parameterization $f_n(\theta_n)$. We will come back to the question of how to pick the reference scheme below. For notational simplicity, we continue using f_n, θ_n, α to denote the parameters in the reference scheme, while we add tildes, $\tilde{f}_n$, $\tilde{\theta}_n$, $\tilde{\alpha}$ for the parameters in some other scheme.

Second, as always we need to pick a scheme in which to evaluate the prediction itself. An obvious and natural choice is to use the same scheme for both, i.e., we would just use the reference scheme $f(\alpha)$, but in principle they could also be different. To obtain the prediction in a different scheme, $\tilde{f}(\tilde{\alpha})$, in terms of the reference parameters $f_n(\theta_n)$, we simply translate from the reference scheme by using eq. (3.8) for the series coefficients of $\tilde{f}(\tilde{\alpha})$. For example, translating the predictions in eq. (3.2), we get

$$\begin{aligned} &\mathrm{N}^{1+1}\mathrm{LO}: \quad \tilde{f}(\tilde{\alpha},\theta_2) = \hat{f}_0 + \hat{f}_1\,\tilde{\alpha} + \big[f_2(\theta_2) - b_0\hat{f}_1\big]\,\tilde{\alpha}^2\,, \\ &\mathrm{N}^{2+1}\mathrm{LO}: \quad \tilde{f}(\tilde{\alpha},\theta_3) = \hat{f}_0 + \hat{f}_1\,\tilde{\alpha} + \big[\hat{f}_2 - b_0\hat{f}_1\big]\,\tilde{\alpha}^2 + \big[f_3(\theta_3) - 2b_0(\hat{f}_2 - b_0\hat{f}_1) - b_1\hat{f}_1\big]\,\tilde{\alpha}^3\,. \end{aligned} \tag{3.9}$$

This makes it clear that the θ_n are always the same parameters and are independent of the scheme of the prediction.

To discuss the residual scheme dependence of the prediction, first consider the $\mathrm{N}^{1+1}\mathrm{LO}$ prediction in eq. (3.9): its residual scheme dependence is of $\mathcal{O}(\tilde{\alpha}^3)$, because the $\mathcal{O}(\tilde{\alpha}^2)$ term

JHEP08(2025)098

includes by construction the correct scheme-dependent term $-b_0\hat{f}_1\,\tilde{\alpha}^2$ that cancels the scheme dependence of $\tilde{\alpha}$ in the previous term. Similarly, at $\mathrm{N}^{2+1}\mathrm{LO}$ (and also $\mathrm{N}^{1+2}\mathrm{LO}$ which is not shown), the $\mathcal{O}(\tilde{\alpha}^3)$ term includes all necessary terms to cancel the $\mathcal{O}(\tilde{\alpha}^3)$ scheme dependence of the lower-order terms, so the residual scheme dependence is pushed to $\mathcal{O}(\tilde{\alpha}^4)$. In general, the residual scheme dependence of the $\mathrm{N}^{m+k}\mathrm{LO}$ prediction is by construction of $\mathcal{O}(\alpha^{m+k+1})$ and formally beyond the smallest included theory uncertainty of $\mathcal{O}(\alpha^{m+k})$. We can thus ignore it for the same reason we can drop the $\mathcal{O}(\alpha^{m+k+1})$ theory uncertainty caused by f_{m+k+1}.[7]

We now come back to the question of how to pick the reference scheme for θ_n. Since θ_n plays the role of an input parameter, defining it in a different scheme merely defines a different (but related) input parameter $\tilde{\theta}_n$ with a different but related true value $\hat{\tilde{\theta}}_n$. Their relation follows from the relation between f_n and $\tilde{f}_n$ in eq. (3.8).[8] Since this relation is exact, it a priori does not matter which parameter we use; we can always translate exactly from one to the other. Hence, choosing a common reference scheme for θ_n is akin to our conventional choice of $\alpha_s(m_Z)$ (defined in a certain reference scheme namely $\overline{\mathrm{MS}}$ at $\mu = m_Z$) as the common input parameter for α_s. We could have just as well chosen $\alpha_s(m_W)$ or $\alpha_s(42\,\mathrm{GeV})$. Since the relation between them is known very precisely, it makes practically no difference which one we decide to extract from data.[9]

The key difference to a purely data-determined parameter like $\alpha_s(m_Z)$ is that for θ_n we also want to be able to obtain constraints based on theory considerations. For this purpose, some reference schemes are better than others. A good reference scheme is one where the f_n are bounded by their natural size, i.e., they don't contain large scheme-induced artifacts, such that the corresponding θ_n are of natural size. We stress that this does not mean that the best reference scheme is necessarily the one where $\hat{f}_n$ is the smallest, as this might just be accidental. Instead, the best reference scheme is the one for which we are *most confident* that the f_n, and thereby the θ_n, are of natural size, because this maximizes the confidence we can ascribe to theory constraints that estimate the natural size of θ_n.

For the scale dependence, this is basically how we would usually choose (or at least should be choosing) the central scale. We choose one for which we are most confident that the f_n do not contain large logarithms of the scale. We usually do not (or at least should not) choose the central scale by minimizing the highest-order $\hat{f}_n$. Hence, by default we can just recycle our "best" conventional or canonical central scales as reference scales.

To discuss the effect of choosing different reference schemes in more detail, let us compare for concreteness the $\mathrm{N}^{1+1}\mathrm{LO}$ predictions with TNPs defined in different reference schemes,

$$\begin{aligned}
&\mathrm{N}^{1+1}\mathrm{LO}: \quad f(\alpha,\theta_2) = \hat{f}_0 + \hat{f}_1\,\alpha + f_2(\theta_2)\,\alpha^2\,,\\
&\mathrm{N}^{1+1}\mathrm{LO}: \quad f(\alpha,\tilde{\theta}_2) = \hat{f}_0 + \hat{f}_1\,\alpha + [\tilde{f}_2(\tilde{\theta}_2) + b_0\hat{f}_1]\,\alpha^2\,,
\end{aligned} \tag{3.10}$$

[7] More precisely, it causes a bias in the central value of our predictions, which is small compared to the nominal theory uncertainty.

[8] Depending on the complexity of $f_n(\theta_n)$, the exact relation between the individual $\theta_{n,i}$ and $\tilde{\theta}_{n,i}$ can be more nontrivial than suggested by eq. (3.8), as it may not be immediately obvious how to distribute the scheme difference between them. There can also be some $\theta_{n,i}$ that are scheme independent, namely those that parameterize new structures in f_n that cannot be generated by the scheme change and are not captured by scale variations.

[9] In contrast, for quark masses the scheme translation can induce a sizeable uncertainty, so the optimal reference scheme for the mass parameter is the scheme of the prediction that is used for its extraction.

JHEP08(2025)098

where the two TNP parameterizations have to satisfy the scheme relation from eq. (3.8),

$$\tilde{f}_2(\tilde{\theta}_2) = f_2(\theta_2) - b_0 \hat{f}_1 \,, \tag{3.11}$$

which determines the exact relation between the two parameters θ_2 and $\tilde{\theta}_2$. As long as they satisfy eq. (3.11), the predictions in eq. (3.10) are literally identical and it does not matter at all which parameter we use.

A dependence on the reference scheme enters when the constraints we put on θ_n vs. $\tilde{\theta}_n$ violate the scheme relation between them. As already discussed above, real measurement constraints always respect this relation, simply because they always constrain the quantity f itself, which is scheme independent. We can see this immediately from eq. (3.10): any constraint we get on f, from data or elsewhere, yields exactly the same constraint on either $f_2(\theta_2)$ or $\tilde{f}_2(\tilde{\theta}_2) + b_0\hat{f}_1$, and thus respects eq. (3.11).

On the other hand, our default theory constraints are scheme dependent because we constrain θ_n directly. It clearly makes a difference whether we decide to constrain $\theta_n = 0 \pm 1$ or $\tilde{\theta}_n = 0 \pm 1$. They yield the same constraint for $f_n(\theta_n)$ or $\tilde{f}_n(\tilde{\theta}_n)$, which means the (absolute) uncertainty on $f(\alpha, \theta_n)$ and $f(\alpha, \tilde{\theta}_n)$ is the same but their central value is shifted by the scheme-dependent terms. Thus, the choice of reference scheme causes a bias (or prior) in the central value of our theory constraint, but also nothing more.

At this point, we need to take a slight detour, as this type of bias is actually not specific to the theory uncertainty but can be the case for any systematic uncertainty. It amounts to the inherent ambiguity that is always present when we have an unknown parameter that lacks any constraints and for which we are therefore forced to pick a reasonable value. In the absence of any external information, there is simply no unbiased way of doing so.

This is where the difference between a "real" vs. "imagined" auxiliary measurement comes in. More precisely, this is how we can *define* this distinction: a real measurement or constraint imposes an unambiguous central value. An imagined one, while also imposing a central value, leaves open the choice on which parameter to impose it. Note that not all theory-based constraints are necessarily of the latter type, e.g., an actual approximate calculation of θ_n will usually apply in a specific scheme and thus resolve the scheme ambiguity. The equivalent constraint for $\tilde{\theta}_n$ would then follow from their scheme relation.

There are standard ways to deal with such biases in practice: first, the bias from choosing a parameter's central value is *not* an additional source of uncertainty. It is a bias that may or may not be covered by the parameter's uncertainty. If it is not, we might decide to enlarge the uncertainty or explicitly state the choice that causes the bias as a precondition or both. We then have several options for treating the bias:

1. If the parameter's bias is small compared to the parameter's uncertainty, we can formally neglect it and move on.

2. Otherwise, if the final analysis or interpretation is insensitive to the bias, i.e., the resulting bias induced in the final result is small compared to its other uncertainties, we can ignore it for practical purposes and move on.

3. Otherwise, the final result is sensitive to the bias. In this case,

JHEP08(2025)098

(a) if possible, we leave the parameter unconstrained and let the data itself constrain it. This removes any bias at the potential cost of reducing the power of the data for determining other parameters.

(b) Otherwise, if possible and still useful, we quote the final result explicitly stating the preconditions under which it is valid.

(c) Otherwise, we have to accept the fact that the analysis or interpretation is not possible or useful with current knowledge.

In cases 1 and 2, we always have the option to further constrain the parameter's uncertainty by the data. Note that these cases require us to be able to quantify the bias, otherwise we are automatically in case 3.

We now return to our discussion at hand. First, the exact scheme choice for the prediction itself is actually an example of case 1. It also causes a small bias in the prediction's central value, but as discussed above, this ambiguity is formally at least one order higher than the theory uncertainty.

The bias caused by the choice of reference scheme of θ_n in our default theory constraint should be covered by its uncertainty on θ_n as long as we are comparing two equally good schemes and the uncertainty is not underestimated. In other words, if the bias is not covered by the uncertainty, the scheme difference $|\theta_n - \tilde{\theta}_n|$ exceeds what we estimated to be θ_n's natural size. This means one of the schemes is not a good one. If we cannot figure out which one, then the uncertainty estimate, i.e., our estimate of θ_n's natural size, is too small. Often however, we really do have a theoretically preferred reference scheme, for instance when there is an obvious canonical scale choice, which effectively reduces the bias to be smaller than the uncertainty.[10]

We should also stress that at the end of the day this bias is not a major issue. For the pre-fit theory predictions we are effectively in case 3b, unless we can argue for case 1. However, at this stage the exact central value is not actually that useful or interesting, what matters more are the uncertainties. We just have to keep in mind when discussing pre-fit predictions that we had to make an explicit choice for the exact central value and we could have made a slightly different one within the uncertainties. The central value actually becomes relevant when the predictions are confronted with data, but at this point the bias can be reduced or even eliminated by the data itself. To be prudent, we can in addition weaken any biased theory constraint, e.g. by taking twice its uncertainty, to reduce its constraining power and put more emphasis on the data as much as desired.

Finally, the above discussion provides another way to highlight the key limitation of scale variations. They can at best provide an estimate of the scheme-induced bias but not of the actual uncertainty, because they cannot actually probe the underlying unknown parameter (the missing f_n) whose central value is implicitly chosen to be zero.

[10] Evaluating the bias is of course subjective, but we should only consider alternative schemes for which we really are as confident as for our reference scheme that the $\tilde{f}_n$ are of natural size. In other words, the bias is not automatically given now by varying the scale by a factor of 2. When we (used to) do scale variations by a factor of 2, it was not to estimate the scheme bias, we just exploited the scheme dependence to guess the theory uncertainty.

JHEP08(2025)098

4 Parameterization guide

This section discusses how the series coefficients f_n can be parameterized in terms of theory nuisance parameters θ_n. It is intended for readers wishing to implement TNPs into their predictions as well as curious readers wishing to use such predictions. It assumes readers are familiar with section 3.1.

As already mentioned in section 3.1, deriving an optimal TNP parameterization $f_n(\theta_n)$ amounts to deriving the correct and relevant theory uncertainty and correlation structure for the prediction at hand. It must thus be regarded as an integral part of performing and providing the prediction itself.[11] This is in general a nontrivial task and requires expert knowledge on the structure of the underlying perturbative series. It is clearly not as easy or convenient as performing scale variations; there is no free lunch.

The particular parameterization strategy or combination of strategies to follow will depend on the case at hand. We hope our discussion here will serve as a useful guideline and starting point for future investigations.

In the next subsection we setup the basic problem and along the way give an outline of the rest of this section. We also provide a brief executive summary for the impatient reader at the beginning of each subsequent subsection.

4.1 Overview and outline

The internal structure of f_n is determined by various dependences on both discrete and continuous parameters, variables, or labels. Typical discrete dependencies are partonic channels, color channels, or any type of discrete quantum numbers. Examples of continuous dependencies are kinematic variables or particle masses. In some cases, f_n is mathematically a continuous function of a parameter which in practice only takes discrete (typically integer) values. Examples are the number of fermions, n_f, or the number of colors, N_c. In section 4.6, we will discuss these and other examples to illustrate our general discussion.

For now, let us denote any one of these variables (discrete or continuous) by x. For the sake of simplicity and without much loss of generality we focus on the case of $f_n(x)$ being a function of a single variable x at a time. The true value of $f_n(x)$ is again denoted by $\hat{f}_n(x)$. The dependence on multiple variables can be treated as a direct generalization as discussed in section 4.5.

Our goal is to construct a TNP parameterization $f_n(x, \theta_n)$ that satisfies the key requirement in eq. (3.3), which now reads

$$\hat{f}_n(x) = f_n(x, \hat{\theta}_n)\,. \tag{4.1}$$

Another goal is that the $\theta_{n,i}$ should be mutually independent, i.e., they should correspond to mutually independent sources of theory uncertainties. As a minimal (but not sufficient) requirement, they must parameterize $f_n(x)$ in a mathematically independent way, so all $\hat{\theta}_{n,i}$ are uniquely determined by eq. (4.1). We will come back to this distinction in section 4.4, where we discuss the parameterization dependence. In section 4.3 we discuss various strategies

[11] To put it more bluntly: it must not be left to users to figure out for themselves what to do as it happens too often right now with scale variations.

JHEP08(2025)098

for deriving suitable parameterizations satisfying these requirements. Before doing so, we first discuss which dependencies we actually need to parameterize in section 4.2 next.

4.2 Correlation requirements

The first question we have to ask ourselves is which parts of the internal structure of f_n we actually need to account for, i.e., which of its internal x dependencies we need to explicitly parameterize. The answer is that it depends on our usage requirements: the dependencies we have to parameterize are in one-to-one correspondence with the correlations we are required to take into account. To see this, we will discuss three different cases:

1. Predictions not requiring x dependence
2. Predictions requiring x dependence without correlations
3. Predictions requiring x dependence with correlations

4.2.1 Predictions not requiring x dependence

In this case 1), we only require predictions for which the x dependence is effectively not needed. There are two basic scenarios for this:

(a) We only require predictions at a given fixed value x_0. For example, we always work in QCD at fixed $N_c = 3$ or fixed $n_f = 5$. Or we only require cross sections at a fixed center-of-mass energy.

In this case, we can consider $f_n(x_0)$ as a scalar coefficient parameterized by a single TNP θ_n,

$$f_n(x_0, \theta_n) = N_n(x_0)\, \theta_n \,, \tag{4.2}$$

where $N_n(x_0)$ is a normalization factor and the true value of θ_n is given by

$$\hat{\theta}_n = \frac{\hat{f}_n(x_0)}{N_n(x_0)} \,. \tag{4.3}$$

(b) We only require predictions summed or integrated over a fixed range $[x_a, x_b]$ in x. For example, we only require a total cross section summed over all partonic channels and integrated over phase space.

In this case, we can consider the integral of $f_n(x)$ (or the sum for discrete x),

$$F_n(x_a, x_b) = \int_{x_a}^{x_b} \mathrm{d}x\, f_n(x) \,, \tag{4.4}$$

as a scalar coefficient parameterized by a single TNP θ_n,

$$F_n(x_a, x_b, \theta_n) = N_n(x_a, x_b)\, \theta_n \,. \tag{4.5}$$

Here, $N_n(x_a, x_b)$ is again a normalization factor and the true value of θ_n is given by

$$\hat{\theta}_n = \frac{\hat{F}_n(x_a, x_b)}{N_n(x_a, x_b)} \,. \tag{4.6}$$

If one of the integration limits is always fixed, say $x_a = 0$, this is the same as case (a) applied to the cumulant function $F_n(x) = \int_0^x \mathrm{d}x\, f_n(x)$.

In either case, we can choose the normalization factor N_n such that θ_n has $\mathcal{O}(1)$ natural size. It may or may not have to depend on the value x_0 or the integration limits (x_a, x_b), depending to what extent the value of x determines the natural size of $f_n(x)$.

4.2.2 Predictions requiring x dependence without correlations

In this case 2), we require predictions at several discrete values of x (e.g. at different n_f), or several bins in x or as a function of x (e.g. a binned or unbinned differential spectrum), but we do not need to have correct correlations in x. Even so, being differential in x forces us to assume some correlations in x, for which we have different options:

(a) We assume the uncertainties to be fully correlated for all x, which means we are happy to neglect any shape uncertainties in x and only care about some overall uncertainty in $f_n(x)$. We can then parameterize $f_n(x)$ in terms of the same single θ_n appearing in case 1) above, so

$$\begin{aligned} &1(a) \qquad f_n(x,\theta_n) = N_n(x_0)\,\theta_n\,\phi_n(x) \qquad \text{with} \qquad \phi_n(x_0) = 1\,, \\ &1(b) \qquad f_n(x,\theta_n) = N_n(x_a,x_b)\,\theta_n\,\phi_n(x) \qquad \text{with} \qquad \int_{x_a}^{x_b} \mathrm{d}x\,\phi_n(x) = 1\,. \end{aligned} \tag{4.7}$$

The normalization factors N_n are the same as in eqs. (4.2) and (4.5). The function $\phi_n(x)$ determines how the uncertainty is distributed over x. Its normalization condition is chosen such that θ_n parameterizes the exact same uncertainty as in cases 1(a) or 1(b) and we assume there are no shape uncertainties in x, i.e., we assume to know the shape perfectly given by $\phi_n(x)$. This also means we are explicitly giving up that eq. (4.1) holds point-by-point in x. Instead, we only require that it holds as in case 1) either at x_0 or integrated over $[x_a, x_b]$.

There are various choices we might consider for $\phi_n(x)$. For example, a constant absolute uncertainty in x is achieved by taking $\phi_n(x)$ to be a constant, $\phi_n(x) = a$. A constant relative uncertainty in x is achieved by taking $\phi_n(x) = a[\hat{f}_0(x) + \cdots + \hat{f}_{n-1}(x)\alpha^{n-1}]$. Another typical choice would be to take $\phi_n(x)$ proportional to the lowest-order x dependence, $\phi_n(x) = a\,\hat{f}_0(x)$. In either case, the proportionality constant a is fixed by the normalization condition for $\phi_n(x)$.

(b) We assume the uncertainties for some set of x values $\{x_i\}$ are fully uncorrelated. This amounts to using case 1) for each x_i with its own independent TNP $\theta_{n,i}$,

$$f_n(x_i,\theta_n) = N_n(x_i)\,\theta_{n,i}\,, \tag{4.8}$$

where $N_n(x_i)$ are again normalization factors, and $\theta_n \equiv \{\theta_{n,i}\}$ now stands for the set of $\theta_{n,i}$. The true values of the $\theta_{n,i}$ are

$$\hat{\theta}_{n,i} = \frac{\hat{f}_n(x_i)}{N_n(x_i)}\,, \tag{4.9}$$

and eq. (4.1) is now satisfied at each x_i.

JHEP08(2025)098

We can now extend this to all x by generalizing case (a) above as follows,

$$f_n(x,\theta_n) = \sum_i N_n(x_i)\,\theta_{n,i}\,\phi_{n,i}(x) \qquad \text{with} \qquad \phi_{n,i}(x_j) = \delta_{ij}\,. \tag{4.10}$$

The functions $\phi_{n,i}(x)$ now determine how the uncertainty due to $\theta_{n,i}$ is distributed away from x_i. Their form is more complicated now due to the additional requirement that they must vanish at all but one x_i. An analogous construction can be used for a set of bins instead of x values.

We stress again, that the above options do not provide correct correlations in x. They should only be used if it is known that correlations in x do not matter or in order to test whether or not this is the case.

4.2.3 Predictions requiring x dependence with correlations

In this case 3), we require x-dependent predictions as in case 2) but now with correct correlations in the uncertainties at different x. In other words, we require predictions with correct shape uncertainties in x.

In this case, we have to explicitly parameterize the correct x dependence of $f_n(x)$. In other words, $f_n(x,\theta_n)$ must parameterize the true functional form of $\hat{f}_n(x)$ such that there are true values $\hat{\theta}_n$ for which it reproduces $\hat{f}_n(x)$ exactly, i.e., eq. (4.1) is satisfied at *any* x. Clearly, this requires us to have some knowledge of the true functional form of $f_n(x)$.

When x is a discrete label, knowing the functional form in x simply means knowing the complete set of possible values $\{x_i\}$, which is basically always the case. We can then assign an independent $\theta_{n,i}$ for each $f_n(x_i)$ as in case 2(b) above.

It gets more complicated when $f_n(x)$ is a continuous function of x, which has in principle infinitely many degrees of freedom. We will discuss several strategies to deal with this situation in section 4.3 next. For the sake of our discussion here, let us consider a simple example: say we know on theoretical grounds that $f_n(x)$ is a polynomial of degree k (as is the case e.g. for $x = n_f$),

$$f_n(x) = f_{n,0} + f_{n,1}\,x + \cdots + f_{n,k}\,x^k\,. \tag{4.11}$$

The scalar coefficients $f_{n,i}$ are parameters of $f_n(x)$ with true but possibly unknown values $\hat{f}_{n,i}$. Without further information, we have to treat them as independent unknown parameters and thus parameterize each with its own TNP, $f_{n,i} = N_{n,i}\,\theta_{n,i}$, such that

$$f_n(x,\theta_n) = \sum_{i=0}^{k} N_{n,i}\,\theta_{n,i}\,x^i\,, \tag{4.12}$$

where $N_{n,i}$ are normalization factors of our choice, and the true values of the $\theta_{n,i}$ are

$$\hat{\theta}_{n,i} = \frac{\hat{f}_{n,i}}{N_{n,i}}\,. \tag{4.13}$$

The key point is that in contrast to case 2), the $\theta_{n,i}$ are now defined to be the actual parameters of the true functional form of $f_n(x)$ and therefore encode the correct correlation structure in x.

4.3 Parameterization strategies

As we have seen, to account for the correct correlations in x we have to parameterize the series coefficient $f_n(x)$ in terms of its correct underlying x dependence. There are different basic strategies for doing so, depending on how much or little we know about the true functional form of $f_n(x)$:

1. We know it well enough to be able to parameterize it explicitly in terms of a small number of parameters.

2. We know it well enough to apply strategy 1) in some well-defined limit and can perform a systematic expansion around that limit.

3. Having insufficient information for strategies 1) or 2), we can still perform an expansion in a suitable complete functional basis.

We now discuss each of these in turn.

4.3.1 Known functional form

If we know the true functional form of $f_n(x)$ well enough, we can parameterize it explicitly. In general, we can imagine $\hat{f}_n(x)$ to be some functional $\hat{\phi}_n$ of x-dependent building blocks $\hat{\phi}_{n,i}(x)$ and scalar coefficients $f_{n,i}$,

$$\hat{f}_n(x) = \hat{\phi}_n\big[\{\hat{\phi}_{n,i}(x)\}, \{\hat{f}_{n,i}\}\big] \,. \tag{4.14}$$

Knowing the true functional form of $f_n(x)$ but not the true $\hat{f}_n(x)$ itself means we know the true $\hat{\phi}_n$ and $\hat{\phi}_{n,i}(x)$ but we do not know the true values $\hat{f}_{n,i}$. We can then parameterize each coefficient $f_{n,i} = N_{n,i}\,\theta_{n,i}$ in terms of its own $\theta_{n,i}$ to obtain the TNP parameterization

$$f_n(x, \theta_n) = \hat{\phi}_n\big[\{\hat{\phi}_{n,i}(x)\}, \{N_{n,i}\,\theta_{n,i}\}\big] \,, \tag{4.15}$$

where as before $N_{n,i}$ are normalization factors of our choice, and the true values are $\hat{\theta}_{n,i} = \hat{f}_{n,i}/N_{n,i}$. A common case is that $\hat{\phi}_n$ is a linear functional, such that

$$f_n(x, \theta_n) = \sum_{i=0}^{k} N_{n,i}\,\theta_{n,i}\,\hat{\phi}_{n,i}(x) \,. \tag{4.16}$$

Common special cases are $\hat{\phi}_{n,i}(x) = x^i$ or $\hat{\phi}_{n,i}(x) = \ln^i x$ corresponding to polynomials in x or $\ln x$ of degree k.

The key point here is that we have to know the true functional form well enough to be able to write eq. (4.14) with a *finite* number of unknown parameters $f_{n,i}$. This is where we need expert knowledge on the structure of the perturbative series of the quantity f. Furthermore, we want to have a "minimal parameterization" in the following sense: we want to choose $\hat{\phi}_n$ and $\hat{\phi}_{n,i}(x)$ such that we have the minimal possible number of a priori unknown parameters $f_{n,i}$. Firstly, this means we should not introduce additional a priori known parameters. For example, we should not needlessly split the $\hat{\phi}_{n,i}(x)$ into smaller pieces at the expense (or for the purpose) of introducing additional fake parameters that would

effectively always be known and which we then pretend to be unknown. Vice versa, we also should not eliminate parameters whose true values we happen to know already but which could a priori be unknown. Instead, we should leave the decision for later whether to use our knowledge of the true value to reduce the uncertainty or not.[12] Note that even if we know in principle the allowed $\hat{\phi}_{n,i}(x)$, this strategy can fail because the number of $\hat{\phi}_{n,i}(x)$ might simply be too large to be practical.

Even a minimal parameterization is not unique. This is easily seen from the linear example in eq. (4.16). We can always choose a different independent combination of the $\hat{\phi}_{n,i}(x)$ and correspondingly use a different combination of the $\theta_{n,i}$ as the independent parameters. This parameterization ambiguity is conceptually analogous to the scheme dependence of the perturbative series discussed in section 3.4 and we will come back to it in section 4.4.

Finally, let us point out that there are cases for which the functional form in x is simple and known, particularly when x is a discrete label (e.g. the partonic channel), and for which it can be of advantage to parameterize the x dependence explicitly even if correlations in x are not required. For example, when the x dependence strongly affects the size of $f_n(x)$, it can be much easier to figure out the natural size of the individual x-independent $\theta_{n,i}$ than of some single overall θ_n.

4.3.2 Supplementary power expansion

If we can identify a suitable small parameter ε, we can perform a supplementary power expansion of $f_n(x)$ in ε,

$$f_n(x) = f_{n0}(x) + f_{n1}(x)\,\varepsilon + f_{n2}(x)\,\varepsilon^2 + \mathcal{O}(\varepsilon^3)\,. \tag{4.17}$$

Whilst we might not know the functional form of $f_n(x)$ well enough to apply strategy 1), we might know the functional form of its $f_{nl}(x)$ series coefficients well enough to apply strategy 1) for each of them. This is clear when the expansion parameter ε is related to x itself, e.g., $\varepsilon = x$ or $\varepsilon = 1 - x$. If f_{nl} is independent of x then this is simply the Taylor expansion of $f_n(x)$ around $x = 0$ or $x = 1$, but it can also be more general. A primary example is the small-p_T expansion we will employ in section 6 in which case $f_{nl}(x)$ are known to be polynomials in $\ln x$. Expanding around some point in x of course only helps us when we are actually close to that point. However, ε does not necessarily have to be x itself in order to simplify the x dependence. It may also be related to another variable y. When $f_n(x, y)$ has a nontrivial two-dimensional structure, expanding in y can simplify not just the y dependence but also the x dependence significantly, and expanding in y can be justified even when expanding in x is not.

The ε series in eq. (4.17) is conceptually completely equivalent to our perturbative series in α and we can apply exactly the same logic for treating its uncertainties. The coefficients f_{nl} are parameters of the series with true but possibly unknown values. As long as the expansion converges, we can keep the first m known coefficients, include the next k terms to parameterize the dominant uncertainties, and truncate the remaining terms since their

[12] Even with this definition of "minimal" there might be corner cases where we might debate whether a parameter is a priori known or unknown. To decide these, we just have to remember that the uncertainties not only reflect our knowledge but also our common sense.

JHEP08(2025)098

uncertainties are formally small compared to the ones we keep. For example,

$$\begin{aligned}
&\mathrm{N}^{0+1}\mathrm{LP}_\varepsilon\text{:} && f_n(x,\theta_{n0}) = f_{n0}(x,\theta_{n0})\,,\\
&\mathrm{N}^{0+2}\mathrm{LP}_\varepsilon\text{:} && f_n(x,\theta_{n0},\theta_{n1}) = f_{n0}(x,\theta_{n0}) + f_{n1}(x,\theta_{n1})\,\varepsilon\,,\\
&\mathrm{N}^{1+1}\mathrm{LP}_\varepsilon\text{:} && f_n(x,\theta_{n1}) = \hat{f}_{n0}(x) + f_{n1}(x,\theta_{n1})\,\varepsilon\,.
\end{aligned} \tag{4.18}$$

where the notation refers to the next-$(m+k)$-leading-power expansion in ε.

We stress that the primary reason for using this expansion is to provide us the formal justification and practical ability to only parameterize the leading-in-ε dependence on x for the purpose of correctly parameterizing the uncertainties in x. Doing so does not force us in any way to perform this expansion also for the known series coefficients, for which we may not want to do so.

In case we happen to know the true value of $\hat{f}_{n0}(x)$ we can include it exactly as in the $\mathrm{N}^{1+1}\mathrm{LP}_\varepsilon$ result in eq. (4.18). In this case, the ε expansion even allows us to include further information and thus reduce the uncertainties, which we would not be able to do otherwise. In fact, this is exactly a case where thanks to our approach we are able to reduce the uncertainty due to $f_n(x)$ by including partial higher-order information. Consequently, we then also have to consider whether or not to include the uncertainty due to $f_{(n+1)0}(x)$.

4.3.3 Generic basis expansion

When we do not have sufficient information to parameterize $f_n(x)$ directly or in some limit, we face the basic mathematical problem of how to best parameterize an unknown function such that we can guarantee that the $\theta_{n,i}$ have true values. We start by expanding $f_n(x)$ in a suitable complete basis $\phi_{n,i}(x)$,

$$f_n(x) = \sum_{i=0}^{\infty} f_{n,i}\,\phi_{n,i}(x)\,. \tag{4.19}$$

Thanks to the Weierstrass approximation theorem this expansion converges for polynomial bases on any bounded interval in x as long as $f_n(x)$ is continuous. This means that formally the $f_{n,i}$ are proper parameters with true values $\hat{f}_{n,i}$. This expansion is not particularly useful yet, because it has infinitely many parameters. To make it useful for our purposes, we have to truncate the series after a few terms to limit the number of parameters.

The key question then is how to justify truncating the series. In principle, we like to apply the same argument as for our perturbative series in α or the ε expansion in strategy 2, namely that the uncertainties due to the truncated terms can be neglected as small compared to the uncertainties from the terms we keep. However, this argument is harder to make now because we lack a parameter like α or ε which would allow us to control the size of the truncated terms and decide where to truncate it without knowing $\hat{f}_n(x)$. Instead, we have to rely more on experience and being able to test it on known coefficients.

Therefore, the most suitable basis we can choose is not necessarily the one which yields the best approximation for $\hat{f}_n(x)$ for a given number of terms, but rather the one for which we are *most confident* that it yields a sufficient approximation for a given number of terms. In other words, we want a basis for which we are confident that it convergences quickly with

JHEP08(2025)098

the first couple of terms to a point where we can safely neglect the remainder. Beyond that point, the remaining series may converge as slowly as it likes. The region of quick convergence should include some safety margin to allow including additional terms in case the first few terms we would keep by default get constrained too strongly.

For simplicity and without loss of generality, let us assume the relevant x range to be $x \in [-1, 1]$. Standard polynomial bases on this interval which are known to converge very fast for sufficiently smooth functions are Legendre and Chebyshev polynomials.[13] An advantage of Legendre polynomials is that they are orthogonal with respect to the unit weight function. An advantage of Chebyshev polynomials is their equioscillation property, which means that all their minima and maxima in the interval $[-1, 1]$ are at ± 1. Even if we do not know the full functional form of $f_n(x)$ we rarely know nothing about it. We can improve the convergence of the series by starting from some ansatz $\phi_n(x)$ and expanding the ratio to $f_n(x)$ to obtain the TNP parameterization,

$$\begin{aligned} \frac{f_n(x)}{\phi_n(x)} &= \sum_{i=0}^{\infty} f_{n,i}\, \phi_{n,i}(x)\,, \\ f_n(x, \theta_n) &= \phi_n(x) \sum_{i=0}^{k} N_{n,i}\, \theta_{n,i}\, \phi_{n,i}(x)\,. \end{aligned} \tag{4.20}$$

Alternatively, we can expand the difference to obtain

$$\begin{aligned} f_n(x) - \phi_n(x) &= \sum_{i=0}^{\infty} f_{n,i}\, \phi_{n,i}(x)\,, \\ f_n(x, \theta_n) &= \phi_n(x) + \sum_{i=0}^{k} N_{n,i}\, \theta_{n,i}\, \phi_{n,i}(x)\,. \end{aligned} \tag{4.21}$$

Note that $\phi_n(x)$ and $\phi_{n,i}(x)$ should be normalized suitably such that the overall size of the uncertainty and the natural size of $\theta_{n,i}$ is determined by their normalization factors $N_{n,i}$.

Another general method to accelerate the convergence is to use a variable transformation to account for some known general behaviour of $f_n(x)$. For example, if $f_n(x)$ is known to have poles or branch cuts in the complex plane, using a variable transformation that maps these to infinity can significantly improve the rate of convergence. One particular option if $\phi_n(x)$ is square-integrable and positive definite, $\phi_n(x) = |\phi_n(x)|$, is to construct a custom orthonormal basis on top of it, as was done in ref. [46] in a different context. The idea is to use the cumulant of $\phi_n(x)$ as a variable transformation as follows. Let us normalize $\phi_n(x)$ such that $\int_{-1}^{1} \mathrm{d}x\, [\phi_n(x)]^2 = 1$ and define

$$\begin{aligned} y(x) &= -1 + 2 \int_{-1}^{x} \mathrm{d}x'\, [\phi_n(x')]^2\,, \\ \phi_{n,i}(x) &= \sqrt{y'(x)}\, p_i[y(x)]\,, \qquad p_i(y) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n n!} \frac{\mathrm{d}^n}{\mathrm{d}y^n} (y^2 - 1)^n\,, \end{aligned} \tag{4.22}$$

[13] Roughly speaking, for differentiable functions that have $p - 1$ continuous derivatives and a pth derivative of bounded variation, Legendre, Chebyshev, and similar polynomial expansions converge algebraically $\sim 1/k^p$. For analytic functions they converge geometrically $\sim 1/\rho^k$ where the constant ρ depends on how far the function can be analytically continued into the complex plane. The corresponding precise mathematically theorems can be found e.g. in ref. [45].

JHEP08(2025)098

Here, $p_i(y)$ are normalized Legendre polynomials, which are orthonormal on $y \in [-1, 1]$, and thus $\phi_{n,i}(x)$ are orthonormal on $x \in [-1, 1]$. Since $\sqrt{y'(x)} = \sqrt{2}\phi_n(x)$, the 0th basis function $\phi_{n,0}(x) = \phi_n(x)$ itself, while the higher basis functions are orthogonal polynomial modulations on top of it. As a result, if $\phi_n(x)$ captures the overall shape of $f_n(x)$, the series is expected to converge much more rapidly than expanding in $p_i(x)$ directly, at least for the first few terms until the detailed shape starts to matter, which is all we really want.

Hence, the key to finding a suitable basis in the above sense is to start from a suitable approximation $\phi_n(x)$. Importantly, the goodness of this approximation is not a fundamental limitation, as it only serves in one or another way as a starting point for a complete expansion. There are various ways we can imagine choosing $\phi_n(x)$:

- Pick $\phi_n(x) = \hat{\phi}_n(x)$, or more generally $\phi_n(x) = \hat{\phi}_n(x, \theta_n)$, where $\hat{\phi}_n(x)$ encodes some known aspect of the true functional form of $f_n(x)$, e.g., it has (or parameterizes) the correct asymptotic behaviour or the correct poles. This essentially supplements strategy 1) in case the information we have is not sufficient for using it standalone.

- Pick $\phi_n(x) = f_{n0}(x, \theta_{n0})$ or $\phi_n(x) = \hat{f}_{n0}(x)$, where $f_{n0}(x)$ is the leading-power limit from strategy 2) which either has a simpler known x dependence or is fully known. This essentially supplements strategy 2) in case it cannot be used standalone, e.g., when the $\varepsilon \to 0$ limit is known but the expansion itself is not or does not apply to all x.

- Use the known lower-order shape $\phi_n(x) = a\hat{f}_0(x)$ or $\phi_n(x) = a[\hat{f}_0(x) + \cdots + \hat{f}_{n-1}(x)\,\alpha^{n-1}]$, with a determined by the appropriate normalization condition on $\phi_n(x)$. This essentially supplements case 2) in section 4.2.2. For the multiplicative case it effectively expands the K factor, which makes sense whenever we are confident that $f_n(x)/\hat{f}_0(x)$ is much flatter in x than $f_n(x)$ itself.

- Use some approximation $\phi_n(x) \approx \hat{f}_n(x)$, which is known to work in similar cases, e.g. a Padé approximation. This supplements any ad hoc approximation method, extending it into a formally complete parameterization.

Before concluding, let us comment that one might naively think that having to know or parameterize the functional form of $f_n(x)$ is a drawback of our approach. It is not. It is simply a necessity for obtaining the correct correlation structure in x. In practice, we almost always have some, even if limited, information about the functional form and it is in fact a key advantage of our approach that all information we have can be systematically incorporated. In contrast, with uncertainties derived from scale variations $f_n(x)$ is silently modelled by some linear combination of lower-order coefficients, see eq. (2.11). If this is indeed believed to be a sufficient correlation model, one can always use the lower-order coefficients to construct the ansatz $\phi_n(x)$ as mentioned in the third bullet point above. This is still much better than scale variations, because it provides explicit control over the assumptions made and furthermore provides a systematic extension to a formally complete parameterization.

4.4 Parameterization dependence

Regardless of the strategy used to derive the TNP parameterization, a given parameterization is never unique. For example, we can always choose some linearly independent combination of

JHEP08(2025)098

the $\theta_{n,i}$ as new independent parameters. The ambiguity in the choice of the parameterization is conceptually analogous to the scheme dependence of the perturbative series, and our discussion here will resemble much of the discussion in section 3.4.

Let us therefore start with an executive summary: it is important to distinguish the uncertainty and correlation *structure*, which is unique and correctly encoded by any valid parameterization, from the actual *values* of the uncertainties and correlations, which are determined by whatever constraints we choose to impose on the parameters. Before any constraints they are simply unknown, which means their uncertainties are infinite and their correlations do not matter, which is a parameterization independent statement. When the parameters are solely constrained by data or other parameterization-independent constraints, the uncertainties and correlations reflect the combined uncertainties and correlations due to all constraints in a parameterization-independent way. A parameterization-dependent bias is only induced when we impose parameterization-dependent constraints.

To discuss the possible parameterization dependence in more detail, let us denote by $f_n(x, \theta_n)$ our default parameterization and by $f'_n(x, \theta'_n)$ some alternative parameterization. For the sake of discussion let us also consider them to still be exact, so before truncating the ε series in strategy 2) or the basis expansion in strategy 3). Different parameterizations must then be equal by definition,

$$f_n(x, \theta_n) = f'_n(x, \theta'_n)\,, \tag{4.23}$$

as they both parameterize the same function $f_n(x)$ and reproduce the same true value $\hat{f}_n(x)$. It follows that from eq. (4.23) the θ'_n are uniquely determined in terms of the θ_n (and vice versa) in exactly the same way the true $\hat{\theta}_n$ and $\hat{\theta}'_n$ are uniquely determined by eq. (4.1).

Note that in principle there could be more $\theta'_{n,i}$ than $\theta_{n,i}$ parameters. If so, it would imply that $f_n(x, \theta_n)$ contains more information on the true functional form in x than $f'_n(x, \theta'_n)$. So from the point of view of $f_n(x, \theta_n)$ some of the $\theta'_{n,i}$ are either known or not independent. Let us therefore assume that both parameterizations are based on the same information and thus have the same number of parameters.

Any valid parameterization, meaning it satisfies eq. (4.1), encodes the correct theory uncertainty and correlation structure. The θ_n and θ'_n play the role of different but related input parameters. If we treat them as unknown parameters to be determined from data, it does not matter at all which one we choose, since there is an exact relation between them. Any constraints imposed by the data are parameterization independent as they always constrain $f_n(x)$, so they respect the relation between θ_n and θ'_n implied by eq. (4.23).

A dependence on the parameterization (only) appears when we make parameterization dependent assumptions, i.e., by imposing theory constraints on the θ_n of a specific parameterization. This also includes the assumption of their mutual independence, which only enters when we choose to impose independent (uncorrelated) theory constraints on them. Doing so for the $\theta_{n,i}$ implies in general some nontrivial correlated uncertainties for $\theta'_{n,i}$ (and vice versa). The point is that the condition for some parameters to be *mathematically* independent is only a necessary but not sufficient condition for them to be *conceptually* independent, i.e. to correspond to independent sources of uncertainties, and thus to be a

JHEP08(2025)098

priori uncorrelated. Whenever we talk about the $\theta_{n,i}$ being mutually independent we really refer to their conceptual independence.

To illustrate this with a simple example, say we know $f_n(x)$ to be a kth-order polynomial. Consider the two equivalent parameterizations

$$\begin{aligned} f_n(x,\theta_n) &= \theta_{n,0} + \theta_{n,1}\,x + \cdots + \theta_{n,k}\,x^k\,, \\ f'_n(x,\theta'_n) &= \theta'_{n,0} + \theta'_{n,1}\,(1-x) + \cdots + \theta'_{n,k}\,(1-x)^k\,. \end{aligned} \tag{4.24}$$

By setting them equal, we can easily derive the exact relation between θ'_n and θ_n, e.g.,

$$\theta'_{n,0} = \theta_{n,0} + \theta_{n,1} + \cdots + \theta_{n,k}\,, \qquad \theta'_{n,1} = -\theta_{n,1} - 2\,\theta_{n,2} - \cdots - k\,\theta_{n,k}\,, \tag{4.25}$$

and so on. A fit to data always chooses the kth-order polynomial that best fits the data, regardless of the specific parameterization, with the post-fit uncertainties and correlations of the parameters reflecting the uncertainties and correlations of the fitted measurements in a parameterization independent way. On the other hand, imposing a theory constraint that the $\theta_{n,i}$ have mutually uncorrelated uncertainties of $\Delta u_{n,i} = 1$ yields an uncertainty for $f_n(x=0)$ of 1 and for $f_n(x=1)$ of $\sqrt{k}$. On the other hand, imposing the same constraint on the $\theta'_{n,i}$ yields an uncertainty for $f_n(x=0)$ of $\sqrt{k}$ and for $f_n(x=1)$ of 1.

Hence, the choice of parameterization in principle induces a bias in the uncertainties and correlations *if* we let it determine which parameters to impose independent theory constraints on. Therefore, we should not choose the independent parameters based on the parameterization, but rather the other way around. We should choose a parameterization for which we are most confident that its $\theta_{n,i}$ can be considered to correspond to independent sources of uncertainty. Furthermore, we can avoid a parameterization bias by imposing theory constraints at the level of $f_n(x)$ itself. For example, we should always choose the central value directly for $f_n(x)$, which by default can just be $f_n(x) = 0$. This is a parameterization-independent condition on the central values of θ_n or θ'_n and thus avoids any parameterization bias in the central value. Similarly, we can impose for example an uncertainty based on the natural size of the integral of $f_n(x)$ or its value at special points. Ultimately, however, this is just another way of choosing what we consider to be the independent sources of uncertainty. The TNPs can only help us to parameterize the independent sources of uncertainty once we have identified them. They cannot decide for us what they are. As soon as we want or need to impose theory constraints we cannot avoid making this decision. This is another place where clearly domain knowledge is required. What we can avoid though is to make an implicit or uninformed decision. If we do not have enough information to decide, we have to limit ourselves to unambiguous parameterization-independent constraints.

Our above discussion implies for strategy 3) that the choice between different linearly related bases (polynomial or otherwise) is actually irrelevant when fitting to data (apart from effects due to numerical stability etc.) or imposing other parameterization-independent constraints. It is only relevant for arguing where we are allowed to truncate and perhaps for arguing which parameters we should consider to be independent. The truncation itself does induce a parameterization-dependent bias, which however can be phrased in terms of the above discussion: we can always think of it as imposing a theory constraint on the parameters of the truncated terms that their central value vanishes with an uncertainty whose net effect we can neglect.

JHEP08(2025)098

4.5 Multiple dependencies

So far, we have assumed that the x-independent coefficients $f_{n,i}$ are scalars so we can parameterize them by a scalar nuisance parameter. This is no longer the case when f_n depends on multiple internal variables whose dependence we are required to parameterize, i.e., which fall into cases 2) or 3) in section 4.2. To generalize our discussion to this situation, it is sufficient to discuss how to extend from the one-dimensional case $f_n(x)$ to the two-dimensional case $f_n(x, y)$. The generalization to further variables then proceeds in exactly the same way.

When we assemble the final prediction from its ingredients there is typically a natural progression of the dependencies from the inner layers to the outer layers, which we can also follow here. For example, the innermost layer could be the color and n_f dependence, then comes the kinematic dependence, next we sum over partonic channels, and eventually at the outermost layer we sum or combine different processes.

To be concrete, let us denote the innermost relevant variable by x and the next outer variable by y. To parameterize the y dependence, we can follow the same strategies discussed in the previous subsections. The only difference is that once the y dependence is stripped away, the y-independent coefficients $f_{n,j} \equiv f_{n,j}(x)$ are not scalars but still functions of x. Each of them we can then parameterize in x in terms of scalar parameters as we have discussed so far for $f_n(x)$. For example, if y is a discrete variable or only needed at fixed values we simply have $f_{n,j}(x) = f_n(x, y_j)$.

The only more complicated case is when x and y are both continuous and appear at the same layer, e.g., when considering a double-differential spectrum in two kinematic variables. If their dependence is separable, $f_n(x, y) = f_n(x)g_n(y)$, we can treat each one-dimensional factor as before. Finally, when we have a genuinely two-dimensional function $f_n(x, y)$ and require correlations in both x and y, we need to parameterize the x and y dependencies simultaneously, for which we can follow the two-dimensional generalization of the strategies in section 4.3.

For strategy 1), we have to consider two-dimensional basic building blocks $\hat{\phi}_{n,ij}(x, y)$. Clearly, finding a minimal parameterization of the true functional form is going to be more difficult now, but it can still be possible if the x and y dependence is separable or if it can be reduced to several one-dimensional functions which only depend on certain combinations of x and y.

For strategy 2), the ε expansion coefficients $f_{nl}(x, y)$ are in general two-dimensional now. This strategy can be quite powerful to make the two-dimensional case more tractable. By expanding in ε, we might be able to simplify one or both dependencies or make them separable or otherwise reduce the problem to the one-dimensional case.

For strategy 3), we have to consider two-dimensional functional bases $\phi_{n,ij}(x, y)$. The approximation of multivariate functions is surprisingly more difficult than the univariate case, and an active area of mathematical research. Finding suitable multivariate parameterizations for an unknown multivariate function is a similarly difficult problem. However, it is ultimately necessary to correctly account for a genuinely multidimensional correlation structure if we lack the ability to apply strategies 1) or 2). The most straightforward is to consider a product basis $\phi_{n,ij}(x, y) = \phi_{n,i}(x)\phi_{n,j}(y)$, which simply amounts to expanding $f_n(x, y)$ in $\phi_{n,j}(y)$ for fixed x and then further expanding the resulting x-dependent series coefficients in $\phi_{n,i}(x)$. Unfortunately, the number of terms quickly proliferates — the curse of dimensionality.

JHEP08(2025)098

However, we can often identify a primary variable x and a secondary variable y, whose correlations might matter less or which is going to be integrated over first. In this case we can mitigate the curse of dimensionality by optimizing the basis in favor of x.

4.6 Examples

In this subsection, we discuss various dependencies to illustrate the general discussion of the previous subsections.

4.6.1 N_c dependence and color structure

When we are only interested in QCD corrections, the dependence on the number of colors, N_c, is an example of case 1) in section 4.2: we always have fixed $N_c = 3$ and do not require correlations between different values of N_c. This means we do not need separate $\theta_{n,i}$ for individual color coefficients but only a single overall one for $f_n(N_c = 3)$.

Nevertheless, if we were to parameterize the N_c dependence, it is a good example for strategy 1) where the functional form is fully known, as we know exactly which color coefficients composed of C_A, C_F, T_F, as well as higher invariants, appear for a given f_n.

When considering QCD and QED corrections, we still do not need the full N_c-dependent structure but effectively two pieces of it. The abelian parts of the QCD coefficients are clearly correlated with the QED coefficients. To correctly account for this correlation we have to separate the abelian and nonabelian parts of the QCD color structure and parameterize each with a separate TNP. The abelian one will then be shared by the QCD and QED coefficients, whereas the nonabelian one only appears in the QCD coefficients. In this way, the partial correlation between QCD and QED coefficients is correctly accounted for.

An analogous discussion applies to QCD and electroweak corrections at sufficiently high energies where the masses m_V of the electroweak gauge bosons can be neglected. When the gauge boson masses cannot be neglected it requires a more detailed investigation to identify the possibly common parts. Effectively one has to consider in addition the dependence on m_V at two fixed points, namely at the physical value of m_V and in the $m_V \to 0$ limit.

4.6.2 n_f dependence

When the number of flavors, n_f, is the same in all considered predictions, as is often the case with $n_f = 5$, we are in case 1) and do not require correlations in n_f and only a single TNP for $f_n(n_f = 5)$. When we do cross flavor thresholds and require $f_n(n_f)$ at different n_f values, we do need to parameterize the n_f dependence to account for the (de)correlation between say $n_f = 5$ and $n_f = 4$.

The n_f dependence is another example where strategy 1) is easily applicable, since $f_n(n_f)$ is a polynomial in n_f of known degree. This actually poses an interesting theoretical question, namely which parts of the n_f dependence are conceptually independent. Neither the naive choice to consider the coefficients of n_f^i as independent nor rewriting n_f in terms of $\beta_0(n_f)$ and considering the coefficients of $\beta_0(n_f)^i$ as independent seem to be supported by empirical evidence. Instead, empirical evidence suggests that the coefficients of $(C_A - T_F n_f)^i$ are independent. This can likely be attributed to the screening effect of quarks, see section 5.2 for some further discussion.

JHEP08(2025)098

4.6.3 Partonic channels

The dependence on different partonic channels is a primary example of a genuinely discrete dependence. Here, strategy 1) is immediately applicable, as we know exactly which partonic channels appear at a given order, and it amounts to separately parameterizing each partonic channel.

One might ask the question when we are actually required to separate the partonic channels. One reason is when we require hadron-collider predictions at different center-of-mass energies, E_{cm}, since the E_{cm} dependence enters via the different parton luminosities for each channel, which can have very different scaling with E_{cm}. Another reason is to capture correlations between different processes that share common partonic channels, see below.

Another important reason to separately parameterize partonic channels is to anticipate new channels that only open up at higher orders but can have sizeable contributions, which is a classic case where scale variations can fail badly. This is an example where parameterizing the dependence, even if not required for correlations, can be of advantage for figuring out the natural size of the TNPs.

4.6.4 Process dependence

Another type of correlation is that between different processes. This tends to be a more complicated dependence to take into account as it requires detailed knowledge of the internal structure. To correctly correlate the process dependence we basically have to map it into the dependence on some internal variables x. Luckily, the most relevant cases, namely closely related processes expected to be strongly correlated, are also the most straightforward. For example, for W vs. Z production, the process dependence essentially maps into the dependence on partonic channels and electroweak gauge couplings and boson masses. We will see an explicit example in section 6.

4.6.5 Continuous dependencies

A typical example of a genuinely continuous dependence is that of a differential spectrum. We will discuss the example of the q_T spectrum in detail in section 6, which is going to involve a repeated application of strategies 1 and 2. Another generic example is the dependence on the partonic momentum fractions $z_{a,b}$ of partonic cross sections in hadronic collisions. Here, if we are only asking about a total cross section, the z_a and z_b dependence is effectively projected onto a single number. If we consider a kinematic distribution that effectively measures the total invariant mass Q of the hard process, then we need the one-dimensional dependence on $z = z_a z_b$. Finally, if we are sensitive to both the total invariant mass and rapidity of the hard process we need the full dependence on z_a and z_b. Suitably parameterizing this dependence is in general nontrivial. Often though, the cross section tends to be dominated by the $z \to 1$ limit, which can be a good starting point by applying strategy 2 with $\varepsilon = 1 - z$. This strategy has already proven very useful in other cases where the dependence on partonic momentum fractions arises, namely to parameterize the unknown parts of beam function matching kernels [47] or QCD splitting functions [13] in terms of TNPs.

JHEP08(2025)098

5 Theory constraints for scalar series

In this section, we discuss TNPs for perturbative series with scalar coefficients f_n and how to obtain robust theory constraints on them, which belongs to the second step of applying the TNP approach as discussed in section 3.1.

We assume that the parameterization of any relevant outer levels of x dependences as discussed in section 4 has happened and has reduced the remaining perturbative series to have scalar coefficients f_n, as will be relevant for the application in section 6. We limit ourselves to QCD corrections at fixed $n_f = 5$. Further investigations beyond this case are of course warranted but are well beyond our scope here and are left to future work.

Hence, the starting point for our discussion in this section is that we have a QCD series in α_s with scalar coefficients f_n that can be parameterized by a single theory nuisance parameter,

$$f_n(n_f = 5, \theta_n) = N_n(n_f = 5)\, \theta_n \,. \tag{5.1}$$

To simplify the notation, we will suppress the $n_f = 5$ argument from here on. The normalization factor N_n accounts for the expected natural size of f_n, i.e., it should be chosen such that we generically expect $|\hat{f}_n| \lesssim N_n$. Consequently, the expected natural size of θ_n is $|\hat{\theta}_n| \lesssim 1$.

5.1 Overview

Not knowing the true value $\hat{\theta}_n$ of θ_n, our goal is to obtain an estimate as in eq. (3.4),

$$\theta_n = u_n \pm \Delta u_n \,, \tag{5.2}$$

based on theoretical arguments. This will be our baseline theory constraint on θ_n, which we use to evaluate the theory uncertainty in the absence of any additional constraint from other sources of information.

Without additional information we will usually just take $u_n = 0$ as our best-guess central value. We then need to assign an uncertainty Δu_n to this choice, which determines the amount by which we vary θ_n and thus the size of the resulting theory uncertainty. When we need a statistical treatment of eq. (5.2), we also need the probability distribution $P(u_n|\theta_n)$ of u_n. For this purpose, we treat u_n as if it came from a measurement with a Gaussian 1σ uncertainty of Δu_n. More precisely, we model our estimator u_n for θ_n as a Gaussian-distributed random variable with mean $\mu = \theta_n$ and standard deviation $\sigma = \Delta u_n$. This is a standard assumption also used for nuisance parameters of experimental systematic uncertainties, whose justification basically stems from the central-limit theorem. In section 5.3, we will find strong empirical evidence that u_n can indeed be considered as a Gaussian-distributed random variable. We will thus refer to the theory uncertainties that result from varying a theory constraint by $\pm\Delta u_n$ as one "theory-σ" uncertainty or 68% "theory CL". Similarly, 95% theory CL refers to varying by $\pm 2\Delta u_n$.

Following our discussion in section 2.1, Δu_n is not given by the distance $|\hat{\theta}_n - u_n|$ of our estimate u_n to the true value $\hat{\theta}_n$. Thus, to estimate Δu_n we *do not* need to estimate a precise value of $\hat{\theta}_n$. (Our best guess for $\hat{\theta}_n$ is already represented by u_n). Rather, Δu_n must reflect our limited knowledge. With the above statistical interpretation this means we need to choose Δu_n such that $|\hat{\theta}_n - u_n| \leq \Delta u_n$ with 68% confidence. For $u_n = 0$, Δu_n is thus determined by the natural size of θ_n, $|\hat{\theta}_n| \lesssim \Delta u_n$, and so with our choice of normalization we have $\Delta u_n \simeq 1$.

JHEP08(2025)098

If we believe to know absolutely nothing about f_n, it would imply to take $\Delta u_n = \infty$ so θ_n would be left to vary unconstrained within $[-\infty, \infty]$. In other words, we would treat θ_n as a truly unknown parameter to be determined from data. In many cases, this is of course too pessimistic as we do have some expectations and plenty of experience of the typical size of higher-order corrections. Therefore, to choose an appropriate Δu_n we proceed in two steps: in the first step in section 5.2, we use theoretical arguments to estimate the expected natural size of f_n. That is, we determine the normalization N_n for which we expect $|\hat{f}_n| \lesssim N_n$ so $|\hat{\theta}_n| \lesssim 1$ and $\Delta u_n \simeq 1$. Based purely on theoretical expectations we can only hope to narrow Δu_n down to an $\mathcal{O}(1)$ factor, perhaps a factor of two at best. Therefore, in the second step in section 5.3 we study the true values $\hat{\theta}_n$ of many known series of a common category. This will provide us with the empirical evidence to verify and further narrow down the value of Δu_n and also to confirm its statistical interpretation in terms of the probability distribution $P(u_n|\theta_n)$.

5.2 Normalization and estimate of natural size

We consider two general categories of perturbative quantities. The first are quantities corresponding to the finite constant terms of matrix elements, which we refer to as matrix-element "constants" and for which we continue to use the generic notation $f(\alpha_s)$. This includes total cross sections and decay rates as well as the constant (nonlogarithmic) terms (RG boundary conditions) of matching coefficients and matrix elements of renormalized operators. The second type are anomalous dimensions, denoted generically as $\gamma(\alpha_s)$, which correspond to the coefficients of $1/\varepsilon$ poles in the bare perturbative series. We distinguish these two categories because we expect and find their perturbative series to behave somewhat differently.

5.2.1 Matrix-element constants

We write the perturbative series for matrix-element constants as

$$f(\alpha_s) = 1 + \sum_{n=1} f_n \left(\frac{\alpha_s}{4\pi}\right)^n , \tag{5.3}$$

which defines their perturbative coefficients f_n. We normalize all quantities such that their leading-order result is $f_0 = 1$, since it only contains overall couplings and prefactors, which are always known, and so does not yet contain nontrivial information about the perturbative series. We choose the normalization N_n^f to parameterize f_n in terms of θ_n^f as

$$f_n(\theta_n^f) = N_n^f\, \theta_n^f \qquad \text{with} \qquad N_n^f = 4^n C_n (n-1)! \,. \tag{5.4}$$

Here, $C_n = C_r C_A^{n-1}$ is the leading color factor of f_n with C_r the one-loop color factor, which depends on the color representation of the external particles, i.e., $C_r = C_F$ for external quarks and $C_r = C_A$ for external gluons. Note that we merely use the leading-color limit to determine the normalization. We do not make a leading-color approximation anywhere. As discussed in section 4.6, we do not need to parameterize the full color structure of the coefficients because here we are only interested in QCD and fixed values of N_c. We will explain the other factors in a moment.

The above discussion applies to tree-level quantities. Considering quantities that are inherently loop induced, their overall normalization is defined to be consistent with that

JHEP08(2025)098

of an associated tree-level quantity. Typical examples would be an off-diagonal partonic channel that has an associated diagonal partonic channel, or a singlet coefficient that has an associated nonsinglet coefficient. Their leading n-loop color factor is then given by the color factor of the first appearing loop order times one power of C_A for each additional loop order.

As an instructive example to understand this choice of N_n^f, let us consider the $q\bar{q}$ vector, $q\bar{q}$ scalar, and gg matching coefficients, corresponding to the infrared-finite parts of the respective QCD form factors, which are known to four loops [48, 49]. The perturbative series of their respective constant terms are denoted as $c_{q\bar{q}V}(\alpha_s)$, $c_{q\bar{q}S}(\alpha_s)$, and $c_{gg}(\alpha_s)$. (The $c_{q\bar{q}V}(\alpha_s)$ coefficient is defined in more detail in section 6.2.) In tables 2, we show the true values $\hat{f}_n/N_n$ for all three matching coefficients successively dividing out the normalization factor N_n^f:

- The first line in each block shows the raw values for $\hat{f}_n$, which grow very large for increasing n. Naively, there would be little hope to directly estimate the correct expected size of these numbers.

- In the second lines, we divide out a factor of 4^n, which basically removes the $1/4^n$ in eq. (5.3). It is clear that the conventional $1/(4\pi)^n$ loop factor is artificial in this regard and a main reason for the quickly increasing magnitude of the coefficients. We could of course have directly expanded eq. (5.3) in terms of α_s/π, which is actually known to be a more appropriate expansion parameter. The reason we did not do so is for the sake of illustration here and because defining the series coefficients with respect to $\alpha_s/(4\pi)$ is the most commonly used convention. Nevertheless, the resulting numbers are still far from $\mathcal{O}(1)$.

- In the third lines, we further divide out the leading color factor $C_r C_A^n$ appearing at n-loop order, which brings the numbers to $\mathcal{O}(1)$ as we might expect.

- Finally, in the fourth and last line, we further divide out a factor of $(n-1)!$, which amounts to $\{1, 1, 2, 6\}$ for $n = \{1, 2, 3, 4\}$ and which is clearly still present in $\hat{f}_3/N_3$ and $\hat{f}_4/N_4$ in the previous line. The appearance of this factor also matches our expectation of the factorial growth of the series coefficients.

The last line in each block in tables 2 corresponds to the nominal N_n^f in eq. (5.4) with the numbers in bold corresponding to the true values $\hat{\theta}_n^f$, which indeed satisfy $|\hat{\theta}_n^f| \lesssim 1$ to well within a factor of two as desired. The above arguments leading to this choice of N_n^f are generic and not specific to the given examples. Hence, we consider it as a very plausible generic expectation for the natural size of f_n, and consequently we can consider $\Delta u_n \simeq 1$ as a plausible uncertainty.

This exercise already teaches us several interesting things and dispels some common lore. First, gluonic quantities do not necessarily have genuinely larger perturbative corrections than quark ones. Once the different overall color factor of $C_r = C_A$ vs. $C_r = C_F$ is accounted for, the remaining normalized coefficients for C_{gg} have the same generic size as those for $C_{q\bar{q}V}$ and $C_{q\bar{q}S}$. In fact, one of the latter is always larger than C_{gg} at each order. Secondly, once normalized to their natural size, the specific size of the coefficient(s) of previous order(s) is not a good indicator for the size of the coefficient(s) at the following order(s). In other words,

JHEP08(2025)098

$f(\alpha_s)$	N_n	$\hat{f}_1/N_1$	$\hat{f}_2/N_2$	$\hat{f}_3/N_3$	$\hat{f}_4/N_4$
$c_{q\bar{q}V}(\alpha_s)$	1	−8.47	−48.6	−1387	−42015
	4^n	−2.12	−3.04	−21.7	−164
	$4^n C_F C_A^{n-1}$	−1.59	−0.76	−1.81	−4.56
	$4^n C_F C_A^{n-1}(n-1)!$	−1.59	−0.76	−0.90	−0.76
$c_{q\bar{q}S}(\alpha_s)$	1	−0.47	+87.1	+2309	+76100
	4^n	−0.12	+5.44	+36.1	+297
	$4^n C_F C_A^{n-1}$	−0.09	+1.36	+3.01	+8.26
	$4^n C_F C_A^{n-1}(n-1)!$	−0.09	+1.36	+1.50	+1.38
$c_{gg}(\alpha_s)$	1	+4.93	−24.0	−4066	−123979
	4^n	+1.23	−1.50	−63.5	−484
	$4^n C_A C_A^{n-1}$	+0.41	−0.17	−2.35	−5.98
	$4^n C_A C_A^{n-1}(n-1)!$	+0.41	−0.17	−1.18	−1.00

Table 2. True values of the series coefficients $\hat{f}_n$ divided by various normalization factors N_n for the quark vector (top block), quark scalar (middle block), and gluon (bottom block) matching coefficients. The numbers in bold in the last line of each block are the $\hat{\theta}_n^f$.

one should not look at this table from left to right but only from top to bottom. Thirdly, the coefficients are not always or mostly positive and may change sign at different orders.

The convention to have $f_0 = 1$ does not yet uniquely fix the overall convention for $f(\alpha_s)$, as we could still raise $f(\alpha_s)$ to some power, which keeps $f_0 = 1$, but changes the f_n. For example, by squaring a series with all $f_n = 1$ we get

$$\left[1 + \sum_{n=1} \alpha^n\right]^2 = 1 + \sum_{n=1} (n+1)\, \alpha^n \,. \tag{5.5}$$

Therefore, by taking the square or square root of $f(\alpha_s)$, the natural size of f_n can change by an $\mathcal{O}(n)$ factor, which we clearly have to account for if we aim for an estimate to within a factor of two or better. We find the normalization in eq. (5.4) to be appropriate for the convention that $f(\alpha_s)$ is raised to an appropriate power such that it effectively scales as a matrix element with two (resolved or Born-level) external QCD partons, as is the case for the matching coefficients considered above, or equivalently a squared matrix element with a single (resolved or Born-level) external QCD parton. This means we consider jet and beam functions as they are, since they can be regarded either as forward $1 \to 1$ matrix elements or 1-parton squared matrix elements. On the other hand, we consider the square root of $0 \to 2$ cross sections and decay rates and also of soft functions with two Wilson lines. We might argue that this is also natural from the point of view of identifying the conceptually independent perturbative corrections, since they fundamentally appear for the matrix element and not its square. A typical example is a large correction to the NLO matrix element whose square then also causes a large NNLO correction to the cross section. By considering the square root of the cross section, this is effectively accounted for.[14]

[14] In the future, instead of just taking the square root for cross-section-like quantities, it might be worth to

JHEP08(2025)098

Finally, there is one more subtlety to consider. The attentive reader might have wondered that the factorial factor in eq. (5.4) is $(n-1)!$ and not just $n!$. As it turns out, an $n!$ factor is indeed appropriate in the pure gauge theory with $n_f = 0$. The factorial growth can be attributed to bubble chains inserted into gluon propagators. Including fermions, the leading n_f dependence comes from replacing a gluon bubble by a fermion loop and generically appears as $C_A - T_F n_f$, which vanishes for $n_f = 6$ (for QCD with $C_A = N_c = 3$ and $T_F = 1/2$). Empirically, we find that this quark screening effect reduces the size of the corrections by $1/n$ turning the $n!$ behaviour for $n_f = 0$ into $(n-1)!$ for $n_f \simeq 6$. This applies when the n_f dependence starts at $n = 2$. In some (but not all) cases with external gluons, the n_f dependence starts at $n = 1$, in which case n must be increased by one in the factorial factor, so we would use $n!$ in eq. (5.4) for $n_f = 5$.

5.2.2 Anomalous dimensions

We write the perturbative series for anomalous dimensions as

$$\gamma(\alpha_s) = \sum_{n=0} \gamma_n \left(\frac{\alpha_s}{4\pi}\right)^{n+1}, \tag{5.6}$$

which defines their coefficients γ_n. Their overall normalization is less obvious than for $f(\alpha_s)$ since they start at loop-level, so γ_0 appears at the same order as f_1 and already contains nontrivial perturbative information. The anomalous dimensions correspond to logarithmic μ derivatives of matrix elements so the ambiguity of raising $f(\alpha_s)$ to some power corresponds to multiplying $\gamma(\alpha_s)$ by some overall factor. We therefore decide to fix the normalization convention for $\gamma(\alpha_s)$, including its overall sign, analogous to that of $f(\alpha_s)$ so it corresponds to the anomalous dimension of some $f(\alpha_s)$, i.e., the derivative with respect to $\ln \mu$ of a matrix element with two external QCD partons.

We then choose the normalization N_n^γ to parameterize γ_n in terms of θ_n^γ as

$$\gamma_n(\theta_n^\gamma) = N_n^\gamma \, \theta_n^\gamma \qquad \text{with} \qquad N_n^\gamma = 4^{n+1} C_{n+1}\,, \tag{5.7}$$

where C_{n+1} is again the leading $(n+1)$-loop color factor, typically given by $C_{n+1} = C_r C_A^n$ with C_r the one-loop color factor determined by the color representation of the external legs. To motivate this normalization, we show the known true values for a few anomalous dimensions in tables 3 successively dividing out the normalization factor N_n^γ. We find a quite similar pattern as before for the constants f_n:

- The first line in each block shows the raw values for $\hat{\gamma}_n$, which quickly grow large as n increases. There would again be little hope to directly estimate the size of these numbers.

- In the second lines, we divide out a factor of 4^{n+1}, which removes the $1/4^{n+1}$ in eq. (5.6). We see again that the conventional $1/(4\pi)^{n+1}$ loop factor artificially enlarges the size of the coefficients.

investigate the option of directly parameterizing and estimating the real and imaginary parts of the underlying complex amplitude. This is clearly more challenging due to the presence of IR divergences and also because in the literature perturbative results are often provided only for the cross section.

JHEP08(2025)098

$\gamma(\alpha_s)$	N_n	$\hat{\gamma}_0/N_0$	$\hat{\gamma}_1/N_1$	$\hat{\gamma}_2/N_2$	$\hat{\gamma}_3/N_3$	$\hat{\gamma}_4/N_4$
β	1	−15.3	−77.3	−362	−9652	−30941
	4^{n+1}	−3.83	−4.83	−5.65	−37.7	−30.2
	$4^{n+1}C_F C_A^n$	−1.28	−0.54	−0.21	−0.47	−0.12
γ_m	1	−8.00	−112	−950	−5650	−85648
	4^{n+1}	−2.00	−7.028	−14.8	−22.1	−83.6
	$4^{n+1}C_F C_A^n$	−1.50	−1.76	−1.24	−0.61	−0.77
$2\Gamma^q_{\text{cusp}}$	1	+10.7	+73.7	+478	+282	(+140000)
	4^{n+1}	+2.67	+4.61	+7.48	+1.10	(+137)
	$4^{n+1}C_F C_A^n$	+2.00	+1.15	+0.62	+0.03	(+1.27)

Table 3. True values of the series coefficients $\hat{\gamma}_n$ divided by various normalization factors N_n for the QCD β function [50–56], the quark-mass anomalous dimension [57–62], and the quark cusp anomalous dimension [63–66]. The numbers in bold in the last line of each block are the $\hat{\theta}_n^\gamma$. The 5-loop result for the quark cusp anomalous dimension [67] in brackets is only known approximately.

- In the third lines, we further divide out the leading n-loop color factor, which yields numbers that are $\lesssim 1$ within a factor of two.

In contrast to the matrix-element constants, no factorial factor appears for the anomalous dimensions, which is not entirely unexpected. However, we still find that for $n_f = 0$ the coefficients are enhanced by a factor of n due to the absence of the quark screening compared to $n_f \simeq 6$. Also, the sign of the higher-order coefficients now tends to be determined by the sign of γ_0 for $n_f \leq 5$, while for $n_f = 6$ the coefficients do change sign at different orders. We leave a more detailed investigation and parameterization of the n_f dependence to the future.

5.3 Validation and statistical interpretation

5.3.1 Statistical model and interpretation

For a real measurement, performing a single measurement corresponds to drawing a value u_n from $P(u_n|\hat{\theta}_n)$ with the measurement's uncertainty Δu_n corresponding to the standard deviation of $P(u_n|\hat{\theta}_n)$. To verify the assigned Δu_n and shape of $P(u_n|\hat{\theta}_n)$ we would repeat the measurement many times, i.e., we would draw a sample of many u_n values from $P(u_n|\hat{\theta}_n)$ for fixed $\hat{\theta}_n$ and study its sample distribution.

With our idealized measurement we do not have the option to repeat the measurement, so we cannot sample $P(u_n|\hat{\theta}_n)$ in u_n for fixed $\hat{\theta}_n$. However, $P(u_n|\theta_n)$ models our entire estimation procedure, which we can apply to all perturbative series f that we consider to belong to a common category. We can thus sample $P(u_n|\hat{\theta}_n)$ over $\hat{\theta}_n$ by applying our estimation procedure to many different parameters θ_n^f of the same category whose true values $\hat{\theta}_n^f$ are known.

Given our estimator u_n for the parameter θ_n with estimated uncertainty Δu_n, we can consider the pull

$$t_n = \frac{\hat{\theta}_n - u_n}{\Delta u_n}\,, \tag{5.8}$$

JHEP08(2025)098

which is invariant under a linear transformation $\theta_n \to a\theta_n + b$, $u_n \to au_n + b$, $\Delta u_n \to a\Delta u_n$. Since $P(u_n|\theta_n)$ should be invariant under such a rescaling, we can consider it to be a function of the pull only,

$$P(u_n|\theta_n) = p\Big(\frac{\theta_n - u_n}{\Delta u_n}\Big)\,. \tag{5.9}$$

In particular, if we model u_n as a Gaussian random variable with mean θ_n and standard deviation Δu_n, then t_n is normally distributed, i.e., $p(t_n)$ is a Gaussian with zero mean and unit variance.

As long as our estimation procedure is deterministic and involves choosing specific values for u_n and Δu_n, we can always perform a linear rescaling and redefine θ_n to have $u_n = 0$ and $\Delta u_n = 1$ so $t_n = \hat{\theta}_n$. In fact, we already did so by choosing our common normalization conventions as discussed in section 5.2, which are thus an integral part of the estimation procedure. The likelihood for θ_n is then given by

$$L(\theta_n) = P(u_n = 0|\theta_n) = p(\theta_n)\,, \tag{5.10}$$

and so we like to learn about the distribution $p(\theta_n)$.

Let us denote the collection of perturbative series f of a given category by F and the corresponding collection of their series coefficients f_n as F_n. In principle, the distribution $p(\theta_n)$ could be specific to each parameter θ_n^f, so to be clear for the moment let us label it and use a generic argument, $p_{\theta_n^f}(x)$. However, since it is primarily a property of our estimation procedure, which is common to all $f_n \in F_n$, we can assume it to be the same for all of their respective θ_n^f. Furthermore, we can naturally identify this common distribution with the distribution of true values $\hat{\theta}_n^f$ of all $f_n \in F_n$, which we denote as $\hat{p}_{F_n}(x)$, so

$$p_{\theta_n^f}(x) \equiv \hat{p}_{F_n}(x) \qquad \forall \theta_n^f \text{ whose } f_n \in F_n\,. \tag{5.11}$$

Although this identification comes natural it is an assumption we make. Intuitively, we can think of it as follows: the collection of series coefficients in F_n is a QCD bag of balls. Each ball has a visible label f_n on it and a not visible number $\hat{\theta}_n^f$ inside it. We now consider a specific coefficient f_n of interest for which we need an estimate. With the identification in eq. (5.11), we think of this situation as having just taken the ball labelled f_n out of the bag, which is not random. But we are not allowed (or able) to look at its number inside it, so we have effectively drawn a random member from the population of hidden $\hat{\theta}_n^f$ numbers in the bag. Knowing that it came out of this bag (and nothing else about it), our best estimate of its value is simply the population mean and its uncertainty the population variance.[15]

For the rest of our discussion, we will work under the premise of this identification. Without it, we would have to live with a stronger assumption of assuming a certain shape for $p_{\theta_n^f}(x)$. We could also be somewhere in the middle and consider the form of $\hat{p}_{F_n}(x)$ only as a motivation for the assumed shape of $p_{\theta_n^f}(x)$ but without making the explicit identification. Ultimately, the precise interpretation is a choice the user of our theory constraint can make.

[15] More precisely, we can obtain an estimate for θ_n based on the likelihood $L(\theta_n) = \hat{p}_{F_n}(\theta_n)$. For a Gaussian (or similar) distribution the maximum likelihood estimate coincides with the mean of the distribution.

JHEP08(2025)098

5.3.2 Distribution of known perturbative series

We now discuss the distribution $\hat{p}_{F_n}(x)$ of the population of true values $\hat{\theta}_n^f$ of all $f_n \in F_n$. We consider two collections of perturbative series belonging to the two broad categories of matrix-element constants (F_f) and anomalous dimensions (F_γ) defined at the beginning of section 5.2.

The true distribution $\hat{p}_{F_n}(x)$ is obviously not known to us, as we would have to know all possible $\hat{\theta}_n^f$. Instead, we can follow the standard procedure of estimating an unknown population distribution by drawing a random sample from the population and using the resulting sample distribution as an approximation of the true population distribution. In our case, we can use the sample $\{\hat{\theta}_n^f\}$ belonging to a subset of known series coefficients $\{\hat{f}_n\} \subset F_n$, which is indeed random because we choose it without having a prior look at the actual values $\hat{\theta}_n^f$. In terms of our QCD bag of balls, by default all balls are locked and we cannot look inside them. While taking out a specific (not random) set of balls which someone has graciously unlocked for us, we are not yet looking at their numbers inside. Hence, just like when we are asking about a specific f_n, this amounts to drawing a random sample of $\hat{\theta}_n^f$ from the population inside the bag.

We might still worry that the sample distribution could be biased by the fact that the perturbative series that are known to high order are naturally simpler to calculate than the ones we do not yet know. Whilst this makes the quantities themselves special in some sense, the only relevant question is whether this also makes their values $\hat{\theta}_n^f$ somehow special and not representative of the full population, which is not necessarily the case. The sample distribution not being representative (yet) can indeed be a valid concern when only a handful and perhaps even closely related series are available. To alleviate this concern and ensure a sample as representative as possible, we have made an effort to include a large variety of different QCD quantities. Furthermore, from our repeated experience of adding new results to the existing samples over time, we do not believe this to be a concern any longer.

A detailed list of the quantities included in our sample is given in appendix A. We have included all four-loop results for matrix-element constants and all four-loop and five-loop results for anomalous dimensions we are aware of (without any claim of completeness), as well as all known three-loop matrix-element constants relevant for q_T and thrust resummation to N^{4}LL. To include a quantity in our sample, we have to be sure that it actually belongs to one of our common categories and roughly obeys our natural size estimate. For this reason we focus on series that are known to at least third order including their n_f dependence, which allows for sufficient sanity checks.

The lower-order coefficients of some quantities are directly related to each other by naive Casimir scaling. Some anomalous dimensions are equivalent due to trivial consistency relations of the form $\gamma_a + \gamma_b = 0$. In these cases, we only include the coefficients once. On the other hand, some anomalous dimensions are related by consistency relations of the form $\gamma_a + \gamma_b + \gamma_c = 0$. For these cases we do include all three series for several reasons. First, each of the γ_i is an actual anomalous dimension of some quantity and should in principle obey our estimate. Second, there is no obvious choice which one of the three to eliminate and we rather introduce a minor correlation into the sample by keeping all three than making an arbitrary selection which might cause some bias.

JHEP08(2025)098

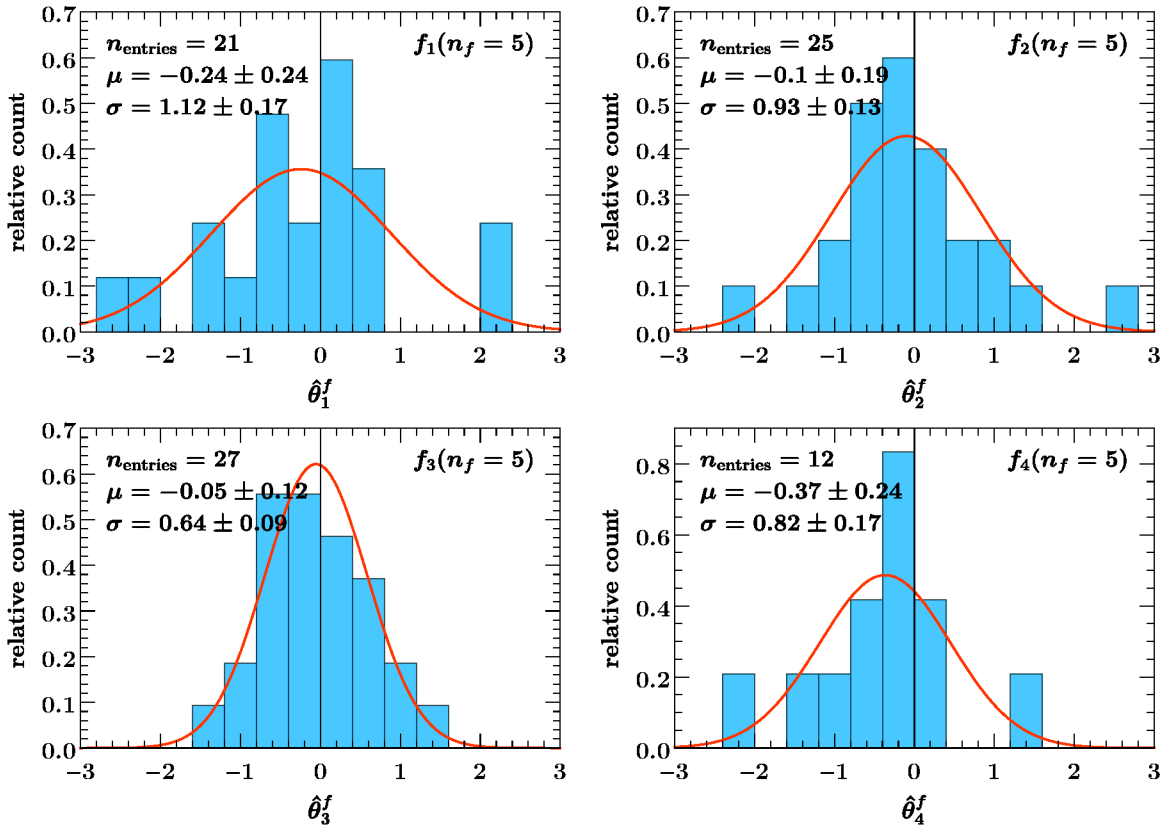

Figure 1. Distribution of true values of theory nuisance parameters for QCD matrix-element constants for $n_f = 5$ at n-loop order for $n = 1$ to 4.

The sample distributions are shown for the matrix-element constants in figures 1 and for the anomalous dimensions in figures 2. The data is shown by the light blue histograms. It is binned only for visualization purposes. For each sample, we perform an unbinned fit to a Gaussian distribution with mean μ and variance σ^2 as free parameters. The result is shown by the orange line and the fitted values for μ and σ are quoted in each plot. The number of entries (n_{entries}) for each sample is also given. The computed sample variance agrees with the fitted Gaussian variance to within a few percent in all cases except at the highest order with few entries where it differs by at most 8%.

We first observe that the standard deviation σ for all samples is consistent with unity, which provides a clear validation of our natural-size estimate in section 5.2. (For the three-loop constants and anomalous dimensions the variance is somewhat smaller than one, which is not a concern.) The mean μ for both matrix-element constants and anomalous dimensions is consistent with zero. The shape of the sample distributions is generally well described by the Gaussian fit for the given number of fitted data points. The one-loop and two-loop anomalous dimensions show some noticeable clusterings away from zero, which we can likely attribute to the fact that the their coefficients do not yet contain sufficient entropy.

Given that the samples for different n all show similar distributions, with in particular the same mean and standard deviation within uncertainties, we can take a step further and

JHEP08(2025)098

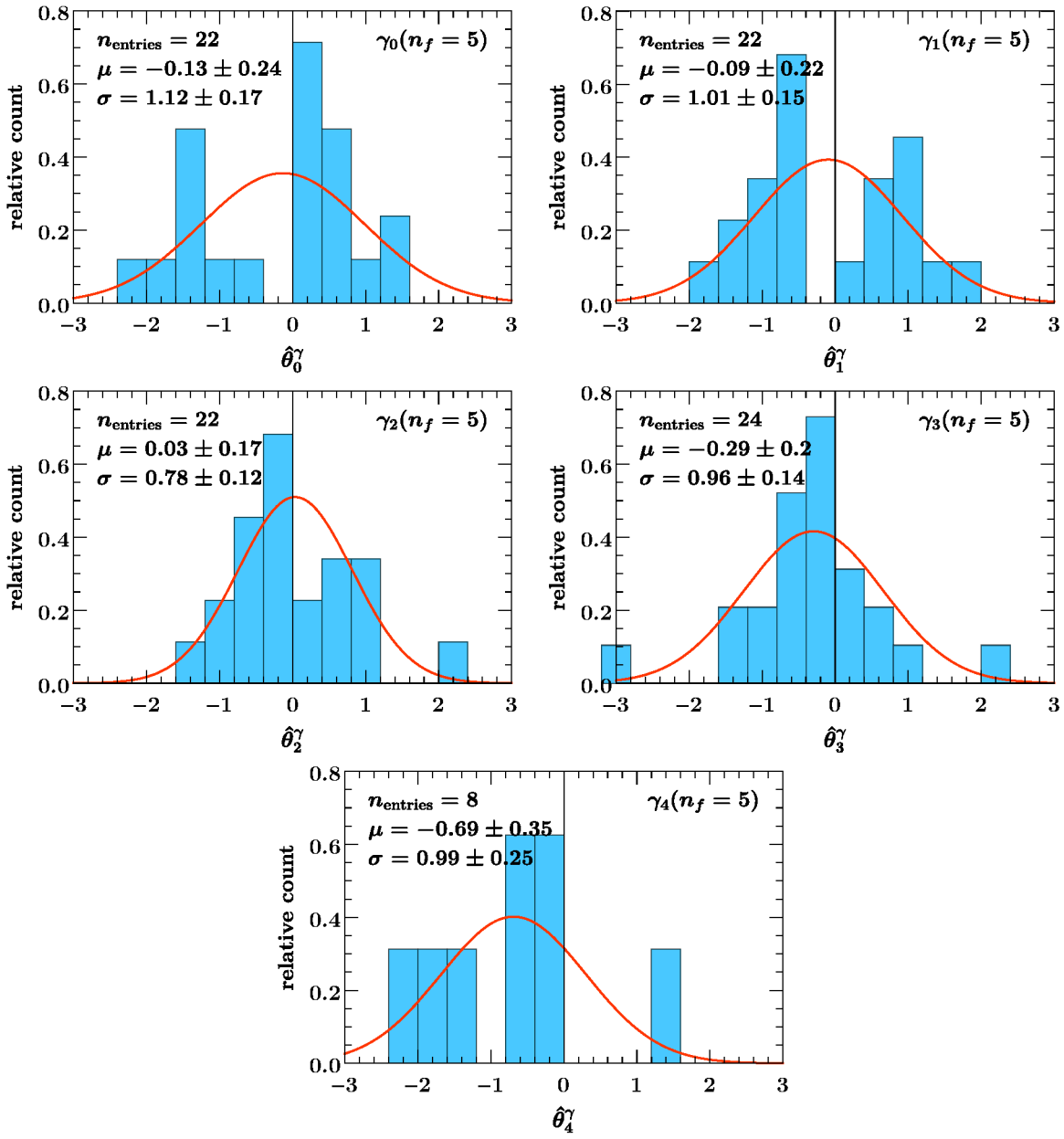

Figure 2. Distribution of true values of theory nuisance parameters for QCD anomalous dimensions for $n_f = 5$ at $(n+1)$-loop order for $n+1 = 1$ to 5.

assume that the populations for different n can be described by a common distribution,

$$\hat{p}_{F_n}(x) \equiv \hat{p}_F(x) \qquad \forall n\,, \tag{5.12}$$

where $\hat{p}_F(x)$ is the distribution of all $\hat{\theta}_n$ in F for any n. This allows us to combine the samples for different n in each category. The resulting distributions of the combined samples for F_f and F_γ are shown in figures 3. Their approximately Gaussian shape is clearly evident. Most importantly, their fitted $\sigma = 0.90 \pm 0.07$ and $\sigma = 1.00 \pm 0.07$ are perfectly consistent with unity, and also agree with the computed sample variances. The means of the distributions

JHEP08(2025)098

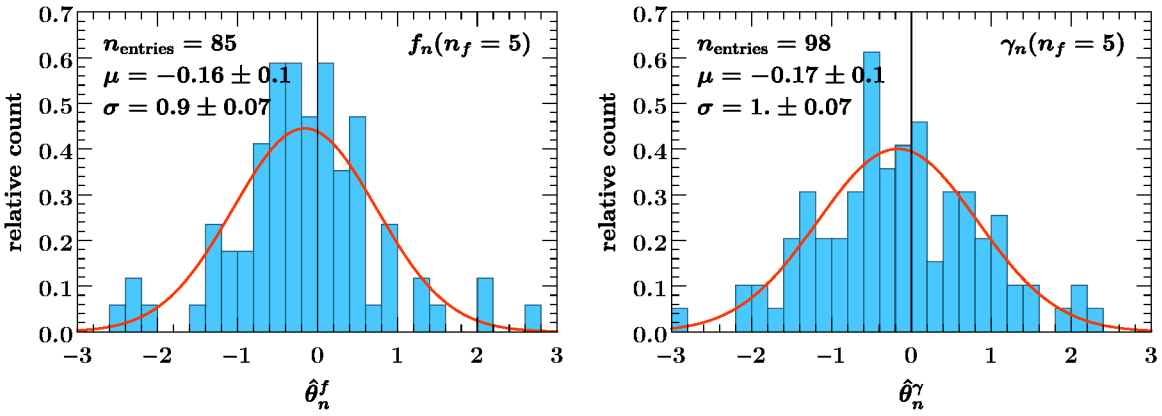

Figure 3. Distribution of true values of theory nuisance parameters for QCD matrix-element constants (left) and anomalous dimensions (right) for $n_f = 5$ combining all available orders.

$\mu = -0.16 \pm 0.10$ and $\mu = -0.17 \pm 0.10$ show a small shift away from zero, but they are still consistent with zero. We could in principle account for this shift by adjusting our estimated central value. However, given that the effect is only marginal, we do not see a good reason for doing so in practice at this point.

In summary, we find very strong and convincing evidence for the robustness of our estimation procedure. Based on a large sample of known series coefficients, and with the identification in eq. (5.11), we find the distribution $p(\theta_n)$ in eqs. (5.9) and (5.10) to have approximately zero mean and unit variance, confirming our original estimate of $u_n = 0$ with $\Delta u_n = 1$. We further find it to be well approximated by a Gaussian in agreement with our original assumption.

5.3.3 Further discussion

The fact that our estimation procedure yields distributions closely resembling Gaussians as in figures 3 speaks for itself. This can be contrasted with the very long-tailed distributions obtained from an analogous exercise using scale variations in ref. [9] or the typically very non-Gaussian distributions produced by the methods of refs. [4–7].

It is also undeniable that the distributions are closer to a Gaussian than a flat box, in contrast to what one might have expected, as such a box-like distribution is sometimes advocated to be more appropriate than a Gaussian for perturbative theory uncertainties (albeit typically in the context of scale variations). We might also ask why to expect the distributions $\hat{p}_{F_n}(\hat{\theta}_n)$ and $\hat{p}_F(\hat{\theta}_n)$ to be sensible or useful in the first place. In fact, even though the $\hat{f}_n \in F_n$ all belong to some common category of perturbative series, we want this category to be as broad as possible to be as useful as possible. This means the distribution of $\hat{f}_n$, which is solely a property of the collection F_n might very well be quite irregular and not very useful by itself. However, as we have stressed already, the distribution $\hat{p}_{F_n}(\hat{\theta}_n)$ is a property of our estimation procedure for F_n. Its goal is precisely to strip the $\hat{f}_n$ of their individuality and reduce them to a generic bunch of numbers of a common more-or-less random origin, namely arising as a more-or-less random sum of Feynman diagrams. It is then perhaps not

JHEP08(2025)098

surprising after all that the resulting population of $\hat{\theta}_n$ can be well described by a Gaussian distribution. We might think of it as the central-limit theorem of Feynman diagrams.[16]

It is also instructive to think about how we would notice if there was something going wrong in our estimation procedure. If we find a Gaussian with different mean or variance, then our estimation procedure itself is sound but our final choice of u_n or Δu_n is off, which we can easily adjust for if necessary. If the resulting distribution is irregular, e.g., with large tails or other undesired features, then this signals that our estimation procedure is suboptimal or missing some important aspect. A long tail would be indicative of underestimating the natural size for some coefficients in F_n. To illustrate this with a simple example, imagine we had not accounted for the overall C_r color factor. As long as our collection F_n contains only quark-like or only gluonic quantities, this would not be much of a problem, we would simply find an uncertainty of C_F or C_A instead of one. However, when F_n contains both, we would end up with a superposition of two Gaussians of different variance. We could still work with this distribution, but it would be suboptimal because it would lead to overestimating the uncertainty for quark-like quantities and underestimating it for gluonic quantities.

This is also why it is prudent to consider separate collections F_n for each n at first, as this allows us to test and identify the appropriate n-dependent normalization. For example, without the $(n-1)!$ in N_n^f in eq. (5.4) we would find distributions of correspondingly larger variance for $n \geq 3$, which is in fact how we became aware of this factor during the course of our investigations.

We conclude this subsection with two more comments. First, when applying our estimation procedure to a known θ_n we should not use our knowledge of its true value $\hat{\theta}_n$, but by including it in the sample of known $\hat{\theta}_n$ we do indirectly use it. However, this is acceptable since the impact of any one coefficient on the sample distribution is minor. Second, when applying our estimation procedure to a new and still unknown θ_n, we still have to make a judgement whether or not it belongs to a particular category. There is a priori no guarantee for that. It might well be the case that there are genuinely different types of quantities than those considered so far that cannot be reduced to fit into an existing category but instead require defining and studying a new category.

5.4 Designing theory estimators

An interesting question to consider is whether it is possible to improve upon our estimator or design alternative estimators, which could be tested using the same procedure as above. We leave this for future investigation and only give some general remarks here. An improved estimator should on average yield a tighter estimate of the natural size but without underestimating it for some subset of series either. In other words, it should yield a reduced variance and ideally still produce a roughly Gaussian distribution.

For example, one could imagine devising an estimator based on the actual leading-color approximation of a series coefficient, i.e., using the large-N_c expansion as a supplementary expansion following strategy 2 in section 4.3. This of course requires performing an actual calculation, but typically the calculation in the large-N_c limit is easier to perform than the

[16] The credit for coining this term goes to Glen Cowan.

JHEP08(2025)098

full calculation. Having a robust estimator based on this limit would allow one to robustly use such approximate results.

Another natural question that arises is whether one could utilize the information from the known lower orders $\hat{f}_{k<n}$ of a given quantity f to improve the estimate for f_n. This effectively amounts to devising alternative parameterizations for f_n involving the information of the known $\hat{f}_{k<n}$ in some way. In our initial attempts we found that this overall leads to less reliable estimates. For example, Padé approximations can work extremely well in some cases and utterly fail in others. The basic problem of relying on lower-order information is that this introduces the same generic pitfall also present for scale variations: it can lead to random significant underestimations when the lower-order coefficient(s) happen to be randomly smaller than their own natural size. This is in fact not unlikely to happen since as we have seen the mean of the distribution of true values is around 0. Examples of this are already present in tables 2 and 3, namely $\hat{f}_1$ of $c_{q\bar{q}S}$, $\hat{f}_2$ of c_{gg}, and notably $\hat{\gamma}_3$ of Γ^q_{cusp}.

One might also consider applying the Bayesian inference models of refs. [4–7] to estimate a given θ_n based on its known lower-order $\hat{\theta}_{k<n}$, as was mentioned already in ref. [6]. In this case, similar care has to be exercised to avoid the above pitfall. Another general option would be to utilize series transformations or series acceleration methods as in ref. [8]. In fact, taking the square root of a quantity can be considered a simple form of a series transformation.

From our experience so far, the most useful way to utilize the known lower-order information is as an important cross check of the estimation procedure rather than as a direct input to it. If a quantity consistently violates its estimated natural size at lower orders, it might indicate that we are not estimating its natural size correctly, which can have various reasons. We might be using the wrong normalization factor or a suboptimal reference scheme, or we might be associating it incorrectly with a given category. The latter can happen when not using the appropriate conventions, e.g. we should be parameterizing $\sqrt{f}$ or f^2 instead of f. It can also happen when we are using a suboptimal parameterization, for example when the scalar series has important internal structures (e.g. new color or partonic channels) which affect its natural size but which we have not explicitly parameterized.

More generally, the question is to what extent the higher-order coefficients are correlated with the lower-order ones and how to best exploit this correlation to our advantage. Such correlations could arise for example from cross terms of lower-order coefficients appearing as part of the higher-order coefficient (an obvious example is again using f^2 instead of f), or we believe the higher-order correction to be a genuinely multiplicative correction on top of the lower-order result (in which case we would parameterize their ratio). In general, such information is specific to a given quantity f. Therefore, all genuine lower-order information that we believe to be relevant should be accounted for explicitly by the specific parameterization of $f_n(\theta_n)$ itself. An optimal parameterization would then be one for which the θ_n are uncorrelated for different n. In this limit, no more information can be gained from the lower-order $\hat{\theta}_{k<n}$ and the easiest and safest approach is to estimate the natural size of θ_n without further reference to $\hat{\theta}_{k<n}$. The boundary between parameterization and estimation is of course somewhat blurry, since as we have seen, the final step of parameterizing the remaining scalar coefficients in terms of scalar TNPs is in fact an important part of the estimation procedure itself.

JHEP08(2025)098

6 Application to transverse-momentum resummation

The q_T spectrum of Z and W bosons produced in hadronic collisions, where $q_T \equiv p_T^{Z,W}$ is the transverse momentum of the produced vector boson, is a benchmark observable of the LHC precision physics program and has been measured to incredible precision by the ATLAS [68–71], CMS [72–74], and LHCb collaborations [75, 76]. In this section, we discuss the application of our approach to precision predictions for the q_T spectrum using resummed perturbation theory.

Correlations within the p_T^W and p_T^ℓ spectra, and depending on the analysis strategy also between p_T^W and p_T^Z, are critical for a precise measurement of the W-boson mass at hadron colliders [1, 77–80]. As shown in ref. [16], the theory correlations in p_T^Z are also critical if one wants to perform fits to the precisely measured small-p_T^Z spectrum to extract nonperturbative parameters [81–83] or the strong coupling constant [84, 85].

In section 6.1 we give a brief account of the aspects of q_T factorization and resummation that are relevant to our discussion. In section 6.2 we identify and discuss the necessary TNPs, and in section 6.3 we present numerical results that illustrate the power of the TNP approach to obtain predictions with proper theory correlations. Finally in section 6.4, we briefly discuss the treatment of subleading effects, which we neglect here for simplicity.

6.1 Aspects of q_T resummation

We denote the four-momentum of the vector boson by q^μ, its invariant mass and rapidity by $Q \equiv \sqrt{q^2}$ and Y, and its transverse momentum by $q_T = |\vec{q}_T|$. The quantity of our interest is the cross section fully differential in Q, Y, and q_T, which we write for brevity as $\mathrm{d}\sigma/\mathrm{d}^4q$.

We start by applying strategy 2 of section 4.3 and expand the cross section in a power series in $\varepsilon \equiv q_T^2/Q^2$,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}^4q} = \frac{\mathrm{d}\sigma^{(0)}}{\mathrm{d}^4q}\Big[1 + \mathcal{O}\Big(\frac{q_T^2}{Q^2}\Big)\Big]\,. \tag{6.1}$$

Compared to our discussion in section 4.3, where we expanded the series coefficient in ε, here it is much more useful to first perform the expansion in ε and only later the perturbative expansion in α_s. The reason is that we actually know the functional form in q_T (and Q) of the leading-power term $\mathrm{d}\sigma^{(0)}$ to all orders in α_s, allowing us to apply strategy 1 and obtain the exact correlations in q_T and Q. Furthermore, we will also resum certain parts of the perturbative series to all orders in α_s, although the precise way of doing so is not of immediate concern to us here, so we will not discuss it but refer the interested reader to refs. [42, 44].

The power expansion in eq. (6.1) converges very well, even better than the q_T^2/Q^2 scaling suggests, such that the power corrections remain below $\lesssim 5\%$ even up to moderately large $q_T \lesssim Q/3$ and even $Q/2$. As a result, the leading-power term $\mathrm{d}\sigma^{(0)}$ dominates and effectively determines the spectrum over this entire small-q_T region, and thus also causes the dominant perturbative uncertainties. We can therefore focus our discussion on $\mathrm{d}\sigma^{(0)}$. In particular, it will serve us to demonstrate a nontrivial example application of the TNP approach. We will comment further on the treatment of the $\mathcal{O}(q_T^2/Q^2)$ power corrections and other subleading effects in section 6.4.

JHEP08(2025)098

The leading-power term $\mathrm{d}\sigma^{(0)}$ is the subject of the q_T factorization and resummation program. We do not intend to provide a detailed review of q_T resummation here. Rather, our focus is on the kinematic and process dependence, which we wish to break down and parameterize in terms of theory nuisance parameters. We use the SCET resummation framework of refs. [42, 44]. We closely follow the notation of those references and refer there for more details and further references. The leading-power cross section can be written as

$$\begin{aligned}\frac{\mathrm{d}\sigma^{(0)}}{\mathrm{d}^4 q} &= \frac{1}{2E_{\mathrm{cm}}^2}\, L_{VV'}(q^2) \sum_{a,b} H_{VV'\,ab}(q^2,\mu) \\ &\quad \times \int \frac{\mathrm{d}^2\vec{b}_T}{(2\pi)^2}\, e^{\mathrm{i}\vec{b}_T\cdot\vec{q}_T}\, \tilde{B}_a(x_a, b_T, \mu, \nu/Q)\, \tilde{B}_b(x_b, b_T, \mu, \nu/Q) \tilde{S}(b_T, \mu, \nu)\,. \end{aligned} \tag{6.2}$$

Here, $VV' = \{\gamma\gamma, \gamma Z, Z\gamma, ZZ, W^+W^+, W^-W^-\}$ labels the produced vector boson including possible interferences. The leptonic tensor $L_{VV'}(q^2)$ contains the vector-boson propagator and decay and receives no QCD corrections, so its q^2 dependence is known. The hard function $H_{VV'\,ab}(Q^2,\mu)$ encodes the production of the vector boson in the underlying hard interaction $ab \to V$, with the sum over a, b running over all relevant combinations of quark and antiquark flavors. The functional form of its q^2 and process dependence is known to all orders. The second line in eq. (6.2) contains all soft and collinear physics at the low scale $\mu \sim q_T$ encoded respectively in the soft function $\tilde{S}$ and beam function $\tilde{B}_{a,b}$. The q_T dependence arises entirely from the second line, and its functional form is fully determined by the functional dependence on its Fourier-conjugate variable $b_T = |\vec{b}_T|$. The functional form of the b_T dependence of the beam and soft functions is in turn known to all orders in α_s. The beam function also depends on the flavor of the (anti)quark participating in the hard interaction and on Q. The functional form of these dependencies is also known to all orders. Finally, the variables $x_{a,b} = (Q/E_{\mathrm{cm}})e^{\pm Y}$ encode the dependence on the rapidity Y and center-of-mass energy E_{cm}. The functional form of the $x_{a,b}$ dependence of the beam function is not known to all orders but depends on their perturbative order.

The factorization in eq. (6.2) is very powerful for our purposes as it predicts the complete functional form in q_T and also in Q for given $x_{a,b}$. Furthermore, it fully parameterizes the exact dependence on the process and partonic channels. We are therefore able to apply strategy 1 and obtain exact correlations in all these dependencies. Although it does not predict the complete functional form in $x_{a,b}$ it still reduces it from a generic two-dimensional dependence to a product of common, universal one-dimensional beam functions.

For simplicity we have limited ourselves to the inclusive q_T spectrum in eq. (6.2). Including the full kinematics of the vector-boson decay products is also possible. Importantly, at leading power doing so only increases the complexity of the leptonic tensor but does not induce any additional sources of QCD uncertainties [42].[17] We can therefore also capture the correlations in leptonic kinematic variables, most notably the lepton transverse momentum p_T^ℓ, or between the q_T spectrum and the q_T-dependent forward-backward asymmetry.

[17] More precisely, leptonic observables can give rise to enhanced power corrections, which for azimuthally symmetric observables can be taken into account in terms of the leading-power QCD contributions, and thus without inducing additional sources of QCD uncertainties. Starting at $\mathcal{O}(q_T^2/Q^2)$ also genuinely new QCD structures can contribute, see ref. [42] for a detailed discussion.

JHEP08(2025)098

In principle, eq. (6.2) could be applied to each coefficient of the perturbative series of $\mathrm{d}\sigma^{(0)}$. However, at each order in α_s, (double) logarithms of q_T/Q appear, which render a fixed-order expansion of $\mathrm{d}\sigma^{(0)}$ unstable. Instead, eq. (6.2) also provides the basis for systematically resumming the unstable logarithmic contributions to all orders in α_s, leading to precise and perturbatively stable predictions. We will not discuss how the resummation is carried out in practice but refer to refs. [42, 44] for details. The key point for us is that a given perturbative resummation order, $\mathrm{N}^n\mathrm{LL}$, is uniquely defined by including all underlying scalar perturbative series discussed below to a specific order in α_s. We then define our generalized counting including TNPs, $\mathrm{N}^{n+k}\mathrm{LL}$, to include the true values for all coefficients relevant for $\mathrm{N}^n\mathrm{LL}$ and in addition for each series the TNP parameterization of the next k terms. In analogy to section 3.2, we also define the approximate $\mathrm{N}^{n+0}\mathrm{LL}$ implementation by absorbing the TNPs appearing at $\mathrm{N}^{n+1}\mathrm{LL}$ as an additive correction to the respective highest coefficients appearing at $\mathrm{N}^n\mathrm{LL}$.

6.2 TNPs for q_T resummation

The perturbative ingredients required in eq. (6.2) are the hard, beam, and soft functions. Their functional dependence on the kinematic variables, except x, is fully predicted to all orders in α_s by their renormalization group equations, which we now discuss in turn. At the end we will be left with a set of (mostly) scalar perturbative series that fully determine the (fixed-order and/or resummed) perturbative series of $\mathrm{d}\sigma^{(0)}$. We will give a summary in section 6.2.4, so readers not interested in the detailed definitions can directly skip there.

6.2.1 Hard function

The leptonic tensors for inclusive $Z \to \ell\ell$ and $W \to \ell\nu$ in eq. (6.2) are given by

$$
\begin{aligned}
L_{ZZ}(q^2) &= \frac{2}{3}\frac{\alpha_{\mathrm{em}}}{q^2}\,(v_\ell^2 + a_\ell^2)\left|\frac{q^2}{q^2 - m_Z^2 + \mathrm{i}\Gamma_Z m_Z}\right|^2, \\
L_{W^+W^+}(q^2) &= \frac{1}{6}\frac{\alpha_{\mathrm{em}}}{q^2}\,\frac{1}{\sin^2\theta_w}\left|\frac{q^2}{q^2 - m_W^2 + \mathrm{i}\Gamma_W m_Z}\right|^2.
\end{aligned}
\tag{6.3}
$$

Their q^2 dependence is known exactly in QCD. The corresponding hard functions have the form

$$
\begin{aligned}
H_{ZZ\,q\bar{q}'}(q^2,\mu) &= \frac{8\pi\alpha_{\mathrm{em}}}{N_c}\delta_{qq'}\Big\{(v_q^2 + a_q^2)|C_q(q^2,\mu)|^2 \\
&\qquad + 2\Re\sum_f\Big[v_q v_f C_q^*(q^2,\mu)C_{vf}(q^2,\mu) + a_q a_f C_q^*(q^2,\mu)C_{af}(q^2,\mu)\Big] + \cdots\Big\}, \\
H_{W^+W^+\,q\bar{q}'}(q^2,\mu) &= \frac{2\pi\alpha_{\mathrm{em}}}{N_c}\frac{|V_{qq'}|^2}{\sin^2\theta_w}|C_q(q^2,\mu)|^2\,.
\end{aligned}
\tag{6.4}
$$

The expressions for the remaining VV' combinations can be found in appendix A of ref. [42]. The v_i and a_i are the usual axial and vector couplings of the Z boson, Q_q is the electromagnetic charge of quark q, and $V_{qq'}$ are the CKM-matrix elements.

The q^2 dependence of the hard function is determined by that of the matching coefficients $C_i(q^2,\mu)$, which correspond to the infrared-finite parts of the respective QCD form factors. Here, $C_q = 1 + \mathcal{O}(\alpha_s^2)$ is the dominant vector nonsinglet coefficient corresponding to diagrams

JHEP08(2025)098

where the vector boson couples to the external quark line. The C_{af} and C_{vf} in eq. (6.4) are axial-singlet and vector-singlet coefficients corresponding to diagrams where the vector boson couples to a closed fermion loop, which only contribute to Z production but not to W production. They have separate perturbative series starting at $\mathcal{O}(\alpha_s^2)$ and $\mathcal{O}(\alpha_s^3)$, respectively, and have to be parameterized separately. In practice, their contributions are very small even at the order they contribute [44], so we can neglect them here for simplicity. In principle they would have to be included (starting at N^{3}LL) to fully account for the correct (de)correlation between W and Z production. The ellipses in $H_{ZZ\,q\bar{q}'}$ denote terms proportional to the square of C_{af} and C_{vf}, which only contribute starting at $\mathcal{O}(\alpha_s^4)$.

The functional form of the q^2 dependence of $C_q(q^2,\mu)$ is known because by dimensional analysis it can only depend on the ratio q^2/μ^2. The q^2 dependence is therefore fully predicted by the μ dependence, which in turn is governed by C_q's renormalization group evolution (RGE) equation,

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln C_q(q^2,\mu) = \Gamma^q_{\text{cusp}}[\alpha_s(\mu)]\ln\frac{-q^2-\mathrm{i}0}{\mu^2} + 2\gamma^q_C[\alpha_s(\mu)]\,. \tag{6.5}$$

The full q^2 and μ dependence of C_q can be reconstructed by solving eq. (6.5) (either order by order in α_s or to all orders to obtain its resummed expression).

The cusp and noncusp anomalous dimensions $\Gamma^q_{\text{cusp}}(\alpha_s)$ and $\gamma^q_C(\alpha_s)$ in eq. (6.5) are already scalar series. Following our conventions for anomalous dimensions in section 5.2, we parameterize

$$\Gamma(\alpha_s) \equiv 2\Gamma^q_{\text{cusp}}(\alpha_s)\,, \qquad \gamma_\mu(\alpha_s) \equiv 2\gamma^q_C(\alpha_s)\,, \tag{6.6}$$

in terms of corresponding TNPs θ^Γ_n and $\theta^{\gamma_\mu}_n$.

The remaining nontrivial part of C_q we need to parameterize is the q^2 and μ-independent constant term, which is not predicted by eq. (6.5) and effectively acts as the boundary condition for solving the differential equation. We can formally define it as the matching coefficient evaluated at the canonical scale $\mu^2 = -q^2$,

$$c_q(\alpha_s) \equiv C_q(q^2,\mu^2=-q^2)\,. \tag{6.7}$$

By choosing the canonical scale proportional to q^2, the perturbative series for $c_q(\alpha_s)$ becomes a scalar series with q^2 and μ independent coefficients. Here, $c_q(\alpha_s)$ is equal to $c_{q\bar{q}V}(\alpha_s)$ in section 5.2, so we parameterize it directly in terms of TNPs θ^H_n, where the label is meant to remind us that they come from the hard function.

Note that the matching coefficient is defined in a certain renormalization scheme, for which we use the standard $\overline{\text{MS}}$ scheme here. Together with the canonical scale choice, which also determines the form of the logarithm in eq. (6.5), this defines the reference scheme for the anomalous dimensions and constant term and their TNPs.

6.2.2 Soft function

The TNP parameterization of the soft function $\tilde{S}(b_T,\mu,\nu)$ proceeds analogously to that of the matching coefficient $C_q(q^2,\mu)$ above. A new element is the soft function's dependence on the additional rapidity renormalization scale ν, which has dimension one. By dimensional

JHEP08(2025)098

analysis, the soft function can only depend on two ratios b_T/μ and μ/ν, so its full b_T dependence is determined by its dependence on μ and ν, which is now governed by a system of RGE equations,

$$\begin{aligned}
\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln\tilde{S}(b_T,\mu,\nu) &= 4\Gamma^q_{\text{cusp}}[\alpha_s(\mu)]\ln\frac{\mu}{\nu}+\tilde{\gamma}_S[\alpha_s(\mu)]\,,\\
\nu\frac{\mathrm{d}}{\mathrm{d}\nu}\ln\tilde{S}(b_T,\mu,\nu) &= \tilde{\gamma}_\nu(b_T,\mu)\,,\\
\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\tilde{\gamma}_\nu(b_T,\mu) &= -4\Gamma^q_{\text{cusp}}[\alpha_s(\mu)]\,.
\end{aligned}\tag{6.8}$$

Here, the rapidity anomalous dimensions $\tilde{\gamma}_\nu(b_T,\mu)$ has a more nontrivial dependence on b_T, which is in turn determined by its own μ dependence governed by its own μ RGE in the last line.

Solving eq. (6.8) now requires two independent boundary conditions, one for $\gamma_\nu(b_T,\mu)$ and one for $\tilde{S}(b_T,\mu,\nu)$ itself. The canonical scale in b_T space is $\mu = b_0/b_T$ with $b_0 = 2e^{-\gamma_E} \approx 1.12291$, which corresponds to $\mu = q_T$ in momentum space. The soft function scales like a squared $2\to 0$ matrix element. Following our conventions in section 5.2, we therefore define the relevant scalar series as

$$\begin{aligned}
\gamma_\nu(\alpha_s) &\equiv \frac{1}{2}\tilde{\gamma}_\nu(b_T,\mu=b_0/b_T)\,,\\
\tilde{s}(\alpha_s) &\equiv \sqrt{\tilde{S}(b_T,\mu=b_0/b_T,\nu=b_0/b_T)}\,,
\end{aligned}\tag{6.9}$$

which we parameterize in terms of corresponding TNPs $\theta_n^{\gamma_\nu}$ and θ_n^S. Note that the reference scheme for the TNPs here corresponds to our choices of using b_T space and its canonical scale, $\overline{\text{MS}}$ renormalization, and rapidity renormalization [86] with the exponential regulator [87].

The other perturbative ingredients we need for the soft function are the cusp and noncusp anomalous dimensions in the first line of eq. (6.8). Following our conventions we would again parameterize $\Gamma(\alpha_s) \equiv 2\Gamma^q_{\text{cusp}}(\alpha_s)$, consistent with eq. (6.6), and $\gamma_S(\alpha_s) = \tilde{\gamma}_S(\alpha_s)/2$. In practice, we do not need TNPs for $\gamma_S(\alpha_s)$, for reasons we will explain in a moment.

6.2.3 Beam functions

The beam function $\tilde{B}_i(x,b_T,\mu,\nu/Q)$ only depends on the combination ν/Q, as indicated by its argument, and thus by dimensional analysis only on b_T/μ. Its b_T and explicit Q dependence is thus governed by its RGE system, which is closely analogous to that of the soft function in eq. (6.8),

$$\begin{aligned}
\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln\tilde{B}_q(x,b_T,\mu,\nu/Q) &= 2\Gamma^q_{\text{cusp}}[\alpha_s(\mu)]\ln\frac{\nu}{Q}+\tilde{\gamma}_B[\alpha_s(\mu)]\,,\\
\nu\frac{\mathrm{d}}{\mathrm{d}\nu}\ln\tilde{B}_q(x,b_T,\mu,\nu/\omega) &= -\frac{1}{2}\tilde{\gamma}_\nu(b_T,\mu)\,,\\
\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\tilde{\gamma}_\nu(b_T,\mu) &= -4\Gamma^q_{\text{cusp}}[\alpha_s(\mu)]\,.
\end{aligned}\tag{6.10}$$

We need again the cusp and rapidity anomalous dimensions, which are the same as before in eq. (6.8), the noncusp beam anomalous dimension $\gamma_B(\alpha_s) \equiv \tilde{\gamma}_B(\alpha_s)$, and the beam function boundary condition.

JHEP08(2025)098

We do not need TNPs for the beam and soft noncusp anomalous dimensions for the following reason. When using the beam and soft function's RGEs to reconstruct their full fixed-order expressions, we only need the known anomalous dimension coefficients (for our considered resummation orders). Their TNPs would only enter in the evolution itself. However, since the beam and soft functions start their evolution at the same canonical scale $\mu = b_0/b_T$, only their total μ anomalous dimension actually enters in the resummation, which by consistency is equal to minus that of the hard function. We therefore only need the single noncusp μ anomalous dimension in eq. (6.6).

Importantly, the RGEs do not depend on x, which implies that the additional x dependence factorizes from the b_T and Q dependencies and only enters via the beam boundary condition, which is now defined at the canonical scales $\mu = b_0/b_T$ and $\nu = Q$,

$$\tilde{b}_i(x, \alpha_s) \equiv \tilde{B}_i(x, b_T, \mu = b_0/b_T, \nu/Q = 1)\,. \tag{6.11}$$

The additional complication for the beam function arises because its x dependence is not predicted by its RGE, so the beam boundary condition is a general one-dimensional function of x. To further break down this dependence, we calculate its series coefficients $\tilde{b}_{i,n}(x)$ in terms of collinear PDFs $f_j(x)$,

$$\tilde{b}_{i,n}(x) = \sum_j \int \frac{\mathrm{d}z}{z}\, \tilde{I}_{ij,n}(z)\, f_j\Big(\frac{x}{z}\Big)\,, \tag{6.12}$$

where $\tilde{I}_{ij,n}(z)$ are perturbatively calculable matching kernels. The x dependence of the beam function is thus determined via the Mellin convolution of the x dependence of the PDFs and the z dependence of the matching kernels. Since the x dependence of the PDFs tends to be quite strong, the mix of contributing PDFs determines the overall size of $\tilde{b}_{i,n}(x)$ as well as playing an important role in determining its shape in x. The $I_{ij,n}(z)$ are perturbative coefficients, so we can in principle estimate their natural size as in section 5.2. (In fact, their moments in x enter into our sample of matrix-element constants). In contrast, it would be quite difficult to estimate the natural size of $\tilde{b}_{i,n}(x)$ directly. eq. (6.12) is thus an example where parameterizing a dependence (here the channel dependence) is beneficial or even necessary for obtaining a natural-size estimate.

Following our discussion in section 4, if we do not require precise correlations in x, one option would be to only parameterize the integral of $I_{ij,n}(z)$, with e.g. a trivial z dependence $\sim \delta(1-z)$. If we do require proper correlations in x, i.e. in Y and/or E_{cm}, we need to properly parameterize the z dependence. At the orders we are working their true expressions are actually known [88, 89]. Therefore, as a starting point we parameterize them using their known functional form in z multiplied by an overall scalar coefficient

$$\tilde{I}_{ij,n}(z, \theta_n^{B_{ij}}) = \frac{3}{2}\, \theta_n^{B_{ij}}\, \hat{\tilde{I}}_{ij,n}(z)\,, \tag{6.13}$$

where we include a factor of 3/2 to be conservative and account for the fact that their true values are typically somewhat below their natural size. Another option would be to explicitly normalize the $\hat{\tilde{I}}_{ij,n}(z)$ in some way.

With the TNP parameterization in eq. (6.13), we effectively treat the shape as exactly known, while the overall normalization is unknown. We prefer this option to using $\delta(1-z)$,

JHEP08(2025)098

because it means we have the exact correlations for the overall normalization uncertainty we do consider. Of course, at the highest known order we cannot do that, and strictly speaking at lower orders we should not be allowed to use the known shape but include some shape uncertainties. In the future, the z dependence of the matching kernels can be explicitly parameterized, for example by using strategy 2 and expanding them in $\varepsilon = 1 - z$, since their $z \to 1$ limit is actually well understood [47].

The dominant partonic channels are $ij = \{qqV, qg\}$, which start at $\mathcal{O}(1)$ and $\mathcal{O}(\alpha_s)$. At higher orders, further singlet channels $ij = \{q\bar{q}V, qqS, qq\Delta S\}$ appear, whose precise definition is given in ref. [47]. Since they only give small corrections even at the order they appear, for our numerical results in section 6.3 we only consider two TNPs for the beam boundary condition, namely a single effective $\theta_n^{B_{qq}}$,

$$\theta_n^{B_{qq}} \equiv \theta_n^{B_{qqV}} \equiv \theta_n^{B_{q\bar{q}V}} \equiv \theta_n^{B_{qqS}} \equiv \theta_n^{B_{qq\Delta S}}\,, \tag{6.14}$$

which collectively varies all qq channels together with $\theta_n^{B_{qg}}$ for the qg channel.

In principle, we also have to include the QCD splitting functions, which govern the evolution of the PDFs, in our counting. In the resummed cross section, the PDFs enter through the beam functions where they are evaluated at the scale of the beam function, which means their evolution contributes to the q_T resummation by resumming single logarithms of b_T. That is, they count as a noncusp anomalous dimension. Constructing TNP parameterizations for the splitting functions can be done similarly to the beam function matching kernels by considering their $z \to 1$ and also $z \to 0$ limits. In fact, in this way TNPs for the four-loop splitting functions have already been considered in ref. [13] including constraints from their known moments. Since varying the splitting functions is rather involved technically, as it requires re-evolving the PDFs, we refrain from doing so here, and leave this for future work. Instead, if needed, this source of uncertainty can be probed for now by conventional μ_F variations.

6.2.4 Summary of TNPs

To summarize, we have a minimum of seven TNPs, corresponding to seven independent perturbative ingredients and thus sources of uncertainty: three anomalous dimensions and four boundary conditions, which belong to the category of matrix-element constants,

$$\theta_n^{\gamma} : \gamma \in \{\Gamma, \gamma_\mu, \gamma_\nu\}\,, \qquad \theta_n^{f} : f \in \{H, S, B_{qq}, B_{qg}\}\,. \tag{6.15}$$

There is actually one piece of perturbative information that we have silently taken for granted so far: the solution of the RGEs also requires the QCD β function, because it governs the μ dependence of $\alpha_s(\mu)$, and its TNP would in principle enter in the resummation at the same loop order as the TNP of the cusp anomalous dimension. In practice however, while the overall μ evolution of $\alpha_s(\mu)$ is important, the higher-order corrections to it tend to be numerically very small. We therefore continue to treat the β function as known to avoid adding significant but unnecessary complexity.

In addition to the above 7 TNPs (or 6 if we are willing to count the beam function as a single one), we have 3 (or 4) more once we account for the full set of partonic channels of the beam function (still without accounting for its functional dependence). In addition, we

JHEP08(2025)098

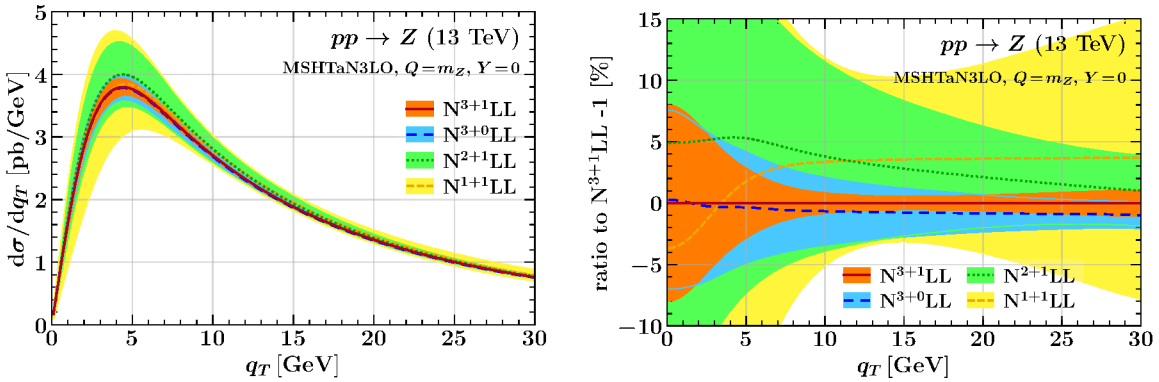

Figure 4. Leading-power $q_T \equiv p_T^Z$ spectrum for inclusive $pp \to Z$ production at the 13 TeV LHC at $\mathrm{N^{1+1}LL}$ (yellow), $\mathrm{N^{2+1}LL}$ (green), $\mathrm{N^{3+0}LL}$ (blue), and $\mathrm{N^{3+1}LL}$ (orange). The results are shown for the absolute spectrum on the left and as the relative difference to the $\mathrm{N^{3+1}LL}$ central value on the right. The bands show the total theory uncertainty at 95% theory CL.

have 2 more once we account for singlet contributions to the hard function, 1 more if we also count the β function, and several more once we account for the splitting functions.

Let us contrast this with a scale-variation based approach to estimate perturbative uncertainties. Even the most sophisticated currently available scale-variation-based setup [44] involves 5 scales (μ_H, μ_B, μ_S, ν_B, ν_S) that are being varied, which thus cannot begin to capture even the minimal space of theory uncertainties and correlations. And in most other approaches to q_T resummation even fewer scales are considered.

6.3 Numerical results

For our numerical results for the leading-power q_T spectrum, we consider inclusive $pp \to V$ production with $V = Z, W^\pm$ at the 13 TeV LHC at fixed invariant mass $Q \equiv \sqrt{q^2} = m_V$ and rapidity $Y = 0$ of the vector boson, unless noted otherwise. We use the `MSHT20an3lo` [13] PDF set with $\alpha_s(m_Z) = 0.118$. All results are obtained with `SCETlib` [90] based on its implementation of q_T resummation up to $\mathrm{N^4LL}$ [42–44, 47], which we have extended to support the required theory nuisance parameter variations. Since we only consider the leading-power spectrum without matching to the full fixed-order result at large q_T, we restrict ourselves to $q_T \leq 30\,\mathrm{GeV}$, where the neglected $\mathcal{O}(q_T^2/Q^2)$ and higher power corrections amount to at most a few-percent correction and the uncertainties associated with the matching procedure are also not yet relevant [44].

In figures 4, we start by presenting the Z q_T spectrum at different subsequent orders up to $\mathrm{N^{3+1}LL}$. The uncertainty bands show the total theory uncertainty at 95% theory CL from varying all TNPs by $\Delta u_n = \pm 2$. Note that the $\mathrm{N^{3+0}LL}$ result is an intermediate order, which is included for illustration and future reference.

In figures 5 we show the breakdown of the theory uncertainty by individual TNPs. The impacts on the spectrum from varying the TNPs up or down are roughly symmetric, so for clarity we always only show the $\Delta u_n = +1$ variation for the anomalous dimensions, hard function, and soft function, and the $\Delta u_n = -1$ variation for the beam function. Since each TNP corresponds to an independent source of uncertainty, which furthermore can be

JHEP08(2025)098

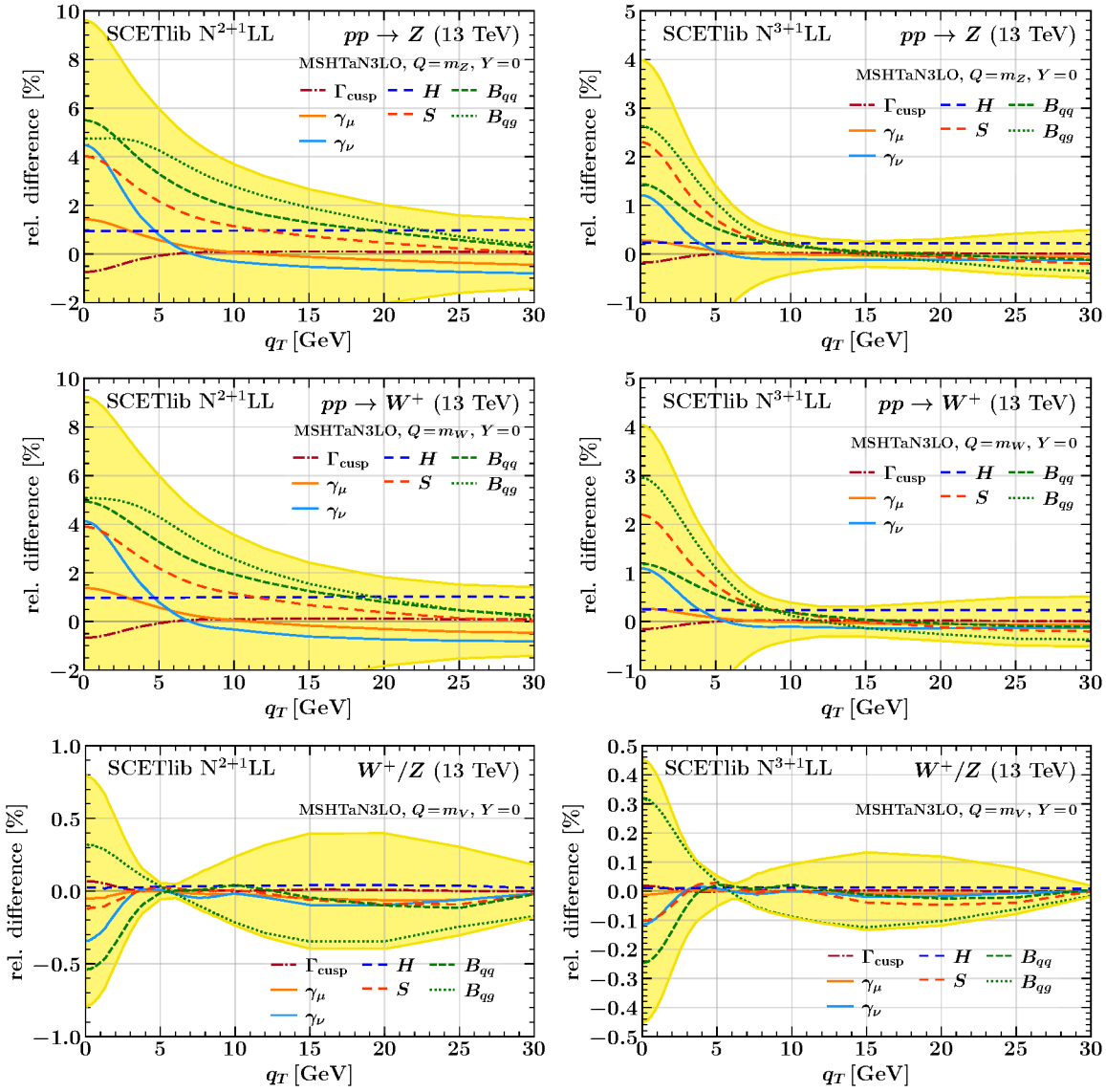

Figure 5. Breakdown of the relative uncertainties in the leading-power $q_T \equiv p_T^{Z,W}$ spectrum at the 13 TeV LHC at N^{2+1}LL (left panels) and N^{3+1}LL (right panels) for $pp \to Z$ (top row), $pp \to W^+$ (middle row), and their ratio (bottom row). The different lines show the impact of varying the corresponding theory nuisance parameter by $+1$ or -1, corresponding to 68% theory CL. The yellow band shows their sum in quadrature. See the text for more details.

considered Gaussian distributed (see section 5), the correct way to combine them into a total uncertainty is to add them in quadrature. This is shown by the light yellow band,[18] corresponding to the total uncertainty at 68% theory CL. In the top and middle rows we show the results for Z and W^+ and in the bottom row their ratio.

Concerning the point-by-point correlations in the shape of the q_T spectrum, each of the TNP variations shown by the different lines in figures 5 reflects a 100% correlated

[18]To be precise, in case of small asymmetries we sum in quadrature the larger of the up and down variations for each TNP for definiteness.

JHEP08(2025)098

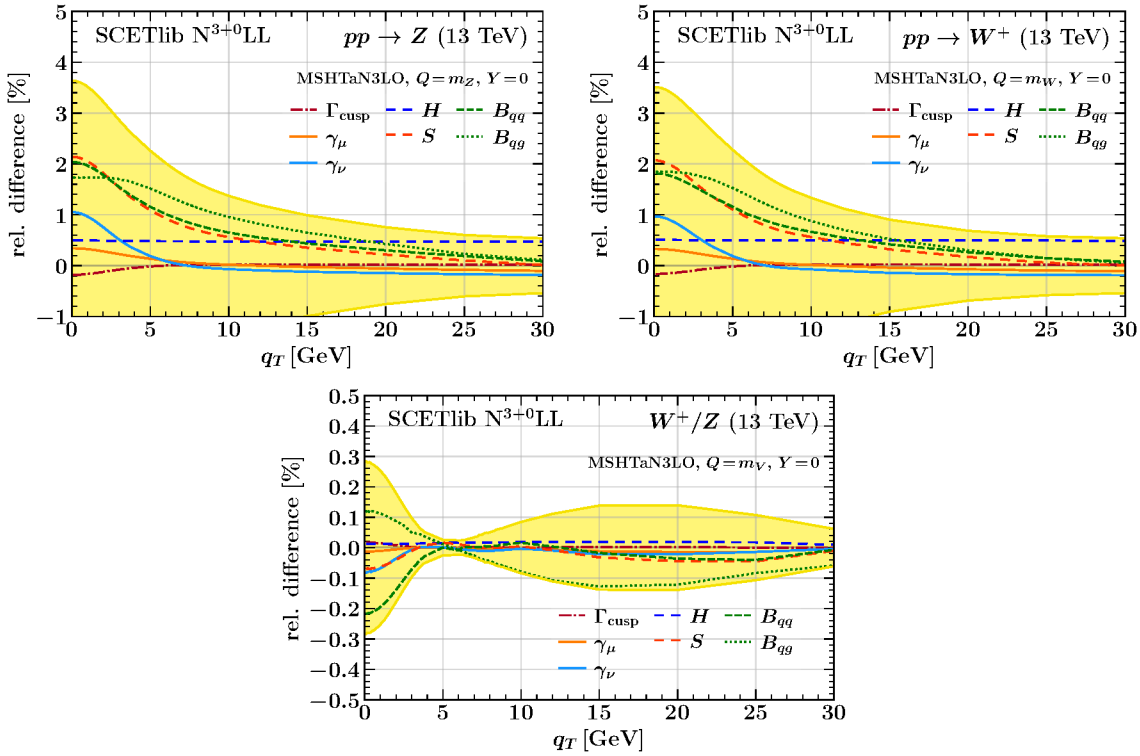

Figure 6. Same as figures 5 but for the approximate implementation at N^{3+0}LL.

uncertainty component across the q_T spectrum. We observe that the different components have quite different shapes and include cases where the correlation is always positive as well as cases switching sign from correlated to anticorrelated at different points in q_T. Hence, as anticipated, it is not possible to correctly model the correlations in the q_T spectrum by a 100% (anti)correlated hypothesis.

Let us briefly compare this to a scale-variation based approach. By scanning over different scale variations in order to account for the shape uncertainties, one is precisely scanning over various (more-or-less arbitrary) 100% (anti)correlated hypotheses. We remind the reader (see section 2.1.2) that when correlations matter, incorrect correlation assumptions can have dramatic consequences on the resulting uncertainties. In ref. [16], we will show explicitly that accounting for the correct point-by-point correlations in q_T is indeed absolutely critical if one wants to exploit precision measurements of the q_T spectrum, particularly its extremely precisely measured shape, for interpretation purposes such as extracting nonperturbative parameters or the strong coupling constant. All current attempts at doing so, including the recent analysis in ref. [85], are based on scale variations and are thus subject to uncontrolled correlation assumptions in the underlying perturbative predictions. Hence, the quoted perturbative theory uncertainties in the extracted parameters of interest cannot be taken at face value but must be interpreted with extreme caution.

As expected, the uncertainties are very similar for the closely related Z and W processes, whose main differences are the different partonic channel combinations and the small difference in their masses. Since the TNPs for both processes are the same, each of the individual

JHEP08(2025)098

impacts are 100% correlated between the processes, and as a result cancel in the ratio to very large extent, roughly by a factor of 10. We stress that whilst a large degree of cancellation is expected and has been encountered many times before, we can now correctly quantify it for the first time, and in particular also its dependence on q_T.

In figures 6 we show the same results using the approximate implementation of the theory nuisance parameters at $\mathrm{N}^{3+0}\mathrm{LL}$. This is the setup utilized for the resummed component of the q_T spectrum in the analysis of ref. [1], where it is also further matched to the fixed-order NLO_1 result for the V+1-parton process. At $\mathrm{N}^{3+0}\mathrm{LL}$, formally the same θ_n enter as at $\mathrm{N}^{3+1}\mathrm{LL}$ but their impacts are only approximately correct. Namely, their shape is approximated by the corresponding one at $\mathrm{N}^{2+1}\mathrm{LL}$, while their overall impact is similar to that at $\mathrm{N}^{3+1}\mathrm{LL}$. Whilst the precise shapes of the components differ between $\mathrm{N}^{3+1}\mathrm{LL}$ and $\mathrm{N}^{3+0}\mathrm{LL}$ their overall qualitative behaviour is similar. The total uncertainties at $\mathrm{N}^{3+0}\mathrm{LL}$ are similar to but somewhat larger than at full $\mathrm{N}^{3+1}\mathrm{LL}$. Notably, the uncertainties on the W/Z ratio, which strongly depend on the detailed correlations, are very similar to those at $\mathrm{N}^{3+1}\mathrm{LL}$. Therefore, we can conclude that the $\mathrm{N}^{3+0}\mathrm{LL}$ result provides a clear improvement over $\mathrm{N}^{2+1}\mathrm{LL}$ and a reasonable approximation to the more correct $\mathrm{N}^{3+1}\mathrm{LL}$ result. Although the $\mathrm{N}^{3+1}\mathrm{LL}$ result should be preferred, the approximate $\mathrm{N}^{3+0}\mathrm{LL}$ result can serve as a viable compromise if the former cannot be utilized for some reason. One such reason could be the availability of the required fixed-order matching at large q_T. Since $\mathrm{N}^{3+0}\mathrm{LL}$ implements the $\mathrm{N}^3\mathrm{LL}$ structure it can be consistently matched to NLO_1, whereas $\mathrm{N}^{3+1}\mathrm{LL}$ implements the full $\mathrm{N}^4\mathrm{LL}$ structure and therefore requires matching to NNLO_1.

In figures 7 we show the ratios of the q_T spectra for Z production at $Q = 1\,\mathrm{TeV}$ vs. $Q = m_Z$ and $Y = 1.6$ vs. $Y = 0$. Figures 8 shows the ratios of the q_T spectra for W^+ vs. W^- and for W^+ at 13 TeV vs. 7 TeV. The cancellation of uncertainties is expectedly most pronounced for W^+/W^-. It is weakest but still present for the case of $Q = 1\,\mathrm{TeV}$ vs. $Q = m_Z$. This is also not unexpected, since the spectrum mostly depends on q_T/Q, so for different Q the q_T spectra are shifted against each other.

We stress that the primary purpose of the various ratios we show is to easily visualize the effect of correlations and the resulting degree of cancellations. When correctly accounting for the theory correlations there is no difference as far as theory uncertainties are concerned in using the ratio or the quantities separately. In a real analysis, one would typically not use ratios but simply perform a combined analysis of all relevant processes, which constrains the TNPs among all of them accounting for all correlations and resulting cancellations. In the limit where one particular process is much more precisely measured than the others, one can think of it as effectively acting as a control process to obtain improved predictions for the others.

An important observation is that the dominant uncertainties that remain in the ratios and tend to cancel the least are those due to the beam functions, in particular for W/Z but also in many cases for the other ratios in figures 7 and 8. This is because the main difference between the processes, which is due to the different combinations of flavor channels, precisely enters via a different relative mix of different beam functions. This motivates a more detailed study of the beam function TNPs.

To conclude this section, we stress again that here we only consider the leading-power contributions. This is warranted as these are by far the dominant contributions to the spectrum

JHEP08(2025)098

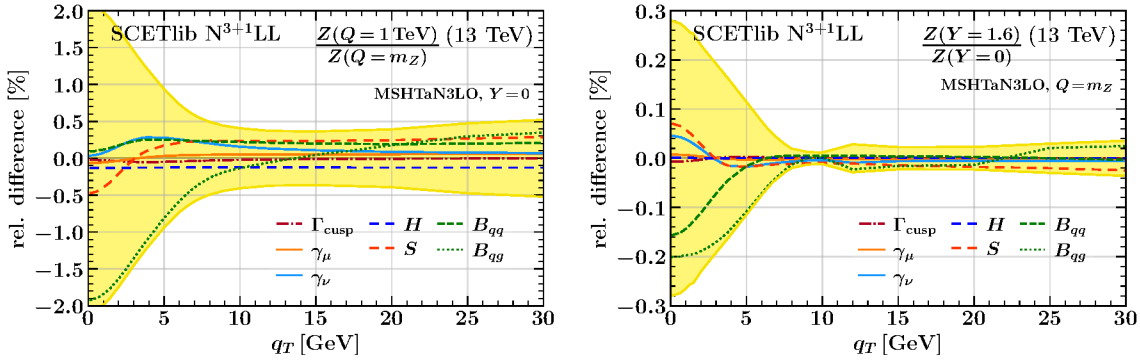

Figure 7. Relative uncertainties in the leading-power $q_T \equiv p_T^Z$ spectrum at the 13 TeV LHC at N^{3+1}LL for the ratio of $pp \to Z$ at $Q = 1$ TeV vs. $Q = m_Z$ (left) and $Y = 1.6$ vs. $Y = 0$ (right). The meaning of the curves is the same as in figures 5.

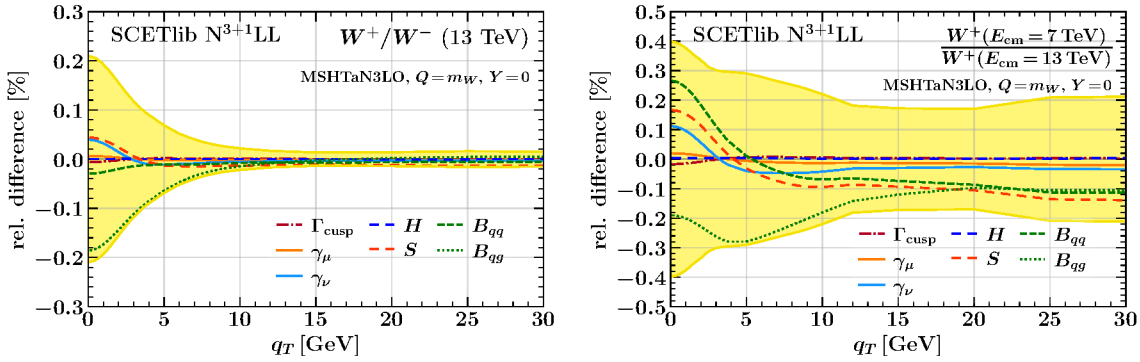

Figure 8. Relative uncertainties in the leading-power $q_T \equiv p_T^W$ spectrum at N^{3+1}LL for the ratio of $pp \to W^+$ vs. $pp \to W^-$ at the 13 TeV LHC (left) and $pp \to W^+$ at the 7 TeV LHC vs. 13 TeV LHC (right). The meaning of the curves is the same as in figures 5.

at small q_T, so the precise correlation and resulting cancellation of their uncertainties is a critical ingredient to any interpretation of precision measurements of the q_T spectrum, which we are able to properly take into account for the first time. The uncertainties in the ratios illustrate the level of precision that can now be reached via the cancellation of the dominant uncertainties in a combined analysis. At the resulting sub-percent level of precision, many other previously subleading effects can become equally or more important and must be accounted for to maintain this level of precision, motivating future work on them. We briefly comment on these in the next subsection.

6.4 Subleading effects

The application of our approach to the leading-power resummed contribution represents a crucial milestone toward a more complete and comprehensive understanding of the q_T spectrum. A complete treatment also requires accounting for several other subdominant effects. These are in particular:

- The neglected subleading power corrections in eq. (6.1) starting at $\mathcal{O}(q_T^2/Q^2)$.

JHEP08(2025)098

- Effects due to finite quark masses of $\mathcal{O}(m_q^2/q_T^2)$.
- QED and electroweak effects.
- Nonperturbative corrections of $\mathcal{O}(\Lambda_{\rm QCD}^2/q_T^2)$.

The nonperturbative corrections already have a parametric nature. In the future, our approach can also be applied systematically to the first three effects following the methodology developed in the previous sections. In the absence of their complete TNP-based treatment, they can still be included from existing results based on conventional methods. In other words, a TNP treatment of even just the leading-power resummed component is already extremely valuable, simply because it contributes by far the dominant uncertainties.

7 Conclusions

The theory nuisance parameter approach developed in this paper holds enormous potential to make perturbative predictions more robust and also more precise:

- It allows for the first time to correctly account for theory correlations, which are important whenever one simultaneously interprets multiple measurements (including different bins in a spectrum).
- The theory uncertainties and correlations are straightforward to propagate, like any other nuisance parameters, into fits, Monte-Carlo generators, multivariate analyses, neural networks, etc.
- In fits to experimental measurements, it is possible and consistent to profile the theory nuisance parameters and thereby constrain them, effectively reducing the theory uncertainties by the measurements, which is not possible with existing methods.
- New structures (e.g. partonic channels or additional logarithmic powers) appearing at higher order are explicitly anticipated and accounted for by the theory uncertainties.
- All new, even partial, higher-order information can have immediate phenomenological impact in reducing theory uncertainties, even if the complete next order is not yet available.
- The theory uncertainties have a well-defined and meaningful statistical interpretation.

Any estimate of a systematic (epistemic) uncertainty will have some level of arbitrariness arising from choices one has to make. An important goal and feature of the TNP approach is to systematically manage this arbitrariness and to minimize its impact on the final uncertainty estimate. In the TNP approach, there are roughly three types of choices involved:

- The perturbative scheme choices used to define the perturbative series. This scheme dependence has been discussed in section 3.4.

JHEP08(2025)098

- The choices required in deriving a suitable TNP parameterization in step 1. This includes which internal dependencies are parameterized, possible required approximations, and given these the actual choice of parameterization. These aspects have been discussed in section 4.
- How to normalize the TNPs and how to constrain their numerical values in step 2. These aspects have been discussed in section 3.3 and section 5.

When applied to color-singlet transverse-momentum (p_T) resummation, the theory nuisance parameters allow one to correctly and fully account for the theory correlations in the shape of the small-p_T spectrum, between different Q values, partonic channels, hard processes (e.g. W and Z production), collider energies, and different resummation-sensitive variables (e.g. p_T^V, p_T^ℓ near the Jacobian peak, or ϕ^*). In this context, our approach opens the door to reaching sub-percent level theoretical precision, which will be able to match the incredible precision already achieved by experimental measurements. To fully reach this level of theoretical precision, a variety of subleading effects must still be accounted for. We thus hope that our results also provide strong motivation for future work in this direction. For expedience, they can at first be included using conventional methods, which does not invalidate the TNP-based treatment of the dominant uncertainties. More importantly, our approach is also not limited to the dominant resummed contribution. It can be systematically and incrementally applied also to subleading effects as they become relevant at any given level of theory precision.

More generally, it will obviously be impossible to equip existing predictions with TNP-based uncertainties all at once. We should stress that this is also not required by the TNP approach. To the contrary, a more practical, incremental adoption, focusing on the dominant sources of uncertainties first is exactly in the spirit of our approach, namely to parameterize and include the sources of uncertainties in the theory predictions in order of their relevance.

Acknowledgments

I would like to thank Maarten Boonekamp for fruitful discussions that originally inspired this work. I am very grateful to my experimental colleagues, Simone Amoroso, Ludovica Aperio Bella, Josh Bendavid, Maarten Boonekamp, Stefano Camarda, Glen Cowan, Andre David, Daniel Froidevaux, Kenneth Long, Kerstin Tackmann, and many others, for numerous discussions and encouragement to pursue this work. I also thank Georgios Billis, Thomas Cridge, Giulia Marinelli, and Johannes Michel for discussions and collaboration on related work. I am indebted to Giulia Marinelli (to be repaid in chocolate muffins) for her generous help in preparing the numerical results in section 6. I thank Lawrence Berkeley National Laboratory for hospitality while parts of this work were carried out. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement 101002090 COLORFREE).

A Sample of known perturbative series

The QCD matrix-element constants and anomalous dimensions included in the samples of known perturbative series in section 5.3 are listed in tables 4 and 5. For definiteness, we

JHEP08(2025)098

give the explicit 1-loop results in the 3rd column. The 4th column shows the included loop orders and the last column gives the original references (starting at 3 loops for brevity). As already mentioned in section 5.3.2, we have made an effort to include all matrix-element constants known to four loops and anomalous dimensions known to four and five loops in QCD. The included quantities that are known at lower orders are certainly not exhaustive, and more can be added in the future.

Following the normalization conventions discussed in sections 5.2.1 and 5.2.2, we consider matching coefficients and jet and beam functions directly, while we consider the square root for decay rates and soft functions, and also include corresponding factors of 2 and 1/2 for the associated anomalous dimensions. For the beam function matching kernels, $\tilde{I}_{ij}(z)$, we reduce their z dependence by considering their z^1 moment (2nd Mellin moment). For $\tilde{I}_{qq}(z)$ we also consider its total integral (1st Mellin moment), which exists since this kernel does not have a $1/z$ singularity. Similarly, for the QCD splitting functions we consider their lowest moments in z, which in some cases are known to five loops.

For scheme-dependent quantities (decoupling constants, Wilson coefficients, beam, jet, and soft functions) we always use the $\overline{\text{MS}}$ scheme and canonical logarithms to define their scalar series for the constant terms or boundary conditions. That is, the constant terms are defined as the remaining nonlogarithmic terms when all logarithmic terms are written in terms of the respective canonical (possibly distributional) logarithms. These choices define the reference scheme for their corresponding TNPs. Some explicit examples with more details can be found in section 6.2. For the soft functions we only consider the quark functions, since the gluon ones are closely related to the quark ones by Casimir scaling. For the threshold and thrust soft functions we consider them both in position and momentum space, since the translation between spaces causes a significant reshuffling of the constant terms with the logarithmic terms such that the constants in either space become largely uncorrelated. On the other hand, for the q_T or b_T soft function we only consider the b_T-space result because the constant terms in q_T and b_T space appear very strongly correlated.

Concerning the anomalous dimensions, the QCD β function is defined as

$$\mu \frac{\mathrm{d}\alpha_s(\mu)}{\mathrm{d}\mu} = -2\alpha_s\,\beta(\alpha_s) \qquad \text{with} \qquad \beta_0 = \frac{11}{3}C_A - \frac{4}{3}T_F n_f\,. \tag{A.1}$$

Since it is the anomalous dimension of the coupling itself, it clearly plays a special role. For example, it is the only anomalous dimension whose n_f dependence starts at one loop. Despite its special role, we include it in our collection for completeness. A closely related anomalous dimension, which fits more naturally into our collection, is the anomalous dimension $\gamma_t(\alpha_s)$ of the ggH Wilson coefficient that arises from integrating out the top quark, which is given to all orders by

$$\gamma_t(\alpha_s) = -2\alpha_s^2 \frac{\mathrm{d}}{\mathrm{d}\alpha_s}\frac{\beta(\alpha_s)}{\alpha_s}\,, \qquad \gamma_{t\,n} = -2n\beta_n\,. \tag{A.2}$$

Note that there are several gluonic anomalous dimensions, whose n_f dependence also starts at one loop. However, this dependence (and similarly the highest power of n_f at higher orders) is always that of $\beta(\alpha_s)$ itself, which we therefore subtract. This can also be understood from the fact that the corresponding quantities are always associated with an explicit power of

JHEP08(2025)098

Name	f	f_1	n	refs.
α_s decoupling	ζ_α	0	1, 2, 3, 4	[91–94]
quark mass decoupling	ζ_m	-	2, 3, 4	[91, 94, 95]
ggH Wilson coefficient	C_1	$-\frac{8}{3}T_F$	1, 2, 3, 4	[91–94, 96]
$q\bar{q}H$ Wilson coefficient	C_2	-	2, 3, 4	[91, 95]
$ggHH$ Wilson coefficient	C_{HH}	$-\frac{8}{3}T_F$	3, 4	[94, 97, 98]
$\gamma^* \to q\bar{q}$ R-ratio (nonsinglet)	$\sqrt{R^{ns}_{q\bar{q}V}}$	$\frac{3}{2}\,C_F$	1, 2, 3, 4	[96, 99–102]
$\gamma^* \to q\bar{q}$ R-ratio (singlet)	$\sqrt{R^{s}_{q\bar{q}V}}$	-	3, 4	[96, 99–102]
$H \to gg$	$\sqrt{R_{gg}}$	$\frac{73}{6}\,C_A - \frac{14}{3}\,T_F n_f$	1, 2, 3, 4	[96, 103, 104]
$H \to q\bar{q}$ (nonsinglet)	$\sqrt{R_{q\bar{q}S}}$	$\frac{17}{2}\,C_F$	1, 2, 3, 4	[96, 105, 106]
quark vector form factor	c_{qqV}	$(-8 + \frac{\pi^2}{6})\,C_F$	1, 2, 3, 4	[48, 107–109]
gluon scalar form factor	c_{gg}	$\frac{\pi^2}{6}\,C_A$	1, 2, 3, 4	[48, 107–109]
quark scalar form factor	c_{qqS}	$(-2 + \frac{\pi^2}{6})\,C_F$	1, 2, 3, 4	[49, 110]
quark jet function	j_q	$(7 - \pi^2)\,C_F$	1, 2, 3	[111]
gluon jet function	j_g	$(\frac{67}{9} - \pi^2)\,C_A - \frac{20}{9}\,T_F n_f$	1, 2, 3	[112]
quark EEC jet function	j_q^{EEC}	$(4 - \frac{4\pi^2}{3})\,C_F$	1, 2, 3	[113]
gluon EEC jet function	j_g^{EEC}	$(\frac{65}{18} - \frac{4\pi^2}{3})\,C_A - \frac{5}{9}\,T_F n_f$	1, 2, 3	[113]
qq b_T beam fct (integral)	$\tilde{I}_{qq,1}$	C_F	1, 2, 3	[88, 89]
qq b_T beam fct (z^1 moment)	$\tilde{I}_{qq,2}$	$\frac{1}{3}\,C_F$	1, 2, 3	[88, 89]
qg b_T beam fct (z^1 moment)	$\tilde{I}_{qg,2}$	$\frac{1}{3}\,T_F$	1, 2, 3	[88, 89]
gg b_T beam fct (z^1 moment)	$\tilde{I}_{gg,2}$	0	1, 2, 3	[88, 89]
gq b_T beam fct (z^1 moment)	$\tilde{I}_{gq,2}$	$\frac{2}{3}\,C_F$	1, 2, 3	[88, 89]
b_T soft function	$\sqrt{\tilde{s}_q}$	$-\frac{\pi^2}{6}\,C_F$	1, 2, 3	[114]
threshold soft fct. (pos. space)	$\sqrt{\tilde{s}_{\text{thr}}}$	$\frac{\pi^2}{6}\,C_F$	1, 2, 3	[115]
threshold soft fct. (mom. space)	$\sqrt{s_{\text{thr}}}$	$-\frac{\pi^2}{6}\,C_F$	1, 2, 3	[115]
thrust soft fct. (pos. space)	$\sqrt{\tilde{s}_\tau}$	$-\frac{\pi^2}{2}\,C_F$	1, 2, 3	[116]
thrust soft fct. (mom. space)	$\sqrt{s_\tau}$	$\frac{\pi^2}{6}\,C_F$	1, 2, 3	[116]
heavy-light soft fct.	$\sqrt{s_{hl}}$	$-\frac{\pi^2}{12}\,C_F$	1, 2, 3	[117]

Table 4. Quantities included in our sample of known matrix-element constants.

α_s at the lowest order, or equivalently, the corresponding operators involve a gluon field strength. It would actually be more natural to always include an appropriate power of the coupling with the field strength, which would then automatically remove the $\beta(\alpha_s)$ piece from the anomalous dimension.

Consistency of the e^+e^- thrust [19, 20] and partonic beam-thrust [141] factorization implies

$$4\gamma^i_C(\alpha_s) + 2\gamma^i_J(\alpha_s) + \gamma^i_S(\alpha_s) = 0\,, \tag{A.3}$$

where $\gamma^i_S(\alpha_s)$ is the (noncusp) anomalous dimension of the (beam)thrust soft function, and we have already used that the anomalous dimensions of the SCET_I inclusive beam and jet function

JHEP08(2025)098

Name	γ	γ_0	n	refs.
QCD β function	-2β	$-\frac{22}{3}C_A + \frac{8}{3}T_F n_f$	$0,1,2,3,4$	[50–56]
ggH Wilson coefficient	γ_t	0	$0,1,2,3,4$	
quark mass	γ_m	$-6C_F$	$0,1,2,3,4$	[57–62]
vector correlator (nonsinglet)	$2\gamma_V^{\rm ns}$	$\frac{8}{3}d_F$	$0,1,2,3,4$	[102]
vector correlator (singlet)	$2\gamma_V^{\rm s}$	-	$3,4$	[102]
scalar correlator	$2\gamma_S$	$4d_F$	$0,1,2,3$	[105]
$P_{\rm ns}^+(z)$ (z^1 moment)	$2\gamma_2^{\rm ns+}$	$-\frac{16}{3}\,C_F$	$0,1,2,3,4$	[67, 118–123]
$P_{\rm ns}^-(z)$ (z^2 moment)	$2\gamma_3^{\rm ns-}$	$-\frac{25}{3}\,C_F$	$0,1,2,3,4$	[67, 118, 121–124]
$P_{gg}(z)$ (z^1 moment)	$2\gamma_2^{gg} + 2\beta$	$-\frac{22}{3}\,C_A$	$0,1,2,3$	[125, 126]
$P_{qg}(z)$ (z^1 moment)	$2\gamma_2^{qg}$	$\frac{8}{3}\,T_F$	$0,1,2,3$	[125–127]
quark cusp	$2\Gamma^q_{\rm cusp}$	$8\,C_F$	$0,1,2,3,(4)$	[63–67]
gluon cusp	$2\Gamma^g_{\rm cusp}$	$8\,C_A$	3	[63–66]
tensor current	γ_T	$2\,C_F$	$0,1,2,3$	[119, 128, 129]
HQET heavy-light current	$\gamma_{\rm HQET}$	$-3\,C_F$	$0,1,2,3$	[130, 131]
quark threshold PDF	γ_f^q	$6\,C_F$	$0,1,2,3$	[63, 122, 123, 132]
gluon threshold PDF	$\gamma_f^g - 2\beta$	0	$0,1,2,3$	[64, 133]
quark collinear	$2\gamma_C^q$	$-6\,C_F$	$0,1,2,3$	[66, 134, 135]
gluon collinear	$2\gamma_C^g + 2\beta$	0	$0,1,2,3$	[66, 135, 136]
heavy-quark collinear	$2\gamma_C^Q$	$-4\,C_F$	$0,1,2$	[117, 137]
quark jet function	γ_J^q	$6\,C_F$	$0,1,2,3$	[111]
gluon jet function	$\gamma_J^g - 2\beta$	0	$0,1,2,3$	
quark soft function	$\gamma_S^q/2$	0	$0,1,2,3$	[115, 122, 132]
gluon soft function	$\gamma_S^g/2$	0	3	[115, 133]
heavy-light soft function	$\gamma_S^Q/2$	$2\,C_F$	$0,1,2$	[117]
quark rapidity	$\tilde\gamma_\nu^q/2$	0	$0,1,2,3$	[114, 138–140]
gluon rapidity	$\tilde\gamma_\nu^g/2$	0	3	[114, 138–140]

Table 5. Quantities included in our sample of known anomalous dimensions.

are equal, $\gamma_B = \gamma_J$ [142]. Consistency of color-singlet threshold factorization [143, 144] implies

$$4\gamma_C^i(\alpha_s) + 2\gamma_f^i(\alpha_s) + \gamma_{\rm thr}^i(\alpha_s) = 0\,. \tag{A.4}$$

Consistency of the generalized threshold factorization [145] implies

$$4\gamma_C^i(\alpha_s) + \gamma_f^i(\alpha_s) + \gamma_J^i(\alpha_s) = 0\,. \tag{A.5}$$

We thus have 3 relations for 5 anomalous dimensions, which means only 2 are independent. In particular, we have

$$\gamma_S^i(\alpha_s) = -\gamma_{\rm thr}^i(\alpha_s) = \gamma_f^i(\alpha_s) - \gamma_J^i(\alpha_s)\,. \tag{A.6}$$

At three loops, γ_C^i and γ_f^i have been known first [63, 64, 134, 136], with γ_J^i, γ_B^i, γ_S^i, $\gamma_{\rm thr}^i$ determined from consistency. Subsequently, $\gamma_{\rm thr}^i$ and γ_J^q have been confirmed by independent

JHEP08(2025)098

explicit calculations [111, 115]. At four loops, γ^i_C and $\gamma^i_{\rm thr}$ are fully known [66, 122, 132, 135], where for the latter the coefficients of some color structures are only known numerically but with sufficient precision for practical purposes. We use these to obtain γ^i_f, γ^i_J, γ^i_S at four loops. In particular, doing so determines the remaining color coefficients in γ^i_f that were only available approximately in refs. [132, 133], see also ref. [146]. To our knowledge, the four-loop γ^g_J had not been considered in the literature so far.

For the cusp, soft, and rapidity anomalous dimensions, we do not include the gluon coefficients up to 3-loop order as they are trivially related to the quark ones by a simple overall Casimir scaling, $\gamma^g_n = C_A/C_F\gamma^q_n$ for $n \leq 2$. At 4-loop order, $n = 3$, the quark and gluon coefficients are still related by generalized Casimir scaling, which however no longer relates the coefficients as a whole, so we include both.

Data Availability Statement. This article has no associated data or the data will not be deposited.

Code Availability Statement. This article has no associated code or the code will not be deposited.

References

[1] CMS collaboration, *High-precision measurement of the W boson mass with the CMS experiment at the LHC*, arXiv:2412.13872 [INSPIRE].

[2] J. Charles, S. Descotes-Genon, V. Niess and L. Vale Silva, *Modeling theoretical uncertainties in phenomenological analyses for particle physics*, *Eur. Phys. J. C* **77** (2017) 214 [arXiv:1611.04768] [INSPIRE].

[3] G. Cowan, *Statistical Models with Uncertain Error Parameters*, *Eur. Phys. J. C* **79** (2019) 133 [arXiv:1809.05778] [INSPIRE].

[4] M. Cacciari and N. Houdeau, *Meaningful characterisation of perturbative theoretical uncertainties*, *JHEP* **09** (2011) 039 [arXiv:1105.5152] [INSPIRE].

[5] E. Bagnaschi, M. Cacciari, A. Guffanti and L. Jenniches, *An extensive survey of the estimation of uncertainties from missing higher orders in perturbative calculations*, *JHEP* **02** (2015) 133 [arXiv:1409.5036] [INSPIRE].

[6] M. Bonvini, *Probabilistic definition of the perturbative theoretical uncertainty from missing higher orders*, *Eur. Phys. J. C* **80** (2020) 989 [arXiv:2006.16293] [INSPIRE].

[7] C. Duhr, A. Huss, A. Mazeliauskas and R. Szafron, *An analysis of Bayesian estimates for missing higher orders in perturbative calculations*, *JHEP* **09** (2021) 122 [arXiv:2106.04585] [INSPIRE].

[8] A. David and G. Passarino, *How well can we guess theoretical uncertainties?*, *Phys. Lett. B* **726** (2013) 266 [arXiv:1307.1843] [INSPIRE].

[9] A. Ghosh et al., *Statistical patterns of theory uncertainties*, *SciPost Phys. Core* **6** (2023) 045 [arXiv:2210.15167] [INSPIRE].

JHEP08(2025)098

[10] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [*Erratum ibid.* **73** (2013) 2501] [arXiv:1007.1727] [INSPIRE].

[11] R.D. Cousins and L. Wasserman, *PHYSTAT Informal Review: Marginalizing versus Profiling of Nuisance Parameters*, arXiv:2404.17180 [INSPIRE].

[12] F.J. Tackmann, *Theory Uncertainties from Nuisance Parameters*, talk given at *SCET 2019 workshop*, UCSD Natural Sciences Building, San Diego, U.S.A., 27 March 2019, https://indico.physics.lbl.gov/event/694/contributions/3344/.

[13] J. McGowan, T. Cridge, L.A. Harland-Lang and R.S. Thorne, *Approximate N^3LO parton distribution functions with theoretical uncertainties: MSHT20aN^3LO PDFs*, *Eur. Phys. J. C* **83** (2023) 185 [*Erratum ibid.* **83** (2023) 302] [arXiv:2207.04739] [INSPIRE].

[14] B. Dehnadi, I. Novikov and F.J. Tackmann, *The photon energy spectrum in $B \to X_s\gamma$ at N^3LL'*, *JHEP* **07** (2023) 214 [arXiv:2211.07663] [INSPIRE].

[15] P. Cal et al., *Jet veto resummation for STXS H+1-jet bins at aNNLL$'$+NNLO*, *JHEP* **03** (2025) 155 [arXiv:2408.13301] [INSPIRE].

[16] T. Cridge, G. Marinelli and F.J. Tackmann, *Theory Uncertainties in the Extraction of α_s from Drell-Yan at Small Transverse Momentum*, arXiv:2506.13874 [INSPIRE].

[17] M.A. Lim and R. Poncelet, *Robust estimates of theoretical uncertainties at fixed-order in perturbation theory*, arXiv:2412.14910 [INSPIRE].

[18] S. Moch, J.A.M. Vermaseren and A. Vogt, *Higher-order corrections in threshold resummation*, *Nucl. Phys. B* **726** (2005) 317 [hep-ph/0506288] [INSPIRE].

[19] T. Becher and M.D. Schwartz, *A precise determination of α_s from LEP thrust data using effective field theory*, *JHEP* **07** (2008) 034 [arXiv:0803.0342] [INSPIRE].

[20] R. Abbate et al., *Thrust at N^3LL with Power Corrections and a Precision Global Fit for $\alpha_s(mZ)$*, *Phys. Rev. D* **83** (2011) 074021 [arXiv:1006.3080] [INSPIRE].

[21] T. Becher, M. Neubert and L. Rothen, *Factorization and N^3LL_p+NNLO predictions for the Higgs cross section with a jet veto*, *JHEP* **10** (2013) 125 [arXiv:1307.0025] [INSPIRE].

[22] M. Bonvini and S. Marzani, *Resummed Higgs cross section at N^3LL*, *JHEP* **09** (2014) 007 [arXiv:1405.3654] [INSPIRE].

[23] A.H. Hoang, D.W. Kolodrubetz, V. Mateu and I.W. Stewart, *C-parameter distribution at N^3LL' including power corrections*, *Phys. Rev. D* **91** (2015) 094017 [arXiv:1411.6633] [INSPIRE].

[24] P.J. Mohr, B.N. Taylor and D.B. Newell, *CODATA Recommended Values of the Fundamental Physical Constants: 2010*, *Rev. Mod. Phys.* **84** (2012) 1527 [arXiv:1203.5425] [INSPIRE].

[25] S. Sturm et al., *High-precision measurement of the atomic mass of the electron*, *Nature* **506** (2014) 467 [arXiv:1406.5590] [INSPIRE].

[26] L. Berthier and M. Trott, *Consistent constraints on the Standard Model Effective Field Theory*, *JHEP* **02** (2016) 069 [arXiv:1508.05060] [INSPIRE].

[27] S. Alte, M. König and W. Shepherd, *Consistent Searches for SMEFT Effects in Non-Resonant Dijet Events*, *JHEP* **01** (2018) 094 [arXiv:1711.07484] [INSPIRE].

[28] M. Trott, *Methodology for theory uncertainties in the standard model effective field theory*, *Phys. Rev. D* **104** (2021) 095023 [arXiv:2106.13794] [INSPIRE].

JHEP08(2025)098

[29] A. Ghosh and B. Nachman, *A cautionary tale of decorrelating theory uncertainties*, *Eur. Phys. J. C* **82** (2022) 46 [arXiv:2109.08159] [INSPIRE].

[30] E. Canonero, A.R. Brazzale and G. Cowan, *Higher-order asymptotic corrections and their application to the Gamma Variance Model*, *Eur. Phys. J. C* **83** (2023) 1100 [arXiv:2304.10574] [INSPIRE].

[31] I.W. Stewart and F.J. Tackmann, *Theory Uncertainties for Higgs and Other Searches Using Jet Bins*, *Phys. Rev. D* **85** (2012) 034011 [arXiv:1107.2117] [INSPIRE].

[32] A. Banfi, G.P. Salam and G. Zanderighi, *NLL+NNLO predictions for jet-veto efficiencies in Higgs-boson and Drell-Yan production*, *JHEP* **06** (2012) 159 [arXiv:1203.5773] [INSPIRE].

[33] S. Gangal and F.J. Tackmann, *Next-to-leading-order uncertainties in Higgs+2 jets from gluon fusion*, *Phys. Rev. D* **87** (2013) 093008 [arXiv:1302.5437] [INSPIRE].

[34] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, *CERN Yellow Rep. Monogr.* **2** (2017) 1 [arXiv:1610.07922] [INSPIRE].

[35] J.R. Andersen et al., *Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report*, arXiv:1803.07977 [INSPIRE].

[36] J.M. Lindert et al., *Precise predictions for $V+$ jets dark matter backgrounds*, *Eur. Phys. J. C* **77** (2017) 829 [arXiv:1705.04664] [INSPIRE].

[37] L.A. Harland-Lang and R.S. Thorne, *On the Consistent Use of Scale Variations in PDF Fits and Predictions*, *Eur. Phys. J. C* **79** (2019) 225 [arXiv:1811.08434] [INSPIRE].

[38] NNPDF collaboration, *Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies*, *Eur. Phys. J. C* **79** (2019) 931 [arXiv:1906.10698] [INSPIRE].

[39] C.F. Berger et al., *Higgs Production with a Central Jet Veto at NNLL+NNLO*, *JHEP* **04** (2011) 092 [arXiv:1012.4480] [INSPIRE].

[40] I.W. Stewart, F.J. Tackmann, J.R. Walsh and S. Zuberi, *Jet p_T resummation in Higgs production at $NNLL' + NNLO$*, *Phys. Rev. D* **89** (2014) 054001 [arXiv:1307.1808] [INSPIRE].

[41] W. Bizon et al., *The transverse momentum spectrum of weak gauge bosons at $N^3LL + NNLO$*, *Eur. Phys. J. C* **79** (2019) 868 [arXiv:1905.05171] [INSPIRE].

[42] M.A. Ebert, J.K.L. Michel, I.W. Stewart and F.J. Tackmann, *Drell-Yan q_T resummation of fiducial power corrections at N^3LL*, *JHEP* **04** (2021) 102 [arXiv:2006.11382] [INSPIRE].

[43] G. Billis et al., *Higgs pT Spectrum and Total Cross Section with Fiducial Cuts at Third Resummed and Fixed Order in QCD*, *Phys. Rev. Lett.* **127** (2021) 072001 [arXiv:2102.08039] [INSPIRE].

[44] G. Billis, J.K.L. Michel and F.J. Tackmann, *Drell-Yan transverse-momentum spectra at N^3LL' and approximate N^4LL with SCETlib*, *JHEP* **02** (2025) 170 [arXiv:2411.16004] [INSPIRE].

[45] L.N. Trefethen, *Approximation Theory and Approximation Practice, Extended Edition*, Society for Industrial and Applied Mathematics (2019) [DOI:10.1137/1.9781611975949].

[46] Z. Ligeti, I.W. Stewart and F.J. Tackmann, *Treating the b quark distribution function with reliable uncertainties*, *Phys. Rev. D* **78** (2008) 114014 [arXiv:0807.1926] [INSPIRE].

[47] G. Billis, M.A. Ebert, J.K.L. Michel and F.J. Tackmann, *A toolbox for q_T and 0-jettiness subtractions at N^3LO*, *Eur. Phys. J. Plus* **136** (2021) 214 [arXiv:1909.00811] [INSPIRE].

JHEP08(2025)098

[48] R.N. Lee et al., *Quark and Gluon Form Factors in Four-Loop QCD*, *Phys. Rev. Lett.* **128** (2022) 212002 [arXiv:2202.04660] [INSPIRE].

[49] A. Chakraborty et al., *Hbb vertex at four loops and hard matching coefficients in SCET for various currents*, *Phys. Rev. D* **106** (2022) 074009 [arXiv:2204.02422] [INSPIRE].

[50] O.V. Tarasov, A.A. Vladimirov and A.Y. Zharkov, *The Gell-Mann-Low Function of QCD in the Three Loop Approximation*, *Phys. Lett. B* **93** (1980) 429 [INSPIRE].

[51] S.A. Larin and J.A.M. Vermaseren, *The Three loop QCD Beta function and anomalous dimensions*, *Phys. Lett. B* **303** (1993) 334 [hep-ph/9302208] [INSPIRE].

[52] T. van Ritbergen, J.A.M. Vermaseren and S.A. Larin, *The four loop beta function in quantum chromodynamics*, *Phys. Lett. B* **400** (1997) 379 [hep-ph/9701390] [INSPIRE].

[53] M. Czakon, *The Four-loop QCD beta-function and anomalous dimensions*, *Nucl. Phys. B* **710** (2005) 485 [hep-ph/0411261] [INSPIRE].

[54] P.A. Baikov, K.G. Chetyrkin and J.H. Kühn, *Five-Loop Running of the QCD Coupling Constant*, *Phys. Rev. Lett.* **118** (2017) 082002 [arXiv:1606.08659] [INSPIRE].

[55] F. Herzog et al., *The five-loop beta function of Yang-Mills theory with fermions*, *JHEP* **02** (2017) 090 [arXiv:1701.01404] [INSPIRE].

[56] T. Luthe, A. Maier, P. Marquard and Y. Schröder, *The five-loop Beta function for a general gauge group and anomalous dimensions beyond Feynman gauge*, *JHEP* **10** (2017) 166 [arXiv:1709.07718] [INSPIRE].

[57] O.V. Tarasov, *Anomalous dimensions of quark masses in the three-loop approximation*, *Phys. Part. Nucl. Lett.* **17** (2020) 109 [arXiv:1910.12231] [INSPIRE].

[58] S.A. Larin, *The renormalization of the axial anomaly in dimensional regularization*, *Phys. Lett. B* **303** (1993) 113 [hep-ph/9302240] [INSPIRE].

[59] K.G. Chetyrkin, *Quark mass anomalous dimension to $O(\alpha_s^4)$*, *Phys. Lett. B* **404** (1997) 161 [hep-ph/9703278] [INSPIRE].

[60] J.A.M. Vermaseren, S.A. Larin and T. van Ritbergen, *The 4-loop quark mass anomalous dimension and the invariant quark mass*, *Phys. Lett. B* **405** (1997) 327 [hep-ph/9703284] [INSPIRE].

[61] P.A. Baikov, K.G. Chetyrkin and J.H. Kühn, *Quark Mass and Field Anomalous Dimensions to $\mathcal{O}(\alpha_s^5)$*, *JHEP* **10** (2014) 076 [arXiv:1402.6611] [INSPIRE].

[62] T. Luthe, A. Maier, P. Marquard and Y. Schröder, *Five-loop quark mass and field anomalous dimensions for a general gauge group*, *JHEP* **01** (2017) 081 [arXiv:1612.05512] [INSPIRE].

[63] S. Moch, J.A.M. Vermaseren and A. Vogt, *The three loop splitting functions in QCD: the Nonsinglet case*, *Nucl. Phys. B* **688** (2004) 101 [hep-ph/0403192] [INSPIRE].

[64] A. Vogt, S. Moch and J.A.M. Vermaseren, *The Three-loop splitting functions in QCD: the Singlet case*, *Nucl. Phys. B* **691** (2004) 129 [hep-ph/0404111] [INSPIRE].

[65] J.M. Henn, G.P. Korchemsky and B. Mistlberger, *The full four-loop cusp anomalous dimension in $\mathcal{N} = 4$ super Yang-Mills and QCD*, *JHEP* **04** (2020) 018 [arXiv:1911.10174] [INSPIRE].

[66] A. von Manteuffel, E. Panzer and R.M. Schabinger, *Cusp and collinear anomalous dimensions in four-loop QCD from form factors*, *Phys. Rev. Lett.* **124** (2020) 162001 [arXiv:2002.04617] [INSPIRE].

JHEP08(2025)098

[67] F. Herzog et al., *Five-loop contributions to low-N non-singlet anomalous dimensions in QCD*, *Phys. Lett. B* **790** (2019) 436 [arXiv:1812.11818] [INSPIRE].

[68] ATLAS collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09** (2014) 145 [arXiv:1406.3660] [INSPIRE].

[69] ATLAS collaboration, *Measurement of the transverse momentum and ϕ^*_η distributions of Drell-Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 291 [arXiv:1512.02192] [INSPIRE].

[70] ATLAS collaboration, *Measurement of the transverse momentum distribution of Drell-Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 616 [arXiv:1912.02844] [INSPIRE].

[71] ATLAS collaboration, *A precise measurement of the Z-boson double-differential transverse momentum and rapidity distributions in the full phase space of the decay leptons with the ATLAS experiment at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J. C* **84** (2024) 315 [arXiv:2309.09318] [INSPIRE].

[72] CMS collaboration, *Measurement of the Rapidity and Transverse Momentum Distributions of Z Bosons in pp Collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. D* **85** (2012) 032002 [arXiv:1110.4973] [INSPIRE].

[73] CMS collaboration, *Measurement of the transverse momentum spectra of weak vector bosons produced in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **02** (2017) 096 [arXiv:1606.05864] [INSPIRE].

[74] CMS collaboration, *Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **12** (2019) 061 [arXiv:1909.04133] [INSPIRE].

[75] LHCb collaboration, *Measurement of forward W and Z boson production in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **01** (2016) 155 [arXiv:1511.08039] [INSPIRE].

[76] LHCb collaboration, *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **09** (2016) 136 [arXiv:1607.06495] [INSPIRE].

[77] CDF collaboration, *High-precision measurement of the W boson mass with the CDF II detector*, *Science* **376** (2022) 170 [INSPIRE].

[78] ATLAS collaboration, *Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **78** (2018) 110 [*Erratum ibid.* **78** (2018) 898] [arXiv:1701.07240] [INSPIRE].

[79] ATLAS collaboration, *Measurement of the W-boson mass and width with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J. C* **84** (2024) 1309 [arXiv:2403.15085] [INSPIRE].

[80] LHCb collaboration, *Measurement of the W boson mass*, *JHEP* **01** (2022) 036 [arXiv:2109.01113] [INSPIRE].

[81] MAP (Multi-dimensional Analyses of Partonic distributions) collaboration, *Unpolarized transverse momentum distributions from a global fit of Drell-Yan and semi-inclusive deep-inelastic scattering data*, *JHEP* **10** (2022) 127 [arXiv:2206.07598] [INSPIRE].

[82] V. Moos, I. Scimemi, A. Vladimirov and P. Zurita, *Extraction of unpolarized transverse momentum distributions from the fit of Drell-Yan data at N^4LL*, *JHEP* **05** (2024) 036 [arXiv:2305.07473] [INSPIRE].

JHEP08(2025)098

[83] MAP (Multi-dimensional Analyses of Partonic distributions) collaboration, *Flavor dependence of unpolarized quark transverse momentum distributions from a global fit*, *JHEP* **08** (2024) 232 [arXiv:2405.13833] [INSPIRE].

[84] S. Camarda, G. Ferrera and M. Schott, *Determination of the strong-coupling constant from the Z-boson transverse-momentum distribution*, *Eur. Phys. J. C* **84** (2024) 39 [arXiv:2203.05394] [INSPIRE].

[85] ATLAS collaboration, *A precise determination of the strong-coupling constant from the recoil of Z bosons with the ATLAS experiment at* $\sqrt{s} = 8$ *TeV*, arXiv:2309.12986 [INSPIRE].

[86] J.-Y. Chiu, A. Jain, D. Neill and I.Z. Rothstein, *A formalism for the Systematic Treatment of Rapidity Logarithms in Quantum Field Theory*, *JHEP* **05** (2012) 084 [arXiv:1202.0814] [INSPIRE].

[87] Y. Li, D. Neill and H.X. Zhu, *An exponential regulator for rapidity divergences*, *Nucl. Phys. B* **960** (2020) 115193 [arXiv:1604.00392] [INSPIRE].

[88] M.A. Ebert, B. Mistlberger and G. Vita, *Transverse momentum dependent PDFs at* N^3LO, *JHEP* **09** (2020) 146 [arXiv:2006.05329] [INSPIRE].

[89] M.-X. Luo, T.-Z. Yang, H.X. Zhu and Y.J. Zhu, *Unpolarized quark and gluon TMD PDFs and FFs at* N^3LO, *JHEP* **06** (2021) 115 [arXiv:2012.03256] [INSPIRE].

[90] M.A. Ebert et al., *SCETlib: a C++ Package for Numerical Calculations in QCD and Soft-Collinear Effective Theory*, *DESY-17-099* (2018).

[91] K.G. Chetyrkin, B.A. Kniehl and M. Steinhauser, *Decoupling relations to* $O(\alpha_s^3)$ *and their connection to low-energy theorems*, *Nucl. Phys. B* **510** (1998) 61 [hep-ph/9708255] [INSPIRE].

[92] Y. Schroder and M. Steinhauser, *Four-loop decoupling relations for the strong coupling*, *JHEP* **01** (2006) 051 [hep-ph/0512058] [INSPIRE].

[93] K.G. Chetyrkin, J.H. Kuhn and C. Sturm, *QCD decoupling at four loops*, *Nucl. Phys. B* **744** (2006) 121 [hep-ph/0512060] [INSPIRE].

[94] M. Gerlach, F. Herren and M. Steinhauser, *Wilson coefficients for Higgs boson production and decoupling relations to* $\mathcal{O}\left(\alpha_s^4\right)$, *JHEP* **11** (2018) 141 [arXiv:1809.06787] [INSPIRE].

[95] T. Liu and M. Steinhauser, *Decoupling of heavy quarks at four loops and effective Higgs-fermion coupling*, *Phys. Lett. B* **746** (2015) 330 [arXiv:1502.04719] [INSPIRE].

[96] F. Herzog et al., *On Higgs decays to hadrons and the R-ratio at* N^4LO, *JHEP* **08** (2017) 113 [arXiv:1707.01044] [INSPIRE].

[97] J. Grigo, K. Melnikov and M. Steinhauser, *Virtual corrections to Higgs boson pair production in the large top quark mass limit*, *Nucl. Phys. B* **888** (2014) 17 [arXiv:1408.2422] [INSPIRE].

[98] M. Spira, *Effective Multi-Higgs Couplings to Gluons*, *JHEP* **10** (2016) 026 [arXiv:1607.05548] [INSPIRE].

[99] S.G. Gorishnii, A.L. Kataev and S.A. Larin, *The* $O(\alpha_s^3)$*-corrections to* $\sigma_{\text{tot}}(e^+e^- \to \text{hadrons})$ *and* $\Gamma(\tau^- \to \nu_\tau + \text{hadrons})$ *in QCD*, *Phys. Lett. B* **259** (1991) 144 [INSPIRE].

[100] L.R. Surguladze and M.A. Samuel, *Total hadronic cross-section in* e^+e^- *annihilation at the four loop level of perturbative QCD*, *Phys. Rev. Lett.* **66** (1991) 560 [*Erratum ibid.* **66** (1991) 2416] [INSPIRE].

[101] P.A. Baikov, K.G. Chetyrkin and J.H. Kuhn, *Order* α_s^4 *QCD Corrections to Z and tau Decays*, *Phys. Rev. Lett.* **101** (2008) 012002 [arXiv:0801.1821] [INSPIRE].

JHEP08(2025)098

[102] P.A. Baikov, K.G. Chetyrkin, J.H. Kuhn and J. Rittinger, *Vector Correlator in Massless QCD at Order $\mathcal{O}(\alpha_s^4)$ and the QED beta-function at Five Loop*, *JHEP* **07** (2012) 017 [arXiv:1206.1284] [INSPIRE].

[103] P.A. Baikov and K.G. Chetyrkin, *Top Quark Mediated Higgs Boson Decay into Hadrons to Order α_s^5*, *Phys. Rev. Lett.* **97** (2006) 061803 [hep-ph/0604194] [INSPIRE].

[104] S. Moch and A. Vogt, *On third-order timelike splitting functions and top-mediated Higgs decay into hadrons*, *Phys. Lett. B* **659** (2008) 290 [arXiv:0709.3899] [INSPIRE].

[105] K.G. Chetyrkin, *Correlator of the quark scalar currents and $\Gamma_{\rm tot}(H \to hadrons)$ at $O(\alpha_s^3)$ in pQCD*, *Phys. Lett. B* **390** (1997) 309 [hep-ph/9608318] [INSPIRE].

[106] P.A. Baikov, K.G. Chetyrkin and J.H. Kuhn, *Scalar correlator at $O(\alpha_s^4)$, Higgs decay into b-quarks and bounds on the light quark masses*, *Phys. Rev. Lett.* **96** (2006) 012003 [hep-ph/0511063] [INSPIRE].

[107] P.A. Baikov et al., *Quark and gluon form factors to three loops*, *Phys. Rev. Lett.* **102** (2009) 212002 [arXiv:0902.3519] [INSPIRE].

[108] R.N. Lee, A.V. Smirnov and V.A. Smirnov, *Analytic Results for Massless Three-Loop Form Factors*, *JHEP* **04** (2010) 020 [arXiv:1001.2887] [INSPIRE].

[109] T. Gehrmann et al., *Calculation of the quark and gluon form factors to three loops in QCD*, *JHEP* **06** (2010) 094 [arXiv:1004.3653] [INSPIRE].

[110] T. Gehrmann and D. Kara, *The $Hb\bar{b}$ form factor to three loops in QCD*, *JHEP* **09** (2014) 174 [arXiv:1407.8114] [INSPIRE].

[111] R. Brüser, Z.L. Liu and M. Stahlhofen, *Three-Loop Quark Jet Function*, *Phys. Rev. Lett.* **121** (2018) 072003 [arXiv:1804.09722] [INSPIRE].

[112] P. Banerjee, P.K. Dhani and V. Ravindran, *Gluon jet function at three loops in QCD*, *Phys. Rev. D* **98** (2018) 094016 [arXiv:1805.02637] [INSPIRE].

[113] M.A. Ebert, B. Mistlberger and G. Vita, *The Energy-Energy Correlation in the back-to-back limit at N^3LO and N^3LL'*, *JHEP* **08** (2021) 022 [arXiv:2012.07859] [INSPIRE].

[114] Y. Li and H.X. Zhu, *Bootstrapping Rapidity Anomalous Dimensions for Transverse-Momentum Resummation*, *Phys. Rev. Lett.* **118** (2017) 022004 [arXiv:1604.01404] [INSPIRE].

[115] Y. Li, A. von Manteuffel, R.M. Schabinger and H.X. Zhu, *Soft-virtual corrections to Higgs production at N^3LO*, *Phys. Rev. D* **91** (2015) 036008 [arXiv:1412.2771] [INSPIRE].

[116] D. Baranowski et al., *Zero-Jettiness Soft Function to Third Order in Perturbative QCD*, *Phys. Rev. Lett.* **134** (2025) 191902 [arXiv:2409.11042] [INSPIRE].

[117] R. Brüser, Z.L. Liu and M. Stahlhofen, *Three-loop soft function for heavy-to-light quark decays*, *JHEP* **03** (2020) 071 [arXiv:1911.04494] [INSPIRE].

[118] S.A. Larin, T. van Ritbergen and J.A.M. Vermaseren, *The Next next-to-leading QCD approximation for nonsinglet moments of deep inelastic structure functions*, *Nucl. Phys. B* **427** (1994) 41 [INSPIRE].

[119] P.A. Baikov and K.G. Chetyrkin, *New four loop results in QCD*, *Nucl. Phys. B Proc. Suppl.* **160** (2006) 76 [INSPIRE].

[120] V.N. Velizhanin, *Four loop anomalous dimension of the second moment of the non-singlet twist-2 operator in QCD*, *Nucl. Phys. B* **860** (2012) 288 [arXiv:1112.3954] [INSPIRE].

JHEP08(2025)098

[121] P.A. Baikov, K.G. Chetyrkin and J.H. Kühn, *Massless Propagators, R(s) and Multiloop QCD*, *Nucl. Part. Phys. Proc.* **261-262** (2015) 3 [arXiv:1501.06739] [INSPIRE].

[122] S. Moch et al., *Four-Loop Non-Singlet Splitting Functions in the Planar Limit and Beyond*, *JHEP* **10** (2017) 041 [arXiv:1707.08315] [INSPIRE].

[123] J. Blümlein, P. Marquard, C. Schneider and K. Schönwald, *The three-loop unpolarized and polarized non-singlet anomalous dimensions from off shell operator matrix elements*, *Nucl. Phys. B* **971** (2021) 115542 [arXiv:2107.06267] [INSPIRE].

[124] V.N. Velizhanin, *Four-loop anomalous dimension of the third and fourth moments of the nonsinglet twist-2 operator in QCD*, *Int. J. Mod. Phys. A* **35** (2020) 2050199 [arXiv:1411.1331] [INSPIRE].

[125] S.A. Larin, P. Nogueira, T. van Ritbergen and J.A.M. Vermaseren, *The three loop QCD calculation of the moments of deep inelastic structure functions*, *Nucl. Phys. B* **492** (1997) 338 [hep-ph/9605317] [INSPIRE].

[126] S. Moch et al., *Low moments of the four-loop splitting functions in QCD*, *Phys. Lett. B* **825** (2022) 136853 [arXiv:2111.15561] [INSPIRE].

[127] J. Ablinger et al., *The three-loop splitting functions $P_{qg}^{(2)}$ and $P_{gg}^{(2,N_F)}$*, *Nucl. Phys. B* **922** (2017) 1 [arXiv:1705.01508] [INSPIRE].

[128] J.A. Gracey, *Three loop $\overline{\rm MS}$ tensor current anomalous dimension in QCD*, *Phys. Lett. B* **488** (2000) 175 [hep-ph/0007171] [INSPIRE].

[129] J.A. Gracey, *Tensor current renormalization in the RI' scheme at four loops*, *Phys. Rev. D* **106** (2022) 085008 [arXiv:2208.14527] [INSPIRE].

[130] K.G. Chetyrkin and A.G. Grozin, *Three loop anomalous dimension of the heavy light quark current in HQET*, *Nucl. Phys. B* **666** (2003) 289 [hep-ph/0303113] [INSPIRE].

[131] A. Grozin, *Anomalous dimension of the heavy-light quark current in HQET up to four loops*, *JHEP* **02** (2024) 198 [arXiv:2311.09894] [INSPIRE].

[132] G. Das, S.-O. Moch and A. Vogt, *Soft corrections to inclusive deep-inelastic scattering at four loops and beyond*, *JHEP* **03** (2020) 116 [arXiv:1912.12920] [INSPIRE].

[133] G. Das, S. Moch and A. Vogt, *Approximate four-loop QCD corrections to the Higgs-boson production cross section*, *Phys. Lett. B* **807** (2020) 135546 [arXiv:2004.00563] [INSPIRE].

[134] S. Moch, J.A.M. Vermaseren and A. Vogt, *The Quark form-factor at higher orders*, *JHEP* **08** (2005) 049 [hep-ph/0507039] [INSPIRE].

[135] B. Agarwal, A. von Manteuffel, E. Panzer and R.M. Schabinger, *Four-loop collinear anomalous dimensions in QCD and N=4 super Yang-Mills*, *Phys. Lett. B* **820** (2021) 136503 [arXiv:2102.09725] [INSPIRE].

[136] S. Moch, J.A.M. Vermaseren and A. Vogt, *Three-loop results for quark and gluon form-factors*, *Phys. Lett. B* **625** (2005) 245 [hep-ph/0508055] [INSPIRE].

[137] A. Grozin, J.M. Henn, G.P. Korchemsky and P. Marquard, *The three-loop cusp anomalous dimension in QCD and its supersymmetric extensions*, *JHEP* **01** (2016) 140 [arXiv:1510.07803] [INSPIRE].

[138] A.A. Vladimirov, *Correspondence between Soft and Rapidity Anomalous Dimensions*, *Phys. Rev. Lett.* **118** (2017) 062001 [arXiv:1610.05791] [INSPIRE].

JHEP08(2025)098

[139] C. Duhr, B. Mistlberger and G. Vita, *Four-Loop Rapidity Anomalous Dimension and Event Shapes to Fourth Logarithmic Order*, *Phys. Rev. Lett.* **129** (2022) 162001 [arXiv:2205.02242] [INSPIRE].

[140] I. Moult, H.X. Zhu and Y.J. Zhu, *The four loop QCD rapidity anomalous dimension*, *JHEP* **08** (2022) 280 [arXiv:2205.02249] [INSPIRE].

[141] I.W. Stewart, F.J. Tackmann and W.J. Waalewijn, *Factorization at the LHC: from PDFs to Initial State Jets*, *Phys. Rev. D* **81** (2010) 094035 [arXiv:0910.0467] [INSPIRE].

[142] I.W. Stewart, F.J. Tackmann and W.J. Waalewijn, *The Quark Beam Function at NNLL*, *JHEP* **09** (2010) 005 [arXiv:1002.2213] [INSPIRE].

[143] G.F. Sterman, *Summation of Large Corrections to Short Distance Hadronic Cross-Sections*, *Nucl. Phys. B* **281** (1987) 310 [INSPIRE].

[144] S. Catani and L. Trentadue, *Resummation of the QCD Perturbative Series for Hard Processes*, *Nucl. Phys. B* **327** (1989) 323 [INSPIRE].

[145] G. Lustermans, J.K.L. Michel and F.J. Tackmann, *Generalized Threshold Factorization with Full Collinear Dynamics*, arXiv:1908.00985 [INSPIRE].

[146] C. Duhr, B. Mistlberger and G. Vita, *Soft integrals and soft anomalous dimensions at N^3LO and beyond*, *JHEP* **09** (2022) 155 [arXiv:2205.04493] [INSPIRE].

JHEP08(2025)098